Locally Testing Direct Products in the High Error Range

Irit Dinur * Elazar Goldenberg †

December 28, 2008

Abstract

Given a function $f : X \to \Sigma$, its ℓ-wise direct product is the function $F = f^\ell : X^\ell \to \Sigma^\ell$ defined by

$$F(x_1, \ldots, x_\ell) = (f(x_1), \ldots, f(x_\ell)).$$

In this paper we study the local testability of the direct product encoding (mapping $f \mapsto f^\ell$). Namely, given an arbitrary function $F : X^\ell \to \Sigma^\ell$, we wish to determine how close it is to f^ℓ for some $f : X \to \Sigma$, by making the smallest possible number of random queries into F (namely, two).

This question has first been studied by Goldreich and Safra and later the following simple two-query test has been studied by Dinur and Reingold: Choose a random pair $\mathbf{x}, \mathbf{x}' \in X^\ell$ that have m coordinates in common. Accept iff $F(\mathbf{x})$ and $F(\mathbf{x}')$ agree on the common coordinates. Dinur and Reingold showed that if the test accepts with sufficiently high probability (close to 1) then F is close to f^ℓ for some f.

In this work we analyze the case of low acceptance probability of the test. We show that even if the test passes with small probability, $\varepsilon > 0$, already F must have a non-trivial structure and in particular must agree with some f^ℓ on nearly ε of the domain. Moreover, we give a structural characterization of all functions F on which the test passes with probability ε. We find a list of functions $f_1, \ldots, f_t$ such that essentially the only way T' will accept on a pair $\mathbf{x}, \mathbf{x}'$, is if both $F(\mathbf{x})$ and $F(\mathbf{x}')$ agree with f_i. This is related to approximate local-decoding of this code, as studied by Impagliazzo et. al. Our result means that both the testing and the approximate local decoding can be done in "one shot" with the minimal possible number (only two) of queries.

Our results hold for values of ε as small as $\ell^{-\Omega(1)}$, and we show that below $1/\ell$ no characterization is possible.

*Weizmann Institute. Research supported by ISF and BSF grants. irit.dinur@weizmann.ac.il

†Hebrew University. elazargo@cs.huji.ac.il

1 Introduction

Given a function $f : X \to \Sigma$, its ℓ-wise direct product is the function $F = f^\ell : X^\ell \to \Sigma^\ell$ defined by

$$F(x_1, \ldots, x_\ell) = (f(x_1), \ldots, f(x_\ell)).$$

We think of $|X|$ as being very large compared to Σ, ℓ (for concreteness one may keep in mind $X = [n]$ and $\Sigma = \{0, 1\}$), and view the mapping $f \mapsto f^\ell$ as an encoding of f. This encoding is useful in various amplification scenarios, for example in PCP constructions where one wants to read many (up to ℓ) values of f without making many queries. Thus F can be viewed as a way to aggregate the answers of f, but there is a caveat: one must be able to test that F does not cheat, i.e., that F is 'faithful' to some underlying f.

In this paper we study the testability of the direct product encoding. Namely, given an arbitrary function $F : X^\ell \to \Sigma^\ell$, we wish to determine how close it is to f^ℓ for some $f : X \to \Sigma$, by making the smallest possible number of random queries into F, namely two.

This question has first been studied by Goldreich and Safra [GS97], who showed that this encoding is testable with a constant number of queries. A very simple two query test for this encoding was analyzed in [DR04], where it was shown that if the test succeeds with probability 99%, then F agrees with some f^ℓ on at least say 95% of the domain. This pretty much pinpoints functions F that pass the test with high probability. One may wonder which are the functions F that pass the test with an arbitrary probability ε? In this paper we answer this question for values of ε as small as $\ell^{-\Omega(1)}$.

One motivation for this question comes from Probabilistically Checkable Proofs (PCPs). It is easy to construct PCPs with small soundness error[1] just by sequential repetition of a PCP with constant soundness error. However, this increases the number of queries made to the proof. One way to reduce the number of queries is by replacing the proof f by its direct product encoding $F = f^\ell$. However, one must be able to test that the encoded proof F does not cheat, i.e., that F is 'faithful' to some underlying f. Moreover, since we are interested in small error, our test must be such that if it passes with probability above ε, then we can already conclude that F is sufficiently close to f^ℓ for some f. A similar property is known to hold for the low degree test and its accompanying encoding [RS97, AS97]. This plays a crucial role in the composition of PCPs based on the low degree test.

Thus our results are analogous to the small-error analysis of the low degree test [RS97, AS97]; and the direct product encoding can be viewed as a combinatorial alternative to the low degree encoding used in small-error PCP constructions. Presumably, one could incorporate this encoding in PCP constructions, but the details are beyond the scope of the current work.

The so-called low-error (or low-acceptance-probability) regime is often more difficult to analyze. One reason is the non-uniqueness of the solution: Clearly F can be a hybrid of $1/\varepsilon$ different legal codewords and still pass the test with probability ε. Thus, this is called the list-decoding regime since one can, at best, guarantee that success of the test implies existence of a list of codewords that have non-trivial agreement with the received word (F in our case).

[1] The soundness error is the probability that the verifier accepts when it should reject.

Let us now formally describe the test T and state our main theorem. Given a function $F : X^\ell \to \Sigma^\ell$, the test T has a parameter m which we fix to be $m = \ell^c$ for some constant $c = 19/75$, and is as follows:

1. Choose $\mathbf{x} \in X^\ell$ uniformly at random.
2. Choose a random set $I \subset [\ell]$, $|I| = m$, and choose a random $\mathbf{x}' \in X^\ell$ conditioned on $\mathbf{x}'_i = \mathbf{x}_i$ for all $i \in I$.
3. Accept iff $F(\mathbf{x})_I = F(\mathbf{x}')_I$.

This test makes two queries into F, at $\mathbf{x}$ and at $\mathbf{x}'$. If $F = f^\ell$ for some function f then clearly the test succeeds with probability one. In fact, even if $F = f_1 \times f_2 \times \cdots \times f_\ell$ for an ℓ-tuple $\vec{f} = (f_1, \ldots, f_\ell)$ of possibly distinct functions $f_i : X \to \Sigma$ (in the sense that $F(x_1, \ldots, x_\ell) = (f_1(x_1), \ldots, f_\ell(x_\ell))$) T still accepts with probability one. Our first theorem states that if T passes with probability ε then it is explained by closeness of F to $f_1 \times \cdots \times f_\ell$ on some $\varepsilon^{O(1)}$ fraction of the domain.

Theorem 1.1 *Let $F : \mathbf{X} \to \Sigma^\ell$. If T accepts F with probability ε, then there exists a tuple $\vec{f} = (f_1, \ldots, f_\ell)$ of functions $f_i : X \to \Sigma$ such that for $\Omega(\varepsilon^5)$ fraction of the tuples $\mathbf{x} \in \mathbf{X}$:*

$$\Pr_{i \in [\ell]}[F(\mathbf{x})_i = f_i(\mathbf{x}_i)] \geq 1 - O(\ell^{-\Omega(1)})$$

Let us make a couple of comments about the above theorem.

- The theorem concludes that on many of the tuples $\mathbf{x}$, $F(\mathbf{x}) \approx \vec{f}(\mathbf{x})$ rather than $F(\mathbf{x}) = \vec{f}(\mathbf{x})$. This weaker conclusion is inherent, as can be seen by taking $F = f^\ell$ and then changing each $F(\mathbf{x})$ arbitrarily in fewer than ℓ/m coordinates. Such a function F will pass the test with high probability, yet is only close to f^ℓ in the above sense.
- A second apparent weakness of this theorem is the fraction $\Omega(\varepsilon^5)$ of tuples that support $\vec{f}$ which fails to fully explain the ε success probability of T. Our second result is a stronger theorem (Theorem 1.3 below) that addresses this issue, and we turn to it shortly.

First, however, let us return to the question of testing whether F is close to the ℓ-th power f^ℓ of a single function $f : X \to \Sigma$ (rather than to $f_1 \times f_2 \times \cdots \times f_\ell$). For this we must consider the modified test T', which is the same as T except that the last step is now:

3′. Choose $s : [\ell] \to [\ell]$ to be a random permutation on $[\ell]$. Denote by $s(\mathbf{x}') \in X^\ell$ the vector defined by $s(\mathbf{x}')_i = \mathbf{x}'_{s(i)}$. Read $F(\mathbf{x})$ and $F(s(\mathbf{x}'))$ and accept iff for every $i \in I$ $F(\mathbf{x})_i = s^{-1}(F(s(\mathbf{x}')))_i$.

Clearly if $F = f^\ell$ then the test accepts always. We prove via reduction from the main theorem that,

Theorem 1.2 *Let $F : \mathbf{X} \to \Sigma^\ell$. If T' accepts F with probability ε , then there exists a function $f : X \to \Sigma$ such that for $\Omega(\varepsilon^6)$ fraction of the tuples $\mathbf{x} \in \mathbf{X}$ it holds:*

$$\Pr_{i \in [\ell]}[F(\mathbf{x})_i = f(\mathbf{x}_i)] \geq 1 - O(\ell^{-\Omega(1)})$$

The proof of this theorem encounters unexpected complications (see Section 4), and it is unclear whether these can be avoided. As previously mentioned, our stronger "structural characterization" below improves this theorem in that the agreement of F with f^ℓ goes from $\varepsilon^{O(1)}$ to $\varepsilon(1 - o(1))$. We remark that both T and T' were essentially considered in [DR04] modulo a slight technical difference, where their high-acceptance-probability (low error) behavior was analyzed.

1.1 The Structural Characterization

Our next result is stronger in that it characterizes (up to lower order terms) functions F on which T' accepts with probability ε. Consider the following "generic" construction of a function F on which T' accepts with probability ε. Choose functions $f_1, \ldots, f_t : X \to \Sigma$. For each function, fix a set $S_i \subseteq \mathbf{X}$ of tuples and set $F(\mathbf{x})$ approximately equal to $f_i(\mathbf{x})$ for all $\mathbf{x} \in S_i$. Outside $\cup S_i$ fix F randomly. Assuming first (for simplicity) that the f_i's are far from each other (hence the S_i's are roughly disjoint), it is easy to check that

$$\sum_i \Pr[\mathbf{x}, \mathbf{x}' \in S_i] \geq \varepsilon \qquad \Longrightarrow \qquad \Pr[T' \text{ accepts } F] \geq \varepsilon.$$

(neglecting an additive $\ell^{-\Omega(1)}$ term).

Our structural characterization can be viewed as an "inverse theorem" in that for any given F it finds functions f_i and supports $S_i \subseteq \mathbf{X}$ such that essentially the only way T' will accept on a pair $\mathbf{x}, \mathbf{x}'$, is if they both belong to S_i for some i,

$$\sum_i \Pr[\mathbf{x}, \mathbf{x}' \in S_i] \geq \varepsilon \qquad \Longleftarrow \qquad \Pr[T' \text{ accepts } F] \geq \varepsilon.$$

(again, neglecting an additive $\ell^{-\Omega(1)}$ term).

We also show that at least one f_i must agree with F on at least $\varepsilon(1 - o(1))$ of the domain. This is proven using the eigenvalues of the transition matrix of T'. The precise statement of our theorem is subtle, essentially since the functions f_i need not be far apart and this, in turn, causes the sets S_i to possibly intersect. An informal version is as follows (the precise statement appears as Theorem 5.1):

Theorem 1.3 (Structural Theorem - Informal Statement) *Let $F : X^\ell \to \Sigma^\ell$ be a function on which T' accepts with probability $\varepsilon > 0$. There is a list of functions $f_1, \ldots, f_t : X \to \Sigma$ and an i such that F agrees with $(f_i)^\ell$ on at least $\varepsilon(1 - o(1))$ of $\mathbf{X}$. Moreover, if T' accepts on $\mathbf{x}, \mathbf{x}'$ then with probability $1 - \ell^{-\Omega(1)}$, (i) $F(\mathbf{x})$ and $F(\mathbf{x}')$ agree with f_i for some $1 \leq i \leq t$; (ii) For each i such that $F(\mathbf{x})$ agrees with f_i also $F(\mathbf{x}')$ agrees with f_i.*

In the above, agreement is taken to be agreement on $1 - \ell^{-\Omega(1)}$ fraction of the coordinates.

One consequence of this result is that if T' accepts on $\mathbf{x}, \mathbf{x}'$ then we can *approximately locally decode* F back to f_i. The theorem guarantees that conditioned on T' accepting on $\mathbf{x}, \mathbf{x}'$, then almost surely there is some i such that for almost all j: the j-th coordinate of $F(\mathbf{x})$ equals $f_i(\mathbf{x}_j)$. Let us make a couple of remarks:

- This is related to the issue of locally list decoding which was defined in [STV01] in the context of hardness amplification. Locally list decoding the direct product encoding was studied in two relatively recent works [IJK06, IJKW08]. In that setting, F is already guaranteed to agree with f^ℓ on an ε fraction of the domain, and the goal is to generate (uniformly) a list of circuits that have oracle access to F, one of which computes f on almost all inputs. Our theorem complements this result in that it removes the need for the assumption about F being ε-close to f^ℓ (rather, we can test whether this holds). Moreover, both testing and decoding can be performed "in one shot" while making the smallest possible number of queries (i.e. two). In addition, it seems that Theorem 1.3 can also be used to give similar local decoding results, but we did not work out the details. We add that [IJKW08] were able to extend their results to derandomized direct products, and it would be extremely interesting to similarly derandomize our testing results.

- By reading $d \ll \ell$ coordinates of $F(\mathbf{x})$ we can obtain several values of f_i and ensure that (whp over the possible d-tuples X^d) nearly all of the d values are consistent with a single f_i while still making only two queries into F. Such "consistent reading" behavior is known for low degree tests (see [RS92, DFK$^+$99]) and is quite useful in composition of PCPs where consistency is a key issue.

1.2 Parallel Repetition

As mentioned above, direct products are often used for amplification. One of the most celebrated such results is the parallel repetition theorem of Raz [Raz98]. Without getting into the details let us mention that our test can be viewed as an ℓ-fold repetition of a single-coordinate 'equality' test. This test has perfect completeness and therefore cannot benefit from any parallel repetition theorem. Even if this wasn't the case, parallel repetition theorems could possibly bound the success of a test but would not provide any *structural information* about functions that pass with "non-negligible" probability.

Even so, one may hope to benefit from the proof techniques. Raz's techniques do not seem to help our setting, in particular since they are "too strong": they guarantee an upper bound that is exponentially small in ℓ, in contrast to the fact that the success probability of T is meaningful only if it is much larger, at least $1/\ell$ (see Section 6 for an appropriate example).

Nevertheless, an earlier proof of a (weaker) parallel repetition theorem due to Feige and Kilian [FK94] turns out to provide the key to our proof. We elaborate on this shortly.

1.3 Our Proof

The analysis of [DR04] in the high-acceptance-probability setting proceeds by defining a *majority* function f based on F, by taking for each x the most popular value among all tuples $\mathbf{x}$ containing

x. It is shown that if T accepts F with high enough (say 99%) probability then $F \approx f^\ell$. In our low-acceptance-probability setting such an approach cannot succeed, as can be seen by the following example: For each $\mathbf{x}$ let $F(\mathbf{x})$ be $(0,\ldots,0)$ with probability $\frac{1}{2}$ and $(1,\ldots,1)$ with probability $\frac{1}{2}$. Then T accepts with probability $\frac{1}{2}$ while the majority function f is a random function, and surely F is far from f^ℓ. Observe however, that this does not contradict our theorem as F is indeed close to two direct product functions: $\mathbf{0}^\ell$ and $\mathbf{1}^\ell$.

Locally testing a code in the list-decoding regime has been studied in the literature, for example in the high-error low degree test of [AS97, RS97]. The low degree polynomial codes have a high relative distance which is crucial for the low-degree test analyses. Indeed, we were set back by the observation that a combinatorial analysis a la Raz-Safra will not work here[2].

The key to our proof comes from the work of Feige and Kilian [FK94] in their analysis of parallel repetition games. They study parallel repetition of so-called miss/match games, and prove a structural dichotomy lemma which easily adapts to our setting. Essentially the lemma says the following: every function $F : X^\ell \to \Sigma^\ell$ either causes our test T to reject whp, or there are *many* exponentially-small sets $E \subset X^\ell$ on which $F \approx f_{1,E} \times \cdots \times f_{\ell,E}$ (however possibly each E has a distinct $\vec{f}_E = (f_{1,E},\ldots,f_{\ell,E})$).

This lemma leverages noticeable success of the test to deduce a certain structure for F. However, deducing structure on tiny (exponentially small) subsets E is not very useful unless, and this is the key point, these subsets can be glued together in a meaningful way. The main technical work in the proof of Theorem 1.1 goes to showing how to go from local to global agreement and to stitch the tiny E's together into one big set that agrees with a *single* direct product. We first show that some noticeable fraction of pairs E, E' intersect non-trivially. Then we deduce that such an intersection implies that $\vec{f}_E \approx \vec{f}_{E'}$. Finally we find a "popular" set E that agrees with sufficiently many of the E''s and proceed to prove that the function f_E agrees with F non-trivially on at least an $\varepsilon^{O(1)}$ fraction of X^ℓ.

We then extend this theorem in two ways to Theorems 1.2 and 1.3. The proofs of both of these theorems encounter unexpected complications on which we elaborate in the corresponding Sections 4 and 5.

Organization of the Paper Section 2 contains basic definitions and lemmas. Sections 3, 4, and 5 contain the proofs of Theorems 1.1, 1.2, and 1.3. We conclude with a discussion of the tightness of the parameters in Section 6. A proof of the Feige-Kilian lemma is included in Appendix A.

2 Preliminaries

There is a certain amount of freedom in choosing the parameters, and we have not made an attempt to optimize them. We require that $|X| > \ell^3$. Also throughout the paper always $2\ell^{-1/75} < \varepsilon < 1/48$ (we require ℓ to be large enough for this to be non-void). We fix the remaining parameters as follows: $\rho = \ell^{-1/75}$, $\eta = \ell^{-7/75}$, $m = \ell^{19/75}$.

[2]There are functions F that pass our test whp, but exhibit many "non-transitive triangles": triples $\mathbf{x},\mathbf{y},\mathbf{z} \in \mathbf{X}$ such that T accepts $\mathbf{x},\mathbf{y}$ and $\mathbf{x},\mathbf{z}$ but rejects $\mathbf{y},\mathbf{z}$.

2.1 General Definitions, Notations, and Tools

Denote $\mathbf{X} = X^\ell$. An element in $\mathbf{X}$ is called an ℓ-tuple or a tuple and is usually denoted by $\mathbf{x}, \mathbf{x}', \mathbf{y}$ etc. . The i-th coordinate of a tuple $\mathbf{x}$ is denoted by $\mathbf{x}_i \in X$. For a $k \in [\ell]$, the notion of a k-block is important in our proof:

Definition 2.1 *A k-block b is a pair $(y, \vec{\imath})$ where $\vec{\imath} = (i_1, \ldots, i_k)$ is a list of indices in increasing order $i_1 < i_2 < \cdots < i_k$ and $y \in X^k$.*

We use the letters b, b' to denote k-blocks, and unless stated otherwise we assume that $b = (y, \vec{\imath}), b' = (y', \vec{\imath}')$. For a pair of blocks b, b' we use the notation $b \cap b' = \emptyset$ if $\vec{\imath} \cap \vec{\imath}' = \emptyset$. The union of disjoint blocks is defined in the natural way: The set of indices is the concatenation of the indices in the original blocks in correct order, and the values of y, y' are concatenated appropriately.

For a tuple $\mathbf{x} \in \mathbf{X}$ and a set of indices $\vec{\imath} = (i_1, \ldots, i_k)$, we denote $\mathbf{x}_b = \mathbf{x}_{\vec{\imath}} = (x_{i_1}, \ldots, x_{i_k})$. We say that a tuple $\mathbf{x}$ contains a block b and denote $b \subset \mathbf{x}$ if $\mathbf{x}_{\vec{\imath}} = y$. For a k-block b, denote by

$$\mathbf{X}_b = \{\mathbf{x} \mid \ \mathbf{x}_{\vec{\imath}} = y\} \quad \text{and similarly} \quad \mathbf{X}_{b,b'} = \{\mathbf{x} \mid \ \mathbf{x}_{\vec{\imath}} = y, \mathbf{x}_{\vec{\imath}'} = y'\}\,.$$

In particular for a 1-block (x, i), $\mathbf{X}_{(x,i)} = \{\mathbf{x} \mid \ \mathbf{x}_i = x\}$. For $\mathbf{x} \in \mathbf{X}_b$ and $F : \mathbf{X} \to \Sigma^\ell$ we define $F(\mathbf{x})_b \stackrel{def}{=} F(\mathbf{x})_{\vec{\imath}}$.

The following lemma shows that for a Boolean function on $\mathbf{X}$ the expectation remains roughly the same when restricting to a random $\mathbf{X}_b$. It is used several times in the course of our proof, and is similar to a lemma in [FK94]. The proof for the precise statement can be found in [OG05]:

Lemma 2.2 *Let X be a set and $n > 1$ an integer, and denote $\mathbf{X} = X^n$. Let $f : \mathbf{X} \to \{0, 1\}$ with expectation $\mu = \mathbf{E}_{\mathbf{x} \in \mathbf{X}}[f(\mathbf{x})]$. For $(x, i) \in X \times [n]$, denote $\tilde{\mu}_{x,i} = \mathbf{E}_{\mathbf{x} \in \mathbf{X}_{(x,i)}}[f(\mathbf{x})]$. Then,*

1. $\Pr_{(x,i)}[|\mu - \tilde{\mu}_{x,i}| \geq 1/\sqrt[3]{n}] \leq 1/\sqrt[3]{n}$
2. $\mathbf{E}_{(x,i)}[(\tilde{\mu}_{x,i})^2] - \mu^2 \leq \frac{\mu}{n}$
3. *For $1 \leq r < n$ and an r-block $b = (y, \vec{\imath})$ denote $\tilde{\mu}_b = \mathbf{E}_{\mathbf{x} \in \mathbf{X}_b}[f(\mathbf{x})]$. Then,* $\Pr_b[|\mu - \tilde{\mu}_b| \geq r/\sqrt[3]{n-r}] \leq r/\sqrt[3]{n-r}$.

We conclude with the following standard bounds.

Lemma 2.3 (Chernoff Bound) *Let $x_1, ..., x_n$ i.i.d Bernoulli random variables having $\Pr[x_i = 1] = p$, then: $\Pr[|\sum x_i - pn| > \varepsilon n] < \exp(-\varepsilon^2 n/2)$.*

Lemma 2.4 (Chebyshev Bound) *Let X be a random variable with expectation μ and variance σ^2, then, for any $c > 0$:*

$$\Pr[|X - \mu| > c] < \frac{\sigma^2}{c^2}.$$

2.2 The Feige-Kilian Dichotomy Lemma

We now turn to describe the Dichotomy Lemma of Feige and Kilian which is the basis for our approach. Without getting into details, in their setting there is a game of questions and answers that is repeated ℓ times. Here a question would be an element of X, and an answer for it would be an element of Σ. More generally, given a k-block (which is essentially a tuple of questions), an answer for it is a tuple $a \in \Sigma^k$.

Fix $F : \mathbf{X} \to \Sigma^\ell$ for the rest of this section. Let $1 \leq k < \ell$ and let $b = (y, \vec{\imath})$ be a k-block. The following definitions are quoted from [FK94].

Definition 2.5 (Live Block) *A k-block b is* alive *if there exists an answer $a \in \Sigma^k$ such that $\Pr_{\mathbf{x}\in\mathbf{X}_b}[F(\mathbf{x})_b = a] \geq \varepsilon$. Such an answer a is called a live answer for b.*

Clearly, each block b can have at most $1/\varepsilon$ live answers.

Definition 2.6 *Let b be a block and $a \in \Sigma^k$, and let $0 \leq \eta < 1/2$. The pair $(x, i) \in X \times ([\ell] \setminus \vec{\imath})$ is $1 - \eta$ determined by (b, a) if there exists $\sigma \in \Sigma$ such that $\Pr_{\mathbf{x}\in\mathbf{X}_{b,(x,i)}}[F(\mathbf{x})_i = \sigma \mid F(\mathbf{x})_b = a] \geq 1 - \eta$.*

Recall that our goal is to find a direct product $g_1 \times \cdots \times g_\ell$ that agrees with F on a noticeable fraction of $\mathbf{X}$, for some $g_i : X \to \Sigma$. In the sequel we follow [FK94] who use notation $g : X \times [\ell] \to \Sigma$ to group together ℓ functions $g(\cdot, i) : X \to \Sigma$. Given such a g, we denote by $\vec{g} : \mathbf{X} \to \Sigma^\ell$ the function defined by

$$\forall \mathbf{x} = (x_1, \ldots, x_\ell) \in \mathbf{X} \qquad \vec{g}(\mathbf{x}) \stackrel{def}{=} (g(x_1, 1), g(x_2, 2), \ldots, g(x_\ell, \ell))$$

Definition 2.7 (Good Block) *A block b is good if b is alive and for every live answer a for it,*

$$\Pr_{(x,i)\in X\times([\ell]\setminus\vec{\imath})}[(x, i) \text{ is } 1 - \eta \text{ determined by } (b, a)] > 1 - \eta.$$

In that case a is called a good answer *for b, and we denote by $g_{b,a} : X \times ([\ell] \setminus \vec{\imath}) \to \Sigma$ the function assigning each (x, i) a value σ that maximizes the probability in Definition 2.6. $g_{b,a}$ is called the function that is $1 - \eta$ determined by (b, a).*

For a good block $b = (y, \vec{\imath})$ and good answer a the function $g_{b,a}$ is only defined on the domain $X \times ([\ell] \setminus \vec{\imath})$. Therefore, we arbitrarily extend each $g_{b,a}$ to the domain $X \times [\ell]$ demanding only $g_{b,a}(y_i, i) = a_i$ for each $(y_i, i) \in (y, \vec{\imath})$.

For two vectors $\mathbf{v}, \mathbf{w} \in \Sigma^\ell$ we write $\mathbf{v} \stackrel{1-\eta}{\approx} \mathbf{w}$ to denote $\Pr_{i\in[\ell]}[\mathbf{v}_i = \mathbf{w}_i] \geq 1 - \eta$.

Claim 2.8 *Let b be a good block with good answer a. Then for any $\rho > 0$,*

$$\Pr_{\mathbf{x}\in\mathbf{X}_b}[F(\mathbf{x}) \stackrel{1-\rho}{\approx} \vec{g}_{b,a}(\mathbf{x}) \mid F(\mathbf{x})_b = a] \geq 1 - \frac{2\eta}{\rho}$$

Proof: We denote $g = g_{b,a}$. For each $\mathbf{x} \in \mathbf{X}_b$ denote $\alpha_{\mathbf{x}} = \Pr_{i\in[\ell]}[F(\mathbf{x})_i = g(\mathbf{x}_i, i)]$. We first prove $\mathbf{E}_{\mathbf{x}\in\mathbf{X}_b}[\alpha_{\mathbf{x}}|F(\mathbf{x})_b = a] \geq 1 - 2\eta$. Let us define

$$D \stackrel{def}{=} \{(x,i) \in X \times ([\ell] \setminus \vec{\imath}) \mid \ (x,i) \text{ is } 1-\eta \text{ determined}\}$$

and note that choosing (x,i) from $X \times [\ell] \setminus \vec{\imath}$, we have $\Pr[(x,i) \in D] \geq 1 - \eta$ since b, a are good. We have

$$\begin{aligned}
\mathbf{E}_{\mathbf{x}\in\mathbf{X}_b}[\alpha_{\mathbf{x}}|F(\mathbf{x})_b = a] &\geq \Pr_{\mathbf{x}\in\mathbf{X}_b, i\in[\ell]\setminus\vec{\imath}}[F(\mathbf{x})_i = g(\mathbf{x}_i, i) \mid F(\mathbf{x})_b = a] \\
&= \Pr_{(x,i),\mathbf{x}\in\mathbf{X}_{b,(x,i)}}[F(\mathbf{x})_i = g(\mathbf{x}_i, i) \mid F(\mathbf{x})_b = a] \\
&\geq \Pr[(x,i) \in D] \cdot \Pr[F(\mathbf{x})_i = g(\mathbf{x}_i, i) \mid F(\mathbf{x})_b = a \text{ and } (x,i) \in D] \\
&\geq (1-\eta)^2 \geq (1-2\eta).
\end{aligned}$$

where the first inequality holds since for all $i \in \vec{\imath}$ equality trivially holds, and (x,i) is chosen from $X \times ([\ell] \setminus \vec{\imath})$. Finally, by Markov's inequality we get

$$\Pr_{\mathbf{x}\in\mathbf{X}_b}[F(\mathbf{x}) \stackrel{1-\rho}{\approx} \vec{g}_{b,a}(\mathbf{x}) \mid F(\mathbf{x})_b = a] = \Pr[\alpha_{\mathbf{x}} \geq 1 - \rho \mid F(\mathbf{x})_b = a] \geq 1 - \frac{2\eta}{\rho}$$

■

We can now state the Dichotomy Lemma:

Lemma 2.9 (Dichotomy Lemma of [FK94]) *Let $F : \mathbf{X} \to \Sigma^\ell$, and let $\varepsilon \geq 2\ell^{-1/75}$, then exactly one of following cases holds:*

1. *(Case 1) The probability that a random k-block is **alive** is at most ε.*
2. *(Case 2) The probability that a random live k-block is **good** is at least $1 - \varepsilon$.*

For completeness of presentation we include a proof of this lemma in Appendix A.

2.3 Agreement

The following definitions will be useful.

Definition 2.10 (Agreement) *Fix a k-block b. Let the agreement on b be defined as*

$$\mathcal{A}_k(b) = \Pr_{\mathbf{x},\mathbf{x}'\in\mathbf{X}_b}[F(\mathbf{x})_b = F(\mathbf{x}')_b]$$

and let

$$\mathcal{A}_k = \mathbf{E}_{b\,:\,|b|=k}[\mathcal{A}_k(b)]$$

Let k_1, k_2 be integers such that $k_1 + k_2 \leq \ell$. Let b_1 be a k_1-block and b_2 be a k_2-block such that $b_1 \cap b_2 = \emptyset$. Define

$$\mathcal{A}_{k_1,k_2}(b_1, b_2) = \Pr_{\mathbf{x},\mathbf{x}'\in\mathbf{X}_{b_1,b_2}}[F(\mathbf{x})_{b_1} = F(\mathbf{x}')_{b_1}]$$

Observation 2.11 *1.* $\Pr[T \text{ accepts}] = \mathcal{A}_m$

2. $\mathcal{A}_{k_1,k_2}(b_1, b_2) \geq \mathcal{A}_{k_1+k_2}(b_1 \cup b_2)$

Lemma 2.12 *Let $s, r > 0$ be integers, $s + r \leq \ell$. Fix an s-block b_1. Then, choosing b_2 disjoint from b_1,*

$$\mathbf{E}_{b_2}[\mathcal{A}_{s,r}(b_1, b_2)] - \mathcal{A}_s(b_1) \leq \frac{r}{\ell - (r+s)}.$$

Proof: Fix $a \in \Sigma^s$ and let $f_a : \mathbf{X}_{b_1} \to \{0, 1\}$ be defined by

$$f_a(\mathbf{x}) = 1 \quad \Leftrightarrow \quad F(\mathbf{x})_{b_1} = a$$

Denote $\mu_a = \mathbf{E}[f_a]$, note that:

$$\mathcal{A}_s(b_1) = \Pr_{\mathbf{x},\mathbf{x}' \in \mathbf{X}_{b_1}}[F(\mathbf{x})_{b_1} = F(\mathbf{x}')_{b_1}] = \sum_{a \in \Sigma^s} \mu_a^2.$$

Let us first prove the inequality for the case $r = 1$:
Given $(x, i) \in X \times ([\ell] \setminus \vec{\imath})$ let us denote $\Pr_{\mathbf{x} \in \mathbf{X}_{b_{1,(x,i)}}}[F(\mathbf{x})_{b_1} = a]$ by $\mu_{a,(x,i)}$:

$$\begin{aligned}
\mathbf{E}_{(x,i) \in X \times ([\ell] \setminus \vec{\imath})}[\mathcal{A}_{s,1}(b_1, (i, x))] - \mathcal{A}_s(b_1) &= \mathbf{E}_{(x,i) \in X \times ([\ell] \setminus \vec{\imath})} \sum_{a \in \Sigma^s} \mu_{a,(x,i)}^2 - \sum_{a \in \Sigma^s} \mu_a^2 \\
&= \sum_{a \in \Sigma^s} \mathbf{E}_{(x,i) \in X \times ([\ell] \setminus \vec{\imath})} \mu_{a,(x,i)}^2 - \mu_a^2 \\
\text{Using Lemma 2.2} &\leq \sum_{a \in \Sigma^s} \mu_a / (\ell - s) \\
\sum_{a \in \Sigma^s} \mu_a = 1 \to &= 1/(\ell - s)
\end{aligned}$$

Turning back to the general case, note that we can choose b_2 sequentially coordinate by coordinate and get a list of blocks in increasing order: $b_{2,1}, \cdots, b_{2,r}$. In order to bound $\mathbf{E}_{b_2}[\mathcal{A}_{s,r}(b_1, b_2)] - \mathcal{A}_s(b_1)$ we look at the telescoping series:

$$\mathbf{E}_{b_{2,r}}[\mathcal{A}_{s,r}(b_1, b_{2,r})] - \mathbf{E}_{b_{2,r-1}}[\mathcal{A}_{s,r-1}(b_1, b_{2,r-1})] + \mathbf{E}_{b_{2,r-1}}[\mathcal{A}_{s,r-1}(b_1, b_{2,r-1})] - \cdots - \mathcal{A}_s(b_1) \quad (1)$$

Note that for every $0 \leq i \leq r - 1$, we can bound

$$\mathbf{E}_{b_{2,r-i}}[\mathcal{A}_{s,r}(b_1, b_{2,r-i})] - \mathbf{E}_{b_{2,r-(i+1)}}[\mathcal{A}_{s,r-(i+1)}(b_1, b_{2,r-(i+1)})]$$

by $\frac{1}{(\ell-(s+i))}$ as we did in the proof for the case $r = 1$. Thus (1) can be bound by $\frac{r}{(l-(s+r))}$ and we are done. ∎

3 Local to Global

In this subsection we prove our first main result.

Theorem 1.1 *Let $F : \mathbf{X} \to \Sigma^\ell$ be a function that the test T accepts with probability 3ε, ($\varepsilon \geq 2\ell^{-1/75}$), then there exists a function $g : X \times [\ell] \to \Sigma$ such that, for $\varepsilon^5/16$ fraction of the tuples $\mathbf{x} \in \mathbf{X}$ it holds:*

$$F(\mathbf{x})_i \overset{1-9\rho}{\approx} \vec{g}(\mathbf{x}).$$

We begin with the following lemma, showing that if the test T accepts with some probability then most live blocks are good.

Lemma 3.1 *If the test T accepts a function F with probability 3ε, ($\varepsilon \geq 2\ell^{-1/75}$), then there exists $m/2 \leq k \leq m$, such that at least ε of the k-blocks are alive, and at least $1-\varepsilon$ of the live k-blocks are good.*

Proof: By Lemma 2.9 either Case 1 or Case 2 apply to F. If Case 1 doesn't apply, then there must be at least ε live k-blocks, of which at least $1-\varepsilon$ are good and the lemma is proven.

It remains to prove that if Case 1 applies to F, then the test T accepts with probability at most 3ε, thereby contradicting the hypothesis of the lemma.

So assume there are at most ε live k-blocks. Let us rewrite in which the way the test T selects the tuples $\mathbf{x}, \mathbf{x}'$:

1. Choose a random k-block b'.
2. Pick a random $m-k$-block b'', such that $b'' \cap b' = \emptyset$, and let $b = (y, \vec{t})$ be the m-block that is obtained from the union of b' and b''.
3. Pick a random $\mathbf{x} \in \mathbf{X}_b$.
4. Pick a random $\mathbf{x}' \in \mathbf{X}_b$.

Clearly b is a random m-block, and the distribution over $\mathbf{x}, \mathbf{x}'$ is identical to the distribution induced by T. Now we examine the probability of T in terms of the agreement.

$$\Pr[T \text{ accepts}] = \mathcal{A}_m = \mathbf{E}_b \mathcal{A}_m(b) = \mathbf{E}_{b',b''} \mathcal{A}_m(b' \cup b'') \leq \mathbf{E}_{b',b''} \mathcal{A}_{k,m-k}(b', b'') \tag{2}$$

Where the first equality and last inequality are obtained from Observation 2.11. We separate the expectation in (2) into two parts: the blocks b' that are alive, and those who are not.

The live blocks can contribute up to ε, since they appear with probability $\leq \varepsilon$.

It is left to bound the contribution of each of the non alive blocks: Let b' be a non-alive k block, then according to Lemma 2.12:

$$\mathbf{E}_{b''}[\mathcal{A}_{k,m-k}(b', b'')] - \mathcal{A}_k[b'] \leq \frac{m-k}{\ell - (m-k)} \leq \frac{m}{\ell/2} < \varepsilon$$

Since b' is non alive block, then $\mathcal{A}_k[b'] < \varepsilon$, and therefore: $\mathbf{E}_{b''}[\mathcal{A}_{k,m-k}(b', b'')] < 2\varepsilon$.

Altogether the expectation in (2) is bounded by 3ε. ∎

From now until the end of the proof, unless stated otherwise, a block is assumed to be a k-block. Let us define an indicator variable $I(\mathbf{x}, b, b', a, a')$ to be equal 1 iff $\mathbf{x} \in \mathbf{X}_{b,b'}$ and $F(\mathbf{x})_b = a$ and $F(\mathbf{x})_{b'} = a'$ and a, a' are good for b, b' respectively. Now set $I(\mathbf{x}, b, b') \stackrel{def}{=} \sum_{a,a'} I(\mathbf{x}, b, b', a, a')$. Clearly $I(\mathbf{x}, b, b')$ is either zero or one and it is one exactly if both $F(\mathbf{x})_b$ is a good answer for b and $F(\mathbf{x})_{b'}$ is a good answer for b'. Let $\mathcal{D}_1$ be a distribution on triples $(b, b', \mathbf{x})$ defined by choosing two random k-blocks b, b' such that $b \cap b' = \emptyset$ and a tuple $\mathbf{x}$ containing b, b'. (We recall that for blocks $b = (y, \vec{\imath})$ and $b' = (y', \vec{\imath}')$ $b \cap b' = \emptyset$ iff $\vec{\imath} \cap \vec{\imath}' = \emptyset$). We first prove that

Lemma 3.2 $\mathbf{E}_{(\mathbf{x},b,b') \sim \mathcal{D}_1}[I(\mathbf{x}, b, b')] \geq \varepsilon^2$.

We will then consider a graph whose vertices are the blocks and whose edges are roughly between pairs (b, b') for which $\Pr_{\mathbf{x}}[I(\mathbf{x}, b, b')]$ is large. We will then choose a block b^* that has maximal degree in this graph and prove that g_{b^*,a^*} is the global function we are seeking for an appropriate good answer a^*.

Proof: Let us examine another distribution $\mathcal{D}_2$ defined by first choosing a uniform tuple $\mathbf{x} \in \mathbf{X}$ and then choosing two blocks $b, b' \subset \mathbf{x}$ independently at random.

We prove Lemma 3.2 in two steps. First we prove that

$$\mathbf{E}_{(\mathbf{x},b,b') \sim \mathcal{D}_2}[I(\mathbf{x}, b, b')] \geq 3\varepsilon^2 \tag{3}$$

and then we argue that $\mathcal{D}_1, \mathcal{D}_2$ are close enough for our needs.

Let $a_{\mathbf{x}} = \Pr_{b \subset \mathbf{x}}[F(\mathbf{x})_b$ is alive for $b]$, and $g_{\mathbf{x}} = \Pr_{b \subset \mathbf{x}}[F(\mathbf{x})_b$ is good for $b]$. Observe that

$$\mathbf{E}_{\mathbf{x}}[(g_{\mathbf{x}})^2] = \Pr_{(b,b',\mathbf{x}) \sim \mathcal{D}_2}[I(\mathbf{x}, b, b')].$$

We would like to connect the probability that T accepts with the expectation of $a_{\mathbf{x}}$. However, the test T checks consistency on blocks of size m, while $a_{\mathbf{x}}$ refers to blocks of size k. Therefore, we consider a new test T_k which acts as follows:

- Choose a random k block b'.
- Pick a random $m - k$ block b'', such that $b' \cap b'' = \emptyset$, and let $b = b' \cup b''$.
- Pick $\mathbf{x}, \mathbf{x}' \in \mathbf{X}_b$ uniformly at random.
- Accept iff $F(\mathbf{x})_{b'} = F(\mathbf{x}')_{b'}$.

Claim 3.3

$$\Pr[T_k \textit{ accepts }] \geq \Pr[T \textit{ accepts }].$$

Proof: Using Observation 2.11, we get:

$$\Pr[T \text{ accepts}] = \mathcal{A}_m = \mathbf{E}_b \mathcal{A}_m(b) = \mathbf{E}_{b',b''} \mathcal{A}_m(b' \cup b'') \leq \mathbf{E}_{b',b''} \mathcal{A}_{k,m-k}(b', b'') = \Pr[T_k \text{ accepts}].$$

■

Let $s_{\mathbf{x}}$ denote the probability of T_k succeeding conditioned on choosing $\mathbf{x}$ as the first tuple.

$$\begin{aligned} s_{\mathbf{x}} &= a_{\mathbf{x}} \cdot \Pr[T_k \text{ succeeds on } b, \mathbf{x} \mid \mathbf{x}, F(\mathbf{x})_b \text{ is alive for } b] \\ &+ (1 - a_{\mathbf{x}}) \cdot \Pr[T_k \text{ succeeds on } b, \mathbf{x} \mid \mathbf{x}, F(\mathbf{x})_b \text{ is not alive for } b] \\ &\leq a_{\mathbf{x}} \cdot 1 + (1 - a_{\mathbf{x}}) \cdot \varepsilon \leq a_{\mathbf{x}} + \varepsilon \end{aligned}$$

So $a_{\mathbf{x}} \geq s_{\mathbf{x}} - \varepsilon$. Now $\mathbf{E}[a_{\mathbf{x}}] \geq \mathbf{E}[s_{\mathbf{x}}] - \varepsilon = \Pr[T_k \text{ succeeds}] - \varepsilon \geq 2\varepsilon$, and therefore $\mathbf{E}[(a_{\mathbf{x}})^2] \geq \mathbf{E}[a_{\mathbf{x}}]^2 \geq 4\varepsilon^2$. Note that from Lemma 3.1 we get that: $\mathbf{E}[g_{\mathbf{x}}] \geq (1-\varepsilon)\mathbf{E}[a_{\mathbf{x}}]$, yielding to $\mathbf{E}[(g_{\mathbf{x}})^2] \geq (1-\varepsilon)^2\mathbf{E}[(a_{\mathbf{x}})]^2 \geq 3\varepsilon^2$ So (3) is established.

Now we have to connect between the distributions $\mathcal{D}_1$ and $\mathcal{D}_2$: Define A as the event of selecting b, b' such that $b \cap b' = \emptyset$.

Claim 3.4 *Fix any values of b_1, b_2 and $\mathbf{x}_0 \supset b_1, b_2$ then:*

$$\Pr_{\mathcal{D}_1}[\mathbf{x} = \mathbf{x}_0 \textit{ and } b = b_1 \textit{ and } b' = b_2] = \Pr_{\mathcal{D}_2}[\mathbf{x} = \mathbf{x}_0 \textit{ and } b = b_1 \textit{ and } b' = b_2 | A]$$

Proof: If $b_1 \cap b_2 \neq \emptyset$, then under both of the distributions the above probability is 0, otherwise: $\Pr_{\mathcal{D}_1}[b = b_1 \text{ and } b' = b_2 \text{ and } \mathbf{x} = \mathbf{x}_0] = \binom{\ell}{k}^{-1} \cdot |X|^{-k} \cdot \binom{\ell-k}{k}^{-1} \cdot |X|^{-k} \cdot |X|^{-(\ell-2k)}$ and $\Pr_{\mathcal{D}_2}[\mathbf{x} = \mathbf{x}_0 \text{ and } b = b_1 \text{ and } b' = b_2 | A] = |X|^{-\ell} \cdot \binom{l}{k}^{-1} \cdot \binom{l-k}{k}^{-1}$. So equality holds. ∎

Let us calculate

$$\Pr_{\mathcal{D}_2}[A] = \frac{\binom{\ell}{k} \cdot \binom{\ell-k}{k}}{\binom{\ell}{k}^2} \geq (1 - 2k/\ell)^k \geq 1 - 2k^2/\ell. \tag{4}$$

Now let us calculate $\Pr_{\mathcal{D}_2}[I(\mathbf{x}, b, b') \mid A]$ using Bayes' rule:

$$\mathbf{E}_{(\mathbf{x},b,b') \sim \mathcal{D}_1}[I(\mathbf{x}, b, b')] = \Pr_{\mathcal{D}_2}[I(\mathbf{x}, b, b') | A] = \frac{\Pr_{\mathcal{D}_2}[I(\mathbf{x}, b, b') \text{ and } A]}{\Pr_{\mathcal{D}_2}[A]} \geq \Pr_{\mathcal{D}_2}[I(\mathbf{x}, b, b')] - \Pr_{\mathcal{D}_2}[\overline{A}]$$

Plugging in (3) and (4) we get a lower bound of ε^2, and we are done. ∎

Our next step is to define a graph on the blocks. Recall that the number of good answers for any block b is at most $1/\varepsilon$, since each good answer is also alive. Let us choose randomly for each good block b a good answer a_b (and an arbitrary answer for the non-good blocks). In expectation over these random choices, (and by Lemma 3.2)

$$\mathbf{E}_{(\mathbf{x},b,b') \sim \mathcal{D}_1}[I(\mathbf{x}, b, b', a_b, a_{b'})] \geq \frac{\varepsilon^2}{1/\varepsilon^2} = \varepsilon^4. \tag{5}$$

Therefore let us fix some deterministic choice of a_b per b that attains this expectation.

For blocks b_1, b_2, such that $b_1 \cap b_2 = \emptyset$ let $a_1 = a_{b_1}, a_2 = a_{b_2}$ and if b_1, b_2 are good then let $g_1 = g_{b_1,a_1}, g_2 = g_{b_2,a_2}$. Define $b_1 \sim b_2$ iff $\mathbf{E}_{\mathbf{x}}[I(\mathbf{x}, b_1, b_2, a_1, a_2)] \geq \varepsilon^4/2$. So by (5) and Markov's inequality $\Pr(b_1 \sim b_2) \geq \varepsilon^4/2$ where b_1, b_2 are random blocks such that $b_1 \cap b_2 = \emptyset$ (as implied by the distribution $\mathcal{D}_1$).

We will prove next that almost always if $b_1 \sim b_2$ then $g_1 \approx g_2$. Let us recall that since b_1 is a good block with good answer a_1 then by Claim 2.8,

$$\Pr_{\mathbf{x}\in\mathbf{X}_{b_1}}\left[F(\mathbf{x}) \overset{1-\rho}{\approx} \vec{g}_1(\mathbf{x}) \;\middle|\; F(\mathbf{x})_{b_1} = a_1\right] \geq 1 - \frac{2\eta}{\rho} \tag{6}$$

Suppose we could replace X_{b_1} by X_{b_1,b_2}, namely, prove that

$$\Pr_{\mathbf{x}\in\mathbf{X}_{b_1,b_2}}\left[F(\mathbf{x}) \overset{1-\rho}{\approx} \vec{g}_1(\mathbf{x}) \;\middle|\; F(\mathbf{x})_{b_1} = a_1\right] \geq 1 - \frac{2\eta}{\rho}. \tag{7}$$

The only difference between (6) and (7) is the domain from which $\mathbf{x}$ is chosen. Similarly suppose this could be done for b_2 and g_2. In that case we would be on our way to showing that in fact $g_1 \approx g_2$ essentially since $b_1 \sim b_2$ implies that on a non-negligible fraction of $\mathbf{x} \in X_{b_1,b_2}$, $F(\mathbf{x})$ agrees both with g_1 and with g_2.

So how do we convert (6) to (7)? The idea is that for a random b_2, $\mathbf{X}_{b_1,b_2}$ is a random restriction of X_{b_1} which cannot change probabilities too much:

Claim 3.5 *Fix a good block b and let $g = g_{b,a_b}$. Then for at least $1 - \frac{2k}{\sqrt[3]{\ell-k}}$ of the blocks b',*

$$\Pr_{\mathbf{x}\in\mathbf{X}_{b,b'}}\left[F(\mathbf{x}) \overset{1-\rho}{\approx} \vec{g}(\mathbf{x}) \;\middle|\; F(\mathbf{x})_b = a_b\right] \geq 1 - \frac{2}{\varepsilon}\left(\frac{2\eta}{\rho} + \frac{k}{\sqrt[3]{\ell-k}}\right) \geq 1 - 6\varepsilon^5$$

We defer the proof to the end of this section.

Constructing a graph on the blocks. We now construct a graph whose vertices are all the k-blocks in two steps. First, place an edge between b_1 and b_2 iff $b_1 \sim b_2$. Using (4) we know that

$$\Pr_{b_1,b_2}[b_1 \sim b_2] \geq \Pr_{b_1,b_2:b_1\cap b_2=\emptyset}[b_1 \sim b_2] - \Pr[b_1 \cap b_2 \neq \emptyset] \geq \varepsilon^4/2 - 2k^2/\ell.$$

Hence the graph is pretty dense. Next, for each block b remove (if exist) edges to all blocks b' which violate Claim 3.5, namely, blocks b' for which $\Pr_{\mathbf{x}\in\mathbf{X}_{b,b'}}\left[F(\mathbf{x}) \overset{1-\rho}{\approx} \vec{g}(\mathbf{x}) \;\middle|\; F(\mathbf{x})_b = a_b\right] < 1 - 6\varepsilon^5$. These are blocks on which the transition from $\mathbf{X}_b$ to $\mathbf{X}_{b,b'}$ causes a big change. Claim 3.5 implies that the fraction of edges removed is at most $\frac{4k}{\sqrt[3]{\ell-k}}$. The final graph has edge density at least $\varepsilon^4/2 - 2k^2/\ell - \frac{4k}{\sqrt[3]{\ell-k}} \geq \varepsilon^4/4$.

Concluding the Proof of Theorem 1.1 Let us fix b^* to be a vertex with maximal degree in this graph, and $g = g_{b^*,a_{b^*}}$ will be our global function. The last step in our proof is to show that

$$\Pr_{\mathbf{x}}[\vec{g}(\mathbf{x}) \overset{1-9\rho}{\approx} F(\mathbf{x})] \geq \varepsilon^5/16.$$

Let b be a neighbor of b^* in the graph. We first show that

$$\Pr_{(x,i)}[g_{b,a_b}(x,i) = g(x,i)] \geq 1 - 4\rho. \tag{8}$$

Indeed, by Claim 3.5 we know that

$$\Pr_{\mathbf{x}\in\mathbf{X}_{b,b^*}}\left[F(\mathbf{x}) \overset{1-\rho}{\approx} \vec{g}(\mathbf{x}) \;\middle|\; F(\mathbf{x})_{b^*} = a^*\right] \geq 1 - 6\varepsilon^5$$

and

$$\Pr_{\mathbf{x}\in\mathbf{X}_{b,b^*}}\left[F(\mathbf{x}) \overset{1-\rho}{\approx} \vec{g}_{b,a_b}(\mathbf{x}) \;\middle|\; F(\mathbf{x})_b = a_b\right] \geq 1 - 6\varepsilon^5.$$

On the other hand, since $b \sim b^*$ we know that $\Pr_{\mathbf{x}\in\mathbf{X}_{b,b^*}}[F(\mathbf{x})_b = a_b \text{ and } F(\mathbf{x})_{b^*} = a_{b^*}] \geq \varepsilon^4/2$. Putting these three equations together we deduce that on at least a fraction $\varepsilon^4/2 - 12\varepsilon^5 \geq \varepsilon^4/4$ of $\mathbf{X}_{b,b^*}$ we have $\vec{g}(\mathbf{x}) \overset{1-\rho}{\approx} F(\mathbf{x}) \overset{1-\rho}{\approx} \vec{g}_{b,a_b}(\mathbf{x})$, so $\vec{g}(\mathbf{x}) \overset{1-2\rho}{\approx} \vec{g}_{b,a_b}(\mathbf{x})$. We now need the following claim.

Claim 3.6 *Let $g_1, g_2 : X \times [t] \to \Sigma$ be two functions, and let $\beta = Pr_{(x,i)}[g_1(x,i) \neq g_2(x,i)] > 0$. The fraction of tuples $\mathbf{x} \in X^t$ for which $|\Pr_i[g_1(\mathbf{x}_i, i) \neq g_2(\mathbf{x}_i, i)] - \beta| \geq \beta/2$ is at most $\frac{4}{\beta t}$.*

Proof: Denote $B_i = B \cap X \times \{i\}$, and $\beta_i = \Pr_{x\in X}[(x,i) \in B_i]$. We define an indicator random variable $I_i(\mathbf{x})$ to equal 1 iff $(\mathbf{x}_i, i) \in B_i$, and let $I(\mathbf{x}) = \sum_{i=1}^t I_i(\mathbf{x})$. Clearly: $\mathbf{E}_{\mathbf{x}}[I_i(\mathbf{x})] = \beta_i$ and $\frac{1}{t}\sum \beta_i = \beta$.

Now we would like to examine the variance and the expectation of $I(\mathbf{x})$ in order to prove Claim 3.6.

$$\mathbf{E}_{\mathbf{x}} I(\mathbf{x}) = \sum \mathbf{E} I_i(\mathbf{x}) = \sum \beta_i = t\beta.$$

Also,

$$\mathbf{E}(I^2) = \mathbf{E}\left[\sum_{i\neq j} I_i(\mathbf{x}) I_j(\mathbf{x}) + \sum_i I_i(\mathbf{x})\right] = \sum_{i\neq j} \beta_i\beta_j + t\beta$$

and

$$\mathbf{Var}[I] = \mathbf{E}(I^2) - (\mathbf{E}I)^2 = t\beta - \sum_i \beta_i^2 \leq t\beta(1-\beta)$$

where the last inequality follows from the fact that $\sum \beta_i^2$ is minimized when all β_i are equal. Using Lemma 2.4 we get that

$$\Pr[|I - t\beta| > t\beta/2] < \frac{t\beta(1-\beta)}{t^2\beta^2/4} < \frac{4}{\beta t}$$

■

We apply Claim 3.6 on the space $\mathbf{X}_{b,b^*}$ (which for our purpose is the same as X^t with $t = \ell - 2k$). We deduce that if $\beta = \Pr[g(x,i) \neq g_{b,a_b}(x,i)] > 4\rho$ then the fraction of tuples $\mathbf{x} \in \mathbf{X}_{b,b^*}$ for which $\vec{g}(\mathbf{x}) \overset{1-2\rho}{\approx} \vec{g}_{b,a_b}(\mathbf{x})$ is at most $\frac{4}{4\rho(\ell-2k)}$, and cannot be as large as $\varepsilon^4/4$. So (8) is established.

Choose now a random block b and a random $\mathbf{x} \in \mathbf{X}_b$ (so clearly $\mathbf{x}$ is uniform in $\mathbf{X}$). b is a neighbor of b^* with probability at least $\varepsilon^4/4$. Conditioned on that and based on Claim 2.8, with probability at least $\varepsilon(1 - \frac{2\eta}{\rho}) > \varepsilon/2$ $\mathbf{x}$ is such that $F(\mathbf{x})_b = a_b$ and also $F(\mathbf{x}) \overset{1-\rho}{\approx} \vec{g}_{b,a_b}(\mathbf{x})$. Again by (8), using Claim 3.6 we know that the fraction of tuples on which $\vec{g}_{b,a_b}(\mathbf{x}) \overset{1-6\rho}{\not\approx} \vec{g}(\mathbf{x})$

is small ($< \frac{4}{\beta(\ell-2k)} < \varepsilon^{74}$). On all other tuples we must have (by the triangle inequality) that $F(\mathbf{x}) \overset{1-7\rho}{\approx} \vec{g}(\mathbf{x})$. Altogether, this holds for at least $\varepsilon^5/16$ of the tuples $\mathbf{x} \in \mathbf{X}$, and Theorem 1.1 is established. ∎

We now finish the proof of Claim 3.5.

Proof: (of Claim 3.5) Let us define a Boolean function $f : \mathbf{X}_b \to \{0,1\}$ by

$$f(\mathbf{x}) = 1 \iff F(\mathbf{x})_b = a_b \text{ and } F(\mathbf{x}) \overset{1-\rho}{\not\approx} \vec{g}(\mathbf{x})$$

Using Claim 2.8 we get:

$$\mu \overset{def}{=} \mathbf{E}_{\mathbf{x}\in\mathbf{X}_b} f(\mathbf{x}) \leq \Pr_{\mathbf{x}\in\mathbf{X}_b}[F(\mathbf{x})_b \neq a] \cdot 0 + \Pr_{\mathbf{x}\in\mathbf{X}_b}[F(\mathbf{x})_b = a] \cdot \frac{2\eta}{\rho} \leq \frac{2\eta}{\rho}.$$

Let b' be a block that is disjoint to b, denote $\tilde{\mu}_{b'} = \mathbf{E}_{\mathbf{x}\in\mathbf{X}_{b,b'}} f(\mathbf{x})$, we can use Lemma 2.2 in order to get that:

$$\Pr_{b' \,:\, b'\cap b=\emptyset}[\tilde{\mu}_{b'} > \frac{2\eta}{\rho} + \frac{k}{\sqrt[3]{\ell-k}}] \leq \Pr[|\tilde{\mu}_{b'} - \mu| \geq \frac{k}{\sqrt[3]{\ell-2k}}] \leq \frac{k}{\sqrt[3]{\ell-2k}}.$$

The function f indicates whether for a tuple $\mathbf{x}$ both $F(\mathbf{x})_b = a_b$ and $F(\mathbf{x}) \overset{1-\rho}{\approx} \vec{g}(\mathbf{x})$. However we want to compute the probability over $\mathbf{x} \in \mathbf{X}_{b,b'}$ $F(\mathbf{x}) \overset{1-\rho}{\approx} \vec{g}(\mathbf{x})$ conditioned on $F(\mathbf{x})_b = a$. In order to do it we compute the probability over $\mathbf{x} \in \mathbf{X}_{b,b'}$ that $F(\mathbf{x})_b = a$, and then use Bayes' Rule. We know that since a is good answer for b, and in particular a live answer, so $\Pr_{\mathbf{x}\in\mathbf{X}_b}[F(\mathbf{x})_b = a] \geq \varepsilon$.

Using the same calculation we have just performed for $\tilde{\mu}_{b'}$, we get that at most $\frac{k}{\sqrt[3]{\ell-k}}$ fraction of the blocks b' such that $b' \cap b = \emptyset$ also have:

$$\Pr_{\mathbf{x}\in\mathbf{X}_{b,b'}}[F(\mathbf{x})_b = a] < \varepsilon - \frac{k}{\sqrt[3]{\ell-k}}. \tag{9}$$

Now let us choose block b' satisfying both $\tilde{\mu}_{b'} < \frac{2\eta}{\rho} + \frac{k}{\sqrt[3]{\ell-k}}$ and (9), (which occurs with probability at least $1 - \frac{2k}{\sqrt[3]{\ell-k}}$) then we get:

$$\Pr_{\mathbf{x}\in\mathbf{X}_{b,b'}}\left[F(\mathbf{x}) \overset{1-\rho}{\approx} \vec{g}(\mathbf{x}) \mid F(\mathbf{x})_b = a_b\right] \geq 1 - \frac{\tilde{\mu}_{b'}}{\Pr_{\mathbf{x}\in\mathbf{X}_{b,b'}}[F(\mathbf{x})_b = a]} > 1 - \frac{\frac{2\eta}{\rho} + \frac{k}{\sqrt[3]{\ell-k}}}{\varepsilon/2}$$

∎

4 From T to T'

In this section we prove Theorem 1.2 that shows that if F passes the test T' then it noticeably agrees not just with $g : X \times [\ell] \to \Sigma$ (as in Theorem 1.1) but rather with f^ℓ for $f : X \to \Sigma$.

Theorem 1.2 *Let $F : \mathbf{X} \to \Sigma^\ell$ be a function that the test T' accepts with probability 5ε , $(\varepsilon \geq 2\ell^{-1/75})$, then there exists a function $f : X \to \Sigma$ such that, for $\varepsilon^6/256$ fraction of the tuples $\mathbf{x} \in \mathbf{X}$ it holds:*

$$F(\mathbf{x})_i \overset{1-23\rho}{\approx} \vec{f}(\mathbf{x}).$$

Let us begin with some notations: Let $s : [\ell] \to [\ell]$ be a permutation. For a vector v we denote $s(v)$ to be the vector defined by $(s(v))_i = v_{s(i)}$. We partition the space $\mathbf{X}$ into equivalence classes such that each class is the set of all permutations of a given $\mathbf{x}$:

$$C(\mathbf{x}) = \{s(\mathbf{x}) \mid \ s \text{ is a permutation } \}.$$

A function $G : \mathbf{X} \to \Sigma^\ell$ is called 'folded' if it is consistent on every equivalence class, i.e. for all $\mathbf{x}, s$: $G(s(\mathbf{x})) = s(G(\mathbf{x}))$.

The proof goes by reduction from T to T'. Namely, we randomly reduce F to a 'folded' function G. We then apply Theorem 1.1 on G and get a function $g : X \times [\ell] \to \Sigma$ that agrees on a non-negligible part of the domain of G, and with a little more work we get a function $g : X \to \Sigma$. For each G we get a (possibly) different g, so it is not immediate to deduce that F too agrees with g on a non-negligible part of the domain. Instead, we first show that the only way F can pass the test with good probability is if it is already somewhat "folded". In other words, on a random equivalence class there are relatively few different values that are supported by at least ε fraction of the class. It is then possible to relate the support of G to the support of F and deduce that F agrees with $\vec{g}$ noticeably.

Erasing Bad Tuples Let C be an equivalence class. Two tuples $\mathbf{x}, s(\mathbf{x}) \in C$ are said to agree if $s(F(\mathbf{x})) = F(s(\mathbf{x}))$. Similarly, $\mathbf{x}, s(\mathbf{x})$ agree on a block b if $F(\mathbf{x})_b = s^{-1}(F(s(\mathbf{x})))_b$ (note the order of $\mathbf{x}, s(\mathbf{x})$ is important). For a block b we say that an answer a is *rare* in C if the fraction of permutations s for which $s^{-1}(F(s(\mathbf{x})))_b = a$ is at most ε.

A tuple $\mathbf{x}$ is called bad if $F(\mathbf{x})_b$ is rare in $C(\mathbf{x})$ for at least $1 - \varepsilon$ of the blocks b.

- Consider T_1 that selects a random $\mathbf{x}$, $b \subset \mathbf{x}$ and already rejects if $F(\mathbf{x})_b$ is rare for b in the class $C(\mathbf{x}')$, otherwise continues like T'. Then $\Pr[T' \text{ accepts } F] - \varepsilon \leq \Pr[T_1 \text{ accepts } F]$.

  Consider T_2 that always rejects if it sees a bad tuple, otherwise continues like T_1. Then $\Pr[T_1 \text{ accepts } F] - \varepsilon \leq \Pr[T_2 \text{ accepts } F]$.

  Now define F_1 as follows: For all bad $\mathbf{x}$, let $F_1(\mathbf{x})$ be a random value in Σ^ℓ. Otherwise, let $F_1(\mathbf{x}) = F(\mathbf{x})$. Then $\Pr[T_2 \text{ accepts } F] \leq \Pr[T' \text{ accepts } F_1]$. We conclude so far that

  $$\Pr[T' \text{ accepts } F_1] \geq \Pr[T' \text{ accepts } F] - 2\varepsilon > 3\varepsilon.$$

- Assume $\mathbf{x}$ is not bad. We next argue that $F(\mathbf{x}) \overset{1-\rho}{\approx} s^{-1}(F(s(\mathbf{x})))$ for at least $\varepsilon/2$ of the permutations s. Indeed by definition, on at least ε of the blocks its answer $F(\mathbf{x})$ agrees with at least ε members of $\mathbf{y} \in C(\mathbf{x})$. Call these blocks B^*. Then the expectation choosing a random $b \in B^*$ and a random $\mathbf{y} \in C(\mathbf{x})$ that $\mathbf{x}$ agrees with $\mathbf{y}$ on b is at least ε. Therefore, $\mathbf{x}$

agrees with at least $\varepsilon/2$ of the tuples $\mathbf{y} \in C$ on at least $\varepsilon/2$ of the blocks in B^* which are at least $\varepsilon^2/2$ of the blocks in total. All in all

$$\Pr_s[\mathbf{x} \text{ agrees with } s(\mathbf{x}) \text{ on at least } \varepsilon^2/2 \text{ of the blocks}] \geq \varepsilon/2.$$

Our claim is complete once we prove,

Claim 4.1 *Let $\mathbf{x}, s(\mathbf{x})$ be tuples such that $\mathbf{x}$ agrees with $s(\mathbf{x})$ on at least $\varepsilon^2/2$ of the blocks, then:*

$$F(\mathbf{x}) \overset{1-\rho}{\approx} s^{-1}(F(s(\mathbf{x}))).$$

Proof: Let $I = \{i \mid \ F(\mathbf{x})_i \neq s^{-1}(F(s(\mathbf{x})))_i\}$. Assume for contradiction that $|I| > \rho\ell$. The probability that over the choices of the blocks $b \subset \mathbf{x}$, that b does not include even single coordinate from I is: $\frac{\binom{\ell-\rho\ell}{m}}{\binom{\ell}{m}} < (1-\frac{m}{l})^{\rho\ell} = ((1-\frac{m}{l})^{\frac{\ell}{m}})^{\rho m} < e^{-\rho m} \ll \varepsilon^2/2$, so we get a contradiction. ∎

Applying Theorem 1.1 Next, we define G from F_1 by choosing a random element $\mathbf{x}$ from each equivalence class C and setting $G(\mathbf{y}) = F_1(\mathbf{x})$ for all $\mathbf{y} \in C(\mathbf{x})$. Clearly

$$\mathbf{E}_G[T \text{ accepts } G] = \Pr[T' \text{ accepts } F_1] \geq 3\varepsilon.$$

Therefore there exists G that passes the test T with probability at least 3ε. By Theorem 1.1 we get a function $g : X \times [\ell] \to \Sigma$ for which

$$\Pr_{\mathbf{x}}[G(\mathbf{x}) \overset{1-7\rho}{\approx} \vec{g}(\mathbf{x})] \geq \varepsilon^5/16$$

Lemma 4.2 *There exists $i \in [\ell]$ such that for $f : X \to \Sigma$ defined by $f(x) \overset{def}{=} g(x,i)$, we have*

$$\Pr_{\mathbf{x}}[G(\mathbf{x}) \overset{1-22\rho}{\approx} \vec{f}(\mathbf{x})] \geq \varepsilon^5/64$$

Proof: We first apply a simple Markov argument implies that since $\vec{g}(\mathbf{x}) \overset{1-7\rho}{\approx} G(\mathbf{x})$ on $\varepsilon^5/16$ of the tuples $\mathbf{x}$, we deduce that for at least $\varepsilon^5/32$ of the equivalence classes, $\Pr_s[G(s(\mathbf{x})) \overset{1-7\rho}{\approx} \vec{g}(s(\mathbf{x}))] \geq \varepsilon^5/32$.

Next, fix such an equivalence class C, and assume that $\mathbf{x} \in C$ is such that $G(\mathbf{x}) \overset{1-7\rho}{\approx} \vec{g}(\mathbf{x})$. Define for each $i \in [\ell]$ the set $B_i = \{j \mid \ g_i(\mathbf{x}_j) \neq g_j(\mathbf{x}_j)\}$. We claim that

$$\Pr_i[|B_i| < 3 \cdot 7\rho \cdot \ell] > \tfrac{1}{2}. \tag{10}$$

Assuming (10), for a random $i \in [\ell]$, we will have $\vec{g}_i(\mathbf{x}) \overset{1-21\rho}{\approx} \vec{g}(\mathbf{x})$ with probability at least a $\frac{1}{2}$. This holds separately for each good equivalence class, so altogether there must be an i for which $\Pr_{\mathbf{x}}[G(\mathbf{x}) \overset{1-22\rho}{\approx} \vec{g}_i(\mathbf{x})] \geq \varepsilon^5/64$ which is what we wanted to prove.

It remains to prove (10). Assuming that it is false, we will show that for nearly all of the permutations s, $\mathbf{y} = s(\mathbf{x})$ hits many B_i's in the sense that $g_j(y_j) \neq g_i(y_j)$ and so $\vec{g}(s(\mathbf{x}))$ must disagree a lot with $G(s(\mathbf{x}))$ $(= s(G(\mathbf{x}))$ since G is folded). This contradicts the choice of C as a class in which $\Pr_s[G(s(\mathbf{x})) \overset{1-7\rho}{\approx} \vec{g}(s(\mathbf{x}))] \geq \varepsilon^5/32$.

For a random permutation $s : [\ell] \to [\ell]$ define an indicator variable $I_{B_i}(s)$ to equal 1 iff $s(i) \in B_i$, and define also the random variable $I_B = \sum_i I_{B_i}$. Clearly $\mathbf{E}[I_{B_i}] := \beta_i = \frac{|B_i|}{\ell}$ and $\mathbf{E}[I_B] = \sum_i \beta_i$. If (10) fails then

$$\beta := \frac{1}{\ell}\mathbf{E}_s[I_B(s)] \geq \tfrac{1}{2} \cdot 21\rho.$$

If s is such that $G(s(\mathbf{x})) \overset{1-7\rho}{\approx} \vec{g}(s(\mathbf{x}))$ then $I_B(s)$ deviates by more than $\frac{21}{2}\rho\ell - 7\rho\ell = \frac{7}{2}\rho\ell$ from its expectation, which we show has very small probability. In particular, it cannot hold for $\varepsilon^5/32$ of the permutations. We will show that $\mathbf{Var}[I_B] < 2\beta\ell$, and plugging in Chebyshev's inequality as in Lemma 2.4 we get that:

$$\Pr[I_B(s) \leq 7\rho\ell] \leq \Pr[|I_B - \mathbf{E}I_B| < 7/2\rho\ell] \leq \frac{\mathbf{Var}(I_B)}{(7/2\rho\ell)^2} < 2(\rho\ell)^{-1} \ll \varepsilon^5/32$$

so we reach a contradiction.

It remains to upper bound $\mathbf{Var}[I_B] = \mathbf{E}[I_B^2] - \mathbf{E}^2[I_B]$. Observe that for $i \neq j$ $\mathbf{E}[I_{B_i} I_{B_j}] \leq \frac{|B_i||B_j|}{\ell(\ell-1)} = \beta_i\beta_j\frac{\ell}{\ell-1}$. Now

$$\begin{aligned}
\mathbf{Var}[I_B] = \mathbf{E}[I_B^2] - \mathbf{E}^2[I_B] &= \sum_{i\neq j} \beta_i\beta_j(\frac{\ell}{\ell-1} - 1) + \sum_i \beta_i - \beta_i^2 \\
&\leq \sum_{i\neq j} \beta_i\beta_j\frac{1}{\ell-1} + \ell\beta(1-\beta) \\
&\leq \frac{1}{\ell-1}\max_j \beta_j \cdot \sum_{i,j,i\neq j} \beta_i + \beta\ell \\
&\leq \frac{1}{\ell-1} \cdot 1 \cdot (\ell-1)\ell\beta + \beta\ell = 2\beta\ell
\end{aligned}$$

■

In order to prove Theorem 1.2 we now claim that

$$\Pr_{\mathbf{x}}[F(\mathbf{x}) \overset{1-23\rho}{\approx} \vec{f}(\mathbf{x})] \geq \frac{\varepsilon^6}{256}$$

We split the analysis to bad and non-bad tuples $\mathbf{x}$ for which $G(\mathbf{x}) \overset{1-22\rho}{\approx} \vec{f}(\mathbf{x})$.

- If $G(\mathbf{x})$ was defined based on a tuple that is not bad, then there are at least $\varepsilon/2$ permutations s for which $s^{-1}(F(s(\mathbf{x}))) \overset{1-\rho}{\approx} F(\mathbf{x})$. For these s's, $F(s(\mathbf{x})) \overset{1-23\rho}{\approx} \vec{f}(\mathbf{y})$.

- If $G(\mathbf{x})$ was defined based on a bad tuple then it is a completely random value. For any fixed $\vec{f}$ the fraction of such tuples $\mathbf{x}$ is expected to be negligible ($\sum_{i=0}^{22\rho\ell}\binom{\ell}{i}|\Sigma|^{-(\ell-i)}$) , so $\Pr_{\mathbf{x}}[G(\mathbf{x}) \overset{1-22\rho}{\approx} \vec{f}(\mathbf{x})$ and $G(\mathbf{x})$ defined based on a non-bad tuple$] \geq \varepsilon^5/128$.

We conclude that

$$\Pr_{\mathbf{x}}[F(\mathbf{x}) \overset{1-23\rho}{\approx} \vec{f}(\mathbf{x})] \geq \frac{\varepsilon}{2}\cdot\Pr_{\mathbf{x}}[G(\mathbf{x}) \overset{1-22\rho}{\approx} \vec{f}(\mathbf{x}) \text{ and } G(\mathbf{x}) \text{ defined based on a non-bad tuple}] = \frac{\varepsilon^6}{256}.$$

5 The Structural Theorem

We have already seen (in Theorem 1.2) that if T' accepts the function F with probability ε, then there exists a function $f : X \to \Sigma$ such that $F(\mathbf{x}) \overset{1-O(\rho)}{\approx} \vec{f}(\mathbf{x})$ for $\Omega(\varepsilon^6)$ fraction of the tuples $\mathbf{x} \in \mathbf{X}$. In this section we fully characterize the structure of all functions F on which T' accepts with probability ε. Consider the following "generic" example for such a function F: Choose functions $f_1, \ldots, f_t : X \to \Sigma$. For each function, fix a set $S_i \subseteq \mathbf{X}$ of tuples and set $F(\mathbf{x}) := \vec{f_i}(\mathbf{x})$ for all $\mathbf{x} \in S_i$. Outside $\cup S_i$ fix F randomly. Assume that the S_i's are pairwise disjoint. It is then easy to check that

$$\sum_i \Pr[\mathbf{x}, \mathbf{x}' \in S_i] \geq \varepsilon \qquad \Longrightarrow \qquad \Pr[T' \text{ accepts } F] \geq \varepsilon.$$

(neglecting an additive $\ell^{-\Omega(1)}$ term).

Our structural characterization can be viewed as an "inverse theorem" in that for any given F it finds functions f_i and supports $S_i \subseteq \mathbf{X}$ (on which $F(\mathbf{x}) \approx \vec{f_i}(\mathbf{x})$) such that essentially the only way T' will accept on a pair $\mathbf{x}, \mathbf{x}'$, is if they both belong to S_i for some i, namely:

$$\sum_i \Pr[\mathbf{x}, \mathbf{x}' \in S_i] \geq \varepsilon \qquad \Longleftarrow \qquad \Pr[T' \text{ accepts } F] \geq \varepsilon.$$

(again, neglecting an additive $\ell^{-\Omega(1)}$ term). In fact, we prove a stronger statement: whenever T' accepts on $\mathbf{x}, \mathbf{x}'$ then (i) there is an i for which $\mathbf{x}, \mathbf{x}' \in S_i$ and (ii) for all j s.t. $\mathbf{x} \notin S_j$ also $\mathbf{x}' \notin S_j$. This implies the following consistency behavior: if we condition on $\mathbf{x} \in S_i$ for a fixed i, then T' will whp only accept pairs $\mathbf{x}, \mathbf{x}'$ for which also $\mathbf{x}' \in S_i$. Two subtle issues need to be addressed:

1. The number of f_i's that agree with F on a non-negligible fraction of $\mathbf{X}$ can be huge, if we allow f_i's that are too close to each other. One would like to "cluster" the similar f_i's and place one representative from each cluster in the final list. This is tricky but possible, as one needs to ensure that the different clusters are "well separated" so that whenever $\mathbf{x}$ supports the cluster of f_i and $\mathbf{x}'$ does not, T' will reject whp.

2. The next subtlety lies with the possible overlap between the S_i's. Even after clustering has been performed, it may happen that S_i and S_j will have a large intersection. In that case the events $\mathbf{x}, \mathbf{x}' \in S_i$ are not disjoint for different i's, and possibly even $\sum_i \Pr[\mathbf{x}, \mathbf{x}' \in S_i] \gg 1$.

A finer statement is obtained by considering all possible intersections

$$R_J = (\cap_{j \in J} S_j) \cap (\cap_{j \notin J} \overline{S}_j) \qquad J \subseteq [t]$$

noting that the R_J's are disjoint. We show that whp if T' accepts on a pair $\mathbf{x}, \mathbf{x}'$ then they both belong to exactly the same R_J, and $J \neq \emptyset$.

For $f : X \to \Sigma$ and $\gamma \in (0,1)$ we denote $\mathrm{supp}_\gamma(f) \stackrel{def}{=} \{\mathbf{x} \in \mathbf{X} \mid F(\mathbf{x}) \stackrel{1-\gamma}{\approx} \vec{f}(\mathbf{x})\}$ and say that $\mathbf{x}$ γ-supports f if $\mathbf{x} \in \mathrm{supp}_\gamma(f)$. Throughout this section we use the following parameters: $\rho_0 = 23\rho, \delta = \rho_0^8$ and $\varepsilon_0 = 2\ell^{-1/75}$.

Theorem 5.1 (Formal Version of Theorem 1.3) *Let $F : \mathbf{X} \to \Sigma^\ell$ be a function that the test T' accepts with probability $\alpha > \ell^{-1/150}$. Then there exist functions $f_1, \ldots, f_t : X \to \Sigma$ (with $t < \ell^{O(1)}$) and radii $\rho_1, \ldots, \rho_t \in [\rho_0, 2\rho_0]$ such that the following holds. Let $S_j = \mathrm{supp}_{\rho_j}(f_j)$, and for each $J \subseteq [t]$, let $R_J = (\cap_{j \in J} S_j) \cap (\cap_{j \notin J} \overline{S}_j)$. Then*

1. $1 \geq \sum_{J \neq \emptyset} \Pr_{\mathbf{x},\mathbf{x}'}[\mathbf{x}, \mathbf{x}' \in R_J \mid T' \text{ accepts on } \mathbf{x}, \mathbf{x}'] \geq 1 - O(\rho_0 + \varepsilon_0)/\alpha = 1 - \ell^{-\Omega(1)}$.
2. *Set $\varepsilon_J = |R_J| / |\mathbf{X}|$. Then $\Pr_{\mathbf{x},\mathbf{x}'}[\mathbf{x}, \mathbf{x}' \in R_J] \approx (\varepsilon_J)^2$, and $\sum_{J \neq \emptyset} (\varepsilon_J)^2 \geq \alpha(1 - \ell^{-\Omega(1)})$. In particular, there is some $J \neq \emptyset$ for which $\varepsilon_J \geq \alpha(1 - \ell^{-\Omega(1)})$.*

Let us make a few remarks.

- Item 1 implies that conditioned on T' accepting, the queried inputs $\mathbf{x}, \mathbf{x}'$ must whp come from the support of the same non-empty collection of functions $\{f_j\}_{j \in J}$. Since both sides of the inequality are roughly 1 this fully explains the success probability of T'.

  Item 2 further claims that the probability that both $\mathbf{x}$ and $\mathbf{x}'$ are chosen in a set R_J is as if they were independent samples, and deduces a nearly tight quantitative lower bound on the possible sizes of the sets R_J.

- We remark that in the informal version we only claimed that

$$\sum_i \Pr_{\mathbf{x},\mathbf{x}'}[\mathbf{x}, \mathbf{x}' \in S_i \mid T' \text{ accepts on } \mathbf{x}, \mathbf{x}'] \geq 1 - O(\rho_0) = 1 - \ell^{-\Omega(1)}.$$

  Since $\cup S_i = \cup R_J$ this follows from the above. However, it is possibly weaker as discussed above, and this would not enable finding any i for which approximately $|S_i| \geq \alpha |\mathbf{X}|$.

We now turn to explain the proof of this theorem. We begin by describing a straightforward way to prove this theorem and where it fails. Suppose one carries out the following iterative algorithm. Choose a function g_1 that is ρ_0-supported by the largest fraction of tuples, and let S_1 be the set of tuples ρ_0-supporting it. By Theorem 1.2, S_1 consists of $\Omega(\alpha^6)$ of the tuples. We can now "erase" F on S_1 (simply by replacing F's value on those tuples by random values) and repeat. If T' accepts the new F with high enough probability (above some threshold, say ε_0), we find g_2 and S_2 and continue. Since the S_i's are essentially disjoint we will halt after at most $1/\varepsilon_0^6$ steps. At this time,

if T' accepts on $\mathbf{x}$ and $\mathbf{x}'$ then each of them must support some g_i (except with small probability) since the iterative procedure terminated. We now need to rule out the case where $\mathbf{x}$ supports g_1 and say $\mathbf{x}'$ does not support g_1 (but supports say, g_2).

However, this need not hold. It is possible to have a large portion of the tuples support exactly one of g_1, g_2 while nearly supporting the other one (say the distance between $F(\mathbf{x})$ and $\vec{g}_1(\mathbf{x})$ is r which falls within the support threshold, and between $F(\mathbf{x})$ and $\vec{g}_2(\mathbf{x})$ is $r+1$ which falls outside the support threshold). Two such tuples $\mathbf{x}, \mathbf{x}'$ might easily cause T' to accept.

The essence of the problem is that contrary to the "usual scenario" in locally testing of codes, the direct product encoding does not have a large enough distance between distinct legal codewords (f^ℓ, g^ℓ may be close for $f \neq g$).

In our solution we manage to gather the functions g_i into clusters, such that each cluster has a representative f_j and a radius $\rho_0 \leq \rho_j \leq 2\rho_0$, and we set $S_j = \mathrm{supp}_{\rho_j}(f_j)$. The S_j's enjoy the property that their boundaries are nearly empty, where the boundary is the set of tuples $\mathbf{x}$ for which $F(\mathbf{x})$ disagrees with $\vec{f_j}(\mathbf{x})$ on $u \in (\rho_j \ell, \rho_j \ell + \delta\ell)$ coordinates. This eliminates the aforementioned obstacle and allows us to complete the proof. The proof of the second item relies on the fact that the transition matrix underlying our test has a large spectral gap to show that $\Pr[\mathbf{x}, \mathbf{x}' \in R_J] \approx (\varepsilon_J)^2$.

5.1 The iterative Construction of f_i's

We use the following iterative procedure. Since we will be considering functions F and G, let us denote $\mathrm{supp}^G_\gamma(f) = \left\{ \mathbf{x} \in X \;\middle|\; G(\mathbf{x}) \overset{1-\gamma}{\approx} \vec{f} \right\}$. We also use $\mathrm{supp}_\gamma(f)$ to implicitly mean $\mathrm{supp}^F_\gamma(f)$.

1. Set $i = 1$. Let $G = F$.
2. As long as T' accepts G with probability at least ε_0:
3. Choose $f_i : X \to \Sigma$ that maximizes $\left|\mathrm{supp}^G_{\rho_0}(f_i)\right|$.
4. Find the smallest radius ρ_i, $\rho_i = \rho_0 + n\delta$ ($n \in \mathbb{N}$), such that, $\left|\mathrm{supp}^F_{\rho_i+\delta}(f_i) \setminus \mathrm{supp}^F_{\rho_i}(f_i)\right| < \rho^7 |\mathbf{X}|$.
5. Randomize each $\mathbf{x} \in \mathrm{supp}^G_{\rho_i}(f_i)$ by setting $G(\mathbf{x})$ to a random value in Σ^ℓ.
6. Set $i = i + 1$ and return to step 2.

5.2 Proof of Item 1 of Theorem 5.1

Before we proceed with the proofs we would like to point out some of the properties that are obtained by the construction:

- Now, since $\left|\mathrm{supp}^F_\gamma(f)\right|$ of any function is bounded by $|\mathbf{X}|$, then in $1/\rho_0^7$ steps of increasing ρ_i by δ we must have that for at least one step r $\left|\mathrm{supp}^F_{r+\delta}(f_i) \setminus \mathrm{supp}^F_r(f_i)\right| < \rho_0^7 |\mathbf{X}|$. Moreover, since $\delta = \rho_0^8$ we get that for each i, ρ_i is in the range $[\rho_0, 2\rho_0]$.

- Note that in each iteration i we can apply Theorem 1.2 and get that

$$\left|\mathrm{supp}^G_{\rho_i}(f_i)\right| \geq \left|\mathrm{supp}^G_{\rho_0}(f_i)\right| \geq \varepsilon_0^6 \left|\mathbf{X}\right| / 256.$$

- In each iteration we "randomize" a fraction of $\Omega(\varepsilon_0^6)$ of the domain of G. It is easy to see that the fraction of tuples that were randomized and are randomized again is negligible. So the number of iterations is bounded by $t = O(1/\varepsilon_0^6)$.

Lemma 5.2 *Assume T' accepts on querying $\mathbf{x}, \mathbf{x}'$, then with probability at least $1 - 2\rho_0^7$ if $\mathbf{x} \in \mathrm{supp}_{\rho_i}(f_i)$ then $\mathbf{x}' \in \mathrm{supp}_{\rho_i}(f_i)$.*

Proof: If $\mathbf{x} \in \mathrm{supp}_{\rho_i}(f_i)$ then we know $F(\mathbf{x}) \overset{1-\rho_i}{\approx} \vec{f_i}(\mathbf{x})$. We argue that except with probability $\exp(-(\Theta(\delta^2 m)))$ over the choices of a random $b \subset \mathbf{x}$, $F(\mathbf{x})_b$ has at most $(\rho_i + \delta/3)m$ coordinates $\vec{\iota_j}$ on which $F(\mathbf{x})_{\vec{\iota_j}} \neq f_i(\mathbf{x}_{\vec{\iota_j}})$. In order to prove it we would like to use Chernoff bounds. However, there is a dependence between choosing each coordinate of b. Therefore, we consider two experiments. In the first experiment we choose the block b coordinate by coordinate, and denote the random variable b_j to equal 1 if on the j-th coordinate $\vec{\iota_j}$ of b $F(\mathbf{x})_{\vec{\iota_j}} \neq f_i(\mathbf{x}_{\vec{\iota_j}})$. In the second experiment we flip a biased coin m times, where the probability for 1 is $\rho_i + m/(\ell - m)$ (which is larger than the probability for b_j equal 1 even conditioned that all the rest are 0) and denote by a_j the random variable indicating whether in the j-th flip we get a 1. Then we argue:

$$\begin{aligned}
\Pr\left[\frac{1}{m}\sum b_i > \rho_i + \delta/3\right] &\leq \Pr\left[\frac{1}{m}\sum a_i > \rho_i + \delta/3\right] \\
&\leq \Pr\left[\frac{1}{m}\left|\sum a_i - \mathbf{E}\sum a_i\right| > \delta/3 - m/(\ell - m) > \delta/4\right] \\
&< \exp(-(\Theta(\delta^2 m)))
\end{aligned}$$

where the last equality follows from a Chernoff bound as in Lemma 2.3.

If $\mathbf{x}' \notin \mathrm{supp}_{\rho_i}(f_i)$ then one of the following cases occurs:

- $\mathbf{x}' \notin \mathrm{supp}_{\rho_i+\delta}(f_i)$: Using a similar calculation we have just performed regarding $\mathbf{x}$, we get, that except with probability $\exp(-\Theta(m\delta^2))$, $F(\mathbf{x}')_b$ has at least $(\rho_i + \delta - \delta/3)m$ coordinates j on which $F(\mathbf{x}')_j \neq f_i(\mathbf{x}'_j)$. But then T' rejects since there are at least $\lfloor \delta\ell/3 \rfloor > 0$ values in b on which $F(\mathbf{x})$ and $F(\mathbf{x}')$ disagree.

- $\mathbf{x}' \in \mathrm{supp}_{\rho_i+\delta}(f_i)$: By construction the fraction of such tuples is at most ρ_0^7.

As a conclusion we get that except with probability $< 2\exp(-\Theta(m\delta^2)) + \rho_0^7 \leq 2\rho_0^7$ if $\mathbf{x} \in \mathrm{supp}_{\rho_i}(f_i)$, then so is $\mathbf{x}'$. ∎

We now prove the first item in the theorem. The event T' accepts on $\mathbf{x}, \mathbf{x}'$ can be partitioned into three cases: (1) $\mathbf{x}, \mathbf{x}' \in R_\emptyset$, (2) $\mathbf{x} \in R_J, \mathbf{x}' \in R_{J'}$ for $J \neq J'$ and (3) $\mathbf{x}, \mathbf{x}' \in R_J$ for $J \neq \emptyset$. Denote these events by E_1, E_2, and E_3 respectively.

1. Both $\mathbf{x}, \mathbf{x}' \in R_\emptyset$. Recall that $R_\emptyset$ consists of those tuples that did not belong to $\mathrm{supp}_{\rho_i}(f_i)$ throughout the algorithm. By the stopping condition of our iterated algorithm we know that having randomized all tuples outside $R_\emptyset$, T' accepts with probability $\leq \varepsilon_0$. Therefore,

$$\Pr[T' \text{ accepts on } \mathbf{x}, \mathbf{x}' \text{ and } \mathbf{x}, \mathbf{x}' \in R_\emptyset] \leq \Pr[T' \text{ accepts on } \mathbf{x}, \mathbf{x}' \text{ in final iteration}] \leq \varepsilon_0.$$

2. There are some $J \neq J'$ such that $\mathbf{x} \in R_J$ and $\mathbf{x}' \in R_{J'}$ (recall $R_J, R_{J'}$ are disjoint). We claim that this probability is at most $O(\rho_0)$. Indeed note that in this situation there must exist some i for which $\mathbf{x} \in \mathrm{supp}_{\rho_i}(f_i)$ while $\mathbf{x}' \notin \mathrm{supp}_{\rho_i}(f_i)$ or vice versa. Lemma 5.2 implies that for each fixed value of i this occurs with probability at most $2\rho_0^7$. Taking a union bound over all possible i's (and switching roles of $\mathbf{x}, \mathbf{x}'$) the probability that there exists i for which the above holds is at most $4t\rho_0^7 = O(\rho_0^7/\varepsilon_0^6) \leq O(\rho_0)$.

3. There is some $J \neq \emptyset$ for which $\mathbf{x}, \mathbf{x}' \in R_J$.

Denote by A the event that T' accepts on $\mathbf{x}, \mathbf{x}'$. Then by the above

$$\Pr[A] = \Pr[A \wedge E_1] + \Pr[A \wedge E_2] + \Pr[A \wedge E_3] \leq \varepsilon_0 + O(\rho_0) + \Pr[A \wedge E_3]$$

So

$$\sum_{J \neq \emptyset} \Pr_{\mathbf{x},\mathbf{x}'}[\mathbf{x}, \mathbf{x}' \in R_J \mid T' \text{ accepts on } \mathbf{x}, \mathbf{x}'] = \Pr[E_3 \mid A] = \frac{\Pr[A \wedge E_3]}{\Pr[A]} \geq 1 - \frac{O(\varepsilon_0 + \rho_0)}{\alpha}.$$

and this completes the proof of item 1.

5.3 Proof of Item 2 of Theorem 5.1

By item 1, the probability that T' accepts $\mathbf{x}, \mathbf{x}'$ but there is no $J \neq \emptyset$ such that $\mathbf{x}, \mathbf{x}' \in R_J$ is at most $O(\rho_0) + \varepsilon_0$.

We now consider the weighted matrix $A_{T'}$ of dimension $|\mathbf{X}| \times |\mathbf{X}|$ defined by

$$A_{T'}(\mathbf{x}, \mathbf{x}') = \Pr[T' \text{ selects } \mathbf{x}' \text{ as the second tuple } \mid T' \text{ selects } \mathbf{x} \text{ as the first tuple}].$$

If $e_\mathbf{x}$ is the vector with 1 in coordinate $\mathbf{x}$ and zeros elsewhere then $A_{T'} e_\mathbf{x}$ is the vector describing the probability that T' chooses $\mathbf{x}'$ having already chosen $\mathbf{x}$. Let λ denote the second largest (in absolute value) eigenvalue of $A_{T'}$.

By an application of the Expander Mixing Lemma as in [HLW06, Theorem 3.6] we get that for any R_J $\Pr[\mathbf{x}, \mathbf{x}' \in R_J] \leq (\varepsilon_J)^2 + \lambda \varepsilon_J$. Summing over all J we get

$$\sum_{J \neq \emptyset} \Pr[\mathbf{x}, \mathbf{x}' \in R_J] \leq \sum_{J \neq \emptyset} (\varepsilon_J)^2 + \lambda$$

By item 1

$$\alpha = \Pr[T' \text{ accepts}] \leq \sum_{J \neq \emptyset} \Pr[\mathbf{x}, \mathbf{x}' \in R_J] + O(\rho_0) + \varepsilon_0 \leq \sum_{J \neq \emptyset} (\varepsilon_J)^2 + \lambda + O(\rho_0) + \varepsilon_0.$$

The following lemma allows us to bound $\frac{1}{\alpha} \cdot (\lambda + \alpha(O(\rho_0) + \varepsilon_0)) \leq \frac{2m}{\alpha \ell} + O(\rho_0) + \varepsilon_0 = \ell^{-\Omega(1)}$. Therefore: $\sum_{J \neq \emptyset} \varepsilon_J^2 \geq \alpha(1 - \ell^{-\Omega(1)})$.

Given a matrix A denote by $\lambda(A)$ the second largest (in absolute value) eigenvalue of A.

Lemma 5.3 $\lambda(A_{T'}) < \frac{2m}{\ell}$.

Proof: Let us bound the second eigenvalue of the matrix A_T (corresponding to the test T rather than T'). We then obtain the lemma since $A_{T'} = A_T B$ for a certain stochastic matrix B, and using the second item in the following claim:

Claim 5.4 *Let G_1, G_2 be d_1, d_2 regular graphs, and let A_1, A_2 be the corresponding adjacency matrices, then:*

1. $\lambda(A_1 + A_2) \leq \lambda(A_1) + \lambda(A_2)$.
2. *Let $\tilde{A}_1 = \frac{1}{d_1} A_1$, and let B be a symmetric stochastic matrix, then:*

$$\lambda(B \cdot \tilde{A}_1) \leq \lambda(\tilde{A}_1)$$

In order to bound $\lambda(A_T)$ we rely on the fact that the adjacency matrices of what is known as the *Hamming scheme* with parameters $n = \ell, q = |X|$ (see [vLW93, Chapter 30]), are closely related to A_T. Let $A_0, \ldots, A_\ell$ be $|\mathbf{X}| \times |\mathbf{X}|$ matrices such that for vectors $\mathbf{x}, \mathbf{y} \in \mathbf{X}$, $A_i(\mathbf{x}, \mathbf{y}) = 1$ iff $\mathbf{x}, \mathbf{y}$ disagree on exactly i coordinates. The following lemma is a summary of what we need from [vLW93, Chapter 30].

Lemma 5.5 *Each A_i has exactly $\ell + 1$ distinct eigenvalues and the $j-th$ eigenvalue of A_i equals:*

$$\sum_{\alpha=0}^{i} (-|X|)^\alpha (|X| - 1)^{i-\alpha} \binom{\ell - \alpha}{i - \alpha} \binom{j}{\alpha}$$

Moreover, the eigen-space of the j-th eigenvalue is the same for all A_i. ∎

A_i is an adjacency matrix of a d_i-regular graph with an edge between $\mathbf{x}, \mathbf{x}'$ if they agree on exactly $\ell - i$ coordinates (so $d_i = \sum_{\mathbf{x}' \in \mathbf{X}} A_i(\mathbf{x}, \mathbf{x}') = \binom{\ell}{i}(|X| - 1)^i$). Hence we can write A_T as a convex combination of the stochastic matrices $\frac{1}{d_i} A_i$:

$$A_T = \sum_{i=0}^{\ell - m} \alpha_i \left(\frac{1}{d_i} A_i \right)$$

where α_i is given by $\alpha_i = \beta_i/(\sum \beta_i)$ for

$$\beta_i = \binom{\ell - i}{m} d_i$$

since if $\mathbf{x}, \mathbf{x}'$ agree on $m' > m$ coordinates then there are $\binom{m'}{m}$ possibilities to reach $\mathbf{x}'$ from $\mathbf{x}$. By item 1 of Claim 5.4,

$$\lambda(A_T) \leq \sum_{i=0}^{\ell-m} \alpha_i \lambda\left(\frac{1}{d_i} A_i\right) \tag{11}$$

In order to use this formula we seem to need the eigenvalues of all of the matrices $A_0, \cdots, A_{\ell-m}$. However, we really need only focus on the matrix $A_{\ell-m}$ since for all $i \geq 1$

$$\beta_{\ell-m-i} \leq \beta_{\ell-m-1} = \frac{\ell - m}{|X| - 1} \cdot \beta_{\ell-m} < \frac{1}{\ell^2} \beta_{\ell-m}$$

where the last inequality follows since $|X| > \ell^3$. Therefore bounding $\lambda(A_i/d_i) \leq 1$ for $i < \ell - m$ we get from (11):

$$\lambda(A_T) \leq \alpha_{\ell-m} \lambda\left(\frac{1}{d_{\ell-m}} A_{\ell-m}\right) + \frac{\ell - m}{\ell^2} \leq \alpha_{\ell-m} \lambda\left(\frac{1}{d_{\ell-m}} A_{\ell-m}\right) + \frac{1}{\ell} \tag{12}$$

We compute $\lambda(A_{\ell-m})$ by Lemma 5.5:

$$\begin{aligned}
\lambda_0(A_{\ell-m}) &= d_{\ell-m} \\
\lambda_1(A_{\ell-m}) &= d_{\ell-m}(1 - \frac{|X|}{|X| - 1} \cdot \frac{\ell - m}{\ell}) \leq d_{\ell-m} \cdot \frac{m}{\ell} \\
\lambda_\ell(A_{\ell-m}) &= d_{\ell-m} \sum_{\alpha=0}^{\ell-m} \left(-\frac{|X|}{|X| - 1}\right)^{\alpha} \binom{\ell - m}{\alpha} = d_{\ell-m} \cdot \left(\frac{-1}{|X| - 1}\right)^{\ell-m}
\end{aligned}$$

So clearly the second largest eigenvalue of $A_{\ell-m}$ is $|\lambda_1(A_{\ell-m})|$, and $\lambda(\frac{1}{d_{\ell-m}} A_{\ell-m}) \leq \frac{m}{\ell}$. Overall we get that $\lambda(A_T)$ is bounded by $(m+1)/\ell \leq 2m/\ell$.

We now argue that $\lambda = \lambda(A_{T'}) \leq \lambda(A_T)$ and by that Lemma 5.3 is done.

We express $A_{T'}$ as a multiplication of stochastic matrix B with A_T. Let us define a matrix B' by $B'(\mathbf{x}, \mathbf{x}') = 1 \iff$ there exists a permutation $s : [\ell] \to [\ell]$ such that $\mathbf{x}' = s(\mathbf{x})$. Let B be the matrix obtained by dividing each row $\mathbf{x}$ of B' by $\sum_{\mathbf{x}' \in \mathbf{X}} B'_{\mathbf{x},\mathbf{x}'}$. Clearly, B is a symmetric stochastic matrix, and $A_{T'} = B \cdot A_T$. Using item 2 of Claim 5.4 we get $\lambda(A_{T'}) \leq \lambda(A_T)$

■

6 Tightness of Parameters

We would like to claim that our main Theorem 5.1 is tight in the sense that with stronger parameters the theorem does not hold.

Lemma 6.1 *There exists a function $F : \mathbf{X} \to \Sigma^\ell$ which T' accepts with probability $\Omega(m/\ell)$ and such that for any $f : X \to \Sigma$ the fraction of tuples $\mathbf{x}$ on which $F(\mathbf{x}) \overset{31/32}{\approx} \vec{f}(\mathbf{x})$ is at most $\ell/|X|$.*

In particular this implies the following constraints on the parameters of Theorem 5.1:

- Since $m \geq 1$, (otherwise the test T' is meaningless), then any variant of Theorem 5.1 does not hold if $\varepsilon = O(1/\ell)$. In particular, one cannot hope for an exponentially small ε.
- If $m = \Theta(\ell)$, then Theorem 5.1 cannot hold with arbitrarily small ε, since the acceptance probability of T' on F is $\Omega(1)$ with this choice of m. We comment that [IJKW08] raised the question of whether passing the consistency test with non-negligible probability imply non-negligible correlation with a direct-product function. This example shows that in their specific parameter setting ($m = \ell/2$) both our test and their test fail to test such a correlation.

Proof: We define F as follows.

- Select $X_0 \subset X$ such that $|X_0| = |X|/\ell$,
- Select a series of functions $\tilde{f} = f_1, \cdots, f_{|X|/\ell} : X \to \Sigma$ such that for every pair of functions f_i, f_j $\Pr_{x \in X}[f_i(x) \neq f_j(x)] > 1/8$. (We can choose at least $\frac{2^{|X|}}{2^{|X|H(1/8)}} > \frac{|X|}{\ell}$ such functions).
- For each $x \in X_0$ let f_x be a distinct such function.
- For each tuple $\mathbf{x}$ select $F(\mathbf{x})$ as follows: If $|\mathbf{x} \cap X_0| \neq 1$ pick $F(\mathbf{x})$ randomly. Otherwise let $x_0 = \mathbf{x} \cap X_0$ and set $F(\mathbf{x}) = (f_{x_0}(\mathbf{x}_1), \cdots, f_{x_0}(\mathbf{x}_\ell))$, in this case we say that x_0 sets the value of $F(\mathbf{x})$.

We would like to analyze the distance of F from the ℓ *direct product code*. Let $f, f' : x \to \Sigma$, we denote $\Pr_{x \in X}[f(x) \neq f(x')]$ by $d(f, f')$.

Let $f : X \to \Sigma^\ell$, we argue that f^ℓ agrees with F on at most $\ell/|X|$ fraction of the tuples. First, we argue that for any f there exists at most one $x \in X_0$ holding $d(f, f_x) < 1/16$.

Second, we argue that if $d(f, f') > 1/16$, then with probability at most $\exp -(\Theta(\ell))$ over $\mathbf{x}$, $\vec{f}(\mathbf{x}) \overset{31/32}{\approx} f'(\mathbf{x})$.

Since the values of $\mathbf{x}$ are either random or fixed according to some $x \in X_0$, and there is no function $f : \mathbf{X} \to \Sigma$ that agrees with more than a single function in $\tilde{f}$, then the only candidates functions f that their encoding f^ℓ is close to F, are the functions f_x ($x \in X_0$). Therefore, in order to analyze the distance of F from the ℓ *direct product code* we have to analyze the distance of F from f_x ($x \in X_0$). Each tuple $\mathbf{x}$ that holds $F(\mathbf{x}) = \vec{f_x}(\mathbf{x})$ must have also $x \in \mathbf{x}$. Therefore there are at most $1 - (1 - 1/|X|)^\ell < \ell/|X|$ fraction of tuples on which $F(\mathbf{x}) = \vec{f_x}(\mathbf{x})$.

Let us compute the acceptance probability of T' applied to F:

We calim that the probability that T' chooses $\mathbf{x}, \mathbf{x}'$ such that $F(\mathbf{x}) = \vec{f}(\mathbf{x})$ and also $F(\mathbf{x}') = \vec{f}(\mathbf{x}')$ is $\Omega(m/\ell)$. If that is the situation, then obviously T' accpets.

The probability that $F(\mathbf{x})$ is not a random value is the probability that $|\mathbf{x} \cap X_0| = 1$ which equals: $\ell \cdot \frac{|X_0|}{|X|} \cdot (\frac{|X| - |X_0|}{|X|})^{\ell - 1} = \ell \cdot \frac{1}{\ell} \cdot (1 - \frac{1}{\ell})^{\ell - 1}$. For large enough ℓ this is greater than some

constant c_1. If there is a unique $x_0 \in \mathbf{x} \cap X_0$ then $F(\mathbf{x}) = \vec{f_{x_0}}(\mathbf{x})$. Conditioned on $x_0 = \mathbf{x} \cap X_0$, what is the probability that also $\mathbf{x}' \cap X_0$ equals x_0? This would lead T' to accept, since then $F(\mathbf{x}') = \vec{f_{x_0}}(\mathbf{x}')$.

The desired probability is the probability that $x_0 \in \mathbf{x}'$ times the probability that any new coordinate in $\mathbf{x}'$ was picked outside X_0. The probability that $x_0 \in \mathbf{x}'$ equals m/ℓ. The probability that all the new coordinates of $\mathbf{x}'$ are drawn outside X_0 equals $\frac{|X|-|X_0|}{|X|}^{\ell-m}$ which is also greater than some constant c_2.

And overall we get that the probability that the test accepts F is at least $c_1 \cdot m/\ell \cdot c_2 = \Omega(m/\ell)$ as claimed. ∎

Acknowledgement

We would like to thank Avi Wigderson for helpful discussions.

References

[AS97] Sanjeev Arora and Madhu Sudan. Improved low degree testing and its applications. In *Proceedings of the Twenty-Ninth Annual ACM Symposium on Theory of Computing*, pages 485–495, El Paso, Texas, 4–6 May 1997.

[DFK+99] I. Dinur, E. Fischer, G. Kindler, R. Raz, and S. Safra. PCP characterizations of NP: Towards a polynomially-small error-probability. In *Proc. 31st ACM Symp. on Theory of Computing*, 1999.

[DR04] Irit Dinur and Omer Reingold. Assignment testers: Towards combinatorial proofs of the PCP theorem. In *Proceedings of the 45th Symposium on Foundations of Computer Science (FOCS)*, 2004.

[FK94] U. Feige and J. Kilian. Two prover protocols–low error at affordable rates. In *Proc. 26th ACM Symp. on Theory of Computing*, pages 172–183, 1994.

[GS97] Oded Goldreich and Shmuel Safra. A combinatorial consistency lemma with application to proving the pcp theorem. In *RANDOM*, pages 67–84, 1997.

[HLW06] Shlomo Hoory, Nathan Linial, and Avi Wigderson. Expander graphs and their applications. *Bull. Amer. Math. Soc.*, 43:439–561, 2006.

[IJK06] Russell Impagliazzo, Ragesh Jaiswal, and Valentine Kabanets. Approximately list-decoding direct product codes and uniform hardness amplification. In *Proc. 47th IEEE Symp. on Foundations of Computer Science*, pages 187–196, 2006.

[IJKW08] Russell Impagliazzo, Ragesh Jaiswal, Valentine Kabanets, and Avi Wigderson. Uniform direct product theorems: Simplified, optimized, and derandomized. In *Proc. 40th ACM Symp. on Theory of Computing*, 2008. to appear.

[OG05] Ryan O'Donnell and Venkatesan Guruswami. Lecture notes from a course on: the PCP theorem and hardness of approximation. 2005.

[Raz98] Ran Raz. A parallel repetition theorem. *SIAM Journal on Computing*, 27(3):763–803, June 1998.

[RS92] R. Rubinfeld and M. Sudan. Testing polynomial functions efficiently and over rational domains. In *Proc. 3rd Annual ACM-SIAM Symp. on Discrete Algorithms*, pages 23–32, 1992.

[RS97] R. Raz and S. Safra. A sub-constant error-probability low-degree test, and a sub-constant error-probability PCP characterization of NP. In *Proc. 29th ACM Symp. on Theory of Computing*, pages 475–484, 1997.

[STV01] Madhu Sudan, Luca Trevisan and Salil Vadhan. Pseudorandom Generators without the XOR Lemma. In *Journal of Computer and System Sciences*, 62(2), pages 236–266,2001.

[vLW93] J. H. van Lint and R. M. Wilson. *A Course in Combinatorics*. Cambridge University Press, 1993.

A The Feige-Kilian Dichotomy Lemma

Lemma 2.9 *Let $F : \mathbf{X} \to \Sigma^\ell$, and let $\varepsilon \geq 2\ell^{-1/75}$, then exactly one of following cases holds:*

1. *(Case 1) The probability that a random k-block is **alive** is at most ε.*
2. *(Case 2) The probability that a random live k-block is **good** is at least $1 - \varepsilon$.*

The proof we present is roughly the same as the original proof, (following more closely the presentation in [OG05]), but using our notations $\mathcal{A}_k, \mathcal{A}_{k_1,k_2}$.

Proof of lemma 2.9: We would like to use the agreement term in order to prove Lemma 2.9. We consider a "thought experiment" regarding F. We imagine picking a random tuple $\mathbf{x} \in \mathbf{X}$ as filling the ℓ coordinates with a value from X at random, one by one. I.e., in the $r - th$ step we pick a value i_r at random from the remaining coordinates and also $x_r \in X$ and set $\mathbf{x}_{i_r}$ to x_r. On each step r, we get a block of size r denoted by b_r. We would like to examine the change in the agreement

during the steps: $m/2, .., m$. In order to do it we now look at the following sequence:

$$\begin{array}{cc}
(\frac{m}{2}) & \mathcal{A}_{m/2}(b_{m/2}) \\
(\frac{m}{2}') & \mathcal{A}_{m/2,1}(b_{m/2}, (i_{m/2+1}, x_{m/2+1})) \\
(\frac{m}{2}+1) & \mathcal{A}_{m/2+1}(b_{m/2+1}) \\
(\frac{m}{2}+1') & \mathcal{A}_{m/2+1,1}(b_{m/2+1}, (i_{m/2+2}, x_{m/2+2})) \\
\vdots & \\
(m-1') & \mathcal{A}_{m-1,1}(b_{m-1}, (i_m, x_m)) \\
(m) & \mathcal{A}_m(b_m)
\end{array}$$

In this process we are examining how the agreement varies in accordance to the following two variants:

1. Pick a new random coordinate to $\mathbf{x}$, on the transition between steps (i) and (i').
2. Require agreement on this coordinate, on the transition between steps (i') and (i+1) (This makes agreement go down).

We would like to argue that there is a "special block size" r in which the expected agreement between the r-th level and the $r+1$ doesn't change a lot, formally:

Proposition A.1 *There exists "some special block size" $m/2 \leq r \leq m$ such that the following holds:*

$$\mathbf{E}_{b_{r^*+1}}[\mathcal{A}_{r^*}(b_{r^*}) - \mathcal{A}_{r^*+1}(b_{r^*+1})] \leq O(1)/m.$$

This special "special block size" r imposes the dichotomy Lemma as it would be explained later.

Proof: Due to Lemma 2.12 for every step i the expected value of the difference of the quantities (i') and (i) is at most $1/(\ell - m) \leq 2/\ell$. Since all these quantities are in the range $(0, 1]$, the total increase between the i' and $i'+1$ from beginning to end is at most $m/2 \cdot 2/\ell < 1$. Therefore the total decrease between the i' and $i+1$ from beginning to end is at most 2. Thus, there exists at least one step $m/2 \leq r \leq m$, where the decrease in the expected agreement is at most $2/(m/2) = 4/m$ (And going from (r^*) makes this the decrease only less).

■

Now we turn back to the proof of Lemma 2.9:

The idea of the proof is by contradiction: We show that if the special block size violates the dichotomy in Lemma 2.9, then the agreement would decrease too much- a contradiction to Proposition A.1.

And, indeed assume that neither case 1 holds nor case 2 holds. We get that while picking a random block b, with probability at least ε^2 it is alive but not good. This means that there is

some particular live answer a such that, conditioned on $F(\mathbf{x})_b = a$, at least η fraction of future coordinates (x, i) are not $(1-\eta)$ determined by (a, b).

This is almost enough to reach a contradiction, besides the fact that live answers a for b that were alive, might become much less probable once the coordinate (x, i) was added. We now show that this event is very rare: Fix a live block b and any corresponding live answer a, so we have $\Pr_{\mathbf{x}\in\mathbf{X}_b}[F(\mathbf{x})_b = a] \geq \varepsilon$. Lemma 2.2 tells as $\Pr_{\mathbf{x}\in\mathbf{X}_{b,(x,i)}}[F(\mathbf{x})_b = a] \geq \varepsilon - 1/\sqrt[3]{\ell - m} \geq \varepsilon/2$ except with probability at most $1/\sqrt[3]{\ell - m}$. Using union bound over all live answers $a \in \Sigma^k$, we get that for a random choice of r^* block b and a random coordinate (x, i), except with probability $1/\varepsilon \cdot 1/\sqrt[3]{\ell - m} \leq \varepsilon^{24} < \eta\varepsilon^2/2 = \varepsilon^9/2$ every answer a that is alive for b still satisfies $\Pr_{\mathbf{x}\in\mathbf{X}_{b,(x,i)}}[F(\mathbf{x})_b = a] \geq \varepsilon/2$. Let us summarize our knowledge:

Proposition A.2 *Assuming the dichotomy doesn't hold, let b be a random r^* block and let (x, i) be an additional coordinate. Then with probability $\geq \eta\varepsilon^2/2$ we get:*

1. *For all $\sigma \in \Sigma$, $Pr_{\mathbf{x}\in\mathbf{X}_{b,(x,i)}}[F(\mathbf{x})_b = a \text{ and } F(\mathbf{x})_i = \sigma] \leq (1-\eta) Pr_{\mathbf{x}\in\mathbf{X}_{b,(x,i)}}[F(\mathbf{x})_b = a]$.*
2. *$Pr_{\mathbf{x}\in\mathbf{X}_{b,(x,i)}}[F(\mathbf{x})_b = a] \geq \varepsilon/2$.*

Assuming item 1 holds, note that in order to maximize: $\sum_{\sigma\in\Sigma} Pr_{\mathbf{x}\in\mathbf{X}_{b,(x,i)}}[F(\mathbf{x})_i = \sigma \text{ and } F(\mathbf{x})_b = a]^2$ there should be two answers: σ, σ' such that one contributes $1-\eta$ to the above probability, the other η and all the rest contribute 0, therefore we get:

$$\begin{aligned}\sum_{\sigma\in\Sigma} Pr_{\mathbf{x}\in\mathbf{X}_{b,(x,i)}}[F(\mathbf{x})_i = \sigma \text{ and } F(\mathbf{x})_b = a]^2 &\leq (1 - 2\eta + 2\eta^2) Pr_{\mathbf{x}\in\mathbf{X}_{b,(x,i)}}[F(\mathbf{x})_b = a]^2 \\ &\leq (1-\eta) Pr_{\mathbf{x}\in\mathbf{X}_{b,(x,i)}}[F(\mathbf{x})_b = a]^2\end{aligned}$$

Therefore, by using item 2 in Proposition A.2:

$$\begin{aligned}Pr_{\mathbf{x}\in\mathbf{X}_{b,(x,i)}}[F(\mathbf{x})_b = a]^2 &- \sum_{\sigma\in\Sigma} Pr_{\mathbf{x}\in\mathbf{X}_{b,(x,i)}}[F(\mathbf{x})_b = a \text{ and } F(\mathbf{x})_i = \sigma]^2 \\ &\geq Pr_{\mathbf{x}\in\mathbf{X}_{b,(x,i)}}[F(\mathbf{x})_b = a]^2 - (1-\eta) Pr_{\mathbf{x}\in\mathbf{X}_{b,(x,i)}}[F(\mathbf{x})_b = a]^2 \\ &= \eta \cdot Pr_{\mathbf{x}\in\mathbf{X}_{b,(x,i)}}[F(\mathbf{x})_b = a]^2 \\ &\geq \eta(\varepsilon/2)^2\end{aligned}$$

We can conclude that whenever the events in Proposition A.2 occur, at least $\eta(\varepsilon/2)^2$ is contributed to:

$$\mathbf{E}_{b_{r^*+1}}[\mathcal{A}_{r^*}(b_{r^*}) - \mathcal{A}_{r^*+1}(b_{r^*+1})]$$

And therefore the total loss in the expected agreement is at least $(\eta\varepsilon^2/2) \cdot \eta(\varepsilon/2)^2 = \eta^2\varepsilon^4/8$. Since $4/m = 4\varepsilon^{19} \leq \eta^2\varepsilon^4/8 = \varepsilon^{18}/8$ we get a contradiction.

■