Electronic Colloquium on Computational Complexity, Report No. 291 (2026)

Reed–Muller Codes are Optimal Binary Linear 3-Query LCCs

Oliver Janzer
oliver.janzer@epfl.ch
EPFL

Peter Manohar*
pmanohar@ias.edu
The Institute for Advanced Study

November 2, 2026

Abstract

We prove that any linear 3-query locally correctable code $C\colon \mathbb{F}_2^k \to \mathbb{F}_2^n$ with constant correction radius must have length $n \geq 2^{\Omega(\sqrt{k})}$. This improves on the prior best lower bound of $2^{\Omega(\sqrt{k/\log k})}$, due to [AG24], and shows that the folklore construction of linear 3-LCCs from Reed–Muller codes, which achieve length $n = 2^{O(\sqrt{k})}$, is optimal up to constant factors in the exponent.

Our proof builds on the approach of [KM24a], further refined in [Yan24, KM24b], that uses "chain derived equations" in combination with Kikuchi graphs. The crux of our argument is to show that a certain edge-coloured Kikuchi graph has a unique colour cycle: a cycle in which some colour is used exactly once. We do this by controlling the distribution of colours at each vertex in the graph using a random walk mixing argument and a novel hypergraph decomposition based on edge expansion of the hypergraph's "flattening" into a graph.

*Supported by NSF Grant DMS-2424441.

ISSN 1433-8092

Contents

1 Introduction **1**

2 Proof overview **2**
2.1 The prior approach of [KM24a, Yan24, KM24b] 3
2.2 A proof sketch of Theorem 1 5

3 Preliminaries **8**

4 Proof of Theorem 1 **10**
4.1 Proof of Lemma 4.2 12

5 The Proof Assuming Edge Expansion: Proof of Lemma 4.1 **13**
5.1 Proof of Lemma 5.3 17
5.2 Proof of Lemma 5.7 18

Statement of AI Use **20**

References **21**

1 Introduction

Locally correctable codes (LCCs) are error-correcting codes that are additionally equipped with a local self-corrector: an algorithm that, when given oracle access to a corrupted codeword, is capable of recovering any symbol of the underlying uncorrupted codeword with good probability while only reading a small number of randomly chosen symbols from the corrupted codeword. Such codes are formally defined as follows.

Definition 1.1 (Locally correctable codes)**.** A code $C\colon \{0,1\}^k \to \{0,1\}^n$ is a (q, δ, ε)-LCC if there is a randomized decoding algorithm $\mathrm{Dec}(\cdot)$ with the following properties. The algorithm $\mathrm{Dec}(\cdot)$ is given oracle access to a string $y \in \{0,1\}^n$, takes an index $u \in [n]$ as input, and outputs an element of $\{0,1\}$ with the following guarantees:

(1) for any y and u, the algorithm $\mathrm{Dec}^y(u)$ reads at most q indices of y,

(2) for all $b \in \{0,1\}^k$, $y \in \{0,1\}^n$ with $\Delta(y, C(b)) \le \delta n$, and $u \in [n]$, $\Pr[\mathrm{Dec}^y(u) = C(b)_u] \ge 1 - \varepsilon$. Here, $\Delta(x, y) \coloneqq |\{i : x_i \neq y_i\}|$ denotes the Hamming distance between x and y.

In this definition, δ and ε are typically taken to be small constants. We write q-LCC to refer to a (q, δ, ε)-LCC where δ, ε are constant.

Locally correctable codes are an important class of error-correcting codes that have numerous intriguing applications throughout theoretical computer science. They were first introduced implicitly in works on program checking [BK95], and were used in constructions of probabilistically checkable proofs (PCPs) and the proof of the PCP theorem [ALM+98, AS98]. Other applications (see the surveys of [Tre04, Yek12, Dvi12] for more details) include sublinear algorithms and property testing [RS96, BLR93], IP=PSPACE [LFKN90, Sha90], worst-case to average-case reductions [BFNW93], constructions of explicit rigid matrices [Dvi10], private information retrieval protocols [IK99, BIW10], connections to incidence geometry [Dvi12], additive combinatorics [BDL13], and the theory of block designs [BIW10].

Because of their importance, constructing better LCCs or understanding their inherent limitations has remained a significant goal in coding theory over the past three decades. The central question of study is to understand the optimal dependence of n, the length of the code, as a function both of k, the length of the underlying message, and q, the number of queries permitted by the decoder; $\delta, \varepsilon \in (0,1)$ are implicitly taken to be small constants. The folklore LCCs based on Reed–Muller codes, or low-degree multivariate polynomials, provide a useful baseline construction. In the regime where q is a constant, degree-$(q-1)$ Reed–Muller codes give a q-LCC $C\colon \{0,1\}^k \to \{0,1\}^n$ with $n = 2^{O(k^{1/(q-1)})}$. Moreover, this code is *linear*, meaning that the encoding map C is an $\mathbb{F}_2$-linear map. For super-constant q, we know that this baseline construction can be improved, as multiplicity codes [KSY11, KMRS16, GKO+17] give constructions of LCCs that have a much smaller blocklength n than Reed–Muller codes. However, for constant q, Reed–Muller codes remain the best-known LCCs, and finding better constructions remains an open question. In fact, the lack of progress on constructing better constant query LCCs has led to the conjecture that Reed–Muller codes may in fact be optimal constant query LCCs.

For the case of $q = 2$, this conjecture is known to be true. The Hadamard code, or equivalently Reed–Muller codes of degree 1, gives a 2-query LCC of length $n = 2^k$, and classic works on lower bounds for locally decodable codes[1] show that $n \geq 2^{\Omega(k)}$, initially for linear codes [GKST02], and later for nonlinear codes [KW04] as well. For the case of $q = 3$, however, the situation was much less understood until very recently. Reed–Muller codes of degree 2 give a 3-LCC with $n = 2^{O(\sqrt{k})}$, while the classic lower bound of [KW04] shows that $n \geq \tilde{\Omega}(k^2)$. Up to polylog$(n)$ factor improvements, this remained the best lower bound until [AGKM23] improved the lower bound to $n \geq \tilde{\Omega}(k^3)$. Then, the work of [KM24a] proved that $n \geq 2^{\Omega(k^{1/8})}$ for any linear 3-LCC, and this lower bound led to a flurry of subsequent improvements [Yan24, AG24, KM24b]. The current best bound for *linear* 3-LCCs is $n \geq 2^{\Omega(\sqrt{k/\log k})}$ due to [AG24], which is tantalizingly close to the $n = 2^{O(\sqrt{k})}$ achieved by Reed–Muller codes, and [KM24b] shows a tight bound of $n \geq 2^{\Omega(\sqrt{k})}$ assuming that the linear 3-LCC is a "block design", which is an additional structural property satisfied by Reed–Muller codes. For nonlinear codes, [KM24b] shows a weaker bound of $n \geq 2^{\Omega(k^{1/5})}$.

As the main result of this paper, we prove a linear 3-LCC lower bound of $n \geq 2^{\Omega(\sqrt{k})}$, showing that the folklore construction of Reed–Muller codes is optimal for linear 3-LCCs.

Theorem 1. *Let $C : \mathbb{F}_2^k \to \mathbb{F}_2^n$ be a linear $(3, \delta, \varepsilon)$-LCC for any $0 \leq \varepsilon < 1/2$. Then, $n \geq 2^{\Omega(\delta^7 \sqrt{k})}$.*

We note that we did not attempt to optimize the dependence on δ in Theorem 1.

Theorem 1 improves on the prior best lower bound of [AG24], which shows a bound of $n \geq 2^{\Omega(\sqrt{k/\log k})}$. However, as we shall discuss in Section 2, the proof of Theorem 1 follows a different approach compared to [AG24]. In [AG24], they reduce the task of proving a linear 3-LCC lower bound to the rainbow cycle problem in extremal combinatorics and then apply known bounds [ABS+25]. In comparison, the proof of Theorem 1 follows the "Kikuchi graph" approach of [KM24a], further refined in [Yan24, KM24b], and directly constructs a certain type of coloured cycle, which we term "unique colour cycle", in the Kikuchi graph. The proof of Theorem 1 introduces several key new ideas, which we will discuss next in Section 2.

2 Proof overview

In this section, we give an overview of the key ideas used to prove Theorem 1, and compare these ideas to the proofs in prior work [KM24a, Yan24, AG24, KM24b]. The proof in this paper, like the proofs in all these previous papers, relies on the following standard combinatorial definition of linear 3-LCCs, which is equivalent to Definition 1.1 up to constant factors [Dvi16]. For the purposes of this proof overview, we will mostly omit the dependence on δ in Theorem 1.

Fact 2.1 (Linear 3-LCC, combinatorial definition; Definition 2.3 and Fact 2.4 in [KM24a])**.** *Let $C : \mathbb{F}_2^k \to \mathbb{F}_2^n$ be a $(3, \delta, \varepsilon)$-locally correctable code where C is an $\mathbb{F}_2$-linear map and $\varepsilon < 1/2$. For each $u \in [n]$, let $\mathbf{v}_u$ denote the u-th row vector of the map C, and note that* $\dim(\mathrm{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}) = k$ *since C*

[1] Local decodability is a weaker notion than local correctability, and only requires that the decoder be capable of recovering the message bits, rather than all codeword bits. For linear codes, it is easy to see that any LCC is an LDC by choosing a systematic encoding map. For nonlinear codes, a similar statement is also true, see [BGT17, Appendix A].

is injective. For $u \in [2n] \setminus [n]$, let $\mathbf{v}_u := \mathbf{0}$. Then, for each $u \in [2n]$, there exists a 3-uniform hypergraph matching $\mathcal{H}_u \subseteq \binom{[2n]\setminus\{u\}}{3}$ with vertex set $[2n] \setminus \{u\}$ such that (1) $|\mathcal{H}_u| \geq \delta n/3$, i.e., the matching has at least $\delta n/3$ hyperedges, and (2) for every $u \in [2n]$ and every $\{u_1, u_2, u_3\} \in \mathcal{H}_u$, it holds that $\mathbf{v}_u = \mathbf{v}_{u_1} + \mathbf{v}_{u_2} + \mathbf{v}_{u_3}$.

Fact 2.1 implies that in order to prove Theorem 1, i.e., to show that every linear $(3, \delta, \varepsilon)$-locally correctable code of length n has $k \leq O_\delta(\log^2 n)$, it suffices to prove the following statement: for every set of n vectors $\{\mathbf{v}_u\}_{u\in[n]}$ in a vector space over $\mathbb{F}_2$ and every collection of hypergraph matchings $\{\mathcal{H}_u\}_{u\in[n]}$, each of size at least $\delta n/3$, where $\mathbf{v}_u = \mathbf{v}_{u_1} + \mathbf{v}_{u_2} + \mathbf{v}_{u_3}$ for all $u \in [n]$ and $\{u_1, u_2, u_3\} \in \mathcal{H}_u$, it holds that $\dim(\mathrm{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}) \leq O_\delta(\log^2 n)$.

The challenges in improving the analysis of [AG24]. The current best bound on $\dim(\mathrm{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_n\})$ is $O(\log^2 n \log\log n)$, shown by [AG24]. One natural route to proving Theorem 1 is thus to optimize the analysis of [AG24] and remove the $O(\log\log n)$ factor. However, as we will now discuss, this appears to be quite challenging, as it requires an analogous improvement on the rainbow cycle problem in extremal combinatorics.

The rainbow cycle problem asks the following question: suppose we are given a graph G with average degree d_{avg} that is *properly edge-coloured*, namely, each edge is assigned a colour and the colours on the edges adjacent to any single vertex are all distinct. How large does d_{avg} need to be to guarantee that G contains a *rainbow cycle*, a cycle that uses each colour at most once? The rainbow cycle problem has been studied extensively in extremal combinatorics [KMSV07, DLS13, Jan23, Tom24, JS24, KLLT24, ABS+25], and the current best bound is $d_{\mathrm{rbw}} \leq O(\log n \log\log n)$ [ABS+25], where d_{rbw} denotes the smallest bound on d_{avg} that guarantees the existence of a rainbow cycle. This is tight up to the $O(\log\log n)$ factor, as the Boolean hypercube (where each coordinate direction denotes a distinct colour) is a properly edge-coloured graph of degree $\log_2 n$ with no rainbow cycle.

Using a very clever reduction, the work of [AG24] shows that $\dim(\mathrm{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}) \leq d_{\mathrm{rbw}} \cdot O_\delta(\log n)$, and thus by [ABS+25] this implies that $k \leq O_\delta(\log^2 n \log\log n)$. To improve their analysis, one thus needs to remove the $O(\log\log n)$ factor from the rainbow cycle bound, which is a challenging task. Another natural approach would be to adapt the analysis of [AG24] to rely on a weaker combinatorial statement than rainbow cycle bounds, ideally one for which an $O(\log n)$ bound would be easier to show. However, the use of rainbow cycle bounds appears intrinsic to the approach of [AG24], so it is not clear how to do this either.

Our approach. Instead, to prove Theorem 1 we will follow the more "direct" approach used in the work [KM24a] and its refinements [Yan24, KM24b]. As we shall see, this will only require us to show the existence of an "odd coloured" cycle in a specific (not properly) edge-coloured Kikuchi graph obtained from the hypergraph matchings in Fact 2.1; an odd coloured cycle is a cycle that uses *some* colour an odd number of times. We will thus analyze this particular Kikuchi graph directly using its specific properties. Before we turn to the key ideas in our proof, we first summarize the ideas of the works [KM24a, Yan24, KM24b] in Section 2.1 below.

2.1 The prior approach of [KM24a, Yan24, KM24b]

The previous proofs of [KM24a, Yan24, KM24b] follow the same blueprint, introduced in [KM24a]. Firstly, rather than analyzing the system of constraints in Fact 2.1, they instead analyze a system of

derived constraints called "chain derivations" formed by applying a specific set of transformations to the original equations. Secondly, they form a Kikuchi graph using the derived constraints, and use this to construct a 2-LDC of length $N = n^{O(\text{polylog}(n))}$. Thus, by applying known 2-LDC lower bounds, this implies that $k \leq O(\log N) = O(\text{polylog}(n))$, which finishes the proof. Below, we explain each of these key steps in more detail.

Chain derived constraints and the bipartite Kikuchi graph. The key idea of [KM24a] is to analyze not the initial system of constraints in Fact 2.1, but a derived system of constraints formed as follows. For each constraint of the form $\mathbf{v}_u = \mathbf{v}_{u_1} + \mathbf{v}_{u_2} + \mathbf{v}_{u_3}$, we can pick another constraint of the form $\mathbf{v}_{u_3} = \mathbf{v}_{u_4} + \mathbf{v}_{u_5} + \mathbf{v}_{u_6}$ using the hypergraph matching $\mathcal{H}_{u_3}$. By combining these two constraints, we get the derived constraint $\mathbf{v}_u = \mathbf{v}_{u_1} + \mathbf{v}_{u_2} + \mathbf{v}_{u_4} + \mathbf{v}_{u_5} + \mathbf{v}_{u_6}$. The definition of r-chains given below captures the set of constraints formed by repeating this process r times.

Definition 2.2 (Chain derived constraints)**.** Let $r \geq 1$ be a positive integer. An r-chain with *head* u is a sequence $C = (u_0, a_1, b_1, u_1, a_2, b_2, u_2, \ldots, u_{r-1}, a_r, b_r, u_r)$ where (1) $u_i, a_i, b_i \in [n]$ for all choices of i, (2) $u_0 = u$, and (3) for all $i = 1, \ldots, r$, we have $\{a_i, b_i, u_i\} \in \mathcal{H}_{u_{i-1}}$. We let $\mathcal{H}_u^{(r)}$ denote the set of r-chains with head u.

Given a chain C with head u, we associate it with the vector $\mathbf{v}_C \coloneqq \mathbf{v}_{u_r} + \sum_{i=1}^r \mathbf{v}_{a_i} + \sum_{i=1}^r \mathbf{v}_{b_i}$. Note that this does not include the vector $\mathbf{v}_u$ for the head of the chain.

It is straightforward to observe that if all the constraints defined in Fact 2.1 are satisfied, then it holds that $\mathbf{v}_C = \mathbf{v}_u$ for any r-chain C with head u.

The next step is to define the following Kikuchi graph.

Definition 2.3 (Bipartite chain Kikuchi graph)**.** Let ℓ be a parameter to be determined later. Let $V_L = \binom{[n]}{\ell} \times [n]$ and let $V_R = \binom{[n]}{\ell}$. For each element $u \in [n]$, the bipartite Kikuchi graph $\mathcal{K}_{r,\ell}(u, \mathcal{H})$ is defined as follows. The left vertex set is V_L and the right vertex set is V_R. For each r-chain C with head u, and each $(S, u') \in V_L$ and $T \in V_R$, there is an edge $((S, u'), T)$ "labeled" by the r-chain C if $S = \{a_1, \ldots, a_r\} \cup R$ and $T = \{b_1, \ldots, b_r\} \cup R$ for some $R \subseteq [n]$ of size $|R| = \ell - r$ disjoint from both $\{a_1, \ldots, a_r\}$ and $\{b_1, \ldots, b_r\}$, and $u' = u_r$ in the chain C.

Note that in the above definition, $\{a_1, \ldots, a_r\}$ must be a set of size r in order for S to have size ℓ. This implicitly requires the chains to have *distinct* $a_1, \ldots, a_r$ (and similarly for $b_1, \ldots, b_r$). This is a minor technicality of Definition 2.3 that does not materially affect the proof in a major way. We note that the Kikuchi graph that we work with when proving Theorem 1 (Definition 5.1) does not have this issue.

The Kikuchi graph is motivated by the following observation. For each set $T \subseteq [n]$, we can associate it with the vector $\mathbf{v}_T \coloneqq \sum_{u \in T} \mathbf{v}_u$, and similarly for each left vertex (S, u'), we associate it with the vector $\mathbf{v}_{(S,u')} \coloneqq \mathbf{v}_S + \mathbf{v}_{u'}$. If we do this, then we observe that for any edge $((S, u'), T)$ in the graph $\mathcal{K}_{r,\ell}(u, \mathcal{H})$, it holds that $\mathbf{v}_{(S,u')} + \mathbf{v}_T = \mathbf{v}_C = \mathbf{v}_u$. The Kikuchi graph thus gives a method for converting the larger arity constraints on the vertex set $[n]$ into arity 2 constraints on the larger vertex sets V_L and V_R.

Extracting a 2-LDC from the Kikuchi graph. To finish the proof, [KM24a, Yan24, KM24b] then show that the $\mathbb{F}_2$-linear encoding map C' given by $x \in \mathbb{F}_2^k \mapsto \{\langle x, \mathbf{v}_S + \mathbf{v}_{u'} \rangle\}_{(S,u') \in V_L} \cup \{\langle x, \mathbf{v}_T \rangle\}_{T \in V_R}$

gives a 2-LDC. This follows by showing that each $\mathcal{K}_{r,\ell}(u,\mathcal{H})$ has a matching of size $\Omega(N)$, where $N = |V_L|$. (Technically, this is not possible since the left and right vertex sets are imbalanced as $|V_L| = |V_R| \cdot n$. However, one can easily overcome this issue by duplicating each right vertex n times and splitting its edges roughly equally across all copies. We ignore this technicality for the proof overview.) It is easy to show that if $\mathcal{K}_{r,\ell}(u,\mathcal{H})$ is d-regular, then it has a matching of size $e(\mathcal{K}_{r,\ell}(u,\mathcal{H}))/2d$, constructed via a simple greedy algorithm, and a similar statement holds if $\mathcal{K}_{r,\ell}(u,\mathcal{H})$ is approximately regular. Hence, the core technical component of all the works [KM24a, Yan24, KM24b] is to show that $\mathcal{K}_{r,\ell}(u,\mathcal{H})$ contains a subgraph $\mathcal{K}'$ that contains nearly all of the edges of $\mathcal{K}_{r,\ell}(u,\mathcal{H})$ and is approximately regular in the sense that $d_{\max}(\mathcal{K}') \leq O(d_{\text{avg}}(\mathcal{K}'))$. This step, sometimes called the "row pruning" step because it "prunes" the rows of the adjacency matrix of the Kikuchi graph, is typically the core technical piece in works using Kikuchi matrices [GKM22, HKM23, AGKM23, BHKL25, JM25].

This core technical component is where the proofs in [KM24a, Yan24, KM24b] diverge. The proof in [KM24b] shows that almost all vertices in $\mathcal{K}_{r,\ell}(u,\mathcal{H})$ have degree $(1+o(1))d_{\text{avg}}$ for some $r \leq O(\log n)$ and $\ell = 2r$, under the assumption that the 3-LCC is a *design LCC*. A design 3-LCC is defined by a 4-uniform hypergraph $\mathcal{H} \subseteq \binom{[n]}{4}$ where each pair $\{w,w'\} \subseteq [n]$ appears in exactly one hyperedge in $\mathcal{H}$, and the 3-uniform hypergraph matchings $\mathcal{H}_u$ are formed by taking all hyperedges in $\mathcal{H}$ that contain u, and then removing u. Note that this means that each $\mathcal{H}_u$ is a *perfect* hypergraph matching on $[n] \setminus \{u\}$. The calculation in [KM24b] showing this $(1+o(1))d_{\text{avg}}$ bound is very precise, relying on sharp bounds on binomial coefficients, and crucially uses that the hypergraph matchings $\mathcal{H}_u$ are perfect matchings and each pair $\{w,w'\}$ appears in exactly one hyperedge. Applying known 2-LDC lower bounds [GKST02] shows that $k \leq O(\log N)$, where $N = |V_L| + |V_R| = O(n^{\ell+1})$. In this case, we have $\ell = O(r) = O(\log n)$, so this implies that $k \leq O(\log^2 n)$.

However, a 3-LCC need not be a design LCC in general, as (1) the hypergraph matchings $\mathcal{H}_u$ need not be perfect, and (2) there may be "heavy pairs" of elements $\{w,w'\}$ that appear in many hyperedges across all the $\mathcal{H}_u$'s. As explained in [KM24b], the analysis of [Yan24] loses an additional $\log n$ factor to handle each of these two irregularities, resulting in a final bound of $k \leq O(\log^4 n)$. The imperfect matchings can cause large fluctuations in the degrees in the Kikuchi graph when $\ell \approx r$ because, e.g., one can form a 3-LCC with matchings of size δn by repeating a design 3-LCC $1/\delta$ times on disjoint blocks of coordinates. The standard deviation of the size of the intersection of S with each block then makes the degree of a random left vertex (S,u') in the Kikuchi graph vary by roughly a factor $(1+\Theta_\delta(\frac{1}{\sqrt{\ell}}))^r$. These fluctuations therefore become bounded by a constant only once one takes $\ell \approx r^2$, rather than the ideal value of $\ell \approx r$, which loses a $\log n$ factor.

The second $\log n$ factor loss comes from the issue of "heavy pairs" $\{w,w'\}$ that appear in many hyperedges across all the $\mathcal{H}_u$'s. This issue is handled by a complicated "chain decomposition" step that forms the other core technical component of [KM24a, Yan24], and is tailored to make the row pruning work.

2.2 A proof sketch of Theorem 1

We now explain the key ideas behind the proof of Theorem 1. From the approach outlined in Section 2.1, we will make use of the first two ideas, the chain derivations and (a slightly modified

version of) Kikuchi graphs. However, we will crucially replace the third component, the reduction to a 2-LDC, with a different combinatorial statement, which we will then prove using several new ideas. The full proof of Theorem 1 is fairly short, and so we will only sketch the main ideas here.

Finding odd-coloured cycles in the chain Kikuchi graph. Let $k = \dim(\text{span}\{\mathbf{v}_1, \dots, \mathbf{v}_n\})$ be the quantity that we wish to bound. Without loss of generality, we may assume that $\mathbf{v}_1, \dots, \mathbf{v}_k$ are linearly independent. We define a bipartite edge-coloured graph $\mathcal{K}_{r,\ell}([k], \mathcal{H})$ with colour set $[k]$ using the chain Kikuchi graphs from Definition 2.3, by simply taking the set of edges in $\mathcal{K}_{r,\ell}([k], \mathcal{H})$ with colour i to be the edges in $\mathcal{K}_{r,\ell}(i, \mathcal{H})$. For simplicity of notation, we will now let $\mathcal{K}$ be $\mathcal{K}_{r,\ell}([k], \mathcal{H})$. The following observation, which is somewhat implicit in prior work [BCG20, AGKM23, KM24a, Yan24, AG24, KM24b], is that to bound k, it suffices to show the existence of an *odd-coloured* cycle in $\mathcal{K}$ if the number of colours k is too large.

Observation 2.4. Let $\mathcal{K}$ be the edge-coloured graph defined above. Suppose that all of the linear equations defined by the matchings $\mathcal{H}_u$ are satisfied by the vectors $\mathbf{v}_1, \dots, \mathbf{v}_n$. Then, $\mathcal{K}$ does not contain an *odd-coloured* cycle, a cycle where some colour is used an odd number of times.

The reason this observation is true is straightforward. Suppose $\mathcal{K}$ did contain an odd-coloured cycle. Then, if we sum the vectors associated to the vertices across all edges in an odd-coloured cycle, we end up with $\mathbf{0}$, since all vertices in the cycle appear in an even number of edges. However, for each edge $((S, u), T)$ with colour i, we have that $\mathbf{v}_{(S,u)} + \mathbf{v}_T = \mathbf{v}_i$, using the linear equations defined by the hypergraph matchings. Hence, if the colour i appears an odd number of times in the cycle, we end up with $\mathbf{v}_i$ on the right-hand side after summing over all the edges. But, $\mathbf{v}_i \notin \text{span}\{\mathbf{v}_j : j \in [k] \setminus \{i\}\}$, so the right-hand side cannot be $\mathbf{0}$, and thus no odd-coloured cycle can exist in $\mathcal{K}$.

It thus remains to show that if k is too large, then $\mathcal{K}$ must contain an odd-coloured cycle. In fact, we will show the existence of a *unique colour* cycle, a cycle where some colour i is used exactly once. Note that, while any rainbow cycle is clearly both a unique colour and odd-coloured cycle, we cannot make black-box use of known rainbow cycle bounds here, as such results require that the host graph is *properly* edge-coloured. Given that the average degree of each colour in $\mathcal{K}$ is large, this is clearly false.

Unique colour cycles exist in graphs with similar degree in each colour. For a vertex $v \in V_L \cup V_R$, let $d_i(v)$ denote the degree of v with colour i, $d(v) = \sum_{i=1}^{k} d_i(v)$ denote the degree of v, and let $c(v)$ denote the *colour degree* at v, which is $|\{i : d_i(v) \geq 1\}|$. By a simple lemma (Fact 3.2), for any edge-coloured graph G, if $c(v) \geq k'$ for all vertices v and G contains no unique colour cycle, then $2^{k'} \leq |V|$. This means that if for some $\ell = O(\log n)$ (so that $|V_L| + |V_R| = n^{O(\log n)}$) we can pass to a subgraph in $\mathcal{K}$ such that all vertices in this subgraph have colour degree $\Omega(k)$, then we will obtain the desired $k \leq O(\log^2 n)$ bound.

It thus remains to understand the colour degree of a typical vertex in $\mathcal{K}$. In fact, one way to argue that the colour degree of a typical vertex is large is to simply show that $d_i(v)$ is approximately the same for almost all i and v. Indeed, this is precisely what the technical proofs in [KM24a, Yan24, KM24b] show, as they argue that $\mathcal{K}_i$, the subgraph consisting only of edges of colour i, is approximately regular for all i (and it is straightforward to see that the average degree of each $\mathcal{K}_i$ is also the same).

These technical "approximate regularity" lemmas can be easily rephrased to show that a typical vertex has $c(v) \geq \eta k$ for either $\eta = 1/\text{polylog}(n)$ ([KM24a, Yan24]) or $\eta = \Omega(1)$ ([KM24b]).

However, crucially for us, it is possible to argue that $c(v)$ is typically $\Omega(k)$ *without* showing that $d_i(v) \approx d_j(v')$ for typical colours i, j, and vertices v, v'. In fact, the two $\log n$ factor losses in previous proofs, coming from (1) imperfect hypergraph matchings, and (2) the existence of heavy pairs, arise precisely because each of these irregularities causes the approximate regularity condition to cease to be true. For example, for any pair $\{w, w'\}$ of vertices in $[n]$, if there are many hyperedges containing both w and w', then we may expect the total degree of the left vertex (S, w') to be unusually large whenever $w \in S$. This is because there are many valid choices for the vertices (u_{r-1}, b_r) in a chain $C = (u_0, a_1, b_1, u_1, a_2, b_2, u_2, \ldots, u_{r-1}, a_r, b_r, u_r)$ with $a_r = w$ and $u_r = w'$.

The central point of our argument is that we will control $c(v)$ by arguing that a typical vertex has $d_i(v) \approx d_j(v)$, i.e., its degree *within* each colour is approximately the same, even if $d_i(v)$ is not close to $d_i(v')$ for a different vertex v'. We do this by analyzing the following expected "local variance" quantity, defined as

$$\mathcal{V} \coloneqq \mathbb{E}_{v \sim V_L \sqcup V_R} \left[\mathbb{E}_{i \sim [k]} \left[\left(d_i(v) - \frac{d(v)}{k} \right)^2 \right] \right]. \tag{1}$$

Notice that if $d_i(v) = d_j(v)$ for all colours i and j, then v does not contribute to $\mathcal{V}$. More generally, if v is a vertex with degree $d(v)$ and has $c(v) \leq \eta k$, then we have

$$\mathbb{E}_{i \sim [k]} \left[\left(d_i(v) - \frac{d(v)}{k} \right)^2 \right] \geq \left(\frac{d(v)}{k} \right)^2 \cdot \left(\frac{1}{\eta} - 1 \right),$$

as the variance is minimized by setting $d_i(v) = d(v)/\eta k$ for ηk choices of i, and 0 otherwise. This means that if $c(v)$ is small, then v has a large contribution (scaling with $d(v)^2$) to Equation (1). In other words, if $\mathcal{V}$ is small, then intuitively most vertices must have large $c(v)$. We make this intuition precise in the proof of Lemma 5.3, where we show that if $\mathcal{V} \leq (|E|/k|V|)^2$, then we can find a subgraph where every vertex v in the subgraph has $c(v) \geq \Omega(k)$.

Bounding the local variance via mixing. To bound $\mathcal{V}$ in Equation (1), we now need to argue that for a typical vertex v, $d_i(v)$ is similar to $d_j(v)$. For reasons that will be explained later, we define our Kikuchi graph $\mathcal{K}$ (Definition 5.1) slightly differently than in previous work (Definition 2.3). First of all, the vertex sets are $V_L = [n]^{r+1}$ and $V_R = [n]^r$, that is, instead of sets of size $\ell \geq r$, we will consider ordered tuples of length exactly r. Our chains will be defined as $C = (u_0, a_1, b_1, u_1, a_2, b_2, u_2, \ldots, u_{r-1}, a_r, b_r, u_r)$, where for each j, we have either $\{a_j, b_j, u_j\} \in \mathcal{H}_{u_{j-1}}$, or $(u_{j-1}, a_j, b_j, u_j) = (x, y, y, x)$ for some $x, y \in [n]$. We say that the tuples of the form (x, y, y, x) are "fake" tuples, and the ones corresponding to hyperedges are "real". We now take an edge of colour $i \in [k]$ in our Kikuchi graph between $(a_1, \ldots, a_r, u_r) \in V_L$ and $(b_1, \ldots, b_r) \in V_R$ for each valid chain of the form $(u_0, a_1, b_1, u_1, a_2, b_2, u_2, \ldots, u_{r-1}, a_r, b_r, u_r)$ with $u_0 = i$. Crucially, we still have $\mathbf{v}_{u_{j-1}} + \mathbf{v}_{a_j} + \mathbf{v}_{b_j} + \mathbf{v}_{u_j} = \mathbf{0}$ for every j, so our new Kikuchi graph still has no unique colour cycle.

By rearranging terms in Equation (1), we can show that for this new Kikuchi graph, bounding $\mathcal{V}$ is equivalent to bounding $\mathbb{E}_{\vec{a}}[\|\mathbf{p}_i(\vec{a}) - \mathbf{p}_{i'}(\vec{a})\|_2^2]$, where $\vec{a} = (a_1, \ldots, a_r)$ is uniformly random and $\mathbf{p}_i(\vec{a})$ is the probability distribution of the "tail vertex" u_r in a random r-chain $C = (u_0, a_1, b_1, u_1, a_2, b_2, u_2, \ldots, u_{r-1}, a_r, b_r, u_r)$ with $u_0 = i$ and given $a_1, \ldots, a_r$.

We can view each r-chain as being constructed one link at a time, where the sequence of steps $u_0 \to u_1 \to \cdots \to u_r$ is simply a random walk where we transition from u_{j-1} to u_j as follows: either we pick u_j with $\{a_j, b_j, u_j\} \in \mathcal{H}_{u_{j-1}}$ for some $b_j \in [n]$ (if such u_j exists), or we take $u_j = u_{j-1}$. By adding a sufficient number of "fake" tuples, we can define the chains such that the probability of staying at u_{j-1} will always be at least $1/2$. Hence, the distributions $\mathbf{p}_i(\vec{a})$ and $\mathbf{p}_{i'}(\vec{a})$ are simply the distribution of the final vertex in this random walk after taking r steps, either starting from the vertex i or the vertex i'. We will be able to show that these distributions are close, as long as the directed graph $D_{\mathcal{H}}$, containing an edge (u, u') for each $\{a, b, u'\} \in \mathcal{H}_u$, is a good edge expander (see Lemma 5.7).

The case where $D_{\mathcal{H}}$ is not an edge expander is more complicated. In essence, we will decompose $D_{\mathcal{H}}$ into $O(1/\delta^2)$ subsets of vertices, where the induced graph on a subset W is an edge expander. This decomposition argument is subtle, and constructing the j-th piece W_j of the decomposition requires "quotienting out" by the vectors $\mathbf{v}_u$ for $u \in W_1 \cup \cdots \cup W_{j-1}$ and ignoring vertices $u \in [n]$ if their vectors $\mathbf{v}_u$ become $\mathbf{0}$ in the quotient space.

In order to apply the mixing argument, we require that the walk $u_0 \to u_1 \to \cdots \to u_r$ corresponding to a chain stays in the piece W in the decomposition, even though the other elements of the chain, namely the a_j's and b_j's, can be arbitrary in $[n]$. Each "real" step in the chain can be viewed as $\{a_j, b_j, u_j\} \in \mathcal{H}_{u_{j-1}}$ where (u_{j-1}, u_j) is the edge in $D_{\mathcal{H}}$ that is within the expanding piece W.

This allows us to bound the dimension of the quotient of each piece

$$\operatorname{span}\{\mathbf{v}_u : u \in W_1 \cup \cdots \cup W_j\} / \operatorname{span}\{\mathbf{v}_u : u \in W_1 \cup \cdots \cup W_{j-1}\}$$

by $O(\log^2 n)$, and conclude using that there are only $O(1/\delta^2)$ pieces.

We give the proof of Theorem 1 in Sections 4 and 5. In Section 4, we begin the proof and execute the decomposition into $O(1/\delta^2)$ expanding pieces. Then, in Section 5 we handle the main technical component of the proof, which is the case where $D_{\mathcal{H}}$ is an edge expander.

3 Preliminaries

For a graph $G = (V, E)$ and $v \in V$, we let $d^G(v)$ denote the degree of v in G. When G is clear from context, we will use $d(v)$ instead of $d^G(v)$ for simplicity. Graphs in this paper may have self-loops and parallel edges. For sets $S, T \subseteq V$, we let $e^G(S, T)$ be the number of edges of the form $\{v, v'\}$ where $v \in S$ and $v' \in T$. When G is clear from context, we will omit the subscript and instead write $e(S, T)$.

Definition 3.1 (Edge-coloured graphs and coloured cycles)**.** An edge-coloured graph G with colour palette $[k]$ is a collection $(V, E_1, \ldots, E_k)$ where V is the vertex set and each $E_i \subseteq \binom{V}{2}$ is an edge multiset. The multiset E_i denotes the edges with colour i. We let $d_i(v)$ denote the degree of v with colour i, i.e., $d_i(v) \coloneqq d^{G_i}(v)$ where $G_i = (V, E_i)$, and we let $d(v) = \sum_{i=1}^{k} d_i(v)$ denote the total degree of v. We let $c(v) \coloneqq |\{i \in [k] : d_i(v) \geq 1\}|$ denote the *colour degree* of the vertex v.

An odd-coloured cycle in G is a cycle where there exists $i \in [k]$ such that the number of edges appearing in the cycle from E_i is odd. A unique colour cycle in G is a cycle where there exists $i \in [k]$

such that exactly one edge from E_i appears in the cycle. A rainbow cycle in G is a cycle where all colours are distinct. Cycles of length two are allowed.

The following fact shows that every edge-coloured graph with no unique colour cycle and large minimum colour degree must have a large number of vertices.

Fact 3.2. *Let $G = (V, E_1, \dots, E_k)$ be an edge-coloured graph. Suppose that $c(v) \geq k'$ for all $v \in V$ and G does not contain a unique colour cycle. Then, $|V| \geq 2^{k'}$, or equivalently $k' \leq \log_2 |V|$.*

Proof. The proof follows by induction on k'. The base case of $k' = 0$ is trivial. Suppose $k' \geq 1$. Because $k' \geq 1$, there exists $i \in [k]$ such that E_i is nonempty. Let $\{v, v'\} \in E_i$ be an arbitrary edge with colour i. Observe that removing all edges in E_i from G must disconnect v and v', as otherwise there is a path from v to v' that uses colours other than i, and then adding in the last edge $\{v, v'\} \in E_i$ produces a unique colour cycle.

Deleting all edges of colour i thus splits the graph into two disconnected pieces, and can only decrease the colour degree of any vertex by 1. By induction, each piece has at least $2^{k'-1}$ vertices, and thus there are at least $2^{k'}$ vertices in the original graph. □

Definition 3.3 (Directed graphs)**.** A directed graph D is a collection (V, F) where V is the set of vertices and $F \subseteq V \times V$ is a multiset of directed edges. We let $d_{\text{in}}(v)$ denote the in-degree of v and $d_{\text{out}}(v)$ denote the out-degree of v. For sets $S, T \subseteq V$, we let $e^D(S, T)$ be the number of directed edges of the form (v, v') where $v \in S$ and $v' \in T$. When D is clear from context, we will omit the subscript and simply write $e(S, T)$.

We say that a directed graph D is *Eulerian* if $d_{\text{in}}(v) = d_{\text{out}}(v)$ for all $v \in V$.

The following result, due to Huang, Ma, Shapira, Sudakov and Yuster [HMS+13], shows that any dense Eulerian directed graph must have a short cycle.

Fact 3.4 (Corollary 1.3 in [HMS+13])**.** *Let $D = (V, F)$ be an Eulerian directed graph with no parallel edges. Then D has a directed cycle of length at most $6|V|^2/|F|$.*

We will need the following definition of edge expansion, which makes sense for dense directed graphs.

Definition 3.5 (Cut-dense directed graphs)**.** Let $D = (V, F)$ be a directed graph. We say that D is λ-cut-dense if for every $S \subseteq V$, it holds that

$$e(S, V \setminus S) \geq \lambda |S||V \setminus S|.$$

Finally, we will use Cheeger's inequality.

Fact 3.6 (Cheeger's inequality)**.** *Let $G = (V, E)$ be an undirected d-regular graph, and let A_G be its adjacency matrix. For $S \subseteq V$ with $1 \leq |S| \leq |V|/2$, define*

$$\phi(S) := \frac{e(S, V \setminus S)}{d|S|},$$

and let $\phi(G) = \min_{1 \leq |S| \leq |V|/2} \phi(S)$. Then, $\lambda_2 \geq \phi(G)^2/2$, where λ_2 is the second smallest eigenvalue of the matrix $\mathrm{I} - \frac{1}{d} A_G$.

4 Proof of Theorem 1

In this section, we prove Theorem 1. Let $C\colon \mathbb{F}_2^k \to \mathbb{F}_2^n$ be a linear $(3,\delta,\varepsilon)$-LCC. By Fact 2.1 and changing δ and n by constant factors, we may assume that we have vectors $\mathbf{v}_1,\dots,\mathbf{v}_n \in \mathbb{F}_2^k$ with $\dim(\mathrm{span}\{\mathbf{v}_1,\dots,\mathbf{v}_n\}) = k$ and 3-uniform hypergraph matchings $\mathcal{H}_1,\dots,\mathcal{H}_n$, each containing at least δn hyperedges, such that for every $u \in [n]$ and $\{u_1,u_2,u_3\} \in \mathcal{H}_u$, it holds that $\mathbf{v}_u = \mathbf{v}_{u_1}+\mathbf{v}_{u_2}+\mathbf{v}_{u_3}$. To prove Theorem 1, we must show that $k = \dim(\mathrm{span}\{\mathbf{v}_1,\dots,\mathbf{v}_n\}) \le O(\delta^{-14}\log^2 n)$.

The following key technical lemma completes the proof under the assumption that the directed graph obtained from "flattening" the hypergraph matchings is a good edge expander.

Lemma 4.1. *Let Ω and Λ be finite sets of size m and n, respectively. For each $u \in \Omega$ and each $a \in \Lambda$, let $\mathbf{v}_u$ and $\mathbf{v}_a$ be elements of an ambient vector space $\mathcal{W}$ over $\mathbb{F}_2$.*

Let $\mathcal{H}$ denote a collection of tuples of the form $(u,\{a,b\},u')$ where $u,u' \in \Omega$ and $\{a,b\} \in \binom{\Lambda}{2}$, where for every pair $(u,a) \in \Omega\times\Lambda$ there is at most 1 tuple of the form $(u,\{a,\star\},\star)$, and for each $(u,\{a,b\},u') \in \mathcal{H}$, it holds that $\mathbf{v}_u + \mathbf{v}_a + \mathbf{v}_b + \mathbf{v}_{u'} = \mathbf{0}$.

Let $D_{\mathcal{H}}$ be the directed graph with vertex set Ω and a directed edge (u,u') for every tuple $(u,\{a,b\},u') \in \mathcal{H}$. Suppose that $D_{\mathcal{H}}$ is an Eulerian directed graph (Definition 3.3) and is λ-cut-dense (Definition 3.5). Then,

$$\dim(\mathrm{span}\{\mathbf{v}_u\}_{u\in\Omega}) \le O\left(\frac{n^2}{\lambda^2 m^2}\cdot \log m \cdot \log n\right).$$

For technical reasons, Lemma 4.1 requires that $D_{\mathcal{H}}$ is Eulerian. However, it is possible to prove a similar statement without this assumption, at the cost of making the dependence on $1/\lambda$ exponential, rather than polynomial, in the dimension bound.

Lemma 4.1 is stated in the general form that we will need for the proof. However, for intuition it is helpful to consider the following setting. Take $\Omega = \Lambda = [n]$, and let the vectors $\mathbf{v}_u$ and $\mathbf{v}_a$ in Lemma 4.1 be the vectors $\mathbf{v}_1,\dots,\mathbf{v}_n$ from Fact 2.1. Then, take $\mathcal{H}$ to denote the tuples obtained by, for every $u \in [n]$ and every hyperedge $\{a,b,u'\} \in \mathcal{H}_u$, picking an arbitrary "special vertex" u' in the hyperedge and adding the tuple $(u,\{a,b\},u')$ to $\mathcal{H}$. Lemma 4.1 essentially states that if the directed graph with the arcs $u \to u'$ obtained from the tuples is a good edge expander, then we have $k \le O_\delta(\log^2 n)$.

We note that this implies[2] the $k \le O(\log^2 n)$ bound obtained in [KM24b] for the case of "design LCCs". Indeed, in a design LCC the matchings are *perfect*, which means that if we pick the special vertex u' randomly from each hyperedge, then the directed graph $D_{\mathcal{H}}$ is essentially a random directed graph with out-degree $n/3$ at each vertex, which is a good expander.

In the remainder of this section, we will prove Theorem 1 assuming Lemma 4.1.

Proof of Theorem 1. Let $\mathcal{W} = \mathrm{span}\{\mathbf{v}_1,\dots,\mathbf{v}_n\}$. We will construct a sequence of subspaces $\{\mathbf{0}\} = \mathcal{W}_0 \subseteq \mathcal{W}_1 \subseteq \cdots \subseteq \mathcal{W}_t = \mathcal{W}$ such that $t \le 8/\delta^2$ and $\dim(\mathcal{W}_i) - \dim(\mathcal{W}_{i-1}) \le O(\delta^{-12}\log^2 n)$. This clearly implies that $k = \dim(\mathcal{W}) \le O(\delta^{-14}\log^2 n)$, which proves Theorem 1.

[2] Up to the assumption that $D_{\mathcal{H}}$ is Eulerian, which is only needed for the better dependence on δ, see above.

We construct $\mathcal{W}_i$ using $\mathcal{W}_{i-1}$. Given $\mathcal{W}_{i-1}$, denote by $\bar{\mathbf{v}}_u$ the image of $\mathbf{v}_u$ in the quotient space $\mathcal{W}/\mathcal{W}_{i-1}$. Let $V = \{u \in [n] : \bar{\mathbf{v}}_u \neq \mathbf{0}\}$ be the subset of indices whose vectors remain nonzero. If $V = \emptyset$, then we stop this process and let $t = i - 1$, and note that $\mathcal{W}_{i-1} = \mathcal{W}$.

Suppose that $V \neq \emptyset$, and let $u \in V$. Recall that we have a hypergraph matching $\mathcal{H}_u$ with at least δn edges such that for every $\{u_1, u_2, u_3\} \in \mathcal{H}_u$, we have that $\mathbf{v}_u + \mathbf{v}_{u_1} + \mathbf{v}_{u_2} + \mathbf{v}_{u_3} = \mathbf{0}$. Note that this relation continues to hold in the quotient space $\mathcal{W}/\mathcal{W}_{i-1}$, i.e., $\bar{\mathbf{v}}_u + \bar{\mathbf{v}}_{u_1} + \bar{\mathbf{v}}_{u_2} + \bar{\mathbf{v}}_{u_3} = \mathbf{0}$. Since $\bar{\mathbf{v}}_u \neq \mathbf{0}$, it follows that at least one of $\bar{\mathbf{v}}_{u_1}$, $\bar{\mathbf{v}}_{u_2}$, or $\bar{\mathbf{v}}_{u_3}$ must also be nonzero. Let $u' \in V$ be the index of one of these nonzero vectors, chosen arbitrarily.

We now construct a directed graph D with vertex set V as follows. For each $u \in V$ and each $\{u_1, u_2, u_3\} \in \mathcal{H}_u$, we add a directed edge (u, u') to D, where u' is the element chosen from the hyperedge $\{u_1, u_2, u_3\}$ as described above, so that $\bar{\mathbf{v}}_{u'} \neq \mathbf{0}$ and thus $u' \in V$. Note that each edge (u, u') in D corresponds to a particular hyperedge $\{u_1, u_2, u'\} \in \mathcal{H}_u$, and since $\mathcal{H}_u$ is a hypergraph matching, D has no parallel directed edges. Furthermore, since $\mathcal{H}_u$ has at least δn hyperedges, we have that $\min_{v \in V} d_{\text{out}}(v) \geq \delta n$.

The following lemma, which we prove in Section 4.1, shows that there exists an Eulerian expanding subgraph of D with a large vertex set.

Lemma 4.2. *Let $D = (V, F)$ be a directed graph with no parallel edges, $|V| \leq n$ and $\min_{v \in V} d_{\text{out}}(v) \geq \delta n$. Then there is an Eulerian subgraph $D' = (V', F')$ of D with $|V'| \geq \delta^2 n/8$ such that D' is λ-cut dense with $\lambda = \delta^4/96$.*

We now use Lemma 4.2 to finish the proof. The directed graph D defined above satisfies the requirements of Lemma 4.2. Let $D' = (V', F')$ be the subgraph of D guaranteed by Lemma 4.2, and let $\mathcal{W}_i = \mathcal{W}_{i-1} + \text{span}\{\mathbf{v}_u : u \in V'\}$. Note that $\dim(\mathcal{W}_i) - \dim(\mathcal{W}_{i-1}) = \dim(\mathcal{W}_i/\mathcal{W}_{i-1}) = \dim(\text{span}\{\bar{\mathbf{v}}_u : u \in V'\})$, where $\bar{\mathbf{v}}_u$ is the image of $\mathbf{v}_u$ in $\mathcal{W}/\mathcal{W}_{i-1}$.

We now apply Lemma 4.1. We let $\Omega = V'$ and $\Lambda = [n]$, and the vectors associated with elements of Ω and Λ are the $\bar{\mathbf{v}}_u$'s, which lie in the vector space $\mathcal{W}/\mathcal{W}_{i-1}$. We let $\mathcal{H}$ denote the tuples $(u, \{a, b\}, u')$ where $u \in V'$, $\{a, b, u'\} \in \mathcal{H}_u$, and the hyperedge $\{a, b, u'\}$ "corresponds to" an edge that is in D'. In other words, for each edge $(u, u') \in F'$, we "remember" the hyperedge $\{a, b, u'\} \in \mathcal{H}_u$ that produced the edge (u, u') in D', and we add the tuple $(u, \{a, b\}, u')$ to $\mathcal{H}$. Notice that we have $u' \in V'$, and $a, b \in [n]$. Furthermore, because $\mathcal{H}_u$ is a matching, the first requirement on $\mathcal{H}$ is satisfied, and the second requirement holds because the linear equation holds in $\mathcal{W}$, and thus also holds in $\mathcal{W}/\mathcal{W}_{i-1}$. The directed graph $D_{\mathcal{H}}$ in the statement of Lemma 4.1 is simply D', and so it is an Eulerian directed graph that is λ-cut-dense. By Lemma 4.1, we thus have that

$$\dim(\mathcal{W}_i/\mathcal{W}_{i-1}) = \dim(\text{span}\{\bar{\mathbf{v}}_u : u \in V'\}) \leq O\left(\frac{n^2}{\lambda^2 |V'|^2} \cdot \log|V'| \cdot \log n\right) = O\left(\delta^{-12} \log^2 n\right),$$

since $|V'| \geq \delta^2 n/8$ and $\lambda = \delta^4/96$.

We have thus shown how to construct $\mathcal{W}_i$ from $\mathcal{W}_{i-1}$. It remains to argue that the number of iterations is bounded by $8/\delta^2$. This follows because by Lemma 4.2, we have that $|V'| \geq \delta^2 n/8$, and since all the vectors $\mathbf{v}_u$ for $u \in V'$ are in $\mathcal{W}_i$, this means that $|V|$ decreases by at least $\delta^2 n/8$ with each iteration. So, the total number of iterations is bounded by $8/\delta^2$, which finishes the proof of Theorem 1. □

4.1 Proof of Lemma 4.2

In this subsection, we prove Lemma 4.2, restated below.

Lemma 4.2. *Let $D = (V, F)$ be a directed graph with no parallel edges, $|V| \le n$ and $\min_{v \in V} d_{\text{out}}(v) \ge \delta n$. Then there is an Eulerian subgraph $D' = (V', F')$ of D with $|V'| \ge \delta^2 n/8$ such that D' is λ-cut dense with $\lambda = \delta^4/96$.*

Proof. Let $d = \lceil \delta n \rceil$. Note that $|V| \ge d$ since each $v \in V$ has at least d outgoing edges and there are no parallel edges. We will construct D' in three steps.

As the first step, we will find an Eulerian subgraph of D that has a large fraction of the edges. To do this, we remove cycles from D until the resulting graph D'' is acyclic, and let F_1 be the union of the removed cycles. Clearly, we have that $D_1 = (V, F_1)$ is Eulerian.

We now show that $|F_1| \ge d^2/2$, so that D_1 has many edges. Because D'' is a directed acyclic graph, we can order its vertices so that each v_i has no outgoing edges to a later vertex in the order. Hence, in D'' the i-th vertex in the order has at most $i-1$ outgoing edges remaining, or in other words, at least $d - (i-1)$ outgoing edges have been removed from D and added to D_1. Summing over the first d vertices in the order implies that at least

$$\sum_{i=1}^{d} (d - (i-1)) \ge d^2/2 \ge \delta^2 n^2/2$$

edges have been removed, and so $|F_1| \ge d^2/2$.

As the second step, we will find an Eulerian subgraph $D_2 = (V, F_2)$ of D_1 where F_2 is the union of *short* directed cycles. We do this by applying Fact 3.4. Let $\ell = 24/\delta^2$. Starting from the edge set F_1, we repeatedly remove a cycle of length $\le \ell$ from D_1 and add it to F_2. Note that by Fact 3.4, such a cycle exists as long as $|F_1| \ge \delta^2 n^2/4$. Hence, $|F_2| \ge \delta^2 n^2/2 - \delta^2 n^2/4 = \delta^2 n^2/4$, and we have that $F_2 = C_1 \sqcup C_2 \sqcup \cdots \sqcup C_t$, where each C_j is a cycle of length at most ℓ.

As the third and final step, we will split the vertex set of D_2 into expanding pieces. However, unlike a standard expander decomposition that removes edges, here we will remove *entire cycles* from F_2 at a time, using the cycle decomposition of F_2 constructed above, as this will ensure that the resulting graph is Eulerian. That is, when we remove an edge e, we will also remove the entire cycle C_j, where C_j is the unique cycle in the partition $F_2 = C_1 \sqcup C_2 \sqcup \cdots \sqcup C_t$ that contains e. The fact that the cycles are *short* will crucially ensure that we do not delete too many edges here.

We now describe this process more formally. Let $\lambda = \delta^4/96$. We will cut our directed graph into a number of disjoint pieces. While there exists a split $U = A \sqcup B$ of a current piece U into nonempty sets where $e^{F_2}(A, B) < \lambda |A||B|$, we split the graph D_2 into two parts as follows. We remove all edges from A to B, along with all other edges that are in the unique cycles C_j that contain these edges according to F_2's short cycle decomposition. Because each cycle has length at most $\ell = 24/\delta^2$, each step of this process deletes at most $\ell\lambda|A||B|$ edges. Note that this process also deletes all edges from B to A, as each such edge lies in some cycle C_j in the cycle decomposition of F_2, and that cycle must also have an edge from A to B. We then replace U by the retained directed graphs on A and B.

As each (unordered) pair of vertices $\{u, u'\}$ in V is separated at most once during this process, we have that $\sum_{\text{splits } A \sqcup B} |A||B| \le \binom{n}{2}$. The total number of edges deleted is thus at most $\ell\lambda n^2/2 = \delta^2 n^2/8$.

So, after this process terminates, we have decomposed D_2 into Eulerian λ-cut-dense pieces where the total number of edges preserved is at least $\delta^2 n^2/8$.

By averaging, there exists some piece $D' = (V', F')$ where $|F'|/|V'| \geq \delta^2 n/8$. Note that as $|F'| \leq |V'|^2$, this implies that $|V'| \geq \delta^2 n/8$, which finishes the proof. □

5 The Proof Assuming Edge Expansion: Proof of Lemma 4.1

In this section, we prove Lemma 4.1, restated below.

Lemma 4.1. *Let Ω and Λ be finite sets of size m and n, respectively. For each $u \in \Omega$ and each $a \in \Lambda$, let $\mathbf{v}_u$ and $\mathbf{v}_a$ be elements of an ambient vector space $\mathcal{W}$ over $\mathbb{F}_2$.*

Let $\mathcal{H}$ denote a collection of tuples of the form $(u, \{a, b\}, u')$ where $u, u' \in \Omega$ and $\{a, b\} \in \binom{\Lambda}{2}$, where for every pair $(u, a) \in \Omega \times \Lambda$ there is at most 1 tuple of the form $(u, \{a, \star\}, \star)$, and for each $(u, \{a, b\}, u') \in \mathcal{H}$, it holds that $\mathbf{v}_u + \mathbf{v}_a + \mathbf{v}_b + \mathbf{v}_{u'} = \mathbf{0}$.

Let $D_{\mathcal{H}}$ be the directed graph with vertex set Ω and a directed edge (u, u') for every tuple $(u, \{a, b\}, u') \in \mathcal{H}$. Suppose that $D_{\mathcal{H}}$ is an Eulerian directed graph (Definition 3.3) and is λ-cut-dense (Definition 3.5). Then,

$$\dim(\operatorname{span}\{\mathbf{v}_u\}_{u \in \Omega}) \leq O\left(\frac{n^2}{\lambda^2 m^2} \cdot \log m \cdot \log n\right).$$

Proof. Let $k = \dim(\operatorname{span}\{\mathbf{v}_u : u \in \Omega\})$, and $\mathscr{C} \subseteq \Omega$ be a set of size k where $\{\mathbf{v}_i : i \in \mathscr{C}\}$ are linearly independent.

We now define a multiset $\mathcal{T} \subseteq \Omega \times \Lambda \times \Lambda \times \Omega$ as follows. For each tuple $(u, \{a, b\}, u') \in \mathcal{H}$, we add the tuples (u, a, b, u'), (u, b, a, u'), (u, a, a, u), and (u, b, b, u) to $\mathcal{T}$. For each pair $(u, a) \in \Omega \times \Lambda$ where no such tuple exists, we add two copies of (u, a, a, u) to $\mathcal{T}$. Note that there is at most one tuple of the form $(u, \{a, b\}, u') \in \mathcal{H}$ for a fixed choice of (u, a), by assumption. Therefore, in the multiset $\mathcal{T}$, for any choice of (u, a), there are exactly 2 tuples of the form $(u, a, \star, \star)$, and similarly for any choice of (u, b), there are exactly 2 tuples of the form $(u, \star, b, \star)$.

We call tuples of the form (u, a, a, u) "fake", and the tuples originating from $\mathcal{H}$ "real". Note that for any tuple $(u, a, b, u') \in \mathcal{T}$, real or fake, it holds that $\mathbf{v}_u + \mathbf{v}_a + \mathbf{v}_b + \mathbf{v}_{u'} = \mathbf{0}$ in $\mathcal{W}$. Additionally, for any choice of (u, a), at least one of the 2 tuples of the form $(u, a, \star, \star)$ is fake.

The proof of Lemma 4.1 revolves around analyzing combinatorial properties of the edge-coloured Kikuchi graph defined below.

Definition 5.1 (r-chain Kikuchi graph)**.** Let $r \geq 1$ be an integer. The graph $\mathcal{K}_r(\mathscr{C}, \mathcal{T})$ is a bipartite graph with left vertices $V_L = \Lambda^r \times \Omega$ and right vertices $V_R = \Lambda^r$. We let $(\vec{a}, u)$ where $\vec{a} = (a_1, \ldots, a_r)$ denote a left vertex in V_L, and $\vec{b} = (b_1, \ldots, b_r)$ denote a right vertex in V_R.

For each sequence $(i, a_1, b_1, u_1, \ldots, u_{r-1}, a_r, b_r, u_r)$ where $i \in \mathscr{C}$ and $(u_{j-1}, a_j, b_j, u_j) \in \mathcal{T}$ for all $j \in [r]$ (treating $u_0 = i$), we add an edge $(((a_1, \ldots, a_r), u_r), (b_1, \ldots, b_r))$ of colour i to $\mathcal{K}_r(\mathscr{C}, \mathcal{T})$. These edges are added with multiplicity according to the multiplicities of the tuples in $\mathcal{T}$.

Because the colour set $\mathscr{C}$ is chosen so that the k vectors $\{\mathbf{v}_i : i \in \mathscr{C}\}$ are linearly independent, it follows that $\mathcal{K}_r(\mathscr{C}, \mathcal{T})$ has no odd-coloured or unique colour cycle. Indeed, this is due to the

following observation, which holds because $(u_{j-1}, a_j, b_j, u_j) \in \mathcal{T}$ implies that $\mathbf{v}_{u_{j-1}} + \mathbf{v}_{a_j} + \mathbf{v}_{b_j} + \mathbf{v}_{u_j} = \mathbf{0}$.

Observation 5.2. For a vertex $(\vec{a}, u) \in V_L$, let $\mathbf{v}_{(\vec{a},u)} := \mathbf{v}_u + \sum_{j=1}^{r} \mathbf{v}_{a_j}$, and for a vertex $\vec{b} \in V_R$, let $\mathbf{v}_{\vec{b}} := \sum_{j=1}^{r} \mathbf{v}_{b_j}$. Then, for any edge $((\vec{a}, u), \vec{b})$ in $\mathcal{K}_r(\mathscr{C}, \mathcal{T})$ of colour $i \in \mathscr{C}$, it holds that $\mathbf{v}_{(\vec{a},u)} + \mathbf{v}_{\vec{b}} = \mathbf{v}_i$.

Thus, if $\mathcal{K}_r(\mathscr{C}, \mathcal{T})$ contains an odd-coloured cycle, then for each edge in the cycle $((\vec{a}, u), \vec{b})$, we have $\mathbf{v}_{(\vec{a},u)} + \mathbf{v}_{\vec{b}} = \mathbf{v}_i$. Summing over all edges in the cycle yields $\mathbf{0}$ on the left, while the right hand side contains some $\mathbf{v}_i$ with an odd coefficient. The only other terms that can appear on the right are $\mathbf{v}_{i'}$ for $i' \in \mathscr{C}$ with $i' \neq i$, which are linearly independent from $\mathbf{v}_i$, and thus the right hand side cannot be $\mathbf{0}$.

So, it remains to show that $\mathcal{K}_r(\mathscr{C}, \mathcal{T})$ must contain an odd-coloured cycle if $k = |\mathscr{C}|$ is too large. We do this by controlling the following variance quantity, which captures the expected local variance of the degrees of each colour at a single vertex.

$$\mathcal{V} := \mathbb{E}_{v \sim V_L \sqcup V_R} \left[\mathbb{E}_{i \sim \mathscr{C}} \left[\left(d_i(v) - \frac{d(v)}{k} \right)^2 \right] \right]. \tag{2}$$

Here, $d_i(v)$ is the degree of the vertex v within the i-th colour in $\mathcal{K}_r(\mathscr{C}, \mathcal{T})$.

The following lemma, which we prove in Section 5.1, shows that if we can obtain a sufficient upper bound on $\mathcal{V}$ in Equation (2), then we are done.

Lemma 5.3. *Let $G = (V, E_1, \ldots, E_k)$ be an edge-coloured graph. Suppose that G has no unique colour cycle. If*

$$\mathcal{V} := \mathbb{E}_{v \sim V} \left[\mathbb{E}_{i \sim [k]} \left[\left(d_i(v) - \frac{d(v)}{k} \right)^2 \right] \right] \leq \left(\frac{|E|}{k|V|} \right)^2,$$

then $k \leq 60 \log_2 |V|$.

The remainder of the proof is therefore dedicated to bounding $\mathcal{V}$. We do this by connecting $\mathcal{V}$ to the behavior of a random walk on the directed transition graph $D_{\mathcal{T}}$, specified below.

Definition 5.4 (Transition graph)**.** Let $D_{\mathcal{T}} = (\Omega, F_{\mathcal{T}})$ be the directed graph with vertex set Ω, and edge set $F_{\mathcal{T}}$ defined by adding an edge (u, u') for each $(u, a, b, u') \in \mathcal{T}$ (with multiplicity). We let $Q_{\mathcal{T}}$ denote the adjacency matrix of $D_{\mathcal{T}}$, i.e., $Q_{\mathcal{T}}(u', u)$ is equal to the number of edges (u, u') in the multiset $F_{\mathcal{T}}$.

We let $G_{\mathcal{T}} = (\Omega, E_{\mathcal{T}})$ be the *undirected* graph with vertex set Ω, defined via the adjacency matrix $A_{\mathcal{T}} := Q_{\mathcal{T}} + Q_{\mathcal{T}}^{\top}$.

The graphs $D_{\mathcal{T}}$ and $G_{\mathcal{T}}$ have the following properties, which we will use.

Proposition 5.5 (Properties of the transition graph)**.** *The following hold for $D_{\mathcal{T}}$ and $G_{\mathcal{T}}$.*

(1) The graph $2D_{\mathcal{H}}$, i.e., the graph where every edge in $D_{\mathcal{H}}$ appears with multiplicity 2, is a subgraph of $D_{\mathcal{T}}$.

(2) The graph $D_{\mathcal{T}}$ is Eulerian, and every vertex $u \in \Omega$ has $d_{\text{in}}(u) = d_{\text{out}}(u) = 2n$. Furthermore, each u has at least n self-loops.

(3) *The graph $G_{\mathcal{T}}$ is $4n$-regular.*

(4) *For each $a \in \Lambda$, let M_a be the matrix where $M_a(u', u) = \frac{1}{2}|\{(u, a, \star, u') \in \mathcal{T}\}|$. Then, $\mathbb{E}_{a\sim\Lambda}[M_a] = \frac{1}{2n}Q_{\mathcal{T}}$, every column of M_a sums to 1 and $M_a(u, u) \geq 1/2$ for all u.*

(5) *Let λ_2 denote the second smallest eigenvalue of the Laplacian $\mathrm{I} - \frac{1}{4n}A_{\mathcal{T}}$. Then, $\lambda_2 \geq \lambda^2 m^2/8n^2$, where $m = |\Omega|$.*

The proof of Proposition 5.5 is straightforward, and is given below.

Proof. The first property follows by definition of $\mathcal{T}$ and the graph $D_{\mathcal{T}}$, because each $(u, \{a, b\}, u') \in \mathcal{H}$ produces 2 real tuples, (u, a, b, u') and (u, b, a, u'), in $\mathcal{T}$.

The second property follows because the graph $D_{\mathcal{H}}$ in the statement of Lemma 4.1 is Eulerian, and the addition of the "fake" tuples to $\mathcal{T}$ makes $d_{\text{out}}(u) = 2n$ for all $u \in \Omega$. We also have $d_{\text{in}}(u) = 2n$, since the addition of the fake tuples preserves the Eulerian property. The fake tuples in $\mathcal{T}$ provide at least n self-loops at each vertex.

The third property follows from the second because $G_{\mathcal{T}}$ is obtained by removing the direction of the edges in $D_{\mathcal{T}}$. Hence, $d^{G_{\mathcal{T}}}(u) = d^{D_{\mathcal{T}}}_{\text{in}}(u) + d^{D_{\mathcal{T}}}_{\text{out}}(u) = 4n$.

The fourth property is by definition of the adjacency matrix. The fact that the column sums of M_a are 1 follows because for every (u, a), there are exactly 2 tuples of the form $(u, a, \star, \star) \in \mathcal{T}$, and $M_a(u, u) \geq 1/2$ follows from the existence of at least one fake tuple (u, a, a, u) in $\mathcal{T}$.

The final property follows from Cheeger's inequality (Fact 3.6) and the fact that the graph $D_{\mathcal{H}}$ is λ-cut-dense (Definition 3.5). Indeed, for any nonempty $S \subseteq \Omega$, we have $\phi(S) = e^{G_{\mathcal{T}}}(S, \Omega \setminus S)/4n|S|$. Because $2D_{\mathcal{H}}$ is a subgraph of $D_{\mathcal{T}}$ and is λ-cut-dense, we have

$$e^{G_{\mathcal{T}}}(S, \Omega \setminus S) \geq 2e^{D_{\mathcal{H}}}(S, \Omega \setminus S) + 2e^{D_{\mathcal{H}}}(\Omega \setminus S, S) \geq 4\lambda|S||\Omega \setminus S| \,.$$

Hence, $\phi(S) \geq \frac{\lambda|\Omega\setminus S|}{n}$, and so $\phi(G_{\mathcal{T}}) \geq \frac{\lambda m}{2n}$. By Fact 3.6, this implies that $\lambda_2 \geq \lambda^2 m^2/8n^2$. □

With Definition 5.4 and Proposition 5.5 in hand, we are ready to bound $\mathcal{V}$. First, we will show that $\mathcal{V}$ is *equal* to a quantity related to a random walk on the graph $D_{\mathcal{T}}$. This quantity is meant to capture the difference in distributions obtained by running a random walk for r steps, starting either at the vertex $i \in \Omega$, or another vertex $i' \in \Omega$. Then, we will bound this quantity using Item (5) in Proposition 5.5, which says that λ_2 of the Laplacian of $G_{\mathcal{T}}$ is somewhat large.

The claim below relates $\mathcal{V}$ to the random walk quantity, and then the lemma bounds the quantity.

Claim 5.6. For $a \in \Lambda$, let M_a be the matrix where $M_a(u', u) = \frac{1}{2}|\{(u, a, \star, u') \in \mathcal{T}\}|$. Then, it holds that

$$\mathcal{V} = \frac{2^{2r}}{2(m+1)} \mathbb{E}_{i,i'\in\mathscr{C}} \left[\mathbb{E}_{a_1,\ldots,a_r\sim\Lambda} \left[\|M_{a_r} \ldots M_{a_1}(e_i - e_{i'})\|_2^2\right]\right] .$$

Lemma 5.7. *For $r \in \mathbb{N}$, let $M^{(1)}, \ldots, M^{(r)} \in \mathbb{R}^{m\times m}$ be independent copies of a random matrix M, where with probability 1, it holds that (1) M is column stochastic: it has nonnegative entries and column sums equal to 1, and (2) $M \geq \mathrm{I}/2$ entrywise. Suppose that $\overline{M} = \mathbb{E}[M]$ is doubly stochastic: all entries are nonnegative*

and all row and column sums are 1*, and suppose that the second largest eigenvalue of* $\frac{1}{2}(\overline{M} + \overline{M}^\top)$ *is at most* $1 - \lambda_2$*. Then,*

$$\max_{u,u' \in [m]} \mathbb{E}_{M^{(1)},\ldots,M^{(r)}} \left[\|M^{(r)} \ldots M^{(1)}(e_u - e_{u'})\|_2^2 \right] \leq 2m(1-\lambda_2)^r \,.$$

The proof of Claim 5.6 is simple, and we give it below. We postpone the proof of Lemma 5.7 to Section 5.2.

Proof of Claim 5.6. First, we observe that each right vertex $\vec{b} \in V_R$ contributes 0 to $\mathcal{V}$. Indeed, we have shown that for each pair (u, b), there are exactly 2 tuples of the form $(u, \star, b, \star) \in \mathcal{T}$, and this implies that $d_i(\vec{b}) = 2^r$ for all i and all $\vec{b}$. Since $|V_L| = |\Omega||V_R| = m|V_R|$, we have that

$$\mathcal{V} = \frac{m}{m+1} \mathbb{E}_{v \sim V_L} \left[\mathbb{E}_{i \sim \mathscr{C}} \left[\left(d_i(v) - \frac{d(v)}{k} \right)^2 \right] \right] .$$

We also have that for any $v \in V_L$,

$$\mathbb{E}_{i \sim \mathscr{C}} \left[\left(d_i(v) - \frac{d(v)}{k} \right)^2 \right] = \frac{1}{2} \mathbb{E}_{i,i' \sim \mathscr{C}} \left[(d_i(v) - d_{i'}(v))^2 \right] .$$

Next, we observe that for any $(\vec{a}, u) \in V_L$, we have that $d_i(\vec{a}, u) = e_u^\top 2^r M_{a_r} \ldots M_{a_1} e_i$. This follows because there are exactly 2 tuples of the form $(u, a, \star, \star) \in \mathcal{T}$, and thus the u-th entry of $2^r M_{a_r} \ldots M_{a_1} e_i$ exactly counts the number of r-chains of the form $(i, a_1, \star, u_1, a_2, \star, \ldots, u_{r-1}, a_r, \star, u)$, which is exactly $d_i(\vec{a}, u)$. Hence,

$$\begin{aligned}
\mathcal{V} &= \frac{m}{2(m+1)} \mathbb{E}_{v \sim V_L} \left[\mathbb{E}_{i,i' \sim \mathscr{C}} \left[(d_i(v) - d_{i'}(v))^2 \right] \right] \\
&= \frac{m}{2(m+1)} \mathbb{E}_{i,i' \sim \mathscr{C}} \left[\mathbb{E}_{v \sim V_L} \left[(d_i(v) - d_{i'}(v))^2 \right] \right] \\
&= \frac{m}{2(m+1)} \mathbb{E}_{i,i' \sim \mathscr{C}} \left[\mathbb{E}_{a_1,\ldots,a_r \sim \Lambda, u \sim \Omega} \left[\left(d_i(\vec{a}, u) - d_{i'}(\vec{a}, u) \right)^2 \right] \right] \\
&= \frac{m}{2(m+1)} \cdot 2^{2r} \frac{1}{m} \cdot \mathbb{E}_{i,i' \sim \mathscr{C}} \left[\mathbb{E}_{a_1,\ldots,a_r \sim \Lambda} \left[\|M_{a_r} \ldots M_{a_1}(e_i - e_{i'})\|_2^2 \right] \right] ,
\end{aligned}$$

which finishes the proof. □

We are now ready to finish the proof of Lemma 4.1. By Claim 5.6, we have

$$\mathcal{V} = \frac{2^{2r}}{2(m+1)} \mathbb{E}_{i,i' \in \mathscr{C}} \left[\mathbb{E}_{a_1,\ldots,a_r \sim \Lambda} \left[\|M_{a_r} \ldots M_{a_1}(e_i - e_{i'})\|_2^2 \right] \right] .$$

We now apply Lemma 5.7 to bound $\mathbb{E}_{a_1,\ldots,a_r \sim \Lambda} \left[\|M_{a_r} \ldots M_{a_1}(e_i - e_{i'})\|_2^2 \right]$. Indeed, we have that each M_{a_j} is an independent copy of the same random matrix, and each M_{a_j} has column sums equal to 1 by Item (4) in Proposition 5.5. Each M_{a_j} is furthermore entrywise at least I/2, as all entries are nonnegative, and the diagonal entries are at least 1/2 because of Item (4) in Proposition 5.5. The mean $\overline{M}$ is equal to $\mathbb{E}_{a \sim \Lambda}[M_a]$, which by Item (4) in Proposition 5.5 is equal to $\frac{1}{2n} Q_{\mathcal{T}}$. By Item (2) in Proposition 5.5, $\frac{1}{2n} Q_{\mathcal{T}}$ is doubly stochastic. By Item (5) in Proposition 5.5, the matrix

$\frac{1}{2}(\overline{M}+\overline{M}^\top) = \frac{1}{4n}A_{\mathcal{T}}$ has second largest eigenvalue at most $1-\lambda_2$ for $\lambda_2 \geq \lambda^2 m^2/8n^2$. Thus, applying Lemma 5.7, we conclude that

$$\mathcal{V} = \frac{2^{2r}}{2(m+1)}\,\mathbb{E}_{i,i'\in\mathscr{C}}\left[\mathbb{E}_{a_1,\ldots,a_r\sim\Lambda}\left[\|M_{a_r}\ldots M_{a_1}(e_i-e_{i'})\|_2^2\right]\right] \leq \frac{2^{2r}}{2(m+1)}\cdot 2m\left(1-\frac{\lambda^2m^2}{8n^2}\right)^r.$$

Finally, in the Kikuchi graph $\mathcal{K}_r(\mathscr{C},\mathcal{T})$, we have $|V| = |V_L| + |V_R| = n^r(m+1)$ and $|E| = |V_R|\cdot k\cdot 2^r = k(2n)^r$. Hence, $|E|/k|V| = 2^r/(m+1)$. Thus, taking $r = c\frac{n^2}{\lambda^2m^2}\cdot\log_2 m$ for a sufficiently large constant c, we get that $\mathcal{V} \leq (|E|/k|V|)^2$, and by Lemma 5.3 we conclude that $k \leq 60\log_2|V| = O(r\log n + \log m) = O(\frac{n^2}{\lambda^2m^2}\cdot\log m\cdot\log n)$. This finishes the proof up to the proofs of Lemma 5.3 and Lemma 5.7, which we show in the subsequent subsections. □

5.1 Proof of Lemma 5.3

In this subsection, we prove Lemma 5.3, restated below.

Lemma 5.3. *Let $G = (V, E_1, \ldots, E_k)$ be an edge-coloured graph. Suppose that G has no unique colour cycle. If*

$$\mathcal{V} \coloneqq \mathbb{E}_{v\sim V}\left[\mathbb{E}_{i\sim[k]}\left[\left(d_i(v)-\frac{d(v)}{k}\right)^2\right]\right] \leq \left(\frac{|E|}{k|V|}\right)^2,$$

then $k \leq 60\log_2|V|$.

Proof. Before we give the detailed proof, we will explain the ideas behind the calculations, which are relatively simple. Using the bound on $\mathcal{V}$, one can show that there are few edges of colour i that touch vertices v where $d_i(v)$ is much larger than $d(v)/k$. After we remove these bad edges, we still have many edges left, as well as a good upper bound on the degrees of individual colours at every vertex, i.e., $d_i(v)$ for all i and v. We then iteratively delete vertices whose remaining degree is small relative to its degree $d(v)$ in the initial graph. This gives us a nonempty subgraph G' of G where every vertex has $d_i^{G'}(v) \leq O(d^{G'}(v)/k)$ for all colours i, and thus each vertex has colour degree $c(v) \geq \Omega(k)$. Applying Fact 3.2, we thus obtain an upper bound on k.

Let us now turn to the details. Call a pair (i,v) "bad" if $d_i(v) > 10d(v)/k$, or in other words, if colour i is overrepresented at v. Delete every edge (v,v') of colour i if either (i,v) or (i,v') is bad. To bound the number of deleted edges, fix a bad pair (i,v). Because (i,v) is bad, we have

$$d_i(v) - \frac{d(v)}{k} > \frac{9}{10}d_i(v).$$

Hence, writing B for the number of bad (i,v), we have

$$\left(\frac{9}{10}\right)^2\left(\sum_{(i,v)\text{ bad}} d_i(v)\right)^2 \leq \left(\sum_{(i,v)\text{ bad}}\left(d_i(v)-\frac{d(v)}{k}\right)\right)^2 \leq B\sum_v\sum_{i=1}^k\left(d_i(v)-\frac{d(v)}{k}\right)^2 = B\cdot\mathcal{V}\cdot k|V|,$$

where in the second step we used the Cauchy–Schwarz inequality. Clearly, each vertex v is in at most $k/10$ bad pairs, so $B \leq |V|k/10$. Hence, we obtain $\left(\frac{9}{10}\right)^2(\sum_{(i,v)\text{ bad}} d_i(v))^2 \leq \frac{|V|k}{10}\cdot\frac{|E|^2}{k|V|} = \frac{|E|^2}{10}$, so the number of deleted edges is at most $\sum_{(i,v)\text{ bad}} d_i(v) \leq |E|/2$.

In the remaining graph, iteratively delete any vertex v whose current degree is at most $d(v)/6$, where $d(v)$ is the degree of v in the original graph G. This procedure deletes at most $\sum_v d(v)/6 = |E|/3$ edges. Thus, the final set of edges E' has size at least $|E| - |E|/2 - |E|/3 = |E|/6 > 0$ edges. Let $G' = (V', E')$ be the final graph after removing all vertices that are not adjacent to any edge in E'.

Every remaining vertex $v \in V'$ satisfies

$$d^{G'}(v) \geq \frac{d(v)}{6}, \qquad d_i^{G'}(v) \leq \frac{10d(v)}{k} \quad \text{for every colour } i.$$

Let the *colour degree* $c(v)$ of a vertex in a graph be the number of distinct colours incident at that vertex. By the above inequalities, each vertex v in G' has $c(v) \geq k/60$.

Applying Fact 3.2 to G', we conclude that $k/60 \leq \log_2|V'| \leq \log_2|V|$, and so $k \leq 60\log_2|V|$. □

5.2 Proof of Lemma 5.7

In this subsection, we prove Lemma 5.7, restated below.

Lemma 5.7. *For $r \in \mathbb{N}$, let $M^{(1)}, \ldots, M^{(r)} \in \mathbb{R}^{m\times m}$ be independent copies of a random matrix M, where with probability 1, it holds that (1) M is column stochastic: it has nonnegative entries and column sums equal to 1, and (2) $M \geq \mathrm{I}/2$ entrywise. Suppose that $\overline{M} = \mathbb{E}[M]$ is doubly stochastic: all entries are nonnegative and all row and column sums are 1, and suppose that the second largest eigenvalue of $\frac{1}{2}(\overline{M} + \overline{M}^\top)$ is at most $1 - \lambda_2$. Then,*

$$\max_{u,u'\in[m]} \mathbb{E}_{M^{(1)},\ldots,M^{(r)}} \left[\|M^{(r)} \ldots M^{(1)}(e_u - e_{u'})\|_2^2\right] \leq 2m(1-\lambda_2)^r\,.$$

Proof. Let $X \in \mathbb{R}^{m\times m}$ be an arbitrary real matrix. For each $i \in [m]$, let $X_{i,\star}$ denote the i-th row of X, and let $X_{\star,j}$ denote the j-th column. Let $\bar{X} \coloneqq \frac{1}{m}\sum_j X_{\star,j}$ denote the average of the columns. We define the following quantity, which intuitively captures how similar the columns of X are:

$$\mathcal{E}(X) \coloneqq \sum_{j=1}^{m} \|X_{\star,j} - \bar{X}\|_2^2\,.$$

We will argue that for any matrix $X \in \mathbb{R}^{m\times m}$, it holds that $\mathbb{E}_M[\mathcal{E}(XM)] \leq (1-\lambda_2)\mathcal{E}(X)$. Intuitively, this means that multiplying by M on the right makes the columns of X become more similar.

This bound immediately implies Lemma 5.7 via the following argument. We have

$$\mathbb{E}_{M^{(r)},\ldots,M^{(2)}}[\mathbb{E}_{M^{(1)}}[\mathcal{E}(M^{(r)}\ldots M^{(2)}M^{(1)})]] \leq (1-\lambda_2)\,\mathbb{E}_{M^{(r)},\ldots,M^{(2)}}[\mathcal{E}(M^{(r)}\ldots M^{(2)})] \leq (1-\lambda_2)^r\mathcal{E}(\mathrm{I}) \leq m(1-\lambda_2)^r\,,$$

by induction on r, since $\mathcal{E}(\mathrm{I}) \leq m$. Letting $Y = M^{(r)}\ldots M^{(1)}$, we have that for any distinct $u, u' \in [m]$,

$$\|Y(e_u - e_{u'})\|_2^2 = \|Y_{\star,u} - Y_{\star,u'}\|_2^2 \leq 2\|Y_{\star,u} - \overline{Y}\|_2^2 + 2\|Y_{\star,u'} - \overline{Y}\|_2^2 \leq 2\mathcal{E}(Y)\,.$$

Hence,

$$\max_{u,u'\in[m]} \mathbb{E}_{M^{(1)},\ldots,M^{(r)}}[\|M^{(r)}\ldots M^{(1)}(e_u - e_{u'})\|_2^2] \leq 2\,\mathbb{E}_{M^{(1)},\ldots,M^{(r)}}[\mathcal{E}(M^{(r)}\ldots M^{(1)})] \leq 2m(1-\lambda_2)^r\,,$$

as required.

It thus remains to show that for any $X \in \mathbb{R}^{m\times m}$, it holds that $\mathbb{E}_M[\mathcal{E}(XM)] \le (1-\lambda_2)\mathcal{E}(X)$.

First, we observe that because M is column stochastic with probability 1, we can without loss of generality assume that $\bar{X} = \mathbf{0}$. In other words, we may replace X with $X - \bar{X}\mathbf{1}^\top$. Indeed, by definition we have that $\mathcal{E}(X) = \mathcal{E}(X - \bar{X}\mathbf{1}^\top)$. Since M is column stochastic with probability 1, it holds that $\mathbf{1}^\top M = \mathbf{1}^\top$, and therefore $\mathcal{E}((X - \bar{X}\mathbf{1}^\top)M) = \mathcal{E}(XM - \bar{X}\mathbf{1}^\top) = \mathcal{E}(XM)$. Note that if $\bar{X} = \mathbf{0}$, then we have $\mathcal{E}(X) = \sum_{j=1}^m \|X_{\star,j}\|_2^2 = \|X\|_F^2 = \sum_{i=1}^m \|X_{i,\star}\|_2^2$.

Next, we observe that for any $y_1, \dots, y_m \in \mathbb{R}^m$, we have that $\sum_{j=1}^m \|y_j - \bar{y}\|_2^2 \le \sum_{j=1}^m \|y_j\|_2^2$ where $\bar{y} = \frac{1}{m}\sum_{j=1}^m y_j$. Therefore, we have

$$\mathbb{E}_M[\mathcal{E}(XM)] \le \mathbb{E}_M\left[\sum_{j=1}^m \|(XM)_{\star,j}\|_2^2\right].$$

Now, by assumption we have that M is entrywise at least I/2 with probability 1. Therefore, letting $R = 2M - \mathrm{I}$, we have that R is also a column stochastic matrix with probability 1. Since $\overline{M} = \mathbb{E}[M]$ is a doubly stochastic matrix, it follows that the mean $\overline{R} = \mathbb{E}_M[R]$ is also a doubly stochastic matrix. Using $M = \frac{1}{2}(\mathrm{I} + R)$, we have

$$\mathbb{E}_M[\mathcal{E}(XM)] \le \frac{1}{4}\,\mathbb{E}_M\left[\sum_{j=1}^m \|X_{\star,j} + (XR)_{\star,j}\|_2^2\right] = \frac{1}{4}\,\mathbb{E}_M\left[\sum_{j=1}^m \left(\|X_{\star,j}\|_2^2 + \|(XR)_{\star,j}\|_2^2 + 2\langle X_{\star,j}, (XR)_{\star,j}\rangle\right)\right].$$

We now bound the terms in the above sum. First, we have $\sum_{j=1}^m \|X_{\star,j}\|_2^2 = \mathcal{E}(X)$, as we have assumed that $\bar{X} = \mathbf{0}$. Second, because R is column stochastic with probability 1 and $\overline{R} = \mathbb{E}_M[R]$ is doubly stochastic, we have

$$\begin{aligned}\mathbb{E}_M\left[\sum_{j=1}^m \|(XR)_{\star,j}\|_2^2\right] &= \mathbb{E}_M\left[\sum_{j=1}^m \left\|\sum_{i=1}^m R_{ij}X_{\star,i}\right\|_2^2\right] \le \mathbb{E}_M\left[\sum_{j=1}^m \sum_{i=1}^m R_{ij}\cdot\|X_{\star,i}\|_2^2\right] \\ &= \sum_{i=1}^m \|X_{\star,i}\|_2^2\left(\sum_{j=1}^m \mathbb{E}_M[R_{ij}]\right) = \sum_{i=1}^m \|X_{\star,i}\|_2^2 = \mathcal{E}(X),\end{aligned}$$

where the inequality is by Jensen's inequality, since $R_{ij} \ge 0$ for all i, j and $\sum_{i=1}^m R_{ij} = 1$ for all j, and the penultimate equality is because $\overline{R} = \mathbb{E}_M[R]$ is doubly stochastic. This bounds the second term in the sum.

We now bound the third and final term. We have

$$\begin{aligned}&\sum_{j=1}^m \langle X_{\star,j}, (XR)_{\star,j}\rangle = \sum_{j=1}^m \langle X_{\star,j}, \sum_{i=1}^m R_{ij}X_{\star,i}\rangle = \sum_{j=1}^m\sum_{k=1}^m\sum_{i=1}^m R_{ij}X_{kj}X_{ki} \\ &= \sum_{k=1}^m\left(\sum_{i=1}^m\sum_{j=1}^m R_{ij}X_{kj}X_{ki}\right) = \sum_{k=1}^m (X_{k,\star})^\top R X_{k,\star} = \sum_{k=1}^m \frac{1}{2}(X_{k,\star})^\top\left(R + R^\top\right)X_{k,\star},\end{aligned}$$

where we view $X_{k,\star}$ as a column vector and the last equality uses that $(X_{k,\star})^\top R X_{k,\star}$ is a scalar. Now, because $\bar{X} = \mathbf{0}$, it follows that each row $X_{k,\star}$ is orthogonal to $\mathbf{1}$. Since $\overline{R} = \mathbb{E}_M[R]$ is a doubly stochastic matrix, the top eigenvector of $\frac{1}{2}(\overline{R} + \overline{R}^\top)$ is $\mathbf{1}$ with eigenvalue 1. Furthermore, because $\frac{1}{2}(\overline{M} + \overline{M}^\top)$ has second eigenvalue at most $1 - \lambda_2$, it follows that $\frac{1}{2}(\overline{R} + \overline{R}^\top)$ has second eigenvalue at most $2(1 - \lambda_2) - 1 = 1 - 2\lambda_2$.

Therefore, by taking expectation over M, we have

$$\mathbb{E}_M\left[\sum_{j=1}^{m} \langle X_{\star,j}, (XR)_{\star,j}\rangle\right] = \sum_{k=1}^{m} \frac{1}{2}(X_{k,\star})^\top \left(\overline{R} + \overline{R}^\top\right) X_{k,\star} \le \sum_{k=1}^{m} \|X_{k,\star}\|_2^2 (1 - 2\lambda_2) = (1 - 2\lambda_2)\mathcal{E}(X)\,,$$

where the inequality uses the second eigenvalue bound on $\frac{1}{2}(\overline{R} + \overline{R}^\top)$ along with the fact that each row $X_{k,\star}$ is orthogonal to $\mathbf{1}$, and the last equality uses that $\mathcal{E}(X) = \|X\|_F^2$ since $\bar{X} = \mathbf{0}$.

Combining all three bounds, we have shown that

$$\mathbb{E}_M[\mathcal{E}(XM)] \le \frac{1}{4}\left(\mathcal{E}(X) + \mathcal{E}(X) + 2(1 - 2\lambda_2)\mathcal{E}(X)\right) = (1 - \lambda_2)\mathcal{E}(X)\,,$$

as required, which finishes the proof. □

Statement of AI Use

This statement is intended solely to describe the role of AI in the research process and should not be read as an endorsement of such systems or of the companies that produce them. In the production of this paper, we have used AI models for the following tasks.

Activity	**AI used**
Asking the research question	[No]
Coming up with the approach	[No]
Proof development	[Yes]
Writing and exposition	[No]
Checking for bugs	[Yes]
Other supporting tasks	[Yes]

Details. The question of proving a $k \le O_\delta(\log^2 n)$ bound for binary linear 3-LCCs is a known problem left open by previous work [KM24a, Yan24, AG24, KM24b]. We began working on this question in September 2024, and we developed the high-level approach used in Section 5 ourselves by January 2025. Specifically, we set out to show that the Kikuchi graph (Definition 5.1) has an odd-coloured cycle by bounding the expected local variance of degrees $\mathcal{V}$ (Equation (2)), and we had proved Lemma 5.3 ourselves, which shows that bounding $\mathcal{V}$ is sufficient to finish the proof. We additionally had some more informal intuition that if we could prove that the directed graph $D_{\mathcal{H}}$ (defined in Lemma 4.1) was an edge expander, then we could show that $\mathcal{V} \le (|E|/k|V|)^2$. However, we did not write down a full formal proof of this, as there was an issue that a random walk on the directed graph $D_{\mathcal{H}}$ could be periodic, which would prevent it from mixing.

On Oct. 5, we gave GPT-6 Astra our notes, which included the main approach of using odd-coloured cycles in Kikuchi graphs with chain-derived constraints, our proof of Lemma 5.3, and the conclusion that showing $\mathcal{V} \leq (|E|/k|V|)^2$ is sufficient to prove Theorem 1. From this input, GPT-6 Astra completed the proof. Over the course of several more interactions, we substantially simplified Astra's completion of our proof approach, obtaining the proof that we have presented in this paper.

From our perspective, the missing key ideas that Astra provided are as follows. Firstly, in Section 5, while nearly all of the major ideas are ours, we crucially did not see the simple idea of adding the "fake tuples" to $\mathcal{T}$. The addition of fake tuples makes the corresponding random walk *lazy*, and this deals with the periodicity issues we were encountering in a very clean manner. It also makes it so any walk can always be continued, even if, say, $\mathcal{H}_u$ does not have a hyperedge containing some $a \in [n]$. Astra also provided a proof of Claim 5.6, and while it is a very nice observation that makes the proof particularly clean, in our view it is not an essential component. In fact, in the initial proof Astra provided, the proof contained a different, more technical statement.

Secondly, we did not see how to execute the decomposition done in Section 4, and this was entirely provided by Astra. In our view, the decomposition has a very intriguing strategy of working in the *quotient spaces* $\mathcal{W}/\mathcal{W}_{i-1}$ and retaining only indices whose vectors remain nonzero, which preserves large out-degree in the remaining vertex set. This was a technical issue that we did not see how to overcome when we thought about using edge-expander decompositions ourselves. The decomposition also has some additional technical components that ensure the resulting directed graph $D_{\mathcal{H}}$ produced is Eulerian. However, as we have noted in Section 4, this is not essential as one can prove a version of Lemma 4.1 without the Eulerian assumption, albeit with an exponential dependence on $1/\lambda$.

Finally, we note that the proof of Fact 3.2 presented here is due to Astra. However, we had proven essentially the same statement ourselves, with just a slightly weaker bound of $k' \leq O(\log|V|)$, which works equally well for its application. We have chosen to use the AI proof in this paper because it is cleaner.

References

[ABS+25] Noga Alon, Matija Bucić, Lisa Sauermann, Dmitrii Zakharov, and Or Zamir. Essentially tight bounds for rainbow cycles in proper edge-colourings. *Proceedings of the London Mathematical Society*, 130(4):e70044, 2025. 2, 3

[AG24] Omar Alrabiah and Venkatesan Guruswami. Near-tight bounds for 3-query locally correctable binary linear codes via rainbow cycles. In *65th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2024, Chicago, IL, USA, October 27-30, 2024*. IEEE, 2024. 1, 2, 3, 6, 20

[AGKM23] Omar Alrabiah, Venkatesan Guruswami, Pravesh K. Kothari, and Peter Manohar. A near-cubic lower bound for 3-query locally decodable codes from semirandom CSP refutation. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing, STOC 2023, Orlando, FL, USA, June 20-23, 2023*, pages 1438–1448. ACM, 2023. 2, 5, 6

[ALM+98] Sanjeev Arora, Carsten Lund, Rajeev Motwani, Madhu Sudan, and Mario Szegedy. Proof verification and the hardness of approximation problems. *Journal of the ACM (JACM)*, 45(3):501–555, 1998. 1

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

[BCG20] Arnab Bhattacharyya, L Sunil Chandran, and Suprovat Ghoshal. Combinatorial lower bounds for 3-query ldcs. In *11th Innovations in Theoretical Computer Science Conference (ITCS 2020)*, volume 151, page 85. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2020. 6

[BDL13] Abhishek Bhowmick, Zeev Dvir, and Shachar Lovett. New bounds for matching vector families. In *Symposium on Theory of Computing Conference, STOC'13, Palo Alto, CA, USA, June 1-4, 2013*, pages 823–832. ACM, 2013. 1

[BFNW93] László Babai, Lance Fortnow, Noam Nisan, and Avi Wigderson. BPP has subexponential time simulations unless EXPTIME has publishable proofs. *Comput. Complex.*, 3:307–318, 1993. 1

[BGT17] Arnab Bhattacharyya, Sivakanth Gopi, and Avishay Tal. Lower bounds for 2-query lccs over large alphabet. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2017)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2017. 2

[BHKL25] Arpon Basu, Jun-Ting Hsieh, Pravesh K. Kothari, and Andrew D. Lin. Improved lower bounds for all odd-query locally decodable codes. In *66th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2025, Sydney, Australia, December 14-17, 2025*, pages 1262–1285. IEEE, 2025. 5

[BIW10] Omer Barkol, Yuval Ishai, and Enav Weinreb. On locally decodable codes, self-correctable codes, and t-private pir. *Algorithmica*, 58(4):831–859, 2010. 1

[BK95] Manuel Blum and Sampath Kannan. Designing programs that check their work. *Journal of the ACM (JACM)*, 42(1):269–291, 1995. 1

[BLR93] Manuel Blum, Michael Luby, and Ronitt Rubinfeld. Self-testing/correcting with applications to numerical problems. *Journal of computer and system sciences*, 47(3):549–595, 1993. 1

[DLS13] Shagnik Das, Choongbum Lee, and Benny Sudakov. Rainbow Turán problem for even cycles. *European Journal of Combinatorics*, 34(5):905–915, 2013. 3

[Dvi10] Zeev Dvir. On matrix rigidity and locally self-correctable codes. In *Proceedings of the 25th Annual IEEE Conference on Computational Complexity, CCC 2010, Cambridge, Massachusetts, USA, June 9-12, 2010*, pages 291–298. IEEE Computer Society, 2010. 1

[Dvi12] Zeev Dvir. Incidence theorems and their applications. *CoRR*, abs/1208.5073, 2012. 1

[Dvi16] Zeev Dvir. Lecture notes on linear locally decodable codes. https://www.cs.princeton.edu/~zdvir/LDCnotes/LDC8.pdf, Fall 2016. 2

[GKM22] Venkatesan Guruswami, Pravesh K. Kothari, and Peter Manohar. Algorithms and certificates for Boolean CSP refutation: smoothed is no harder than random. In *STOC '22: 54th Annual ACM SIGACT Symposium on Theory of Computing, Rome, Italy, June 20 - 24, 2022*, pages 678–689. ACM, 2022. 5

[GKO+17] Sivakanth Gopi, Swastik Kopparty, Rafael Mendes de Oliveira, Noga Ron-Zewi, and Shubhangi Saraf. Locally testable and locally correctable codes approaching the gilbert-varshamov bound. In *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2017, Barcelona, Spain, Hotel Porta Fira, January 16-19*, pages 2073–2091. SIAM, 2017. 1

[GKST02] Oded Goldreich, Howard J. Karloff, Leonard J. Schulman, and Luca Trevisan. Lower bounds for linear locally decodable codes and private information retrieval. In *Proceedings of the 17th Annual IEEE Conference on Computational Complexity, Montréal, Québec, Canada, May 21-24, 2002*, pages 175–183. IEEE Computer Society, 2002. 2, 5

[HKM23] Jun-Ting Hsieh, Pravesh K. Kothari, and Sidhanth Mohanty. A simple and sharper proof of the hypergraph Moore bound. In *Proceedings of the 2023 ACM-SIAM Symposium on Discrete Algorithms, SODA 2023, Florence, Italy, January 22-25, 2023*, pages 2324–2344. SIAM, 2023. 5

[HMS+13] Hao Huang, Jie Ma, Asaf Shapira, Benny Sudakov, and Raphael Yuster. Large feedback arc sets, high minimum degree subgraphs, and long cycles in Eulerian digraphs. *Combinatorics, Probability and Computing*, 22(6):859–873, 2013. 9

[IK99] Yuval Ishai and Eyal Kushilevitz. Improved upper bounds on information-theoretic private information retrieval (extended abstract). In *Proceedings of the Thirty-First Annual ACM Symposium on Theory of Computing, May 1-4, 1999, Atlanta, Georgia, USA*, pages 79–88. ACM, 1999. 1

[Jan23] Oliver Janzer. Rainbow Turán number of even cycles, repeated patterns and blow-ups of cycles. *Israel Journal of Mathematics*, 253(2):813–840, 2023. 3

[JM25] Oliver Janzer and Peter Manohar. A $k^{\frac{q}{q-2}}$ lower bound for odd query locally decodable codes from bipartite kikuchi graphs. In *66th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2025, Sydney, Australia, December 14-17, 2025*, pages 75–85. IEEE, 2025. 5

[JS24] Oliver Janzer and Benny Sudakov. On the Turán number of the hypercube. In *Forum of Mathematics, Sigma*, volume 12, page e38. Cambridge University Press, 2024. 3

[KLLT24] Jaehoon Kim, Joonkyung Lee, Hong Liu, and Tuan Tran. Rainbow cycles in properly edge-colored graphs. *Combinatorica*, 44(4):909–919, 2024. 3

[KM24a] Pravesh K. Kothari and Peter Manohar. An exponential lower bound for linear 3-query locally correctable codes. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing, STOC 2024, Vancouver, BC, Canada, June 24-28, 2024*, pages 776–787. ACM, 2024. 1, 2, 3, 4, 5, 6, 7, 20

[KM24b] Pravesh K. Kothari and Peter Manohar. Exponential lower bounds for smooth 3-lccs and sharp bounds for designs. In *65th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2024, Chicago, IL, USA, October 27-30, 2024*. IEEE, 2024. 1, 2, 3, 4, 5, 6, 7, 10, 20

[KMRS16] Swastik Kopparty, Or Meir, Noga Ron-Zewi, and Shubhangi Saraf. High-rate locally-correctable and locally-testable codes with sub-polynomial query complexity. In *Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2016, Cambridge, MA, USA, June 18-21, 2016*, pages 202–215. ACM, 2016. 1

[KMSV07] Peter Keevash, Dhruv Mubayi, Benny Sudakov, and Jacques Verstraëte. Rainbow Turán problems. *Combinatorics, Probability and Computing*, 16(1):109–126, 2007. 3

[KSY11] Swastik Kopparty, Shubhangi Saraf, and Sergey Yekhanin. High-rate codes with sublinear-time decoding. In *Proceedings of the 43rd ACM Symposium on Theory of Computing, STOC 2011, San Jose, CA, USA, 6-8 June 2011*, pages 167–176. ACM, 2011. 1

[KW04] Iordanis Kerenidis and Ronald de Wolf. Exponential lower bound for 2-query locally decodable codes via a quantum argument. *Journal of Computer and System Sciences*, 69(3):395–420, 2004. 2

[LFKN90] Carsten Lund, Lance Fortnow, Howard J. Karloff, and Noam Nisan. Algebraic methods for interactive proof systems. In *31st Annual Symposium on Foundations of Computer Science, St. Louis, Missouri, USA, October 22-24, 1990, Volume I*, pages 2–10. IEEE Computer Society, 1990. 1

[RS96] Ronitt Rubinfeld and Madhu Sudan. Robust characterizations of polynomials with applications to program testing. *SIAM Journal on Computing*, 25(2):252–271, 1996. 1

[Sha90] Adi Shamir. Ip=pspace. In *31st Annual Symposium on Foundations of Computer Science, St. Louis, Missouri, USA, October 22-24, 1990, Volume I*, pages 11–15. IEEE Computer Society, 1990. 1

[Tom24] István Tomon. Robust (rainbow) subdivisions and simplicial cycles. *Advances in Combinatorics*, 2024:1, 37 pp, 2024. 3

[Tre04] Luca Trevisan. Some applications of coding theory in computational complexity. *arXiv preprint cs/0409044*, 2004. 1

[Yan24] Tal Yankovitz. A stronger bound for linear 3-lcc. In *65th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2024, Chicago, IL, USA, October 27-30, 2024*. IEEE, 2024. 1, 2, 3, 4, 5, 6, 7, 20

[Yek12] Sergey Yekhanin. Locally decodable codes. *Foundations and Trends in Theoretical Computer Science*, 6(3):139–255, 2012. 1

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il