Towards Constant Rate Split State Non-malleable Codes: Locally Testable Non-Malleable Codes Against Decision Trees*

Irem Ergun† Supervisor: Silas Richelson‡

Abstract

In this work, we work on a conjecture to construct locally testable non-malleable codes against decision trees. We believe that this work will pave the way for a construction (based on low-degree polynomials) of constant rate split state non-malleable codes with optimal error, which is a big open problem. We describe our contributions, which are the categorization of tampering of planes by decision trees and a significant progress towards showing that none of these types of tampering is problematic: they either reduce to affine tampering, which we can easily deal with or they are not even performing mauling attacks. We then give a detailed explanation of the remaining work we have to do and explain why we believe our ideas are powerful.

1 Introduction

A *coding scheme* is a pair $(\mathsf{Enc}, \mathsf{Dec})$ of functions $\mathsf{Enc} : \Gamma^k \to \Gamma^n$ and $\mathsf{Dec} : \Gamma^n \to \Gamma^k \cup \{\perp\}$ such that $\mathsf{Dec}\big(\mathsf{Enc}(m)\big) = m$. The quantity k/n is called the *rate* of the code. Given $\mathbf{x}, \mathbf{y} \in \Gamma^n$, the *distance* between $\mathbf{x}$ and $\mathbf{y}$ is $\Pr_{i\sim[n]}\big[\mathbf{x}_i \neq \mathbf{y}_i\big]$. The *distance of the code* is the minimum distance between any two distinct valid codewords. We say $(\mathsf{Enc}, \mathsf{Dec})$ is an *error-correcting code* [Ham50] if there exists $\delta > 0$ such that $\mathsf{Dec}(\mathbf{y}) = m$ for all $\mathbf{y}$ which are within distance δ of some valid codeword $\mathbf{x} = \mathsf{Enc}(m)$. The state of the art today is codes with constant rate and which can decode from a constant fraction of errors [RS60, Jus72].

Locally-testable codes (LTCs) [FS95, GS06] are a type of error correcting code that support a very efficient, randomized test, which reads only a few symbols from the code and outputs a bit indicating whether or not it thinks the codeword is valid. Intuitively, $\mathsf{Test}(\mathbf{x}) = 1$ for all valid codewords $\mathbf{x} \in \Gamma^n$; and if $\mathbf{y}$ is very far from being valid, then $\mathsf{Test}(\mathbf{y}) = 0$ should occur with high probability. In addition to their obviously useful test feature, LTCs share many similarities with probabilistically checkable proofs, and this connection proves useful in many cases. [ALM+98, AS98].

*Many thanks to Sourya Roy for his numerous contributions.

†UC Riverside. Email: iergu001@ucr.edu.

‡UC Riverside. Email: silas@cs.ucr.edu.

As we have stated earlier, error correcting codes are very powerful objects. However, they only provide security against a honest but noisy channel and they are not secure against active adversarial behavior. *Non-malleable codes* (NMCs) [DPW18], on the other hand, provide security against a channel which actively tampers codewords using a function $f : \Gamma^n \to \Gamma^n$,which is called the tampering function. This model was initially motivated by applications to leakage and tamper resilient cryptography [DPW18, AGM+15, CDM+20]. However, since then, it has been immensely useful for many different applications, for example to secure protocol design [GPR16, GR19], complexity theory [DJMW12], and pseudorandomness [CGL16, CZ16]. Given a message $m \in \Gamma^k$ and $f : \Gamma^n \to \Gamma^n$, the *tampering distribution* outputs $\big(\mathsf{Dec} \circ f \circ \mathsf{Enc}\big)(m) \in \Gamma^k \cup \{\perp\}$. Roughly speaking, we say that $(\mathsf{Enc}, \mathsf{Dec})$ is *non-malleable* against a function family $\mathcal{F} \subset \big\{f : \Gamma^n \to \Gamma^n\big\}$ if for all $f \in \mathcal{F}$ and $m \in \Gamma^k$, the tampering distribution either outputs m (the case when there is no tampering present) or is statistically independent of m (the case when there is tampering present).

Naturally, one would like to have non-malleable codes that are secure against all kinds of tampering functions. However, this is not possible if the adversary has access to all bits of the encoded message. Because the encoding and decoding procedures are efficient, the adversary can decode, tamper and then encode. It is trivial to see that this tampering cannot be prevented or detected. So, the primary goal becomes constructing non-malleable codes that are secure against stronger adversaries and the secondary goal is to keep the rate of the code as close to 1 as possible. There is an obvious trade-off here as the length of the codeword increases, we can have more security (while having decodability) because we can basically "hide" the message within the codeword more easily. So, in an ideal situation, we would like to keep the error of the code exponentially small while having the rate constant.

Split-state non-malleable codes model is the most popular model (with a lot of useful applications [ADKO15, AKO17, BDSG+18, ADN+19, BGW19]), where security against arbitrary behavior of the tampering function is achieved. In this model, the tampering function is actually composed of two independent functions, each of which has access to a different half of the input. There are many works done on this topic, which we describe in section 2.3. However, the work done so far fails the construct the optimal codes in terms of security and rate. We aim to construct codes that are optimal in terms of security against split-state tampering and rate.

2 Preliminaries and Conjecture

2.1 Locally Testable Codes and Non-Malleable Codes

Definition 1 (Locally Testable Code). *Fix $q \in \mathbb{N}$ and $\varepsilon > 0$. We say that a code $(\mathsf{Enc}, \mathsf{Dec})$, is a $(q, \varepsilon)-$*locally testable code *(LTC) if there exists a randomized algorithm Test which reads q symbols of a supposed codeword $\mathbf{y} \in \Gamma^n$ (the symbols are indexed by $I \subset [n]$ of size $|I| = q$) and outputs a bit such that 1) $\mathsf{Test}(\mathbf{x}) = 1$ with probability 1 for all valid codewords $\mathbf{x} \in \Gamma^n$, and 2) there exists a constant $c > 0$ such that for all $\mathbf{y} \in \Gamma^n$ with $\mathsf{dist}(\mathbf{y}) \geq \varepsilon$,*

$$\Pr_I\Big[\mathsf{Test}(\mathbf{y}; I) = 0\Big] \geq c \cdot \mathsf{dist}(\mathbf{y}),$$

where $\mathsf{dist}(\mathbf{y})$ denotes the distance between $\mathbf{y}$ and the nearest valid codeword.

We formally define non-malleable codes via non-malleable reductions [ADKO15]. Intuitively, a non-malleable reduction from $\mathcal{F}$ to $\mathcal{G}$ guarantees that the tampering of codewords by functions in $\mathcal{F}$ is captured by tampering messages by functions in $\mathcal{G}$. The key feature of non-malleable reductions is that they compose well. For example, if $(\mathsf{Enc}_{\mathcal{F}}, \mathsf{Dec}_{\mathcal{F}})$ is a non-malleable reduction from $\mathcal{F}$ to $\mathcal{G}$ and $(\mathsf{Enc}_{\mathcal{G}}, \mathsf{Dec}_{\mathcal{G}})$ is a non-malleable code against $\mathcal{G}$, then $\big(\mathsf{Enc}_{\mathcal{F}} \circ \mathsf{Enc}_{\mathcal{G}}, \mathsf{Dec}_{\mathcal{G}} \circ \mathsf{Dec}_{\mathcal{F}}\big)$ is a non-malleable code against $\mathcal{F}$.

Definition 2 (Non-Malleable Reductions). *Fix $\varepsilon > 0$ and tampering function families*

$$\mathcal{F} \subset \big\{f : \Gamma^n \to \Gamma^n\big\} \text{ and } \mathcal{G} \subset \big\{g : \Gamma^k \to \Gamma^k \cup \{\perp\}\big\}.$$

*We say that a coding scheme $(\mathsf{Enc}, \mathsf{Dec})$ is an $\varepsilon-$*non-malleable reduction *from $\mathcal{F}$ to $\mathcal{G}$ if for all $f \in \mathcal{F}$ there exists a distribution G_f on $\mathcal{G}$ such that $\Delta\big((\mathsf{Dec} \circ f \circ \mathsf{Enc})(m), \mathsf{G}_f(m)\big) \leq \varepsilon$ for all $m \in \Gamma^k$, where $\mathsf{G}_f(m)$ is the distribution which draws $g \sim \mathsf{G}_f$ and outputs $g(m)$ (Δ denotes statistical distance). A* non-malleable code *is a non-malleable reduction to the family of "trivial" tampering functions, containing only the identity and constants.*

Tampering Function Families. We identify five types of tampering.

- **Coordinate-Wise:** In the coordinate-wise tampering model, the each symbol of the codeword $c \in \Gamma^n$ is tampered independently. Thus, $\{f_i\}_{i\in[n]} \in \mathcal{F}_{coord}$ is a sequence of $f_i : \Gamma \to \Gamma$ which tampers via $\{f_i\}_i : \{c_i\} \longmapsto \{f_i(c_i)\}$.

- **Decision Trees:** The decision tree model is a generalization of the coordinate-wise tampering model. Specifically, if $\{f_i\} \in \mathcal{F}^r_{dtree}$, then every f_i is a decision tree of depth r. A decision tree from this function family is a $|\Gamma|-$ary tree of depth r, whose leaves have labels from Γ and its each non-leaf node has label i for some $i \in [n]$. Evaluating the decision tree is as follows: for the node with label i, query c_i and descend to the c_i-th child of the node with label i and continue until a leaf node is reached. The value in the resulting leaf node is the output of that decision tree. So, the tampering distribution is

$$(Dec \circ \{f_i\}_i \circ Enc)(m) = Dec(\{f_i(c)\}_i) = Dec(\{\tilde{c}_i\}_i) = \tilde{m}.$$

- **Affine:** We say that $\mathsf{T} : \Gamma \to \Gamma$ is *affine* if $\exists\ (s, \Phi_0) \in \mathbb{F} \times \Gamma$ such that $\mathsf{T}(\Phi) = s \cdot \Phi + \Phi_0$.

- **Split-State:** In the split-state model, the tampering function is actually composed of two independent functions (f,g), and each of these functions take (a fixed) half[1] of the encoded message as input and (f,g) tamper these parts independently. Formally, $\mathcal{F}_{split} = \{(f, g) \mid f, g : \{0,1\}^n \to \{0,1\}^n\}$. So, the tampering distribution for $(f, g) \in \mathcal{F}_{split}$ and $m \in \{0,1\}^k$ outputs: $(Dec \circ (f, g) \circ Enc)(m) = Dec(f(L), g(R)) = Dec(\tilde{L}, \tilde{R}) = \tilde{m}$ where $L, R \in \{0,1\}^n$ and $(\tilde{L}, \tilde{R}) = f(L), g(R)$.

[1]Functions (f, g) can take arbitrary number of bits as input, as long as their inputs are non-overlapping. For simplicity, we will say that they take half of the encoded message as input.

2.2 Sampler Graphs

For the sake of completeness, we added added some useful definitions about sampler graphs in this paper, which are taken from [RR19] and we shortened the content. We encourage the reader to read section 2 and 5 of [RR19].

Notations. For a finite set S, $s \sim S$ indicates that s is drawn uniformly from S. For a bipartite graph $(A \cup B, E)$ and $a \in A$, $B(a)$ denotes the uniform distribution on the neighborhood of a in B: $\{b \in B : (a,b) \in E\}$. The neighborhood distribution $A(b)$ for $b \in B$ is defined analogously. For all bipartite graphs used in this work, the edge relations are natural. For example, A might be the set of lines in $\mathbb{F}^k$ ($\mathbb{F}$ a finite field), B the set of points in $\mathbb{F}^k$, and the edge relation captures incidence: $(a,b) \in E$ iff $b \in a$. For this reason, we simplify notations by suppressing E and denoting bipartite graphs as A/B instead of $(A \cup B, E)$, and writing $a \sim b$ instead of $(a,b) \in E$.

Definition 3 (Biregularity). *Let A/B be a bipartite graph and fix $\eta > 0$. We say that A/B is $\eta-$*biregular *if the distribution which draws $a \sim A$, $b \sim B(a)$, and outputs (a,b) is within statistical distance η of the distribution which gives the same output by drawing $b \sim B$, $a \sim A(b)$.*[2]

Biregularity ensures that for any $B' \subset B$ of size $|B'| = \lambda \cdot |B|$, the expectation (over $a \sim A$) of $\Pr_{b \sim B(a)}[b \in B']$ is close to λ. We say that A/B is *sampling* if a concentration bound holds.

Definition 4 (Sampler Graph [Zuc97]**).** *Fix $\varepsilon, \delta > 0$. We say that the bipartite graph A/B is $(\varepsilon,\delta)-$*sampling *if for all subsets $B' \subset B$ of size $|B'| = \lambda \cdot |B|$,*

$$\Pr_{a \sim A}\left[\left|\Pr_{b \sim B(a)}\left[b \in B'\right] - \lambda\right| > \varepsilon\right] \leq \delta.$$

Double Samplers. A triple (A,B,C) is called a *double sampler* if B/C is sampling and for all $c \in C$, $A(c)/B(c)$ is sampling. Double samplers have been used implicitly in several works prior to their formalization in [DK17]. We use them implicitly in this work as well. The construction in [DK17] is of a double sampler of linear size (*i.e.*, $|A| \approx |B| \approx |C|$) based on high-dimensional expanders. The double samplers used in this work are built from elementary means and are not linear size (our double samplers have $|A| \gg |B| \gg |C|$). Importantly, a random object in our parameter regime is a double sampler with good probability, while this is not true in the linear size regime.

Fact 1 (Properties of Samplers). *Fix $\varepsilon, \varepsilon', \delta, \delta', \eta > 0$. Suppose $A/B/C$ are such that $B(a)/C(a)$ is $\eta-$biregular and $C(a,b) = C(b)$ for all $a \in A$ and $b \in B(a)$. The following hold.*

1. *If B/C is $(\varepsilon',\delta')-$sampling and A/B is $\eta-$biregular, then A/C is $(\varepsilon,\delta)-$sampling, where $\delta \geq \varepsilon^{-1} \cdot (2\eta + \varepsilon' + \delta')$.*
2. *If A/B is $(\varepsilon',\delta')-$sampling and B/C is $\eta-$biregular, then A/C is $(\varepsilon,\delta)-$sampling, where $\varepsilon \geq 3\varepsilon' + 2\eta$ and $\delta \geq \delta'/\varepsilon'$.*

[2] This is related to the usual notion of biregularity; specifically, if A/B is biregular in the usual sense, then it is $0-$biregular in the sense of Definition 3.

2.3 Prior Work

The split-state model for non-malleable codes was conceived in [DPW18], its existence is proved with parameters $n = O(k)$ and $\varepsilon = 2^{-\Omega(k)}$. Many explicit constructions followed and the state of the art today is represented by two works [Li18, AO19]. Li constructs an $[n, k, \varepsilon]_{NM}$-code with $n = O\big(k \cdot \frac{logk}{loglogk}\big), \varepsilon = 2^{-\Omega(k)}$; Aggarwal and Obremski get $n = O(k)$ and $\varepsilon = 2^{-k^{\alpha}}$ for a constant $0 < \alpha < 1$. Both constructions use the "alternating extraction" technique of [DP07], which is both a very powerful method for proving non-malleability, and the source of the sub-optimalities. For this reason, it seems as though substantially new ideas will be required in order to obtain optimal split-state non-malleable codes.

Using low-degree polynomials comes to mind as they have been used in many influential works on coding theory [RS60, Ric64, Jus72, WB86, Zuc97, Yek08, TS17], but not on non-malleable codes due to its complex nature, at least until [RR19]. Richelson and Roy show that a Reed-Muller type code, which is a code that is constructed by using low-degree polynomials, is non-malleable against the family of coordinate-wise tampering functions. Although this work itself is not sufficient to get optimal split-state non-malleable codes, it offers a fresh perspective on constructing split-state non-malleable codes, which we build upon in this work.

2.4 Our Conjecture and Contributions

We work towards getting optimal split-state non-malleable codes by using low-degree polynomials, just like [RR19] did. However, the tampering function family we work with is decision trees function family, which is a natural generalization of the coordinate-wise tampering function family. We conjecture that the techniques that are used to construct locally testable non-malleable codes can also be used to construct constant rate split-state non-malleable codes. Our contributions are:

1. We categorize the types of tampering of planes that can be done by decision trees according to how it tampers two planes intersecting at various surfaces.

2. We show that some of these tampering types are not actually mauling whereas the others reduce to affine tampering.

3. We describe a complete and novel roadmap to come up with locally testable non-malleable codes for decision trees and explain why our ideas are powerful.

3 Categorization of Plane Tampering

In this section, we give basic version of the code (which is a type of Reed-Solomon code) that was shown to be non-malleable against coordinate-wise tampering in [RR19]. Then, we describe the kinds of tampering that can be achieved by decision trees of depth 1 on this type of codes, and show that these are either tampering in an affine fashion or not mauling. Note that we can have non-malleability against affine tampering easily by composing this code with an inner code that is non-malleable against affine tampering.

3.1 The Code

Notation: Let $\mathbb{F}$ be a finite field and let $t \geq 4$ and $d \geq 2$ be the dimension and degree parameters respectively. Let A be the set of affine 3-planes and $\mathsf{C} = \mathbb{F}^t$. Let Γ be the set of t-variate polynomials and Γ_{A} be the set of 3-variate polynomials of degree at most d over $\mathbb{F}$.

- Enc(m): For $m \in \mathbb{F}$, draw $\Phi \sim \Gamma$ such that $\Phi(0) = m$ and output $\{(\mathsf{a}, \alpha)\}_{\mathsf{a} \in \mathsf{A}}$, where $\alpha = \Phi \mid_{\mathsf{a}}$.
- Dec($\{(\mathsf{a}, \alpha)\}_{\mathsf{a} \in \mathsf{A}}$): Find $\Phi \in \Gamma$ such that $(\mathsf{a}, \alpha) = (\mathsf{a}, \Phi \mid_{\mathsf{a}})$ for all $\mathsf{a} \in \mathsf{A}$. If such Φ exists, output $\Phi(0)$. Otherwise, output $\perp$.[3]

3.2 Types of Plane Tampering

Notation and context: Let A be the set of 3-planes in $\mathbb{F}^t$, B_1 the set of lines in $\mathbb{F}^t$. Recall that decision trees are actually a set of functions that tamper with the coordinates of the code independently. Each function that is responsible for tampering with one coordinate of the code can actually read as many bits as its depth. Here, we deal with decision trees of depth 1 and the coordinates correspond to planes. Formally, the tampering function is in the following form: $\{\mathsf{f}_{\mathsf{a}}\}_{\mathsf{a} \in \mathsf{A}}$. Each f_{a} has access to a and another plane of its choosing. For the sake of simplicity, we will use $\mathsf{f} : \mathsf{A} \to \mathsf{A}$ to define the one and only decision tree that is responsible for tampering all planes independently. We will also use tilde to denote that the corresponding geometric object is tampered by the tampering function.

We perform a case analysis on the types of tampering on planes. We classify types of tampering that can be done into four major groups according to how the decision tree affects the agreement of two different planes. We give the high level classification here. Then, in the following sections, we officially define these cases and show that these types of tampering not mauling or are tampering in an affine way. Let $\mathsf{b}_1 \sim \mathsf{B}_1,\ \mathsf{a}, \mathsf{a}' \sim \mathsf{A}(\mathsf{b}_1)$, meaning that a, a' are two 3-affine planes such that they intersect at a line. Depending on the surface of intersection of $\tilde{\mathsf{a}}, \tilde{\mathsf{a}}'$, we divide the types of tampering into four categories.

1. **Semi-constant case:** This is the case when $\tilde{\mathsf{a}} \cap \tilde{\mathsf{a}}' \in \mathsf{B}_2$. The agreement of the planes increases after being tampered, which can only mean that f is tampering in a semi-constant fashion, which is described in section 4.

2. **Affine case:** This is the case when $\tilde{\mathsf{a}} \cap \tilde{\mathsf{a}}' \in \mathsf{B}_1$, meaning that amount of agreement is the same after the planes are tampered. We work towards showing that tampering of this type is actually tampering in an affine fashion in section 5 and we make significant progress. However, this is the most complex case of this analysis and requires more work as described in section 6.

3. **Random case:** This is the case when $\tilde{\mathsf{a}} \cap \tilde{\mathsf{a}}' \in \emptyset$. We conjecture that this is a simple case and say that this type of tampering is not mauling as querying random planes will not let the decision tree learn anything about the message.

[3] For the sake of brevity, we do not describe the test procedure as we will not be using it throughout this work.

4 Increased Agreement

Notational Setup: Throughout this section, A denotes the set of $3-$dimensional planes in $\mathbb{F}^k$, B_2 denotes the set of $2-$dimensional planes in $\mathbb{F}^k$, B_1 denotes the set of lines in $\mathbb{F}^k$ and C denotes the set of points in $\mathbb{F}^k$. For planes which have more than 3 dimensions in $\mathbb{F}^k$, we use G_i to denote i-dimensional planes, where $i > 3$. We slightly abuse notation and use $\mathsf{f}(\mathsf{a}, \mathsf{a}') := \mathsf{a} \cap \mathsf{a}'$ and $\mathsf{f}(\mathsf{a}, \mathsf{a}', \mathsf{a}'') := \mathsf{a} \cap \mathsf{a}' \cap \mathsf{a}''$.

Here we analyze the cases where planes that get tampered with the tampering function intersect more than they intersected before being tampered. We claim that whatever happens in this case, we always end up in a "semi-constant" tampering situation, where the tampering function has very low entropy. Towards that end, we prove Lemma 1, restated below in a quantitative form.

Lemma 1 (Restated). *Suppose ε is a non-trivial fraction. Suppose $\mathsf{f} : \mathsf{A} \to \mathsf{A}$ is such that*

$$\Pr_{\substack{\mathsf{b}_1,\mathsf{b}_2 \sim \mathsf{B}_1 \\ \mathsf{a}' \sim \mathsf{A}(\mathsf{b}_1,\mathsf{b}_2)}} \left[\Pr_{\substack{\mathsf{a} \sim \mathsf{A}(\mathsf{b}_1) \\ \mathsf{a}'' \sim \mathsf{A}(\mathsf{b}_2)}} \left[\tilde{\mathsf{a}} \cap \tilde{\mathsf{a}}' \in \mathsf{B}_2 \& \tilde{\mathsf{a}}' \cap \tilde{\mathsf{a}}'' \in \mathsf{B}_2 \right] \geq \varepsilon^2/64 \right] \geq \varepsilon/8,$$

where $(\tilde{\mathsf{a}}, \tilde{\mathsf{a}}'\tilde{\mathsf{a}}'') = (\mathsf{f}(\mathsf{a}), \mathsf{f}(\mathsf{a}'), \mathsf{f}(\mathsf{a}''))$. Then, one of the following three cases occur:

A) There exists a fixed $\mathsf{p} \in \mathsf{B}_2$ such that $\Pr_{\mathsf{a}'}\left[\mathsf{p} \in \tilde{\mathsf{a}}'\right] \geq \varepsilon^*$.

B) There exists a fixed $\mathsf{V} \in \mathsf{G}_4$ and a line $\ell \subset \mathsf{V}$ such that $\Pr_{\mathsf{a}'}\left[\ell \in \tilde{\mathsf{a}}' \in \mathsf{V}\right] \geq \varepsilon''$.

C) There exists a fixed $\mathsf{R} \in \mathsf{G}_5$ and a line $t \subset \mathsf{R}$ such that $\Pr_{\mathsf{a}'}\left[t \in \tilde{\mathsf{a}}' \in \mathsf{R}\right] \geq \varepsilon'$.

Proof. We begin our proof with a case analysis. We break our initial assumption down into three cases by categorizing on the relationship between $\tilde{\mathsf{a}}$ and $\tilde{\mathsf{a}}''$. Thus, we have with probability greater than or equal to $\varepsilon/8$ over $\mathsf{b}_1, \mathsf{b}_2 \sim \mathsf{B}_1, \mathsf{a}' \sim \mathsf{A}(\mathsf{b}_1, \mathsf{b}_2)$:

1. $\Pr_{\substack{\mathsf{a} \sim \mathsf{A}(\mathsf{b}_1) \\ \mathsf{a}'' \sim \mathsf{A}(\mathsf{b}_2)}} \left[\tilde{\mathsf{a}} \cap \tilde{\mathsf{a}}' \in \mathsf{B}_2 \& \tilde{\mathsf{a}}' \cap \tilde{\mathsf{a}}'' \in \mathsf{B}_2 \& \tilde{\mathsf{a}} = \tilde{\mathsf{a}}'' \right] \geq \varepsilon/192$
2. $\Pr_{\substack{\mathsf{a} \sim \mathsf{A}(\mathsf{b}_1) \\ \mathsf{a}'' \sim \mathsf{A}(\mathsf{b}_2)}} \left[\tilde{\mathsf{a}} \cap \tilde{\mathsf{a}}' \in \mathsf{B}_2 \& \tilde{\mathsf{a}}' \cap \tilde{\mathsf{a}}'' \in \mathsf{B}_2 \& \tilde{\mathsf{a}} \cap \tilde{\mathsf{a}}'' \in \mathsf{B}_2 \right] \geq \varepsilon/192$
3. $\Pr_{\substack{\mathsf{a} \sim \mathsf{A}(\mathsf{b}_1) \\ \mathsf{a}'' \sim \mathsf{A}(\mathsf{b}_2)}} \left[\tilde{\mathsf{a}} \cap \tilde{\mathsf{a}}' \in \mathsf{B}_2 \& \tilde{\mathsf{a}}' \cap \tilde{\mathsf{a}}'' \in \mathsf{B}_2 \& \tilde{\mathsf{a}} \cap \tilde{\mathsf{a}}'' \in \mathsf{B}_1 \right] \geq \varepsilon/192$

We observe that if f is such that it maps two random 3-planes to the same plane, f is a constant function with good probability, which is not mauling.
We now analyze the second and third cases and show correspondence between these cases and the cases of the lemma. We start with the second case. By our starting assumption, we have

$$\Pr_{\substack{\mathsf{b}_1,\mathsf{b}_2 \sim \mathsf{B}_1 \\ (\mathsf{a},\mathsf{a}'',\mathsf{a}')}} \left[\tilde{\mathsf{a}} \cap \tilde{\mathsf{a}}' \in \mathsf{B}_2 \& \tilde{\mathsf{a}}' \cap \tilde{\mathsf{a}}'' \in \mathsf{B}_2 \& \tilde{\mathsf{a}} \cap \tilde{\mathsf{a}}'' \in \mathsf{B}_2 \right] \geq \varepsilon^3/1536,$$

where $\mathsf{a}' \sim \mathsf{A}(\mathsf{b}_1, \mathsf{b}_2), \mathsf{a} \sim \mathsf{A}(\mathsf{b}_1), \mathsf{a}'' \sim \mathsf{A}(\mathsf{b}_2)$. We further categorize this event into subcases according to the relationship between $\mathsf{f}(\mathsf{a}, \mathsf{a}')$ and $\mathsf{f}(\mathsf{a}', \mathsf{a}'')$:

- $\Pr_{\substack{\mathsf{b}_1,\mathsf{b}_2\sim\mathsf{B}_1\\(\mathsf{a},\mathsf{a}'',\mathsf{a}')}}\left[\tilde{\mathsf{a}}\cap\tilde{\mathsf{a}}'\in\mathsf{B}_2\&\tilde{\mathsf{a}}'\cap\tilde{\mathsf{a}}''\in\mathsf{B}_2\&\tilde{\mathsf{a}}\cap\tilde{\mathsf{a}}''\in\mathsf{B}_2\&\mathsf{f}(\mathsf{a},\mathsf{a}')=\mathsf{f}(\mathsf{a}',\mathsf{a}'')\right]\geq\varepsilon^3/9216$
- $\Pr_{\substack{\mathsf{b}_1,\mathsf{b}_2\sim\mathsf{B}_1\\(\mathsf{a},\mathsf{a}'',\mathsf{a}')}}\left[\tilde{\mathsf{a}}\cap\tilde{\mathsf{a}}'\in\mathsf{B}_2\&\tilde{\mathsf{a}}'\cap\tilde{\mathsf{a}}''\in\mathsf{B}_2\&\tilde{\mathsf{a}}\cap\tilde{\mathsf{a}}''\in\mathsf{B}_2\&\mathsf{f}(\mathsf{a},\mathsf{a}')\neq\mathsf{f}(\mathsf{a}',\mathsf{a}'')\right]\geq\varepsilon^3/3072$

We first show the first subcase corresponds to the case A of Lemma 1.

$$\begin{aligned}\Pr_{\substack{\mathsf{b}_1,\mathsf{b}_2\sim\mathsf{B}_1\\\mathsf{a}'\sim\mathsf{A}(\mathsf{b}_1,\mathsf{b}_2)}}&\left[\exists\mathsf{a}\sim\mathsf{A}(\mathsf{b}_1),\mathsf{a}''\sim\mathsf{A}(\mathsf{b}_2)\text{ s.t. }\tilde{\mathsf{a}}\cap\tilde{\mathsf{a}}'\in\mathsf{B}_2\&\tilde{\mathsf{a}}'\cap\tilde{\mathsf{a}}''\in\mathsf{B}_2\&\tilde{\mathsf{a}}\cap\tilde{\mathsf{a}}''\in\mathsf{B}_2\&\mathsf{f}(\mathsf{a},\mathsf{a}')=\mathsf{f}(\mathsf{a}',\mathsf{a}'')\right]\\&\geq\Pr_{\substack{\mathsf{b}_1,\mathsf{b}_2\sim\mathsf{B}_1\\(\mathsf{a},\mathsf{a}'',\mathsf{a}')}}\left[\tilde{\mathsf{a}}\cap\tilde{\mathsf{a}}'\in\mathsf{B}_2\&\tilde{\mathsf{a}}'\cap\tilde{\mathsf{a}}''\in\mathsf{B}_2\&\tilde{\mathsf{a}}\cap\tilde{\mathsf{a}}''\in\mathsf{B}_2\&\mathsf{f}(\mathsf{a},\mathsf{a}')=\mathsf{f}(\mathsf{a}',\mathsf{a}'')\right]\geq\varepsilon^3/9216\end{aligned}$$

We fix $\mathsf{p}=\mathsf{f}(\mathsf{a},\mathsf{a}'')\in\mathsf{B}_2$ and immediately see that $\Pr_{\substack{\mathsf{b}_1,\mathsf{b}_2\sim\mathsf{B}_1\\(\mathsf{a},\mathsf{a}'',\mathsf{a}')}}\left[\exists\mathsf{p}\in\mathsf{B}_2\text{ s.t. }\mathsf{p}\in\tilde{\mathsf{a}}'\right]\geq\varepsilon^3/9216$.
Now, we show the second subcase corresponds to case B of Lemma 1.

$$\begin{aligned}\Pr_{\substack{\mathsf{b}_1,\mathsf{b}_2\sim\mathsf{B}_1\\\mathsf{a}'\sim\mathsf{A}(\mathsf{b}_1,\mathsf{b}_2)}}&\left[\exists\mathsf{a},\mathsf{a}''\text{ s.t. }\tilde{\mathsf{a}}\cap\tilde{\mathsf{a}}'\in\mathsf{B}_2\&\tilde{\mathsf{a}}'\cap\tilde{\mathsf{a}}''\in\mathsf{B}_2\&\tilde{\mathsf{a}}\cap\tilde{\mathsf{a}}''\in\mathsf{B}_2\&\mathsf{f}(\mathsf{a},\mathsf{a}')\neq\mathsf{f}(\mathsf{a}',\mathsf{a}'')\neq\mathsf{f}(\mathsf{a},\mathsf{a}'')\right]\\&\geq\Pr_{\substack{\mathsf{b}_1,\mathsf{b}_2\sim\mathsf{B}_1\\(\mathsf{a},\mathsf{a}'',\mathsf{a}')}}\left[\tilde{\mathsf{a}}\cap\tilde{\mathsf{a}}'\in\mathsf{B}_2\&\tilde{\mathsf{a}}'\cap\tilde{\mathsf{a}}''\in\mathsf{B}_2\&\tilde{\mathsf{a}}\cap\tilde{\mathsf{a}}''\in\mathsf{B}_2\&\mathsf{f}(\mathsf{a},\mathsf{a}')\neq\mathsf{f}(\mathsf{a}',\mathsf{a}'')\neq\mathsf{f}(\mathsf{a},\mathsf{a}'')\right]\\&\geq\varepsilon^3/3072\end{aligned}$$

We fix $\mathsf{V}=span(\tilde{\mathsf{a}},\tilde{\mathsf{a}}''),\ell=\mathsf{f}(\mathsf{a},\mathsf{a}',\mathsf{a}'')$ and note that since $\mathsf{f}(\mathsf{a},\mathsf{a}'')\in\mathsf{B}_2$, $\mathsf{V}\in\mathsf{G}_4$. Then, we observe that

$$\begin{aligned}&\ell=\mathsf{f}(\mathsf{a},\mathsf{a}',\mathsf{a}'')=\mathsf{f}(\mathsf{a},\mathsf{a}')\cap\mathsf{f}(\mathsf{a}',\mathsf{a}'')\Rightarrow\ell\in\mathsf{B}_1\Rightarrow span(\mathsf{f}(\mathsf{a},\mathsf{a}'),\mathsf{f}(\mathsf{a}',\mathsf{a}''))\in\mathsf{A}\\&\Rightarrow\tilde{\mathsf{a}}'=span(\mathsf{f}(\mathsf{a},\mathsf{a}'),\mathsf{f}(\mathsf{a}',\mathsf{a}''))\subset span(\mathsf{f}(\mathsf{a}),\mathsf{f}(\mathsf{a}''))=\mathsf{V}.\end{aligned}$$

For the first if statement, we used the fact that ℓ is the intersection of two 2-planes in a 3-plane, hence a line (because $\mathsf{f}(\mathsf{a},\mathsf{a}')\neq\mathsf{f}(\mathsf{a}',\mathsf{a}'')$). Thus, we showed that

$$\Pr_{(\mathsf{b}_1,\mathsf{b}_2,\mathsf{a}')}\left[\exists\mathsf{V}\in\mathsf{G}_4,\ell\in\mathsf{B}_1\text{ s.t. }\ell\subset\mathsf{f}(\mathsf{a}')\subset\mathsf{V}\right]\geq\varepsilon^3/3072.$$

Lastly, we need to show that the third case corresponds to case C of Lemma 1 to finish the proof. We observe that the way used to show the correspondence between the second subcase of the second case and case B of the lemma can also be used to show correspondence between the third case and case C of the lemma. We fix $\mathsf{R}=span(\tilde{\mathsf{a}},\tilde{\mathsf{a}}'')$ and $t=\mathsf{f}(\mathsf{a},\mathsf{a}',\mathsf{a}'')$. The only detail that is different is the dimension of $\mathsf{R}=span(\tilde{\mathsf{a}},\tilde{\mathsf{a}}'')$, which is 5 in this case, instead of 4. Thus, for the third case, we have $\Pr_{(\mathsf{b}_1,\mathsf{b}_2,\mathsf{a}')}\left[\exists\mathsf{R}\in\mathsf{G}_5,t\in\mathsf{B}_1\text{ s.t. }t\subset\tilde{\mathsf{a}}'\subset\mathsf{R}\right]\geq\varepsilon^3/1536$. □

5 Global Agreement

This section is for showing that the tampering functions which preserve agreement of planes are tampering in an affine fashion. The proof of this powerful statement has not been completed, yet significant progress has been made. Here we prove the lemma stated below in a quantitative form and we move from high error regime to low error regime. We refer the reader to section 6 for an organized analysis of the remaining work.

Lemma 2. *Suppose* $\varepsilon \geq ??$[4]*,* $\nu = ??$ *and fix parameters ,* $\eta = ??$*,* $\delta = ??$*,* $\tau = \tau(\delta, \varepsilon, \eta, \nu)$*, Suppose* $\mathsf{f} : \mathsf{A} \to \mathsf{A}, \mathsf{h} : \mathsf{C} \to \mathsf{C}$ *are such that*

$$\Pr_{\substack{\mathsf{a}\sim\mathsf{A} \\ \mathsf{c}\sim\mathsf{C}(\mathsf{a})}}\left[\tilde{\mathsf{c}} \in \tilde{\mathsf{a}}\right] = 6\varepsilon \tag{1}$$

, and where $(\tilde{\mathsf{a}}, \tilde{\mathsf{c}}) = \big(\mathsf{f}(\mathsf{a}), \mathsf{h}(\mathsf{c})\big)$*. Then there exists a set* $\mathsf{A}' \subset \mathsf{A}$ *of size at least* $|\mathsf{A}'| \geq 2\varepsilon \cdot |\mathsf{A}|$ *and a function* $\mathsf{g} : \mathsf{B} \to \mathsf{B}$ *such that:* $\Pr_{\substack{\mathsf{a}\sim\mathsf{A}' \\ \mathsf{b}\sim\mathsf{B}(\mathsf{a})}}\left[\tilde{\mathsf{b}} \in \tilde{\mathsf{a}}\right] \geq 1 - \zeta(\eta, \tau, \delta)$*,*

We begin by introducing the notation and ideas needed to prove Lemma 2 in Section 5.1. The actual proof appears in Section 5.2, conditioned on two claims which we state in Section 5.1 and prove in Section 5.3.

5.1 Proof Setup.

Notations. In this section, A denotes the set of $3-$dimensional planes in $\mathbb{F}^k$, B_2 denotes the set of $2-$dimensional planes in $\mathbb{F}^k$, B_1 denotes the set of lines in $\mathbb{F}^k$ and C denotes the set of points in $\mathbb{F}^k$. We will take advantage of the sampling properties of the quadruple $\mathsf{A}/\mathsf{B}_2/\mathsf{B}_1/\mathsf{C}$.

We say that c is *good* if the following holds:

1. $\Pr_{\mathsf{a}\sim\mathsf{A}(\mathsf{c})}\left[\tilde{\mathsf{c}} \in \tilde{\mathsf{a}}\right] \geq 4\varepsilon$.
2. $\mathbf{E}_i := \Pr_{\substack{\mathsf{b}\sim\mathsf{B}_i(\mathsf{c}) \\ \mathsf{a},\mathsf{a}'\sim\mathsf{A}(\mathsf{b})}}\left[\tilde{\mathsf{c}} \in \tilde{\mathsf{a}} \cap \tilde{\mathsf{a}}' \in \mathsf{B}_i\right] \geq \varepsilon^2$, for $i \in \{1, 2\}$.

Local Functions. For $\mathsf{c} \in \mathsf{C}$, $i \in \{1, 2\}$, let $\mathsf{g}_\mathsf{c} : \mathsf{B}_i(\mathsf{c}) \to \mathsf{B}_i$ be the randomized function that does the following on input $b \in B_i(c)$: It first draws $\mathsf{a}, \mathsf{a}' \sim \mathsf{A}(\mathsf{b})$ and then, it outputs $\tilde{\mathsf{b}} = \tilde{\mathsf{a}} \cap \tilde{\mathsf{a}}'$ if $\tilde{\mathsf{b}} \in \mathsf{B}_i$ else outputs $\tilde{\mathsf{b}} \sim \mathsf{B}_i$.

Definition 5 (Excellent). *Let* $\eta = ??$*. We say that* $\mathsf{c} \in \mathsf{C}$ *is* excellent *if* c *is* good *and the following holds:*

$$\Pr_{\substack{\mathsf{b}\sim\mathsf{B}_i(\mathsf{c}) \\ \mathsf{a}_i,\mathsf{a}_i'\sim\mathsf{A}(\mathsf{b}) \text{ for } i=1,2}}\left[\tilde{\mathsf{a}}_1 \cap \tilde{\mathsf{a}}_1' \neq \tilde{\mathsf{a}}_2 \cap \tilde{\mathsf{a}}_2' | \tilde{\mathsf{c}} \in \tilde{\mathsf{a}}_i \cap \tilde{\mathsf{a}}_i' \in \mathsf{B}_i\right] \leq \eta \quad \text{for } i \in \{1, 2\}.$$

Claim 1. *There exists a set* $\mathsf{C}' \subset \mathsf{c}$ *such that the following hold: 1)* $|\mathsf{C}'| \geq \varepsilon^3|\mathsf{C}|$*; 2) every* $\mathsf{c} \in \mathsf{C}'$ *is excellent; 3)*

$$\Pr_{\mathsf{c},\mathsf{c}'\sim\mathsf{C}'}\left[\Pr_{\mathsf{a}\sim\mathsf{A}(\mathsf{c},\mathsf{c}')}\left[\tilde{\mathsf{c}}, \tilde{\mathsf{c}}' \in \tilde{\mathsf{a}}\right] \geq \varepsilon^5\right] \geq 1 - \sigma,$$

where $\sigma = \sigma(\delta, \varepsilon, \eta, \nu)$*.*

The Global Function. Let $\mathsf{g} : \mathsf{B} \to \mathsf{B}$ be the randomized function where $\mathsf{g}(\mathsf{b})$ draws $\mathsf{c} \sim \mathsf{C}'(b)$ and outputs $\mathsf{g}_\mathsf{c}(\mathsf{b})$ The following is also proved in Section 5.3.

Claim 2. *We have* $\Pr_{\substack{\mathsf{b}\sim\mathsf{B}_2 \\ \mathsf{c}_1,\mathsf{c}_2\sim\mathsf{C}'(\mathsf{b})}}\left[\tilde{\mathsf{b}}_1 = \tilde{\mathsf{b}}_2\right] \geq 1 - \tau$*, where* $\tau := 1 - \Big(4\eta \cdot \varepsilon^{-10} + 2\gamma + \delta\Big)$*,* $\tilde{\mathsf{b}}_i = \mathsf{g}_{\mathsf{c}_i}(\mathsf{b})$ *for* $i \in \{1, 2\}$*.*

[4]Some parameters cannot be fixed yet as they depend on some future work. We leave them as ??.

5.2 Proof of Lemma 2

Notational Convention.

Proof. Shorthand $\mathsf{B} := \mathsf{B}_2$. Let $\mathsf{A}' \subset \mathsf{A}$ be the set of $\mathsf{a} \in \mathsf{A}$ such that $\Pr_{\mathsf{b},\mathsf{c},\mathsf{a}'}\left[\tilde{\mathsf{c}} \in \tilde{\mathsf{a}} \cap \tilde{\mathsf{a}}' \in \mathsf{B}\right] \geq \varepsilon$ where $\mathsf{b} \sim \mathsf{B}(\mathsf{a}), \mathsf{c} \sim \mathsf{C}'(\mathsf{b}), \mathsf{a}' \sim \mathsf{A}(\mathsf{b})$. We have,

$$\begin{aligned}
\mathbb{E}_{\mathsf{a}\sim\mathsf{A}}\left[\Pr_{\mathsf{b},\mathsf{c},\mathsf{a}'}\left[\tilde{\mathsf{c}} \in \tilde{\mathsf{a}} \cap \tilde{\mathsf{a}}' \in \mathsf{B}\right]\right] &\geq \mathbb{E}_{\mathsf{a}\sim\mathsf{A},\mathsf{c}\sim\mathsf{C}'(\mathsf{a})}\left[\Pr_{\substack{\mathsf{b}\sim\mathsf{B}(\mathsf{c})\\ \mathsf{a}'\sim\mathsf{A}(\mathsf{b})}}\left[\tilde{\mathsf{c}} \in \tilde{\mathsf{a}} \cap \tilde{\mathsf{a}}' \in \mathsf{B}\right]\right] - \delta \\
&\geq \mathbb{E}_{\mathsf{c}\sim\mathsf{C}',\mathsf{a}\sim\mathsf{A}(\mathsf{c})}\left[\Pr_{\substack{\mathsf{b}\sim\mathsf{B}(\mathsf{c})\\ \mathsf{a}'\sim\mathsf{A}(\mathsf{b})}}\left[\tilde{\mathsf{c}} \in \tilde{\mathsf{a}} \cap \tilde{\mathsf{a}}' \in \mathsf{B}\right]\right] - 2\delta \\
&\geq 4\varepsilon^2 - 2\delta \geq 3\varepsilon^2
\end{aligned}$$

Here, we have used sampling of $\mathsf{B}(\mathsf{a})/\mathsf{C}(\mathsf{a})$, A/C and the fact that $\mathsf{c} \in \mathsf{C}'$ are good. It follows that: $|\mathsf{A}'| \geq 2\varepsilon^2|\mathsf{A}|$. Now, we show that the first property holds.

$$\begin{aligned}
\Pr_{\substack{\mathsf{a}\sim\mathsf{A}'\\ \mathsf{b}\sim\mathsf{B}(\mathsf{a})}}\left[\tilde{\mathsf{b}} \in \tilde{\mathsf{a}}\right] = \Pr_{\substack{\mathsf{a}\sim\mathsf{A}'\\ \mathsf{b}\sim\mathsf{B}(\mathsf{a})\\ \mathsf{c}\sim\mathsf{C}'(\mathsf{b})}}\left[\tilde{\mathsf{b}} \in \tilde{\mathsf{a}} \mid \tilde{\mathsf{c}} \in \tilde{\mathsf{a}}\right] &\geq \Pr_{\substack{\mathsf{a}\sim\mathsf{A}'\\ \mathsf{b}\sim\mathsf{B}(\mathsf{a})\\ \mathsf{c}\sim\mathsf{C}'(\mathsf{b})}}\left[\tilde{\mathsf{b}} = \mathsf{g}_{\mathsf{c}}(\mathsf{b}) \in \tilde{\mathsf{a}} \mid \tilde{\mathsf{c}} \in \tilde{\mathsf{a}}\right] \\
&\geq \Pr_{\substack{\mathsf{a}\sim\mathsf{A}'\\ \mathsf{b}\sim\mathsf{B}(\mathsf{a})\\ \mathsf{c}\sim\mathsf{C}'(\mathsf{b})}}\left[\mathsf{g}_{\mathsf{c}}(\mathsf{b}) \in \tilde{\mathsf{a}} \mid \tilde{\mathsf{c}} \in \tilde{\mathsf{a}}\right] - \Pr_{\substack{\mathsf{a}\sim\mathsf{A}'\\ \mathsf{b}\sim\mathsf{B}(\mathsf{a})\\ \mathsf{c}\sim\mathsf{C}'(\mathsf{b})}}\left[\tilde{\mathsf{b}} \neq \mathsf{g}_{\mathsf{c}}(\mathsf{b}) \mid \tilde{\mathsf{c}} \in \tilde{\mathsf{a}}\right]
\end{aligned}$$

We now bound these two terms. We start with the second term.

$$\begin{aligned}
\Pr_{\substack{\mathsf{a}\sim\mathsf{A}'\\ \mathsf{b}\sim\mathsf{B}(\mathsf{a})\\ \mathsf{c}\sim\mathsf{C}'(\mathsf{b})}}\left[\tilde{\mathsf{b}} \neq \mathsf{g}_{\mathsf{c}}(\mathsf{b}) \mid \tilde{\mathsf{c}} \in \tilde{\mathsf{a}}\right] &\leq \frac{\Pr_{\substack{\mathsf{a}\sim\mathsf{A}'\\ \mathsf{b}\sim\mathsf{B}(\mathsf{a})\\ \mathsf{c}\sim\mathsf{C}'(\mathsf{b})}}\left[\tilde{\mathsf{b}} \neq \mathsf{g}_{\mathsf{c}}(\mathsf{b})\right]}{\Pr_{\substack{\mathsf{a}\sim\mathsf{A}'\\ \mathsf{c}\sim\mathsf{C}'(\mathsf{b})}}\left[\tilde{\mathsf{c}} \in \tilde{\mathsf{a}}\right]} \leq \frac{1}{\varepsilon} \cdot \Pr_{\substack{\mathsf{a}\sim\mathsf{A}'\\ \mathsf{b}\sim\mathsf{B}(\mathsf{a})\\ \mathsf{c}\sim\mathsf{C}'(\mathsf{b})}}\left[\tilde{\mathsf{b}} \neq \mathsf{g}_{\mathsf{c}}(\mathsf{b})\right] \\
&\leq \frac{1}{\varepsilon} \cdot \Pr_{\substack{\mathsf{b}\sim\mathsf{B}\\ \mathsf{c}\sim\mathsf{C}'(\mathsf{b})}}\left[\tilde{\mathsf{b}} \neq \mathsf{g}_{\mathsf{c}}(\mathsf{b})\right] + \delta \leq \frac{\tau}{\varepsilon} + \delta.
\end{aligned}$$

The second inequality follows from the definition of the set A', the third inequality follows from sampling of A/B and the last inequality follows from Claim 2.
Finally, we bound the first term and finish the proof.

$$\begin{aligned}
\Pr_{\substack{\mathsf{a}\sim\mathsf{A}'\\ \mathsf{b}\sim\mathsf{B}(\mathsf{a})\\ \mathsf{c}\sim\mathsf{C}'(\mathsf{b})}}\left[\mathsf{g}_{\mathsf{c}}(\mathsf{b}) \in \tilde{\mathsf{a}} \mid \tilde{\mathsf{c}} \in \tilde{\mathsf{a}}\right] &= \Pr_{\substack{\mathsf{a}\sim\mathsf{A}'\\ \mathsf{b}\sim\mathsf{B}(\mathsf{a})\\ \mathsf{c}\sim\mathsf{C}'(\mathsf{b})\\ \mathsf{a}_1,\mathsf{a}_1'\sim\mathsf{A}(\mathsf{b})}}\left[\tilde{\mathsf{a}}_1 \cap \tilde{\mathsf{a}}_1' \in \tilde{\mathsf{a}} \mid \tilde{\mathsf{c}} \in \tilde{\mathsf{a}} \& \tilde{\mathsf{c}} \in \tilde{\mathsf{a}}_1 \cap \tilde{\mathsf{a}}_1' \in \mathsf{B}\right] \\
&= \Pr_{\substack{\mathsf{a}\sim\mathsf{A}'\\ \mathsf{b}\sim\mathsf{B}(\mathsf{a})\\ \mathsf{c}\sim\mathsf{C}'(\mathsf{b})\\ \mathsf{a}_1,\mathsf{a}_1',\mathsf{a}'\sim\mathsf{A}(\mathsf{b})}}\left[\tilde{\mathsf{a}}_1 \cap \tilde{\mathsf{a}}_1' \in \tilde{\mathsf{a}} \mid \tilde{\mathsf{c}} \in \tilde{\mathsf{a}} \cap \tilde{\mathsf{a}}' \in \mathsf{B} \& \tilde{\mathsf{c}} \in \tilde{\mathsf{a}}_1 \cap \tilde{\mathsf{a}}_1' \in \mathsf{B}\right] \\
&\geq \Pr_{\substack{\mathsf{c}\sim\mathsf{C}'\\ \mathsf{b}\sim\mathsf{B}(\mathsf{c})\\ \mathsf{a},\mathsf{a}_1,\mathsf{a}_1',\mathsf{a}'\sim\mathsf{A}(\mathsf{b})}}\left[\tilde{\mathsf{a}}_1 \cap \tilde{\mathsf{a}}_1' \in \tilde{\mathsf{a}} \mid \tilde{\mathsf{c}} \in \tilde{\mathsf{a}} \& \tilde{\mathsf{c}} \in \tilde{\mathsf{a}}_1 \cap \tilde{\mathsf{a}}_1' \in \mathsf{B} \& \tilde{\mathsf{a}} \in \mathsf{A}'\right] - \delta \\
&\geq \Pr_{\substack{(\mathsf{c},\mathsf{b})\\ (\mathsf{a},\mathsf{a}')\\ (\mathsf{a}_1,\mathsf{a}_1')}}\left[\tilde{\mathsf{a}}_1 \cap \tilde{\mathsf{a}}_1' = \tilde{\mathsf{a}} \cap \tilde{\mathsf{a}}' \mid \tilde{\mathsf{c}} \in \tilde{\mathsf{a}} \cap \tilde{\mathsf{a}}' \& \tilde{\mathsf{c}} \in \tilde{\mathsf{a}}_1 \cap \tilde{\mathsf{a}}_1' \in \mathsf{B} \& \tilde{\mathsf{a}} \in \mathsf{A}'\right] - \delta \\
&\geq 1 - \frac{\eta}{2\varepsilon^2} - \delta.
\end{aligned}$$

The first inequality follows from the sampling of A/C, the last inequality follow definition of being excellent and the fact that $|\mathsf{A}'| \geq 2\varepsilon^2|\mathsf{A}|$.

□

5.3 Proving the Claims

Claim 3. $\Pr_{\mathsf{c}\sim\mathsf{C}}\big[\mathsf{c}\ \textit{is excellent}\big] \geq \varepsilon/2$.

Proof. Let, $\mathsf{C}_0 \subset \mathsf{C}$ be the set of c's such that $\Pr_{\mathsf{a}\sim\mathsf{A}(\mathsf{c})}\big[\tilde{\mathsf{c}} \in \tilde{\mathsf{a}}\big] \geq 4\varepsilon$. From the main assumption it easily follows that $|\mathsf{C}_0| \geq 2\varepsilon|\mathsf{C}|$. Now for all $\mathsf{c} \in \mathsf{C}_0$, $i \in \{1,2\}$, we have,

$$\begin{aligned}\Pr_{\substack{\mathsf{b}\sim\mathsf{B}_i(\mathsf{c})\\ \mathsf{a},\mathsf{a}'\sim\mathsf{A}(\mathsf{b})}}\big[\tilde{\mathsf{c}} \in \tilde{\mathsf{a}} \cap \tilde{\mathsf{a}}' \in \mathsf{B}_i\big] &\geq \Pr_{\substack{\mathsf{b}\sim\mathsf{B}_i(\mathsf{c})\\ \mathsf{a},\mathsf{a}'\sim\mathsf{A}(\mathsf{b})}}\big[\tilde{\mathsf{c}} \in \tilde{\mathsf{a}} \cap \tilde{\mathsf{a}}'\big] - \Pr_{\substack{\mathsf{b}\sim\mathsf{B}_i(\mathsf{c})\\ \mathsf{a},\mathsf{a}'\sim\mathsf{A}(\mathsf{b})}}\big[\tilde{\mathsf{c}} \in \tilde{\mathsf{a}} \cap \tilde{\mathsf{a}}' \notin \mathsf{B}_i\big]\\ &\geq 16\varepsilon^2 - \Pr_{\substack{\mathsf{b}\sim\mathsf{B}_i(\mathsf{c})\\ \mathsf{a},\mathsf{a}'\sim\mathsf{A}(\mathsf{b})}}\big[\tilde{\mathsf{c}} \in \tilde{\mathsf{a}} \cap \tilde{\mathsf{a}}' \notin \mathsf{B}_i\big]\end{aligned}$$

where the second inequality follows from the definition of the set C_0 and Jensen's inequality. It follows that :

$$\begin{aligned}\Pr_{\mathsf{c}\sim\mathsf{C}_0}\Big[\Pr_{\substack{\mathsf{b}\sim\mathsf{B}_i(\mathsf{c})\\ \mathsf{a},\mathsf{a}'\sim\mathsf{A}(\mathsf{b})}}\Big[\tilde{\mathsf{c}} \in \tilde{\mathsf{a}} \cap \tilde{\mathsf{a}}' \in \mathsf{B}_i\Big] \leq \varepsilon^2\Big] &\leq \Pr_{\mathsf{c}\sim\mathsf{C}_0}\Big[\Pr_{\substack{\mathsf{b}\sim\mathsf{B}_i(\mathsf{c})\\ \mathsf{a},\mathsf{a}'\sim\mathsf{A}(\mathsf{b})}}\big[\tilde{\mathsf{c}} \in \tilde{\mathsf{a}} \cap \tilde{\mathsf{a}}' \notin \mathsf{B}_i\big] > \varepsilon^2\Big]\\ &\leq \mathbb{E}_{\mathsf{c}\sim\mathsf{C}_0}\Big[\Pr_{\substack{\mathsf{b}\sim\mathsf{B}_i(\mathsf{c})\\ \mathsf{a},\mathsf{a}'\sim\mathsf{A}(\mathsf{b})}}\big[\tilde{\mathsf{c}} \in \tilde{\mathsf{a}} \cap \tilde{\mathsf{a}}' \notin \mathsf{B}_i\big]\Big] \cdot \varepsilon^{-2}\\ &\leq \nu/\varepsilon^2\end{aligned}$$

For the last two inequalities we have used Markov and **entropy assumption(?)**. Thus, it follows that

$$\Pr_{\mathsf{c}\sim\mathsf{C}}\big[\mathsf{c}\ \textit{is good}\ \big] \geq 2\varepsilon - 2\nu/\varepsilon^2 \geq \varepsilon.$$

Finally, we have,

$$\begin{aligned}&\Pr_{\mathsf{c}\sim\mathsf{C}}\big[\mathsf{c}\ \textit{is excellent}\big]\\ &\geq \Pr_{\mathsf{c}\sim\mathsf{C}}\big[\mathsf{c}\ \textit{is good}\big] - \sum_{i=1,2}\Pr_{\mathsf{c}\sim\mathsf{C}}\Big[\Pr_{\substack{\mathsf{b}\sim\mathsf{B}_i(\mathsf{c})\\ \mathsf{a}_i,\mathsf{a}_i'\sim\mathsf{A}(\mathsf{b})}}\big[\tilde{\mathsf{a}}_1 \cap \tilde{\mathsf{a}}_1' \neq \tilde{\mathsf{a}}_2 \cap \tilde{\mathsf{a}}_2' | \tilde{\mathsf{c}} \in \tilde{\mathsf{a}}_i \cap \tilde{\mathsf{a}}_i' \in \mathsf{B}_i\big] > \eta\ \big|\mathsf{c}\ \textit{is good}\Big]\\ &\geq \varepsilon - \sum_{i=1,2}\Pr_{\mathsf{c}\sim\mathsf{C}_0}\Big[\Pr_{\substack{\mathsf{b}\sim\mathsf{B}_i(\mathsf{c})\\ \mathsf{a}_i,\mathsf{a}_i'\sim\mathsf{A}(\mathsf{b})}}\big[\tilde{\mathsf{a}}_1 \cap \tilde{\mathsf{a}}_1' \neq \tilde{\mathsf{a}}_2 \cap \tilde{\mathsf{a}}_2' \,\&\, \tilde{\mathsf{c}} \in \tilde{\mathsf{a}}_i \cap \tilde{\mathsf{a}}_i' \in \mathsf{B}_i\big] > \eta \cdot \varepsilon^2\Big]\\ &\geq \varepsilon - \sum_{i=1,2}\mathbb{E}_{\mathsf{c}\sim\mathsf{C}_0}\Big[\Pr_{\substack{\mathsf{b}\sim\mathsf{B}_i(\mathsf{c})\\ \mathsf{a}_i,\mathsf{a}_i'\sim\mathsf{A}(\mathsf{b})}}\big[\tilde{\mathsf{a}}_1 \cap \tilde{\mathsf{a}}_1' \neq \tilde{\mathsf{a}}_2 \cap \tilde{\mathsf{a}}_2' \,\&\, \tilde{\mathsf{c}} \in \tilde{\mathsf{a}}_i \cap \tilde{\mathsf{a}}_i' \in \mathsf{B}_i\big]\Big] \cdot (\eta\varepsilon^2)^{-1} \geq \varepsilon/2\end{aligned}$$

For the second inequality we use the fact that c is a *good* point. For the last inequality we use Markov and **entropy assumption(?)**.

□

Claim 1 (Restated). There exists a set $\mathsf{C}' \subset \mathsf{c}$ such that the following hold: 1) $|\mathsf{C}'| \geq \varepsilon^3|\mathsf{C}|$; 2) every $\mathsf{c} \in \mathsf{C}'$ is excellent; 3)

$$\Pr_{\mathsf{c}_1,\mathsf{c}_2\sim\mathsf{C}'}\Big[\Pr_{\mathsf{a}\sim\mathsf{A}(\mathsf{c}_1,\mathsf{c}_2)}\big[\tilde{\mathsf{c}}_1,\tilde{\mathsf{c}}_2 \in \tilde{\mathsf{a}}\big] \geq \varepsilon^5\Big] \geq 1-\sigma.$$

where $\sigma = 2\delta\cdot\varepsilon^{-6} + 4(\eta+\nu)\cdot\varepsilon^{-12}$

Proof. By Claim 3, it suffices to construct a large subset of $\mathsf{C}'_o :=$, the set of excellent points, such that the third property holds. For this purpose, we equip C'_0 with a graph structure: $\mathsf{c}_1,\mathsf{c}_2 \in \mathsf{C}'_0$ are adjacent if $q(\mathsf{c}_1,\mathsf{c}_2) := \Pr_{\mathsf{a}\sim\mathsf{A}(\mathsf{c}_1,\mathsf{c}_2)}\big[\tilde{\mathsf{c}}_1,\tilde{\mathsf{c}}_2 \in \tilde{\mathsf{a}}\big] \geq \varepsilon^2$. Our final set will be the neighborhood, $\mathsf{N}(\mathsf{c}) := \{\mathsf{c} \in \mathsf{C}'_0 : q(\mathsf{c},\mathsf{c}') \geq \varepsilon^2\}$ of some $\mathsf{c}' \in \mathsf{C}'_0$. In order for this to work, c' should satisfy: 1) $|\mathsf{N}(\mathsf{c}')|$ must be large; 2) $\Pr_{\mathsf{c}_1,\mathsf{c}_2\sim\mathsf{N}(\mathsf{c}')}\big[q(\mathsf{c}_1,\mathsf{c}_2) < \varepsilon^5\big]$ must be small. We show there exists such $\mathsf{c}' \in \mathsf{C}'_0$. Specifically, we prove

1. $\mathbb{E}_{\mathsf{c},\mathsf{c}'\sim\mathsf{C}'_0}\big[q(\mathsf{c},\mathsf{c}')\big] \geq 3\varepsilon^2$; and
2. $\Pr\limits_{\substack{\mathsf{c}'\sim\mathsf{C}'_0\\ \mathsf{c},\mathsf{c}''\sim\mathsf{N}(\mathsf{c}')}}\big[q(\mathsf{c}_1,\mathsf{c}_2) \geq \varepsilon^5 \mid |\mathsf{N}(\mathsf{c}')| > \varepsilon^3|\mathsf{C}|\big] \geq 1-\sigma.$

It follows from the first point that $\Pr_{\mathsf{c}'\sim\mathsf{C}'_0}\big[|\mathsf{N}(\mathsf{c}')| \geq \varepsilon^3|\mathsf{c}|\big] > \varepsilon^2$ (using $|\mathsf{C}'_0| \geq \varepsilon|\mathsf{c}|$). Thus, the two points together guarantee the existence of some $\mathsf{c}' \in \mathsf{C}'_0$ such that $|\mathsf{N}(\mathsf{c}')| \geq \varepsilon^3|\mathsf{c}|$ and $\Pr_{\mathsf{c},\mathsf{c}''\sim\mathsf{N}(\mathsf{c}')}\big[\mathsf{q}(\mathsf{c},\mathsf{c}'') \geq \varepsilon^5\big] \geq 1-\sigma$. Setting $\mathsf{c}' = \mathsf{N}(\mathsf{c}')$ for such a $\mathsf{c}' \in \mathsf{C}'_0$ completes the proof. So it remains to establish the above two bounds.

$$\begin{aligned}
\mathop{\mathbb{E}}_{\mathsf{c},\mathsf{c}'\sim\mathsf{C}'_0}\big[q(\mathsf{c},\mathsf{c}')\big] &\geq \mathop{\mathbb{E}}_{\mathsf{a}\sim\mathsf{A}}\Big[\Pr_{\mathsf{c}\sim\mathsf{C}'_0(\mathsf{a})}\big[\tilde{\mathsf{c}} \in \tilde{\mathsf{a}}\big]^2\Big] - \delta \geq \mathop{\mathbb{E}}_{\mathsf{a}\sim\mathsf{A}}\Big[\Pr_{\mathsf{c}\sim\mathsf{C}'_0(\mathsf{a})}\big[\tilde{\mathsf{c}} \in \tilde{\mathsf{a}}\big]\Big]^2 - \delta \\
&\geq \mathop{\mathbb{E}}_{\mathsf{c}\sim\mathsf{C}'_0}\big[\mu_{\mathsf{c}}\big]^2 - 3\delta \geq 16\varepsilon^2 - 3\delta \geq 3\varepsilon^2.
\end{aligned}$$

We have used the sampling of A/C^2, Jensen's inequality, the sampling of A/C, and the fact that $\mu_{\mathsf{c}} \geq 4\varepsilon$ for all $\mathsf{c} \in \mathsf{C}'_0$. Now, we show the second bound. Towards that end, we fix $\mathsf{c} \sim \mathsf{C}'_0, \mathsf{c}_1,\mathsf{c}_2 \sim \mathsf{N}(\mathsf{c})$ (by fixing these, we fix $\mathsf{b}_i \in \mathsf{B}_1$, for i=1, 2 such that b_i is the line containing the points $(\mathsf{c},\mathsf{c}_i)$) and define four quantities, shorthanded as $\mathsf{val}_1,\mathsf{val}_2,\mathsf{val}_3,\mathsf{val}_4$;

- $\mathsf{val}_1 := |\Pr_{\mathsf{a}\sim\mathsf{A}(\mathsf{c}_1,\mathsf{c}_2)}\big[\tilde{\mathsf{c}}_1,\tilde{\mathsf{c}}_2 \in \tilde{\mathsf{a}}\big] - \Pr_{\mathsf{a}\sim\mathsf{A}(\mathsf{c},\mathsf{c}_1,\mathsf{c}_2)}\big[\tilde{\mathsf{c}}_1,\tilde{\mathsf{c}}_2 \in \tilde{\mathsf{a}}\big]|$;
- $\mathsf{val}_2 := |\Pr_{\mathsf{a}\sim\mathsf{A}(\mathsf{c})}\big[\tilde{\mathsf{c}} \in \tilde{\mathsf{a}}\big] - \Pr_{\mathsf{a}\sim\mathsf{A}(\mathsf{c},\mathsf{c}_1,\mathsf{c}_2)}\big[\tilde{\mathsf{c}} \in \tilde{\mathsf{a}}\big]|$;
- $\mathsf{val}_3 := \sum_{i=1}^2 \Pr_{\mathsf{a}_i,\mathsf{a}'_i\sim\mathsf{A}(\mathsf{b}_i)}\big[\tilde{\mathsf{c}} \in \tilde{\mathsf{a}}_i \cap \tilde{\mathsf{a}}'_i \;\&\; \tilde{\mathsf{a}}_i \cap \tilde{\mathsf{a}}'_i \neq \tilde{\mathsf{b}}_i\big]$
- $\mathsf{val}_4 := \sum_{i=1}^2 \Pr\limits_{\substack{\mathsf{a}\sim\mathsf{A}(\mathsf{c}_1,\mathsf{c}_2,\mathsf{c})\\ \mathsf{a}_i\sim\mathsf{A}(\mathsf{b}_i)}}\big[\tilde{\mathsf{c}} \in \tilde{\mathsf{a}} \cap \tilde{\mathsf{a}}_i \;\&\; \tilde{\mathsf{a}}_i \cap \tilde{\mathsf{a}} \neq \tilde{\mathsf{b}}_i\big]$

Where $\tilde{\mathsf{b}}_i = \mathsf{g}_{\mathsf{c}}(\mathsf{b}_i)$. We show that each val_i is very small with very high probability over $(\mathsf{c},\mathsf{c}_1,\mathsf{c}_2)$. These bounds will be used in the computation which follows.

$$\Pr_{(\mathsf{c},\mathsf{c}_1,\mathsf{c}_2)}\big[\mathsf{val}_1 > \delta\big] \leq \varepsilon^{-3}\cdot\max_{\mathsf{c}_1,\mathsf{c}_2\in\mathsf{C}}\Big\{\Pr_{\mathsf{c}\sim\mathsf{C}}\Big[\Big|\mathbb{E}_{\mathsf{a}'\sim\mathsf{A}(\mathsf{c},\mathsf{c}_1,\mathsf{c}_2)}\big[f_1(\mathsf{a}')\big] - \mathbb{E}_{\mathsf{a}'\sim\mathsf{A}(\mathsf{c}_1,\mathsf{c}_2)}\big[f_1(\mathsf{a}')\big]\Big| > \delta\Big]\Big\},$$

where $f_1(\mathsf{a}') = 1$ if $\tilde{\mathsf{c}}_1, \tilde{\mathsf{c}}_2 \in \tilde{\mathsf{a}}'$, 0 otherwise. Thus, $\Pr_{(\mathsf{c},\mathsf{c}_1,\mathsf{c}_2)}\big[\mathsf{val}_1 > \delta\big] \leq \delta/\varepsilon^3$, by the sampling of $\mathsf{A}(\mathsf{c}_1,\mathsf{c}_2)/\mathsf{C}$ for all $\mathsf{c}_1, \mathsf{c}_2 \in \mathsf{C}$. Similarly, $\Pr_{(\mathsf{c},\mathsf{c}_1,\mathsf{c}_2)}\big[\mathsf{val}_2 > \delta\big] \leq \delta/\varepsilon^6$, follows from the same reasoning using the sampling of $\mathsf{A}(\mathsf{c})/\mathsf{C}^2$ and the function $f_2(\mathsf{a}') = 1$ if $\tilde{\mathsf{c}} \in \tilde{\mathsf{a}}'$.
Now, we bound val_3 and val_4. For val_3, we have

$$\begin{aligned}
\Pr_{\mathsf{c},\mathsf{c}_1,\mathsf{c}_2}\big[\mathsf{val}_3 > \varepsilon^6\big] &\leq \varepsilon^{-6} \cdot \mathbb{E}_{\mathsf{c},\mathsf{c}_1,\mathsf{c}_2}\big[2 \cdot \Pr_{\mathsf{a}_1,\mathsf{a}_1'\sim\mathsf{A}(\mathsf{b}_1)}\big[\tilde{\mathsf{c}} \in \tilde{\mathsf{a}}_1 \cap \tilde{\mathsf{a}}_1' \;\&\; \tilde{\mathsf{a}}_1 \cap \tilde{\mathsf{a}}_1' \neq \tilde{\mathsf{b}}_1\big]\big] \\
&\leq 2\varepsilon^{-12} \cdot \mathbb{E}_{\substack{\mathsf{c}\sim\mathsf{C}_0' \\ \mathsf{b}_1\sim\mathsf{B}_1(\mathsf{c})}}\Big[\Pr_{\mathsf{a}_1,\mathsf{a}_1'\sim\mathsf{A}(\mathsf{b}_1)}\big[\tilde{\mathsf{c}} \in \tilde{\mathsf{a}}_1 \cap \tilde{\mathsf{a}}_1' \;\&\; \tilde{\mathsf{a}}_1 \cap \tilde{\mathsf{a}}_1' \neq \tilde{\mathsf{b}}_1\big]\Big] \\
&\leq 2\varepsilon^{-12} \cdot \Big(\Pr_{\substack{\mathsf{c},\mathsf{b}_1 \\ \mathsf{a}_1,\mathsf{a}_1'}}\Big[\tilde{\mathsf{a}}_1 \cap \tilde{\mathsf{a}}_1' \neq \tilde{\mathsf{b}}_1 | \tilde{\mathsf{c}} \in \tilde{\mathsf{a}}_1 \cap \tilde{\mathsf{a}}_1' \in \mathsf{B}_1\Big] + \Pr_{\substack{\mathsf{c},\mathsf{b}_1 \\ \mathsf{a}_1,\mathsf{a}_1'}}\Big[\tilde{\mathsf{c}} \in \tilde{\mathsf{a}}_1 \cap \tilde{\mathsf{a}}_1' \notin \mathsf{B}_1\Big]\Big) \\
&\leq 2\varepsilon^{-12} \cdot (\eta + \nu)
\end{aligned}$$

Where the second last probability is over $\mathsf{c} \sim \mathsf{C}_0', \mathsf{b}_1 \sim \mathsf{B}_1(\mathsf{c}), \mathsf{a}_1, \mathsf{a}_1' \sim \mathsf{A}(\mathsf{b}_1)$. The argument for val_4 works identically as above. Now, we have:

$$\begin{aligned}
q(\mathsf{c}_1,\mathsf{c}_2) &\geq \Pr_{\mathsf{a}\sim\mathsf{A}(\mathsf{c},\mathsf{c}_1,\mathsf{c}_2)}\big[\tilde{\mathsf{c}}_1, \tilde{\mathsf{c}}_2 \in \tilde{\mathsf{a}}\big] - \mathsf{val}_1 \\
&\geq \Pr_{\substack{\mathsf{a}\sim\mathsf{A}(\mathsf{c},\mathsf{c}_1,\mathsf{c}_2) \\ \mathsf{a}_i,\mathsf{a}_i'\sim\mathsf{A}(\mathsf{b}_i),\ i=1,2}}\big[\tilde{\mathsf{c}}, \tilde{\mathsf{c}}_i \in \tilde{\mathsf{a}}_i \cap \tilde{\mathsf{a}}_i' \;\&\; \tilde{\mathsf{a}}_i \cap \tilde{\mathsf{a}}_i' = \mathsf{g}_{\mathsf{c}}(\mathsf{b}_i) \;\&\; \tilde{\mathsf{a}} \cap \tilde{\mathsf{a}}_i = \mathsf{g}_{\mathsf{c}}(\mathsf{b}_i) \;\&\; \tilde{\mathsf{c}} \in \tilde{\mathsf{a}}\big] - \mathsf{val}_1 \\
&\geq \Pr_{\substack{\mathsf{a}\sim\mathsf{A}(\mathsf{c},\mathsf{c}_1,\mathsf{c}_2) \\ \mathsf{a}_i,\mathsf{a}_i'\sim\mathsf{A}(\mathsf{b}_i),\ i=1,2}}\big[\tilde{\mathsf{c}} \in \tilde{\mathsf{a}} \;\&\; \tilde{\mathsf{c}}, \tilde{\mathsf{c}}_i \in \tilde{\mathsf{a}}_i \cap \tilde{\mathsf{a}}_i'\big] - (\mathsf{val}_3 + \mathsf{val}_4 + \mathsf{val}_1) \\
&\geq \Pr_{\mathsf{a}\sim\mathsf{A}(\mathsf{c},\mathsf{c}_1\mathsf{c}_2)}\big[\tilde{\mathsf{c}} \in \tilde{\mathsf{a}}\big] \cdot \Big(\Pr_{\mathsf{a}_1,\mathsf{a}_1'\sim\mathsf{A}(\mathsf{c},\mathsf{c}_1)}\big[\tilde{\mathsf{c}}, \tilde{\mathsf{c}}_1 \in \tilde{\mathsf{a}}_1 \cap \tilde{\mathsf{a}}_1'\big]\Big)^2 - (\mathsf{val}_3 + \mathsf{val}_4 + \mathsf{val}_1) \\
&\geq 4\varepsilon^5 - (\mathsf{val}_1 + \mathsf{val}_2 + \mathsf{val}_3 + \mathsf{val}_4)
\end{aligned}$$

The result follows:

$$\Pr_{\substack{\mathsf{c}'\sim\mathsf{C}_0' \\ \mathsf{c},\mathsf{c}''\sim\mathsf{N}(\mathsf{c}')}}\big[q(\mathsf{c},\mathsf{c}'') \geq \varepsilon^5 \mid |\mathsf{N}(\mathsf{c}')| > \varepsilon^3|\mathsf{C}|\big] \geq \Pr_{\mathsf{c}_0,\mathsf{c}_1,\mathsf{c}_2}\big[\mathsf{val}_1 + \mathsf{val}_2 + \mathsf{val}_3 + \mathsf{val}_4 \leq 3\varepsilon^5\big] \geq 1 - \sigma.$$

□

Claim 2 (Restated). Let C' be the set as in claim 1. We have, $\Pr_{\substack{\mathsf{b}\sim\mathsf{B}_2 \\ \mathsf{c}_1,\mathsf{c}_2\sim\mathsf{C}'(\mathsf{b})}}\big[\tilde{\mathsf{b}}_1 = \tilde{\mathsf{b}}_2\big] \geq 1 - \tau$, where $\tau := 1 - \Big(4\eta \cdot \varepsilon^{-10} + 2\gamma + \delta\Big)$.

Proof. We will use B to denote B_2. For any $\mathsf{c}_1, \mathsf{c}_2$, the event E holds if:

$$\Pr_{\mathsf{b}\sim\mathsf{B}_2(\mathsf{c}_1,\mathsf{c}_2)}\Big[\Pr_{\mathsf{a},\mathsf{a}'\sim\mathsf{A}(\mathsf{b})}\big[\tilde{\mathsf{c}}_1, \tilde{\mathsf{c}}_2 \in \tilde{\mathsf{a}} \cap \tilde{\mathsf{a}}' \in \mathsf{B}_2\big] \geq \varepsilon^{10}\Big] \geq 1 - \gamma$$

. Now, we have,

$$\begin{aligned}
\Pr_{\substack{\mathsf{b}\sim\mathsf{B} \\ \mathsf{c}_1,\mathsf{c}_2\sim\mathsf{C}'(\mathsf{b})}}\big[\tilde{\mathsf{b}}_1 = \tilde{\mathsf{b}}_2\big] &\geq \Pr_{\substack{\mathsf{c}_1,\mathsf{c}_2\sim\mathsf{C}' \\ \mathsf{b}\sim\mathsf{B}(\mathsf{c},\mathsf{c}')}}\big[\tilde{\mathsf{b}}_1 = \tilde{\mathsf{b}}_2\big] - \delta \geq \Pr_{\mathsf{c}_1,\mathsf{c}_2\sim\mathsf{C}'}\big[\mathsf{E}\big] \cdot \Pr_{\substack{\mathsf{c}_1,\mathsf{c}_2\sim\mathsf{C}' \\ \mathsf{b}\sim\mathsf{B}(\mathsf{c},\mathsf{c}')}}\Big[\tilde{\mathsf{b}}_1 = \tilde{\mathsf{b}}_2 \Big| \mathsf{E}\Big] - \delta \\
&\geq \mathbb{E}_{\mathsf{c}_1,\mathsf{c}_2\sim\mathsf{C}'}\Big[\Pr_{\mathsf{b}\sim\mathsf{B}(\mathsf{c},\mathsf{c}')}\big[\tilde{\mathsf{b}}_1 = \tilde{\mathsf{b}}_2\big]\Big|\mathsf{E}\Big] - \gamma - \delta
\end{aligned}$$

The first inequality follows from sampling of C^2/B. The last inequality follows from fact 2. We bound the inner probability of the first term above. For any $\mathsf{c}_1, \mathsf{c}_2$ such that E holds we have,

$$\begin{aligned}\Pr_{\mathsf{b}\sim\mathsf{B}(\mathsf{c}_1,\mathsf{c}_2)}\left[\tilde{\mathsf{b}}_1 = \tilde{\mathsf{b}}_2\right] &\geq \Pr_{\mathsf{b}\sim\mathsf{B}(\mathsf{c}_1,\mathsf{c}_2)}\left[\exists \mathsf{a}, \mathsf{a}' \sim \mathsf{A}(\mathsf{b}) \text{ s.t. } \tilde{\mathsf{b}}_1 = \tilde{\mathsf{a}} \cap \tilde{\mathsf{a}}' = \tilde{\mathsf{b}}_2 \;\&\; \mathsf{c}_1, \mathsf{c}_2 \in \tilde{\mathsf{a}} \cap \tilde{\mathsf{a}}'\right] \\ &\geq \Pr_{\substack{\mathsf{b}\sim\mathsf{B}(\mathsf{c}_1,\mathsf{c}_2)\\ \mathsf{a},\mathsf{a}'\sim\mathsf{A}(\mathsf{b})}}\left[\tilde{\mathsf{b}}_1 = \tilde{\mathsf{a}} \cap \tilde{\mathsf{a}}' = \tilde{\mathsf{b}}_2 \middle| \mathsf{c}_1, \mathsf{c}_2 \in \tilde{\mathsf{a}} \cap \tilde{\mathsf{a}}' \in \mathsf{B}_2\right] - \gamma \\ &\geq 1 - (\mathsf{val}_1 + \mathsf{val}_2 + \gamma).\end{aligned}$$

where $\mathsf{val}_i := \Pr_{\mathsf{b},\mathsf{a},\mathsf{a}'}\left[\tilde{\mathsf{b}}_i \neq \tilde{\mathsf{a}} \cap \tilde{\mathsf{a}}' \mid \mathsf{c}_1, \mathsf{c}_2 \in \tilde{\mathsf{a}} \cap \tilde{\mathsf{a}}' \in \mathsf{B}_2\right]$ for $i = 1, 2$. In the second line, we have used the fact that event E holds. We finish by bounding the terms val_1; bounding val_2 works identically. We have,

$$\mathsf{val}_1 \leq \frac{\Pr_{\substack{\mathsf{b}\sim\mathsf{B}(\mathsf{c}_1,\mathsf{c}_2)\\ \mathsf{a},\mathsf{a}'\sim\mathsf{A}(\mathsf{b})}}\left[\tilde{\mathsf{b}}_1 \neq \tilde{\mathsf{a}} \cap \tilde{\mathsf{a}}' \mid \mathsf{c}_1 \in \tilde{\mathsf{a}} \cap \tilde{\mathsf{a}}' \in \mathsf{B}_2\right]}{\Pr_{\substack{\mathsf{b}\sim\mathsf{B}(\mathsf{c}_1,\mathsf{c}_2)\\ \mathsf{a},\mathsf{a}'\sim\mathsf{A}(\mathsf{b})}}\left[\tilde{\mathsf{c}}_1, \tilde{\mathsf{c}}_2 \in \tilde{\mathsf{a}} \cap \tilde{\mathsf{a}}' \in \mathsf{B}_2\right]} \leq \frac{\eta}{\varepsilon^{10} \cdot (1-\gamma)} \leq 2\eta \cdot \varepsilon^{-10}$$

The last inequality follows from the excellence of c_1 and fact 2. The claims then follows.

□

Fact 2. *If $\mathsf{C}' \subset \mathsf{C}$ be the set such that: $\Pr_{\mathsf{c}_1,\mathsf{c}_2\sim\mathsf{C}'}\left[\Pr_{\mathsf{a}\sim\mathsf{A}(\mathsf{c}_1,\mathsf{c}_2)}\left[\tilde{\mathsf{c}}_1, \tilde{\mathsf{c}}_2 \in \tilde{\mathsf{a}}\right] \geq 3\varepsilon^5\right] \geq 1 - \sigma$, then we have*

$$\Pr_{\mathsf{c}_1,\mathsf{c}_2\sim\mathsf{C}'}\left[\Pr_{\mathsf{b}\sim\mathsf{B}_2(\mathsf{c}_1,\mathsf{c}_2)}\left[\Pr_{\mathsf{a},\mathsf{a}'\sim\mathsf{A}(\mathsf{b})}\left[\tilde{\mathsf{c}}_1, \tilde{\mathsf{c}}_2 \in \tilde{\mathsf{a}} \cap \tilde{\mathsf{a}}' \in \mathsf{B}_2\right] \geq \varepsilon^{10}\right] \geq 1 - \gamma\right] \geq 1 - \gamma$$

Where $\gamma = \max\{\delta + \sqrt{(\delta+\nu)\cdot\varepsilon^{-10}}, \sigma + \sqrt{(\delta+\nu)\cdot\varepsilon^{-10}}\}$

Proof. We will use B to denote B_2. For $\mathsf{c}_1, \mathsf{c}_2 \in \mathsf{C}'$, let $\mathsf{A}' \subset \mathsf{A}(\mathsf{c}_1, \mathsf{c}_2)$ be set of $\mathsf{a} \in \mathsf{A}(\mathsf{c}_1, \mathsf{c}_2)$ such that $\tilde{\mathsf{c}}_1, \tilde{\mathsf{c}}_2 \in \tilde{\mathsf{a}}$. From sampling of $\mathsf{A}(\mathsf{c}_1, \mathsf{c}_2)/\mathsf{B}(\mathsf{c}_1, \mathsf{c}_2)$, we have,

$$\Pr_{\mathsf{b}\sim\mathsf{B}(\mathsf{c}_1,\mathsf{c}_2)}\left[\left|\Pr_{\mathsf{a}\sim\mathsf{A}(\mathsf{b})}\left[\tilde{\mathsf{a}} \in \mathsf{A}'\right] - \Pr_{\mathsf{a}\sim\mathsf{A}(\mathsf{c}_1,\mathsf{c}_2)}\left[\tilde{\mathsf{a}} \in \mathsf{A}'\right]\right| > \delta\right] \leq \delta$$

It follows that: $\Pr_{\mathsf{c}_1,\mathsf{c}_2\sim\mathsf{C}'}\left[\Pr_{\mathsf{b}\sim\mathsf{B}(\mathsf{c}_1,\mathsf{c}_2)}\left[\Pr_{\mathsf{a},\mathsf{a}'\sim\mathsf{A}(\mathsf{b})}\left[\tilde{\mathsf{c}}_1, \tilde{\mathsf{c}}_2 \in \tilde{\mathsf{a}} \cap \tilde{\mathsf{a}}'\right] \geq 4\varepsilon^{10}\right] \geq 1-\delta\right] \geq 1-\sigma$. Now,

$$\begin{aligned}\Pr_{\mathsf{c}_1,\mathsf{c}_2,\mathsf{b}}\left[\Pr_{\mathsf{a},\mathsf{a}'}\left[\tilde{\mathsf{c}}_1, \tilde{\mathsf{c}}_2 \in \tilde{\mathsf{a}} \cap \tilde{\mathsf{a}}' \notin \mathsf{B}\right] > \varepsilon^{10}\right] &\leq \mathbb{E}_{\mathsf{c}_1,\mathsf{c}_2,\mathsf{b}}\left[\Pr_{\mathsf{a},\mathsf{a}'}\left[\tilde{\mathsf{c}}_1, \tilde{\mathsf{c}}_2 \in \tilde{\mathsf{a}} \cap \tilde{\mathsf{a}}' \notin \mathsf{B}\right]\right] \cdot \varepsilon^{-10} \\ &\leq \left(\mathbb{E}_{\substack{\mathsf{c}_1\sim\mathsf{C}'\\ \mathsf{b}\sim\mathsf{B}(\mathsf{c}_1)}}\left[\Pr_{\mathsf{a},\mathsf{a}'\sim\mathsf{A}(\mathsf{b})}\left[\tilde{\mathsf{c}}_1, \tilde{\mathsf{c}}_2 \in \tilde{\mathsf{a}} \cap \tilde{\mathsf{a}}' \notin \mathsf{B}\right]\right] + \delta\right) \cdot \varepsilon^{-10} \\ &\leq (\delta + \nu) \cdot (\varepsilon^{-10})\end{aligned}$$

The second line uses Markov and sampling of C/B. The last inequality follows from the **Entropy assumption(?)**. It follows,

$$\Pr_{\mathsf{c}_1,\mathsf{c}_2\sim\mathsf{C}'}\left[\Pr_{\mathsf{b}\sim\mathsf{B}(\mathsf{c}_1,\mathsf{c}_2)}\left[\Pr_{\mathsf{a},\mathsf{a}'\sim\mathsf{A}(\mathsf{b})}\left[\tilde{\mathsf{c}}_1, \tilde{\mathsf{c}}_2 \in \tilde{\mathsf{a}} \cap \tilde{\mathsf{a}}' \in \mathsf{B}\right] \geq \varepsilon^{10}\right] \geq 1 - \gamma\right] \geq 1 - \gamma$$

Where $\gamma = \max\{\delta + \sqrt{(\delta+\nu)\cdot\varepsilon^{-10}}, \sigma + \sqrt{(\delta+\nu)\cdot\varepsilon^{-10}}\}$

□

6 Future Work

In this section, we describe the remaining parts of our conjecture.

6.1 The entropy assumption from section 5

Throughout section 5, we assume that the function $\mathsf{h} : \mathsf{C} \to \mathsf{C}$ possesses significant entropy and use this assumption multiple times in our proofs. Formally, the assumption is:

$$\left[\Pr_{\substack{\mathsf{b}\sim\mathsf{B}_1 \\ \mathsf{a},\mathsf{a}'\sim\mathsf{A}(\mathsf{b}) \\ \mathsf{c}\sim\mathsf{C}}}\left[\tilde{\mathsf{c}} \in \tilde{\mathsf{a}} \cap \tilde{\mathsf{a}}' \notin \mathsf{B}_i\right]\right] \leq \nu$$

The reason why we believe that this assumption holds is the following. We first fix a, a', thus also fix $\tilde{\mathsf{a}}, \tilde{\mathsf{a}}'$. We then draw c, thus fix $\tilde{\mathsf{c}}$, which is a random point that is not any way related to $\tilde{\mathsf{a}}$ or $\tilde{\mathsf{a}}'$. In addition, since $\tilde{\mathsf{a}}$ and $\tilde{\mathsf{a}}'$ are affine planes, they can agree on at most point if they do not agree on a bigger surface. Thus, $\tilde{\mathsf{a}}$ and $\tilde{\mathsf{a}}'$ intersect a point and not at a line[5], the probability that the randomly picked $\tilde{\mathsf{c}}$ being that point of intersection is $\mathcal{O}(|\mathbb{F}|)$.

6.2 Show linearity in section 5

So far, we have showed that if incidence preserving property holds with non-trivial probability, namely $\Pr_{\substack{\mathsf{a}\sim\mathsf{A} \\ \mathsf{c}\sim\mathsf{C}(\mathsf{a})}}\left[\tilde{\mathsf{c}} \in \tilde{\mathsf{a}}\right] = 6\varepsilon$, there is a non-trivial subset of 3-planes where this incidence preserving property holds with very high probability. This is not enough to show that the tampering functions that keep the amount of agreement the same are affine. We plan to use a form of fundamental theorem of projective geometry, which is defined and proved in [NR20], which basically states that if a function preserves incidences with high probability, then the function acts in a linear fashion. [6]

6.3 Finishing the proof

Even though we strongly believe that constant and random cases are not mauling, we still have to show that this is the case formally. We also might want to prove a master theorem that shows equivalence between tests consisting of planes of different sizes. For example, we want to show equivalence between situations where $\mathsf{a}, \mathsf{a}' \in \mathsf{A}$, $\mathsf{a} \cap \mathsf{a}' \in \mathsf{B}_1 \& \tilde{\mathsf{a}} \cap \tilde{\mathsf{a}}' \in \mathsf{B}_2$ and $\mathsf{a}, \mathsf{a}' \in \mathsf{B}_2$, $\mathsf{a} \cap \mathsf{a}' \in \mathsf{C} \& \tilde{\mathsf{a}} \cap \tilde{\mathsf{a}}' \in \mathsf{B}_1$, and categorize both of this situations as increased agreement. This way, it would be easier for us connect different parts of our proof.

6.4 Dealing with tampering of polynomials

So far, we have not considered the case where polynomials get tampered and we focused only on tampering of planes. However, we are optimistic the functions that tamper with the planes

[5] Any bigger surface is not considered since we are in the case where tampering functions keep agreement the same

[6] Going from linear to affine is trivial, so we use those terms interchangeably.

are independent of the polynomials and the job we have done so far about tampering of planes is legitimate. But we still need to work on the cases where polynomials get tampered. We are again optimistic about dealing with cases about tampering of polynomials because polynomials are more powerful objects compared to planes and we believe that it will be easier to deal with the tampering of them.

6.5 Using a locking scheme to improve security

The code (or the type of codes) given in section 3 is not non-malleable against decision trees because the decision trees can choose to query a plane going through the origin (and the corresponding 3-variate polynomial) and recover the message. But if decision trees are given random access to the codeword, then decision trees learn nothing about the underlying message, except with probability $\mathcal{O}(|\mathbb{F}|^{-1})$, which is the probability of that random plane going through the origin. Towards this end, [KPT97] describes a method to randomize the access patterns to the code by composing it with an outer code, called a locking scheme. Using the locking scheme, we "lock" each symbol of the code so that if the "key" is known, the underlying symbol can be recovered with a few queries and if the key is not known, the underlying symbol cannot be obtained from the locked symbol. If we combine this type of locking scheme with a random permutations π of planes A, namely if the decision tree asks to query the polynomial at location a, it receives instead the key for the symbol $\pi(\mathsf{a})$.

6.6 Designing an improved locking scheme

The locking scheme described in the previous section is sufficient for proving zero-knowledge. However, non-malleability is a stronger notion than zero-knowledge. Thus, we will have to design a new locking scheme that is suitable for our needs. We expect to be able to design the locking scheme by directly modifying [KPT97]'s locking scheme. If that does not work, we might use techniques from the area of non-malleable commitments [DDN, GRRV14, GPR16].

6.7 Showing non-malleability against decision trees

In this section, we give the big picture of what we have done so far and what we plan to do. So far, we have categorized the tampering of planes and finished working on one of the cases, which is the case when tampering increases agreement of planes. We also have accomplished significant amount of work on the hardest case, which is the one where agreement stays the same after tampering, and reduced that problem to basically fundamental theorem of projective geometry. We also have some intuition about the remaining two cases.
We first need to finish the proof of plane tampering case analysis by proving the remaining parts. We still need to deal with possible polynomial tampering but we believe that it is going to be simpler than plane tampering as polynomials are stronger objects. We also need to come up with a locking scheme that is tailored to our needs. Finally, we will compose these codes together and we will get a locally testable non-malleable code against decision trees with depth $d \geq n^{1/4-o(1)}$, which is better than the current state-of-the-art for non-malleable codes against decision trees [BGW19].

7 Discussion

We believe that there is either a direct connection between decision tree tampering and split-state tampering or we can use the methods we used to get locally testable non-malleable codes against decision trees to get constant-rate split-state non-malleable codes. The reason for our optimism is because our construction is a fresh perspective on split-state non-malleable codes that use low-degree polynomials and linearity tests instead of randomness extractors, which are functions which convert a non-uniform source of randomness that possesses a certain amount of entropy into a random variable. The methods that use randomness extractors for constructing split-state non-malleable codes put a restriction on the tampering functions: they cannot possess a lot of entropy. It is proven in [RR19] that a tampering function that tampers with the type of code in section 3 in a coordinate-wise fashion is actually tampering in an affine fashion. This is a much stronger restriction than the restriction imposed by not possessing a lot of entropy. Non-malleability against decision tree tampering is one step further down this road and we believe that if we can achieve non-malleability against decision trees, it will lead us to construct constant-rate split state non-malleable codes. Moreover, we believe that it will also give us a new construction, called non-malleable PCPs by using a known connection between locally testable codes and PCPs. We will further explore this path and possible applications of this new concept.

References

[ADKO15] Divesh Aggarwal, Yevgeniy Dodis, Tomasz Kazana, and Maciej Obremski. Non-malleable reductions and applications. In *Proceedings of the forty-seventh annual ACM symposium on Theory of Computing*, pages 459–468, 2015.

[ADN+19] Divesh Aggarwal, Nico Döttling, Jesper Buus Nielsen, Maciej Obremski, and Erick Purwanto. Continuous non-malleable codes in the 8-split-state model. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pages 531–561. Springer, 2019.

[AGM+15] Shashank Agrawal, Divya Gupta, Hemanta K Maji, Omkant Pandey, and Manoj Prabhakaran. Explicit non-malleable codes against bit-wise tampering and permutations. In *Annual Cryptology Conference*, pages 538–557. Springer, 2015.

[AKO17] Divesh Aggarwal, Tomasz Kazana, and Maciej Obremski. Inception makes non-malleable codes stronger. In *Theory of Cryptography Conference*, pages 319–343. Springer, 2017.

[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 (JACM)*, 45(3):501–555, 1998.

[AO19] Divesh Aggarwal and Maciej Obremski. Inception makes non-malleable codes shorter as well! *IACR Cryptol. ePrint Arch.*, 2019:399, 2019.

[AS98] Sanjeev Arora and Shmuel Safra. Probabilistic checking of proofs: A new characterization of np. *Journal of the ACM (JACM)*, 45(1):70–122, 1998.

[BDSG+18] Marshall Ball, Dana Dachman-Soled, Siyao Guo, Tal Malkin, and Li-Yang Tan. Non-malleable codes for small-depth circuits. In *2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 826–837. IEEE, 2018.

[BGW19] Marshall Ball, Siyao Guo, and Daniel Wichs. Non-malleable codes for decision trees. In *Annual International Cryptology Conference*, pages 413–434. Springer, 2019.

[CDM+20] Sandro Coretti, Yevgeniy Dodis, Ueli Maurer, Björn Tackmann, and Daniele Venturi. Non-malleable encryption: simpler, shorter, stronger. *Journal of Cryptology*, pages 1–50, 2020.

[CGL16] Eshan Chattopadhyay, Vipul Goyal, and Xin Li. Non-malleable extractors and codes, with their many tampered extensions. In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, pages 285–298, 2016.

[CZ16] Eshan Chattopadhyay and David Zuckerman. Explicit two-source extractors and resilient functions. In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, pages 670–683, 2016.

[DDN] D Dolev, C Dwork, and M Naor. Non-malleable cryptography extended abstract. In *Proceedings of the 23rd Annual ACM Symposium on Theory of Computing*.

[DJMW12] Yevgeniy Dodis, Abhishek Jain, Tal Moran, and Daniel Wichs. Counterexamples to hardness amplification beyond negligible. In *Theory of Cryptography Conference*, pages 476–493. Springer, 2012.

[DK17] Irit Dinur and Tali Kaufman. High dimensional expanders imply agreement expanders. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 974–985. IEEE, 2017.

[DP07] Stefan Dziembowski and Krzysztof Pietrzak. Intrusion-resilient secret sharing. In *48th Annual IEEE Symposium on Foundations of Computer Science (FOCS'07)*, pages 227–237. IEEE, 2007.

[DPW18] Stefan Dziembowski, Krzysztof Pietrzak, and Daniel Wichs. Non-malleable codes. *Journal of the ACM (JACM)*, 65(4):1–32, 2018.

[FS95] Katalin Friedl and Madhu Sudan. Some improvements to total degree tests. In *Proceedings Third Israel Symposium on the Theory of Computing and Systems*, pages 190–198. IEEE, 1995.

[GPR16] Vipul Goyal, Omkant Pandey, and Silas Richelson. Textbook non-malleable commitments. In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, pages 1128–1141, 2016.

[GR19] Vipul Goyal and Silas Richelson. Non-malleable commitments using goldreich-levin list decoding. In *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 686–699. IEEE, 2019.

[GRRV14] Vipul Goyal, Silas Richelson, Alon Rosen, and Margarita Vald. An algebraic approach to non-malleability. In *2014 IEEE 55th Annual Symposium on Foundations of Computer Science*, pages 41–50. IEEE, 2014.

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

[Ham50] Richard W Hamming. Error detecting and error correcting codes. *The Bell system technical journal*, 29(2):147–160, 1950.

[Jus72] Jørn Justesen. Class of constructive asymptotically good algebraic codes. *IEEE Transactions on Information Theory*, 18(5):652–656, 1972.

[KPT97] Joe Kilian, Erez Petrank, and Gábor Tardos. Probabilistically checkable proofs with zero knowledge. In *Proceedings of the twenty-ninth annual ACM symposium on Theory of computing*, pages 496–505, 1997.

[Li18] Xin Li. Non-malleable extractors and non-malleable codes: Partially optimal constructions. *arXiv preprint arXiv:1804.04005*, 2018.

[Ric64] C Richard. Singleton, maximum distance q-ary codes. *IEEE Trans Inf Theory*, 10(2):116–118, 1964.

[RR19] Silas Richelson and Sourya Roy. Locally testable non-malleable codes. In *Electronic Colloquium on Computational Complexity (ECCC)*, volume 26, page 117, 2019.

[RS60] Irving S Reed and Gustave Solomon. Polynomial codes over certain finite fields. *Journal of the society for industrial and applied mathematics*, 8(2):300–304, 1960.

[TS17] Amnon Ta-Shma. Explicit, almost optimal, epsilon-balanced codes. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 238–251, 2017.

[WB86] Lloyd R Welch and Elwyn R Berlekamp. Error correction for algebraic block codes, December 30 1986. US Patent 4,633,470.

[Yek08] Sergey Yekhanin. Towards 3-query locally decodable codes of subexponential length. *Journal of the ACM (JACM)*, 55(1):1–16, 2008.

[Zuc97] David Zuckerman. Randomness-optimal oblivious sampling. *Random Structures & Algorithms*, 11(4):345–367, 1997.