UNIVERSITY OF CALIFORNIA
RIVERSIDE

A Study of Pseudorandomness and its Applications to Coding Theory

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Computer Science

by

Sourya Roy

March 2022

Dissertation Committee:

Dr. Silas Richelson, Co-Chairperson
Dr. Amey Bhangale, Co-Chairperson
Dr. Marek Chrobak
Dr. Kevin Costello
Dr. Samet Oymak

The Dissertation of Sourya Roy is approved:

Committee Co-Chairperson

Committee Co-Chairperson

University of California, Riverside

Acknowledgments

I would like to extend my deepest gratitude to my advisors: Silas and Amey, for their amazing support and guidance. Doing my thesis under their mentorship has been a remarkable learning experience that I truly enjoyed. Honestly, I consider myself very fortunate to be advised by them. I would also like to sincerely thank my other committee members : Marek Chrobak, Kevin Costello and Samet Oymak. They all encouraged me and offered valuable advice at different stages of my PhD.

Because of Neal Young's graduate algorithms course, I decided to do research in theoretical computer science in the first place. I was deeply inspired by his style of teaching. Additionally, Neal has always been a great source of support and he was available whenever I needed to discuss anything. Thank you Neal, for everything.

I'm also extremely grateful to Prahladh Harsha(TIFR) and Shachar Lovett(UCSD). The second chapter of this dissertation is based on a joint work with Prahladh. So Prahladh's help was crucial for completion of this dissertation. He gave many useful feedback on my other projects as well. I had also many stimulating discussions with Shachar in the last two years. He suggested many interesting research questions and gave insightful comments on write-ups. I immensely enjoyed interacting with him.

I'm deeply indebted to my friend Sujoy Paul for all his help. I am really glad that he was my roommate during most part of my PhD. I will always cherish his friendship and all the joyful moments that we shared.

Special thanks to my friends in India: Saikat, Arya, Debajyoti, Wriddhi, Soham, Satadal and others. They all have been great source of support. I would also like to thank

my friends at UCR: Dripta, Miraj, Irem, Parker and Huong. Thanks should also go to Fernando Jeronimo(Princeton) and Tushant Mittal(UChicago) for useful discussions.

Without office-work related guidance from our graduate supervisor Vanda Yamaguchi, everything would have been lot more difficult. Many thanks to her for all the help.

Finally, I would like to thank my family for their constant support and love. Without them, none of this would have been possible.

Chapter one and Chapter three of this dissertation is based on joint works with Silas Richelson. Chapter two contains a reprint of material as it appears in Proceedings of 13th Innovations in Theoretical Computer Science (ITCS) 2022. Amey Bhangale, Prahladh Harsha, Sourya Roy. *Mixing of 3-term progressions in Quasirandom Groups.*

To my family and friends.

ABSTRACT OF THE DISSERTATION

A Study of Pseudorandomness and its Applications to Coding Theory

by

Sourya Roy

Doctor of Philosophy, Graduate Program in Computer Science
University of California, Riverside, March 2022
Dr. Silas Richelson, Co-Chairperson
Dr. Amey Bhangale, Co-Chairperson

Pseudo-randomness is an indispensable tool in theoretical computer science. In this dissertation, we aim to study several questions related to pseudo-randomness and its applications in designing codes. First, we give an alternate proof of Ta-Shma's breakthrough result on near-optimal binary error correcting code construction. While Ta-Shma's original analysis was entirely linear algebraic, our approach is more combinatorial in nature. Additionally, using our techniques, we give an alternate proof of the expander hitting set lemma. In our second work, we show the mixing of three term arithmetic progressions in quasi-random groups and fully resolve a question by Gowers. Our proof is elementary and uses only basic non-abelian Fourier analysis. Finally, we propose a generalization of locally testable codes that are resilient against adversarial channels in a certain information theoretic sense. We call these codes 'locally testable, non-malleable' and give a construction of such objects. Our construction heavily uses properties of certain pseudo-random objects called sampler graphs and tools from low degree testing literature. This establishs a connection between cryptographic non-malleability and polynomial codes.

Contents

List of Figures

Chapter 1

Introduction

Pseudorandomness is a recurring theme in theoretical computer science. Informally, we call a combinatorial object or phenomena, pseudorandom if they resembles purely random objects even though they emerges from processes that are either completely deterministic or use far less randomness. Pseudorandomness has numerous applications in almost all areas in computer science theory including approximation algorithms, circuit complexity, cryptography etc. This dissertation contains three works in pseudorandomness and its applications in coding theory. Though on a surface level these works may seem a bit disjoint from each other, but at their core two closely related tools or objects from pseudorandomness play the key role. The first one is the notion of mixing over graphs. One popular example from theoretical computer science is *Expander mixing lemma*(EML) that says: if A is an expander graph and $f_1, f_2 : A \to \mathbb{R}$, then $\mathbb{E}_{(x,y)\sim A}[f_1(x)f_2(y)] \approx \mathbb{E}_{x,y\sim A}[f_1(x)f_2[y]$ where $(x, y) \sim A$ denotes sampling random edge (x, y) from graph A and $x, y \sim A$ denotes independently sampling x and y uniformly from the vertex set of A. The first two chapters

in this dissertation, revolve around these types of mixing properties. The other notion from the pseudorandomness literature that we will use heavily in our work is the idea of *sampler* graphs. Very informally, for a sampler graph A for all most every $a \in A$ we have $\mathbb{E}_{a' \sim N(a)}[f(a')] \approx \mathbb{E}_{a' \sim A}[f(a')]$ for $f : A \to \mathbb{R}$ where $N(a)$ is the neighborhood of the vertex $a \in A$. Note that this guarantee given by sampler graphs is even stronger than the mixing property we discussed before. We crucially use properties of such sampler graphs in chapter three proofs. Below we briefly describe the three chapters in this dissertation.

In chapter one, we focus on deterministic construction of good binary error correcting codes. Specifically, we revisit the state of the art binary code construction by Ta-Shma[60] and give a new combinatorial proof of the construction. The main tool we use for this is the *expander mixing lemma*. We demonstrate that repeated application of EML suffices to prove the result which contrasts the original proof that relies heavily on intricate use of elementary linear algebra. In chapter two, We continue to explore similar mixing phenomenon as in EML though along a different direction: arithmetic progressions in non-abelian groups. Informally, we study the following question: is there any finite group G such that $\left|\mathbb{E}_{x,y \sim G}[f_1(x)f_2(xy)f_3(xy^2)] - \mathbb{E}_{x,y,z \sim G}[f_1(x)f_2(y)f_3(z)]\right|$ is negligible for all bounded functions $f_1, f_2, f_3 : G \to \mathbb{C}$. . This can be interpreted as mixing over 3-uniform hyper-graphs where the vertex set is the finite group G and hyper-edges are of the form (x, xy, xy^2) for $x, y \in G$. We show that such mixing indeed happens when the underlying finite group is quasirandom(informally, these are highly non-abelian groups). Such mixing behavior was conjectured by Gowers[43]. Previously, Tao[61] and Peluse[53] proved the conjecture for restricted classes of quasirandom groups using tools from algebraic geome-

tries and representation theory respectively. We settle the conjecture in its full generality. Moreover, our proof is completely elementary and short.

Finally, in the third chapter we describe our work on locally testable codes(LTC). Locally testable codes are error correcting codes with fast testing algorithms which can distinguish between codewords and strings that are far from all codewords. LTCs[42, 40] are well studied combinatorial objects and share deep connections with probabilistically checkable proofs(PCP)[42, 40]. In this work, we study how LTCs can be made secure against 'active adversaries' that are allowed to transform a transmitted codeword to another codeword of a different message. Towards this we combine the notion of locally testable codes with Non-malleable codes(NMCs)[37] and define locally testable, non-malleable codes (LTNMC) as a generalization of LTCs. Informally, non-malleable codes gives the guarantee that after decoding a tampered codeword either the original message or something completely unrelated(in some appropriate sense) is retrieved. In the same vein, LTNMCs provide the following guarantee: : if any tampered codeword passes the local test with good probability then it is close to a valid codeword encoding the original message or an unrelated message. To motivate our definition, we generalize it further to a natural notion of non-malleability (NM) for probabilistically checkable proofs. We do this by strengthening the definition of zero-knowledge for PCPs[47].

We instantiate our definition of LTNMC by giving an explicit construction of LTNMC in the co-ordinate wise tampering model which allows adversary to tamper each co-ordinate of a codeword independently. We achieve this by first showing that a known locally testable Reed-Muller-type code is also non-malleable against co-ordinate wise tampering.

Roughly, we show that if a tampered Reed-Muller type codeword passes the test then the tampering function must be close to an affine transformation on the polynomial space of codewords. Then we compose this polynomial code with an inner non-malleable code against affine tampering to get the final LTNMC. Our proof uses makes heavy use of sampler graphs and techniques from low degree testing literature. As additional contribution, we describe a new (standard) non-malleable code against affine tampering. Our non-malleable code against affine is arguably simpler than known constructions, and achieves better parameters.

Chapter 2

Analyzing Ta-Shma's Code via the Expander Mixing Lemma

2.1 Introduction

Error correcting codes (ECCs) allow a sender to encode a message so that the receiver can recover the full message even if several codeword bits are lost or flipped during transmission. ECCs are incredibly useful, both in theory and in practice [57, 59, 25] (and many, many more). Formally, a binary code is a map $\mathcal{C} : \{0,1\}^k \to \{0,1\}^n$ which sends a message $m \in \{0,1\}^k$ to the codeword $\mathcal{C}(m) \in \{0,1\}^n$. Two important parameters of a code are the *distance* and *rate*, which are respectively measures of the code's quality and efficiency. *Rate* is the ratio k/n, the number of message bits per codeword bit while *distance* refers to the minimum fraction of coordinates (in $[n]$) on which two distinct codewords disagree. One of the holy grails in coding theory is to find the best tradeoff between the distance

and rate of a binary code. It is known that codes with optimal distance $\delta = 1/2$ must have exponentially small rate [54]. The Gilbert-Varshamov (GV) bound [39, 62] states for any $\delta \in (0, 1/2)$, there exists a code C_n with blocklength n and distance d with rate $1 - H(\delta) - o_n(1)$ where $H(\cdot)$ is Shannon's binary entropy function. Unfortunately, this is a probabilistic (or greedy) construction and we do not know of explicit binary codes matching this bound. For distances δ close to $1/2$, the GV bound states that there exists a code with distance $(1-\varepsilon)/2$ and rate $\Omega(\varepsilon^2)$. On the other hand, it is known that any code with distance $(1-\varepsilon)/2$ must have rate $\mathcal{O}\big(\varepsilon^2 \cdot \log(1/\varepsilon)\big)$ [9]. Constructing an explict code matching the GV bound even for these distance parameters is a major open problem.

A few years ago, in a breakthrough result, Ta-Shma [60] described an explicit construction which got very close: he constructed a family of codes $\{C_n\}_n$ with rate $\Omega(\varepsilon^{2+o_\varepsilon(1)})$ and distance $(1-\varepsilon)/2$. The core of his construction is an amplification procedure which increases the distance of the code using certain special types of random walks on expander graphs. Specifically, Ta-Shma encodes a message $m \in \{0,1\}^k$ as follows.

1. Use a "base code" $\mathcal{C}_0 : \{0,1\}^k \to \{0,1\}^n$ with a good (but not optimal) rate/distance tradeoff, to encode message $m \in \{0,1\}^k$ into a n-bit codeword $\mathcal{C}_0(m)$ which we will equivalently interpret as function $f : [n] \to \{0,1\}$.

2. Identify the coordinate set $[n]$ with the vertices of an expander graph A.[1]

3. Let $W \subset A^t = [n]^t$ be a *special* subset of the set of all t-length walks in A. Define $g : W \to \{0,1\}$ by $g(a_1, \ldots, a_t) = f(a_1) \oplus \cdots \oplus f(a_t)$, where $\oplus$ is the bit XOR. Output $g \in \{0,1\}^{|W|}$.

[1] We abuse notation by refering to A both as the graph and the vertex set.

The ingenious component in TaShma's construction is the choice of the subset W. As we will soon see, choosing W to be the set of all t-length walks in A does not yield an optimal distance/rate tradeoff. TaShma, instead, uses a derandomized subset of walks, resulting from taking an *s-wide replacement product walk* on A. In the ordinary replacement product, another expander B is chosen with $|B| = \deg(A)$ so that given $a \in A$, each $b \in B$ corresponds to some $a' \in N(a)$. A t-length replacement product walk in A chooses a random $a \sim A$ and a $(t-1)$-length walk $(b_1, \ldots, b_{t-1})$ in B and outputs the walk $(a_1, \ldots, a_t)$ in A where $a_1 = a$ and a_{i+1} is the b_i-th neighbor of a_i for $i = 1, \ldots, t-1$. Note the set of replacement product walks in A is a proper subset of the set of all walks. The s-wide replacement product is a parametrized version of the ordinary replacement product. We explain the s-wide replacement product in detail in Section 2.2.

2.1.1 Our Contribution

In this note, we rederive the analysis of TaShma [60] using repeated applications of the *Expander Mixing Lemma*. TaShma's original analysis, as well as subsequent developments, convey a strongly linear algebraic viewpoint. In this writeup, we take the expander mixing lemma as our starting point and proceed from there in a combinatorial fashion. Thus, we demonstrate that no linear algebra is needed for the analysis of Ta-Shma's code beyond that which is needed to prove the expander mixing lemma. We would like to be forthcoming and stress that **our analysis is completely equivalent to Ta-Shma's original analysis**. So if you are hoping to read about a new code with improved parameters, you should read something else. This paper is for those researchers who have had difficulty penetrating the intuition behind Ta-Shma's construction. We believe that this alternate

perspective will appeal to a wider audience and make it easier for the scientific community to innovate on Ta-Shma's breakthrough work.

Our proof is the same as the original proof insofar as a random walk on a graph can be modelled both as a random process and as a linear operator. The original analysis takes the linear operator view, we take the random process view. In theory, the linear operator view is convenient for quantitatively reason about random walks because it reduces the task to understanding repeated multiplication by a fixed matrix. However, when analyzing replacement product walks from the linear operator perspective, the adjacency matrices of the outer and inner expander graphs have to be combined using some kind of tensor product. The situation is worse for the $s-$wide replacement product since then one has to keep track of s different tensor product matrices and the iterated matrix product needs to alternate over these s matrices. Thus, it seems there are diminishing returns in terms of the simplicity afforded by the linear operator perspective when the set of all random walks is to be derandomized. By using the random process view, we are able to express the same ideas in a much simpler way. This, in turn, makes it easier to see what is going on in certain key steps of the argument.

2.1.2 Techniques: Expander Mixing Lemma and consequences

Notation. We refer to graphs by their vertex sets, and use $\sim$ to indicate two vertices that are connected by an edge. So for example, if A is a graph and $a, a' \in A$ are vertices, we write $a \sim a'$ if there is an edge between a and a'. We write RW_A^t (resp. $\mathsf{RW}_A^t(a)$) for the distribution which outputs a $t-$length random walk in A (resp. a $t-$length random walk in A which begins at a). Given distributions $\mathcal{D}$ and $\mathcal{D}'$, $\mathcal{D} \equiv \mathcal{D}'$ denotes that they are same.

In order to get a sense for our technique, let us analyze the distance amplification procedure resulting from taking a random walk on an expander. Typically expander graphs are defined via the second largest eigenvalue of the adjacency matrix of the graph; in this paper we will use the following equivalent definition (similar definitions have been used in other works, *e.g.*, [33]).

Definition 1 *We say that a graph A is a $\lambda-$*expander *if for all $f, g : A \to \mathbb{R}$, the following holds:*

$$\left|\mathbb{E}_{a\sim a'}\big[f(a)\cdot g(a')\big] - \mu_f\mu_g\right| \leq \lambda\sigma_f\sigma_g,$$

where μ_f and σ_f are the expectation and standard deviation of the random variable $f(a)$ (namely, $\mu_f = \mathbb{E}_a\big[f(a)\big]$ and $\sigma_f^2 + \mu_f^2 = \mathbb{E}_a\big[f(a)^2\big]$, and similarly for μ_g and σ_g).

Now consider the distance amplification framework above instantiated with A being a constant degree, $d-$regular $\lambda-$expander, and W being the set of all $t-$length random walks in A. Note that $|W| = n \cdot d^{t-1}$, and so the rate of the resulting code is $\mathcal{O}(d^{-t})$. If A is Ramanujan (*i.e.*, an expander with the best possible relationship between λ and d) then $\lambda \approx 2/\sqrt{d}$ which makes the rate $\mathcal{O}\big((\lambda/2)^{2t}\big)$. Regarding the distance, note that for any $n-$bit string $f : [n] \to \{0,1\}$, if the fraction of non-zero coordinates is $\frac{1-\varepsilon}{2}$, then $\varepsilon = -\mathbb{E}_{v\sim[n]}\big[(-1)^{f(v)}\big]$. For this reason, we show that the amplification framework above decreases *bias*, where

$$\mathrm{Bias}(f) := \left|\mathbb{E}_{v\sim[n]}\big[(-1)^{f(v)}\big]\right|.$$

The claim below shows that when W is the set of all $t-$length walks in A, a regular Ramanujan expander graph with expansion λ, and when $\mathrm{Bias}(f) \leq \sqrt{\lambda}$, then $\mathrm{Bias}(g) \leq$

$\frac{1}{2} \cdot (4\lambda)^{t/2}$. It follows that if the distance of the amplified code is $\frac{1-\varepsilon}{2}$, then the rate is $\Omega(\varepsilon^4 \cdot 8^{-2t})$. For any constant $\alpha > 0$, it is possible to choose parameters so that $\varepsilon^\alpha \leq 8^{-2t}$, in which case the rate is $\Omega(\varepsilon^{4+\alpha})$.

Claim 1 *Let A be a regular $\lambda-$expander, $f : A \to \{0,1\}$ a function of bias $\left|\mathbb{E}_a\left[(-1)^{f(a)}\right]\right| \leq \sqrt{\lambda}$. For $k \geq 1$, define $h_k : A \to \mathbb{R}$ as*

$$h_k(a) := \mathbb{E}_{(a_1,\ldots,a_k)\sim \mathsf{RW}_A^k(a)}\left[(-1)^{f(a_1)\oplus\cdots\oplus f(a_k)}\right].$$

Let $\varepsilon_k := \left|\mathbb{E}_a\left[h_k(a)\right]\right|$ and σ_k be such that $\sigma_k^2 + \varepsilon_k^2 = \mathbb{E}_a\left[h_k(a)^2\right]$. Then for all $k \geq 1$:

$$\varepsilon_k \leq \frac{1}{2} \cdot (4\lambda)^{k/2};\ \sigma_k \leq \sqrt{\mathbb{E}_a\left[h_k(a)^2\right]} \leq (4\lambda)^{\frac{k-1}{2}}.$$

We will actually prove the following slight generalization of Claim 1, which will be more useful in our analysis later on. Note Claim 1 is recovered from Claim 2 by letting H be the constant function which always outputs 1, and noting that $\hat{\varepsilon}_1 \leq \sqrt{\lambda}$ and $\hat{\sigma}_1 \leq 1$.

Claim 2 *Let A be a regular $\lambda-$expander, $f : A \to \{0,1\}$ a function of bias $\left|\mathbb{E}_a\left[(-1)^{f(a)}\right]\right| \leq \sqrt{\lambda}$, and $H : A \to \mathbb{R}$ any function. For $k \geq 1$, let $\hat{h}_k : A \to [0,1]$ be defined by*

$$\hat{h}_k(a) = \mathbb{E}_{(a_1,\ldots,a_k)\sim \mathsf{RW}^k(a)}\left[(-1)^{f(a_1)\oplus\cdots\oplus f(a_k)} \cdot H(a_k)\right].$$

Let $\hat{\varepsilon}_k := \left|\mathbb{E}_a\left[\hat{h}_k(a)\right]\right|$ and $\hat{\sigma}_k$ such that $\hat{\sigma}_k^2 + \hat{\varepsilon}_k^2 = \mathbb{E}_a\left[\hat{h}_k(a)^2\right]$. Then for $k \geq 2$,

$$\hat{\varepsilon}_k \leq 2^{k-2} \cdot (\lambda^{\frac{k-1}{2}}\hat{\varepsilon}_1 + \lambda^{\frac{k}{2}}\hat{\sigma}_1);\ \text{and } \hat{\sigma}_k \leq \sqrt{\mathbb{E}_a\left[\hat{h}_k(a)^2\right]} \leq 2^{k-2} \cdot (\lambda^{\frac{k-2}{2}}\hat{\varepsilon}_1 + \lambda^{\frac{k-1}{2}}\hat{\sigma}_1).$$

Proof. The key observation is that for $k \geq 2$, $\hat{h}_k(a) = (-1)^{f(a)} \cdot \mathbb{E}_{a' \sim N(a)}\left[\hat{h}_{k-1}(a')\right]$. This lets us bound $\hat{\varepsilon}_k$ and $\hat{\sigma}_k$ in terms of $\hat{\varepsilon}_{k-1}$ and $\hat{\sigma}_{k-1}$ using the expander mixing lemma (Definition 1) as follows:

$\cdot\ \hat{\varepsilon}_k = \left|\mathbb{E}_a\left[\hat{h}_k(a)\right]\right| = \left|\mathbb{E}_{a \sim a'}\left[(-1)^{f(a)} \cdot \hat{h}_{k-1}(a')\right]\right| \leq \sqrt{\lambda}\hat{\varepsilon}_{k-1} + \lambda\hat{\sigma}_{k-1}$;

$\cdot\ \hat{\sigma}_k^2 \leq \hat{\sigma}_k^2 + \hat{\varepsilon}_k^2 = \mathbb{E}_a\left[\hat{h}_k(a)^2\right] = \mathbb{E}_a\left[\mathbb{E}_{a' \sim N(a)}\left[\hat{h}_{k-1}(a')\right]^2\right] = \mathbb{E}_{a' \sim_{A^2} a''}\left[\hat{h}_{k-1}(a') \cdot \hat{h}_{k-1}(a'')\right]$

$\leq \hat{\varepsilon}_{k-1}^2 + \lambda^2\hat{\sigma}_{k-1}^2$,

where $a' \sim_{A^2} a''$ indicates that (a', a'') is a uniform edge in A^2 (a λ^2-expander). We have used that the distribution which draws $a \sim A$, $a', a'' \sim N(a)$ and outputs (a', a'') is identical to the uniform edge distribution on A^2. The claim follows by induction. ∎

2.1.3 Improving the rate via s-wide replacement product walks

The rate of the above code is roughly ε^4, which is too low. In order for it to have rate $\approx \varepsilon^2$, we would have needed $\varepsilon_t \leq \lambda^t$ rather than what we got which was $\varepsilon_t \leq \lambda^{t/2}$ (actually we got something weaker, we are oversimplifying to clarify the discussion). The recursive formulas which appeared in the proof were:

$\cdot\ \varepsilon_k \leq \mathrm{Bias}(f) \cdot \varepsilon_{k-1} + \lambda\sigma_{k-1} \leq \sqrt{\lambda}\varepsilon_{k-1} + \lambda\sigma_{k-1}$ (we assumed $\mathrm{Bias}(f) \leq \sqrt{\lambda}$);

$\cdot\ \sigma_k \leq \varepsilon_{k-1} + \lambda\sigma_{k-1}$ (implied by $\sigma_k^2 \leq \varepsilon_{k-1}^2 + \lambda^2\sigma_{k-1}^2$).

The problem here is the bound $\sigma_k \leq \varepsilon_{k-1} + \lambda\sigma_{k-1}$, specifically the ε_{k-1} term on the right since we are moving from a $k-$th level term to a $(k-1)-$th level term without gaining a factor of λ. Plugging this into the first equation gives $\varepsilon_k \leq \sqrt{\lambda}\varepsilon_{k-1} + \lambda\varepsilon_{k-2} + \lambda^2\sigma_{k-2}$, where the first two terms are problematic (we are moving from level k to level $k-1$ and $k-2$

but gaining only one factor of $\sqrt{\lambda}$ and λ, respectively). The first problematic term could be fixed by choosing λ such that $\text{Bias}(f) \leq \lambda$; but the second problematic term cannot be easily fixed. This phenomenon was observed in [60] where the problem is summarized by saying "one out of every two steps works".

A natural idea for derandomizing W is to work with a set of replacement (or zig-zag) product walks. Unfortunately this yields no improvement as the "one out of every two steps works" problem persists. Ben-Aroya and Ta-Shma [16] solved this problem in a different context by using an expander graph B on a slightly larger vertex set of size d^s for $s \geq 2$, and by analyzing the resulting walk s steps at a time. This is called the s-wide replacement product. Ta-Shma was then able to successfully argue that "$s-4$ out of every s steps work". When interpreted in our language, this observation translates to a recursive formula like $\varepsilon_k \leq \lambda^{s-4} \cdot \varepsilon_{k-s}$, where we move from a $k-$th level term to a $(k-s)-$th level term, while gaining $(s-4)$ factors of λ. Gaining s factors of λ would have let us solve to the optimal $\varepsilon_k \leq \lambda^k$, obtaining rate of $\approx \varepsilon^2$; gaining $(s-4)$ factors of λ lets us solve instead to $\varepsilon_k \leq \lambda^{k(1-4/s)}$ which is almost as good when s is large.

2.2 Preliminaries

Random Walks on Graphs. Let A be the vertex set of a graph. Given $a, a' \in A$, we write $a \sim a'$ if a and a' are connected by an edge. For $a \in A$, let $N(a) \subset A$ denote the *neighborhood* of A, *i.e.*, $N(a) := \{a' \in A : a \sim a'\}$. For an integer $d \geq 1$, we say that A is

$d-$regular if $|N(a)| = d$ for all $a \in A$. For an integer $k \geq 1$, let

$$\mathsf{RW}_A^k := \{(a_1, \ldots, a_k) \in A^k : a_i \sim a_{i+1} \ \forall \ i = 1, \ldots, k-1\}$$

denote the set of $k-$length random walks in A. Similarly, for $a \in A$, $\mathsf{RW}_A^k(a)$ is the set of $k-$length random walks in A which begin at a, so $\mathsf{RW}_A^k(a) := \{(a_1, \ldots, a_k) \in \mathsf{RW}_A^k : a_1 = a\}$. We will often view RW_A^k as a distribution, where $(a_1, \ldots, a_k) \sim \mathsf{RW}_A^k$ means that $a_1 \sim A$ is drawn uniformly and then $a_{i+1} \sim N(a_i)$ is drawn for $i = 1, \ldots, k-1$.

Expander Graphs. Graph expansion is usually defined as the second largest eigenvalue of the graph's adjacency matrix,[2] *i.e.*,

$$\lambda := \max_{x, y \perp \mathbb{1}} \frac{|\langle x, My \rangle|}{|x||y|}, \tag{2.1}$$

where the max is over all nonzero $x, y \in \mathbb{R}^{|A|} - \{0\}$ which are perpendicular to the all 1s vector $\mathbb{1}$. Our Definition 1 can be recovered from (2.1) for any $f, g : A \to \mathbb{R}$ by setting $x, y \in \mathbb{R}^{|A|}$ to be $x_a = f(a) - \mu_f$ and $y_a = g(a) - \mu_g$.

Cayley Graphs. Given a finite group G and a subset $U \subseteq G$, the Cayley graph $\mathrm{Cayley}(G, U)$ has vertex set G with $g \sim g'$ iff $g^{-1}g' \in U$. Note that $\mathrm{Cayley}(G, U)$ is $|U|-$regular; additionally, if U is closed under inversion, then $\mathrm{Cayley}(G, U)$ is undirected. Cayley graphs play a key role in many explicit constructions of expander graphs. Ta-Shma's original construction used two Cayley graphs as explicit expander constructions. The first Cayley graph was

[2] The *adjacency matrix* of the graph A is $M \in \{0,1\}^{|A| \times |A|}$, where $M(a, a') = 1$ iff $a \sim a'$.

over $\mathbb{F}_2^k$, and the second was over $\mathsf{PGL}_2(\mathbb{F}_q)$, the projective general linear group over a large finite field. The use of this second Cayley graph put restrictions on some of the parameters, which required some care in order to navigate. Subsequently to Ta-Shma's original paper, new constructions of expanders based on Cayley graphs have been given. We will use a new construction, due to Alon [8], instead of the $\mathsf{PGL}_2(\mathbb{F}_q)$ construction as it will give us more flexibility.

Theorem 1 *We have the following expander constructions from [8] and [10], respectively.*

The Outer Graph: *For all integers $n, d \in \mathbb{N}$ there is an explicit construction of a $d-$regular Cayley graph with $n \cdot (1 + o_n(1))$ vertices and expansion $\lambda \leq \frac{8}{\sqrt{d}}$.*

The Inner Graph: *For all integers $r, \ell \in \mathbb{N}$ such that $\ell \leq r/2$, there exists an explicit[3] construction of an undirected $2^{2\ell}-$regular Cayley graph over $\mathbb{F}_2^r$ which is a $(r-1)2^{-\ell}-$expander.*

The Shifted Neighborhood Distribution. Let B be a Cayley graph on $\mathbb{F}_2^{ms}$, and let $d = 2^m$. For any $b = \big(b[1], \ldots, b[s]\big) \in B \cong [d]^s$, let $\mathsf{shift}(b) = \big(b[2], \ldots, b[s], b[1]\big) \in B$ be the element obtained by circularly shifting the coordinates of b. Given $b \in B$, the *shifted neighborhood distribution of* b, denoted $\tilde{N}(b)$, draws $u \sim U$ (the generator set of the Cayley graph) and outputs $\mathsf{shift}(b+u)$ (note $b+u$ is a random neighbor of b in B). It is clear that the expansion of B is not affected by using the shifted neighborhood distribution instead of

[3]This Cayley graph construction is actually *fully explicit*, in the sense that given any vertex, the $i-$th neighbor can be computed in polylogarithmic time.

the original neighborhood distribution. Indeed,

$$\left|\mathbb{E}_{\substack{b\sim B\\ b'\sim\tilde{N}(b)}}\left[f(b)\cdot g(b')\right]-\mu_f\mu_g\right| = \left|\mathbb{E}_{\substack{b\sim B\\ b'\sim N(b)}}\left[f(b)\cdot \tilde{g}(b')\right]-\mu_f\mu_{\tilde{g}}\right| \le \lambda\sigma_f\sigma_{\tilde{g}} = \lambda\sigma_f\sigma_g,$$

where $\tilde{g} = g\circ\mathsf{shift}$; clearly $(\mu_{\tilde{g}},\sigma_{\tilde{g}}) = (\mu_g,\sigma_g)$. Let $\tilde{\mathsf{RW}}_B^k$ denote the set of $k-$length shifted random walks in B. We prove the following claim about $\tilde{\mathsf{RW}}_B^k$, when k is small.

Claim 3 *For all $k \le s$, the distribution that chooses $(b_1,\ldots,b_k) \sim \tilde{\mathsf{RW}}_B^k$ and outputs the tuple $(b_1[1], b_2[1], \ldots, b_k[1]) \in [d]^k$ is identical to the uniform distribution on $[d]^k$.*

Proof. It suffices to prove the claim for $k = s$, since when $k < s$, the distribution $\tilde{\mathsf{RW}}_B^k$ is identical to the distribution which draws $(b_1,\ldots,b_s) \sim \tilde{\mathsf{RW}}_B^s$ and outputs $(b_1,\ldots,b_k)$. Note that $\tilde{\mathsf{RW}}_B^s$ draws $u_1,\ldots,u_{s-1}\sim U$, $b_1\sim B$ and outputs $(b_1,\ldots,b_s)\in B^s$, where $b_i = \mathsf{shift}(b_{i-1}+u_{i-1})$ for $i = 2,\ldots,s$. This means that for all $i = 1,\ldots,s$, $b_i[1] = b_1[i] + \sum_{j<i} u_j[i-j+1]$ (addition over $\mathbb{F}_2^m$). Uniformity of $\big(b_1[1], b_2[1], \ldots, b_t[1]\big)$ follows from the uniformity of $b_1 = \big(b_1[1],\ldots,b_1[s]\big)\sim[d]^s$. ■

2.2.1 The s-wide Replacement Product

Let A and B denote, respectively, the outer and inner graphs promised by Theorem 1. So A is a $d-$regular graph on (roughly) n vertices, while B is a Cayley graph over $\mathbb{F}_2^{ms}$, where $2^m = d$, so that vertices of B are identified with $s-$tuples of elements in $[d]$: $b = \big(b[1],\ldots,b[s]\big)\in[d]^s$. Given $a\in A$, a vertex $b\in B$ can be identified with an $s-$tuple of neighbors of a since $|N(a)| = d$. Define the *rotation map* $\phi : A\times B\to A$ via $\phi(a,b) = a'$ where a' is the $b[1]-$th neighbor of a. Since ϕ only depends on the first coordinate of b, we write $\phi(a,\hat{b})$ where $\hat{b}$ is shorthand for $b[1]$. For any $k\ge 1$, the $k-$length $s-$wide replacement

walk distribution, denoted $s\mathsf{RW}^k_{A,B}$ draws $a \sim A$ and $(b_1, \ldots, b_{k-1}) \sim \tilde{\mathsf{RW}}^{k-1}_B$, and outputs $(a_1, \ldots, a_k) \in A^k$ where $a_1 = a$ and $a_{i+1} = \phi(a_i, \hat{b}_i)$ for $i = 1, \ldots, k-1$. Since the graphs A and B will be fixed throughout this paper, we write $s\mathsf{RW}^k$ rather than $s\mathsf{RW}^k_{A,B}$. Given $a \in A$, the distribution $s\mathsf{RW}^k(a)$ outputs a sample from $s\mathsf{RW}^k$ conditioned on $a_1 = a$. Likewise, given $(a, b) \in A \times B$, $s\mathsf{RW}^k(a, b)$ outputs a sample from $s\mathsf{RW}^k$ conditioned on $(a_1, b_1) = (a, b)$. The $s-$wide replacement walk is shown in Figure 2.1.

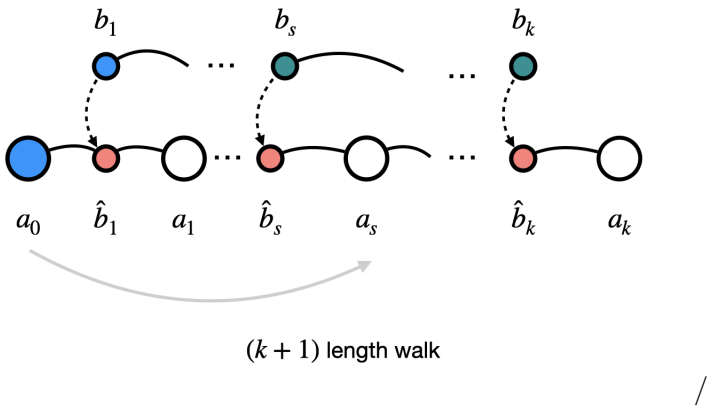

Figure 2.1: Illustration of s-wide random walk on A using a graph B.

For our graphs A and B (specifically, since A is $d-$regular and B is a Cayley graph over $\mathbb{F}_2^{ms} \cong [d]^s$) the next fact follows immediately from Claim 3.

Fact 1 (Pseudorandomness) *For all $k = 1, 2, \ldots, s, s+1$ and all $a \in A$, $s\mathsf{RW}^k(a) \equiv \mathsf{RW}^k_A(a)$.*

Following Ta-Shma's nomenclature, we will refer to the fact above as the *pseudorandomness* property. This property will play a crucial role in our proofs below as it will allow us to transform a short $s-$wide walk into a pure random walk on A, thus eliminating the

dependency on the graph B.

Local Invertibility. Since A is undirected, its edge relation is symmetric. This means that whenever $a, a' \in A$ and $b \in B$ are such that $a' = \phi(a, \hat{b})$, there must exist some $\hat{b}' \in [d]$ such that $a = \phi(a', \hat{b}')$. In this case we say that $(\hat{b}, \hat{b}')$ are inverses with respect to the $A-$edge (a, a'). Local invertibility in our context means that these inverse relations are independent of the A edges. So, specifically, for all $\hat{b}$ there exists $\hat{b}'$ such that $(\hat{b}, \hat{b}')$ are inverses with respect to all A edges. This means, for example that for all $a \in A$, if you walk to $a' = \phi(a, \hat{b})$ and then continue to $a'' = \phi(a', \hat{b}')$, then $a'' = a$. This property is easy to establish in our situation because A is a Cayley graph.

Practically speaking, what this means for us is that $s-$wide replacement walks can be "started in the middle". For standard random walks, the distribution RW_A^k which outputs $(a_1, \ldots, a_k)$ is identical to the distribution which first chooses $a_i \sim A$ randomly, and then draws $(a_i, a_{i+1}, \ldots, a_k) \sim \mathsf{RW}_A^{k-i+1}(a_i)$ and $(a_i, a_{i-1}, \ldots, a_1) \sim \mathsf{RW}_A^{i}(a_i)$, outputting $(a_1, \ldots, a_k)$. This follows from the regularity of A. Likewise, because of local invertibility, the $s-$wide replacement walk distribution $s\mathsf{RW}^k$ is identical to the following "start in the middle" version which draws $a_i \sim A$ and $b_i \sim B$, then draws $(b_i, \ldots, b_{k-1}) \sim \tilde{\mathsf{RW}}_B^{k-i}(b_i)$ and $(b_i, \ldots, b_1) \sim \tilde{\mathsf{RW}}_B^{i}(b_i)$ (in this case the shifted neighborhood distribution needs to shift the other way), then sets $a_{j+1} = \phi(a_j, \hat{b}_j)$ for $j = i, \ldots, k-1$ and $a_{j-1} = \phi(a_j, \hat{b}'_j)$ for $j = i, \ldots, 2$, where $\hat{b}'_j$ is the inverse of $\hat{b}_j$; finally $(a_1, \ldots, a_k)$ is output.

2.3 Main theorem

Theorem 2 *For every $\varepsilon > 0$ there exists an explicit linear code $\{\mathcal{C}_k\}_k$ that has distance $\geq \frac{1}{2} - \varepsilon$ and rate $= \Omega(\varepsilon^{2+o(1)})$.*

Proof. Fix $k \in \mathbb{N}$. The construction of $\mathcal{C}_k$ uses the following building blocks.

- **The Base Code:** Let $\mathcal{C}_0 : \{0,1\}^k \to \{0,1\}^{n_0}$ be an explicit code of bias ε_0 and rate R_0. We use the construction in [9], so that $R_0 = \mathcal{O}(\varepsilon_0^{-3})$.

- **The Outer Graph:** Let A be the d_A-regular Cayley graph with expansion λ_A. We use the construction of Theorem 1, so that $\lambda_A \leq 8/\sqrt{d_A}$ and $|A| = n_0 \cdot \big(1 + o_{n_0}(1)\big)$.

- **The Inner Graph:** Let B be a d_B-regular Cayley graph over $\mathbb{F}_2^r$ with expansion λ_B. We use the construction of Theorem 1 so that $\lambda_B = (r-1) \cdot 2^{-\ell}$ and $d_B = 2^{2\ell}$ for integers $\ell, r \in \mathbb{N}$ such that $\ell \leq r/2$.

The building blocks carry several parameters which we now connect. In order to set up the $s-$wide replacement product, define additional parameters $s, m \in \mathbb{N}$ such that $r = ms$, and let $d_A = 2^m$, so $B \simeq [d_A]^s$. It will be important for our analysis to have $\lambda_A \leq \lambda_B^2$; in order to arrange this, set $m = s$ and $\ell = s/5$. This gives

$$\lambda_A \leq \frac{8}{\sqrt{d_A}} = 8 \cdot 2^{-m/2} = \frac{8}{2^{\ell/2}} \cdot 2^{-2\ell} \leq (ms-1)^2 \cdot 2^{-2\ell} = \lambda_B^2,$$

where the final inequality holds whenever $s \geq 2$. We will also require $\varepsilon_0 \leq \lambda_B/2$ which we ensure by setting $\varepsilon_0 = \frac{s^2-1}{2} \cdot 2^{-s/5}$. At this point, all parameters so far have been defined in terms of s; we will specify s later. Note that our setup allows us to use B to take $s-$wide

replacement walks in A. We now describe the code. Given $x \in \{0,1\}^k$, $\mathcal{C}_k(x)$ is computed as follows.

- Compute $\mathcal{C}_0(x) \in \{0,1\}^{n_0}$, and define $f : A \to \{0,1\}$ by setting

$$f(a) = \begin{cases} \mathcal{C}_0(x)_i, & a = \iota(i) \\ 0, & \text{otherwise} \end{cases}$$

  where $\iota : [n_0] \hookrightarrow A$ is some fixed embedding.

- Define $g : s\mathsf{RW}^t \to \{0,1\}$ by setting $g(a_0, \ldots, a_t) = f(a_0) \oplus \cdots \oplus f(a_t)$. Output $g \in \{0,1\}^{s\mathsf{RW}^t}$.

The rate of $\mathcal{C}_k$ is

$$\mathsf{Rate}_k = \frac{k}{|s\mathsf{RW}^t|} \geq \frac{k}{|A|} \cdot \frac{1}{|B|} \cdot \frac{1}{d_B^{t-1}} = \Omega(\varepsilon_0^{-3}) \cdot 2^{-s^2} \cdot d_B^{-(t-1)} = \Omega\big(s^{-6} \cdot 2^{-s^2}\big) \cdot d_B^{-(t-1)}.$$

To bound the bias of $\mathcal{C}_k$, we use the following lemma which is proved in the next section.

Lemma 1 (Bias Reduction of Wide Replacement Product Walks) *Let integers $s, t \in \mathbb{N}$ and graphs A and B be as above; so in particular A and B are λ_A and λ_B expanders with $\lambda_A \leq \lambda_B^2$. Let $f : A \to \{0,1\}$ be any function such that $\big|\mathbb{E}_a\big[(-1)^{f(a)}\big]\big| \leq \lambda_B$. Then*

$$\left|\mathbb{E}_{(a_0,\ldots,a_t)\sim s\mathsf{RW}^t}\left[(-1)^{f(a_0)\oplus\cdots\oplus f(a_t)}\right]\right| \leq (2\lambda_B)^{t(1-4/s)}.$$

Note that the function $f : A \to \{0,1\}$ defined in the first step of computing $\mathcal{C}_k(x)$ satisfies

$$\left|\mathbb{E}_a\left[(-1)^{f(a)}\right]\right| \leq 2 \cdot \left|\mathbb{E}_{i\sim[n_0]}\left[(-1)^{\mathcal{C}_0(x)_i}\right]\right| \leq 2\varepsilon_0 \leq \lambda_B,$$

and so Lemma 1 ensures that $\mathsf{Bias}_k \leq (2\lambda_B)^{t(1-4/s)}$. Putting the calculations of Rank_k and Bias_k together and using $\lambda_B = (s^2-1)/\sqrt{d_B}$ gives

$$\mathsf{Rate}_k = \Omega\Big(s^{-6} \cdot (s^2-1)^{-2t} \cdot 2^{-2t-s^2+2s/5} \cdot (2\lambda_B)^{8t/s}\Big) \cdot \mathsf{Bias}_k^2 = \Omega\Big(s^{-5t} \cdot (2\lambda_B)^{8t/s}\Big) \cdot \mathsf{Bias}_k^2,$$

where the right most equality holds whenever $6\log s \leq 2s/5$ (implied by $s \geq 100$) and $t \geq s^2$. Note, therefore, that for $\eta \in (0,1/2)$, $\mathsf{Rate}_k = \Omega\big(\mathsf{Bias}_k^{2+\eta}\big)$ holds whenever $(2\lambda_B)^{t(\eta-4\eta/s-8/s)} \leq s^{-5t}$ which, if $\eta \geq 24/s$ is implied by $(2\lambda_B)^{\eta/2} \leq s^{-5}$. Finally, by plugging in $\lambda_B = (s^2-1)\cdot 2^{-s/5}$, we see that this holds whenever $\eta s \geq 60\log s$.

So finally, let us prove the theorem. Suppose that we are given $\varepsilon > 0$ and $\eta \in (0,1/2)$, and we want to construct $\mathcal{C}_k$ such that $\mathsf{Bias}_k \leq \varepsilon$ and $\mathsf{Rate}_k = \Omega\big(\mathsf{Bias}^{2+\eta}\big)$. We let $\mathcal{C}_k$ be the construction defined above with s chosen large enough so that $\eta s \geq 60\log s$; this ensures $\mathsf{Rate}_k = \Omega\big(\mathsf{Bias}_k^{2+\eta}\big)$ as noticed above. Finally, let us choose t large enough so that $t \geq s^2$ and $(2\lambda_B)^{t(1-4/s)} \leq \varepsilon$; this ensures $\mathsf{Bias}_k \leq \varepsilon$, as desired. ■

2.4 Proof of Lemma 1

In this section we prove the key bias reduction lemma that was the core of Theorem 4. Our proof will be by induction, just like Claim 2, so we will need to modify the statement of Lemma 1 so it adheres to an inductive argument.

2.4.1 Lemma Statement

Let A and B be the graphs from Section 2.3. Write λ instead of λ_B for the expansion of B and recall that $\lambda_A \leq \lambda^2$. Let $f : A \to \{0,1\}$ be a function such that $\left|\mathbb{E}_a\left[(-1)^{f(a)}\right]\right| \leq \lambda$. For any $k \geq 0$, define $g_k : A \times B \to \mathbb{R}$ by

$$g_k(a,b) = \mathbb{E}_{(a_0,\ldots,a_k)\sim s\mathsf{RW}^k(a,b)}\left[(-1)^{f(a_0)\oplus\cdots\oplus f(a_k)}\right]. \tag{2.2}$$

Let $\varepsilon_k = \left|\mathbb{E}_{a,b}\left[g_k(a,b)\right]\right|$ and let σ_k be such that $\sigma_k^2 + \varepsilon_k^2 = \mathbb{E}_{a,b}\left[g_k(a,b)^2\right]$. We prove the following.

Lemma 2 (Implies Lemma 1) *Assume the above setup. For all $k \geq 0$*

$$\varepsilon_k \leq (2\lambda)^{k(1-4/s)};\ \sigma_k \leq (2\lambda)^{(k-2)(1-4/s)}.$$

As mentioned, we prove Lemma 2 by induction. The following two claims combine to easily prove Lemma 2; we will prove them in Sections 2.4.3 and 2.4.4.

Claim 4 (Base Case.) *Assume the above setup. For all $k = 0, 1, \ldots, s$:*

$$\varepsilon_k \leq \frac{1}{2}\cdot(2\lambda)^{k+1};\ \sigma_k \leq 2\cdot(2\lambda)^{k-1}.$$

Claim 5 (Induction Step.) *Assume the above setup. For all $k > s$:*

- $\varepsilon_k \leq \frac{1}{2}(2\lambda)^s(\varepsilon_{k-s} + 3\sigma_{k-s})$;
- $\sigma_k^2 \leq \frac{1}{2}(2\lambda)^{s-2}(\varepsilon_{k-2} + \lambda\sigma_{k-1})\big(\varepsilon_{k-s} + (2+\lambda)\sigma_{k-s}\big) + \lambda^s\sigma_{k-s}\sigma_{k-1} + \lambda^2\sigma_{k-1}^2$

Proof of Lemma 2. Claim 4 clearly establishes the base cases since $\frac{1}{2}\cdot(2\lambda)^{k+1} \le (2\lambda)^{k(1-4/s)}$ and $2\cdot(2\lambda)^{k-1} \le (2\lambda)^{(k-2)(1-4/s)}$. For the first part of the induction step, we have

$$\begin{aligned}\varepsilon_k &\le \frac{1}{2}\cdot(2\lambda)^s\cdot(\varepsilon_{k-s}+3\sigma_{k-s}) \le \frac{1}{2}\cdot(2\lambda)^s\cdot\left[(2\lambda)^{(k-s)(1-4/s)}+3\cdot(2\lambda)^{(k-s-2)(1-4/s)}\right]\\ &= 8\lambda^4\cdot\left[(2\lambda)^{k(1-4/s)}+3\cdot(2\lambda)^{(k-2)(1-4/s)}\right] \le 2\lambda^2(4\lambda^2+3)\cdot(2\lambda)^{k(1-4/s)} \le (2\lambda)^{k(1-4/s)}.\end{aligned}$$

The bound $2\lambda^2(4\lambda^2+3)\le 1$ holds because $\lambda \le 1/3$. The second part of the induction step is similar:

$$\begin{aligned}\sigma_k^2 &\le \frac{1}{2}\cdot(2\lambda)^{s-2}\cdot(\varepsilon_{k-2}+\lambda\sigma_{k-1})\big(\varepsilon_{k-s}+(2+\lambda)\sigma_{k-s}\big)+\lambda^s\sigma_{k-s}\sigma_{k-1}+\lambda^2\sigma_{k-1}^2\\ &\le \frac{1}{2}\cdot(2\lambda)^2\cdot\left[(2\lambda)^{(k-2)(1-4/s)}+\lambda(2\lambda)^{(k-3)(1-4/s)}\right]\cdot\left[(2\lambda)^{k(1-4/s)}+(2+\lambda)(2\lambda)^{(k-2)(1-4/s)}\right]\\ &+\lambda^s(2\lambda)^{(k-s-2)(1-4/s)}(2\lambda)^{(k-3)(1-4/s)}+\lambda^2(2\lambda)^{2(k-3)(1-4/s)}\\ &= 2\lambda^2(2\lambda)^{(2k-2)(1-4/s)}+2\lambda^3(2\lambda)^{(2k-3)(1-4/s)}+(4\lambda^2+2\lambda^3)(2\lambda)^{(2k-4)(1-4/s)}+\\ &+(4\lambda^3+2\lambda^4)(2\lambda)^{(2k-5)(1-4/s)}+2^{4-s}\lambda^4(2\lambda)^{(2k-5)(1-4/s)}+\lambda^2(2\lambda)^{(2k-6)(1-4/s)}\\ &\le \left[2\lambda^2+2\lambda^3+(4\lambda^2+2\lambda^3)+(2\lambda^2+\lambda^3)+2^{3-s}\lambda^3+\frac{1}{4}\right]\cdot(2\lambda)^{(2k-4)(1-4/s)}\\ &\le (2\lambda)^{(2k-4)(1-4/s)},\end{aligned}$$

where the last bound has used $8\lambda^2+6\lambda^3\le 3/4$ which holds because $\lambda\le 1/4$. ∎

2.4.2 Key Intuition

In this section we zoom in on some of the key steps in the coming proofs in order to give extra explanations and intuitions.

$s-$wide Replacement Product Walks in A. Recall that a random $s-$wide replacement product walk in A (*i.e.*, a random sample from $s\mathsf{RW}^k$) is produced as follows:

1. choose base points $(a, b) \sim A \times B$;
2. generate $(b_1, \ldots, b_k) \in B^k$ as follows:
   (*i*) set $b_1 = b$;
   (*ii*) for $i \geq 2$, draw $b_i \sim N(b_{i-1})$ and set $b_i = \mathsf{shift}(b_i)$, where shift cycles the coordinates of an element of $B \simeq [d]^s$, so $\mathsf{shift}\big(b_i[1], \ldots, b_i[s]\big) = \big(b_i[2], \ldots, b_i[s], b_i[1]\big)$.
3. generate and output $(a_0, \ldots, a_k) \in A^{k+1}$ as follows:
   (*i*) set $a_0 = a$;
   (*ii*) for $i \geq 1$, set $a_i = \phi(a_{i-1}, \hat{b}_i)$ where $\hat{b}_i = b_i[1] \in [d]$ denotes the first coordinate of $b_i \in [d]^s$, and where ϕ is the rotation map of A.

Pseudorandomness. As mentioned in Section 2.2, when $k \leq s$ the distributions $s\mathsf{RW}^k$ and RW_A^{k+1} are identical. That is, a random $k-$step $s-$wide replacement product walk in A is just a random $(k+1)-$step random walk in A. The following is an example of how this concept manifests itself in the next section. Let $\varepsilon_k(a) = \mathbb{E}_b\big[g_k(a, b)\big]$.

$$\varepsilon_k(a) = \mathbb{E}_{(a_0,\ldots,a_k)\sim s\mathsf{RW}^k(a)}\Big[(-1)^{f(a_0)\oplus\cdots\oplus f(a_k)}\Big] = \mathbb{E}_{(a_0,\ldots,a_k)\sim \mathsf{RW}_A^{k+1}}\Big[(-1)^{f(a_0)\oplus\cdots\oplus f(a_k)}\Big]$$

So, $\varepsilon_k(a) = h_{k+1}(a)$ whenever $k \le s$, where h_{k+1} is the function defined and analyzed in Claim 1. This will be used crucially later.

The Ignore First Step Trick. This refers to a key step in the proof that for all $k \ge 1$,

$$\sigma_k^2 \le \mathbb{E}_a\left[\varepsilon_{k-1}(a)^2\right] + \lambda^2 \sigma_{k-1}^2. \tag{2.3}$$

This bound is useful as it reduces the task of bounding σ_k^2 to the task of bounding $\mathbb{E}_a\left[\varepsilon_{k-1}(a)^2\right]$, which will turn out to be much easier. The proof of (2.3) requires other ideas as well. Recall from the previous paragraph the definition of $\varepsilon_k(a)$; additionally let $\sigma_k(a)$ be such that $\sigma_k(a)^2 + \varepsilon_k(a)^2 = \mathbb{E}_b\left[g_k(a,b)^2\right]$.

$$\begin{aligned}
\sigma_k^2 \le \sigma_k^2 + \varepsilon_k^2 &= \mathbb{E}_{a,b}\left[g_k(a,b)^2\right] \\
&= \mathbb{E}_{a,b}\left[\mathbb{E}_{b' \sim N(b)}\left[g_{k-1}(a',b')\right]^2\right] \\
&= \mathbb{E}_{\substack{a \sim A \\ b \sim_{B^2} b'}}\left[g_{k-1}(a,b) \cdot g_{k-1}(a,b')\right] \le \mathbb{E}_a\left[\varepsilon_{k-1}(a)^2\right] + \lambda^2 \mathbb{E}_a\left[\sigma_{k-1}(a)^2\right] \\
&\le \mathbb{E}_a\left[\varepsilon_{k-1}(a)^2\right] + \lambda^2 \sigma_{k-1}^2.
\end{aligned}$$

The second line holds because $g_k(a,b) = (-1)^{f(a)} \cdot \mathbb{E}_{b' \sim N(b)}\left[g_{k-1}(a',b')\right]$, where $a' = \phi(a,\hat{b})$; the first inequality on the second line follows from the expander mixing lemma (Definition 1) on B^2 (a λ^2-expander); the final inequality has used $\mathbb{E}_a\left[\sigma_{k-1}(a)^2\right] \le \sigma_{k-1}^2$ which holds because

$$\mathbb{E}_a\left[\sigma_{k-1}(a)^2 + \varepsilon_{k-1}(a)^2\right] = \mathbb{E}_{a,b}\left[g_{k-1}(a,b)^2\right] = \sigma_{k-1}^2 + \varepsilon_{k-1}^2,$$

and $\varepsilon_{k-1}^2 \leq \mathbb{E}_a\left[\varepsilon_{k-1}(a)^2\right]$ (Jensen's inequality). The ignore first step trick is the reasoning behind the final equation on the first line. The observation is that the distribution which draws $(a, b) \sim A \times B$ and $b', b'' \sim N(b)$ and outputs (a', b', b'') where $a' = \phi(a, \hat{b})$ is identical to the distribution which draws $a' \sim A$ and a random edge $b' \sim_{B^2} b''$ in B^2 and outputs (a', b', b''). See Figure 2.2 for intuition.

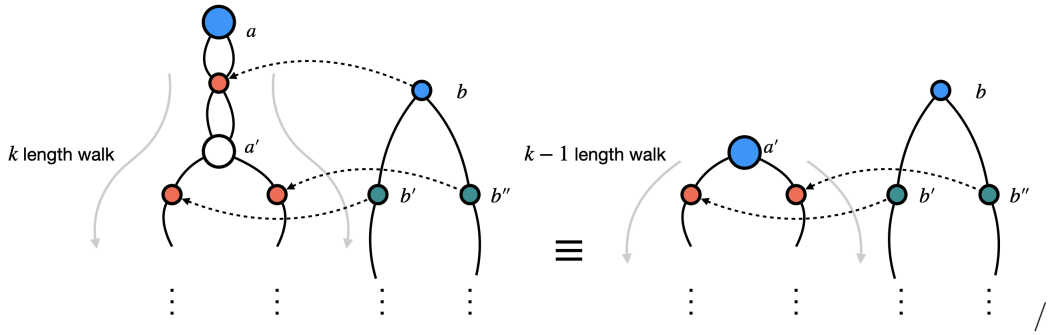

Figure 2.2: "Ignore first step" trick.

Starting the Replacement Walk in the Middle. A useful feature of random walks on an undirected $d-$regular graph is that the steps can be generated out of order. Specifically, the vertices in a $k-$step random walk can be generated by choosing $a_i \sim A$ first for any $i \in [k]$ and then drawing two walks $(a_i, a_{i+1}, \ldots, a_k) \sim \mathsf{RW}_A^{k-i+1}(a_i)$, $(a_i, a_{i-1}, \ldots, a_1) \sim \mathsf{RW}_A^{i}(a_i)$ and outputting $(a_1, \ldots, a_k)$. Replacement product walks also have this feature, though correctly formulating it requires precision. We will use that the following distribution is identical to $s\mathsf{RW}^k$ for any $i \in \{0, 1 \ldots, k-1\}$:

1. $a_i \sim A$ and a random edge $b_i \sim b_{i+1}$ in B; set $b_{i+1} = \mathsf{shift}(b_{i+1})$;
2. generate $(b_1, \ldots, b_k) \in B^k$ as follows:
   (*i*) for $j \geq i+2$, draw $b_j \sim N(b_{j-1})$ and set $b_j = \mathsf{shift}(b_j)$;

(*ii*) for $j \leq i-1$, draw $b_j \sim N(b_{j+1})$ and set $b_j = \mathsf{shift}^{-1}(b_j)$;

3. generate and output $(a_0, \ldots, a_k) \in A^{k+1}$ as follows:

   (*i*) for $i \geq i+1$, set $a_i = \phi(a_{i-1}, \hat{b}_i)$ where $\hat{b}_i = b_i[1] \in [d]$ denotes the first coordinate of $b_i \in [d]^s$, and where ϕ is the rotation map of A;

   (*ii*) for $j \leq i-1$, set $a_j = \phi^{-1}(a_{j+1}, \hat{b}_j)$ where $\phi^{-1}(a, \hat{b}) = \phi(a, \hat{b}')$ where $\hat{b}'$ is the local inverse of $\hat{b}$.

An example of how this is used is the first step of the bound for ε_k when $k > s$:

$$\begin{aligned} \varepsilon_k &= \left| \mathbb{E}_{(a_0,\ldots,a_k)\sim s\mathsf{RW}^k} \left[(-1)^{f(a_s)} \cdot (-1)^{f(a_0)\oplus\cdots\oplus f(a_s)} \cdot (-1)^{f(a_s)\oplus\cdots\oplus f(a_k)} \right] \right| \\ &= \left| \mathbb{E}_{\substack{a_s\sim A \\ b_s\sim b_{s+1}}} \left[(-1)^{f(a_s)} \cdot \overleftarrow{g}_s(a_s, b_s) \cdot g_{k-s}(a_s, b_{s+1}) \right] \right|, \end{aligned}$$

where $\overleftarrow{g}_s(a, b)$ indicates that the repalcement walk is drawn in the "backwards" fashion according to Steps 2(ii) and 3(ii) above. Equivalently, $\overleftarrow{g}_s(a, b)$ is the expectation of $(-1)^{f(a_0)\oplus\cdots\oplus f(a_s)}$ over $(a_0, \ldots, a_s) \sim s\mathsf{RW}^s$ conditioned on $(a_s, b_s) = (a, b)$.

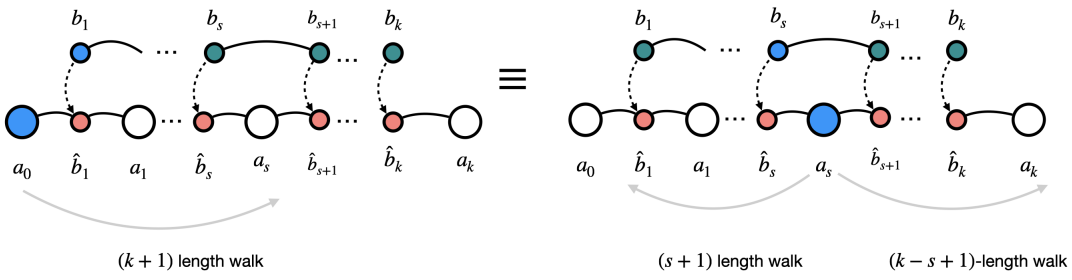

Figure 2.3: Starting the Replacement Walk in the Middle.

2.4.3 Bounding the ε_k Terms

In this section we bound the ε_k terms in Claims 4 and 5, thereby proving half of each claim. We bound the σ_k terms in the next section.

The Base Case. This follows directly from the pseudorandomness property, and the analysis already done in Section 2.1.2 (Claim 1). Specifically, when $k \leq s$, we have

$$\varepsilon_k = \left|\mathbb{E}_a\left[\varepsilon_k(a)\right]\right| = \left|\mathbb{E}_a\left[h_{k+1}(a)\right]\right| \leq \frac{1}{2} \cdot (2\lambda)^{k+1},$$

where $\varepsilon_k(a) = h_{k+1}(a)$ by pseudorandomness (h_{k+1} is the function defined in Claim 1).

The Induction Step. Fix $k > s$. We have

$$\varepsilon_k = \left|\mathbb{E}_{\substack{a\sim A\\ b\sim b'}}\left[(-1)^{f(a)} \cdot \overleftarrow{g}_s(a,b) \cdot g_{k-s}(a,b')\right]\right| \leq \left|\mathbb{E}_{a\sim A}\left[(-1)^{f(a)} \cdot \overleftarrow{\varepsilon}_s(a) \cdot \varepsilon_{k-s}(a)\right]\right| + \lambda\sigma_s\sigma_{k-s},$$

where the equality holds by starting the replacement walk in the middle, and the inequality is the expander mixing lemma (Definition 1) on B. We are using the shorthand $\overleftarrow{\varepsilon}_s(a)$ for $\mathbb{E}_b\left[\overleftarrow{g}_s(a,b)\right]$, and we have used Cauchy-Schwarz to bound the standard deviation terms, just as we did in the computation in the "ignore first step trick" paragraph in Section 2.4.2. Specifically,

$$\mathbb{E}_a\left[\overleftarrow{\sigma}_s(a) \cdot \sigma_{k-s}(a)\right] \leq \sqrt{\mathbb{E}_a[\overleftarrow{\sigma}_s(a)^2]}\sqrt{\mathbb{E}_a[\sigma_{k-s}(a)^2]} \leq \overleftarrow{\sigma}_s\sigma_{k-s} = \sigma_s\sigma_{k-s}.$$

By pseudorandomness, $(-1)^{f(a)} \cdot \overleftarrow{\varepsilon}_s(a) = (-1)^{f(a)} \cdot h_{s+1}(a) = \mathbb{E}_{a' \sim N(a)}\big[h_s(a')\big] = \mathbb{E}_{a' \sim N(a)}\big[\varepsilon_{s-1}(a')\big]$, and so we get the desired bound on ε_k via the expander mixing lemma on A, as follows:

$$\begin{aligned}\varepsilon_k &\leq \Big|\mathbb{E}_{a \sim a'}\big[\varepsilon_{s-1}(a) \cdot \varepsilon_{k-s}(a')\big]\Big| + \lambda\sigma_s\sigma_{k-s} \leq \varepsilon_{s-1}\varepsilon_{k-s} + \lambda^2\sigma_{s-1}\sigma_{k-s} + \lambda\sigma_s\sigma_{k-s} \\ &\leq \frac{1}{2}(2\lambda)^s(\varepsilon_{k-s} + 3\sigma_{k-s}).\end{aligned}$$

2.4.4 Bounding the σ_k Terms

The Base Case. We have already noted that when $1 \leq k \leq s$, $\varepsilon_{k-1}(a) = h_k(a)$ by pseudorandomness. Thus, $\mathbb{E}_a\big[\varepsilon_{k-1}(a)^2\big] = \mathbb{E}_a\big[h_k(a)^2\big] \leq (2\lambda)^{2k-2}$, by Claim 1. It follows from the first step trick that $\sigma_k^2 \leq (2\lambda)^{2k-2} + \lambda^2\sigma_{k-1}^2$, which implies $\sigma_k \leq (2\lambda)^{k-1} + \lambda\sigma_{k-1}$. Iterating this bound gives

$$\sigma_k \leq \lambda^{k-1} \cdot \big(2^{k-1} + 2^{k-2} + \cdots + 2 + 1\big) \leq 2 \cdot (2\lambda)^{k-1}.$$

The Induction Step. Fix $k > s$. As mentioned in the "ignore first step trick" paragraph in Section 2.4.2, $\sigma_k^2 \leq \mathbb{E}_a\big[\varepsilon_{k-1}(a)^2\big] + \lambda^2\sigma_{k-1}^2$ holds and so it suffices to bound $\mathbb{E}_a\big[\varepsilon_{k-1}(a)^2\big]$. By starting the replacement walk in the middle, we get

$$\mathbb{E}_a\big[\varepsilon_{k-1}(a)^2\big] = \mathbb{E}_{\substack{a_{s-1} \sim A \\ b_{s-1} \sim b_s}}\Big[(-1)^{f(a_{s-1})} \cdot g_{k-s}(a_{s-1}, b_s) \cdot G(a_{s-1}, b_{s-1})\Big],$$

where $G : A \times B \to \mathbb{R}$ is defined by $G(a,b) := \mathbb{E}_{(a_0,\ldots,a_{s-1})}\big[(-1)^{f(a_{s-1}) \oplus \cdots \oplus f(a_0)} \cdot \varepsilon_{k-1}(a_0)\big]$, where the expectation is over $(a_0, \ldots, a_{s-1})$ drawn as follows:

· set $b_{s-1} = b$; for $1 \le i \le s-2$, draw $b_i \sim N(b_{i+1})$ and then set $b_i = \mathsf{shift}^{-1}(b_i)$;

· set $a_{s-1} = a$; for $0 \le i \le s-2$ set $a_i = \phi^{-1}(a_{i+1}, \hat{b}_{i+1})$.

The expander mixing lemma (Definition 1) on B gives

$$\mathbb{E}_a\big[\varepsilon_{k-1}(a)^2\big] \le \mathbb{E}_a\Big[(-1)^{f(a)} \cdot \varepsilon_{k-s}(a) \cdot \mu_G(a)\Big] + \lambda\sigma_{k-s}\sigma_G,$$

where $\mu_G := \mathbb{E}_{a,b}\big[G(a,b)\big]$, $\mu_G(a) := \mathbb{E}_b\big[G(a,b)\big]$ and σ_G is such that $\sigma_G^2 + \mu_G^2 = \mathbb{E}_{a,b}\big[G(a,b)^2\big]$. By pseudorandomness, $\mu_G(a) = \mathbb{E}_{(a_0,\ldots,a_{s-1})\sim \mathsf{RW}_A^s(a)}\big[(-1)^{f(a_0)\oplus\cdots\oplus f(a_{s-1})} \cdot \varepsilon_{k-1}(a_{s-1})\big] = \hat{h}_s(a)$, where $\hat{h}_s : A \to \mathbb{R}$ is given by $\hat{h}_s(a) = \mathbb{E}_{(a_1,\ldots,a_s)\sim \mathsf{RW}_A^s}\big[(-1)^{f(a_1)\oplus\cdots\oplus f(a_s)} \cdot \varepsilon_{k-1}(a_s)\big]$. Note this is the function defined in Claim 2, instantiated with $H(a) = \varepsilon_{k-1}(a)$. We have $(-1)^{f(a)} \cdot \mu_G(a) = \mathbb{E}_{a'\sim N(a)}\big[\hat{h}_{s-1}(a')\big]$, and so by the expander mixing lemma on A and Claim 2 we have

$$\begin{aligned}
\mathbb{E}_a\big[\varepsilon_{k-1}(a)^2\big] &\le \mathbb{E}_{a\sim a'}\big[\varepsilon_{k-s}(a) \cdot \hat{h}_{s-1}(a')\big] + \lambda\sigma_{k-s}\sigma_G \\
&\le \varepsilon_{k-s} \cdot 2^{s-3}(\lambda^{s-2} \cdot \hat{\varepsilon}_1 + \lambda^{s-1}\hat{\sigma}_1) + \lambda^2\sigma_{k-s} \cdot 2^{s-3}(\lambda^{s-3}\hat{\varepsilon}_1 + \lambda^{s-2}\hat{\sigma}_1) + \lambda\sigma_{k-s}\sigma_G,
\end{aligned}$$

where $\hat{\varepsilon}_1$ and $\hat{\sigma}_1$ are the notations from Claim 2. In our case, $\hat{\varepsilon}_1 = \mathbb{E}_a\big[(-1)^{f(a)} \cdot \varepsilon_{k-1}(a)\big] = \varepsilon_{k-2}$, and $\hat{\sigma}_1 = \sqrt{\mathbb{E}_a[\varepsilon_{k-1}(a)^2] - \hat{\varepsilon}_1^2} \le \sqrt{\mathbb{E}_{a,b}[g_{k-1}(a,b)^2] - \hat{\varepsilon}_1^2} = \sqrt{\sigma_{k-1}^2 + \varepsilon_{k-1}^2 - \varepsilon_{k-2}^2} \le \sigma_{k-1}$. We have used Jensen's inequality and that $\varepsilon_{k-2} \ge \varepsilon_{k-1}$. Using these values and remembering the bound $\sigma_k^2 \le \mathbb{E}_a\big[\varepsilon_{k-1}(a)^2\big] + \lambda^2\sigma_{k-1}^2$ gives

$$\sigma_k^2 \le \frac{1}{2}(2\lambda)^{s-2}(\varepsilon_{k-2} + \lambda\sigma_{k-1})(\varepsilon_{k-s} + \lambda\sigma_{k-s}) + \lambda\sigma_{k-s}\sigma_G + \lambda^2\sigma_{k-1}^2. \tag{2.4}$$

This is almost the required bound except we still need to simplify σ_G. For this purpose, let us add a parameter to our notation for G, writing G_{s-1} instead of G, since it is an expectation over a length $(s-1)$ "backwards" replacement walk. For $r \le s-1$, let $\mu_r := \mathbb{E}_{a,b}\big[G_r(a,b)\big]$, let $\mu_r(a) := \mathbb{E}_b\big[G_r(a,b)\big]$ and τ_r such that $\tau_r^2 + \mu_r^2 = \mathbb{E}_{a,b}\big[G_r(a,b)^2\big]$. We need to bound$\tau_{s-1}$. By the ignore first step trick and expander mixing lemma on B^2,

$$\tau_{s-1}^2 \le \mathbb{E}_{a,b}\big[G_{s-1}(a,b)^2\big] = \mathbb{E}_{\substack{a\sim A \\ b\sim_{B^2} b'}}\Big[G_{s-2}(a,b)\cdot G_{s-2}(a,b')\Big] \le \mathbb{E}_a\big[\mu_{s-2}(a)^2\big] + \lambda^2\tau_{s-2}^2.$$

We have already seen that $\mu_{s-2}(a) = \hat{h}_{s-1}(a)$, and so by Claim 2 and our computation of $\hat{\varepsilon}_1$ and $\hat{\sigma}_1$ above, $\tau_{s-1}^2 \le (2\lambda)^{2s-6}(\varepsilon_{k-2}+\lambda\sigma_{k-1})^2 + \lambda^2\tau_{s-2}^2$, which implies $\tau_{s-1} \le (2\lambda)^{s-3}(\varepsilon_{k-2}+\lambda\sigma_{k-1}) + \lambda\tau_{s-2}$. Iterating this bound (and using $\tau_0 \le \sigma_{k-1}$) gives

$$\tau_{s-1} \le \lambda^{s-3}(\varepsilon_{k-2}+\lambda\sigma_{k-1})(2^{s-3}+2^{s-4}+\cdots)+\lambda^{s-1}\tau_0 \le 2\cdot(2\lambda)^{s-3}(\varepsilon_{k-2}+\lambda\sigma_{k-1})+\lambda^{s-1}\sigma_{k-1}.$$

Plugging this into (2.4) gives the desired bound:

$$\sigma_k^2 \le \frac{1}{2}(2\lambda)^{s-2}(\varepsilon_{k-2}+\lambda\sigma_{k-1})\big(\varepsilon_{k-s}+(2+\lambda)\sigma_{k-s}\big) + \lambda^s\sigma_{k-s}\sigma_{k-1} + \lambda^2\sigma_{k-1}^2.$$

2.5 Expander Hitting Set Lemma

Just for fun, we include a new proof of the classical expander hitting set lemma.

Lemma 3 *Let A be a $\lambda-$expander, and let $S \subset A$ be a set of size $|S| = \rho|A|$. Then for all*

$t \geq 1$,

$$\Pr_{(a_1,\ldots,a_t)\sim \mathsf{RW}^t}\Big[a_i \in S \;\forall\; i = 1,\ldots,t\Big] \leq \rho \cdot \big(\rho + \lambda(1-\rho)\big)^{t-1}.$$

Proof. Let $\mathbb{1}_S : A \to \{0,1\}$ be the indicator function of S. For $k \geq 1$, define $g_k : A \to \mathbb{R}$ by

$$g_k(a) = \Pr_{(a_1,\ldots,a_k)\sim \mathsf{RW}^k(a)}\Big[a_i \in S \;\forall\; i = 1,\ldots,k\Big].$$

Let $\varepsilon_k := \mathbb{E}_a\big[g_k(a)\big]$ and σ_k be so $\sigma_k^2 + \varepsilon_k^2 = \mathbb{E}_a\big[g_k(a)^2\big]$. Our proof is by induction on t; it is clear that the lemma holds in the base case. For $k \geq 2$, note that $g_k(a) = \mathbb{1}_S(a) \cdot \mathbb{E}_{a'\sim N(a)}\big[g_{k-1}(a')\big]$ holds, and so

$$\sigma_k^2 + \varepsilon_k^2 = \mathbb{E}_a\big[g_k(a)^2\big] = \mathbb{E}_{\substack{a\sim A \\ a',a''\sim N(a)}}\Big[\mathbb{1}_S(a) \cdot g_{k-1}(a') \cdot g_{k-1}(a'')\Big] = \varepsilon_{2k-1}.$$

We have used that $\mathbb{1}_S(a)^2 = \mathbb{1}_S(a)$ holds for all $a \in A$, and that choosing $a \sim A$ and then two $(k-1)$ length walks starting at a is identical to simply choosing a random walk of length $(2k-1)$. Now, fix $t \geq 2$ and $k, \ell \geq 1$ such that $t = k + \ell$. We have

$$\begin{aligned}
\varepsilon_t &= \mathbb{E}_{(a_1,\ldots,a_t)\sim \mathsf{RW}^t}\Big[\mathbb{1}_S(a_1)\cdots\mathbb{1}_S(a_t)\Big] = \mathbb{E}_{a\sim a'}\big[g_k(a)\cdot g_\ell(a')\big] \leq \varepsilon_k\varepsilon_\ell + \lambda\sigma_k\sigma_\ell \\
&\leq \sqrt{\varepsilon_k^2 + \lambda\sigma_k^2}\cdot\sqrt{\varepsilon_\ell^2 + \lambda\sigma_\ell^2} = \sqrt{(1-\lambda)\varepsilon_k^2 + \lambda\varepsilon_{2k-1}}\cdot\sqrt{(1-\lambda)\varepsilon_\ell^2 + \lambda\varepsilon_{2\ell-1}}
\end{aligned}$$

where the last inequality on the first line is the expander mixing lemma on A and the first inequality on the second line is Cauchy-Schwarz. Note that if $2k-1 < t$ then we can use

induction to bound the terms on the right hand side:

$$(1-\lambda)\varepsilon_k^2 + \lambda\varepsilon_{2k-1} \leq \rho \cdot \big(\rho + \lambda(1-\rho)\big)^{2k-2} \cdot \big[(1-\lambda)\rho + \lambda\big] = \rho \cdot \big(\rho + \lambda(1-\rho)\big)^{2k-1}.$$

Therefore, if t is even, we can set $k = \ell = t/2$ to obtain $\varepsilon_t \leq \rho \cdot \big(\rho + \lambda(1-\rho)\big)^{t-1}$, as desired. This does not fully work if t is odd since if we set $k = \lceil t/2 \rceil$ and $\ell = \lfloor t/2 \rfloor$, then $2k - 1 = t$ and so we cannot use induction to bound ε_{2k-1}. However, we can bound ε_k, ε_ℓ, $\varepsilon_{2\ell-1}$ by induction; this gives

$$\varepsilon_t^2 \leq \Big((1-\lambda)\rho^2\big(\rho + \lambda(1-\rho)\big)^{2k-2} + \lambda\varepsilon_t\Big) \cdot \Big(\rho\big(\rho + \lambda(1-\rho)\big)^{2\ell-1}\Big) = 2A \cdot \varepsilon_t + B,$$

where $A = \frac{\lambda\rho}{2} \cdot \big(\rho + \lambda(1-\rho)\big)^{t-2}$ and $B = (1-\lambda)\rho^3\big(\rho + \lambda(1-\rho)\big)^{2t-3}$. Collecting the terms in this way allows us to proceed by completing the square. We get $\varepsilon_t \leq A + \sqrt{A^2 + B}$ and we complete the proof by showing that $A + \sqrt{A^2 + B} = \rho\big(\rho + \lambda(1-\rho)\big)^{t-1}$. For this last calculation, set the shorthand $\Phi := \rho + \lambda(1-\rho)$. We have

$$A + \sqrt{A^2 + B} = \rho \cdot \Phi^{t-2} \cdot \left[\frac{\lambda}{2} + \sqrt{\frac{\lambda^2}{4} + \rho(1-\lambda)\Phi}\right] = \rho \cdot \Phi^{t-1},$$

where the final equation holds because $\Phi = \lambda/2 + \sqrt{\lambda^2/4 + \rho(1-\lambda)\Phi}$, which is verified by a simple calculation. ■

Chapter 3

Mixing of 3-term progressions in Quasirandom Groups

3.1 Introduction

In this work, we revisit a conjecture by Gowers [43] about mixing of three term progressions in quasirandom finite groups. Gowers initiated the study of quasirandom groups while refuting a conjecture of Babai and Sós [14] regarding the size of the largest product-free set in a given finite group. A finite group is said to be D-quasirandom for a positive integer D if all its non-trivial irreducible representations are at least D-dimensional. The quasirandomness property of groups can be used to show that certain "objects" related to the group "mix" well. For instance, the quasirandomness of the group $\mathrm{PSL}_2(\mathbb{F}_q)$ can be used to give an alternate (and weaker) proof [29] that the Ramanujan graphs of Lubotzky, Philips and Sarnak [51] are expanders.

Gowers proved that for any D-quasirandom group G and any three subsets $A, B, C \subset G$ satisfying $|A|\cdot|B|\cdot|C| \geq |G|^3/D$, there exist $x \in A, y \in B, z \in C$ such that $x \cdot y = z$. More generally, he proved that the number of such triples $(x, y, z) \in A \times B \times C$ such that $x \cdot y = z$ is at least $(1-\eta)|A| \cdot |B| \cdot |C|/|G|$ provided $|A| \cdot |B| \cdot |C| \geq |G|^3/\eta^2 D$. In other words the set of triples of the form (x, y, xy) mix well in a quasirandom group. Gowers' proof of this result was the inspiration and the first step towards the recent optimal inapproximability result for satisfiable kLIN over non-Abelian groups [18]. After proving the well-mixing of triples of the form (x, y, xy) in quasirandom groups, Gowers conjectured a similar statement for triples of the form (x, xy, xy^2). More precisely, he conjectured the following statement: Let G be a D-quasirandom group and $f_1, f_2, f_3 : G \to \mathbb{C}$ such that $\|f_i\|_\infty \leq 1$, then

$$\left|\mathbb{E}_{x,y\sim G}\big[f_1(x)f_2(xy)f_3(xy^2)\big] - \prod_{i=1,2,3}\mathbb{E}_{x\sim G}\big[f_i(x)\big]\right| = o_D(1)\ , \tag{3.1}$$

where the expression $o_D(1)$ goes to zero as D increases.

When D is small, one hope to bound the left-hand side expression above by any meaningful quantity. Consider G to be the Abelian group $\mathbb{Z}/n\mathbb{Z}$ which is 1-quasirandom and set $f_i = \mathbf{1}_B$ for all $i \in [3]$ where $B = \{1, \ldots, \lfloor \delta n \rfloor\}$ for any $\delta \in (0, 1/3)$. It is easy to observe that the first term in the left-hand side of (3.1) is $\Omega(\delta^2)$ while the second term is δ^3. A more interesting example is when the group is S_n. In this case, let $f_i = \mathbf{1}_{B_i}$, where $B_1 = A_n, B_2 = S_n$and $B_3 = S_N \setminus A_n$. Now, the $f_i's$ have density $1/2, 1, 1/2$ respectively. Note that there in no 3-term progression in (B_1, B_2, B_3) and therefore the first term in the left-hand side of (3.1) is 0. Although S_n is a non-Abelian group, it does have a non-trivial representation of dimension 1. Thus the conjecture essentially asks if the group is very "non-

Abelian" (more precisely, is D-quasirandom for large D), then do these counterexamples go away.The conjecture can be naturally extended to k-term progressions and product of k functions for $k > 3$. However, in this note we will focus on the three term case.

For the specific case of 3-term progressions, Tao [61] proved the conjecture for the group $\mathrm{SL}_d(\mathbb{F}_q)$ for bounded d using algebraic geometric machinery. In particular, he proved that the left-hand side expression in (3.1) can be bounded by $O(1/q^{1/8})$ when $d = 2$ and $O_d(1/q^{1/4})$ for larger d. Tao's approach relied on algebraic geometry and was not amenable to other quasirandom groups. Later, Peluse [53] proved the conjecture for all non-Abelian finite simple groups. She used basic facts from non-Abelian Fourier analysis to prove that the left-hand side expression in (3.1) can be bounded by $\sum_{1\neq\rho\in\hat{G}} 1/d_\rho$ where $\hat{G}$ represents the set of irreducible unitary representation of G and d_ρ the dimension of the irreducible representation ρ. This latter quantity is the *Witten zeta function* ζ_G of the group G minus one and can be bounded for *simple* finite quasirandom groups using a result due to Liebeck and Shalev [50, 49].

In this paper, we show that a slight variation of Peluse's argument can be used to prove the conjecture for *all quasirandom groups* with *better* error parameters. More surprisingly, the proof stays completely elementary and short. Specifically, we prove the following statement:

Theorem 4 *Let G be a D-quasirandom finite group, i.e, its all non-trivial irreducible representations are at least D-dimensional. Let $f_1, f_2, f_3 : G \to \mathbb{C}$ such that $\|f_i\|_\infty \leq 1$ then*

$$\left|\mathbb{E}_{x,y\sim G}\left[f_1(x)f_2(xy)f_3(xy^2)\right] - \prod_{i=1,2,3}\mathbb{E}_{x\sim G}\left[f_i(x)\right]\right| \leq \left(\frac{2}{\sqrt{D}}\right)^{\frac{1}{4}}.$$

The only tools that we use the prove the theorem are: the Cauchy-Schwarz inequality and basic non-abelian Fourier analysis.

3.2 Preliminaries

We begin by recalling some basic representation theory and non-Abelian Fourier analysis. See the monograph by Diaconis [30, Chapter 2] for a more detailed treatment (with proofs).

We will be working with a finite group G and complex-valued functions $f\colon G \to \mathbb{C}$ on G. All expectations will be with respect to the uniform distribution on G. The *convolution* between two function $f, h\colon G \to \mathbb{C}$, denoted by $f * h$, is defined as follows:

$$(f * h)(x) := \mathbb{E}_y[f(xy^{-1})h(y)].$$

For any $p \geq 1$, the p-norm of any function $f\colon G \to \mathbb{C}$ is defined as

$$\|f\|_p^p := \mathbb{E}_x[|f(x)|^p].$$

For any element $g \in G$, the *conjugacy class of* g, denoted by $C(g)$, refers to the set $\{x^{-1}gx | x \in G\}$. Observe that the conjugacy classes form a partition of the group G. A function $f\colon G \to \mathbb{C}$ is said to be a *class function* if it is constant on conjugacy classes.

For any $b \in G$ we use $\Delta_b f(x) := f(x) \cdot f(xb)$. For any set $S \subset G$, $\mu_S\colon G \to \mathbb{R}$ denotes the scaled density function $\frac{|G|}{|S|}\mathbb{1}_S$. The scaling ensures that $\mathbb{E}_x[\mu_S(x)] = 1$.

Given a complex vector space V, we denote the vector space of linear operators on

V by $\mathrm{End}(V)$. This space is endowed with the following inner product and norm (usually referred to as the *Hilbert-Schmidt* norm):

$$\text{For } A, B \in \mathrm{End}(V), \quad \langle A, B\rangle_{\mathrm{HS}} := \mathrm{Trace}(A^*B) \quad \text{and} \quad \|A\|_{\mathrm{HS}}^2 := \langle A, A\rangle_{\mathrm{HS}} = \mathrm{Trace}(A^*A).$$

This norm is known to be submultiplicative (i.e, $\|AB\|_{\mathrm{HS}} \leq \|A\|_{\mathrm{HS}} \cdot \|B\|_{\mathrm{HS}}$).

Representations and Characters: A *representation* $\rho\colon G \to \mathrm{End}(V)$ is a homomorphism from G to the set of linear operators on V for some finite-dimensional vector space V over $\mathbb{C}$, i.e., for all $x, y \in G$, we have $\rho(xy) = \rho(x)\rho(y)$. The dimension of the representation ρ, denoted by d_ρ, is the dimension of the underlying $\mathbb{C}$-vector space V. The *character* of a representation ρ, denoted by $\chi_\rho\colon G \to \mathbb{C}$, is defined as $\chi_\rho(x) := \mathrm{Trace}(\rho(x))$.

The representation $1\colon G \to \mathbb{C}$ satisfying $1(x) = 1$ for all $x \in G$ is the *trivial* representation. A representation $\rho\colon G \to \mathrm{End}(V)$ is said to *reducible* if there exists a non-trivial subpsace $W \subset V$ such that for all $x \in G$, we have $\rho(x)W \subset W$. A representation is said to be *irreducible* otherwise. The set of all irreducible representations of G (upto equivalences) is denoted by $\hat{G}$.

For every representation $\rho\colon G \to \mathrm{End}(V)$, there exists an inner product $\langle \cdot, \cdot \rangle_V$ over V such that every $\rho(x)$ is unitary (i.e, $\langle \rho(x)u, \rho(x)v\rangle_V = \langle u, v\rangle_V$ for all $u, v \in V$ and $x \in G$). Hence, we might wlog. assume that all the representations we are considering are unitary.

The following are some well-known facts about representations and characters.

Proposition 1 *1. The group G is Abelian iff $d_\rho = 1$ for every irreducible representation ρ in $\hat{G}$.*

2. *For any finite group* G, $\sum_{\rho \in \hat{G}} d_\rho^2 = |G|$.

3. *[orthogonality of characters] For any* $\rho, \rho' \in \hat{G}$ *we have:* $\mathbb{E}_x\left[\chi_\rho(x)\overline{\chi_{\rho'}(x)}\right] = \mathbb{1}[\rho = \rho']$.

Definition 2 (quasirandom groups) *A non-Abelian group* G *is said to be* D-quasirandom *for some positive integer* D *if all its non-trivial irreducible representations* ρ *satisfy* $d_\rho \geq D$.

Any group G having a non-trivial Abelian subgroup is 1-quasirandom. For instance, the symmetric group S_n is 1-quasirandom, while the alternating group A_n is $\Omega(n)$-quasirandom. The special linear group $\mathrm{SL}_2(\mathbb{F}_p)$ for prime p is $(p-1)/2$-quasirandom. If G, G' are D-quasirandom, so is $G \times G'$.

Non-Abelian Fourier analysis: Given a function $f\colon G \to \mathbb{C}$ and an irreducible representation $\rho \in \hat{G}$, the Fourier transform is defined as follows:

$$\hat{f}(\rho) := \mathbb{E}_x[f(x)\rho(x)].$$

The following proposition summarizes the basic properties of Fourier transform that we will need.

Proposition 2 *For any* $f, h\colon G \to \mathbb{C}$, *we have the following*

1. *[Fourier transform of trivial representation]*

$$\hat{f}(1) = \mathbb{E}_x[f(x)].$$

2. *[Convolution]*

$$\widehat{f * h}(\rho) = \hat{f}(\rho) \cdot \hat{h}(\rho).$$

3. *[Fourier inversion formula]*

$$f(x) = \sum_{\rho \in \hat{G}} d_\rho \cdot \langle \hat{f}(\rho), \rho(x) \rangle_{\mathrm{HS}}.$$

4. *[Parseval's identity]*

$$\|f\|_2^2 = \sum_{\rho \in \hat{G}} d_\rho \cdot \|\hat{f}(\rho)\|_{\mathrm{HS}}^2.$$

5. *[Fourier transfrom of class functions] For any class function $f \colon G \to \mathbb{C}$, the Fourier transform satisfies*

$$\hat{f}(\rho) = c \cdot I_{d_\rho}$$

for some constant $c = c(f, \rho) \in \mathbb{C}$. In other words, the Fourier transform is a scaling of the Identity operator I_{d_ρ}.

The following claim (also used by Peluse [53]) observes that the scaled density function $\mu_{gC(g)}$ has a very simple Fourier transform since it is a translate of the class function $\mu_{C(g)}$

Claim 6 *For any $g \in G$ and $\rho \in \hat{G}$ we have:*

$$\hat{\mu}_{gC(g)}(\rho) = \frac{\chi_\rho(g)}{d_\rho} \cdot \rho(g)$$

where $C(g)$ refers to the conjugacy class of g. Moreover, $\|\hat{\mu}_{gC(g)}\|_{\mathrm{HS}}^2 = \frac{|\chi_\rho(g)|^2}{d_\rho}$

Proof. We begin by observing that

$$\begin{aligned}
\hat{\mu}_{gC(g)}(\rho) &= \mathbb{E}_x\left[\mu_{gC(g)}(x)\cdot\rho(x)\right] \\
&= \mathbb{E}_x\left[\mu_{gC(g)}(gx)\cdot\rho(gx)\right] \\
&= \mathbb{E}_x\left[\mu_{gC(g)}(gx)\cdot\rho(g)\cdot\rho(x)\right] \\
&= \rho(g)\cdot\mathbb{E}_x\left[\mu_{C(g)}(x)\cdot\rho(x)\right] \\
&= \rho(g)\cdot\hat{\mu}_{C(g)}(\rho).
\end{aligned}$$

On the other hand, as $\mu_{C(g)}$ is a class function, we have $\hat{\mu}_{C(g)}(\rho) = c\cdot I_{d_\rho}$ for some constant $c\in\mathbb{C}$. The constant c can be determined by taking trace on either side of $c\cdot I_{d_\rho} = \hat{\mu}_{C(g)} = \mathbb{E}_x[\mu_{C(g)}(x)\cdot\rho(x)]$ and noting that $\text{Trace}(\rho(x)) = \chi_\rho(g)$ as follows:

$$c\cdot d_\rho = \mathbb{E}_x\left[\mu_{C(g)}(x)\cdot\chi_\rho(g)\right] = \mathbb{E}_x\left[\mu_{C(g)}(x)\right]\cdot\chi_\rho(g) = \chi_\rho(g).$$

Hence, $c = \frac{\chi_\rho(g)}{d_\rho}$ and $\hat{\mu}_{gC(g)} = \frac{\chi_\rho(g)}{d_\rho}\cdot\rho(g)$. Lastly we have,

$$\begin{aligned}
\|\hat{\mu}_{gC(g)}\|_{\text{HS}}^2 &= \left\|\frac{\chi_\rho(g)}{d_\rho}\cdot\rho(g)\right\|_{\text{HS}}^2 \\
&= \frac{|\chi_\rho(g)|^2}{d_\rho^2}\cdot\text{Trace}\left(\rho(g)^*\cdot\rho(g)\right) \\
&= \frac{|\chi_\rho(g)|^2}{d_\rho^2}\cdot d_\rho \qquad \text{(By unitariness of } \rho(g)\text{)} \\
&= \frac{|\chi_\rho(g)|^2}{d_\rho}.
\end{aligned}$$

■ The key property of D-quasirandom groups that we will be using is the following inequality

due to Babai, Nikolov and Pyber, the proof of which we provide for the sake of completeness.

Lemma 5 ([13]) *If G is a D-quasirandom group and $f_1, f_2 \colon G \to \mathbb{C}$ such that either f_1 or f_2 is mean zero then*

$$\|f_1 * f_2\|_2 \leq \frac{1}{\sqrt{D}} \cdot \|f_1\|_2 \cdot \|f_2\|_2.$$

Proof.

$$\begin{aligned}
\|f_1 * f_2\|^2 &= \sum_{\rho \in \hat{G}} d_\rho \|\widehat{f_1 * f_2}(\rho)\|_{\mathrm{HS}}^2 \\
&= \sum_{\rho \in \hat{G}} d_\rho \|\hat{f}_1(\rho) \cdot \hat{f}_2(\rho)\|_{\mathrm{HS}}^2 \\
&\leq \sum_{\rho \in \hat{G}} d_\rho \|\hat{f}_1(\rho)\|_{\mathrm{HS}}^2 \cdot \|\hat{f}_2(\rho)\|_{\mathrm{HS}}^2 && \text{(By submultiplicativity of norm)} \\
&= \sum_{1 \neq \rho \in \hat{G}} d_\rho \|\hat{f}_1(\rho)\|_{\mathrm{HS}}^2 \cdot \|\hat{f}_2(\rho)\|_{\mathrm{HS}}^2 && \text{(By mean zeroness)} \\
&\leq \frac{1}{D} \cdot \sum_{1 \neq \rho \in \hat{G}} d_\rho^2 \|\hat{f}_1(\rho)\|_{\mathrm{HS}}^2 \cdot \|\hat{f}_2(\rho)\|_{\mathrm{HS}}^2 && \text{(By } D\text{-quasirandomness)} \\
&\leq \frac{1}{D} \left(\sum_{1 \neq \rho \in \hat{G}} d_\rho \|\hat{f}_1(\rho)\|_{\mathrm{HS}}^2 \right) \cdot \left(\sum_{1 \neq \rho \in \hat{G}} d_\rho \|\hat{f}_2(\rho)\|_{\mathrm{HS}}^2 \right) \\
&\leq \frac{1}{D} \cdot \|f_1\|_2^2 \cdot \|f_2\|_2^2 \,.
\end{aligned}$$

■

The following is a simple corollary of Theorem 5.

Corollary 1 *If G is D-quasirandom; $f \colon G \to \mathbb{C}$ has zero mean and $\|f\|_\infty \leq 1$ then*

$$\mathbb{E}_b \big[|\mathbb{E}_x \Delta_b f(x)| \big] \leq \frac{1}{\sqrt{D}}.$$

Proof. Let $f'(x) := f(x^{-1})$. We have,

$$\begin{aligned}
\mathbb{E}_b\big[|\mathbb{E}_x \Delta_b f(x)|\big] &= \mathbb{E}_b\big[|\mathbb{E}_x f(x) f(xb)|\big] \\
&= \mathbb{E}_b\Big[\big|\mathbb{E}_x f'(x^{-1}) f(xb)\big|\Big] \\
&= \mathbb{E}_b\Big[\big|f' * f(b)\big|\Big] \\
&\leq \mathbb{E}_b\Big[\big|f' * f(b)\big|^2\Big]^{1/2} && \text{(By Cauchy-Schwarz inequality)} \\
&= \|f' * f\|_2 \\
&\leq \frac{1}{\sqrt{D}} \cdot \|f'\|_2 \cdot \|f\|_2 && \text{(By Theorem 5)} \\
&\leq \frac{1}{\sqrt{D}}. && \text{(Since } \|f\|_2 \leq \|f\|_\infty \leq 1\text{).}
\end{aligned}$$

■

3.3 Proof of Theorem 4

The following proposition is where we deviate from Peluse's proof [53]. We give an elementary proof for *every* quasirandom group while Peluse proved the same result for *simple* finite groups using the result of Liebeck and Shalev [50, 49] to bound the Witten zeta function ζ_G for *simple* finite groups.

Proposition 3 *Let G be a D-quasirandom group. Let $f\colon G \to \mathbb{C}$ such that $\|f\|_\infty \leq 1$, $\mathbb{E}[f] = 0$ and f_b is the mean zero component of the function $\Delta_b f$ (i.e., $f_b(x) = \Delta_b f(x) - \mathbb{E}_x[\Delta_b f(x)]$). Then*

$$\mathbb{E}_{g,b}\Big[\Big|\mathbb{E}_x\big[\Delta_b f(x) \cdot (f_{g^{-1}bg} * \mu_{g^{-1}C(g^{-1})})(x)\big]\Big|\Big] \leq \frac{1}{\sqrt{D}}\,.$$

Proof. Let us denote the expression on the L.H.S. as Γ. We use simple manipulations and previously stated facts to simplify the expression.

$$\begin{aligned}
\Gamma^2 &\leq \mathbb{E}_{g,b}\left[\left(\|\Delta_b f\|_2\right)\cdot\left(\|f_{g^{-1}bg} * \mu_{g^{-1}C(g^{-1})}\|_2\right)\right]^2 && \text{(By Cauchy-Schwarz inequality)}\\
&\leq \mathbb{E}_{g,b}\left[\|f_{g^{-1}bg} * \mu_{g^{-1}C(g^{-1})}\|_2\right]^2 && \text{(Since } \|\Delta_b f\|_2 \leq 1\text{)}\\
&\leq \mathbb{E}_{g,b}\left[\|f_{g^{-1}bg} * \mu_{g^{-1}C(g^{-1})}\|_2^2\right] && \text{(By Cauchy Schwarz inequality)}\\
&= \mathbb{E}_{g,b}\left[\sum_{1\neq\rho\in\hat{G}} d_\rho \cdot \|\hat{f}_{g^{-1}bg}(\rho)\cdot\hat{\mu}_{g^{-1}C(g^{-1})}(\rho)\|_{\mathrm{HS}}^2\right]\\
& && \text{(By Parseval's identity \& } \hat{f}_{g^{-1}bg}(1) = 0\text{)}\\
&\leq \mathbb{E}_{g,b}\left[\sum_{1\neq\rho\in\hat{G}} d_\rho \cdot \|\hat{f}_{gbg^{-1}}(\rho)\|_{\mathrm{HS}}^2\cdot\|\hat{\mu}_{g^{-1}C(g^{-1})}(\rho)\|_{\mathrm{HS}}^2\right] && \text{(By submultiplicativity of norm)}\\
&= \mathbb{E}_{g,b}\left[\sum_{1\neq\rho\in\hat{G}} \|\hat{f}_{g^{-1}bg}(\rho)\|_{\mathrm{HS}}^2\cdot|\chi_\rho(g)|^2\right] && \text{(By Claim 6)}\\
&= \sum_{1\neq\rho\in\hat{G}} \mathbb{E}_g\left[|\chi_\rho(g)|^2\cdot\mathbb{E}_b\left[\left\|\hat{f}_{gbg^{-1}}(\rho)\right\|_{\mathrm{HS}}^2\right]\right].
\end{aligned}$$

Now using the fact that gbg^{-1} is uniformly distributed in G for a fixed g and a uniformly random b in G, we can simplify the above expression as follows.

$$
\begin{aligned}
\Gamma^2 &\leq \sum_{1\neq\rho\in\hat{G}} \mathbb{E}_g\Big[|\chi_\rho(g)|^2 \cdot \mathbb{E}_b\Big[\big\|\hat{f}_b(\rho)\big\|_{\mathrm{HS}}^2\Big]\Big] \\
&= \sum_{1\neq\rho\in\hat{G}} \mathbb{E}_b\Big[\big\|\hat{f}_b(\rho)\big\|_{\mathrm{HS}}^2\Big] \cdot \mathbb{E}_g\Big[|\chi_\rho(g)|^2\Big] \\
&= \sum_{1\neq\rho\in\hat{G}} \mathbb{E}_b\Big[\big\|\hat{f}_b(\rho)\big\|_{\mathrm{HS}}^2\Big] && \text{(By orthogonality of } \chi_\rho\text{)} \\
&= \mathbb{E}_b\Big[\sum_{1\neq\rho\in\hat{G}} \big\|\hat{f}_b(\rho)\big\|_{\mathrm{HS}}^2\Big].
\end{aligned}
$$

Finally, we use the fact that all the terms in the summation are non-negative and the group G is a D-quasirandom group.

$$
\begin{aligned}
\Gamma^2 &\leq \frac{1}{D} \cdot \mathbb{E}_b\Big[\sum_{1\neq\rho\in\hat{G}} d_\rho \cdot \big\|\hat{f}_b(\rho)\big\|_{\mathrm{HS}}^2\Big] \\
&= \frac{1}{D} \cdot \mathbb{E}_b\Big[\|f_b\|_2^2\Big] && \text{(By Parseval's identity)} \\
&\leq \frac{1}{D}, && \text{(Because } \|f_b\|_2^2 \leq 1\text{).}
\end{aligned}
$$

The proof of this lemma is similar to the proof of the BNP inequality (Theorem 5). The key difference being that we have a complete characterization of the Fourier transform of $\mu_{gC(g)}$ from Claim 6 which we use to give a sharper bound. ■

We are now ready to prove the main Theorem 4. This part of the proof is similar to the corresponding expression that appears in the paper of Peluse [53], which is in turn inspired by Tao's adaptation of Gowers' repeated Cauchy-Schwarzing trick to the nonebelian setting. We, however, present the entire proof for the sake of completeness.

Proof of Theorem 4. Let us denote the L.H.S. of the expression by Θ_{f_1,f_2,f_3}. Without loss of generality we assume $\mathbb{E}[f_3] = 0$. Now we have,

$$\begin{aligned}
\Theta^4_{f_1,f_2,f_3} &= \Big|\mathbb{E}_{x,y}\big[f_1(x)f_2(xy)f_3(xy^2)\big]\Big|^4 \\
&= \Big|\mathbb{E}_{x,z}\big[f_1(xz^{-1})f_2(x)f_3(xz)\big]\Big|^4 \qquad \text{(Change of variables: } x \leftarrow xy, z \leftarrow y\text{)} \\
&\le \Big|\mathbb{E}_{x,z_1,z_2}\big[f_1(xz_1^{-1})f_1(xz_2^{-1})f_3(xz_1)f_3(xz_2)\big]\Big|^2 \\
&\qquad\qquad \text{(Cauchy-Schwarz over } x\text{; } \|f_2\|_\infty = 1 \text{ and expansion)} \\
&= \Big|\mathbb{E}_{y,z,a}\big[f_1(y)f_1(ya)f_3(yz^2)f_3(yza^{-1}z)\big]\Big|^2 \\
&\qquad\qquad \text{(Change of variables: } y \leftarrow xz_1^{-1}, z \leftarrow z_1, a \leftarrow z_1z_2^{-1}\text{)} \\
&= \Big|\mathbb{E}_{y,z,a}\big[\Delta_a f_1(y) \cdot \Delta_{z^{-1}a^{-1}z}\ f_3(yz^2)\big]\Big|^2 \\
&\le \Big|\mathbb{E}_{y,a,z_1,z_2}\big[\Delta_{z_1^{-1}a^{-1}z_1}\ f_3(yz_1^2) \cdot \Delta_{z_2^{-1}a^{-1}z_2}\ f_3(yz_2^2)\big]\Big|, \\
&\qquad\qquad \text{(Cauchy-Schwarz over } y, a\text{; } \|f_1\|_\infty \le 1 \text{).}
\end{aligned}$$

Now, using the following change of variables, $z \leftarrow z_1,\ x \leftarrow yz_1^2,\ b \leftarrow z_1^{-1}a^{-1}z_1,\ g \leftarrow z_1^{-1}z_2$, we get

$$\begin{aligned}
\Theta^4_{f_1,f_2,f_3} &\le \Big|\mathbb{E}_{x,b,z,g}\big[\Delta_b\ f_3(x) \cdot \Delta_{g^{-1}bg}\ f_3(xz^{-1}gzg)\big]\Big| \\
&= \Big|\mathbb{E}_{x,b,g}\Big[\Delta_b\ f_3(x) \cdot \mathbb{E}_z[\Delta_{g^{-1}bg}\ f_3(xz^{-1}gzg)]\Big]\Big| \\
&= \Big|\mathbb{E}_{x,b,g}\Big[\Delta_b\ f_3(x) \cdot \mathbb{E}_a[\Delta_{g^{-1}bg}\ f_3(xa^{-1}) \cdot \frac{|G|}{|C(g^{-1})|}1_{g^{-1}C(g^{-1})}(a)]\Big]\Big| \\
&= \Big|\mathbb{E}_{x,b,g}\Big[\Delta_b\ f_3(x) \cdot \mathbb{E}_a[\Delta_{g^{-1}bg}\ f_3(xa^{-1}) \cdot \mu_{g^{-1}C(g^{-1})}(a)]\Big]\Big| \\
&= \Big|\mathbb{E}_{x,b,g}\Big[\Delta_b\ f_3(x) \cdot \Delta_{g^{-1}bg}\ f_3 * \mu_{g^{-1}C(g^{-1})}(x)\Big]\Big|.
\end{aligned}$$

The second equality follows because after g, x, b have been fixed we only use z to compute $z^{-1}gz$ and the map that takes $z \in G$ to $z^{-1}gz \in C(g)$ is surjective where each member in the range has preimage of size $\frac{|G|}{|C(g^{-1})|} = |\text{Centralizer}(g)|$. We now separate the function $\Delta_{g^{-1}bg}\ f_3$ from its the mean zero part as follows: Let $\Delta_{g^{-1}bg}\ f_3 = f'_{g^{-1}bg} + f_{g^{-1}bg}$ where $f'_{g^{-1}bg} = \mathbb{E}_x[\Delta_{g^{-1}bg}\ f_3(x)]$ and $f_{g^{-1}bg}(x) = \Delta_{g^{-1}bg}\ f_3(x) - f'_{g^{-1}bg}$.

$$
\begin{aligned}
\Theta^4_{f_1,f_2,f_3} &\leq \left|\mathbb{E}_{x,b,g}\Big[\Delta_b\ f_3(x) \cdot (f_{g^{-1}bg} + f'_{g^{-1}bg}) * \mu_{g^{-1}C(g^{-1})}(x)\Big]\right| \\
&\leq \mathbb{E}_{b,g}\left[\left|\mathbb{E}_x\big[\Delta_b\ f_3(x) \cdot f_{g^{-1}bg} * \mu_{g^{-1}C(g^{-1})}(x)\big]\right|\right] \\
&\qquad\qquad + \mathbb{E}_{b,g}\left[\left|\mathbb{E}_x\big[\Delta_b\ f_3(x) \cdot f'_{g^{-1}bg} * \mu_{g^{-1}C(g^{-1})}(x)\big]\right|\right] \\
&\leq \frac{1}{\sqrt{D}} + \mathbb{E}_{b,g}\Big[\big|\mathbb{E}_x\big[\Delta_b\ f_3(x)\big]\big| \cdot \|f'_{g^{-1}bg} * \mu_{g^{-1}C(g^{-1})}\|_\infty\Big] \\
&\qquad\qquad \text{(Using ?? 3 to bound the first expectation)} \\
&= \frac{1}{\sqrt{D}} + \mathbb{E}_{b,g}\Big[\big|\mathbb{E}_x\big[\Delta_b\ f_3(x)\big]\big| \cdot |f'_{g^{-1}bg}|\Big] \\
&\leq \frac{1}{\sqrt{D}} + \mathbb{E}_b\Big[\big|\mathbb{E}_x\big[\Delta_b\ f_3(x)\big]\big|\Big] \qquad \text{(Using } |f'_{g^{-1}bg}| \leq 1\text{)} \\
&\leq \frac{2}{\sqrt{D}}\,, \qquad \text{(By corollary 1 and } \|f_3\|_\infty \leq 1\text{).}
\end{aligned}
$$

■

Chapter 4

Locally Testable Non-malleable Codes

4.1 Introduction

A *coding scheme* is a pair $\big(\{\mathsf{Enc}\}_k, \{\mathsf{Dec}\}_k\big)$ of function ensembles where $\mathsf{Enc} : \Gamma^k \to \Gamma^n$ is a possibly randomized encoder, $\mathsf{Dec} : \Gamma^n \to \Gamma^k \cup \{\perp\}$ is the decoder and $\mathsf{Dec}\big(\mathsf{Enc}(m)\big) = m$ holds with probability 1 for all $m \in \Gamma^k$. We say $\mathbf{x} \in \Gamma^n$ is a *valid codeword* if $\mathbf{x} = \mathsf{Enc}(m)$ for some $m \in \Gamma^k$ (and some choice of randomness for Enc). 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. When a code's distance is bounded away from zero, one can try to design decoding-type algorithms with extra features such as error-correcting or local decoding/testing capabilities. In this paper we will focus on *locally testable codes*

(LTCs). These are codes for which there is a Test algorithm which reads a small number of symbols and decides if a given string is close to some, or far from every, valid codeword. The requirement is that Test should reject an input string with probability proportional to its distance from the nearest codeword. We study what happens in a general tampering model where the adversery applies a tampering function f to a transmitted codeword. Is it possible to make LTCs resilient to such attacks? What type of security guarantees can we hope to obtain in this scenerio? What can be said about f if codewords tampered by f pass the test with good probability?

Dziembowski, Pietrzak and Wichs asked similar questions in the realm of error-correcting codes, which led to their influential definition of *non-malleable codes* (NMCs) [37]. Since their introduction, NMCs have found numerous applications in cryptography [7, 26, 44, 2, 19], pseudorandomness [21, 24], and complexity [34]. NMCs provide security against a family $\mathcal{F}$ of tampering functions[1] by guaranteeing that if a transmitted codeword is tampered by any $f \in \mathcal{F}$, then the resulting decoded message is either the original untampered message (such is the case if f is the identity function) or else is independent of the original message. So for example, it should not be possible to tamper by $f \in \mathcal{F}$ and modify a codeword encoding m to a codeword encoding $m + 1$.

4.1.1 Our Contributions

LTCs and NMCs are both generalizations of error-correcting codes along different axes. In this work we combine the notions and define *locally testable, non-malleable codes*

[1] No security can be obtained if $\mathcal{F}$ is the set of all functions, since in this case f could decode the original codeword, tamper the underlying message arbitrarily and then re-encode.

(LTNMCs). Roughly speaking, these are LTCs which have the following non-malleability guarantee: any tampered codeword which passes the test with good probability is close to a valid codeword which either encodes the original message, or else encodes an unrelated message. The new definition appears in Section 4.2 and the remainder of the paper is devoted to discussing, motivating and instantiating the new primitive.

Motivating LTNMCs. LTCs generalize ECCs by adding extra *functionality*, while NMCs generalize ECCs by adding extra *security*. For this reason, it is an interesting challenge to try to achieve both notions simultaneously. However, we believe that NMLTCs are well motivated and we give two possible connections as evidence.

1. Non-malleable interactive proofs [35] have been studied extensively in cryptography as a strengthening of zero-knowledge (ZK) interactive proofs (IPs). ZK for probabilistically checkable proofs (PCPs) was defined and constructed in [47]and subsequently used for applications in hardness of approximation [46, 38]. To our knowledge, no definition of non-malleable PCPs has been given, but the notion certainly makes sense and could also give applications in hardness of approximation. In Section 4.3 we define a notion of non-malleability for PCPs by strengthening the existing notion of ZK, analogously to how one obtains NM for IPs by strengthening ZK for IPs. Our definition of LTNMCs is essentially the "combinatorial" analogue of NMPCPs, just as LTCs are combinatorial analogues of PCPs. The new definition of non-malleability for PCPs might be of independent interest.

2. Another motivation for the study of LTNMCs is for building standard NMCs. While NMCs need not inherently have good distance, many constructions do [37, 4, 3, 6, 5].

Moreover, the decoding algorithms in these works all contain a test subroutine and decoding only occurs if the test passes (if the test fails, the decoder outputs $\perp$). The proofs of non-malleability begin by categorizing the functions which pass the test, and then proceed to prove non-malleability against each such category of tampering functions. LTNMCs explicitly incorporate such a test, and could provide a useful abstraction for designing better NMCs in the future.

Constructing LTNMCs. We instantiate our new notion by proving that a Reed-Muller-type code is non-malleable against the family of coordinate-wise tampering function, *i.e.*, against an adversery that tampers each coordinate of the codeword independently. Our construction has three main parts.

1. We prove that when the Reed-Muller-type LTC of Raz and Safra [55] (*i.e.*, the "planes table" which is the set of restrictions of the message polynomial on every 3-dimensional affine sub-spaces of the ambient space) is tampered by a coordinate-wise tampering function then either the tampered codeword is far from a valid codeword (and so fails the local test with high probability) or else is close to a valid codeword which encodes an affine function of the original message. In NMC terminology, we show that the planes table is a (locally testable) *non-malleable reduction* from coordinate-wise tampering to affine tampering.

2. We describe an elementary construction of a (standard) NMC against the family of affine tampering functions. Such codes were previously known [4, 1, 22], but our construction is much simpler those in prior work. When the message space is large,

our construction is more efficient than the one in [4, 1] as our encoding algorithm does not require drawing large random primes. Our code achieves a better rate/error tradeoff than the construction of [22].

3. We combine the codes from the above points into a concatenated code, obtaining a LTNMC against coordinate-wise tampering via a composition theorem for Reed-Muller-type codes. The local test of our composed code works by decoding a symbol of the outer code and checking validity using the inner code. This idea has been used previously to analyze the composition of LTCs and PCPs [12].

Theorem 3 (Main Construction.) *There exists an explicit locally testable, non-malleable code against an adversery who tampers each coordinate of the codeword independently.*

In order to streamline the introduction of the new definition, we have deferred the "technical overview" where we give a high level view of the construction until Section 4.4.

4.1.2 Other Relevant Prior Work

Non-Malleable Codes. Since the introduction of non-malleable codes in 2010 [37], an immense research effort has focused on giving constructions which are secure against richer classes of tampering functions, and with better rate [36, 4, 3, 21, 48, 15] (and many, many more). Our work uses some of the machinery developed in this area. In particular, our main definition of a LTNMC is inspired by the definition of a NM reduction in [3], which both simplifies and generalizes the original definition in [37].

Sampler-Based Decoding. Our work fits into a recent line of work on sampler-based decoding [45, 52, 17, 31, 32] (and more). In these works, sampling properties of a code's index set are exploited in order to give non-trivial decoding algorithms. Our work builds on techniques developed in these papers in order to "decode" a coordinate-wise tampering function which respects codeword proximity, to a small list of affine functions.

Locally Decodable Non-Malleable Codes. A few works combine the notions of local decodability with non-malleability [27, 20, 28]. These works give constructions of non-malleable codes which admit local decode/update subroutines. Our work differs in several ways from these. First, the codes in these works achieve super-constant locality, whereas our main construction achieves constant locality. More significantly, the constructions in prior work achieve local decodeability and updateability by separately encoding each element of the message, so they do not support a local test of proximity to a valid codeword. Finally, the techniques differ significantly; our techniques are similar to those used in the LTC literature.

4.2 Defining Locally Testable Non-malleable Codes

4.2.1 Coding Theory Background

Throughout this section $k \in \mathbb{N}$ is fixed, and $(\mathsf{Enc}, \mathsf{Dec})$ is a coding scheme for messages in Γ^k. Recall this means $\mathsf{Enc} : \Gamma^k \to \Gamma^n$ is possibly randomized, $\mathsf{Dec} : \Gamma^n \to \Gamma^k \cup \{\perp\}$, and $\mathsf{Dec}\big(\mathsf{Enc}(m)\big) = m$ holds with probability 1 for all $m \in \Gamma^k$. Given $\mathbf{y} \in \Gamma^n$, we write $\mathsf{dist}(\mathbf{y})$ for the minimum of $\mathsf{dist}(\mathbf{x}, \mathbf{y}) = \Pr_{i \sim [n]}\big[\mathbf{x}_i \neq \mathbf{y}_i\big]$ over all valid codewords

$\mathbf{x}$. We write $\mathsf{agr}(\mathbf{y})$ for the maximum of $\Pr_{i\sim[n]}\big[\mathbf{x}_i = \mathbf{y}_i\big]$ over all valid codewords $\mathbf{x}$ (so $\mathsf{agr}(\mathbf{y}) + \mathsf{dist}(\mathbf{y}) = 1\ \forall\ \mathbf{y}$).

Definition 3 (LTCs) *Fix $\varepsilon > 0$ and $q \in \mathbb{N}$. We say that* $(\mathsf{Enc}, \mathsf{Dec})$*, is a* $(q, \varepsilon)-$locally testable code *if there exists a distribution $\mathcal{T}$ supported on the set of $q-$local functions from Γ^n to $\{0,1\}$ (*i.e.*, functions which read q coordinates from their input) such that the following hold:*

1. $\forall\ m \in \Gamma^k$, $\mathbb{E}_{\mathsf{Enc}(m),\mathcal{T}}\big[\mathsf{Test}(\mathsf{Enc}(m))\big] = 1$;
2. $\exists$ *a constant $c > 0$ st if $\mathbf{y} \in \Gamma^n$ has $\mathbb{E}_{\mathcal{T}}\big[\mathsf{Test}(\mathbf{y})\big] \geq \varepsilon$, then* $\mathsf{agr}(\mathbf{y}) \geq c \cdot \mathbb{E}_{\mathcal{T}}\big[\mathsf{Test}(\mathbf{y})\big]$.

Point 1, sometimes called *completeness*, says that valid codewords always pass the test. Point 2, sometimes called *soundness*, says that any $\mathbf{y} \in \Gamma^n$ which passes the test with good probability must have non-negligible agreement with some valid codeword. The next definition enhances the soundness requirement by demanding that for every $\mathbf{y} \in \Gamma^n$, almost all of $\mathbf{y}$'s test passing probability comes from its agreement with a short list of valid codewords.

Definition 4 (List-Decoding for LTCs) *Fix $\varepsilon > 0$ and $q, \ell \in \mathbb{N}$. We say that* $(\mathsf{Enc}, \mathsf{Dec})$ *is a* $(q, \varepsilon, \ell)-$list decodeable LTC *if there exists a $q-$local distribution $\mathcal{T}$ (*i.e.*, every test function in the support of $\mathcal{T}$ is $q-$local) such that the following two points hold:*

1. $\forall\ m \in \Gamma^k$, $\mathbb{E}_{\mathsf{Enc}(m),\mathcal{T}}\big[\mathsf{Test}(\mathsf{Enc}(m))\big] = 1$;
2. $\forall\ \mathbf{y} \in \Gamma^n\ \exists$ *a list $\mathsf{L}_{\mathbf{y}} = \{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(\ell)}\} \subset \Gamma^n$ of valid codewords of size $|\mathsf{L}_{\mathbf{y}}| \leq \ell$ such that*

$$\Pr_{\mathcal{T}}\Big[\mathsf{Test}(\mathbf{y}) = 1\ \&\ \mathbf{y}_I \notin \{\mathbf{x}^{(j)}_I : \mathbf{x}^{(j)} \in \mathsf{L}_{\mathbf{y}}\}\Big] \leq \varepsilon,$$

where $I \subset [n]$ with $|I| = q$ are the coordinates read by Test.

Non-malleable codes [37] (NMCs) provide meaningful security guarantees even in situations where error correction is impossible. Intuitively, $(\mathsf{Enc}, \mathsf{Dec})$ is non-malleable against a tampering family $\mathcal{F} \subset \{f : \Gamma^n \to \Gamma^n\}$ if for all $f \in \mathcal{F}$ and $m \in \Gamma^k$, the distribution

$$\mathsf{Tamper}_f(m) := \big(\mathsf{Dec} \circ f \circ \mathsf{Enc}\big)(m)$$

(randomness over Enc) is either equal to, or independent of, m. In order to define this formally let $\mathcal{G}_{\mathsf{trivial}} \subset \big\{g : \Gamma^k \to \Gamma^k \cup \{\perp\}\big\}$ consist of all constant functions as well as the identity. These are the trivial message tampering functions. Intuitively, $(\mathsf{Enc}, \mathsf{Dec})$ is non-malleable against $\mathcal{F}$ if tampering codewords by any $f \in \mathcal{F}$ tampers the underlying message by functions in $\mathcal{G}_{\mathsf{trivial}}$.

Definition 5 (Non-Malleable Codes) *Fix $\varepsilon > 0$ and let $\mathcal{F} \subset \{f : \Gamma^n \to \Gamma^n\}$ be a family of tampering functions. We say that* $(\mathsf{Enc}, \mathsf{Dec})$ *is* $\varepsilon-$non-malleable against $\mathcal{F}$ *if for all $f \in \mathcal{F}$ there exists a distribution S_f on $\mathcal{G}_{\mathsf{trivial}}$ such that for all $m \in \Gamma^k$,*

$$\Delta\big(\mathsf{Tamper}_f(m), \mathsf{S}_f(m)\big) \leq \varepsilon,$$

(Δ denotes statistical distance) where $\mathsf{S}_f(m)$ draws $g \sim \mathsf{S}_f$ and outputs $g(m) \in \Gamma^k \cup \{\perp\}$.

4.2.2 The New Definition and Discussion

In order to define LTNMCs we need to redefine the set of trivial tampering functions so that they map codewords to codewords. Let $\mathcal{G}_{\mathsf{trivial}} \subset \big\{g : \Gamma^n \to \Gamma^n\big\}$ consisting of

all constant functions as well as the identity. We will also need to redefine the tampered distribution so it makes sense in the context of LTNMCs. Given $f \in \mathcal{F}$, $\mathcal{T}$ a $q-$local distribution and $m \in \Gamma^k$, $\mathsf{Tamper}_{f,\mathcal{T}}(m)$ is the distribution on $\mathrm{Supp}(\mathcal{T}) \times \Gamma^q$ which draws $\mathbf{x} \sim \mathsf{Enc}(m)$, $\mathsf{Test} \sim \mathcal{T}$ and outputs $\big(\mathsf{Test}, f(\mathbf{x})_I\big)$ where $I \subset [n]$ with $|I| = q$ are the coordinates read by Test.

Definition 6 (LTNMCs) *Fix $\varepsilon > 0$, $\ell \in \mathbb{N}$ and let $\mathcal{F} \subset \{f : \Gamma^n \to \Gamma^n\}$ be a family of tampering functions. We say* $(\mathsf{Enc}, \mathsf{Dec})$ *is a* locally testable, non-malleable code against $\mathcal{F}$ *if there exists a $q-$local distribution $\mathcal{T}$ such that the following three points hold:*

1. *for all $m \in \Gamma^k$, $\mathbb{E}_{\mathsf{Enc}(m),\mathcal{T}}\big[\mathsf{Test}(\mathsf{Enc}(m))\big] = 1$;*

2. *for all $f \in \mathcal{F}$ there exists a distribution S_f on $\mathrm{Supp}(\mathcal{T}) \times \Gamma^q$ such that for all $m \in \Gamma^k$,*

$$\Delta\big(\mathsf{Tamper}_{f,\mathcal{T}}(m), \mathsf{S}_f\big) \leq \varepsilon.$$

3. *for all $f \in \mathcal{F}$ there exists $\mathsf{L}_f = \{g^{(1)}, \ldots, g^{(\ell)}\} \subset \mathcal{G}_{\mathsf{trivial}}$ of size $|\mathsf{L}_f| = \ell$ such that for all $m \in \Gamma^k$,*

$$\mathrm{Pr}_{\mathbf{x}\sim\mathsf{Enc}(m),\mathcal{T}}\Big[\mathsf{Test}\big(f(\mathbf{x})\big) = 1 \;\&\; f(\mathbf{x})_I \notin \big\{g_I^{(j)}(\mathbf{x}) : g^{(j)} \in \mathsf{L}_f\big\}\Big] \leq \varepsilon,$$

where $I \subset [n]$ with $|I| = q$ are the coordinates read by Test.

The following remarks attempt to justify several aspects of the definition.

1. Intuition for Understanding Points 2 and 3: Note that point 2 is local in nature, promising that the view of the tampered codeword seen by the testing procedure does

not depend on the original encoded message. This ensures that the testing procedure cannot operate differently given tampered encodings of different original messages. So in particular, the chance that the tampered codeword passes the test is the same regardless of the message inside the original codeword. Point 3 is more interesting as it ensures that a global property of the tampered codeword is independent of the original message. Specifically, if the tampered codeword's chances of passing the test are high due to good agreement with a list of valid codewords, then this agreement list does not depend on the original message.[2]

2. The logic behind the trivial tampering functions: The idea that the "trivial" tampering functions are constants and the identity is standard to non-malleability. The reasoning is that there are two resources the tampering adversary has at its disposal to make $f(\mathbf{x})$ pass the test with probability 1. The first is if the tampering adversary completely ignores the incoming codeword, overwriting whatever it receives with a constant: $f(\mathbf{x}) = \tilde{\mathbf{x}}$ for all $\mathbf{x} \in \Gamma^n$. Though $f(\mathbf{x})$ can be made to pass the test with probability 1 in this case (*e.g.*, if $\tilde{\mathbf{x}}$ is a valid codeword), this adversary is *not* performing a mauling attack since the tampered codeword does not depend on the original codeword. The tampering adversary is not content with simply destroying the ability for the sender to communicate with the receiver; this would be too easy. Instead, the adversary's goal is to, by tampering, obtain a codeword which is correlated in some way with the original.

The second option which is always available to the tampering adversary is to use

[2]Readers who are familiar with non-malleable interactive proofs from cryptography might recognize point 2 as requiring that the transcript of the tampered protocol can be simulated without the witness of the original protocol (this usually follows trivially from zero-knowledge of the original protocol), while point 3 is analogous to requiring that the witness used in the tampered protocol is independent of the original witness.

the identity function, which amounts to acting as an honest channel between the sender and receiver. In this case also, $f(\mathbf{x})$ passes the test with probability 1 since $f(\mathbf{x}) = \mathbf{x}$, a valid codeword. However, the goal of non-malleability is not to rule out this behavior, rather it is to show that the *only* meaningful option for the tampering adversary is to play honestly and avoid tampering altogether. If desired, security against a passive adversary can often be obtained via other means.

3. The requirement that $f(\mathbf{x})_I \in \{g^{(j)}(\mathbf{x})_I : g^{(j)} \in \mathsf{L}_f\}$: List decoding soundness guarantees in the "high soundness regime" for LTCs typically have a "unique decoding" flavor. For example, a common type of statement is: if $\mathbf{y} \in \Gamma^n$ has $\mathbb{E}_{\mathcal{T}}\big[\mathsf{Test}(\mathbf{y})\big] \geq 1 - \varepsilon$ then there exists a unique valid codeword $\mathbf{x} \in \Gamma^n$ such that $\mathsf{agr}(\mathbf{x}, \mathbf{y}) \geq 1 - \varepsilon'$. In the "low soundness regime", when all we know is that $\mathbb{E}_{\mathcal{T}}\big[\mathsf{Test}(\mathbf{y})\big] \geq \varepsilon$, it is not possible to obtain such a strong conclusion. For example, $\mathbf{y}$ can ensure $\mathbb{E}_{\mathcal{T}}\big[\mathsf{Test}(\mathbf{y})\big] \geq \varepsilon$ by agreeing with $1/\varepsilon$ different valid codewords, each on about $n\varepsilon$ coordinates. In our case, the tampering function can likewise alternate between $1/\varepsilon$ different $g \in \mathcal{G}_{\mathsf{trivial}}$, agreeing with each on roughly an $\varepsilon-$fraction of inputs. Additionally, the tampering function can alternate at the coordinate level, agreeing with each of the g's at an $\varepsilon-$fraction of coordinates on all inputs. By forcing $f(\mathbf{x})_I \in \{g^{(j)}(\mathbf{x})_I : g^{(j)} \in \mathsf{L}_f\}$ whenever $\mathsf{Test}\big(f(\mathbf{x})\big) = 1$ we are essentially saying that if f passes the test then it *must* be alternating among the trivial tampering functions of L_f in these fashions.

4.2.3 Fitting LTNMCs into the Coding Theory Tree

In this section, we briefly discuss how LTNMCs relate to nearby items in the coding theory tree and we also mention a few naive attempts at building LTNMCs by combining known coding objects.

First, it is clear that any LTNMC is also a LTC. On the other hand, LTNMCs do not seem to immediately give NMCs. Essentially, this is because the tester for LTNMC does not distinguish between the case when the tampered codeword is valid and when it is very close but not equal to a valid codeword. Thus, the definition 6 does not prevent "selective bot attacks" where the probability of decoding failure varies very slightly with the message. For example, a tampering function might be able to tamper an encoding of $m = 0$ to a valid codeword, and an encoding of $m = 1$ to a valid codeword except with a single incorrect symbol. In this case the tester will not notice the difference, but a decoding algorithm will have to output 0 in one case and $\perp$ in the other.

In the other direction, NMCs also do not readily give LTNMC because they might not be locally testable. One might try composed a NMC with an outer LTC to obtain a code with a local tester and (hopefully) some non-malleability properties. However, in order to show that the concatenated code is non-malleable, one basically has to show that if the outer LTC is tampered, the resulting tampering on the inner NMC is precisely what it is secure against. Thus, this requires the outer LTC to already have some non-malleability guarantees.

One notable exception to this is the case of linear (or affine) tampering. If an LTC has an encoding algorithm which is linear and the inner NMC is non-malleable against

affine tampering, then the composed code will be a LTNMC against affine tampering as well, since an affine attack on the outer code translates (by linearity) to an affine attack on the inner one.

4.3 Non-malleable PCPs

In Section 4.2 we approached the definition of LTNMCs from the perspective of unifying LTCs and NMCs. In this section, we re-approach LTNMCs from a completely different angle. This time we use the the standard cryptographic notions of zero-knowledge (ZK) and non-malleability (NM) for interactive proofs in order to define a notion of non-malleability for PCPs. In the IP setting, NM is a strengthening of ZK to handle the case of an active adversary who is able to tamper protocol messages. We begin with the definition of ZK PCPs due to Kilian, Petrank and Tardos [47] and modify it so it can deal with an active adversary. When our new definition of NM PCPs is relaxed to the setting of LTCs we recover our Definition 6 from Section 4.2.

The main takeaway from this exercise in definition tracing is that Definition 6 is the right way to formulate non-malleability for LTCs. The new definition we give for NM PCPs might also be of some independent interest. However, we stress that we do not give a construction of a NM PCP, nor do we give an application to hardness of approximation. These are both left for future work.

4.3.1 ZK and NM for Interactive Proofs

All interactive proofs will take place between a prover P and verifier V. Both P and V will share a common input x and P will use an additional secret input w to prove that $x \in \mathsf{L}$ for some language L. Throughout the interaction, P and V will exchange messages according to the protocol description, obtaining a final transcript τ. At the end, V outputs a bit indicating whether or not it accepted P's proof; P gives no output.

Definition 7 (Interactive Proof System) *Let $\varepsilon > 0$. We say that a protocol $\langle P, V \rangle$ satisfying the above syntax is an* interactive proof system for a language L *if the following completeness and $\varepsilon-$soundness properties hold.*

Completeness: *For all (x, w) such that w is a witness to $x \in \mathsf{L}$, if P and V follow the protocol specifications and if P uses (x, w) as input then $\mathbb{E}\big[V(x, \tau)\big] = 1$.*

Soundness: *For all $x \notin \mathsf{L}$, and for all adversarial P^* who possibly deviate from the protocol specifications, $\mathbb{E}\big[V(x, \tau)\big] \leq \varepsilon$.*

Definition 8 (Zero-Knowledge for IPs [41]) *Let $\varepsilon > 0$. We say that the interactive proof system $\langle P, V \rangle$ is $\varepsilon-$zero-knowledge if for all efficient adversarial V^* who possibly deviate from the protocol and for all (x, w) such that w witnesses $x \in \mathsf{L}$, there exists an efficiently sampleable distribution $\mathsf{S}_x^{V^*}$ which outputs a "simulated transcript" τ such that for all efficient distinguishers D,*

$$\left| \Pr_{\tau \sim \mathsf{R}_{(x,w)}^{V^*}} \Big[D(\tau) = 1 \Big] - \Pr_{\tau \sim \mathsf{S}_x^{V^*}} \Big[D(\tau) = 1 \Big] \right| \leq \varepsilon,$$

where $\mathsf{R}^{V^*}_{(x,w)}$ *is the "real" distribution which outputs* τ*, the transcript of interaction obtained when* V^* *interacts with an honest prover using inputs* (x, w)*.*

The Setup for Non-Malleability. Non-malleability for IPs involves an adversarial "man-in-the-middle" M^* who plays in two protocol executions, one where it plays "on the left" as the verifier against an honest P using input (x, w) and one where it plays "on the right" as the prover against an honest verifier using input $\tilde{x} \neq x$ (of M^*'s choice). For non-malleability, we redefine the real distribution $\mathsf{R}^{M^*}_{(x,w)}$ as the distribution which outputs $(\tau, \tilde{x}, \tilde{w}, \tilde{\tau})$ obtained as follows:

- τ is the transcript of the left protocol where P uses input (x, w);
- $\tilde{x}$ is the statement used in the right protocol;
- $\tilde{\tau}$ is the transcript of the right protocol where V uses input $\tilde{x}$;
- if $V(\tilde{x}, \tilde{\tau}) = 1$ then $\tilde{w}$ is a witness to $\tilde{x} \in \mathsf{L}$ (which exists with probability $1 - \varepsilon$ by soundness).

Definition 9 (Non-Malleability for IPs [35]) *Let* $\varepsilon > 0$*. We say that the interactive proof system* $\langle P, V\rangle$ *is* $\varepsilon-$*non-malleable if for all efficient adversarial* M^**, there exists an efficiently sampleable distribution* $\mathsf{S}^{M^*}_x$ *which outputs* $(\tau, \tilde{x}, \tilde{w}, \tilde{\tau})$ *such that for all efficient distinguishers* D*,*

$$\left|\Pr_{\mathsf{R}^{M^*}_{(x,w)}}\left[D(\tau, \tilde{x}, \tilde{w}, \tilde{\tau}) = 1\right] - \Pr_{\mathsf{S}^{M^*}_x}\left[D(\tau, \tilde{x}, \tilde{w}, \tilde{\tau}) = 1\right]\right| \leq \varepsilon.$$

Note Definition 9 requires that M^*'s view during the protocols $(\tau, \tilde{\tau})$ can be simulated

without knowledge of w. This actually holds whenever $\langle P, V \rangle$ is zero-knowledge, and is analogous to Point 2 of Definition 6. Additionally, Definition 9 requires that the witness $\tilde{w}$ that M^* is using to prove $\tilde{x} \in \mathsf{L}$ can also be simulated without w. This is analogous to Point 3 from 6.

4.3.2 ZK and NM for PCPs

Probabilistically checkable proofs (PCPs) also take place between a prover P and verifier V, except the syntax is different from above. Here, P uses (x, w) where w witnesses $x \in \mathsf{L}$ to (non-interactively) produce the PCP transcript $\tau \in \Gamma^n$. Given τ, V draws $\mathsf{Test} \sim \mathcal{T}$ from a $q-$local distribution and outputs $\mathsf{Test}(x, \tau)$ (Test can read all of x but queries τ in only q places).

Definition 10 (Probabilistically Checkable Proof System [11]) *Let $\varepsilon > 0$ and $\ell \in \mathbb{N}$. We say that the protocol above is a* probabilistically checkable proof system *if the following completeness and $\varepsilon-$soundness properties hold.*

Completeness: *For all (x, w) such that w is a witness to $x \in \mathsf{L}$, if P and V follow the protocol specifications and if P uses (x, w) as input then $\mathbb{E}_{\mathcal{T}}\big[\mathsf{Test}(x, \tau)\big] = 1$.*

Soundness: *For all $x \notin \mathsf{L}$, and for all $\tau \in \Gamma^n$ (possibly adversarially computed), there exists $\mathsf{L}_\tau = \{\tau^{(1)}, \ldots, \tau^{(\ell)}\} \subset \Gamma^n$ of size $|\mathsf{L}_\tau| \leq \ell$ such that*

$$\Pr_{\mathcal{T}}\Big[\mathsf{Test}(\tau) = 1 \;\&\; \tau_I \notin \{\tau_I^{(j)} : \tau^{(j)} \in \mathsf{L}_\tau\}\Big] \leq \varepsilon.$$

The adversaries in the definitions of ZK and NM for PCPs are modeled as decision

trees of bounded depth. Let $\mathcal{F}^r_{\mathsf{dt}}$ denote the set of decision trees of depth at most r.

Definition 11 (Zero-Knowledge for PCPs [47]) *Let $\varepsilon > 0$ and $r \in \mathbb{N}$. We say that a PCP system is $\varepsilon-$zero-knowledge if for all depth r decision trees $f \in \mathcal{F}^r_{\mathsf{dt}}$, there exists a distribution S^f_x on Γ^r such that*

$$\Delta\big(\mathsf{R}^f_{(x,w)}, \mathsf{S}^f_x\big) \leq \varepsilon,$$

where $\mathsf{R}^f_{(x,w),\mathcal{T}}$ draws a random PCP $\tau \in \Gamma^n$ for proving $x \in \mathsf{L}$ using w, and then outputs the r coordinates of τ which f reads.

Remark. In order to make the adversary as strong as possible, we model it as a depth r decision tree so it can make its r queries to τ adaptively. Since checking the proof requires reading q coordinates, ZK is only interesting when $r \geq q$ since in this case the adversary reads enough of the proof to verify its validity, though by ZK is unable to learn anything else about the witness (other than that it exists).

Setup for Non-Malleability. Similar to non-malleability for IPs, here the man-in-the-middle adversary receives a proof τ proving $x \in \mathsf{L}$ using witness w and computes a new statement $\tilde{x} \neq x$ and proof $\tilde{\tau} = f(\tau)$ that $\tilde{x} \in \mathsf{L}$ using some $f \in \mathcal{F}^r_{\mathsf{dt}}$.

Definition 12 (Non-Malleability for PCPs) *Let $\varepsilon > 0$ and $\ell, r \in \mathbb{N}$. We say that a PCP system with $q-$local test distribution $\mathcal{T}$ is $(\varepsilon, \ell)-$non-malleable if the following two points hold:*

1. *for every depth r decision tree $f \in \mathcal{F}^r_{\mathsf{dt}}$, there exists a distribution $\mathsf{S}^f_{x,\mathcal{T}}$ on $\mathrm{Supp}(\mathcal{T}) \times \Gamma^q$*

such that for all (x, w) *and proof* τ *that* $x \in \mathsf{L}$ *using* w,

$$\Delta\big(\mathsf{R}^f_{(x,w),\mathcal{T}}, \mathsf{S}^f_{x,\mathcal{T}}\big) \leq \varepsilon,$$

where $\mathsf{R}^f_{(x,w),\mathcal{T}}$ *is the distribution which draws* $\mathsf{Test} \sim \mathcal{T}$ *and outputs* $(\mathsf{Test}, \tilde{\tau}_I)$*, where* $I \subset [n]$ *with* $|I| = q$ *are the coordinates read by* Test*.*

2. *for every* $x \in \mathsf{L}$ *and* $f \in \mathcal{F}^r_{\mathsf{dt}}$*, there exists a list* $\mathsf{L}_{f,x} = \{\tilde{\tau}^{(1)}, \ldots, \tilde{\tau}^{(\ell)}\} \subset \Gamma^n$ *of size* $|\mathsf{L}_f| \leq \ell$ *such that*

$$\Pr_{\mathcal{T}}\Big[\mathsf{Test}(\tilde{\tau}) = 1 \;\&\; \tilde{\tau}_I \notin \big\{\tilde{\tau}^{(j)}_I : \tilde{\tau}^{(j)} \in \mathsf{L}_{f,x}\big\}\Big] \leq \varepsilon,$$

where $\tilde{\tau} = f(\tau)$ *and* $I \subset [n]$ *with* $|I|$ *are the coordinates read by* Test*.*

Remark. Note the syntactic differences and the similarities between Definitions 12 and 9. In both definitions, it is required that the adversary's view *and* witness used in the right protocol execution are independent of the witness used in the left. In Definition 12 these requirements are separated into two points whereas in Definition 9 there is a single requirement. The reason for this is that in the IP setting, the view and right witness have to be generated together since all parties must be efficient, and the right witness might not be efficiently computable given the right view. We do not have this constraint in the PCP setting so we can consider the requirements on the adversary's view and witness separately. This is convenient as the list-decoding-type PCP syntax for soundness is more complicated than soundness in IPs. Finally, note that Definition 12 is remarkably similar to Definition 6

from Section 4.2. The main differences are that in Definition 6 we relax the requirement that the tampering function is a decision tree, instead letting it be any function from a prescribed tampering family $\mathcal{F}$. Second, in Definition 6 we have to allow for the case when f is copying the codeword; this is impossible in Definition 12 because we require the statements proved on the left and right to be different: $\tilde{x} \neq x$. This leads to us using the trivial tampering family $\mathcal{G}_{\mathsf{trivial}}$ in Definition 6 but not in Definition 12.

4.4 Constructing LTNMCs

4.4.1 Our Outer Code and the Non-Malleable Affine Agreement Theorem

Notations Let $\mathbb{F}$ be a finite field, and let $k \geq 4$ and $d \geq 2$ be dimension and degree parameters, respectively. We denote by A the set of affine 3-planes in $\mathbb{F}^k$, and $\mathsf{C} = \mathbb{F}^k$. Given $\mathsf{c} \in \mathsf{C}$, we write $\mathsf{A}(\mathsf{c})$ for the set of planes $\mathsf{a} \in \mathsf{A}$ which contain c. Let Γ and Γ_{A} be, respectively, the sets of k-variate and 3-variate polynomials over $\mathbb{F}$ of degree at most d; let $\Gamma_{\mathsf{C}} = \mathbb{F}$. Given a polynomial $\Phi \in \Gamma$ and a plane $\mathsf{a} \in \mathsf{A}$, we will write $\alpha = \Phi|_{\mathsf{a}} \in \Gamma_{\mathsf{A}}$ the 3-variate restriction of Φ to a. Likewise, we write $\gamma = \Phi|_{\mathsf{c}} \in \mathbb{F}$ for the evaluation of Φ at c.

As discussed in the introduction, our final code will be the concattenation of an outer and an inner code. The outer code is a polynomial-based code of Reed-Muller type. The core of the construction is showing that any coordinate-wise tampering of the outer code which passes the local test with good probability corresponds to tampering the codeword (and underlying message) according to an affine transformation. The inner code is a (standard) non-malleable code against affine tampering. We combine the two using a composition

theorem to get a non-malleable code against coordinate-wise tampering.

Outer Code. The outer code is a version of the "planes table" code of [55].

- $\underline{\mathsf{Enc}(m)}$**:** For $m \in \mathbb{F}$, draw $\Phi \sim \Gamma$ such that $\Phi(\mathbf{0}) = m$ and output $\{\Phi|_{\mathsf{a}}\}_{\mathsf{a}\in\mathsf{A}} \in \Gamma_{\mathsf{A}}^{|\mathsf{A}|}$. We will often write codewords as $\big\{(\mathsf{a},\alpha)\big\}_{\mathsf{a}\in\mathsf{A}}$ with the understanding that $\alpha = \Phi|_{\mathsf{a}}$.

- $\underline{\mathsf{Dec}\big(\{(\mathsf{a},\alpha)\}_{\mathsf{a}\in\mathsf{A}}\big)}$**:** Given $\big\{(\mathsf{a},\alpha)\big\}_{\mathsf{a}\in\mathsf{A}}$, find $\Phi \in \Gamma$ such that $(\mathsf{a},\alpha) = (\mathsf{a},\Phi|_{\mathsf{a}})$ for all $\mathsf{a} \in \mathsf{A}$.[3] If such Φ exists, output $m = \Phi(\mathbf{0})$, otherwise output $\perp$.[4]

- $\underline{\mathsf{Test}\big(\{(\mathsf{a},\alpha)\}_{\mathsf{a}\in\mathsf{A}}\big)}$**:** Draw $\mathsf{c} \sim \mathsf{C}$, $\mathsf{a},\mathsf{a}' \sim \mathsf{A}(\mathsf{c})$; read (a,α) and (a',α'), and output 1 if $\alpha|_{\mathsf{c}} = \alpha'|_{\mathsf{c}}$ ($\alpha|_{\mathsf{c}}$ denotes the evaluation of α at c), 0 otherwise.

Codeword Tampering. We consider the family of coordinate-wise tampering function

$$\mathcal{F} := \big\{\{\mathsf{f}_{\mathsf{a}}\}_{\mathsf{a}\in\mathsf{A}} \big| \mathsf{f}_{\mathsf{a}} : \Gamma_{\mathsf{A}} \to \Gamma_{\mathsf{A}}\big\}$$

Given $\{\mathsf{f}_{\mathsf{a}}\}_{\mathsf{a}} \in \mathcal{F}$, and a codeword $\{(\mathsf{a},\alpha)\}_{\mathsf{a}}$, we write the tampered codeword as $\{(\mathsf{a},\tilde{\alpha})\}_{\mathsf{a}}$, with the understanding that $(\mathsf{a},\tilde{\alpha}) = (\mathsf{a},\mathsf{f}_{\mathsf{a}}(\alpha))$ for all $\mathsf{a} \in \mathsf{A}$.

The following theorem is the technical core of the entire construction.

Theorem 4 (Non-Malleable Affine Agreement) *Fix $\varepsilon = |\mathbb{F}|^{-\mathcal{O}(1)}$. Suppose $\{\mathsf{f}_{\mathsf{a}}\}_{\mathsf{a}\in\mathsf{A}} \in \mathcal{F}$ is such that*

$$\Pr_{\Phi,(\mathsf{c},\mathsf{a},\mathsf{a}')}\Big[\mathsf{Test}\big(\{(\mathsf{a},\tilde{\alpha})\}_{\mathsf{a}}\big) = 1\Big] \geq \varepsilon,$$

[3] Such Φ, if it exists, can be found in time $\mathsf{poly}\big(|\mathbb{F}|\big)$ by interpolation.

[4] As written, decoding runs in time $\mathsf{poly}(|\mathbb{F}|)$, which is exponential in the message length. However, local decoding algorithms exist which run in time $\mathsf{poly}\big(\lambda, \log|\mathbb{F}|, 1/\delta\big)$ and output m (or a list containing m) with probability $1 - 2^{-\lambda}$ whenever the input is within distance δ of a valid encoding of m. See for example [58].

where the probability is over $\Phi \sim \Gamma$, *and* $\mathsf{c} \sim \mathsf{C}$, $\mathsf{a}, \mathsf{a}' \sim \mathsf{A}(\mathsf{c})$,[5] *and where* $\{(\mathsf{a}, \alpha)\}_{\mathsf{a}} = \{(\mathsf{a}, \Phi|_{\mathsf{a}})\}_{\mathsf{a}}$, *and* $\{(\mathsf{a}, \tilde{\alpha})\}_{\mathsf{a}} = \{(\mathsf{a}, \mathsf{f}_{\mathsf{a}}(\alpha))\}_{\mathsf{a}}$. *Then there exists an affine map* $\mathsf{T} : \Gamma \to \Gamma$ *such that*

$$\Pr_{\Phi, \mathsf{a}}\left[\tilde{\alpha} = \mathsf{T}(\Phi)|_{\mathsf{a}}\right] = \Omega(\varepsilon).$$

4.4.2 High Level Map of the Analysis

In this section we describe, from a high level, how to prove Theorem 4, , which is where most of the new ideas are required. Before diving into this, however, we briefly mention the steps required to use Theorem 4 to get a complete analysis for a LTNMC against coordinate-wise tampering. The first step is strengthening Theorem 4 from giving an "agreement guarantee" to giving a stronger "list-decoding guarantee". Specifically, we can use Theorem 4 2 to show that for any $\{\mathsf{f}_{\mathsf{a}}\}_{\mathsf{a}} \in \mathcal{F}$, there exists a short list of affine maps $\{\mathsf{T}^{(1)}, \ldots, \mathsf{T}^{(l)}\}$ which explain nearly all of the test-passing probability. This part is mostly standard. d. For example, it is analogous to how one strengthens the agreement guarantee for the planes table LTC to a list decoding guarantee. The main theorem for the outer code is the following, it is proven assuming Theorem 4 in Appendix A.3.

Theorem 5 ((Non-Malleability of the Outer Code) *Fix* $\varepsilon = |\mathbb{F}|^{-\mathcal{O}(1)}$ *and* $l = 4/\varepsilon$. . *Suppose* $\{\mathsf{f}_{\mathsf{a}}\}_{\mathsf{a}} \in \mathcal{F}$ *is such that*

$$\Pr_{\Phi, (\mathsf{c}, \mathsf{a}, \mathsf{a}')}\left[\mathsf{Test}\big(\{(\mathsf{a}, \tilde{\alpha})\}_{\mathsf{a}}\big) = 1\right] \geq \varepsilon,$$

where the probability is over $\Phi \sim \Gamma$, *and* $\mathsf{c} \sim \mathsf{C}$, $\mathsf{a}, \mathsf{a}' \sim \mathsf{A}(\mathsf{c})$,[6] *and where* $\{(\mathsf{a}, \alpha)\}_{\mathsf{a}} =$

[5] Equivalently this probability is over $m \sim \mathbb{F}$, $\{(\mathsf{a}, \alpha)\} \sim \mathsf{Enc}(m)$ and over $(\mathsf{c}, \mathsf{a}, \mathsf{a}')$ the randomness of Test

[6] Equivalently this probability is over $m \sim \mathbb{F}$, $\{(\mathsf{a}, \alpha)\} \sim \mathsf{Enc}(m)$ and over $(\mathsf{c}, \mathsf{a}, \mathsf{a}')$ the randomness of Test

$\{(\mathsf{a}, \Phi|_{\mathsf{a}})\}_{\mathsf{a}}$, *and* $\{(\mathsf{a}, \tilde{\alpha})\}_{\mathsf{a}} = \{(\mathsf{a}, \mathsf{f}_{\mathsf{a}}(\alpha))\}_{\mathsf{a}}$. *Then there exists a list* $\mathsf{L}_{\{\mathsf{f}_{\mathsf{a}}\}} = \{\mathsf{T}^{(1)}, \ldots, \mathsf{T}^{(l)}\}$ *of affine maps* $\mathsf{T}^{(j)} : \Gamma \to \Gamma$ *of size at most* $|\mathsf{L}_{\{\mathsf{f}_{\mathsf{a}}\}}| \leq l$ *such that*

$$\mathrm{Pr}_{\Phi,(\mathsf{c},\mathsf{a},\mathsf{a}')}\left[\tilde{\alpha}|_{\mathsf{c}} = \tilde{\alpha'}|_{\mathsf{c}} \;\&\; \tilde{\alpha} \notin \{\mathsf{T}^{(j)}(\Phi)|_{\mathsf{a}}\} : \mathsf{T}^{(j)} \in \mathsf{L}_{\{\mathsf{f}_{\mathsf{a}}\}}\right] = \mathcal{O}(\varepsilon)$$

. The second step is to construct the "inner code". For this we will use a new elementary construction of a (standard) NMC against affine tampering. As mentioned in the introduction, prior constructions for non-malleable codes against affine tampering are known [4, 1, 22, 7] but our construction is much simpler. When the message space is large, our construction is more efficient than the one in [4, 1, 15] as our encoding algorithm does not require drawing large random primes. Our code achieves a better rate/error tradeoff than the construction of [22].]. The inner code construction and analysis appears in Section 4.6.1. 1. The final step is a composition theorem to combine the outer and inner codes into a concattenated code which is non-malleable against coordinate-wise tampering. This part appears in Section 4.6.2.

Proving Theorem 2. Proving Theorem 4 involves essentially analyzing a new low-degree test, similar to the planes table which was analyzed in [55, 17]. In the setting of those works, each plane is labeled with a low degree polynomial (defined on the plane) such that the intersections agree with non-negligible probability. In our work, each plane $\mathsf{a} \in \mathsf{A}$ is labeled with a function $\mathsf{f}_{\mathsf{a}} : \Gamma_{\mathsf{A}} \to \Gamma_{\mathsf{A}}$ mapping low-degree polynomials to low-degree polynomials with the following modified agreement guarantee: if $\mathsf{c} \in \mathsf{a} \cap \mathsf{a}'$ and $\alpha|_{\mathsf{c}} = \alpha'|_{\mathsf{c}}$ then with good probability $\tilde{\alpha}|_{\mathsf{c}} = \tilde{\alpha'}|_{\mathsf{c}}$ where $(\tilde{\alpha}, \tilde{\alpha'}) = (\mathsf{f}_{\mathsf{a}}(\alpha), \mathsf{f}'_{\mathsf{a}}(\alpha'))$. Unsurprisingly, the analysis of this test

borrows significantly from the analysis of the analysis of the basic planes table. However, several new ideas are needed as well. At a very high level (and slighly inaccurately), the ideas from the planes table analysis (specifically the analysis of[17]) lets us move from a non-negligible agreement guarantee, namely $\Pr_{\Phi,(\mathsf{c},\mathsf{a},\mathsf{a}')}\left[\mathsf{Test}\left(\{(\mathsf{a},\tilde{\alpha})\}_{\mathsf{a}}\right)=1\right]\geq\varepsilon$, to a high agreement guarantee on a small but non-negligible fraction of planes. . This moves us from the low soundness regime to the high soundness regime which is much easier. We then use new ideas to complete the analysis. We describe this further momentarily. We briefly mention that the low-to-high soundness conversion method of [17] 7] uses crucially the sampling properties of the incidence "planes vs points" graph. Due to the differences in our setting we need to use the sampling properties of the "incidence×agreement" graph which we establish. More information about this part appears in appendixA.1.

Finally, we walk through the proof of the affine agreement theorem assuming we are in the high soundness regime. As mentioned, this is where the bulk of our new ideas are needed. Suppose for the moment that $\{\mathsf{f}_{\mathsf{a}}\}_{\mathsf{a}}\in\mathcal{F}$ is such that the test passes with probability 1, instead of with probability ε (we will show how to remove this assumption below). If this is the case then for all $\Phi\in\Gamma$, there exists $\tilde{\Phi}\in\Gamma$ such that $\tilde{\alpha}=\tilde{\Phi}|_{\mathsf{a}}$ for all $\mathsf{a}\in\mathsf{A}$. So in this case, $\{\mathsf{f}_{\mathsf{a}}\}_{\mathsf{a}}$ defines a map $F:\Gamma\to\Gamma$, via $F(\Phi)=\tilde{\Phi}$. We must show that F is affine. The key point is that because $\{\mathsf{f}_{\mathsf{a}}\}$ acts coordinate-wise on $\{(\mathsf{a},\alpha)\}_{\mathsf{a}}$, it must be that for every $\mathsf{a}\in\mathsf{A}$, if $\Phi,\Phi'\in\Gamma$ are such that $\Phi|_{\mathsf{a}}=\Phi'|_{\mathsf{a}}$ then $F(\Phi)|_{\mathsf{a}}=F(\Phi')|_{\mathsf{a}}$. . So in words, F maps polynomials which agree at a to polynomials which agree at a for all $\mathsf{a}\in\mathsf{A}$. We show that such F must be affine.

In order to illustrate how this proof works, let us simplify the situation by changing the parameters. Instead of working with $F : \Gamma \to \Gamma$, where Γ is the set of degree d polynomials, let us assume instead that F maps linear polynomials to linear polynomials, and has the following modified property: for all $\mathsf{c} \in \mathsf{C}$, F maps polynomials which agree at c to polynomials which agree at c. Now, suppose we fix three points $\mathsf{c}, \mathsf{c}_1, \mathsf{c}_2$ on a line ine in our space and draw eight random field elements $\gamma, \gamma', \gamma'', \gamma''', \gamma_1, \gamma_1', \gamma_2, \gamma_2'$. Now, we draw four random m linear polynomials such that 1) first polynomial evaluates to $\gamma, \gamma_1, \gamma_2'$ at $\mathsf{c}, \mathsf{c}_1, \mathsf{c}_2$ 2 respectively. 2) second polynomial evaluates $\gamma', \gamma_1', \gamma_2'$ at $\mathsf{c}, \mathsf{c}_1, \mathsf{c}_2$ 3) third polynomial evaluates to $\gamma'', \gamma_1, \gamma_2$ 4) fourth polynomial evaluates to $\gamma''', \gamma_1', \gamma_2$ at $\mathsf{c}, \mathsf{c}_1, \mathsf{c}_2$. These lines are graphed in Figure 4.1 on the left. . Note, that, $\gamma, \gamma', \gamma'', \gamma'''$ share the linear relationship $\gamma - \gamma' = \gamma'' - \gamma'''$. By modified property of F, these four lines are mapped to four other lines which share the same intersecting structure (the right side of Figure4.1. This forces the same linear relationship $\tilde{\gamma} - \tilde{\gamma}' = \tilde{\gamma}'' - \tilde{\gamma}'''$ to hold on the right. It follows that that c-th co-ordinate function of F is affine. . The same argument shows that every coordinate function of F is affine, from which it follows that F itelf is affine.

We now discuss the first part: how to get the above ideas to work assuming just that the test passes with non-negligible probability ε instead of probability 1. The first point is that we don't actually need to assume that the test passes with probability 1, everything we discussed above works if the test passes with probability $1 - \delta$ for sufficiently small δ. The key is to show that if the test passes with probability ε then there are small subsets $\Gamma' \subset \Gamma$ and $\mathsf{A}' \subset \mathsf{A}$ of non-negligible weight such that, conditioned on $\Phi \in \Gamma'$ and $\mathsf{a}, \mathsf{a}' \in \mathsf{A}'$

, the test passes with very high probability $1-\delta$. Theorems of this type have been proven before in the context of LTCs and PCPs, however, our situation is a bit different because our test chooses $\Phi \sim \Gamma$, whereas usually in LTCs, Φ is fixed and only $(\mathsf{c}, \mathsf{a}, \mathsf{a}')$ are chosen.

It is useful to cast our situation in the same terminology as standard low-degree theorems. We think of the tampering function data $\{\mathsf{f}_\mathsf{a}\}_\mathsf{a}$ as assigning the 3-variate low-degree polynomial $\tilde{\alpha} = \mathsf{f}_\mathsf{a}(\alpha)$ to the plane/polynomial pair (a, α). Thus $\{\mathsf{f}_\mathsf{a}\}_\mathsf{a}$ is a a planes/polynomials table, *i.e.*, it is like the planes table from [55] which assigns a polynomial to each plane, except that the index set now consists of all plane/polynomial pairs. In [55] it is shown that for any planes table which passes the test with probability ε , the incidence graph[7] splits into "near cliques" of weight roughly ε , where the test passes with high probability whenever both planes chosen belong to the same clique. Recently, several works [45], [52], [17], [31], [32]prove similar theorems for various types of tables where they appeal only to the sampling structure of the incidence graph. Thus, the first part of our proof of the non-malleable affine agreement theorem works by demonstrating that the incidence×agreement graph is a good sampler; then we use the machinery developed in prior work to get our result. The incidence×agreement graph is the graph whose vertices are pairs (a, α) and where $\big((\mathsf{a}, \alpha), (\mathsf{a}', \alpha')\big) \in E$ iff $\mathsf{a} \cap \mathsf{a}' \in \mathsf{C}$ and if $\alpha|_{\mathsf{a}\cap\mathsf{a}'} = \alpha'|_{\mathsf{a}\cap\mathsf{a}'}$ see appendix A.3 for formal proofs.

[7] The graph whose vertices are the planes, and edges indicate that the two planes intersect in a point.

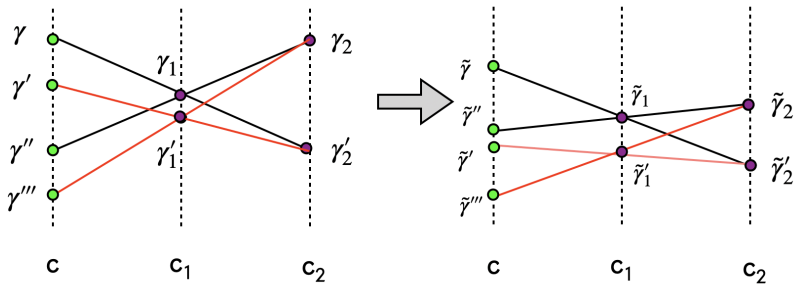

Figure 4.1: Affine agreement testing

4.4.3 Proof of Theorem 4

Notation. We have already defined A and C to be the sets of 3−planes and points over $\mathbb{F}^m$ for a finite field $\mathbb{F}$, and we have already let Γ, Γ_{A} be the sets of $m-$variate and 3−variate polynomials of degree at most d, also $\Gamma_{\mathsf{C}} = \mathbb{F}$. For convenience we define $\overline{\mathsf{A}} = \mathsf{A} \times \Gamma_{\mathsf{A}}$ and $\overline{\mathsf{C}} = \mathsf{C} \times \Gamma_{\mathsf{C}}$. We have already defined $\mathsf{A}(\mathsf{c})$ as the set of planes $\mathsf{a} \in \mathsf{A}$ which contain c. Now, given $(\mathsf{c}, \gamma) \in \overline{\mathsf{C}}$ we similarly define $\overline{\mathsf{A}}(\overline{\mathsf{c}})$ to be the set of $(\mathsf{a}, \alpha) \in \overline{\mathsf{A}}$ such that $\mathsf{c} \in \mathsf{a}$ and $\alpha|_{\mathsf{c}} = \gamma$.

Theorem 4 (Restated). *Fix $\varepsilon = |\mathbb{F}|^{-\mathcal{O}(1)}$. Suppose $\{\mathsf{f}_{\mathsf{a}}\}_{\mathsf{a} \in \mathsf{A}} \in \mathcal{F}$ is such that*

$$\Pr_{\Phi, (\mathsf{c}, \mathsf{a}, \mathsf{a}')} \Big[\mathsf{Test}\big(\{(\mathsf{a}, \tilde{\alpha})\}_{\mathsf{a}}\big) = 1 \Big] \geq 6\varepsilon,$$

where the probability is over $\Phi \sim \Gamma$, and $\mathsf{c} \sim \mathsf{C}$, $\mathsf{a}, \mathsf{a}' \sim \mathsf{A}(\mathsf{c})$,[8] *and where $\big\{(\mathsf{a}, \alpha)\big\}_{\mathsf{a}} = \big\{(\mathsf{a}, \Phi|_{\mathsf{a}})\big\}_{\mathsf{a}}$, and $\big\{(\mathsf{a}, \tilde{\alpha})\big\}_{\mathsf{a}} = \big\{(\mathsf{a}, \mathsf{f}_{\mathsf{a}}(\alpha))\big\}_{\mathsf{a}}$. Then there exists an affine map $\mathsf{T} : \Gamma \to \Gamma$ such*

[8] Equivalently this probability is over $m \sim \mathbb{F}$, $\{(\mathsf{a}, \alpha)\} \sim \mathsf{Enc}(m)$ and over $(\mathsf{c}, \mathsf{a}, \mathsf{a}')$ the randomness of Test

that

$$\Pr_{\Phi,\mathsf{a}}\left[\tilde{\alpha} = \mathsf{T}(\Phi)|_{\mathsf{a}}\right] = \varepsilon.$$

In this section, we separate the proof into two parts by stating two lemmas which combine to immediately prove the theorem.

Proof of Theorem 4. Suppose $\varepsilon = |\mathbb{F}|^{-\Omega(1)}$ is chosen so it satisfies Lemmas 6 and 7, below. Let $\{\mathsf{f}_{\mathsf{a}}\}_{\mathsf{a}} \in \mathcal{F}$ be such that

$$\Pr_{\Phi,(\mathsf{c},\mathsf{a},\mathsf{a}')}\left[\tilde{\alpha}|_{\mathsf{c}} = \tilde{\alpha}'|_{\mathsf{c}}\right] \geq 6\varepsilon. \tag{4.1}$$

By Lemma 6 below, there exists a function $\mathsf{h} : \overline{\mathsf{C}} \to \Gamma_{\mathsf{C}}$ such that

$$\Pr_{(\mathsf{a},\Phi)\sim \mathsf{A}\times\Gamma}\left[\Pr_{\overline{\mathsf{c}}\sim\overline{\mathsf{C}}(\overline{\mathsf{a}})}\left[\tilde{\alpha}|_{\mathsf{c}} = \tilde{\gamma}\right] \geq 1-\zeta\right] \geq 2\varepsilon, \tag{4.2}$$

where $\tilde{\gamma} = \mathsf{h}(\overline{\mathsf{c}})$, $\overline{\mathsf{a}} = (\mathsf{a}, \Phi|_{\mathsf{a}})$, and where $\zeta = |\mathbb{F}|^{-\Omega(1)}$ is specified precisely in Section A.3. By Lemma 7, there exists an affine map $\mathsf{T} : \Gamma \to \Gamma$ such that

$$\Pr_{(\mathsf{a},\Phi)\sim\mathsf{A}\times\Gamma}\left[\tilde{\alpha} = \mathsf{T}(\Phi)|_{\mathsf{a}}\right] \geq \varepsilon. \tag{4.3}$$

■

Lemma 6 (Global Agreement) *There exists $\varepsilon = |\mathbb{F}|^{-\Omega(1)}$ such that whenever $\{\mathsf{f}_{\mathsf{a}}\}_{\mathsf{a}} \in \mathcal{F}$ is such that (4.1) holds, there exists $\mathsf{h} : \overline{\mathsf{C}} \to \Gamma_{\mathsf{C}}$ such that (4.2) holds.*

Lemma 7 (Affine Agreement) *There exists $\varepsilon = |\mathbb{F}|^{-\Omega(1)}$ such that whenever $\{\mathsf{f}_{\mathsf{a}}\}_{\mathsf{a}} \in \mathcal{F}$ and $\mathsf{h} : \overline{\mathsf{C}} \to \Gamma_{\mathsf{C}}$ are such that (4.2) holds, there exists an affine $\mathsf{T} : \Gamma \to \Gamma$ such that (4.3)*

holds.

Lemma 6 is proved in Appendix A.3 using a sampler-based decoding argument similar to ones which have appeared in several recent works. Lemma 7 is proved in Section 4.5. Most of the ideas involved in this proof, including the new linearity test which is analyzed are new to this work.

4.5 Affine Agreement

In this section we prove Lemma 7, restated in an expanded form below. We begin here by reducing Lemma 7 to Claims 7, 8 and 9, which we will prove in Section 4.5.2 after gathering some background on linearity/low-degree tests in Section 4.5.1. Recall that a function $\mathsf{T} : \Gamma \to \Gamma$ is *affine* if there exists $u \in \mathbb{F}$ and $\Phi_0 \in \Gamma$ such that $\mathsf{T}(\Phi) = u \cdot \Phi + \Phi_0$.

Lemma 7 (Restated). *Suppose* $\{\mathsf{f_a}\}_{\mathsf{a}} \subset \{\mathsf{f} : \Gamma_{\mathsf{A}} \to \Gamma_{\mathsf{A}}\}$, $\mathsf{h} : \Gamma_{\mathsf{C}} \to \Gamma_{\mathsf{C}}$ *and* $\mathsf{G} \subset \mathsf{A} \times \Gamma$ *are such that* $|\mathsf{G}| \geq 2\varepsilon \cdot |\mathsf{A} \times \Gamma|$, *and*

$$\Pr_{\substack{(\mathsf{a},\Phi)\sim\mathsf{G} \\ \bar{\mathsf{c}}\sim\mathsf{C}(\bar{\mathsf{a}})}}\left[\tilde{\gamma} \sim \tilde{\alpha}\right] \geq 1 - \zeta, \tag{4.4}$$

where (ε, ζ) *are as in Lemma 6. Then there exists an affine map* $\mathsf{T} : \Gamma \to \Gamma$ *such that*

$$\Pr_{(\mathsf{a},\Phi)\sim\mathsf{G}}\left[\tilde{\alpha} = \mathsf{T}(\Phi)\big|_{\mathsf{a}}\right] \geq 1/2.$$

Claim 7 *Let* $\left(\varepsilon, \zeta, \{\mathsf{f_a}\}, \mathsf{h}, \mathsf{G}\right)$ *be as in the hypothesis of Lemma 7, so that (4.4) holds. Then there exist affine maps* $\{\mathsf{T_c}\}_{\mathsf{c}\in\mathsf{C}}$ *with* $\mathsf{T_c} : \Gamma_{\mathsf{C}} \to \Gamma_{\mathsf{C}}$ *such that* $\Pr_{\bar{\mathsf{c}}\sim\bar{\mathsf{C}}}\left[\tilde{\gamma} = \mathsf{T_c}(\gamma)\right] \geq 1 - \zeta$.

holds, where $\xi_7^2 := 32(d+1)(\zeta+\delta)$.

Claim 8 *Let* $(\varepsilon, \zeta, \{\mathsf{f_a}\}, \mathsf{h}, \mathsf{G})$ *be as in the hypothesis of Lemma 7, so that (4.4) holds, and let* $\{\mathsf{T_c}\}$ *be the family of affine maps promised by Claim 7. For each* $\mathsf{c} \in \mathsf{C}$, *let* $u_{\mathsf{c}}, v_{\mathsf{c}} \in \mathbb{F}$ *be the scalars defining* $\mathsf{T_c}$, *so* $\mathsf{T_c}(\gamma) := u_{\mathsf{c}} \cdot \gamma + v_{\mathsf{c}}$. *Then there exists* $u \in \mathbb{F}$ *such that* $\Pr_{\mathsf{c}\sim\mathsf{C}}\left[u_{\mathsf{c}} = u\right] \geq 1 - \xi_8$, *where* $\xi_8 := (d+2)(\zeta+\delta) + 4\xi_7 + 2/|\mathbb{F}|$.

Claim 9 *Let* $(\varepsilon, \zeta, \{\mathsf{f_a}\}, \mathsf{h}, \mathsf{G})$ *be as in the hypothesis of Lemma 7, so that (4.4) holds, and let* $\{\mathsf{T_c}\}$ *be the family of affine maps promised by Claim 7, with* $\mathsf{T_c}(\gamma) := u_{\mathsf{c}} \cdot \gamma + v_{\mathsf{c}}$, *as in Claim 8. Then there exists* $\Phi_0 \in \Gamma$ *such that* $\Pr_{\mathsf{c}\sim\mathsf{C}}\left[v_{\mathsf{c}} = \Phi_0(\mathsf{c})\right] \geq 1 - \xi_9$, *where* $\xi_9^2 := 8(d+3)^2(\zeta + \xi_7 + \xi_8)$.

Proof of Lemma 7 Assuming Claims 7, 8 and 9. Let $(\varepsilon, \zeta, \{\mathsf{f_a}\}, \mathsf{h}, \mathsf{G})$ be as in the hypothesis of Lemma 7, so that (4.4) holds, and let $\{\mathsf{T_c}\}$ be the family of affine maps promised by Claim 7. Define the affine map $\mathsf{T} : \Gamma \to \Gamma$ by $\mathsf{T}(\Phi) := u \cdot \Phi + \Phi_0$, where $u \in \mathbb{F}$ and $\Phi_0 \in \Gamma$ are the quantities guaranteed by Claims 8 and 9, respectively. We have

$$\frac{3}{4} \leq \Pr_{\substack{(\mathsf{a},\Phi)\sim\mathsf{G} \\ \mathsf{c}\sim\mathsf{C}(\mathsf{a})}}\left[\tilde{\gamma} \sim \tilde{\alpha} \;\&\; \tilde{\gamma} = \mathsf{T_c}(\gamma) \;\&\; u_{\mathsf{c}} = u \;\&\; v_{\mathsf{c}} = \Phi_0(\mathsf{c})\right] \leq \Pr_{\substack{(\mathsf{a},\Phi)\sim\mathsf{G} \\ \mathsf{c}\sim\mathsf{C}(\mathsf{a})}}\left[\tilde{\alpha}|_{\mathsf{c}} = \mathsf{T}(\Phi)|_{\mathsf{c}}\right].$$

This follows from (4.4), Claims 7, 8, 9 and the sampling of $\mathsf{A} \times \Gamma / \overline{\mathsf{C}}$. We have used the loose bound $1/4 \leq (\zeta + \xi_7 + \xi_8 + \xi_9 + \delta)$ where $\zeta > 0$ (resp. ξ_7, ξ_8, ξ_9) are the quantities from the statement of Lemma 7 (resp. Claims 7, 8, and 9), and $\delta > 0$ is the sampling parameter. It follows that $\Pr_{(\mathsf{a},\Phi)\sim\mathsf{G}}\left[\tilde{\alpha} = \mathsf{T}(\Phi)|_{\mathsf{a}}\right] \geq 1/2$, since whenever $\tilde{\alpha}$ and $\mathsf{T}(\Phi)|_{\mathsf{a}}$ agree on half of the $\mathsf{c} \in \mathsf{C}(\mathsf{a})$, they must be equal as they are both low degree. The lemma follows. ■

4.5.1 Linearity Testing Background

In this section we state three facts which we use in the next section to prove the claims. Throughout this section we use notations consistent with the rest of the paper. Additionally, in this section we use B as the set of lines in $\mathbb{F}^k$ and Γ_{B} is the set of univariate polynomials over $\mathbb{F}$ of degree at most d. Recall $\mathsf{T} : \Gamma_{\mathsf{C}} \to \Gamma_{\mathsf{C}}$ is *affine* if there exist coefficients $u, v \in \mathbb{F}$ such that $\mathsf{T}(x) = u \cdot x + v$ for all $x \in \Gamma_{\mathsf{C}}$. The first fact is standard and can be proved using linear algebraic methods.

Fact 2 (Linear Dependence of Polynomial Evaluations) *Suppose* $|\mathbb{F}| \geq d + 2$. *For any* $\mathsf{b} \in \mathsf{B}$ *and distinct* $\mathsf{c}_0, \ldots \mathsf{c}_{d+1} \in \mathsf{C}(\mathsf{b})$, *there exist non-zero coefficients* $r_0, r_1, \ldots, r_{d+1} \in \mathbb{F}$ *such that for all* $\beta \in \Gamma_{\mathsf{B}}$,

$$\sum_{i=0}^{d+1} r_i \cdot \beta|_{\mathsf{c}_i} = 0.$$

The second and third facts are proved in [56]. The second fact gives a sufficient condition for a function $f : \mathbb{F}^k \to \mathbb{F}$ being close to a multivariate low-degree polynomial.

Fact 3 (Robust Characterization of Low-Degree Functions) *Fix* $\kappa > 0$ *such that* $\kappa \leq \frac{1}{2(d+2)^2}$. *If* $f : \mathsf{C} \to \mathbb{F}$ *is such that*

$$\Pr_{\mathsf{b} \sim \mathsf{B}}\Big[\exists\ \beta \in \Gamma_{\mathsf{B}} \text{ st } \Pr_{\mathsf{c} \sim \mathsf{C}(\mathsf{b})}\big[f(\mathsf{c}) = \beta|_{\mathsf{c}}\big] \geq 1 - \kappa\Big] \geq 1 - \kappa,$$

then there exists $\Phi \in \Gamma$ *such that* $\Pr_{\mathsf{c} \sim \mathsf{C}}\big[f(\mathsf{c}) = \Phi(\mathsf{c})\big] \geq 1 - 2(d+3)\kappa$.

Fact 4 (Testing Affine Maps over Large Fields in High Soundness Regime) *Fix* $\kappa >$

0 *such that* $\kappa \leq \frac{1}{18}$. *If* $f : \Gamma_{\mathsf{C}} \to \Gamma_{\mathsf{C}}$ *is such that*

$$\Pr_{x,y,z\sim\Gamma_{\mathsf{C}}}\Big[f(x) + f(y+z) = f(x+y) + f(z)\Big] \geq 1 - \kappa,$$

then there exists an affine $\mathsf{T} : \Gamma_{\mathsf{C}} \to \Gamma_{\mathsf{C}}$ *such that* $\Pr_{x\sim\Gamma_{\mathsf{C}}}\big[f(x) = \mathsf{T}(x)\big] \geq 1 - 2\kappa$.

4.5.2 Proving the Claims

In this section we restate and prove the claims used to prove Lemma 7.

Notation. Throughout this section, we assume $\{\mathsf{f}_{\mathsf{a}}\}_{\mathsf{a}} \subset \{\mathsf{f} : \Gamma_{\mathsf{A}} \to \Gamma_{\mathsf{A}}\}$, $\mathsf{h} : \Gamma_{\mathsf{C}} \to \Gamma_{\mathsf{C}}$ and $\mathsf{G} \subset \mathsf{A} \times \Gamma$ with $|\mathsf{G}| \geq 2\varepsilon \cdot |\mathsf{A} \times \Gamma|$ are such that (4.4) holds. Namely, we assume that the hypotheses of Lemma 7. We also use $\tilde{\gamma} = \mathsf{h}(\bar{\mathsf{c}})$ throughout.

Claim 7 (Restated). *There exist affine maps* $\{\mathsf{T}_{\mathsf{c}}\}_{\mathsf{c}\in\mathsf{C}}$ *such that* $\Pr_{\bar{\mathsf{c}}\sim\bar{\mathsf{C}}}\big[\tilde{\gamma} = \mathsf{T}_{\mathsf{c}}(\gamma)\big] \geq 1 - \xi_7$.

Proof. Consider the following distribution, $\mathcal{D}$ on $\mathsf{C} \times \Gamma_{\mathsf{C}}^3$. Ultimately, the output of $\mathcal{D}$ is just uniform, however the internal choices of $\mathcal{D}$ help in our analysis. $\mathcal{D}$ works as follows:

1. draw $\mathsf{b} \sim \mathsf{B}$ and distinct $\mathsf{c}_0, \mathsf{c}_1, \ldots, \mathsf{c}_{d+1} \sim \mathsf{C}(\mathsf{b})$; let $r_0, \ldots, r_{d+1} \in \mathbb{F}$ be the coefficients guaranteed by Fact 2;

2. draw $\gamma_0^0, \gamma_0^1, \ldots, \gamma_d^0, \gamma_d^1 \sim \Gamma_{\mathsf{C}}$; let $\bar{\mathsf{c}}_{i,k} = (\mathsf{c}_k, \gamma_k^i)$, and $\tilde{\gamma}_k^i = \mathsf{h}(\bar{\mathsf{c}}_{i,k})$ for $i = 0, 1$ and $k = 0, \ldots, d$;

3. for $i, j \in \{0,1\}$, let $\beta^{i,j} \in \Gamma_{\mathsf{B}}$ be the unique polynomial that agrees with γ_0^i at c_0 and γ_k^j at c_k for all $k = 1, \ldots, d$; let $\overline{\mathsf{b}}_{i,j} = (\mathsf{b}, \beta^{i,j})$;

4. for $i, j \in \{0,1\}$, draw $(\mathsf{a}_{i,j}, \Phi^{i,j}) \sim \mathsf{G}(\overline{\mathsf{b}}_{i,j})$ and set $\tilde{\alpha}^{i,j} = \mathsf{f}_{\mathsf{a}_{i,j}}(\Phi^{i,j}|_{\mathsf{a}_{i,j}})$ and $\tilde{\beta}^{i,j} = \tilde{\alpha}^{i,j}|_{\mathsf{b}}$;

5. let $(\tilde{\gamma}, \tilde{\gamma}', \tilde{\gamma}'', \tilde{\gamma}''') = \big(\mathsf{h}(\mathsf{c}_{d+1}, \gamma), \mathsf{h}(\mathsf{c}_{d+1}, \gamma'), \mathsf{h}(\mathsf{c}_{d+1}, \gamma''), \mathsf{h}(\mathsf{c}_{d+1}, \gamma''')\big)$, where

$$(\gamma, \gamma', \gamma'', \gamma''') = \Big(\beta^{0,0}|_{\mathsf{c}_{d+1}}, \beta^{1,0}|_{\mathsf{c}_{d+1}}, \beta^{0,1}|_{\mathsf{c}_{d+1}}, \beta^{1,1}|_{\mathsf{c}_{d+1}}\Big);$$

here $\beta|_{\mathsf{c}}$ denotes the evaluation of the polynomial β at the point c;

6. output $(\mathsf{c}, x, y, z) = (\mathsf{c}_{d+1}, \gamma, \gamma' - \gamma, \gamma'')$.

Note that the output of $\mathcal{D}$ is uniform on $\mathsf{C} \times \Gamma_{\mathsf{C}}^3$. Indeed, c_{d+1} drawn in Step 1 is uniform since B/C is biregular. Moreover, given any fixed $\gamma_1^1, \ldots, \gamma_k^1$, γ'' varies uniformly as γ_0^0 does. Then, given any fixing of $(\gamma_0^0, \gamma_1^1, \ldots, \gamma_k^1)$, γ varies uniformly as $(\gamma_1^0, \ldots, \gamma_k^0)$ does. Finally, given any fixing of γ_0^0 and $(\gamma_1^0, \gamma_1^1, \ldots, \gamma_k^0, \gamma_k^1)$, γ' varies uniformly as γ_0^1 does.

Now, let $\mathbf{E}$ be the event: $\tilde{\gamma}_0^i \sim \tilde{\beta}^{i,j} \sim \tilde{\gamma}_k^j \ \forall\ (i,j,k) \in \{0,1\}^2 \times \{1, \ldots, d\}$, where the $\tilde{\gamma}_0^i$, $\tilde{\beta}^{i,j}$, and $\tilde{\gamma}_k^j$ are the internal values drawn during steps 2 and 4. By the assumptions of Lemma 7 and the sampling of $\mathsf{A} \times \Gamma/\overline{\mathsf{B}}$, we have $\Pr_{\overline{\mathsf{b}}, \overline{\mathsf{c}}, (\mathsf{a}, \Phi)}\big[\tilde{\gamma} \sim \tilde{\alpha}\big] \geq 1 - \zeta - \delta$, where the probability is over $\overline{\mathsf{b}} \sim \overline{\mathsf{B}}$, $\overline{\mathsf{c}} \sim \overline{\mathsf{C}}(\overline{\mathsf{b}})$, $(\mathsf{a}, \Phi) \sim \mathsf{G}(\overline{\mathsf{b}})$. It follows from the union bound that $\Pr_{\mathcal{D}}\big[\mathbf{E}\big] \geq 1 - \xi_7^2/8$ (substituting $\xi_7^2 = 32(d+1)(\zeta + \delta)$), since each $(\overline{\mathsf{b}}_{i,j}, \overline{\mathsf{c}}_{i,0}, \mathsf{a}_{i,j}, \Phi^{i,j})$ and $(\overline{\mathsf{b}}_{i,j}, \overline{\mathsf{c}}_{j,k}, \mathsf{a}_{i,j}, \Phi^{i,j})$ are, individually, drawn in this way for each $(i,j,k) \in \{0,1\}^2 \times \{0, \ldots, d\}$.

We complete the proof by showing that whenever the sampling of $(\mathsf{c}, x, y, z) \sim \mathcal{D}$ is such that $\mathbf{E}$ occurs, it holds that $\mathsf{h}(\mathsf{c}, x) + \mathsf{h}(\mathsf{c}, y + z) = \mathsf{h}(\mathsf{c}, x + y) + \mathsf{h}(\mathsf{c}, z)$. Together

with Fact 4, this implies that there is a family of affine maps $\{\mathsf{T}_\mathsf{c}\}_{\mathsf{c}\in\mathsf{C}}$ such that

$$\Pr_{\mathsf{c}\sim\mathsf{C}}\left[\Pr_{\gamma\sim\Gamma_\mathsf{C}}\left[\tilde{\gamma}=\mathsf{T}_\mathsf{c}(\gamma)\right]\geq 1-\frac{\xi_{7}}{2}\right]\geq 1-\frac{\xi_{7}}{2},$$

which implies the claim.

So it suffices to show that

$$\gamma-\gamma'=\gamma''-\gamma'''\text{ and }\tilde{\gamma}-\tilde{\gamma}'=\tilde{\gamma}''-\tilde{\gamma}'''$$

both hold whenever $\mathbf{E}$ occurs (the first equality always holds, the second holds whenever $\mathbf{E}$ occurs). This follows from Fact 2. The first equality holds since the $\beta^{i,j}$ are low-degree and for all (i,j,k), γ_0^i and γ_k^j are the evaluations of $\beta^{i,j}$ at c_0 and c_k, respectively. Thus Fact 2 gives

$$\begin{aligned}
&r_0\cdot\gamma_0^0+\left(\textstyle\sum_{k=1}^d r_k\cdot\gamma_k^0\right)+r_{d+1}\cdot\gamma=0; \quad && r_0\cdot\gamma_0^1+\left(\textstyle\sum_{k=1}^d r_k\cdot\gamma_k^0\right)+r_{d+1}\cdot\gamma'=0;\\
&r_0\cdot\gamma_0^0+\left(\textstyle\sum_{k=1}^d r_k\cdot\gamma_k^1\right)+r_{d+1}\cdot\gamma''=0; && r_0\cdot\gamma_0^1+\left(\textstyle\sum_{k=1}^d r_k\cdot\gamma_k^1\right)+r_{d+1}\cdot\gamma'''=0,
\end{aligned}$$

which simplifies to $\gamma-\gamma'=\gamma''-\gamma'''$ since $r_{d+1}\neq 0$. Likewise, for the second equality, the $\tilde{\beta}^{i,j}$ are low degree and when $\mathbf{E}$ occurs, the $\tilde{\gamma}_0^i$ and $\tilde{\gamma}_k^j$ are the evaluations of $\tilde{\beta}^{i,j}$ at c_0 and c_k. As above, this implies $\tilde{\gamma}-\tilde{\gamma}'=\tilde{\gamma}''-\tilde{\gamma}'''$. ■

Claim 8 (Restated). *Let $\{\mathsf{T}_\mathsf{c}\}$ be the family of affine maps promised by Claim 7; for each $\mathsf{c}\in\mathsf{C}$, let $\mathsf{T}_\mathsf{c}(\gamma):=u_\mathsf{c}\cdot\gamma+v_\mathsf{c}$ for $u_\mathsf{c},v_\mathsf{c}\in\mathbb{F}$. Then there exists $u\in\mathbb{F}$ such that $\Pr_{\mathsf{c}\sim\mathsf{C}}\left[u_\mathsf{c}=u\right]\geq 1-\xi_{8}$, where $\xi_{8}=(d+2)(\zeta+\delta)+4\xi_{7}+2/|\mathbb{F}|$.*

Proof. We prove that $\Pr_{\mathsf{c},\mathsf{c}'\sim\mathsf{C}}\big[u_{\mathsf{c}} = u_{\mathsf{c}'}\big] \geq 1 - \xi$[8] which suffices since

$$\Pr_{\mathsf{c},\mathsf{c}'\sim\mathsf{C}}\big[u_{\mathsf{c}} = u_{\mathsf{c}'}\big] = \sum_{u\in\mathbb{F}} \mathsf{p}_u^2 \leq \max\big\{\mathsf{p}_u : u \in \mathbb{F}\big\},$$

where $\mathsf{p}_u := \Pr_{\mathsf{c}\sim\mathsf{C}}\big[u_{\mathsf{c}} = u\big]$ is shorthand. As in the previous proof, we describe a distribution $\mathcal{D}'$ on C^2:

1. draw $\mathsf{b} \sim \mathsf{B}$ and distinct $\mathsf{c}_0, \mathsf{c}_1, \ldots, \mathsf{c}_{d+1} \sim \mathsf{C}(\mathsf{b})$; let $r_0, \ldots, r_{d+1} \in \mathbb{F}$ be the coefficients guaranteed by Fact 2; let $u_0, u_{d+1} \in \mathbb{F}$ denote the linear terms of $\mathsf{T}_{\mathsf{c}_0}$ and $\mathsf{T}_{\mathsf{c}_{d+1}}$, respectively;

2. draw $\gamma_0^0, \gamma_0^1, \gamma_k \sim \Gamma_{\mathsf{C}}$ for $k = 1, \ldots, d$; let $\bar{\mathsf{c}}_{i,0} = (\mathsf{c}_0, \gamma_0^i)$ for $i = 0, 1$ and $\bar{\mathsf{c}}_k = (\mathsf{c}_k, \gamma_k)$ for $k = 1, \ldots, d$; let $\tilde{\gamma}_0^i = \mathsf{h}(\bar{\mathsf{c}}_{i,0})$ and $\tilde{\gamma}_k = \mathsf{h}(\bar{\mathsf{c}}_k)$;

3. for $i \in \{0, 1\}$, let $\beta^i \in \Gamma_{\mathsf{B}}$ be the unique polynomial that agrees with γ_0^i at c_0 and γ_k at c_k for all $k = 1, \ldots, d$; let $\bar{\mathsf{b}}_i = (\mathsf{b}, \beta^i)$;

4. for $i \in \{0, 1\}$, draw $(\mathsf{a}_i, \Phi^i) \sim \mathsf{G}(\bar{\mathsf{b}}_i)$ and set $\tilde{\alpha}^i = \mathsf{f}_{\mathsf{a}_i}(\Phi^i|_{\mathsf{a}_i})$ and $\tilde{\beta}^i = \tilde{\alpha}^i|_{\mathsf{b}}$;

5. let $(\tilde{\gamma}, \tilde{\gamma}') = \big(\mathsf{h}(\mathsf{c}_{d+1}, \gamma), \mathsf{h}(\mathsf{c}_{d+1}, \gamma')\big)$, where $(\gamma, \gamma') = \big(\beta^0|_{\mathsf{c}_{d+1}}, \beta^1|_{\mathsf{c}_{d+1}}\big)$;

6. output $(\mathsf{c}, \mathsf{c}') = (\mathsf{c}_0, \mathsf{c}_{d+1})$.

Note that $\mathcal{D}'$ outputs two random points on a random line, which is within statistical distance $2/|\mathbb{F}|$ of uniform on C^2. Let $\mathbf{E}'$ be the event:

1. $\tilde{\gamma}_0^i \sim \tilde{\beta}^i \sim \tilde{\gamma}_k \;\forall\; (i, k) \in \{0, 1\} \times \{1, \ldots, d\}$; and

2. $(\tilde{\gamma}_0^0, \tilde{\gamma}_0^1, \tilde{\gamma}, \tilde{\gamma}') = \big(\mathsf{T}_{\mathsf{c}_0}(\gamma_0^0), \mathsf{T}_{\mathsf{c}_0}(\gamma_0^1), \mathsf{T}_{\mathsf{c}_{d+1}}(\gamma), \mathsf{T}_{\mathsf{c}_{d+1}}(\gamma')\big)$

The first condition occurs with probability at least $1-(d+2)(\zeta+\delta)$; as in the proof of Claim 7, this follows from (4.4), the sampling of $\mathsf{A}\times\Gamma/\overline{\mathsf{B}}$, and a union bound. The second condition occurs with probability at least $1-4\xi_{7}$, by Claim 7. Upon substituting $\xi_{8}=(d+2)(\zeta+\delta)+4\xi_{7}+2/|\mathbb{F}|$, we get $\Pr_{(\mathsf{c},\mathsf{c}')\sim\mathsf{C}^2}\left[\mathbf{E}'\right]\geq\Pr_{\mathcal{D}'}\left[\mathbf{E}'\right]-2/|\mathbb{F}|\geq 1-\xi_{8}$. As in the proof of Claim 7, Fact 2 gives

$$r_0\cdot(\gamma_0^0-\gamma_0^1)+r_{d+1}\cdot(\gamma-\gamma')=0=r_0\cdot(\tilde{\gamma}_0^0-\tilde{\gamma}_0^1)+r_{d+1}\cdot(\tilde{\gamma}-\tilde{\gamma}').$$

Substituting $(\tilde{\gamma}_0^0-\tilde{\gamma}_0^1)=u_0\cdot(\gamma_0^0-\gamma_0^1)$ and $(\tilde{\gamma}-\tilde{\gamma}')=u_{d+1}\cdot(\gamma-\gamma')$ gives $r_{d+1}(u_{d+1}-u_0)(\gamma-\gamma')=0$ which means $u_{d+1}=u_0$ since $r_{d+1}\neq 0$ and $\gamma\neq\gamma'$. Thus, $\Pr_{\mathsf{c},\mathsf{c}'\sim\mathsf{C}}\left[u_{\mathsf{c}}=u_{\mathsf{c}'}\right]\geq 1-\xi_{8}$.

■

Claim 9 (Restated). *Let $\{\mathsf{T}_{\mathsf{c}}\}$ be the family of affine maps promised by Claim 7. Then there exists $\Phi_0\in\Gamma$ with $\Pr_{\mathsf{c}\sim\mathsf{C}}\left[\mathsf{T}_{\mathsf{c}}(\mathbf{0})=\Phi_0(\mathsf{c})\right]\geq 1-\xi_{9}$, where $\xi_{9}^2=8(d+3)^2(\zeta+\xi_{7}+\xi_{8})$.*

Proof. Let $v:\mathsf{C}\to\mathbb{F}$ as a function mapping $\mathsf{c}\mapsto v_{\mathsf{c}}=\mathsf{T}_{\mathsf{c}}(\mathbf{0})$. Let $\xi:=\frac{\xi_{9}}{2(d+3)}$. We will show that

$$\Pr_{\mathsf{b}\sim\mathsf{B}}\left[\exists\ \tilde{\beta}'\in\Gamma_{\mathsf{B}}\text{ st }\Pr_{\mathsf{c}\sim\mathsf{C}(\mathsf{b})}\left[v_{\mathsf{c}}=\tilde{\beta}'|_{\mathsf{c}}\right]\geq 1-\xi\right]\geq 1-\xi. \tag{4.5}$$

The claim then follows from Fact 3. Towards establishing (4.5), note that

$$\Pr_{\substack{(\mathsf{a},\Phi)\sim\mathsf{G}\\ \mathsf{b}\sim\mathsf{B}(\mathsf{a})\\ \mathsf{c}\sim\mathsf{C}(\mathsf{b})}}\left[v_{\mathsf{c}}=\tilde{\beta}|_{\mathsf{c}}-u\cdot\beta|_{\mathsf{c}}\right]\geq 1-(\zeta+\xi_{7}+\xi_{8})\geq 1-\xi(\xi-\delta),$$

where $\beta=\Phi|_{\mathsf{b}}$ and $\tilde{\beta}=\tilde{\alpha}|_{\mathsf{b}}$; we have used $\xi(\xi-\delta)\geq\xi^2/2=\zeta+\xi_{7}+\xi_{8}$. This follows

immediately from (4.4) and Claims 7 and 8. By an averaging argument,

$$\Pr_{\substack{(\mathsf{a},\Phi)\sim\mathsf{G}\\ \mathsf{b}\sim\mathsf{B}(\mathsf{a})}}\Big[\Pr_{\mathsf{c}\sim\mathsf{C}(\mathsf{b})}\big[v_{\mathsf{c}}=\tilde{\beta}'|_{\mathsf{c}}\big]\geq 1-\xi\Big]\geq 1-\xi+\delta,$$

where $\tilde{\beta}'=\tilde{\beta}-u\cdot\beta$. The bound (4.5) now follows from the sampling of $\mathsf{A}\times\Gamma/\mathsf{B}$. ■

4.6 A Locally Testable, Non-Malleable Code

In this section, we give a construction of a locally testable non-malleable code against coordinate wise tampering. We take the outer code, $(\mathsf{E}_{\mathsf{LTNM}},\mathsf{D}_{\mathsf{LTNM}},\mathsf{T}_{\mathsf{LTNM}})$ from section 4.4 and compose it with a new non-malleable code, $(\mathsf{E}_{\mathsf{aff}},\mathsf{D}_{\mathsf{aff}})$, against affine tampering to get the final code.

4.6.1 A Simple Non-malleable Code against Affine Tampering

We begin with a new constant rate, non-malleable code against affine tampering. This result is not new, several prior works [4, 23, 48, 22] give such codes, however, our construction is considerably simpler than those prior.

Notations. Let $\mathbb{F}$ be a finite field and $\mathbb{K}/\mathbb{F}$ a degree 3 extension, so $\mathbb{K}=\mathbb{F}[x]/\big(p(x)\big)$ for an irreducible cubic polynomial $p(x)=x^3-e_2x^2-e_1x-e_0$. Thus $\mathbb{K}$ is a 3-dimensional $\mathbb{F}-$vector space with basis $\{1,\sigma,\sigma^2\}$, where $\sigma\in\mathbb{K}$ is a root of $p(x)$. The 'multiplication by

σ' map $\mathbb{F}^3 \to \mathbb{F}^3$ is linear, specified over this basis by the matrix

$$\Sigma = \begin{bmatrix} 0 & 0 & e_0 \\ 1 & 0 & e_1 \\ 0 & 1 & e_2 \end{bmatrix} \in \mathbb{F}^{3\times 3}.$$

Our code makes use of an $\varepsilon-$*high entropy encoding*, (E, D), with codeword space $\mathbb{F}$, such that for all m, c^*, $\Pr_{c\sim\mathsf{E}(m)}\big[c = c^*\big] \leq \varepsilon$. Such codes can be trivilally constructed by appending a message with a random string of length $\log\big(1/\varepsilon\big)$.

Construction. Let (E, D) be an $\varepsilon-$high entropy code with message space $\mathcal{M}$ and codeword space $\mathbb{F}$, and let $m \in \mathcal{M}$.

• $\underline{\mathsf{E}_{\mathsf{aff}}(m)}$**:** Draw $r \sim \mathbb{F}; w \sim \mathsf{E}(m)$ and output $w + r \cdot \sigma + wr \cdot \sigma^2 \in \mathbb{K}$.

• $\underline{\mathsf{D}_{\mathsf{aff}}(c)}$**:** Parse $c = c_0 + c_1 \cdot \sigma + c_2 \cdot \sigma^2$; if $c_0 \cdot c_1 = c_2$, output $m = \mathsf{D}(c_0)$; if not, output $\perp$.

Theorem 6 *Fix $\varepsilon > 0$, and let (E, D) be an $\varepsilon-$high entropy code with message space $\mathcal{M}$ and codeword space $\mathbb{F}$. Then $(\mathsf{E}_{\mathsf{aff}}, \mathsf{D}_{\mathsf{aff}})$ is a $(2\varepsilon + 2/|\mathbb{F}|)-$non-malleable code against affine tampering functions.*

Proof. Fix an affine map f given by $f(x) = sx + t$ where $s, t, x \in \mathbb{K}$ and fix any message $m \in \mathcal{M}$. Parse $s = s_0 + s_1 \cdot \sigma + s_2 \cdot \sigma^2$ and $t = t_0 + t_1 \cdot \sigma + t_2 \cdot \sigma^2$. To prove the theorem, we exhibbit a trivial tampering function g_f (*i.e.*, either constant or the identity) such that the tampering distribution $\big(\mathsf{D}_{\mathsf{aff}} \circ f \circ \mathsf{E}_{\mathsf{aff}}\big)(m)$ outputs $g_f(m)$ with probability at least $1 - 2\varepsilon - 2/|\mathbb{F}|$. The trivial function g_f is f if f is either the identity or a constant function mapping to a valid codeword, and is the constant $\perp$ function otherwise.

Specifically, if $(s,t) = (1,0)$, g_f is the identity; if $s = 0$ and $t_0 \cdot t_1 = t_2$, g_f is the constant function mapping everything to t; otherwise g_f is the constant $\perp$ function. The key point, is that for all $m \in \mathcal{M}$, the distribution $f\big(\mathsf{E}_{\mathsf{aff}}(m)\big)$ draws $w \sim \mathsf{E}(m)$, $r \sim \mathbb{F}$ and outputs

$$S\begin{bmatrix} w \\ r \\ wr \end{bmatrix} + \begin{bmatrix} t_0 \\ t_1 \\ t_2 \end{bmatrix} = \begin{bmatrix} t_0 + s_0 w + e_0 s_2 r + (e_0 s_1 + e_0 e_2 s_2) wr \\ t_1 + s_1 w + (s_0 + e_1 s_2) r + (e_1 s_1 + s_2 e_0 + s_2 e_1 e_2) wr \\ t_2 + s_2 w + (s_1 + e_2 s_2) r + (s_0 + e_2 s_1 + s_2 e_2^2 + s_2 e_1) wr \end{bmatrix} =: \begin{bmatrix} \mathsf{C}_0(w,r) \\ \mathsf{C}_1(w,r) \\ \mathsf{C}_2(w,r) \end{bmatrix},$$

where $S \in \mathbb{F}^{3\times 3}$ is the 'multiplication by s' matrix: $S = s_0 \cdot \mathbb{1} + s_1 \cdot \Sigma + s_2 \cdot \Sigma^2$. In the above, we have defined bilinear (*i.e.*, of the form $a + bx + cy + dxy$) polynomials $\mathsf{C}_0, \mathsf{C}_1, \mathsf{C}_2 \in \mathbb{F}[x,y]$. Note that if $\mathsf{C}_0(x,y) \cdot \mathsf{C}_1(x,y) \not\equiv \mathsf{C}_2(x,y)$ as polynomials, then $\mathsf{C}_0(w,r) \cdot \mathsf{C}_1(w,r) = \mathsf{C}_2(w,r)$ holds with probability at most $2\varepsilon + 2/|\mathbb{F}|$, in which case $\big(\mathsf{D}_{\mathsf{aff}} \circ f \circ \mathsf{E}_{\mathsf{aff}}\big)(m) = \perp$ with high probability. This follows immediately from Schwartz-Zippel and the low entropy property of (E, D). Therefore, in order to prove the theorem, it suffices to show that if $\mathsf{C}_0(x,y) \cdot \mathsf{C}_1(x,y) \equiv \mathsf{C}_2(x,y)$ holds, then either $s = 0$ or $(s,t) = (1,0)$. We assume $\mathsf{C}_0(x,y) \cdot \mathsf{C}_1(x,y) \equiv \mathsf{C}_2(x,y)$ holds, and we prove the following three items:

1. either $s_1 = 0$ or $s_2 = 0$;
2. $s_1 = 0 \Leftrightarrow s_2 = 0$;
3. if $s_1 = s_2 = 0$ then either $s_0 = 0$ or $s_0 = 1$ and $t_0 = t_1 = t_2 = 0$.

The third point is easiest: if $\mathsf{C}_0(x,y) \cdot \mathsf{C}_1(x,y) \equiv \mathsf{C}_2(x,y)$ and $s_1 = s_2 = 0$ then plugging gives

$$(t_0 + s_0 x) \cdot (t_1 + s_0 y) = t_2 + s_0 xy,$$

from which it follows that either $s_0 = 0$ or $s_0 = 1$ and $t_i = 0$ for all $i = 0, 1, 2$. To prove the first point, note that if $\mathsf{C}_0(x, y) \cdot \mathsf{C}_1(x, y) \equiv \mathsf{C}_2(x, y)$, then $s_0 \cdot s_1 = 0$ (since the x^2 coefficient in C_2 is zero). If $s_1 = 0$ we are done; if $s_0 = 0$ then $e_0 e_1 s_2^2 = 0$ (since y^2 coefficient in C_2 is zero), which implies $e_1 s_2 = 0$ since $e_0 \neq 0$ (else $p(x)$ is reducible). If $s_2 = 0$ we are done; if $e_1 = 0$ then $e_0^2 s_2^2 = 0$ (since xy^2 coefficient in C_2 is zero). Again, $e_0 \neq 0$ so $s_2 = 0$ so the first point follows.

Finally, for the second point, assume $s_1 = 0$. Then $s_0 s_2 \cdot (e_0 + e_1 e_2) = 0$ since the coefficient of $x^2 y = 0$ in C_2. Note $e_0 \neq -e_1 e_2$ since otherwise $p(x)$ is reducible: $p(x) = (x - e_2)(x^2 - e_1)$. However, if $s_0 = 0$ then, as shown in the proof of the first point, $s_2 = 0$; therefore $s_1 = 0$ implies $s_2 = 0$. Conversely, if $s_2 = 0$ then $e_0 s_0 s_1 = 0$ (coefficient of xy^2 in C_2 is zero), so $s_0 s_1 = 0$. If $s_0 = 0$ then $e_0 s_1^2 = 0$ (coefficient of $x^2 y$ in C_2 is zero). Thus $s_2 = 0$ implies $s_1 = 0$, and we are done. ■

Remark. In our LTNM code in the next section, we will use $(\mathsf{E}_{\mathsf{aff}}, \mathsf{D}_{\mathsf{aff}})$ to encode a random $w \in \mathbb{F}$ and so the high entropy encoding is not necessary. The precise claim we use is stated below. The proof is the same as above since if $\mathsf{C}_0(x, y) \cdot \mathsf{C}_1(x, y) \not\equiv \mathsf{C}_2(x, y)$ as polynomials, then $\mathsf{C}_0(w, r) \cdot \mathsf{C}_1(w, r) = \mathsf{C}_2(w, r)$ holds with probability at most $4/|\mathbb{F}|$ over $w, r \sim \mathbb{F}$.

Claim 10 *Let $f : \mathbb{K} \to \mathbb{K}$ be affine of the form $f(x) = sx + t$ for $s, t \in \mathbb{K}$ such that $s \neq 0$ and $(s, t) \neq (1, 0)$. Then* $\Pr_{w,r \sim \mathbb{F}}\Big[\mathsf{D}_{\mathsf{aff}}\big(f(w + r \cdot \sigma + wr \cdot \sigma^2)\big) \neq \perp\Big] \leq 4/|\mathbb{F}|$.

4.6.2 A LTNM Code via Composition

Composition Overview. The local test of our main construction from Section 4.4 passes whenever codewords are tampered by a coordinate-wise affine function. Thus, in order to use our main construction to build a fully LTNM code, we must modify the test in such a way so that it fails whenever a non-trivial affine tampering function is used. We do this in two steps. First, we modify the local tester so that it locally decodes a specified polynomial evaluation. Second, the tester checks that the evaluation recovered is a valid codeword of $(\mathsf{E}_{\mathsf{aff}}, \mathsf{D}_{\mathsf{aff}})$, if not it outputs $\perp$. Essentially, the reason this works is that the local decoder will output $\perp$ unless the codeword is tampered with an affine function, in which case the evaluation recovered is an affine function of the original evaluation. If the original evaluation is a random valid codeword of $(\mathsf{E}_{\mathsf{aff}}, \mathsf{D}_{\mathsf{aff}})$ then by Claim 10, the recovered evaluation is a valid codeword only if the affine tampering function is trivial.

Notations. As in the previous section, let $\mathbb{K}/\mathbb{F}$ be a degree 3 extension with $\mathbb{F}-$basis $\{1, \sigma, \sigma^2\}$. Let $k \geq 5$ and $d \geq 2$. As in the rest of the paper, let A be the set of 3−planes in $\mathbb{F}^k$ and $\mathsf{C} = \mathbb{F}^k$. In this section, we use B and $\bar{\mathsf{A}}$ to denote the set of lines and 4−planes respectively(note, the second usage is different from rest of the paper where we used $\bar{\mathsf{A}}$ to denote $\mathsf{A} \times \Gamma_{\mathsf{A}}$). Let $\mathsf{p} = (1, 0, \ldots, 0) \in \mathbb{F}^k$.

Construction. Let $\mathsf{E}_{\mathsf{aff}}()$ denote the procedure which draws $w, r \sim \mathbb{F}$, and outputs the value $w + r \cdot \sigma + wr \cdot \sigma^2 \in \mathbb{K}$; let $\mathsf{D}_{\mathsf{aff}}$ be the decoding algorithm from the previous section. Let $m \in \mathbb{K}$ be a message.

• $\underline{\mathsf{Enc}(m)}$**:** Draw $v \sim \mathsf{E}_{\mathsf{aff}}()$; and $\Phi \sim \Gamma$ such that $\Phi(\mathbf{0}) = m$ and $\Phi(\mathsf{p}) = v$; output

$\{(\mathsf{a}, \Phi_{\mathsf{a}})\}_{\mathsf{a}\in\mathsf{A}}$.

- $\underline{\mathsf{Dec}(\{(\mathsf{a},\alpha)\}_{\mathsf{a}\in\mathsf{A}})}$**:** Find $\Phi \in \Gamma$ such that $(\mathsf{a}, \alpha) = (\mathsf{a}, \Phi|_{\mathsf{a}})$ for all $\mathsf{a} \in \mathsf{A}$. If such Φ exists, and if $\mathsf{D}_{\mathsf{aff}}(\Phi(\mathsf{p})) \neq \perp$, output $m = \Phi(\mathbf{0})$, otherwise output $\perp$.

- $\underline{\mathsf{Test}(\{(\mathsf{a},\alpha)\}_{\mathsf{a}\in\mathsf{A}})}$**:** Draw $\mathsf{b} \sim \mathsf{B}(\mathsf{p})$, $\mathsf{c}_1, \mathsf{c}_2, \mathsf{c}_3 \sim \mathsf{C}(\mathsf{b})$, $\mathsf{c}, \mathsf{c}' \sim \mathsf{C}$, $\mathsf{a}_1 \sim \mathsf{A}(\mathsf{c}, \mathsf{c}_1)$, $\mathsf{a}_2 \sim \mathsf{A}(\mathsf{c}, \mathsf{c}', \mathsf{c}_2)$, $\mathsf{a}_3 \sim \mathsf{A}(\mathsf{c}', \mathsf{c}_3)$. Read (a_1, α_1), (a_2, α_2), (a_3, α_3) and do the following.

  1) Check that $\alpha_1|_{\mathsf{c}} = \alpha_2|_{\mathsf{c}}$ and $\alpha_2|_{\mathsf{c}'} = \alpha_3|_{\mathsf{c}'}$; if not output 0; if so use interpolation to recover $\beta \in \Gamma_{\mathsf{B}}$, the unique degree 2 polynomial such that $\beta|_{\mathsf{c}_i} = \alpha_i|_{\mathsf{c}_i}$ for $i = 1, 2, 3$; let $v = \beta|_{\mathsf{p}}$.

  2) If $\mathsf{D}_{\mathsf{aff}}(v) \neq \perp$, output 1; otherwise output 0.

Theorem 7 *Let ℓ, ε as in theorem 5. Then the code $(\mathsf{Enc}, \mathsf{Dec}, \mathsf{Test})$ above is a $(\ell, \varepsilon')-$locally testable, non-malleable code against $\mathcal{F}$, the family of coordinate-wise tampering functions where $\varepsilon' = \mathcal{O}(\varepsilon^{1/2})$.*

Proof. Fix a tampering function $\mathsf{f} = \{\mathsf{f}_{\mathsf{a}}\}_{\mathsf{a}} \in \mathcal{F}$. Let $\mathcal{G}$ be the family of affine maps. We prove that $(\mathsf{Enc}, \mathsf{Dec}, \mathsf{Test})$ is LTNMC by showing conditions of Definition 6 holds. The first condition is trivial. It is also not difficult to see that the distribution S_{f} that draws $m' \sim \mathbb{K}$ and outputs $\mathsf{Tamper}_{\mathsf{f},\mathcal{T}}(m')$ satisfies the second condition. Therefore, it remains to exhibit a list $\mathsf{L}_{\mathsf{f}} \subset \mathcal{G}_{\mathsf{trivial}}$ of size at most $|\mathsf{L}_{\mathsf{f}}| \leq \ell$ such that $\mathsf{val} \leq \mathcal{O}(\varepsilon^{1/2})$ where

$$\mathsf{val} := \Pr_{\Phi,\mathsf{rand}}\Big[\mathsf{Test} \text{ passes } \& \ (\tilde{\alpha}_1, \tilde{\alpha}_2, \tilde{\alpha}_3) \notin \big\{\big(\mathsf{h}_{\mathsf{a}_1}(\Phi|_{\mathsf{a}_1}), \mathsf{h}_{\mathsf{a}_2}(\Phi|_{\mathsf{a}_2}), \mathsf{h}_{\mathsf{a}_3}(\Phi|_{\mathsf{a}_3})\big) : \{\mathsf{h}_{\mathsf{a}}\}_{\mathsf{a}} \in \mathsf{L}_{\mathsf{f}}\big\}\Big],$$

where $\tilde{\alpha}_i = \mathsf{f}_{\mathsf{a}_i}(\Phi|_{\mathsf{a}_i})$. In the course of the proof of Theorem 5 from Section 4.4.2(see

appendix A.2), a similar list $\mathsf{L}'_{\mathsf{f}} \subset \mathcal{G}$ of size at most $|\mathsf{L}'_{\mathsf{f}}| \leq \ell$ was constructed such that

$$\Pr_{\Phi,(\mathsf{c},\mathsf{a}_1,\mathsf{a}_2)}\Big[\tilde{\alpha}_1|_{\mathsf{c}} = \tilde{\alpha}_2|_{\mathsf{c}} \;\&\; (\tilde{\alpha}_1,\tilde{\alpha}_2) \notin \big\{\big(\mathsf{g}_{\mathsf{a}_1}(\Phi|_{\mathsf{a}_1}), \mathsf{g}_{\mathsf{a}_2}(\Phi|_{\mathsf{a}_2})\big) : \{\mathsf{g}_{\mathsf{a}}\}_{\mathsf{a}} \in \mathsf{L}'_{\mathsf{f}}\big\}\Big] \leq \varepsilon,$$

where this probability is over $\Phi \sim \Gamma$ and $\mathsf{c} \sim \mathsf{C}$, $\mathsf{a}_1, \mathsf{a}_2 \sim \mathsf{A}(\mathsf{c})$. Our list $\mathsf{L}_{\mathsf{f}} \subset \mathcal{H}$ is the set of trivial $\{\mathsf{g}_{\mathsf{a}}\}_{\mathsf{a}} \in \mathsf{L}'_{\mathsf{f}}$. The quantity val can now be bounded

$$\mathsf{val} \leq \Pr_{\Phi,\mathsf{rand}}\big[\mathbf{E}_1 \vee \mathbf{E}'_1 \vee \mathbf{E}_2 \vee \mathbf{E}_3\big]$$

for the following events:

$\mathbf{E}_1$: $\tilde{\alpha}_1|_{\mathsf{c}} = \tilde{\alpha}_2|_{\mathsf{c}} \;\&\; (\tilde{\alpha}_1,\tilde{\alpha}_2) \notin \big\{\big(\mathsf{g}_{\mathsf{a}_1}(\Phi|_{\mathsf{a}_1}), \mathsf{g}_{\mathsf{a}_2}(\Phi|_{\mathsf{a}_2})\big) : \{\mathsf{g}_{\mathsf{a}}\}_{\mathsf{a}} \in \mathsf{L}'_{\mathsf{f}}\big\}$;

$\mathbf{E}'_1$: $\tilde{\alpha}_2|_{\mathsf{c}'} = \tilde{\alpha}_3|_{\mathsf{c}'} \;\&\; (\tilde{\alpha}_2,\tilde{\alpha}_3) \notin \big\{\big(\mathsf{g}'_{\mathsf{a}_2}(\Phi|_{\mathsf{a}_2}), \mathsf{g}'_{\mathsf{a}_3}(\Phi|_{\mathsf{a}_3})\big) : \{\mathsf{g}'_{\mathsf{a}}\}_{\mathsf{a}} \in \mathsf{L}'_{\mathsf{f}}\big\}$;

$\mathbf{E}_2$: the $\{\mathsf{g}_{\mathsf{a}}\}_{\mathsf{a}}, \{\mathsf{g}'_{\mathsf{a}}\}_{\mathsf{a}} \in \mathcal{G}$ which agree with f from $\mathbf{E}_1$ and $\mathbf{E}'_1$ are distinct and such that $\mathsf{g}_{\mathsf{a}_2}(\Phi|_{\mathsf{a}_2}) = \mathsf{g}'_{\mathsf{a}_2}(\Phi|_{\mathsf{a}_2})$;

$\mathbf{E}_3$: the same $\{\mathsf{g}_{\mathsf{a}}\}_{\mathsf{a}} \in \mathcal{G}$ results from $\mathbf{E}_1$ and $\mathbf{E}'_1$; this $\{\mathsf{g}_{\mathsf{a}}\}_{\mathsf{a}} \in \mathcal{G}$ is non-trivial, but the affine check passes: $\mathsf{D}_{\mathsf{aff}}(\tilde{v}) \neq \perp$.

The marginal distribution on a_2 from rand is uniform, so $\Pr_{\Phi,\mathsf{rand}}\big[\mathbf{E}_2\big] = \mathcal{O}\big(|\mathbb{F}|^{-1}\big)$. By Claim 10, $\Pr_{\Phi,\mathsf{rand}}\big[\mathbf{E}_3\big] \leq 4/|\mathbb{F}|$. We prove $\Pr_{\Phi,\mathsf{rand}}\big[\mathbf{E}_1\big] \leq \varepsilon^{1/2} + \mathcal{O}\big(|\mathbb{F}|^{-1}\big)$. The same holds for $\mathbf{E}'_1$, and the result follows. Towards bounding $\Pr_{\Phi,\mathsf{rand}}\big[\mathbf{E}_1\big]$, note that drawing $\Phi \sim \Gamma$ uniformly, rather than uniformly subject to $\Phi(\mathbf{0}) = m$ and $\Phi(\mathsf{p}) = v$ changes the probability

by at most $\mathcal{O}\big(|\mathbb{F}|^{-1}\big)$. Therefore, in the calculation below, we assume $\Phi \sim \Gamma$. We have

$$\begin{aligned}\Pr_{\Phi,\mathsf{rand}} \quad \big[\mathbf{E}_1\big]^2 &= \mathbb{E}_{\Phi,\mathsf{c}\sim\mathsf{C},\mathsf{a}_2\sim\mathsf{A}(\mathsf{c})}\Big[\Pr_{\mathsf{a}_1\sim\mathsf{rand}(\mathsf{c},\mathsf{a}_2)}\big[\mathbf{E}_1\big]\Big]^2 \leq \mathbb{E}_{\Phi,\mathsf{c},\mathsf{a}_2}\Big[\Pr_{\mathsf{a}_1\sim\mathsf{rand}(\mathsf{c},\mathsf{a}_2)}\big[\mathbf{E}_1\big]^2\Big] \\ \leq \qquad &\mathbb{E}_{\Phi,\mathsf{c},\mathsf{a}_2}\Big[\Pr_{\mathsf{a}_1,\mathsf{a}_3\sim\mathsf{rand}(\mathsf{c},\mathsf{a}_2)}\big[\tilde{\alpha}_1|_{\mathsf{c}} = \tilde{\alpha}_2|_{\mathsf{c}} = \tilde{\alpha}_3|_{\mathsf{c}} \;\&\; (\tilde{\alpha}_1,\tilde{\alpha}_2,\tilde{\alpha}_3) \notin \mathsf{L}'_{\mathsf{f}}\big]\Big] + \mathcal{O}(|\mathbb{F}|^{-1}),\end{aligned}$$

where "$(\tilde{\alpha}_1,\tilde{\alpha}_2,\tilde{\alpha}_3) \notin \mathsf{L}'_{\mathsf{f}}$" is shorthand for

$$(\tilde{\alpha}_1,\tilde{\alpha}_2,\tilde{\alpha}_3) \notin \big\{\big(\mathsf{g}_{\mathsf{a}_1}(\Phi|_{\mathsf{a}_1}),\mathsf{g}_{\mathsf{a}_2}(\Phi|_{\mathsf{a}_2}),\mathsf{g}_{\mathsf{a}_3}(\Phi|_{\mathsf{a}_3})\big) : \{\mathsf{g}_{\mathsf{a}}\}_{\mathsf{a}} \in \mathsf{L}'_{\mathsf{f}}\big\}$$

and the $\mathcal{O}(|\mathbb{F}|^{-1})$ term in the second line accounts for the case when there are $\{\mathsf{g}_{\mathsf{a}}\}_{\mathsf{a}}, \{\mathsf{g}'_{\mathsf{a}}\}_{\mathsf{a}} \in \mathsf{L}'_f$ such that $\mathsf{g}_{\mathsf{a}_2}(\Phi|_{\mathsf{a}_2}) = \mathsf{g}'_{\mathsf{a}_2}(\Phi|_{\mathsf{a}_2})$ holds. Note that if $\tilde{\alpha}_1 = \mathsf{g}_{\mathsf{a}_1}(\Phi|_{\mathsf{a}_1})$, and $\tilde{\alpha}_2 \neq \mathsf{g}_{\mathsf{a}_2}(\Phi|_{\mathsf{a}_2})$, then $\tilde{\alpha}_1|_{\mathsf{c}} = \tilde{\alpha}_2|_{\mathsf{c}}$ occurs with probability $\mathcal{O}\big(|\mathbb{F}|^{-1}\big)$. It follows that

$$\Pr_{\Phi,\mathsf{rand}}\big[\mathbf{E}_1\big]^2 \leq \Pr_{\substack{\Phi,\mathsf{c},\mathsf{a}_2\\ \mathsf{a}_1,\mathsf{a}_3\sim\mathsf{rand}(\mathsf{c},\mathsf{a}_2)}}\Big[\tilde{\alpha}_1|_{\mathsf{c}} = \tilde{\alpha}_3|_{\mathsf{c}} \;\&\; (\tilde{\alpha}_1,\tilde{\alpha}_3) \notin \mathsf{L}'_{\mathsf{f}}\Big] + \mathcal{O}\big(|\mathbb{F}|^{-1}\big).$$

Therefore, it suffices to show that for all $\mathsf{c} \in \mathsf{C}$, the distribution which draws $\mathsf{a}_2 \sim \mathsf{A}(\mathsf{c})$, $\mathsf{a}_1,\mathsf{a}_3 \sim \mathsf{rand}(\mathsf{c},\mathsf{a}_2)$ and outputs $(\mathsf{a}_1,\mathsf{a}_3)$ is within statistical distance $\mathcal{O}\big(|\mathbb{F}|^{-1}\big)$ of uniform on $\mathsf{A}(\mathsf{c})^2$. The distribution $\mathsf{rand}(\mathsf{c},\mathsf{a}_2)$ draws $\mathsf{c}_2 \sim \mathsf{C}(\mathsf{a}_2)$, $\mathsf{c}_1 \sim \mathsf{C}(\mathsf{b})$, where b is the line through p and c_2, and outputs $\mathsf{a}_1 \sim \mathsf{A}(\mathsf{c},\mathsf{c}_1)$. This is equivalent to drawing $\mathsf{c}_1 \sim \mathsf{C}(\bar{\mathsf{a}}_2)$ and outputting $\mathsf{a}_1 \sim \mathsf{A}(\mathsf{c},\mathsf{c}_1)$, where $\bar{\mathsf{a}}_2$ is the $4-$plane containing a_2 and p. Thus the distribution which draws $\mathsf{a}_2 \sim \mathsf{A}(\mathsf{c})$ and then $\mathsf{a}_1,\mathsf{a}_3 \sim \mathsf{rand}(\mathsf{c},\mathsf{a}_2)$, outputting $(\mathsf{a}_1,\mathsf{a}_2,\mathsf{a}_3)$ can be equivalently described by drawing $\mathsf{a}_1,\mathsf{a}_3 \sim \mathsf{A}(\mathsf{c})$, $\mathsf{c}_i \sim \mathsf{C}(\mathsf{a}_i)$ for $i = 1,3$, $\bar{\mathsf{a}}_2 \sim \overline{\mathsf{A}}(\mathsf{c},\mathsf{p},\mathsf{c}_1,\mathsf{c}_3)$ (*i.e.*, a random $4-$plane containing $\mathsf{c},\mathsf{p},\mathsf{c}_1,\mathsf{c}_3$), $\mathsf{a}_2 \sim \mathsf{A}(\mathsf{c},\bar{\mathsf{a}}_2)$ and outputting $(\mathsf{a}_1,\mathsf{a}_2,\mathsf{a}_3)$. In the previous

calculation we have ignored error terms of size $\mathcal{O}\big(|\mathbb{F}|^{-1}\big)$. Thus the marginal distribution on $(\mathsf{a}_1, \mathsf{a}_3)$ is $\mathcal{O}\big(|\mathbb{F}|^{-1}\big)-$close to uniform on $\mathsf{A}(\mathsf{c})$, and the result follows. ■

Chapter 5

Conclusions

5.1 Thesis Summary

Over the last few decades many different notions of pseudorandomness and pseudorandom objects have been studied. One important example of such objects is the notion of expander graphs which has found many applications in theoretical computer science. Recently, Ta-Shma[60] constructed binary codes based on random walks on expander graphs, that achieves almost optimal rate-bias trade-off. The proof of his construction uses linear algebra in a elementary but intricate manner. As our first work in this dissertation, we give an alternate proof of Ta-Shma's[60] construction using only repeated applications of the expander mixing lemma. Our proof is more combinatorial and arguably simpler compared to Ta-Shma's original analysis. Additionally, we showed that our techniques can be used to give an alternate proof of the expander hitting set lemma. In our second chapter, we proved a new mixing result that roughly says: in certain non-abelian groups, a random three term progression tuple behaves like a random tuple whose each element is sampled

independently randomly. Our result resolved a more than a decade old conjecture in additive combinatorics, by Gowers[43]. As an immediate consequence of our result, it follows that there are finite groups in which all three term progression free subsets are small. In our third work, we proposed a new code. This code is called locally testable, non-malleable code(LTNMC). Informally, LTNMCs come with efficient testing algorithms that gives the following guarantee: if any tampered codeword passes the test with high probability then it must be encoding either the original message or a statistically unrelated one. We gave a construction of such code in a popularly studied adversary model. Along the process, we also gave a new and efficient non-malleable code against affine tampering functions.

Bibliography

[1] Divesh Aggarwal. Affine-evasive sets modulo a prime. *Inf. Process. Lett.*, 115(2):382–385, 2015.

[2] Divesh Aggarwal, Shashank Agrawal, Divya Gupta, Hemanta K. Maji, Omkant Pandey, and Manoj Prabhakaran. Optimal computational split-state non-malleable codes. In Eyal Kushilevitz and Tal Malkin, editors, *Theory of Cryptography - 13th International Conference, TCC 2016-A, Tel Aviv, Israel, January 10-13, 2016, Proceedings, Part II*, volume 9563 of *Lecture Notes in Computer Science*, pages 393–417. Springer, 2016.

[3] Divesh Aggarwal, Yevgeniy Dodis, Tomasz Kazana, and Maciej Obremski. Non-malleable reductions and applications. In Rocco A. Servedio and Ronitt Rubinfeld, editors, *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing, STOC 2015, Portland, OR, USA, June 14-17, 2015*, pages 459–468. ACM, 2015.

[4] Divesh Aggarwal, Yevgeniy Dodis, and Shachar Lovett. Non-malleable codes from additive combinatorics. In David B. Shmoys, editor, *Symposium on Theory of Computing, STOC 2014, New York, NY, USA, May 31 - June 03, 2014*, pages 774–783. ACM, 2014.

[5] 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.

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

[7] Shashank Agrawal, Divya Gupta, Hemanta K. Maji, Omkant Pandey, and Manoj Prabhakaran. Explicit non-malleable codes against bit-wise tampering and permutations. In *Advances in Cryptology - CRYPTO 2015 - 35th Annual Cryptology Conference, Santa Barbara, CA, USA, August 16-20, 2015, Proceedings, Part I*, pages 538–557, 2015.

[8] Noga Alon. Explicit expanders of every degree and size. *Combinatorica*, pages 1–17, 2021.

[9] Noga Alon, Jehoshua Bruck, Joseph Naor, Moni Naor, and Ron M Roth. Construction of asymptotically good low-rate error-correcting codes through pseudo-random graphs. *IEEE Transactions on information theory*, 38(2):509–516, 1992.

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

[11] Sanjeev Arora and Shmuel Safra. Probabilistic checking of proofs; A new characterization of NP. In *33rd Annual Symposium on Foundations of Computer Science, Pittsburgh, Pennsylvania, USA, 24-27 October 1992*, pages 2–13. IEEE Computer Society, 1992.

[12] Sanjeev Arora and Shmuel Safra. Probabilistic checking of proofs: A new characterization of NP. *J. ACM*, 45(1):70–122, 1998.

[13] László Babai, Nikolay Nikolov, and László Pyber. Product growth and mixing in finite groups. In *Proc. 19th Annual ACM-SIAM Symp. on Discrete Algorithms (SODA)*, pages 248–257, 2008.

[14] László Babai and Vera T. Sós. Sidon sets in groups and induced subgraphs of Cayley graphs. *Eur. J. Comb.*, 6(2):101–114, 1985.

[15] Marshall Ball, Dana Dachman-Soled, Siyao Guo, Tal Malkin, and Li-Yang Tan. Non-malleable codes for small-depth circuits. In Mikkel Thorup, editor, *59th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2018, Paris, France, October 7-9, 2018*, pages 826–837. IEEE Computer Society, 2018.

[16] Avraham Ben-Aroya and Amnon Ta-Shma. A combinatorial construction of almost-ramanujan graphs using the zig-zag product. *SIAM Journal on Computing*, 40(2):267–290, 2011.

[17] Amey Bhangale, Irit Dinur, and Inbal Livni Navon. Cube vs. cube low degree test. In Christos H. Papadimitriou, editor, *8th Innovations in Theoretical Computer Science Conference, ITCS 2017, January 9-11, 2017, Berkeley, CA, USA*, volume 67 of *LIPIcs*, pages 40:1–40:31. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2017.

[18] Amey Bhangale and Subhash Khot. Optimal inapproximability of satisfiable k-LIN over non-abelian groups. In *Proc. 53rd ACM Symp. on Theory of Computing (STOC)*, pages 1615–1628, 2021.

[19] Nishanth Chandran, Vipul Goyal, Pratyay Mukherjee, Omkant Pandey, and Jalaj Upadhyay. Block-wise non-malleable codes. In Ioannis Chatzigiannakis, Michael Mitzenmacher, Yuval Rabani, and Davide Sangiorgi, editors, *43rd International Colloquium on Automata, Languages, and Programming, ICALP 2016, July 11-15, 2016, Rome, Italy*, volume 55 of *LIPIcs*, pages 31:1–31:14. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2016.

[20] Nishanth Chandran, Bhavana Kanukurthi, and Srinivasan Raghuraman. Information-theoretic local non-malleable codes and their applications. *IACR Cryptology ePrint Archive*, 2015:1056, 2015.

[21] 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.

[22] Eshan Chattopadhyay and Xin Li. Non-malleable codes and extractors for small-depth circuits, and affine functions. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 1171–1184. ACM, 2017.

[23] Eshan Chattopadhyay and David Zuckerman. Non-malleable codes against constant split-state tampering. In *55th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2014, Philadelphia, PA, USA, October 18-21, 2014*, pages 306–315. IEEE Computer Society, 2014.

[24] 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.

[25] Lijie Chen, Ce Jin, and R Ryan Williams. Hardness magnification for all sparse np languages. In *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1240–1255. IEEE, 2019.

[26] Sandro Coretti, Yevgeniy Dodis, Björn Tackmann, and Daniele Venturi. Non-malleable encryption: Simpler, shorter, stronger. In *Theory of Cryptography - 13th International Conference, TCC 2016-A, Tel Aviv, Israel, January 10-13, 2016, Proceedings, Part I*, pages 306–335, 2016.

[27] Dana Dachman-Soled, Feng-Hao Liu, Elaine Shi, and Hong-Sheng Zhou. Locally decodable and updatable non-malleable codes and their applications. In Yevgeniy Dodis and Jesper Buus Nielsen, editors, *Theory of Cryptography - 12th Theory of Cryptography Conference, TCC 2015, Warsaw, Poland, March 23-25, 2015, Proceedings, Part I*, volume 9014 of *Lecture Notes in Computer Science*, pages 427–450. Springer, 2015.

[28] Dana Dachman-Soled, Feng-Hao Liu, Elaine Shi, and Hong-Sheng Zhou. Locally decodable and updatable non-malleable codes and their applications. *J. Cryptology*, 33(1):319–355, 2020.

[29] Giuliana Davidoff, Peter Sarnak, and Alain Valette. *Elementary Number Theory, Group Theory and Ramanujan Graphs.* London Mathematical Society Student Texts. Cambridge University Press, 2003.

[30] Persi Diaconis. *Group representations in probability and statistics*, volume 11 of *IMS Lecture Notes Monogr. Ser.* Institute of Mathematical Statistics, 1998.

[31] Irit Dinur, Prahladh Harsha, Tali Kaufman, Inbal Livni Navon, and Amnon Ta-Shma. List decoding with double samplers. In Timothy M. Chan, editor, *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2019, San Diego, California, USA, January 6-9, 2019*, pages 2134–2153. SIAM, 2019.

[32] Irit Dinur, Prahladh Harsha, Tali Kaufman, and Noga Ron-Zewi. From local to robust testing via agreement testing. In Avrim Blum, editor, *10th Innovations in Theoretical Computer Science Conference, ITCS 2019, January 10-12, 2019, San Diego, California, USA*, volume 124 of *LIPIcs*, pages 29:1–29:18. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2019.

[33] 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.

[34] 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.

[35] Danny Dolev, Cynthia Dwork, and Moni Naor. Non-malleable cryptography (extended abstract). In *Proceedings of the 23rd Annual ACM Symposium on Theory of Computing, May 5-8, 1991, New Orleans, Louisiana, USA*, pages 542–552, 1991.

[36] Stefan Dziembowski, Tomasz Kazana, and Maciej Obremski. Non-malleable codes from two-source extractors. In Ran Canetti and Juan A. Garay, editors, *Advances in Cryptology - CRYPTO 2013 - 33rd Annual Cryptology Conference, Santa Barbara, CA, USA, August 18-22, 2013. Proceedings, Part II*, volume 8043 of *Lecture Notes in Computer Science*, pages 239–257. Springer, 2013.

[37] Stefan Dziembowski, Krzysztof Pietrzak, and Daniel Wichs. Non-malleable codes. *J. ACM*, 65(4):20:1–20:32, 2018.

[38] Uriel Feige and Joe Kilian. Zero knowledge and the chromatic number. In Steven Homer and Jin-Yi Cai, editors, *Proceedings of the Eleventh Annual IEEE Conference on Computational Complexity, Philadelphia, Pennsylvania, USA, May 24-27, 1996*, pages 278–287. IEEE Computer Society, 1996.

[39] E. N. Gilbert. A comparison of signalling alphabets. *The Bell System Technical Journal*, 31(3):504–522, 1952.

[40] Oded Goldreich. Short locally testable codes and proofs: A survey in two parts. In *Property testing*, pages 65–104. Springer, 2010.

[41] Oded Goldreich, Silvio Micali, and Avi Wigderson. Proofs that yield nothing but their validity and a methodology of cryptographic protocol design (extended abstract). In *27th Annual Symposium on Foundations of Computer Science, Toronto, Canada, 27-29 October 1986*, pages 174–187. IEEE Computer Society, 1986.

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

[43] William Timothy Gowers. Quasirandom groups. *Comb. Probab. Comput.*, 17(3):363–387, 2008.

[44] Vipul Goyal, Omkant Pandey, and Silas Richelson. Textbook non-malleable commitments. In *Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2016, Cambridge, MA, USA, June 18-21, 2016*, pages 1128–1141, 2016.

[45] Russell Impagliazzo, Valentine Kabanets, and Avi Wigderson. New direct-product testers and 2-query pcps. *SIAM J. Comput.*, 41(6):1722–1768, 2012.

[46] Joe Kilian and Moni Naor. On the complexity of statistical reasoning (extended abtract). In *Third Israel Symposium on Theory of Computing and Systems, ISTCS 1995, Tel Aviv, Israel, January 4-6, 1995, Proceedings*, pages 209–217. IEEE Computer Society, 1995.

[47] Joe Kilian, Erez Petrank, and Gábor Tardos. Probabilistically checkable proofs with zero knowledge. In *Proceedings of the Twenty-Ninth Annual ACM Symposium on the Theory of Computing, El Paso, Texas, USA, May 4-6, 1997*, pages 496–505, 1997.

[48] Xin Li. Improved two-source extractors, and affine extractors for polylogarithmic entropy. In Irit Dinur, editor, *IEEE 57th Annual Symposium on Foundations of Computer Science, FOCS 2016, 9-11 October 2016, Hyatt Regency, New Brunswick, New Jersey, USA*, pages 168–177. IEEE Computer Society, 2016.

[49] Martin W Liebeck and Aner Shalev. Character degrees and random walks in finite groups of Lie type. *Proc. Amer. Math. Soc.*, 90(1):61–86, 2004.

[50] Martin W Liebeck and Aner Shalev. Fuchsian groups, coverings of Riemann surfaces, subgroup growth, random quotients and random walks. *Journal of Algebra*, 276(2):552–601, 2004.

[51] Alexander Lubotzky, Ralph Phillips, and Peter Sarnak. Ramanujan graphs. *Combinatorica*, 8(3):261–277, 1988.

[52] Dana Moshkovitz. Low-degree test with polynomially small error. *Computational Complexity*, 26(3):531–582, 2017.

[53] Sarah Peluse. Mixing for three-term progressions in finite simple groups. *Math. Proc. Cambridge Philos. Soc.*, 165(2):279–286, 2018.

[54] Morris Plotkin. Binary codes with specified minimum distance. *IRE Transactions on Information Theory*, 6(4):445–450, 1960.

[55] Ran Raz and Shmuel Safra. A sub-constant error-probability low-degree test, and a sub-constant error-probability PCP characterization of NP. In Frank Thomson Leighton

and Peter W. Shor, editors, *Proceedings of the Twenty-Ninth Annual ACM Symposium on the Theory of Computing, El Paso, Texas, USA, May 4-6, 1997*, pages 475–484. ACM, 1997.

[56] Ronitt Rubinfeld and Madhu Sudan. Robust characterizations of polynomials with applications to program testing. *SIAM J. Comput.*, 25(2):252–271, 1996.

[57] Adi Shamir. How to share a secret. *Communications of the ACM*, 22(11):612–613, 1979.

[58] Madhu Sudan. Decoding of reed solomon codes beyond the error-correction bound. *J. Complexity*, 13(1):180–193, 1997.

[59] Madhu Sudan, Luca Trevisan, and Salil Vadhan. Pseudorandom generators without the xor lemma. *Journal of Computer and System Sciences*, 62(2):236–266, 2001.

[60] 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.

[61] Terrence Tao. Mixing for progressions in nonabelian groups. *Forum of Mathematics, Sigma*, 1:e2, 2013.

[62] R. R. Varshamov. Estimate of the number of signals in error correcting codes. *Docklady Akad. Nauk, S.S.S.R.*, 117:739–741, 1957.

[63] Avi Wigderson and David Zuckerman. Expanders that beat the eigenvalue bound: explicit construction and applications. In S. Rao Kosaraju, David S. Johnson, and Alok Aggarwal, editors, *Proceedings of the Twenty-Fifth Annual ACM Symposium on Theory of Computing, May 16-18, 1993, San Diego, CA, USA*, pages 245–251. ACM, 1993.

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

Appendix A

A.1 Sampler Graph Preliminaries

A.1.1 Basic definitions and Facts

Definition 13 (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 $b \sim B$, $a \sim A(b)$.*[1]

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 14 (Sampler Graph [64]) *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 $|B'| = \lambda \cdot |B|$,*

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

[1] 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 13.

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 [33]. We use them implicitly in this work as well. The construction in [33] 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 5 (Properties of Samplers) *Suppose A/B is $\eta-$biregular and $(\varepsilon, \delta)-$sampling. We have the following.*

1. *For any $\rho > 0$ and $f : B \to [0, 1]$,*

$$\Pr_{a\sim A}\left[\left|\mathbb{E}_{b\sim B(a)}\big[f(b)\big] - \mathbb{E}_{b\sim B}\big[f(b)\big]\right| > \varepsilon + 2\rho\right] \leq \delta/\rho.$$

2. *For any $\rho > 0$, B/A is $\big(\rho, 2(\varepsilon + \delta + \eta)/\rho\big)-$sampling.*

3. *For any $B' \subset B$ of size $|B'| = \lambda \cdot |B|$ with $\lambda > \varepsilon$,*

$$\Delta\Big(\big\{(a, b) : \substack{b\sim B' \\ a\sim A(b)}\big\}, \big\{(a, b) : \substack{a\sim A \\ b\sim B'(a)}\big\}\Big) \leq \delta + \eta/\varepsilon,$$

where $B'(a)$ denotes the distribution which draws $b \sim B(a)$ and outputs if $b \in B'$, else resamples (or if $B(a) \cap B' = \emptyset$, $B'(a)$ outputs an arbitrary $b \in B$).

The facts above are all well-known. See, for example, [64, 45, 17] for proofs of points 1, 2,

and 3, respectively.

Fact 6 (Extending Sampling via Biregularity) *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 $\delta \geq \delta'/\varepsilon'$.*

Proof. Assume $A/B/C$ are such that for all $a \in A$, $B(a)/C(a)$ is $\eta-$biregular, and also that $C(a,b) = C(b)$. Let $C' \subset C$ be a subset of size $|C'| = \lambda \cdot |C|$. By $\eta-$biregularity,

$$\left|\Pr_{c \sim C(a)}(c \in C') - \lambda\right| \leq \left|\mathbb{E}_{b \sim B(a)}\left[\Pr_{c \sim C(b)}(c \in C')\right] - \lambda\right| + \eta$$

holds for all $a \in A$. Now, let $\mathsf{val} := \Pr_{a \sim A}\left[\left|\Pr_{c \sim C(a)}(c \in C') - \lambda\right| > \varepsilon\right]$ be the quantity we have to bound. For the first point we have

$$\mathsf{val} \leq \varepsilon^{-1} \cdot \left(\mathbb{E}_{\substack{a \sim A \\ b \sim B(a)}}\left[\left|\Pr_{c \sim C(b)}(c \in C') - \lambda\right|\right] + \eta\right) \leq \varepsilon^{-1} \cdot (2\eta + \varepsilon' + \delta'),$$

by Markov's inequality, the $\eta-$biregularity of A/B and the $(\varepsilon', \delta')-$sampling of B/C. For the second point we have

$$\mathsf{val} \leq \Pr_{a \sim A}\left[\left|\mathbb{E}_{b \sim B(a)}\left[\lambda(b)\right] - \mathbb{E}_{b \sim B}\left[\lambda(b)\right]\right| > \varepsilon - 2\eta \geq 3\varepsilon'\right] \leq \delta'/\varepsilon',$$

where $\lambda(b) := \Pr_{c \sim C(b)}(c \in C')$. We have used the $\eta-$biregularity of B/C to say that $\mathbb{E}_{b \sim B}\big[\lambda(b)\big]$ is in $\lambda \pm \eta$, and the $(\varepsilon', \delta')-$sampling of A/B combined with the first point of Fact 5. ■

Fact 7 (Replacement Product) *Let $\varepsilon, \varepsilon', \delta, \delta' > 0$ be such that $\delta \cdot (\varepsilon - 5\varepsilon') \geq 2\delta'/\varepsilon'$. Suppose $A/B/C$ is such that:*

- *A/C, B/C and $B(a)/C(a)$ are $0-$biregular for all $a \in A$; and*
- *A/C and $A(c)/B(c)$ are $(\varepsilon', \delta')-$sampling for all $c \in C$.*

Then A/B is $(\varepsilon, \delta)-$sampling.

The replacement product was originally proved in [63] in the context of seeded randomness extractors (which are equivalent to sampler graphs). We give the proof ported over to the language of samplers in Appendix A.4 for completeness.

A.1.2 Why Samplers Play a Role

Here we briefly discuss how sampler graphs serve as an important component in our analysis. We begin by recalling the 'plane vs plane' low degree testing model from PCP literature [55, 17]. In this model, a test algorithm gets oracle access to a 'planes' table where to each plane, $\mathsf{a} \in \mathsf{A}$, the table contains a polynomial, α, defined on that plane. Then the test algorithm's task is to decide if the table is close to any global low degree polynomial Φ. The final step is then to prove an agreement theorem that says if the test passes with good probability, then there exists a polynomial that agrees with the table on many planes. In literature, these agreement theorems are proven using essentially

two ingredients: sampling properties of planes and facts about low degree polynomials. Now, its easy to see that our tampering and testing model is very similar to the 'plane vs plane' model. The only difference is that in our model as we are looking at coordinate-wise tampering $f_a(\alpha) = \tilde{\alpha}$, we have a 'plane×polynomial table' where to (a, α) the table contains a polynomial $\tilde{\alpha}$. Thus, to prove an agreement theorem in our setting, we wind up using sampling of 'planes×polynomials' [see section A.1.3 below] and the same facts about polynomials.

A.1.3 Incidence × Agreement Samplers

Sampler graphs play a big role in the proofs. In this section we list all the graphs whose sampling will be used, and various properties of sampler graphs. All of the graphs are what we call "incidence × agreement" graphs, such as $\overline{\mathsf{A}}/\overline{\mathsf{C}}$ from last section. We begin with some notation.

Notation. Recall $\mathbb{F}$ is a finite field, $k \geq 4$, $d \geq 2$, A is the set of 3−planes in $\mathbb{F}^k$, $\mathsf{C} = \mathbb{F}^k$, Γ and Γ_{A} are the sets of $k-$variate and $3-$variate polynomials of degree at most d over $\mathbb{F}$, respectively, $\Gamma_{\mathsf{C}} = \mathbb{F}$. This defines an incidence × agreement bipartite graph $\overline{\mathsf{A}}/\overline{\mathsf{C}}$ where $\overline{\mathsf{A}} = \mathsf{A} \times \Gamma_{\mathsf{A}}$, $\overline{\mathsf{C}} = \mathsf{C} \times \Gamma_{\mathsf{C}}$ and the edge relation is "incidence × agreement": $\overline{\mathsf{a}} = (\mathsf{a}, \alpha) \sim (\mathsf{c}, \gamma) = \overline{\mathsf{c}}$ iff $\mathsf{c} \in \mathsf{a}$ and $\alpha|_{\mathsf{c}} = \gamma$. For $r = 1, 2$, let B_r denote the set of affine $r-$dimensional planes in $\mathbb{F}^k$, let Γ_{B_r} be the set of $r-$variate polynomials of degree at most d over $\mathbb{F}$, and let $\overline{\mathsf{B}}_r = \mathsf{B}_r \times \Gamma_{\mathsf{B}_r}$. At various points during the proof, we will use that $\overline{\mathsf{A}}/\overline{\mathsf{B}}_r/\overline{\mathsf{C}}$ is a double sampler. The incidence × agreement edge relation extends naturally to $\overline{\mathsf{A}}/\overline{\mathsf{B}}_r$, $\overline{\mathsf{B}}_r/\overline{\mathsf{C}}$, and $\overline{\mathsf{B}}_2/\overline{\mathsf{B}}_1$. For example, if $\overline{\mathsf{a}} = (\mathsf{a}, \alpha) \in \overline{\mathsf{A}}$ and $\overline{\mathsf{b}} = (\mathsf{b}, \beta) \in \overline{\mathsf{B}}_2$, then $\overline{\mathsf{a}} \sim \overline{\mathsf{b}}$ iff $\mathsf{b} \subset \mathsf{a}$ and $\alpha|_{\mathsf{b}} = \beta$.

We begin by listing the incidence $\times$ agreement samplers we will need in the remainder of the paper and proving they are sampling. In the claim statement below, $\overline{\mathsf{A}}(\overline{\mathsf{c}})$, for $\overline{\mathsf{c}} \in \overline{\mathsf{C}}$, denotes the set of $\overline{\mathsf{a}} \in \overline{\mathsf{A}}$ such that $\overline{\mathsf{a}} \sim \overline{\mathsf{c}}$. In the proof which follows, we use $\overline{\mathsf{A}}(\overline{\mathsf{c}})$ to mean either this set, or the uniform distribution on this set; in all cases, our intention should be clear from the context.

Claim 11 *The following graphs are all $\mathcal{O}\big(|\mathbb{F}|^{-1}\big)-$biregular and $\big(12 \cdot |\mathbb{F}|^{-1/15}, |\mathbb{F}|^{-1/15}\big)-$sampling:*

(1) $\overline{\mathsf{B}}_1/\overline{\mathsf{C}}$ (2) $\overline{\mathsf{A}}(\overline{\mathsf{c}})/\overline{\mathsf{B}}_2(\overline{\mathsf{c}})\ \forall\ \overline{\mathsf{c}} \in \overline{\mathsf{C}}$ (3) $\overline{\mathsf{A}}/\overline{\mathsf{C}}$ (4) $\overline{\mathsf{A}}(\overline{\mathsf{c}}, \overline{\mathsf{c}}')/\overline{\mathsf{C}}$

(5) $\overline{\mathsf{A}}/\overline{\mathsf{C}}^2$ (6) $\overline{\mathsf{A}}(\overline{\mathsf{c}})/\overline{\mathsf{C}}^2\ \forall\ \overline{\mathsf{c}} \in \overline{\mathsf{C}}$ (7) $\mathsf{A} \times \Gamma/\overline{\mathsf{C}}$ (8) $\overline{\mathsf{B}}_2(\overline{\mathsf{c}})/\overline{\mathsf{C}}$

(9) $\overline{\mathsf{A}}(\overline{\mathsf{c}})/\overline{\mathsf{B}}_1(\overline{\mathsf{c}}) \times \overline{\mathsf{B}}_1(\overline{\mathsf{c}})$ (10) $\mathsf{A} \times \Gamma/\overline{\mathsf{B}}_1$ (11) $\overline{\mathsf{A}}(\overline{\mathsf{b}})/\overline{\mathsf{C}}$

Proof. It is easy to see that all of the graphs in the Claim statement are $\mathcal{O}\big(|\mathbb{F}|^{-1}\big)-$biregular, as per Definition 13. By symmetry, graphs (1), (2), (3), (5), (7), (10) are actually $0-$biregular. The others have a slight error introduced by the fact, for example, that the distribution which draws $\overline{\mathsf{a}} \sim \overline{\mathsf{A}}(\overline{\mathsf{c}})$ and outputs a random element of $\overline{\mathsf{C}}(\overline{\mathsf{a}})$ is more likely to output $\overline{\mathsf{c}}$ than $\overline{\mathsf{c}}' \neq \overline{\mathsf{c}}$. However, an easy calculation shows that the statistical distance between the required distributions is $\mathcal{O}\big(|\mathbb{F}|^{-1}\big)$; the same is true for all examples in the list. The rest of the proof is divided into two stages. First, we use a pairwise independence argument to show that $\overline{\mathsf{B}}_1/\overline{\mathsf{C}}$, $\overline{\mathsf{B}}_2/\overline{\mathsf{C}}$, $\overline{\mathsf{A}}(\overline{\mathsf{b}}_1)/\overline{\mathsf{B}}_2(\overline{\mathsf{b}}_1)$ for all $\overline{\mathsf{b}}_1 \in \overline{\mathsf{B}}_1$ and $\overline{\mathsf{B}}_2(\overline{\mathsf{c}})/\overline{\mathsf{B}}_1(\overline{\mathsf{c}})$, $\overline{\mathsf{A}}(\overline{\mathsf{c}})/\overline{\mathsf{B}}_1(\overline{\mathsf{c}})$ for all $\overline{\mathsf{c}} \in \overline{\mathsf{C}}$ are $(|\mathbb{F}|^{-1/5}, |\mathbb{F}|^{-3/5})-$sampling. Then we reduce the sampling of every graph above to the sampling of these five.

We phrase the pairwise independence argument for a generic bipartite graph A/B. The key feature we need involves a set X which parametrizes the neighborhoods $B(a)$ for

all $a \in A$. Given $x \in X$ and $a \in A$, we write the $x-$th neighbor of a as $a(x) \in B$, so X parametrizes neighborhoods as $B(a) = \{a(x) : x \in X\}$ for all $a \in A$. The property we require is that for all $x_1 \neq x_2 \in X$, the random variable $\big(a(x_1), a(x_2)\big)$ (randomness over $a \sim A$) is uniform on B^2. For $\overline{\mathsf{B}}_1/\overline{\mathsf{C}}$, $X = \mathbb{F}$ since $\overline{\mathsf{C}}(\overline{\mathsf{b}})$ is parametrized by the points on the line b. Likewise, for $\overline{\mathsf{B}}_2/\overline{\mathsf{C}}$, $X = \mathbb{F}^2$. For $\overline{\mathsf{A}}(\overline{\mathsf{b}}_1)/\overline{\mathsf{B}}_2(\overline{\mathsf{b}}_1)$,the neighborhood $\overline{\mathsf{B}}_2(\overline{\mathsf{b}}_1, \overline{\mathsf{a}})$ is paramterized by all possible planes in a through b_1, so we have $|X| = |\mathbb{F}| + 1$. For $\overline{\mathsf{B}}_2(\overline{\mathsf{c}})/\overline{\mathsf{B}}_1(\overline{\mathsf{c}})$, $X = \mathbb{F} \cup \{\infty\}$, since $\overline{\mathsf{B}}_1(\overline{\mathsf{c}}, \overline{\mathsf{b}}_2)$ is parametrized by all possible slopes of a line in b_2 through c. Finally, for $\overline{\mathsf{A}}(\overline{\mathsf{c}})/\overline{\mathsf{B}}_1(\overline{\mathsf{c}})$ the neighborhood $\overline{\mathsf{B}}_1(\overline{\mathsf{c}}, \overline{\mathsf{a}})$ is paramterized by all possible lines in a through c, so we have $|X| = |\mathbb{F}|^2 + |\mathbb{F}| + 1$. In all cases, independence follows from the fact that for every $b_1 \in B$, the distribution which draws $a \sim A(b_1)$ and outputs $b_2 \sim B(a) \setminus \{b_1\}$ is the uniform distribution on B.

So now, let A/B be a bipartite graph which satisfies the pairwise independent parametrized neighborhood property described above. Let $B' \subset B$ be a subset of size $|B'| = \lambda \cdot |B|$. For $b \in B$, let $\mathbb{1}_{B'}(b)$ indicate whether $b \in B'$ or not, and let $\hat{\mathbb{1}}_{B'}(b) := \mathbb{1}_{B'}(b) - \lambda$. Note $\mathbb{E}_{b\sim B}\big[\hat{\mathbb{1}}_{B'}(b)\big] = 0$. Finally, define $f : A \to [0,1]$ by $f(a) := \mathbb{E}_{b\sim B(a)}\big[\hat{\mathbb{1}}_{B'}(b)\big]$. We will show $\mathbb{E}_{a\sim A}\big[f(a)^2\big] \leq |\mathbb{F}|^{-1}$. This suffices by Markov's inequality:

$$\Pr_{a\sim A}\left[\left|\Pr_{b\sim B(a)}\big(b \in B'\big) - \lambda\right| > |\mathbb{F}|^{-1/5}\right] \leq \Pr_{a\sim A}\Big[f(a)^2 > |\mathbb{F}|^{-2/5}\Big] \leq |\mathbb{F}|^{2/5} \cdot \mathbb{E}_{a\sim A}\big[f(a)^2\big].$$

We use the pairwise independence property to conclude:

$$\begin{aligned}
\mathbb{E}_{a\sim A}\big[f(a)^2\big] &= \mathbb{E}_{a\sim A}\Big[\mathbb{E}_{x_1,x_2\sim X}\big[\hat{\mathbb{1}}_{B'}\big(a(x_1)\big) \cdot \hat{\mathbb{1}}_{B'}\big(a(x_2)\big)\Big] \\
&\leq \frac{1}{|X|} + \mathbb{E}_{b_1,b_2\sim B}\Big[\hat{\mathbb{1}}_{B'}(b_1) \cdot \hat{\mathbb{1}}_{B'}(b_2)\Big] = \frac{1}{|X|}.
\end{aligned}$$

For the reductions in the second phase, we use the generic facts about samplers stated in Section A.1.1. Since $\overline{\mathsf{B}}_2/\overline{\mathsf{C}}$ and $\overline{\mathsf{B}}_2(\overline{\mathsf{c}})/\overline{\mathsf{B}}_1(\overline{\mathsf{c}})$ for all $\overline{\mathsf{c}} \in \overline{\mathsf{C}}$ are each $\left(|\mathbb{F}|^{-1/5}, |\mathbb{F}|^{-3/5}\right)-$sampling, $\overline{\mathsf{B}}_1/\overline{\mathsf{C}}$ and $\overline{\mathsf{B}}_2/\overline{\mathsf{B}}_1$ are both $\left(7\cdot|\mathbb{F}|^{-1/5}, |\mathbb{F}|^{-1/5}\right)-$sampling (we have already shown sampling of $\overline{\mathsf{B}}_1/\overline{\mathsf{C}}$ with better parameters, sampling of $\overline{\mathsf{B}}_2/\overline{\mathsf{B}}_1$ follows from Fact 7. We have also shown that $\overline{\mathsf{A}}(\overline{\mathsf{b}}_1)/\overline{\mathsf{B}}_2(\overline{\mathsf{b}}_1)$ for all $\overline{\mathsf{b}}_1 \in \overline{\mathsf{B}}_1$ and $\overline{\mathsf{A}}(\overline{\mathsf{c}})/\overline{\mathsf{B}}_1(\overline{\mathsf{c}})$ for all $\overline{\mathsf{c}} \in \overline{\mathsf{C}}$ are both $\left(|\mathbb{F}|^{-1/5}, |\mathbb{F}|^{-3/5}\right)-$sampling. This fact combined with Fact 7 proves sampling of $\overline{\mathsf{A}}(\overline{\mathsf{c}})/\overline{\mathsf{B}}_2(\overline{\mathsf{c}})$ for all $\overline{\mathsf{c}} \in \overline{\mathsf{C}}$. The first point of Fact 6 says that any time we have Z such that $Z/\overline{\mathsf{B}}_1$ or $Z/\overline{\mathsf{B}}_2$ is $\mathcal{O}\left(|\mathbb{F}|^{-1}\right)-$biregular, then $Z/\overline{\mathsf{C}}$ or $Z/\overline{\mathsf{B}}_1$ is $\left(3\cdot|\mathbb{F}|^{-1/15}, 3\cdot|\mathbb{F}|^{-2/15}\right)-$sampling. This proves the sampling of all graphs except for (5), (6) and (9): $\overline{\mathsf{A}}/\overline{\mathsf{C}}^2$ and $\overline{\mathsf{A}}(\overline{\mathsf{c}})/\overline{\mathsf{C}}^2$ for all $\overline{\mathsf{c}} \in \overline{\mathsf{C}}$, so it remains to prove sampling of these. Note $\overline{\mathsf{A}}(\overline{\mathsf{c}})/\overline{\mathsf{B}}_1$ for all $\overline{\mathsf{c}} \in \overline{\mathsf{C}}$ and $\overline{\mathsf{A}}/\overline{\mathsf{B}}_1$ are $\left(3\cdot|\mathbb{F}|^{-1/15}, 3\cdot|\mathbb{F}|^{-2/15}\right)$-samplers, since $\overline{\mathsf{A}}(\overline{\mathsf{c}})/\overline{\mathsf{B}}_2$ and $\overline{\mathsf{A}}/\overline{\mathsf{B}}_2$ are $\mathcal{O}\left(|\mathbb{F}|^{-1}\right)-$biregular. Thus we can use the second point Fact 6 to get $\left(12\cdot|\mathbb{F}|^{-1/15}, |\mathbb{F}|^{-1/15}\right)-$sampling of graphs (5) and (6) because $\overline{\mathsf{B}}_1/\overline{\mathsf{C}}^2$ is $\mathcal{O}\left(|\mathbb{F}|^{-1}\right)-$biregular. Sampling of $\overline{\mathsf{A}}(\overline{\mathsf{c}})/\overline{\mathsf{B}}_1(\overline{\mathsf{c}}) \times \overline{\mathsf{B}}_1(\overline{\mathsf{c}})\ \forall\ \overline{\mathsf{c}} \in \overline{\mathsf{C}}$ follows similarly. ■

Notational Conventions and Example Use. Our proofs in the next sections rely heavily, and often implicitly, on the fact that the graphs of Claim 11 are samplers, and on the properties of sampler graphs stated in Fact 5. To facilitate readability, from here on, we reserve the quantity $\delta > 0$ for the loss introduced any time a sampling argument is used. As an example of how this looks in the body of the paper, let $\overline{\mathsf{C}}' \subset \overline{\mathsf{C}}$ be a set with $|\overline{\mathsf{C}}'| \geq \lambda \cdot |\overline{\mathsf{C}}|$, and let $\mathbf{E}$ be some event. Then we might deduce: $\mathbb{E}_{\overline{\mathsf{c}},\overline{\mathsf{c}}' \sim \overline{\mathsf{C}}'}\left[\Pr_{\overline{\mathsf{a}} \sim \overline{\mathsf{A}}(\overline{\mathsf{c}},\overline{\mathsf{c}}')}(\mathbf{E})\right] \geq \mathbb{E}_{\overline{\mathsf{a}} \sim \overline{\mathsf{A}}}\left[\Pr_{\overline{\mathsf{c}},\overline{\mathsf{c}}' \sim \overline{\mathsf{C}}'(\overline{\mathsf{a}})}(\mathbf{E})\right] - \delta$, "because of the sampling of $\overline{\mathsf{A}}/\overline{\mathsf{C}}^2$." Formally, we are using the third point of Fact 5, the fact that $\overline{\mathsf{A}}/\overline{\mathsf{C}}^2$ is $\eta'-$biregular, $(\varepsilon', \delta')-$sampling with $\lambda > \varepsilon'$ and

that $\delta \geq \delta' + \eta'/\varepsilon'$.

Setting the Sampling Parameter. In the example use mentioned above, $\eta' = \mathcal{O}\big(|\mathbb{F}|^{-1}\big)$ and $\varepsilon', \delta' = \mathcal{O}\big(|\mathbb{F}|^{-1/15}\big)$. Thus, $\delta = \mathcal{O}\big(|\mathbb{F}|^{-1/15}\big)$ is sufficient for $\delta \geq \delta' + \eta'/\varepsilon'$ to hold. In general, each sampler property use will put a lower bound on δ, and so we simply set δ large enough so that they all hold. Explicitly, $\delta = 3 \cdot |\mathbb{F}|^{-1/60}$ is sufficient for our purposes.

We conclude this section with a claim listing two sampler-based facts which will be useful in the calculations in the next section.

Claim 12 *Let the notations be as above, and let* $\delta = 3 \cdot |\mathbb{F}|^{-1/60}$ *and* $\eta = \mathcal{O}\big(|\mathbb{F}|^{-1}\big)$. *Let* $\overline{\mathsf{C}}' \subset \overline{\mathsf{C}}$ *be a subset of size* $|\overline{\mathsf{C}}'|/|\overline{\mathsf{C}}| \geq 12 \cdot |\mathbb{F}|^{-1/15}$. *We have the following.*

1. $$\left\{ (\overline{\mathsf{c}}, \overline{\mathsf{b}}, \overline{\mathsf{c}}') \;\middle|\; \begin{array}{c} \overline{\mathsf{a}} \sim \overline{\mathsf{A}} \\ \overline{\mathsf{c}} \sim \overline{\mathsf{C}}(\overline{\mathsf{a}}) \\ \overline{\mathsf{c}}' \sim \overline{\mathsf{C}}'(\overline{\mathsf{a}}) \\ \mathsf{b} \sim \mathsf{B}_2(\mathsf{a}, \mathsf{c}, \mathsf{c}') \end{array} \right\} \approx_\delta \left\{ (\overline{\mathsf{c}}, \overline{\mathsf{b}}, \overline{\mathsf{c}}') \;\middle|\; \begin{array}{c} \overline{\mathsf{c}} \sim \overline{\mathsf{C}} \\ \overline{\mathsf{b}} \sim \overline{\mathsf{B}}_2(\overline{\mathsf{c}}) \\ \overline{\mathsf{c}}' \sim \overline{\mathsf{C}}'(\overline{\mathsf{b}}) \end{array} \right\},$$

   where in the first distribution $\overline{\mathsf{b}} = (\mathsf{b}, \alpha|_{\mathsf{b}})$, *where* $\overline{\mathsf{a}} = (\mathsf{a}, \alpha)$.

2. $$\left\{ (\overline{\mathsf{a}}, \overline{\mathsf{b}}, \overline{\mathsf{c}}') \;\middle|\; \begin{array}{c} \overline{\mathsf{a}} \sim \overline{\mathsf{A}} \\ \overline{\mathsf{c}} \sim \overline{\mathsf{C}}(\overline{\mathsf{a}}) \\ \overline{\mathsf{c}}' \sim \overline{\mathsf{C}}'(\overline{\mathsf{a}}) \\ \mathsf{b} \sim \mathsf{B}_2(\mathsf{c}, \mathsf{c}') \end{array} \right\} \approx_\delta \left\{ (\overline{\mathsf{a}}, \overline{\mathsf{b}}, \overline{\mathsf{c}}') \;\middle|\; \begin{array}{c} \overline{\mathsf{c}}' \sim \overline{\mathsf{C}}' \\ \overline{\mathsf{b}} \sim \overline{\mathsf{B}}_2(\overline{\mathsf{c}}') \\ \overline{\mathsf{a}} \sim \overline{\mathsf{A}}(\overline{\mathsf{b}}) \end{array} \right\},$$

   where in the first distribution $\overline{\mathsf{b}} = \overline{\mathsf{a}}|_{\mathsf{b}}$.

In both (1) and (2) above, $\approx_\delta$ denotes that the two distributions are within statistical distance δ of one another.

Proof. For the first part, we have

$$\left\{\begin{array}{c}\overline{\mathsf{a}} \sim \overline{\mathsf{A}} \\ \overline{\mathsf{c}} \sim \overline{\mathsf{C}}(\overline{\mathsf{a}}) \\ \overline{\mathsf{c}}' \sim \overline{\mathsf{C}}'(\overline{\mathsf{a}})\end{array}\right\} \approx_{\delta/3} \left\{\begin{array}{c}\overline{\mathsf{c}}' \sim \overline{\mathsf{C}}' \\ \overline{\mathsf{a}} \sim \overline{\mathsf{A}}(\overline{\mathsf{c}}') \\ \overline{\mathsf{c}} \sim \overline{\mathsf{C}}(\overline{\mathsf{a}})\end{array}\right\} \approx_{\eta} \left\{\begin{array}{c}\overline{\mathsf{c}}' \sim \overline{\mathsf{C}}' \\ \overline{\mathsf{b}} \sim \overline{\mathsf{B}}_2(\overline{\mathsf{c}}') \\ \overline{\mathsf{c}} \sim \overline{\mathsf{C}}(\overline{\mathsf{b}})\end{array}\right\} \approx_{\delta/3} \left\{\begin{array}{c}\overline{\mathsf{b}} \sim \overline{\mathsf{B}}_2 \\ \overline{\mathsf{c}}' \sim \overline{\mathsf{C}}'(\overline{\mathsf{b}}) \\ \overline{\mathsf{c}} \sim \overline{\mathsf{C}}(\overline{\mathsf{b}})\end{array}\right\},$$

where each distribution outputs $(\overline{\mathsf{c}}, \overline{\mathsf{b}}, \overline{\mathsf{c}}')$ and where $\overline{\mathsf{b}} = \overline{\mathsf{a}}|_{\mathsf{b}}$ for $\mathsf{b} \sim \mathsf{B}_2(\mathsf{a}, \mathsf{c}, \mathsf{c}')$ is implied in the first two distributions. The first relation follows from sampling of $\overline{\mathsf{A}}/\overline{\mathsf{C}}$; the second follows from the $\eta-$biregularity of $\mathsf{B}_2(\overline{\mathsf{a}}, \overline{\mathsf{c}}')/\overline{\mathsf{C}}(\overline{\mathsf{a}})$ for all $\overline{\mathsf{a}} \in \overline{\mathsf{A}}$ and $\overline{\mathsf{c}}' \in \overline{\mathsf{C}}(\overline{\mathsf{a}})$, and the $0-$biregularity of $\overline{\mathsf{A}}(\overline{\mathsf{c}}')/\overline{\mathsf{B}}(\overline{\mathsf{c}}')$ for all $\overline{\mathsf{c}}' \in \overline{\mathsf{C}}$; the third follows from the sampling of $\overline{\mathsf{B}}_2/\overline{\mathsf{C}}$. Finally, the last distribution is identical to the desired distribution on the right of point 1 because of the $0-$biregularity of $\overline{\mathsf{B}}_2/\overline{\mathsf{C}}$. We work similarly for the second point:

$$\left\{\begin{array}{c}\overline{\mathsf{a}} \sim \overline{\mathsf{A}} \\ \overline{\mathsf{c}} \sim \overline{\mathsf{C}}(\overline{\mathsf{a}}) \\ \overline{\mathsf{c}}' \sim \overline{\mathsf{C}}'(\overline{\mathsf{a}})\end{array}\right\} \approx_{\delta/2} \left\{\begin{array}{c}\overline{\mathsf{c}}' \sim \overline{\mathsf{C}}' \\ \overline{\mathsf{a}} \sim \overline{\mathsf{A}}(\overline{\mathsf{c}}') \\ \overline{\mathsf{c}} \sim \overline{\mathsf{C}}(\overline{\mathsf{a}})\end{array}\right\} \approx_{\eta} \left\{\begin{array}{c}\overline{\mathsf{c}}' \sim \overline{\mathsf{C}}' \\ \overline{\mathsf{a}} \sim \overline{\mathsf{A}}(\overline{\mathsf{c}}') \\ \overline{\mathsf{b}} \sim \overline{\mathsf{B}}_2(\overline{\mathsf{a}}, \overline{\mathsf{c}}')\end{array}\right\} \equiv \left\{\begin{array}{c}\overline{\mathsf{c}}' \sim \overline{\mathsf{C}}' \\ \overline{\mathsf{b}} \sim \overline{\mathsf{B}}_2(\overline{\mathsf{c}}') \\ \overline{\mathsf{a}} \sim \overline{\mathsf{A}}(\overline{\mathsf{b}})\end{array}\right\},$$

where each distribution outputs $(\overline{\mathsf{a}}, \overline{\mathsf{b}}, \overline{\mathsf{c}}')$ and where $\overline{\mathsf{b}} = \overline{\mathsf{a}}|_{\mathsf{b}}$ (as above, $\mathsf{b} \sim \mathsf{B}_2(\mathsf{a}, \mathsf{c}, \mathsf{c}')$ is implicit in the first two distributions). We have used the sampling of $\overline{\mathsf{A}}/\overline{\mathsf{C}}$, $\eta-$biregularity of $\overline{\mathsf{B}}_2(\overline{\mathsf{a}}, \overline{\mathsf{c}}')/\overline{\mathsf{C}}(\overline{\mathsf{a}})$ for all $\overline{\mathsf{a}} \in \overline{\mathsf{A}}$ and $\overline{\mathsf{c}}' \in \overline{\mathsf{C}}(\overline{\mathsf{a}})$, and $0-$biregularity of $\overline{\mathsf{A}}(\overline{\mathsf{c}}')/\overline{\mathsf{B}}_2(\overline{\mathsf{c}}')$ for all $\overline{\mathsf{c}}' \in \overline{\mathsf{C}}$.

■

A.2 Missing Proofs

We first recall the theorems.

Theorem 5 (Restated). (Non-Malleability of the Outer Code) *Fix* $\varepsilon = |\mathbb{F}|^{-\mathcal{O}(1)}$ *and* $l = 4/\varepsilon$. . *Suppose* $\{\mathsf{f}_\mathsf{a}\}_\mathsf{a} \in \mathcal{F}$ *is such that*

$$\Pr_{\Phi,(\mathsf{c},\mathsf{a},\mathsf{a}')}\Big[\mathsf{Test}\big(\{(\mathsf{a},\tilde{\alpha})\}_\mathsf{a}\big) = 1\Big] \geq \varepsilon,$$

where the probability is over $\Phi \sim \Gamma$, *and* $\mathsf{c} \sim \mathsf{C}$, $\mathsf{a},\mathsf{a}' \sim \mathsf{A}(\mathsf{c})$,[2] *and where* $\big\{(\mathsf{a},\alpha)\big\}_\mathsf{a} = \big\{(\mathsf{a},\Phi|_\mathsf{a})\big\}_\mathsf{a}$, *and* $\big\{(\mathsf{a},\tilde{\alpha})\big\}_\mathsf{a} = \big\{(\mathsf{a},\mathsf{f}_\mathsf{a}(\alpha))\big\}_\mathsf{a}$. *Then there exists a list* $\mathsf{L}_{\{\mathsf{f}_\mathsf{a}\}} = \{\mathsf{T}^{(1)},\ldots,\mathsf{T}^{(l)}\}$ *of affine maps* $\mathsf{T}^{(j)} : \Gamma \to \Gamma$ *of size at most* $|\mathsf{L}_{\{\mathsf{f}_\mathsf{a}\}}| \leq l$ *such that*

$$\Pr_{\Phi,(\mathsf{c},\mathsf{a},\mathsf{a}')}\Big[\tilde{\alpha}|_\mathsf{c} = \tilde{\alpha'}|_\mathsf{c} \;\&\; \tilde{\alpha} \notin \{\mathsf{T}^{(j)}(\Phi)|_\mathsf{a}\} : \mathsf{T}^{(j)} \in \mathsf{L}_{\{\mathsf{f}_\mathsf{a}\}}\Big] = \mathcal{O}(\varepsilon)$$

Theorem 4 (Restated). *Fix* $\varepsilon = |\mathbb{F}|^{-\mathcal{O}(1)}$. *Suppose* $\{\mathsf{f}_\mathsf{a}\}_{\mathsf{a}\in\mathsf{A}} \in \mathcal{F}$ *is such that*

$$\Pr_{\Phi,(\mathsf{c},\mathsf{a},\mathsf{a}')}\Big[\mathsf{Test}\big(\{(\mathsf{a},\tilde{\alpha})\}_\mathsf{a}\big) = 1\Big] \geq \varepsilon,$$

where the probability is over $\Phi \sim \Gamma$, *and* $\mathsf{c} \sim \mathsf{C}$, $\mathsf{a},\mathsf{a}' \sim \mathsf{A}(\mathsf{c})$,[3] *and where* $\big\{(\mathsf{a},\alpha)\big\}_\mathsf{a} = \big\{(\mathsf{a},\Phi|_\mathsf{a})\big\}_\mathsf{a}$, *and* $\big\{(\mathsf{a},\tilde{\alpha})\big\}_\mathsf{a} = \big\{(\mathsf{a},\mathsf{f}_\mathsf{a}(\alpha))\big\}_\mathsf{a}$. *Then there exists an affine map* $\mathsf{T} : \Gamma \to \Gamma$ *such that*

$$\Pr_{\Phi,\mathsf{a}}\big[\tilde{\alpha} = \mathsf{T}(\Phi)|_\mathsf{a}\big] = \Omega(\varepsilon).$$

[2] Equivalently this probability is over $m \sim \mathbb{F}$, $\{(\mathsf{a},\alpha)\} \sim \mathsf{Enc}(m)$ and over $(\mathsf{c},\mathsf{a},\mathsf{a}')$ the randomness of Test

[3] Equivalently this probability is over $m \sim \mathbb{F}$, $\{(\mathsf{a},\alpha)\} \sim \mathsf{Enc}(m)$ and over $(\mathsf{c},\mathsf{a},\mathsf{a}')$ the randomness of Test

Proof of Theorem 5 Assuming Theorem 4. Let ε be as in Theorem 4 above and fix $\mathsf{f} = \{\mathsf{f}_\mathsf{a}\}_\mathsf{a} \in \mathcal{F}$. We will show that there exists $\mathsf{L}_\mathsf{f} \subset \mathcal{G}$ of size at most ℓ such that

$$\Pr_{\mathbf{\Phi},(\mathsf{c},\mathsf{a},\mathsf{a}')}\Big[\tilde{\alpha}|_\mathsf{c} = \tilde{\alpha}'|_\mathsf{c} \;\&\; (\tilde{\alpha},\tilde{\alpha}') \notin \big\{\big(\mathsf{g}_\mathsf{a}(\alpha), \mathsf{g}_{\mathsf{a}'}(\alpha')\big) : \{\mathsf{g}_\mathsf{a}\}_\mathsf{a} \in \mathsf{L}_\mathsf{f}\big\}\Big] < 6\varepsilon, \tag{A.1}$$

where $(\tilde{\alpha},\tilde{\alpha}') = \big(\mathsf{f}_\mathsf{a}(\alpha), \mathsf{f}_{\mathsf{a}'}(\alpha')\big)$ for $(\alpha,\alpha') = \big(\mathbf{\Phi}|_\mathsf{a}, \mathbf{\Phi}|_{\mathsf{a}'}\big)$, and where $\mathbf{\Phi} \sim \Gamma$.[4] Towards this end, let $\mathsf{L}_\mathsf{f} := \big\{\{\mathsf{g}_\mathsf{a}\}_\mathsf{a} \in \mathcal{G} :\ \Pr_{(\mathbf{\Phi},\mathsf{a})\sim\Gamma\times\mathsf{A}}\big[\tilde{\alpha} = \mathsf{g}_\mathsf{a}(\alpha)\big] \geq \varepsilon/2\big\}$.

Small List Size. Assume for contradiction that $|\mathsf{L}_f| \geq \ell = 4/\varepsilon + 1$, and so contains a set $\big\{\{\mathsf{g}^1_\mathsf{a}\}_\mathsf{a}, \ldots, \{\mathsf{g}^\ell_\mathsf{a}\}_\mathsf{a}\big\}$. By inclusion-exclusion,

$$\begin{aligned} 1 \;&\geq\; \Pr_{(\mathbf{\Phi},\mathsf{a})\sim\Gamma\times\mathsf{A}}\Big[\tilde{\alpha} \in \big\{\mathsf{g}^i_\mathsf{a}(\alpha) : i = 1,\ldots,\ell\big\}\Big] \\ &\geq\; \frac{\ell\cdot\varepsilon}{2} - \sum_{1\leq i<j\leq\ell} \Pr_{\mathbf{\Phi},\mathsf{a}}\Big[\mathsf{g}^i_\mathsf{a}(\alpha) = \mathsf{g}^j_\mathsf{a}(\alpha)\Big] > 2 - \binom{\ell}{2}\cdot\left(\frac{1}{|\Gamma|} + \frac{d}{|\mathbb{F}|}\right). \end{aligned}$$

The last inequality used $\ell\varepsilon > 4$, and the bound on $\Pr_{\mathbf{\Phi},\mathsf{a}}\big[\mathsf{g}^i_\mathsf{a}(\mathbf{\Phi}|_\mathsf{a}) = \mathsf{g}^j_\mathsf{a}(\mathbf{\Phi}|_\mathsf{a})\big]$ from point 2 above. The right hand side simplifies to $2 - o(1) > 1$, a contradiction.

Proximity Implies List Decoding. Suppose $\{\mathsf{f}_\mathsf{a}\}$ is such that (A.1) does not hold. Define $\{\mathsf{f}'_\mathsf{a}\}_\mathsf{a} \in \mathcal{F}$ as follows: $\mathsf{f}'_\mathsf{a}(\alpha) = \mathsf{f}_\mathsf{a}(\alpha)$, unless $\mathsf{f}_\mathsf{a}(\alpha) = \mathsf{g}_\mathsf{a}(\alpha)$ for some $\{\mathsf{g}_\mathsf{a}\}_\mathsf{a} \in \mathsf{L}_\mathsf{f}$ in which case $\mathsf{f}'_\mathsf{a}(\alpha)$ outputs a random $\tilde{\alpha} \notin \big\{\mathsf{g}_\mathsf{a}(\alpha) : \{\mathsf{g}_\mathsf{a}\}_\mathsf{a} \in \mathsf{L}_\mathsf{f}\big\}$. Note

$$\Pr_{\mathbf{\Phi},(\mathsf{c},\mathsf{a},\mathsf{a}')}\big[\mathsf{f}'_\mathsf{a}(\alpha)|_\mathsf{c} = \mathsf{f}'_{\mathsf{a}'}(\alpha')|_\mathsf{c}\big] \geq 6\varepsilon$$

[4] as noted in point 3 above, the difference in probability caused by drawing $\mathbf{\Phi} \sim \Gamma$ such that $\mathbf{\Phi}(\mathbf{0}) = m$ instead is negligible.

since (A.1) does not hold. Therefore, by Theorem 4, there exists an affine $\mathsf{T} : \Gamma \to \Gamma$ such that $\Pr_{\Phi,\mathsf{a}}\left[\mathsf{f}'_{\mathsf{a}}(\Phi|_{\mathsf{a}}) = \mathsf{T}(\Phi)|_{\mathsf{a}}\right] \geq \varepsilon$. Thus $\Pr_{\Phi,\mathsf{a}}\left[\mathsf{f}_{\mathsf{a}}(\Phi|_{\mathsf{a}}) = \mathsf{T}(\Phi)|_{\mathsf{a}}\right] \geq \varepsilon - \ell/|\Gamma_{\mathsf{A}}| \geq \varepsilon/2$, and so the coordinate-wise version of T is in L_{f}. This is a contradiction since by construction, for every $\{\mathsf{g}_{\mathsf{a}}\}_{\mathsf{a}} \in \mathsf{L}_{\mathsf{f}}$, $\mathsf{f}'_{\mathsf{a}}(\alpha) \neq \mathsf{g}_{\mathsf{a}}(\alpha)$ holds for all $\mathsf{a} \in \mathsf{A}$ and $\alpha \in \Gamma_{\mathsf{A}}$. ■

A.3 Global Agreement

In this section we prove Lemma 6, restated below in a quantitative form.

Lemma 6 (Restated). *Suppose $\varepsilon \geq \mathbb{F}^{-1/1000}$, and fix parameters $\eta = |\mathbb{F}|^{-9/10}$, $\delta = 3 \cdot |\mathbb{F}|^{-1/60}$, and $\tau = \mathcal{O}\big(\delta/\varepsilon^6 + \eta/\varepsilon^{11}\big)$. Suppose $\{\mathsf{f}_{\mathsf{a}}\}_{\mathsf{a}} \subset \{\mathsf{f} : \Gamma_{\mathsf{A}} \to \Gamma_{\mathsf{A}}\}$ is such that*

$$\Pr_{\Phi,(\mathsf{c},\mathsf{a},\mathsf{a}')}\left[\tilde{\alpha}|_{\mathsf{c}} = \tilde{\alpha}'|_{\mathsf{c}}\right] = 6\varepsilon \tag{A.2}$$

where the probability is over $\Phi \sim \Gamma$, $\mathsf{c} \sim \mathsf{C}$, $\mathsf{a}, \mathsf{a}' \sim \mathsf{A}(\mathsf{c})$, and where $(\tilde{\alpha}, \tilde{\alpha}') = \big(\mathsf{f}_{\mathsf{a}}(\Phi|_{\mathsf{a}}), \mathsf{f}_{\mathsf{a}'}(\Phi|_{\mathsf{a}'})\big)$. Then there exists a set $\mathsf{G} \subset \mathsf{A} \times \Gamma$ of size at least $|\mathsf{G}| \geq 2\varepsilon \cdot |\mathsf{A} \times \Gamma|$ and a function $\mathsf{h} : \overline{\mathsf{C}} \to \Gamma_{\mathsf{C}}$ such that: $\Pr_{\substack{(\mathsf{a},\Phi)\sim\mathsf{G} \\ \mathsf{c}\sim\mathsf{C}(\mathsf{a})}}\left[\tilde{\gamma} \sim \tilde{\alpha}\right] \geq 1 - \zeta$, where $\tilde{\gamma} = \mathsf{h}(\mathsf{c}, \Phi|_{\mathsf{c}})$ and $\zeta := \varepsilon^{-2} \cdot (\tau + \delta) + \varepsilon^{-1} \cdot (\eta + \delta)$.

Remark 8 *Many different parameters are introduced during the course of our analysis which are all $\mathcal{O}\big(|\mathbb{F}|^{-1}\big)$. We encourage the reader to think of two levels of parameters: level one consists of ε only; all other parameters are in level 2 and are much smaller. The level two parameters are each defined to be smaller than ε^c for some constant $c = \mathcal{O}(1)$ which arises during our analysis. So in the above theorem, for example, in order for τ to be level 2, it must be that $\delta \ll \varepsilon^6$ and $\eta \ll \varepsilon^{11}$; additionally, for ζ to be level 2, $\tau \ll \varepsilon^2$ is required. We remark that the analysis prioritizes modularity and succinctness, rather than optimizing*

constants. As a result, the small constant $1/1000$ *is suboptimal.*

We begin by introducing the notation and ideas needed to prove Lemma 6 in Section A.3.1. The actual proof appears in Section A.3.2, conditioned on two claims which we state in Section A.3.1 and prove in Section A.3.3.

A.3.1 Proof Setup.

Notations. In this section B denotes the set of 2−dimensional planes in $\mathbb{F}^k$, and Γ_{B} is the set of 2−variate polynomials over $\mathbb{F}$ of degree at most d, and $\overline{\mathsf{B}} = \mathsf{B} \times \Gamma_{\mathsf{B}}$. The sets $\overline{\mathsf{A}}, \overline{\mathsf{C}}, \Gamma$ are as usual. We will take advantage of the sampling properties of the triple $\overline{\mathsf{A}}/\overline{\mathsf{B}}/\overline{\mathsf{C}}$. When considering two polynomials whose domains intersect, we write $\sim$ to indicate that they agree on the intersection. For example, given $\tilde{\alpha}, \tilde{\alpha}' \in \Gamma_{\mathsf{A}}$ defined on $\mathsf{a}, \mathsf{a}' \in \mathsf{A}(\overline{\mathsf{c}})$ we write $\tilde{\alpha} \sim \tilde{\alpha}'$ if $\tilde{\alpha}|_{\mathsf{c}} = \tilde{\alpha}'|_{\mathsf{c}}$.

We say that $(\mathsf{c}, \gamma, \tilde{\gamma})$ is *good* if $\Pr_{(\mathsf{a},\Phi)}[\tilde{\alpha} \sim \tilde{\gamma}] \geq 4\varepsilon$, where the probability is over $\mathsf{a} \sim \mathsf{A}(\mathsf{c})$ and $\Phi \sim \Gamma(\overline{\mathsf{c}})$. We say $\overline{\mathsf{c}} = (\mathsf{c}, \gamma)$ is *good* if there exists $\tilde{\gamma}$ such that $(\mathsf{c}, \gamma, \tilde{\gamma})$ is. Note that $\Pr_{\overline{\mathsf{c}} \sim \overline{\mathsf{C}}}[\overline{\mathsf{c}} \text{ good}] \geq 2\varepsilon$. To see this, let $\mathsf{p}_{\mathsf{c},\gamma,\tilde{\gamma}} := \Pr_{(\mathsf{a},\Phi)}[\tilde{\alpha} \sim \tilde{\gamma}]$. Then (A.2) gives

$$6\varepsilon = \mathbb{E}_{\overline{\mathsf{c}} \sim \overline{\mathsf{C}}}\left[\sum_{\tilde{\gamma}} \mathsf{p}_{\mathsf{c},\gamma,\tilde{\gamma}} \cdot \Pr_{\mathsf{a}' \sim \mathsf{A}(\mathsf{c})}[\tilde{\alpha}' \sim \tilde{\gamma}]\right] \leq \mathbb{E}_{\overline{\mathsf{c}} \sim \overline{\mathsf{C}}}\left[\max_{\tilde{\gamma}} \{\mathsf{p}_{\mathsf{c},\gamma,\tilde{\gamma}}\}\right].$$

We have used that $\sum_{\tilde{\gamma}} \Pr_{\mathsf{a}' \sim \mathsf{A}(\mathsf{c})}[\tilde{\alpha}' \sim \tilde{\gamma}] = 1$ for all $\overline{\mathsf{c}}$.

Local Functions. Let $\mathsf{h}_0 : \overline{\mathsf{C}} \to \Gamma_{\mathsf{C}}$ be the randomized function which sends $\overline{\mathsf{c}} = (\mathsf{c}, \gamma)$ to a random $\tilde{\gamma}$ such that $(\mathsf{c}, \gamma, \tilde{\gamma})$ is good if such $\tilde{\gamma}$ exists, and to an arbitrary $\tilde{\gamma} \in \Gamma_{\mathsf{C}}$ if not. For $\overline{\mathsf{c}} \in \overline{\mathsf{C}}$, let $\mathsf{g}_{\overline{\mathsf{c}}} : \overline{\mathsf{B}}(\overline{\mathsf{c}}) \to \Gamma_{\mathsf{B}}$ be the randomized function where $\mathsf{g}_{\overline{\mathsf{c}}}(\overline{\mathsf{b}})$ is the distribution on

Γ_{B} which draws $\overline{\mathsf{a}} \sim \overline{\mathsf{A}}(\overline{\mathsf{b}})$ such that $\tilde{\alpha} \sim \mathsf{h}_0(\overline{\mathsf{c}})$, and outputs $\tilde{\beta} = \tilde{\alpha}|_{\mathsf{b}}$. Additonally, we define $\hat{\mathsf{g}}_{\overline{\mathsf{c}}} : \overline{\mathsf{B}}_1(\overline{\mathsf{c}}) \to \Gamma_{\mathsf{B}_1}$ as follows: for $\overline{l} = (l, \lambda) \in \overline{\mathsf{B}}_1(\overline{\mathsf{c}})$ the distribution $\mathsf{g}_{\overline{\mathsf{c}}}(\overline{l})$ draws $\overline{\mathsf{a}} \sim \overline{\mathsf{A}}(\overline{l})$ such that $\tilde{\alpha} \sim \mathsf{h}_0(\overline{\mathsf{c}})$, and outputs $\tilde{\lambda} = \tilde{\alpha}|_l$. Note that, here we are denoting an element of $\overline{\mathsf{B}}_1$ as $\overline{l} = (l, \lambda)$.

Definition 15 (Well-Defined) *Let $\eta = |\mathbb{F}|^{-9/10}$. We say:*

1. $\mathsf{g}_{\overline{\mathsf{c}}}$ *is* well-defined *if:* $\Pr_{\substack{\overline{\mathsf{b}} \sim \overline{\mathsf{B}}(\overline{\mathsf{c}}) \\ \overline{\mathsf{a}}, \overline{\mathsf{a}}' \sim \overline{\mathsf{A}}(\overline{\mathsf{b}})}} \left[\tilde{\alpha} \approx \tilde{\alpha}' \middle| \tilde{\alpha} \sim \mathsf{h}_0(\overline{\mathsf{c}}) \sim \tilde{\alpha}' \right] \geq 1 - \eta$, *where $\tilde{\alpha} \approx \tilde{\alpha}'$ indicates that $\tilde{\alpha}|_{\mathsf{b}} = \tilde{\alpha}'|_{\mathsf{b}}$*

2. $\hat{\mathsf{g}}_{\overline{\mathsf{c}}}$ *is* well-defined *if:* $\Pr_{\substack{\overline{l} \sim \overline{\mathsf{B}}_1(\overline{\mathsf{c}}) \\ \overline{\mathsf{a}}, \overline{\mathsf{a}}' \sim \overline{\mathsf{A}}(\overline{l})}} \left[\tilde{\alpha} \approx \tilde{\alpha}' \middle| \tilde{\alpha} \sim \mathsf{h}_0(\overline{\mathsf{c}}) \sim \tilde{\alpha}' \right] \geq 1 - \eta$, *where $\tilde{\alpha} \approx \tilde{\alpha}'$ indicates that $\tilde{\alpha}|_l = \tilde{\alpha}'|_l$*

Previous work [45, 17] refers to the good $\overline{\mathsf{c}} \in \overline{\mathsf{C}}$ for which $\mathsf{g}_{\overline{\mathsf{c}}}$ is well-defined as *excellent*; the fact that the excellent points comprise a non-negligible fraction of $\overline{\mathsf{C}}$ is a crucial component of the proofs in these papers. We require one extra property from our specialized subset of $\overline{\mathsf{C}}$ which simplifies the remainder of our proof greatly. The following is proved in Section A.3.3.

Claim 13 *There exists a set $\overline{\mathsf{C}}' \subset \overline{\mathsf{C}}$ such that the following hold: 1) $|\overline{\mathsf{C}}'| \geq \varepsilon^3 |\overline{\mathsf{C}}|$; 2) every $\overline{\mathsf{c}} \in \overline{\mathsf{C}}'$ is good and such that both $\mathsf{g}_{\overline{\mathsf{c}}}$ and $\hat{\mathsf{g}}_{\overline{\mathsf{c}}}$ are well-defined; 3)*

$$\Pr_{\overline{\mathsf{c}}, \overline{\mathsf{c}}' \sim \overline{\mathsf{C}}'} \left[\Pr_{\overline{\mathsf{a}} \sim \overline{\mathsf{A}}(\overline{\mathsf{c}}, \overline{\mathsf{c}}')} \left[\mathsf{h}_0(\overline{\mathsf{c}}) \sim \tilde{\alpha} \sim \mathsf{h}_0(\overline{\mathsf{c}}') \right] \geq \varepsilon^5 \right] \geq 1 - \sigma,$$

where $\sigma := \delta/\varepsilon^3 + \delta/\varepsilon^6 + \eta/\varepsilon^{11}$.

Intuitively, the extra property captured by (3) demands that the set of excellent points can be partitioned into large sets of *mutually compatible* points; the set $\overline{\mathsf{C}}'$ is any member of this partition.

The Global Function. Let $\mathsf{h} : \overline{\mathsf{C}} \to \Gamma_{\mathsf{C}}$ be the randomized function where $\mathsf{h}(\overline{\mathsf{c}})$ draws $\overline{\mathsf{b}} \sim \overline{\mathsf{B}}(\overline{\mathsf{c}})$, $\overline{\mathsf{c}}' \sim \overline{\mathsf{C}}'(\overline{\mathsf{b}})$ and outputs $\tilde{\beta}|_{\mathsf{c}}$ where $\tilde{\beta} = \mathsf{g}_{\overline{\mathsf{c}}'}(\overline{\mathsf{b}})$. The following is also proved in Section A.3.3.

Claim 14 *We have* $\Pr_{(\overline{\mathsf{c}},\overline{\mathsf{b}},\overline{\mathsf{c}}')}\left[\mathsf{h}(\overline{\mathsf{c}}) \sim \tilde{\beta}\right] \geq 1-\tau$, *where* $\tau := \left(\sigma+2\varepsilon^{-5}(\eta+\delta)+2\delta\right)$, $\tilde{\beta} = \mathsf{g}_{\overline{\mathsf{c}}'}(\overline{\mathsf{b}})$ *and the probability is over* $\overline{\mathsf{c}} \sim \overline{\mathsf{C}}$, $\overline{\mathsf{b}} \sim \overline{\mathsf{B}}(\overline{\mathsf{c}})$, $\overline{\mathsf{c}}' \sim \overline{\mathsf{C}}'(\overline{\mathsf{b}})$.

A.3.2 Proof of Lemma 6

Notational Convention. Let $\mathsf{h}_0, \mathsf{h} : \overline{\mathsf{C}} \to \Gamma_{\mathsf{C}}$ be the functions defined in Section A.3.1. In this section if we write $\tilde{\gamma}$ when working with $\overline{\mathsf{c}} \in \overline{\mathsf{C}}$, it should be understood that $\tilde{\gamma} = \mathsf{h}(\overline{\mathsf{c}})$. We will always refer to $\mathsf{h}_0(\overline{\mathsf{c}})$ explicitly.

Proof. Suppose $(\varepsilon, \{\mathsf{f}_{\mathsf{a}}\})$ are such that (A.2) holds; let $\overline{\mathsf{C}}' \subset \overline{\mathsf{C}}$ be the set guaranteed by Claim 13. We define G to be the set of $(\mathsf{a}, \Phi) \in \mathsf{A} \times \Gamma$ such that $\Pr_{\overline{\mathsf{c}} \sim \overline{\mathsf{C}}'(\overline{\mathsf{a}})}\left[\tilde{\alpha} \sim \mathsf{h}_0(\overline{\mathsf{c}})\right] \geq \varepsilon$. We have,

$$\mathbb{E}_{(\mathsf{a},\Phi)\sim \mathsf{A}\times\Gamma}\left[\Pr_{\overline{\mathsf{c}} \sim \overline{\mathsf{C}}'(\overline{\mathsf{a}})}\left[\tilde{\alpha} \sim \mathsf{h}_0(\overline{\mathsf{c}})\right]\right] \geq \mathbb{E}_{\overline{\mathsf{c}} \sim \overline{\mathsf{C}}'}\left[\Pr_{\substack{\mathsf{a}\sim\mathsf{A}(\mathsf{c}) \\ \Phi\sim\Gamma(\overline{\mathsf{c}})}}\left[\tilde{\alpha} \sim \mathsf{h}_0(\overline{\mathsf{c}})\right]\right] - \delta \geq 3\varepsilon$$

We have used the sampling of $\mathsf{A} \times \Gamma / \overline{\mathsf{C}}$ for the first inequality, and that all $\overline{\mathsf{c}} \in \overline{\mathsf{C}}'$ are good for the second (and $4\varepsilon - \delta \geq 3\varepsilon$). It follows that $|\mathsf{G}| \geq 2\varepsilon|\mathsf{A} \times \Gamma|$. Thus, it remains to prove that $\Pr_{(\mathsf{a},\Phi),\mathsf{c}}\left[\tilde{\gamma} \sim \tilde{\alpha}\right] \geq 1 - \zeta$, where the probability is over $(\mathsf{a}, \Phi) \sim \mathsf{G}$, $\mathsf{c} \sim \mathsf{C}(\mathsf{a})$ and where $\tilde{\gamma} = \mathsf{h}(\mathsf{c}, \Phi|_{\mathsf{c}})$, where h is the global function defined in Section A.3.1.

So let $\mathsf{p} := \Pr_{(\mathsf{a},\Phi),\mathsf{c}}[\tilde{\gamma} \sim \tilde{\alpha}]$ be the probability we are trying to bound. We have

$$\mathsf{p} \geq \Pr_{\substack{(\mathsf{a},\Phi)\\ \mathsf{b},\mathsf{c},\overline{\mathsf{c}}'}}\left[\tilde{\gamma} \sim \tilde{\beta} \sim \tilde{\alpha} \middle| \tilde{\alpha} \sim \mathsf{h}_0(\overline{\mathsf{c}}')\right] \geq \Pr_{\substack{(\mathsf{a},\Phi)\\ \mathsf{b},\mathsf{c},\overline{\mathsf{c}}'}}\left[\tilde{\gamma} \sim \tilde{\beta} \middle| \tilde{\alpha} \sim \mathsf{h}_0(\overline{\mathsf{c}}')\right] - \Pr_{\substack{(\mathsf{a},\Phi)\\ \mathsf{b},\mathsf{c},\overline{\mathsf{c}}'}}\left[\tilde{\beta} \not\sim \tilde{\alpha} \middle| \tilde{\alpha} \sim \mathsf{h}_0(\overline{\mathsf{c}}')\right],$$

where the probabilities are over $(\mathsf{a},\Phi) \sim \mathsf{G}$, $\mathsf{c} \sim \mathsf{C}(\mathsf{a})$, $\overline{\mathsf{c}}' \sim \overline{\mathsf{C}}'(\overline{\mathsf{a}})$, $\mathsf{b} \sim \mathsf{B}(\mathsf{a},\mathsf{c},\mathsf{c}')$, and where $\tilde{\beta} = \mathsf{g}_{\overline{\mathsf{c}}'}(\overline{\mathsf{b}})$, for $\overline{\mathsf{b}} = (\mathsf{b}, \Phi|_{\mathsf{b}})$. We conclude by bounding both probabilities on the right; denoted RHS_1 and RHS_2, respectively. We have

$$\begin{aligned} 1 - \mathsf{RHS}_1 &= \Pr_{\substack{(\mathsf{a},\Phi)\sim\mathsf{G}\\ \mathsf{b},\mathsf{c},\overline{\mathsf{c}}'}}\left[\tilde{\gamma} \not\sim \tilde{\beta} \middle| \tilde{\alpha} \sim \mathsf{h}_0(\overline{\mathsf{c}}')\right] \leq \frac{\Pr_{(\mathsf{a},\Phi),\mathsf{b},\mathsf{c},\overline{\mathsf{c}}'}\left[\tilde{\gamma} \not\sim \tilde{\beta}\right]}{\min_{(\mathsf{a},\Phi)\in\mathsf{G}}\left\{\Pr_{\overline{\mathsf{c}}'\sim\overline{\mathsf{C}}'(\overline{\mathsf{a}})}\left[\tilde{\alpha} \sim \mathsf{h}_0(\overline{\mathsf{c}}')\right]\right\}} \\ &\leq \frac{\varepsilon^{-2}}{2} \cdot \Pr_{\substack{\overline{\mathsf{a}}\sim\overline{\mathsf{A}}\\ \mathsf{b},\mathsf{c},\overline{\mathsf{c}}'}}\left[\tilde{\gamma} \not\sim \tilde{\beta}\right] < \varepsilon^{-2} \cdot \left(\Pr_{\substack{\overline{\mathsf{c}}\sim\overline{\mathsf{C}}\\ \overline{\mathsf{b}}\sim\overline{\mathsf{B}}(\overline{\mathsf{c}})\\ \overline{\mathsf{c}}'\sim\overline{\mathsf{C}}'(\overline{\mathsf{b}})}}\left[\tilde{\gamma} \not\sim \tilde{\beta}\right] + \delta\right) \leq \varepsilon^{-2} \cdot (\tau + \delta). \end{aligned}$$

The first inequality on the second line used the definition of G and that $|\mathsf{G}| \geq 2\varepsilon \cdot |\mathsf{A} \times \Gamma|$; the second used Claim 12, point 1; and the last used Claim 14. Finally,

$$\begin{aligned} \mathsf{RHS}_2 &\leq \frac{\varepsilon^{-1}}{2} \cdot \Pr_{\substack{\overline{\mathsf{a}}\sim\overline{\mathsf{A}}\\ \overline{\mathsf{c}}'\sim\overline{\mathsf{C}}'(\overline{\mathsf{a}})\\ \overline{\mathsf{b}}\sim\overline{\mathsf{B}}(\overline{\mathsf{c}}',\overline{\mathsf{a}})}}\left[\tilde{\beta} \not\sim \tilde{\alpha} \middle| \tilde{\alpha} \sim \mathsf{h}_0(\overline{\mathsf{c}}')\right] \\ &\leq \varepsilon^{-1} \cdot \left(\max_{\overline{\mathsf{c}}'\in\overline{\mathsf{C}}'}\left\{\Pr_{\substack{\overline{\mathsf{b}}\sim\overline{\mathsf{B}}(\overline{\mathsf{c}}')\\ \overline{\mathsf{a}}\sim\overline{\mathsf{A}}(\overline{\mathsf{b}})}}\left[\tilde{\beta} \not\sim \tilde{\alpha} \middle| \tilde{\alpha} \sim \mathsf{h}_0(\overline{\mathsf{c}}')\right]\right\} + \delta\right) \leq \varepsilon^{-1}(\eta + \delta). \end{aligned}$$

We have used Claim 12 point 2 and the fact that $\mathsf{g}_{\overline{\mathsf{c}}'}$ is well-defined for all $\overline{\mathsf{c}}' \in \overline{\mathsf{C}}'$. The result follows. ■

A.3.3 Proving the Claims

Starting Assumption and Notational Conventions. Throughout this section, we assume the hypotheses of Lemma 6, namely $(\varepsilon, \{\mathsf{f}_{\mathsf{a}}\})$ are such that $\Pr_{\Phi,(\mathsf{c},\mathsf{a},\mathsf{a}')}[\tilde{\alpha} \sim \tilde{\alpha}'] = 6\varepsilon$

(*i.e.*, such that (A.2) holds). Let $\mathsf{h}_0, \mathsf{h} : \overline{\mathsf{C}} \to \Gamma_{\mathsf{C}}$ be the functions defined in Section A.3.1. In this section if we write $\tilde{\gamma}$ when working with $\overline{\mathsf{c}} \in \overline{\mathsf{C}}$, it should be understood that $\tilde{\gamma} = \mathsf{h}_0(\overline{\mathsf{c}})$. We will refer to $\mathsf{h}(\overline{\mathsf{c}})$ explicitly (note, this is opposite to the convention of Section A.3.2). Given $\overline{\mathsf{c}}, \overline{\mathsf{c}}' \in \overline{\mathsf{C}}$ set $\mu_{\overline{\mathsf{c}}}$, $\mathsf{p}(\overline{\mathsf{c}})$, $\mathsf{r}(\overline{\mathsf{c}})$ and $\mathsf{q}(\overline{\mathsf{c}}, \overline{\mathsf{c}}')$ to:

$$\Pr_{\substack{\mathsf{a}\sim\mathsf{A}(\mathsf{c})\\ \Phi\sim\Gamma(\overline{\mathsf{c}})}}\left[\tilde{\gamma} \sim \tilde{\alpha}\right];\ \Pr_{\substack{\overline{\mathsf{b}}\sim\overline{\mathsf{B}}(\overline{\mathsf{c}})\\ \overline{\mathsf{a}}\sim\overline{\mathsf{A}}(\overline{\mathsf{b}})}}\left[\tilde{\beta} \sim \tilde{\alpha} \middle| \tilde{\gamma} \sim \tilde{\alpha}\right]; \Pr_{\substack{\overline{l}\sim\overline{\mathsf{B}}_1(\overline{\mathsf{c}})\\ \overline{\mathsf{a}}\sim\overline{\mathsf{A}}(\overline{l})}}\left[\tilde{\lambda} \sim \tilde{\alpha} \middle| \tilde{\gamma} \sim \tilde{\alpha}\right];\ \Pr_{\overline{\mathsf{a}}\sim\overline{\mathsf{A}}(\overline{\mathsf{c}},\overline{\mathsf{c}}')}\left[\tilde{\gamma} \sim \tilde{\alpha} \sim \tilde{\gamma}'\right].$$

In Section A.3.1 we called $\overline{\mathsf{c}} \in \overline{\mathsf{C}}$ such that $\mu_{\overline{\mathsf{c}}} \geq 4\varepsilon$ *good*. Also for $\overline{\mathsf{c}} \in \overline{\mathsf{C}}$ we defined local functions $\mathsf{g}_{\overline{\mathsf{c}}} : \overline{\mathsf{B}}(\overline{\mathsf{c}}) \to \Gamma_{\mathsf{B}}$, $\hat{\mathsf{g}}_{\overline{\mathsf{c}}} : \overline{\mathsf{B}}_1(\overline{\mathsf{c}}) \to \Gamma_{\mathsf{B}_1}$ and said that $\mathsf{g}_{\overline{\mathsf{c}}}$ was *well-defined* if $\mathsf{p}(\overline{\mathsf{c}}) \geq 1 - \eta$ and $\hat{\mathsf{g}}_{\overline{\mathsf{c}}}$ was *well-defined* if $\mathsf{r}(\overline{\mathsf{c}}) \geq 1 - \eta$, where $\eta = |\mathbb{F}|^{-9/10}$. In the remainder of this section we prove three claims; the first two combine to prove Claim 13, the last is Claim 14.

Claim 15 *There exists a set $\overline{\mathsf{C}}'_0 \subset \overline{\mathsf{C}}$ such that the following hold: 1) $|\overline{\mathsf{C}}'_0| \geq \varepsilon|\overline{\mathsf{C}}|$; 2) $\mu_{\overline{\mathsf{c}}} \geq 4\varepsilon$ for every $\overline{\mathsf{c}} \in \overline{\mathsf{C}}'_0$; 3) $\mathsf{p}(\overline{\mathsf{c}}) \geq 1 - \eta$; 4) $\mathsf{r}(\overline{\mathsf{c}}) \geq 1 - \eta$ for every $\overline{\mathsf{c}} \in \overline{\mathsf{C}}'_0$.*

Proof. Let $\overline{\mathsf{C}}'_0 \subset \overline{\mathsf{C}}$ be the set of $\overline{\mathsf{c}} \in \overline{\mathsf{C}}$ for which $\mu_{\overline{\mathsf{c}}} \geq 4\varepsilon$ and $\mathsf{p}(\overline{\mathsf{c}}) \geq 1 - \eta$ (*i.e.*, $\overline{\mathsf{c}} \in \overline{\mathsf{C}}'_0$ if $\overline{\mathsf{c}}$ is good and such that $\mathsf{g}_{\overline{\mathsf{c}}}$ is well-defined). We bound $|\overline{\mathsf{C}}'_0|$ using three observations. First, as noted in Section A.3.1, $\Pr_{\overline{\mathsf{c}}\sim\overline{\mathsf{C}}}\left[\mu_{\overline{\mathsf{c}}} \geq 4\varepsilon\right] \geq 2\varepsilon$. Second, for all $\overline{\mathsf{c}} \in \overline{\mathsf{C}}$ such that $\mu_{\overline{\mathsf{c}}} \geq 4\varepsilon$:

$$\Pr_{\substack{\overline{\mathsf{b}}\sim\overline{\mathsf{B}}(\overline{\mathsf{c}})\\ \overline{\mathsf{a}},\overline{\mathsf{a}}'\sim\overline{\mathsf{A}}(\overline{\mathsf{b}})}}\left[\tilde{\alpha} \sim \tilde{\gamma} \sim \tilde{\alpha}'\right] = \mathbb{E}_{\overline{\mathsf{b}}\sim\overline{\mathsf{B}}(\overline{\mathsf{c}})}\left[\mu_{\overline{\mathsf{c}}}(\overline{\mathsf{b}})^2\right] \geq \Pr_{\overline{\mathsf{b}}\sim\overline{\mathsf{B}}(\overline{\mathsf{c}})}\left[\left|\mu_{\overline{\mathsf{c}}}(\overline{\mathsf{b}}) - \mu_{\overline{\mathsf{c}}}\right| \leq \varepsilon\right] \cdot 9\varepsilon^2 \geq \varepsilon^2,$$

where $\mu_{\overline{\mathsf{c}}}(\overline{\mathsf{b}}) := \Pr_{\overline{\mathsf{a}}\sim\overline{\mathsf{A}}(\overline{\mathsf{b}})}\left[\tilde{\alpha} \sim \tilde{\gamma}\right]$ is shorthand. We have used the sampling of $\overline{\mathsf{A}}(\overline{\mathsf{c}})/\overline{\mathsf{B}}(\overline{\mathsf{c}})$ to (crudely) lower bound $\Pr_{\overline{\mathsf{b}}\sim\overline{\mathsf{B}}(\overline{\mathsf{c}})}\left[\left|\mu_{\overline{\mathsf{c}}}(\overline{\mathsf{b}}) - \mu_{\overline{\mathsf{c}}}\right| \leq \varepsilon\right]$. Similarly, using sampling of $\overline{\mathsf{A}}(\overline{\mathsf{c}})/\overline{\mathsf{B}}_1(\overline{\mathsf{c}})$ we

get $\Pr_{\substack{\bar{l}\sim\overline{\mathsf{B}}_1(\bar{\mathsf{c}}) \\ \bar{\mathsf{a}},\bar{\mathsf{a}}'\sim\overline{\mathsf{A}}(\bar{l})}}\left[\tilde{\alpha}\sim\tilde{\gamma}\sim\tilde{\alpha}'\right]\geq\varepsilon^2$. Now, Event_1 be: $\Pr_{\substack{\bar{\mathsf{b}}\sim\overline{\mathsf{B}}(\bar{\mathsf{c}}) \\ \bar{\mathsf{a}},\bar{\mathsf{a}}'\sim\overline{\mathsf{A}}(\bar{\mathsf{b}})}}\left[\tilde{\alpha}\not\approx\tilde{\alpha}'\ \&\ \tilde{\alpha}\sim\tilde{\gamma}\sim\tilde{\alpha}'\right]>\eta\varepsilon^2$ and Event_2 to be the same event except over the distribution $\bar{l}\sim\mathsf{B}_1(\bar{\mathsf{c}})$ and $\bar{\mathsf{a}},\bar{\mathsf{a}}'\sim\overline{\mathsf{A}}(\bar{l})$. By Markov's inequality and Schwartz-Zippel:

$$\Pr_{\bar{\mathsf{c}}\sim\overline{\mathsf{C}}}\left[\mathsf{Event}_1\right]+\Pr_{\bar{\mathsf{c}}\sim\overline{\mathsf{C}}}\left[\mathsf{Event}_2\right]\leq\frac{2d}{\eta\varepsilon^2|\mathbb{F}|}.$$

Putting these together gives

$$\begin{aligned}&\frac{|\overline{\mathsf{C}}_0'|}{|\overline{\mathsf{C}}|}\\&\geq\Pr_{\bar{\mathsf{c}}\sim\overline{\mathsf{C}}}\left[\mu_{\bar{\mathsf{c}}}\geq 4\varepsilon\right]-\Pr_{\bar{\mathsf{c}}\sim\overline{\mathsf{C}}}\left[\mathsf{Event}_1\right]-\Pr_{\bar{\mathsf{c}}\sim\overline{\mathsf{C}}}\left[\mathsf{Event}_2\right]\\&\geq 2\varepsilon-\frac{2d}{\eta\varepsilon^2|\mathbb{F}|}\geq\varepsilon.\end{aligned}$$

■

Claim 13 (Restated). There exists a set $\overline{\mathsf{C}}'\subset\overline{\mathsf{C}}$ such that the following hold: 1) $|\overline{\mathsf{C}}'|\geq\varepsilon^3|\overline{\mathsf{C}}|$; 2) $\mu_{\bar{\mathsf{c}}}\geq 4\varepsilon$ for every $\bar{\mathsf{c}}\in\overline{\mathsf{C}}'$; 3) $\mathsf{p}(\bar{\mathsf{c}})\geq 1-\eta$ for every $\bar{\mathsf{c}}\in\overline{\mathsf{C}}'$; 4) $\Pr_{\bar{\mathsf{c}},\bar{\mathsf{c}}'\sim\overline{\mathsf{C}}'}\left[\mathsf{q}(\bar{\mathsf{c}},\bar{\mathsf{c}}')\geq\varepsilon^5\right]\geq 1-\sigma$, where $\sigma:=\delta/\varepsilon^3+(2\delta/+\eta)/\varepsilon^{11}$.

Proof. By Claim 15 it suffices to construct a large subset of $\overline{\mathsf{C}}_0'$ such that the fourth property holds. For this purpose, we equip $\overline{\mathsf{C}}_0'$ with a graph structure: $\bar{\mathsf{c}},\bar{\mathsf{c}}'\in\overline{\mathsf{C}}_0'$ are adjacent if $\mathsf{q}(\bar{\mathsf{c}},\bar{\mathsf{c}}')\geq\varepsilon^2$. Our final set $\overline{\mathsf{C}}'$ will be the neighborhood, $\mathsf{N}(\bar{\mathsf{c}}'):=\left\{\bar{\mathsf{c}}\in\overline{\mathsf{C}}_0':\mathsf{q}(\bar{\mathsf{c}},\bar{\mathsf{c}}')\geq\varepsilon^2\right\}$ of some $\bar{\mathsf{c}}'\in\overline{\mathsf{C}}_0'$. In order for this to work, $\bar{\mathsf{c}}'$ should satisfy: 1) $\left|\mathsf{N}(\bar{\mathsf{c}}')\right|$ must be large; 2) $\Pr_{\bar{\mathsf{c}},\bar{\mathsf{c}}''\sim\mathsf{N}(\bar{\mathsf{c}}')}[\mathsf{q}(\bar{\mathsf{c}},\bar{\mathsf{c}}'')<\varepsilon^5]$ must be small. We show there exists such a $\bar{\mathsf{c}}'\in\overline{\mathsf{C}}_0'$. Specifically we prove

1. $\mathbb{E}_{\overline{\mathsf{c}},\overline{\mathsf{c}}' \sim \overline{\mathsf{C}}_0'}\left[\mathsf{q}(\overline{\mathsf{c}},\overline{\mathsf{c}}')\right] \geq 3\varepsilon^2$; and

2. $\Pr_{\substack{\overline{\mathsf{c}}' \sim \overline{\mathsf{C}}_0' \\ \overline{\mathsf{c}},\overline{\mathsf{c}}'' \sim \mathsf{N}(\overline{\mathsf{c}}')}}\left[\mathsf{q}(\overline{\mathsf{c}},\overline{\mathsf{c}}'') \geq \varepsilon^5 \middle| |\mathsf{N}(\overline{\mathsf{c}}')| > \varepsilon^3|\overline{\mathsf{C}}|\right] \geq 1-\sigma.$

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

For the first, we have

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

We have used the sampling of $\overline{\mathsf{A}}/\overline{\mathsf{C}}^2$, Jensen's inequality, the sampling of $\overline{\mathsf{A}}/\overline{\mathsf{C}}$, and the fact that $\mu_{\overline{\mathsf{c}}} \geq 4\varepsilon$ for all $\overline{\mathsf{c}} \in \overline{\mathsf{C}}_0'$. Establishing the second bound is more involved. Towards this end, we define three quantities, shorthanded as $\mathsf{val}_1, \mathsf{val}_2, \mathsf{val}_3$; each is a function of $(\overline{\mathsf{c}},\overline{\mathsf{c}}',\overline{\mathsf{c}}'')$:

- $\mathsf{val}_1 := \left|\Pr_{\overline{\mathsf{a}}' \sim \overline{\mathsf{A}}(\overline{\mathsf{c}},\overline{\mathsf{c}}',\overline{\mathsf{c}}'')}\left[\tilde{\gamma} \sim \tilde{\alpha}' \sim \tilde{\gamma}''\right] - \mathsf{q}(\overline{\mathsf{c}},\overline{\mathsf{c}}'')\right|;$
- $\mathsf{val}_2 := \mathbb{E}_{\substack{\bar{l} \sim \overline{\mathsf{B}}_1(\overline{\mathsf{c}},\overline{\mathsf{c}}') \\ \bar{l}'' \sim \overline{\mathsf{B}}_1(\overline{\mathsf{c}}'',\overline{\mathsf{c}}')}}\left[\left|\Pr_{\overline{\mathsf{a}}' \sim \overline{\mathsf{A}}(\bar{l},\bar{l}'')}[\tilde{\gamma}' \sim \tilde{\alpha}'] - \mu_{\overline{\mathsf{c}}'}\right|\right]$
- $\mathsf{val}_3 := \Pr_{\substack{\bar{l} \sim \overline{\mathsf{B}}_1(\overline{\mathsf{c}},\overline{\mathsf{c}}') \\ \overline{\mathsf{a}}' \sim \overline{\mathsf{A}}(\bar{l},\overline{\mathsf{c}}'') \\ \overline{\mathsf{a}} \sim \overline{\mathsf{A}}(\bar{l})}}\left[\tilde{\alpha} \not\approx \tilde{\alpha}' \middle| \tilde{\alpha} \sim \tilde{\gamma}' \sim \tilde{\alpha}'\right] + \Pr_{\substack{\bar{l}'' \sim \overline{\mathsf{B}}_1(\overline{\mathsf{c}},\overline{\mathsf{c}}') \\ \overline{\mathsf{a}}' \sim \overline{\mathsf{A}}(\bar{l}'',\overline{\mathsf{c}}) \\ \overline{\mathsf{a}} \sim \overline{\mathsf{A}}(\bar{l}'')}}\left[\tilde{\alpha}' \not\approx \tilde{\alpha}'' \middle| \tilde{\alpha}' \sim \tilde{\gamma}' \sim \tilde{\alpha}''\right].$

We show that each val_i is small with very high probability over $(\overline{\mathsf{c}},\overline{\mathsf{c}}',\overline{\mathsf{c}}'')$ drawn as follows: $\overline{\mathsf{c}}' \sim \overline{\mathsf{C}}_0'$ such that $|\mathsf{N}(\overline{\mathsf{c}}')| \geq \varepsilon^3|\overline{\mathsf{C}}|$, $\overline{\mathsf{c}},\overline{\mathsf{c}}'' \sim \mathsf{N}(\overline{\mathsf{c}}')$. These bounds will be used in the

computation which follows. We have

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

where $f_1(\overline{\mathsf{a}}') = 1$ if $\tilde{\gamma} \sim \tilde{\alpha}' \sim \tilde{\gamma}''$, 0 otherwise. Thus $\Pr_{(\overline{\mathsf{c}},\overline{\mathsf{c}}',\overline{\mathsf{c}}'')}\big[\mathsf{val}_1 > \delta\big] \leq \delta/\varepsilon^3$, by the sampling of $\overline{\mathsf{A}}(\overline{\mathsf{c}},\overline{\mathsf{c}}'')/\overline{\mathsf{C}}$ for all $\overline{\mathsf{c}},\overline{\mathsf{c}}'' \in \overline{\mathsf{C}}$. Likewise, for val_2, we have

$$\begin{aligned}\Pr_{(\overline{\mathsf{c}},\overline{\mathsf{c}}',\overline{\mathsf{c}}'')}\big[\mathsf{val}_2 > 2\varepsilon^5\big] &= \varepsilon^{-6} \max_{\overline{\mathsf{c}}'\in\overline{\mathsf{C}}} \left\{ \Pr_{\overline{\mathsf{c}},\overline{\mathsf{c}}''} \Big[\mathsf{val}_2 \geq 2\varepsilon^5\Big] \right\} \\ &= \varepsilon^{-11} \max_{\overline{\mathsf{c}}'\in\overline{\mathsf{C}}} \left\{ \mathbb{E}_{\bar{l},\bar{l}''\sim\overline{\mathsf{B}}_1(\overline{\mathsf{c}}')} \Big[\big| \Pr_{\overline{\mathsf{a}}'\sim\overline{\mathsf{A}}(\bar{l},\bar{l}'')}[\tilde{\gamma}' \sim \tilde{\alpha}'] - \mu_{\overline{\mathsf{c}}'} \big| \Big] \right\}\end{aligned}$$

It follows $\Pr_{(\overline{\mathsf{c}},\overline{\mathsf{c}}',\overline{\mathsf{c}}'')}\big[\mathsf{val}_2 > 2\varepsilon^5\big] \leq 2\delta/\varepsilon^{11}$ from the sampling of $\overline{\mathsf{A}}(\overline{\mathsf{c}}')/\overline{\mathsf{B}}_1(\overline{\mathsf{c}}') \times \overline{\mathsf{B}}_1(\overline{\mathsf{c}}')$ for all $\overline{\mathsf{c}}' \in \overline{\mathsf{C}}$ and the function $f_2(\overline{\mathsf{a}}') = 1$ iff $\tilde{\gamma}' \sim \tilde{\alpha}'$. Finally,

$$\begin{aligned}\Pr_{(\overline{\mathsf{c}},\overline{\mathsf{c}}',\overline{\mathsf{c}}'')}\big[\mathsf{val}_3 > 2\varepsilon^5\big] &\leq \varepsilon^{-6} \cdot \max_{\overline{\mathsf{c}}'\in\overline{\mathsf{C}}'_0} \Big\{ \Pr_{\overline{\mathsf{c}},\overline{\mathsf{c}}''\sim\overline{\mathsf{C}}}\big[\mathsf{val}_3 > 2\varepsilon^5\big] \Big\} \leq \frac{\varepsilon^{-11}}{2} \cdot \max_{\overline{\mathsf{c}}'\in\overline{\mathsf{C}}'_0} \Big\{ \mathbb{E}_{\overline{\mathsf{c}},\overline{\mathsf{c}}''\sim\overline{\mathsf{C}}}\big[\mathsf{val}_3\big] \Big\} \\ &= \frac{\varepsilon^{-11}}{2} \cdot \max_{\overline{\mathsf{c}}'\in\overline{\mathsf{C}}'_0} \Big\{ 2 \cdot \big(1 - \mathsf{p}(\overline{\mathsf{c}}')\big) \Big\} \leq \eta/\varepsilon^{11}.\end{aligned}$$

Now we show how these values figure into deriving the bound we need. The key point is that they let us bound $\mathsf{q}(\overline{\mathsf{c}},\overline{\mathsf{c}}'')$ in terms of $\mathsf{q}(\overline{\mathsf{c}},\overline{\mathsf{c}}') \cdot \mathsf{q}(\overline{\mathsf{c}}',\overline{\mathsf{c}}'') \cdot \mu_{\overline{\mathsf{c}}'}$, which is large when $\overline{\mathsf{c}},\overline{\mathsf{c}}'' \in \mathsf{N}(\overline{\mathsf{c}}')$

and $\overline{\mathsf{c}}' \in \overline{\mathsf{C}}'_0$. We have:

$$\begin{aligned}
\mathsf{q}(\overline{\mathsf{c}}, \overline{\mathsf{c}}'') &= \Pr_{\overline{\mathsf{a}}' \sim \overline{\mathsf{A}}(\overline{\mathsf{c}}, \overline{\mathsf{c}}'')}\left[\tilde{\gamma} \sim \tilde{\alpha}' \sim \tilde{\gamma}''\right] \geq \Pr_{\overline{\mathsf{a}}' \sim \overline{\mathsf{A}}(\overline{\mathsf{c}}, \overline{\mathsf{c}}', \overline{\mathsf{c}}'')}\left[\tilde{\gamma} \sim \tilde{\alpha}' \sim \tilde{\gamma}''\right] - \mathsf{val}_1 \\
&\geq \Pr_{\substack{\overline{\mathsf{a}}' \sim \overline{\mathsf{A}}(\overline{\mathsf{c}}, \overline{\mathsf{c}}', \overline{\mathsf{c}}'') \\ \overline{\mathsf{a}} \sim \overline{\mathsf{A}}(\overline{l}) \\ \overline{\mathsf{a}}'' \sim \overline{\mathsf{A}}(\overline{l}'')}}\left[\tilde{\gamma} \sim \tilde{\alpha} \sim \tilde{\gamma}' \sim \tilde{\alpha}'' \sim \tilde{\gamma}'' \;\&\; \tilde{\alpha} \approx \tilde{\alpha}' \approx \tilde{\alpha}'' \;\&\; \tilde{\gamma}' \sim \tilde{\alpha}'\right] - \mathsf{val}_1 \\
&\geq \Pr_{\substack{\overline{\mathsf{a}}' \sim \overline{\mathsf{A}}(\overline{\mathsf{c}}, \overline{\mathsf{c}}', \overline{\mathsf{c}}'') \\ \overline{\mathsf{a}} \sim \overline{\mathsf{A}}(\overline{l}) \\ \overline{\mathsf{a}}'' \sim \overline{\mathsf{A}}(\overline{l}'')}}\left[\tilde{\gamma} \sim \tilde{\alpha} \sim \tilde{\gamma}' \;\&\; \tilde{\gamma}' \sim \tilde{\alpha}' \;\&\; \tilde{\gamma}' \sim \tilde{\alpha}'' \sim \tilde{\gamma}''\right] - \mathsf{val}_1 - \mathsf{val}_3 \\
&= \mathsf{q}(\overline{\mathsf{c}}, \overline{\mathsf{c}}') \cdot \mathsf{q}(\overline{\mathsf{c}}', \overline{\mathsf{c}}'') \cdot \mu_{\overline{\mathsf{c}}'} - \mathsf{quantity} \\
&\geq \mathsf{q}(\overline{\mathsf{c}}, \overline{\mathsf{c}}') \cdot \mathsf{q}(\overline{\mathsf{c}}', \overline{\mathsf{c}}'') \cdot \mu_{\overline{\mathsf{c}}'} - \mathsf{val}_1 - \mathsf{val}_2 - \mathsf{val}_3 \geq 4\varepsilon^5 - \mathsf{val}_1 - \mathsf{val}_2 - \mathsf{val}_3.
\end{aligned}$$

Where $\mathsf{quantity}$ is $\mathbb{E}_{\overline{\mathsf{a}}, \overline{\mathsf{a}}''}\left[\mathbb{1}_{\tilde{\gamma} \sim \tilde{\alpha} \sim \tilde{\gamma}'} \cdot \mathbb{1}_{\tilde{\gamma}'' \sim \tilde{\alpha}'' \sim \tilde{\gamma}'}(\Pr_{\overline{\mathsf{a}}' \sim \overline{\mathsf{A}}(\overline{l}, \overline{l}'')}[\tilde{\gamma}' \sim \tilde{\alpha}'] - \mu_{\overline{\mathsf{c}}'})\right] - \mathsf{val}_1 - \mathsf{val}_3$. In the probability subscript in the second line, $\overline{l}$ and $\overline{l}''$ are the restrictions of $\overline{\mathsf{a}}'$ to the lines spanned by $(\mathsf{c}, \mathsf{c}')$ and $(\mathsf{c}', \mathsf{c}'')$, respectively. The result follows:

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

■

Claim 14 (Restated). *We have*

$$\Pr_{\substack{\overline{\mathsf{c}} \sim \overline{\mathsf{C}} \\ \overline{\mathsf{b}}_1 \sim \overline{\mathsf{B}}(\overline{\mathsf{c}}) \\ \overline{\mathsf{c}}'_1 \sim \overline{\mathsf{C}}'(\overline{\mathsf{b}}_1)}}\left[\mathsf{h}(\overline{\mathsf{c}}) \sim \tilde{\beta}_1\right] \geq 1 - \tau,$$

where $\tilde{\beta} = \mathsf{g}_{\overline{\mathsf{c}}'}(\overline{\mathsf{b}})$*, and where* $\tau := \left(\sigma + 2\varepsilon^{-5}(\eta + \delta) + 2\delta\right)$*. Recall* $\mathsf{h}(\overline{\mathsf{c}})$ *is the distribution on* Γ_{C} *which draws* $\overline{\mathsf{b}}'_2 \sim \overline{\mathsf{B}}(\overline{\mathsf{c}})$*,* $\overline{\mathsf{c}}'_2 \sim \overline{\mathsf{C}}'(\overline{\mathsf{b}}_2)$ *and outputs* $\mathsf{g}_{\overline{\mathsf{c}}'_2}(\overline{\mathsf{b}}_2)|_{\mathsf{c}}$*.*

Proof. We show $\Pr_{(\bar{\mathsf{c}},\bar{\mathsf{c}}'_1,\bar{\mathsf{c}}'_2,\bar{\mathsf{b}}_1,\bar{\mathsf{b}}_2)}[\tilde{\beta}_1 \sim \tilde{\beta}_2] \geq 1 - \big(\sigma + 2\varepsilon^{-5}(\eta+\delta)\big)$, where the probability is over $\bar{\mathsf{c}} \sim \overline{\mathsf{C}}$, $\bar{\mathsf{c}}'_1, \bar{\mathsf{c}}'_2 \sim \overline{\mathsf{C}}'$, $\bar{\mathsf{b}}_1 \sim \overline{\mathsf{B}}(\bar{\mathsf{c}}, \bar{\mathsf{c}}'_1)$, $\bar{\mathsf{b}}_2 \sim \overline{\mathsf{B}}(\bar{\mathsf{c}}, \bar{\mathsf{c}}'_2)$ and where $\tilde{\beta}_1 \sim \tilde{\beta}_2$ means that $\mathsf{g}_{\bar{\mathsf{c}}'_1}(\bar{\mathsf{b}}_1)$ and $\mathsf{g}_{\bar{\mathsf{c}}'_2}(\bar{\mathsf{b}}_2)$ agree at $\bar{\mathsf{c}}$. The result then follows by the sampling of $\overline{\mathsf{B}}(\bar{\mathsf{c}})/\overline{\mathsf{C}}$ for all $\bar{\mathsf{c}} \in \overline{\mathsf{C}}$. We have

$$\Pr_{(\bar{\mathsf{c}},\bar{\mathsf{c}}'_1,\bar{\mathsf{c}}'_2,\bar{\mathsf{b}}_1,\bar{\mathsf{b}}_2)}\Big[\tilde{\beta}_1 \sim \tilde{\beta}_2\Big] \geq \mathbb{E}_{\bar{\mathsf{c}}'_1,\bar{\mathsf{c}}'_2 \sim \overline{\mathsf{C}}'}\left[\Pr_{\substack{(\bar{\mathsf{c}},\bar{\mathsf{b}}_1,\bar{\mathsf{b}}_2) \\ \bar{\mathsf{a}} \sim \overline{\mathsf{A}}(\bar{l}_1,\bar{l}_2)}}\Big[\tilde{\beta}_1 \sim \tilde{\alpha} \sim \tilde{\beta}_2 \Big| \tilde{\gamma}'_1 \sim \tilde{\alpha} \sim \tilde{\gamma}'_2\Big]\right].$$

where $\bar{l}_1 = (l_1, \lambda_1)$, $l_1 =$ line spanned by $\mathsf{c}, \mathsf{c}'_1$ and $\lambda_1 = \beta_1|_{l_1}$; similary $\bar{l}_2 = (l_2, \lambda_2)$ with $l_2 =$ line $(\mathsf{c}, \mathsf{c}'_2)$ and $\lambda_2 = \beta_2|_{l_2}$. Let $\mathsf{val} := \Pr_{(\bar{\mathsf{c}},\bar{\mathsf{b}}_1,\bar{\mathsf{b}}_2,\bar{\mathsf{a}})}[\tilde{\beta}_1 \sim \tilde{\alpha} \sim \tilde{\beta}_2 | \tilde{\gamma}'_1 \sim \alpha \sim \tilde{\gamma}'_2]$ be shorthand for the quantity inside the expectation. We have

$$\begin{aligned}
\mathsf{val} &\geq 1 - \left[\Pr_{(\bar{\mathsf{c}},\bar{\mathsf{b}}_1,\bar{\mathsf{b}}_2,\bar{\mathsf{a}})}\Big[\tilde{\beta}_1 \not\sim \tilde{\alpha} \Big| \tilde{\gamma}'_1 \sim \tilde{\alpha} \sim \tilde{\gamma}'_2\Big] + \Pr_{(\bar{\mathsf{c}},\bar{\mathsf{b}}_1,\bar{\mathsf{b}}_2,\bar{\mathsf{a}})}\Big[\tilde{\beta}_2 \not\sim \tilde{\alpha} \Big| \tilde{\gamma}'_1 \sim \tilde{\alpha} \sim \tilde{\gamma}'_2\Big]\right] \\
&\geq 1 - \frac{1}{\mathsf{q}(\bar{\mathsf{c}}'_1, \bar{\mathsf{c}}'_2)} \cdot \left[\Pr_{\substack{\bar{\mathsf{c}} \sim \overline{\mathsf{C}}, \bar{\mathsf{b}}_1 \sim \overline{\mathsf{B}}(\bar{\mathsf{c}},\bar{\mathsf{c}}'_1) \\ \bar{\mathsf{a}} \sim \overline{\mathsf{A}}(\bar{l}_1,\bar{\mathsf{c}}'_2)}}\Big[\tilde{\beta}_1 \not\sim \tilde{\alpha} \Big| \tilde{\gamma}'_1 \sim \tilde{\alpha}\Big] + \Pr_{\substack{\bar{\mathsf{c}} \sim \overline{\mathsf{C}}, \bar{\mathsf{b}} \sim \overline{\mathsf{B}}(\bar{\mathsf{c}},\bar{\mathsf{c}}'_2) \\ \bar{\mathsf{a}} \sim \overline{\mathsf{A}}(\bar{\mathsf{c}}'_1,\bar{l}_2)}}\Big[\tilde{\beta}_2 \not\sim \tilde{\alpha} \Big| \tilde{\gamma}'_2 \sim \tilde{\alpha}\Big]\right]
\end{aligned}$$

By definition of $\overline{\mathsf{C}}'$, we have $\Pr_{\bar{\mathsf{c}}'_1,\bar{\mathsf{c}}'_2 \sim \overline{\mathsf{C}}'}[\mathsf{q}(\bar{\mathsf{c}}'_1, \bar{\mathsf{c}}'_2) < \varepsilon^5] \leq \sigma$ and also

$$\begin{aligned}
\mathbb{E}_{\bar{\mathsf{c}}'_1,\bar{\mathsf{c}}'_2 \sim \overline{\mathsf{C}}'}\left[\Pr_{\substack{\bar{\mathsf{c}} \sim \overline{\mathsf{C}}, \bar{\mathsf{b}}_1 \sim \overline{\mathsf{B}}(\bar{\mathsf{c}},\bar{\mathsf{c}}'_1) \\ \bar{\mathsf{a}} \sim \overline{\mathsf{A}}(\bar{l}_1,\bar{\mathsf{c}}'_2)}}\Big[\tilde{\beta}_1 \not\sim \tilde{\alpha} \Big| \tilde{\gamma}'_1 \sim \tilde{\alpha}\Big]\right] &\leq \max_{\bar{\mathsf{c}}'_1 \in \overline{\mathsf{C}}'}\left\{\Pr_{\substack{\bar{\mathsf{c}} \sim \overline{\mathsf{C}}, \bar{\mathsf{b}}_1 \sim \overline{\mathsf{B}}(\bar{\mathsf{c}},\bar{\mathsf{c}}'_1) \\ \bar{\mathsf{a}} \sim \overline{\mathsf{A}}(\bar{l}_1)}}\Big[\tilde{\beta}_1 \not\sim \tilde{\alpha} \Big| \tilde{\gamma}'_1 \sim \tilde{\alpha}\Big] + \delta\right\} \\
&= \max_{\bar{\mathsf{c}}'_1 \in \overline{\mathsf{C}}'}\left\{\Pr_{\substack{\bar{\mathsf{c}} \sim \overline{\mathsf{C}}, \bar{l}_1 \sim \overline{\mathsf{B}}_1(\bar{\mathsf{c}},\bar{\mathsf{c}}'_1) \\ \bar{\mathsf{a}}, \bar{\mathsf{a}}'' \sim \overline{\mathsf{A}}(\bar{l}_1)}}\Big[\tilde{\alpha} \not\approx \tilde{\alpha}' \Big| \tilde{\gamma}'_1 \sim \tilde{\alpha}\Big] + \delta\right\} \\
&= \max_{\bar{\mathsf{c}}'_1 \in \overline{\mathsf{C}}'}\left\{1 - \mathsf{p}(\bar{\mathsf{c}}'_1) + \delta\right\} \leq \eta + \delta.
\end{aligned}$$

We have used the sampling of $\overline{\mathsf{A}}(\bar{l})/\overline{\mathsf{C}}$ for all $\bar{l} \in \overline{\mathsf{B}}_1$, and that $\mathsf{p}(\bar{\mathsf{c}}'_1) \geq 1 - \eta$ since $\bar{\mathsf{c}}'_1 \in \overline{\mathsf{C}}'$.

The result follows:

$$\mathbb{E}_{\overline{c}'_1,\overline{c}'_2\sim\overline{C}'}\big[\mathsf{val}\big] \geq \big(1-\sigma\big)\cdot\big(1-2\varepsilon^{-5}(\eta+\delta)\big) \geq 1-\big(\sigma+2\varepsilon^{-5}(\eta+\delta)\big).$$

■

A.4 Sampler Replacement

In the body we used the following fact with $(\varepsilon',\delta')=(\varepsilon,\delta)$ and $\rho=\zeta=\varepsilon$.

Fact 7 (Restated). *Let $\varepsilon,\delta,\varepsilon',\delta',\varepsilon^*,\delta^*,\rho,\zeta>0$ be such that $\delta^*(\varepsilon^*-\varepsilon-\varepsilon'-2\rho-\zeta)\geq \delta'/\zeta+\delta/\rho$. Suppose $A/B/C$ is such that:*

- *A/C, B/C and $B(a)/C(a)$ are $0-$biregular for all $a\in A$; and*
- *A/C is $(\varepsilon,\delta)-$sampling and $A(c)/B(c)$ is $(\varepsilon',\delta')-$sampling for all $c\in C$.*

Then A/B is $(\varepsilon^*,\delta^*)-$sampling.

Proof. Fix $\varepsilon,\delta,\varepsilon',\delta',\varepsilon^*,\rho,\zeta>0$ and $A/B/C$ as in the statement. Let $B'\subset B$ be a set of size $|B'|=\lambda\cdot|B|$, and let $A'\subset A$ be the set of $a\in A$ such that $\big|\Pr_{b\sim B(a)}(b\in B')-\lambda\big|>\varepsilon^*$, let $\nu=|A'|/|A|$. We must show that $\nu\leq\big(\delta'/\zeta+\delta/\rho\big)/(\varepsilon^*-\varepsilon-\varepsilon'-2\rho-\zeta)$. We have

$$\begin{aligned}\varepsilon^* &< \mathbb{E}_{a\sim A'}\Big[\big|\Pr_{b\sim B(a)}(b\in B')-\lambda\big|\Big] \leq \mathbb{E}_{a\sim A'}\Big[\Big|\mathbb{E}_{c\sim C(a)}\big[\Pr_{b\sim B(a,c)}(b\in B')\big]-\lambda\Big|\Big]\\ &\leq \mathbb{E}_{\substack{a\sim A'\\ c\sim C(a)}}\Big[\big|\Pr_{b\sim B(a,c)}(b\in B')-\lambda(c)\big|\Big]+\mathbb{E}_{a\sim A'}\Big[\Big|\mathbb{E}_{c\sim C(a)}\big[\lambda(c)\big]-\mathbb{E}_{c\sim C}\big[\lambda(c)\big]\Big|\Big],\end{aligned}$$

where for $c\in C$, $\lambda(c):=\Pr_{b\sim B(c)}(b\in B')$. We have used the biregularity of $B(a)/C(a)$

for all $a \in A$ and that $\mathbb{E}_{c\sim C}\big[\lambda(c)\big] = \lambda$, which follows from biregularity of B/C. Let RHS_1 and RHS_2 be the two expectations on the right hand side of the equation above. We bound RHS_1 and RHS_2 separately. Note,

$$\mathsf{RHS}_2 \leq \varepsilon + 2\rho + \nu^{-1} \cdot \Pr_{a\sim A}\left[\left|\mathbb{E}_{c\sim C(a)}\big[\lambda(c)\big] - \mathbb{E}_{c\sim C}\big[\lambda(c)\big]\right| > \varepsilon + 2\rho\right] \leq \varepsilon + 2\rho + \nu^{-1} \cdot \delta/\rho.$$

Thus, it suffices to show that $\mathsf{RHS}_1 \leq \zeta + \varepsilon' + \nu^{-1} \cdot \delta'/\zeta$. Let $C' \subset C$ be the set of $c \in C$ such that $\Pr_{\substack{a\sim A' \\ c'\sim C(a)}}(c' = c) < \zeta/|C|$. Clearly, $\Pr_{\substack{a\sim A' \\ c\sim C(a)}}(c \in C') < \zeta$. Also, whenever $c \notin C'$, we have

$$\nu \cdot \zeta \leq \nu \cdot |C| \cdot \Pr_{\substack{a\sim A \\ c'\sim C(a)}}\big[c' = c \big| a \in A'\big] = \Pr_{\substack{c'\sim C \\ a\sim A(c')}}\big[a \in A' \big| c' = c\big] = \Pr_{a\sim A(c)}\big[a \in A'\big].$$

We have used the biregularity of A/C. This gives

$$\begin{aligned}
\mathsf{RHS}_1 &< \zeta + \varepsilon' + \max_{c\notin C'}\left\{\Pr_{a\sim A(c)}\left[\left|\Pr_{b\sim B(a,c)}(b \in B') - \lambda(c)\right| > \varepsilon'\right] \Big/ \Pr_{a\sim A(c)}(a \in A')\right\} \\
&\leq \zeta + \varepsilon' + \nu^{-1} \cdot \delta'/\zeta,
\end{aligned}$$

and the result follows. ■