THE UNIVERSITY OF CHICAGO

SOME RESULTS IN LOW-DEPTH CIRCUIT COMPLEXITY

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF COMPUTER SCIENCE

BY
YUAN LI

CHICAGO, ILLINOIS
AUGUST 2017

TABLE OF CONTENTS

ACKNOWLEDGMENTS

First and foremost, I would like to thank my advisors Andy Drucker and Janos Simon for guiding me through my PhD, for their constant support in the past years, for their encouragement in my academic pursuits, and for giving me the freedom to explore problems I like.

I feel fortunate to collaborate with Andy Drucker, Alexander Razborov, and Ben Rossman. I would like to thank Andy for sharing a lot of interesting problems and ideas, for teaching me taking notes (to help iterate on simple ideas/observations, etc.), for his positive attitudes towards difficulties, and for his availability and patience to discuses problems, where a lot of his comments lead to fruitful directions. I would like to thank Alexander Razborov for his advice during my first 2-3 years, from which I learned a lot, including a bird's-eye view of theory to look at theory, and the courage to attack important problems. I am very thankful to Ben Rossman, for his generosity, and for many beautiful results he has proved in circuit complexity.

Many thanks to Laci Babai for his help in the past six years. I remember he commented on every email I sent when solving challenge problems in his Discrete Mathematics class; he corrected a lot of my wrong pronunciations; he talked about research problems passionately in the summer REU program; and he introduced me to Andy Drucker when my research got badly stuck. I want to thank Ketan Mulmuley for supporting me exploring problems in Geometric Complexity Theory, although I didn't manage to do something there.

I want to thank Haris Angelidakis, Bob Bartlett, Charisee Chiw, Julia Chuzhoy, Gokalp Demirci, Mrinalkanti Ghosh, Pooya Hatami, Margaret Jaffrey, David Kim, Denis Pankratov, John Santerre, Jiajun Shen, Li-Yang Tan, Shubhendu Trivedi, Joseph Tsang, Madhur Tulsiani, Pratik Worah, Angela Wu, and Liwen Zhang for their help during my time in the University of Chicago.

Finally, I want to thank my parents Wenhong Li, Bing Gong and my girlfriend Yunfan Zhang for their love and support.

ABSTRACT

Circuit complexity is a branch of computational complexity theory in which we study complexity measures including size and depth, where the computation models are circuits instead of Turing machines. Little is known for general circuits, while there are nontrivial results in some restricted circuit models. This dissertation consists of three contributions related to low-depth circuit complexity.

The first contribution answers the following question: what is the minimum depth required to compute good codes by linear-size circuits? Good codes have both constant code rate and constant relative distance, and size of the circuit is defined to be the number of wires instead of gates since the fan-in is unbounded. We prove the answer is $\Theta(\alpha(n))$, where $\alpha(n)$ is the inverse Ackermann function. The lower bound applies to *unrestricted* circuits, and the proof is a graph-theoretic argument relying on a lemma by Raz and Shpilka (2003), and a connection between good codes and *densely regular* graphs by Gál *et al.* (2013). The upper bound is inspired by the recursive construction of superconcentrators; we prove a similar recursion exists. The upper bound tightens the previous result $O(\log^* n)$ by Gál *et al.*

In the algebraic setting over large field, we show a close connection between superconcentrators and good codes. For example, we prove *any* superconcentrator with n inputs and $n + \Theta(n)$ outputs computes a good code, by replacing each vertex with an addition gate and assigning the coefficient for each edge uniformly at random. We also show the potential application of the above "superconcentrator codes" in Network Coding.

The second contribution is about conservative circuits and routing networks. Our original motivation is to study the circuit complexity of the (cyclic) Shift operator, which takes $n+\log n$ input bits and outputs n bits. We propose the definition of *Expansive Routing Family (ERF)* networks based on some entropy property satisfied by the Shift operator, with the aim to extend lower bounds from conservative circuits to a more general model which allows arbitrary preprocessing and a final layer of postprocessing. However, it turns out there exist small-size ERF networks. For depth 2 and 3, we obtain tight bounds $\Theta\left(n(\log n/\log\log n)^2\right)$

and $\Theta(n \log\log n)$ respectively; for depth $d \geq 4$, we prove lower bound $\Omega_d\left(\lambda_d(n) \cdot n\right)$.

We propose the research challenge to develop a powerful and broadly-applicable set of techniques for both upper bounding and lower bounding the wire complexity of routing networks for given specific demands. Towards this challenge, we significantly generalize the Pippenger-Yao lower bound for shifts based on Shannon entropy, which can be applied to any multirequests; and for constant depth d, we construct size-$O\left(dn^{1+\frac{1}{d}}\right)$ routing network realizing all shifts, where the size is optimal up to a constant factor.

The third contribution is about the AC^0 complexity of subgraph isomorphism. Let $\textsc{Subgraph}(P)$ denote the problem of deciding whether a given n-vertex graph G contains a subgraph isomorphic to P. Let $C(P)$ denotes smallest possible exponent $C(P)$ for which $\textsc{Subgraph}(P)$ possesses bounded-depth circuits of size $n^{C(P)+o(1)}$. Motivated by the previous research in the area, we also consider its "colorful" version, and the *average-case* version $\textsc{Subgraph}_{\mathsf{ave}}(P)$ under the Erdős-Rényi random graphs. Let us define $C_{\mathsf{col}}(P)$ and $C_{\mathsf{ave}}(P)$ analogously to $C(P)$.

For the average-case version, we give a *characterization* of $C_{\mathsf{ave}}(P)$ in purely combinatorial terms up to a multiplicative factor of 2. The lower bound closely follows Rossman's techniques [73]. For the worst-case colored version, we prove $C_{\mathsf{col}}(P) = \Omega\left(\frac{tw(P)}{\log tw(P)}\right)$, where $tw(P)$ denotes the *tree width* of P. The lower bound is obtained in the average case, and is tight up to a logarithmic factor.

We also prove some structural results suggesting that the colorful version of the subgraph isomorphism problem is much better structured and well-behaved than the standard (worst-case, uncolored) one. This suggests that new techniques may be required to solve the worst-case uncolored version.

The first two contributions are joint work with Andrew Drucker [23, 24], and the third contribution is joint with Alexander Razborov and Benjamin Rossman [54].

CHAPTER 1

INTRODUCTION

1.1 Background

Circuit complexity is a branch of complexity theory, where the computational models are *circuits* instead of Turing machines. Generally, a circuit C is a directed acyclic graph with some vertices specified as inputs and outputs, where inputs have no incoming edges and outputs may have outgoing edges; and each non-input vertex (*gate*) v computes a function from $\{0,1\}^{\deg(v)^+} \to \{0,1\}$ over the Boolean setting, where $\deg(v)^+$ denotes the *indegree* of v, which is called the *fan-in* of the gate. If C has n inputs and m outputs, we can a define function (or *operator*, if $m > 1$) $\{0,1\}^n \to \{0,1\}^m$ computed by C in the natural way.

Different types of circuits impose different restrictions on the functions allowed to be computed by individual gates, and the topology of the underlying graph including its size, depth, fan-in, and so on. For example, the *complete binary basis* model is the class of circuits where each gate has fan-in at most 2, and can compute *any* unary or binary function. Size and depth are two major complexity measures, where *size* is defined to be the number of gates or the number of wires (edges), depending on the circuit model, and *depth* is the length of the longest path from input to output.

In some cases, the types of allowed gates does not matter. For example, using only a constant number of AND, OR, NOT gates, we can simulate any binary/unary function, and thus both size and depth only vary by a constant factor in these two models, provided fan-out is unrestricted.

In general, circuits are universal computation models, because any function $f : \{0,1\}^n \to \{0,1\}$ can be computed by circuit of size $O\left(\frac{2^n}{n}\right)$ over the complete binary basis. Also we can simulate Turing machines using circuits. Let $(f_n)_{n\geq 1}$, where $f_n : \{0,1\}^n \to \{0,1\}$, be a decision problem computable by some Turing machine in *time* $T(n)$. It is known that f_n can be computed by size-$O(T(n) \log T(n))$ circuit. On the other hand, circuits are more powerful

in the sense of non-uniformity, that is, functions $(f_n)_{n\geq 1}$ which are efficiently computable by small-size circuits may not be Turing machine computable since the circuit may vary in n arbitrarily as n grows.

Given a decision problem encoded by a sequence of Boolean functions $(f_n)_{n\geq 1}$, where $f_n : \{0,1\}^n \to \{0,1\}$, and fixing a circuit model, we are interested in finding a sequence of circuits $(\mathsf{C}_n)_{n\geq 1}$ computing $(f_n)_{n\geq 1}$, while minimizing certain complexity measures. And we would be interested in both *lower bounds* and *upper bounds*. The ultimate goal in circuit complexity is to prove that some function in NP cannot be computed by polynomial-size circuits, which will imply $P \neq NP$. However, it seems that we are very far from this goal, in terms of both results proved and techniques available.

Upper bounds may also be interesting, because the existence of small circuits may shed light on the design of efficient algorithms. Sometimes, there exist surprisingly small circuits satisfying certain properties, which may help us better understand the lower bound approach.

Keeping the ultimate goal in mind, the reality is somewhat embarrassing. For the complete binary basis model, only $3n - o(n)$ lower bounds were known, e.g., [13], which was recently improved to $\left(3 + \frac{1}{86}\right) n - o(n)$ [30]. For the obvious reason, researchers turn their attention to *restricted* circuit models, and some highly nontrivial results have been obtained. Such restricted models include

- *formulas*, where each gate has fan-out at most one. In other words, intermediate computation results cannot be reused;
- *monotone circuits*, where each gate is of fan-in 2 and is either AND or OR, and thus the circuits can only compute monotone functions;
- AC^0 *circuits*, where the size is polynomial in n, and depth of circuits is a constant, [1] and only AND, OR, NOT gates (with unbounded-fan) are allowed;

1. The constant can be arbitrarily large but must remain fixed as the number of inputs n goes to infinity.

- $AC^0[m]$ *circuits*, which are AC^0 circuits adding MOD_m gates, where MOD_m outputs 1 if and only if the sum of the inputs is a multiple of m;
- ACC^0 *circuits*, which are the collection of $\mathrm{AC}^0[m]$ circuits for all $m \geq 2$;
- *arithmetic circuits*, where each gate computes either $+$ or $\times$ over some underlying field;
- *conservative circuits*, where the inputs have two parts, say x and i, and all gates become *projection* gates after i is fixed;
- *unrestricted circuit of bounded depth*, where each gate can compute *any* function, and the size is defined to be the number of wires since fan-in is unbounded. Observe that any single-output function can be computed by one gate and n wires, and thus we study the wire complexity of *operators* in this model.

One hope is that understanding restricted circuit models will help us understand the unrestricted case better, and we may be able to remove the restrictions gradually. On the other hand, key results in this area also involve beautiful insights, which are easy to state and appreciate. In the next section, we review some (far from comprehensive) known results in this field classified by the methods.

1.2 Methods in Circuit Complexity

Counting

In the counting method (due to Shannon), we count the number of small-size circuits, and the number of different functions, and we conclude that there *exists* some function which cannot be computed by a small-size circuit. For the complete binary basis, Shannon (1949) proved that most n-variable Boolean functions require circuit size at least $\frac{2^n}{n}$ for sufficiently large n [80]. For formulas, Riordan and Shannon (1942) proved that most n-variable Boolean functions require size $\Omega\left(\frac{2^n}{\log n}\right)$ [72]. Similar arguments can be carried to other circuit models, for example, linear (arithmetic) circuits [49].

Like many other difficult problems, it is easy to show a *random* object satisfies the desired property, while it is challenging to prove that some *explicit* object satisfies the desired property. (In theoretical computer science, we have several precise notions of explicitness, e.g., a sequence of functions $(f_n)_n$ may be considered explicit if the truth table of f_n can be generated by a Turing machine on input n, for all n.)

Gate Elimination

Gate elimination refers to the following method: for *any* circuit computing $f_n \in \mathcal{C}$, where $\mathcal{C}$ denotes a class of non-constant target functions, we can *set* one variable to constant such that the circuit gets simplified, that is, the number of gates (or other sophisticated complexity measure) is decreased by $\geq c$ per variable fixed, and the remaining function f_{n-1} still belongs to $\mathcal{C}$. By induction, we conclude that f_n has circuit lower bound $cn - o(n)$.

This method has been used to prove circuit lower bound since 1960s. To name a few, Paul [62] proved a $2.5n$ lower bound for a storage access function, and Blum [13] proved $3n - o(n)$ lower bound for a similar function. Recently, Find *et al.* [30] proved a $\left(3 + \frac{1}{86}\right) n - o(n)$ lower bound for *affine dispersers*, where explicit affine dispersers had been constructed by Ben-Sasson and Koparrty [11]. All the results above are proved under the complete binary basis. For other restricted basis, slightly better lower bounds are known, for example, $5n - o(n)$ lower bound was known when parity and its negation are forbidden [44].

One limitation of the gate elimination method is that, when simplifying the circuit, only *local* information about the circuit is used. It appears likely that the gate elimination-type methods cannot achieve superlinear lower bounds for *general* circuits.

Random Restrictions and Switching Lemma

Instead of setting one variable to constant, what if a fraction of variables are set to constants at random? This motivates the definition of *random restrictions*, which have been successfully applied to many circuit models. The argument is similar as the gate elimination

method, that is, by repeatedly applying random restrictions, we argue that the circuit can be simplified, i.e., the size or the depth is shrinking, and finally reach a contradiction. The random restriction method dates back to at least 1961, when Subbotovskaya used the idea to prove an $n^{\frac{3}{2}}$ lower bound for the formula size of parity function, that is, she showed that the size of any DeMorgan formula can be considerably reduced by setting some variables to constants randomly.

This method finds its most successful application in AC^0 circuits. Note that AC^0 circuits can be rearranged into *layered* circuits, where each layer consists of only AND or OR gates alternatively, and all NOT gates are pushed to the inputs. After applying a random restriction, if we can show that an AND of ORs can be rewritten as an OR of ANDs (or vice versa), then the depth of the circuit can be reduced by one. Proving the validity of this idea (in a so-called "switching lemma"), Furst, Saxe and Sipser [32] were the first to prove the parity function is not in AC^0, and it was improved by Ajtai [4], Yao [86], and Håstad [41]. The lemma proved by Håstad [41] is the most powerful, and has been widely applied for the AC^0 circuit model, which has become well understood.

Not too long ago, Rossman invented a *top-down argument* based on the switching lemma to prove average-case lower bounds for the k-clique [73] and st-connectivity problem [75]. Compared to the previous bottom-up argument, which repeatedly applies switching lemma to eliminate the last layer, the top-down argument involves some completely new ideas.

Approximation

The approximation argument goes like this:

(1) First, define a class of *good functions*, and define a distance measure over all Boolean function;

(2) Then, show that small-size circuit is *close to* good functions, while the candidate hard function is *far from* good.

In 1985, Razborov [70] proved the first superpolynomial lower bound of the clique function for monotone circuits, which is a landmark result. Later, it was improved to exponential by Alon and Boppana [2].

Besides monotone circuits, approximation method has also been successfully applied to $\mathrm{AC}^0[p]$ circuit, where p is a prime. Razborov [71] proved that the majority function is not in $\mathrm{AC}^0[2]$, and Smolensky [81] extended the result showing that the MOD_q function is not in $\mathrm{AC}^0[p]$, when p and q are coprime.

Graph-theoretic Methods

A graph property is some property of a circuit that depends only on its graph structure. Graph-theoretic arguments, articulated by Valiant in [84], consist of two steps:

(1) Show that, in order to compute a given function/operator, the circuit must satisfy some graph property P;

(2) Show that any graph satisfying property P must have a large number of edges.

Valiant observed that any circuit computing cyclic convolution must be a so called *superconcentrator*. A graph is a superconcentrator if for any equal-size subsets of inputs and outputs, there exists a family of vertex-disjoint paths connecting them. However, he showed the existence of superconcentrators with linear number of edges (in the number of its inputs and outputs).

For constant depth $d \geq 4$, Dolev *et al.* [21] proved that the minimum size of depth-d superconcentrator is $\Theta(\lambda_d(n) \cdot n)$, where $\lambda_2(n) = \Theta(\log(n))$, $\lambda_4(n) = \Theta(\log^*(n))$, and so on. For depth 3, Alon and Pudlák proved that the minimal size is $\Theta(n \log \log n)$. For depth 2, Radhakrishnan and Ta-Shama [78] proved that the size is $\Theta(n \log^2 n / \log \log n)$.

All the lower bounds for superconcentrators above imply *wire* lower bounds for bounded-depth circuits (with unbounded fan-in and *unrestricted* [2] gates) for certain operators includ-

2. Unrestricted gates can compute any function. Thus, under this model, we consider multi-output functions, called operators, and measure the size by the number of wires instead of gates.

ing cyclic convolution and the discrete Fourier transform. Using similar techniques, Raz and Shpilka [77] proved that for depth-d unrestricted circuits, matrix multiplication over $GF(2)$ requires $\Omega_d\left(\lambda_d(n) \cdot n\right)$ wires, where the proof is based on a lemma strengthening Pudlák's [65] and Dolev *et al.*'s [21].

Communication Complexity

Karchmer and Wigderson [52] related circuit depth with the complexity of the following communication game:

- Given function f, there are two players Zero and One, where player Zero receives an input x such that $f(x) = 0$ and player One receives an input y such that $f(y) = 1$;
- Two players send bits to each other, until they agree on an $i \in [n]$ such that $x_i \neq y_i$.

The minimum number of bits required is denoted by $C_{KW}(f)$ in the min-max way. They showed that $C_{KW}(f)$ is *exactly* the minimum depth among all circuits (over the complete binary basis) computing f. Using this connection, they proved that *st*-connectivity of an undirected graph on n vertices requires depth $\Omega(\log^2 n)$ for all *monotone* circuits.

Of course, our selections of results and methods are far from complete. Interested readers can read Jukna's book [50] to learn more about circuit complexity. In the next section, we will describe our contributions in the area.

1.3 Our Contributions

The thesis consists of three themes of results organized in three separate chapters, where the first two are joint work with Drucker [23, 24], and the third is joint with Razborov and Rossman [54].

Chapter 2: Good Codes

In Chapter 2, we answer the following question: what is the minimum depth required to compute good codes by a linear-size circuit? Since the fan-in is unbounded, size is defined to be the number of wires. A code $C : \{0,1\}^n \to \{0,1\}^m$ is called *good code* if $m = O(n)$ and the Hamming distance between any two distinct code words is at least $\Omega(n)$.

For the lower bound, we show that depth $\Omega(\alpha(n))$ is required by linear-size unrestricted circuits, where $\alpha(n)$ is the inverse of Ackermann function. The proof is a graph-theoretic argument relying on a lemma by Raz and Shpilka [77], and a connection between good codes and *densely regular graphs* by Gál *et al.* [34]. Let us mention that in [34], tight size lower bounds are proved for all *constant* depths.

For the upper bound, we show there *exist* good codes computable by linear-size linear (arithmetic) circuits of depth $O(\alpha(n))$, where the previous best result is of depth $O(\log^* n)$ [34], and is of depth $O(\log n)$ for explicit codes [82]. The proof is inspired by the recursive construction of superconcentrators [21]; we prove a similar recursion holds for good codes. This upper bound suggests that all the known explicit circuits for good codes are far from optimal in terms of *depth*.

In the algebraic setting over large field, we show a close connection between superconcentrators and good codes. For example, *any* superconcentrator with n inputs and cn outputs, $c > 1$, computes a good code with distance at least $(c-1)n$, by replacing each vertex with an addition gate and assigning the coefficient for each edge uniformly at random. It is known that any circuit computing good codes should satisfy some superconcentrator-type property [34, 82], our result shows a connection in the reverse direction. It suggests that there is a close connection between *graph-theoretic properties* of the circuit and *algebraic properties* of the function computed by that circuit. Hopefully, there are more such connections to be discovered.

We show that the above "arithmetic circuit" with the underlying graph being a superconcentrator satisfies some nice entropy properties (called *lossless decodable*), which may have

potential application in Network Coding, especially given that fact that there are explicit constructions of linear-size logarithmic-depth superconcentrators [1].

Chapter 3: Conservative Circuits and Routing Networks

Research in this chapter is motivated by studying the circuit complexity of the (cyclic) Shift operator, which takes two inputs $x \in \{0,1\}^n$ and $i \in \{0,1,\ldots,n-1\}$, and outputs $(x_{1+i}, x_{2+i}, \ldots, x_{n+i}) \in \{0,1\}^n$, where the indices are computed modulo n. Although this is a very simple function, its circuit complexity is poorly understood. For example, it is unknown whether it can be computable by $O(n)$-size, and $O(\log n)$-depth circuit. Even for constant depths, the gap between lower bounds and upper bounds is significant.

In [69], Pippenger and Yao (in a somewhat different theoretical context) proved that for constant depth d, the smallest *conservative circuit* computing the Shift operator is of size $\widetilde{O}\left(n^{1+\frac{1}{d}}\right)$. In order to extend the lower bounds to more general circuit model, we propose to study the *semi-conservative circuit*, where we allow arbitrary *preprocessing* on x and i individually, and then combined together by a conservative circuit with a final layer of arbitrary postprocessing. We observe that the Shift operator satisfies a nice entropy property (in the sense of Jukna [50]), and propose the definition of the *Expansive Routing Family (ERF)* network, of which lower bounds will imply circuit lower bounds for the semi-conservative circuits.

However, it turns out the ERF does not give good lower bounds comparable to $\Omega\left(n^{1+\frac{1}{d}}\right)$. For depth 2, we prove that the size of the smallest ERF network is $\Theta\left(n(\log n/\log\log n)^2\right)$; for depth 3, we prove the size is $\Theta(n\log\log n)$. For fixed depth $d \geq 4$, we prove lower bound $\Omega\left(\lambda_d(n)\cdot n\right)$ replying on the lemma by Raz and Shipilka [77]. Both the lower bounds and upper bounds are inspired by superconcentrators, but the upper bound constructions for routing networks are more complicated.

We propose the research challenge to develop a powerful and broadly-applicable set of techniques for both *upper bounding* and *lower bounding* the wire complexity of routing net-

work for given specific demands. This challenge seems natural and important, but is little studied in its full generality. Towards this challenge, we generalize the lower bound in [69] based on the concept of *entropy*, which can be applied to any multirequests, not only shifts; for the upper bound, we construct size-$O\left(dn^{1+\frac{1}{d}}\right)$ routing networks of constant depth d realizing all shift permutations, which matches the lower bound $dn^{1+\frac{1}{d}}$ by Pipppenger and Yao [69] up to a constant factor.

Chapter 4: Subgraph Isomorphism

In Chapter 4, we study the AC^0 complexity of subgraph isomorphism, where the subgraph (called a "pattern") is fixed. Let $\textsc{Subgraph}(P)$ denote the problem of deciding whether a given graph G contains a subgraph isomorphic to P.

Let $C(P)$ denotes smallest possible exponent $C(P)$ for which $\textsc{Subgraph}(P)$ possesses bounded-depth circuits of size $n^{C(P)+o(1)}$. Motivated by the previous research in the area, we also consider its "colorful" version $\textsc{Subgraph}_{\mathsf{col}}(P)$ in which the target graph G is $V(P)$-colored, and the *average-case* version $\textsc{Subgraph}_{\mathsf{ave}}(P)$ under the distribution $G(n, n^{-\theta(P)})$, where $\theta(P)$ is the *threshold exponent* of P. Let us define $C_{\mathsf{col}}(P)$ and $C_{\mathsf{ave}}(P)$ analogously to $C(P)$.

For the average-case version, we give a *characterization* of $C_{\mathsf{ave}}(P)$ in purely combinatorial terms up to a multiplicative factor of 2, which is a graph parameter called $\kappa(P)$. The lower bound closely follows Rossman's techniques in [73]. The upper bound is by observing the *dual form* of $\kappa(P)$, and it suggests that Rossman's techniques are tight for any patterns.

For the worst-case colored version, we prove $C_{\mathsf{col}}(P) = \Omega\left(\frac{tw(P)}{\log tw(P)}\right)$, where $tw(P)$ denotes the *tree width* of P. The lower bound is obtained in the average case, where the distribution for *colored random graphs* is carefully chosen tailored to P. Since this problem can be solved in $n^{tw(P)+O(1)}$ time [8] (and also by AC^0 circuit of the same size [9]), the lower bound is tight up to a logarithmic factor.

To compare the colored with uncolored version, we prove that if Q is a minor of P then

$\textsc{Subgraph}_{\mathsf{col}}(Q)$ is reducible to $\textsc{Subgraph}_{\mathsf{col}}(P)$ via a linear-size monotone projection. At the same time, we show that there is *no* monotone projection that reduces $\textsc{Subgraph}(M_3)$ to $\textsc{Subgraph}(P_3 + M_2)$ (P_3 is a path on 3 vertices, M_k is a matching with k edges, and "+" stands for the disjoint union). This result suggests that the colorful version of the subgraph isomorphism problem is much better structured and well-behaved than the standard (worst-case, uncolored) one. Thus different techniques may be required to solve the worst-case uncolored case.

CHAPTER 2
MINIMUM DEPTH REQUIRED TO COMPUTE GOOD CODES IN LINEAR SIZE

2.1 Introduction

Studying the complexity of encoding/decoding is a fundamental task in theoretical computer science. Given a class of codes with certain parameters, we are interested in understanding the complexity of computing them, where the complexity measures include time, space, parallelism, etc. Results of this kind may shed light on the designing of (explicit) codes to be used in practice. In this chapter, we consider the class of codes with constant code rate, and constant error-correcting rate, which are sometimes called *good codes*.

Bazzi and Mitter [14] proved that linear time and sublinear space is not sufficient to compute good codes in the computational model of *algebraic branching programs*. For AC^0 circuits, i.e., polynomial-size circuits consists of AND/OR/NOT gates with unbounded fan-in, it is not difficult to prove that they cannot compute good codes using the switching lemma. In fact, Lovett and Viola [53] proved that the statistical distance between the distribution sampled by any AC^0 circuit with the uniform distributions over any good code is inverse-polynomially close to one. Beck *et al.* [12] strengthened their results by showing the distance is exponentially close to one.

In [34], Gál *et al.* studied the complexity of computing good codes by circuits consisting of arbitrary gates with unbounded fan-in, which can be regarded as the most general and powerful circuit model. For all constant depths, they obtained tight bounds on the size of the circuits. For depth $d = 2$, the size is $\Theta\left(n(\log n/\log\log n)^2\right)$; for $d = 3$, the size is $\Theta(n \log\log n)$; for $d \geq 4$, the size is $\Theta_d\left(\lambda_d(n) \cdot n\right)$. The lower bound belongs to graph-theoretical methods, articulated by Valiant in [84]. Specifically, they proved that any circuit computing good codes should be *densely regular* (viewed as a directed acyclic graph); the lower bound follows from the edge lower bounds of densely regular graphs by Pudlák [65].

Their upper bounds are recursive constructions based on the concept of *range detectors*, and the circuits consist of XOR gates only.

Gelfand *et al.* proved that there *exist* good codes computable in linear time [22]. For the first time, Sipser and Spielman constructed good codes which are encodable by explicit linear-size logarithmic-depth circuits, also decodable by $O(n \log n)$-size circuits [82, 83]. Later, Guruswami and Indyk gave an improved construction with better code rate and error-correction rate [33]. Gál *et al.* [34] proved that if the depth is $\log^* n$, then there *exist* linear-size circuits computing good codes.

We believe it is interesting to answer the following question: what is the minimum depth required to compute good codes by linear-size circuits? Note that sublinear size is obviously impossible. We prove the answer is $\Theta(\alpha(n))$, where $\alpha(n)$ is the inverse Ackermman function, which is an extremely slow-growing function.

In the algebraic setting over large field $\mathbb{F}$, we prove that *any* superconcentrator with n inputs and cn outputs, $c > 1$, computes a good code with minimum distance $\geq (c-1) \cdot n$, if we replace all vertices by addition gates over $\mathbb{F}$, and choose the coefficients on each edge uniformly at random, assuming the size of the field $\mathbb{F}$ is large enough. It is known that any circuit computing good codes should satisfy some superconcentrator-type properties [34, 82]. Our result shows a partial result in the inverse direction, and it suggests that there is a close connection between *graph-theoretic properties* and *algebraic properties*.

Furthermore, we show the above (linear) arithmetic circuit with the underlying graph being a superconcentrator may be used in Network Coding. Because information gets "mixed" rapidly, such network codings have some nice entropy properties (called *lossless decodable* and will be defined in Section 2.5). Since there are small-size explicit constructions of superconcentrators [1], we hope such network codes will have some impact in practice.

2.2 Depth Lower Bound

Let $\mathbb{F}$ be any finite field. In this section, we will prove, in order to compute good codes $C : \mathbb{F}^n \to \mathbb{F}^{cn}$ using $O(n)$ wires, depth $\Omega(\alpha(n))$ is required under the *arbitrary gates* model. That is, each gate can compute any function $\mathbb{F}^s \to \mathbb{F}$, where s is the fan-in of the gate, and s is unbounded. The proof is based on the connection between good codes and *densely regular* graphs by Gál *et al.* [34], and a lemma by Raz and Shpilka [77].

Definition 1 (Definition 2.3 in [77])**.** *For a function f, define $f^{(i)}$ to be the composition of f with itself i times. For a function $f : \mathbb{N} \to \mathbb{N}$ such that $f(n) < n$ for all $n > 0$, define*

$$f^*(n) := \min\{i : f^{(i)}(n) \leq 1\}.$$

Let

$$\begin{aligned} \lambda_1(n) &:= \lfloor \sqrt{n} \rfloor \ , \\ \lambda_2(n) &:= \lceil \log n \rceil \ , \\ \lambda_d(n) &:= \lambda^*_{d-2}(n) \ . \end{aligned}$$

Definition 2 (Inverse Ackermann Function)**.** *For any positive integer n, let*

$$\alpha(n) := \min\{d : \lambda_d(n) \leq d\} \ .$$

In the literature, there is another definition of the inverse Ackermann function, which varies by at most a multiplicative constant factor.

Definition 3. *(Pudlák [65]) Let G be a directed acyclic graph with n inputs and n outputs. Let $0 < \epsilon, \delta$ and $0 \leq \mu \leq 1$. We say G is* (ϵ, δ, μ)-densely regular *if for every $k \in [\mu n, n]$, there are probability distributions $\mathcal{X}$ and $\mathcal{Y}$ on k-element subsets of inputs and outputs resp.,*

such that for every $i \in [n]$,

$$\Pr_{\mathbf{X}\in\mathcal{X}}[i \in \mathbf{X}] \leq \frac{k}{\delta n} , \qquad \Pr_{\mathbf{Y}\in\mathcal{Y}}[i \in \mathbf{Y}] \leq \frac{k}{\delta n} ,$$

and the expected number of vertex-disjoint paths from $\mathbf{X}$ *to* $\mathbf{Y}$ *is at least* ϵk *for randomly chosen* $\mathbf{X} \in \mathcal{X}$ *and* $\mathbf{Y} \in \mathcal{Y}$.

Code $C : \mathbb{F}^n \to \mathbb{F}^{cn}$ is called a (ρ, δ)-*good code* if $\rho \leq \frac{1}{c}$ and the Hamming distance between any two distinct code words is at least δn.

Corollary 4. *(Corollary 15 in Gál* et al. *[34]) Let* $0 < \rho, \delta < 1$ *be constants and* C *be a circuit computing a* (ρ, δ)*-good code. If we extend the circuit by* $(1-\rho)n$ *dummy inputs, then its underlying graph is* $\left(\rho\delta, \rho, \frac{1}{n}\right)$*-densely regular.*

The following lemma is a powerful strengthening the lower bounds of superconcentrators in [65] and [21].

Lemma 5 (Lemma 1.1 in Raz and Shpilka [77])**.** *For any* $0 < \epsilon < \frac{1}{400}$ *and any layered directed acyclic graph* G *of depth* d*, with more than* n *vertices and less than* $\epsilon \cdot n \cdot \lambda_d(n)$ *edges, the following is satisfied: for some* k *such that* $\sqrt{n} \leq k = o(n)$*, there exist subsets* $I \subseteq I_G, O \subseteq O_G$ *and* $V \subseteq V_G$ *such that*

- $|I|, |O| \leq 5\epsilon \cdot d \cdot n$;
- $|V| = k$;
- *The number of directed paths from* $I_G \setminus I$ *to* $O_G \setminus O$*, that do not pass through vertices in* V*, is at most* $\epsilon \cdot \frac{n^2}{k}$.

Corollary 6. *For fixed constants* $0 < \epsilon', \delta \leq 1$*, if the directed acyclic graph* G *with* n *inputs,* n *outputs and* $O(n)$ *edges is* $\left(\epsilon', \delta, \frac{1}{n}\right)$*-densely regular, then depth*

$$d \geq \alpha(n) - 2 ,$$

for n large enough, say, $n \geq N$, where N only depends on ϵ', δ and the hidden constant in $O(n)$.

Proof. Let S denote the number of edges, where $S \leq tn$, and $t > 0$ is a constant. It is clear that we can convert G into a layered graph G' such that G' is still $\left(\epsilon, \delta, \frac{1}{n}\right)$-densely regular and $|E(G')| \leq dS$.

Let

$$\epsilon := \frac{\delta^2 \epsilon'}{500d} \ .$$

We claim $|E(G')| \geq \epsilon \cdot n \cdot \lambda_d(n)$, which will imply $S \geq \frac{\delta^2 \epsilon'}{500d^2} \cdot \lambda_d(n) \cdot n$. Combining it with the condition $S \leq tn$, we have

$$\lambda_d(n) \leq \frac{500t}{\delta^2 \epsilon'} \cdot d^2 \ .$$

Thus,

$$\begin{aligned}
\lambda_{d+2}(n) &= \min\left\{i : \lambda_d^{(i)}(n) \leq 1\right\} \\
&\leq \min\left\{i : \lambda_d^{(i-1)}\left(\frac{500d^2t}{\delta^2\epsilon'}\right) \leq 1\right\} \\
&\leq \log\left(\frac{500t}{\delta^2\epsilon'} \cdot d^2\right) - 1 \\
&\leq d + 2 \ ,
\end{aligned}$$

assuming d is large enough, only depending on δ, ϵ, t. Thus, $d \geq \alpha(n) - 2$ for sufficiently large n.

Assume for contradiction that $|E(G')| < \epsilon n \cdot \lambda_d(n)$. By Lemma 5, there exists integer k satisfying $\sqrt{n} \leq k = o(n)$, and there exist $I \subseteq I_G$, $O \subseteq O_G$, and $V \subseteq V_G$ such that the conditions in Lemma 5 are satisfied.

Consider $\frac{10k}{\epsilon'}$-element inputs $\mathbf{X}$ and outputs $\mathbf{Y}$ drawn from the distribution $\mathcal{X}, \mathcal{Y}$ in

Definition 3. By definition, we know

$$\mathbb{E}_{\mathbf{X},\mathbf{Y}}\left[p(\mathbf{X},\mathbf{Y})\right] \geq \epsilon' \cdot \frac{10k}{\epsilon'} = 10k \ ,$$

where $p(\mathbf{X},\mathbf{Y})$ denotes the maximum number of vertex-disjoint paths connecting $\mathbf{X}$ and $\mathbf{Y}$ in graph G'.

On the other hand,

$$p(\mathbf{X},\mathbf{Y}) \leq |\mathbf{X} \cap I| + |\mathbf{Y} \cap O| + |V| + p(\mathbf{X} \setminus I, \mathbf{Y} \setminus O, V) \ , \tag{2.1}$$

where $p(\mathbf{X} \setminus I, \mathbf{Y} \setminus O, V)$ denotes the maximum number of vertex-disjoint paths connecting $\mathbf{X} \setminus I$ and $\mathbf{Y} \setminus O$, avoiding vertices V. Let us estimate the right hand side of (2.1).

$$\mathbb{E}_{\mathbf{X}}\left[\mathbf{X} \cap I\right] = \sum_{u \in I} \Pr_{\mathbf{X}}[u \in \mathbf{X}] \leq 5\epsilon dn \cdot \frac{10k}{\epsilon' \delta n} \leq \frac{k}{10} \ . \tag{2.2}$$

Similarly,

$$\mathbb{E}_{\mathbf{Y}}\left[\mathbf{Y} \cap O\right] \leq \frac{k}{10} \ . \tag{2.3}$$

By Lemma 5, we know that

$$\sum_{\substack{x \in I_{G'} \setminus I \\ y \in O_{G'} \setminus O}} p(x,y,V) \leq \epsilon \cdot \frac{n^2}{k},$$

therefore,

$$\begin{aligned} \mathbb{E}_{\mathbf{X},\mathbf{Y}}\left[p(\mathbf{X} \setminus I, \mathbf{Y} \setminus O, V)\right] &\leq \sum_{\substack{x \in I_{G'} \setminus I \\ y \in O_{G'} \setminus O}} \Pr[x \in \mathbf{X}, y \in \mathbf{Y}] \cdot p(x,y,V) \\ &\leq \left(\frac{k}{\delta n}\right)^2 \cdot \epsilon \cdot \frac{n^2}{k} \\ &\leq \frac{\epsilon' k}{500d} \ . \end{aligned} \tag{2.4}$$

Thus, combining (2.1), (2.2), (2.3) and (2.4), we have

$$\mathbb{E}_{\mathbf{X},\mathbf{Y}}[p(\mathbf{X},\mathbf{Y})] \leq \frac{k}{10} + \frac{k}{10} + k + \frac{\epsilon' k}{500d} < 2k,$$

contradicting with $\mathbb{E}[p(\mathbf{X},\mathbf{Y})] \geq 10k$. □

The following depth lower bound is immediate from Corollary 4 and 6. Since Corollary 4 holds for the unrestricted circuit model, i.e., arbitrary gates of unbounded fan-in over any finite field, [1] the following corollary also holds for the unrestricted circuit model.

Corollary 7. *Let $\rho, \delta > 0$ be constants. Any circuit computing (ρ, δ)-good codes with n inputs using cn edges, $c > 0$ constant, requires depth at least $\alpha(n) - 2$ for sufficiently large n, where this sufficiently large number only depends on ρ, δ, c.*

2.3 Recursive Construction

We will construct an $O(n)$-size $O(\alpha(n))$-depth linear arithmetic circuit over $\mathbb{F}_2$ computing good code, where $\mathbb{F}_2$ can be replaced by any finite field. The construction is non-explicit, and is inspired by the construction of superconcentrators [21].

Let $S_d(n)$ denote the minimal size of a linear (XOR) circuit $\mathsf{C} : \mathbb{F}_2^n \to \mathbb{F}_2^{20n}$ with minimum distance $\geq n$, that is, for any nonzero $x \in \mathbb{F}_2^n$, $\mathrm{wt}(C(x)) \geq n$. We will prove the following two recursions in the two subsections:

- $S_d(n) \leq S_{d-2}\left(\frac{n}{20}\right) + O(n)$;
- $S_d(n) \leq 2^{O(d)} \cdot \lambda_d(n) \cdot n$.

Together they will imply the desired upper bound. Gál *et al.* [34] show that good codes can be computed by depth-d, constant $d \geq 3$, circuit using $O_d(\lambda_d(n) \cdot n)$ wires. However, the

1. Corollary 4 in [34] is stated in the boolean case. Their proof works step by step for any finite field $\mathbb{F}$, that is, each gate can compute any function from $\mathbb{F}^s$ to $\mathbb{F}$, where s is the fan-in of the gate, and s is unbounded.

hidden constant in $O_d\left(\lambda_d(n)\cdot n\right)$ is very large, which is not good enough for our purpose. (In fact, it grows faster than any primitive function, which is implicit in the proof of Lemma 26 in [34].)

2.3.1 First Recursion

The goal of this subsection is to prove $S_d(n) \le S_{d-2}\left(\frac{n}{20}\right)+O(n)$. We propose the definitions of *linear condenser* and *linear amplifier*, which are special cases of *range detectors* [34]. However, we find it more clear to have separate definitions.

Definition 8 (Linear Condenser)**.** *Linear mapping $h:\mathbb{F}_2^n\to\mathbb{F}_2^m$ is called* (n,m,α)-linear condenser *if any nonzero vector $x\in\mathbb{F}_2^n$ of weight $\le\alpha$ maps to a nonzero $h(x)$.*

Lemma 9. *For every $n\ge 1$, there exists $\left(n,\frac{n}{10},\frac{n}{1000}\right)$-linear condenser computable by depth-1 circuit using $O(n)$ wires.*

Proof.

Claim 10. *Let $D=8$. There exists bipartite graph $G(U=[n],V=\left[\frac{n}{10}\right],E)$ such that*

- $\deg(u)=D$ *for each $u\in U$;*
- *for any $X\subseteq U$ of size at most $\frac{n}{1000}$, the size of $\Gamma(X)$ is greater than $\frac{D|X|}{2}$.*

Proof. Connect $u\in U$ to D uniformly random vertices in V. For convenience of analysis, we allow repetition. Let $X\subseteq U$ of be any subset of size $k\le\frac{n}{1000}$.

$$\begin{aligned}&\Pr\left[|\Gamma(X)|\le\frac{Dk}{2}\right]\le\binom{Dk}{\frac{1}{2}\cdot Dk}\left(\frac{\frac{1}{2}\cdot Dk}{\frac{1}{10}\cdot n}\right)^{\frac{1}{2}\cdot Dk}\\ \le\ & 2^{Dk-\frac{1}{2}Dk\log\left(\frac{n}{5Dk}\right)}\le 2^{-\frac{1}{4}Dk\log\left(\frac{n}{5Dk}\right)},\end{aligned}$$

since $D=8$. Taking a union bound, the probability that there exists such X of size k is at most

$$\binom{n}{k}2^{-\frac{1}{4}Dk\log\left(\frac{n}{5Dk}\right)}\le 2^{k\log\left(\frac{en}{k}\right)-\frac{1}{4}Dk\log\left(\frac{n}{5Dk}\right)}\le 2^{-k}.$$

Summing over all $2 \leq k \leq \frac{n}{1000}$, we conclude such graph exists. $\square$

Fix such a graph $G(U = [n], V = \left[\frac{n}{10}\right], E)$ with inputs U, and for each $v \in V$, replace v by an XOR gate, i.e., an addition gate over $\mathbb{F}_2$. It suffices to prove that, for every $X \subseteq U = [n]$ of size $\leq \frac{1}{1000} \cdot n$, there exists $v \in V$ such that $|\Gamma(v) \cap X| = 1$. Assume for contradiction that such v does not exist, which implies that for each $v \in \Gamma(X)$, v has at least two neighbors in X. Thus, the total number of edges leaving X is $> \frac{D|X|}{2} \cdot 2 = D|X|$. Contradiction. $\square$

Definition 11 (Linear Amplifier)**.** *A linear mapping $h : \mathbb{F}_2^n \to \mathbb{F}_2^m$ is an (n, m, α, β)-*linear amplifier *if it sends any $x \in \mathbb{F}_2^n$ of Hamming weight at least α to an output $y \in \mathbb{F}_2^m$ of Hamming weight at least β.*

Bipartite graph $G = (V_1 = [n], V_2 = [m], E)$ is a (k, ϵ)-*disperser graph*, if for every $X \subseteq V_1$ of cardinality k, $|\Gamma(X)| > (1 - \epsilon)m$.

Theorem 12 (Theorem 1.10 in [78])**.** *For every $1 < k \leq n, m \geq 0$ and $\epsilon > 0$ there exists a (k, ϵ)-disperser graph $G = (V_1 = [n], V_2 = [m], E)$ with left-degree*

$$D \;=\; \left\lceil \frac{1}{\epsilon} \ln\left(\left(\frac{n}{k}\right) + 1\right) + \frac{m}{k}\left(\ln\left(\frac{1}{\epsilon}\right) + 1\right)\right\rceil \;.$$

Lemma 13. *For any $n, m \geq 3n$, there exists an $\left(n, m, \frac{n}{1000}, \frac{m}{10}\right)$-linear amplifier computable by depth-1 circuit with $O(m)$ edges.*

Proof. By Theorem 12, there exists a bipartite graph $G = (V_1 = [n], V_2 = [m], E)$ with $O(m)$ edges such that for every $X \subseteq V_1$ of cardinality at least $\frac{1}{1000} \cdot n$, $|\Gamma(X)| \geq \frac{9}{10} \cdot m$. Fix such a graph G, and convert it into a random circuit $\mathbf{C}$ with inputs V_1 and outputs V_2 such that the coefficient on each edge is chosen uniformly at random, and all gates in V_2 are XOR gates.

Fix any $x \in \mathbb{F}^n$ of weight at least $\frac{1}{1000} \cdot n$. Let $X = \mathrm{supp}(x) = \{i \in [n] : x_i = 1\}$. By definition, $|\Gamma(X)| \geq \frac{9}{10} \cdot m$, that is, there exist at least $\frac{9}{10} \cdot m$ vertices, which are incident to at least one nonzero input. We claim the outputs V_2 *restricted to* $\Gamma(X)$ are uniformly

distributed. To prove this claim, for each $v \in \Gamma(X)$, we can fix some incident edge $e_v \in v \times X$. Since the order of choosing random coefficients does not matter, when e_v is fixed at last, it is clear that the output of v is uniformly distributed. Therefore

$$\Pr\left[\mathrm{wt}\left(\mathbf{C}(x)\right) < \frac{m}{10}\right] \;\leq\; \frac{\binom{|\Gamma(X)|}{\frac{m}{10}}}{2^{|\Gamma(X)|}} \;\leq\; 2^{-\frac{m}{3}} \;.$$

Finally, taking a union bound over all x, we conclude there exists such a *deterministic* circuit. □

Recall that $S_d(n)$ denotes the minimal size of a linear arithmetic circuit $\mathsf{C} : \mathbb{F}_2^n \to \mathbb{F}_2^{20n}$ with minimum distance $\geq n$, that is, for any nonzero $x \in \mathbb{F}_2^n$, $\mathrm{wt}(\mathsf{C}(x)) \geq n$.

Lemma 14. *For any $d \geq 3$, and for all n,*

$$S_d(n) \;\leq\; S_{d-2}\left(\frac{n}{20}\right) + O(n) \;,$$

where the constant in $O(n)$ is an absolute constant.

Proof. The circuit consists of two parts on the same inputs and *disjoint* outputs. The left part is an $\left(n, 10n, \frac{n}{1000}, n\right)$-amplifier of size $O(n)$, where the existence of such amplifier is proved in Lemma 13. The right part is the concatenation of an $\left(n, \frac{n}{20}, \frac{n}{1000}\right)$-condenser, a circuit of depth $d-2$ computing good code with $\frac{n}{20}$ inputs and n outputs, and an $\left(n, 10n, \frac{n}{20}, n\right)$-amplifier. By Lemma 9, there exists $\left(n, \frac{n}{20}, \frac{n}{1000}\right)$-linear condenser of size $O(n)$, and by Lemma 13, there exists $\left(n, 10n, \frac{n}{20}, n\right)$-amplifier of size $O(n)$. Therefore, the total size is of the circuit is

$$S_{d-2}\left(\frac{n}{20}\right) + O(n) \;.$$

Let us verify that it computes a good code with minimum distance $\geq n$. For any nonzero $x \in \mathbb{F}_2^n$, if $\mathrm{wt}(x) \geq \frac{n}{1000}$, then the left part will amplifier the weight to $\geq n$. If $\mathrm{wt}(x) < \frac{n}{1000}$, then the output of the condenser will be nonzero, i.e., the input to the good code in the

middle is nonzero. By definition, the output of the good code in the middle is of weight at least $\frac{n}{20}$. Through the amplifier, the weight of the output will be at least n, as desired. $\square$

The recursion $S_d(n) \leq S_{d-2}\left(\frac{n}{20}\right) + O(n)$ already implies that there exist linear-size logarithmic-depth circuits computing good codes. This recursion will be combined with the second recursion (proved in the next subsection) to achieve depth $O(\alpha(n))$.

2.3.2 Second Recursion

The goal of this subsection is to prove Lemma 24, that is, $S_d(n) \leq 2^{O(d)} \cdot \lambda_d(n) \cdot n$, which qualitatively improves the hidden constant in [34]. Compared with [34], our presentation is somewhat different, which follows the framework of superconcentrators in [21].

Definition 15. *Linear function $C : \mathbb{F}_2^n \to \mathbb{F}_2^{20n}$ is called* (a, b)-partial good code *if for all $x \in \mathbb{F}_2^n$ with $\mathrm{wt}(x) \in [a, b]$, the weight of $C(x)$ is at least n.*

By definition, a $(1, 20n)$-partial good code is a good code. Let $S_d(n, a, b)$ denote the minimal size of a linear (arithmetic) circuit computing some (a, b)-partial good code. Over the field $\mathbb{F}_2$, linear circuits consist of XOR gates only.

We will prove $S_d(n, a, b)$ satisfies a recursion similar to that obeyed by superconcentrators, and the proof follows the framework in [21]. Here are some intuitive explanations why such a recursion holds for $S_d(n, a, b)$.

1. Circuits computing partial good codes with different range of parameters can be combined, i.e., if circuit C_1 computes some (a, b)-partial good code, and circuit C_2 computes some (b, c)-partial good code, then there exists circuit C_3 computing (a, c)-partial codes such that $\mathrm{size}(\mathsf{C}_3) \leq \mathrm{size}(\mathsf{C}_1) + \mathrm{size}(\mathsf{C}_2) + O(n)$.

2. Large weights are easy. For linear codes C, in order to satisfy minimum distance condition, it is enough to make sure $\mathrm{wt}(C(x))$ is large for every nonzero x. Intuitively, if x has larger weight, it is easier to "amplify".

3. A composition lemma, that is Lemma 21, holds, which says that, if the weight of inputs are small, we can reduce the number of inputs n by putting a "condenser" at the top, and an "amplifier" at the bottom. A similar composition lemma holds for partial superconcentrators, where both the top part and the bottom part are dispersers.

Lemma 16. *For any $t \geq 2$ and any $1 \leq w_1 \leq w_2 \leq \cdots \leq w_t \leq n$,*

$$S_d\left(n, w_1, w_t\right) \;\leq\; \sum_{i=1}^{t-1} S_{d-1}\left(n, w_i, w_{i+1}\right) + O\left((t-1)n\right) \;, \tag{2.5}$$

and

$$S_d(n, w_1, w_t) \;\leq\; O\left(\sum_{i=1}^{t-1} S_d\left(n, w_i, w_{i+1}\right)\right) \;, \tag{2.6}$$

where the constants in both big-O notations are absolute constants.

Proof.

Claim 17. *For any (a, b)-partial good code $C : \mathbb{F}_2^n \to \mathbb{F}_2^{20n}$, there exists a "linear amplifier" $H : \mathbb{F}^{20n} \to \mathbb{F}_2^{20n}$ computable by some depth-1 size-$O(n)$ circuit such that the weight of $H(C(x))$ is at least $5n$ for all nonzero $x \in \mathbb{F}_2^n$.*

Proof. (of the Claim) The proof is similar to Lemma 13, so we sketch the proof. Let $G(U = [20n], V = [20n], E)$ be a size-$O(n)$ disperser such that for all $X \subseteq U$ of size n, $|\Gamma(X)| \geq 19n$. The existence of such disperser is guaranteed by Theorem 12. We convert the graph G into a circuit $\mathbf{C}$ using randomness as follows: let U be the inputs, and let V be the set of addition gates where the coefficient of each edge is uniformly at random in $\mathbb{F}_2$. We claim $\mathbf{C}$ satisfies the desired property with nonzero probability. Fix a nonzero $y \in \mathbb{F}_2^n$ of weight at least n, we know that outputs restricted to $\Gamma(X)$ is uniformly distributed int $\mathbb{F}_2^{|\Gamma(X)|}$, and thus

$$\Pr\left[\mathrm{wt}(\mathbf{C}(y)) < 5n\right] \;\leq\; \frac{\binom{19n}{\leq 5n}}{2^{19n}} \;\leq\; 2^{-3n} \;.$$

Note that there are at most 2^n such y's in total, and thus the proof is complete by a union

bound. □

Let $\mathsf{C}_i : \mathbb{F}_2^n \to \mathbb{F}_2^{20n}$ be a linear circuit computing a (w_i, w_{i+1})-partial good code of depth $d-1$ and size $S_{d-1}(n, w_i, w_{i+1})$, where $i = 1, 2, \ldots, t-1$. Let $\mathsf{H}_i : \mathbb{F}_2^{20n} \to \mathbb{F}_2^{20n}$ be a "linear amplifier" of depth 1 and size $O(n)$ such that the weight of $(\mathsf{C}_i \circ \mathsf{H}_i)(x)$ is at least $5n$ for all nonzero $x \in \mathbb{F}_2^n$. The existence of such H_i is proved in the above claim. The overall circuit $\mathbf{C}$ is constructed by identifying the outputs of H_i, and for each output u of H_i, with probability $\frac{1}{2}$, remove *all* the edges incident to u in H_i.

This random process will generate a random circuit $\mathbf{C}$. In order to prove there exists a deterministic circuit computing a (w_1, w_t)-partial good code, it suffices to prove for each $x \in \mathbb{F}_2^n$ of weight $\mathrm{wt}(x) \in [w_1, w_t]$,

$$\Pr[\mathbf{C}(x) < n] \;<\; 2^{-n} \;.$$

Since $\mathrm{wt}(x) \in [w_1, w_t]$, there exists i such that $\mathrm{wt}(x) \in [w_i, w_{i+1}]$. By definition, vector $y := (\mathsf{C}_i \circ \mathsf{H}_i)(x)$ is of weight $\geq 5n$, which implies that the outputs of $\mathbf{C}$, restricted to $\mathrm{supp}(y)$, is uniformly distributed in $\mathbb{F}_2^{|\mathrm{supp}(y)|}$. Thus,

$$\Pr[\mathbf{C}(x) < n] \;\leq\; \frac{\binom{5n}{\leq n}}{2^{5n}} \;<\; 2^{-n} \;.$$

The total size of $\mathbf{C}$ is bounded by $\sum_i S_{d-1}(n, w_i, w_{i+1}) + O((t-1)n)$.

It is clear that (2.5) implies (2.6), because $S_d(n, w_i, w_{i+1}) \geq n$ for all i, and the disperser graph in Claim 17 has constant left-degree. □

An (m, n, ℓ, k, r, s)-*range detector* is a linear circuit that has m inputs, n outputs, and on any input of weight between ℓ and k, it outputs a string with weight between r and s [34]. Note that a range detector with certain parameters is by definition also a condenser, amplifier, good code, and partial good code. However, it will be clearer to have these separate definitions.

Lemma 18 (Lemma 23 in [34]). *There exists $c \geq 6$ such that for all $c \leq r \leq n$ and $1 \leq a \leq \frac{n}{r^{1.5}}$, there exists an*

$$\left(n, \frac{n}{r}, a, \frac{n}{r^{1.5}}, a, \frac{n}{r}\right) \text{-range detector}$$

of depth 1 and size $6n$.

From the work Gál *et al.* [34], we have:

Lemma 19 (Lemma 27 in [34]). *For any $r \leq n$, $S_2\left(n, \frac{n}{r}, n\right) \leq O(n \log^2 r)$.*

Lemma 20 (Corollary 29 in [34]). *For any $d \geq 2$, and for any $r \leq n$,*

$$S_d\left(n, \frac{n}{r}, n\right) \;\leq\; O_d\left(\lambda_d(r) \cdot n\right),$$

where the constant in $O_d(\lambda_d(r) \cdot n)$ only depends on d.

However, the hidden constant in $O_d\left(\lambda_d(r) \cdot n\right)$ grows very fast in d, like Ackermann function, which is implicit in the proof of Lemma 26 in [34]. For the purpose of constructing linear-size $O(\alpha(n))$-depth circuits computing good codes, we cannot apply their result directly; more careful analysis is required.

Lemma 21. *For all $a \leq \frac{n}{r^2}$,*

$$S_d\left(n, a, \frac{n}{2r}\right) \;\leq\; S_{d-2}\left(\frac{n}{2r}, a, \frac{n}{2r}\right) + O(n)\ ,$$

where the constant in O is an absolute constant.

Proof. Let us construct a depth-d circuit computing some (a, b)-partial good code as the concatenation of three circuits as follows. The top part is an $\left(n, \frac{n}{2r}, a, \frac{n}{r^2}, a, \frac{n}{2r}\right)$-range detector of size $3n$. The existence of such range detector is proved in Lemma 18. The middle part is an $\left(a, \frac{n}{2r}\right)$-partial good code with $\frac{n}{2r}$ inputs of size $S_{d-2}\left(\frac{n}{2r}, a, b\right)$ and depth $d-2$. The

bottom part is a depth-1 $O(n)$-size $\left(\frac{10n}{r}, 20n, \frac{n}{100r}, 2n\right)$-amplifier, whose existence is proved in Lemma 13.

Let $x \in \mathbb{F}_2^n$ have weight $\mathrm{wt}(x) \in \left[a, \frac{n}{2r}\right]$. By the definition of range detector, the output of the top part has weight $\geq a$; through the $\left(a, \frac{n}{2r}\right)$-partial good code in the middle, the output therefore has weight $\geq \frac{n}{2r}$; so after the bottom amplifier, the output has weight $\geq n$, as desired. □

Lemma 22. *For any $r \leq n$,*

$$S_4\left(n, \frac{n}{r^2}, \frac{n}{r}\right) \;\leq\; O(n)\ ,$$

where the constant in $O(n)$ is an absolute constant.

Proof. Applying Lemma 21 with $r' = \frac{\sqrt{r}}{2}$, we have

$$\begin{aligned} S_4\left(n, \frac{n}{r^2}, \frac{n}{r}\right) &\leq S_2\left(\frac{n}{\sqrt{r}}, \frac{n}{r^2}, \frac{n}{\sqrt{r}}\right) + O(n) \\ &\leq \frac{n}{\sqrt{r}} \log r^2 + O(n) \qquad \text{By Lemma 19} \\ &= O(n)\ . \end{aligned}$$

□

Let $A_d(i) = \min\{n : \lambda_d(n) \geq i\}$, where $\lambda_d(n)$ is defined in Definition 1. Function $A_d(i)$ grows like Ackermann function. Note that there are various definitions with slight differences in the literature.

Proposition 23. *1. For all $n, d \geq 1$,*

$$\lambda_d\left(A_d(n)\right) \;=\; n\ . \tag{2.7}$$

2. For all $n, d \geq 1$,

$$A_d\left(\lambda_d(n)\right) \;\leq\; n\ . \tag{2.8}$$

3. For all $i \geq 0$ and $d \geq 3$,

$$A_d(i+1) \;\leq\; A_{d-2}^{(i)}(2) \;. \tag{2.9}$$

Proof. 1. By definition,

$$\begin{aligned} \lambda_d(A_d(n)) &= \lambda_d\left(\min\{m : \lambda_d(m) \geq n\}\right) \\ &= \lambda_d\left(\min\{m : \lambda_d(m) = n\}\right) \\ &= n \;, \end{aligned}$$

where the second last step works because the image of λ_d can take any positive integer value.

2. By definition,

$$A_d\left(\lambda_d(n)\right) \;=\; \min\{m : \lambda_d(m) \geq \lambda_d(n)\} \;\leq\; n \;.$$

3. We prove by induction on i. The base case $i = 0$ is obvious, since $A_d^{(0)}(2) = 2$ and $A_d(1) \leq 2$. Assuming the case $i - 1$ is true, we have

$$A_d(i) \;\leq\; A_{d-2}^{(i-1)}(2) \;. \tag{2.10}$$

Since $A_{d-2}(i)$ is a non-decreasing function in i, applying A_{d-2} to both sides of (2.10), the right hand side becomes $A_{d-2}^{(i)}(2)$, while the left hand side becomes

$$\begin{aligned} A_{d-2}\left(A_d(i)\right) &= A_{d-2}\left(\min\{n : \lambda_d(n) \geq i\}\right) \\ &= \min\{A_{d-2}(n) : \lambda_d(n) \geq i\} \\ &= \min\{A_{d-2}(n) : \lambda_{d-2}^*(n) \geq i\} \;. \end{aligned}$$

Let $m := A_{d-2}(n)$, and by (2.7), $n = \lambda_{d-2}(m)$, thus we have

$$\begin{aligned} A_{d-2}\left(A_d(i-1)\right) &\leq \min\{m : \lambda^*_{d-2}(\lambda_{d-2}(m)) \geq i\} \\ &= \min\{m : \lambda^*_{d-2}(m) \geq i+1\} \\ &= A_d(i+1)\ . \end{aligned}$$

Therefore, $A_d(i+1) \leq A^{(i)}_{d-2}(2)$, which completes the induction step. □

Lemma 24. *There exists an absolute constant $c > 1$ such that, for all $k \geq 2$, and for any $r \leq n$,*

$$S_{2k}\left(n, \frac{n}{A_{2(k-1)}(r)}, \frac{n}{r}\right) \ \leq\ c^{2k-1} \cdot n\ , \tag{2.11}$$

and

$$S_{2k}\left(n, \frac{n}{r}, n\right) \ \leq\ c^{2k} \cdot \lambda_{2k}(r) \cdot n\ . \tag{2.12}$$

Proof. First, let us show that (2.11) implies (2.12). By Lemma 16,

$$\begin{aligned} & S_{2k}\left(n, \frac{n}{r}, n\right) \\ \leq\ & c_1 \left(S_{2k}\left(n, \frac{n}{2}, n\right) + \sum_{i=1}^{h} S_{2k}\left(n, \frac{n}{A^{(i)}_{2(k-1)}(2)}, \frac{n}{A^{(i-1)}_{2(k-1)}(2)}\right)\right)\ , \end{aligned}$$

where h is the *minimum* integer such that $A^{(h)}_{2(k-1)}(2) \geq r \Rightarrow A^{(h-1)}_{2(k-1)}(2) < r$. By Proposition 23 (3), $A_{2k}(h) < r$, which implies $h < \lambda_{2k}(r)$ by Proposition 23 (1), i.e., $h \leq \lambda_{2k}(r) - 1$. Thus

$$S_{2k}\left(n, \frac{n}{r}, n\right) \ \leq\ c_1\left(2n + c^{2k-1} \cdot (\lambda_{2k}(r) - 1) \cdot n\right) \ \leq\ c^{2k} \cdot \lambda_{2k}(r) \cdot n\ ,$$

if c is large enough, say, $c \geq \max(c_1, 2)$.

Now, let us prove (2.11) by induction on k. The base case $k = 2$ is true by Lemma 20 and 21. For the induction step, let us assume both (2.11) and (2.12) are true for $k-1$, we

will prove (2.11) for k. First, observe that

$$\begin{aligned} S_{2k}\left(n, \frac{n}{A_{2(k-1)}(r)}, \frac{n}{r}\right) &\leq c_1\left(S_{2k}\left(n, \frac{n}{A_{2(k-1)}(r)}, \frac{n}{4r^2}\right) + S_{2k}\left(n, \frac{n}{4r^2}, \frac{n}{r}\right)\right) \\ &\leq c_1\left(S_{2k}\left(n, \frac{n}{A_{2(k-1)}(r)}, \frac{n}{4r^2}\right) + c_2 \cdot n\right) \ , \end{aligned}$$

where the last step is by Lemma 22. Applying Lemma 21,

$$\begin{aligned} S_{2k}\left(n, \frac{n}{A_{2(k-1)}(r)}, \frac{n}{4r^2}\right) &\leq S_{2(k-1)}\left(\frac{n}{2r}, \frac{n}{A_{2(k-1)}(r)}, \frac{n}{4r^2}\right) + c_3 \cdot n \\ &\leq c^{2k-2} \cdot \frac{n}{2r} \cdot \lambda_{2(k-1)}\left(\frac{A_{2(k-1)}(r)}{2r}\right) + c_3 \cdot n \\ &\leq c^{2k-2} \cdot \frac{n}{2} + c_3 \cdot n \ , \end{aligned}$$

where c_3 is the absolute constant in Lemma 21. Therefore,

$$\begin{aligned} S_{2k}\left(n, \frac{n}{A_{2(k-1)}(r)}, \frac{n}{r}\right) &\leq c_1\left(c^{2k-2} \cdot \frac{n}{2} + c_3 \cdot n + c_2 \cdot n\right) \\ &\leq c^{2k-1} \cdot n \ , \end{aligned}$$

if c is large enough, say, $c \geq \max\left(c_1, 2(c_2 + c_3)\right)$. □

Putting two recursions together (Lemma 14 and 24), we have the following result.

Theorem 25. *There exists an absolute constant $c > 1$ such that, for all $d \geq c \cdot \alpha(n)$, we have $S_d(n) = O(n)$.*

Proof. Let $d = \alpha(n) + i$, where $i \geq \alpha(n) \log_{20} c$. By applying Lemma 14 for i times, and

then applying Lemma 24, we have

$$\begin{aligned} S_d(n) &\leq S_{\alpha(n)}\left(\frac{n}{20^i}\right) + O(n) \\ &\leq \frac{n}{20^i} \cdot c^{\alpha(n)} + O(n) \\ &\leq O(n) \ . \end{aligned}$$

□

2.4 Superconcentrator Codes

Spielman observed that any circuit computing good codes must satisfy some superconcentrator-like properties [82], namely, for some constant $\delta > 0$, there exist vertex-disjoint paths connecting any chosen δn inputs to some subset of $(1-\delta)m$ outputs, where an arbitrary size-δm subset of outputs are removed; Gál *et al.* [34] also proved a similar result that any circuit computing good codes should be *densely regular*, which is a graph property generalizing that of superconcentrators.

In this section, we will prove some results in the reverse direction. For example, by replacing each vertex with an addition gate, and assigning the coefficient of each edge uniformly at random, *any* superconcentrator computes a good code, assuming the size of the field is large enough. The results suggest that there is a close connection between the *graph-theoretic property* of the circuits and the *algebraic property* of the functions computed by the circuits.

Let us propose the definition of *superconcentrator codes*.

Definition 26 (Superconcentrator Codes)**.** *Linear code $C : \mathbb{F}^n \to \mathbb{F}^m$ is called a* superconcentrator code *if all minors of its generator matrix is nonzero. (Recall that a minor of a matrix is the determinant of some smaller square matrix.)*

Recall that the generator matrix of a linear code $C : \mathbb{F}^n \to \mathbb{F}^m$ is the $n \times m$ matrix M such that $C(x) = xM$ for all $x \in \mathbb{F}^n$. By definition, linear code C is a superconcentrator

code if $\det M_{X,Y} \neq 0$ for all equal-size subsets $X \subseteq [n]$ and $Y \subseteq [m]$, where $M_{X,Y}$ denotes the submatrix of M with rows indexed by X, and columns indexed by Y.

Superconcentrator codes are *Maximum Distance Separable* (MDS) codes, and thus they meet the Singleton bound. It is well known that, code $C : \mathbb{F}^n \to \mathbb{F}^m$ is a MDS code if and only if every set of n columns of its generator matrix are linearly independent [59]; for superconcentrators, we require that every square submatrix has full rank, which is a more strict requirement. Following lemma is a characterization of such codes.

Lemma 27. *Let $C : \mathbb{F}^n \to \mathbb{F}^m$ be a linear code, where $m \geq n$. Code C is a superconcentrator code if and only if, for any nonzero $x \in \mathbb{F}^n$,*

$$\mathrm{wt}(C(x)) > m - \mathrm{wt}(x) \ ,$$

where $\mathrm{wt}(x)$ *is defined as the number of nonzero coordinates of x.*

Proof. For the "only if" direction, assume for contradiction that there exists some nonzero $x \in \mathbb{F}^n$ such that $\mathrm{wt}(C(x)) \leq m - \mathrm{wt}(x)$. Let $X := \mathrm{supp}(x) \subseteq [n]$, and let Y be any subset of $[m] \setminus \mathrm{supp}(C(x))$ of size $\mathrm{wt}(x)$ such that $|X| = |Y|$. Let $M \in \mathbb{F}^{n \times m}$ be the generator matrix of C. Thus, $C(x) = xM$, which implies that

$$C(x)|_Y \ = \ (xM)|_Y \ = \ xM_{[n],Y} \ ,$$

where $M_{[n],Y}$ denotes the $n \times |Y|$ submatrix indexed by columns Y. Since $x|_{[n]\setminus X}$ is the all-zero vector, we have

$$C(x)|_Y \ = \ x|_X M_{X,Y} \ .$$

Note that $C(x)|_Y = \overline{0}$ by the definition of Y. On the other hand, $(x|_X)\, M_{X,Y}$ is nonzero, because $x|_X$ is nonzero and $M_{X,Y}$ is of full rank. Contradiction.

For the "if" direction, we prove the contraposition, that is, if C is not a superconcentrator code, then there exists a nonzero x such that $\mathrm{wt}(C(x)) \leq m - \mathrm{wt}(x)$. Since C is not

a superconcentrator code, there exist equal-size subsets $X \subseteq [n]$ and $Y \subseteq [m]$ such that $\operatorname{rank} M_{X,Y} < |X|$, where M denotes the generator matrix of C. Since $\operatorname{rank} M_{X,Y} < |X|$, there exists a nonzero $x' \in \mathbb{F}^{|X|}$ such that $x'M_{X,Y} = \overline{0}$. Extend x' to vector $x \in \mathbb{F}^n$ by adding zeros on the coordinates outside of X, that is, $x' = x|_{\mathrm{supp}(x)}$. It is clear that $C(x) = xM = x'M_{X,[m]}$ and $C(x)|_Y = x'M_{X,Y} = 0$, which implies that $\mathrm{wt}(C(x)) \leq m - |X|$, as desired. □

By the above lemma, linear code $C : \mathbb{F}^n \to \mathbb{F}^m$ is a superconcentrator code if and only if

$$\mathrm{dist}\,(C(x), C(y)) > m - \mathrm{dist}(x, y)$$

for all distinct $x, y \in \mathbb{F}^n$, where $\mathrm{dist}(x, y)$ denotes the Hamming distance between x and y, i.e., $\mathrm{dist}(x, y) = \mathrm{wt}(x - y)$. This condition much stronger than that of good codes.

The following corollary is immediate, which implies that a superconcentrator code is a good code if $m - n = \Theta(n)$.

Corollary 28. *Let linear code $C : \mathbb{F}^n \to \mathbb{F}^m$ be a superconcentrator code. Then C has minimum distance at least $m - n + 1$.*

The following two results justify the definition of "supconcentrator code".

Proposition 29. *Any unrestricted* [2] *circuit computing a superconcentrator code must be a superconcentrator.*

Proof. We prove by contradiction. Suppose the underlying graph G of the circuit $\mathsf{C} : \mathbb{F}^n \to \mathbb{F}^m$ is not a superconcentrator. Thus, there exist a subset of inputs $X \subseteq [n]$, and a subset of outputs $Y \subseteq [m]$ such that

- $|X| = |Y|$;
- There do not exist $|X|$ vertex-disjoint paths connecting X and Y.

2. For unrestricted circuits, each gate can compute any function $\mathbb{F}^s \to \mathbb{F}$, where s is the fan-in of the gate, and s is unbounded.

By Menger's theorem, there exists a subset $Z \subseteq V(G)$ of size $|X| - 1$ whose removal will disconnect X and Y.

Let us fix the inputs to x_i, for all $i \in [n] \setminus X$, to 0. The key observation is that, any output in Y can be written as a function in the outputs of the gates in Z, which implies that the total number of different $(y_j)_{j \in Y}$ is at most $|\mathbb{F}|^{|Z|}$, when all inputs to x_i, $i \in [n] \setminus X$, are fixed.

On the other hand, for linear code C with the generator matrix M, it is not difficult to prove

$$\left|\{C(x)|_Y : x \in \mathbb{F}^k \text{ s.t. } X|_{[k]\setminus X} = 0\}\right| \;=\; |\mathbb{F}|^{\operatorname{rank} M_{X,Y}} \;.$$

Since circuit C computes a superconcentrator code, we have $\operatorname{rank} M_{X,Y} = |X| > |Z|$. Contradiction. □

The other direction is also true, that is, any superconcentrator can compute a superconcentrator code, when each vertex is replaced by an addition gate over some large enough field, and the coefficients are chosen uniformly at random.

Theorem 30. *Let G be any superconcentrator with n inputs and m outputs. Let*

$$\mathbf{C}_G : \mathbb{F}^n \to \mathbb{F}^m$$

be a linear arithmetic circuit by replacing each vertex of G with an addition gate, and choosing the coefficient of each edge uniformly at random in field $\mathbb{F}$. With probability $1 - o_{|\mathbb{F}|;G}(1)$, $\mathbf{C}_G$ computes a superconcentrator code.

Proof. Let us denote the inputs by $x_1, x_2, \ldots, x_n$, and outputs by $y_1, y_2, \ldots, y_m$. For each edge $e \in E(G)$, there is a distinct coefficient $\mathbf{r}_e$ associated with it. Let $\mathbf{M}$ be the $n \times m$ generator matrix of the linear code $\mathbf{C}_G$, where each entry is a polynomial in $\mathbf{r}_e$'s. By definition,

$$\mathbf{M}_{i,j} = \sum_p \prod_{e \in p} \mathbf{r}_e,$$

where p enumerates all paths from x_i to y_j. It is clear that $\deg(\mathbf{M}_{i,j})$ at most the length of the longest path from x_i to y_j, where $\mathbf{M}_{i,j}$ is viewed as a polynomial in $\mathbb{F}[\{\mathbf{r}_e\}_{e\in E(G)}]$.

Claim 31. *For any equal-size subsets $X \subseteq [n]$ and $Y \subseteq [m]$,*

$$\Pr\left[\det(\mathbf{M}_{X,Y}) = 0\right] \;\leq\; \frac{d|X|}{|\mathbb{F}|}\ ,$$

where d is the depth of the superconcentrator.

Proof. First note that $\det(\mathbf{M}_{X,Y})$ is a polynomial of degree at most $d|X|$, because each entry is of degree at most d. Since G is a superconcentrator, there exist $|X|$ vertex-disjoint paths connecting X and Y. If we set all coefficients $\mathbf{r}_e$ on those $|X|$ paths to be 1, otherwise 0, then the polynomial $\det(\mathbf{M}_{X,Y})$ evaluates to ± 1, which implies that $\det(\mathbf{M}_{X,Y})$ is a nonzero polynomial. The claim follows from Schwartz-Zippel Lemma. □

Taking a union bound over all equal-size subsets $X \subseteq [n]$ and $Y \subseteq [m]$, we have

$$\begin{aligned}\Pr[\mathbf{C}_G \text{ computes a superconcentrator code}] \;&\geq\; 1 - \sum_i \binom{n}{i}\binom{m}{i}\frac{di}{|\mathbb{F}|}\\ &\geq\; 1 - 2^{m+n}\frac{dn}{|\mathbb{F}|}\ ,\end{aligned}$$

which completes the proof. □

The following corollary is immediate from Theorem 30 and Corollary 28.

Corollary 32. *Let G be any superconcentrator with n inputs and cn outputs, where constant $c > 1$. Let*

$$\mathbf{C}_G : \mathbb{F}^n \to \mathbb{F}^m$$

be a linear arithmetic circuit by replacing each vertex of G with an addition gate, and choosing the coefficient of each edge uniformly at random in field $\mathbb{F}$. With probability $1 - o_{|\mathbb{F}|;G}(1)$, $\mathbf{C}_G$ computes a good code with distance at least $(c-1)n$.

In Theorem 30, in order to make random coefficients work, we require $|\mathbb{F}| \gg dn2^{m+n}$, which depends on the number of inputs and outputs, as well as the depth. In terms of the *existence* of the coefficients, the following theorem gives a better bound on $|\mathbb{F}|$, which does not depend on the depth.

Theorem 33. *Let G be an (n, m)-superconcentrator, where $m \geq n$. If $|\mathbb{F}| > \left(\frac{2em}{n}\right)^n$, then there exists an assignment of coefficients in $\mathbb{F}$ such that the linear arithmetic circuit $\mathsf{C}_G : \mathbb{F}^n \to \mathbb{F}^m$ computes a superconcentrator code.*

Proof. Enumerate all the edges of G in a topological order, say, $e_1, e_2, \ldots, e_\ell$, where $\ell = |E(G)|$. Let

$$(V(G), \emptyset) = G_0 \subseteq G_1 \subseteq \cdots G_\ell = G$$

be a sequence of subgraphs of G such that $V(G_k) = V(G)$ and $E(G_k) = E(G_{k-1}) \cup \{e_k\}$, for $k = 1, 2, \ldots, \ell$. For each G_k, we will define the circuit C_{G_k} by converting each vertex to an addition gate, and choosing the coefficient for the edge e_k, while all other coefficients are the same as $\mathsf{C}_{G_{k-1}}$. Before stating the induction hypothesis, we need some notations.

Let $X \subseteq [n]$ be a subset of inputs, and $Y \subseteq [m]$ be a subset of outputs such that $|X| = |Y|$. Since G is a superconcentrator, there exist $|X|$ vertex-disjoint paths connecting X and Y; we *fix* one collection of such paths, denoted by $P_{X,Y} = \{p_1, p_2, \ldots, p_{|X|}\}$. Consider the collection of paths $P_{X,Y}$ on graph G_k, which becomes a collection of *partial paths* (because some edges are missing), that is,

$$P_{X,Y}^{(k)} \;=\; \left\{p_1^{(k)}, p_2^{(k)}, \ldots, p_{|X|}^{(k)}\right\} \;,$$

where $p_t^{(k)} = p_t \cap E(G_k)$ for $t = 1, 2, \ldots, |X|$. Let $E_{X,Y}^{(k)} = \{v_1^{(k)}, v_2^{(k)}, \ldots, v_{|X|}^{(k)}\}$ denotes the set of *end vertices*, that is, $v_t^{(k)}$ is the last vertex on the path $p_t^{(k)}$. Let $M_{X,Y}^{(k)}$ be the $|X| \times |Y|$ matrix such that

$$M_{i,j}^{(k)} \;=\; \sum_{p} \prod_{e \in E(p)} r_e \;,$$

where p ranges over all paths from input x_i to $v_j^{(k)}$ on graph G_k, and r_e denotes the coefficient associated with the edge e. Note that, when $k = \ell$, $M_{i,j}^{(k)}$ is exactly the (i,j)th entry of the submatrix $M_{X,Y}$ of the generator matrix (of the linear code computed by C_{G_k}).

The induction hypothesis is the following: for all $k = 0, 1, \ldots, \ell$, and for all equal-size $X \subseteq [n]$ and $Y \subseteq [m]$, the matrix $M_{X,Y}^{(k)}$ is of full rank.

Induction basis: $k = 0$. It is clear that, for any equal-size subsets X and Y, matrix $M_{X,Y}^{(0)}$ is the identity matrix (up to a row permutation).

Induction step: assuming the induction hypothesis is true for $k-1$, we will prove it is true for k. That is, when $|\mathbb{F}| > \left(\frac{2em}{n}\right)^n$, we can always choose the coefficient r_{e_k} such that $M_{X,Y}^{(k)}$ is of full rank for all pairs of X and Y.

Consider any $X \subseteq [n]$ and $Y \subseteq [m]$ of equal size s, and consider the collection of paths $P_{X,Y}^{(k-1)}$. Recall that graph G_k is obtained from G_{k-1} by adding the edge e_k, where $e_k = (u, v)$. There are three possibilities:

1. e_k extends some path in $P_{X,Y}^{(k-1)}$, that is, $p_t^{(k)} = p_t^{(k-1)} \cup e_k$ for some t.
2. v is the end vertex of some path in $P_{X,Y}^{(k-1)}$, i.e., $v \in E_{X,Y}^{(k-1)}$;
3. otherwise.

We will prove, in any of the three cases above, there is *at most* one choice for r_{e_k} such that $M_{X,Y}^{(k)}$ is *not* of full rank. Note that the total number of equal-size X, Y pairs is bounded

$$\sum_{i=1}^{n} \binom{n}{i}\binom{m}{i} \;\le\; 2^n \binom{m}{\le n} \;\le\; \left(\frac{2em}{n}\right)^n \;<\; |\mathbb{F}| \;.$$

Thus we can avoid all bad choices when $|\mathbb{F}|$ is large enough.

For notational convenience, let $X = \{1, 2, \ldots, s\}$, $Y = \{1, 2, \ldots, s\}$, and without loss of generality assume path $p_t^{(\ell)}$ connects input x_t and output y_t, for $t = 1, 2, \ldots, s$, where

$$P_{X,Y}^{(\ell)} = \left\{p_1^{(\ell)}, p_2^{(\ell)}, \ldots, p_s^{(\ell)}\right\} \;.$$

Case 1: e_k extends some path in $P_{X,Y}^{(k-1)}$, that is, $p_t^{(k)} = p_t^{(k-1)} \cup e_k$ for some $t \in [s]$. Without loss of generality, assume $t = 1$. Let us compare $M_{X,Y}^{(k)}$ with $M_{X,Y}^{(k-1)}$; it is clear that only the tth column may change. When X is fixed, for each $v \in V(G_{k-1})$, consider the column vector

$$\psi_X^{(k-1)}(v) \ \in \ \mathbb{F}^s \ ,$$

where the ith coordinate is defined to be the accumulated coefficient between x_i and v, that is,

$$\psi_X^{(k-1)}(v)|_i \ = \ \sum_p \prod_{e \in E(p)} r_e \ ,$$

where p enumerates all paths from x_i and v in graph G_{k-1}.

Let $E_{X,Y}^{(k-1)} = \{v_1^{(k-1)}, v_2^{(k-1)}, \ldots, v_s^{(k-1)}\}$ and $e_k = (u, v)$. Adopting the above notations, and comparing $M_{X,Y}^{(k)}$ with $M_{X,Y}^{(k-1)}$ column by column, we have

$$\begin{aligned}
\psi_X^{(k)}(v_1^{(k)}) &= r_{e_k} \cdot \psi_X^{(k-1)}(v_1^{(k-1)}) + \sum_{e'=(w,v) \in E(G_k)} r_{e'} \cdot \psi_X^{(k)}(w) \ , \\
\psi_X^{(k)}(v_2^{(k)}) &= \psi_X^{(k-1)}(v_2^{(k-1)}) \ , \\
\cdots & \quad \cdots \\
\psi_X^{(k)}(v_s^{(k)}) &= \psi_X^{(k)}(v_s^{(k-1)}) \ .
\end{aligned}$$

By the induction hypothesis, matrix $M_{X,Y}^{(k-1)}$ is of full rank, that is,

$$\psi_X^{(k-1)}(v_1), \psi_X^{(k-1)}(v_2), \ldots, \psi_X^{(k-1)}(v_s) \ \in \ \mathbb{F}^s$$

is a basis of $\mathbb{F}^s$. Thus, vector $\sum_{e'=(w,v)} r_{e'} \cdot \psi_X^{(k)}(w)$ can be written as a unique linear combination in the basis, say,

$$\sum_{e'=(w,v) \in E(G_k)} r_{e'} \cdot \psi_X^{(k)}(w) = \sum_{p=1}^{s} \alpha_p \cdot \psi_X^{(k-1)}(v_p) \ ,$$

where $\alpha_1, \alpha_2, \ldots, \alpha_s \in \mathbb{F}$. Hence,

$$\psi_X^{(k)}(v_1) = (r_{e_k} + \alpha_1)\psi_X^{(k-1)}(v_1) + \sum_{p=2}^{s} \alpha_p \cdot \psi_X^{(k-1)}(v_p) \ ,$$

which implies that $\psi_X^{(k)}(v_1), \psi_X^{(k)}(v_2), \ldots, \psi_X^{(k)}(v_s)$ are linearly independent if and only if $\alpha_1 + r_{e_k} = 0$. In other words, $M_{X,Y}^{(k-1)}$ is of full rank if $r_{e_k} \neq -\alpha_1$.

Case 2: v is the end vertex of some path in $P_{X,Y}^{(k-1)}$, i.e., $v \in E_{X,Y}^{(k-1)}$, where $e_k = (u, v)$. Let $E_{X,Y}^{(k-1)} = \{v_1^{(k-1)}, v_2^{(k-1)}, \ldots, v_s^{(k-1)}\}$, and without loss of generality, assume $v = v_1^{(k-1)}$. Comparing $M_{X,Y}^{(k)}$ with $M_{X,Y}^{(k-1)}$ column by column, we have

$$\begin{aligned}
\psi_X^{(k)}(v_1^{(k)}) &= \psi_X^{(k-1)}(v_1^{(k-1)}) + r_{e_k} \cdot \psi_X^{(k)}(u) \ , \\
\psi_X^{(k)}(v_2^{(k)}) &= \psi_X^{(k-1)}(v_2^{(k-1)}) \ , \\
\cdots &\quad \cdots \\
\psi_X^{(k)}(v_s^{(k)}) &= \psi_X^{(k-1)}(v_s^{(k-1)}) \ .
\end{aligned}$$

By induction hypothesis, $\psi_X^{(k-1)}(v_1^{(k-1)}), \psi_X^{(k-1)}(v_2^{(k-1)}), \ldots, \psi_X^{(k-1)}(v_s^{(k-1)}) \in \mathbb{F}^s$ are linearly independent, and thus forms a basis. So $\psi_X^{(k)}(u)$ can be written as a linear combination in the basis as follows:

$$\psi_X^{(k)}(u) \ = \ \sum_{p=1}^{s} \beta_p \cdot \psi_X^{(k-1)}(v_p^{(k-1)}) \ ,$$

where $\beta_1, \beta_2, \ldots, \beta_s \in \mathbb{F}$. Therefore,

$$\psi_X^{(k)}(v_1^{(k)}) \ = \ (r_{e_k}\beta_1 + 1) \cdot \psi_X^{(k-1)}(v_1^{(k-1)}) + \sum_{p=2}^{s} r_e \beta_p \cdot \psi_X^{(k-1)}(v_p^{(k-1)}) \ .$$

It is clear that $\psi_X^{(k)}(v_1^{(k)}), \psi_X^{(k)}(v_2^{(k)}), \ldots, \psi_X^{(k)}(v_s^{(k)})$ are linearly dependent if and only if $r_{e_k}\beta_1 + 1 = 0$, which implies that there is *at most* one choice for r_{e_k} such that $M_{X,Y}^{(k)}$ is not of full rank.

Case 3: otherwise. In this case, $M^{(k)}_{X,Y} = M^{(k-1)}_{X,Y}$, regardless of r_{e_k}.

In summary, for each X, Y pair, there is at most one choice for r_{e_k} such that $M^{(k)}_{X,Y}$ is not of full rank. Whenever $|\mathbb{F}|$ is larger than the total number of X, Y pairs, we can always avoid the bad choices for all possible X, Y pairs. Therefore, we complete the induction proof. □

The following corollary is similar to prove.

Corollary 34. *Let G be an (n, m)-superconcentrator, where $m \geq n$. If $|\mathbb{F}| > \binom{m}{n}$, then there exists an assignment of coefficients in $\mathbb{F}$ such that the linear arithmetic circuit $\mathsf{C}_G : \mathbb{F}^n \to \mathbb{F}^m$ computes an MDS code.*

Proof. The proof is almost the same as Theorem 33; we sketch the difference. Instead of considering all equal-size subsets $X \subseteq [n]$ and $Y \subseteq [m]$, we fix $X := [n]$, and let $Y \subseteq [m]$ be a subset of size n. The total number of such X, Y pairs is bounded by $\binom{m}{n}$. □

Following [21], an (m, ℓ, n)-concentrator (of depth 1) is a bipartite graph

$$G(U = [m], V = [\ell], E)$$

such that for any $X \subseteq V$ of size at most n, there are $|X|$ vertex-disjoint edges connecting X to U. Let $\mathbf{H}_G : \mathbb{F}^\ell \to \mathbb{F}^k$ denotes the random linear map such that

$$y_j = \sum_{(i,j) \in E} \mathbf{r}_{i,j} x_i \ ,$$

where $\mathbf{r}_{i,j} \in \mathbb{F}$ is chosen uniformly at random. In other words, bipartite graph G is converted into a linear arithmetic circuit, by replacing each vertex in V with an addition gate and choosing the coefficient for each edge uniformly at random. The following result says that a superconcentrator code concatenated with a random "concentrator map" is a good code with high probability.

Lemma 35. *Let $C : \mathbb{F}^n \to \mathbb{F}^m$ be a superconcentrator code, and let $G(U = [m], V = [\ell], E)$ be an (m, ℓ, n)-concentrator, where $\ell = cn$ for some constant $c > 2$. Then*

$$C \circ \mathbf{H}_G : \mathbb{F}^n \to \mathbb{F}^\ell$$

is a good code with probability $1 - o_n(1)$.

Proof. Let $C(x_1, x_2, \ldots, x_n) = (z_1, z_2, \ldots, z_m)$, and let $\mathbf{H}_G(z_1, z_2, \ldots, z_m) = (y_1, y_2, \ldots, y_\ell)$. We will prove that, for each fixed $x \in \mathbb{F}^n$ of weight w, there exists a decomposition of outputs

$$\widetilde{Y}_1 \dot{\cup} \widetilde{Y}_2 \dot{\cup} \cdots \dot{\cup} \widetilde{Y}_t = \{y_1, y_2, \ldots, y_\ell\}$$

such that

- $|\widetilde{Y}_k| \leq w$ for all $k = 1, 2, \ldots, t-1$, and $0 \leq |\widetilde{Y}_t| < w$;
- $\left(\widetilde{\mathbf{Y}}_1(x), \widetilde{\mathbf{Y}}_2(x), \ldots, \widetilde{\mathbf{Y}}_{t-1}(x)\right) \in \mathbb{F}^{\ell - |\widetilde{Y}_t|}$ is uniformly distributed.

Partition $\widetilde{Y}_1 \dot{\cup} \widetilde{Y}_2 \dot{\cup} \cdots \dot{\cup} \widetilde{Y}_t = \{y_1, y_2, \ldots, y_n\}$ will be generated by the following procedure. Let $X := \mathrm{supp}(x)$ and $x' := x|_X \in \mathbb{F}^w$. Let $Y_1 \subseteq \{y_1, y_2, \ldots, y_\ell\}$ be any subset of size w. By the definition of (m, ℓ, n)-concentrator, there exists set $Z_1 \subseteq \{z_1, z_2, \ldots, z_m\}$ of size w such that Y_1 and Z_1 are connected by w vertex-disjoint edges (which is a matching); denote this bijection (matching) by $\phi_1 : Z_1 \to Y_1$. For fixed inputs x, it is clear that

$$Z_1(x) = x' M_{X,Z_1} \ \in \ \mathbb{F}^w \ ,$$

where $M \in \mathbb{F}^{n \times m}$ is the generator matrix of the linear code C. Since C is a superconcentrator code, M_{X,Z_1} is of full rank w, and thus $Z_1(x) \neq \overline{0}$. Let z_1' denotes the maximal subvector of $Z_1(x)$ such that all coordinates are nonzero, i.e., $Z_1' = \mathrm{supp}(Z_1(x))$ and $z_1' := Z_1(x)|_{Z_1'}$. Let $\widetilde{Y}_1 = \phi_1(Z_1')$. Repeat this procedure for $\widetilde{Y}_2$, and so on; everything is the same except that Y_2 is a subset of the *remaining* vertices in $\{y_1, y_2, \ldots, y_\ell\}$ of size w.

Claim 36. *For each $k = 1, 2, \ldots, t-1$, output $\widetilde{\mathbf{Y}}_k(x) \in \mathbb{F}^{|\widetilde{Y}_k|}$ is uniformly distributed.*

Proof. Let $\mathbf{R}$ be the $m \times \ell$ matrix such that

$$\mathbf{R}_{i,j} = \begin{cases} \mathbf{r}_{i,j}, & (i,j) \in E(G), \\ 0, & \text{otherwise.} \end{cases} \tag{2.13}$$

Thus, $y = z\mathbf{R}$ defines the linear mapping $\mathbf{H}_G$. Using this notation,

$$\begin{aligned} \widetilde{\mathbf{Y}}_k(x) &= z\mathbf{R}_{[m],\widetilde{Y}_k} \\ &= z'_k \mathbf{D}_{Z'_k,\widetilde{Y}_k} + z\left(\mathbf{R}_{[m],\widetilde{Y}_k} - \mathbf{D}_{Z'_k,\widetilde{Y}_k}\right) , \end{aligned} \tag{2.14}$$

where $\mathbf{D}_{Z'_k,\widetilde{Y}_k}$ denotes the diagonal matrix (up to a row permutation) corresponding to the matching ϕ_k.

Let us *fix* all the coefficients in $\mathbf{R}_{[m],\widetilde{Y}_k} - \mathbf{D}_{Z'_k,\widetilde{Y}_k}$ first, which are all the coefficients on the edges incident to $\widetilde{Y}_k$ except the matching ϕ_k. Then (2.14) becomes

$$\widetilde{\mathbf{Y}}_k(x) = z'_k \mathbf{D}_{Z'_k,\widetilde{Y}_k} + u \ .$$

Note that each coordinate of z'_k is nonzero, and $\mathbf{D}_{Z'_k,\widetilde{Y}_k}$ is a diagonal matrix (up to a permutation) such that each diagonal entry is a distinct random variable uniformly distributed in $\mathbb{F}$. Thus, vector $z'_k \mathbf{D}_{Z'_k,\widetilde{Y}_k} \in \mathbb{F}^{|\widetilde{Y}_k|}$ is uniformly distributed, which implies that $z'_k \mathbf{D}_{Z'_k,\widetilde{Y}_k} + u$ is also uniformly distrusted, because vector u is fixed. [3] □

Claim 37. $\left(\widetilde{\mathbf{Y}}_1(x), \widetilde{\mathbf{Y}}_2(x), \ldots, \widetilde{\mathbf{Y}}_{t-1}(x)\right) \in \mathbb{F}^{\ell - |\widetilde{Y}_t|}$ *is uniformly distributed.*

Proof. Because edges incident to $\widetilde{Y}_i$ and $\widetilde{Y}_j$ are disjoint for distinct i and j. □

3. We use the facts that $(\mathbb{F}, +)$ and $(\mathbb{F}, \times)$ are groups.

The rest is a counting argument. For fixed $x \in \mathbb{F}^n$ of weight w, we have shown that

$$\left(\widetilde{\mathbf{Y}}_1(x), \widetilde{\mathbf{Y}}_2(x), \ldots, \widetilde{\mathbf{Y}}_{t-1}(x)\right) \ \in \ \mathbb{F}^{\ell - |\widetilde{Y}_t|}$$

is uniformly distributed, and $|\widetilde{Y}_t| < w$. Let $\delta > 0$ be a constant to be determined later, which only depends on c. Thus,

$$\begin{aligned} \Pr\left[\mathrm{wt}(C \circ \mathbf{H}_G(x)) < \delta n\right] &\leq \frac{\binom{cn}{\delta n}|\mathbb{F}|^{\delta n}}{\mathbb{F}^{cn-w+1}} \\ &\leq 2^{n\left(cH\left(\frac{\delta}{c}\right) - (c-1-\delta)\log|\mathbb{F}|\right)} , \end{aligned}$$

where $H\left(\frac{\delta}{c}\right)$ is the binary entropy function. Applying a union bound over all nonzero x, we have

$$\Pr\left[\exists x \neq \overline{0} \text{ such that } \mathrm{wt}(C \circ \mathbf{H}_G(x)) < \delta n\right] \ \leq \ 2^{n\left(cH\left(\frac{\delta}{c}\right) - (c-2-\delta)\log|\mathbb{F}|\right)} .$$

Let $\delta > 0$ be a sufficiently small constant such that $cH\left(\frac{\delta}{c}\right) - (c-2-\delta)\log|\mathbb{F}| < 0$, and the proof is complete. □

Remark 38. *The above lemma is also true with a weaker assumption that $C : \mathbb{F}^n \to \mathbb{F}^m$ is a linear MDS codes, that is, any set of n columns of its generator matrix is linearly independent.*

Corollary 39. *Let G be an (n, cn)-superconcentrator, where constant $c > 2$, and there are m vertices incident to the cn outputs.* [4] *If $|\mathbb{F}| > \binom{m}{n}$, then there exists an assignment of coefficients in $\mathbb{F}$ such that the linear arithmetic circuit $\mathsf{C}_G : \mathbb{F}^n \to \mathbb{F}^{cn}$ computes a good code.*

Proof. Note that the linear arithmetic circuit $\mathsf{C}_G : \mathbb{F}^n \to \mathbb{F}^{cn}$ can be decomposed into two parts:

$$\mathsf{C}_1 : \mathbb{F}^n \to \mathbb{F}^m \text{ and } \mathsf{C}_2 : \mathbb{F}^m \to \mathbb{F}^{cn} ,$$

4. If the superconcentrator is layered, then m is the number of vertices on the second last layer.

such that $\mathsf{C}_G = \mathsf{C}_1 \circ \mathsf{C}_2$. Apply Corollary 34 so that C_1 computes an MDS code, which is possible since $|\mathbb{F}| > \binom{m}{n}$. For C_2, we choose the coefficients uniformly at random; By Lemma 35, we claim $\mathsf{C}_1 \circ \mathsf{C}_2$ computes a good code. □

What if the inputs and outputs of the superconcentrator codes are switched? It turns out that it computes a so-called *resilient function*.

Definition 40 (Definition 2 in Chor *et al.* [19])**.** *Let $\mathbb{F}$ be a finite field, and let $f : \mathbb{F}^n \to \mathbb{F}^m$ be a function on n variables $x_1, x_2, \ldots, x_n$. The function is said to be* uniformly distributed with respect to $T \subseteq [n]$ *if the random variable $f(x_1, x_2, \ldots, x_n)$ is uniformly distributed in $\mathbb{F}^m$, when $\{x_i : i \notin T\}$ is a set of independent uniformly distributed random variables, and $\{x_i : i \in T\}$ is a set of constants.*

A function $f : \mathbb{F}^n \to \mathbb{F}^m$ is said to be t-resilient *if for every $T \subseteq [n]$ of cardinality t, the function is uniformly distributed with respect to T.*

The following lemma is proved in [19] for the case $\mathbb{F} = \mathbb{F}_2$, and proved in [36] for the case $\mathbb{F} = \mathbb{F}_p$ for any prime p in a quantitative version using Fourier analysis. For other finite field $\mathbb{F}_q$, where $q = p^m$ for prime p and integer $m \geq 2$, there is a reduction to the case $\mathbb{F}_p$ as we will show.

Lemma 41. *For any finite field $\mathbb{F}$, a set $\{x_i\}_{i=1}^n$ of random variables in $\mathbb{F}$ is uniformly and independently distributed if and only if for all not-all-zero $a_1, a_2, \ldots, a_n \in \mathbb{F}$,*

$$a_1 x_1 + a_2 x_2 + \ldots + a_n x_n \in \mathbb{F}$$

is uniformly distributed.

Proof. (Reduction from $\mathbb{F}_q$ to $\mathbb{F}_p$) The "only if" direction is trivial.

For the "if" direction, let $\mathbb{F} = \mathbb{F}_q$, where $q = p^m$ for some prime p and integer $m \geq 2$, and we assume for the case $\mathbb{F} = \mathbb{F}_p$, the lemma is proved. [5] Fix an irreducible polynomial

5. Interested readers can refer to [36] for a three-line proof.

$f(t)$ of degree m over $\mathbb{F}_p$, and it is well known that $\mathbb{F}_q$ is isomorphic to $\mathbb{F}_p[t]/(f(t))$.

Case 1: $n = 1$. By condition, for any $a \in \mathbb{F}_q$, ax is uniformly distributed in $\mathbb{F}_q$. In the quotient ring $\mathbb{F}_p[t]/(f(t))$, we write $a = a_0 + a_1 t + \ldots + a_{m-1}t^{m-1}$ and $x = x_0 + x_1 t + \ldots + x_{m-1}t^{m-1}$. (At this point, we observe that that x is uniformly distributed in $\mathbb{F}_q$ if and only if $(x_0, x_1, \ldots, x_{m-1})$ is uniformly distributed in $\mathbb{F}_p^m$. This fact will be used shortly.) Thus, in the quotient ring $\mathbb{F}_p[t]/(f(t))$, the product ax can be written as

$$\begin{aligned} ax &= a_0x_0 \\ &\quad + (a_0x_1 + a_1x_0) \cdot t \\ &\quad + (a_0x_2 + a_1x_1 + a_2x_0) \cdot t^2 \\ &\quad + \ldots \\ &\quad + a_{m-1}x_{m-1} \cdot t^{2m-2} \ . \end{aligned}$$

Note that for all $i \geq m$, polynomial $t^i \equiv r_i(t) \mod f(t)$ for a unique polynomial $r_i(t)$ of degree $< m$, and thus ax can be written as

$$ax = L_0(a, x) + L_1(a, x) \cdot t + \ldots + L_{m-1}(a, x) \cdot t^{m-1} \ ,$$

where $L_i(a, x)$ is a *bilinear form* in a and x, viewed as vectors in $\mathbb{F}_p^m$. Thus we have $L_i(a + b, x) = L_i(a, x) + L_i(b, x)$ for all $a, b \in \mathbb{F}_q$.

Claim 42. *Bilinear form $L_0(a, x)$ is distinct for different $a \in \mathbb{F}_q$.*

Proof. Suppose for contradiction's sake that there exist distinct $a, b \in \mathbb{F}_q$ such that $L_0(a, x) = L_0(b, x)$. By linearity, $L_0(a - b, x) = L_0(a, x) - L_0(b, x) = 0$, which implies that the first coordinate of $(a - b)x$ is always 0, and thus $(a - b)x$ is not uniformly distributed in $\mathbb{F}_q$. Contradiction. □

Given the above claim, we count the number of different $a \in \mathbb{F}_q$ and different linear

forms $L_0(a,x)$ in x, which implies that for any $y \in \mathbb{F}_p^m$, there exists some $a \in \mathbb{F}_q$ such that $L_0(a,x) = y^T x$. Since ax is uniformly distributed in $\mathbb{F}_q$, we claim $L_0(a,x) = y^T x$ is uniformly distributed in $\mathbb{F}_p$. Applying the Lemma for the case $\mathbb{F} = \mathbb{F}_p$, we know that $(x_0, x_1, \ldots, x_{m-1})$ is uniformly distributed in $\mathbb{F}_p^m$, which implies that x is uniformly distributed in $\mathbb{F}_q$.

Case 2: $n \geq 2$. The proof is similar to Case 1, and let us sketch the outline. Let us write each $x_i \in \mathbb{F}_q$ as

$$x_i = (x_{i,0}, x_{i,1}, \ldots, x_{i,m-1}) \in \mathbb{F}_p^m,$$

where $x_{i,j}$ is the coefficient of t^j when viewing x_i as a polynomial in $\mathbb{F}_p[t]/(f(t))$. Using similar counting argument, we can show that for any fixed $(a_{i,j}) \in \mathbb{F}_p^{mn}$, the sum

$$\sum_{i,j} a_{i,j} x_{i,j}$$

is uniformly distributed in $\mathbb{F}_p$. Since the Lemma is true for $\mathbb{F} = \mathbb{F}_p$, we claim $(x_{i,j}) \in \mathbb{F}_p^{mn}$ is uniformly distributed, and thus $(x_1, x_2, \ldots, x_n) \in \mathbb{F}_q^n$ is uniformly distributed.

□

The next theorem is stated in [19] for the case $\mathbb{F} = \mathbb{F}_2$. For general finite field $\mathbb{F}$, the proof is the same, equipped with the above lemma. We reproduce the proof here.

Theorem 43 (Theorem 2 in [19])**.** *Let $M \in \mathbb{F}^{m \times n}$ be a matrix. M is a generator matrix of a linear error correcting code with distance at least $t+1$ if and only if $f(x) = xM^T$ is t-resilient.*

Proof. First, let us prove that if the linear code given by the generator matrix M has distance at least $t+1$, then $xM^T : \mathbb{F}^m \to \mathbb{F}^n$ is t-resilient. In order to prove xM^T is t-resilient, by Lemma 41, it suffices to prove that for any nonzero $y \in \mathbb{F}^n$, and for any $T \subseteq [m]$ of size t, $xM^T y^T$ is uniformly distributed in $\mathbb{F}$, where x_i, for $i \in T$, are set to constants arbitrarily, and x_i, for all $i \notin T$, are uniformly distributed. Note that $xM^T y^T = yMx^T$. Since the linear code given by the generator matrix M has distance at least $t+1$, vector yM has weight at

least $t+1$, which implies that the inner product yMx^T contains at least one random variable x_i with $i \notin T$. Thus, yMx^T is uniformly distributed.

For the other direction, we will prove that if the linear code given by the generator matrix M has distance $\leq t$, then xM^T is not t-resilient. Since the linear code given by the generator matrix M has distance $\leq t$, there exists some nonzero $y \in \mathbb{F}^n$ such that $\mathrm{wt}(yM) \leq t$. Let us consider

$$xM^T y^T = yMx^T \ .$$

We can set all x_i with $i \in \mathrm{supp}(yM)$ to be zero, while other x_i's are uniformly distributed, so that yMx^T is identically zero, which implies that xM^T is not t-resilient. □

The following corollary follows from Corollary 32 and Theorem 43.

Corollary 44. *Let G be a superconcentrator with cn inputs and n outputs, where constant $c > 1$. Let*

$$\mathbf{C}_G : \mathbb{F}^{cn} \to \mathbb{F}^n$$

be the linear arithmetic circuit by replacing each vertex with an addition gate over $\mathbb{F}$, and choosing the coefficient on each edge uniformly at random. With probability $1 - o_{|\mathbb{F}|;G}(1)$, $\mathbf{C}_G$ computes a $(c-1)\,n$-resilient function.

2.5 Superconcentrators in Network Coding

Network coding is an approach to improve the capacity for network transmission, which in some setting outperforms more traditional network communications based on routing and replication. In [3], Ahlswede *et al.* proved that for the so-called *multicast task*, network coding can achieve the information-theoretic upper bound, i.e., the max-flow min-cut bound. In [55], Li *et al.* proved that linear encoding suffices to achieve the optimum rate in the above multicast scenario. In [42], Ho *et al.* proved that even *random* linear encoding can achieve the optimal capacity, when the size of the underlying finite field is large enough.

As far as we know, most literature in network coding usually studied the network capacity with an eye to a *fixed* network topology, and little attention is paid to the *designing* of the network. In this section, we show that if the network graph is a superconcentrator, random linear network coding satisfies some nice entropy property, so called *lossless decodable* (which will defined in Section 2.5.1). Considering there exist small-size explicit superconcentrators [1], our result may have some impact in practice.

2.5.1 Lossless Decodable Network Coding

We will briefly formulate (linear) network coding following [3, 55], and then define *lossless decodable* network coding schemes.

Let $\mathbb{F}$ be a finite field. Let $G(U \cup V \cup W, E)$ be a network (directed acyclic graph) with inputs (or called *sources*) $U = [n]$ and outputs (or called *sinks*) $V = [m]$.

Definition 45. *Given network $G(U \cup V \cup W, E)$, an* encoding scheme *is a collection of function $\{f_v\}_{v \in U \cup W \cup V}$, where*

$$f_v : \mathbb{F}^{\deg^-(v)} \to \mathbb{F} \ ,$$

where $\deg^-(v)$ *denotes the* indegree *of v, i.e., the number of incoming edges to v. (With slight abuse of notation, let* $\deg^-(u) := 1$ *for all sources $u \in U$.)*

Note that we perform encoding on all vertices including sources and sinks.

Given encoding scheme $\{f_v\}_{v \in U \cup W \cup V}$, we can define the function *computed* at each node in the natural way as follows:

1. Let $x_1, x_2, \ldots, x_n$ be formal variables over $\mathbb{F}$;
2. For input vertex $u \in U = [n]$, let $F_u := f_u(x_u)$;
3. For vertex $v \in W \cup V$, let

$$F_v \ := \ f_v(F_{u_1}, F_{u_2}, \ldots, F_{u_d}) \ ,$$

where $u_1, u_2, \ldots, u_d$ enumerate all vertices $u \to v$ in some predetermined order.

Note that each $F_u = F_u(x_1, x_2, \ldots, x_m)$ is a function from $\mathbb{F}^m$ to $\mathbb{F}$.

Say encoding scheme $\{f_v\}_{v \in V(G)}$ is a *linear encoding scheme* if all f_v's are (homogenous) linear functions; thus each F_v is also linear by definition. A *random linear encoding scheme* is a linear encoding scheme where all coefficients are chosen uniformly at random in $\mathbb{F}^* := \mathbb{F} \setminus \{0\}$.

Definition 46. *Given network $G(U \cup V \cup W, E)$ with inputs $U = [n]$ and outputs $V = [m]$, and s (not necessarily linear) encoding schemes $F^{(1)}, F^{(2)}, \ldots, F^{(s)}$. For any $I \subseteq [s]$, $J \subseteq [m]$, $K \subseteq [n]$, and any* partial assignment $\rho : \{x_k : k \in [n] \setminus K\} \to \mathbb{F}$, *let*

$$\begin{aligned} & \mathrm{Ent}_F(I, J, K; \rho) \\ := \quad & \log_{|\mathbb{F}|} \left| \left\{ \left(F_j^{(i)}(x) \right)_{i \in I, j \in J} \in \mathbb{F}^{|I| \cdot |J|} : x \in \mathbb{F}^n \text{ consistent with } \rho \right\} \right| , \end{aligned} \tag{2.15}$$

and let

$$\mathrm{Ent}_F(I, J, K) \ := \ \min_{\rho} \mathrm{Ent}_F(I, J, K; \rho) \ , \tag{2.16}$$

where ρ ranges over all partial assignments $\rho : \{x_k : k \in [n] \setminus K\} \to \mathbb{F}$.

In other words, when restricting everything outside of K to ρ, the quantity $\mathrm{Ent}_F(I, J, K; \rho)$ measures the total information received by sinks J under the encoding schemes in I. By definition, it is clear that $\mathrm{Ent}_F(I, J, K; \rho) \leq \min(|I| \cdot |J|, |K|)$.

The following proposition is not difficult to prove.

Proposition 47. *Suppose $F^{(1)}, F^{(2)}, \ldots, F^{(s)}$ are linear encoding schemes. For any $I \subseteq [s], J \subseteq [m]$ and $K \subseteq [n]$,*

$$\mathrm{Ent}_F(I, J, K) \ = \ \mathrm{rank} \left\{ F_j^{(i)}\big|_K : i \in I, j \in J \right\} \ ,$$

where $F_j^{(i)} \in \mathbb{F}^m$ denotes the vector representation of the linear function $F_j^{(i)}(x_1, \ldots, x_n)$, and $F_j^{(i)}\big|_K \in \mathbb{F}^{|K|}$ denotes the vector restricted to the coordinates indexed by K.

Definition 48 (Lossless Decodable)**.** *Let $G(U \cup V \cup W, E)$ be a network with inputs $U = [n]$ and outputs $V = [m]$, and s encoding schemes $F^{(1)}, F^{(2)}, \ldots, F^{(s)}$. Let $\mathrm{Ent}_F(I, J, K)$ be defined as in Definition 46. Encoding schemes $F^{(1)}, F^{(2)}, \ldots, F^{(s)}$ are called* lossless decodable *if*

$$\mathrm{Ent}_F(I, J, K) \;=\; \min(|I| \cdot |J|, |K|)$$

for all $I \subseteq [s]$, $J \subseteq [m]$, and $K \subseteq [n]$.

We consider the following application scenario:

1. Network $G(U \cup V \cup W)$ performs s encoding schemes *in turn*, where the encoding schemes are given by $F^{(1)}, F^{(2)}, \ldots, F^{(s)}$.

2. Each source $u \in U(G) = [n]$ carries commodity $x_u \in \mathbb{F}$ to transmit during s different encoding schemes. (Or x_u could be a vector of any *fixed* length, and the encoding function is applied to x_u independently on each coordinate.)

3. Through the network, each sink v receives $F_v^{(i)}(x_1, \ldots, x_n)$, and assume the functions $F_v^{(i)}$ for all $v \in V(G)$, $i \in [s]$, are known to the receivers. [6]

4. We are only given access to a set of sinks $J \subseteq V(G) = [m]$ under encoding schemes indexed by $I \subseteq [s]$ with the aim to decode x_k for all $k \in K$; assume we know x_k for all $k \notin K$, denoted by $\rho : \{x_k : k \in [n] \setminus K\} \to \mathbb{F}$.

5. $|I| \cdot |J| = |K|$.

Given I, J, K and ρ as above, regardless of the computational complexity, we can decode all x_k for $k \in K$ if and only if $\mathrm{Ent}_F(I, J, K; \rho) \geq |K|$. This is the reason we call them lossless decodable.

The next proposition asserts that, the underlying network of any lossless decodable encoding schemes must be a superconcentrator.

6. This assumption looks very strong. However, for linear encoding schemes, it suffices to carry an encoding vector of length n on every edge [55].

Proposition 49. *Let $G(U \cup V \cup W, E)$ be a network with inputs $U = [n]$ and outputs $V = [m]$ such that there exist lossless decodable encoding schemes $F^{(1)}, F^{(2)}, \ldots, F^{(s)}$, $s \geq 1$, then G is a superconcentrator.*

Proof. Without loss of generality, assume $s = 1$, and let $F_1, F_2, \ldots, F_m$ denote the linear functions computed by the sinks. Since $F = F^{(1)}$ is lossless decodable, we have

$$\mathrm{Ent}_F(\{1\}, J, K) \;=\; \min(|J|, |K|) \tag{2.17}$$

for any $J \subseteq [m]$ and $K \subseteq [n]$. Fix any $J \subseteq U(G) = [m]$ and $K \subseteq V(G) = [n]$ with equal size.

Assume for contradiction that there does *not* exist $|J|$ vertex-disjoint paths connecting J and K. By Menger's theorem, there exists a set of vertices of size less than $|J|$, say X, whose removal disconnects J and K.

The following claim is easy to prove.

Claim 50. *Fix any F_j, $j \in J$. Let X be a subset of vertices whose removal disconnects j and U. (X may intersect with U.) Then F_j can be written a function in F_u, $u \in X$. In other words, F_j is completely determined by the values of F_u, for all $u \in X$.*

By the above claim, the values of $(F_j)_{j \in J}$ are completely determined by the values of $(F_u)_{u \in X \cup (U \setminus K)}$. When the input values of x_u for all $u \in U \setminus K$ are known, $(F_j)_{j \in J}$ can be determined by the values of $(F_u)_{u \in X}$, which implies that, for any fixed partial assignment $\rho : \{x_k : k \in U \setminus K\} \to \mathbb{F}$,

$$\left| \left\{ \big(F_j(x)\big)_{j \in J} : x \in \mathbb{F}^m \text{ consistent with } \rho \right\} \right| \;\leq\; |\mathbb{F}|^{|X|} \;\leq\; |\mathbb{F}|^{|J|-1} \,.$$

This is a contradiction with (2.17). □

2.5.2 Random Linear Network Coding on Superconcentrators

For linear encoding schemes, *lossless decodability* corresponds to certain minor conditions of the encoding matrix. Given network $G(U \cup V \cup W, E)$ with inputs $U = [n]$, outputs $V = [m]$, and linear encoding schemes $F^{(1)}, F^{(2)}, \ldots, F^{(s)}$, let $M \in \mathbb{F}^{sm \times n}$ be the matrix, whose rows are indexed by $[s] \times [m]$, and columns indexed by $[n]$, and

$$M\left((i,j);[n]\right) \;:=\; F_i^{(j)} \;\in \mathbb{F}^n \quad .$$

That is, $M_{(i,j),k}$ is the kth coordinate of $F_i^{(j)}$, and the (i,j)th row corresponds to the coefficients of $F_j^{(i)}$.

By Proposition 47, the lossless decodable condition is equivalent to the condition

$$\operatorname{rank} M(I \times J; K) \;=\; \min\left(|I| \cdot |J|, |K|\right)$$

for all $I \subseteq [s]$, $J \subseteq [m]$ and $K \subseteq [n]$.

Next, we will prove random linear encoding schemes on *any* superconcentrator are lossless decodable when the size of the field $\mathbb{F}$ is large enough.

Theorem 51. *Let $G(U \cup V \cup W, E)$ be a superconcentrator with inputs $U = [n]$ and outputs $V = [m]$. Let $\mathbf{F}^{(1)}, \mathbf{F}^{(2)}, \ldots, \mathbf{F}^{(s)}$ be s random linear encoding schemes, then they are lossless decodable with probability $1 - o_{|\mathbb{F}|;G,s}(1)$.*

Proof. Recall that $\mathbf{M}$ is an $sm \times n$ matrix, whose rows are indexed by $(i,j) \in [s] \times [m]$, columns indexed by $[n]$, and $\mathbf{M}_{(i,j),k}$ is the kth coordinate of the vector $\mathbf{F}_j^{(i)}$.

Fix $I \subseteq [s]$, $J \subseteq [m]$, and $K \subseteq [n]$, satisfying $|I| \cdot |J| = |K|$. For notational convenience, assume $I = \{1, 2, \ldots, |I|\}$ without loss of generality. Let $S_1, S_2, \ldots, S_{|I|}$ be a decomposition of K such that $|S_1| = |S_2| = \ldots = |S_{|I|}| = |J|$.

Claim 52. *For each $i \in I$, submatrix*

$$\mathbf{M}(\{i\} \times J; S_i) \ \in \ \mathbb{F}^{|J| \times |J|}$$

is of full rank with probability $\geq 1 - \frac{d|J|}{|\mathbb{F}^*|}$, where d is the depth of G.

Proof. Observe that $\mathbf{M}(\{i\} \times J; S_i)$ is a $|J| \times |J|$ matrix, where each entry is of degree at most d. Since G is a superconcentrator, there exist $|J|$ vertex-disjoint paths connecting J and S_i; in the ith encoding scheme $\mathbf{F}^{(i)}$, we can set coefficients on those $|J|$ paths to 1 and 0 everywhere else, which implies that there *exists* an assignment of random coefficients such that $\det \mathbf{M}(\{i\} \times J; S_i)$ is nonzero. Therefore, we claim $\det \mathbf{M}(\{t\} \times J; S_i)$ is a nonzero polynomial (in the random coefficients associated with the linear encoding schemes) of degree $\leq d|J|$. The Claim follows by applying Schwartz-Zippel Lemma. □

Claim 53.

$$\Pr\left[\det \mathbf{M}(I \times J; K) = 0\right] \ \leq \ \sum_{i=1}^{|I|} \Pr\left[\det \mathbf{M}(\{i\} \times J; S_i) = 0\right] + \frac{1}{|\mathbb{F}^*|} \ .$$

Proof. Recall that we let $I = \{1, 2, \ldots, |I|\}$ for notational convenience. Let

$$\mathbf{M}(I \times J; K) = \begin{bmatrix} \mathbf{M}\left(\{1\} \times J; S_1\right) & \mathbf{Y}(1,2) & \ldots & \mathbf{Y}(1,|I|) \\ \mathbf{Y}(2,1) & \mathbf{M}\left(\{2\} \times J; S_2\right) & \ldots & \mathbf{Y}(2,|I|) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{Y}(|I|,1) & \mathbf{Y}(|I|,2) & \ldots & \mathbf{M}\left(\{|I|\} \times J; S_{|I|}\right) \end{bmatrix} ,$$

and we think of each entry as a (multilinear) polynomial in the random coefficients, which are not chosen yet.

Let $\mathbf{Y}(i,j) = \mathbf{R}(i,j)\mathbf{M}(i,j)$ for $i \neq j$, where $\mathbf{M}(i,j)$ is the $|J| \times |J|$ matrix corresponding to the linear encoding coefficients of inputs S_j in outputs J during the encoding scheme $\mathbf{F}^{(i)}$ *excluding* the encodings applied at the sources, and $\mathbf{R}(i,j) \in \mathbb{F}^{|J| \times |J|}$ is a diagonal matrix

corresponding to the random coefficients applied to the sources S_j during the encoding scheme $\mathbf{F}^{(i)}$.

Observe that the diagonal entry (k,k) in $\mathbf{R}(i,j)$ *only* appears in the kth row of submatrix $\mathbf{Y}(i,j)$, which implies that this variable will have degree one in the polynomial $\det \mathbf{M}(I \times J; K)$. Expand the determinant of $\mathbf{M}(I \times J; K)$ as

$$\prod_{i=1}^{|I|} \det \mathbf{M}\left(\{i\} \times J; S_i\right) + \mathbf{T} \ ,$$

where the term $\mathbf{T}$ is a sum of products (of determinants of submatrices), and each product contains at least one $\mathbf{R}(i,j)$ for $i \neq j$. Thus, we can write

$$\mathbf{T} = \mathbf{r}_1 \mathbf{T}_1 + \mathbf{r}_2 \mathbf{T}_2 + \ldots$$

where $\mathbf{r}_1, \mathbf{r}_2, \ldots$ enumerate all the indeterminates in $\mathbf{R}(i,j)$ for all $i \neq j$. Note that $\mathbf{r}_i$ does not appear in $\mathbf{T}_j$ for all j, and thus $\mathbf{T}$ is *linear* in every $\mathbf{r}_i$. Write

$$\mathbf{T} \ = \ \mathbf{r}_1 \left(\mathbf{T}_1 + \frac{\mathbf{r}_2}{\mathbf{r}_1} \cdot \mathbf{T}_2 + \frac{\mathbf{r}_3}{\mathbf{r}_1} \cdot \mathbf{T}_3 + \ldots \right) \ .$$

Choosing $\mathbf{r}_1, \mathbf{r}_2, \ldots \in \mathbb{F}^*$ uniformly at random is equivalent to choosing $\frac{\mathbf{r}_2}{\mathbf{r}_1}, \frac{\mathbf{r}_3}{\mathbf{r}_1}, \ldots$ uniformly at random in $\mathbb{F}^*$, and choosing $\mathbf{r}_1 \in \mathbb{F}^*$ at last. Consider the last step, that is, the step right before choosing $\mathbf{r}_1$. If $\mathbf{T}' := \mathbf{T}_1 + \frac{\mathbf{r}_2}{\mathbf{r}_1} \cdot \mathbf{T}_2 + \frac{\mathbf{r}_3}{\mathbf{r}_1} \cdot \mathbf{T}_3 + \ldots$ is zero, then $\mathbf{T} = 0$ with probability 1, which implies

$$\begin{aligned}
& \Pr\left[\det \mathbf{M}\left(I \times J; K\right) = 0 \ \mid \ \mathbf{T}' = 0\right] \\
= \ & \Pr\left[\prod_{i=1}^{|I|} \det \mathbf{M}\left(\{i\} \times J; S_i\right) = 0 \ \mid \ \mathbf{T}' = 0\right] \\
\leq \ & \sum_{i=1}^{|I|} \Pr\left[\det \mathbf{M}\left(\{i\} \times J; S_i\right) = 0 \ \mid \ \mathbf{T}' = 0\right] \ . \qquad (2.18)
\end{aligned}$$

If $\mathbf{T}' \neq 0$, then $\mathbf{T}$ is uniformly distributed in $\mathbb{F}^*$, and thus

$$\Pr\left[\mathbf{T} + \prod_{i=1}^{|I|} \det \mathbf{M}\left(\{i\} \times J; S_i\right) = 0 \;\mid\; \mathbf{T}' \neq 0\right] \;\leq\; \frac{1}{|\mathbb{F}^*|}\;, \tag{2.19}$$

because the value of $\prod_{i=1}^{|I|} \det(\mathbf{M}\left(\{i\} \times J; S_i\right))$ is already determined when $\mathbf{r}_1$ is the only coefficient left to be chosen. Combining (2.18) and (2.19), we have

$$\begin{aligned}
& \Pr\left[\mathbf{M}(I \times J; K) = 0\right] \\
= \;& \Pr\left[\mathbf{M}(I \times J; K) = 0 \;\mid\; \mathbf{T}' = 0\right] \Pr[\mathbf{T}' = 0] + \Pr\left[\mathbf{M}(I \times J; K) = 0 \;\mid\; \mathbf{T}' \neq 0\right] \Pr[\mathbf{T}' \neq 0] \\
\leq \;& \sum_{i=1}^{|I|} \Pr\left[\det \mathbf{M}\left(\{i\} \times J; S_i\right) = 0 \;\mid\; \mathbf{T}' = 0\right] \Pr[\mathbf{T}' = 0] + \frac{1}{|\mathbb{F}^*|} \\
\leq \;& \sum_{i=1}^{|I|} \Pr\left[\det \mathbf{M}\left(\{i\} \times J; S_i\right) = 0\right] + \frac{1}{|\mathbb{F}^*|}\;.
\end{aligned}$$

□

Putting the above two claims together, we have

$$\Pr\left[\mathbf{M}\left(I \times J; K\right) \text{ has full rank}\right] \;\geq\; 1 - \frac{d|K| + 1}{|\mathbb{F}^*|}\;.$$

Applying a union bound over all such I, J, K, where there are at most 2^{m+n+s} such triples, we conclude that the random linear encoding schemes $\mathbf{F}^{(1)}, \mathbf{F}^{(2)}, \ldots, \mathbf{F}^{(s)}$ are lossless decodable with probability

$$\geq 1 - \frac{dn + 1}{|\mathbb{F}^*|} \cdot 2^{m+n+s} \;=\; 1 - o_{|\mathbb{F}|;G,s}(1)\;.$$

□

CHAPTER 3

CONSERVATIVE CIRCUITS AND ROUTING NETWORKS

3.1 Introduction

One important goal of circuit complexity is to understand complexity of *joint computation*, in cases where we have multiple computational tasks to perform simultaneously; is it possible to combine computations to make them more efficient? For example, given an explicit operator $F : \{0,1\}^n \to \{0,1\}^m$, what is the *wire complexity* of computing F when *arbitrary gates* are allowed? Note that any operator with n inputs and m outputs can be computed by circuits using mn wires when arbitrary gates are allowed.

A good candidate example to study is the Shift operator. It is a simplified variant of cyclic convolution, and sort of a special case of matrix multiplication, while it still inherits some interesting properties. Our original motivation is to understand the wire complexity of the Shift operator in the *unrestricted* model.

One approach to prove lower bounds is based on *information flow arguments*. Much work in circuit complexity is guided by the heuristic idea that information is a "substance", that needs to be routed through a circuit from input gates to appropriate output gates [34, 65, 77]. The notion is that each gate and wire in the circuit can only carry one unit of this substance, and so *congestion* problems that occur in sparse networks might explain the computational complexity of certain natural problems.

There are at least restricted models of computation, notably the *conservative circuits* model [68], where this heuristic intuition is made into a firm *requirement* on the circuit's behavior. For conservative circuits, lower bounds may be proved (see [68, 69]) that are much stronger than those known for general circuits (with unrestricted gates). In the conservative model, it is known that for constant depth d, the wire complexity of the Shift operator is $\widetilde{O}\left(n^{1+\frac{1}{d}}\right)$.

Conservative circuits are equivalent to *routing networks* in a certain sense [68, 69]. Com-

pared with well-known network coding instances, each node is only allowed to route instead of encoding. By contrast, our goal is to *design* a communication network, minimizing complexity parameters like depth and number of wires. Very little is known about the actual comparative strengths of the routing-based paradigm and of general network-coding solutions.

We want to know what kind of property of the Shift operator captures its wire complexity, and hope to extend known arguments to a more general model, called *semi-conservative circuits*, which allows arbitrary preprocessing and a final layer of postprocessing. This motivates the definition of "Expansive Routing Families" (ERFs). It turns out there exist small-size ERF networks of depth 2 and 3, and we determine the asymptotically optimal sizes. For depth 2, the size is $\Theta\left(n(\log n/\log\log n)^2\right)$; for depth 3, the size is $\Theta(n\log\log n)$; for higher depth $d \geq 4$, we prove lower bound $\Omega_d\left(\lambda_d(n)\cdot n\right)$. However, we do not know whether these bounds are tight or not.

The existence of small-size ERF networks eliminates the possibility to prove strong lower bounds from *this* connectivity requirement. On the other hand, the definition of the ERF networks and the constructions may be interesting on its own right. In general, we propose the research challenge to develop a powerful and broadly-applicable set of techniques for both upper bounding and lower bounding the wire complexity of routing networks for given specific demands. This challenge has essentially been studied before in various theoretical contexts, but is little studied in its full generality. For example, both upper bounds and lower bounds are proved for routing networks realizing shifts or all permutations [68, 69]; Riis proposed a *guessing game* approach to study such problems, which applies to unrestricted computation, not necessarily routings [79].

3.1.1 Organization

In Section 3.2.1, we describe one formulation of the *conservative circuits* model [68]. This model aims to make precise the idea of an algorithm that handles certain pieces of information

as "atomic", manipulating them without inspecting their values or combining them with other atomic information sources. We also describe some relaxations of the conservative circuits model, on which we have made partial progress. The basic notion we are exploring is that of circuits which are allowed to separately "preprocess" certain parts of the input, but which must then handle this information in an essentially "conservative" way. (We call these circuits *semi-conservative*.) We believe such circuits should to some extent inherit the provable limitations of ordinary conservative circuits.

In Section 3.2.2, we formally define *routing networks*, which are naturally related to conservative circuits. Roughly spaeking, a routing network with a specific routing scheme is a circuit where all gates are projection gates. We will see the relationship between the entropy function computed by the conservative circuit and the information set of the routing schemes. In Section 3.2.3, we formulate the definition of Expansive Routing Family (ERF) networks, based on some entropy property satisfied by the Shift operator.

In Section 3.3.1, we prove a lower bound on the size of depth-2 ERF networks, which relies on the edge lower bounds of disperser graphs [78]. In Section 3.3.2, we prove an $\Omega_d(\lambda_d(n) \cdot n)$ lower bound on the size of depth-d ERF networks for all constant $d \geq 3$, where the proof uses a powerful lemma of Raz and Shpilka [77]. In Section 3.3.3, we show how these results imply the same circuit lower bounds for computing the Shift operator under the constant-depth semi-conservative circuits model.

In Section 3.4, we turn to the study of upper bounds, that is, the construction of ERF networks. All the constructions are *probabilistic* (for both graphs and routing schemes), and are inspired by superconcentrators [6, 21, 78]. Compared with superconcentrators, they seem more delicate, and their sizes are the same up to a constant factor for depth 3; for depth 2, it is even smaller than superconcentrators by a factor of $O(\log \log n)$. Specifically, we obtain depth-2 upper bound $O(n(\log n/\log \log n)^2)$ in Section 3.4.1; we have depth-3 upper bound $O(n \log \log n)$ in Section 3.4.4.

In Section 3.5, we propose the challenge to determine the number of edges required in

bounded-depth networks implementing certain families of "routings" between source and sink nodes. This problem is a fairly general graph-theoretic problem, and it is the combinatorial essence of circuit complexity in the conservative-circuit model. Despite its natural significance, however, the problem seems to have received little attention in recent years. Towards this challenge, in Section 3.5.1, we give a lower bound based on the *entropy concept*. The lower-bound criterion we give is a significant generalization of work of [68, 69] on so-called *shifter* and *connector* graphs. In Section 3.5.2, using probabilistic arguments, we construct depth-d routing networks of size $O\left(dn^{1+\frac{1}{d}}\right)$ realizing all n shifts, for fixed d. The size is asymptomatically optimal, and the method is related to previous works [68, 64], but simpler.

In Section 3.5.3, we list some related open problems may be worth investigating.

3.2 Preliminaries

3.2.1 Conservative Circuits and Relaxations

The idea of a conservative circuit is most naturally motivated by specific computational problems, for example the *Boolean (cyclic) Shift* operator. For input parameter n, this operator takes as input a string $x \in \{0,1\}^n$ and a number $i \in \mathbb{Z}_n$ (described by $\log_2 n$ bits). The output is a string $y \in \{0,1\}^n$. We consider x, y to have coordinates indexed by $\mathbb{Z}_n$, and the operator is defined by

$$\text{Shift}(x,i)_j = y_j := x_{j+i \mod n} \,.$$

This operator appears very simple, yet after decades of study it is unknown whether Shift can be computed by (logarithmic-depth) linear-sized Boolean circuits.

Now one possible hunch about this problem is that inspecting the values $x_1, \ldots, x_n$ is not really "useful" for the circuit; all that matters is that they get "routed" through the circuit

to their destination output coordinate, and the choice of routing ought not to depend on the values $x_1, \ldots, x_n$ themselves, but only on the shift-amount i.

To formalize this requirement, let us say that a circuit $\mathsf{C}(x, i)$ is *conservative with respect to* x if, after fixing any setting to i, all remaining gates in the circuit become *projection gates*: that is, they simply take on the value of some particular input variable. Formally, a projection gate g is just a "dictator function" $g(y_1, \ldots, y_k) = y_a$, for some $a \in [k]$, where this value a is determined in some way by the shift amount i given to C. It is worth noting that all operators computable by conservative circuits have a special structure, that is, after fixing any setting to i, all outputs are projections of the inputs.

Next we propose a relaxation of the conservative-circuit model. Our model makes sense for computing operators $F : \{0,1\}^n \to \{0,1\}^m$ with multiple output bits. (For single-output circuits the model is trivially all-powerful.)

Say that an (unbounded-fanin, Boolean-valued-gates) circuit

$$\mathsf{C}(x^{(1)}, \ldots, x^{(k)}, i) \ ,$$

is *semi-conservative* with respect to the designated input blocks $x^{(1)}, \ldots, x^{(k)}$, if C has the form

$$\mathsf{C} = \mathsf{C}'(\mathsf{C}_1(x^{(1)}), \ldots, \mathsf{C}_k(x^{(k)}), i) \ ,$$

for some subcircuits $\mathsf{C}', \mathsf{C}_1, \ldots, \mathsf{C}_k$ (where the understanding is that C' takes as inputs only the output gates of $\mathsf{C}_1, \ldots, \mathsf{C}_k$ and i); and, where we have the following property:

- for any fixed setting to i, all intermediate gates in C' become projection gates. (We allow the output gates of C' to compute arbitrary Boolean functions.)

The natural complexity measure for such circuits is their number of wires. For an operator $F : \{0,1\}^n \to \{0,1\}^{\Theta(n)}$ we can always compute F by a semi-conservative circuit of $O(n^2)$ wires, and for randomly chosen operators this is optimal (this follows from [49]).

We also define the model of *semi-conservative circuits with preprocessing*, which is the same as above, except we do not "charge" for the wires in the subcircuits $\mathsf{C}_1, \ldots, \mathsf{C}_k$, but only for those in C'.

A research question is to extend known lower bounds for conservative circuits to these relaxations, that is, semi-conservative model or semi-conservative models with preprocessing. For the Shift operator, in Section 3.3.3, we manage to prove wire lower bounds $\Omega_d\left(n \cdot \lambda_d(n)\right)$ for depth $d \geq 3$; $\Omega\left(n\left(\frac{\log n}{\log\log n}\right)^2\right)$ for depth $d = 2$, under the semi-conservative model with preprocessing. The lower bound follow from some required connectivity properties which we formalize in the "Expansive Routing Family" (ERF), that is, Definition 55.

Unfortunately, the ERF property cannot yield lower bounds comparable to $\Omega\left(n^{1+\frac{1}{d}}\right)$. This is because for depth 2 and 3, there exist ERF networks of much smaller size. (We determine the asymptotic size requirements for these depths.)

On the other hand, in the course of this work, we have become convinced that there are interesting and powerful techniques to lower bound and upper bound the size of information-routing networks. Such techniques (which ought to apply to many natural patterns of routing requests) could be of interest to designers of communication networks and to approximation-algorithms readers as well as to complexity theorists.

3.2.2 Networks and Routings

The definition of conservative circuits is naturally related to routing networks. That is, after fixing the setting to i (recall that the circuit inputs have two disjoint parts, x and i), the circuit is equivalent to a network with a *specific* routing scheme. The formal definition of routing networks will be given shortly. Roughly speaking, a network with a routing scheme is a circuit where all gates are *projection* gates. This framework has been studied in the past. For example, in [68, 69], both lower bounds and upper bounds (on the number of edges) are proved for routing networks realizing shifts or all permutations; the book "Switching Networks" contains some papers studying such routing problems with an eye to some specific

connectivity patterns [25].

A *network*, for our purposes, is just a directed acyclic graph $G = (V, E)$, with vertex set

$$V(G) = (U, V, W).$$

Here, U and V are disjoint sets of so-called *input* and *output* vertices, respectively. The vertices U are always source vertices in G, while V are sink vertices. The remaining nodes, W, are the *intermediate* vertices (W may contain sources and sinks). Important parameters of G are $n := |U|, m := |V|$, and the *depth* $d(G)$, defined as the length of the longest directed path in G. We will freely call G an (n, m)-*network* or n-*network* when $n = m$.

Let $e \to v$ denote the condition that edge e is incoming on vertex v, and similarly let $v \to e$ denote that e is outgoing from v.

We will be interested in "routings" in G, where each input vertex $x_i \in X$ has a unique "commodity" of its own, call it commodity i. Formally, a *routing* in G is a function

$$\mathfrak{R} : E(G) \to [n] \cup \{\perp\} \ .$$

We think of $\mathfrak{R}(e)$ as the (unique) commodity routed along edge e; if $\mathfrak{R}(e) = \perp$ then e is unused. We require a routing to obey some basic constraints:

1. (Input vertices provide their own commodities) For any input vertex $u_i \in U$,

$$\{\mathfrak{R}(e) : u_i \to e\} \subseteq \{i, \perp\} \ ; \tag{3.1}$$

2. (Intermediate vertices have unit capacity) For any vertex $w \in W$,

$$|\{\mathfrak{R}(e) : e \to w \text{ and } \mathfrak{R}(e) \neq \perp\}| \ \leq \ 1 \ ; \tag{3.2}$$

3. (Free copying, but no free synthesis or garbage disposal) For any $w \in W$,

$$\{\mathfrak{R}(e) \cap [n]\}_{w \to e} \;=\; \{\mathfrak{R}(e') \cap [n]\}_{e' \to w} \; . \tag{3.3}$$

While there is some resemblance between our notion of a routing and the multicommodity flow problems most typically studied in network design problems (particularly those with integrality constraints), there are also important differences, notably the fact that we allow free "copying" of our commodities.

Before drawing the connection between conservative circuits and routing networks, we need some definitions. Following [50], for an operator $F : \{0,1\}^{n+n'} \to \{0,1\}^m$, define the *entropy of* F as

$$\mathrm{Ent}(F) := \log_2 \left(|\, \mathrm{Range}(F)| \right) \; . \tag{3.4}$$

(This is a combinatorial measure, held distinct from the Shannon entropy.) Now suppose that F has two designated disjoint input blocks: $F = F(x,i)$, where $(x,i) \in \{0,1\}^n \times \{0,1\}^{n'}$. For $I \subseteq \{0,1\}^{n'}, J \subseteq [m]$, define the operator $F_{I,J} : \{0,1\}^n \to \{0,1\}^{|I| \cdot |J|}$ by

$$F_{I,J}(x) := \big(F_j(x,i)\big)_{i \in I, j \in J} \, . \tag{3.5}$$

For an operator $F(x,y) : \{0,1\}^{n+n'} \to \{0,1\}^m$, the quantities $\mathrm{Ent}(F_{I,J})$ are a fairly natural measure of "information flow" from the input, into the output gates indexed by J, in any augmented circuit computing F. For semi-conservative circuits, however, we have greater control on *how* information travels through the circuit, and this opens the possibility that we might exploit the quantities $\mathrm{Ent}(F_{I,J})$ in more effectively for lower-bound purposes.

Now let us draw a simple connection between semi-conservative circuits and multirequests. Suppose $\mathsf{C}(x,i,z)$ is a semi-conservative circuit for F, augmented with respect to x

and auxiliary advice operator F_{adv}, so that

$$\mathsf{C}(x, i, F_{\text{adv}}(x)) \equiv F(x, i) \ .$$

Let $z = (z_1, \ldots, z_p)$. Let G be the directed acyclic graph associated with C.

For each setting to i, every gate g in C (other than the input and output gates) becomes a projection gate. Let $e_i(g) \in E(G)$ denote the incoming edge (i.e., wire) to g carrying the input that g projects under the setting i. Say that an edge e ending at g is a *projecting edge for* i if $e = e_i(g)$. We define a *routing* $\mathfrak{R}_i$ in G, by letting $\mathfrak{R}_i(e) = w$, if e is a projecting edge for i and carries the input variable $w \in \{x_1, \ldots, x_n, z_1, \ldots, z_p\}$. (We are abusing notation slightly.) Otherwise, set $\mathfrak{R}_i(e) = \perp$. (Formally, a projecting edge e for i *carries* w if there is a path of projecting edges for i, beginning at w and ending with e itself. In this case e will, in fact, transmit the value of w on input (x, i, z) to the circuit C.)

$\mathfrak{R}_i$ automatically obeys properties 1 and 2 of our definition of routings, but it might not obey the "no free garbage disposal" rule. To fix this, say that a projecting edge e for i is *live* if there is a path of projecting edges for i, beginning at an input in $\{x_1, \ldots, x_n, z_1, \ldots, z_p\}$ and ending at an output gate, and that contains e.

Define $\mathfrak{R}_i^*(e) := \mathfrak{R}_i(e)$ if e is live, otherwise $\mathfrak{R}_i^*(e) := \perp$. It is now easy to verify that $\mathfrak{R}^*$ is a valid routing according to our definition. $\mathfrak{R}_i^*$ exactly fulfills some $(n+p, n)$-request; call this $\mathfrak{r}_i$. Let $\mathfrak{R}$ be the multirequest containing all of the requests $\mathfrak{r}_i$.

On an input (x, i, z), it follows readily from the definitions that the value of any output gate v_j with $j \in J$ is determined by the values of inputs $w \in \{x_1, \ldots, x_n, z_1, \ldots, z_p\}$ carried by wires (edges) incident to v_j that are live and projecting edges with respect to i; that is, by the values of the set of inputs

$$\left\{w \in \{x_1, \ldots, x_n, z_1, \ldots, z_p\} : \mathfrak{R}_i^*(e) = w\right\}_{e \to v_j} \ = \ \mathfrak{r}_i^{-1}(j) \ . \tag{3.6}$$

In particular, this holds when $z = F_{\text{adv}}(x)$.

Now fix a subset $J \subseteq [m]$ of outputs, and let $I \subseteq \{0,1\}^{n'}$ be a collection of settings to i. It follows from our work above that $F_{I,J}(x)$ can be determined by the values of

$$\left\{w \in \{x_1, \ldots, x_n, z_1, \ldots, z_p\} : \mathfrak{R}_i^*(e) = w\right\}_{i \in I, j \in J, e \to v_j} \;=\; \bigcup_{i \in I, j \in J} \mathfrak{r}_i^{-1}(j) \;, \tag{3.7}$$

a set we denote $\mathfrak{R}_I^{-1}(J) \subseteq [n+p]$. Thus, we have:

Proposition 54. *Suppose $\mathsf{C}(x,i,z)$ is as augmented as above. Then, for all $I \subseteq \{0,1\}^{n'}$ and $J \subseteq [m]$,*

$$\mathrm{Ent}(F_{I,J}) \;\leq\; \left|\mathfrak{R}_I^{-1}(J)\right| \;. \tag{3.8}$$

For conservative circuits without preprocessing, we have strict equality in (3.8), that is, $\mathrm{Ent}(F_{I,J}) = \left|\mathfrak{R}_I^{-1}(J)\right|$.

3.2.3 Expansive Routing Families

Now if F is an operator for which we have good *lower bounds* on the quantities $\mathrm{Ent}(F_{I,J})$, then this tells us that the collection of routings $\{\mathfrak{R}_i^*(e)\}$ implemented by the network G has a good "expansion" property. The hope is that such property entails a lower bound on the number of edges for G, if G is of bounded depth.

What are the "upper limits" of this approach? Trivially, we always have $\mathrm{Ent}(F_{I,J}) \leq |I| \cdot |J|$, since $F_{I,J}$ outputs only $|I| \cdot |J|$ bits. For Shift operator, we have

$$\mathrm{Ent}(F_{I,J}) \;=\; |I - J| \;,$$

where I, J are viewed as subsets of $\mathbb{Z}_n$, and $I - J := \{x - y : x \in I, y \in J\}$. For most sets I, J, we have $\mathrm{Ent}(F_{I,J}) = \Omega(\min(n, |I| \cdot |J|))$. This motivates the definition of the "Expansive Routing Family" (ERF).

Definition 55 (Expansive Routing Family (ERF))**.** *Given a network G with n sources and n*

sinks, and n routing schemes, it is called an Expansive Routing Family (ERF) *with constant $\delta > 0$ if*

$$\left|\mathfrak{R}_I^{-1}(J)\right| \;\geq\; \delta \min\left(|I| \cdot |J|, n\right)$$

for all $I, J \subseteq [n]$. Whenever $\delta \geq \frac{1}{2}$, we can simply say it is an Expansive Routing Family, and G is called an ERF network.

At this point, let us compare Expansive Routing Family (ERF) with the Strong Multiscale Entropy (SME) property, which is due to Jukna ([51], Chapter 13) generalizing a lower bound of Cherukhin [20]. Roughly speaking, if operator $F : \{0,1\}^n \to \{0,1\}^m$ satisfies the *Strong Multiscale Entropy (SME)* property, then for every $p \geq \sqrt{n}$ there exists an equal-size partition of $[n]$, denoted by $I_1, I_2, \ldots, I_p$, and an equal-size partition of $[m]$, denoted by $J_1, J_2, \ldots, J_{m/p}$, such that

$$\mathrm{Ent}(\widetilde{F}_{I_i,J_j}) \;\geq\; \Omega(n)$$

for all i and j, where

$$\widetilde{F}_{I,J} \;:=\; (F_{I,i,j})_{i \in I, j \in J}\ ,$$

where $F_{I,i,j}(x) := F_j(x[I : i])$, and $x[I : i]$ denotes the vector obtained from x by setting the ith bit to 1, and setting the i'th bit to 0 for all $i' \in I \setminus \{i\}$. It is worth noting that the definition of $\widetilde{F}_{I,J}$ is different from the definition of $F_{I,J}$ in (3.5), in which $I \subseteq \{0,1\}^{s_2}$ represents the "controlling" part.

The significance of the Strong Multiscale Entropy property is that it implies circuit lower bounds. For example, Jukna proved that $n \times n$ matrix multiplication over $\mathbb{F}_2$ requires at least n^3 wires for depth-2 circuits ([51], Chapter 13); Cherukhin proved lower bounds of the form $\Omega_d\left(n \cdot \lambda_{d-1}(n)\right)$ on the number of wires needed to compute cyclic convolutions in depth $d \geq 2$ [20]. On the other hand, Drucker exhibited an explicit operator satisfying the SME property that is computable in depth d with $O\left(\lambda_{d-1}(n) \cdot n\right)$ wires, for $d = 2, 3$, and for even $d \geq 6$ [27].

Remark 56. *In the above definition, why do we not insist on $\delta = 1$? Because in this case $\Omega(n^2)$ edges are needed regardless of the depth. The proof is the following:* fix *any $I \subseteq [n]$ of size 2, and consider $J \subseteq [n]$ of size $\frac{n}{2}$. In order to be an ERF network with $\delta = 1$, we should have $\mathfrak{R}_I^{-1}(J) = [n]$. In words, for* every *source, it gets routed to* some *sink in J, under* some *routing scheme in I.*

Without loss of generality, assume G is layered, and denote the sources by U, sinks by V, and the layer next to V by W. We will show the number of edges between W and V is $\Omega(n^2)$. Let $\overline{D}$ be the average degree of $v \in V$. It suffices to show $\overline{D} = \Omega(n)$.

Consider a new bipartite graph

$$G' \;=\; G'_I(U, V, E)$$

as follows (depending on the choice of $I \subseteq [n]$): edge $(u, v) \in E$ if u gets routed to v under some routing scheme in I. There are $\overline{D}n$ edges between W and V in graph G, and each edge can carry at most one unit in one routing scheme. So, in $|I|$ schemes, the total number of units routed is at most $\overline{D}n|I|$, which implies $|E(G')| \leq \overline{D}n|I|$. In average, each $u \in U$ is routed to $\leq \overline{D}|I|$ sinks, that is, the average degree of $u \in U$ in graph G' is $\leq \overline{D}|I|$. Hence there exists some $u^* \in U$ with degree $\leq \overline{D}|I|$. By our definition of the ERF network with $\delta = 1$, any $J \subseteq V$ of size $\frac{n}{2}$ should cover U, i.e., $\Gamma(J) = U$ in graph G'. Therefore, vertex $u^* \in U$ is allowed to miss at most $|J| - 1$ vertices, which implies

$$\overline{D}|I| \geq \deg(u^*) \geq n - |J| + 1 = \frac{n}{2} + 1 \ ,$$

which implies $\overline{D} \geq \frac{n}{4}$.

Furthermore, to justify the definition of ERF, we will prove the choice of the constant $\delta = \frac{1}{2}$ does not matter. As the next lemma shows, we can "amplify" the constant arbitrarily close to 1, at the cost of a multiplicative constant factor in size (both vertices and edges),

without increasing the depth.

Lemma 57 (Amplification Lemma)**.** *Let G be a depth-d size-s ERF network with constant $\delta > 0$, that is, for all $I, J \subseteq [n]$,*

$$\left|\mathfrak{R}_I^{-1}(J)\right| \;\geq\; \delta \min(|I| \cdot |J|, n) \;.$$

For any constant $\epsilon > 0$, there exists an ERF network G' with constant $1-\epsilon$ of size $O_{\delta,\epsilon}(s)$ and depth d, where the constant in $O_{\delta,\epsilon}(s)$ only depends on δ and ϵ.

Proof. Assume $G = (U \cup W \cup V, E)$, where $|U| = |V| = n$, and U are sources, V sinks. Let $t = t(\delta, \epsilon)$ be a constant to be determined later, which only depends on δ and ϵ. Roughly speaking, the network G' is a disjoint union of t copies of G on the same U and V, where each copy is obtained by randomly permuting n sources. To be specific,

1. Assume U is indexed by $[n]$, and V is also indexed by $[n]$;
2. Make t copies of G (including the routing schemes), and denote by $G_1, G_2, \ldots, G_t$;
3. For each G_i, we randomly permute the labels of U_i, where $G_i = (U_i \cup W_i \cup V_i, E_i)$;
4. Finally, we take a disjoint union of G_i and identify U_i and V_i according to their labels;
5. The routing scheme of G' is the *union* of the routing scheme of G_i. (This is possible because outputs can receive all commodities from the incoming neighbors.)

It is clear that the depth does not change, and the size (= number of edges) is at most ts. We shall determine the value of t, and prove $\left|\mathfrak{R}'^{-1}_I(J)\right| \geq (1-\epsilon)\min(ij, n)$, for all $I, J \subseteq [n]$, where $\mathfrak{R}'$ denotes the routing scheme for G', $i := |I|$, and $j := |J|$.

Let $I \subseteq [n]$, and $J \subseteq V = [n]$. Since G is an ERF network with constant δ, we have $\left|\mathfrak{R}_I^{-1}(J)\right| \geq \delta \min(ij, n)$. Let $A := \mathfrak{R}_I^{-1}(J) \subseteq [n]$, and let π_i be the random permutation applied to G_i. Then,

$$\mathbf{B} \;:=\; \mathfrak{R}'^{-1}_I(J) \;=\; \bigcup_i \pi_i(A) \;.$$

By union bound, it suffices to prove that

$$\sum_{I,J} \Pr\left[|\mathbf{B}| < (1-\epsilon)\min(ij, n)\right] \;\to\; 0\ . \tag{3.9}$$

Case 1: $ij \le \frac{n}{2}$. Let $\ell := ij$, and thus $|A| \ge \ell\delta$. Since each π_i is a random permutation on $[n]$, $\pi_i(A)$ will be a uniformly random subset of $[n]$ of size $|A| \ge \delta\ell$, and hence $\mathbf{B}$ contains t uniformly random subsets of $[n]$, each of size $\delta\ell$. Then,

$$\begin{aligned}
&\log \Pr\left[|\mathbf{B}| < (1-\epsilon)\ell\right] \\
&\le \log\left[\binom{n}{(1-\epsilon)\ell}\left(\frac{(1-\epsilon)\ell}{n}\right)^{t\delta\ell}\right] && \text{Counting argument} \\
&\le \ell\log\frac{en}{\ell} - t\delta\ell\log\frac{n}{(1-\epsilon)\ell} && \binom{n}{(1-\epsilon)\ell} \le \binom{n}{\ell} \le (en/\ell)^{\ell} \\
&\le 2\ell\log\frac{n}{\ell} - t\delta\ell\log\frac{n}{\ell} \\
&= -(t\delta - 2)\ell\log\frac{n}{\ell}\ .
\end{aligned}$$

Case 2: $ij > \frac{n}{2}$. Again let $\ell := ij$.

$$\begin{aligned}
&\Pr[|\mathbf{B}| < (1-\epsilon)\min(\ell, n)] \\
&\le \binom{n}{(1-\epsilon)\ell}\left(\frac{(1-\epsilon)\min(\ell,n)}{n}\right)^{t\delta\ell} && \text{Counting argument} \\
&< 2^n\left(\frac{(1-\epsilon)\min(\ell,n)}{n}\right)^{t\delta\ell} && \binom{n}{(1-\epsilon)\ell} < 2^n \\
&< 2^n(1-\epsilon)^{t\delta n/2}. && n/2 < \ell \le n
\end{aligned}$$

Set $t = t(\delta, \epsilon)$ such that $t\delta - 2 \ge 8$ and $(1-\epsilon)^{\frac{t\delta}{2}} < 2^{-6}$. Split the summation of (3.9)

into two parts: $ij < \frac{n}{2}$ and $ij \geq \frac{n}{2}$. The second part goes to 0, because if $\ell \geq \frac{n}{2}$,

$$\Pr\left[|\mathbf{B}| < (1-\epsilon)\min(\ell, n)\right] \ < \ 2^{-5n},$$

and we are summing over I, J, and there are at most $2^n \cdot 2^n = 2^{2n}$ such I, J in total. For the first part,

$$\sum_{\substack{I,J \\ |I|\cdot|J|\leq n/2}} \Pr\left[|\mathbf{B}| < (1-\epsilon)ij\right] \ < \ \sum_{\substack{i,j \\ ij\leq n/2}} \binom{n}{i}\binom{n}{j} 2^{-8ij\log\frac{n}{ij}} \ . \tag{3.10}$$

Applying the following lemma (Lemma 58) with $\epsilon = 8$, we complete the proof. □

The following lemma will be used many times when applying union bounds for analyzing the routing networks.

Lemma 58. *For any constant $\epsilon > 0$,*

$$\sum_{i,j} \binom{n}{i}\binom{n}{j} 2^{-\epsilon ij\log\frac{n}{ij}} \ \to \ 0 \ ,$$

where we sum over all such positive integer i, j such that $ij \leq n2^{-8/\epsilon}$ and $j \geq \frac{8}{\epsilon}$.

Proof. First,

$$\sum_{i,j} \binom{n}{i}\binom{n}{j} 2^{-\epsilon ij\log\frac{n}{ij}} \ \leq \ 2\sum_{i\geq j} 2^{2i\log\frac{n}{i} - \epsilon ij\log\frac{n}{ij}} \ .$$

Split the sum into the following two cases.

Case 1. $j \leq i \leq \log n$.

$$\begin{aligned} & 2i \log \frac{n}{i} - \epsilon ij \log \frac{n}{ij} & \\ \leq\ & 2i \log n - \frac{\epsilon}{2} ij \log n & \frac{n}{ij} \geq \sqrt{n} \\ \leq\ & -2i \log n\ . & j > \frac{8}{\epsilon} \end{aligned}$$

Thus,

$$\sum_{j \leq i \leq \log n} \binom{n}{i} \binom{n}{j} 2^{-\epsilon ij \log \frac{n}{ij}} \ \leq\ \sum_{i,j} 2^{-2i \log n} \ \to\ 0\ .$$

Case 2. $\log n < j \leq i$.

$$\begin{aligned} 2i \log \frac{n}{i} - \epsilon ij \log \frac{n}{ij} \ &\leq\ \epsilon i \left(\frac{2}{\epsilon} \log \frac{n}{i} - j \log \frac{n}{ij} \right) & \\ &\leq\ \epsilon i \log \left[\left(\frac{ij}{n} \right)^{j - \frac{2}{\epsilon}} j^{\frac{2}{\epsilon}} \right] & \\ &\leq\ \epsilon i \log \left(2^{-(j - \frac{2}{\epsilon}) \frac{8}{\epsilon}} n^{\frac{2}{\epsilon}} \right) & \log \left(\frac{n}{ij} \right) \geq \frac{8}{\epsilon} \\ &\leq\ \epsilon i \log \left(2^{-\frac{4j}{\epsilon}} n^{\frac{2}{\epsilon}} \right) & \\ &\leq\ -2i \log n\ . & \end{aligned}$$

Thus,

$$\sum_{\log n < j \leq i} \binom{n}{i} \binom{n}{j} 2^{-\epsilon ij \log \frac{n}{ij}} \ \leq\ \sum_{i,j} 2^{-2i \log n} \to 0\ .$$

□

3.3 Lower Bounds

In this section, we will prove lower bounds on the size of ERF networks. The tightness of the lower bounds for depth 2 and 3 will be shown in the next section.

3.3.1 Depth 2

We will prove $\Omega(n(\log n/\log\log n)^2)$ lower bound on the number of edges for depth-2 ERF network. The proof relies on Theorem 1.5 in [78], which is a lower bound for disperser graphs. The idea is to show any depth-2 ERF network must contain $\Omega(\log n/\log\log n)$ disjoint copies of dispersers, and each disperser graph has $\Omega(n\log n/\log\log n)$ edges.

Definition 59 (disperser graphs)**.** *A bipartite graph $G = (V_1 = [N], V_2 = [M], E)$ is a (K,ϵ)-disperser graph, if for every $X \subseteq V_1$ of cardinality K, $|\Gamma(X)| > (1-\epsilon)M$. The size of G is $|E(G)|$.*

Theorem 60. *(lower bounds for disperser graphs, Theorem 1.5 in [78]). Let $G = (V_1 = [N], V_2 = [M], E)$ be a (K,ϵ)-disperser. Denote by $\overline{D}$ the average degree of a vertex in V_1. Assume that $K < N$ and $\lceil\overline{D}\rceil \le (1-\epsilon)M/2$ (i.e., G is non trivial). If $\frac{1}{M} \le \epsilon \le \frac{1}{2}$, then $\overline{D} = \Omega\left(\frac{1}{\epsilon}\cdot\log(N/K)\right)$; if $\epsilon > \frac{1}{2}$, then $\overline{D} = \Omega\left(\frac{1}{\log(1/(1-\epsilon))}\cdot\log(N/K)\right)$.*

Theorem 61. *Let $G = (U \cup W \cup V, E)$, where $|V| = n$, be a depth-2 ERF network. Then*

$$|E| \;\ge\; \Omega\left(n\left(\frac{\log n}{\log\log n}\right)^2\right) .$$

Proof. Let $E = E_1 \cup E_2$, where E_1 is the set of edges from U to W, and E_2 from W to V. Assume $|E_2| \le n\log^2 n/2$, otherwise there is nothing to prove. Let $V' \subseteq V$ be the set of vertices with degree at most $\log^2 n$. Since $|E_2| \le n\log^2 n/2$, we have $|V'| \ge n/2$.

Restricting G on $U \cup W \cup V'$, denoted by G', it is clear that G' is still an ERF network, where $|U| = n$ and $|V'| \ge n/2$. For each $i = 0, 1, \ldots, \log n/(12\log\log n)$, let

$$W_i \;:=\; \left\{w \in W : \deg^+(w) \ge \frac{n}{(\log n)^{6i+3}}\right\} ,$$

where $\deg^+(w)$ denotes the outdegree of w (= the number of edges from w to V) in graph

G'. Since $|E_2| \leq n \log^2 n/2$, we have

$$|W_i| \;\leq\; \frac{|E_2|}{n/(\log n)^{6i+3}} \;\leq\; \frac{1}{2} \cdot (\log n)^{6i+5} \;. \tag{3.11}$$

Claim 62. *Let $k = (\log n)^{6i}$. Graph G' restricted to vertices $V' \cup (W_i \setminus W_{i-1})$ is a (k, ϵ) disperser, where $1 - \epsilon = \Omega(1/(\log^5 n))$.*

Proof. Let $J \subseteq V'$ be any subset of size k. Note that each vertex in V' has degree at most $\log^2 n$, and thus $|\Gamma(J)| \leq k \log^2 n$. Since $\left|\mathfrak{R}_I^{-1}(J)\right| \geq \frac{1}{2} \cdot \min(|I| \cdot |J|, n)$, regardless of the routing scheme, the following condition should be always satisfied: [1]

$$\sum_{w \in \Gamma(J)} \min(|I|, \deg^-(w)) \;\geq\; \Omega(\min(|I| \cdot |J|, n)) \;, \tag{3.12}$$

where $\deg^-(w)$ is the indegree of w (= number of edges from U to w). Setting $|I| = \frac{n}{k}$, (3.12) becomes

$$\sum_{w \in \Gamma(J)} \min\left(\frac{n}{k}, \deg^-(w)\right) \;\geq\; \Omega(n) \;. \tag{3.13}$$

Split the sum in the right hand side of (3.13) into two parts: $w \in \Gamma(J) \setminus W_i$ and $w \in \Gamma(J) \cap W_i$. For the first part,

$$\begin{aligned}
\sum_{w \in \Gamma(J) \setminus W_i} \deg^-(w) \;&\leq \frac{n}{(\log n)^{6i+3}} \cdot |\Gamma(J)| \\
&\leq \frac{n}{(\log n)^{6i+3}} \cdot k \log^2 n && |\Gamma(J)| \leq k \log^2 n \\
&= \frac{n}{\log n} \;. && k = \log^{6i} n
\end{aligned}$$

Thus

$$\sum_{w \in \Gamma(J) \cap W_i} \min\left(\frac{n}{k}, \deg^-(w)\right) \;\geq\; \Omega(n) - \frac{n}{\log n} \;=\; \Omega(n) \;,$$

1. This is the only place we use the property of ERF for proving the lower bound for depth-2 networks.

which implies

$$|\Gamma(J) \cap W_i| \;\geq\; \frac{\Omega(n)}{n/k} \;=\; \Omega(k) \;.$$

Observing that $|W_{i-1}| \leq \frac{1}{2} \cdot (\log n)^{6i-1} = o(k)$ by (3.11), we have

$$|\Gamma(J) \cap (W_i \setminus W_{i-1})| = \Omega(k) - o(k) = \Omega(k) = \Omega((1-\epsilon) \cdot |W_i \setminus W_{i-1}|) \;,$$

which proves the claim. □

By Theorem 60, the number of edges incident on V' and $W_i \setminus W_{i-1}$ is at least

$$|V'| \cdot \Omega\left(\frac{1}{\log(.5 \log^5 n)} \log\left(\frac{n}{k}\right)\right) \;=\; \Omega\left(\frac{n \log n}{\log\log n}\right) \;.$$

Summing over $i = 0, 1, \ldots, \log n/(12 \cdot \log\log n)$ gives the desired conclusion. □

Remark 63. *In the proof, only a weaker condition is required: for each k, there* exists *I of size k, such that $|\mathfrak{R}_I^{-1}(J)| \geq \Omega(\min(|I| \cdot |J|, n))$ holds for all $J \subseteq [n]$.*

3.3.2 Depth ≥ 3

In this subsection, we will prove lower bounds on the size of ERF networks for constant depth ≥ 3. In fact, the lower bound holds for the unrestricted circuits model instead of conservative circuits. By the connection between conservative circuits and routing networks in Proposition 54, the lower bound for routing networks follows immediately.

Definition 64 (Definition 2.3 in [77])**.** *Let*

$$\begin{aligned} \lambda_1(n) &:= \lfloor \sqrt{n} \rfloor \;, \\ \lambda_2(n) &:= \lceil \log n \rceil \;, \\ \lambda_d(n) &:= \lambda^*_{d-2}(n) \;. \end{aligned}$$

Lemma 65 (Lemma 1.1 in [77]). *Let $d > 1$ be a constant, and let $G = ((U, V, W), E)$ be a directed, layered, [2] acyclic graph of depth d with at least n vertices in total. Here the vertex set of G has a designated subset U of source vertices and a designated subset V of sink vertices, along with the rest of the vertices (denoted W).*

Let $\varepsilon \in (0, 1/400)$ be a constant. If G has fewer than $\varepsilon n \cdot \lambda_d(n)$ edges, then there exist sets $U_{\text{bad}} \subseteq U, V_{\text{bad}} \subseteq V$, and $X_{bad} \subseteq U \cup V \cup W$, such that:

1. *$|U_{\text{bad}}|,\ |V_{\text{bad}}| \leq 5\varepsilon dn$;*
2. *Letting $k := |X_{\text{bad}}|$, we have $\sqrt{n} \leq k = o(n)$;*
3. *The number of directed paths in G that begin in $U \setminus U_{\text{bad}}$ and end in $V \setminus V_{\text{bad}}$, and that are disjoint from X_{bad}, is at most $\varepsilon n^2/k$.*

Our proof of Theorem 66 is an application of a powerful lemma of Raz and Shpilka [77] (building on work of Pudlák [65] and Dolev *et al.* [21]). In a *layered* directed acyclic graph of depth d, the vertices are partitioned into layers $0, 1, \ldots, d$, such that all edges go between layers ℓ and $\ell + 1$, for some $0 \leq \ell < d$.

Theorem 66. *Let $d \geq 2$ be a constant. If operator $F : \{0,1\}^{n+\log n} \to \{0,1\}^n$ satisfies*

$$\text{Ent}(F_{I,J}) \ \geq \ \Omega(\min(|I| \cdot |J|, n))$$

for all $I, J \subseteq [n]$, then any depth-d (unrestricted) circuit with preprocessing computing F has at least $\Omega_d(n \cdot \lambda_d(n))$ wires, where the constant in Ω_d is $\Omega\left(\frac{1}{d^2}\right)$ (only depending on d).

Proof. It is enough to prove a lower bound of the form $\Omega\left(\frac{1}{d} \cdot n \cdot \lambda_d(n)\right)$ for *layered* circuits, since a general depth-d circuit can easily be made layered, with at most a factor-d increase in the number of wires.

2. In the lemma's statement in [77], the authors apparently forgot to stipulate that G be layered. This was an isolated typo, however, and in the proof of the lemma and all its applications in [77], the layeredness requirement is obeyed.

Let $\delta \in (0,1)$ be a constant such that $\mathrm{Ent}(F_{I,J}) > \delta \cdot \min(|I| \cdot |J|, n)$. Fix

$$\varepsilon \ := \ \delta^2 \min\left(\frac{1}{400}, \frac{1}{150d}\right) \ .$$

Assume for contradiction's sake that G has fewer than $\varepsilon n \cdot \lambda_d(n)$ edges, where G is the underlying graph of the circuit C. Let $U_{\mathrm{bad}}, V_{\mathrm{bad}}, X_{\mathrm{bad}}$ be the vertex sets produced by Lemma 65 for G.

Let $V = \{v_1, \ldots, v_n\}$. For $J \subseteq [n]$, let $V_J := \{v_j \in V : j \in J\}$. Also, let

$$U_J^* \subseteq U$$

be defined as the set of vertices $u \in U$ for which there exists a directed path in G beginning at u, that avoids $U_{\mathrm{bad}} \cup V_{\mathrm{bad}} \cup W_{\mathrm{bad}}$ and that ends in V_J. (So, e.g., $U_J^* = \emptyset$ if $V_J \subseteq V_{\mathrm{bad}}$.) Note that any path in G from U to a vertex in V_J, must intersect the vertex set

$$Q \ := \ U_{\mathrm{bad}} \cup (V_{\mathrm{bad}} \cap V_J) \cup X_{\mathrm{bad}} \cup U_J^* \ .$$

Thus on an input $(x, i, F_{\mathrm{adv}}(x))$ to C, all outputs at vertices in V_J can be determined by specifying the values $g(x, i, F_{\mathrm{adv}}(x))$ of gates $g \in Q$ when C is given input (x, i, F_{adv}).

As $\mathsf{C}(x, i, F_{\mathrm{adv}}(x)) \equiv F(x, i)$, it follows that $F_{I,J}(x)$ can be determined from the values

$$(g(x, i, F_{\mathrm{adv}}(x)))_{i \in I, g \in U_{\mathrm{bad}} \cup U_J^* \cup (V_{\mathrm{bad}} \cap V_J) \cup X_{\mathrm{bad}}} \ .$$

Now, the value $g(x, i, F_{\mathrm{adv}}(x))$ is either *independent* of i, *or* can be determined from i itself, whenever $g \in U_{\mathrm{bad}} \cup U_J^*$. It follows that

$$\begin{aligned} \mathrm{Ent}\left(F_{I,J}\right) \ &\leq \ |U_{\mathrm{bad}}| + |U_J^*| + |I| \cdot (|V_{\mathrm{bad}} \cap V_J| + |X_{\mathrm{bad}}|) \\ &\leq \ 5\varepsilon dn + |U_J^*| + |I| \cdot (|V_{\mathrm{bad}} \cap V_J| + k) \ . \end{aligned} \tag{3.14}$$

Now we will choose $\mathbf{I}, \mathbf{J} \subseteq [n]$ *randomly*. For each $i \in [n]$, independently include i in $\mathbf{I}$ with probability $\frac{\delta}{100k}$. Also, independently include i in $\mathbf{J}$ with probability $\frac{10k}{\delta n}$. (Recall that $k = o(n)$; to prove our asymptotic result, we may assume n is large enough to satisfy $\frac{10k}{n} < 1$.) The sets $\mathbf{I}, \mathbf{J}$ need not be disjoint, and membership decisions for $\mathbf{I}$ are independent of those for $\mathbf{J}$.

First we lower-bound $\mathbb{E}[\min(|\mathbf{I}| \cdot |\mathbf{J}|, n)]$. By Chernoff bound, $|\mathbf{I}| \geq (1-\mu) \cdot \mathbb{E}[|\mathbf{I}|]$ with high probability for any constant $\mu > 0$. Similarly, $|\mathbf{J}| \geq (1-\mu) \cdot \mathbb{E}[|\mathbf{J}|]$ with high probability. Thus

$$|\mathbf{I}| \cdot |\mathbf{J}| \;\geq\; (1-\mu)^2 \cdot \mathbb{E}\left[|\mathbf{I}|\right] \cdot \mathbb{E}\left[|\mathbf{J}|\right] \;=\; (1-\mu)^2 \cdot \frac{n}{10}$$

with high probability. Let constant $\mu > 0$ be sufficiently small, say $\mu = .001$, we get $\mathbb{E}[\mathrm{Ent}(F_{\mathbf{I},\mathbf{J}})] > .09\delta n$.

Next, we will upper-bound the expected value of the quantity on the right-hand-side of Eq. (3.16). Of course, we have

$$\mathbb{E}[|\mathbf{I}|] \;=\; n \cdot \frac{\delta}{100k} \;=\; \frac{\delta n}{100k}\ ,$$

and

$$\mathbb{E}\left[|V_{\mathrm{bad}} \cap V_{\mathbf{J}}|\right] \;=\; |V_{\mathrm{bad}}| \cdot \frac{10k\delta}{n} \;\leq\; 50\varepsilon dk\ .$$

Furthermore, $\mathbf{I}$ and $V_{\mathrm{bad}} \cap V_{\mathbf{J}}$ are *independent* as random sets, so we have

$$\mathbb{E}\left[|\mathbf{I}| \cdot |V_{\mathrm{bad}} \cap V_{\mathbf{J}}|\right] \;\leq\; \frac{\delta n}{100k} \cdot (50\varepsilon dk) \;=\; \frac{1}{2} \cdot \varepsilon dn\ .$$

We turn to analyze the set $U^*_{\mathbf{J}}$. For $j \in [n]$, let P_j denote the number of directed paths from U to $v_j \in V$ that are disjoint from the vertex set $U_{\mathrm{bad}} \cup V_{\mathrm{bad}} \cup W_{\mathrm{bad}}$. By Lemma 65, we have

$$\sum_{j \in [n]} P_j \;\leq\; \frac{\varepsilon n^2}{k}\ .$$

Now we clearly have $|U^*_{\mathbf{J}}| \leq \sum_{j \in \mathbf{J}} P_j$ in any outcome. It follows that

$$\mathbb{E}\left[|U^*_{\mathbf{J}}|\right] \;\leq\; \mathbb{E}\left[\sum_{j \in \mathbf{J}} P_j\right] \;\leq\; \frac{10k}{\delta n} \cdot \frac{\varepsilon n^2}{k} \;=\; \frac{10\varepsilon n}{\delta} \;.$$

Combining our bounds, we find

$$\begin{aligned} .09\delta n < \mathbb{E}\left[\mathrm{Ent}(F_{\mathbf{I},\mathbf{J}})\right] &< 5\varepsilon dn + \frac{10\varepsilon n}{\delta} + \frac{\varepsilon dn}{2} + \frac{n\delta}{100k} \cdot k \\ &< .01\delta n + \frac{11\varepsilon dn}{\delta} < .09\delta n \;, \end{aligned}$$

a contradiction, where we used the inequalities $d \geq 2$ and $\varepsilon < \frac{\delta^2}{150d}$. Thus C must have at least $\varepsilon n \cdot \lambda_d(n)$ wires. This proves Theorem 66. □

The following theorem is immediate from the proof of the above theorem and Proposition 54. We exclude the case $d = 2$ because it is subsumed by (and weaker than) Theorem 61.

Theorem 67. *Let $d \geq 3$ be a constant. Let $G = (U \cup W \cup V, E)$, where $|U| = |V| = n$, be depth-d ERF network. Then*

$$|E| \;\geq\; \Omega_d\left(n \cdot \lambda_d(n)\right) \;,$$

where the constant in Ω_d is $\Omega\left(\frac{1}{d^2}\right)$.

Proof. The proof is essentially the same. The only difference is that all values of the outputs can be determined by the values of the gates in Q *and* the input $i \in \{0,1\}^{\log n}$. This will increase the right hand side of (3.16) by $\log n$, which will not affect desired the contradiction. □

3.3.3 Semi-conservative Circuit Lower Bounds for Shift

Proposition 54 and the depth-2 size lower bound for the ERF networks (Theorem 61) imply the same lower bound for depth-2 semi-conservative circuits (with preprocessing) computing

any operator $F : \{0,1\}^{n+\log n} \to \{0,1\}^n$ satisfying

$$\mathrm{Ent}(F_{I,J}) \;\geq\; \Omega\left(\min(|I| \cdot |J|, n)\right)$$

for all $I, J \subseteq [n]$. However, for the Shift operator, $\mathrm{Ent}(\mathrm{Shift}_{I,J}) \geq |I - J|$, which is weaker than the above inequality. Fortunately, for the depth-2 lower bound, only a weaker condition is required (see Remark 63), which is true for the Shift operator.

Lemma 68. *For any $k \in [n]$, there exists set $I \subseteq [n]$ of size $\frac{n}{k}$, such that for any $J \subseteq [n]$ of size k, $|I - J| \;\geq\; \Omega(n)$, where the constant in Ω is an absolute constant (say, .01), and $I - J$ is computed modulo n.*

Proof. Let $\mathbf{I} \subseteq [n]$ be a set of $\frac{n}{k}$ uniformly random elements of $[n]$. If we can prove, for *any* $J \subseteq [n]$ of size k, $\Pr_{\mathbf{I}}\left[|\mathbf{I} - J| < \frac{n}{100}\right] < 2^{-n}$, then the lemma follows by applying a union bound.

Let $\mathbf{I} := \{\mathbf{i}_1, \mathbf{i}_2, \ldots, \mathbf{i}_{n/k}\}$, where each i_t is an independent uniformly random element of $[n]$. For any fixed element $x \in [n]$,

$$\Pr_{\mathbf{I}}\left[x \notin I - J\right] \;=\; \Pr_{\mathbf{I}}\left[x \notin \mathbf{i}_t - J \text{ for all } t = 1, \ldots, \frac{n}{k}\right] \;.$$

Note that $\mathbf{i}_t$'s are independent, and thus events $x \notin \mathbf{i}_t - J \Leftrightarrow \mathbf{i}_t \notin x + J$ are independent. Therefore, for *any* $J \subseteq [n]$ of size k, and for *any* $x \in [n]$,

$$\begin{aligned}
\Pr_{\mathbf{I}}\left[x \notin \mathbf{I} - J\right] &= \prod_{t=1}^{\frac{n}{k}} \Pr_{\mathbf{i}_t}\left[x \notin \mathbf{i}_t - J\right] \\
&= \Pr_{\mathbf{i}}\left[\mathbf{i} \notin x + J\right]^{\frac{n}{k}} \\
&= \left(1 - \frac{k}{n}\right)^{\frac{n}{k}} \;\leq\; e^{-1} \;.
\end{aligned}$$

Event $|\mathbf{I} - J| < \frac{n}{100}$ implies that there exist at least $\frac{99}{100} \cdot n$ elements in $[n]$ uncovered by

$\mathbf{I} - J$. By union bound,

$$\begin{aligned}\Pr_{\mathbf{I}}\left[|\mathbf{I}-J| < \frac{n}{100}\right] &\leq \binom{n}{\frac{99}{100}\cdot n} \Pr_{\mathbf{I}}\left[x \notin \mathbf{I}-J\right]^{\frac{99}{100}\cdot n} \\ &\leq 2^{nH(\frac{1}{100})-\frac{99}{100}n\log e} < 2^{-n}\ .\end{aligned}$$

We have shown that there *exists* I of size $\leq \frac{n}{k}$ satisfying our condition. Any superset of such I of size $\frac{n}{k}$ would be the desired set I, and thus the lemma is proved. □

The following theorem is immediate from the above lemma, Proposition 54, and the proof of Theorem 61.

Theorem 69. *Any depth-2 semi-conservative circuit* [3] *for computing* $\mathrm{Shift}(x,i) : \{0,1\}^{n+\log n} \to \{0,1\}^n$*, augmented with respect to* x*, requires* $\Omega\left(n(\log n/\log\log n)^2\right)$ *wires.*

For general constant depths, Theorem 66 can be extended to the Shift operator, as we will show next. In contrast with semi-conservative circuits, the computation model here is that of unrestricted circuits with preprocessing, which is much more powerful. The proof is essentially the same with slight modifications.

For ordinary circuits in the arbitrary-gates model, the lower bound above was already known for $d > 2$, and followed from lower bounds for the size of depth-d superconcentrators [21, 65].

Let us recall the computation model. Consider a function $F(x,i)$ with two input blocks x, i, where x consists of n bits. A *circuit for* F*, augmented with respect to* x*,* is a circuit $\mathsf{C}(x,i,z)$ with new auxiliary input variables $z \in \{0,1\}^m$, such that there exists some "advice" operator $F_{\mathrm{adv}} : \{0,1\}^n \to \{0,1\}^m$ (not necessarily efficiently computable) for which we have the identity

$$\mathsf{C}(x,i,F_{\mathrm{adv}}(x)) \equiv F(x,i)\ .$$

3. We only charge depth for the conservative part.

Theorem 70. *For any constant $d > 1$, any depth-d circuit for* $\mathrm{Shift}(x,i) : \{0,1\}^{n+\log n} \to \{0,1\}^n$*, augmented with respect to x (and in the arbitrary-gates model), requires* $\Omega\left(\frac{1}{d^2} \cdot n \cdot \lambda_d(n)\right)$ *wires.*

Proof. The proof is essentially the same as Theorem 66. Let us sketch the difference.

For the Shift operator,

$$\mathrm{Ent}(\mathrm{Shift}_{I,J}) \;\geq\; |I - J| \;, \tag{3.15}$$

with set arithmetic taken mod n.

Fix $\varepsilon := \min\left(\frac{1}{400}, \frac{1}{150d}\right)$. Assume for the sake of contradiction that G has fewer than $\varepsilon n \cdot \lambda_d(n)$ edges, where G is the underlying graph of the circuit $\mathsf{C}(x, i, F_{\mathrm{adv}}(x))$ computing $\mathrm{Shift}(x,i)$. Let $U_{\mathrm{bad}}, V_{\mathrm{bad}}, X_{\mathrm{bad}}$ be the vertex sets produced by Lemma 65 for G. Sets V_J, U_J^* and Q are defined as before. Similarly, on an input $(x, i, F_{\mathrm{adv}}(x))$ to C, all outputs at vertices in V_J can be determined by specifying the values $g(x, i, F_{\mathrm{adv}}(x))$ of gates $g \in Q$ when C is given input (x, i, F_{adv}), and it follows that

$$\begin{aligned} \mathrm{Ent}\left(\mathrm{Shift}_{I,J}\right) &\leq |U_{\mathrm{bad}}| + |U_J^*| + |I| \cdot (|V_{\mathrm{bad}} \cap V_J| + |X_{\mathrm{bad}}|) \\ &\leq 5\varepsilon d n + |U_J^*| + |I| \cdot (|V_{\mathrm{bad}} \cap V_J| + k) \;. \end{aligned} \tag{3.16}$$

Now we will choose $\mathbf{I}, \mathbf{J} \subseteq [n]$ *randomly*. For each $i \in [n]$, independently include i in $\mathbf{I}$ with probability $\frac{1}{100k}$. Also, independently include i in $\mathbf{J}$ with probability $\frac{10k}{n}$.

First we lower-bound $\mathbb{E}[|\mathbf{I} - \mathbf{J}|]$, where the calculation is modulo n. For $\ell \in [n]$, there are n ways to write $\ell = i - j$: one valid choice of j for each choice of i. Each such representation satisfies $(i,j) \in \mathbf{I} \times \mathbf{J}$ with probability $\frac{1}{100k} \cdot \frac{10k}{n} = \frac{1}{10n}$, and these events are independent. Thus,

$$\Pr[\ell \in \mathbf{I} - \mathbf{J}] \;=\; 1 - \left(1 - \frac{1}{10n}\right)^n \;>\; 1 - e^{-.1} \;>\; .09 \;,$$

so that $\mathbb{E}[|\mathbf{I} - \mathbf{J}|] > .09n$. By Eq. (3.15), we get $\mathbb{E}[\mathrm{Ent}(\mathrm{Shift}_{\mathbf{I},\mathbf{J}})] > .09n$.

Next, we will upper-bound the expected value of the quantity on the right-hand-side of

Eq. (3.16). Of course, we have

$$\mathbb{E}[|\mathbf{I}|] \;=\; n \cdot \frac{1}{100k} \;=\; \frac{n}{100k},$$

and

$$\mathbb{E}[|V_{\text{bad}} \cap V_{\mathbf{J}}|] \;=\; |V_{\text{bad}}| \cdot \frac{10k}{n} \;\leq\; 50\varepsilon dk\ .$$

Furthermore, $\mathbf{I}$ and $V_{\text{bad}} \cap V_{\mathbf{J}}$ are *independent* as random sets, so we have

$$\mathbb{E}[|\mathbf{I}| \cdot |V_{\text{bad}} \cap V_{\mathbf{J}}|] \;\leq\; \frac{n}{100k} \cdot (50\varepsilon dk) \;=\; \frac{\varepsilon dn}{2}\ ,$$

and

$$\mathbb{E}[|U^*_{\mathbf{J}}|] \;\leq\; \mathbb{E}\left[\sum_{j \in \mathbf{J}} P_j\right] \;\leq\; \frac{10k}{n} \cdot \frac{\varepsilon n^2}{k} \;=\; 10\varepsilon n\ .$$

Combining our bounds, we find

$$\begin{aligned} .09n < \mathbb{E}\left[\text{Ent}(\text{Shift}_{\mathbf{I},\mathbf{J}})\right] &< 5\varepsilon dn + 10\varepsilon n + \frac{\varepsilon dn}{2} + \frac{n}{100k} \cdot k \\ &< .01n + 11\varepsilon dn < .09n\ , \end{aligned}$$

a contradiction, where we used the inequalities $d \geq 2$ and $\varepsilon < \frac{1}{150d}$. □

3.4 Upper Bounds

In this section, we will show the existences of small-size ERF networks for depth 2 and 3, where, in both cases, the size is optimal up to a constant factor. The constructions are inspired by superconcentrators [6, 21, 78]. Both the networks and the routing schemes are constructed using probabilistic methods.

Compared with superconcentrators, there are a lot of similarities in the structure of the network. In terms of the size, for depth 3, they are asymptotically the same; for depth 2,

superconcentrators are larger by an $O(\log \log n)$ factor.

3.4.1 Depth 2

In this subsection, we will construct depth-2 ERF networks of size $O(n(\log n/\log\log n)^2)$, which matches the lower bound up to a multiplicative constant factor.

The following definition is auxiliary.

Definition 71. *Given (n,n)-network G with n routing schemes, say it is a* (j_1, j_2)-partial Expansive Routing Family (ERF) network *if*

$$\left|\mathfrak{R}_I^{-1}(J)\right| \;\geq\; \frac{1}{2}\cdot \min(|I|\cdot|J|, n)$$

for all $I, J \subseteq [n]$ with $j_1 \leq |J| \leq j_2$. Ignoring (j_1, j_2), we can simply say G is a partial ERF network.

First, let us define random depth-2 routing network $\mathbf{H}_r$.

1. Let $\mathbf{H}_r = (U \cup W \cup V, E_1 \cup E_2)$, where $|U| = |V| = n$ and $|W| = \frac{n}{r}$. Bipartite graph $(W \cup V, E_2)$ is chosen as a disperser such that for any $J \subseteq V$ of size $\frac{n}{2r}$, $|\Gamma(J)| \geq \frac{n}{2r}$. It is known that such disperser graphs can be constructed using $O(n \log r)$ edges [63, 78];

2. Bipartite graph $(U \cup W, E_1)$ is constructed as follows: for each $v \in W$, connect v to Dr randomly chosen vertices in U (allowing repetition), where D is an absolute constant to be determined later;

3. The n routing schemes are defined randomly and independently: for each $v \in W$, it selects a random neighbor to route.

Lemma 72. *For any r, with probability $1 - o_n(1)$, $\mathbf{H}_r$ is an $\left(\frac{n}{2r}, \frac{n}{r}\right)$-partial ERF network.*

Proof. Fix $I, J \subseteq [n]$ such that $\frac{n}{2r} \leq j \leq \frac{n}{r}$ and $ij \leq 2^{-7}\cdot n$, where $i := |I|$ and $j := |J|$. Let us estimate $\Pr\left[\left|\mathfrak{R}_I^{-1}(J)\right| < \frac{ij}{8}\right]$. For each $w \in \Gamma(J) \subseteq W$, let $\mathbf{d}_w$ denote the number of

distinct neighbors being routed during I. Since for each routing scheme, w chooses a random neighbor to route independently, we have

$$\Pr\left[\mathbf{d}_w < \frac{i}{2}\right] \;\le\; \binom{i}{\frac{i}{2}} \left(\frac{\frac{1}{2}\cdot i}{Dr}\right)^{\frac{i}{2}} \;\le\; 2^{-\frac{i}{2}\log(\frac{2r}{i})} \;,$$

assuming constant D is large enough. Let $\mathcal{B}$ denote the event that there exist $\frac{j}{2}$ vertices among the first j vertices of $\Gamma(J)$ such that $\mathbf{d}_w < \frac{i}{2}$. We have

$$\Pr[\mathcal{B}] \;\le\; \binom{j}{\frac{j}{2}} \Pr\left[\mathbf{d}_w < \frac{i}{2}\right]^{\frac{j}{2}} \;\le\; 2^{\frac{ij}{8}\log\left(\frac{n}{ij}\right)}. \tag{3.17}$$

Conditioning on $\neg\mathcal{B}$, there exist $\frac{j}{2}$ vertices in $\Gamma(J)$ with $\mathbf{d}_w \ge \frac{i}{2}$, which implies $\mathfrak{R}_I^{-1}(J)$ contains at least $\frac{ij}{4}$ uniformly random elements of $[n]$. Therefore,

$$\Pr\left[\left|\mathfrak{R}_I^{-1}(J)\right| < \frac{ij}{8} \;\middle|\; \neg\mathcal{B}\right] \;\le\; \binom{\frac{1}{4}\cdot ij}{\frac{1}{8}\cdot ij} \left(\frac{\frac{1}{8}\cdot ij}{n}\right)^{\frac{1}{8}\cdot ij} \;\le\; 2^{-\frac{1}{8}\cdot ij\log\left(\frac{n}{ij}\right)} \;. \tag{3.18}$$

Putting (3.17) and (3.18) together, we have

$$\Pr\left[\left|\mathfrak{R}_I^{-1}(J)\right| < \frac{ij}{8}\right] \;\le\; 2^{-\frac{1}{16}\cdot ij\log\left(\frac{n}{ij}\right)} \;.$$

Finally, we take a union bound over all possible I and J and apply Lemma 58 with $\epsilon = \frac{1}{16}$. Note that the constant in the ERF property can always be amplified by Lemma 57. □

Let $\mathbf{G}_\gamma = (U \cup W \cup V, E)$ be the following *random* depth-2 routing network, where $\gamma \ge 0$ is *not* necessarily an integer:

1. Let $U = [n]$ be the sources, and let $V = [n]$ be the sinks;
2. Let $W = \left[\frac{n}{(\log n)^\gamma}\right]$ be the intermediate layer;
3. The graph is layered, that is, edges only go from U to W, or from W to V;

4. Bipartite graph (W, V, E_2) is defined as follows: for each $v \in V$, connect it to $O(\log n/\log\log n)$ *uniformly random* vertices in W, where the hidden constant is to be determined later. For convenience of analysis, we allow repetition. It is obvious that $|E_2| \leq O(n \log n/\log\log n)$;

5. Bipartite graph (U, W, E_1) is defined as follows: for each $w \in W$, connect it to $(\log n)^{\gamma+0.6}$ *uniformly random* vertices in U. It is obvious that $|E_1| \leq n(\log n)^{0.6}$;

6. The total number of edges $|E| = |E_1| + |E_2| \leq O(n \log n/\log\log n)$;

7. For the routing schemes of U, basically there is no choice; route each $u \in U$ to all its neighbors. The routing scheme for W is defined as follows: for each $w \in W$, choose among its neighbors *uniformly at random.*

The goal is to prove $\mathbf{G}_\gamma$ is an $\left(n/(\log n)^{\gamma+0.5}, n/(\log n)^{\gamma+0.4}\right)$-partial ERF network with high probability. By taking the union of $O(\log n/\log\log n)$ partial ERF networks and identifying the sources and sinks, we will obtain a $\left(1, \frac{n}{(\log n)^{0.4}}\right)$-partial ERF network, where the remaining range will be taken care of by $O(\log\log n)$ copies of $\mathbf{H}_r$.

Let us prove (U, W, E_2) is a disperser first.

Lemma 73. *With probability* $1 - o_n(1)$*, for all* $J \subseteq V$ *of size in* $\left[\frac{n}{(\log n)^{\gamma+0.5}}, \frac{n}{(\log n)^{\gamma+0.4}}\right]$*,*

$$|\mathbf{\Gamma}(J)| \;\geq\; |J| \;.$$

Proof. It suffices to prove, for any $J \subseteq V$ of size $j := n/(\log n)^{\gamma+0.5}$,

$$|\mathbf{\Gamma}(J)| \;\geq\; \frac{n}{(\log n)^{\gamma+0.4}} \;.$$

In fact, a disperser (in Definition 59) with $1 - \epsilon = 1/(\log n)^{0.4}$, and $K = n/(\log n)^{\gamma+0.5}$

would suffice. The following is a standard probabilistic argument. For fixed J of size j,

$$\Pr\left[|\mathbf{\Gamma}(J)| < \frac{n}{(\log n)^{\gamma+0.4}}\right] \le \binom{|W|}{n/(\log n)^{\gamma+0.4}} \left(\frac{n/(\log n)^{\gamma+0.4}}{|W|}\right)^{jC\log n/\log\log n} .$$

Therefore,

$$\begin{aligned} & \log \Pr\left[\exists J \text{ such that } |\mathbf{\Gamma}(J)| < \frac{n}{(\log n)^{\gamma+0.4}}\right] \\ \le\ & \log\left[\binom{n}{j}\binom{|W|}{n/(\log n)^{\gamma+0.4}}\right] - 0.4jC\log n \\ \le\ & (4-0.4C)j\log n \\ \to\ & -\infty\ , \end{aligned}$$

if constant C is large enough, say, $C \ge 20$. □

Now we are ready to prove $\mathbf{G}_\gamma$ is a partial ERF network. For depth-2 routing network, there is only one intermediate layer, so the analysis is straightforward.

Lemma 74. *With probability* $1 - o_n(1)$, $\mathbf{G}_\gamma$ *is an* $\left(\frac{n}{(\log n)^{\gamma+0.5}}, \frac{n}{(\log n)^{\gamma+0.4}}\right)$*-partial ERF network.*

Proof. Let $J \subseteq V$ has size $|J| \in \left[n/(\log n)^{\gamma+0.5}, n/(\log n)^{\gamma+0.4}\right]$, and let $I \subseteq [n]$ be any subset such that $ij \le 2^{-80} \cdot n$, where $i := |I|$ and $j := |J|$. By Lemma 57, let us bound $\Pr\left[|\mathfrak{R}_I^{-1}(J)| < \frac{ij}{100}\right]$.

For $w \in \Gamma(J)$, let $\mathbf{d}_w$ denote the number of "different" [4] neighbors being routed during in routing schemes I. Recall that in each routing scheme, w independently choose a uniformly random neighbor to route, and there are $D := C(\log n)^{\gamma+0.5} > i$ neighbors in total. We have

$$\Pr\left[\mathbf{d}_w < \frac{i}{2}\right] \ \le\ \binom{i}{\frac{i}{2}}\left(\frac{\frac{1}{2}\cdot i}{D}\right)^{\frac{i}{2}} \ \le\ 2^{-\frac{i}{2}\log\left(\frac{D}{i}\right)} \ \le\ 2^{-\frac{i}{2}\log\left(\frac{n}{ij}\right)} ,$$

4. The same neighbor with different *labels* are considered to be different. Recall that each $v \in V$ is connected to $C\log n/\log\log n$ uniformly random vertices in W, instead of a random subset of fixed size. For $v \in V$, we *label* its neighbors by $[C\log n/\log\log n]$.

because $\frac{n}{j} \le D$. Among $|\Gamma(J)| \ge |J|$ vertices, let $\mathbf{B}$ denote the number vertices in W such that $\mathbf{d}_w < \frac{i}{2}$. By similar calculation,

$$\Pr\left[\mathbf{B} > \frac{j}{2}\right] \;\le\; \binom{j}{\frac{j}{2}} \left(2^{-\frac{i}{2}\log\left(\frac{n}{ij}\right)}\right)^{\frac{j}{2}} \;\le\; 2^{-\frac{ij}{5}\log\left(\frac{n}{ij}\right)} \;. \tag{3.19}$$

Conditioning on the event $\mathbf{B} \le \frac{j}{2}$, set $\mathfrak{R}_I^{-1}(J)$ contains at least $\frac{ij}{4}$ *independent* and *uniformly random* elements in $[n]$. Therefore,

$$\begin{aligned} &\Pr\left[\left|\mathfrak{R}_I^{-1}(J)\right| < \frac{ij}{100} \;\mid\; \mathbf{B} \le \frac{j}{2}\right] \\ \le\; &\binom{\frac{1}{4}\cdot ij}{\frac{1}{100}\cdot ij} \left(\frac{\frac{1}{100}\cdot ij}{n}\right)^{\frac{ij}{4}} \le 2^{-\frac{ij}{5}\log\left(\frac{n}{ij}\right)} \;. \end{aligned} \tag{3.20}$$

Combining (3.19) and (3.20),

$$\begin{aligned} \Pr\left[\left|\mathfrak{R}_I^{-1}(J)\right| < \frac{ij}{100}\right] \;&\le\; \Pr\left[\left|\mathfrak{R}_I^{-1}(J)\right| < \frac{ij}{100} \;\mid\; \mathbf{B} \le \frac{j}{2}\right] + \Pr\left[\mathbf{B} > \frac{j}{2}\right] \\ &\le\; 2^{-\frac{ij}{5}\log\left(\frac{n}{ij}\right)} + 2^{-\frac{ij}{5}\log\left(\frac{n}{ij}\right)} \\ &\le\; 2^{-\frac{ij}{10}\log\left(\frac{n}{ij}\right)} \;. \end{aligned}$$

Finally, we apply a union bound over all such I, J using Lemma 58. □

The final construction for depth-2 ERF network is immediate from Lemma 72 and 74.

Theorem 75. *There exist depth-2 ERF networks of size*

$$O\left(n\left(\frac{\log n}{\log\log n}\right)^2\right) \;.$$

Proof. Take the *union* of G_γ for $\gamma = 0, 0.1, 0.2, \ldots, O(\log n/\log\log n)$, with the union of H_r for $r = 1, 2, 2^2, \ldots, O\left((\log n)^{0.4}\right)$, and *identify* their sources and sinks. The total size is $O(n\log n/\log\log n)$. The conclusion follows from Lemma 72 and 74. □

3.4.2 Composition Lemma

Inspired by the composition step for superconcentrators [21], we will prove a composition lemma for ERF networks, which will be used in the construction for depth-3 ERF networks, and it may have other applications.

The following lemma serves to prove a dichotomy in the composition lemma (Lemma 77) we will prove, for which the motivation will be clear shortly.

Lemma 76. *Let A be a $(0,1)$-matrix of size $m \times n$, and the total number of ones in A is i. Let j be a positive integer such that $ij \leq mn$. Let r_k denote the number of ones on the kth row, $k = 1, 2, \ldots, m$, and let c_k be the number of ones on the kth column, $k = 1, 2, \ldots, n$. Either*

$$\sum_{k=1}^{m} \min\left(r_k, \frac{n}{j}\right) \geq \frac{i}{2} \tag{3.21}$$

or

$$\sum_{k=1}^{n} \min\left(c_k, \frac{ij}{n}\right) \geq \frac{i}{2} \tag{3.22}$$

is true.

Proof. Assume (3.22) is not true, and we will prove (3.21). Combining our assumption $\sum_k \min(c_k, \frac{ij}{n}) < \frac{i}{2}$ and the fact $\sum_k c_k = i$, we have

$$\begin{aligned} \sum_{c_k \geq \frac{ij}{n}} c_k &= \sum_{k=1}^{n} c_k - \sum_{c_k < \frac{ij}{n}} c_k \\ &\geq i - \sum_{k=1}^{n} \min\left(c_k, \frac{ij}{n}\right) \\ &\geq \frac{i}{2}. \end{aligned} \tag{3.23}$$

Let $C := \left\{k : c_k \geq \frac{ij}{n}\right\}$. Since $\sum_k c_k = i$, we have $|C| \leq \frac{n}{j}$. By (3.23), we claim the $m \times |C|$ submatrix indexed by C contains at least 0.5 fraction of ones. Let r'_k denote the number of ones in submatrix indexed by columns C. By definition, it is obvious that $r'_k \leq |C| \leq \frac{n}{j}$ and

$r'_k \leq r_k$. Thus,

$$\sum_{k=1}^{m} \min\left(r_k, \frac{n}{j}\right) \;\geq\; \sum_{k=1}^{m} \min\left(r'_k, \frac{n}{j}\right) \;\geq\; \sum_{k=1}^{m} r'_k \;\geq\; \frac{i}{2}\;,$$

which proves (3.21). □

A *conservative routing scheme* is a routing scheme where each output can receive at most one commodity; in order words, like intermediate gates, outputs are also projection gates. A collection of routing schemes is called a *Conservative Expansive Routing Family* (CERF) if each routing scheme is conservative, and it is an Expansive Routing Family (ERF). Let us define (a, b)-*partial CERF* similarly.

Roughly speaking, the following composition lemma says that, we can obtain an ERF network by adding two layers to a CERF network with smaller inputs/outputs so that the size only increases by a multiplicative constant factor, and the depth is increased by 2. It is worth noting that the routing network in the middle must be conservative; removing that restriction will lead to the constructions of ERF networks of depth d and size $O_d\left(\lambda_d(n) \cdot n\right)$, which seems unlikely to be true.

Lemma 77 (Composition Lemma)**.** *Let M be an $\left(a, \frac{n}{4r^{1.1}}\right)$-partial CERF network with constant δ of depth d and $\frac{n}{2r}$ inputs/outputs. Let N be a depth-$(d+2)$ n-network obtained by*

- *adding one bottom layer to M, denoted by*

$$Q\left(U_Q = \left[\frac{n}{2r}\right], V_Q = [n], E_Q\right)\;,$$

which is a bipartite graph [5] with $O_\delta(n)$ edges such that $|\Gamma(J)| \geq \frac{256}{\delta} \cdot |J|$ for any $J \subseteq V_Q$ of size $|J| \leq \frac{n}{4r^{1.1}}$;

5. The existence of such disperser graph can be shown by a standard probabilistic argument. See [63] for a general result.

- *adding one top layer to M, denoted by*

$$P\left(U_P = [n], V_P = \left[\frac{n}{2r}\right], E_P\right) \ ,$$

  which is a random bipartite graph such that each vertex in V_P is connected to $2r$ random vertices in U_P;

- *duplicating the entire construction, and identifying the corresponding inputs and outputs.*

Then with probability $1 - o_n(1)$, there exist n routing schemes for N which is an $\left(a, \frac{n}{4r^{1.1}}\right)$-partial ERF with constant $2^{-\frac{256}{\delta}}$. [6]

Proof. We are duplicating the entire construction twice, because we will define two different (random) routing schemes separately. And we will prove, for each $I, J \subseteq [n]$, at least one routing scheme will succeed with high probability.

Let $\mathfrak{R}_1, \mathfrak{R}_2, \ldots, \mathfrak{R}_n$ denote the routing schemes for N. With slight abuse of notation, we use the same notation for both routing schemes. Let bijection $\psi : [n] \to [2r] \times \left[\frac{n}{2r}\right]$ be $\psi(i) := (\psi_1(i), \psi_2(i))$, where

$$\psi_1(i) := \left\lfloor \frac{i-1}{\frac{n}{2r}} \right\rfloor + 1 \text{ and } \psi_2(i) := i - \left\lfloor \frac{i-1}{\frac{n}{2r}} \right\rfloor \cdot \frac{n}{2r} \ .$$

Scheme I: Define routing schemes $\mathfrak{R}_1, \mathfrak{R}_2, \ldots, \mathfrak{R}_n$ for each part, i.e., M, P, Q, separately as follows.

- Let

$$M \ = \ \left(U_M = \left[\frac{n}{2r}\right] \cup V_M = \left[\frac{n}{2r}\right] \cup W_M, E\right)$$

  be the $\frac{n}{2r}$-routing network in the middle. By our condition, there are $\frac{n}{2r}$ routing

6. Note that by Lemma 57, we can amplify the constant $2^{-\frac{256}{\delta}}$ arbitrarily close to 1.

schemes associated with M, denoted by $\mathfrak{M}_1, \mathfrak{M}_2, \ldots, \mathfrak{M}_{\frac{n}{2r}}$. Let

$$\mathfrak{R}_k|_M \ := \ \mathfrak{M}_{\psi_2(k)} \ .$$

In words, we simply *duplicate* the routing schemes of M for $2r$ times.

- Let

$$P \ = \ \left(U_P = [n], V_P = \left[\frac{n}{2r}\right], E\right)$$

be the bipartite graph on the top. For each vertex $u \in U_P$, route the information carried by u to all its neighbors. Fix each $v \in V_P$, among its $2r$ neighbors, choose $2r$ vertices uniformly at random, denoted by $u_1, u_2, \ldots, u_{2r} \in U_M$. Let

$$\mathfrak{R}_k(e) \ := \ u_{\psi_1(k)}$$

for all $v \to e$. In other words, $[n]$ is decomposed into $2r$ intervals of equal length, and in the kth interval, v always routes the *same* neighbor, which is randomly chosen.

- Let $Q = \left(U_Q = \left[\frac{n}{2r}\right], V_Q = [n], E\right)$ be the disperser graph on the bottom. Basically there is no choice, that is, for each vertex $u \in U_Q$, route the information carried by u to all its neighbors. (Note that the routing schemes for M are conservative, that is, each sink only outputs one commodity.)

Scheme II: Define the routing schemes $\mathfrak{R}_1, \mathfrak{R}_2, \ldots, \mathfrak{R}_n$ for each part separately. For M and Q, they are exactly the same as Scheme I.

- For each vertex $u \in U_P$, route the information carried by u to all its neighbors. For each $v \in V_P$, each time we choose a uniformly random neighbor to route.

In Scheme I, $v \in U_P$ will route the same neighbor within each interval of length $\frac{n}{2r}$; in Scheme II, each time v chooses a random neighbor to route independently.

Let $I \subseteq [n]$ and $J \subseteq [n]$ such that $j \in \left[a, \frac{n}{4r^{1.1}}\right]$ and $ij \leq 2^{-\frac{256}{\delta}} \cdot n$, where $i := |I|$ and $j := |J|$. Let

$$I \;=\; I_1 \dot{\cup} I_2 \dot{\cup} \cdots \dot{\cup} I_{2r} \;, \tag{3.24}$$

where $I_k \subseteq \left[\frac{n}{2r} \cdot (k-1) + 1, \frac{n}{2r} \cdot k\right]$. Let A be the $2r \times \frac{n}{2r}$ matrix such that

$$A_{s,t} = \begin{cases} 1, & \text{if } t \in \psi_2(I_s), \\ 0, & \text{otherwise .} \end{cases}$$

Applying Lemma 76 with $m := 2r$, $n := \frac{n}{2r}$, $i := i$, and $j := j$, we have either

$$\sum_{k=1}^{2r} \min\left(r_k j, \frac{n}{2r}\right) \;\geq\; \frac{ij}{2} \tag{3.25}$$

or

$$\sum_{k=1}^{\frac{n}{2r}} \min\left(c_k, \frac{2rij}{n}\right) \;\geq\; \frac{i}{2} \;, \tag{3.26}$$

where $c_k := |\{s : k \in I_t\}|$ and $r_k := |I_k|$. If (3.25) is true, we apply Scheme I; otherwise apply Scheme II.

Case 1: (3.25) is true. We apply Scheme I. For those $I, J \subseteq [n]$ satisfying (3.25), we will prove $\mathfrak{R}$ is an $\left(a, \frac{n}{4r^{1.1}}\right)$-partial ERF with probability close to 1. Specifically, we will bound $\Pr\left[\left|\mathfrak{R}_I^{-1}(J)\right| < \frac{\delta}{16} \cdot ij\right]$, and take a union bound at the end.

Note that $|\Gamma(J)| \geq |J| = j$ by the definition of disperser Q. By the construction of our routing schemes, $\{\mathfrak{R}_i|_M\}_{i \in I_k}$ are exactly the routing schemes of M *indexed* by $\psi_2(I_k)$. Since $\mathfrak{M}$ is an $\left(a, \frac{n}{4r^2}\right)$-partial CERF with constant δ,

$$\left|\mathfrak{M}_{I_k}^{-1}(\Gamma(J))\right| \;\geq\; \delta \cdot \min\left(r_k j, \frac{n}{2r}\right) \;.$$

Observe that $\mathfrak{R}_I^{-1}(J) = \bigcup_k \mathfrak{R}_{I_k}^{-1}(J)$. By the construction of $\mathfrak{R}$ (the second bullet point in

the definition of Scheme I), set $\mathfrak{R}_{I_k}^{-1}(J)$ contains exactly $\left|\mathfrak{M}_{I_k}^{-1}\left(\Gamma(J)\right)\right|$ uniformly random elements of $[n]$. By (3.25), the union of sets $\mathfrak{R}_{I_k}^{-1}(J)$, $k \in [2r]$, contains at least

$$\sum_{k=1}^{2r} \delta \cdot \min\left(r_k j, \frac{n}{2r}\right) \;\geq\; \frac{\delta}{2} \cdot ij \;.$$

uniformly random elements of $[n]$. Therefore,

$$\begin{aligned}\Pr\left[\left|\mathfrak{R}_I^{-1}(J)\right| < \frac{\delta}{16} \cdot ij\right] \;&\leq\; \binom{\frac{\delta}{2} \cdot ij}{\frac{\delta}{16} \cdot ij} \left(\frac{\frac{\delta}{16} \cdot ij}{n}\right)^{(\frac{1}{2}-\frac{1}{16})\cdot\delta ij} \\ &\leq\; 2^{-\frac{\delta}{5}\cdot ij \log\left(\frac{n}{ij}\right)} \;.\end{aligned}$$

To finish the proof of the case, we apply a union bound over all such I, J, and use Lemma 58 by setting $\epsilon = \frac{5}{\delta}$.

Case 2: assume (3.25) is not true, and thus (3.26) is true by Lemma 76. We apply scheme II. Let $I = I_1 \dot\cup I_2 \dot\cup \ldots \dot\cup I_{2r}$, where $I_k \;\subseteq\; \left[\frac{n}{2r} \cdot (k-1) + 1, \frac{n}{2r} \cdot k\right]$.

Claim 78. *Given* (3.26)*, there exist disjoint subsets $L_1, L_2, \ldots, L_\ell$ of I such that*

- $\ell = \frac{2rij}{n}$*;*
- *for each L_k, the second coordinates of points in $\psi(L_k)$ are different, that is, $|\psi_2(L_k)| = |L_k|$;*
- $\left|\bigcup_{k=1}^{\ell} L_k\right| \geq \frac{i}{2}$*;*
- $|L_1| = |L_2| = \ldots = |L_\ell| \geq \frac{n}{4rj}$.

Proof. (of the Claim) Initially, let $L_1, L_2, \ldots, L_\ell$ be all empty.

Note that $c_k = |\{s \in [2r] : k \in \psi_2(I_s)\}|$. In other words, for each $s \in [2r]$, it appears in $\bigcup_{k=1}^{2r} \psi(I_k)$ as the second coordinate for exactly c_k times. Distribute $\min(c_k, \ell)$ of those into *distinct* sets among $L_1, L_2, \ldots, L_\ell$, and make their sizes as close as possible. In total, $\bigcup_{k=1}^{\ell} L_k$ contains $\sum_k \min(c_k, \ell) \geq \frac{i}{2}$ elements by (3.26). □

It is obvious that $\mathfrak{R}_I^{-1}(J)$ contains $\mathfrak{R}_{L_k}^{-1}(J)$ for all $k = 1, 2, \ldots, \ell$. Consider $\mathfrak{R}_{L_k}^{-1}(J)$, which contains all the information routed by

$$\mathfrak{M}_{\psi_2(L_k)}^{-1}(\Gamma(J)) \subseteq U_M = V_P .$$

Since $\mathfrak{M}$ is an $\left(a, \frac{n}{4r^2}\right)$-partial CERF with constant δ, we have

$$\left|\mathfrak{M}_{\psi_2(L_k)}^{-1}(\Gamma(J))\right| \geq \delta \cdot |\psi_2(L_k)| \cdot |\Gamma(J)| \geq \frac{\delta n}{4r} .$$

By the definition of routing schemes in P, during L_k, for each vertex in $\mathfrak{M}_{\psi_2(L_k)}^{-1}(\Gamma(J))$, *at least* one random neighbor is routed. For the purpose of bounding $\Pr\left[|\mathfrak{R}_I^{-1}(J)| < \frac{\delta}{16} \cdot ij\right]$, without loss of generality, assume

- $\left|\mathfrak{M}_{\psi_2(L_k)}^{-1}(\Gamma(J))\right| = \frac{\delta n}{4r}$;
- $\mathfrak{M}_{\psi_2(L_1)}^{-1}(\Gamma(J)) = \ldots = \mathfrak{M}_{\psi_2(L_\ell)}^{-1}(\Gamma(J))$.

For convenience, let $W := \mathfrak{M}_{\psi_2(L_1)}^{-1}(\Gamma(J))$. For each vertex $v \in W$, at least ℓ random neighbors (not necessarily distinct) are routed. Let $\mathbf{d}_v$ be the number distinct neighbors of v being routed during time I. We have

$$\Pr\left[\mathbf{d}_v < \frac{\ell}{4}\right] \leq \binom{\ell}{\frac{1}{4} \cdot \ell} \left(\frac{\frac{1}{4} \cdot \ell}{2r}\right)^{\frac{3}{4} \cdot \ell} \leq 2^{-\frac{\ell}{2} \log(\frac{n}{ij})} .$$

Let $\mathcal{B}$ denote the event that there exist $\frac{1}{2} \cdot |W|$ vertices with $\mathbf{d}_v < \frac{1}{4} \cdot \ell$. By a union bound, we have

$$\Pr[\mathcal{B}] \leq \binom{|W|}{\frac{1}{2} \cdot |W|} \Pr\left[\mathbf{d}_v < \frac{\ell}{4}\right]^{\frac{1}{2} \cdot |W|} \leq 2^{-\frac{\delta}{16} \cdot ij \log\left(\frac{n}{ij}\right)} . \tag{3.27}$$

Conditioning on $\neg\mathcal{B}$, there exist $\frac{1}{2} \cdot |W|$ vertices, where each contributes at least $\frac{1}{2} \cdot \ell$ random elements to $\mathfrak{R}_I^{-1}(J)$. In total, $\mathfrak{R}_I^{-1}(J)$ contains at least $\frac{1}{4} \cdot \ell |W| = \frac{\delta}{8} \cdot ij$ random elements of

$[n]$. Thus

$$\Pr\left[\left|\mathfrak{R}_I^{-1}(J)\right| < \frac{\delta}{16}\cdot ij \;\mid\; \neg\mathcal{B}\right] \;\leq\; \binom{\frac{\delta}{8}\cdot ij}{\frac{\delta}{16}\cdot ij}\left(\frac{\frac{\delta}{16}\cdot ij}{n}\right)^{\frac{\delta}{16}\cdot ij} \leq\; 2^{-\frac{\delta}{16}\cdot ij\log\left(\frac{n}{ij}\right)}. \tag{3.28}$$

Combining (3.28) and (3.27), we conclude

$$\Pr\left[\left|\mathfrak{R}_I^{-1}(J)\right| < \frac{\delta}{16}\cdot ij\right] \;\leq\; 2^{-\frac{\delta}{32}\cdot ij\log\left(\frac{n}{ij}\right)}.$$

Finally, we take a union and apply Lemma 58 by setting $\epsilon = \frac{\delta}{32}$, and the proof of this case is complete. □

In the construction of depth-3 ERF networks, we will apply the Composition Lemma to construct a depth-3 $\left(1, n^{\frac{9}{20}}\right)$-partial ERF network of size $O(n)$ as follows.

Proposition 79. *There exists depth-1 size-$O(n^2)$ routing network $G(U,V,E)$ with n inputs and n outputs which is a $(16,n)$-partial CERF network with constant $\frac{1}{3}$.*

Proof. The graph is a complete bipartite graph, and for each $v \in V$, it outputs a uniformly random neighbor. Thus, for fixed $I, J \subseteq [n]$ with $|I|\cdot|J| \leq n$, $\mathfrak{R}_I^{-1}(J)$ contains $|I|\cdot|J|$ uniformly random elements. It is easy to verify

$$\Pr\left[\left|\mathfrak{R}_I^{-1}(J)\right| \leq \frac{|I|\cdot|J|}{3}\right] \;\leq\; 2^{-\frac{1}{2}\cdot ij\log\left(\frac{n}{ij}\right)}.$$

The proof is complete by applying a union bound using Lemma 58. □

For the range $|J| \leq n^{\frac{9}{20}}$, we can apply the Composition Lemma with $r = n^{\frac{1}{2}}$ and the above proposition to obtain partial ERF network of size $O(n)$.

Corollary 80. *There exist depth-3 $\left(1, n^{\frac{9}{20}}\right)$-partial ERF networks of size $O(n)$.*

Proof. We apply Lemma 77 with $r = n^{\frac{1}{2}}$, where the middle part is a $\left(16, \frac{n}{2r}\right)$-partial CERF network of size $O(n)$; the existence of such routing network is proved in Proposition 79. Finally, the constant can always be amplified by Lemma 57. □

3.4.3 Negative Association of Random Variables

In this subsection, we review some known facts about negative association, which will be used in the construction of depth-3 ERF networks. Negative association is one version of negative dependence of random variables. Once some random variables are negatively associated, we can apply marginal probability bound or Chernoff bound as for independent variables. Interested readers may refer to [26, 48] for proofs and applications.

Definition 81 (Negative Association)**.** *Let* $\overline{\mathbf{X}} := (\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n)$ *be a vector of random variables.* $\overline{\mathbf{X}}$ *are* negatively associated *if for every two disjoint index sets,* $I, J \subseteq [n]$*,*

$$\mathbb{E}\left[f(\mathbf{X}_i, i \in I) \cdot g(\mathbf{X}_j, j \in J)\right] \;\leq\; \mathbb{E}\left[f(\mathbf{X}_i, i \in I)\right] \cdot \mathbb{E}\left[g(\mathbf{X}_j, j \in J)\right]$$

for all functions $f : \mathbb{R}^{|I|} \to \mathbb{R}$ *and* $g : \mathbb{R}^{|J|} \to \mathbb{R}$ *that are both non-decreasing or both non-increasing.*

Proposition 82 (Proposition 7 in [26])**.**

1. *If* $\overline{\mathbf{X}}$ *and* $\overline{\mathbf{Y}}$ *are negatively associated, and are mutually independent, then the augmented vector* $(\mathbf{X}, \mathbf{Y}) = (\mathbf{X}_1, \ldots, \mathbf{X}_n, \mathbf{Y}_1, \ldots, \mathbf{Y}_m)$ *satisfies negative association condition.*

2. *Let* $\overline{\mathbf{X}} = (\mathbf{X}_1, \ldots, \mathbf{X}_n)$ *be negatively associated. Let* $I_1, \ldots, I_k \subseteq [n]$ *be disjoint index sets, for some positive integer* k*. For* $j \in [k]$*, let* $h_j : \mathbb{R}^{|I_k|} \to \mathbb{R}$ *be functions that are all non-decreasing or all non-increasing, and define,* $\mathbf{Y}_j := h_j(\mathbf{X}_i, i \in I_j)$*. Then the vector* $\overline{\mathbf{Y}} := (\mathbf{Y}_1, \ldots, \mathbf{Y}_k)$ *also satisfies the negative association condition. That is, non-decreasing (or non-increasing) functions of disjoint subsets of negatively associated variables are also negatively associated.*

Lemma 83 (Zero-One Lemma)**.** *If* $\mathbf{X}_1, \ldots, \mathbf{X}_n$ *are zero-one random variables such that* $\sum_i \mathbf{X}_i = 1$*, then* $\mathbf{X}_1, \ldots, \mathbf{X}_n$ *are negatively associated.*

Proposition 84 (Marginal Probability Bounds)**.** *Let* $\mathbf{X}_1, \ldots, \mathbf{X}_n$ *be distributed to satisfy the negative association condition. Then*

$$\Pr[\mathbf{X}_1 \leq t_1 \wedge \ldots \wedge \mathbf{X}_n \leq t_n] \;\leq\; \prod_{i \in [n]} \Pr[\mathbf{X}_i \leq t_i] \;.$$

Theorem 85 (Chernoff Bounds)**.** *Let* $\mathbf{X} = \sum_{i=1}^{n} \mathbf{X}_i$*, where* $\mathbf{X}_i = 1$ *with probability* p_i *and* $\mathbf{X}_i = 0$ *with probability* $1 - p_i$*, and all* $\mathbf{X}_i$ *are independent. Let* $\mu = \mathbb{E}[\mathbf{X}] = \sum_{i=1}^{n} p_i$*. Then*

1. *(Upper Tail)* $\Pr[\mathbf{X} \geq (1+\delta)\mu] \leq e^{-\frac{\delta^2}{2+\delta} \cdot \mu}$ *for all* $\delta > 0$*;*
2. *(Lower Tail)* $\Pr[\mathbf{X} \leq (1-\delta)\mu] \leq e^{-\frac{\delta^2}{2} \cdot \mu}$ *for all* $0 < \delta < 1$*.*

Proposition 86 (Chernoff Bound)**.** *The Chernoff-Hoeffding bounds are applicable to sums of variables that satisfy the negative association condition.*

Definition 87 (Permutation Distribution)**.** *A permutation distribution is generated by taking all permutations of a given vector.*

Proposition 88. *Permutation distribution satisfies negative association condition.*

3.4.4 Depth 3

The construction of depth-3 ERF network is inspired by superconcentrators [6]. However, the analysis is quite different.

For depth 3, we will construct routing network of size $O(n \log\log n)$ by taking the union of $O(\log\log n)$ partial ERF networks, where each partial ERF network is of size $O(n)$. For the partial ERF network except the one in Corollary 80, the overall construction is simple and straightforward. Denote the input layer by $V_0 = [n]$, the next two layers by $V_1 = \left[\frac{n}{r^{2/3}}\right]$, $V_2 = \left[\frac{n}{r^{2/3}}\right]$, and the output layer by $V_3 = [n]$. The graph is constructed randomly as follows:

each vertex in V_3 is connected to $O(1)$ uniformly random chosen vertices in V_2; each vertex in V_2 is connected to $\min\left(|V_1|, r^{\frac{2}{3}}\right)$ random vertices in V_1; each vertex in V_1 is connected to $r^{\frac{2}{3}}$ random vertices in V_0. The routing schemes are also defined randomly in a natural way: for any vertex $u \in V_1 \cup V_2$, always pick a uniformly random neighbor to route. To make the proof module, we will present the construction step by step.

Before starting the proof, we need the following "balls into bins" lemma.

Lemma 89. *Let $q \geq 37, p$ be positive integers, where q is sufficiently large. Let $\mathbf{B}_1, \mathbf{B}_2, \ldots, \mathbf{B}_p \subseteq [pq]$ be a partition of $[pq]$ such that, independently for each $i \in [pq]$,*

$$\Pr\left[i \in \mathbf{B}_k\right] \;=\; \frac{1}{p}\;,$$

for $k = 1, 2, \ldots, p$. The probability that there exist $\frac{1}{2} \cdot p$ sets $\mathbf{B}_k$, each with size at least $\frac{1}{2} \cdot q$, is at least

$$1 - 2^{-\frac{1}{16} \cdot pq}\;.$$

Proof. Let $\mathbf{S}_i := |\mathbf{B}_i|$. By definition, it is clear that $\mathbf{S}_i = \mathbf{X}_1 + \mathbf{X}_2 + \ldots + \mathbf{X}_{pq}$, where $\mathbf{X}_k \in \{0, 1\}$ is the indicator variable that $k \in \mathbf{B}_i$. Since $\Pr[\mathbf{X}_k = 1] = \frac{1}{p}$, and $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_{pq}$ are independent, we have $\mathbb{E}[\mathbf{B}_i] = q$. By Chernoff bound, $\Pr\left[\mathbf{S}_i < \frac{q}{2}\right] \leq e^{-\frac{q}{8}}$.

Since random variables $\mathbf{S}_1, \mathbf{S}_2, \ldots, \mathbf{S}_p$ are negatively associated (Theorem 13 in [26]), then by Proposition 84, for any set $I \subseteq [p]$,

$$\Pr_{i \in I}\left[\mathbf{S}_i \leq \frac{q}{2}\right] \;\leq\; \prod_{i \in I} \Pr\left[\mathbf{S}_i \leq \frac{q}{2}\right]\;.$$

Applying a union bound, among p random variables $\mathbf{S}_1, \mathbf{S}_2, \ldots, \mathbf{S}_p$, the probability that there exist $\frac{1}{2} \cdot p$ variables with value less than $\frac{1}{2} \cdot q$, is bounded by

$$\binom{p}{\frac{1}{2} \cdot p} \left(e^{-\frac{q}{8}}\right)^{\frac{p}{2}} \;<\; 2^{p - \frac{1}{16} \cdot pq \log e} \;<\; 2^{-\frac{1}{16} \cdot pq}\;,$$

assuming $q \geq 37$. $\square$

In words, the above lemma says, throwing pq balls into p bins, with probability $\geq 1 - 2^{pq/16}$, there exist $\frac{1}{2} \cdot p$ bins, each containing at least $\frac{1}{2} \cdot q$ balls.

Lemma 90. *Given $r \leq n^{\frac{3}{4}}$, where r is sufficiently large (larger than some absolute constant), there exists bipartite graph $G(V_1, V_2, E)$ such that*

- $|V_1| = |V_2| = \frac{n}{r^{2/3}}$;
- $\deg(v) = D := r^{2/3}$ *for all* $v \in V_2$;
- *for any* $J \subseteq V_2$ *with* $|J| = \frac{n}{r^{1+\delta}} \in \left[\frac{n}{r^{1+\frac{1}{8}}}, \frac{n}{r}\right]$, *there exist* $\frac{1}{2} \cdot |V_1|$ *vertices in* V_1 *such that each has*

$$\frac{1}{2} \cdot \frac{D \cdot |J|}{|V_1|}$$

neighbors in J .

Proof. We connect each $v \in V_2$ to D uniformly random elements. For convenience of analysis, we allow repetition; in case $\deg(v) < D$, we simply add some extra edges to make the degree exactly D, which will not violate the desired property.

It is equivalent to throwing $D|J|$ balls into $|V_1|$ bins, and we shall estimate the probability that there exists $\frac{1}{2} \cdot |V_1|$ bins with at least $\frac{1}{2} \cdot \frac{D \cdot |J|}{|V_1|}$ balls. Applying Lemma 89, we claim, for fixed J, the third bullet point holds with probability at least $1 - 2^{\frac{1}{16} \cdot D|J|}$. Finally, we apply a union bound, and the proof is completed by showing

$$\sum_{\frac{n}{r^{9/8}} \leq j \leq \frac{n}{r}} \binom{|V_1|}{j} 2^{-\frac{1}{16} \cdot jD} \ \to \ 0 \ .$$

$\square$

Lemma 91. *Fix n, r, where $r \leq n^{\frac{3}{4}}$, and r is sufficiently large (larger than an absolute constant). There exists bipartite graph $G = (V_1, V_2, E)$ such that*

- $|V_1| = |V_2| = \frac{n}{r^{2/3}}$;
- $\deg(v) = r^{\frac{2}{3}}$ *for all* $v \in V_2$;
- *for each edge* $e \in E \subseteq V_1 \times V_2$, *there is a set* $R_e \subseteq [n]$ *associated, and* $(R_e)_{e \to v}$ *is a partition of* $[n]$ *for each fixed* $v \in V_2$, *that is,* $\dot{\bigcup}_{e \to v} R_e = [n]$;
- *fixing* $J \subseteq V_2$ *of size* $\frac{n}{r^{1+\delta}}$, *where* $\delta \in \left[0, \frac{1}{8}\right]$, *and fixing* $I \subseteq [n]$ *of size* $\leq \frac{n}{|J|}$, *when* $|I| \geq 32 \cdot r^{\frac{1}{3}+\delta}$, *there exist* $\frac{1}{8} \cdot |V_1|$ *good vertices in* V_1; *when* $|I| < 32 \cdot r^{\frac{1}{3}+\delta}$, *there exist* $\frac{1}{256} \cdot |I| \cdot |J|$ *good vertices in* V_1, *where vertex* $u \in V_1$ *is called good (with respect to* I, J*) if*

$$\left|I_u^{(J)}\right| \;\geq\; \frac{1}{32} \cdot \frac{|I| \cdot |J|}{|V_1|} \;,$$

where

$$I_u^{(J)} \;:=\; \left(\bigcup_{e \in u \times J} R_e\right) \cap I \;.$$

Proof. The construction is given below:

1. The bipartite graph between V_1 and V_2 is given by Lemma 90;
2. For each $v \in V_2$, let us define the sets associated with its $r^{\frac{2}{3}}$ incident edges. Independently for each $p \in [n]$,

$$\Pr\left[p \in \mathbf{R}_e\right] \;=\; r^{-\frac{2}{3}} \;.$$

   In other words, each element of $[n]$ is uniformly random distributed in $(\mathbf{R}_e)_{e \to v}$, for fixed $v \in V_2$.

We will prove our construction will succeed with nonzero probability.

Case 1: $|I| \geq r^{\frac{1}{4}}$**.** Fix $J \subseteq V_2$ of size $\frac{n}{r^{1+\delta}}$, where $\delta \in \left[0, \frac{1}{8}\right]$, and fix $I \subseteq [n]$ of size $|I| \in \left[r^{\frac{1}{4}}, r^{1+\delta}\right]$. Let $i := |I|$ and $j := |J|$. For each $u \in V_1$, let $\mathbf{X}_u$ be the indicator variable that u is good, i.e.,

$$\left|\mathbf{I}_u^{(J)}\right| \;\geq\; \frac{1}{32} \cdot \frac{|I| \cdot |J|}{|V_1|} \;.$$

Claim 92. $(\mathbf{X}_u)_{u\in V_1}$ *is negatively associated.*

Proof. (of the Claim) For each $v \in V_2$, let $\mathbf{Z}^u_{p,q}$ be the indicator variable that $p \in \mathbf{R}_e$, where e is the qth incoming edge of v. By Zero-One Lemma, each row of $\mathbf{Z}^u$, i.e., $\{\mathbf{Z}^u_{p,q}\}_q$, is negatively associated. Since different rows are independent, by Proposition 82 (1), the matrix $\mathbf{Z}^u$ is negatively associated. Since for different u, $\mathbf{Z}^u$ are independent, again by Proposition 82 (1), $(\mathbf{Z}^u)_{u\in V_1}$ is negatively associated.

Observe that, when the graph between V_1 and V_2 is fixed, $\mathbf{X}_u$ is a non-decreasing function on some entries of $(\mathbf{Z}^u)_{u\in V_1}$, and the set of entries are disjoint for different u. The Claim follows from Proposition 82 (2). □

Assume u has ℓ neighbors in J. By our construction and Lemma 90, there exist .5 fraction of vertices in V_1 such that $\ell \geq \frac{1}{2}\cdot r^{\frac{1}{3}-\delta}$. Let us calculate the mean of $\mathbf{X}_u$. For each $p \in I$,

$$\Pr\left[p \in \mathbf{I}_u^{(J)}\right] = 1 - \Pr\left[p \notin \mathbf{I}_u^{(J)}\right] = 1 - \left(1 - r^{-\frac{2}{3}}\right)^{\ell} .$$

Thus,

$$\mathbb{E}\left[\left|\mathbf{I}_u^{(J)}\right|\right] = i \cdot \left(1 - \left(1 - r^{-\frac{2}{3}}\right)^{\ell}\right) \geq i \cdot \left(1 - e^{-r^{-\frac{2}{3}}\ell}\right) \geq \frac{i}{4}\cdot r^{-\frac{1}{3}-\delta} , \tag{3.29}$$

assuming $\ell \geq \frac{1}{2}\cdot r^{\frac{1}{3}-\delta}$. By Chernoff bound,

$$\mathbb{E}[\mathbf{X}_u] = 1 - \Pr\left[\left|\mathbf{I}_u^{(J)}\right| < \frac{i}{32}\cdot r^{-\frac{1}{3}-\delta}\right] \geq 1 - e^{-\frac{i}{12}\cdot r^{-\frac{1}{3}-\delta}} . \tag{3.30}$$

Note that $\frac{|I|\cdot|J|}{|V_1|} = i \cdot r^{-\frac{1}{3}-\delta}$.

Subcase 1.1: $r^{\frac{1}{4}} \leq i < 32\cdot r^{\frac{1}{3}+\delta}$. We have $\mathbb{E}[\mathbf{X}_u] \geq \frac{i}{64}\cdot r^{-\frac{1}{3}-\delta}$ by (3.30). Let $\mathbf{X} := \sum_{u\in V_1}\mathbf{X}_u$, and thus

$$\mathbb{E}[\mathbf{X}] \;\geq\; \frac{i}{64}\cdot r^{-\frac{1}{3}-\delta}\cdot\frac{1}{2}\cdot\frac{n}{r^{\frac{2}{3}}} \;=\; \frac{ij}{128} .$$

Applying Chernoff bound again, we have $\Pr\left[\mathbf{X} < \frac{ij}{256}\right] \le e^{-\frac{ij}{1024}}$. Applying a union bound over all such i, j, where $r^{\frac{1}{4}} \le i \le 32 \cdot r^{\frac{1}{3}+\delta}$ and $\frac{n}{r^{9/8}} \le j \le \frac{n}{r}$, we have

$$\begin{aligned}\sum_{I,J} \Pr\left[\mathbf{X} < \frac{ij}{256}\right] &\le \sum_{i,j} \binom{n}{i}\binom{n}{j} 2^{-\frac{ij}{1024}} \\ &\le \sum_{i,j} 2^{i\log\left(\frac{en}{i}\right)+j\log\left(\frac{en}{j}\right)-\frac{ij}{1024}} ,\end{aligned}$$

which tends to 0 since $j \gg \log n$. (Note that $\log\left(\frac{en}{j}\right) \le \log\left(er^{9/8}\right) < \frac{r^{\frac{1}{4}}}{2000}$ when r is sufficiently large.)

Subcase 1.2: $i > 32 \cdot r^{\frac{1}{3}+\delta}$. By Lemma 90, there exist $\frac{1}{2} \cdot |V_1|$ vertices in V_1, denoted by V_1', such that each has at least $\frac{1}{2} \cdot r^{\frac{1}{3}-\delta}$ neighbors in J. For any $u \in V_1'$, by (3.29) and Chernoff bound, we have

$$\Pr\left[\left|\mathbf{I}_u^{(J)}\right| < \frac{i}{32} \cdot r^{-\frac{1}{3}-\delta}\right] \le 2^{-\frac{i}{12}\cdot r^{-\frac{1}{3}-\delta}} .$$

The probability that there exist $\frac{3}{8} \cdot |V_1|$ "bad" (= not good) vertices in V_1' is bounded by

$$\binom{\frac{1}{2}\cdot|V_1|}{\frac{1}{8}\cdot|V_1|} \Pr\left[\left|\mathbf{I}_u^{(J)}\right| < \frac{i}{32} \cdot r^{-\frac{1}{3}-\delta}\right]^{\frac{3}{8}\cdot|V_1|} \le 2^{-\frac{1}{256}\cdot ij} .$$

Applying a union bound over all such I, J, we have

$$\sum_{I,J} \Pr\left[\mathbf{X} < \frac{|V_1|}{8}\right] \le \sum_{i,j} \binom{n}{i}\binom{n}{j} 2^{-\frac{ij}{256}} \le \sum_{i,j} 2^{i\log\left(\frac{en}{i}\right)+j\log\left(\frac{en}{j}\right)-\frac{ij}{256}} ,$$

which tends to 0 because $j \gg \log n$.

Case 2: $i < r^{\frac{1}{4}}$**.** Fix $J \subseteq V_2$ of size $\frac{n}{r^{1+\delta}}$, $\delta \in \left[0, \frac{1}{8}\right]$, and fix $I \subseteq [n]$ of size less than $r^{\frac{1}{4}}$ (and thus less than $32 \cdot r^{\frac{1}{3}}$).

For each $v \in J$, let $\mathbf{b}_v$ denote the number of incident edges e such that $\mathbf{R}_e \cap I \neq \emptyset$, i.e.,

$\mathbf{b}_v := |\{e \to v : \mathbf{R}_e \cap I \neq \emptyset\}|$. Thinking of this process as throwing $|I| = i$ balls into $r^{2/3}$ bins, by union bound, we have

$$\Pr\left[\mathbf{b}_v < \frac{i}{4}\right] \;<\; \binom{i}{\frac{1}{4}\cdot i}\left(\frac{\frac{1}{4}\cdot i}{r^{\frac{2}{3}}}\right)^{\frac{3}{4}\cdot i} \;<\; 2^{-\frac{i}{3}\log\left(\frac{r^{\frac{2}{3}}}{i}\right)}.$$

Let $\mathbf{J}' := \left\{v \in J : \mathbf{b}_v \geq \frac{i}{4}\right\}$. Note that $\mathbf{b}_v$ is independent for each v. By union bound, we have

$$\Pr\left[\left|\mathbf{J}'\right| < \frac{|J|}{2}\right] \;<\; \binom{|J|}{\frac{1}{2}\cdot|J|}\Pr\left[\mathbf{b}_v < \frac{i}{2}\right]^{\frac{|J|}{2}} \;<\; 2^{-\frac{1}{7}ij\log\left(\frac{r^{\frac{2}{3}}}{i}\right)}.$$

Let $\mathcal{A}$ denote the event that there exist at least $\frac{1}{2}\cdot|J|$ vertices in J such that $\mathbf{b}_v \geq \frac{i}{4}$. Rewriting the above inequality,

$$\Pr[\neg\mathcal{A}] \;\leq\; 2^{-\frac{1}{7}\cdot ij\log\left(\frac{r^{\frac{2}{3}}}{i}\right)} \;\leq\; 2^{-\frac{1}{21}\cdot ij\log\left(\frac{n}{ij}\right)}. \tag{3.31}$$

Conditioning on event $\mathcal{A}$, there exist $\frac{1}{8}\cdot ij$ edges such that $\mathbf{R}_e \cap I \neq \emptyset$, where each is connected to V_1 uniformly random, and they are negatively associated. [7] Therefore,

$$\begin{aligned}\Pr\left[\left|\left\{\mathbf{I}_u^{(J)} \neq \emptyset : u \in V_1\right\}\right| < \frac{ij}{16} \;\middle|\; \mathcal{A}\right] \;&\leq\; \binom{\frac{1}{8}\cdot ij}{\frac{1}{16}\cdot ij}\left(\frac{\frac{1}{16}\cdot ij}{|V_1|}\right)^{\frac{ij}{16}} \\ &\leq\; 2^{-\frac{1}{16}\cdot ij\log\left(\frac{|V_1|}{ij}\right)} \\ &\leq\; 2^{-\frac{1}{200}\cdot ij\log\left(\frac{n}{ij}\right)},\end{aligned} \tag{3.32}$$

assuming $|I| < r^{\frac{1}{4}}$.

Combining (3.31) and (3.32), we have

$$\Pr\left[\left|\left\{\mathbf{I}_u^{(J)} \neq \emptyset : u \in V_1\right\}\right| < \frac{ij}{8}\right] \;\leq\; 2^{-\frac{1}{300}\cdot ij\log\left(\frac{n}{ij}\right)}.$$

7. For different $v \in V_2$, $\mathbf{R}_e$'s are clearly independent. For the same v, indicator variables $|\mathbf{R}_e \cap I| > 0$ are negatively associated as we have shown before.

Applying a union bound over all possible I, J and Lemma 58, we complete the proof of this case. □

Now we are ready to construct $\left(\frac{n}{r^{1+1/8}}, \frac{n}{r}\right)$-partial ERF networks with $O(n)$ edges.

Lemma 93. *For any $r < n^{\frac{3}{4}}$, where r is sufficiently large (larger than some absolute constant). There exists depth-3*

$$\left(\frac{n}{r^{1+\frac{1}{8}}}, \frac{n}{r}\right) \text{-partial}$$

ERF network with $O(n)$ edges.

Proof. The depth-3 network $G = (V_0 \cup V_1 \cup V_2 \cup V_3, E)$ is constructed as follows:

- The bipartite graph between V_2 and V_3 is a disperser such that for any subset $J \subseteq V_3$ of size $|J| \in \left[\frac{n}{r^{1+1/8}}, \frac{n}{r}\right]$, inequality $|\Gamma(J)| \geq |J|$ always holds. The existence of such graph can be shown by a standard probabilistic argument [63], that is, connecting each vertex in V_3 to $O(1)$ randomly chosen neighbors in V_2;

- The bipartite graph and routing schemes between V_1 and V_2 satisfy the conditions in Lemma 91;

- The bipartite graph between V_0 and V_1 is constructed randomly: for each vertex $v \in V_1$, connect it to $r^{\frac{2}{3}}$ random neighbors [8] in V_0;

- The routing schemes are randomly defined in a natural way: for each vertex in V_1, among its $r^{\frac{2}{3}}$ incoming edges, always choose a random one to route (uniformly and independently for each scheme in $[n]$).

For any $\delta \in \left[0, \frac{1}{8}\right]$, let us fix $J \subseteq V_3$ of size $\frac{n}{r^{1+\delta}}$, and $I \subseteq [n]$ of size $\leq r^{1+\delta}$. Let $i := |I|$, $j := |J|$ as usual. We will estimate the probability that $|\mathfrak{R}_I^{-1}(J)| < \frac{1}{2^{15}} \cdot ij$, and take a union bound. Finally, we can amplify the constant arbitrarily close to 1 using Lemma 57.

8. For convenience of analysis, we allow repetition.

Since the bipartite graph between V_2 and V_3 is a disperser, we have $|\Gamma(J)| \geq |J| = \frac{n}{r^{1+\delta}}$. By Lemma 91, when $|I| \geq 32 \cdot r^{\frac{1}{3}+\delta}$, there exist $\frac{1}{8} \cdot |V_1|$ good vertices; when $|I| < 32 \cdot r^{\frac{1}{3}+\delta}$, there exist $\frac{1}{256} \cdot ij$ good vertices.

Case 1: $|I| \geq 32 \cdot r^{\frac{1}{3}+\delta}$. By Lemma 91, there exist $\frac{1}{8} \cdot |V_1|$ good vertices such that

$$\left|I_u^{(J)}\right| \;\geq\; \frac{1}{32} \cdot \frac{|I| \cdot |J|}{|V_1|} \;=\; \frac{i}{32} \cdot r^{-\frac{1}{3}-\delta} \;.$$

Let edge $e = (w, u) \in V_0 \times V_1$. Let $\mathbf{Y}_e$ be the indicator variable that e is routed during $I_u^{(J)} \subseteq I$. (Note that each time u picks a random neighbor to route.) Let $\mathbf{Y}_u := \sum_{e \to u} \mathbf{Y}_e$, and let $\mathbf{Y} := \sum_{u \in V_1} \mathbf{Y}_u$.

Claim 94. $(\mathbf{Y}_e)_{e \in V_0 \times V_1}$ *is negatively associated.*

Proof. For fixed $u \in V_1$, $(\mathbf{Y}_e)_{e \to u}$ satisfies the condition of Lemma 83 (Zero-One Lemma), and thus is negatively associated. Since $(\mathbf{Y}_e)_{e \to u}$ are independent for different u, by Proposition 82 (1), $(\mathbf{Y}_e)_{e \in V_0 \times V_1}$ is negatively associated. □

Claim 95. $(\mathbf{Y}_u)_{u \in V_1}$ *is negatively associated.*

Proof. Apply Proposition 82 (2). □

For each good $u \in V_1$, $\left|I_u^{(J)}\right| \geq \frac{i}{32} \cdot r^{-\frac{1}{3}-\delta}$; for each routing scheme in $I_u^{(J)}$, we pick a random neighbor of u to route. By Proposition 84, we have

$$\begin{aligned} \Pr\left[\mathbf{Y}_u < \frac{i}{64} \cdot r^{-\frac{1}{3}-\delta}\right] \;&\leq\; \binom{\frac{i}{32} \cdot r^{-\frac{1}{3}-\delta}}{\frac{i}{64} \cdot r^{-\frac{1}{3}-\delta}} \cdot \left(\frac{\frac{i}{64} \cdot r^{-\frac{1}{3}-\delta}}{r^{\frac{2}{3}}}\right)^{\frac{i}{64} \cdot r^{-\frac{1}{3}-\delta}} \\ &\leq\; 2^{-\frac{i}{64} \cdot r^{-\frac{1}{3}-\delta} \log\left(\frac{n}{ij}\right)} \;. \end{aligned}$$

Since $(\mathbf{Y}_u)_{u \in V_1}$ is negatively associated, among $\frac{1}{8} \cdot |V_1|$ good vertices, the probability that

there exist $\frac{15}{128} \cdot |V_1|$ vertices with $\mathbf{Y}_u < \frac{i}{64} \cdot r^{-\frac{1}{3}-\delta}$ is bounded by

$$\binom{\frac{1}{8} \cdot |V_1|}{\frac{1}{128} \cdot |V_1|} \cdot \Pr\left[\mathbf{Y}_u < \frac{i}{64} \cdot r^{-\frac{1}{3}-\delta}\right]^{\frac{15}{128} \cdot |V_1|} \leq 2^{-\frac{ij}{2^{12}} \log\left(\frac{n}{ij}\right)},$$

which implies

$$\Pr\left[\mathbf{Y} < \frac{ij}{2^{13}}\right] \leq 2^{-\frac{ij}{2^{12}} \log\left(\frac{n}{ij}\right)}. \tag{3.33}$$

By definition, $\left|\mathfrak{R}_I^{-1}(J)\right|$ equals the number of distinct elements by choosing $\mathbf{Y}$ uniformly random elements in $[n]$. Thus,

$$\begin{aligned}\Pr\left[\left|\mathfrak{R}_I^{-1}(J)\right| < \frac{ij}{2^{14}} \mid \mathbf{Y} \geq \frac{ij}{2^{13}}\right] &\leq \binom{\frac{1}{2^{13}} \cdot ij}{\frac{1}{2^{14}} \cdot ij} \left(\frac{\frac{1}{2^{14}} \cdot ij}{n}\right)^{\frac{ij}{2^{14}}} \\ &\leq 2^{-\frac{ij}{2^{14}} \log\left(\frac{n}{ij}\right)}. \end{aligned} \tag{3.34}$$

Combining (3.33) and (3.34), we have

$$\Pr\left[\left|\mathfrak{R}_I^{-1}(J)\right| < \frac{ij}{2^{14}}\right] \leq 2^{-\frac{ij}{2^{15}} \log\left(\frac{n}{ij}\right)}.$$

Finally, apply a union bound over all such I, J and the proof of this case follows from Lemma 58.

Case 2: $|I| < 32 \cdot r^{\frac{1}{3}+\delta}$**.** By Lemma 91, there exist $\frac{1}{256} \cdot ij$ good vertices. Each good vertex contributes *at least* one (uniformly random) element to $\mathfrak{R}_I^{-1}(J)$, and thus

$$\Pr\left[\left|\mathfrak{R}_I^{-1}(J)\right| < \frac{ij}{512}\right] \leq \binom{\frac{1}{256} \cdot ij}{\frac{1}{512} \cdot ij} \left(\frac{\frac{1}{512} \cdot ij}{n}\right)^{\frac{ij}{512}} < 2^{-\frac{1}{512} \log\left(\frac{n}{ij}\right)}.$$

To finish the proof of this case, we apply a union bound by using Lemma 58. □

The final construction for depth-3 ERF networks is by taking the union of $O(\log \log n)$ partial ERF networks, each of size $O(n)$.

Theorem 96. *There exist depth-3 ERF networks of size $O(n \log\log n)$.*

Proof. Let c be a sufficiently large constant. Taking

$$r = c, c^{9/8}, c^{(9/8)^2}, \ldots, c^{(9/8)^{O(\log\log n)}},$$

and applying the Lemma 93, we obtain $\left(n^{\frac{9}{20}}, n\right)$-partial ERF network of depth 3 and size $O(n \log\log n)$. Combining it with Corollary 80, we prove this theorem. □

3.5 The Challenge

Motivated by the study of semi-conservative circuits computing the Shift operator, we are interested in bounding the number of wires needed to implement an Expansive Routing Family (in bounded depth). At the same time, this leads us to a more general study of routing networks, which may not imply any circuit lower bound, but seems interesting on its own. In this section, we will formulate this challenge, and present some of our progress towards this challenge.

Routings in our work are intended to fulfill demands for commodity flow from certain inputs to certain outputs. Collections of demands are specified as a *request* $\mathfrak{r}$. This is a mapping $\mathfrak{r} : [n] \to \mathcal{P}([m])$, where $\mathfrak{r}(i) \subseteq [m]$ gives the set of indices $j \in [m]$ for which $y_j \in Y$ wants to receive commodity i from $x_i \in X$. We also call $\mathfrak{r}$ an (n, m)-*request*. For $J \subseteq [m]$, we also define

$$\mathfrak{r}^{-1}(J) = \{i \in [n] : \mathfrak{r}(i) \cap J \neq \emptyset\} .$$

We say that a routing $\mathfrak{R}$ *(exactly) fulfills* the request $\mathfrak{r}$, and write $\mathfrak{R} \models \mathfrak{r}$, if for every $j \in [m]$, we have

$$\{\mathfrak{R}(e)\}_{e \to y_j} \cap [n] \ = \{i : j \in \mathfrak{r}(i)\} .$$

An (n, m)-*multirequest* $\mathscr{R} = (\mathfrak{r}_1, \ldots, \mathfrak{r}_T)$ is just a sequence of (n, m)-requests. We say that G implements $\mathscr{R}$ if G implements $\mathfrak{r}_t$ for every $t \in [T]$.

We now come to some definitions of central interest. Given $n, m \geq 0$ and a (n, m)-multirequest $\mathscr{R}$, define $s(\mathscr{R})$ as the minimum number of edges in any network G (with $|X| = n, |Y| = m$) that implements $\mathscr{R}$. We also define a bounded-depth version: for $d > 0$, define $s_d(\mathscr{R})$ as the least number of edges in any depth-d network G that implements $\mathscr{R}$.

The following general research challenge seems important, yet is poorly-understood and is not studied in its full generality:

Challenge 97. *Develop a powerful and broadly-applicable set of techniques for lower-bounding and upper-bounding $s(\mathscr{R})$ and $s_d(\mathscr{R})$.*

In some interesting cases, the quantities defined above have been effectively lower-bounded. For example, set $m = n$ and let $\mathscr{R}_{perm}$ denote the set of all *permutation requests*. (A permutation request $\mathbf{r}_\pi$ is specified by a permutation $\pi \in S_n$: we have $\mathfrak{r}_\pi(i) = \{\pi(i)\}$.)

Let $\mathscr{R}_{shift} \subset \mathscr{R}_{perm}$ denote the set of requests specified by *cyclic shift* permutations $\{\tau_j\}_{j \in \mathbb{Z}_n}$, where

$$\tau_j(i) := i + j \mod n \ .$$

It was shown by Pippenger and Valiant [68] that $s(\mathscr{R}_{shift}) = \Omega(n \log n)$. (This implies the same lower bound for the wire complexity of conservative circuits to compute the Shift operator.) Later, this result was refined to a statement [9] for all depths by Pippenger and Yao [69]: for any $d > 0$,

$$s_d(\mathscr{R}_{shift}) \geq d \cdot n^{1+\frac{1}{d}} \ . \tag{3.35}$$

This is complemented in [69] by a nearly matching upper bound on the larger class $\mathscr{R}_{perm}$, namely

$$s_d(\mathscr{R}_{perm}) = \widetilde{O}\left(n^{1+\frac{1}{d}}\right)$$

for constant d.

We will give a generalization of the lower bound in Eq. (3.35), based on entropy concepts,

9. In the proof, they assume the network is layered.

that can be applied to any multirequest. Although it does not always yield strong bounds, it does provide at least one interesting new result.

3.5.1 Entropy Lower Bound

Before giving our generalization, let us show a simple proof of a slightly weaker statement than Eq. (3.35), that is,

$$s_d(\mathscr{R}_{shift}) \geq \Omega\left(n^{1+\frac{1}{d}}\right) , \tag{3.36}$$

where the proof is implicit in Pudlák [65].

Recall that a graph with n inputs and n outputs is an $f(r)$-*grate* if after removing any r vertices, there are at least $f(r)$ input-output pairs remaining connected [84].

Proposition 98. *If directed acyclic graph G is an $f(r)$-grate of depth d, then for every r,*

$$|E(G)| > r\left(\frac{f(r)}{n}\right)^{\frac{1}{d}} .$$

Proof. The proof closely follows Proposition 2 in [65]. Let $S := |E(G)|$. There are at most r vertices with degree $\geq \frac{S}{r}$, denoted by X. By the definition of $f(r)$-grate, Removing X, there are at least $f(r)$ distinct connected input-output pairs. Note that after removing X, all vertices have degree at most $\frac{S}{r}$, and thus the number of distinct connected input-output pairs is at most

$$n\left(\frac{S}{r}\right)^{d} \geq f(r),$$

which implies the desired result. □

Observe that any routing network realizing all shift permutations must an $r(n-r)$-grate. (Because any removed r vertices can contribute at most rn input commodities in n routing schemes.) Equation (3.36) follows by applying Proposition 98 with $r = \frac{n}{2}$.

Say that a request $\mathfrak{r}_t$ is k-*uniform* if $|\mathfrak{r}_t(i)| = k$ for all i. An (n, m)-multirequest $\mathscr{R} = (\mathfrak{r}_1, \ldots, \mathfrak{r}_T)$ is k-uniform if each $\mathfrak{r}_t$ is k-uniform. We will develop our lower bound method

for k-uniform multirequests. The method can be applied to non-uniform multirequests as well, but the numerical estimates involved become messier.

Given a multirequest $\mathscr{R}$, let us fix attention to a particular index $i \in [n]$. Consider the scenario in which a request index $t \in [T]$ is chosen at random according to some distribution $\mathcal{D}$. Let $\mathbf{S}_i$ denote the random variable $\mathfrak{r}_t(i) \subseteq [m]$, and let $h_i = h_{i,\mathcal{D}} := H(\mathbf{S}_i)$ denote the (Shannon) entropy of this random variable. With these definitions in hand, our result is as follows:

Theorem 99. *Take any $n, m, k > 0$. For any k-uniform (n, m)-multirequest $\mathscr{R} = (\mathfrak{r}_1, \ldots, \mathfrak{r}_T)$ and any $d > 1$,*

$$s_d(\mathscr{R}) \geq \frac{dk}{4} \cdot \sum_{i=1}^{n} 2^{\frac{h_i}{dk}} - 2dkn \ .$$

As an example application, let $k := 1$, $m := n$, and take $\mathscr{R} := \mathscr{R}_{shift}$, with $\mathcal{D}$ as the uniform distribution; then $h_i = \log_2 n$ for all i, and we get $s_d(\mathscr{R}_{shift}) \geq (1/4)dn^{1+\frac{1}{d}} - 2dn$, off by essentially only a factor of 4 from the lower bound in [69]. (Their lower bound can be perfectly recovered from our proof, by using more careful estimates for this specific case.)

As another corollary of our theorem, we get new super-linear lower bounds on the number of edges in *asymmetric generalized connectors*. For instance, suppose we want a network G with $\sqrt{n}$ input vertices, n output vertices, and we want to be able to simultaneously route the input values $x_1, \ldots, x_{\sqrt{n}}$ along disjoint trees in G to arbitrary disjoint subsets $A_1, \ldots, A_{\sqrt{n}} \subseteq [n]$ of the output vertices. Theorem 99 can be used to show that G must have $n^{1+\Omega(1/d)}$ edges.

Proof of Theorem 99. Suppose that G implements $\mathscr{R}$, and let $(\mathfrak{R}_1, \ldots, \mathfrak{R}_T)$ be routings in G where $\mathfrak{R}_t \models \mathfrak{r}_t$.

Fix an index $i \in [n]$. Recall that $t \in [T]$ is chosen according to distribution $\mathcal{D}$. Define the random edge-set

$$\mathbf{E}_i \ := \ \{e \in E(G) : \mathfrak{R}_t(e) = i\} \ ,$$

with associated vertex set $\mathbf{V}_i$ consisting of all vertices adjacent to an $e \in \mathbf{E}_i$. Now $\mathfrak{R}_t$ is

a routing that exactly fulfills the k-uniform request $\mathfrak{r}_t$. It follows that $\mathbf{G}_i := (\mathbf{V}_i, \mathbf{E}_i)$ is a directed tree with k leaves, with root vertex $x_i \in X$.

For $\ell \in \{0, \ldots, d\}$, define

$$\mathbf{Z}_{i,\ell} \;:=\; \{v \in V_i : v \text{ is at distance} \leq \ell \text{ from } x_i \text{ within } G_i\} \;.$$

Thus, $\{x_i\} = \mathbf{Z}_{i,0} \subseteq \mathbf{Z}_{i,1} \subseteq \ldots \subseteq \mathbf{Z}_{i,d}$, and $\mathbf{Z}_{i,d} \cap \text{leaves} = \mathbf{S}_i = \mathfrak{r}_t(i)$ (using the fact $\mathfrak{R}_t \models \mathfrak{r}_t$).

It follows that $H(\mathbf{Z}_{i,d}) \geq H(\mathbf{S}_i) = h_i$. On the other hand, we can derive an upper bound on $H(\mathbf{Z}_{i,d})$, as follows. First, for $\ell \in \{0, \ldots, d-1\}$, define the random variables

$$\partial_{i,\ell} \;:=\; \{e \in E(G) : e = (v, v') \text{ for some } v \in (\mathbf{Z}_{i,\ell} \setminus \mathbf{Z}_{i,\ell-1})\} \;, \quad \mathbf{K}_{i,\ell} \;:=\; \left|\partial_{i,\ell}\right| \;,$$

determined by $Z_{i,\ell}$. (We use the convention $\mathbf{Z}_{i,-1} := \emptyset$, so that $\partial_{i,0}$ is just the outgoing edges of x_i.) Note that for $0 \leq \ell < d$, once we've conditioned on $\mathbf{Z}_{i,0}, \ldots, \mathbf{Z}_{i,\ell}$ (which determine $\partial_{i,\ell}$), then the set $\mathbf{Z}_{i,\ell+1}$ can be determined from $\partial_{i,\ell} \cap \mathbf{E}_i$, a set of at most k edges within $\partial_{i,\ell}$. It follows that

$$H(\mathbf{Z}_{i,\ell+1} | \mathbf{Z}_{i,0}, \ldots, \mathbf{Z}_{i,\ell}) \;\leq\; \mathbb{E}\left[\log_2 \binom{\mathbf{K}_{i,\ell}}{\leq k}\right] \;,$$

where $\binom{b}{\leq c} := \sum_{0 \leq s \leq c} \binom{b}{s}$. (It is not assumed in this definition that $c \leq b$.) We will use the fact that for $\alpha \in (0, \frac{1}{2}]$,

$$\binom{b}{\leq \alpha b} \;\leq\; 2^{H(\alpha) b} \;, \tag{3.37}$$

where $H(\alpha) := \alpha \log_2\left(\frac{1}{\alpha}\right) + (1-\alpha)\log_2\left(\frac{1}{1-\alpha}\right)$ is the binary entropy function.

Applying the chain rule for entropy, we have

$$\begin{aligned} H(\mathbf{Z}_{i,d}) &= H(\mathbf{Z}_{i,0}) + \sum_{\ell=1}^{d} H(\mathbf{Z}_{i,\ell}|\mathbf{Z}_{i,0},\ldots,\mathbf{Z}_{i,\ell-1}) \\ &\leq \sum_{\ell=1}^{d} \mathbb{E}\left[\log_2 \binom{\mathbf{K}_{i,\ell}}{\leq k}\right]. \end{aligned} \tag{3.38}$$

Define $\mathbf{K}'_{i,\ell} := \max(\mathbf{K}_{i,\ell}, 2k) \leq \mathbf{K}_{i,\ell} + 2k$. Note that $\binom{\mathbf{K}_{i,\ell}}{\leq k} \leq \binom{\mathbf{K}'_{i,\ell}}{\leq k}$. Using Eqs. (3.37) and (3.38),

$$\begin{aligned} H(\mathbf{Z}_{i,d}) &\leq \sum_{\ell=1}^{d} \mathbb{E}\left[\log_2 \left(2^{H\left(\frac{k}{\mathbf{K}'_{i,\ell}}\right)\mathbf{K}'_{i,\ell}}\right)\right] \\ &= \sum_{\ell=1}^{d} \mathbb{E}\left[H\left(\frac{k}{\mathbf{K}'_{i,\ell}}\right)\cdot \mathbf{K}'_{i,\ell}\right] \\ &\leq \sum_{\ell=1}^{d} \mathbb{E}\left[\left(\frac{k}{\mathbf{K}'_{i,\ell}}\cdot \log_2\left(\frac{\mathbf{K}'_{i,\ell}}{k}\right) + \log_2\left(\frac{1}{1-k/\mathbf{K}'_{i,\ell}}\right)\right)\cdot \mathbf{K}'_{i,\ell}\right] \\ &\leq \sum_{\ell=1}^{d} \mathbb{E}\left[k\cdot \log_2\left(\frac{\mathbf{K}'_{i,\ell}}{k}\right) + \mathbf{K}'_{i,\ell}\cdot \log_2\left(1+2k/\mathbf{K}'_{i,\ell}\right)\right] \\ &= \sum_{\ell=1}^{d} \mathbb{E}\left[k\cdot \log_2 \mathbf{K}'_{i,\ell} - k\cdot \log_2 k + 2k\right] \\ &\leq k\cdot\left(\sum_{\ell=1}^{d} \mathbb{E}\left[\log_2 \mathbf{K}'_{i,\ell}\right]\right) + dk(2-\log_2 k) \\ &\leq dk\cdot \log_2\left(\mathbb{E}\left[\frac{1}{d}\sum_{\ell=1}^{d}\mathbf{K}'_{i,\ell}\right]\right) + dk(2-\log_2 k)\ , \end{aligned}$$

the last step by Jensen's inequality. Rearranging,

$$\mathbb{E}\left[\sum_{\ell=1}^{d}\mathbf{K}'_{i,\ell}\right] \geq d\cdot 2^{\frac{h_i}{dk}+\log_2 k-2} = \frac{dk}{4}\cdot 2^{\frac{h_i}{dk}},$$

which implies

$$\mathbb{E}\left[\sum_{\ell=1}^{d} \mathbf{K}_{i,\ell}\right] \;\geq\; \frac{dk}{4} \cdot 2^{\frac{h_i}{dk}} - 2dk \;. \tag{3.39}$$

Now fix any choice of the random variable $t \in [T]$, which determines all the trees G_i and the random variables $Z_{i,0}, \ldots, Z_{i,d}$. Note that for $0 \leq \ell < \ell' \leq d$, the edge-sets $\partial_{i,\ell}, \partial_{i,\ell'}$ are disjoint. Also, $\partial_{i,\ell} \cap \partial_{i',\ell'} = \emptyset$ when $i \neq i'$, since the intersection $V_i \cap V_{i'}$ contains only vertices in Y (which are sinks). So with probability 1 over t , we have

$$\sum_{i=1}^{n}\sum_{\ell=1}^{d} K_{i,\ell} \;\leq\; |E(G)| \;.$$

It then follows from Eq. (3.39) that

$$|E(G)| \;\geq\; \frac{dk}{4}\sum_{i=1}^{n} 2^{\frac{h_i}{dk}} \;-\; 2dkn \;.$$

As G was an arbitrary depth-d network implementing $\mathscr{R}$, this proves the Theorem. □

3.5.2 Routing Networks for Shifts

In [69], Pippenger and Yao proved the following:

Theorem 100. *[69] For any d, there exists a depth-d routing network realizing all permutations of size at most*

$$256d(d-1)n\,(2n\ln(2n))^{\frac{1}{d}} + 2(512)^d n(2\ln n)^{d-1}.$$

At the same time, they show that $dn^{1+\frac{1}{d}}$ is a lower bound for realizing all shift permutations. Note that there are only n shift permutations, while the number of all permutations is $n!$. It would be interesting to remove the $(\log n)^{\frac{1}{d}}$ term for routing networks realizing all shift permutations.

In this subsection, we will remove $(\log n)^{\frac{1}{d}}$ for constant d, and our analysis is simpler,

which is based on the definition of being *H-decomposable with small leftover size*. Our argument is inspired by the construction of semilinear circuit by Pudák *et al.* [64].

Let $G(U \cup W \cup V, E)$ be layered depth-d graph with inputs $U(G) = \{u_1, u_2, \ldots, u_n\}$, outputs $V(G) = \{v_1, v_2, \ldots, v_n\}$, and intermediate vertices W. Let H_k be a layered depth-d graph with inputs $U(H_k) = \{u_1, u_2, \ldots, u_k\}$ and outputs $V(H_k) = \{v_1, v_2, \ldots, v_k\}$.

Definition 101. *Given the above definitions, G is H_k-*decomposable with leftover size $\leq s$ with respect to $\pi \in S_n$ *if there exists a vertex decomposition*

$$\begin{aligned} U_1 \dot{\cup} U_2 \dot{\cup} \cdots \dot{\cup} U_\ell &\subseteq U, \\ V_1 \dot{\cup} V_2 \dot{\cup} \cdots \dot{\cup} V_\ell &\subseteq V, \\ W_1 \dot{\cup} W_2 \dot{\cup} \cdots \dot{\cup} W_\ell &\subseteq W, \end{aligned}$$

such that the followings are satisfied:

- $|U_i| = |V_i|$ *and* $\pi(U_i) = V_i$ *for all* $i \in [\ell]$*;*
- *for each* $i \in [\ell]$*, the subgraph induced by* $U_i \cup W_i \cup V_i$ *contains* H_k *as a subgraph with respect to* π*, that is, if* ψ *maps* $u_r \in V(G)$ *to* $u_s \in V(H_k)$*, then* ψ *maps* $v_{\pi(r)}$ *to* $v_s \in V(H_k)$*, where one-on-one map* $\psi : U_i \cup W_i \cup V_i \to V(H_k)$ *denotes the subgraph isomorphism. For convenience, we say* ψ *is* an embedding of H_k into G with respect to π*;*
- $|U| - k\ell \leq s$*, and* $|V| - k\ell \leq s$*.*

For fixed depth d, $\mathbf{G}_p(n)$ denotes the *layered* depth-d random graph

$$\mathbf{G}_p(n) \;=\; \mathbf{G}_p(\underbrace{n, n, \cdots, n}_{d+1}) \;,$$

where the probability that there exists an edge between any pair of vertices in adjacent layers is p.

Lemma 102. *Let H be a path of length d. Let $p = n^{-\theta}$, where $\theta = \frac{d-1}{d} - \alpha_d$, and $0 < \alpha_d < \frac{1}{d^2}$. Then,* [10]

$$\Pr\left[\textit{there exists no embedding of } H \textit{ into } \mathbf{G}_p(n)\right] \leq e^{-(1-o(1))\mu},$$

where $\mu = n^{1+d\alpha_d}$.

Proof. Let $H(1), H(2), \ldots$ enumerate all possible embeddings of H into $\mathbf{G} = \mathbf{G}_p(n)$. By Janson's inequality,

$$\Pr\left[\text{there exists no embedding of } H \text{ into } \mathbf{G}\right] \;\leq\; e^{-\mu + \frac{1}{1-\epsilon}\cdot\frac{\Delta}{2}} \;,$$

where

$$\begin{aligned}
\epsilon &= \Pr[H(i) \subseteq \mathbf{G}] \;=\; p^d \;=\; o(1) \;, \\
\mu &= \mathbb{E}[\mathbf{X}(H, \mathbf{G})] \;=\; n^{1+d\alpha_d} \;, \\
\Delta &= \sum_{i \sim j} \Pr[H(i) \subseteq \mathbf{G} \wedge H(j) \subseteq \mathbf{G}] \;,
\end{aligned}$$

where random variable $\mathbf{X}(H, \mathbf{G})$ denotes the number of embeddings of H into $\mathbf{G}$. As in [47],

$$\Delta \;\leq\; \sum_{L \subset H} \frac{\mu^2}{\mathbb{E}[\mathbf{X}(L, \mathbf{G})]} \cdot X(L, H)^2 \;,$$

where the sum is over all subgraphs L with at least one edge which may occur in $H(i) \cap H(j)$ for $i \neq j$. Since H is of constant size,

$$\Delta = O\left(\frac{\mu^2}{\min_L \mathbb{E}\left[\mathbf{X}(L, \mathbf{G})\right]}\right),$$

10. By symmetry of layered random graph $\mathbf{G}_p(n)$, permutation $\pi \in S_n$ does not affect the probability, and thus we assume π is the identity without loss of generality.

and thus it suffices to prove $\mathbb{E}\left[\mathbf{X}(L, \mathbf{G})\right] = \omega(\mu)$. It is clear that

$$\mathbb{E}\left[\mathbf{X}(L, \mathbf{G})\right] \;=\; n^{v'(L)-\theta e(L)} \;, \tag{3.40}$$

where $v'(L)$ denotes the number of vertices in L excluding output. Note that L is a subgraph of H, which is a path of length d. So L is a disjoint union of segments (subpaths). For a segment of length d', its contribution to the linear form in the exponent of (3.40) is

$$d' + 1 - \theta d' \;=\; 1 + d'\left(\frac{1}{d} + \alpha_d\right) \;>\; 1 + d\alpha_d \;,$$

assuming $\alpha_d < \frac{1}{d^2}$. Therefore $\mathbb{E}\left[\mathbf{X}(L, \mathbf{G})\right] = \omega(\mu)$, and $\Delta = o(\mu)$, and the conclusion follows from Janson's inequality. □

For fixed depth, the following upper bound is tight up to a constant factor, because of the size lower bound $dn^{1+\frac{1}{d}}$ proved in [69].

Theorem 103. *For any fixed $d \geq 1$, there exists depth-d routing network of size $O\left(dn^{1+\frac{1}{d}}\right)$ realizing all n shift permutations.*

Proof. Consider depth-d layered random graph $\mathbf{G}_p(n)$, where $p = n^{-\theta}$, $\theta = \frac{d-1}{d} - \alpha_d$, and

$$\alpha_d \;=\; \frac{2}{d} \cdot \log_{\frac{n}{2}} d \;.$$

It is clear that the expected number of edges in $\mathbf{G}_p(n)$ is

$$dn^{1+\frac{1}{d}+\alpha_d} \;=\; dn^{1+\frac{1}{d}} \cdot d^{\frac{2}{d}} \cdot 2^{\frac{2}{d}\cdot\log_{\frac{n}{2}} d} \;=\; O\left(dn^{1+\frac{1}{d}}\right) \;.$$

We claim, for any shift permutation $\pi \in S_n$, $\mathbf{G}_p(n)$ is H-decomposable with leftover size $\leq \frac{n}{2}$ with probability $1 - o_n(1)$.

The idea is to remove the embeddings one by one until the leftover has size $\leq \frac{n}{2}$. Note that there are n shift permutations in total, and the number of all possible leftover subgraphs of

$\mathbf{G}_p(n)$ with inputs size $> n/2$ is bounded by $\binom{n}{\leq n/2}^d$, and thus the total number is bounded by 2^{dn}. Applying a union bound over all possible π and all possible subgraphs, and using Lemma 102, we claim that probability that $\mathbf{G}_p(n)$ is H-decomposable with leftover size $\leq n/2$ is at least

$$1 - e^{-(1-o(1))\cdot\left(\frac{n}{2}\right)^{1+d\alpha_d}} \cdot 2^{dn} \geq 1 - 2^{-dn\left(\frac{2d}{3}-1\right)} \to 1 \ ,$$

as desired. (We assume $d \geq 2$, because the case $d = 1$ is trivial.)

We have shown that, for any given shift permutation $\pi \in S_n$, graph $\mathbf{G}_p(n)$ can route at least $\frac{n}{2}$ pairs, while there are $\leq \frac{n}{2}$ pairs remaining. To take care of the rest, we put two dispersers at the top and the bottom. That is, the top disperser is a bipartite graph

$$\left(U = [n], V = \left[\frac{1.1n}{r}\right], E\right)$$

such that for any $X \subseteq U$ of size $\frac{n}{r}$, the set of neighbors $\Gamma(X)$ has size at least $\frac{n}{r}$, which implies that there is a matching for X. Using probabilistic method, it is known that such disperser of size $O(n \log r)$ exists (for example, see Theorem 12), and the bipartite graph is regular with right-degree $O(r \log r)$; the disperser at the bottom is symmetric. The middle part is a routing network with $\frac{1.1n}{r}$ inputs, and $\frac{1.1n}{r}$ outputs, which can route $\frac{n}{2r}$ pairs using the previous construction and analysis. So, the total size is $O(n \log r) + O\left(d\left(\frac{1.1n}{r}\right)^{1+\frac{1}{d}}\right)$, and the depth is $d + 2$. Let us decrease the depth by 2 by *expanding* all inputs and outputs (of the middle part), which will increase the size of the adjacent layers by a factor of at most $O(r \log r)$. So, the size of this depth-d routing network (parameterized by r) is bounded by

$$O\left(d\left(\frac{1.1n}{r}\right)^{1+\frac{1}{d}}\right) + O\left(\left(\frac{1.1n}{r}\right)^{1+\frac{1}{d}}\right) \times O(r \log r) = O\left(d\left(\frac{1.1n}{r}\right)^{1+\frac{1}{d}}\right) + O\left(\frac{(1.1n)^{1+\frac{1}{d}}}{r^{\frac{1}{2d}}}\right) \ ,$$

Let $r = 2, 2^1, \ldots, 2^{\log n}$, and take the union of all the routing networks, where the total size

is

$$\sum_{i\geq 0} O\left(d\left(\frac{1.1n}{2^i}\right)^{1+\frac{1}{d}}\right) + O\left(\frac{(1.1n)^{1+\frac{1}{d}}}{2^{\frac{i}{2d}}}\right) \;=\; O\left(dn^{1+\frac{1}{d}}\right) \;.$$

□

The above proof works for any set of n permutations, not necessarily shift permutations. The following corollary is an immediate generalization.

Corollary 104. *For any fixed $d \geq 1$, there exists depth-d routing network of size*

$$O\left(dn^{1+\frac{1}{d}}\left(d + \frac{\log s}{n}\right)^{\frac{1}{d}}\right)$$

realizing any given s permutations in S_n.

Proof. Let

$$\alpha_d = \frac{\log\left(\frac{3}{4}\left(2d + \frac{2\log s}{n}\right)\right)}{d\log\left(\frac{n}{2}\right)} \;,$$

and rest of the proof is the almost same as Theorem 103. □

3.5.3 Open Problems

We end this chapter by listing a few open problems.

Question 105. *For depth $d \geq 4$, what is the size of the smallest depth-d ERF networks?*

For depths 2 and 3, our bounds are tight. For depth $d \geq 4$, the answer is somewhere between $\Omega_d\left(\lambda_d(n)\cdot n\right)$ and $O(n\log\log n)$.

Question 106. *Are conservative circuits (nearly-)optimal for computing the Shift operator?*

This is a major unsolved problem. More generally, are conservative circuits optimal for *every* operator for which they can actually perform the computation? It might be instructive to try to cook up a counterexample. In [69] it was shown that Shift can be computed

conservatively with $\widetilde{O}\left(n^{1+\frac{1}{d}}\right)$ edges in constant depth d, and that any conservative solution requires $dn^{1+\frac{1}{d}}$ edges. No asymptotic improvements for the non-conservative case are known. It was shown in [64] that for depth-2 circuits with a certain *restricted structure*, non-conservative circuits can provably outperform conservative ones, but even here the savings shown is only a sub-logarithmic factor.

Question 107. *Explicit constructions of ERF networks.*

The explicitness here has twofold meanings — both the network graphs and the routing schemes are explicit.

Question 108. *What is the wire complexity of the Shift operator in the unrestricted model?*

For the Shift operator, the upper bound is $\tilde{O}(n^{1+1/d})$ while the best known lower bound is $\Omega(n \cdot \lambda_d(n))$ for constant depth $d \geq 3$. Currently, it is not clear which one is closer to the truth.

CHAPTER 4

AC0 COMPLEXITY OF SUBGRAPH ISOMORPHISM

4.1 Introduction

The *subgraph isomorphism problem* takes as its input two graphs H and G and asks to determine whether or not G contains a subgraph (not necessarily induced) isomorphic to H. This is one of the most basic NP-complete problems that includes Clique and Hamiltonian Cycle as special cases, and little more can be said about its complexity in full generality.

A significant body of research, motivated both by the framework of parameterized complexity and practical applications, has been devoted to the case when the graph H is fixed and possesses some useful structure (see e.g. the sources [8, 28, 29, 57, 58, 61] related to the subject). To stress its nature in this situation, the graph H is traditionally called a *pattern* and designated by the letter P; we also follow this convention and denote by Subgraph(P) the corresponding restriction of the general subgraph isomorphism problem.

The sources above (among many others) provide quite non-trivial improvements on the obvious time bound $O(n^{|V(P)|})$ in many cases of interest. But for *unconditional* lower bounds we, given our current state of knowledge, have to resort to restricted models, and, indeed, a substantial amount of work has been done here in the context of both bounded-depth circuits and monotone circuits. In this chapter we focus on the former model.

As for upper bounds, it was observed by Amano [9] that the color-coding algorithm by Alon, Yuster and Zwick [8] can be adapted to our context and gives AC^0 circuits for Subgraph(P) of size[1] $\widetilde{O}(n^{tw(P)+1})$, where $tw(P)$ is the treewidth of the pattern P. Our work is motivated by the following natural question:

How tight is this bound?

Or, in other words,

1. "$\widetilde{O}$" is the "soft" version of the "big-O" notation that ignores not only constant but polylogarithmic multiplicative factors as well.

Question 1. *Is it possible to give good general lower bounds on the* AC^0 *complexity of* $\textsc{Subgraph}(P)$ *in terms of the treewidth of* P *only?*

Prior to our work, Rossman [73] answered this question in affirmative for the case of a k-clique by proving a lower bound of $\Omega(n^{k/4})$ on the AC^0 complexity of $\textsc{Subgraph}(K_k)$. Generalizing Rossman's method, Amano [9] gave a general lower bound that holds for arbitrary patterns P. It in particular implied an $n^{\Omega(k)}$ lower bound (and, thus, an affirmative answer to Question 1) for the $k \times k$ grid $G_{k,k}$: this result is very interesting since $G_{k,k}$ is the "canonical" example of a sparse graph with large treewidth.

Before discussing our results, it will be convenient to introduce the following handy notation: given a pattern P, we let $C(P)$ be the minimal real number $c \geq 0$ for which $\textsc{Subgraph}(P)$ is solvable on n-vertex graphs by AC^0 circuits of size $n^{c+o(1)}$. In this notation, the previous results mentioned above can be stated as $C(P) \leq tw(P) + 1$ ([8, 9], P any pattern), $C(K_k) \geq \frac{k}{4}$ [73] and $C(G_{k,k}) \geq \Omega(k)$ [9].

Our contributions.

We explicitly formulate and study two modifications that already played a great role in the previous research. The first of them is the *colorful* P*-subgraph isomorphism problem* $\textsc{Subgraph}_{\mathsf{col}}(P)$ in which the target graph G comes with a coloring $\chi : V(G) \to V(P)$ (that w.l.o.g. can and will be assumed to be a graph homomorphism), and we are looking only for properly colored P-subgraphs. Let $C_{\mathsf{col}}(P)$ be defined analogously to $C(P)$. Then the very first thing done by the algorithm of Alon, Yuster and Zwick is a simple reduction from $\textsc{Subgraph}(P)$ to $\textsc{Subgraph}_{\mathsf{col}}(P)$ thus establishing $C(P) \leq C_{\mathsf{col}}(P)$. After that they work exclusively with the colorful version that leads to

$$C(P) \leq C_{\mathsf{col}}(P) \leq tw(P) + 1.$$

We settle in the affirmative (up to a logarithmic factor) our motivating Question 1 for

the colorful version by proving the following

Theorem 109. $C_{\mathsf{col}}(P) \geq \Omega\left(\frac{tw(P)}{\log tw(P)}\right)$.

By previous work of Marx [57], it was known that $\textsc{Subgraph}_{\mathsf{col}}(P)$ has no $n^{o(tw(P)/\log tw(P))}$ time algorithm unless the Exponential Time Hypothesis fails. Theorem 109 establishes the same lower bound *unconditionally* for AC^0 circuits. (We say more about Marx's result and related work of Alon and Marx [5] in Section 4.6.)

We show that the colorful version is quite well-behaved by proving that it is minor-monotone: if Q is a minor of P, then $C_{\mathsf{col}}(Q) \leq C_{\mathsf{col}}(P)$ (Theorem 153).[2] Whether a similar result holds for $C(P)$ is open, but we give a strong evidence (Theorem 158) that even if this is true, the proof will most likely require totally different techniques. One possible interpretation is that perhaps the colorful version is in fact a cleaner and more natural model to study than the standard (uncolored) version. We also observe that if the pattern P is a core (i.e., every homomorphism from P to P is an automorphism), then $C(P) = C_{\mathsf{col}}(P)$ and thus our lower bound from Theorem 109 transfers to the uncolored case. What happens to $C(P)$ at the opposite side of the spectrum, say, for bipartite patterns P, remains wide open.

All lower bounds surveyed above, including our proof of Theorem 109, were actually achieved in the context of average-case complexity. Prior to our work, the only distribution that was considered for this purpose is the *Erdős-Rényi* model $G(n, n^{-\theta(P)})$, where $\theta(P)$ is the uniquely defined *threshold exponent* for which the probability of containing a copy of P is bounded away from 0 and 1 (see [47] or Section 4.2.4 below). Accordingly, we define $C_{\mathsf{ave}}(P)$ analogously to $C(P)$, but only require that our circuit outputs the correct answer a.a.s. (asymptotically almost surely) when the input is drawn from $G(n, n^{-\theta(P)})$. Clearly,

2. It is worth observing that this fact, along with the recent result [17] by Chekura and Chuzhoy and Amano's bound $C_{\mathsf{col}}(G_{k,k}) \geq \Omega(k)$ [9] already implies the weaker bound $C_{\mathsf{col}}(P) \geq tw(P)^{\Omega(1)}$. But the exponent given by this approach will be disappointingly small.

$C_{\mathsf{ave}}(P) \leq C(P)$ so the whole picture now looks like

$$C_{\mathsf{ave}}(P) \leq C(P) \leq C_{\mathsf{col}}(P) \approx tw(P),$$

where $\approx$ means approximation within a logarithmic factor. Also, $C_{\mathsf{ave}}(K_k) \geq k/4$ [73] and $C_{\mathsf{ave}}(G_{k,k}) \geq \Omega(k)$ [9] where K_k is the complete graph on k vertices and $G_{k,k}$ is the k-by-k grid.

We explicitly define a combinatorial parameter $\kappa(P)$ and prove the following

Theorem 110. $\kappa(P) \leq C_{\mathsf{ave}}(P) \leq 2\kappa(P) + O(1)$.

In other words, we give lower and upper bounds on the average-case AC^0 complexity for an arbitrary pattern P, matching within a quadratic factor. The proof of Theorem 110 exploits a duality in the definition of $\kappa(P)$, which has equivalent min-max and max-min formulations (the former suited to upper bounds and the latter to lower bounds). The lower bound $C_{\mathsf{ave}}(P) \geq \kappa(P)$ generalizes the proof of $C_{\mathsf{ave}}(K_k) \geq \frac{k}{4}$ in Rossman [73] and improves a previous lower bound of Amano [9] for general patterns P.

Let us say a few words about the proof of Theorem 109. Being itself a worst-case lower bound, it is obtained as the maximum of a family of average-case lower bounds with respect to P-colored random graphs. These random graphs generalize Erdős-Rényi random graphs in the P-colored setting by allowing different edge probabilities according to the color classes of vertices, and we believe that this generalization may be of independent interest. Each P-colored random graph in this family is parameterized by a point in a certain convex polytope, denoted $\theta_{\mathsf{col}}(P)$. We rely on results of [31, 57] that characterize the treewidth of P in terms of the existence of an appropriate concurrent flow on P, which we convert to a suitable point in $\theta_{\mathsf{col}}(P)$.

Finally, it is also worth noting that lower bounds in Theorems 109, 110 (and hence all our structural conclusions) hold even if we allow circuits of a super-constant depth $d(n)$, as long as $d(n) \leq o(\log n / \log\log n)$.

The chapter is organized as follows. In Section 4.2 we give the necessary definitions and preliminaries; in particular, in Section 4.2.6 we present the parameters $\kappa(P)$ and $\kappa_{\mathsf{col}}(P)$ that are our main technical tools. Section 4.3 is devoted to the proof of Theorem 110, and it also paves way to the proof of Theorem 109 that, up to a certain point, goes in parallel to the former. The proof of Theorem 109 is completed in Section 4.4. Section 4.5 contains our structural results about the behavior of $\textsc{Subgraph}(P)$ and $\textsc{Subgraph}_{\mathsf{col}}(P)$ with respect to minors and subgraphs. The chapter is concluded with a brief discussion and a list of open problems in Section 4.6.

4.2 Definitions and Preliminaries

Let $[k] := \{1, \ldots, k\}$, and let $\binom{X}{2}$ be the family of all 2-element subsets of X.

4.2.1 Graphs

We start off with terminology and notation for graphs. Throughout this chapter, *graphs* are finite simple graphs $G = (V(G), E(G))$ where $E(G)$ is a subset of $\binom{V(G)}{2}$. We often write $v(G)$ for $|V(G)|$ and $e(G)$ for $|E(G)|$.

A graph H is a *subgraph* of G, denoted $H \subseteq G$, if $V(H) \subseteq V(G)$ and $E(H) \subseteq E(G)$. For arbitrary G and H, $G + H$ and $G \times H$ respectively denote the *disjoint union* and *Cartesian product* of graphs G and H (where $E(G \times H) := \{\{(v, v'), (w, w')\} : \{v, w\} \in E(G)$ and $\{v', w'\} \in E(H)\}$).

A *homomorphism* from G to H is a function $\varphi : V(G) \to V(H)$ such that $\{\varphi(v), \varphi(w)\} \in E(H)$ for all $\{v, w\} \in E(G)$. A graph G is a *core* if every homomorphism from G to G is an automorphism.

The *treewidth* of G is denoted by $tw(G)$ (for the definition and background, see e.g. [15]). Relevant facts about treewidth will be stated where needed.

K_k is a *clique* on k vertices, and $G_{k,k}$ is a $k \times k$ *grid*. These graphs have treewidth

$tw(K_k) = k-1$ and $tw(G_{k,k}) = k$.

4.2.2 Monotone Projections

Definition 111. *Let I, J be arbitrary sets.*

1. *For a function $p : J \to I \cup \{0,1\}$ and $x \in \{0,1\}^I$, we write $p^*(x)$ for the unique $y \in \{0,1\}^J$ such that $y_j = x_{p(j)}$ if $p(j) \in I$, and $y_j = p(j)$ if $p(j) \in \{0,1\}$.*

2. *For boolean functions $f : \{0,1\}^I \to \{0,1\}$ and $g : \{0,1\}^J \to \{0,1\}$, we say that f is* reducible via a monotone projection *to g, denoted $f \leq_{\mathrm{mp}} g$, if there exists $p : J \to I \cup \{0,1\}$ such that $f(x) = g(p^*(x))$ for all $x \in \{0,1\}^I$. (Note that $\leq_{\mathrm{mp}}$ is transitive.)*

Any decision problem L can be represented as a sequence of Boolean functions $\{L^n\}$ in n variables. We say that L_1 is *reducible via a monotone projection* to another decision problem L_2 if for any n there exists[3] $m(n)$ such that $L_1^n \leq_{\mathrm{mp}} L_2^{m(n)}$. If in addition $m(n) \leq O(n)$, we call this projection *linear*.

4.2.3 Subgraph Isomorphism Problems

Throughout this chapter, the letters P, Q represent arbitrary fixed graphs that should be intuitively thought of as "patterns". G stands for a (large) "input" graph for the P-subgraph isomorphism problem. Subgraphs of G (not necessarily induced) which are isomorphic to P will be called *P-subgraphs*.

We also consider *P-colored graphs*, defined as pairs (G, χ) where G is a graph and $\chi : V(G) \to V(P)$ is a homomorphism. We usually suppress χ and simply refer to G as *P-colored graph*. In this setting, given a sub-pattern $Q \subseteq P$ (not necessarily induced), a *Q-subgraph* of G is a subgraph of G (again, not necessarily induced) that is isomorphic to Q and consistent with χ in the sense that every vertex $v \in V(Q)$ is mapped to a vertex in $\chi^{-1}(v)$.

3. Uniformity issues do not play any role in this chapter.

We consider two versions ("uncolored" and "colored") of the P-subgraph isomorphism problem:

- $\textsc{Subgraph}(P)$ is the problem, given a graph G, of determining whether or not G contains a P-subgraph.
- $\textsc{Subgraph}_{\mathsf{col}}(P)$ is the problem, given a P-colored graph (G, χ), of determining whether or not G contains a (properly colored) P-subgraph.

This problem is also known in the literature as the "partitioned" or "colorful" variant (see e.g. [8, 5]), and in this chapter we mostly adopt the latter term.

It will be convenient to introduce a notation for the AC^0 complexity of these problems. (Recall that AC^0 is the class of problems solvable by polynomial-size constant-depth boolean circuits over $\{\neg, \wedge, \vee\}$ with unbounded fan-in.)

Definition 112. *Let $C(P)$ (resp. $C_{\mathsf{col}}(P)$) denote the infimum of all real numbers $c > 0$ such that $\textsc{Subgraph}(P)$ (resp. $\textsc{Subgraph}_{\mathsf{col}}(P)$) is solvable (in the worst-case) on n-vertex graphs by AC^0 circuits of size[4] $O(n^c)$.*

Note that if $\textsc{Subgraph}(P)$ is reducible to $\textsc{Subgraph}(Q)$ via a linear monotone projection then $C(P) \leq C(Q)$, and this remains true if we add the subscript col to both sides.

Lemma 113.

1. *$C(P) \leq C_{\mathsf{col}}(P) \leq tw(P) + 1$.*
2. *If P is a core, then $C(P) = C_{\mathsf{col}}(P)$.*

Proof. (1): The second inequality $C_{\mathsf{col}}(P) \leq tw(P) + 1$ is by the color-coding algorithm of Alon, Yuster and Zwick [8] (adapted to the P-colored setting), which can be implemented in AC^0 as observed by Amano [9]. The first inequality $C(P) \leq C_{\mathsf{col}}(P)$ is also implicitly proved

4. In this chapter, the size of all constant-depth circuits is measured by the number of *gates*.

there by reducing $\textsc{Subgraph}(P)$ to $\textsc{Subgraph}_{\mathsf{col}}(P)$: the reduction searches through logarithmically many different colorings $\chi_1, \chi_2, \ldots : V(G) \to V(P)$ of the same target graph G, picked at random. An easy counting argument shows that a.a.s. every P-subgraph of G will be properly colored with respect to at least one of the colorings χ_i.

(2): This observation goes back at least to Grohe [37]. If P is a core, then $(G, \chi) \mapsto G$ is a reduction from $\textsc{Subgraph}_{\mathsf{col}}(P)$ to $\textsc{Subgraph}(P)$. To see why, it suffices to show that every P-subgraph of G is properly colored with respect to every homomorphism $\chi : G \to P$. Suppose H is a P-subgraph of G. Then $H = \varphi(P)$ for some one-to-one homomorphism $\varphi : P \to G$. Since P is a core, the homomorphism $\chi \circ \varphi : P \to P$ is an automorphism of P. It follows that the homomorphism $\chi|_{V(H)} : H \to P$ is one-to-one. Since $|E(H)| = |E(P)|$, it must be an isomorphism, that is H is properly colored with respect to χ. □

4.2.4 The Average Case

We now define the random graphs which appear in our average-case lower bounds for $\textsc{Subgraph}(P)$ and $\textsc{Subgraph}_{\mathsf{col}}(P)$. In the uncolored setting, we consider the Erdős-Rényi random graph $G(n, p(n))$ for an appropriately chosen threshold function $p(n)$. Also, in what follows we assume that P is non-empty, that is contains at least one edge.

Definition 114.

1. *The* threshold exponent *of P is defined by* $\theta(P) := \min\limits_{\substack{Q \subseteq P \\ Q \neq \emptyset}} \frac{v(Q)}{e(Q)}$.
2. *P is* balanced *if* $\frac{v(P)}{e(P)} = \theta(P)$.
3. *P is* strictly balanced *if* $\frac{v(Q)}{e(Q)} > \theta(P)$ *for every nonempty proper subgraph* $Q \subset P$.
4. *Let* $\mathsf{Bal}(P) := \bigcup \left\{ Q \subseteq P : \frac{v(Q)}{e(Q)} = \theta(P) \right\}$.

Lemma 115.

1. *P is balanced if and only if* $P = \mathsf{Bal}(P)$.

2. *For every P, $\mathsf{Bal}(P)$ is balanced and $\theta(\mathsf{Bal}(P)) = \theta(P)$.*

Proof. It suffices to show that $\mathcal{B} := \left\{Q \subseteq P : Q \neq \emptyset \wedge \frac{v(Q)}{e(Q)} = \theta(P)\right\}$ is closed under unions (in fact, it is closed under intersections as well). For all $Q_1, Q_2 \in \mathcal{B}$, we have

$$\begin{aligned} v(Q_1 \cup Q_2) + v(Q_1 \cap Q_2) &= v(Q_1) + v(Q_2) \\ &= \theta(P)e(Q_1) + \theta(P)e(Q_2) \\ &= \theta(P)e(Q_1 \cup Q_2) + \theta(P)e(Q_1 \cap Q_2). \end{aligned} \tag{4.1}$$

By definition of $\theta(P)$,

$$v(Q_1 \cup Q_2) \geq \theta(P)e(Q_1 \cup Q_2) \quad \text{and} \quad v(Q_1 \cap Q_2) \geq \theta(P)e(Q_1 \cap Q_2). \tag{4.2}$$

Together (4.1) and (4.2) imply that equality holds in (4.2), that is, $Q_1 \cup Q_2$ and $Q_1 \cap Q_2$ are both in $\mathcal{B}$. □

Recall that $G(n,p)$ is the Erdős-Rényi random graph with vertex set $[n]$, in which each $e \in \binom{[n]}{2}$ occurs as an edge independently with probability p. The next lemma states that $p = n^{-\theta(P)}$ is a threshold function for $\textsc{Subgraph}(P)$ and that detecting P-subgraphs on $G(n, n^{-\theta(P)})$ is equivalent to detecting $\mathsf{Bal}(P)$-subgraphs. (Lemma 116(1) is a standard fact about random graphs (see [47]); Lemma 116(2) was proved in [16].)

Lemma 116.

1. $\Pr[\, G(n, n^{-\theta(P)}) \text{ has a } P\text{-subgraph} \,]$ *is bounded away from* 0 *and* 1.

2. *Asymptotically almost surely, if $G(n, n^{-\theta(P)})$ contains a $\mathsf{Bal}(P)$-subgraph, then it contains a P-subgraph.*

With slight abuse of notation, we denote by $\textsc{Subgraph}_{\mathsf{ave}}(P)$ the algorithmic problem of solving $\textsc{Subgraph}(P)$ on $G(n, n^{-\theta(P)})$ correctly a.a.s, that is with probability that tends to 1 as n tends to ∞. (We remark that our results are unchanged if $n^{-\theta(P)}$ is replaced by

any other threshold function $p(n) \in \Theta(n^{-\theta(P)})$.) Similarly to Definition 112, let $C_{\mathsf{ave}}(P)$ be the smallest $c > 0$ for which this problem can be solved by AC^0-circuits of size $n^{c+o(1)}$.

Remark 117. *Obviously, $C_{\mathsf{ave}}(P) \leq C(P)$, but the gap between them can be arbitrarily large. Assume e.g. that $P = K_4 + G_{k,k}$ where $k \to \infty$. Then $\mathsf{Bal}(P) = K_4$ and thus Lemma 116(2) implies that $C_{\mathsf{ave}}(P) = C_{\mathsf{ave}}(K_4) \leq 4$. On the other hand, $\textsc{Subgraph}(G_{k,k})$ is reduced to $\textsc{Subgraph}(P)$ via an obvious linear monotone projection that takes G to $K_4 + G$. This proves $C(P) \geq C(G_{k,k}) \geq \Omega(k)$ by the result from [9].*

One might argue that this example is not "fair" since it heavily exploits the fact that the pattern P is highly unbalanced. It is, however, possible to give nearly the same separation (albeit, more complicated) with a strictly balanced pattern P. Say, let $d > 0$ be a sufficiently large constant, and $V(P) = [k]$, where $k \gg d$. We start building $E(P)$ with the clique on the set $[d]$, and then for every $i \in \{d+1, \ldots, k\}$ pick at random d different vertices $j_1, \ldots, j_d < i$ and add all d edges $\{j_\nu, i\}$. Then P will be strictly balanced since every subgraph with $v \geq d$ vertices has at most $\frac{d(d-1)}{2} + d(v-d) = \frac{d(2v-d-1)}{d}$ edges. Taking as a union sequence (see Definition 123 below) the natural sequence according to the order in which P was built, we conclude (see Definition 124) that $\kappa(P) \leq O(d)$. Hence $C_{\mathsf{ave}}(P) \leq O(d)$ by Theorem 110. On the other hand, randomness in selecting the edges implies that $tw(P) \geq \Omega(k)$ and that P is a core. From the latter fact we conclude that $C(P) = C_{\mathsf{col}}(P)$, and from the former, by Theorem 109, that $C_{\mathsf{col}}(P) \geq \Omega(k/\log k)$.

We now move on to the notion of average case complexity for $\textsc{Subgraph}_{\mathsf{col}}(P)$. In contrast to the uncolored setting, there is no single most natural distribution on P-colored random graphs. Instead, we consider a family of P-colored random graphs, denoted $G_{\alpha,\beta}(n)$, which are parameterized by certain pairs of functions $\alpha : V(P) \to [0,1]$ and $\beta : E(P) \to [0,2]$ called "threshold pairs". (Note: Unlike $G(n,p)$, the vertex set of $G_{\alpha,\beta}(n)$ is not $[n]$, but rather consists of $|V(P)|$ disjoint parts of different sizes.)

Definition 118. ***(P-colored random graph $G_{\alpha,\beta}(n)$)***

1. *A* threshold pair *for P is a pair (α, β) of functions $\alpha : V(P) \to [0,1]$ and*[5] *$\beta : E(P) \to [0,2]$ such that*

   - $\alpha(P) = \beta(P)$,
   - $\alpha(Q) \geq \beta(Q)$ *for all* $Q \subseteq P$,

   where $\alpha(Q) := \sum_{v \in V(Q)} \alpha(v)$ and $\beta(Q) := \sum_{e \in E(Q)} \beta(e)$.

2. *$\theta_{\mathsf{col}}(P)$ denotes the set of threshold pairs for P. Note that $\theta_{\mathsf{col}}(P)$ is a polytope in $\mathbb{R}^{V(P) \cup E(P)}$ and its section $\{\beta : (1, \beta) \in \theta_{\mathsf{col}}(P)\}$ is a polytope in $\mathbb{R}^{E(P)}$. We view elements of $\theta_{\mathsf{col}}(P)$ as the "P-colored" analogue of $\theta(P)$ (see Remark 119 below).*

3. *We say that $(\alpha, \beta) \in \theta_{\mathsf{col}}(P)$ is* nontrivial *if α and β are not identically zero.*

4. *We say that $(\alpha, \beta) \in \theta_{\mathsf{col}}(P)$ is* strictly balanced *if $\alpha(Q) > \beta(Q)$ for every nonempty proper subgraph $Q \subset P$.*

5. *For all $(\alpha, \beta) \in \theta_{\mathsf{col}}(P)$, let $G_{\alpha,\beta}(n)$ denote the random graph with vertex set $\{(v, i) : v \in V(P), 1 \leq i \leq \lfloor n^{\alpha(v)} \rfloor\}$ where each $\{(v, i), (w, j)\}$ with $\{v, w\} \in E(P)$ is an edge, independently, with probability $n^{-\beta(\{v,w\})}$. The P-coloring of $G_{\alpha,\beta}(n)$ is the obvious one: $(v, i) \mapsto v$.*

Remark 119. *Note that if P is a balanced pattern, then the pair of constant functions $(\alpha \equiv 1,\ \beta \equiv \theta(P))$ is a threshold pair for P; moreover, P is strictly balanced if and only if this (α, β) is strictly balanced. Thus, Definition 118 is indeed a generalization of threshold exponent for balanced patterns. The following lemma makes the analogy even more clear, justifies the terminology "threshold pair" and refines Lemma 116(1).*

Lemma 120. *For every pattern P and nontrivial threshold pair $(\alpha, \beta) \in \theta_{\mathsf{col}}(P)$,*

1. $\liminf_{n \to \infty} [\, G_{\alpha,\beta}(n)$ *contains no P-subgraph*$\,] \geq \dfrac{1}{e}$,

5. n^{α} and $n^{-\beta}$ will determine the number of vertices in the colored parts and edge densities between them, respectively, whence comes our choice of normalization.

2. $\liminf_{n\to\infty}[\, G_{\alpha,\beta}(n) \text{ contains exactly one } P\text{-subgraph}\,] \geq \frac{1}{e^{|E(P)|}}$.

The proof is included in the next subsection, that is, Section 4.2.5. With a bit of work, it is possible to completely characterize the asymptotic distribution of the number of P-subgraphs in $G_{\alpha,\beta}(n)$; this distribution is a function of independent Poisson random variables (in the uncolored setting, see [16] for a characterization of the asymptotic number of P-subgraphs in $G(n, n^{-\theta(P)})$).

In the context of $\textsc{Subgraph}_{\mathsf{col}}(P)$, we speak of the *average-case complexity with respect to* $G_{\alpha,\beta}(P)$, meaning the size of an AC^0 circuit which solves $\textsc{Subgraph}_{\mathsf{col}}(P)$ on $G_{\alpha,\beta}(P)$ with probability that tends to 1 as n tends to ∞. We do not introduce any special notation like $C_{\alpha,\beta}(P)$ as this concept is intended to be auxiliary.

4.2.5 Proof of Lemma 120

Fix a pattern P and a nontrivial threshold pair $(\alpha, \beta) \in \theta_{\mathsf{col}}(P)$. We can assume w.l.o.g. that $\beta(e) > 0$ for all $e \in E(P)$ (as the edges with $\beta(e) = 0$ can be removed). Following the approach of Bollobás and Wierman [16], we fix a chain of (necessarily induced) subgraphs

$$\emptyset = Q_0 \subset Q_1 \subset \cdots \subset Q_{t-1} \subset Q_t = P$$

satisfying

- $\alpha(Q_i) = \beta(Q_i)$ for all $0 \leq i \leq t$, and
- $\alpha(R) > \beta(R)$ for all $1 \leq i \leq t$ and $Q_{i-1} \subset R \subset Q_i$.

Call such a sequence $(Q_0, \ldots, Q_t)$ an (α, β)-*grading* of P. Clearly, at least one (α, β)-grading exists. Note that $1 \leq t \leq |E(P)|$, since (α, β) is nontrivial. (It is known that t is the same for all (α, β)-gradings; however, we will not use this fact.)

Let $\mathbf{G} := G_{\alpha,\beta}(P)$. For $0 \leq i \leq t$, define random variable $\mathbf{X}_i$ as the number of Q_i-subgraphs in $\mathbf{G}$. Obviously, $\mathbf{X}_0 = 1$ (with probability 1). For $1 \leq i \leq t$, let $\mathcal{L}(\mathbf{X}_i \mid \mathbf{X}_{i-1} = 1)$

denote the distribution of $\mathbf{X}_i$ conditioned on the event $\mathbf{X}_{i-1} = 1$. We prove Lemma 120 by showing the following

Lemma 121. *For all* $1 \leq i \leq t$, $\mathcal{L}(\mathbf{X}_i \mid \mathbf{X}_{i-1} = 1)$ *is asymptotically the Poisson distribution* Po(1). *In particular,*

$$\lim_{n\to\infty} \Pr[\,\mathbf{X}_i = 0 \mid \mathbf{X}_{i-1} = 1\,] = \lim_{n\to\infty} \Pr[\,\mathbf{X}_i = 1 \mid \mathbf{X}_{i-1} = 1\,] = \frac{1}{e}.$$

The first inequality of Lemma 120 follows immediately, as we have

$$\begin{aligned}
\liminf_{n\to\infty} \Pr[\,\mathbf{G} \text{ has no } P\text{-subgraph}\,] &\geq \liminf_{n\to\infty} \Pr[\,\mathbf{G} \text{ has no } Q_1\text{-subgraph}\,] \\
&= \liminf_{n\to\infty} \Pr[\,\mathbf{X}_1 = 0 \mid \mathbf{X}_0 = 1\,] \\
&= \frac{1}{e}.
\end{aligned}$$

For the second inequality, we have

$$\begin{aligned}
\liminf_{n\to\infty} \Pr[\,\mathbf{G} \text{ has a unique } P\text{-subgraph}\,] &= \liminf_{n\to\infty} \Pr[\,\mathbf{X}_t = 1\,] \\
&\geq \liminf_{n\to\infty} \Pr[\,\mathbf{X}_0 = \cdots = \mathbf{X}_t = 1\,] \\
&= \liminf_{n\to\infty} \prod_{1\leq i\leq t} \Pr[\,\mathbf{X}_i = 1 \mid \mathbf{X}_{i-1} = 1\,] \\
&= \frac{1}{e^t} \;\geq\; \frac{1}{e^{|E(P)|}}.
\end{aligned}$$

In the remainder of this appendix we give the proof of Lemma 121. We will use the following result on Poisson approximation. Before stating it, recall that the *total variation distance* $d_{TV}(\mathbf{X}, \mathbf{Y})$ between two random variables $\mathbf{X}$ and $\mathbf{Y}$ with values in the same set (in particular, real-valued variables) is given by

$$d_{TV}(\mathbf{X}, \mathbf{Y}) := \sup_A |\Pr[\,\mathbf{X} \in A\,] - \Pr[\,\mathbf{Y} \in A\,]|.$$

$\{0,1\}$-valued random variables $\mathbf{I}_1, \ldots, \mathbf{I}_m$ on the same probability space are *positively related* if for any given $i \in [m]$ one can find $\{0,1\}$-valued random variables $\mathbf{J}_{ji}$ $(j \neq i)$ such that $\mathbf{J}_{ji} \geq \mathbf{I}_j$ and this tuple is distributed identically with the tuple $\mathbf{I}_j$ $(j \neq i)$ conditioned by the event $\mathbf{I}_i = 1$.

Lemma 122 (Theorem 6.24 in [47])**.** *Suppose $\mathbf{I}_1, \ldots, \mathbf{I}_m$ are positively related $\{0,1\}$-valued random variables, and let $\mathbf{k} := \sum_i \mathbf{I}_i$. Then*

$$d_{TV}(\mathbf{k}, \mathrm{Po}(\mathrm{E}[\mathbf{k}])) \leq \frac{\mathrm{Var}[\mathbf{k}]}{\mathrm{E}[\mathbf{k}]} - 1 + 2 \max_i \mathrm{E}[\mathbf{I}_i].$$

Proof of Lemma 121. Fix $i \in \{1, \ldots, t\}$ and let $Q := Q_i$ and $Q' := Q_{i-1}$ and $\mathbf{X} := \mathbf{X}_i$ and $\mathbf{X}' := \mathbf{X}_{i-1}$. To show that $\mathcal{L}(\mathbf{X} \mid \mathbf{X}' = 1)$ is asymptotically Po(1), we would like to sample $\mathbf{G}$ conditioned on $\mathbf{X}' = 1$ (i.e. the event that $\mathbf{G}$ contains a unique Q'-subgraph). However, it will be convenient to condition on the entire $V(Q')$-colored part of $\mathbf{G}$ (i.e. the induced subgraph of G on the vertices which map to $V(Q')$ under the vertex-coloring of $\mathbf{G}$). We shall therefore fix an arbitrary $V(Q')$-colored graph G' such that

- G' equals the Q'-colored part of G for some G in the support of $\mathbf{G}$, and
- G' contains a unique Q'-subgraph, which we will denote by H'.

We denote by $\mathbf{G}|G'$ the random graph $\mathbf{G}$ conditioned on the event that the $V(Q')$-colored part of $\mathbf{G}$ equals G'. Note that $\mathbf{G}|G'$ is a product distribution on the unrestricted edges.

Let $\mathcal{Q}$ $(= \mathcal{Q}(H'))$ be the set of potential Q-subgraphs which extend H'. For each $H \in \mathcal{Q}$, let $\mathbf{I}_H$ be the indicator variable for the event that $\mathbf{G}|G'$ contains H. These random variables are positively related: just let $\mathbf{J}_H$ be the characteristic function of the event that $\mathbf{G}$ contains $E(H') \setminus E(H)$. Let $\mathbf{k} := \sum_{H \in \mathcal{Q}} \mathbf{I}_H$. We will show that $\mathbf{k}$ is asymptotically Po(1) using Lemma 122. Since the event $\{\mathbf{X}' = 1\}$ is the disjoint union of events $\{G'$ is the Q'-colored part of $\mathbf{G}\}$ over all G', it follows that $\mathcal{L}(\mathbf{X} \mid \mathbf{X}' = 1)$ is asymptotically Po(1) by the convexity of d_{TV}.

We will now calculate the expectation of $\mathbf{k}$. First, we have

$$|\mathcal{Q}| = \prod_{v \in V(Q) \setminus V(Q')} \lfloor n^{\alpha(v)} \rfloor = (1 - o(1)) n^{\alpha(Q) - \alpha(Q')}.$$

For each $H \in \mathcal{Q}$, we have

$$\mathrm{E}[\mathbf{I}_H] = n^{-\beta(Q) + \beta(Q')}.$$

(Note for the record that this is $o(1)$ since $\beta(Q') < \beta(Q)$ by the fact that $Q_0, \ldots, Q_t$ is an (α, β)-grading.) Therefore,

$$\mathrm{E}[\mathbf{k}] = (1 - o(1)) n^{\alpha(Q) - \alpha(Q')} n^{-\beta(Q) + \beta(Q')} = 1 - o(1),$$

using the fact that $\alpha(Q) = \beta(Q)$ and $\alpha(Q') = \beta(Q')$. In particular, $|\mathrm{E}[\mathbf{k}] - \mathrm{E}[\mathbf{k}]^2| \leq o(1)$.

We next calculate $\mathrm{Var}[\mathbf{k}]$. For $H, K \in \mathcal{Q}$, let $U := \chi(V(H) \cap V(K))$ be the set of P-colors of vertices in the intersection of $V(H)$ and $V(K)$. Note that

$$V(Q') \subseteq U \subseteq V(Q).$$

Thus,

$$\mathrm{E}[\, \mathbf{I}_H \mathbf{I}_K \,] = n^{-2\beta(Q) + \beta(Q') + \beta(U)}$$

where $\beta(U) := \sum_{e \in E(P) \cap \binom{U}{2}} \beta(e)$. For all $V(Q') \subseteq U \subseteq V(Q)$, we have

$$\#\{(H, K) \in \mathcal{Q} \times \mathcal{Q} : \chi(V(H) \cap V(K)) = U\} = (1 - o(1)) n^{2\alpha(Q) - \alpha(Q') - \alpha(U)}.$$

Therefore,

$$\begin{aligned}\mathrm{Var}[\mathbf{k}] &= \sum_{H,K\in\mathcal{Q}\,:\,H\neq K} \mathrm{E}[\,\mathbf{I}_H\mathbf{I}_K\,] + (\mathrm{E}[\mathbf{k}] - \mathrm{E}[\mathbf{k}]^2)\\ &= \sum_{U\,:\,V(Q')\subseteq U\subset V(Q)} (1-o(1))n^{2\alpha(Q)-\alpha(Q')-\alpha(U)}n^{-2\beta(Q)+\beta(Q')+\beta(U)} \pm o(1)\\ &= \sum_{U\,:\,V(Q')\subseteq U\subset V(Q)} (1-o(1))n^{\beta(U)-\alpha(U)} \pm o(1).\end{aligned}$$

Note that $\beta(U) < \alpha(U)$ for all $V(Q') \subset U \subset V(Q)$ (otherwise, letting R denote the induced subgraph of Q on U, we would have $\alpha(R) = \beta(R)$, contradicting the fact that $Q_0, \ldots, Q_t$ is an (α, β)-grading). It follows that

$$\mathrm{Var}[\mathbf{k}] = 1 \pm o(1).$$

Plugging the bounds $\mathrm{E}[\mathbf{k}] = 1 - o(1)$ and $\mathrm{Var}[\mathbf{k}] = 1 \pm o(1)$ and $\mathrm{E}[\mathbf{I}_H] = o(1)$ into Lemma 122, we have

$$d_{TV}\big(\mathcal{L}(\mathbf{k}), \mathrm{Po}(\mu)\big) \leq \frac{\mathrm{Var}[\mathbf{k}]}{\mathrm{E}[\mathbf{k}]} - 1 + 2\max_{H\in\mathcal{Q}} \mathrm{E}[\mathbf{I}_H] = o(1).$$

Finally, since $d_{TV}\big(\mathrm{Po}(1), \mathrm{Po}(1-o(1))\big) = o(1)$, we conclude that $\mathbf{k}$ is asymptotically Po(1), which completes the proof. □

4.2.6 Parameters $\kappa(P)$ and $\kappa_{\mathsf{col}}(P)$

We now introduce the parameters $\kappa(P)$ and $\kappa_{\mathsf{col}}(P)$ which figure in our lower bounds. The definitions, which might appear unmotivated at first glance, are derived from the lower bound technique of [73], which we explain in the next section.

Definition 123. ***(Union sequences and hitting sets)*** *A* union sequence *for P is a sequence $Q_1, \ldots, Q_t$ of subgraphs of P such that $Q_t = P$ and for all $1 \leq k \leq t$, either Q_k is a single vertex or a single edge or $Q_k = Q_i \cup Q_j$ for some $1 \leq i < j < k$. A* hitting set for

union sequences *(or* hitting set *for short) is a set $\mathcal{H}$ of subgraphs of P such that $\mathcal{H}$ contains at least one element from every union sequence.*

Definition 124. ***(Parameters $\kappa(P)$, $\kappa_{\alpha,\beta}(P)$ and $\kappa_{\mathsf{col}}(P)$)***

1. *If P is balanced, then $\kappa(P)$ is defined by*

$$\kappa(P) := \min_{\text{union seq. } Q_1,\ldots,Q_t} \max_{i\in[t]} v(Q_i) - \theta(P)e(Q_i).$$

   For P which is not balanced, we define $\kappa(P) := \kappa(\mathsf{Bal}(P))$.

2. *For $(\alpha,\beta) \in \theta_{\mathsf{col}}(P)$, let*

$$\kappa_{\alpha,\beta}(P) := \min_{\text{union seq. } Q_1,\ldots,Q_t} \max_{i\in[t]} \alpha(Q_i) - \beta(Q_i).$$

3. *Let $\kappa_{\mathsf{col}}(P) := \max_{(\alpha,\beta)\in\theta_{\mathsf{col}}(P)} \kappa_{\alpha,\beta}(P)$ (the maximum exists since $\kappa_{\alpha,\beta}(P)$, viewed as a function of α,β for a fixed P, is continuous).*

Remark 125. *Later on we will see that in this definition we could restrict ourselves to threshold pairs with $\alpha \equiv 1$ (Corollary 145). But since arbitrary threshold pairs appear quite naturally in our lower bound proofs in Section 4.4.2, we prefer to give this more general definition at once.*

The next lemma is key to linking our upper and lower bounds on the average-case AC^0 complexity of $\textsc{Subgraph}(P)$.

Lemma 126. (Minimax principle for $\kappa(P)$ and $\kappa_{\alpha,\beta}(P)$)

1. *If P is balanced, then*

$$\kappa(P) = \max_{\mathcal{H}} \min_{Q\in\mathcal{H}} v(Q) - \theta(P)e(Q),$$

   where $\mathcal{H}$ ranges over hitting sets for P.

2. *Similarly,* $\kappa_{\alpha,\beta}(P) = \max_{\mathcal{H}} \min_{Q \in \mathcal{H}} \alpha(Q) - \beta(Q)$ *for all* $(\alpha, \beta) \in \theta_{\mathsf{col}}(P)$.

Proof. The argument is the same for (1) and (2). Let $f(Q) := v(Q) - \theta(P)e(Q)$ (the proof works for any real-valued objective function). First, we will prove that $\max_{\mathcal{H}} \min_{Q \in \mathcal{H}} f(Q) \leq \kappa(P)$. Since $\mathcal{H}$ is a hitting set, for any union sequence $\{Q_i\}$, there exists some $Q_i \in \mathcal{H}$. It follows that $\min_{Q \in \mathcal{H}} f(Q) \leq \max_i f(Q_i)$, and thus $\min_{Q \in \mathcal{H}} f(Q) \leq \kappa(P)$ as $\{Q_i\}$ is taken arbitrarily.

On the other hand, let us prove $\kappa(P) \leq \max_{\mathcal{H}} \min_{Q \in \mathcal{H}} f(Q)$. Enumerate all union sequences $\{Q_i^{(j)}\}$, $j = 1, 2, \ldots$ (each $\{Q_i^{(j)}\}$ is a finite sequence). For each j, take a subgraph $S^{(j)}$ in $\{Q_i^{(j)}\}$ with maximal $f(Q_i^{(j)})$. Let $\mathcal{S} = \{S^{(1)}, S^{(2)}, \ldots\}$. It is easily seen that $\mathcal{S}$ is a hitting set, as every union sequence has some element in it. By definition,

$$\max_{\mathcal{H}} \min_{Q \in \mathcal{H}} f(Q) \geq \min_{S^{(j)} \in \mathcal{S}} f(S^{(j)}) = \min_j \max_i f(Q_i^{(j)}) = \kappa(P),$$

which completes the proof. □

4.3 Average-Case AC^0 Complexity

In this section, we prove Theorem 110 ($\kappa(P) \leq C_{\mathsf{ave}}(P) \leq 2\kappa(P) + O(1)$), which gives a combinatorial characterization of the AC^0-complexity of $\textsc{Subgraph}_{\mathsf{ave}}(P)$ up to a quadratic factor. More generally, we prove a family of average-case lower and upper bounds for the average-case colorful P-subgraph isomorphism problem:

Theorem 127. *For every pattern* P *and* $(\alpha, \beta) \in \theta_{\mathsf{col}}(P)$*, the average-case* AC^0*-complexity of* $\textsc{Subgraph}_{\mathsf{col}}(P)$ *on the* P*-colored random graph* $G_{\alpha,\beta}(n)$ *is between* $n^{\kappa_{\alpha,\beta}(P) - o(1)}$ *and* $n^{2\kappa_{\alpha,\beta}(P) + O(1)}$.

Rather than proving Theorem 110 and Theorem 127 separately, to avoid redundancy we present a proof of the latter only. For balanced P the proof of Theorem 110 looks exactly like the proof of Theorem 127 in the special case where $\alpha \equiv 1$ and $\beta \equiv \theta(P)$ (see Remark

119). The general case is reduced to the balanced one since for an arbitrary pattern P we have $\kappa(P) = \kappa(\mathsf{Bal}(P))$ (by definition of $\kappa(P)$) and $C_{\mathsf{ave}}(P) = C_{\mathsf{ave}}(\mathsf{Bal}(P))$ (by Lemma 116(2)).

Theorem 127 also plays a key role in our other main result, Theorem 109 (the worst-case lower bound $C_{\mathsf{col}}(P) \geq \Omega(tw(P)/\log tw(P))$). Since the worst-case AC^0-complexity of $\textsc{Subgraph}_{\mathsf{col}}(P)$ is lower-bounded by the average-case AC^0-complexity of $\textsc{Subgraph}_{\mathsf{col}}(P)$ on $G_{\alpha,\beta}(n)$ for every $(\alpha,\beta) \in \theta_{\mathsf{col}}(P)$ (and $G_{\alpha,\beta}(n)$ is supported on graphs with $n^{1+o(1)}$ vertices), Theorem 127 directly implies:

Corollary 128. $C_{\mathsf{col}}(P) \geq \kappa_{\mathsf{col}}(P)$.

In Section 4.4, we will show that $\kappa_{\mathsf{col}}(P) \geq \Omega(tw(P)/\log tw(P))$; together with Corollary 128, this proves Theorem 109.

The remainder of this section contains the proof of Theorem 127. The $n^{2\kappa_{\alpha,\beta}(P)+O(1)}$ upper bound is proved in Section 4.3.1, followed by the $n^{\kappa_{\alpha,\beta}(P)-o(1)}$ lower bound in Section 4.3.2.

4.3.1 Upper Bound

Fix a pattern P and a threshold pair $(\alpha,\beta) \in \theta_{\mathsf{col}}(P)$. For a P-colored graph G and $Q \subseteq P$, let $\mathrm{sub}(Q,G)$ denote the number of (colored) Q-subgraphs of G. We write $\mathbf{G}$ for the P-colored random graph $G_{\alpha,\beta}(n)$. Note that $\mathrm{E}[\,\mathrm{sub}(Q,\mathbf{G})\,] \leq n^{\alpha(Q)-\beta(Q)}$.

We begin with a sketch of the algorithm that determines, correctly with high probability, whether or not $\mathbf{G}$ contains a P-subgraph. We first establish that, in typical instances of $\mathbf{G}$, the number $\mathrm{sub}(Q,\mathbf{G})$ is highly concentrated around its mean (in particular, we have $\mathrm{sub}(Q,\mathbf{G}) \leq n^{\alpha(Q)-\beta(Q)+1}$ for all $Q \subseteq P$ with very high probability). For such typical instances of $\mathbf{G}$, the basic idea is "search" for a P-subgraph by exhaustively computing the lists L_i of all Q_i-subgraphs in $\mathbf{G}$ for all elements Q_i of an optimal (that is, with $\max_i \alpha(Q_i) - \beta(Q_i) = \kappa_{\alpha,\beta}(P)$) union sequence $Q_1,\ldots,Q_t = P$. Each list L_i will have size at most

$n^{\alpha(Q_i)-\beta(Q_i)+1}$; for $Q_k = Q_i \cup Q_j$ in the union sequence, the list L_k is obtained by merging lists L_i and L_j at a cost of roughly $|L_i| \cdot |L_j|$, which is at most $n^{2\kappa_{\alpha,\beta}(P)+2}$. Non-emptiness of the final list L_t is equivalent to $\mathbf{G}$ containing a P-subgraph. To implement this algorithm by AC^0 circuits, we first invert the role of randomness by designing a random AC^0 circuit which solves the problem in the worst-case over the set of "typical instances"; lists L_i are constructed by random hashing.

We now turn to the details of the construction. Let $\mathcal{G}_{\alpha,\beta}(n)$ denote the support of $\mathbf{G}$, that is, the set of P-colored graphs with vertex set $\{(v,i) : v \in V(P),\ 1 \le i \le \lfloor n^{\alpha(v)} \rfloor\}$ and the vertex-coloring $(v,i) \mapsto v$. Let also

$$\mathcal{G}'_{\alpha,\beta}(n) := \{G \in \mathcal{G}_{\alpha,\beta}(n) : \mathrm{sub}(Q,G) \le n^{\alpha(Q)-\beta(Q)+1} \text{ for all } Q \subseteq P\}.$$

The next lemma says that $\mathbf{G}$ is extremely unlikely to contain significantly more than $n^{\alpha(Q)-\beta(Q)}$ Q-subgraphs for any $Q \subseteq P$. It is proved by a straightforward application of Markov's inequality.

Lemma 129. $\Pr[\,\mathbf{G} \notin \mathcal{G}'_{\alpha,\beta}(n)\,] = o(1)$.

We wish to construct a *deterministic* AC^0-circuit C which solves $\textsc{Subgraph}_{\mathsf{col}}(P)$ correctly on $\mathbf{G}$ with probability $1 - o(1)$. We will invert the role of randomness and instead construct a *random* AC^0-circuit $\mathbf{C}$ which solves $\textsc{Subgraph}_{\mathsf{col}}(P)$ correctly with probability $1 - o(1)$ on *every* $G \in \mathcal{G}'_{\alpha,\beta}(n)$. That is, we will show

Lemma 130. *There exists a random* AC^0 *circuit* $\mathbf{C}$ *of size* $n^{2\kappa_{\alpha,\beta}(P)+O(1)}$ *and depth*[6] $O(e(P))$ *such that for every* $G \in \mathcal{G}'_{\alpha,\beta}(n)$,

$$\Pr[\,\mathbf{C}(G) = 1 \Leftrightarrow \mathrm{sub}(P,G) \ge 1\,] = 1 - o(1).$$

6. In fact, the depth is linear in the *height* of the optimal union sequence, where the height is defined as the length of the longest path from the root to a leaf in the directed acyclic graph induced by a union sequence.

The upper bound of Theorem 127 follows as a corollary of Lemmas 129 and 130.

Proposition 131. *There exists a* AC^0 *circuit* C *of size* $n^{2\kappa_{\alpha,\beta}(P)+O(1)}$ *such that*

$$\Pr[\,\mathsf{C}(\mathbf{G}) = 1 \Leftrightarrow \mathrm{sub}(P, \mathbf{G}) \geq 1\,] = 1 - o(1).$$

Proof. Lemmas 129 and 130 imply that $\Pr[\,\mathbf{C}(\mathbf{G}) = 1 \Leftrightarrow \mathrm{sub}(P, \mathbf{G}) \geq 1\,] = 1 - o(1)$. Now Proposition 131 follows by a straightforward application of Yao's Principle [85]. □

The random circuit $\mathbf{C}$.

It remains to define the randomized AC^0-algorithm solving $\textsc{Subgraph}_{\mathsf{col}}(P)$ with high probability on every $G \in \mathcal{G}'_{\alpha,\beta}(n)$. We first describe the algorithm informally. We then check that this algorithm can be implemented by circuits of size $n^{2\kappa_{\alpha,\beta}(P)+O(1)}$ and depth $O(e(P))$.

By definition of $\kappa_{\alpha,\beta}(P)$, there exists a union sequence $Q_1, \ldots, Q_t$ with $Q_t = P$ such that $\kappa_{\alpha,\beta}(P) = \max_{i\in[t]} \alpha(Q_i) - \beta(Q_i)$. On a high level, the idea behind the algorithm was already sketched above: given a graph $G \in \mathcal{G}'_{\alpha,\beta}(n)$ (the input), we will compute a sequence $L_1, \ldots, L_t$ of lists, where L_k contains all of the Q_k-subgraphs of G (with high probability). Many entries in L_k will be *blank* (signified by $\emptyset$); by construction, every non-blank entry of L_k will contain the description of a Q_k-subgraph of G (as a string of length $\alpha(Q_k)\log n$). Blank and non-blank entries will in general be interleaved.

Some notation: we write ℓ_k for the number of entries in the list L_k. For $a \in [\ell_k]$, we write $L_k(a)$ for the contents of the ath entry in L_k (either $\emptyset$ or a Q_k-subgraph of G). We say that L_k is *good* (with respect to G and the randomness of the algorithm) if L_k contains all Q_k-subgraphs of G exactly once.

Lists $L_1, \ldots, L_t$ are computed, in order, as follows. For $k \in [t]$, assume that $L_1, \ldots, L_{k-1}$ have been computed and are good.

If Q_k is a single vertex v, the construction is trivial: L_k simply lists all (v, i), $1 \leq i \leq \lfloor n^{\alpha(v)} \rfloor$.

In the case that Q_k is a single edge of P, let L_k have $\ell_k := n^{\alpha(Q_k)}$ entries, indexed by the potential Q_k-subgraphs of G. For $a \in [\ell_k]$, the ath entry $L_k(a)$ will contain the ath potential Q_k-subgraph iff it is a Q_k subgraph of G; otherwise $L_k(a)$ is blank. Clearly L_k is good.

If Q_k is not a single edge, then by the definition of union sequence, $Q_k = Q_i \cup Q_j$ for some $1 \le i < j < k$. We compute L_k in three steps as follows.

1. Let M_k be the $\ell_i \times \ell_j$ array where, for $a \in [\ell_i]$ and $b \in [\ell_j]$,

$$M_k(a,b) := \begin{cases} L_i(a) \cup L_j(b) & \text{if non-blank } L_i(a), L_j(b) \text{ are consistent on } V(Q_i) \cap V(Q_j), \\ \emptyset & \text{otherwise.} \end{cases}$$

   (Note that, since L_i and L_j are good, M_k contains each Q_k-subgraph of G exactly once. That is, M_k satisfies the "good" condition that we want for L_k.)

2. We hash M_k down to a smaller number of entries to obtain the list L_k. Let $\mathrm{Supp}(M_k) \subseteq [\ell_i] \times [\ell_j]$ denote the set of nonempty entries of M_k. Let $m_k := \lceil n^{\alpha(Q_k)-\beta(Q_k)+1} \rceil$ and note that $m_k \ge \#\{Q_k\text{-subgraphs of } G\} = |\mathrm{Supp}(M_k)|$. Let $\mathbf{h}_k$ be a uniform random function

$$\mathbf{h}_k : [\ell_i] \times [\ell_j] \to [m_k].$$

   (Restricted to the $\le m_k$ nonempty entries of M_k, this gives a uniform random packing of $\le m_k$ balls into m_k bins.)

3. Let $\ell_k := m_k \ln m_k$. Indexing entries of L_k by pairs $(p,q) \in [m_k] \times [\ln m_k]$ (rather than elements of $[\ell_k]$), let

$$L_k(p,q) := \begin{cases} \text{the } q\text{th element of } \mathbf{h}_k^{-1}(p) \cap \mathrm{Supp}(M_k) & \text{if } |\mathbf{h}_k^{-1}(p) \cap \mathrm{Supp}(M_k)| \ge q, \\ \emptyset & \text{otherwise.} \end{cases}$$

Note that L_k is good if and only if

$$\bigwedge_{p\in[m_k]} |\mathbf{h}_k^{-1}(p) \cap \mathrm{Supp}(M_k)| \le \ln m_k. \tag{4.3}$$

After computing the final list L_t, the algorithm outputs 1 iff L_t has non-blank entries. Note that the output of the algorithm will be correct provided L_t is good.

To analyze the success probability of the algorithm, note the following elementary fact about balls-into-bins, established by a simple union bound.[7]

$$\begin{cases} \textit{For any } \widetilde{m} \le m\textit{, the maximum load of a random function of } \widetilde{m} \textit{ balls to } m \textit{ bins is} \\ \le \ln m \text{ with probability } \ge 1 - 1/m. \end{cases}$$

From this fact, we have

$$\begin{aligned} \Pr_{\mathbf{h}_k}[\, L_k \text{ is not good} \mid L_1, \ldots, L_{k-1} \text{ are good}] &\le \Pr_{\mathbf{h}_k}[\bigvee_{p\in[m_k]} |\mathbf{h}_k^{-1}(p) \cap \mathrm{Supp}(M_k)| > \ln m_k\,] \\ &\le \frac{1}{m_k} \;\le\; \frac{1}{n}. \end{aligned}$$

It follows that

$$\begin{aligned} \Pr_{\mathbf{h}_1,\ldots,\mathbf{h}_t}[\,\text{erroneous output}\,] &= \sum_{k\in[t]} \Pr_{\mathbf{h}_1,\ldots,\mathbf{h}_k}[\,L_k \text{ is not good}, L_1, \ldots, L_{k-1} \text{ are good}\,] \\ &\le \sum_{k\in[t]} \Pr_{\mathbf{h}_1,\ldots,\mathbf{h}_k}[\,L_k \text{ is not good} \mid L_1, \ldots, L_{k-1} \text{ are good}] \\ &\le tn^{-1} \;\le\; o(1). \end{aligned}$$

Therefore, the algorithm correctly solves $\textsc{Subgraph}_{\mathsf{col}}(P)$ with high probability for every $G \in \mathcal{G}'_{\alpha,\beta}(P)$.

7. A tighter analysis than we require shows that the maximum load is $\le 3\ln m/\ln\ln m$ with probability $\ge 1 - 1/m$.

It remains to show that this algorithm can be implemented by a random circuit $\mathbf{C}$ of size $n^{2\kappa_{\alpha,\beta}(P)+O(1)}$ and depth $O(e(P))$. We will make an additional assumption about the random functions $\mathbf{h}_1, \ldots, \mathbf{h}_t$:

$$\left|\mathbf{h}_k^{-1}(p)\right| \leq \frac{2\ell_i\ell_j}{m_k} \quad \text{for all } k \in [t] \text{ and } p \in [m_k]. \tag{4.4}$$

That is, $|\mathbf{h}_k^{-1}(p)|$ is at most twice its expectation for all k and p. By Chernoff and union bounds, (4.4) holds with probability $1 - \exp(-n^{\Omega(1)})$. So even with this assumption, the error probability of the circuits we describe remains $o(1)$.

Let us now *fix*[8] any particular hash functions $h_1, \ldots, h_t$ such that (4.4) holds. We will design constant-depth circuits of size $n^{2\kappa_{\alpha,\beta}(P)+O(1)}$ computing the lists that correspond to our particular choice of $h_1, \ldots, h_t$ whenever all these lists are good, that is satisfy (4.3). (If (4.3) fails for at least one k, this is the error case and we do not care what the algorithm does.)

We describe the sub-circuit which computes the list L_k given lists $L_1, \ldots, L_{k-1}$. If Q_k is a single vertex, this is trivial, and in the case when Q_k is a single edge, the list L_k is clearly computable by a depth-2 circuit of size $\widetilde{O}(n^{\alpha(Q_k)})$ (the $\widetilde{O}()$ coming from the fact that it takes $\alpha(Q_k)\log n$ gates to encode each entry of L_k). In the case that $Q_k = Q_i \cup Q_j$, first note that we can compute the array M_k by a circuit of size $\widetilde{O}(n\ell_i\ell_j)$ and depth $O(1)$ (sitting on top of the sub-circuits which compute lists L_i and L_j); this is because checking that $L_i(a)$ and $L_j(b)$ agree on all vertices of $V(Q_i) \cap V(Q_j)$ requires only $O(n)$ size and depth 2. Having computed M_k, computing the entries $L_k(p, q)$ requires finding the qth element in the 0-1 string that represents the characteristic function of $h_k^{-1}(p) \cap \mathrm{Supp}(M_k)$ within the known set $h_k^{-1}(p)$. This string has length $|h_k^{-1}(p)| \leq O\left(\frac{\ell_i\ell_j}{m_k}\right)$ and, by (4.3), has at most $\ln m_k = O(\log n)$ 1 entries. Hence, by [43, Theorem 6] applied with $n := |h_k^{-1}(p)|$, $k := \ln m_k$, $\gamma := 2$ and $m := 3$, this can be done by a constant-depth circuit of size $n^{o(1)}\frac{\ell_i\ell_j}{m_k}$. As there are altogether

8. It is important for our argument that the random functions $\mathbf{h}_1, \ldots, \mathbf{h}_t$ are generated in advance and that we do not attempt to make our construction uniform in $\mathbf{h}_1, \ldots, \mathbf{h}_t$.

$\ell_k \leq \widetilde{O}(m_k)$ pairs (p, q), we get a constant-depth circuit which computes L_k (given L_i and L_j) with total size $n^{o(1)}\ell_i\ell_j \leq n^{\alpha(Q_i)-\beta(Q_i)+\alpha(Q_j)-\beta(Q_j)+O(1)} \leq n^{2\kappa_{\alpha,\beta}(P)+O(1)}$.

After computing all lists $L_1, \ldots, L_t$, we have a circuit of size $n^{2\kappa_{\alpha,\beta}(P)+O(1)}$. Finally, note that the depth of this circuit will be $O(d)$ where d is the height of the poset where $i, j \prec k$ iff $Q_k = Q_i \cup Q_j$ for $i, j < k$. Clearly, $d \leq e(P)$ as long as all graphs in the sequence are pairwise distinct.

4.3.2 Lower Bound

This subsection gives the proof of the lower bound in Theorem 127 (the average-case AC^0-complexity of $\textsc{Subgraph}_{\mathsf{col}}(P)$ on $G_{\alpha,\beta}(n)$ is at least $n^{\kappa_{\alpha,\beta}(P)-o(1)}$). The argument closely follows the technique of [73, 74].

It will be convenient to work with an alternative characterization of AC^0 as boolean circuits with fan-in 2. We distinguish between "type-I" and "type-II" AC^0 circuits as follows.

1. polynomial-size constant-depth $\{\mathrm{AND}_\infty, \mathrm{OR}_\infty, \mathrm{NOT}\}$-circuits with unbounded fan-in (this is the standard definition of AC^0),

2. polynomial-size $\{\mathrm{AND}_2, \mathrm{OR}_2, \mathrm{NOT}\}$-circuits with fan-in 2 and arbitrary depth, but $O(1)$ alternations between AND and OR gates (where w.l.o.g. NOT gates are on the bottom level).

The conversion from type-I to type-II replaces each AND_∞ (resp. OR_∞) gate with a binary tree of AND_2 (resp. OR_2) gates. This conversion can result in a quadratic blow-up in size (= number of gates), as a type-I circuit with g gates and $w \leq O(g^2)$ wires becomes a type-II circuit with $O(w)$ gates. Therefore, a lower bound of S on the size of type-II circuits implies a lower bound of $\Omega(\sqrt{S})$ on the size of type-I circuits.

We will first prove an $n^{\kappa_{\alpha,\beta}(P)-o(1)}$ lower bound on the size of type-II circuits solving $\textsc{Subgraph}_{\mathsf{col}}(P)$ in the average-case on $G_{\alpha,\beta}(n)$. This implies a weaker $n^{\kappa_{\alpha,\beta}(P)/2-o(1)}$

lower bound for type-I circuits. The stronger $n^{\kappa_{\alpha,\beta}(P)-o(1)}$ lower bound for type-I circuits is shown by an additional argument in Section 4.3.4.

Let $\mathcal{G} := \mathcal{G}_{\alpha,\beta}(n)$ denote the support of the random P-colored graph $G_{\alpha,\beta}(n)$, that is, the set of P-colored graphs with vertex set $\{(v,i) : v \in V(P), 1 \le i \le \lfloor n^{\alpha(v)} \rfloor\}$ and the vertex-coloring $(v,i) \mapsto v$. (The following definitions are also used in the uncolored setting, where $\mathcal{G}$ is the set of graphs with vertex $\{1,\dots,n\}$.) We identify $\mathcal{G}$ with the hypercube $\{0,1\}^{\mathcal{E}}$ where $\mathcal{E}$ is the set of potential edges $\{(v,i),(w,j)\}$ with $\{v,w\} \in V(P)$.

Definition 132. *Let f be any function with domain $\mathcal{G}$ (and arbitrary range), and let H be any graph in $\mathcal{G}$. The* f-sensitive subgraph *of H, denoted $\mathsf{Sens}(f,H)$, is defined as the unique minimal subgraph $S \subseteq H$ such that $f(H') = f(H' \cap S)$ for every $H' \subseteq H$. We say that f is* sensitive over *H if $\mathsf{Sens}(f,H) = H$.*

For all f and H, observe that

$$f \text{ is sensitive over } \mathsf{Sens}(f,H) \text{ (i.e. } \mathsf{Sens}(f,\mathsf{Sens}(f,H)) = \mathsf{Sens}(f,H)), \tag{4.5}$$

$$\text{if } f : \mathcal{G} \to \{0,1\} \text{ is the AND or OR of } f_1, f_2, \text{ then } \mathsf{Sens}(f,H) \subseteq \mathsf{Sens}(f_1,H) \cup \mathsf{Sens}(f_2,H). \tag{4.6}$$

We say that a single-output boolean circuit whose variables encode potential edges in a graph is *sensitive over H* if its output function is so.

Lemma 133. *Let C be a boolean circuit with fan-in 2, computing a function $\mathcal{G} \to \{0,1\}$, such that C is sensitive over some nonempty graph H. Then there exists a union sequence $H_1,\dots,H_t = H$ and a sequence $\mathsf{C}_1,\dots,\mathsf{C}_t$ of sub-circuits of C such that C_i is sensitive over H_i for all $i \in \{1,\dots,t\}$.*

Proof. We argue by induction on boolean circuits with fan-in 2. In the base case, C is a variable (corresponding to a possible edge). The assumption that C is sensitive over H implies that H is a single edge. Therefore, H itself is a union sequence of length 1 which satisfies the condition of the lemma.

For the induction step, note that if $\mathsf{C} = \mathrm{NOT}(\mathsf{C}')$, then C' is sensitive over H; therefore, the lemma holds by the induction hypothesis for C'. Finally, suppose C is the AND or OR of sub-circuits C_1 and C_2. If C_1 or C_2 is sensitive over H, then appealing to the induction hypothesis, we are done. So we will assume that neither C_1 nor C_2 are sensitive over H. Let $H_i := \mathsf{Sens}(\mathsf{C}_i, H)$ for $i = 1, 2$. Then C_i is sensitive over H_i by observation (4.5). By observation (4.6),

$$H \;=\; \mathsf{Sens}(\mathsf{C}, H) \;\subseteq\; \mathsf{Sens}(\mathsf{C}_1, H) \cup \mathsf{Sens}(\mathsf{C}_2, H) \;=\; H_1 \cup H_2.$$

Hence $H = H_1 \cup H_2$. By the induction hypothesis, there exist union sequence $S_1, \ldots, S_s = H_1$ and $T_1, \ldots, T_t = H_2$ which satisfy the condition in the lemma with respect to C_1, H_1 and C_2, H_2 respectively. Then $S_1, \ldots, S_s, T_1, \ldots, T_t, H$ is a union sequence which satisfies the condition in the lemma with respect to C and H. □

Definition 134. *If f is a function with domain $\mathcal{G}$ (and arbitrary range) and G is a graph in $\mathcal{G}$, then let $f^{\cup G}$ denote the function $f^{\cup G}(H) := f(G \cup H)$.*

Note that if a boolean circuit C computes a function f on $\mathcal{G}$, then the circuit $\mathsf{C}^{\cup G}$ that substitutes 1 for variables corresponding to edges in G computes $f^{\cup G}$.

We now fix a pattern P and a threshold pair $(\alpha, \beta) \in \theta_{\mathsf{col}}(P)$. Without loss of generality, we assume that $\beta(e) > 0$ for all $e \in E(P)$ (otherwise we replace P with the subgraph with edge set $\{e \in E(P) : \beta(e) > 0\}$). We continue to write $\mathbf{G}$ for the P-colored random graph $G_{\alpha,\beta}(n)$. Independently, let $\mathbf{P} \in \mathcal{G}$ be a uniform random "planted" P-subgraph (viewed as element of $\mathcal{G}$). That is, after ignoring isolated vertices, $\mathbf{P}$ is the P-subgraph with vertex set $\{(v, i_v) : v \in V(P)\}$ where i_v is uniform random in $\{1, \ldots, \lfloor n^{\alpha(v)} \rfloor\}$. For a subgraph $Q \subseteq P$, let $\mathbf{Q} \in \mathcal{G}$ denote the corresponding subgraph of $\mathbf{P}$. (While $\mathbf{P}$ and $\mathbf{G}$ are independent, $\mathbf{Q}$ is completely determined by $\mathbf{P}$ and hence is also independent of $\mathbf{G}$.)

We next state two technical lemmas.

Lemma 135. *Suppose $f : \mathcal{G} \to \{0,1\}$ solves $\textsc{Subgraph}_{\mathsf{col}}(P)$ in the average-case on $\mathbf{G}$, that is,*

$$\Pr_{\mathbf{G}}[\, f(\mathbf{G}) = 1 \Leftrightarrow \mathbf{G} \text{ has a } P\text{-subgraph}\,] = 1 - o(1). \tag{4.7}$$

Then

$$\liminf_{n\to\infty} \Pr_{\mathbf{G},\mathbf{P}}[\, f^{\cup\mathbf{G}} \text{ is sensitive over } \mathbf{P}\,] > 0. \tag{4.8}$$

The full proof of Lemma 135 is included in the next subsection, i.e., Section 4.3.3. We give a brief outline here. First, we show that the event "$E(\mathbf{P}) \cap E(\mathbf{G}) = \emptyset$ and $\mathbf{P}$ is the unique P-subgraph in $\mathbf{G} \cup \mathbf{P}$" holds with probability bounded away from 0. (In particular, we have $\Pr[\, E(\mathbf{P}) \cap E(\mathbf{G}) = \emptyset\,] = 1 - o(1)$ and $\Pr[\, \mathbf{P}$ is the unique P-subgraph in $\mathbf{G} \cup \mathbf{P}\,] = \Omega(1)$, as shown in Lemma 138(3).) Note that, if f computes $\textsc{Subgraph}_{\mathsf{col}}(P)$ exactly, then this event implies that $f^{\cup\mathbf{G}}$ is the AND function over the edges of $\mathbf{P}$ (i.e. for all $Q \subseteq P$, $f(\mathbf{G} \cup \mathbf{Q}) = 1$ iff $Q = P$) and is therefore sensitive over $\mathbf{P}$. However, if f merely agrees with $\textsc{Subgraph}_{\mathsf{col}}(P)$ on $\mathbf{G}$ a.a.s. (as in the hypothesis of Lemma 135), then we require an additional argument bounding the total variation distance between $\mathbf{G}$ and $\mathbf{G} \cup \mathbf{Q}$ for subgraphs $Q \subseteq P$ (see Appendix 4.3.3 for details).[9]

The second technical lemma relies on Håstad's Switching Lemma [41] and its proof closely follows Proposition 3.11 of [74]. (The reader who wishes to skip the technical details is encouraged to jump ahead to Theorem 137.)

Lemma 136. *Suppose $f : \mathcal{G} \to \{0,1\}$ is AC^0-computable. Then for every $Q \subseteq P$,*

$$\Pr_{\mathbf{G},\mathbf{Q}}[\, f^{\cup\mathbf{G}} \text{ is sensitive over } \mathbf{Q}\,] \leq n^{-\alpha(Q)+\beta(Q)+o(1)}.$$

9. If we are content with *worst-case* lower bounds for $\textsc{Subgraph}_{\mathsf{col}}(P)$ (as opposed to *average-case* lower bounds with respect to $\mathbf{G}$), then it suffices to prove Lemma 135 in the special case where f computes $\textsc{Subgraph}_{\mathsf{col}}(P)$ exactly. For this, we merely require the bound $\Pr[\, E(\mathbf{P}) \cap E(\mathbf{G}) = \emptyset$ and $\mathbf{P}$ is the unique P-subgraph in $\mathbf{G} \cup \mathbf{P}\,] = \Omega(1)$ (Lemma 138(3)), thus avoiding most of the work in Appendix 4.3.3 that deals with total variation distance. An even easier way to establish the worst-case lower bound is to prove (4.8) for $\mathbf{G}$ in $G_{\alpha,\beta'}(G)$, where (α, β') is a sub-threshold pair, that is, $\beta'(e) = \beta(e) + \epsilon$ for an arbitrarily small $\epsilon > 0$, which follows from a union bound.

Proof. Let $\mathcal{E}$ be the set of potential edges of graphs in $\mathcal{G}$. We identify $\mathcal{G}$ with the hypercube $\{0,1\}^{\mathcal{E}}$ and we view f as a boolean function $\{0,1\}^{\mathcal{E}} \to \{0,1\}$. For $e \in \mathcal{E}$, let $\widehat{e} \in E(P)$ be the corresponding edge of P (under the $V(P)$-coloring $(v,i) \mapsto v$ of graphs in $\mathcal{G}$).

Let $\delta > 0$ be an arbitrarily small constant, which is independent of n but may depend on P, α, β. We generate a random restriction $\rho : \mathcal{E} \to \{0,1,\star\}$ where, independently for all $e \in \mathcal{E}$,

$$\Pr_{\rho}[\,\rho(e) = \star\,] = n^{-\beta(\widehat{e})-\delta}, \qquad \Pr_{\rho}[\,\rho(e) = 1 \mid \rho(e) \neq \star\,] = n^{-\beta(\widehat{e})}. \tag{4.9}$$

Let H_ρ denote the P-colored graph with edge set $E(H_\rho) = \rho^{-1}(\star)$. Note that H_ρ has distribution $G_{\alpha,\beta+\delta}(n)$. In particular, $\mathrm{E}_\rho[\,\mathrm{sub}(Q, H_\rho)\,] = n^{\alpha(Q)-\beta(Q)-\delta|E(Q)|}$. We assume δ is sufficiently small so that $\alpha(Q) - \beta(Q) - \delta|E(Q)| > 0$ (here we assume $\alpha(Q) - \beta(Q) > 0$ since otherwise the lemma is trivial). Using the lower-tail version of Janson's Inequality [47], it can be shown that

$$\Pr_{\rho}[\,\mathrm{sub}(Q, H_\rho) < \tfrac{1}{2} n^{\alpha(Q)-\beta(Q)-\delta|E(Q)|}\,] = n^{-\omega(1)}. \tag{4.10}$$

That is, with very high probability, H_ρ contains at least half the expected number of Q-subgraphs. (Since P and $\delta > 0$ are fixed, $n^{-\omega(1)}$ is $O(n^{-c})$ for every constant $c = c(P,\delta)$ which may depend on P and δ.)

Let L_ρ denote the subgraph of H_ρ with edge set

$$E(L_\rho) = \{e \in \rho^{-1}(\star) : \text{restricted function } f\restriction_\rho : \{0,1\}^{\rho^{-1}(\star)} \to \{0,1\} \text{ depends}^{10} \text{ on coordinate } e\}.$$

Note that in the language of graphs predominantly used in this proof, we have $E(L_\rho) = \mathsf{Sens}(f\restriction_\rho, \rho^{-1}(\{1,\star\}))$.

We now use the fact that f is computed by an AC^0-circuit (in particular, a type-I AC^0-circuit). A bottom-up depth-reduction argument using Håstad's Switching Lemma [41]

10. A function $g : \{0,1\}^I \to \{0,1\}$ *depends* on a coordinate $i \in I$ if there exists $x \in \{0,1\}^I$ such that $g(x) \neq g(x^{(i)})$ where $x^{(i)}$ is x with its ith coordinate flipped.

shows that

$$\Pr_{\rho}[\,|L_\rho| > n^\delta\,] \le S \cdot (5n^{-\delta/d}\delta\log n)^{\delta\log n} = n^{-\omega(1)} \tag{4.11}$$

where S and d are the size and depth of the circuit defining f. (This bound is $n^{-\omega(1)}$ since $S = n^{O(1)}$ and $d = O(1)$ and $\delta = \Omega(1)$.) Interested readers can refer to [73] or [74] for a proof of (4.11).

Let $\mathcal{A} = \mathcal{A}(\rho)$ denote the event that $\mathrm{sub}(Q, H_\rho) \ge \frac{1}{2}n^{\alpha(Q)-\beta(Q)-\delta|E(Q)|}$ and $|L_\rho| \le n^\delta$. Note that $\Pr[\,\neg\mathcal{A}\,] = n^{-\omega(1)}$ (by (4.10) and (4.11)) and

$$\mathcal{A} \implies \frac{\mathrm{sub}(Q, L_\rho)}{\mathrm{sub}(Q, H_\rho)} \le 2n^{-\alpha(Q)+\beta(Q)+\delta|E(Q)|+\delta|V(Q)|}.$$

We now generate a pair $(\mathbf{G'}, \mathbf{Q'})$ of random variables by the following two-step process.

- Independently, for all $e \in \mathcal{E}$, let

$$\Pr[\,e \in E(\mathbf{G'}) \mid \rho\,] = \begin{cases} \rho(e) & \text{if } \rho(e) \in \{0,1\}, \\ n^{-\beta(\widehat{e})} & \text{if } \rho(e) = \star. \end{cases} \tag{4.12}$$

- If $\mathcal{A}(\rho)$ holds (in particular, $\mathrm{sub}(Q, H_\rho) \neq \emptyset$), we let $\mathbf{Q'}$ be a uniform random Q-subgraph of H_ρ. Otherwise, we let $\mathbf{Q'} := \perp$ ($\perp$ stands for "undefined").

We claim that $(\mathbf{G'}, \mathbf{Q'})$ under the condition $\mathbf{Q'} \neq \perp$ is distributed identically with $(\mathbf{G}, \mathbf{Q})$.

Indeed, by inspecting the definitions (4.9) and (4.12), we see that $\Pr_{\mathbf{G'}}[\,e \in \mathbf{G'}\,] = n^{-\beta(\widehat{e})}$. This implies that, firstly, $\mathbf{G'} \sim G_{\alpha,\beta}(n)$ and, secondly, $\mathbf{G'}$ and H_ρ are independent. As $\mathbf{Q'}$ is a function of H_ρ, it is independent of $\mathbf{G'}$ as well. Finally, $\mathbf{Q'}$ conditioned by the event $\mathbf{Q'} \neq \perp$ is distributed identically with $\mathbf{Q}$ simply by symmetry.

This observation, along with the crucial fact $\Pr[\,\neg\mathcal{A}\,] = n^{-\omega(1)}$ mentioned above, implies that we can rephrase the inequality we are proving as

$$\Pr_{\rho,\mathbf{G'},\mathbf{Q'}}[\,f^{\cup\mathbf{G'}} \text{ is sensitive over } \mathbf{Q'} \mid \mathcal{A}(\rho)\,] \le n^{-\alpha(Q)+\beta(Q)+o(1)}.$$

We now fix an arbitrary ρ such that $\mathcal{A}$ holds.

Since $\mathbf{Q}' \subseteq H_\rho = \{e : \rho(e) = \star\}$, it follows from definitions that if $f^{\cup \mathbf{G}'}$ is sensitive over $\mathbf{Q}'$, then $\mathbf{Q}' \subseteq L_\rho$. We now have

$$\Pr_{\mathbf{G}',\mathbf{Q}'}[\, f^{\cup \mathbf{G}'} \text{ is sensitive over } \mathbf{Q}' \,] \leq 2n^{-\alpha(Q)+\beta(Q)+\delta|E(Q)|+\delta|V(Q)|}.$$

The lemma follows, since we can choose $\delta > 0$ arbitrarily small relative to $\frac{|E(Q)|}{\beta(Q)}$ and $\frac{|V(Q)|}{\alpha(Q)}$. □

Finally, the main result of this subsection (the lower bound of Theorem 127):

Theorem 137. *Suppose* C *is a type-II* AC^0 *circuit that solves* $\textsc{Subgraph}_{\mathsf{col}}(P)$ *in the average-case on* $G_{\alpha,\beta}(P)$*. Then* C *has size at least* $n^{\kappa_{\alpha,\beta}(P)-o(1)}$.

Proof. For contradiction, assume C has size $\leq n^{\kappa_{\alpha,\beta}(P)-\varepsilon}$ and d alternations for constants $\varepsilon, d > 0$ (independent of n). By Lemma 126, there exists a hitting set $\mathcal{H}$ for P such that $\kappa_{\alpha,\beta}(P) = \min_{Q\in\mathcal{H}} \alpha(Q)-\beta(Q)$. Note that every sub-circuit C' of C is computable by a type-I AC^0 circuit of depth d (by combining all adjacent AND and OR gates in C'). Therefore, by Lemma 136,

$$\begin{aligned}
& \Pr_{\mathbf{G},\mathbf{P}}[\, \bigvee_{Q\in\mathcal{H}} \mathsf{C}'^{\cup \mathbf{G}} \text{ is sensitive over } \mathbf{Q} \,] \\
\leq\ & \sum_{Q\in\mathcal{H}} \Pr_{\mathbf{G},\mathbf{Q}}[\, \mathsf{C}'^{\cup \mathbf{G}} \text{ is sensitive over } \mathbf{Q} \,] \\
\leq\ & \sum_{Q\in\mathcal{H}} n^{-\alpha(Q)+\beta(Q)+o(1)} \\
\leq\ & |\mathcal{H}| \cdot \max_{Q\in\mathcal{H}} n^{-\alpha(Q)+\beta(Q)+o(1)} \\
=\ & n^{-\kappa_{\alpha,\beta}(P)+o(1)} \quad (\text{since } |\mathcal{H}| \leq 2^{|E(P)|} = n^{o(1)}).
\end{aligned}$$

Taking a union bound over the $\leq n^{\kappa_{\alpha,\beta}(P)-\varepsilon}$ sub-circuits of C, we have

$$\Pr_{\mathbf{G},\mathbf{P}}[\bigvee_{\text{sub-circuits } \mathsf{C}'} \bigvee_{Q \in \mathcal{H}} \mathsf{C}'^{\cup \mathbf{G}} \text{ is sensitive over } \mathbf{Q}] \leq n^{-\varepsilon+o(1)} = o(1). \tag{4.13}$$

We now derive a contradiction to (4.13). By Lemma 133 (with respect to the circuit $\mathsf{C}^{\cup \mathbf{G}}$), if $\mathsf{C}^{\cup \mathbf{G}}$ is sensitive over $\mathbf{P}$, then there exist $Q \in \mathcal{H}$ and a sub-circuit C' of C such that $\mathsf{C}'^{\cup \mathbf{G}}$ is sensitive over $\mathbf{Q}$. It follows that

$$\begin{aligned}\Pr_{\mathbf{G},\mathbf{P}}[\bigvee_{\text{sub-circuits } \mathsf{C}'} \bigvee_{Q \in \mathcal{H}} \mathsf{C}'^{\cup \mathbf{G}} \text{ is sensitive over } \mathbf{Q}] &\geq \Pr_{\mathbf{G},\mathbf{P}}[\mathsf{C}^{\cup \mathbf{G}} \text{ is sensitive over } \mathbf{P}] \\ &= \Omega(1) \quad \text{(by Lemma 135).}\end{aligned} \tag{4.14}$$

Inequalities (4.13) and (4.14) give a contradiction, which completes the proof. □

4.3.3 Proof of Lemma 135

In this section we continue to assume that $\beta(e) > 0$ for all $e \in P$ (see the paragraph before the statement of Lemma 135). This assumption in particular implies that $E(\mathbf{G}) \cap E(\mathbf{P}) = \emptyset$ almost surely. Thus we only have to prove that with constant probability $\mathbf{G} \cup \mathbf{P}$ does not contain any P-subgraphs other than $\mathbf{P}$ itself (a formal argument is included at the end of this section).

Lemma 138.

1. *For every P-colored graph G in the support of $\mathbf{G}$ and every subgraph $Q \subseteq P$,*

$$\frac{\Pr[\mathbf{G} \cup \mathbf{Q} = G]}{\Pr[\mathbf{G} = G]} = (1 + o(1)) \frac{\mathrm{sub}(Q, G)}{n^{\alpha(Q)-\beta(Q)}}.$$

2. *If $\mathcal{A}$ is a property of P-colored graphs which holds a.a.s. for $\mathbf{G}$, then $\mathcal{A}$ holds a.a.s. for $\mathbf{G} \cup \mathbf{Q}$ for every $Q \subseteq P$.*

3. $\liminf_{n\to\infty} \Pr[\,\mathrm{sub}(P, \mathbf{G}\cup\mathbf{P}) = 1\,] > 0$. *That is,* $\mathbf{G}\cup\mathbf{P}$ *has a unique* P*-subgraph (namely* $\mathbf{P}$*) with probability bounded away from* 0.

Proof. (1): Noting that the number of possible Q-subgraphs in $\mathbf{G}$ is $\prod_{v\in V(Q)} \lfloor n^{\alpha(v)} \rfloor = (1-o(1))n^{\alpha(Q)}$, we have

$$\Pr[\,\mathbf{G}\cup\mathbf{Q} = G\,] = (1+o(1))n^{-\alpha(Q)} \sum_{K\in \mathrm{Sub}(Q,G)} \sum_{H\,:\,G\setminus K\subseteq H\subseteq G} \Pr[\,\mathbf{G} = H\,].$$

For every $K \in \mathrm{Sub}(Q,G)$ and H such that $G\setminus K \subseteq H \subseteq G$, we have

$$\begin{aligned}\Pr[\,\mathbf{G} = H\,] &= \Pr[\,\mathbf{G} = G\,] \cdot \prod_{e\in E(K\setminus H)} \frac{1-n^{-\beta(\widehat{e})}}{n^{-\beta(\widehat{e})}} \\ &= (1-o(1))\Pr[\,\mathbf{G} = G\,] \cdot \begin{cases} n^{\beta(Q)} & \text{if } H = G\setminus K, \\ n^{\beta(Q)-\Omega(1)} & \text{otherwise (since } \beta \text{ positive).}\end{cases}\end{aligned}$$

(Above, $\widehat{e}$ is the edge in P corresponding to e under the vertex-coloring of G.) Since $n^{-\Omega(1)}$ dominates $2^{|E(Q)|}-1$ (i.e. the number of summands where $H \neq G\setminus K$), statement (1) follows.

(2): Suppose $\mathcal{A}$ holds a.a.s. with respect to $\mathbf{G}$ (i.e. $\limsup_{n\to\infty} \Pr[\,\mathbf{G}\notin\mathcal{A}\,] = 0$) and let $Q\subseteq P$. Let $c>0$ be an arbitrary (large) constant. We split up the event $\{\mathbf{G}\cup\mathbf{Q}\notin\mathcal{A}\}$ as follows:

$$\begin{aligned}\Pr[\,\mathbf{G}\cup\mathbf{Q}\notin\mathcal{A}\,] \leq{}& \Pr[\,\mathrm{sub}(Q,\mathbf{G}\cup\mathbf{Q}) \geq cn^{\alpha(Q)-\beta(Q)}\,] \\ &+ \Pr[\,\mathbf{G}\cup\mathbf{Q}\notin\mathcal{A} \text{ and } \mathrm{sub}(Q,\mathbf{G}\cup\mathbf{Q}) \leq cn^{\alpha(Q)-\beta(Q)}\,].\end{aligned}$$

We bound each of the righthand terms separately.

First, note that

$$\begin{aligned}\mathrm{E}[\,\mathrm{sub}(Q, \mathbf{G} \cup \mathbf{Q})\,] &= \sum_{R \subseteq Q} \mathrm{E}[\,|\{H \in \mathrm{Sub}(Q, \mathbf{G} \cup \mathbf{Q}) : \chi(H \cap \mathbf{Q}) = R\}|\,] \\ &\leq \sum_{R \subseteq Q} n^{\alpha(Q)-\alpha(R)} \cdot n^{-\beta(Q \setminus R)} \\ &\leq 2^{|E(Q)|} \cdot n^{\alpha(Q)-\beta(Q)}\end{aligned}$$

(the latter inequality holds since $\alpha(R) \geq \beta(R)$ for all R). So by Markov's inequality,

$$\Pr[\,\mathrm{sub}(Q, \mathbf{G} \cup \mathbf{Q}) \geq cn^{\alpha(Q)-\beta(Q)}\,] \leq \frac{2^{|E(Q)|}}{c}.$$

Second, we have

$$\begin{aligned}&\Pr[\,\mathbf{G} \cup \mathbf{Q} \notin \mathcal{A} \text{ and } \mathrm{sub}(Q, \mathbf{G} \cup \mathbf{Q}) \leq cn^{\alpha(Q)-\beta(Q)}\,] \\ &= \sum_{G\,:\,G \notin \mathcal{A} \text{ and } \mathrm{sub}(Q,G) \leq cn^{\alpha(Q)-\beta(Q)}} \Pr[\,\mathbf{G} \cup \mathbf{Q} = G\,] \\ &= \sum_{G\,:\,G \notin \mathcal{A} \text{ and } \mathrm{sub}(Q,G) \leq cn^{\alpha(Q)-\beta(Q)}} (1 + o(1)) \Pr[\,\mathbf{G} = G\,] \frac{\mathrm{sub}(Q, G)}{n^{\alpha(Q)-\beta(Q)}} \text{ (by (1))} \\ &\leq \sum_{G\,:\,G \notin \mathcal{A} \text{ and } \mathrm{sub}(Q,G) \leq cn^{\alpha(Q)-\beta(Q)}} (1 + o(1)) c \Pr[\,\mathbf{G} = G\,] \\ &\leq (1 + o(1)) c \Pr[\,\mathbf{G} \notin \mathcal{A}\,].\end{aligned}$$

Since $\liminf_{n\to\infty} \Pr[\,\mathbf{G} \notin \mathcal{A}\,] = 0$, it follows that

$$\liminf_{n\to\infty} \Pr[\,\mathbf{G} \cup \mathbf{Q} \notin \mathcal{A}\,] \leq \frac{2^{|E(Q)|}}{c}.$$

Since c may be chosen arbitrarily large, we conclude that $\mathcal{A}$ holds a.a.s. with respect to $\mathbf{G} \cup \mathbf{Q}$.

(3): Note that for $Q = P$ and $\mathrm{sub}(P, G) = 1$ Lemma 138(1) simplifies to $\Pr[\,\mathbf{G} \cup \mathbf{P} = G\,] = (1 + o(1)) \Pr[\,\mathbf{G} = G\,]$. Thus, we have

$$\begin{aligned}\liminf_{n\to\infty} \Pr[\,\mathrm{sub}(P, \mathbf{G} \cup \mathbf{P}) = 1\,] &= \liminf_{n\to\infty} \Pr[\,\mathrm{sub}(P, \mathbf{G}) = 1\,] \\ &> 0 \text{ (by Lemma 120).}\end{aligned}$$

□

Proof of Lemma 135. Let $h : \mathcal{G}_{\alpha,\beta}(n) \to \{0,1\}$ denote the $\textsc{Subgraph}_{\mathsf{col}}(P)$ function, that is, $h(G) = 1 \Leftrightarrow G$ contains a P-subgraph. Assume $f : \mathcal{G}_{\alpha,\beta}(n) \to \{0,1\}$ solves $\textsc{Subgraph}_{\mathsf{col}}(P)$ in the average-case on $\mathbf{G}$, that is,

$$\Pr[\,f(\mathbf{G}) = h(\mathbf{G})\,] = 1 - o(1).$$

By Lemma 138(2),

$$\Pr[\,f(\mathbf{G} \cup \mathbf{Q}) = h(\mathbf{G} \cup \mathbf{Q}) \text{ for all } Q \subseteq P\,] = 1 - o(1).$$

Since the event "$f^{\cup\mathbf{G}}$ is sensitive over $\mathbf{P}$" depends only on the values of $f(\mathbf{G} \cup \mathbf{Q})$ for $Q \subseteq P$, we have

$$\Pr[\,f^{\cup\mathbf{G}} \text{ is sensitive over } \mathbf{P} \Leftrightarrow h^{\cup\mathbf{G}} \text{ is sensitive over } \mathbf{P}\,] = 1 - o(1). \tag{4.15}$$

As we already indicated above,

$$h^{\cup\mathbf{G}} \text{ is sensitive over } \mathbf{P} \Leftrightarrow E(\mathbf{G}) \cap E(\mathbf{P}) = \emptyset \text{ and } \mathrm{sub}(P, \mathbf{G} \cup \mathbf{P}) = 1 \tag{4.16}$$

(with probability 1). To see why, first assume $E(\mathbf{G}) \cap E(\mathbf{P}) = \emptyset$ and $\mathrm{sub}(P, \mathbf{G} \cup \mathbf{P}) = 1$. It follows that $h^{\cup\mathbf{G}}(\mathbf{P} - \{e\}) = 0$ for all $e \in E(\mathbf{P})$. Since $h^{\cup\mathbf{G}}(\mathbf{P}) = 1$, this shows that $h^{\cup\mathbf{G}}$ is sensitive over $\mathbf{P}$. For the opposite direction, consider the case that there exists $e \in E(\mathbf{G}) \cap E(\mathbf{P})$. Note that e does appear in $\mathsf{Sens}(h^{\cup G}, \mathbf{P})$. Therefore, $h^{\cup\mathbf{G}}$ is not sensitive

over $\mathbf{P}$. Finally, consider the case that $\mathrm{sub}(P, \mathbf{G} \cup \mathbf{P}) > 1$. Then $\mathbf{G} \cup \mathbf{P}$ contains a P-subgraph other than $\mathbf{P}$; this P-subgraph necessarily does not include some edge $e \in E(\mathbf{P})$. Note that $\mathrm{sub}(P, \mathbf{G} \cup (\mathbf{P} - \{e\})) \geq 1$, which means that e does appear in $\mathsf{Sens}(h^{\cup G}, \mathbf{P})$. So again $h^{\cup \mathbf{G}}$ is not sensitive over $\mathbf{P}$.

From (4.15) and (4.16), we have

$$\begin{aligned}\Pr[\, f^{\cup \mathbf{G}} \text{ is sensitive over } \mathbf{P}\,] &\geq \Pr[\, E(\mathbf{G}) \cap E(\mathbf{P}) = \emptyset \text{ and } \mathrm{sub}(P, \mathbf{G} \cup \mathbf{P}) = 1\,] - o(1) \\ &\geq \Pr[\, \mathrm{sub}(P, \mathbf{G} \cup \mathbf{P}) = 1\,] - \Pr[\, E(\mathbf{G}) \cap E(\mathbf{P}) \neq \emptyset\,] - o(1).\end{aligned}$$

Since β is positive,

$$\Pr[\, E(\mathbf{G}) \cap E(\mathbf{P}) \neq \emptyset\,] \leq \sum_{e \in E(P)} n^{-\beta(e)} = o(1).$$

Completing the proof, we have

$$\liminf_{n \to \infty} \Pr[\, f^{\cup \mathbf{G}} \text{ is sensitive over } \mathbf{P}\,] \geq \liminf_{n \to \infty} \Pr[\, \mathrm{sub}(P, \mathbf{G} \cup \mathbf{P}) = 1\,] > 0$$

by Lemma 138(3). □

Remark 139. *We get analogous version of Lemma 135 and Lemma 138 in the uncolored setting where P is balanced and $\mathbf{G} = G(n, n^{-\theta(P)})$. The analysis is essentially the same as we get the colored setting with respect to the threshold pair $(\alpha, \beta) = (1, \theta(P))$, that is, $\alpha(Q) = |V(Q)|$ and $\beta(Q) = \theta(P)|E(Q)|$. However, all instances of $n^{\alpha(Q)}$ become $\binom{n}{|V(Q)|}$ in the uncolored setting; for instance, the expected number of Q-subgraphs in $\mathbf{G}$ is $\binom{n}{|V(Q)|} n^{-\theta(P)|E(Q)|}$. Modulo this change, the proofs in this section adapt straightforwardly to the uncolored setting.*

4.3.4 Unbounded Fan-In

As discussed at the beginning of this section, our lower bound for type-II circuits (Theorem 137) implies an $n^{\frac{1}{2}\kappa_{\alpha,\beta}(P)-o(1)}$ lower bound on the size of type-I (that is, bounded-depth unbounded-fanin AC^0) circuits. We now modify the proof of Theorem 137 to obtain the stronger $n^{\kappa_{\alpha,\beta}(P)-o(1)}$ lower bound for type-I circuits. (The argument presented below follows Section 3.4 of [74].)

Theorem 140. *Suppose* C *is a type-I* AC^0 *circuit that solves* $\textsc{Subgraph}_{\mathsf{col}}(P)$ *in the average-case on* $G_{\alpha,\beta}(P)$*. Then* C *has size at least* $n^{\kappa_{\alpha,\beta}(P)-o(1)}$*.*

Our proof requires a slightly stronger version of Lemma 136.

Lemma 141. *Suppose* $f : \mathcal{G} \to \{0,1\}^m$ *is* AC^0*-computable where* $m = n^{o(1)}$ *(that is,* f *is computed by an* $O(1)$*-depth,* $n^{O(1)}$*-size circuit with* $n^{o(1)}$ *output gates). Then for every* $Q \subseteq P$*,*

$$\Pr_{\mathbf{G},\mathbf{Q}}\left[\, f^{\cup\mathbf{G}} \text{ is sensitive over } \mathbf{Q} \,\right] \leq n^{-\alpha(Q)+\beta(Q)+o(1)}.$$

Proof. We modify the proof of Lemma 136 as follows. Let boolean functions $f_1, \ldots, f_m : \mathcal{G} \to \{0,1\}$ be the coordinates of the output of f. Consider the random restriction ρ and the graph H_ρ as before (that is, as defined in the proof of Lemma 136 with $E(H_\rho) = \rho^{-1}(\star)$). Also as before, let L_ρ denote the subgraph of H_ρ with edge set

$$E(L_\rho) = \{e \in \rho^{-1}(\star) : f\restriction_\rho : \{0,1\}^{\rho^{-1}(\star)} \to \{0,1\}^m \text{ depends on coordinate } e\}.$$

Let us now also consider graphs $L_{1,\rho}, \ldots, L_{m,\rho} \subseteq H_\rho$ defined by

$$E(L_{i,\rho}) = \{e \in \rho^{-1}(\star) : f_i\restriction_\rho : \{0,1\}^{\rho^{-1}(\star)} \to \{0,1\} \text{ depends on coordinate } e\}.$$

Clearly, we have $L_\rho = L_{1,\rho} \cup \cdots \cup L_{m,\rho}$.

Recall that $\delta > 0$ is an arbitrary constant (which we allow to depend on P and (α,β), but will be independent of n). For each $i \in \{1,\ldots,m\}$, the argument in Appendix **??** gives the bound

$$\Pr_\rho[\,|L_{i,\rho}| > n^\delta\,] \leq S \cdot (5n^{-\delta/d}\delta\log n)^{\delta\log n} = n^{-\omega(1)}$$

where S and d are the size and depth of the circuit defining f. (This bound is $n^{-\omega(1)}$ since $S = n^{O(1)}$ and $d = O(1)$ and $\delta = \Omega(1)$.) Similarly, we have $\Pr_\rho[\,|L_{i,\rho}| > n^{\delta/2}\,] \leq n^{-\omega(1)}$ (replacing δ with $\delta/2$ above). Since $m = n^{o(1)}$, for all n sufficiently large, we get the bound

$$\Pr_\rho[\,|L_\rho| > n^\delta\,] \leq \Pr_\rho[\,\bigvee_{i=1}^{m} |L_{i,\rho}| > n^{\delta/2}\,] \leq \sum_{i=1}^{m} \Pr_\rho[\,|L_{i,\rho}| > n^{\delta/2}\,] = m \cdot n^{-\omega(1)} = n^{-\omega(1)}. \tag{4.17}$$

With (4.17) in place of (4.11), the rest of the proof is identical to that of Lemma 136. □

To prove Theorem 140, we shall once again convert type-I circuits into type-II circuits. However, this time we replace each unbounded fan-in gate with a *unbalanced* tree of fan-in 2 gates.

Definition 142. *For every type-I circuit* C *computing a boolean function* $\mathcal{G} \to \{0,1\}$*, we shall define an equivalent type-II circuit* $\widetilde{\mathsf{C}}$*. In addition, if the output of* C *is an AND/OR gate of fan-in* m*, then we shall also define auxiliary functions* $g^{\mathsf{C}} : \mathcal{G} \to \{0,1\}^m$ *and* $f^{\mathsf{C}} : \mathcal{G} \to \{0,1\}^{\lceil \log(m+1) \rceil}$*. The definition is inductive:*

- *If* C *is an input (variable or constant), then* $\widetilde{\mathsf{C}} := \mathsf{C}$*.*
- *If* $\mathsf{C} = \mathrm{NOT}(\mathsf{C}')$*, then* $\widetilde{\mathsf{C}} := \mathrm{NOT}(\widetilde{\mathsf{C}'})$*.*
- *If* $\mathsf{C} = \mathrm{OR}(\mathsf{C}_1, \ldots, \mathsf{C}_m)$*, then*

$$\widetilde{\mathsf{C}} := \widetilde{\mathsf{C}}^{(m)} \text{ where } \widetilde{\mathsf{C}}^{(i)} := \begin{cases} \widetilde{\mathsf{C}_1} & \text{if } i = 1, \\ \mathrm{OR}(\widetilde{\mathsf{C}}^{(i-1)}, \widetilde{\mathsf{C}_i}) & \text{if } i \in \{2, \ldots, m\}, \end{cases}$$

(that is, $\widetilde{\mathsf{C}} = \mathrm{OR}(\ldots \mathrm{OR}(\mathrm{OR}(\widetilde{\mathsf{C}_1}, \widetilde{\mathsf{C}_2}), \widetilde{\mathsf{C}_3}) \ldots, \widetilde{\mathsf{C}_m})$ *and, for all* $i \in [m]$*,*

$$g_i^{\mathsf{C}}(G) := C_1(G) \vee \cdots \vee C_i(G) \text{ (= the boolean function computed by } \widetilde{\mathsf{C}}^{(i)})$$

and

$$f^{\mathsf{C}}(G) := \begin{cases} 0 & \text{if } \mathsf{C}(G) = 0, \\ \text{binary}(i) & \text{if } \mathsf{C}_i(G) = 1 \text{ and } \mathsf{C}_1(G) = \cdots = \mathsf{C}_{i-1}(G) = 0, \end{cases}$$

where $\text{binary}(i) \in \{0,1\}^{\lceil \log(m+1) \rceil}$ *is the binary representation of* i*. (That is,* $f^{\mathsf{C}}(G)$ *reports that* $\mathsf{C}(G) = 0$ *or points to the first 1-value among* $\mathsf{C}_1(G), \ldots, \mathsf{C}_m(G)$*.)*

- *If* $\mathsf{C} = \text{AND}(\mathsf{C}_1, \ldots, \mathsf{C}_m)$*, then the definition is dual to the case of* OR*.*

We make some straightforward observations.

1. Note that $|\{\text{gates in } \widetilde{\mathsf{C}}\}| = |\{\text{wires in } \mathsf{C}\}| = O(|\{\text{gates in } \mathsf{C}\}|^2)$. (Indeed, the conversion $\mathsf{C} \mapsto \widetilde{\mathsf{C}}$ incurs a quadratic blow-up in size in the worst case. Our proof of Theorem 140 will avoid taking a union bound over gates in $\widetilde{\mathsf{C}}$.)

2. The operation $\mathsf{C} \mapsto \widetilde{\mathsf{C}}$ commutes with the operation $\mathsf{C} \mapsto \mathsf{C}^{\cup G}$. That is, we have $\widetilde{\mathsf{C}}^{\cup G} = \widetilde{\mathsf{C}^{\cup G}}$.

3. If $\mathsf{C} = \text{OR}(\mathsf{C}_1, \ldots, \mathsf{C}_m)$, then g^{C} reports the vector of values of sub-circuits $\widetilde{\mathsf{C}}^{(1)}, \ldots, \widetilde{\mathsf{C}}^{(m)}$. Note that $g^{\mathsf{C}}(G)$ can acheive only $m+1$ possible values:

$$(\underbrace{0, \ldots, 0}_{m}),\ (\underbrace{0, \ldots, 0}_{m-1}, 1),\ (\underbrace{0, \ldots, 0}_{m-2}, 1, 1),\ \ldots,\ \text{or } (\underbrace{1, \ldots, 1}_{m}).$$

4. Function $g^{\mathsf{C}} : \mathcal{G} \to \{0,1\}^m$ are $f^{\mathsf{C}} : \mathcal{G} \to \{0,1\}^{\lceil \log(m+1) \rceil}$ information-theoretically equivalent (that is, there is a bijection $\varphi : \text{Range}(g^{\mathsf{C}}) \xrightarrow{\cong} \text{Range}(f^{\mathsf{C}})$ such that $f^{\mathsf{C}}(H) = \varphi(g^{\mathsf{C}}(H))$ for all $H \in \mathcal{G}$). Consequently, for every graph $H \in \mathcal{G}$, we have $\mathsf{Sens}(g^{\mathsf{C}}, H) = \mathsf{Sens}(f^{\mathsf{C}}, H)$ and therefore

$$g^{\mathsf{C}} \text{ is sensitive over } H \iff f^{\mathsf{C}} \text{ is sensitive over } H.$$

5. We have $\mathsf{Sens}(g^{\mathsf{C}}, H) = \bigcup_{i\in[m]} \mathsf{Sens}(\widetilde{\mathsf{C}}^{(i)}, H)$. (In general, $\mathsf{Sens}(h, H) = \mathsf{Sens}(h_1, H) \cup \cdots \cup \mathsf{Sens}(h_t, H)$ for any multi-output function $h : \mathcal{G} \to \{0,1\}^t$.) In particular,

$$\bigvee_{i\in[m]} \big(\widetilde{\mathsf{C}}^{(i)} \text{ is sensitive over } H\big) \implies g^{\mathsf{C}} \text{ is sensitive over } H.$$

6. If C has depth $d = O(1)$ and size $\mathrm{poly}(n)$, then f^{C} is computable by a circuit of depth $d + 2 = O(1)$ and size $\mathrm{poly}(n)$ with $\lceil \log(m+1) \rceil = O(\log n) = n^{o(1)}$ output gates.

Observations (iii)-(vi) apply equally if $\mathsf{C} = \mathrm{AND}(\mathsf{C}_1, \ldots, \mathsf{C}_m)$ (with only the roles of 0 and 1 exchanged in observation (iii)).

Lemma 143. *Suppose C is a type-I AC^0 circuit with top fan-in $m \geq 2$ computing a boolean function $\mathcal{G} \to \{0,1\}$. Then for every $Q \subseteq P$, we have*

$$\Pr_{\mathbf{G},\mathbf{Q}}\Big[\bigvee_{i\in[m]} \big((\widetilde{\mathsf{C}}^{(i)})^{\cup\mathbf{G}} \textit{ is sensitive over } \mathbf{Q}\big)\Big] \leq n^{-\alpha(Q)+\beta(Q)+o(1)}.$$

Proof. By observations (ii), (iv) and (v), we have the implication

$$\bigvee_{i\in[m]} \big((\widetilde{\mathsf{C}}^{(i)})^{\cup\mathbf{G}} \text{ is sensitive over } \mathbf{Q}\big) \implies (f^{\mathsf{C}})^{\cup\mathbf{G}} \text{ is sensitive over } \mathbf{Q}.$$

By observation (vi) and Lemma 141, we have

$$\Pr_{\mathbf{G},\mathbf{Q}}\big[(f^{\mathsf{C}})^{\cup\mathbf{G}} \text{ is sensitive over } \mathbf{Q}\big] \leq n^{-\alpha(Q)+\beta(Q)+o(1)}.$$

The lemma follows from these two facts. □

We now present the proof of Theorem 140.

Proof of Theorem 140. Let C be a type-I circuit of depth $d = O(1)$ that solves $\textsc{Subgraph}_{\mathsf{col}}(P)$ in the average-case on $G_{\alpha,\beta}(P)$. Toward a contradiction, assume C has size $\leq n^{\kappa_{\alpha,\beta}(P)-\varepsilon}$

for some constant $\varepsilon > 0$ (independent of n). By Lemma 126, we may fix a hitting set $\mathcal{H}$ for P such that $\kappa_{\alpha,\beta}(P) = \min_{Q\in\mathcal{H}} \alpha(Q) - \beta(Q)$.

Let $\mathcal{S}$ (resp. $\widetilde{\mathcal{S}}$) be the set of subcircuits of C (resp. $\widetilde{\mathsf{C}}$). Note that, for each $\mathsf{D} \in \widetilde{\mathcal{S}}$, either

- $\mathsf{D} = \widetilde{\mathsf{C}'}$ for some $\mathsf{C}' \in \mathcal{S}$ with top fan-in 1 (i.e. C' is an input or has a NOT gate on top), or
- $\mathsf{D} = \widetilde{\mathsf{C}'}^{(i)}$ for some $\mathsf{C}' \in \mathcal{S}$ with top fan-in $m \geq 2$ and some $i \in [m]$.

Applying Lemma 133 to the type-II circuit $\widetilde{\mathsf{C}}^{\cup\mathbf{G}}$, we have that if $\widetilde{\mathsf{C}}^{\cup\mathbf{G}}$ is sensitive over $\mathbf{P}$, then there exist $Q \in \mathcal{H}$ and a sub-circuit $\mathsf{D} \in \widetilde{\mathcal{S}}$ such that $\mathsf{D}^{\cup\mathbf{G}}$ is sensitive over $\mathbf{Q}$. Therefore,

$$\begin{aligned}
\Omega(1) &\leq \Pr_{\mathbf{G},\mathbf{P}}\big[\,\mathsf{C}^{\cup\mathbf{G}} \text{ is sensitive over } \mathbf{P}\,\big] \quad \text{(by Lemma 135)} \\
&= \Pr_{\mathbf{G},\mathbf{P}}\big[\,\widetilde{\mathsf{C}}^{\cup\mathbf{G}} \text{ is sensitive over } \mathbf{P}\,\big] \\
&\leq \Pr_{\mathbf{G},\mathbf{P}}\Big[\,\bigvee_{Q\in\mathcal{H}} \bigvee_{\mathsf{D}\in\widetilde{\mathcal{S}}} \mathsf{D}^{\cup\mathbf{G}} \text{ is sensitive over } \mathbf{Q}\,\Big] \\
&\leq \sum_{Q\in\mathcal{H}} \Pr_{\mathbf{G},\mathbf{Q}}\Big[\,\bigvee_{\mathsf{D}\in\widetilde{\mathcal{S}}} \mathsf{D}^{\cup\mathbf{G}} \text{ is sensitive over } \mathbf{Q}\,\Big] \\
&\leq |\mathcal{H}| \cdot \max_{Q\in\mathcal{H}} \Pr_{\mathbf{G},\mathbf{Q}}\Big[\,\bigvee_{\mathsf{D}\in\widetilde{\mathcal{S}}} \mathsf{D}^{\cup\mathbf{G}} \text{ is sensitive over } \mathbf{Q}\,\Big].
\end{aligned}$$

For each $Q \in \mathcal{H}$, by Lemmas 136 and 143, we have

$$\begin{aligned}
&\Pr_{\mathbf{G},\mathbf{Q}}\Big[\,\bigvee_{\mathsf{D}\in\widetilde{\mathcal{S}}} \mathsf{D}^{\cup\mathbf{G}} \text{ is sensitive over } \mathbf{Q}\,\Big] \\
&\leq \sum_{\mathsf{C}'\in\mathcal{S}} \begin{cases} \Pr_{\mathbf{G},\mathbf{Q}}\big[\,(\mathsf{C}')^{\cup\mathbf{G}} \text{ is sensitive over } \mathbf{Q}\,\big] & \text{if } C' \text{ has top fan-in 1,} \\ \Pr_{\mathbf{G},\mathbf{Q}}\big[\,\bigvee_{i\in[m]} (\widetilde{\mathsf{C}'}^{(i)})^{\cup\mathbf{G}} \text{ is sensitive over } \mathbf{Q}\,\big] & \text{if } C' \text{ has top fan-in } m \geq 2, \end{cases} \\
&\leq \text{size}(\mathsf{C}) \cdot n^{-\alpha(Q)+\beta(Q)+o(1)}.
\end{aligned}$$

It follows that

$$\begin{aligned}\Omega(1) &\leq \text{size}(\mathsf{C}) \cdot |\mathcal{H}| \cdot \max_{Q \in \mathcal{H}} n^{-\alpha(Q)+\beta(Q)+o(1)} \\ &= \text{size}(\mathsf{C}) \cdot |\mathcal{H}| \cdot n^{-\kappa_{\alpha,\beta}(P)+o(1)} \\ &= n^{-\varepsilon+o(1)}\end{aligned}$$

(since $|\mathcal{H}| = O(1)$ and using our assumption that $\text{size}(\mathsf{C}) = n^{\kappa_{\alpha,\beta}(P)-\varepsilon}$). This is the desired contradiction, which completes the proof of the theorem. □

4.4 Bounds on $\kappa_{\mathsf{col}}(P)$

In the previous section, we proved that $C_{\mathsf{col}}(P) \geq \kappa_{\mathsf{col}}(P)$, that is, $n^{\kappa_{\mathsf{col}}(P)-o(1)}$ is a lower bound on the AC^0 complexity of $\textsc{Subgraph}_{\mathsf{col}}(P)$. In Section 4.4.2 below we will complete the proof of Theorem 109 by showing that $\kappa_{\mathsf{col}}(P) \geq \Omega(tw(P)/\log tw(P))$. But, as a warm-up, let us do a simple combinatorial *upper* bound on $\kappa_{\mathsf{col}}(P)$ (and also present a useful construction in Lemma 144 that we will need for lower bound proofs).

4.4.1 Upper Bound

We have already established that $\kappa_{\mathsf{col}}(P) \leq C_{\mathsf{col}}(P)$ (Corollary 128) and $C_{\mathsf{col}}(P) \leq tw(P)+1$ (Lemma 113(1)). By these lower and upper bounds in circuit complexity, it follows that $\kappa_{\mathsf{col}}(P) \leq tw(P)+1$. In this subsection, we give a direct proof that $\kappa_{\mathsf{col}}(P) \leq tw(P)+1$.

We need the following fact, which shows that the max in the definition $\kappa_{\mathsf{col}}(P) := \max_{(\alpha,\beta)\in\theta_{\mathsf{col}}(P)} \kappa_{\alpha,\beta}(P)$ is always achieved by some $(\alpha,\beta) \in \theta_{\mathsf{col}}(P)$ with $\alpha \equiv 1$.

Lemma 144. *Assume that $(\alpha,\beta) \in \theta_{\mathsf{col}}(P)$ and define β' by the formula*

$$\beta'(\{v,w\}) := \beta(\{v,w\}) + \frac{1-\alpha(v)}{d_P(v)} + \frac{1-\alpha(w)}{d_P(w)},$$

where $d_P(v)$ is the degree of the vertex v. Then $(1, \beta') \in \theta_{\mathsf{col}}(P)$ and, moreover, $v(Q) - \beta'(Q) \geq \alpha(Q) - \beta(Q)$ for any $Q \subseteq P$.

Proof. First, $\beta(\{v, w\}) \leq \alpha(v) + \alpha(w)$ (by Definition 118(i) applied to $Q := \{\{u, v\}\}$) and $d_P(v), d_P(w) \geq 1$. Hence $\beta'(\{v, w\}) \leq 2$. Now, for all $Q \subseteq P$ we have

$$\begin{aligned} v(Q) - \beta'(Q) &= v(Q) - \sum_{\{v,w\} \in E(Q)} \Big(\beta(\{v, w\}) + \frac{1 - \alpha(v)}{d_P(v)} + \frac{1 - \alpha(w)}{d_P(w)} \Big) \\ &= \sum_{v \in V(Q)} \Big(\underbrace{1 - \frac{d_Q(v)}{d_P(v)}(1 - \alpha(v))}_{\geq \alpha(v)} \Big) - \sum_{\{v,w\} \in E(Q)} \beta(\{v, w\}) \\ &\geq \alpha(Q) - \beta(Q) \end{aligned}$$

with equality when $Q = P$. □

Corollary 145. *For all P, there exists $\beta : E(P) \to [0, 2]$ such that $(1, \beta) \in \theta_{\mathsf{col}}(P)$ and $\kappa_{\mathsf{col}}(P) = \kappa_{1,\beta}(P)$.*

Proof. Let $(\alpha, \beta) \in \theta_{\mathsf{col}}(P)$ be such that $\kappa_{\mathsf{col}}(P) = \kappa_{\alpha,\beta}(P)$. Then the element $(1, \beta') \in \theta_{\mathsf{col}}(P)$ constructed from (α, β) as in Lemma 144, has the desired property. □

Proposition 146. *$\kappa_{\mathsf{col}}(P) \leq tw(P) + 1$.*

Proof. In fact, we will prove $\kappa_{\mathsf{col}}(P) \leq bw(P)$ where $bw(P)$ is the branch-width of P (it is known that $bw(P) \leq tw(P) + 1$ by [76]). Recall that a *branch decomposition* of P is a pair (T, b) where T is a ternary tree and b is a bijection from Leaves(T) to E. Each edge in T determines a partition of Leaves(T) (and hence of E) into two sets. The *width* of (T, b) is the maximum of $|V(E_1) \cap V(E_2)|$ over partitions $E = E_1 \uplus E_2$ determined by the edges of T, and the *branch-width* $bw(P)$ is the minimum possible width of a branch decomposition of P.

Suppose $bw(P) = k$. This means that there exists a branch decomposition (T, b) of width k. Fix an arbitrary root of T and let $x_1, \ldots, x_t$ be a post-order traversal of nodes

in T (in particular, x_t is the root). For every $i \in [t]$, let Q_i and $\overline{Q_i}$ be the subgraphs of P with $E(Q_i) := \{b(y) : y \text{ is a leaf of } T \text{ lying below } x_i\}$ and $E(\overline{Q_i}) := E(P) \setminus E(Q_i)$ and $V(Q_i) := \bigcup_{e \in E(Q_i)} e$ and $V(\overline{Q_i}) := \bigcup_{e \in E(\overline{Q_i})} e$. Note that $Q_1, \ldots, Q_t$ is a union sequence for P. Since (T, b) has width $\leq k$, we have $|V(Q_i) \cap V(\overline{Q_i})| \leq k$ for all $i \in [t]$.

By Corollary 145, there exists $\beta : E(P) \to [0, 2]$ such that $(1, \beta) \in \theta_{\mathsf{col}}(P)$ and $\kappa_{\mathsf{col}}(P) = \kappa_{1,\beta}(P)$. For all $i \in [t]$, we have

$$\begin{aligned}
& v(Q_i) - \beta(Q_i) \\
\leq & v(Q_i) - \beta(Q_i) + v(\overline{Q_i}) - \beta(\overline{Q_i}) && \text{(since } v(\overline{Q_i}) \geq \beta(\overline{Q_i})\text{)} \\
= & v(Q_i) + v(\overline{Q_i}) - v(P) && \text{(since } \beta(Q_i) + \beta(\overline{Q_i}) = \beta(P) = v(P)\text{)} \\
= & |V(Q_i) \cap V(\overline{Q_i})| \ \leq \ k.
\end{aligned}$$

Therefore, $\kappa_{\mathsf{col}}(P) = \kappa_{1,\beta}(P) \leq \max_{i \in [t]} v(Q_i) - \beta(Q_i) \leq k \leq bw(P) \leq tw(P) + 1$. □

4.4.2 Lower Bounds

We give two lower bounds on $\kappa_{\mathsf{col}}(P)$. The first applies to all patterns P.

Theorem 147. $\kappa_{\mathsf{col}}(P) \geq \Omega\left(\frac{tw(P)}{\log tw(P)}\right)$.

Together with the fact that $C_{\mathsf{col}}(P) \geq \kappa_{\mathsf{col}}(P)$ (Corollary 128), this completes the proof of our main theorem (Theorem 109).

Our proof of Theorem 147 uses a characterization of treewidth from Marx [57] (based on results of Feige et al [31]): for every P with $tw(P) = k$, there is a subset $W \subseteq V(P)$ of size $|W| = \Omega(k)$ and a concurrent flow on P which routes $\Omega(1/k \log k)$ flow between every pair of distinct vertices in W (Lemma 151). Given such a concurrent flow on P, we construct a corresponding threshold pair $(\alpha, \beta) \in \theta_{\mathsf{col}}(P)$ and show that $\kappa_{\alpha,\beta}(P)$ gives the desired bound.

We also include a lower bound on $\kappa_{\mathsf{col}}(P)$ in terms of the expansion of P (Theorem 152), which improves Theorem 147 in the case where P is a constant-degree expander.

Definition 148.

1. *Let* $\mathsf{Paths}(P)$ *denote the set of* paths *in* P *(i.e. subgraphs of* P *isomorphic to an (undirected, simple) path of length* ≥ 1*).*

2. *Let* $\mathsf{Flows}(P)$ *denote the set of* concurrent flows *on* P *with node-capacity* 1*, that is, functions* $f : \mathsf{Paths}(P) \to [0,1]$ *such that for all* $v \in V(P)$*,* $\sum_{\substack{\pi \in \mathsf{Paths}(P) \\ v \in V(\pi)}} f(\pi) \leq 1$*.*

3. *For* $f \in \mathsf{Flows}(P)$ *and disjoint* $S, T \subseteq V(P)$*, let* $f(S,T)$ *denote the total flow that* f *sends between* S *and* T*, that is,*

$$f(S,T) := \sum_{\substack{\pi \in \mathsf{Paths}(P) \\ \pi \text{ has endpoints in } S \text{ and } T}} f(\pi).$$

   For two distinct vertices v, w*, we let* $f(v,w) := f(\{v\},\{w\})$*.*

4. *For* $\pi \in \mathsf{Paths}(P)$*, define* $\alpha_\pi : V(P) \to [0,1]$ *and* $\beta_\pi : E(P) \to [0,2]$ *by*

$$\alpha_\pi(v) := \begin{cases} 1/2 & \text{if } v \text{ is an endpoint of } \pi, \\ 1 & \text{if } v \text{ is an interior vertex of } \pi, \\ 0 & \text{if } v \notin V(\pi), \end{cases} \qquad \beta_\pi(e) := \begin{cases} 1 & \text{if } e \in E(\pi), \\ 0 & \text{if } e \notin E(\pi). \end{cases}$$

5. *For* $f \in \mathsf{Flows}(P)$*, define* $\alpha_f : V(P) \to [0,1]$ *and* $\beta_f : E(P) \to [0,2]$ *by*

$$\alpha_f(v) := \sum_{\pi \in \mathsf{Paths}(P)} f(\pi) \cdot \alpha_\pi(v), \qquad \beta_f(e) := \sum_{\pi \in \mathsf{Paths}(P)} f(\pi) \cdot \beta_\pi(e).$$

Lemma 149. $(\alpha_f, \beta_f) \in \theta_{\mathsf{col}}(P)$ *for all* $f \in \mathsf{Flows}(P)$*.*

Proof. Clearly, $\alpha_\pi(P) = \beta_\pi(P)$ $(= |E(\pi)|)$ and $\alpha_\pi(Q) \geq \beta_\pi(Q)$ for all $Q \subseteq P$ and $\pi \in \mathsf{Paths}(P)$. $(\alpha_f, \beta_f) \in \theta_{\mathsf{col}}(P)$ follows by convexity. □

Lemma 150. *For all $Q \subseteq P$ and $f \in \mathsf{Flows}(P)$,*

$$\alpha_f(Q) - \beta_f(Q) \geq \frac{1}{2} f(V(Q), \overline{V(Q)}).$$

Proof. Note that $f(S,T) = \sum_{\pi \in \mathsf{Paths}(P)} f(\pi) \cdot \chi_\pi(S,T)$ where

$$\chi_\pi(S,T) := \begin{cases} 1 & \text{if } \pi \text{ has one endpoint in } S \text{ and another in } T, \\ 0 & \text{otherwise.} \end{cases}$$

Therefore, it suffices to show, for all $\pi \in \mathsf{Paths}(P)$, that

$$\alpha_\pi(Q) - \beta_\pi(Q) \geq \frac{1}{2} \chi_\pi(V(Q), \overline{V(Q)}). \tag{4.18}$$

If both endpoints of π belong to the same set among $V(Q)$, $\overline{V(Q)}$, then $\frac{1}{2}\chi_\pi(V(Q), \overline{V(Q)}) = 0$ while $\alpha_\pi(Q) - \beta_\pi(Q) \geq 0$ by Lemma 149 (since $(\alpha_\pi, \beta_\pi) \in \theta_{\mathsf{col}}(P)$); so (4.18) holds. On the other hand, if π has one endpoint in $V(Q)$ and another in $\overline{V(Q)}$, then $\frac{1}{2}\chi_\pi(V(Q), \overline{V(Q)}) = \frac{1}{2}$, while

$$\alpha_\pi(Q) - \beta_\pi(Q) \geq \frac{1}{2} |\{\text{edges of } \pi \text{ that cross between } V(Q) \text{ and } \overline{V(Q)}\}| \geq \frac{1}{2},$$

so again (4.18) holds. □

Our lower bound on $\kappa_{\mathsf{col}}(P)$ relies on a characterization of treewidth in terms of concurrent flows:

Lemma 151. *If P has treewidth k, then there exists $W \subseteq V(P)$ with $|W| \geq \frac{2k}{3}$ and $f \in \mathsf{Flows}(P)$ such that $f(v,w) \geq \frac{1}{ck \log k}$ for all distinct $v, w \in W$ where $c > 0$ is a universal constant.*

Proof. This lemma is implicit in [31, 57]. Utilizing the notation from the latter paper, [57,

Lemma 3.2(a)] implies that there exists $W \subseteq V(P)$ with $|W| > \frac{2k}{3}$ that has no balanced $\frac{k}{3}$-separator. By [57, Lemma 3.3], $\alpha^w(G) \geq \left(\frac{4}{3}k+1\right)^{-1}$. By the contrapositive of [57, Theorem 3.5], all solutions to the linear program (LP2) are $\Omega\left(\frac{1}{k \log k}\right)$, hence its dual (LP1) has a solution $\alpha \geq \Omega\left(\frac{1}{k \log k}\right)$, and it gives us the required concurrent flow. □

Proof of Theorem 147. Suppose $tw(P) = k$ and fix $W \subseteq V(P)$ and $f \in \mathsf{Flows}(P)$ as in Lemma 151. Let $\mathcal{H}$ be the set of subgraphs $Q \subseteq P$ such that $\frac{2k}{9} \leq |W \cap V(Q)| \leq \frac{4k}{9}$. Clearly $\mathcal{H}$ is a hitting set for P (i.e. every union sequence for P contains a graph in this set). For every $Q \in \mathcal{H}$, we have

$$\begin{aligned}\alpha_f(Q) - \beta_f(Q) &\geq \frac{f(V(Q), \overline{V(Q)})}{2} \geq \frac{f(W \cap V(Q), W \setminus V(Q))}{2} \\ &\geq \frac{|W \cap V(Q)| \cdot |W \setminus V(Q)|}{2ck \log k} \geq \frac{4k}{81c \log k} = \Omega\left(\frac{k}{\log k}\right).\end{aligned}$$

Therefore, $\kappa_{\mathsf{col}}(P) \geq \kappa_{\alpha_f, \beta_f}(P) = \Omega(k/\log k)$. □

Tight lower bound for expanders.

We conclude this section by giving a second lower bound on $\kappa_{\mathsf{col}}(P)$ in terms of edge expansion; this sometimes gives the optimal $\Omega(tw(P))$ lower bound in the case that P is a good expander such as K_k or $G_{k,k}$. Let $\Delta(P)$ denote the maximum degree of P. For $S \subseteq V(P)$, let $e_P(S, \overline{S}) := |\{\{v, w\} \in E(P) : v \in S \text{ and } w \in V(P) \setminus S\}|$. Recall that the *edge expansion* of P is defined by

$$h(P) := \min_{S : \emptyset \subset S \subset V(P)} \frac{e_P(S, \overline{S})}{\min\{|S|, |\overline{S}|\}}.$$

Theorem 152. $\kappa_{\mathsf{col}}(P) \geq \dfrac{h(P)v(P)}{3\Delta(P)}$.

Proof. Let us apply the construction from Lemma 144 to the pair $(0, 0) \in \theta_{\mathsf{col}}(P)$. This gives us the function $\beta : E(P) \to [0, 2]$ defined by

$$\beta(\{v, w\}) := \frac{1}{d_P(v)} + \frac{1}{d_P(w)}$$

such that $(1, \beta) \in \theta_{\mathsf{col}}(P)$.

Consider the hitting set $\mathcal{H}$ consisting of subgraphs $Q \subseteq P$ such that $\frac{1}{3}v(P) \leq v(Q) \leq \frac{2}{3}v(P)$. For every $Q \in \mathcal{H}$ the calculation in the proof of Lemma 144 gives us

$$\begin{aligned} v(Q) - \beta(Q) &= \sum_{v \in V(Q)} \left(1 - \frac{d_Q(v)}{d_P(v)}\right) \\ &\geq \frac{1}{\Delta(P)} \sum_{v \in V(Q)} (d_P(v) - d_Q(v)) \\ &= \frac{e_P(V(Q), \overline{V(Q)})}{\Delta(P)} \\ &\geq \frac{h(P) \min\{v(Q), v(P) - v(Q)\}}{\Delta(P)} \\ &\geq \frac{h(P)v(P)}{3\Delta(P)}. \end{aligned}$$

Completing the proof,

$$\kappa_{\mathsf{col}}(P) \geq \kappa_{1,\beta}(P) \geq \min_{Q \in \mathcal{H}} (v(Q) - \beta(Q)) \geq \frac{h(P)v(P)}{3\Delta(P)}. \qquad \square$$

4.5 Minor-Monotonicity and Monotone Projections

In this section, we prove that $\kappa_{\mathsf{col}}(P)$ and $C_{\mathsf{col}}(P)$ are minor-monotone graph parameters. First, a few definitions.

Recall that a *minor* of G is any graph that can be obtained from G by a sequence of vertex deletions, edge deletions, and edge contractions. A real-valued graph parameter f is *minor-monotone* if $f(G) \leq f(G')$ whenever G is a minor of G'.

Theorem 153. *κ_{col} and C_{col} are minor-monotone.*

The algorithmic problem $\textsc{Subgraph}_{\mathsf{col}}(P)$ was defined in Section 4.2.3 in such a way that the coloring $\chi : G \to P$ is a part of its input. We first observe that the parameter $C_{\mathsf{col}}(P)$ does not change if we consider instead its more structured version $\textsc{Subgraph}_{\mathsf{col},n}(P)$ in which

(cf. Definition 118) we demand that the target graph G has the vertex set $V(P) \times [n]$, and χ is the projection onto the first coordinate. An easy AC^0-reduction from $\textsc{Subgraph}_{\mathsf{col}}(P)$ to $\textsc{Subgraph}_{\mathsf{col},n}(P)$ works as follows. Assume that we are given an input (G, χ) to the problem $\textsc{Subgraph}_{\mathsf{col}}(P)$, and assume w.l.o.g. that $V(G) = [n]$. We map it to the pair (G', χ'), where $V(G') := V(P) \times [n]$, χ' is the projection onto the first coordinate, and $E(G')$ is defined as follows: $E(G') := \{\{(\chi(i), i), (\chi(j), j)\} : \{i, j\} \in E(G)\}$. (Thus, all vertices (v, i) with $v \neq \chi(i)$ remain isolated.) Hence Theorem 153 readily follows from the following lemma[11].

Lemma 154. *Suppose P is a minor of P'. Then*

1. *for every $(\alpha, \beta) \in \theta_{\mathsf{col}}(P)$, there exists $(\alpha', \beta') \in \theta_{\mathsf{col}}(P')$ such that $\kappa_{\alpha,\beta}(P) \leq \kappa_{\alpha',\beta'}(P')$,*
2. $\textsc{Subgraph}_{\mathsf{col},n}(P) \leq_{\mathrm{mp}} \textsc{Subgraph}_{\mathsf{col},n}(P')$.

Proof. It suffices to show that the lemma holds in the two cases where P is a subgraph of P', and where P is obtained from P' by contracting a single edge $\{x, y\}$ where x, y have no common neighbors. (Otherwise, we perform necessary edge deletions before contraction).

Subgraph Case. Suppose P is a subgraph of P'.

For (1): Consider any $(\alpha, \beta) \in \theta_{\mathsf{col}}(P)$. Define $\alpha' : V(P') \to [0, 1]$ and $\beta' : E(P') \to [0, 2]$ by

$$\alpha'(v) := \begin{cases} \alpha(v) & \text{if } v \in V(P), \\ 0 & \text{otherwise,} \end{cases} \qquad \beta'(e) := \begin{cases} \beta(e) & \text{if } e \in E(P), \\ 0 & \text{otherwise.} \end{cases}$$

It is easily seen that $(\alpha', \beta') \in \theta_{\mathsf{col}}(P')$ and $\kappa_{\alpha,\beta}(P) = \kappa_{\alpha',\beta'}(P')$.

11. A somewhat similar (de)construction recently appeared in [18].

For (2): The monotone projection p is defined as follows:

$$p(\{(v,i),(w,j)\}) := \begin{cases} \{(v,i),(w,j)\} & \text{if } \{v,w\} \in E(P), \\ 1 & \text{if } \{v,w\} \in E(P') \setminus E(P). \end{cases}$$

Thus, p^* takes an input G to the problem $\textsc{Subgraph}_{\mathsf{col},n}(P)$ and converts it into an input G' to $\textsc{Subgraph}_{\mathsf{col},n}(P')$ by filling in complete bipartite graphs between $\{v\} \times [n]$ and $\{w\} \times [n]$ for all new edges $\{v,w\} \in E(P') \setminus E(P)$.

Contraction Case. Now suppose P is obtained from P' by contracting a single edge $\{x,y\}$ where x, y have no common neighbors. Let z label the contracted vertex in P, so that $V(P) \setminus V(P') = \{z\}$ and $V(P') \setminus V(P) = \{x,y\}$. Let $\rho : V(P') \to V(P)$ be the function $x, y \mapsto z$ and $v \mapsto v$ for all $v \in V(P') \setminus \{x,y\}$. For $e = \{v,w\} \in E(P') \setminus \{x,y\}$, let $\rho(e) := \{\rho(v), \rho(w)\} \in E(P)$ ($\rho(\{x,y\})$ is undefined).

For (1): Consider any $(\alpha, \beta) \in \theta_{\mathsf{col}}(P)$. Define $\alpha' : V(P') \to [0,1]$ and $\beta' : E(P') \to [0,2]$ by

$$\alpha'(v) := \alpha(\rho(v)), \qquad \beta'(e) := \begin{cases} \alpha(z) & \text{if } e = \{x,y\}, \\ \beta(\rho(e)) & \text{otherwise.} \end{cases}$$

We now check that $(\alpha', \beta') \in \theta_{\mathsf{col}}(P')$ and $\kappa_{\alpha',\beta'}(P') \geq \kappa_{\alpha,\beta}(P)$. For that consider the mapping $\widehat{\rho} : Q' \mapsto \rho(Q' \setminus \{\{x,y\}\})$ that takes subgraphs of P' to subgraphs of P. It is easy to see that $\alpha'(Q') - \beta'(Q') \geq \alpha(\widehat{\rho}(Q')) - \beta(\widehat{\rho}(Q'))$, and that this is tight for $Q' = P'$: in the only non-trivial case $\{x,y\} \in E(Q')$ we have $\alpha(\widehat{\rho}(Q')) = \alpha'(Q') - \alpha(z)$ and $\beta(\widehat{\rho}(Q')) = \beta'(Q') - \alpha(z)$. This proves the first claim $(\alpha', \beta') \in \theta_{\mathsf{col}}(P')$. To see that $\kappa_{\alpha,\beta}(P) \leq \kappa_{\alpha',\beta'}(P')$, it suffices to observe that $\widehat{\rho}$ takes union sequences for P' into union sequences for P and thus $\widehat{\rho}^{-1}$ maps hitting sets for P into hitting sets for P'.

For (2): This time the monotone projection p is defined by

$$p(\{(v,i),(w,j)\}) := \begin{cases} 1 & \text{if } \{v,w\} = \{x,y\} \text{ and } i = j, \\ 0 & \text{if } \{v,w\} = \{x,y\} \text{ and } i \neq j, \\ \{(\rho(v),i),(\rho(w),j)\} & \text{otherwise.} \end{cases}$$

(That is, p^* duplicates $\{z\} \times [n]$ into two sets $\{x\} \times [n]$, $\{y\} \times [n]$ and then plants a perfect matching between twins.) This p is clearly a monotone projection from $\textsc{Subgraph}_{\mathsf{col},n}(P)$ to $\textsc{Subgraph}_{\mathsf{col},n}(P')$. □

4.5.1 Negative Results in the Uncolored Setting

In the colored setting, we have seen that $\textsc{Subgraph}_{\mathsf{col}}(P)$ is minor-monotone via linear-size monotone projections. Highlighting a difference between the uncolored and colored settings, we now show that there is *no* monotone projection whatsoever that reduces $\textsc{Subgraph}(M_3)$ to $\textsc{Subgraph}(P_3 + M_2)$ (where P_3 is a path on 3 vertices and M_k is a matching with k edges). While it remains an open problem whether $C(P)$ is (even approximately) minor-monotone under general AC^0 reductions, this result strongly suggests that the colorful version of the subgraph isomorphism problem is much better structured and well-behaved than the standard (uncolored) one.

We begin with some properties of $P_3 + M_2$-free graphs.

Lemma 155. *Every $P_3 + M_2$-free graph G satisfies one of the following conditions:*

1. *G has $\leq C$ edges for some absolute constant C,*
2. *G is a matching,*
3. *G contains a vertex of degree ≥ 6.*

Note: Lemma 155 is true with any integer replacing 6 in (iii), for an appropriate constant C in (i). The choice of 6 is what we need in the proof of Theorem 158 later on.

Proof. Assume G is $P_3 + M_2$-free, not a matching, and has maximum degree ≤ 5. We will show that G has $O(1)$ edges. Since G is not a matching, it contains a vertex u of degree ≥ 2. Since G has maximum degree ≤ 5, there is a constant C' such that if G has $> C'$ non-isolated vertices, then it contains non-isolated vertices v, w such that any two of u, v, w have distance ≥ 3; in that case, we would have $P_3 + M_2$-subgraph of G by taking any two edges containing u plus any two edges containing v and w respectively. Therefore, G has $\leq C'$ non-isolated vertices. It follows that G has $\leq 5C'/2$ edges. □

Lemma 156. *If G is $P_3 + M_2$-free and contains an M_4-subgraph, then G is a matching.*

Proof. Suppose G contains an M_4-subgraph H, but G is not a matching. We will show that G contains a $P_3 + M_2$-subgraph. Since G is not a matching, it contains a P_3-subgraph K. If K is vertex-disjoint from H, then K plus any two edges from H is a $P_3 + M_2$-subgraph of G. Now assume that K contains a vertex in H. Then there is a P_3-subgraph K' which contains an edge in H. This K' is vertex-disjoint from at least two edges in H; then K' plus these two edges is a $P_3 + M_2$-subgraph of G. □

Lemma 157. *Suppose G contains a $P_3 + M_2$-subgraph and a vertex u of degree ≥ 6. Then G contains a $P_3 + M_2$-subgraph in which u is the degree-2 vertex.*

Proof. Let H be any $P_3 + M_2$-subgraph of G. H contains an M_2-subgraph H' which does not include the vertex u. Since u has degree ≥ 6, it has two distinct neighbors v and w such that $\{u, v, w\} \cap V(H') = \emptyset$. Then H' plus edges $\{u, v\}$ and $\{u, w\}$ is a $P_3 + M_2$-subgraph of G in which u is the degree-2 vertex. □

Now the main result of this subsection:

Theorem 158. $\textsc{Subgraph}(M_3)$ *is not a monotone projection of* $\textsc{Subgraph}(P_3 + M_2)$.

Proof. Toward a contradiction, assume there exists a monotone projection $p : \binom{[N]}{2} \to \binom{[n]}{2} \cup \{0, 1\}$ from $\textsc{Subgraph}(M_3)$ on n-vertex graphs to $\textsc{Subgraph}(P_3 + M_2)$ on N-vertex

graphs for some $n, N \in \mathbb{N}$ where $n \geq C + 2$ with C the constant from Lemma 155. That is, for every graph G with vertex set $[n]$, we have

$$G \text{ contains an } M_3\text{-subgraph} \Leftrightarrow p^*(G) \text{ contains a } P_3 + M_2\text{-subgraph}$$

where $p^*(G)$ is the graph with edge set $p^{-1}(E(G) \cup \{1\})$. Note that since the predicate in the left-hand side essentially depends on all variables $e \in \binom{[n]}{2}$, so must the function p^* or, in other words, $p^{-1}(e)$ is non-empty for any $e \in \binom{[n]}{2}$.

For $a \in [n]$, let S_a denote the n-vertex star centered at a (i.e., with edge set $\{e \in \binom{[n]}{2} : a \in e\}$). Let $F_a := p^{-1}(S_a)$ (so that $p^*(S_a)$ is the disjoint union of F_a and $p^{-1}(1)$). Over the next few claims, we will show that F_a are stars of degree ≥ 6 with distinct centers. Since all $p^{-1}(e)$ are non-empty, F_a contains at least $n - 1$ ($> C$) edges.

Since S_a is M_3-free, $p^*(S_a)$ is $P_3 + M_2$-free, hence F_a is $P_3 + M_2$-free. By Lemma 155, it follows that either F_a is a matching or F_a contains a vertex of degree ≥ 6. The next claim eliminates the first possibility.

Claim 159. *For every $a \in [n]$, F_a is not a matching.*

Proof. (of the Claim) For contradiction, assume F_a is a matching for some $a \in [n]$. Consider any $b \in [n]$. Note that $S_a \cup S_b$ is M_3-free, hence $p^*(S_a \cup S_b)$ is $P_3 + M_2$-free. Since $p^*(S_a \cup S_b) \supseteq F_a \cup F_b$ and F_a contains a M_4-subgraph, Lemma 156 implies that $F_a \cup F_b$ is a matching. In particular, F_b is a matching for any $b \in [n]$, and we can repeat the above argument with $a := b$ to conclude that $F_b \cup F_c$ is a matching for *all* $b, c \in [n]$. Therefore, the entire pre-image $p^{-1}(K_n)$ is a matching, where K_n is the complete graph on vertices $[n]$.

Since K_n contains an M_3-subgraph, $p^*(K_n)$ ($= p^{-1}(K_n) \cup p^{-1}(1)$) contains a $P_3 + M_2$-subgraph. It follows that either $p^{-1}(1)$ contains a path of length 2, or $p^{-1}(1)$ contains an edge with an endpoint in $V(p^{-1}(K_n))$. In both cases we get a contradiction, as it follows that $p^*(S_c)$ contains a $P_3 + M_2$-subgraph for some $c \in [n]$, even though S_c is M_3-free. (If $p^{-1}(1)$ contains a P_3-subgraph, then any $c \in [n]$ will do; if $p^{-1}(1)$ contains an edge with an

endpoint $v \in V(p^{-1}(K_n))$, then any $c \in [n]$ with $v \in V(F_c)$ will do.) $\square$

For all $a \in [n]$, we have established that F_a is $P_3 + M_2$-free, has $> C$ edges and is not a matching. By Lemma 155, we conclude that F_a contains at least one vertex of degree ≥ 6. Let us now fix a function $z : [n] \to [N]$ such that $z(a)$ is a vertex of degree ≥ 6 in F_a for all $a \in [n]$.

Claim 160. *z is (≤ 2)-to-1.*

Proof. (of the Claim) For contradiction, assume there exist distinct $a, b, c \in [n]$ such that $v := z(a) = z(b) = z(c)$. By Lemma 157, $p^*(S_a \cup S_b \cup S_c)$ contains a $P_3 + M_2$-subgraph in which v is the degree-2 vertex. Let $e, f \in \binom{[N]}{2}$ be the two edges in this subgraph which are not adjacent to v. Without loss of generality, $\{e, f\} \subseteq p^*(S_a \cup S_b)$. Since v has degree ≥ 6 in $p^*(S_a \cup S_b)$, we can find a different path of length 2 through v which is vertex-disjoint from edges e and f. Therefore, $p^*(S_a \cup S_b)$ contains a $P_3 + M_2$-subgraph. Since $S_a \cup S_b$ is M_3-free, this contradicts our assumption about p. $\square$

Claim 161. *F_a is a star with center $z(a)$ for all $a \in [n]$.*

Proof. (proof of the Claim) For contradiction, assume F_a is not a star with center $z(a)$. Then F_a contains an edge e with $z(a) \notin e$. Since z is (≤ 2)-to-1, there exists $b \in [n]$ such that $z(b) \notin \{z(a)\} \cup e$. We may find a $P_3 + M_2$-subgraph within $F_a \cup F_b$ by taking e together with a disjoint path of length 2 through $z(a)$ and a disjoint edge containing $z(b)$. This contradicts the fact that $p^*(S_a \cup S_b)$ is $P_3 + M_2$-free. $\square$

Claim 162. *z is 1-to-1.*

Proof. (of the Claim) For contradiction, assume $v := z(a) = z(b)$ for some $a \neq b$. Let $c \in [n] \setminus \{a, b\}$. Then $z(c) \neq v$ and $p^*(S_a \cup S_b \cup S_c) = F_a \cup F_b \cup F_c \cup p^{-1}(1)$ contains a $P_3 + M_2$-subgraph H. We may assume that H contains edges $\{v, u\} \in E(F_a) \setminus E(F_b)$ and $\{v, w\} \in E(F_b) \setminus E(F_a)$ since otherwise H would be a subgraph of either $p^*(S_a \cup S_c)$ or $p^*(S_b \cup S_c)$ contradicting $P_3 + M_2$-freeness of these graphs. Note that $u \neq w$. Since v has

degree ≥ 6 in F_a, we can find an edge $\{v, w'\} \in E(F_a)$ such that $w' \notin V(H)$. Let H' be the graph obtained by substituting the edge $\{v, w'\}$ for $\{v, w\}$. Then H' is a $P_3 + M_2$-subgraph of $p^*(S_a \cup S_c)$, which is again a contradiction. □

At this point, we have established that graphs F_a ($a \in [n]$) are stars of degree ≥ 6 with distinct centers.

Claim 163. $|p^{-1}(e)| = 1$ *for all* $e \in \binom{[n]}{2}$.

Proof. (of the Claim) Suppose $e = \{a, b\}$. Since F_a and F_b are stars with different centers and $p^{-1}(e) \subseteq F_a \cap F_b$, we conclude $|p^{-1}(e)| \leq 1$. Since $p^{-1}(e)$ is nonempty, it follows that $|p^{-1}(e)| = 1$. □

Claim 164. $p^{-1}(1)$ *is nonempty.*

Proof. (of the Claim) Let G be any copy of M_3 (i.e. any three disjoint edges) among n-vertices. Then $p^{-1}(G)$ has only three edges by Claim 163. Since $p^*(G) = p^{-1}(G) \cup p^{-1}(1)$ has $P_3 + M_2$-subgraph, it contains at least 4 edges. Therefore, $p^{-1}(1)$ is nonempty. □

Fix any edge e in $p^{-1}(1)$ and any $a \neq b \in [n]$ such that $z(a), z(b) \notin e$. Then $p^*(S_a \cup S_b)$ contains a $P_3 + M_2$-subgraph, even though $S_a \cup S_b$ is M_3-free. This, finally, is the contradiction which completes the proof of Theorem 158. □

4.6 Conclusion and Open Problems

With the results of this chapter, the state of knowledge on the average/worst-case AC^0 complexity of the uncolored/colorful P-subgraph isomorphism problem now stands as

$$\Omega(\tfrac{tw(P)}{\log tw(P)}) \leq \kappa_{\mathsf{col}}(P) \leq C_{\mathsf{col}}(P) \leq tw(P)+1$$

$$\vee|$$

$$C(P)$$

$$\vee|$$

$$\kappa(P) \leq C_{\mathsf{ave}}(P) \leq 2\kappa(P)+O(1).$$

We have examples showing that the gap between $C_{\mathsf{ave}}(P)$ and $C(P)$ (i.e. the average-case vs. worst-case AC^0 complexity of $\textsc{Subgraph}(P)$) can be arbitrarily large (see Remark 117). We do not know of any gap between $C(P)$ and $C_{\mathsf{col}}(P)$. Equivalently, we can ask whether $C(P)$ is bounded from below by any function of $tw(P)$. Restating Question 1 from the introduction:

Question 165. *Is it possible to give general lower bounds on the worst-case* AC^0 *complexity of* $\textsc{Subgraph}(P)$ *(uncolored P-subgraph isomorphism) in terms of the treewidth of P only?*

When P is a core, we know that $C(P) = C_{\mathsf{col}}(P) = \widetilde{\Theta}(tw(P))$. At the opposite end of the spectrum, Question 1 is wide open for bipartite patterns P.

The next two questions seek to improve the parameters in our main results.

Question 166. *Can the upper bound $C_{\mathsf{ave}}(P) \leq 2\kappa(P)+O(1)$ of Theorem 110 be improved to $\kappa(P)+O(1)$?*

Question 167. *Can the* $\log tw(P)$ *factor be eliminated from our lower bounds on $\kappa_{\mathsf{col}}(P)$ (Theorem 109) or at least $C_{\mathsf{col}}(P)$?*

We are able to answer Question 167 affirmatively in the special case where P is a constant-degree expander (Theorem 152).

Another question raised by this work is whether the AC^0 complexity of $\textsc{Subgraph}(P)$ is monotone with respect to minors or subgraphs. In contrast to the colorful setting, we showed that monotone projections (the simplest form of reduction) fail to give any reduction whatsoever from $\textsc{Subgraph}(Q)$ to $\textsc{Subgraph}(P)$, even when Q is only a subgraph of P.

Question 168. *Is $C(P)$ minor-monotone or at least monotone under subgraphs?*

More modestly, if Q is a minor (or subgraph) of P, is there a reduction from $\textsc{Subgraph}(Q)$ to $\textsc{Subgraph}(P)$ by AC^0-circuits of size $O(n^c)$ for a constant c independent of P and Q? That would imply $C(Q) \leq O(C(P))$; currently we do not know if $C(Q)$ can be bounded by *any* function in $C(P)$.

Finally, it would be interesting to investigate the relationship between $\kappa_{\mathsf{col}}(P)$ and the complexity of $\textsc{Subgraph}_{\mathsf{col}}(P)$ beyond AC^0. In particular, we recall the result of Marx [57] that $\textsc{Subgraph}_{\mathsf{col}}(P)$ has no $n^{o(tw(P)/\log tw(P))}$-time algorithm unless the Exponential Time Hypothesis (ETH) fails. Follow-up work of Alon and Marx [5] looked at the question of removing the $\log tw(P)$ factor loss in the exponent of this result (toward the goal of showing that $n^{\Theta(tw(P))}$ is the true complexity of $\textsc{Subgraph}_{\mathsf{col}}(P)$, at least assuming the ETH). Alon and Marx specifically identified constant-degree expanders as a case where "substantially different methods" are needed to eliminate the $\log tw(P)$ factor loss incurred by the reduction of [57]. In light of our lower bounds $C_{\mathsf{col}}(P) \geq \kappa_{\mathsf{col}}(P) = \Omega(|V(P)|)$ when P is a constant-degree expander, it becomes interesting to ask:

Question 169. *Can it be shown that $\textsc{Subgraph}_{\mathsf{col}}(P)$ has no $n^{o(\kappa_{\mathsf{col}}(P))}$-time algorithm unless the ETH fails?*

REFERENCES

[1] Noga Alon and Michael Capalbo. Smaller explicit superconcentrators. In *Proceedings of the 14th Annual ACM-SIAM Symposium on Discrete Algorithms* (SODA 03), pages 340-346, 2003.

[2] Noga Alon and Ravi B. Boppana. The monotone circuit complexity of boolean functions. *Combinatorica,* 7(1):1-22, 1987.

[3] Rudolf Ahlswede, Ning Cai, Shuo-Yen Robert Li, and Raymond W. Yeung. Network information flow. *IEEE Transcations on Information Theory*, 46(4):1204-1216, 2000.

[4] Miklós Ajtai, Σ_1^1-formula on finite structures. *Annals Pure Applied Logic* 24, pp. 1-48, 1983.

[5] Noga Alon and Dániel Marx. Sparse balanced partitions and the complexity of subgraph problems. *SIAM Journal on Discrete Mathematics* 25 (2), 631-644, 2011.

[6] Noga Alon and Pavel Pudlák. Superconcentrators of depth 2 and 3; odd levels help (rarely). *J. Comput. System Sci.,* 48 (1994), pp. 194-202.

[7] Noga Alon and Joel H. Spencer. *The Probabilistic Method*. John Wiley & Sons, Inc., 2004.

[8] Noga Alon, Raphael Yuster and Uri Zwick. Color-coding. *J. ACM,* 42(4), 844-856, 1995.

[9] Kazuyuki Amano. k-Subgraph isomorphism on AC^0 circuits. *Computational Complexity,* 19(2):183-210, 2010.

[10] Andrew D. Barbour, Lars Holst and Svante Janson. *Poisson Approximation.* Oxford University Press, Oxford, UK, 1992.

[11] Eli Ben-Sasson and Swastik Kopparty. Affine dispersers from subspace polynomials. In *Proceedings of the Annual Symposium on Theory of Computing (STOC)*, vol. 679, pp. 65-74, 2009.

[12] Chris Beck, Russell Impagliazzo, and Shachar Lovett. Large deviation bounds for decision trees and sampling lower bounds for AC0-circuits. In *Proceedings of the 53rd IEEE Symposium on Foundations of Computer Science,* 2012.

[13] Norbert Blum. A Boolean function requiring $3n$ network size. *Theoretical Computer Science*, 28 (1984), 337-345.

[14] Louay M.J. Bazzi and Sanjoy K. Mitter. Endcoding Complexity Versus Minimum Distance. *IEEE Transcations on Information Theory*, vol. 51, no. 6, pp. 2103-2112, 2005.

[15] Hans L. Bodlaender. Discovering treewidth. In *Proceedings of the 31st International Conference on Current Trends in theory and Practice of Computer Science*, volume 3381 of *Lecture Notes in Computer Science*, pages 1–16. Springer, 2005.

[16] Béla Bollobás and John C. Wierman. Subgraph counts and containment probabilities of balanced and unbalanced subgraphs in a large random graph. *Annals of the New York Academy of Sciences,* 576: 63-70, 1989.

[17] Chandra Chekuri and Julia Chuzhoy. Polynomial bounds for the grid-minor theorem. In *Proceedings of the 46th Annual Symposium on the Theory of Computing (STOC)*, pages 60-69, 2014.

[18] Hubie Chen, Moritz Müler. One hierarchy spawns another: graph deconstructions and the complexity classification of conjunctive queries. In *Proceedings of the Joint CSL-LICS Meeting*, pages 1-10, 2014.

[19] Benny Chor, Oded Goldreich, Johan Hastad, Joel Freidmann, Steven Rudich, and Roman Smolensky. The bit extraction problem and t-resilient functions. In *26th, IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 396-407, 1985.

[20] Dmitriy Yu. Cherukhin. Lower bounds for Boolean circuits with finite depth and arbi-

trary gates. *Electronic Colloquium on Computational Complexity (ECCC)*, TR08-032, 2008.

[21] Danny Dolev, Cynthhia Dwork, Nicholas Pippenger, and Avi Wigderson. Superconcentrators, generalizers and genearlized connectors with limited depth. In *Proceedings ACM Symposium on Theory of Computing (STOC),* ACM, New York, 1983, pp. 42-51.

[22] Roland L. Dobrushin, S. I. Gelfand, and Mark S. Pinsker. On the complexity of coding. *Proc. 2nd Internat. Symp. on Information Theory,* pages 174-184, 1973.

[23] Andrew Drucker and Yuan Li. Conservative circuits and routing nxetowrks. Manuscript, 2017.

[24] Andrew Drucker and Yuan Li. Minimum depth required to compute good codes in linear size. Manuscript, 2017.

[25] Ding-Zhu Du and Hung Q. Ngo (Editors). Switching Networks: Recent Advances, Kluwer Academic Publishers, 2001.

[26] Devdatt Dubhashi and Desh Ranjan. Balls and bins: a study in negative dependence. *Random Struct. Algorithms,* 13:99-124, September 1998.

[27] Andrew Drucker. Limitations of Lower-bound methods for the wire complexity of boolean operators. In *IEEE Conference on Computational Complexity (CCC)*, 2012.

[28] David Eppstein. Subgraph isomorphism in planar graphs and related problems. *Journal of Graph Algorithms and Applications*, 3(3):1–27, 1999.

[29] David Eppstein. Diameter and treewidth in minor-closed graph families. *Algorithmica*, 27:275–291, 2000.

[30] Magnus G. Find, Alexander Golovnev, Edward A. Hirsch, and Alexander S. Kulikov. A better-than-3n lower bound for the circuit complexity of an explicit function. In *57th Annual IEEE Symposium on Foundations of Computer Science, 2016.*

[31] Uriel Feige, Mohammadtaghi Hajiaghayi and James R. Lee. Improved approximation algorithms for minimum weight vertex separators. *SIAM Journal on Computing*, 38(2), 629-657, 2008.

[32] Merrick Furst, James B. Saxe, and Michael Sipser, Parity, circuits, and the polynomial time hierarchy. *Proceedings of 22nd Annual IEEE Symposium on Foundations of Computer Science*, 1981, pp. 260-270.

[33] Venkatesan Guruswami and Piotr Indyk. Linear-time encodable/decodable codes with near-optimal rate. *IEEE Transactions on Information Theory,* 51(10):3393-3400, 2005.

[34] Anna Gál, Kristoffer Arnsfelt Hansen, Michal Koucký, Pavel Pudlák, Emanuele Viola. Tight bounds on computing error-correcting codes by bounded-depth circuits with arbitrary gates. *IEEE Transcations on Information Theory,* vol. 59, no. 10, Oct. 2013.

[35] Martin Grohe and Dániel Marx. On tree width, bramble size, and expansion. *J. Comb. Theory, Ser. B,* 99(1):218-228, 2009.

[36] Oded Goldreich. Three XOR-Lemmas — an exposition. *Electronic Colloquium on Computational Complexity,* TR 95-050, 1995.

[37] Martin Grohe. The complexity of homomorphism and constraint satisfaction problems seen from the other side. *Journal of the ACM*, 54(1):1-24, 2007.

[38] Oded Goldreich and Avishay Tal. Matrix rigidity of random toeplitz matrices. *Electronic Colloquium on Computational Complexity (ECCC),* TR15-079, 2015.

[39] Venkatesan Guruswami, Christopher Umans, and Salil P. Vadhan. Unbalanced expanders and randomness extractors from parvaresh-vardy codes. *J. ACM,* vol. 56, no. 4, 2009.

[40] Oded Goldreich and Awi Wigderson. On the size of depth-three boolean circuits for

computing multilinear functions. *Electronic Colloquium on Computational Complexity (ECCC),* 20:43, 2013.

[41] Johan Håstad. *Computational limitations of small-depth circuits.* MIT press, 1987.

[42] Tracey Ho, Muriel Mdard, Ralf Koetter, David R. Karger, Michelle Effros, Jun Shi, and Ben Leong. A random linear network coding approach to multicast. *IEEE Transactions on Information Theory*, vol. 52, no. 10, pp. 4413-4430, 2006.

[43] Johan Håstad, Ingo Wegener, Norbert Wurm and Sang-Zin Yi. Optimal depth, very small size circuit for symmetric functions in AC^0. *Information and Computation,* 108(2):200-211, 1994.

[44] Kazuo Iwama and Hiroki Morizumi. An explciit lower bound of $5n - o(n)$ for boolean circuits. In *Springer LNCS*, vol. 2420, pp. 353-364.

[45] Svante Janson. Poisson approximation for large deviations. *Random Structures and Algorithms,* 1(2):221-230, 1990.

[46] Svante Janson. Coupling and poisson approximation. *Acta Appl. Math.* 34, 7-15, 1994.

[47] Svante Janson, Tomasz Luczak and Andrzej Rucinski. *Random Graphs.* Wiley-Interscience, 2000.

[48] Kumar Joag-Dev, and Frank Proschan, Negative association of random variables with applications. *Annals of Statistics,* 11:4, pp. 286-295, 1983.

[49] Stasys Jukna and Georg Schnitger. Circuits with arbitrary gates for random operator. arXiv:1004.5236, 2010.

[50] Stasys Jukna. Entropy of operators or why matrix multiplication is hard for depth-two circuits. *Theory of Comput. Syst.*, 46(2):301-310, 2010.

[51] Stasys Jukna. Boolean Function Complexity: Advances and Frontiers. Algorithms and Combinatorics Series, #27. Springer-Verlag New York, LLC, 2012.

[52] Mauricio Karchmer and Avi Wigderson. Monotone circuits for connectivity require super-logarithmic depth. *SIAM J. Discrete Mathematics,* 3(2):255-265, 1990.

[53] Shachar Lovett and Emanuele Viola. Bounded-depth circuits cannot sample good codes. In *Conference on Computational Complexity (CCC)*, 2011.

[54] Yuan Li, Alexander Razborov, and Benjamin Rossman. On the AC0 Complexity of Subgraph Isomorphism. In *Proceedings of the 55th Annual IEEE Symposium on Foundations of Computer Science (FOCS),* 2014, pages 344-353. Extended version to appear in *SIAM Journal on Computing.*

[55] Shuo-Yen Robert Li, Raymond W. Yeung, and Ning Cai. Linear network coding. *IEEE Trans. Inf. Theory,* vol. 49, no. 2, pp. 371-381, Feb. 2003.

[56] Satyanarayana V. Lokam, Complexity lower bounds using linear algebra. *Foundations and Trends in Theoretical Computer Science*, 4(1-2), 1-155.

[57] Dániel Marx. Can you beat treewidth? *Theory of Computing*, 6, 85-112, 2010.

[58] Dániel Marx and Michał Pilipczuk. Everything you always wanted to know about the parameterized complexity of Subgraph Isomorphism (but were afraid to ask). In *Proc. 31st International Symposium on Theoretical Aspects of Computer Science*, 542-553, 2014.

[59] Florence J. MacWilliams and Neil Sloane. The theory of error correcting codes. New York: North-Holland, 1977.

[60] Kotaro Nakagawa and Osamu Watanabe. Gap between two combinatorial measures for constant depth circuit complexity of subgraph isomorphism. Technical Report, Tokyo Institute of Technology, 2011.

[61] Jaroslav Nešetřil and Patrice Ossona de Mendez. Linear time low tree-width partitions and algorithmic consequences. In *Proc.* 38*th ACM Symposium on the Theory of Computing*, 391–400, 2006.

[62] Wolfgang J. Paul. A 2.5n-lower bund on the combinational complexity of Boolean functions. *Proc. 7th ACM Symp. on Theory of Computing*, 1975, 27-36.

[63] Mark S. Pinsker. On the complexity of a concentrator. *Proc. 7th Internat. Teletraffic Conf.,* Stockholm, 1973, pp. 318/1-318/4.

[64] Pavel Pudlák, Vojtech Rödl, and Jiri Sgall. Boolean circuits, tensor ranks, and communication complexity. *SIAM J. Comput.,* 26(3):605-633, 1997.

[65] Pavel Pudlák. Communication in bounded depth circuits. *Combinatorica,* 14 (1994), pp. 203-216.

[66] Jürgen Plehn and Bernd Voigt. Finding minimally weighted subgraphs. *Graph-Theoretic Concepts in Computer Science*. Springer Berlin Heidelberg, 1991.

[67] Pavel Pudlák and Petr Savický. On shifting networks. Theoretical Computer Science 116 (1993), 415-419.

[68] Nicholas Pippenger and Leslie G. Valiant. Shifting graphs and their applications. *J. ACM*, 23:423-432, July 1976.

[69] Nicholas Pippenger and Andrew C.-C. Yao, Rearrangeable networks with limited depth. *SIAM Journal on Algebraic and Discrete Methods,* 3, 1982.

[70] Alexander Razborov. Lower bounds on the monontone complexity of some boolean functions. *Dokl. Akad. Nauk. SSSR,* 281(4):798-801, 1985.

[71] Alexander Razborov. Lower bounds on the size of bounded depth networks over a complete basis with logical addition. In *Matematicheskie Zametki,* 37(6):887-900, 1985.

[72] John Riordan and Claude E. Shannon. The number of two terminal series-parallel networks. *J. of Math. and Phys.* vol. 21, pp. 83-93, 1942.

[73] Benjamin Rossman. On the constant-depth complexity of k-clique. *Proc. 40th ACM Symposium on Theory of Computing,* 721-730, 2008.

[74] Benjamin Rossman. Average-case complexity of detecting cliques. Ph.D. thesis, MIT, 2010.

[75] Benjamin Rossman. Formulas vs. circuits for small disance connectivity. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing (STOC)*, pages 203-212, 2014.

[76] Neil Robertson and Paul D. Seymour. Graph minors X. Obstructions to tree-decomposition. *Journal of Combinatorial Theory* 52(2): 153-190, 1991.

[77] Ran Raz and Amir Shpilka. Lower bounds for matrix product in bounded depth circuits with arbitrary gates. *SIAM J. Comput.*, 32(2):488-513, 2003.

[78] Jaikumar Radhakrishnan and Amnon Ta-Shama. Bounds for dispersers, extractors, and depth-two superconcentrators, *SIAM J. Discrete Math.*, vol. 13, no. 1, pp. 2-24, 2000.

[79] Søren Riis. Information flows, graphs and their guessing numbers. Electron. J. Combin., vol. 14, no. 44, pp. 1-17, 2007.

[80] Claude E. Shannon. The synthesis of two-terminal switching circuits. *Bell Systems Technical Jounal.* 28, pp. 59-98.

[81] Roman Smolensky. Algebraic methods in the theory of lower bounds for Boolean circuit complexity. In *STOC*, pp. 77-82, 1987.

[82] Daniel A. Spielman. Linear-time encodable and decodable error-correcting codes. *IEEE Transactions on Information Theory*, vol. 42, no. 6, pp. 1723-1731, 1996.

[83] Michael Sipser and Daniel A. Spielman. Expander codes. *IEEE Transactions on Information Theory,* 42(6):1710-1722, 1996.

[84] Leslie G. Valiant. Graph-theorectic arguments in low-level complexity. In *Proceedings of the 6th Symposium on Mathematical Foundations of Computer Science,* pages 162-176, 1977.

[85] Andrew Yao. Probabilistic computations: Toward a unified measure of complexity. *Proc. 18th IEEE Symposium on Foundations of Computer Science,* 222-227, 1977.

[86] Andrew Yao. Separating the polynomial-time hierarchy by oracles. *Proceedings 26th Annual IEEE Symposium on Foundations of Computer Science*, 1985, pp. 1-10.