Quasipolynomial-Time Canonical Form for Steiner Designs

László Babai[*]
laci@cs.uchicago.edu

John Wilmes[†]
wilmesj@math.uchicago.edu

Departments of Mathematics and Computer Science
University of Chicago
Chicago, IL 60637

ABSTRACT

A Steiner 2-design is a finite geometry consisting of a set of "points" together with a set of "lines" (subsets of points of uniform cardinality) such that each pair of points belongs to exactly one line. In this paper we analyse the individualization/refinement heuristic and conclude that after individualizing $O(\log n)$ points (assigning individual colors to them), the refinement process gives each point an individual color. The following consequences are immediate: (a) isomorphism of Steiner 2-designs can be tested in $n^{O(\log n)}$ time, where n is the number of lines; (b) a canonical form of Steiner 2-designs can be computed within the same time bound; (c) all isomorphisms between two Steiner 2-designs can be listed within the same time bound; (d) the number of automorphisms of a Steiner 2-design is at most $n^{O(\log n)}$ (a fact of interest to finite geometry and group theory.)

The best previous bound in each of these four statements was moderately exponential, $\exp(\widetilde{O}(n^{1/4}))$ (Spielman, STOC'96). Our result removes an exponential bottleneck from Spielman's analysis of the Graph Isomorphism problem for strongly regular graphs.

The results extend to Steiner t-designs for all $t \geq 2$.

Strongly regular (s. r.) graphs have been known as hard cases for graph isomorphism testing; the best previously known bound for this case is moderately exponential, $\exp(\widetilde{O}(n^{1/3}))$ where n is the number of vertices (Spielman, STOC'96). Line graphs of Steiner 2-designs enter as a critical subclass via Neumaier's 1979 classification of s. r. graphs.

Previously, $n^{O(\log n)}$ isomorphism testing and canonical forms for Steiner 2-designs was known for the case when the lines of the Steiner 2-design have bounded length (Babai and Luks, STOC'83). That paper relied on Luks's group-theoretic divide-and-conquer algorithms and did not yield a subexponential bound on the number of automorphisms.

To analyse the individualization/refinement heuristic, we develop a new structure theory of Steiner 2-designs based on the analysis of controlled growth and on an addressing scheme that produces a hierarchy of increasing sets of pairwise independent, uniformly distributed points. This scheme represents a new expression of the structural homogeneity of Steiner 2-designs that allows applications of the second moment method.

We also address the problem of reconstruction of Steiner 2-designs from their line-graphs beyond the point of unique reconstructability, in a manner analogous to list-decoding, and as a consequence achieve an $\exp(\widetilde{O}(n^{1/6}))$ bound for isomorphism testing for this class of s. r. graphs.

Results, essentially identical to our main results, were obtained simultaneously by Xi Chen, Xiaorui Sun, and Shang-Hua Teng, building on a different philosophy and combinatorial structure theory than the present paper. They do not claim an analysis of the individualization/refinement algorithm but of a more complex combinatorial algorithm.

We comment on how this paper fits into the overall project of improved isomorphism testing for strongly regular graphs (the ultimate goal being subexponential ($\exp(n^{o(1)})$) time). In the remaining cases we need to deal with s. r. graphs satisfying "Neumaier's claw bound," permitting the use of a separate set of asymptotic structural tools. In joint work (in progress) with Chen, Sun, and Teng, we address that case and have already pushed the overall bound below $\exp(\widetilde{O}(n^{1/4}))$. The present paper is a methodologically distinct and stand-alone part of the overall project.

Categories and Subject Descriptors

F.2.2 [**Analysis of Algorithms and Problem Complexity**]: Nonnumerical Algorithms and Problems

Keywords

algorithms, graph isomorphism, canonical forms, combinatorial designs, strongly regular graphs

1. INTRODUCTION

A *Steiner t-design* $S(t, k, v)$ is a pair $\mathfrak{X} = (\mathcal{P}, \mathcal{L})$ where $\mathcal{P}$ is a set of v *points* and $\mathcal{L}$ a collection of k-subsets of $\mathcal{P}$, called *lines*, with the property that for any set T of t points there is a unique line containing T ($t \leq k < v$).

1.1 The main result

In this paper we prove the following main result. For more detailed statements, see Theorem 6, Observation 5, and Theorem 33.

[*] Partially supported by NSF grant CCF-1017781.

[†] Partially supported by an NSF Graduate Research Fellowship under grant DGE-1144082.

STOC'13, June 1–4, 2013, Palo Alto, California, USA.
Copyright 2013 ACM 978-1-4503-2029-0/13/06 ...$15.00.

THEOREM 1. *Isomorphism of Steiner t-designs can be tested in time $n^{O(\log n)}$ where n is the number of lines.*

Using the number of lines as our parameter is justified by the fact that this is the length of the input. We have $n \le \binom{v}{t}$, so for bounded t we also have a $v^{O(\log v)}$ bound.

The essence of the difficulty is in solving the problem for $t = 2$; our solution then extends to arbitrary t by a standard "derived design" argument.

While the complexity of testing isomorphism of Steiner designs, a class of much-studied objects in combinatorics, coding theory, cryptography, and statistics [5, 8, 16], should be of interest in its own right, strong motivation also comes from the problem of testing isomorphism of strongly regular graphs, a notoriously hard case of the Graph Isomorphism problem, as we explain below.

The question arises how "final" our main result is; in particular, whether a polynomial-time algorithm or even an $n^{o(\log n)}$ algorithm could be expected for testing isomorphism of Steiner 2-designs. For the simplest case, Steiner triple systems (STS), the easy $n^{O(\log n)}$ bound was established in the late 1970s, and it has not been improved in the intervening more than three decades. Given this history, an $n^{o(\log n)}$ algorithm, even for STS, would count as a major breakthrough. As to the number of automorphisms, the $n^{O(\log n)}$ bound is tight, as shown by certain STSs (projective spaces over $\mathbb{F}_2$, affine spaces over $\mathbb{F}_3$).

Results essentially identical with our main results (for the essential case $t = 2$) were obtained simultaneously by Xi Chen, Xiaorui Sun, and Shang-Hua Teng and appear in these proceedings [7]. We give a brief comparison of the two papers at the end of Section 2.

1.2 Strongly regular graphs

The *line graph* of a Steiner 2-design $\mathfrak{X}$ is a graph with vertex set $\mathcal{L}$; two elements of this set will be adjacent in the line graph if the corresponding lines intersect.

A graph $G = (V, E)$ is *strongly regular* if there exist parameters (ρ, λ, μ) such that (i) every vertex of G has valency ρ; (ii) every pair of adjacent vertices has λ common neighbors; and (iii) every pair of non-adjacent vertices has μ common neighbors. The pentagon and the Petersen graph are strongly regular; and so are the line graphs of Steiner 2-designs and certain graphs derived from sets of orthogonal Latin squares ("Latin square graphs," cf. [12, 11, 15]).

Although perceived as hard cases, strongly regular graphs (s. r. graphs) are not known to be graph-isomorphism complete, and testing their isomorphism has been slightly ahead of progress on the general Graph Isomorphism problem.

In 1980, Babai [1] (cf. [2]) used a simple combinatorial argument to test isomorphism of s. r. graphs in time

$$\exp(O((n/\rho)\log^2 n)), \tag{1}$$

where n is the number of vertices and $\rho < n/2$ is the valency. (Note: the complement of a s. r. graph is s. r., thus the assumption on ρ.) As $\rho \ge \sqrt{n-1}$, this gives an algorithm in moderately exponential, $\exp(O(\sqrt{n}\log^2 n))$ time for all ρ.

At STOC'96, Daniel Spielman presented a much more involved but still purely combinatorial argument yielding an $\exp(O(n^{1/3}\log^2 n))$ time bound [15]. This should be compared with $\exp(O(\sqrt{n \log n}))$, the best known bound for testing isomorphism of general graphs [4, 3, 18] based on Luks's group-theoretic method [10].

A naive approach to graph isomorphism testing consists in assigning unique colors to a small set S of vertices ("individualizing" the members of S) and then applying a color refinement procedure [14, 17]. The cost of this procedure is $n^{s+O(1)}$ where $s = |S|$. While the seminal paper of Cai, Furer, and Immerman [6] demonstrates that this method fails badly on certain pairs of non-isomorphic graphs (it requires $s = \Omega(n)$ to distinguish them), the jury is still out on the question whether or not this method could succeed for strongly regular graphs possibly in as little as quasipolynomial $(\exp((\log n)^{O(1)}))$ time. In this paper we make a step in this direction.

Both the Babai [1] and the Spielman [15] papers were based on the individualization/refinement heuristic. Spielman's analysis organizes Neumaier's [12] classification of connected s. r. graphs of valency $\rho < n/2$ as follows:

(a) Latin square graphs (settled by Miller [11] in $n^{O(\log n)}$)

(b) line graphs of Steiner 2-designs satisfying $\rho < f(n)$, for a certain function $f(n) \sim n^{3/4}$ (solved by Spielman in time $\exp(O(n^{1/4}\log^2 n))$)

(c) $\rho = (n-1)/2$ (settled in [1] in $n^{O(\log n)}$)

(d) graphs, satisfying a certain eigenvalue inequality referred to as "Neumaier's claw bound" (solved by Spielman in time $\exp(O(n^{1/4}\log^2 n))$ for $\rho = o(n^{2/3})$ and in time $\exp(O(n^{1/3}\log^2 n))$ by eq. (1) [1] for $\rho = \Omega(n^{2/3})$).

In this paper we address case (b) with no constraint on the valency ρ. The solution consists of two parts:

(i) testing isomorphism of Steiner 2-designs

(ii) reconstructing a Steiner 2-design from its line graph.

Our main result is a solution of problem (i) in $n^{O(\log n)}$ time (Theorem 1). For problem (ii), Spielman gave a polynomial-time reconstruction algorithm [15] assuming $\rho < f(n)$ for a certain function $f(n) \sim n^{3/4}$. Under this constraint on ρ, reconstruction is unique, but this does not hold for larger valencies. We address this situation in Section 9 with the following overall result:

THEOREM 2. *Isomorphism of line graphs of Steiner 2-designs can be tested in* $\exp(O(n^{1/6}\log^3 n))$ *time.*

However, for the purposes of the larger project of testing isomorphism of strongly regular graphs in subexponential time, Theorem 1 along with Spielman's reconstruction algorithm already suffice to eliminate case (b) as a bottleneck: for $\rho < f(n)$ we find canonical forms for the line graphs in $n^{O(\log n)}$ time; for larger ρ the line graphs fall in case (d).

1.3 Number of automorphisms

Our analysis also proves the following mathematical result, of interest in its own right.

THEOREM 3. *Let $\mathfrak{X}$ be an $S(t,k,v)$ with $t < k$. Then $\mathfrak{X}$ has at most $n^{O(\log n)}$ automorphisms, where n is the number of lines.*

Even for the case $t = 2$, no bound better than Spielman's $\exp(O(n^{1/4}\log^2 n))$ appears to have been known previously. We note that this upper bound does not hold for the trivial case $t = k$ in which case all t-subsets are lines; in that case, all the $v!$ permutations are automorphisms. For $t < k$ we are in fact able to list all automorphisms (and all isomorphisms) within the time bound stated in Theorem 1.

1.4 Canonical form

A function F from a class $\mathcal{F}$ of finite structures to itself is called a *canonical form* if for every $X, Y \in \mathcal{F}$ we have $X \cong F(X)$ and $X \cong Y$ if and only if $F(X) = F(Y)$. Isomorphism of members of $\mathcal{F}$ can be decided by two applications of a canonical form function and comparison of the outputs.

THEOREM 4. *A canonical form of Steiner t-designs can be computed in time $n^{O(\log n)}$ where n is the number of lines.*

Using the number of lines as our parameter is justified by the fact that it is a lower bound on the length of the input. Since the number of lines is $n \leq \binom{v}{t}$, for bounded t we also have an equivalent $v^{O(\log v)}$ bound.

The case $t = k$ being trivial (all t-subsets are lines), we assume $t < k$ throughout the paper.

1.5 Previous $n^{O(\log n)}$ bound for short lines

In 1983, Babai and Luks [4] showed that Steiner 2-designs with lines of bounded length admit computation of a canonical form (and thus, isomorphism testing) in time $n^{O(\log n)}$. However, that result does not imply a subexponential bound on the number of automorphisms. The structural property of the automorphism group G of a Steiner 2-design proved in that paper is that G has a subgroup H of index $n^{O(\log n)}$ such that H is "nice" in the sense that every composition factor of H is a subgroup of the symmetric group S_k where k is the number of points per line. This property permits an efficient application of Luks's group-theoretic divide-and-conquer algorithm [10] to H.

On the other hand, [4] also applies to (v, k, λ)-designs (BIBDs) with bounded k and λ, where λ is the number of lines through any given pair of points. Our argument is so tight that it does not even extend to the case $\lambda = 2$. We propose as an open question whether or not BIBDs with $\lambda = 2$ have a quasipolynomially bounded number $(\exp(\log^{O(1)} n))$ of automorphisms.

Huber [9] recently extended the cited result from [4] to t-designs. Instead of using the simple trick (derived designs) given in Section 8 to reduce the problem to the case $t = 2$, he reapplies the method of [4] to this case. Again, for the case of Steiner t-designs, for bounded k and t we obtain an $n^{O(\log n)}$ bound on the order of the automorphism group; but our method does not extend to the non-Steiner cases, not even to the case of $\lambda = 2$.

2. INDIVIDUALIZATION / REFINEMENT

For the rest of the paper except for Section 8 we consider the case $t = 2$ (Steiner 2-designs). We introduce further terminology and notation. We say that a Steiner 2-design $\mathfrak{X}$ is an $S(2, k, v)$ if it has v points and k points per line. For distinct $p, q \in \mathcal{P}$, we denote the unique line containing p and q by $\overline{pq} \in \mathcal{L}$. An *element* of $\mathfrak{X}$ is a point or a line. We have an *incidence relation* between elements of $\mathfrak{X}$: points are incident with the lines containing them. Note that an $S(2, 2, v)$ is simply a complete graph on v vertices, so we will *assume $k \geq 3$ throughout.*

A *coloring* γ of $\mathfrak{X}$ is a map $\gamma : \mathcal{P} \sqcup \mathcal{L} \to \mathcal{C}$ from the elements of $\mathfrak{X}$ to an arbitrary ordered set $\mathcal{C}$, called the set of "colors," such that for any $x \in \mathcal{P}$ and $y \in \mathcal{L}$ we have $\gamma(x) < \gamma(y)$. The set range$(\gamma|_{\mathcal{P}})$ is called the set of *point-colors* of $\mathfrak{X}$. We refer to colored Steiner 2-designs $(\mathfrak{X}, \gamma)$ as *systems.* Isomorphisms of systems preserve color by definition. For $x \in \mathcal{P} \sqcup \mathcal{L}$, the color class of x is the set $\gamma^{-1}(\gamma(x))$, i. e., the set of elements of the same color as x. The *multiplicity* of a color $c \in \mathcal{C}$ in a system is the cardinality of the color class $\gamma^{-1}(c)$.

Given a system $(\mathfrak{X}, \gamma)$, a *refinement step* produces a new coloring γ^* of $\mathfrak{X}$, defined as follows. For $x \in \mathcal{P} \sqcup \mathcal{L}$, let $\Gamma(x)$ denote the set of elements incident with x. (If x is a point then these elements are lines; if x is a line then these elements are points.) We set $\gamma^*(x) = (\gamma(x); |\Gamma(x) \cap \gamma^{-1}(c)| : c \in \mathcal{C})$. Let $\mathcal{C}^*$ denote the range of γ^*, ordered lexicographically, and for each $c^* \in \mathcal{C}^*$, let $\delta(c^*)$ denote the rank of c^* in this lexicographic ordering. Set $\gamma'(x) = \delta(\gamma^*(x))$. So γ' is again a coloring, by the set $\mathcal{C}' = \{1, \ldots, |\mathcal{C}^*|\}$ of colors. It is clear that for elements $x, y \in \mathcal{P} \sqcup \mathcal{L}$, we have $\gamma(x) < \gamma(y) \Rightarrow \gamma'(x) < \gamma'(y)$; in particular, the partition into γ'-classes is a refinement of the partition into γ-classes. If these two partitions are the same (γ' does not yield a proper refinement of γ), we say that γ is a *stable* coloring.

The *refinement process* repeats the refinement step until a stable coloring is reached. This stable coloring is called the *stable refinement* of the original coloring. The process clearly terminates after at most $v + |\mathcal{L}| = O(v^2)$ refinement steps.

We say that a set is *stable* with respect to the coloring γ if it is a union of color classes of the stable refinement of γ.

Individualization of an element assigns a unique color to the element, and individualization of a set S of elements assigns unique colors to each element of S. *Depth-d stabilization* is the process of individualizing d vertices and refining to a stable coloring. We say that depth-d stabilization *achieves* or *yields a certain type of coloring* if there exists a set S such that $|S| \leq d$ and after individualizing S, the stable refinement of the resulting coloring is of the stated type. For instance, we say that depth-d stabilization *completely splits the system* $(\mathfrak{X}, \gamma)$ if there exists a set S of cardinality d such that after individualization of S, every color in the stable refinement is unique (has multiplicity one).

The following standard observations are the basis of using depth-d stabilization in isomorphism testing and related problems.

OBSERVATION 5. *If some set of d points splits the system $(\mathfrak{X}, \gamma)$ completely then*

(a) $|\operatorname{Aut}(\mathfrak{X}, \gamma)| \leq v^d$;

(b) a canonical form of $(\mathfrak{X}, \gamma)$ can be computed in $v^{d+O(1)}$ steps;

(c) isomorphism of $(\mathfrak{X}, \gamma)$ and any system $(\mathfrak{X}_1, \gamma_1)$ can be decided, and the set of isomorphisms listed, within $v^{d+O(1)}$ steps.

In the light of the above, we look for small sets that split the system completely.

THEOREM 6. *Let $\mathfrak{X}$ be an $S(2, k, v)$ with $k \geq 3$. For some $d = O(\log v)$, depth-d stabilization completely splits $\mathfrak{X}$.*

Henceforth we assume $\mathfrak{X} = (\mathcal{P}, \mathcal{L})$ is an $S(2, k, v)$ equipped with a stable coloring γ (except in Section 8 where we discuss the extension of the results to t-designs).

2.1 Brief outline of the proof

Along the way to proving Theorem 6, we first achieve the following four targets. Each target is achieved after $O(\log v)$ individualizations. We denote by $R = (v - 1)/(k - 1)$ the number of lines incident on a point.

Target 1. *Ensure that there exists a stable set of size at least $R/4$ and maximum point-color multiplicity at most $2^{-24}R$.*

Target 2. *Ensure that the maximum point-color multiplicity in $\mathfrak{X}$ is at most $2^{-18}R$.*

Target 3. *Ensure that the maximum point-color multiplicity in $\mathfrak{X}$ is at most $2^{-14}(\log(k-1)/\log v)^2R$.*

Target 4. *Completely split $\mathfrak{X}$.*

We outline the procedure that yields each target.

For a stable set A, we define the *granularity* of A (with respect to the current coloring) as $|A|/m(A)$ where $m(A)$ is the largest color multiplicity within A. The first target is achieved by an iterative process that individualizes points in stages. In each round, we double the granularity of some subset until Target 1 is achieved. The inductive step is described in the combination of the following two lemmas.

Lemma 7. *Suppose $A \subseteq \mathcal{P}$ is a stable set with granularity g. Then for some $d = O(\log v/\log k)$, depth-d stabilization yields a stable set B of size $|B| > R/4$ and granularity $\geq g/4$.*

Lemma 8. *Suppose $B \subseteq \mathcal{P}$ is a stable set of size $|B| > R/4$ with granularity g. Then either B satisfies Target 1 or for some $d = O(1)$, depth-d stabilization yields a stable set C of granularity $\geq 8g$.*

So our iteration will move from A to B above and either terminate there or replace A by $A' = C$. The number of iterations is $\leq \log k + O(1)$ since once the granularity reaches $2^{24}k$, the set B will have reached Target 1. So the total number of points individualized in this phase is $O(\log v)$.

Lemma 7 is proved via the strictly controlled growth of a tower of "cones" (Sections 3 and 4). A "cone" over a set A with apex p is the union of lines connecting p to the points of A. We show that for $|A| = o(v/k)$ and for almost all choices of the apex p, the size of the cone is about $(k-1)|A|$ (Lemma 12).

Thus, given a small stable set A, we obtain a large stable set B with a comparable granularity by individualizing a sequence of $\log v/\log(k-1)$ apexes.

The size of B can only be controlled up to a factor of about $k-1$; we make it greater than $R/4$ (and $\leq v$, of course). If the largest color-class in B has size $\leq 2^{-24}R$ then B satisfies Target 1. Otherwise, B is "large" (the size of the largest color class times the granularity) and therefore a random line ℓ has a good chance of intersecting B in many color-classes (Lemma 15), so after individualizing ℓ some subset of ℓ will have granularity $\Omega($ granularity of $B)$. Applying this to a constant number of random lines ℓ we obtain a subset C with any constant times the granularity of B, proving Lemma 8.

The following lemma accomplishes the next phase.

Lemma 9. *There exists $d = O(\log k)$ such that if Target 1 has been achieved, then depth-d stabilization achieves Target 2.*

This in fact is the case $\delta = 2^{-24}$ of Lemma 16.

The idea is the following. Assume the set A satisfies Target 1. We extend the fine coloring of A to a $2/15$ fraction of the space, and subsequently to the entire space, by individualizing a logarithmic number of additional points, and considering how lines through them intersect the color-classes of A, using estimates provided by Lemma 18.

We now move to Target 3.

Lemma 10. *There exists $d = O(\log v)$ such that if Target 2 is achieved, then depth-d stabilization achieves Target 3.*

We need to refine the already fine coloring of Target 2. We accomplish this by showing that for a tower of cones in such a finely colored system, the point-color classes within the tower grow more slowly than the tower itself (Lemma 29). Thus, we obtain a stable set with improved granularity. We then use Lemmas 7 and 16 to extend this finer coloring to the entire space, completing the proof of Lemma 10.

Finally we reach Target 4.

Lemma 11. *There exists $d = O(\log v/\log k)$ such that if Target 3 is achieved, then depth-d stabilization achieves Target 4.*

We prove this in Sec. 7 via the analysis of a stochastic process that produces increasing numbers of pairwise independent, uniformly distributed points. We note that this pairwise independence is already used in reaching Target 3 (Lemma 10).

The stochastic process is defined in Sec. 5. The idea is an addressing scheme, which we sketch. A *depth-d label* is a string $x \in [k-1]^d$. We assign a point $f_d(x)$ to each label x; this assignment depends on a sequence of d randomly chosen apexes in our process of building a tower of cones. We show that (1) for every x, the point $f_d(x)$ is uniformly distributed over $\mathcal{P}$; (2) for $x \neq y$, the points $f_d(x)$ and $f_d(y)$ are independent (Lemma 21). From these, via Chebyshev's inequality, it follows that for some $d = \Theta(\log v/\log k)$, the labels are nearly uniformly distributed among the points (Corollary 23). Because of the fine partition, we infer that many labels correspond to unique colors after individualizing the apexes (Observation 31); so, by the foregoing, each point is likely to receive such a label; therefore its color is unique.

Note that having reached Target 4 at the cost of $O(\log v)$ individualizations proves Theorem 6.

2.2 Brief comparison with the work of Chen, Sun, and Teng

Both papers prove items (a), (b), (c), and (d) from the first paragraph of our Abstract. Our paper achieves this by analysing the classical individualization/refinement algorithm. The Chen–Sun–Teng [7] paper, while inspired by this intuitive method, analyses (in their words) a "considerably more complicated" combinatorial algorithm.

The two papers differ in the philosophy and techniques used in the analysis. The paper by Chen et al. [7] builds distinguishers for each pair of points, following the general framework of the papers by Babai [1] and Spielman [15]. Our paper takes an overall approach to gradually building an increasing number of color classes of controlled sizes, until in a final step each point gets individualized. As noted in the previous subsection, this final step turns on a concentration inequality via Chebyshev, using large families of pairwise independent random points defined from the geometry of the space. We believe that the pairwise independence of our variables provides a new and useful expression of the structural homogeneity of Steiner 2-designs.

3. CONES

In this section, we introduce *cones*, which are the fundamental combinatorial objects used in the proof. We give some elementary estimates on cones, and use them to prove Lemmas 7 and 8.

For a pair (p, q) of distinct points, we define the *truncated line* $\ell(p, q) = \overline{pq} \setminus \{q\}$; so $|\ell(p, q)| = k - 1$. We also define the (degenerate) truncated line $\ell(p, p)$ to be the singleton set $\{p\}$. For $p \in \mathcal{P}$ and $A \subseteq \mathcal{P}$, define the *cone with base A and apex p* as $C_p(A) = \bigcup_{a \in A} \ell(a, p)$. Note that if A is a fixed stable set, then $|C_p(A)|$ is determined by the color of p in the stable refinement. Furthermore, if A is stable, then after individualizing p and refining to a stable coloring, $C_p(A)$ is also stable.

The next lemma controls the size of our cones. By a *random* member of a nonempty finite set S we mean a member of S chosen from the uniform distribution.

LEMMA 12. *Let $A \subseteq \mathcal{P}$ be nonempty and choose a point $p \in \mathcal{P}$ at random. Then*

$$\mathbb{E}_p(|C_p(A)|) \geq (k-1)\left(1 - \frac{(k-2)|A|}{v}\right)|A|.$$

PROOF.

$$\begin{aligned}
\mathbb{E}_p(|C_p(A)|) &= \frac{1}{v}\sum_{p \in \mathcal{P}} |C_p(A)| \\
&\geq \frac{k-1}{v}\sum_{p \in \mathcal{P}} |\{\ell \in \mathcal{L} \colon p \in \ell, \quad \ell \cap A \setminus \{p\} \neq \emptyset\}| \\
&\geq \frac{k-1}{v}\sum_{p \in \mathcal{P} \setminus A} |\{\ell \in \mathcal{L} \colon p \in \ell \text{ and } |\ell \cap A| = 1\}| \\
&= \frac{k-1}{v}\sum_{a \in A} |\{p \in \mathcal{P} \setminus A \colon \{a\} = \overline{ap} \cap A\}| \\
&\geq \frac{k-1}{v}\sum_{a \in A} v - |A|(k-2) \\
&\geq (k-1)\left(1 - \frac{(k-2)|A|}{v}\right)|A|. \quad \square
\end{aligned}$$

We now observe a duality between the apex and the members of a cone. Its consequence that the apex of a large cone belongs to many cones over the same set will be used in Lemma 18.

LEMMA 13. *Let $A \subseteq \mathcal{P}$ and fix $q \in \mathcal{P}$. If $p \in \mathcal{P}$ is chosen at random, then*

$$P_p[q \in C_p(A)] \geq \left(\frac{k-2}{k-1}\right) \cdot P_p[p \in C_q(A)] \tag{2}$$

Furthermore, if $q \notin A$, then

$$P_p[q \in C_p(A)] \leq P_p[p \in C_q(A)]. \tag{3}$$

PROOF. Clearly if $q \in A$ then $P[q \in C_p(A)] = 1$, so suppose $q \notin A$. Let $B \subseteq A$ be such that for every line ℓ of the form $\overline{qa}$ for some $a \in A$, there is a unique point $b \in \ell \cap B$. Thus, $C_q(B) = C_q(A)$ and $|B| = |C_q(A)|/(k-1)$. Furthermore, for any $u \in C_q(A) \setminus B$ we have $q \in C_u(A)$. It follows that

$$\begin{aligned}
P_p[q \in C_p(A)] \geq P_p[p \in C_q(A) \setminus B] &= \frac{|C_q(A)|}{v} - \frac{|C_q(A)|}{v(k-1)} \\
&= \left(\frac{k-2}{k-1}\right) \cdot P_p[p \in C_q(A)].
\end{aligned}$$

Inequality (3) holds since $p \in \ell(q, a)$ when $q \in \ell(p, a)$. $\square$

In the next lemma, we bound the expected intersection of a fixed set of points with a random cone over a fixed base.

LEMMA 14. *Let $A, B \subset \mathcal{P}$. Then if $p \in \mathcal{P}$ is chosen at random, we have*

$$\mathbb{E}(|C_p(A) \cap B \setminus A|) \leq \frac{(k-1)|A||B|}{v}.$$

PROOF. By (3) of Lemma 13, we have

$$\begin{aligned}
\mathbb{E}(|C_p(A) \cap B \setminus A|) &= \sum_{x \in B \setminus A} P[x \in C_p(A)] \\
&\leq \sum_{x \in B \setminus A} P[p \in C_x(A)] \\
&\leq \sum_{x \in B \setminus A} \frac{(k-1)|A|}{v}. \quad \square
\end{aligned}$$

To prove Lemma 7, we will construct an iterated "tower" of cones, foreshadowing the construction of Section 5. Recall that the granularity of a stable set A is $|A|/m(A)$, where $m(A)$ is the maximum point-color multiplicity in A.

PROOF OF LEMMA 7. Define m_0 as the maximum point-color multiplicity in $A_0 = A$. We define recursively a sequence $p_1, \dots, p_d$ of points in $\mathcal{P}$ as follows. When A_i is defined and $|A_i| \leq (\log(k-1)/\log v)R$, by Lemma 12 let p_{i+1} be such that

$$|C_{p_{i+1}}(A_i)| > (k-1)\left(1 - \frac{\log(k-1)}{\log v}\right)|A_i|,$$

and define $A_{i+1} = C_{p_{i+1}}(A_i)$. For $s = \log v/\log(k-1)$, we have

$$(k-1)^s\left(1 - \frac{\log(k-1)}{\log v}\right)^s \sim \frac{v}{\mathrm{e}},$$

and thus for some $d \lesssim s$ we have $|A_d| > (\log(k-1)/\log v)R$.

Note that after individualizing $p_1, \dots, p_i$, the set A_i is stable, so let m_i denote the maximum point-color multiplicity in the stable refinement of A_i. Observe that $m_i \leq (k-1)m_{i-1}$, so $m_i \leq (k-1)^i m_0$. It follows that that the granularity of A_d is

$$\frac{|A_d|}{m_d} \geq \left(1 - \frac{\log(k-1)}{\log v}\right)^d \cdot \frac{|A_0|}{m_0} \gtrsim \frac{|A_0|}{\mathrm{e}m_0}.$$

Thus, for v sufficiently large, the granularity of A_d is at least $1/3$ the granularity of A.

If $|A_d| > R/4$, we are done. Otherwise, suppose $B \subset \mathcal{P}$ is such that $A_d \subseteq B$ and $|B| \leq R/4$. Then by Lemmas 12 and 14, for a random point p,

$$\begin{aligned}
\mathbb{E}(|C_p(A_d) \setminus B|) &= \mathbb{E}(|C_p(A_d)| - |A_d| - |C_p(A_d) \cap B \setminus A_d|) \\
&> \frac{3}{4}(k-1)|A_d| - |A_d| - \frac{1}{4}|A_d| \\
&\geq \frac{1}{8}(k-1)|A_d|.
\end{aligned}$$

Thus, for some $d' = O(\log v/(k \log k))$, there exist d' points $p_1, \dots, p_{d'}$ such that if $B = \bigcup C_{p_i}(A_d)$ then $|B| \geq R/4$. Furthermore, the maximum point-color multiplicity r in B is at most $(k-1)m_d$, while by Lemma 12 we may ensure that $|B| > (3/4)(k-1)|A_d|$. Hence, the granularity of B is at least $1/4$ the granularity of A. $\square$

We now turn our attention to the proof of Lemma 8. The following estimate will ensure that if A is a large stable set, then after individualizing a random line ℓ, some subset of ℓ is likely to have comparable granularity to A.

LEMMA 15. *Let $A \subseteq \mathcal{P}$ be a stable set with point-color multiplicity at most m, and let $\lambda > 1$. Let ℓ be a random line. Then after individualizing ℓ, the expected number of points in $\ell \cap A$ whose color multiplicity is at most $1 + \lambda m/R$ is at least $(k/v)(1 - 1/\lambda)|A|$.*

PROOF. For $p \in A$, let $s(p)$ denote the size of the color class containing p. Of the R lines incident with p, there are at most $(s(p) - 1)R/(\lambda m)$ lines which intersect this color class in more than $\lambda m/R$ other points. Now let $\vartheta(p)$ be the indicator variable for the event that $p \in \ell$ and at most $\lambda m/R$ other points with the same color are on ℓ. Thus, since $|\mathcal{L}| = (v/k)R$, the expected number of points on ℓ whose color multiplicity is at most $1 + \lambda m/R$ is

$$\begin{aligned}\mathbb{E}\left(\sum_{p\in A}\vartheta(p)\right) &\geq \sum_{p\in A}\frac{1}{|\mathcal{L}|}\left(R - \frac{(s(p)-1)}{\lambda m}R\right)\\ &= \frac{k}{v}\left(\sum_{p\in A}1 - \frac{s(p)-1}{\lambda m}\right)\\ &\geq \frac{k|A|}{v}\left(1 - \frac{m-1}{\lambda m}\right)\\ &> \frac{k|A|}{v}\left(1 - \frac{1}{\lambda}\right). \quad \square\end{aligned}$$

PROOF OF LEMMA 8. Let B be as in the statement of the lemma, a stable set of size $|B| > R/4$ with granularity g. Suppose B does not satisfy Target 1. Then the maximum point-color multiplicity m in B is at least $2^{-24}R$, and so $|B| > 2^{-24}gR$. We may assume that v is sufficiently large that $R > 2^{29}$.

We claim that after individualizing at most 2^{30} lines, we obtain a stable set $C \subset B$ with size $|C| > 2^{28}|B|/R$ and maximum point-color multiplicity $m' < 2^{25}m/R$, and hence granularity $> 8g$. Indeed, if we have a stable set $C \subset B$ of size less than $2^{28}|B|/R < |B|/2$, then by Lemma 15, there exists a line ℓ such that there are at least $(1/2)(k/v)|B \setminus C| > |B|/(4R)$ points in $\ell \cap B \setminus C$ whose color class intersects ℓ in at most $1 + 2m/R < 2^{25}m/R$ points. Thus, by individualizing ℓ and adding its small point-color classes to C, we augment the size of C by at least $|B|/(4R)$. So, by individualizing at most 2^{30} lines, we obtain the desired set B. $\square$

4. MODERATE COLOR MULTIPLICITIES

In this section, we prove the following Lemma 16, which lets us extend a fine coloring of a large subset of the space to the entire space at the cost of a logarithmic number of individualizations.

LEMMA 16. *Assume $\delta < 2^{-14}$ and suppose A is a stable set with $|A| \geq R/4$ and point-color multiplicity at most δR. For some $d = O(\log(k/\delta))$, depth-d stabilization yields a coloring with no point-color of multiplicity greater than $64\delta R$.*

We use the following notion of a *ζ-split* to quantify progress in extending a fine coloring in the proof of Lemma 16.

Definition 17. Let $0 < \zeta < 1$, $p \in \mathcal{P}$, and $B \subseteq \mathcal{P}$ a color class. Then B is ζ-split by p if after individualizing p, no point-color in the stable refinement of B has multiplicity greater than $(1 - \zeta)|B|$.

In particular, if by individualizing a random point, we have a positive probability of obtaining a ζ-split of any large color class for some ζ not too small, then by individualizing a logarithmic number of points, we will see in the proof of Lemma 16 that we can eliminate all large color classes. The following lemma shows that we indeed have a positive probability of obtaining a ζ-split under appropriate conditions.

LEMMA 18. *Let A be a stable set with maximum point-color multiplicity δR and let $B \subseteq \mathcal{P}$ be a color class with $|C_q(A)| \geq \varepsilon v$ for some (hence, every) $q \in B$. Then there exists a $\zeta > \varepsilon/4$ such that if $|B| \geq 8\delta R$ and $p \in \mathcal{P}$ is chosen at random, then*

$$P[B \text{ is } \zeta\text{-split by } p] > \varepsilon/4 - 32\delta.$$

To prove the lemma, we use the following well-known fact.

FACT 19. *Let X be a set partitioned into b blocks, let $r = |X|/b$ be the average size of the blocks, and for every $x \in X$ let $\sigma(x)$ denote the size of the block containing x. Then for every $\varepsilon > 0$, if $x \in X$ is a random element, then*

$$P[\sigma(x) \leq \varepsilon r] \leq \varepsilon.$$

PROOF OF LEMMA 18. By Lemma 13, for any $q \in B$ we have $P[q \notin C_p(A)] < 1 - \varepsilon(k-2)/(k-1)$. Thus, $\mathbb{E}(|B \setminus C_p(A)|) < (1 - \varepsilon(k-2)/(k-1))|B|$, and so for any $0 < \zeta < 1$, Markov's inequality gives

$$\begin{aligned}P[|B \cap C_p(A)| < \zeta|B|] &= P[|B \setminus C_p(A)| \geq (1-\zeta)|B|]\\ &< \frac{1 - \varepsilon(k-2)/(k-1)}{1-\zeta}\end{aligned}$$

which is at most $1 - \varepsilon/4$ for some $\varepsilon/4 < \zeta \leq 1/2$.

Let $\mathcal{A}$ denote the collection of point-color classes in A, and let

$$\iota = P[\exists S \in \mathcal{A} \text{ such that } |C_p(S) \cap B| \geq (1-\zeta)|B|].$$

We will show that $\iota < 32\delta$. The lemma then follows, since with probability at least $1 - (1 - \varepsilon/4) - \iota$, at least a ζ-fraction of B is covered by $C_p(A)$, but there is no $S \in \mathcal{A}$ such that $C_p(S)$ covers a $(1 - \zeta)$-fraction of B.

Suppose $p \in \mathcal{P}$ and $S \in \mathcal{A}$ are such that $|C_p(S) \cap B| \geq (1 - \zeta)|B|$. Let $D = C_p(S) \cap B \setminus \{p\}$. For some $s \leq |S|$, we have D partitioned into s blocks by the lines of the form $\overline{pq}$ for $q \in S \setminus \{p\}$. By Fact 19, at least half the points $q \in D$ have $|\overline{pq} \cap D| > |D|/(2s)$. Hence, at least a $(1-\zeta)/2$-fraction of the points $q \in B \setminus \{p\}$ satisfy

$$|\overline{pq} \cap B| > \frac{(1-\zeta)|B|}{2s} \geq \frac{(1-\zeta)|B|}{2\delta R} \qquad (4)$$

as well. This property (which does not depend on S) holds for at least an ι-fraction of the points $p \in \mathcal{P}$, so that for a random pair $(p, q) \in \mathcal{P} \times B$ of distinct points, the probability that (4) holds is at least $\iota(1 - \zeta)/2$. It follows that there exists $q \in B$ such that for at least $\iota(1 - \zeta)(v - 1)/2$ points $p \in \mathcal{P} \setminus \{q\}$, equation (4) holds, and hence at least $\iota(1-\zeta)R/2$ lines containing q intersect B in more than $(1-\zeta)|B|/(2\delta R)$ points. Since $|B| \geq 4\delta R/(1 - \zeta)$, we have

$$|B| > 1 + \frac{\iota(1-\zeta)R}{2}\left(\frac{(1-\zeta)|B|}{2\delta R} - 1\right) > \frac{\iota(1-\zeta)^2|B|}{8\delta}.$$

Thus $\iota < 32\delta$, so the lemma follows. $\square$

PROOF OF LEMMA 16. Choose $p \in \mathcal{P}$ at random. By Lemma 12, we have $\mathbb{E}(|C_p(A)|) > 3v/16$ for v sufficiently large. Since $|C_p(A)| \leq v$ for all $p \in \mathcal{P}$, therefore $P(|C_p(A)| > v/16) > 2/15$. Let $U \subseteq \mathcal{P}$ be the collection of points p such that $|C_p(A)| > v/16$, so $|U| > 2v/15$. Note that U is a stable set.

For a collection $\mathcal{B}$ of subsets of $\mathcal{P}$, let $\rho(\mathcal{B}) = \sum_{B \in \mathcal{B}} |B|^2$. Let $\mathcal{B}$ be the collection of point-color classes of cardinality at least $8\delta R$ in U, and note that $\rho(\mathcal{B}) \leq v^2$. Choose $p \in \mathcal{P}$ at random, and let $\mathcal{B}_p$ be the collection of color classes of cardinality at least $8\delta R$ in U after individualizing p and refining to a stable coloring. For $B \in \mathcal{B}$, let X_B be the collection of subsets of B in $\mathcal{B}_p$. By Lemma 18, B is 2^{-6}-split by p with probability at least 2^{-7} (using the assumption on δ). Thus,

$$\begin{aligned}\mathbb{E}(\rho(X_B)) &\leq (1-2^{-7})|B|^2 + 2^{-7}(2^{-12} + (1-2^{-6})^2)|B|^2 \\ &< (1-2^{-13})|B|^2 .\end{aligned}$$

It follows that $\mathbb{E}(\rho(\mathcal{B}_p)) \leq (1-2^{-13})\rho(\mathcal{B})$, so there exists some p with $\rho(\mathcal{B}_p) \leq (1-2^{-13})\rho(\mathcal{B})$. Thus, by individualizing at most $O(\log(k/\delta))$ points, we guarantee that $\rho(\mathcal{B})$ is reduced by a factor of at least $k^2/(64\delta^2)$, and hence no point-color in U has multiplicity greater than $8\delta v/k < 8\delta R$.

Now since $|U| > 2v/15$, we have $\mathbb{E}(|C_p(U)|) > 2v/15$ for every point p. Thus, the result follows by repeating the argument of the previous paragraph with the set $\mathcal{P}$ in place of U and the set U in place of A. □

5. TOWERS

In the proof of Lemma 7, we constructed a "tower" of cones. In this section, we will analyze random towers of cones via an addressing scheme. This scheme produces rapidly increasing families of pairwise independent points and will be an essential tool in our proofs of Lemmas 10 and 11.

By a *numbering* of the truncated line $\ell(p,q)$ we mean a bijection from $[k-1]$ to $\ell(p,q)$ if $p \neq q$, and the constant map $[k-1] \to \{p\}$ if $p = q$.

Let $(p_0, \ldots, p_d)$ be a sequence of points, not necessarily distinct. A *d-dimensional tower* generated by $(p_0, \ldots, p_d)$ is a sequence $f = (f_0, \ldots, f_d)$ of maps $f_j : [k-1]^j \to \mathcal{P}$ such that $f_0 = p_0$ and for every $(x_1, \ldots, x_{j-1}) \in [k-1]^{j-1}$ the map $f_{(x_1,\ldots,x_{j-1})}(y) = f_j(x_1, \ldots, x_{j-1}, y)$ is a numbering of $\ell(f_{j-1}(x_1, \ldots, x_{j-1}), p_j)$. Note that $\mathrm{range}(f_{j+1}) = C_{p_j}(\mathrm{range}(f_j))$. In particular, $\mathrm{range}(f_j) \subseteq \mathrm{range}(f_{j+1})$ and $\mathrm{range}(f_j)$ does not depend on the particular numberings chosen along the way. Furthermore, $\mathrm{range}(f_j)$ is stable after individualization of $p_0, \ldots, p_j$.

We call the elements of $[k-1]^d$ *labels*, and say that x is a *label of p* if $f_d(x) = p$, the value of d being clear from the context. A point $p \in \mathcal{P}$ may have multiple labels or no labels at all. For a label $x = (x_1, \ldots, x_d) \in [k-1]^d$, let $x^j = (x_1, \ldots, x_j) \in [k-1]^j$ denote the prefix of x in $[k-1]^j$; so $x = x^d$.

A *random tower* over $(p_0, \ldots, p_d)$ is defined by choosing the numbering $f_{(x_1,\ldots,x_j)}$ at random for every $0 \leq j \leq d$ and $(x_1, \ldots, x_j) \in [k-1]^j$. The rest of this section refers to a d-dimensional random tower f over a random sequence $(p_0, \ldots, p_d)$ of points in $\mathcal{P}$.

LEMMA 20. *For any $x \in [k-1]^d$, the point $f_d(x)$ is uniformly distributed over $\mathcal{P}$.*

PROOF. The claim is obvious for $d = 0$, so suppose for induction on d that $f_{d-1}(x)$ is uniformly distributed over $\mathcal{P}$. For any $p \in \mathcal{P}$, it is clear under either of the conditions $p_d = p$ and $p_d \neq p$ that $P[f_d(x) = p] = 1/v$, so it is true overall. □

Now we establish the key pairwise independence property.

LEMMA 21. *For any distinct $x, y \in [k-1]^d$ with $d \geq 1$, the points $f_d(x), f_d(y)$ are independent random variables.*

PROOF. Let $x = (x_1, \ldots, x_d)$ and $y = (y_1, \ldots, y_d)$ be distinct, and let j be the length of the common prefix of x and y. Define a $(d-j)$-dimensional tower g over $(f_j(x), p_{j+1}, \ldots, p_d)$ by $g_k(z) = f_k((x^j, z))$. Then letting $x' = (x_{j+1}, \ldots, x_d)$ and $y' = (y_{j+1}, \ldots, y_d)$, we have $g_k(x') = f_{k+j}(x')$ and $g_k(y') = f_{k+j}(y')$ for all $0 \leq k \leq d-j$. But by Lemma 20, we have $f_j(x)$ uniformly distributed over $\mathcal{P}$, and clearly $f_j(x)$ is independent of $p_{j+1}, \ldots, p_d$, so g is a random $(d-j)$-dimensional tower over a random sequence of points. Thus, without loss of generality, we may assume $j = 0$.

Fix $p, q \in \mathcal{P}$ and let E denote the event that $f_d(x) = p$ and $f_d(y) = q$. Since $f_d(x)$ and $f_d(y)$ are uniformly distributed over $\mathcal{P}$ by Lemma 20, we need to show that $P[E] = 1/v^2$. We now proceed by induction on d. For the base case, suppose $d = 1$. If $p = q$, then E occurs if and only if $p_0 = p_1 = p$, so $P[E] = 1/v^2$. If $p \neq q$ then a necessary condition for E is that $p, q \in \ell(p_0, p_1)$, and under this condition the probability of E is $1/((k-1)(k-2))$. To ensure $p, q \in \ell(p_0, p_1)$, we must have $p_0, p_1 \in \overline{pq}$, and furthermore, we must have $p_1 \neq p_0, p, q$. Thus, assuming $p_0 \in \overline{pq}$, if $p_0 \neq p, q$ then there are $k-3$ possibilities for p_1, and otherwise there are $k-2$ possibilities for p_1. Altogether, we have

$$P[E] = \frac{1}{(k-1)(k-2)} \left(\frac{k-2}{v} \cdot \frac{k-3}{v} + \frac{2}{v} \cdot \frac{k-2}{v} \right) = \frac{1}{v^2}.$$

Now for $d > 1$, the inductive hypothesis says that $f_{d-1}(x)$ and $f_{d-1}(y)$ are independent. Fix p_d. The random, independent points $f_{d-1}(x)$ and $f_{d-1}(y)$ determine two random, independent truncated lines through p_d, and the points $f_d(x)$ and $f_d(y)$ are random points from these truncated lines. Thus, after fixing p_d, we have $f_d(x)$ and $f_d(y)$ independent, and so they are independent overall. □

For any $p \in \mathcal{P}$ and $x \in [k-1]^d$, let $\vartheta_p(x)$ be the indicator variable for the event $f_d(x) = p$. We note that for every fixed p and x we have $\mathbb{E}(\vartheta_p(x)) = 1/v$ by Lemma 20. Furthermore, by Lemma 21, for every $p \in \mathcal{P}$ the set $\{\vartheta_p(x) : x \in [k-1]^d\}$ of variables is pairwise independent. Let X_p denote the number of labels of p, so

$$X_p = \sum_{x \in [k-1]^d} \vartheta_p(x). \tag{5}$$

We have the following estimate of the number of labels f gives to points of $\mathcal{P}$.

COROLLARY 22. *For any $p \in \mathcal{P}$, we have $\mathrm{Var}(X_p) < \mathbb{E}(X_p) = (k-1)^d/v$.*

PROOF. The statement $\mathbb{E}(X_p) = (k-1)^d/v$ is immediate from Lemma 20. By the pairwise independence of the $\vartheta_p(x)$,

$$\begin{aligned}\mathrm{Var}(X_p) &= \sum_{x \in [k-1]^d} \mathrm{Var}(\vartheta_p(x)) + \sum_{x \neq y \in [k-1]^d} \mathrm{Cov}(\vartheta_p(x), \vartheta_p(y)) \\ &= \sum_{x \in [k-1]^d} (1/v)(1-1/v) < \mathbb{E}(X_p) . \quad \square\end{aligned}$$

The next corollary is used in the proof of Lemma 11 to show that a good fraction of the space gets unique colors.

Corollary 23. *For all $\beta > 0$ we have*

$$P\left[(\exists p \in \mathcal{P})\left(\left|X_p - \frac{(k-1)^d}{v}\right| \geq \beta(k-1)^{d/2}\right)\right] < 1/\beta^2.$$

Proof. Noting that for every p, $\mathrm{Var}(X_p) < \mathbb{E}(X_p) = (k-1)^d/v$, by Chebyshev we obtain $P(|X_p - (k-1)^d/v| \geq \beta\sqrt{v}\sqrt{(k-1)^d/v}) < 1/(v\beta^2)$. The conclusion follows by the union bound. □

Corollary 24. *The expected number of points which get at least one label is*

$$\mathbb{E}(|\mathrm{range}(f_d)|) \geq (k-1)^d - (k-1)^{2d}/(2v).$$

Proof. Fix a point p. By the inclusion-exclusion principle and Lemmas 20 and 21,

$$\begin{aligned} P[p \in \mathrm{range}(f_d)] \geq & \sum_{x\in[k-1]^d} P[f_d(x) = p] \\ & - \sum_{x\neq y\in[k-1]^d} P[f_d(x) = f_d(y) = p] \\ & = \frac{(k-1)^d}{v} - \binom{(k-1)^d}{2}\frac{1}{v^2}. \end{aligned}$$

Thus, letting $\eta(p)$ denote the indicator variable for the event $p \in \mathrm{range}(f_d)$, we have

$$\mathbb{E}(|\mathrm{range}(f_d)|) = \sum_{p\in\mathcal{P}} \mathbb{E}(\eta(p)) \geq (k-1)^d - \frac{(k-1)^{2d}}{2v}. \quad \Box$$

6. SIMPLE POINTS

For a truncated line $\ell = \ell(x, y)$ and a coloring γ, we say that p is a *simple point* of ℓ with respect to γ if p is the only point of its color on ℓ. Note in particular that the single point of a degenerate line is always simple. A point $p \in \ell$ is a *multiple point* of ℓ if it is not simple. We note that if a line has a unique color, the next refinement step assigns unique colors to each of the simple points of the line.

We claim that if the multiplicity of each color is much smaller than R then, after stabilization, most points of most lines will be simple. Here is a formal statement.

Proposition 25. *Assume the multiplicity of each color is not greater than $1 + \delta R$. Then for a random line ℓ the expected number of multiple points of ℓ is $\leq \delta k$. It follows that for a random line ℓ, the probability that ℓ has more than $\sqrt{\delta}k$ multiple points is less than $\sqrt{\delta}$.*

Proof. Let the maximum color multiplicity be m. Then for a point p, the number of lines of which p is a multiple point is at most $m - 1$. So the total number of pairs (p, ℓ) such that p is a multiple point of ℓ is at most $v(m-1)$. These incidences are distributed among the $(v/k)R$ lines, so the average number of multiple points per line is at most $(m-1)k/R \leq \delta k$. The last statement follows by Markov. □

For the remainder of this section, we fix a d-dimensional random tower f over a random sequence of points $(p_0, \ldots, p_d)$.

Corollary 26. *Fix $x \in [k-1]^d$. For $1 \leq j \leq d$, the probability that $f_j(x)$ is not a simple point of $\ell(f_{j-1}(k), p_j)$ is at most $2\sqrt{\delta}$.*

Proof. If $f_{j-1}(x) = p_j$, then $f_j(x)$ is a simple point of $\ell(f_{j-1}(x), p_j)$. Otherwise, since $f_{j-1}(x)$ and p_j are distinct random points by Lemma 20, they determine a random line ℓ. By Proposition 25, the probability that ℓ has at least $\sqrt{\delta}k$ multiple points is at most $\sqrt{\delta}$. If ℓ has fewer than $\sqrt{\delta}k$ multiple points, then since p_j and $f_{j-1}(x)$ are random points of ℓ, so is $f_j(x)$, so the probability that $f_j(x)$ is a multiple point of $\ell(f_{j-1}(x), p_j)$ is at most $\sqrt{\delta}$. Thus, the overall probability that $f_j(x)$ is not a simple point is at most $2\sqrt{\delta}$. □

For a label $x \in [k-1]^d$, define the *tower simplicity* $s(x)$ of x to be the number of indices $1 \leq j \leq d$ such that $f_j(x)$ is a simple point of $\ell(f_{j-1}(x), p_j)$.

Corollary 27. *Assume the multiplicity of each point-color is not greater than $1 + \delta R$, and fix $x \in [k-1]^d$. Then $\mathbb{E}(s(x)) \geq (1 - 2\sqrt{\delta})d$.*

Proof. For $1 \leq j \leq d$, let X_j be the indicator variable for the event that $f_j(x)$ is a simple point of $\ell(f_{j-1}(x), p_j)$. Thus, $s(x) = \sum_{j=1}^d X_j$, and so the proposition follows from Corollary 26. □

Proposition 28. *Fix $x \in [k-1]^d$. Then after individualizing the points $p_0, \ldots, p_d$ and refining to a stable coloring, the color class containing $f_d(x)$ has at most $(k-1)^{d-s(x)}$ points.*

Proof. By induction on d, with the base case $d = 0$ clear since p_0 is individualized. Let A be the color class containing $f_{d-1}(x)$ and B the color class containing $f_d(x)$ after individualizing $p_1, \ldots, p_{d-1}$ and refining to a stable coloring (but before individualizing p_d). If $f_d(x) = p_d$, then $f_d(x)$ gets a unique color after individualizing p_d. More generally, if $f_d(x)$ is a simple point of $\ell(f_{d-1}(x), p_d)$, then after individualizing p_d and refining to a stable coloring, a point p gets the same color as $f_d(x)$ only if (i) $p \in \ell(q, p_d)$ for some $q \in A$; (ii) p is a simple point of $\ell(q, p_d)$; and (iii) $p \in B$. Thus, each of the at most $|A|$ truncated lines $\ell(q, p_d)$ with $q \in A$ contributes at most one point to the color class containing $f_d(x)$ in the stable refinement, so the number of points in that class is at most

$$|A| \leq (k-1)^{d-1-s(x^{d-1})} = (k-1)^{d-s(x)}.$$

On the other hand, if $f_d(x)$ is not a simple point of the truncated line $\ell(f_{d-1}(x), p_d)$, then since $f_d(x) \in C_{p_d}(A)$ and $C_{p_d}(A)$ is a stable set with $|C_{p_d}(A)| \leq (k-1)|A|$, it follows that the color class containing $f_d(x)$ in the stable refinement has size at most $(k-1)^{d-s(x)}$, as desired. □

Lemma 29. *Let $\alpha > 1$, $\delta > 0$, and define $\varepsilon = (1 + 64(\alpha-1)\sqrt{\delta})/\alpha$. Suppose $v > (k-1)^\alpha$, the multiplicity of each point-color is not greater than $1 + \delta R$, and $\varepsilon < (\alpha-1)/\alpha$. Then for some $d = O(\log v)$, depth-d stabilization achieves maximum point-color multiplicity $O(v^\varepsilon)$.*

Proof. Let d be such that

$$\frac{v}{k-1} < (k-1)^d \leq v,$$

and let f_d be a random d-dimensional tower over a random sequence of points $\sigma = (p_0, \ldots, p_d)$. Define $T = \mathrm{range}(f_d)$. Then by Corollary 24, we have $\mathbb{E}(|T|) > (k-1)^d/2$. Since

$|T| \leq (k-1)^d$ for any d-dimensional tower, by Markov we have

$$P\left[|T| \leq \frac{(k-1)^d}{4}\right] \leq P\left[(k-1)^d - |T| \geq \frac{3(k-1)^d}{4}\right] \leq \frac{2}{3}\,.$$

Fix $x \in [k-1]^d$. By Corollary 27, we have $\mathbb{E}_{\sigma,f_d}(s(x)) \geq (1-2\sqrt{\delta})d$. Thus, since $s(x) \leq d$, by Markov the probability over σ that $\mathbb{E}_{f_d}(d-s(x)) \geq 8\sqrt{\delta}d$ is at most $1/4$. Therefore, with probability at least $1-2/3-1/4 = 1/12$, for a random sequence σ of d points, a random tower over σ has $|T| > (1/4)(k-1)^d$ and $\mathbb{E}_{f_d}(d-s(x)) < 8\sqrt{\delta}$. Let σ be such a sequence, and individualize the points of the sequence and refine to a stable coloring. Again by Markov, with probability at least $7/8$, we have $d-s(x) < 64\sqrt{\delta}d$. Since f_d is a random tower, it follows that $7/8$ of the labels $x \in [k-1]^d$ have $d-s(x) < 64\sqrt{\delta}d$. Now for a point $p \in T$, by Proposition 28, there are at most $(k-1)^{d-s}$ points in the color class containing p, where $s = \max\{s(x) : f_d(x) = p\}$. Thus, if p belongs to a color class of size more than $(k-1)^{64\sqrt{\delta}d}$, then p has a label $x \in [k-1]^d$ with $d-s(x) \geq 64\sqrt{\delta}$; hence, at most $(1/8)(k-1)^d$ points $p \in T$ belong to color classes of size more than $(k-1)^{64\sqrt{\delta}d}$ in the stable refinement. Therefore, there is a stable subset A of T with $|A| \geq (1/8)(k-1)^d$ in which the maximum point-color multiplicity is $m \leq (k-1)^{64\sqrt{\delta}d}$. It follows that

$$\frac{|A|}{m} \geq \frac{1}{8}(k-1)^{(1-64\sqrt{\delta})d} \geq \frac{1}{8}\left(\frac{v}{k-1}\right)^{1-64\sqrt{\delta}}$$
$$\geq \frac{1}{8}v^{\frac{\alpha-1}{\alpha}(1-64\sqrt{\delta})}\,.$$

Now by Lemma 7, by individualizing $O(\log v/\log k)$ additional points and refining to a stable coloring, we obtain a stable set A' with $|A'| \geq v/(4(k-1))$ and maximum point-color multiplicity m' satisfying $|A'|/m' \geq |A|/(4m)$. Hence, since $|A'| \leq v$, we have

$$m' \leq 32v^{1-\frac{\alpha-1}{\alpha}(1-64\sqrt{\delta})} = 32v^{\varepsilon}\,.$$

Since $v^{\varepsilon} = o(v^{(\alpha-1)/\alpha}) = o(R)$, then by Lemma 16, for some $d = O(\log v)$, depth-d stabilization achieves maximum point-color multiplicity $O(m') = O(v^{\varepsilon})$. □

Proof of Lemma 10. When $v \leq (k-1)^4$, Target 2 already gives Target 3, since $(\log(k-1)/\log v)^2 \geq 1/16$. Thus, we assume $v > (k-1)^4$ and the maximum point-color multiplicity in $\mathfrak{X}$ is at most $2^{-18}R$.

By Lemma 29, for $\varepsilon = 11/32$ and for some $d' = O(\log v)$, depth-d' stabilization achieves maximum point-color multiplicity $O(v^{\varepsilon})$. Since $v^{\varepsilon} = o((\log(k-1)/\log v)^2 R)$, the lemma follows. □

7. COMPLETE SPLITTING

We continue to fix a random d-dimensional tower f over a random sequence of points $(p_0, \ldots, p_d)$. Let us say that the label $x \in [k-1]^d$ is *simple* if $f_j(x^j)$ is a simple point of $\ell(f_{j-1}(x), p_j)$ for every j $(1 \leq j \leq d)$, i. e., if $s(x) = d$. The following two observations follow immediately from Proposition 28 and Corollary 26, respectively.

Observation 30. *All points with at least one simple label from f receive unique colors after individualizing $p_0, \ldots, p_d$ and a refining to a stable coloring.*

Observation 31. *Assume the multiplicity of each point-color is not greater than $1+\delta R$, and fix $x \in [k-1]^d$. Then the probability that x becomes a simple label is at least $1-2\sqrt{\delta}d$.*

Lemma 32. *If more than $(k-2)+R$ points of the space have unique colors, then the next two rounds of refinement assigns unique colors to all points.*

Proof. Let B be the set of points with unique colors and let $p \in \mathcal{P} \setminus B$. If for a line ℓ we have $|B \cap \ell| \geq 2$ then ℓ will receive a unique color; and if two such lines both contain p then p is will receive a unique color in the subsequent round. So if in the end the point p did not receive a unique color then at most one line containing p has more than one point in B, therefore $|B| \leq (k-1)+R-1$, where the first term accounts for the points in B on the possible exceptional line; the number of remaining lines containing p is $R-1$. □

Proof of Lemma 11. Let d be minimal such that $(k-1)^{d/2} \geq 4v$, so in particular $d < 4\log v/\log(k-1)$. Assume the multiplicity of each point-color is not greater than $2^{-14}(\log(k-1)/\log v)^2 R$. Let f be a random d-dimensional tower over a random sequence of points $(p_0, \ldots, p_d)$.

By Observation 31, the expected proportion of labels that are not simple is at most $2^{-6}(\log(k-1)/\log v)d < 1/16$, so by Markov, with probability at least $1/2$, the actual proportion is at most $1/8$.

Setting $\beta = 2$ in Corollary 23, we see that with probability at least $3/4$, the labels are distributed nearly uniformly over $\mathcal{P}$ in the sense that each point has at least half its expected number $(k-1)^d/v$ of labels, since $2(k-1)^{d/2} \leq (k-1)^d/(2v)$.

Thus, with probability at least $1/4$, both of the above events occur. Let $(p_0, \ldots, p_d)$ be a sequence of points such that for a random tower f over $(p_0, \ldots, p_d)$, at most $1/8$ of the labels are not simple, and every point in $\mathcal{P}$ gets at least $(k-1)^d/(2v)$ labels. Then the number of points which do not get simple labels is at most $(1/8)(k-1)^d/((k-1)^d/(2v)) < v/4$. Since at least $3/4$ of the points get at least one simple point, at least this number get unique colors after individualizing $p_0, \ldots, p_d$ and refining to a stable coloring. Thus, by Lemma 32, the system is split completely. □

8. EXTENSION TO STEINER t-DESIGNS

We now generalize Theorem 6 to Steiner t-designs. Define $n = |\mathcal{L}|$. Recall that a *nontrivial* $S(t,k,v)$ has parameters $t < k < v$.

Theorem 33. *Let $\mathfrak{X} = (\mathcal{P}, \mathcal{L})$ be a nontrivial $S(t,k,v)$. There exists a set T of cardinality $O(\log n)$ that splits $\mathfrak{X}$ completely.*

We note that the bound on $|T|$ is logarithmic in the input length, conservatively estimated from below by $n = |\mathcal{L}|$.

Theorem 33 immediately implies Theorems 1, 3, and 4.

The proof of Theorem 33 is by reduction to the $t = 2$ case.

Let $\mathfrak{X} = (\mathcal{P}, \mathcal{L})$ be an $S(t,k,v)$ Steiner t-design. Let $A \subset \mathcal{P}$ such that $|A| \leq t-2$. The pair $\mathfrak{X}_A = (\mathcal{P}_A, \mathcal{L}_A)$, where $\mathcal{P}_A = \mathcal{P} \setminus A$ and $\mathcal{L}_A = \{\ell \setminus A \mid \ell \in \mathcal{L}, A \subset \ell\}$, is an $S(t-|A|, k-|A|, v-|A|)$ design, called the *derived design* at A.

Proof of Theorem 33. We pick any $A \subset V$ such that $|A| = t-2$ and consider the derived Steiner 2-design $\mathfrak{X}_A$. By Theorem 6, there exists a set $S \subset V \setminus A$ of size $O(\log v)$ such that S completely splits $\mathfrak{X}_A$. Let $T = A \cup S$. Then T completely splits $\mathfrak{X}$. We have $v \leq n$ by Fisher's inequality, so to complete the proof we need to show that $t = O(\log n)$. Now $n \geq 2^{\lfloor t/2 \rfloor}$ is known; it can be deduced from the results of [13]. (In fact we can show $n \geq 2^t$.) □

9. RECONSTRUCTION

We conclude this paper by returning to the problem of isomorphism testing for s. r. graphs. Line-graphs of Steiner 2-designs are an important special case of the problem, as we noted in Section 1.2. In order to apply Theorem 1 in this context, it is necessary to first reconstruct a Steiner 2-design from its line-graph.

However, the *unique* reconstruction of a Steiner 2-design from its line graph is not always possible; in particular, there exist non-isomorphic finite projective planes of the same order, but the line-graph of every finite projective plane is the complete graph. In fact, while $k \sim v^{1/2}$ for finite projective planes, unique reconstructability fails already for $k \sim v^{1/3}$. We overcome the non-uniqueness obstacle by borrowing the basic idea of "list decoding" in the theory of error-correcting codes: we show that the number of reconstructions can be controlled and produce a moderate-length list that includes all reconstructions. Theorem 2 is a consequence.

We sketch the idea. The details are left to the full paper.

Definition 34. Given a regular graph G, a *reconstruction system* $\mathcal{C}$ is a collection of cliques in G such that for some positive numbers v, k, the following hold:

- every vertex of G belongs to k cliques in $\mathcal{C}$
- every pair of cliques in $\mathcal{C}$ shares exactly one vertex.

A regular graph G is the line-graph of a Steiner 2-design if and only if there exists a reconstruction system $\mathcal{C}$ in G.

Say a clique C is *feasible* if $C \in \mathcal{C}$ for some reconstruction system $\mathcal{C}$. A set A is a *seed* of C if C consists of A and all the common neighbors of the set A. Assume now that G is the line graph of an $S(2, k, v)$ design $\mathfrak{X} = (\mathcal{P}, \mathcal{L})$, and let $\delta = k(k-1)/(v-1) = v/n$.

LEMMA 35. *If G is not complete, then each feasible clique has a seed of size $\alpha = \lceil 1 - \log v / \log \delta \rceil$.*

PROOF. Fix $p \in \mathcal{P}$ and $\ell \in \mathcal{L}$ such that $p \notin \ell$. Let $s \geq 1$. Let $S = \{\ell_1, \dots, \ell_s\}$ be a collection of lines through p chosen independently at random. The probability that ℓ intersects each ℓ_i is δ^s. The result follows by the union bound. □

As a consequence of Lemma 35, we can produce a list containing all feasible cliques (and possibly some additional cliques) by considering every subset of $V(G)$ of size α.

THEOREM 36. *Let $\varepsilon > 0$ and let G be a regular graph on n vertices with valency $\rho = O(n^{1/(1+\varepsilon)})$. The number of reconstruction systems of G is $\leq \exp(O(\rho^2 n^{-3/2} \log^2 n/\varepsilon))$ and we can find them all in $\exp(O(\log n + \rho^2 n^{-3/2} \log^2 n/\varepsilon))$ time.*

We outline the proof. To prove Theorem 36 from Lemma 35, we observe that if C is the set of lines through a point in a Steiner 2-design $\mathfrak{X}$, then every other point appears on exactly one line of C. On the other hand, if C is a set of lines of $\mathfrak{X}$ which do not all pass through a point, but which does appear in the list of feasible cliques we obtain from Lemma 35, then with probability bounded away from zero (in terms of ρ), a random point of $\mathfrak{X}$ does not lie on any line of C. Theorem 36 then follows by considering all subsets of our list of feasible cliques of size $O(\rho^2 n^{-3/2} \log n)$.

Finally, Theorem 2 follows from Theorem 36: for line-graphs with valency $\geq n^{5/6}$, we use eq. (1), and for line-graphs with smaller valency, by Theorem 36, we can find every reconstruction system in time $\exp(O(n^{1/6} \log^2 n))$, compute a canonical form for each resulting Steiner 2-design in quasipolynomial time, and take the lexicographically least form. □

10. REFERENCES

[1] L. Babai. On the complexity of canonical labeling of strongly regular graphs. *SIAM J. Comput.*, 9(1):212–216, 1980.

[2] L. Babai. On the order of uniprimitive permutation groups. *Annals of Math.*, 113(3):553–568, 1981.

[3] L. Babai, W. M. Kantor, and E. M. Luks. Computational complexity and the classification of finite simple groups. In: *24th FOCS*, pages 162–171, 1983.

[4] L. Babai and E. M. Luks. Canonical labeling of graphs. In: *15th STOC*, pages 171–183, 1983.

[5] R. C. Bose. Combinatorial problems of experimental design. I. Incomplete block designs. *Proc. Sympos. Pure Math.*, 34:47–68, 1978.

[6] J.-Y. Cai, M. Fürer, and N. Immerman. An optimal lower bound on the number of variables for graph identification. *Combinatorica*, 12(4):389–410, 1992.

[7] X. Chen, X. Sun, and S.-H. Teng. Multi-stage design for quasipolynomial-time isomorphism testing of Steiner 2-systems. In: *45th STOC*, 2013.

[8] C. J. Colbourn and P. C. van Oorschot. Applications of combinatorial designs in computer science. *ACM Comput. Surv.*, 21(2):223–250, 1989.

[9] M. Huber. Computational complexity of reconstruction and isomorphism testing for designs and line graphs. *J. Comb. Theory A*, 118(2):341–349, 2011.

[10] E. M. Luks. Isomorphism of graphs of bounded valence can be tested in polynomial time. *J. Comput. Syst. Sci.*, 25(1):42–65, 1982.

[11] G. L. Miller. On the $n^{\log n}$ isomorphism technique: A preliminary report. In: *10th STOC*, pages 51–58, 1978.

[12] A. Neumaier. Strongly regular graphs with smallest eigenvalue $-m$. *Arch. Math.*, 33(4):392–400, 1979.

[13] D. K. Ray-Chaudhuri and R. M. Wilson. On t-designs. *Osaka J. Math.*, 12(3):737–744, 1975.

[14] R. C. Read and D. G. Corneil. The graph isomorphism disease. *J. Graph Theory*, 1(4):339–363, 1977.

[15] D. A. Spielman. Faster isomorphism testing of strongly regular graphs. In: *28th STOC*, pages 576–584, 1996.

[16] J. H. van Lint and R. M. Wilson. *A Course in Combinatorics.* Cambridge U. Press, 2nd ed., 2001.

[17] B. Weisfeiler, editor. *On construction and identification of graphs*, volume 558 of *Lecture Notes in Mathematics.* Springer-Verlag, 1976.

[18] V. N. Zemlyachenko, N. M. Korneenko, and R. I. Tyshkevich. Graph isomorphism problem. *Zapiski Nauchnykh Seminarov LOMI*, 118:83–158, 215, 1982.