Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 107 (2010)

Derandomized Parallel Repetition via Structured PCPs

Irit Dinur* Or Meir†

May 12, 2010

Abstract

A PCP is a proof system for NP in which the proof can be checked by a probabilistic verifier. The verifier is only allowed to read a very small portion of the proof, and in return is allowed to err with some bounded probability. The probability that the verifier accepts a false proof is called the soundness error, and is an important parameter of a PCP system that one seeks to minimize. Constructing PCPs with sub-constant soundness error and, at the same time, a minimal number of queries into the proof (namely two) is especially important due to applications for inapproximability.

In this work we construct such PCP verifiers, i.e., PCPs that make only two queries and have sub-constant soundness error. Our construction can be viewed as a combinatorial alternative to the "manifold vs. point" construction, which is the only construction in the literature for this parameter range. The "manifold vs. point" PCP is based on a low degree test, while our construction is based on a direct product test.

Our construction of a PCP is based on extending the derandomized direct product test of Impagliazzo, Kabanets and Wigderson (STOC 09) to a derandomized parallel repetition theorem. More accurately, our PCP construction is obtained in two steps. We first prove a derandomized parallel repetition theorem for specially structured PCPs. Then, we show that any PCP can be transformed into one that has the required structure, by embedding it on a de-Bruijn graph.

*Weizmann Institute of Science, ISRAEL. Email: irit.dinur@weizmann.ac.il. Research supported in part by the Israel Science Foundation and by the Binational Science Foundation and by an ERC grant.

†Weizmann Institute of Science, ISRAEL. Research supported in part by the Israel Science Foundation (grant No. 460/05) and by the Adams Fellowship Program of the Israel Academy of Sciences and Humanities. Email: or.meir@weizmann.ac.il.

ISSN 1433-8092

1 Introduction

The PCP theorem [AS98, ALM+98] says that every language in NP can be verified by a polynomial-time verifier that uses $O(\log n)$ random bits and queries the proof in a constant number of locations. The verifier is guaranteed to always accept a correct proof, and to accept a false proof with bounded probability (called the *soundness error*). Following the proof of the PCP theorem, research has been directed towards strengthening the PCP theorem in terms of the important parameters, such as the proof length, the number of queries, and the soundness error.

In parallel, there is a line of work attempting to expand the variety of techniques at our disposal for constructing PCPs. Here the aim is to gain a deeper and more intuitive understanding of why PCP theorems hold. One of the threads in this direction is replacing algebraic constructions by combinatorial ones. This is motivated by the intuition that algebra is not an essential component of PCPs, indeed the definition of PCPs involves no algebra at all. Of course, one may also hope that the discovery of new techniques may lead to new results.

For the "basic" PCP theorem [AS98, ALM+98] there have been alternative combinatorial proofs [DR06, Din07]. It is still a challenge to match stronger PCP theorems with combinatorial constructions. Such is the work of the second author [Mei09] on PCPs with efficient verifiers. In this paper we seek to do so for PCPs in the small soundness error regime.

In this work we give a new construction of a PCP with sub-constant soundness error and two queries. This setting is particularly important for inapproximability, as will be discussed shortly below. Formally, we prove

Theorem 1.1 (Two-query PCP with small soundness)**.** *Every language $L \in NP$ has a two-query PCP system with perfect completeness, soundness error $1/\mathrm{poly}\log n$ and alphabet size $2^{\mathrm{poly}\log n}$. Furthermore, the verifier in this PCP system makes only 'projection' queries.*

This theorem matches the parameters of the folklore "manifold vs. point" construction which has been the only construction in the literature for this parameter range. The technical heart of that construction is a sub-constant error low degree test [RS97, AS03], see full details in [MR08].

Our proof of Theorem 1.1 is based on the elegant derandomized direct product test of [IKW09]. In a nutshell, our construction is based on applying this test to obtain a "derandomized parallel repetition theorem". While it is not clear how to do this for an arbitrary PCP, it turns out to be possible for PCPs with certain structure. We show how to convert any PCP to a PCP with the required structure, and then prove a "derandomized parallel repetition theorem" for such PCPs, thereby getting Theorem 1.1. The derandomized parallel repetition theorem relies on a reduction from the derandomized direct product test of [IKW09].

The Moshkovitz-Raz Construction. Recently, Moshkovitz and Raz [MR08] constructed even stronger PCPs. Their PCPs have nearly linear proof length, two queries, sub-constant error probability, and hold for *all smaller* alphabet sizes. Being able to reduce the alphabet size has strong consequences for inapproximability, see [MR08] for details. The technique of [MR08] (as evident in the later simplification of [DH09]) is essentially based on composition of existing PCP constructs. In fact, their main building block is the "manifold vs. point" construction mentioned above.

Our construction can be extended to yield a so-called decodable PCP [DH09], which is an object slightly stronger than a PCP. This can be plugged into the scheme of [DH09] to give a new proof of the main result of [MR08][1] (namely, an analog of Theorem 1.1 that works for *all smaller* alphabet

[1] Admittedly, the construction will have polynomial rather than nearly-linear length as in [MR08]. This we leave for future work.

sizes). This will give a completely different construction, one that makes no use of low degree polynomials.

Organization of the introduction In the following sections three sections we outline the background and main ideas of this work. We start by describing the parallel repetition technique in general and its relation with direct product tests. We proceed to describe our technique of derandomized parallel repetition. We then describe our notion of "PCPs with linear structure", to which the derandomized parallel repetition is applied.

After the foregoing outline, we discuss relevant works and possible future directions, and describe the organization of this work.

Parallel repetition and Direct Products

A natural approach to constructing PCPs with small soundness error is by *sequential repetition*. That is, we start with a PCP verifier that has a (large) constant soundness, and invoke it multiple times. Obviously, if the verifier is invoked for k times then the soundness of the resulting PCP system is exponentially decreasing with k. However, this technique also increases the query complexity of the PCP by a factor of k, and in particular the resulting PCP can not have query complexity of 2. Since our focus is on constructing PCPs that make only two queries, we can not afford using sequential repetition.

In order to decrease the soundness error while maintaining the query complexity, one may use *parallel repetition*. For the rest of this discussion, we consider only PCPs that use only two queries. Let us briefly recall what parallel repetition means in this context. As in the case of sequential repetition, one starts out with a PCP with constant soundness error, and then amplifies the rejection probability by repetition of the verifier. However, in order to save on queries, the prover is expected to give the k-wise direct product encoding of the original proof. Formally, if $\pi : [n] \to \Sigma$ describes the original proof then its direct product encoding, denoted by $\pi^{\otimes k}$, is the function $\pi^{\otimes k} : [n]^k \to \Sigma^k$ defined by

$$\pi^{\otimes k}(x_1, \ldots, x_k) = (\pi(x_1), \ldots, \pi(x_k)).$$

The new verifier will simulate the original verifier on k independent runs, but will read only *two* symbols from the new proof, which together contain answers to k independent runs of the original verifier.

The challenge in analyzing the soundness error of this verifier stems from the fact that the proof Π is adversarial, and is not guaranteed to be a direct product encoding $\pi^{\otimes k}$ of any underlying proof π, as intended in the construction. Raz's celebrated parallel repetition theorem [Raz98] states that the soundness error of this verifier does go down exponentially with k, and this is clearly the best possible.

The main difficulty in proving the parallel repetition theorem stems from the fact that the parallel-repetition proof is not necessarily a legal (direct-product) encoding of another proof. One may try to simplify the analysis by augmenting the parallel repetition with a direct product test. That is, making the verifier *test* that the given proof Π is a direct product encoding of some string π, and only then running the original parallel repetition verifier. This can sometimes be done without even incurring extra queries. Motivated by this approach Goldreich and Safra [GS00] suggested and studied the following question:

DP testing: Given a function $F : [n]^k \to \Sigma^k$ test that it is close to $f^{\otimes k}$ for some $f : [n] \to \Sigma$.

Let us now describe a two query direct product test. From now on let us make the simplifying assumption that the function $F : [n]^k \to \Sigma^k$ to be tested is given as a function of k-sized subsets rather than tuples, meaning that $F(x_1, \ldots, x_k)$ is the same for any permutation of $x_1, \ldots, x_k$. The test chooses two random k-subsets $B_1, B_2 \in \binom{[n]}{k}$ that intersect on a subset $A = B_1 \cap B_2$ of a certain prescribed size and accept if and only if $F(B_1)_{|A} = F(B_2)_{|A}$. This test was analyzed in [GS00, DR06, DG08, IKW09].

Derandomized Direct Product Testing

Recall that our goal is to construct PCPs with sub-constant soundness error. Note, however, that since the parallel repetition increases the proof length exponentially in k (and the randomness of the verifier grows k-fold), one can only afford to make a constant number of repetitions if one wishes to maintain polynomial proof length. On the other hand, obtaining sub-constant soundness error requires a super-constant number of repretitions.

This leads to the derandomization question, addressed already 15 years ago [FK95]. Can one recycle randomness of the verifier in the parallel repetition scheme without losing too much in soundness error?

Motivated by this question, Impagliazzo, Kabanets, and Wigderson [IKW09] introduced an excellent method for analyzing the direct product test which allowed them to derandomize it. Namely, they exhibited a relatively small collection of subsets $\mathcal{K} \subset \binom{[n]}{k}$, and considered the restriction of the direct product encoding $f^{\otimes k}$ to this collection. They then showed that this form of derandomized direct product can be tested using the above test. The collection $\mathcal{K}$ is as follows: identify $[n]$ with a vector space $\mathbb{F}^m$, let $k = |\mathbb{F}|^d$ for constant d, and let $\mathcal{K}$ be the set of all d-dimensional linear subspaces.

One would like to use the derandomized direct product of [IKW09] to obtain a derandomized parallel repetition theorem. Recall that the parallel repetition verifier works by simulating k independent runs of the original verifier on π, and querying the (supposed) direct product Π on the resulting k-tuples of queries. However, in the derandomized setting, the k-tuples of queries generated by the verifier may fall outside $\mathcal{K}$. This is the main difficulty that we address in this work.

This is where the structure of the PCP comes to our aid. We show that for PCPs with a certain linear structure, the k-tuples of queries can be made in a way that is compatible with the derandomized direct product test of [IKW09]. This allows us to prove a derandomized parallel repetition theorem for the particular case of PCPs with linear structure. Our main theorem is proved by constructing PCPs with linear structure (discussed next), and applying the derandomized parallel repetition theorem.

PCPs with Linear Structure

We turn to discuss PCPs with linear structure. The *underlying graph structure* of a two-query PCP is a graph defined as follows. The vertices are the proof coordinates, and the edges correspond to all possible query pairs of the verifier. (See also Section 2.3). We say that a graph has *linear structure* if the vertices can be identified with a vector space $\mathbb{F}^m$ and the edges, which clearly can be viewed as a subset of $\mathbb{F}^{2m}$, form a linear subspace of $\mathbb{F}^{2m}$ (see also Definition 3.1). A two-query PCP has linear structure if its underlying graph has linear structure.

As mentioned above, an additional contribution to this work is the construction of PCPs with linear structure. That is, we prove the following result.

Theorem 1.2 (PCPs with linear structure)**.** *Every language $L \in NP$ has a PCP system with linear structure, using $O(\log n)$ randomness, constant alphabet size, and such that the PCP has*

perfect completeness and soundness error $1 - 1/\text{poly}\log n$.

We believe that Theorem 1.2 is interesting in its own right: For known PCPs, the underlying graph structure is quite difficult to describe, mostly due to the fact that PCP constructions are invariably based on composition. In principle, however, the fact that a PCP is a "complex" object need not prevent the underlying graph from being simple. In analogy, certain Ramanujan expanders [LPS88] are Cayley graphs that are very easy to describe, even if the proof of their expansion is not quite so easy. It is therefore interesting to study whether there exist PCPs with simple underlying graphs.

Philosophically, the more structured the PCP, the stronger is the implied statement about the class NP, and the easier it is to exploit for applications. Indeed, the structure of a PCP system has been used in several previous works. For example, Khot constructs [Kho06] a PCP with pseudorandom structure in order to establish the hardness of minimum bisection. Dinur [Din07] imposes an expansion structure on a PCP to obtain amplification.

We prove Theorem 1.2 by embedding a given PCP into the de Bruijn graph and relying on the algebraic structure of this graph. We remark that the de Bruijn graph has been used in constructions of PCPs before, e.g. [PS94, BFLS91], in similar contexts. We believe that structured PCPs are an object worthy of further study. One may view their applicability towards proving Theorem 1.1 as supporting evidence. An interesting question which we leave open is whether Theorem 1.2 can be strengthened so as to get *constant* soundness error. By simply plugging such a PCP into our derandomized parallel repetition theorem one would get a direct proof of the aforementioned result of [MR08], *without* using two-query composition.

Related Work and Future directions

Our final construction of a two-query PCP has exponential relation between the alphabet size (which is $2^{\text{poly}\log n}$) and the error probability (which is $1/\text{poly}\log n$). In general, one can hope for a polynomial relation, and this is the so-called "sliding scale" conjecture of [BGLR93]. Our approach is inherently limited to an exponential relation both because of a lower bound on direct product testing from [DG08], and, more generally, because of the following lower bound of Feige and Kilian [FK95] on parallel repetition of games. Feige and Kilian prove that for every PCP system and $k = O(\log n)$, if one insists on the parallel repetition using only $O(\log n)$ random bits, then the soundness error must be at least $1/\text{poly}\log n$ (and not $1/\text{poly}(n)$ as one might hope). Our work matches the [FK95] lower bound by exhibiting a derandomized parallel repetition theorem, albeit only for PCPs with linear structure, that achieves a matching upper bound of $1/\text{poly}\log n$ on the soundness error.

Nevertheless, for three queries we are in a completely different ball-game, and no lower bound is known. It would be interesting to find a derandomized direct product test with three queries with lower soundness error, and to try and adapt it to a PCP. We note that there are "algebraic" constructions [RS97, DFK$^+$99] that make only three queries and have much better relationship between the error and the alphabet size.

It has already been mentioned that while our result matches the soundness error and alphabet size of the [MR08] result, it does not attain nearly linear proof length. Improving our result in this respect is another interesting direction.

Organization

In Section 2, we give the required preliminaries for this work, including a description of the derandomized direct product test of [IKW09]. In Section 3 we prove Theorem 1.1 based on our main

1. Choose a uniformly distributed d_1-subspace $B \subseteq \mathbb{F}^m$.
2. Choose a uniformly distributed d_0-subspace $A \subseteq B$.
3. Accept if and only if $\Pi(B)_{|A} = \Pi(A)$.

Figure 1: The P-test

lemmas. The construction of PCPs with linear structure is given in Section 4. In Section 5 we prove the "derandomized parallel repetition" theorem for PCPs with linear structure, by reducing it to the analysis of a specialized variant of the test of [IKW09]. Finally, we analyze the specialized direct product in Section 6.

2 Preliminaries

Let $g : U \to \Sigma$ be an arbitrary function, and let $A \subset U$ be a subset. We denote by $g_{|A}$ the restriction of g (as a function) to A. Given two functions $f, g : U \to \Sigma$ we denote $f \overset{\alpha}{\approx} g$ ($f \overset{\alpha}{\not\approx} g$) to mean that they differ on at most (more than) α fraction of the elements of U.

We refer to a d-dimensional linear subspace of an underlying vector space simply as a *d-subspace*. For two linear subspaces A_1 and A_2 we denote by $A_1 + A_2$ the smallest linear subspace containing both of them. We say that A_1, A_2 are *disjoint* if and only if $A_1 \cap A_2 = \{0\}$. If A_1 and A_2 are disjoint, we use $A_1 \oplus A_2$ to denote $A_1 + A_2$.

Let $G = (V, E)$ be a directed graph. For each edge $e \in E$ we denote by $\text{left}(e)$ and $\text{right}(e)$ the left and right endpoints of e respectively. That is, if we view the edge $e \in E$ as a pair in $V \times V$, then $\text{left}(e)$ and $\text{right}(e)$ are the first and second elements of the pair e respectively. Given a set of edges $E_0 \subseteq E$, we denote by $\text{left}(E_0)$ and $\text{right}(E_0)$ the set of left endpoints and right endpoints of the edges in E_0 respectively.

2.1 Direct product testing [IKW09]

Let us briefly describe the setting in which we use the derandomized direct product test of [IKW09]. In [IKW09] the main derandomized direct product test is a so-called "V-test". We consider a variation of this test that appears in [IKW09, Section 6.3] to which we refer as the "P-test" (P for projection).

Given a string $\pi \in \Sigma^\ell$, we define its (derandomized) P-direct product Π as follows: We identify $[\ell]$ with $\mathbb{F}^m$, where $\mathbb{F}$ is a finite field and $m \in \mathbb{N}$, and think of π as an assignment that maps the points in $\mathbb{F}^m$ to Σ. We also fix $d_0 < d_1 \in \mathbb{N}$. Now, we define to be Π the assignment that assigns each d_0- and d_1-subspace W of $\mathbb{F}^m$ to the function $\pi_{|W} : W \to \Sigma$ (recall that $\pi_{|W}$ is the restriction of π to W).

We now consider the task of testing whether a given assignment Π is the P-direct product of some string $\pi : \mathbb{F}^m \to \Sigma$. In those settings, we are given an assignment to subspaces, i.e. a function Π that on input a d_0-subspace $A \subset \mathbb{F}^m$ (respectively d_1-subspace $B \subset \mathbb{F}^m$), answers with a function $a : A \to \Sigma$ (respectively, $b : \mathbb{F}^m \to \Sigma$). We wish to test whether Π is a P-direct product of some $\pi : \mathbb{F}^m \to \Sigma$, and to this end we invoke the P-test, described in Figure 1.

It is easy to see that if Π is a P-direct product then the P-test always accepts. Furthermore, it can be shown that if Π is "far" from being a P-direct product, then the P-test rejects with high probability. Formally, we have the following result.

Theorem 2.1 ([IKW09]). *There exists a universal constant $h \in \mathbb{N}$ such that the following holds: Let $\varepsilon \geq h \cdot d_0 \cdot |\mathbb{F}|^{-d_0/h}$, $\alpha \stackrel{\text{def}}{=} h \cdot d_0 \cdot |\mathbb{F}|^{-d_0/h}$. Assume that $d_1 \geq h \cdot d_0$, $m \geq h \cdot d_1$. Suppose that an assignment Π passes the P-test with probability at least ε. Then, there exists an assignment π such that*

$$\Pr\left[\Pi(B)_{|A} = \Pi(A) \quad \text{and } \Pi(B) \stackrel{\alpha}{\approx} \pi_{|B} \quad \text{and } \Pi(A) \stackrel{\alpha}{\approx} \pi_{|A}\right] = \Omega(\varepsilon^4) \tag{1}$$

where the probability is over A, B chosen as in the P-test.

Theorem 2.1 can be proved using the techniques of [IKW09]. For completeness, the proof is given in Appendix A.

Working with randomized assignments. As noticed by [IKW09], Theorem 2.1 works in even stronger settings. Suppose that Π is a randomized function, i.e., a function of both its input and some additional randomness. Then, Theorem 2.1 still holds for Π, where the probability in (1) is over both the choice of A and B, *and over the internal randomness of* Π. We will rely on this fact in a crucial way in this work.

2.2 Sampling tools

The following is a standard definition, in graph terms, see e.g. [IJKW08].

Definition 2.2 (Sampler Graph)**.** A bipartite graph $G = (L, R, E)$ is said to be an (ε, δ)-sampler if, for every function $f : L \to [0,1]$, there are at most $\delta |R|$ vertices $u \in R$ for which

$$\left|\mathbb{E}_{v \in N(u)}[f(v)] - \mathbb{E}_{v \in L}[f(v)]\right| > \varepsilon.$$

Observe that if G is an (ε, δ)-sampler, and if $F \subset L$, then by considering the function $f \equiv 1_F$ we get that there are at most $\delta |R|$ vertices $u \in R$ for which

$$\left|\Pr_{v \in N(u)}[v \in F] - \Pr_{v \in L}[v \in F]\right| > \varepsilon.$$

We have the following result

Lemma 2.3 (Subspace-point sampler [IJKW08])**.** *Let $d' < d$ be natural numbers, let V be a linear space over a finite field $\mathbb{F}$, and let W be a fixed d'-dimensional of V. Let G be the bipartite graph whose left vertices are all points V and whose right vertices are all d-subspaces of V that contain W. We place an edge between a d-subspace X and $x \in V$ iff $x \in X$. Then G is an $(\tau + \frac{1}{|\mathbb{F}|^{d-d'}}, \frac{1}{|\mathbb{F}|^{d-d'-2} \cdot \tau^2})$-sampler for every $\tau > 0$.*

Proof Fix a function $f : V \to [0,1]$. We show that for a uniformly distributed d-subspace $X \subseteq V$ that contains W it holds with probability at least $1 - \frac{1}{|\mathbb{F}|^{d-d'-2} \cdot \tau^2}$ that

$$\left|\mathbb{E}_{x \in X}[f(x)] - \mathbb{E}_{v \in V}[f(v)]\right| \leq \tau + \frac{1}{|\mathbb{F}|^{d-d'}}$$

Let $\overline{W}$ be a fixed subspace of V for which $V = W \oplus \overline{W}$. Let $f_W : \overline{W} \to [0,1]$ be the function that maps each vector $\overline{w}$ of $\overline{W}$ to $\mathbb{E}_{v \in \overline{w} + W}[f(v)]$, and observe that $\mathbb{E}_{v \in V}[f(v)] = \mathbb{E}_{\overline{w} \in \overline{W}}[f_W(\overline{w})]$. Furthermore, observe that every d-subspace X that contains W can be written as $X = W \oplus U$ where U is a $(d - d')$-subspace of $\overline{W}$, and moreover that $\mathbb{E}_{x \in X}[f(x)] = \mathbb{E}_{u \in U}[f_W(u)]$. Thus, it suffices to

prove that for a uniformly distributed $(d-d')$-subspace U of $\overline{W}$ it holds with probability at least $1-\frac{1}{|\mathbb{F}|^{d-d'-2}\cdot\tau^2}$ that

$$\left|\mathbb{E}_{u\in U}\left[f_W(u)\right]-\mathbb{E}_{\overline{w}\in\overline{W}}\left[f_W(\overline{w})\right]\right|\leq\tau+\frac{1}{|\mathbb{F}|^{d-d'}} \tag{2}$$

To that end, let U be a uniformly distributed $(d-d')$-subspace of $\overline{W}$. Let S_1 be a set of $Q=\frac{|\mathbb{F}|^{d-d'}-1}{|\mathbb{F}|-1}$ vectors of U such that every two vectors in S_1 are linearly independent (it is easy to construct such a set). For every $\alpha\in\mathbb{F}^*$ let S_α be the set obtained by multiplying every vector in S_1 by α. Observe that all the sets S_α have the property that every two vectors in S_α are linearly independent, and that the sets S_α form a partition of $U\setminus\{0\}$. We will show that for every $\alpha\in\mathbb{F}^*$ it holds with probability at least $1-\frac{1}{|\mathbb{F}|^{d-d'-1}\cdot\tau^2}$ that

$$\left|\mathbb{E}_{u\in S_\alpha}\left[f_W(u)\right]-\mathbb{E}_{\overline{w}\in\overline{W}}\left[f_W(\overline{w})\right]\right|\leq\tau$$

and the required result will follow by taking the union bound over all $\alpha\in\mathbb{F}^*$, and by noting that the vector 0 contributes at most $\frac{1}{|\mathbb{F}|^{d-d'}}$ to the difference in Inequality 2.

Fix $\alpha\in\mathbb{F}^*$, and let $s_1,\ldots,s_Q$ be the vectors in S_α. It is a known fact that $s_1,\ldots,s_Q$ are pair-wise independent and uniformly distributed vectors of $\overline{W}$ (over the random choice of U). This implies that $f_W(s_1),\ldots,f_W(s_Q)$ are pair-wise independent random variables with expectation $\mathbb{E}_{\overline{w}\in\overline{W}}\left[f_W(\overline{w})\right]$, and therefore by the Chebyshev inequality it follows that

$$\Pr\left[\left|\frac{1}{Q}\sum_{i=1}^{Q}f_W(s_i)-\mathbb{E}_{\overline{w}\in\overline{W}}\left[f_W(\overline{w})\right]\right|>\tau\right]\leq\frac{1}{Q\cdot\tau^2}\leq\frac{1}{|\mathbb{F}|^{d-d'-1}\cdot\tau^2}$$

as required. ■

2.3 Constraint graphs and PCPs

As discussed in the introduction, the focus of this work is on claims that can be verified by reading a small number of symbols of the proof. A PCP system for a language L is an oracle machine M, called a verifier, that has oracle access to a proof π over an alphabet Σ. The verifier M reads the input x, tosses r coins, makes at most q "oracle" queries into π, and then accepts or rejects. If x is in the language then it is required that M accepts with probability 1 for some π, and otherwise it is required that M accepts with probabiltiy at most ε for every π. More formally:

Definition 2.4. Let $r,q:\mathbb{N}\to\mathbb{N}$, and let Σ be a function that maps the natural numbers to finite alphabets. A $(r,q)_\Sigma$-*PCP verifier* M is a probabilistic polynomial time oracle machine that when given input $x\in\{0,1\}^*$, tosses at most $r(|x|)$ coins, makes at most $q(|x|)$ *non-adaptive* queries to an oracle that is a string over $\Sigma(|x|)$, and outputs either "accept" or "reject". We refer to r, q, and Σ as the *randomness complexity*, *query complexity*, and *proof alphabet* of the verifier respectively.

Remark 2.5. Note that for an $(r,q)_\Sigma$-PCP verifier M and an input x, we can assume without loss of generality that the oracle is a string of length at most $2^{r(|x|)}\cdot q(|x|)$, since this is the maximal number of different queries that M can make.

Definition 2.6. Let r, q and Σ be as in Definition 2.4, let $L\subseteq\{0,1\}^*$ and let $\varepsilon:\mathbb{N}\to[0,1)$. We say that $L\in\mathbf{PCP}_{\varepsilon,\Sigma}[r,q]$ if there exists an $(r,q)_\Sigma$-PCP verifier M that satisfies the following requirements:

- **Completeness:** For every $x \in L$, there exists $\pi \in \Sigma^{(|x|)^*}$ such that $\Pr[M^\pi(x) \text{ accepts}] = 1$.
- **Soundness:** For every $x \notin L$ and for every $\pi \in \Sigma^{(|x|)^*}$ it holds that $\Pr[M^\pi(x) \text{ accepts}] \leq \varepsilon$.

One possible formulation of the the PCP theorem is as follows.

Theorem 2.7 (PCP Theorem [AS98, ALM+98])**.** *There exist universal constant $\varepsilon \in (0,1)$ and a finite alphabet Σ such that* $\mathbf{NP} \subseteq \mathbf{PCP}_{\varepsilon,\Sigma}[O(\log n), 2]$.

PCPs that have query complexity 2 correspond to graphs in a natural way: Consider the action of an $(r,2)_\Sigma$-verifier M on some fixed string x, and let $r \stackrel{\text{def}}{=} r(|x|), \Sigma \stackrel{\text{def}}{=} \Sigma(|x|)$. The verifier M is given access to some proof string π of length ℓ, and may make 2^r possible tests on this string, where each such test consists of making two queries to π and deciding according to the answers. We now view the action of M as a graph in the following way. We consider the graph G whose vertices are the coordinates in $[\ell]$, and that has an edge for each possible test of the verifier M. The endpoints of an edge e of G are the coordinates that are queried by M in the test that corresponds to e. We also associate an edge e with a constraint $c_e \in \Sigma \times \Sigma$, which contains all the pairs of answers that make M accept when performing the test that corresponds to e. We think of π as an assignment that assigns the vertices of G values in Σ, and say that π *satisfies* an edge (u,v) if $(\pi(u), \pi(v)) \in c_{(u,v)}$. If $x \in L$, then it is required that there exists some assignment π that satisfies all the edges of G, and otherwise it is required that every assignment satisfies at most ε fraction of the edges. This correspondence is called the FGLSS correspondence [FGL+96]. We turn to state it formally:

Definition 2.8 (Constraint graph)**.** A (directed) constraint graph is a directed graph $G = (V, E)$ together with an alphabet Σ, and, for each edge $(u,v) \in E$, a binary constraint $c_{u,v} \subseteq \Sigma \times \Sigma$. The *size* of G is the number of edges of G. The graph is said to have *projection constraints* if every constraint $c_{u,v}$ has an associated function $f_{u,v} : \Sigma \to \Sigma$ such that $c_{u,v}$ is satisfied by (a,b) iff $f_{u,v}(a) = b$.
Given an assignment $\pi : V \to \Sigma$, we define

$$\mathrm{SAT}(G,\pi) = \Pr_{(u,v)\in E}[(\pi(u),\pi(v)) \in c_{u,v}] \qquad \text{and} \qquad \mathrm{SAT}(G) = \max_\pi(\mathrm{SAT}(G,\pi)).$$

We also denote $\mathrm{UNSAT}(G,\pi) = 1 - \mathrm{SAT}(G,\pi)$ and similarly $\mathrm{UNSAT}(G) = 1 - \mathrm{SAT}(G)$.

Remark 2.9. Note that Definition 2.8 uses *directed graphs*, while the common definition of constraint graphs refers to undirected graphs.

Remark 2.10. Note that if the graph G is bipartite and all edges are directed from, say, left to right, then this is simply a label cover instance with projection constraints [AL96].

Proposition 2.11 (FGLSS correspondence [FGL+96])**.** *The following two statements are equivalent:*

- $L \in \mathbf{PCP}_{\varepsilon,\Sigma}[r, 2]$.
- *There exists a polynomial-time transformation that transforms strings $x \in \{0,1\}^*$ to constraint graphs G_x of size $2^{r(|x|)}$ with alphabet $\Sigma^{(|x|)}$ such that: (1) if $x \in L$ then $\mathrm{SAT}(G_x) = 1$, and (2) if $x \notin L$ then $\mathrm{SAT}(G_x) \leq \varepsilon$.*

Given a PCP system for L, we refer to the corresponding family of graphs $\{G_x\}$ where x ranges over all possible instances as its underlying graph family*. If the graphs $\{G_x\}$ have projection constraints then we say that the PCP system has the* projection property*.*

Using the [FGL+96] correspondence, we can rephrase the PCP theorem in the terminology of constraint graphs:

Theorem 2.12 (PCP Theorem for constraint graphs)**.** *There exist universal constant $\varepsilon \in (0,1)$ and a finite alphabet Σ such that for every language $L \in \mathbf{NP}$ the following holds: There exists a polynomial time reduction that on input $x \in \{0,1\}^*$, outputs a constraint graph G_x such that if $x \in L$ then $\mathrm{SAT}(G_x) = 1$ and otherwise $\mathrm{SAT}(G_x) \leq \varepsilon$.*

2.4 Basic facts about random subspaces

In this section we present two useful propositions about random subspaces. The following proposition says that a uniformly distributed subspace is disjoint from every fixed subspace with high probability.

Proposition 2.13. *Let $d, d' \in \mathbb{N}$ such that $d > 2d'$, and let V be a d-dimensional space. Let W_1 be a uniformly distributed d'-subspace of V, and let W_2 be a fixed d'-subspace of V. Then,*

$$\Pr[W_1 \cap W_2 = \{0\}] \geq 1 - 2 \cdot d' / |\mathbb{F}|^{d-2\cdot d'} .$$

Proof Suppose that W_1 is chosen by choosing random basis vectors $v_1, \ldots, v_{d'}$ one after the other. It is easy to see that $W_1 \cap W_2 \neq \{0\}$ only if $v_i \in \mathrm{span}\,(W_2 \cup \{v_1, \ldots, v_{i-1}\})$ for some $i \in [d']$. For each fixed i, the vector v_i is uniformly distributed in $V \backslash \mathrm{span}\,\{v_1, \ldots, v_{i-1}\}$, and therefore the probability that $v_i \in \mathrm{span}\,(W_2 \cup \{v_1, \ldots, v_{i-1}\})$ for a fixed i is at most

$$\begin{aligned} \frac{|\mathrm{span}\,(W_2 \cup \{v_1, \ldots, v_{i-1}\})|}{|V \backslash \mathrm{span}\,\{v_1, \ldots, v_{i-1}\}|} &= \frac{|\mathbb{F}|^{d'+i-1}}{|\mathbb{F}|^d - |\mathbb{F}|^{i-1}} \\ &\leq \frac{2 \cdot |\mathbb{F}|^{d'+i-1}}{|\mathbb{F}|^d} \\ &\leq \frac{2 \cdot |\mathbb{F}|^{2\cdot d'-1}}{|\mathbb{F}|^d} \\ &\leq \frac{2}{|\mathbb{F}|^{d-2\cdot d'}} \end{aligned} \tag{3}$$

where Inequality 3 can be observed by noting that $|\mathbb{F}|^{i-1} \leq |\mathbb{F}|^{d-1} \leq \frac{1}{2} \cdot |\mathbb{F}|^d$. By the union bound, the probability that this event occurs for some $i \in [d']$ is at most $\frac{2\cdot d'}{|\mathbb{F}|^{d-2\cdot d'}}$. It follows that the probability that $W_1 \cap W_2 \neq \{0\}$ is at most $\frac{2\cdot d'}{|\mathbb{F}|^{d-2\cdot d'}}$ as required. ■

The following proposition says that the span of d' uniformly distributed vectors is with high probability a uniformly distributed d'-subspace.

Proposition 2.14. *Let V be a d-dimensional space over a finite field $\mathbb{F}$, let $w_1, \ldots, w_{d'}$ be independent and uniformly distributed vectors of V, and let $W = \mathrm{span}\{w_1, \ldots, w_{d'}\}$. Then, with probability at least $1 - d' / |\mathbb{F}|^{d-d'}$ it holds that $\dim W = d'$. Furthermore, conditioned on the latter event, W is a uniformly distributed d'-subspace of V.*

Proof The fact that $\dim W = d'$ with probability at least $1 - d' / |\mathbb{F}|^{d-d'}$ can be proved in essentially the same way as Proposition 2.13. To see that conditioned on the latter event it holds that the subspace W is uniformly distributed, observe that since $w_1, \ldots, w_{d'}$ were originally chosen to be uniformly distributed, all the possible d'-sets of linearly independent vectors have the same probability to occur. ■

Finally, the following proposition shows the equivalence of two different ways of choosing subspaces $A_1, A_2 \subseteq B$ where A_1 and A_2 are disjoint.

Proposition. *Let V be a linear space over a finite field $\mathbb{F}$, and let $d_0, d_1 \in \mathbb{N}$ be such that $d_0 < d_1 < \dim V$. The following two distributions over d_0-subspaces A_1, A_2 and a d_1-subspace B are the same:*

1. *Choose B to be a uniformly distributed d_1-subspace of V, and then choose A_1 and A_2 to be two uniformly distributed and disjoint d_0-subspaces of B.*
2. *Choose A_1 and A_2 to be two uniformly distributed and disjoint d_0-subspaces of V, and then choose B to be a uniformly distributed d_1-subspace of V that contains A_1 and A_2.*

Proof Observe that choosing A_1, A_2, B under the first distribution amounts to choosing d_1 uniformly distributed and linearly independent vectors in V (those vectors will serve as the basis of B), and then choosing two disjoint subsets of those vectors to serve as the basis of A_1 and as the basis of A_2. On the other hand, choosing A_1, A_2 and B under the second distribution amounts to choosing d_0 uniformly distributed and linearly independent vectors in V to serve as the basis of A_1, then choosing another d_0 uniformly distributed and linearly independent vectors in V to serve as the basis of A_2 while making sure that this basis is also linearly independent from the basis of A_1, and then completing the basis of A_1 and the basis of A_2 to a basis of B. It is easy to see that those two distributions over a set of d_1 vectors and its two disjoint subsets are identical. ■

3 Main theorem

In this section we prove the main theorem (Theorem 1.1). To that end, we use the PCP theorem for graphs (Theorem 2.12) to reduce the problem of deciding membership of a string x in the language L to the problem of checking the satisfiability of a constraint graph with constant soundness error. We then show that every constraint graph can be transformed into one that has "linear structure", defined shortly below. This is done in Lemma 3.2, which directly proves Theorem 1.2. Finally, in Lemma 3.3 we prove a derandomized parallel repetition theorem for constraint graphs with linear structure. Theorem 1.1 follows by combining the two lemmas. We begin by defining the notion of a graph with linear structure.

Definition 3.1. We say that a directed graph G has a *linear structure if it satisfies the following conditions:*

1. *The vertices of G can be identified with the linear space $\mathbb{F}^m$, where $\mathbb{F}$ is a finite field and $m \in \mathbb{N}$.*
2. *We identify the set of pairs of vertices $(\mathbb{F}^m)^2$ with the linear space $\mathbb{F}^{2m}$. Using this identification, the edges E of G are required to form a linear subspace of $\mathbb{F}^{2m}$.*
3. *We require that $\text{left}(E) = \text{right}(E) = \mathbb{F}^m$. In other words, this means that every vertex of G is both the left endpoint of some edge and the right point of some edge.*

The following lemmas are proved in Sections 4 and 5 respectively.

Lemma 3.2 (PCP with Linear Structure)**.** *There exists a polynomial time procedure that satisfies the following requirements:*

- ***Input:***

- *A constraint graph G of size n over alphabet Σ.*
- *A finite field $\mathbb{F}$ of size q.*

- ***Output:*** *A constraint graph $G' = (\mathbb{F}^m, E')$ such that the following holds:*
  - *G' has a linear structure.*
  - *The size of G' is at most $O\left(q^2 \cdot n\right)$.*
  - *G' has alphabet $\Sigma^{O(\log_q(n))}$.*
  - *If G is satisfiable then G' is satisfiable.*
  - *If $\mathrm{UNSAT}(G) \geq \rho$ then $\mathrm{UNSAT}(G') \geq \Omega\left(\frac{1}{q \cdot \log_q(n)} \cdot \rho\right)$.*

Lemma 3.3 (Derandomized Parallel Repetition)**.** *There exist a universal constant h and a polynomial time procedure that satisfy the following requirements:*

- ***Input:***
  - *A finite field $\mathbb{F}$ of size q*
  - *A constraint graph $G = (\mathbb{F}^m, E)$ over alphabet Σ that has a linear structure.*
  - *A parameter $d_0 \in \mathbb{N}$ such that $d_0 < m/h^2$.*
  - *A parameter $\rho \in (0,1)$ such that $\rho \geq h \cdot d_0 \cdot q^{-d_0/h}$.*

- ***Output:*** *A constraint graph G' such that the following holds:*
  - *G' has size $n^{O(d_0)}$.*
  - *G' has alphabet $\Sigma^{q^{O(d_0)}}$.*
  - *If G is satisfiable then G' is satisfiable.*
  - *If $\mathrm{SAT}(G) < 1 - \rho$ then $\mathrm{SAT}(G') < h \cdot d_0 \cdot q^{-d_0/h}$.*
  - *G' has the projection property*

We turn to prove the main theorem from the above lemmas.

Theorem (1.1, restated)**.** *Every language $L \in NP$ has a two-query PCP system with perfect completeness, soundness error $1/\mathrm{poly}\log n$ and alphabet size $2^{\mathrm{poly}\log n}$. Furthermore, the verifier in this PCP system makes only 'projection' queries.*

Proof Fix $L \in \mathbf{NP}$ and let $c \in \mathbb{N}$ be a constant to be chosen later. We show that L has a two-query PCP system with perfect completeness, soundness error $1/\log n$ and alphabet size $2^{\mathrm{poly}\log n}$, which has the projection property. By the [FGL+96] correspondence (Proposition 2.11), it suffices to show a polynomial time procedure that on input $x \in \{0,1\}^*$, outputs a constraint graph G' of size $\mathrm{poly}(n)$ such that the following holds: If $x \in L$ then G' is satisfiable (i.e. $\mathrm{SAT}(G') = 1$), and if $x \notin L$ then $\mathrm{SAT}(G') \leq O(1/\log|x|)$. The procedure begins by transforming x, using the PCP theorem for constraint graphs (Theorem 2.12), to a constraint graph G of size $n = \mathrm{poly}|x|$ such that if $x \in L$ then $\mathrm{SAT}(G) = 1$ and if $x \notin L$ then $\mathrm{SAT}(G) \leq \varepsilon$, where $\varepsilon \in [0,1)$ is a universal constant that does not depend on x. Let $n = \mathrm{poly}(|x|)$ be the size of G, and let $\rho = 1 - \varepsilon$.

Next, the procedure sets q to be the least power of 2 that is at least $\log^c(n)$, and sets $\mathbb{F}$ be the finite field of size q. Note that $q = O(\log^c n)$. The procedure now invokes Lemma 3.2 on input G

and $\mathbb{F}$, thus obtaining a new constraint graph G_1. Note that by Lemma 3.2 if $\mathrm{UNSAT}(G) \geq \rho$, then $\rho_1 \stackrel{\text{def}}{=} \mathrm{UNSAT}(G_1) \geq \Omega\left(\frac{1}{q \cdot \log_q(n)} \cdot \rho\right)$.

Finally, the procedure sets d_0 to be an arbitrary constant such that $\rho_1 \geq h \cdot d_0 \cdot q^{-d_0/h}$. Note that this is indeed possible, since $\log_q(1/\rho_1)$ is a constant that depends only on ρ. Finally, the procedure invokes Lemma 3.3 on input G_1, $\mathbb{F}$, ρ_1, and d_0, and outputs the resulting constraint graph G'.

It remains to analyze the parameters of G'. By defining $p(k) = k^{O(c \cdot d_0)}$, we get that G' has size at most $p(n)$ and alphabet $\Sigma^{q^{O(d_0)}} = \Sigma^{p(\log n)}$. Furthermore, if $\mathrm{UNSAT}(G) \geq \rho$, then $\mathrm{UNSAT}(G_1) \geq \rho_1$. Therefore, by Lemma 3.3 and by the choice of d_0 , it holds that $\mathrm{SAT}(G') \leq O(1/q^{\Omega(1)})$. Since $O(1/q^{\Omega(1)}) = O(1/\log^{\Omega(c)} n) = O(1/\log^{\Omega(c)} |x|)$, we make G' have soundness of $O(1/\log n)$ by choosing c to be sufficiently large. ■

Remark 3.4. Recall that [MR08] prove a stronger version of the main theorem, saying that for every soundness error $s > 1/\mathrm{poly}\log n$ it holds that **NP** has a PCP system with soundness s and alphabet size $\exp(\mathrm{poly}(1/s))$. If one could prove a stronger version of Lemma 3.2 in which the soundness of G' is $\rho/\mathrm{poly}(q)$ and the alphabet size is $|\Sigma|^{\mathrm{poly}(q)}$ then the desired stronger version would follow using the same proof as above, without using a composition technique as in [MR08, DH09].

The reduction described in Theorem 1.1 is polynomial but not nearly-linear size. In fact, the construction of graphs with linear structure (Lemma 3.2) is nearly linear size (taking an instance of size n to an instance of size $q^2 \cdot n$). The part that incurs a polynomial and not nearly-linear blow-up is the reduction in Lemma 3.3 that relies on the derandomized direct product. It is possible that a more efficient derandomized direct product may lead to a nearly-linear size construction in total.

4 PCPs with Linear Structure

In this section we prove Lemma 3.2, which implies Theorem 1.2 by combining it with the PCP theorem (Theorem 2.12). The lemma which says that every constraint graph can be transformed into one that has linear structure. To this end, we use a family of structured graphs called de-Bruijn graphs. We show that de-Bruijn graphs have linear structure, and that every constraint graph can be embedded in some sense on a de-Bruijn graph. This embedding technique is a variant of a technique introduced by Babai et. al. [BFLS91] and Polishchuk and Spielman [PS94] for embedding circuits on de-Bruijn graphs. We begin by defining de-Bruijn graphs.

Definition 4.1. Let Λ be a finite alphabet and let $m \in \mathbb{N}$. The *de Bruijn graph* $\mathcal{DB}_{\Lambda,m}$ is the directed graph whose vertices set is Λ^m such that each vertex $(\alpha_1, \ldots, \alpha_t) \in \Lambda^m$ has outgoing edges to all the vertices of the form $(\alpha_2, \ldots, \alpha_t, \beta)$ for every $\beta \in \Lambda$.

Remark 4.2. We note that previous works used the special case of Definition 4.1 for $\Lambda = \{0,1\}$. In this work we use the more general definition.

Lemma 3.2 follows easily from the following two propositions. Proposition 4.3 says that any constraint graph can be embedded on a de Bruijn graph, and is proved in Section 4.1. Proposition 4.4 says that de Bruijn graphs have linear structure.

Proposition 4.3. *There exists a polynomial time procedure that satisfies the following requirements:*

- ***Input:***
  - *A constraint graph G of size n over alphabet Σ.*

- *A finite alphabet Λ whose size is a power of 2.*
- *A natural number m such that $|\Lambda|^m \geq 2 \cdot n$*

- ***Output:*** *A constraint graph G' such that the following holds:*
  - *The underlying graph of G' is the augmented wrapped de Bruijn graph $\mathcal{DB}_{\Lambda,m}$.*
  - *The size of G' is $|\Lambda|^{m+1}$.*
  - *G' has alphabet $\Sigma^{O(1)}$.*
  - *If G is satisfiable then G' is satisfiable.*
  - *If $\mathrm{UNSAT}(G) \geq \rho$ then $\mathrm{UNSAT}(G') \geq \Omega\left(\frac{n}{|\Lambda|^{m+1} \cdot l} \cdot \rho\right)$.*

Proposition 4.4. *Let $\mathbb{F}$ be a finite field and let $m \in \mathbb{N}$. Then, the de Bruijn graph $\mathcal{DB}_{\mathbb{F},m}$ has linear structure.*

Proof Items 1 and 3 of the definition of linear structure (Definition 3.1) follow immediately from the definition of de Bruijn graphs. To see that Item 2 holds, observe that in order for a tuple in $\mathbb{F}^{2m}$ to be an edge of $\mathcal{DB}_{\mathbb{F},m}$, it only needs to satisfy equality constraints, which are in turn linear constraints. Thus, the set of edges of $\mathcal{DB}_{\mathbb{F},m}$ form a linear subspace of $\mathbb{F}^{2m}$. ■

Lemma 3.2 is obtained by invoking Proposition 4.3 with $\Lambda = \mathbb{F}$, $m = \lceil \log_q (2 \cdot n) \rceil$ and combining it with Proposition 4.4.

4.1 Embedding constraint graphs on de Bruijn graphs

In this section we prove Proposition 4.3. The section is organized as follows: In Section 4.1.1 we give the required background on the routing properties of de Bruijn graphs. Then, in Section 4.1.2, we give an outline of the proof of Proposition 4.1.2. Finally, we give the full proof in Section 4.1.3.

4.1.1 de Bruijn graphs as routing networks

The crucial property of de Bruijn graphs that we use is that de Bruijn graph is a permutation routing network. To explain the intuition that underlies this notion, let us think of the vertices of the de Bruijn graph as computes in a network, such that two computers can communicate if and only if they are connected by an edge. Furthermore, sending a message from a computer to its neighbor takes one unit of time. Suppose that each computer in the network wishes to send a message to some other computer in the network, and furthermore each computer needs to receive a message from exactly one computer (that is, the mapping from source computer to target computer is a permutation). Then, the routing property of the de Bruijn network says that we can find paths in the network that have the following properties:

1. Each path corresponds to a message that needs to be sent. The path starts at the computer that wishes to send the message and ends at the computer to which the message is sent.

2. If all the messages are sent simultaneously along their corresponding paths, then at each unit of time, every computer needs to deal with exactly one message.

3. The paths are of length exactly $2 \cdot m$. This means that if all the messages are sent simultaneously along their corresponding paths, then after $2 \cdot m$ units of time all the packets will reach their destination.

Formally, this property can be stated as follows.

Fact 4.5 (). *Let $\mathcal{DB}_{\Lambda,m}$ be a de-Brujin graph. Then, given a permutation μ on the vertices of $\mathcal{DB}_{\Lambda,m}$ one can find a set of undirected paths of length $l = 2 \cdot m$ that connect each vertex v to $\mu(v)$ and that have the following property: For every $j \in [l]$, each vertex v is the j-th vertex of exactly one path. Furthermore, finding the paths can be done in time that is polynomial in the size of $\mathcal{DB}_{\Lambda,m}$.*

Remark 4.6. Fact 4.5 is proved in [Lei92] for the special case of $\Lambda = \{0,1\}$. The proof of the general case essentially follows the original proof, and replaces the looping algorithm of Benes with the fact that the edges of d-regular bipartite graphs can be decomposed into d perfect matchings. For completeness, we give the proof to the general case in Appendix B.

Remark 4.7. Note that the paths mentioned in Fact 4.5 are undirected. That is, if a vertex u appears immediately after a vertex v in path, then either (u,v) or (v,u) are edges of $\mathcal{DB}_{\Lambda,m}$.

4.1.2 Proof overview

In this section we give an overview of the proof of Proposition 4.3. Suppose we are given as input a constraint graph G which we want to embed on $\mathcal{DB} = \mathcal{DB}_{\Lambda,m}$. Recall that the size of G is at most $|\Lambda|^m$, so we may identify the vertices of G with some of the vertices of $\mathcal{DB}$.

Handling maximal degree 1 As a warm up, assume that G has maximal degree 1, i.e., G is a matching. In this case, we set the alphabet of the constraint graph G' to Σ^l for $l = 2m$. Fix any assignment π to G. We begin by describing how to construct a corresponding assignment π' to G'. We think of the vertices of G as computers, such that each vertex v wants to send the value $\pi(v)$ as a message to its unique neighbor in G. Using the routing property of the de Bruijn graph, we find paths for routing those messages along the edges of G'. Recall that if all the messages are sent simultaneously along those paths, then every computer has to deal with one packet at each unit of time, for l units of time. We now define the assignment π' to assign each vertex v of G' a tuple in Σ^l whose i-th element is the message with which v deals at the i-th unit of time.

The constraints of G' are defined to check that the routing is done correctly. That is, if the computer u is supposed to send a message to a vertex v between the j-th unit of time and the $(j+1)$-th unit of time, then the constraint of the edge betwen u and v will check that $\pi'(u)_j = \pi'(v)_{j+1}$. Furthermore, for each edge (u,v) of G, the constraints of G' check that the values $\pi'(v)_l$ and $\pi'(v)_1$ satisfy the edge (u,v). This condition should hold because if π' was constructed correctly according to π then $\pi'(v)_l = \pi(u)$ and $\pi'(v)_1 = \pi(v)$. It should be clear that the constraints of G' "simulate" the the constraints of G.

Handling arbitrary degree graphs Using the expander replacement technique of Papadimitriou and Yannakakis [PY91], we may assume that G is d-regular for some universal constant d. The d-regularity of G implies that the edges of G can be partitioned to d disjoint matchings $G_1, \ldots, G_d$ in polynomial time(see, e.g., [Cam98, Proposition 18.1.2]). Now, we set the alphabet of G' to be $\left(\Sigma^l\right)^d$, and handle each of the matchings G_i as before, each time using a "different part" of the alphabet symbols. In other words, the alphabet of G' consists of d-tuples of Σ^l, and so the constraints used to handle each matching G_i will refer to the i-th coordinates in those tuples. Finally, for vertex v, its constraints will also check that the message it sends in each of the d routings is the same. In other words, if $\pi'(v) = (\sigma_1, \ldots, \sigma_d) \in \left(\Sigma^l\right)^d$ then the constraints will check that $(\sigma_1)_1 = \ldots = (\sigma_d)_1$. As before, the constraints of resulting graph G' "simulate" the constraints of the original graph G.

Remark 4.8. Observe that the foregoing proof used only the routing property of de Bruijn graphs, and will work for any graph satisfying this property. In other words, Proposition 4.3 holds for any graph for which Fact 4.5 holds.

4.1.3 Detailed proof

We turn to present the full proof of Proposition 4.3. We use the following version of the expander-replacement technique of [PY91].

Lemma 4.9 ([Din07, Lemma 3.2])**.** *There exist universal constants $c, d \in \mathbb{N}$ and a polynomial time procedure that when given as input a constraint graph G of size n outputs a constraint graph G' of size $2 \cdot d \cdot n$ over alphabet Σ such that the following holds:*

- *G' has $2 \cdot n$ vertices and is d-regular.*
- *If G is satisfiable then so is G'.*
- *If $\mathrm{UNSAT}(G) \geq \rho$ then $\mathrm{UNSAT}(G') \geq \rho/c$.*

We turn to proving Proposition 4.3. When given as input a constraint graph G, a finite alphabet Λ and a natural number m such that $|\Lambda^m| \geq 2 \cdot n$, the procedure of Proposition 4.3 acts as follows. Let $l = 2 \cdot m$. The procedure begins by invoking Lemma 4.9 on G, resulting in a d-regular constraint graph G_1 over $2 \cdot n$ vertices. Then, the vertices of G_1 are identified with a subset of the vertices of the first layer of $\mathcal{DB} = \mathcal{DB}_{\Lambda,m}$ (note that this is possible since $|\Lambda^m| \geq 2 \cdot n$).

Next, the procedure partitions the edges of G_1 to d disjoint matchings, and extends those matchings to permutations $\mu_1, \ldots, \mu_d$ on the vertices of $\mathcal{DB}$ in the following way: Given a vertex v of $\mathcal{DB}$, if v is identified with a vertex of G_1 then μ_i maps v to its unique neighbor via the i-th matchning, and otherwise μ_i maps v to v. The procedure then applies Fact 4.5 to each permutation μ_i resulting in a set of paths $\mathcal{P}_i$ of length l. Let $\mathcal{P} = \bigcup \mathcal{P}_i$.

Finally, the procedure constructs G' by associating the edges of $\mathcal{DB}'$ with constraints in the following way. We set the alphabet of G' to be $\Sigma^{l \cdot d}$, viewed as $\left(\Sigma^l\right)^d$. If $\sigma \in \left(\Sigma^l\right)^d$, and we denote $\sigma = (\sigma_1, \ldots, \sigma_d)$, then we denote by $\sigma_{i,j}$ the element $(\sigma_i)_j \in \Sigma$. To define the constraints, let us consider their action on an assignment π' of G'. An edge (u, v) of $\mathcal{DB}'$ is associated with the constraint that accepts unless one of the following conditions holds:

1. There exists $i \in [d]$ such that the values $\left(\pi'(\mu_i(u))_{i,l}, \pi'(\mu_i(u))_{i,1}\right)$ do not satisfy the edge $(u, \mu_i(u))$ of G.
2. It either does not hold that $\pi'(u)_{1,1} = \ldots = \pi'(u)_{d,1}$ or that $\pi'(v)_{1,1} = \ldots = \pi'(v)_{d,1}$.
3. There exists $i \in [d]$ and $j \in [l]$ such that u and v are the j-th and $(j+1)$-th vertices of a path in $p \in \mathcal{P}_i$ respectively, but $\pi'(u)_{i,j} \neq \pi'(v)_{i,i+1}$.
4. Same as Condition 3, but when v is the j-th vertex of p and u is its $(j+1)$-th vertex.

The size of G' is indeed $|\Lambda|^{t+1}$: The graph is $|\Lambda|$-regular and contains $|\Lambda|^t$ vertices.. Furthermore, if G is satisfiable, then so is G': The satisfiability of G implies the satisfiability of G_1, so there exists a satisfying assignment π_1 for G_1. We construct a satisfying assignment π' from π_1 by assigning each vertex v of G' a value $\pi'(v)$, such that for each $i \in [d]$, if v is the j-th vertex of a path $p \in \mathcal{P}_i$ that connects the vertices u and $\mu_i(u)$, then we set $\pi'(v)_{i,j} = \pi(u)$. Note that this is well defined, since every vertex is the j-th vertex of exactly one path in $\mathcal{P}_i$.

It remains to analyze the soundness of G'. Suppose that $\mathrm{UNSAT}(G) \geq \rho$. Then, by Lemma 4.9 it holds that $\mathrm{UNSAT}(G_1) \geq \rho/c$. Let π' be an assignment to G' that minimizes the fraction of violated edges of G'. Without loss of generality, we may assume that for every vertex v of the $\mathcal{DB}$ it holds that $\pi'(v)_{1,1} = \ldots = \pi'(v)_{d,1}$: If there is a vertex v that does not match this condition, all of the edges attached to v are violated and therefore we can modify the value assigned to v by π' to match this condition without increasing the fraction of violated edges of π'. Define an assignment π_1 to G_1 by setting $\pi_1(v) = \pi'(v)_{1,1}$ (when v is viewed as a vertex of $\mathcal{DB}$).

Since $\mathrm{UNSAT}(G_1) \geq \rho/c$, it holds that π_1 violates at least ρ/c fraction of the edges of G_1, or in other words π_1 violates at least $\rho \cdot 2 \cdot n \cdot d/c$ edges of G_1. Thus, there must exist a permutation μ_i such that π_1 violates at least $\rho \cdot 2 \cdot n/c$ edges of G_1 of the form $(u, \mu_i(u))$. Fix such an edge $(u, \mu_i(u))$ and consider the corresponding path $p \in \mathcal{P}_i$. Observe that π' must violate at least one of the edges of p: To see it, note that if π' would satisfy all the edges on p, then it would imply that $\pi'(\mu_i(u))_{i,l} = \pi_1(u)$ and that $\pi'(\mu_i(u))_{i,1} = \pi_1(\mu_i(u))$, but the last two values violate the edge $(u, \mu_i(u))$ of G_1, and therefore π' must violate the last edge of p - contradiction. It follows that for each of the $\rho \cdot 2 \cdot n/c$ edges of μ_i that are violated by π_1 it holds that π' violates at least one edge of their corresponding path, and thus by averaging there must exist $j \in [l]$ such that for at least $\rho \cdot 2 \cdot n/c \cdot l$ edges of μ_i it holds that π' violates the j-th edge of their corresponding path.

Now, by the definition of the paths in $\mathcal{P}_i$, no edge of G' can be the j-th edge of two distinct paths in $\mathcal{P}_i$, and therefore it follows that there at least $\rho \cdot 2 \cdot n/c \cdot l$ edges of G' are violated by π'. Finally, there are $|\Lambda|^{m+1}$ edges in G', and this implies that π' violates a fraction of the edges of G' that is at least

$$\frac{\rho \cdot 2 \cdot n/c \cdot l}{|\Lambda|^{m+1}} = \Omega\left(\frac{n}{|\Lambda|^{m+1} \cdot l} \cdot \rho\right)$$

as required.

4.2 De Bruijn graphs have linear structure

In this section we prove Proposition 4.4.

Proposition 4.10 (4.4, restated)**.** *Let $\mathbb{F}$ be a finite field of size q and let t be a natural number. Then, the augmented de Bruijn graph $\mathcal{DB}'_{\mathbb{F},m}$ has linear structure.*

Let $m = t+1$. We begin by identifying the vertices of $\mathcal{DB}' = \mathcal{DB}'_{\mathbb{F},m}$ with the vectors of a vector space $\mathbb{F}^m$ as follows. Let γ denote the generator of the multiplicative group of $\mathbb{F}$, and note that this group is of size $q-1$. Recall that the vertices in each layer of $\mathcal{DB}'$ are identified with $\mathbb{F}^t$. Now, for each $i \in [q-1]$, we identify the vertices of the (non-dummy) i-th layer with vectors of $\mathbb{F}^m$ using the mapping $(\alpha_1, \ldots, \alpha_t) \in \mathbb{F}^t \to (\gamma^i, \alpha_1, \ldots, \alpha_t) \in \mathbb{F}^m$. We identify the vertices of the dummy layer with vectors of $\mathbb{F}^m$ using the mapping $(\alpha_1, \ldots, \alpha_t) \in \mathbb{F}^t \to (0, \alpha_1, \ldots, \alpha_t) \in \mathbb{F}^m$.

Next, let E denote the edges of $\mathcal{DB}'$ and view E as a subset of $\mathbb{F}^{2m}$ as in Definition 3.1. Observe that E is indeed a linear subspace of $\mathbb{F}^{2m}$. To see it, note that a vector $v \in \mathbb{F}^{2m}$ is in E if and only if v is either of the form $(\gamma^i, \alpha_1, \ldots, \alpha_t, \gamma^{i+1}, \alpha_2, \ldots, \alpha_t, \beta)$ or of the form $(0, \alpha_1, \ldots, \alpha^t, 0, \alpha_2, \ldots, \alpha_t, \beta)$. In other words, v is in E if and only if (i) $v_{m+1} = \gamma \cdot v_1$, and (ii) for each $i = 1, \ldots, t-1$, it holds that $v_{1+i} = v_{(m+1)+(i-1)}$. Since those are linear conditions, it follows that the edges form a linear subspace. It is easy to see that $\mathrm{left}(E) = \mathrm{right}(E) = \mathbb{F}^m$, and this concludes the proof.

Remark 4.11. The only reason for including the dummy layer in the graph $\mathcal{DB}'$ is in order to be able to identify the vectors of the form $(0, \alpha_1, \ldots, \alpha_t) \in \mathbb{F}^m$ with vertices of $\mathcal{DB}'$.

Remark 4.12. Note that the assumption that $q-1$ equals the number of layers l is not essential to the foregoing construction, and we could in fact work with any field size. To see it, observe that

we only used this assumption in order to have an element $\gamma \in \mathbb{F}$ whose order equals l, so we can use powers of γ to represent the index of a layer. However, if q was smaller than l, we could have taken some square matrix over $\mathbb{F}$ whose order equals l, and use powers of A to represent the index of a layer.

5 Derandomized Parallel Repetition of Constraint Graphs with Linear Structure

In this section we prove Lemma 3.3, restated below, by implementing a form of derandomized parallel repetition on graphs that have linear structure.

Lemma 5.1 (3.3, restated)**.** *There exist a universal constant h and a polynomial time procedure that satisfy the following requirements:*

- ***Input:***
  - *A finite field $\mathbb{F}$ of size q*
  - *A constraint graph $G = (\mathbb{F}^m, E)$ over alphabet Σ that has a linear structure.*
  - *A parameter $d_0 \in \mathbb{N}$ such that $d_0 < m/h^2$.*
  - *A parameter $\rho \in (0,1)$ such that $\rho \geq h \cdot d_0 \cdot q^{-d_0/h}$.*
- ***Output:*** *A constraint graph G' such that the following holds:*
  - *G' has size $n^{O(d_0)}$.*
  - *G' has alphabet $\Sigma^{q^{O(d_0)}}$.*
  - *If G is satisfiable then G' is satisfiable.*
  - *If $\mathrm{SAT}(G) < 1 - \rho$ then $\mathrm{SAT}(G') < h \cdot d_0 \cdot q^{-d_0/h}$.*
  - *G' has the projection property*

We begin by describing the construction of G'. Let $G = (\mathbb{F}^m, E)$ be the given constraint graph, let d_0 be the parameter from Lemma 3.3, and let $d_1 = O(d_0)$ be chosen later. The graph G' is bipartite. The right vertices of G' are identified with all the $2d_0$-subspaces of $\mathbb{F}^m$ (the vertex-space of G). The left vertices of G' are identified with all the $2d_1$-subspaces of the edge space E of G. If a satisfying assignment π for G exists, then one can extend it to a satisfying assignment Π for G' as follows: Π labels each $2d_0$-subspace A with $\pi_{|A}$, and each $2d_1$-subspace E_0 of edges with the values assigned by π to the endpoints of all the edges in F.

The edges of G' are constructed such that they simulate the action of the "E-test" described in Figure 2.

The rest of this section is organized as follows. In Section 5.1 we introduce a specialized direct product test that we use in order to analyze the E-test. Then, in Section 5.2 we analyze the completeness, size, and alphabet of G'. Finally, in Section 5.3, we analyze the soundness of G'.

1. Let F_L and F_R to be random d_1-subspaces of E, and let

$$B_L \stackrel{\text{def}}{=} \text{left}(F_L), \quad B_R \stackrel{\text{def}}{=} \text{right}(F_R), \quad F \stackrel{\text{def}}{=} F_L + F_R.$$

   F_L and F_R are chosen to be uniformly and independently distributed d_1-subspaces of E conditioned on $\dim(F) = 2d_1$, $\dim(B_L) = d_1$, $\dim(B_R) = d_1$, and $B_L \cap B_R = \{0\}$.

2. Let A_L and A_R be uniformly distributed d_0-subspaces of B_L and B_R respectively, and let

$$A \stackrel{\text{def}}{=} A_L + A_R.$$

3. Accept if and only if $\Pi(F)_{|(A_L,A_R)} = \Pi(A)_{|(A_L,A_R)}$ and the assignment $\Pi(F)$ satisfies the edges in F.

Figure 2: The E-test

1. Choose two uniformly distributed and disjoint d_1-subspaces B_1, B_2 of $\mathbb{F}^m$.
2. Choose two uniformly distributed d_0-subspaces $A_1 \subseteq B_1$, $A_2 \subseteq B_2$.
3. Accept if and only if $\Pi(B_1, B_2)_{|(A_1,A_2)} = \Pi(A_1 + A_2)_{|(A_1,A_2)}$.

Figure 3: The S-test

5.1 The specialized direct product test

In order to analyze the E-test, we introduce a variant of the direct product test of [IKW09] that is specialized to our needs. We refer to this variant as the *specialized direct product test*, abbreviated the "S-test".

We begin with some notation: Given two functions $f : U \to \Sigma$, $g : V \to \Sigma$ and two subsets $S \subseteq U$, $T \subseteq V$ we denote by $(f, g)_{|(S,T)}$ the pair $\left(f_{|S}, g_{|T}\right)$, and abbreviate $f_{|(S,T)} = (f, f)_{|(S,T)}$. Given two pairs of functions $f_1, f_2 : U \to \Sigma$ and $g_1, g_2 : V \to \Sigma$, we denote by $(f_1, g_1) \overset{\alpha}{\approx} (f_2, g_2)$ the fact that both $f_1 \overset{\alpha}{\approx} f_2$ and $g_1 \overset{\alpha}{\approx} g_2$, and otherwise we denote $(f_1, g_1) \overset{\alpha}{\not\approx} (f_2, g_2)$.

Now, given an string $\pi : \mathbb{F}^m \to \Sigma$, we define its *S-direct product* Π (with respect to $d_0, d_1 \in \mathbb{N}$) as follows: Π assigns each $2d_0$-subspace $A \subseteq \mathbb{F}^m$ the function $\pi_{|A}$, and assigns each pair of disjoint d_1-subspaces (B_1, B_2) the pair of functions $(\pi_{|B_1}, \pi_{|B_2})$.

We turn to consider the task of testing whether a given assignment Π is the S-direct product of some string $\pi : \mathbb{F}^m \to \Sigma$. In our settings, we are given an assignment Π that assigns each $2d_0$-subspace A to a function $a : A \to \Sigma$ and each pair of disjoint d_1-subspaces (B_1, B_2) to a pair of functions $b_1 : B_1 \to \Sigma$, $b_2 : B_2 \to \Sigma$. We wish to check whether Π is a S-direct product of some $\pi : \mathbb{F}^m \to \Sigma$, and to this end we invoke the S-test, described in Figure 3.

It is easy to see that if Π is a S-direct product then the S-test always accepts. Furthermore, it can be shown that if Π is "far" from being a S-direct product, then the S-test rejects with high probability. As in the P-test, this holds even if Π is a randomized assignment. Formally, we have the following result.

Theorem 5.2. *There exist universal constant $h', c \in \mathbb{N}$ such that the following holds: Let $\varepsilon \geq h' \cdot d_0 \cdot q^{-d_0/h'}$, $\alpha \stackrel{\text{def}}{=} h' \cdot d_0 \cdot q^{-d_0/h'}$. For every $d_0 \in \mathbb{N}$, $d_1 \geq h' \cdot d_0$, and $m \geq h' \cdot d_1$, suppose that a (possibly randomized) assignment Π passes the S-test with probability at least ε. Then there*

exists an assignment $\pi : \mathbb{F}^m \to \Sigma$ for which the following holds. Let B_1, B_2 be uniformly distributed and disjoint d_1-subspaces of $\mathbb{F}^m$, let A_1 and A_2 be uniformly distributed d_0-subspaces of B_1 and B_2 respectively, and denote $A = A_1 + A_2$. Then:

$$\Pr\left[\Pi(B_1,B_2)_{|(A_1,A_2)} = \Pi(A)_{|(A_1,A_2)} \quad \text{and } \Pi(B_1,B_2) \overset{\alpha}{\approx} \pi_{|(B_1,B_2)}\right] = \Omega(\varepsilon^c)$$

We defer the proof of Theorem 5.2 to Section 6.

5.2 The completeness, size, and alphabet of G'

Completeness is immediate: if G is satisfiable then so is G'.

Let us verify the size and alphabet-size of G'. We choose $d_1 = h \cdot d_0$, where h is the universal constant from Lemma 3.3 to be chosen later. The size of G' is at most the number of $2d_1$-subspaces of E multiplied by the number of $2d_0$-subspaces of $\mathbb{F}^m$, which is $|E|^{2d_1} \cdot |\mathbb{F}^m|^{2d_0}$. It holds that $d_0 < d_1$, and furthemore the linear structure of G' implies that $\dim E \geq m$ (by Item 3 of Definition 3.1), so it follows that $|\mathbb{F}^m|^{2d_0} \leq |E|^{2d_1}$ and thus $|E|^{2d_1} \cdot |\mathbb{F}^m|^{2d_0} \leq |E|^{4d_1}$. Finally, observe that the size of G is $n = |E|$, so it follows that the size of G' is at most $n^{4d_1} = n^{O(d_0)}$, as required.

It suffices to set the alphabet of G' to $\Sigma^{2 \cdot q^{2 \cdot d_1}}$, since each $2d_1$-subspace $F \subseteq E$ contains q^{2d_1} edges and each has two endpoints. Furthermore, the labels assigned by Π to $2d_0$-subspaces A of $\mathbb{F}^m$ do not require a larger alphabet. The alphabet of G' is therefore $\Sigma^{2 \cdot q^{2 \cdot d_1}} = \Sigma^{q^{O(d_0)}}$, as required.

5.3 The soundness of the derandomized parallel repetition

In this section we prove the soundness of G': namely, that if $\mathrm{SAT}(G) < 1 - \rho$, then $\mathrm{SAT}(G') \leq \varepsilon \overset{\text{def}}{=} h \cdot d_0 \cdot q^{-d_0/h}$, where h is the universal constant from Lemma 3.3. We will choose h to be sufficiently large such that the various inequalities in the following proof will hold. To this end, we note that throughout all the following proof, increasing the choice of h does not break any of our assumptions on h, so we can always choose a larger h to satisfy the required inequalities.

Let h' and c be the universal constants whose existence is guaranteed by Theorem 5.2, and let α denote the corresponding value from Theorem 5.2. We will choose the constant h to be at least h'.

Let Π be an assignment to G' that maximizes the acceptance probability of the E-test. Without loss of generality, we may assume that for every d_1-subspace $F \subseteq E$ it holds that the assignment $\Pi(F)$ satisfies the edges in F, since we can always modify Π to an assignment that satisfies this property and has at least the same acceptance probability.

Notation 5.3. Let us denote by $\mathcal{T}$ the event in which the E-test accepts Π. By our assumption on Π, the event $\mathcal{T}$ is equivalent to the event $\Pi(F)_{|(A_L,A_R)} = \Pi(A)_{|(A_L,A_R)}$. With a slight abuse of notation, for a subspace $F \subseteq E$ and an assignment $\pi : \mathbb{F}^m \to \Sigma$, we denote by $\Pi(F) \overset{\alpha}{\approx} \pi$ the claim that for at least $1 - \alpha$ fraction of the edges e of F it holds that $\Pi(F)$ is consistent with π on both the endpoints of e, and otherwise we denote $\Pi(F) \overset{\alpha}{\not\approx} \pi$.

Our proof is based on two steps:

- We will show (in Proposition 5.4 below) that if the test accepts with probability ε, then it is "because" Π is consistent with some underlying assignment $\pi : \mathbb{F}^m \to \Sigma$. This is done essentially by observing that the E-test "contains" an S-test, and reducing to the analysis of the S-test.

- On the other hand, we will show (in Proposition 5.5 below) that for every assignment $\pi : \mathbb{F}^m \to \Sigma$ the probability that the test accepts while being consistent with π is negligible. This is showed roughly as follows: Any fixed assignment π is rejected by at least ρ fraction of G's edges. Furthermore, the subspace F queried by the test is approximately a uniformly distributed subspace of E, and hence a good sampler of E. It follows F must contain $\approx \rho$ fraction of edges of G that reject π, and therefore $\Pi(F)$ must be inconsistent with π.

We have reached a contradiction and therefore conclude that the E-test accepts with probability less than ε. We now state two said propositions.

Proposition 5.4. *There exists $\varepsilon_0 = \Omega(\varepsilon^c)$ such that the following holds: If $\Pr[\mathcal{T}] \geq \varepsilon$, then there exists an assignment $\pi : \mathbb{F}^m \to \Sigma$ such that $\Pr\left[\mathcal{T} \text{ and } \Pi(F) \overset{4\cdot\alpha}{\approx} \pi\right] \geq \varepsilon_0$.*

Proposition 5.5. *Let ε be as in Lemma 5.4. Then, for every assignment $\pi : \mathbb{F}^m \to \Sigma$ it holds that $\Pr\left[\mathcal{T} \text{ and } \Pi(F) \overset{4\cdot\alpha}{\approx} \pi\right] < \varepsilon_0$.*

Clearly the two propositions together imply that $\Pr[\mathcal{T}] \leq \varepsilon$, as required.

Before turning to the proofs of Propositionss 5.4 and 5.5 let us state a useful claim that says that if we take a random d-subspace of edges and project it to its left endpoints (respectively, right endpoints), we get a random d-subspace of vertices with high probability.

Claim 5.6. *Let $d \in \mathbb{N}$ and let E_a be a uniformly distributed d-subspace of E. Then, $\Pr[\dim(\text{left}(E_a)) = d] \geq 1 - d/q^{m-d}$, and conditioned on $\dim(\text{left}(E_a)) = d$, it holds that $\text{left}(E_a)$ is a uniformly distributed d-subspace of $\mathbb{F}^m$. The same holds for $\text{right}(E_a)$.*

More generally, let E_b be a fixed subspace of E such that $\dim(E_b) > d$ and $\dim(\text{left}(E_b)) > d$. Let E_a be a uniformly distributed d-subspace of E_b. Then, $\Pr[\dim(\text{left}(E_a)) = d] \geq 1 - d/q^{\dim(\text{left}(E_b))-d}$, and conditioned on $\dim(\text{left}(E_a)) = d$, it holds that $\text{left}(E_a)$ is a uniformly distributed d-subspace of $\text{left}(E_b)$. Again, the same holds for $\text{right}(E_a)$.

We defer the proof of to Appendix C

5.3.1 Proof of Proposition 5.4

Suppose that $\Pr[\mathcal{T}] \geq \varepsilon$. We prove Proposition 5.4 by arguing that the E-test contains an "implicit S-test" and applying Theorem 5.2.

The implicit S-test We extend Π to pairs of disjoint d_1-subspaces of $\mathbb{F}^m$ in a randomized manner as follows: Given a pair of disjoint d_1-subspaces B_1 and B_2, we choose F_1 and F_2 to be uniformly distributed and disjoint d_1-subspaces of E such that $\text{left}(F_1) = B_1$ and $\text{right}(F_2) = B_2$, and set $\Pi(B_1, B_2) = \Pi(F_1 + F_2)_{|(B_1,B_2)}$. Now, observe that

$$\Pr[\mathcal{T}] = \Pr\left[\Pi(B_1, B_2)_{|(A_1,A_2)} = \Pi(A_1 + A_2)_{|(A_1,A_2)}\right]$$

for uniformly distributed and disjoint d_1-subspaces B_1 and B_2 and unifromly distributed d_0-subspaces A_1 and A_2 of B_1 and B_2 respectively. The reason is that the subspaces B_L and B_R of the E-test are distributed like a pair of disjoint uniformly distributed d_1-subspaces of $\mathbb{F}^m$, and that conditioned on a specific choice of B_L and B_R, the subspaces F_L and F_R are distributed like disjoint uniformly distributed d_1-subspaces of E for which $\text{left}(F_L) = B_L$ and $\text{right}(F_R) = B_R$. It thus follows the E-test performs in a way an S-test on the extended assignment Π.

Next, we note that by choosing h to be sufficiently large, the foregoing "implicit S-test" matches the requirements of Theorem 5.2, and we can thus apply this theorem. It follows that there exists an assignment $\pi : \mathbb{F}^m \to \Sigma$ such that

$$\Pr\left[\Pi(B_L, B_R)_{(A_L,A_R)} = \Pi(A)_{|(A_L,A_R)} \text{ and } \Pi(B_L, B_R) \overset{\alpha}{\approx} \pi_{(B_L,B_R)}\right] \geq \Omega(\varepsilon^c)$$

or in other words

$$\Pr\left[\mathcal{T} \text{ and } \Pi(F)_{|(B_L,B_R)} \overset{\alpha}{\approx} \pi_{|(B_L,B_R)}\right] \geq \Omega(\varepsilon^c) \tag{4}$$

We turn to show that

$$\Pr\left[\mathcal{T} \text{ and } \Pi(F) \overset{4\alpha}{\approx} \pi\right] \geq \Omega(\varepsilon^c).$$

We will prove that if F is such that $\Pi(F) \overset{4\alpha}{\not\approx} \pi$, then for a random choice of B_L, B_R conditioned on F, it is highly unlikely that Inequality 4 still holds. Formally, we will prove the following.

Claim 5.7. *For every fixed $2d_0$-subspace F_0 of E such that $\Pi(F_0) \overset{4\alpha}{\not\approx} \pi_i$, it holds that*

$$\Pr\left[\Pi(F)_{|(B_L,B_R)} \overset{\alpha}{\approx} \pi_{|(B_L,B_R)} \middle| F = F_0\right] \leq 1/\left(q^{d_1-2} \cdot \alpha^2\right)$$

We defer the proof of Claim 5.7 to the end of this section. Claim 5.7 immediately implies the following.

Corollary 5.8. *It holds that*

$$\Pr\left[\Pi(F)_{|(B_L,B_R)} \overset{\alpha}{\approx} \pi_{|(B_L,B_R)} \middle| \Pi(F) \overset{4\alpha}{\not\approx} \pi\right] \leq 1/\left(q^{d_1-2} \cdot (\alpha/2)^2\right)$$

By combining Corollary 5.8 with Inequality 4, and by choosing h to be sufficiently large, it follows that

$$\Pr\left[\mathcal{T} \text{ and } \Pi(F)_{|(B_L,B_R)} \overset{\alpha}{\approx} \pi_{|(B_L,B_R)} \text{ and } \Pi(F) \overset{4\alpha}{\approx} \pi\right] \geq \Omega(\varepsilon^c),$$

This implies that

$$\Pr\left[\mathcal{T} \text{ and } \Pi(F) \overset{4\alpha}{\approx} \pi\right] \geq \Omega(\varepsilon^c)$$

Setting ε_0 to be the latter lower bound finishes the proof.

Proof of Claim 5.7 Observe that the assumption $\Pi(F_0) \overset{4\alpha}{\not\approx} \pi$ implies that one of the following holds

$$\Pi(F_0)_{|\text{left}(F_0)} \overset{2\alpha}{\not\approx} \pi_{|\text{left}(F_0)}$$

$$\Pi(F_0)_{|\text{right}(F_0)} \overset{2\alpha}{\not\approx} \pi_{|\text{right}(F_0)}$$

Without loss of generality, assume that the first holds. Now, when conditioning on $F = F_0$, it holds that F_L is a uniformly distributed d_1-subspace of F_0 satisfying $\dim(\text{left}(F_L)) = d_1$. By Claim 5.6 (with $E_b = F_0$ and $E_a = F_L$), under the conditioning on $\dim(\text{left}(F_L)) = d_1$, it holds that $B_L \overset{\text{def}}{=} \text{left}(F_L)$ is a uniformly distributed d_1-subspace of $\text{left}(F_0)$. Therefore, by Lemma 2.3, the event $\Pi(F)_{|B_L} \overset{\alpha}{\not\approx} \pi_{|B_L}$ occurs with probability at least

$$1 - 1/\left(q^{d_1-2} \cdot \left(\alpha - q^{-d_1}\right)^2\right) \geq 1 - 1/\left(q^{d_1-2} \cdot (\alpha/2)^2\right)$$

as required. ■

5.3.2 Proof of Proposition 5.5

Fix an assignment $\pi : \mathbb{F}^m \to \Sigma$. By assumption it holds that $\mathrm{SAT}(G) < 1-\rho$, and therefore π must violate a set E^* of edges of G of density at least ρ. Below we will show that at least $\rho/2$ fraction of the edges in F are in E^* with probability greater than $1-\varepsilon_0$. Now, observe that $\Pi(F)$ cannot be consistent with π on the edges in E^*, and hence whenever the latter event occurs it holds that $\Pi(F) \overset{\rho/2}{\not\approx} \pi$. However, for sufficiently large choice of h, it holds that $\rho/2 > 4 \cdot \alpha$ and therefore the probability that $\Pi(F) \overset{4\cdot\alpha}{\approx} \pi$ is less than ε_0, as required.

It remains to show that

$$\Pr\left[\frac{|F \cap E^*|}{|F|} \geq \rho/2\right] > 1 - \varepsilon_0$$

We prove the above inequality by showing that F is close to being a uniformly distribured $2d_1$-subspace of E, and then applying Lemma 2.3. To this end, let F'_L and F'_R be uniformly distributed d_1-subspaces of F, and let $F' = F'_L + F'_R$. Let us denote by $\mathcal{E}_1$ the event in which $\dim(F') = 2d_1$, and by $\mathcal{E}_2$ the event in which $\mathrm{left}(F'_L)$ and $\mathrm{right}(F'_R)$ are disjoint and are of dimension d_1. Observe that conditioned on $\mathcal{E}_1$ and $\mathcal{E}_2$ the subspace F' is distributed exactly like the subspace F. It therefore holds that

$$\begin{aligned}
\Pr\left[\frac{|F \cap E^*|}{|F|} \geq \rho/2\right] &= \Pr\left[\frac{|F' \cap E^*|}{|F'|} \geq \rho/2 \,\middle|\, \mathcal{E}_1 \text{ and } \mathcal{E}_2\right] \\
&\geq \Pr\left[\frac{|F' \cap E^*|}{|F'|} \geq \rho/2 \text{ and } \mathcal{E}_2 \,\middle|\, \mathcal{E}_1\right] \\
&\geq \Pr\left[\frac{|F' \cap E^*|}{|F'|} \geq \rho/2 \,\middle|\, \mathcal{E}_1\right] - \Pr\left[\neg\mathcal{E}_2 | \mathcal{E}_1\right] \\
&\geq \Pr\left[\frac{|F' \cap E^*|}{|F'|} \geq \rho/2 \,\middle|\, \mathcal{E}_1\right] - \frac{\Pr[\neg\mathcal{E}_2]}{\Pr[\mathcal{E}_1]}
\end{aligned}$$

Now, observe that conditioned on $\mathcal{E}_1$, the subspace F' is a uniformly distributed $2d_1$-subspace of E. Thus, by Lemma 2.3 it holds that

$$\Pr\left[\frac{|F' \cap E^*|}{|F'|} \geq \rho/2 \,\middle|\, \mathcal{E}_1\right] \geq 1 - 1/q^{2d_1-2} \cdot \left(\rho/2 - q^{-2d_1}\right)^2 \geq 1 - 1/q^{2d_1-2} \cdot (\rho/3)^2$$

Moreover, by Proposition 2.13 it holds that

$$\begin{aligned}
\Pr[\mathcal{E}_1] &\geq 1 - 2d_1/q^{\dim E - 2d_1} \\
&\geq 1 - 2d_1/q^{m-2d_1} \\
&\geq \frac{1}{2}
\end{aligned}$$

Finally, we upper bound $\Pr[\mathcal{E}_2]$. By Claim 5.6 (with $E_b = E$ and $E_a = F'_L, F'_R$) it holds that $\dim(\mathrm{left}(F'_L)) = \dim(\mathrm{right}(F'_R)) = d_1$ with probability at least $1 - 2 \cdot d_1/q^{m-d_1}$. Furthermore, conditioned on the latter event, it holds that $\mathrm{left}(F'_L)$ and $\mathrm{right}(F'_R)$ are uniformly distributed d_1-subspaces of $\mathbb{F}^m$, and it is also easy to see that those subspaces are independent. By Proposition 2.13, this implies that conditioned on $\dim(\mathrm{left}(F'_L)) = \dim(\mathrm{right}(F'_R)) = d_1$ the subspaces $\mathrm{left}(F'_L)$ and $\mathrm{right}(F'_R)$ are disjoint with probability at least $1 - 2d_1/q^{m-2\cdot d_1}$, and hence $\Pr[\mathcal{E}_2] \geq 1 - 4d_1/q^{m-2\cdot d_1}$ as required.

1. Choose two uniformly distributed d_1-subspaces B_1, B_2 of $\mathbb{F}^m$.
2. Choose two uniformly distributed d_0-subspaces $A_1 \subseteq B_1$, $A_2 \subseteq B_2$.
3. Accept if and only if $\Pi(B_1, B_2)_{|(A_1,A_2)} = \Pi(A_1, A_2)$.

Figure 4: The P^2-test

We conclude that that

$$\begin{aligned}
\Pr\left[\frac{|F \cap E^*|}{|F|} \geq \rho/2\right] &\geq \Pr\left[\frac{|F' \cap E^*|}{|F'|} \geq \rho/2 \middle| \mathcal{E}_1\right] - \frac{\Pr[\neg\mathcal{E}_2]}{\Pr[\mathcal{E}_1]} \\
&\geq 1 - 1/q^{2\cdot d_1 - 2} \cdot (\rho/3)^2 - \frac{4 \cdot d_1/q^{m-2\cdot d_1}}{1/2} \\
&= 1 - 1/q^{2\cdot d_1 - 2} \cdot (\rho/3)^2 - 8 \cdot d_1/q^{m-2\cdot d_1} \\
&> 1 - \varepsilon_0
\end{aligned}$$

where the last inequality holds for sufficiently large choice of h. This concludes the proof.

6 The Analysis of the Specialized Direct Product Test

In this section we provide the analysis of the S-test and prove Theorem 5.2. The proof proceeds in two steps. First, in Section 6.1, we define and analyze an intermediate direct product test, which we call the P^2-test. Then, in Section 6.2, we reduce the analysis of the S-test to the P^2-test.

In the rest of this section, we let $\mathbb{F}$ be a finite field of size q and let $d_0, d_1 \in \mathbb{N}$.

6.1 The P^2-test

In this section we define and analyze the P^2-test. Infornally, the P^2-test consists of two P-tests that are performed simultaneously. Details follow.

Given two strings $\pi_1, \pi_2 : \mathbb{F}^m \to \Sigma$, we define their *$P^2$-direct product* Π (with respect to $d_0, d_1 \in \mathbb{N}$) as follows: Π assigns each pair of d_0-subspaces (A_1, A_2) the pair of functions $(\pi_{1|A_1}, \pi_{2|A_2})$, and assigns each pair of d_1-subspaces (B_1, B_2) to the pair of functions $(\pi_{1|B_1}, \pi_{2|B_2})$. We consider the task of testing whether a given assignment Π is the P^2-direct product of some pair of strings $\pi_1, \pi_2 : \mathbb{F}^m \to \Sigma$. That is, we are given an assignment Π , and in order to check whether Π is a P^2-direct product, we invoke the P^2-test, described in Figure 4.

It is easy to see that if Π is a P^2-direct product then the P^2-test always accepts. Again, it can be shown that if Π is "far" from being a P^2-direct product, then the P^2-test rejects with high probability, and that this holds even if Π is a randomized assignment. Formally, we have the following result.

Theorem 6.1. *There exist universal constants $h, c \in \mathbb{N}$ such that the following holds: Let $\varepsilon \geq h \cdot d_0 \cdot q^{-d_0/h}$, $\alpha \stackrel{\text{def}}{=} h \cdot d_0 \cdot q^{-d_0/h}$. Assume that $d_1 \geq h \cdot d_0$, $m \geq h \cdot d_1$. Suppose that an assignment Π passes the P^2-test with probability at least ε. Then, there exist two assignments π_1 and π_2 to $\mathbb{F}^m$ such that for uniformly distributed B_1, B_2, A_1, A_2 as in the P^2-test it holds that*

$$\Pr\left[\Pi(B_1, B_2)_{|(A_1,A_2)} = \Pi(A_1, A_2) \quad \text{and } \Pi(A_1, A_2) \stackrel{\alpha}{\approx} (\pi_{1|A_1}, \pi_{2|A_2}) \quad \text{and } \Pi(B_1, B_2) \stackrel{\alpha}{\approx} (\pi_{1|B_1}, \pi_{2|B_2})\right]$$

is at least $\Omega(\varepsilon^c)$.

In the rest of this section we prove Theorem 6.1. We denote by $\mathcal{P}$ the event in which the P^2-test accepts, that is, that $\Pi(B_1, B_2)_{|(A_1,A_2)} = \Pi(A_1, A_2)$. The core of the proof is the following lemma:

Lemma 6.2. *There exist universal constants $h', c' \in \mathbb{N}$ such that the following holds: Let $\varepsilon \geq h' \cdot d_0 \cdot q^{-d_0/h'}$, $\alpha' \stackrel{\text{def}}{=} h' \cdot d_0 \cdot q^{-d_0/h'}$. Assume that $d_1 \geq h' \cdot d_0$, $m \geq h' \cdot d_1$. If Π passes the P^2-test with probability at least ε then there exists an assignment $\pi_2 : \mathbb{F}^m \to \Sigma$ such that*

$$\Pr\left[\mathcal{P} \text{ and } \Pi(A_1, A_2)_{|A_2} \stackrel{\alpha'}{\approx} \pi_{2|A_2} \text{ and } (B_1, B_2)_{|B_2} \stackrel{\alpha'}{\approx} \pi_{2|B_2}\right] \geq \Omega(\varepsilon^{c'})$$

and symmetrically, there exists a function $\pi_1 : \mathbb{F}^m \to \Sigma$ such that

$$\Pr\left[\mathcal{P} \text{ and } \Pi(A_1, A_2)_{|A_1} \stackrel{\alpha'}{\approx} \pi_{1|A_1} \text{ and } (B_1, B_2)_{|B_1} \stackrel{\alpha'}{\approx} \pi_{1|B_1}\right] \geq \Omega(\varepsilon^{c'})$$

We prove Lemma 6.2 in Section 6.1.1. We turn to derive Theorem 6.1 from Lemma 6.2.

Proof of Theorem 6.1 We will choose h to be larger than the constant h' of Lemma 6.2, so we can apply this lemma. Let $\pi_2 : \mathbb{F}^m \to \Sigma$ be the assignment guaranteed by Lemma 6.2, and let Π' be an assignment that is obtained from Π as follows:

1. For every pair (A_1, A_2) for which $\Pi(A_1, A_2)_{|A_2} \stackrel{\alpha'}{\approx} \pi_{2|A_2}$, set $\Pi'(A_1, A_2) = \Pi(A_1, A_2)$.
2. For every other pair (A_1, A_2), set $\Pi'(A_1, A_2) = \perp$, where $\perp$ is some special value on which the test never accepts.
3. Set the pairs (B_1, B_2) similarly.

The probability ε' that the assignment Π' passes the P^2-test is at least $\Omega(\varepsilon^{c'})$ by the definition of π_2. By choosing h to be sufficiently larger than the corresponding constants of Lemma 6.2, we can make sure that ε' satisfies the requirements of Lemma 6.2. Therefore, we can deduce by Lemma 6.2 that there exists an assignment $\pi_1 : \mathbb{F}^m \to \Sigma$ such that

$$\Pr\left[\mathcal{P} \text{ and } \Pi'(A_1, A_2)_{|A_1} \stackrel{\alpha'}{\approx} \pi_{1|A_1} \text{ and } \Pi'(B_1, B_2)_{|B_1} \stackrel{\alpha'}{\approx} \pi_{1|B_1}\right] \geq \Omega((\varepsilon')^{c'}) = \Omega(\varepsilon^{(c')^2}).$$

We now choose $c = (c')^2$. Since the test never accepts when Π' answers $\perp$, we deduce that

$$\Pr\left[\mathcal{P} \text{ and } \Pi(A_1, A_2) \stackrel{\alpha'}{\approx} (\pi_{1|A_1}, \pi_{2|A_2}) \text{ and } \Pi(B_1, B_2) \stackrel{\alpha'}{\approx} (\pi_{1|B_1}, \pi_{2|B_2})\right] \geq \Omega(\varepsilon^c)$$

Choosing h such that $\alpha \geq \alpha'$ completes the proof. ■

Remark 6.3. If Π is randomized, then the definition of Π' in the foregoing proof should be slightly changed to consider the internal randomness of Π. That is, we define Π' to be a randomized assignment, and obtain it from Π as follows. For every pair (A_1, A_2) and every internal randomness ω of Π, let us denote by (a_1, a_2) the output of Π on (A_1, A_2) and randomness ω. We define the output of Π' on (A_1, A_2) and randomness ω to be (a_1, a_2) if $a_2 \stackrel{\alpha'}{\approx} \pi_{2|A_2}$, and define it to be $\perp$ otherwise. The definition for pairs (B_1, B_2) is again similar.

6.1.1 The proof of Lemma 6.2

We prove Lemma 6.2 only for the assignment π_2, and the conclusion π_1 can be proved analogously. The proof proceeds in three steps. First, we rely on Theorem 2.1 to find for each pair of A_1, B_1 a direct product function that agrees (on average) with a good fraction of $\Pi(A_1, \cdot)$ and $\Pi(B_1, \cdot)$. Then, we show that for each A_1 separately, the number of distinct such functions is bounded. Next, we show that there is a single function π such that the probability that the test accepts and $\Pi(A_1, A_2)_{|A_2} \approx \pi_{|A_2}$ is non-negligible (Apriory there could have been a different π for each A_1). Finally, we extend the latter result for d_1-subspaces B_1, B_2. Let h_1 be the universal constant whose existence is guaranteed in Theorem 2.1, and let α_1 be the corresponding value from Theorem 2.1.

Step 1. Consider the bipartite graph corresponding to the P-test, that is, the graph whose left vertices are d_0-subspaces and whose right vertices are d_1-subspaces, and such that a d_0-subspace A_1 is connected to a d_1-subspace B_1 by an edge if and only if $A_1 \subseteq B_1$. . We label an edge (A_1, B_1) by $\pi : \mathbb{F}^m \to \Sigma$ if

$$\Pr_{A_2,B_2}\left[\mathcal{P} \text{ and } \Pi(B_1,B_2)_{|B_2} \overset{\alpha_1}{\approx} \pi_{|B_2} \text{ and } \Pi(A_1,A_2)_{|A_2} \overset{\alpha_1}{\approx} \pi_{|A_2}\right] \geq \Omega\left(\varepsilon^4\right)$$

If no such π exists then do not label the edge.

Fix A_1, B_1. We will choose the universal constant h' to be at least $2 \cdot h_1$. If the probability of passing the P^2-test conditioned on A_1, B_1 is at least $\varepsilon/2$, then we claim that the edge is labeled. Indeed, define an assignment $\Pi_{(A_1,B_1)}$ by

$$\Pi_{(A_1,B_1)}(A_2) = \Pi(A_1,A_2)_{|A_2} \quad \text{and } \Pi_{(A_1,B_1)}(B_2) = \Pi(B_1,B_2)_{|B_2}$$

If $\Pi_{(A_1,B_1)}$ passes the P-test with probability at least $\varepsilon/2$, then by Theorem 2.1 there is an assignment π as needed (since $h' \geq 2 \cdot h_1$).

Furthermore, observe that by averaging at least $\varepsilon/2$ of the edges (A_1, B_1) have conditional success at least $\varepsilon/2$, so (A_1, B_1) is labeled.

Step 2. Fix B_1 and let $L(B_1)$ be the labels on edges touching B_1. Consider the following "pruning" process: arbitrarily choose a label $\pi \in L(B_1)$ and remove all elements in $L(B_1)$ that are within relative Hamming distance $3\alpha_1$ of π. Repeat until no more labels can be removed. Let $L'(B_1)$ denote the remaining set of labels. The set $L'(B_1)$ has the following properties

- Every pair of labels in $L'(B_1)$ are at least $3\alpha_1$ apart, and
- Every $f \in L(B_1)$ is $3\alpha_1$-close to some label in $L'(B_1)$.

We prove that $|L'(B_1)| \leq O(1/\varepsilon^4)$, using an argument in the spirit of the Johnson bound: Suppose $L'(B_1) = \{\pi_1, \pi_2, \ldots\}$ is non-empty. For every $\pi_i \neq \pi_j \in L'(B)$ let us denote

$$\begin{aligned} p_i &\overset{\text{def}}{=} \Pr_{B_2}\left[\Pi(B_1,B_2)_{|B_2} \overset{\alpha_1}{\approx} \pi_{i|B_2}\right] \\ p_{i,j} &= \Pr_{B_2}\left[\Pi(B_1,B_2)_{|B_2} \overset{\alpha_1}{\approx} \pi_{i|B_2} \text{ and } \Pi(B_1,B_2)_{|B_2} \overset{\alpha_1}{\approx} \pi_{j|B_2}\right] \end{aligned}$$

By the definition of the labels π_i, we know that for some universal constant η it holds that $p_i \geq \eta \cdot \varepsilon^4$ for every π_i. We upper bound the fractions $p_{i,j}$: We know that for every $\pi_i \neq \pi_j$ it holds that

$\pi_i \overset{3\cdot\alpha_1}{\not\approx} \pi_j$. It follows that

$$\begin{aligned} p_{i,j} &\leq \Pr_{B_2}\left[\pi_{i|B_2} \overset{2\cdot\alpha_1}{\approx} \pi_{j|B_2}\right] \\ &\leq 1/\left(q^{d_1-2}\cdot\left(\alpha_1 - q^{-d_1}\right)^2\right) \\ &\leq \frac{1}{2}\cdot\eta^2\cdot\varepsilon^8 \end{aligned}$$

where the second inequality follows by Lemma 2.3 and the third inequality holds for sufficiently large choice of h'. Now, by the inclusion-exclusion principle that

$$\begin{aligned} \sum_i p_i - \sum_{i\neq j} p_{i,j} &\leq 1 \\ \left|L'(B_1)\right|\cdot\left(\eta\cdot\varepsilon^4\right) - \frac{1}{2}\left|L'(B_1)\right|^2\cdot\left(\frac{1}{2}\cdot\eta^2\cdot\varepsilon^8\right) &\leq 1 \end{aligned}$$

The last inequality immediately implies that $|L'(B_1)| \leq 2/\left(\eta\cdot\varepsilon^4\right) = O(1/\varepsilon^4)$.

We define $L(A_1)$ similarly, and prune it to $L'(A_1)$. Imagine now choosing a random $\pi_{A_1} \in L'(A)$ for each A_1 and a random $\pi_{B_1} \in L'(B_1)$ for each B_1. An edge (A_1, B_1) is called alive if it is labeled by a function π that is $3\alpha'$-close to both π_{A_1} and π_{B_1}. We expect at least $1/\left|L'(A)\right|\left|L'(B)\right| = \Omega(\varepsilon^8)$ fraction of edges to be alive. Fix a choice of π_{A_1} and π_{B_1} for each A_1 and B_1 in a way that attains this expectation.

Step 3. Let $\mathcal{D}_1$ be the distribution of choosing a random d_1-subspace B_1 and two neighbors A_1, A_1' of it in the graph. Let $\mathcal{D}_2$ be the distribution of choosing two d_0-spaces A_1, A_1' independently and a random B_1 that is a common neighbor of them in the graph. The statistical distance between $\mathcal{D}_1$ and $\mathcal{D}_2$ is small:

Claim 6.4. *For every $\kappa \in \mathbb{N}$, if the constant h' is sufficiently large then the distributions $\mathcal{D}_1$ and $\mathcal{D}_2$ are δ-close for $\delta < \varepsilon^{24}/\kappa$.*

We defer the proof of this claim to Section 6.1.2. Now choose a random triplet A_1, A_1', B_1 according to $\mathcal{D}_1$. We lower bound the probability that both edges (A_1, B_1) and (A_1', B_1) are alive. This certainly holds if (i) $\Omega(\varepsilon^8)$ fraction of the edges adjacent to B are alive, and (ii) both both edges (A_1, B_1) and (A_1', B_1) are alive. Part (i) holds with probability $\Omega(\varepsilon^8)$ and conditioned on this, Part (ii) holds with probability at least $\Omega(\varepsilon^{16})$. Altogether

$$\Pr_{(B_1,A_1,A_1')\sim\mathcal{D}_1}\left[(A_1,B_1),(A_1',B_1) \text{ are both alive}\right] = \Omega(\varepsilon^{24}).$$

Finally, if we let δ be the statistical distance of $\mathcal{D}_1$ and $\mathcal{D}_2$, and apply Claim 6.4 with sufficiently large choices of κ and h', then we have that

$$\Pr_{(B_1,A_1,A_1')\sim\mathcal{D}_2}\left[(A_1,B_1),(A_1',B_1) \text{ are both alive}\right] \geq \Omega(\varepsilon^{24}) - \delta = \Omega(\varepsilon^{24}).$$

Now fix A_1 such that the above holds when conditioning on A_1. This means that for at least $\Omega(\varepsilon^{24})$ fraction of the d_0-subspaces A_1' there exists a d_1-subspace B_1 such that both the edges (A_1, B_1) and (A_1', B_1) are alive. For each such A_1', it holds that the label of (A_1', B_1) is $3\alpha_1$-close to π_{B_1}, which in turn is $3\alpha_1$-close to the label of the edge (A_1, B_1), which is $3\alpha_1$-close to π_{A_1}. Thus, the label of

(A'_1, B_1) is is $9\alpha_1$-close to π_{A_1}. Let us denote by $\pi_{(A'_1,B_1)}$ the label of the edge (A'_1, B_1). Recall that by the definition of $\pi_{(A'_1,B_1)}$ it holds that

$$\Pr_{A_2,B_2}\left[\mathcal{P} \text{ and } \Pi\left(A'_1, A_2\right)_{|A_2} \overset{\alpha_1}{\approx} \pi_{(A'_1,B_1)|A_2}\right] \geq \Omega\left(\varepsilon^4\right) \tag{5}$$

Since $\pi_{(A'_1,B_1)} \overset{9\cdot\alpha_1}{\approx} \pi_A$ it holds by Lemma 2.3 that for a uniformly distributed d_0-subspace A_2:

$$\Pr_{A_2}\left[\pi_{(A'_1,B_1)|A_2} \overset{10\cdot\alpha_1}{\not\approx} \pi_{A_1|A_2}\right] \leq \frac{1}{q^{d_0-2}\cdot\left(\alpha_1 - q^{-d_0}\right)^2}$$

The latter expression can be made smaller than any constant times ε^4 by choosing h' to be sufficiently large. By substracting that expression from Inequality 5, we obtain that

$$\Pr_{A_2,B_2}\left[\mathcal{P} \text{ and } \Pi\left(A'_1, A_2\right)_{|A_2} \overset{\alpha_1}{\approx} \pi_{(A'_1,B_1)|A_2} \text{ and } \pi_{(A'_1,B_1)|A_2} \overset{10\cdot\alpha_1}{\approx} \pi_{A_1|A_2}\right] \geq \Omega\left(\varepsilon^4\right)$$

By letting $\pi_2 = \pi_{A_1}$ and choosing $c' = 28$, we have by the triangle inequality

$$\Pr_{A'_1,A_2}\left[\mathcal{P} \text{ and } \Pi\left(A'_1, A_2\right)_{|A_2} \overset{11\cdot\alpha_1}{\approx} \pi_{2|A_2}\right] \geq \Omega(\varepsilon^{24})\cdot\Omega\left(\varepsilon^4\right) = \Omega(\varepsilon^{c'}) \tag{6}$$

Step 4 It remains to show that the assignment Π agrees with both π_1 and π_2 on a non-negligible fraction of the B's. To this end, we observe that

$$\Pr\left[\mathcal{P} \text{ and } \Pi\left(A_1, A_2\right)_{|A_2} \overset{11\cdot\alpha_1}{\approx} \pi_{2|A_2} \middle| \Pi\left(B_1, B_2\right)_{|B_2} \overset{12\cdot\alpha_1}{\not\approx} \pi_{2|B_2}\right] \leq \frac{1}{q^{d_0-2}\cdot\left(\alpha_1/2\right)^2} \tag{7}$$

To see it, note that it suffices to prove that

$$\Pr\left[\Pi\left(B_1, B_2\right)_{|A_2} \overset{11\cdot\alpha_1}{\approx} \pi_{2|A_2} \middle| \Pi\left(B_1, B_2\right)_{|B_2} \overset{12\cdot\alpha_1}{\not\approx} \pi_{2|B_2}\right] \leq \frac{1}{q^{d_0-2}\cdot\left(\alpha_1 - q^{-d_0}\right)^2} \leq \frac{1}{q^{d_0-2}\cdot\left(\alpha_1/2\right)^2}$$

The latter inequality is an immediate corollary of Lemma 2.3.

Now, by choosing h' to be sufficiently large so that the upper bound in Inequality 7 is sufficiently smaller than $\varepsilon^{c'}$, and by combining Inequality 6 with Inequality 7, we obtain that

$$\Pr\left[\mathcal{P} \text{ and } \Pi\left(A_1, A_2\right)_{|A_2} \overset{11\cdot\alpha_1}{\approx} \pi_{2|A_2} \text{ and } \Pi\left(B_1, B_2\right)_{|B_2} \overset{12\cdot\alpha_1}{\approx} \pi_{2|B_2}\right] \geq \Omega(\varepsilon^{c'})$$

By setting h' such that $\alpha' \geq 12\cdot\alpha_1$ this concludes the proof of Lemma 6.2.

6.1.2 Proofs of Auxiliary Claim

Proof of Claim 6.4 Fix $\kappa \in \mathbb{N}$. In order to prove the claim, consider the event J which holds if and only if A and A' are disjoint. We argue that

$$\mathcal{D}_1 \overset{\delta/2}{\approx} \mathcal{D}_1|J = \mathcal{D}_2|J \overset{\delta/2}{\approx} D_2.$$

The fact that $\mathcal{D}_1|J = \mathcal{D}_2|J$ is exactly Proposition 2.4. We show that $\mathcal{D}_1 \overset{\delta/2}{\approx} \mathcal{D}_1|J$ and $\mathcal{D}_2 \overset{\delta/2}{\approx} \mathcal{D}_2|J$. The statistical distance between $\mathcal{D}_1$ and $\mathcal{D}_1|J$ (respectively, $\mathcal{D}_2$ and $\mathcal{D}_2|J$) is exactly the probability that the event J does not occur under $\mathcal{D}_1$ (respectively $\mathcal{D}_2$). It follows immediately from Proposition 2.13 that $\Pr_{\mathcal{D}_1}[\neg J] \leq 2\cdot d_0/q^{d_1-2\cdot d_0}$ and $\Pr_{\mathcal{D}_2}[\neg J] \leq 2\cdot d_0/q^{m-2\cdot d_0}$. Both the latter expressions can indeed be made smaller than ε^{24}/κ by choosing sufficiently large h', as required. ■

6.2 The proof of Theorem 5.2

In the rest of this section we prove Theorem 5.2, restated below.

Theorem (5.2, restated)**.** *There exists a universal constants $h, c \in \mathbb{N}$ such that the following holds: Let $\varepsilon \geq h \cdot d_0 \cdot q^{-d_0/h}$, $\alpha \stackrel{\text{def}}{=} h \cdot d_0 \cdot q^{-d_0/h}$. Assume that $d_1 \geq h \cdot d_0$, $m \geq h \cdot d_1$. Suppose that a (possible randomized) assignment Π passes the S-test with probability at least ε. There exists an assignment $\pi : \mathbb{F}^m \to \Sigma$ for which the following holds. Let B_1, B_2 be uniformly distributed and disjoint d_1-subspaces of $\mathbb{F}^m$, let A_1 and A_2 be uniformly distributed d_0-subspaces of B_1 and B_2 respectively, and denote $A = A_1 + A_2$. Then:*

$$\Pr\left[\Pi\left(B_1, B_2\right)_{|(A_1,A_2)} = \Pi\left(A\right)_{|(A_1,A_2)} \quad \text{and } \Pi\left(B_1, B_2\right) \stackrel{\alpha}{\approx} \pi_{|(B_1,B_2)}\right] = \Omega\left(\varepsilon^c\right)$$

Remark 6.5. Note that in the foregoing restatement of Theorem 5.2 we denote the first universal constant by h, while in its original statement it was denoted by h'.

The intuition that undelies the proof is the following. Consider an adversary the chooses the proof Π. Since the S-test essentially contains a P^2-test, the adversary must choose the assignment Π such that for random d_0-subspaces A_1 and A_2, the assignment $\Pi\left(A_1 + A_2\right)_{|(A_1,A_2)}$ is consistent with two assignments π_1, π_2 on A_1, A_2 respectively. On the other hand, given the sum $A_1 + A_2$, the adversary can not deduce the choices of A_1 and A_2, and therefore he must label both of A_1 and A_2 with the same assignment in order to make the S-test accept. We conclude that π_1 and π_2 must be essentially the same. Details follow.

In the rest of this section we prove Theorem 5.2. Let h' be the universal constant whose existence guaranteed in Theorem 6.1, and let α' be the corresponding value from Theorem 6.1. We choose c to be the same constant as in Theorem 6.1, and will choose the universal constant h to be at least h'.

Fix an assignment Π that passes the S-test with probability at least ε. We define a new assignment Π' that assigns values to pairs of d_0-subspaces and to pairs of d_1-subspaces of $\mathbb{F}^m$ (not necessarily disjoint) by choosing $\Pi'\left(B_1, B_2\right)$ (respectively $\Pi'\left(A_1, A_2\right)$) to be equal to $\Pi\left(B_1, B_2\right)$ (respectively $\Pi\left(A_1 + A_2\right)$) if B_1 and B_2 (respectively A_1 and A_2) are disjoint, and choosing Π' to be arbitrary otherwise. Observe that the assignment Π' passes the P^2-test whenever B_1 and B_2 are disjoint and Π passes the S-test. Furthermore, the probability that two uniformly distributed d_1-subspaces B_1 and B_2 of $\mathbb{F}^m$ are not disjoint is at most $d_1/q^{m-2\cdot d_1}$ by Proposition 2.13, and therefore Π' passes the P^2-test with probability at least $\varepsilon - d_1/q^{m-2\cdot d_1}$. For a sufficiently large choice of h, the latter probability is at least $\Omega\left(\varepsilon\right)$, and also matches the requirements of Theorem 6.1, so we can apply this theorem. It follows that there exist assignments $\pi_1, \pi_2 : \mathbb{F}^m \to \Sigma$ such that for uniformly distributed (not necessarily disjoint) B_1, B_2, $A_1 \subseteq B_1$, $A_2 \subseteq B_2$ it holds that

$$\begin{aligned} & \Pr[\Pi'\left(B_1, B_2\right)_{|(A_1,A_2)} = \Pi'\left(A_1, A_2\right) \\ & \quad \text{and } \Pi'\left(A_1, A_2\right) \stackrel{\alpha'}{\approx} \left(\pi_{1|A_1}, \pi_{2|A_2}\right) \\ & \quad \text{and } \Pi'\left(B_1, B_2\right) \stackrel{\alpha'}{\approx} \left(\pi_{1|B_1}, \pi_{2|B_2}\right)] \\ = \;& \Omega\left(\varepsilon^c\right) \end{aligned} \tag{8}$$

The probability that B_1 and B_2 are not disjoint is at most $d_1/q^{m-2\cdot d_1}$, and the latter expression can be made smaller than any constant factor times ε^c by choosing h to be sufficiently large. Thus, Inequality 8 also holds for uniformly distributed *disjoint* B_1 and B_2. We now argue that

Claim 6.6. *For sufficiently large choice of h, it holds that $\pi_1 \stackrel{5\cdot\alpha'}{\approx} \pi_2$.*

We defer the proof of Claim 6.6 to the end of this section. We turn to prove the theorem. By Inequality 8 it holds for uniformly distributed and disjoint d_1-subspaces B_1 and B_2 of $\mathbb{F}^m$ that

$$\Pr\left[\Pi'(B_1,B_2)_{|(A_1,A_2)} = \Pi'(A_1,A_2) \quad \text{and } \Pi(B_1,B_2) \overset{\alpha'}{\approx} \left(\pi_{1|B_1},\pi_{2|B_2}\right)\right] \geq \Omega\left(\varepsilon^c\right)$$

By Claim 6.6 it holds that $\pi_1 \overset{5\cdot\alpha'}{\approx} \pi_2$. Since B_2 is a uniformly distributed d_1-subspace of $\mathbb{F}^m$, this implies by Lemma 2.3 that

$$\Pr\left[\pi_{1|B_2} \overset{6\cdot\alpha'}{\approx} \pi_{2|B_2}\right] \geq 1 - \frac{1}{q^{d_1-2}\cdot\left(\alpha' - q^{-d_1}\right)^2} \geq 1 - \frac{1}{q^{d_1-2}\cdot\left(\alpha'/2\right)^2}$$

We conclude that

$$\begin{aligned}
& \Pr\left[\Pi'(B_1,B_2)_{|(A_1,A_2)} = \Pi'(A_1,A_2) \quad \text{and } \Pi(B_1,B_2) \overset{7\cdot\alpha'}{\approx} \left(\pi_{1|B_1},\pi_{1|B_2}\right)\right] \\
\geq\ & \Pr\left[\Pi'(B_1,B_2)_{|(A_1,A_2)} = \Pi'(A_1,A_2) \quad \text{and } \Pi(B_1,B_2) \overset{\alpha'}{\approx} \left(\pi_{1|B_1},\pi_{2|B_2}\right) \quad \text{and } \pi_{1|B_2} \overset{6\cdot\alpha'}{\approx} \pi_{2|B_2}\right] \\
=\ & \Omega\left(\varepsilon^c\right) - \frac{1}{q^{d_1-2}\cdot\left(\alpha'/2\right)^2} \\
=\ & \Omega\left(\varepsilon^c\right)
\end{aligned}$$

where the last equality holds for sufficiently large choice of h. the theorem now follows by defining $\pi = \pi_1$ and setting h to be sufficiently large such that $\alpha = 7\cdot\alpha'$.

Proof of Claim 6.6 For the sake of contradition, assume that $\pi_1 \overset{5\cdot\alpha'}{\not\approx} \pi_2$. Let A be a uniformly distributed $2\cdot d_0$-subspace A of $\mathbb{F}^m$ and let A_1 and A_2 be uniformly disributed and disjoint d_0-subspaces of A. By Lemma 2.3, it holds that

$$\Pr\left[\pi_{1|A} \overset{4\cdot\alpha'}{\not\approx} \pi_{2|A}\right] \geq 1 - \frac{1}{q^{2\cdot d_0-2}\cdot\left(\alpha' - q^{-2d_0}\right)^2} \geq 1 - \frac{1}{q^{2\cdot d_0-2}\cdot\left(\alpha'/2\right)^2}$$

If $\pi_{1|A} \overset{4\cdot\alpha'}{\not\approx} \pi_{2|A}$ then by the triangle inequality it either holds that $\Pi(A) \overset{2\cdot\alpha'}{\not\approx} \pi_{1|A}$ or that $\Pi(A) \overset{2\cdot\alpha'}{\not\approx} \pi_{2|A}$. Since A_1 is a uniformly distributed d_0-subspace of A, it holds by Lemma 2.3 that

$$\Pr\left[\Pi(A)_{|A_1} \overset{\alpha'}{\not\approx} \pi_{1|A_1} \,\middle|\, \Pi(A) \overset{2\cdot\alpha'}{\not\approx} \pi_{1|A}\right] \geq 1 - \frac{1}{q^{2\cdot d_0-2}\cdot\left(\alpha'/2\right)^2}$$

A similar claim can be made for π_2 and A_2. Now, if either $\Pi(A)_{|A_1} \overset{\alpha'}{\not\approx} \pi_{1|A_1}$ or $\Pi(A)_{|A_2} \overset{\alpha'}{\not\approx} \pi_{2|A_2}$ then by definition it holds that $\Pi(A)_{|(A_1,A_2)} \overset{\alpha'}{\not\approx} \left(\pi_{1|A_1},\pi_{2|A_2}\right)$. We conclude that

$$\Pr\left[\Pi(A)_{|(A_1,A_2)} \overset{\alpha'}{\not\approx} \left(\pi_{1|A_1},\pi_{2|A_2}\right) \,\middle|\, \pi_{1|A} \overset{4\cdot\alpha'}{\not\approx} \pi_{2|A}\right] \geq 1 - \frac{1}{q^{2\cdot d_0-2}\cdot\left(\alpha'/2\right)^2}$$

and therefore by lifting the conditioning and substituing $A = A_1 + A_2$ we obtain that for a uniformly distributed and disjoint d_0-subspaces A_1 and A_2 of $\mathbb{F}^m$ it holds that

$$\Pr\left[\Pi(A_1+A_2)_{|(A_1,A_2)} \overset{\alpha'}{\approx} \left(\pi_{1|A_1},\pi_{2|A_2}\right)\right] \leq \frac{2}{q^{2\cdot d_0-2}\cdot\left(\alpha'/2\right)^2}$$

On the other hand, by the definition of Π', Inequality 8 implies that for uniformly distributed and disjoint d_0-subspaces A_1 and A_2 of $\mathbb{F}^m$ it holds that

$$\Pr\left[\Pi\left(A_1+A_2\right)_{|(A_1,A_2)} \overset{\alpha'}{\approx} \left(\pi_{1|A_1},\pi_{2|A_2}\right)\right] \geq \Omega\left(\varepsilon^c\right)$$

By choosing h to be sufficiently large, the latter lower bound can be made larger than $2/\left(q^{2\cdot d_0-2}\cdot\left(\alpha'\right)^2\right)$, and this is a contradiction. ■

Acknowledgement. We would like to thank Eli Ben Sasson for a useful discussion.

References

[AL96] Sanjeev Arora and Carsten Lund. *Hardness of Approximations.* PW Publishing, 1996.

[ALM+98] Sanjeev Arora, Carsten Lund, Rajeev Motwani, Madhu Sudan, and Mario Szegedy. Proof verification and intractability of approximation problems. *Journal of ACM*, 45(3):501–555, 1998. Preliminary version in FOCS 1992.

[AS98] Sanjeev Arora and Shmuel Safra. Probabilistic checkable proofs: A new characterization of NP. *Journal of ACM volume*, 45(1):70–122, 1998. Preliminary version in FOCS 1992.

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

[BFLS91] László Babai, Lance Fortnow, Leonid A. Levin, and Mario Szegedy. Checking computations in polylogarithmic time. In *STOC*, pages 21–31, 1991.

[BGLR93] Mihir Bellare, Shafi Goldwasser, Carsten Lund, and Alexander Russell. Efficient probabilistically checkable proofs and applications to approximations. In *STOC*, pages 294–304, 1993.

[Cam98] Peter J. Cameron. *Combinatorics: Topics, Techniques, Algorithms.* Cambridge University Press, Cambridge CB2 2RU, MA, USA, 1998.

[DFK+99] Irit Dinur, Eldar Fischer, Guy Kindler, Ran Raz, and Shmuel Safra. PCP characterizations of NP: Towards a polynomially-small error-probability. In *STOC*, pages 29–40, 1999.

[DG08] Irit Dinur and Elazar Goldenberg. Locally testing direct product in the low error range. In *FOCS*, pages 613–622, 2008.

[DH09] Irit Dinur and Praladh Harsha. Composition of low-error 2-query PCPs using decodable PCPs. In *FOCS*, 2009.

[Din07] Irit Dinur. The PCP Theorem by gap amplification. *Journal of ACM*, 54(3):241–250, 2007. Preliminary version in STOC 2006.

[DR06] Irit Dinur and Omer Reingold. Assignment testers: Towards combinatorial proof of the PCP theorem. *SIAM Journal of Computing*, 36(4):155–164, 2006.

[FGL+96] Uriel Feige, Shafi Goldwasser, László Lovász, Shmuel Safra, and Mario Szegedy. Interactive proofs and the hardness of approximating cliques. *J. ACM*, 43(2):268–292, 1996.

[FK95] Uriel Feige and Joe Kilian. Impossibility results for recycling random bits in two-prover proof systems. In *STOC*, pages 457–468, 1995.

[GS00] Oded Goldreich and Shmuel Safra. A combinatorial consistency lemma with application to proving the PCP theorem. *SIAM J. Comput.*, 29(4):1132–1154, 2000.

[IJKW08] Russell Impagliazzo, Ragesh Jaiswal, Valentine Kabanets, and Avi Wigderson. Uniform direct product theorems: simplified, optimized, and derandomized. In *STOC*, pages 579–588, 2008.

[IKW09] Russell Impagliazzo, Valentine Kabanets, and Avi Wigderson. New direct-product testers and 2-query PCPs. In *STOC*, pages 131–140, 2009.

[Kho06] Subhash Khot. Ruling out ptas for graph min-bisection, dense k-subgraph, and bipartite clique. *SIAM J. Comput.*, 36(4):1025–1071, 2006.

[Lei92] F. Thomson Leighton. *Introduction to parallel algorithms and architectures: array, trees, hypercubes*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1992.

[LPS88] Alexander Lubotzky, R. Phillips, and P. Sarnak. Ramanujan graphs. *Combinatorica*, 8(3):261–277, 1988.

[Mei09] Or Meir. Combinatorial PCPs with efficient verifiers. In *FOCS*, 2009.

[MR08] Dana Moshkovitz and Ran Raz. Two query PCP with sub-constant error. In *FOCS*, 2008. Full version is available as ECCC TR08-071.

[PS94] Alexander Polishchuk and Daniel A. Spielman. Nearly-linear size holographic proofs. In *STOC*, pages 194–203, 1994.

[PY91] Christos H. Papadimitriou and Mihalis Yannakakis. Optimization, approximation, and complexity classes. *J. Comput. Syst. Sci.*, 43(3):425–440, 1991.

[Raz98] Ran Raz. A parallel repetition theorem. *SIAM J. Comput.*, 27(3):763–803, 1998.

[RS97] Ran Raz and Shmuel Safra. A sub-constant error-probability low-degree test, and a sub-constant error-probability PCP characterization of NP. In *STOC*, pages 475–484, 1997.

[Spi95] Daniel A. Spielman. *Computationally efficient error-correcting codes and holographic proofs*. PhD thesis, MIT, 1995.

A Proof of Theorem 2.1

In this section we prove Theorem 2.1, restated below. Let $\mathbb{F}$ be a finite field of size q, let $m, d_0, d_1 \in \mathbb{N}$, and consider a (possible randomized) assignment Π that assigns values to d_0- and d_1-subspaces of $\mathbb{F}^m$.

Theorem A.1 (2.1, restated)**.** *There exists a universal constant $h \in \mathbb{N}$ such that the following holds: Let $\varepsilon \geq h \cdot d_0 \cdot q^{-d_0/h}$, $\alpha \stackrel{\text{def}}{=} h \cdot d_0 \cdot q^{-d_0/h}$. Assume that $d_1 \geq h \cdot d_0$, $m \geq h \cdot d_1$. Suppose that an assignment Π passes the P-test with probability at least ε. Then, there exists an assignment π such that*

$$\Pr\left[\Pi(B)_{|A} = \Pi(A) \quad \text{and } \Pi(B) \stackrel{\alpha}{\approx} \pi_{|B} \quad \text{and } \Pi(A) \stackrel{\alpha}{\approx} \pi_{|A}\right] = \Omega(\varepsilon^4)$$

where the probability is over A, B chosen as in the P-test.

We begin by recalling the required preliminaries from [IKW09], and then turn to prove Theorem 2.1.

Definition A.2 (Good)**.** Let A be a d_0-subspace of $\mathbb{F}^m$ and let $\varepsilon \in (0,1)$. We say that A is *ε-good* (with respect to an assignment Π) if for a uniformly distributed d_1-dimensional subspace B that contains A it holds that

$$\Pr\left[\Pi(B)_{|A} = \Pi(A)\right] \geq \varepsilon$$

where the randomness is over the choice of B and over the randomness of Π.

Definition A.3 (Plurality function)**.** Let A be a d_0-subspace of $\mathbb{F}^m$. We denote by $\pi_A : \mathbb{F}^m \to \Sigma$ the *plurality function* of A (with respect to Π). In other words, for every $x \in \mathbb{F}^m$ we define $\pi_A(x)$ to be the value $v \in \Sigma$ that maximizes

$$\Pr_{B \supseteq A}\left[\Pi(B)_{|x} = v \,\middle|\, \Pi(B)_{|A} = \Pi(A)\right]$$

where B is a uniformly distributed d_1-dimensional subspace that contains A.

Definition A.4 (DP-consistent)**.** Let A be a d_0-subspace of $\mathbb{F}^m$ and let $\alpha, \gamma \in (0,1)$. We say that A is *$(\varepsilon,\alpha,\gamma)$-direct product consistent* (abbreviated *$(\varepsilon,\alpha,\gamma)$ -DP-consistent*) if A is ε-good and it holds that

$$\Pr_{B \supseteq A}\left[\Pi(B) \overset{\alpha}{\approx} \pi_{A|B} \,\middle|\, \Pi(B)_{|A} = \Pi(A)\right] \geq 1 - \gamma$$

The following lemma is a direct corollary of the proofs of [IKW09, Lemma 4.2] and [IKW09, Lemma 4.4].

Lemma A.5. *There exists a universal constant $h_0 \in \mathbb{N}$ such that the following holds: Let $\varepsilon \geq h_0 \cdot q^{-(d_1/h_0 - d_0)}$ and $\alpha, \gamma \in (0,1)$. The probability that a uniformly distributed A is ε-good but not $(\varepsilon,\alpha,\gamma)$-DP-consistent is at most $O\big(1/\left(\alpha \cdot \gamma \cdot \varepsilon^2 \cdot q^{d_0-2}\right)\big)$.*

Proof of Theorem 2.1

We will choose the universal constant h to be larger than h_0 (where h_0 is the constant from Lemma A.5). Assume that the P-test accepts with probability at least ε as in the statement of the theorem. Let $\varepsilon_1 = \frac{1}{3} \cdot \varepsilon$ and $\gamma_1 = \varepsilon_1^3/h$. Choose $\alpha_1 = O\left(1/\varepsilon_1^3 \cdot \gamma_1 \cdot q^{d_0-2}\right)$ such that the probability in Lemma A.5 that A is ε_1-good but not $(\varepsilon_1,\alpha_1,\gamma_1)$-DP-consistent is at most ε_1, which is indeed possible for sufficiently large choice of h. We will later choose $\alpha = O(\alpha_1)$, by choosing again h to be sufficiently large.

We consider the following sequence of events. Let A_1, A_2 denote random d_0-subspaces, and let B denote a random d_1-subspace, and define events $\mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3$ as follows:

1. $\mathcal{S}_1(A_1, A_2, B)$: A_1 and A_2 are $(\varepsilon_1,\alpha_1,\gamma_1)$-DP-consistent and $\Pi(B)_{|A_1} = \Pi(A_1)$, $\Pi(B)_{|A_2} = \Pi(A_2)$.

2. $\mathcal{S}_2(A_1, A_2, B)$: The event $\mathcal{S}_1(A_1, A_2, B)$ occurs and $\pi_{A_1|B} \overset{2\alpha_1}{\approx} \pi_{A_2|B}$ (recall that π_{A_1} and π_{A_2} are the plurality assignments of A_1 and A_2 respectively).

3. $\mathcal{S}_3(A_1, A_2)$: A_1 and A_2 are $(\varepsilon_1,\alpha_1,\gamma_1)$-DP-consistent and $\pi_{A_1} \overset{3\alpha_1}{\approx} \pi_{A_2}$.

In the next three claims we choose A_1, A_2 and B according to the following distribution: choose A_1 and A_2 to be uniformly distributed and disjoint d_0-spaces A_1, A_2, and a choose B to be a uniformly distributed d_1-subspace that contains them. We show that the probability of events $\mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3$ under this distribution is non-negligible.

Claim A.6. $Pr[\mathcal{S}_1] \geq \Omega\left(\varepsilon_1^3\right)$.

Proof Let B' be a uniformly distributed d_1-subspace of $\mathbb{F}^m$ and let A' be a d_0-uniformly distributed subspace of B'. We begin by lower bounding the probability

$$\Pr\left[\Pi\left(B'\right)_{|A'} = \Pi\left(A'\right) \quad \text{and } A' \text{ is } (\varepsilon_1, \alpha_1, \gamma_1)\text{-DP-consistent}\right] \tag{9}$$

To this end, let us denote by $\mathcal{P}$ the event that $\Pi\left(B'\right)_{|A'} = \Pi\left(A'\right)$, by $\mathcal{D}$ the event that A' is $(\varepsilon_1, \alpha_1, \gamma_1)$-DP-consistent, and by $\mathcal{G}$ the event that A' is ε_1-good. Observe that$\Pr\left[\mathcal{P} \text{ and } \neg\mathcal{G}\right] \leq \Pr\left[\mathcal{P} | \neg\mathcal{G}\right] \leq \varepsilon_1$. Furthermore, A' is a uniformly distributed d_0-subspace of $\mathbb{F}^m$ and thus by Lemma A.5 and our choice of α_1, it holds that $\Pr\left[\mathcal{G} \text{ and } \neg\mathcal{D}\right] \leq \varepsilon_1$. Finally, it holds that the probability in (9) is

$$\begin{aligned}
\Pr\left[\mathcal{P} \text{ and } \mathcal{D}\right] &\geq \Pr\left[\mathcal{P} \text{ and } \mathcal{G} \text{ and } \mathcal{D}\right] \\
&= \Pr\left[\mathcal{P} \text{ and } \mathcal{G}\right] - \Pr\left[\mathcal{P} \text{ and } \mathcal{G} \text{ and } \neg\mathcal{D}\right] \\
&= \Pr\left[\mathcal{P}\right] - \Pr\left[\mathcal{P} \text{ and } \neg\mathcal{G}\right] - \Pr\left[\mathcal{P} \text{ and } \mathcal{G} \text{ and } \neg\mathcal{D}\right] \\
&\geq \Pr\left[\mathcal{P}\right] - \Pr\left[\mathcal{P} \text{ and } \neg\mathcal{G}\right] - \Pr\left[\mathcal{G} \text{ and } \neg\mathcal{D}\right] \\
&\geq \varepsilon - \varepsilon_1 - \varepsilon_1 \\
&\geq \varepsilon_1
\end{aligned}$$

So the probability in (9) is at least ε_1. By averaging, this implies that for $\Omega\left(\varepsilon_1\right)$ fraction of the d_1-subspaces B' it holds that at least $\Omega\left(\varepsilon_1\right)$ fraction of the d_0-subspaces A' of B' are $(\varepsilon_1, \alpha_1, \gamma_1)$-DP-consistent and satisfy $\Pi\left(B'\right)_{|A'} = \Pi\left(A'\right)$.

Now, observe that by Proposition 2.4, the distribution over A_1, A_2, B is equivalent to choosing B to be a uniformly distributed d_1-subspace of $\mathbb{F}^m$ and then choosing A_1 and A_2 to be disjoint uniformly distributed d_0-subspaces of B. With probability at least $\Omega\left(\varepsilon_1\right)$ it holds for B that at least $\Omega\left(\varepsilon_1\right)$ fraction of the d_0-subspaces A of B are $(\varepsilon_1, \alpha_1, \gamma_1)$-DP-consistent and satisfy $\Pi\left(B\right)_{|A} = \Pi\left(A\right)$. We condition on the latter event, and claim that under this conditioning the event $\mathcal{S}_1(A_1, A_2, B)$ occurs with probability at least $\Omega\left(\varepsilon_1^2\right)$. To see it, consider two uniformly distributed (*not necessarily disjoint*) d_0-subspaces A_1' and A_2' of B. Then, by our conditioning, it holds that $\mathcal{S}_1(A_1', A_2', B)$ occurs with probability at least $\Omega\left(\varepsilon_1^2\right)$. Furthermore, by Proposition 2.13 it holds with probability at least $1 - 2 \cdot d_0/q^{d_1 - 2 \cdot d_0}$ that A_1' and A_2' are disjoint. It therefore follows under the foregoing conditioning on B that

$$\begin{aligned}
\Pr\left[\mathcal{S}_1(A_1, A_2, B)\right] &= \Pr\left[\mathcal{S}_1(A_1', A_2', B) \middle| A_1', A_2' \text{ are disjoint}\right] \\
&\geq \Pr\left[\mathcal{S}_1(A_1', A_2', B) \text{ and } A_1', A_2' \text{ are disjoint}\right] \\
&\geq \Pr\left[\mathcal{S}_1(A_1', A_2', B)\right] - \Pr\left[A_1', A_2' \text{ are disjoint}\right] \\
&\geq \Omega\left(\varepsilon_1^2\right) - 2 \cdot d_0/q^{d_1 - 2 \cdot d_0} \\
&\geq \Omega\left(\varepsilon_1^2\right)
\end{aligned}$$

where the last inequlaity holds for sufficiently large h. Lifting the conditioning on B, we get that for a uniformly distributed d_1-subspace B of $\mathbb{F}^m$ and two disjoint uniformly distributed d_0-subspaces A_1 and A_2 of B, it holds with probability at least $\Omega\left(\varepsilon_1^3\right)$ that both A_1 and A_2 are $(\varepsilon_1, \alpha_1, \gamma_1)$-DP-consistent and that $\Pi\left(B\right)_{|A_1} = \Pi\left(A_1\right)$, $\Pi\left(B\right)_{|A_2} = \Pi\left(A_2\right)$, as required. ■

Claim A.7. $\Pr[\mathcal{S}_2] \geq \Omega\left(\varepsilon_1^3\right)$.

Proof Let $\mathcal{E}_1$ be the event in which A_1 is $(\varepsilon_1, \alpha_1, \gamma_1)$-DP-consistent, $\Pi(B)_{|A_1} = \Pi(A_1)$ and $\Pi(B) \overset{\alpha_1}{\not\approx} \pi_{A_1|B}$, and let $\mathcal{E}_2$ be the corresponding event for A_2. We begin by noting that the probabilities of both $\mathcal{E}_1$ and $\mathcal{E}_2$ are upper bounded by γ_1. To see it for $\mathcal{E}_1$, note that conditioned on A_1 being $(\varepsilon_1, \alpha_1, \gamma_1)$-DP-consistent and on $\Pi(B)_{|A_1} = \Pi(A_1)$ it holds that B is a uniformly distributed d_1-subspace satisfying $\Pi(B)_{|A_1} = \Pi(A_1)$, and therefore it holds that $\Pi(B) \overset{\alpha_1}{\not\approx} \pi_{A_1|B}$ with probability at most γ_1 (by the DP-consistency of A_1). The probability of $\mathcal{E}_2$ can be upper bounded similarly.

It now follows by Claim A.6 that

$$\begin{aligned}
\Pr[\mathcal{S}_2] &= \Pr\left[\mathcal{S}_1 \text{ and } \pi_{A_1|B} \overset{2\alpha_1}{\approx} \pi_{A_2|B}\right] \\
&\geq \Pr[\mathcal{S}_1 \text{ and } \neg\mathcal{E}_1 \text{ and } \neg\mathcal{E}_2] \\
&\geq \Pr[\mathcal{S}_1] - \Pr[\mathcal{E}_1] - \Pr[\mathcal{E}_2] \\
&\geq \Omega\left(\varepsilon_1^3\right) - 2 \cdot \gamma_1 \\
&\geq \Omega\left(\varepsilon_1^3\right)
\end{aligned}$$

where the last inequality holds for sufficiently large choice of h. The required result follows. ■

Claim A.8. $\Pr[\mathcal{S}_3] \geq \Omega\left(\varepsilon_1^3\right)$.

Proof Let us say that A_1 and A_2 are "agree on a random B" if both A_1 and A_2 are $(\varepsilon_1, \alpha_1, \gamma_1)$-DP-consistent and $\Pr_{B \supset A_1, A_2}\left[\pi_{A_1|B} \overset{2\cdot\alpha_1}{\approx} \pi_{A_2|B}\right] \geq \Omega\left(\varepsilon_1^3\right)$. By Claim A.7 and by averaging, we know that with probability at least $\Omega\left(\varepsilon_1^3\right)$ it holds that A_1 and A_2 agree on a random B. We show that for every A_1 and A_2 that are $(\varepsilon_1, \alpha_1, \gamma_1)$-DP-consistent such that $\pi_{A_1} \overset{3\cdot\alpha_1}{\not\approx} \pi_{A_2}$ it holds that A_1 and A_2 do not agree on a random B. This will imply that if A_1 and A_2 agree on a random B then it must hold that $\pi_{A_1} \overset{3\cdot\alpha_1}{\approx} \pi_{A_2}$. Since we know that the probability of A_1 and A_2 to agree on a random B is at least $\Omega\left(\varepsilon_1^3\right)$ the required result will follow.

Fix A_1 and A_2 to be any $(\varepsilon_1, \alpha_1, \gamma_1)$-DP-consistent disjoint d_0-subspaces such that $\pi_{A_1} \overset{3\cdot\alpha_1}{\not\approx} \pi_{A_2}$. Now, by Lemma 2.3 and by sufficiently large choice of h, the probability that a uniformly distributed d_1-subspace B that contains A_1 and A_2 contains at most $2\cdot\alpha_1 \leq 3\cdot\alpha_1 - 1/q^{d_0-2} - 1/q^{d_1-2\cdot d_0}$ fraction of coordinates on which π_{A_1} and π_{A_2} disagree is at most $1/\left(q^{d_1-4\cdot d_0-6}\right)$, and the latter expression can be made smaller than any constant factor times ε_1^3. Thus, it holds that $\Pr_{B \supset A_1, A_2}\left[\pi_{A_1|B} \overset{2\cdot\alpha_1}{\approx} \pi_{A_2|B}\right]$ can be made sufficiently small such that A_1 and A_2 do not agree on a random B, as required. ■

We now find a global assignment π and show that it agrees with Π on many B's, and then on many A's.

Claim A.9. *There exists an assignment* $\pi : \mathbb{F}^m \to \Sigma$ *such that* $\Pr_B[\Pi(B) \overset{5\cdot\alpha_1}{\approx} \pi_{|B} \text{ and } \Pi(B)_{|A} = \Pi(A)] \geq \Omega\left(\varepsilon_1^4\right)$.

Proof By Claim A.8 and by averaging, we get that for at least $\Omega\left(\varepsilon_1^3\right)$ fraction of the d_0-subspaces A_1 it holds that A_1 is $(\varepsilon_1, \alpha_1, \gamma_1)$-DP-consistent and

$$\Pr_{A_2 : A_2 \text{ is disjoint from } A_1}\left[A_2 \text{ is } (\varepsilon_1, \alpha_1, \gamma_1)\text{-DP-consistent and } \pi_{A_1} \overset{3\cdot\alpha_1}{\approx} \pi_{A_2}\right] \geq \Omega\left(\varepsilon_1^3\right)$$

Fix such d_0-subspace A_1, and set $\pi = \pi_{A_1}$. Consider choosing a uniformly distributed d_0-space A_2 and a uniformly distributed d_1-space $B \supset A_2$. We show that $\Pi(B) \overset{5\cdot\alpha_1}{\approx} \pi_{|B}$ with probability at least $\Omega\left(\varepsilon_1^4\right)$.

Let us denote by $\mathcal{D}$ the event in which A_2 is disjoint from A_1, by $\mathcal{P}$ the event in which $\Pi\left(B\right)_{|A_2} = \Pi\left(A_2\right)$, and by $\mathcal{C}$ the event in which A_2 is $(\varepsilon_1, \alpha_1, \gamma_1)$-DP-consistent and $\pi_{A_1} \overset{3\cdot\alpha_1}{\approx} \pi_{A_2}$.

By Proposition 2.13, it holds that $\Pr\left[\mathcal{D}\right] \geq 1 - 2 \cdot d_0/q^{m-2\cdot d_0} \geq \frac{1}{2}$ (where the second inequality holds for sufficiently large h). Furthermore, conditioned on $\mathcal{D}$, the subspace A_2 is a uniformly distributed d_0-subspace of $\mathbb{F}^m$ *that is disjoint from* A_1, and thus by the choice of A_1 it holds that $\Pr\left[\mathcal{C}|\mathcal{D}\right] \geq \Omega\left(\varepsilon_1^3\right)$. Lifting the conditioning, it follows that $\Pr\left[\mathcal{C}\right] \geq \Omega\left(\varepsilon_1^3\right)$. Next, observe that B is distributed uniformly over the d_1-subspaces that contain A_2, and thus (since in particular A_2 is ε_1-good) $\Pr\left[\mathcal{P}|\mathcal{C}\right] \geq \varepsilon_1$. It therefore holds that $\Pr\left[\mathcal{C} \text{ and } \mathcal{P}\right] \geq \Omega\left(\varepsilon_1^4\right)$

Now, let us condition on the events $\mathcal{C}$ and $\mathcal{P}$. By Lemma 2.3 and for sufficiently large h, it holds with probability at least $1 - 1/\left(q^{d_1-3\cdot d_0-6}\right) \geq \frac{3}{4}$ that B contains at most $4 \cdot \alpha_1 \geq 3\alpha_1 + 1/q^{d_0-2} + 1/q^{d_1-2\cdot d_0}$ fraction of coordinates on which π_{A_1} and π_{A_2} disagree. Furthermore, by the DP-consistency of A_2 and for sufficiently large choice of h, it holds with probability at least $1-\gamma_1 \geq \frac{3}{4}$ that $\Pi\left(B\right) \overset{\alpha_1}{\approx} \pi_{A_2|B}$. By the union bound and the triangle inequality, it follows that with probability at least $\frac{1}{2}$ it holds that $\Pi\left(B\right)$ disagrees with $\pi_{A_1|B}$ on at most $5 \cdot \alpha_1$ fraction of the coordinates. Lifting the conditioning on $\mathcal{C}$ and $\mathcal{P}$, we obtain that with probability at least $\Omega\left(\varepsilon_1^4\right)$ it holds that $\Pi\left(B\right) \overset{5\cdot\alpha_1}{\approx} \pi_{A_1|B}$, and $\Pi\left(B\right) = \Pi\left(A\right)$ as required. ■

Finally, we turn to prove the theorem. Let π be the assignment whose existence is guaranteed by the previous claim. Let us denote by $\mathcal{P}$ the event in which $\Pi\left(B\right)_{|A} = \Pi\left(A\right)$ (i.e., the P-test accepts A and B), by $\mathcal{E}_1$ the event in which $\Pi(B) \overset{5\cdot\alpha_1}{\approx} \pi_{|B}$, by $\mathcal{E}_2$ the event in which $\Pi(A) \overset{6\cdot\alpha_1}{\approx} \pi_{|A}$, and by $\mathcal{E}_3$ the event in which $\Pi\left(B\right)_{|A} \overset{6\cdot\alpha_1}{\approx} \pi_{|A}$. Using this notation, it suffices to prove that

$$\Pr\left[\mathcal{P} \text{ and } \mathcal{E}_1 \text{ and } \mathcal{E}_2\right] = \Omega\left(\varepsilon_1^4\right)$$

By the definition of π, it holds that

$$\Pr\left[\mathcal{P} \text{ and } \mathcal{E}_1\right] = \Omega\left(\varepsilon_1^4\right)$$

The subspace A is a uniformly distributed d_0-subspace of B, and therefore it holds by Lemma 2.3 that

$$\Pr\left[\neg\mathcal{E}_3 \,|\mathcal{E}_1\right] = O\left(1/q^{d_0/2-2}\right)$$

This implies that

$$\begin{aligned}
\Pr\left[\mathcal{P} \text{ and } \mathcal{E}_1 \text{ and } \mathcal{E}_3\right] &= \Pr\left[\mathcal{P} \text{ and } \mathcal{E}_1\right] - \Pr\left[\mathcal{P} \text{ and } \mathcal{E}_1 \text{ and } \neg\mathcal{E}_3\right] \\
&\geq \Pr\left[\mathcal{P} \text{ and } \mathcal{E}_1\right] - \Pr\left[\neg\mathcal{E}_3|\mathcal{E}_1\right] \\
&= \Omega\left(\varepsilon_1^4\right) - O\left(1/q^{d_0/2-2}\right) \\
&= \Omega\left(\varepsilon_1^4\right)
\end{aligned}$$

where the last inequality holds for sufficiently large h. Now, observe that whenever both the events $\mathcal{P}$ and $\mathcal{E}_3$ occur, the event $\mathcal{E}_2$ also occurs. It follows that

$$\Pr\left[\mathcal{P} \text{ and } \mathcal{E}_1 \text{ and } \mathcal{E}_2\right] \geq \Pr\left[\mathcal{P} \text{ and } \mathcal{E}_1 \text{ and } \mathcal{E}_3\right] = \Omega\left(\varepsilon_1^4\right)$$

as required.

B Routing on de Bruijn graphs

In this section we prove the routing property of de Bruijn graph given in Fact 4.5. Recall the following.

Definition (4.1, restated)**.** Let Λ be a finite alphabet and let $m \in \mathbb{N}$. The *de Bruijn graph* $\mathcal{DB}_{\Lambda,m}$ is the directed graph whose vertices set is Λ^m such that each vertex $(\alpha_1, \ldots, \alpha_t) \in \Lambda^m$ has outgoing edges to all the vertices of the form $(\alpha_2, \ldots, \alpha_t, \beta)$ for every $\beta \in \Lambda$.

Fact (4.5, restated)**.** *Let $\mathcal{DB}_{\Lambda,m}$ be a de-Brujin graph. Then, given a permutation μ on the vertices of $\mathcal{DB}_{\Lambda,m}$ one can find a set of undirected paths of length $2m$ that connect each vertex v to $\mu(v)$ and that have the following property: For every $j \in [2m]$, each vertex v is the j-th vertex of exactly one path. Furthermore, finding the paths can be done in time that is polynomial in the size of $\mathcal{DB}_{\Lambda,m}$.*

We actually prove the following slightly stronger result.

Claim B.1. *Let $\mathcal{DB}_{\Lambda,m}$ be a de-Brujin graph and let $i \in [m]$. Then, given a permutation μ on Λ^i one can find a set of undirected paths of length $2 \cdot i$ that connect each vertex $(\alpha_1, \ldots, \alpha_m)$ of $\mathcal{DB}_{\Lambda,m}$ to the vertex $(\alpha_1, \ldots, \alpha_{m-i}, \mu(\alpha_{m-i+1}, \ldots, \alpha_m))$ and that have the following two properties: For every $j \in [l]$, each vertex v is the j-th vertex of exactly one path. Furthermore, finding the paths can be done in time that is polynomial in the size of $\mathcal{DB}_{\Lambda,m}$.*

The proof works by induction on i. For $i = 0$ the claim is obvious. Assume that the claim holds for some $0 \leq i < m$. We prove that the claim holds for $i + 1$. Let $\mathcal{DB} = \mathcal{DB}_{\Lambda,m}$, amd let μ be a permutation on Λ^{i+1}. For convenience, let us define the action of μ on each $(\alpha_1, \ldots, \alpha_m) \in \mathbb{F}^m$ as $\mu(\alpha_1, \ldots, \alpha_m) = (\alpha_1, \ldots, \alpha_{m-i-1}, \mu(\alpha_{m-i}, \ldots, \alpha_m))$. Let G be the directed graph whose vertices are the set Λ^m and whose edges are all the pairs of the form $(v, \mu(v))$. Let G' be the graph that is obtained from G by contracting each $|\Lambda|$ vertices of G that agree on their last coordinate to one vertex. Clearly, every vertex in G' has in-degree and out-degree exactly $|\Lambda|$, and each edge of G' corresponds to an edge of G. Furthermore, observe that the vertices of G' can be identified with the vertices of Λ^{m-1}.

The $|\Lambda|$-regularity of G implies that the edges of G' can be partitioned to $|\Lambda|$ perfect matchings $\{G'_\sigma\}_{\sigma \in \Lambda}$ in polynomial time (see, e.g., [Cam98, Proposition 18.1.2]). Fix a matching G'_σ, and consider an edge e' in G'_σ. Observe that if e is coming out of a vertex $(\alpha_1, \ldots, \alpha_{m-1})$ of G', then it must enter a vertex of the form $\left(\alpha_1, \ldots, \alpha_{m-i}, \alpha'_{m-i+1}, \ldots, \alpha'_{m-1}\right)$. Thus, we can define a permutation ν_σ on Λ^i that maps $(\alpha_{m-i}, \ldots, \alpha_{m-1})$ to $\left(\alpha'_{m-i}, \ldots, \alpha'_{m-1}\right)$ for each such edge e' - since G'_σ is a matching, this is well defined. We invoke the induction hypothesis on the graph $\mathcal{DB} = \mathcal{DB}_{\Lambda,m}$ to find a set of paths $\mathcal{P}_\sigma$ of length $2 \cdot i$ for each permutation ν_σ.

We now construct the required paths for μ as follows. Let $v = (\alpha_1, \ldots, \alpha_m) \in \Lambda^m$, and suppose that $\mu(\alpha_{m-i}, \ldots, \alpha_m) = \left(\alpha'_{m-i}, \ldots, \alpha'_m\right)$. We wish to construct a path p in $\mathcal{DB}$ that connects v to $\mu(v)$. The edge $(v, \mu(v))$ corresponds to some edge e' in G', so let G'_β be the matching to which e' belongs. We turn to construct the path p. The first edge in the path p connects $v = (\alpha_1, \ldots, \alpha_m)$ to the vertex $(\beta, \alpha_1, \ldots, \alpha_{m-1})$. The next $2 \cdot i$ edges of p will be the edges of the path in $\mathcal{P}_\beta$ that connects $(\beta, \alpha_1, \ldots, \alpha_{m-1})$ to $\left(\beta, \alpha_1, \ldots, \alpha_{m-i-1}, \alpha'_{m-i}, \ldots, \alpha'_{m-1}\right)$. Finally, the last edge of p will go from the vertex $\left(\beta, \alpha_1, \ldots, \alpha_{m-i-1}, \alpha'_{m-i}, \ldots, \alpha'_{m-1}\right)$ to the vertex $\left(\alpha_1, \ldots, \alpha_{m-i-1}, \alpha'_{m-i}, \ldots, \alpha'_m\right) = \mu(v)$. Observe that p indeed connects v to $\mu(v)$ and is of length $2 \cdot (i + 1)$

It remains to show that for each $j \in [2i + 2]$ it holds that every vertex v is the j-th vertex of exactly one. The cases of $j = 1$ and $j = 2 \cdot i + 2$ are trivial. We analyze the case of $j = 2$, and the rest of the cases will follow from the induction hypothesis. Let $u = (\beta, \alpha_1, \ldots, \alpha_{m-1}) \in \Lambda^m$. We show that u is the second vertex of a unique path p by constructing p. Let e' be the outgoing edge

of the vertex $(\alpha_1, \ldots, \alpha_{m-1})$ that belongs to the matching G'_β - observe that there is a unique choice of such e'. The edge e' of G' corresponds to some unique edge $(v, \mu(v))$ of G. Now, by construction, the only path p such that u is the second vertex of p is the path that connects v to $\mu(v)$. The required result follows.

C Proof of Claim 5.6

In this section, we prove Claim 5.6, restated below. Recall that $G = (\mathbb{F}^m, E)$ is a graph with linear structure and in particular E is a linear supspace of edges.

Claim (5.6, restated)**.** *Let $d \in \mathbb{N}$ and let E_a be a uniformly distributed d-subspace of E. Then,* $\Pr[\dim(\text{left}(E_a)) = d] \geq 1 - d/q^{m-d}$*, and conditioned on* $\dim(\text{left}(E_a)) = d$*, it holds that* $\text{left}(E_a)$ *is a uniformly distributed d-subspace of $\mathbb{F}^m$. The same holds for* $\text{right}(E_a)$*.*

More generally, let E_b be a fixed subspace of E such that $\dim(E_b) > d$ *and* $\dim(\text{left}(E_b)) > d$*. Let E_a be a uniformly distributed d-subspace of E_b. Then,* $\Pr[\dim(\text{left}(E_a)) = d] \geq 1 - d/q^{\dim(\text{left}(E_b))-d}$*, and conditioned on* $\dim(\text{left}(E_a)) = d$*, it holds that* $\text{left}(E_a)$ *is a uniformly distributed d-subspace of* $\text{left}(E_b)$*. Again, the same holds for* $\text{right}(E_a)$*.*

Proof We prove the proposition only for special case in which $E_b = E$ and only for $\text{left}(E_a)$. The proof of the general case and of the case of for $\text{right}(E_a)$ is analogous. Let $e_1, \ldots, e_d$ be independent and uniformly distributed vectors of E, and let $E'_a = \text{span}\{e_1, \ldots, e_d\}$. We prove Proposition 5.6 by showing that E_a is distributed similarly to E'_a, and analyzing the distribution of E'_a.

Observe that by Proposition 2.14, it holds that conditioned on $\dim(E'_a) = d$, the subspace E'_a is a uniformly distributed d-subspace of E. It therefore holds that

$$\begin{aligned}
\Pr[\dim(\text{left}(E_a)) = d] &= \Pr\left[\dim\left(\text{left}\left(E'_a\right)\right) = d \middle| \dim\left(E'_a\right) = d\right] \\
&\geq \Pr\left[\dim\left(\text{left}\left(E'_a\right)\right) = d \text{ and } \dim\left(E'_a\right) = d\right] \\
&= \Pr\left[\dim\left(\text{left}\left(E'_a\right)\right) = d\right]
\end{aligned}$$

where the last equality holds since clearly $\dim(\text{left}(E'_a)) = d$ implies $\dim(E'_a) = d$. Now, since $\text{left}(\cdot)$ is a linear function, it holds that $\text{left}(e_1), \ldots \text{left}(e_d)$ are independent and uniformly distributed vectors of $\text{left}(E) = \mathbb{F}^m$, and therefore by Proposition 2.14 it holds that $\Pr[\dim(\text{left}(E'_a)) = d] \geq 1 - d/q^{m-d}$. It thus follows that $\Pr[\dim(\text{left}(E_a)) = d] \geq 1 - d/q^{m-d}$, as required.

It remains to show that conditioned on $\Pr[\dim(\text{left}(E_a)) = d]$ it holds that $\text{left}(E_a)$ is a uniformly distributed d-subspace of $\mathbb{F}^m$. To see it, observe that for every fixed d-subspace D of $\mathbb{F}^m$, it holds that

$$\begin{aligned}
\Pr[\text{left}(E_a) = D | \dim(\text{left}(E_a)) = d] &= \Pr\left[\text{left}\left(E'_a\right) = D \middle| \dim\left(E'_a\right) = d \text{ and } \dim\left(\text{left}\left(E'_a\right)\right) = d\right] \\
&= \Pr\left[\text{left}\left(E'_a\right) = D \middle| \dim\left(\text{left}\left(E'_a\right)\right) = d\right]
\end{aligned}$$

where the first equality again holds since conditioned on $\dim(E'_a) = d$ it holds that E'_a is a uniformly distributed d-subspace, and the second equality again holds since $\dim(\text{left}(E'_a)) = d$ implies $\dim(E'_a) = d$. Now, it holds that $\text{left}(E'_a)$ is the span of d uniformly distributed vectors of $\mathbb{F}^m$, and therefore by Proposition 2.14 it holds that conditioned on $\dim(\text{left}(E'_a)) = d$ the subspace $\text{left}(E'_a)$ is a uniformly distributed d-subspace of $\text{left}(E_b)$. This implies that the probability

$$\Pr\left[\text{left}\left(E'_a\right) = D \middle| \dim\left(\text{left}\left(E'_a\right)\right) = d\right]$$

is the same for all possible choices of D, and therefore the probability

$$\Pr[\text{left}(E_a) = D | \dim(\text{left}(E_a)) = d]$$

is the same for all possible choices of D, as required. ■

ECCC ISSN 1433-8092

http://eccc.hpi-web.de