Foundations and Trends® in
Theoretical Computer Science
Vol. 9, No. 1 (2013) 1–123

DOI: 10.15610400000063

now
the essence of knowledge

Complexity of Linear Boolean Operators

Stasys Jukna
Vilnius University, Lithuania and Frankfurt University, Germany
jukna@online.de

Igor Sergeev
Lomonosov Moscow State University, Russia
isserg@gmail.com

see the list of errata and supplementary materials on the book homepage:
http://web.vu.lt/mif/s.jukna/Knizka

Contents

Abstract

Given a linear boolean operator, how to compute it by a small circuit using only unbounded fanin addition gates? Since this is one of the simplest and most basic circuit models, the question was considered by many authors since early 1950s. This led to a variety of upper and lower bound arguments—ranging from algebraic (determinant and matrix rigidity), to combinatorial (Ramsey properties, coverings and decompositions) to graph-theoretic (the superconcentrator method).

We give a throughout survey of the research in this direction, and prove some new results to fill out the picture. The focus is on the cases when the addition operation is either the boolean OR or XOR, but the model in which arbitrary boolean functions are allowed as gates is considered as well.

S. Jukna and I. Sergeev. *Complexity of Linear Boolean Operators*. Foundations and Trends® in Theoretical Computer Science, vol. 9, no. 1, pp. 1–123, 2013.

DOI: 10.15610400000063.

1

Introduction

Let $(S, +)$ be a commutative semigroup, that is, a set S closed under a binary "sum" operation $+$ which is associative and commutative. Our goal is to simultaneously compute a given system

$$y_i = \sum_{j \in T_i} x_j, \qquad i = 1, \ldots, m \tag{1.1}$$

of m sums by only using the sum operation of the semigroup. By identifying the subsets T_i with their characteristic 0/1 vectors, this system turns to a linear operator $y = Ax$ for a boolean matrix A.

A natural computational model towards this goal is that of *addition circuits* over $(S, +)$. Such a circuit is a directed acyclic graph with n input nodes $x_1, \ldots, x_n$ of zero fanin, and m output nodes $y_1, \ldots, y_m$ of zero fanout. Each non-input node computes the sum of its inputs over $(S, +)$. There are no restrictions on the fanin or fanout of gates. The *size* of a circuit is the total number of edges in it, and the *depth* is the length of (the number of edges in) a longest path.

We will concentrate on the most basic semigroups—the OR semigroup $(\{0, 1\}, \vee)$, and the XOR group $(\{0, 1\}, \oplus)$. Thus, OR circuits allow "cancellations" $x + x = x$ (partial sums can be "merged"), whereas XOR circuits allow cancellations $x + x = 0$ (partial sums can be "anni-

hilated"). We also consider a restricted model of SUM circuits where the system of sums (1.1) is computed over the semiring $(\mathbb{N}, +)$. In this model none of these two types of cancellations can be used. Note that the XOR and OR (and its "dual" AND) are the only commutative semigroups over $S = \{0, 1\}$.

We stress that, given a boolean matrix A, the goal of all these three types of circuits is the *same*: to compute the system of sums (1.1) defined by A. The only difference is in what type of cancellations a circuit can use to achieve this goal. OR circuits constitute the simplest *monotone* model, whereas XOR circuits constitute the simplest group model (necessarily non-monotone since the group is finite). SUM circuits are "universal" in the sense that every such circuit for A is an addition circuit for A over any semigroup $(S, +)$.

The model of OR circuits was first considered by Lupanov [62] by inventing the model of rectifier circuits. XOR circuits were first considered by Nechiporuk in [70]. SUM circuits were first explicitly introduced by Pippenger [81]. SUM circuits of fanin-2 are also known as "vector addition chains" (see, for example, Knuth [55, Sect. 4.6.3]).

It is important to note that computing an operator $y = Ax$ for a boolean matrix $A = (a_{ij})$ by an addition circuit actually means to "encode" the matrix A by paths in a directed acyclic graph. Namely, if p_{ij} denotes the number of paths from the input node x_j to the output node y_i in such a circuit for A then the circuit *implements* (or *encodes*) the matrix A in the following sense:

- SUM circuit: $p_{ij} = a_{ij}$.
- OR circuit: $p_{ij} > 0$ if $a_{ij} = 1$, and $p_{ij} = 0$ if $a_{ij} = 0$.
- XOR circuit: p_{ij} is odd if $a_{ij} = 1$, and p_{ij} is even if $a_{ij} = 0$.

Thus, SUM circuits constitute the most restricted model in which there cannot be more than one path between the same pair of input and output nodes. Also, unlike XOR circuits, SUM and OR circuits are *monotone* models: increasing values of inputs cannot decrease the values of outputs. For these circuits, large (almost quadratic) explicit[1] lower bounds, without any restriction on the circuit-depth are known.

[1]Intuitively, a matrix or a boolean function being "explicit" means being "explic-

However, XOR circuits are a "Waterloo" of circuit complexity: here superlinear lower bounds are only known for constant-depth circuits (and these are barely-superlinear even for depth 5, say).

In this text we survey the most important complexity-theoretic questions about the addition circuit model:

Q1: What is the maximum complexity of implementing a boolean $n \times n$ matrix? Answer: it is about $n^2/\log n$ in all three models (Chapter 2).

Q2: What are the best known explicit lower bounds for the three complexity measures? Answer: for SUM and OR circuits, we have near-optimal explicit examples of boolean $n \times n$ matrices with a lower bound of $n^{2-o(1)}$ (§ 3.4). On the other hand, we have nothing super-linear for XOR circuits, except for constant depth d, and these degrade badly as d grows. For depth 2, the strongest known lower bound is about $n(\ln n/\ln\ln n)^2$, and is about $n \ln\ln n$ for depth 3 (§ 3.7, § 3.8 and Chapter 6).

Q3: How large a gap can occur between the SUM, OR and XOR complexities of a given boolean $n \times n$ matrix A? Answer: the largest possible gap in each of the three models is $O(n/\log n)$ (Chapter 2). The largest known SUM/OR gap is $\Omega(\sqrt{n}/\log^2 n)$, OR/XOR gap is $\Omega(n/\log^2 n)$, and the largest known gap between the OR complexity of a matrix A and its complement is $\Omega(n/\log^3 n)$ (Chapter 5).

Q4: What are the most important known lower bound techniques for handling specific matrices, what are their limitations? A variety of techniques are described in Chapter 3 and Chapter 6. They give a flexible toolkit for lower-bounding the SUM and OR complexities, their bounded-depth analogues, and the depth-2 XOR complexity. Each of presented lower-bound techniques uses some property of matrices and gives some lower bound based on only these properties. Is the technique "optimal" in the sense that one cannot derive a larger bound by only using the same properties? We show various examples of this kind, indicating where progress on lower bounds gets stuck (Table 4.1, § 6.1 and § 6.2).

---

itly constructed", not just being "shown to exist". A more rigorous definition of the term "explicit" can be found, for example, in the book [50, Section 1.5.1].

Q5: XOR circuits are the "natural" way to compute linear operators over $\mathbb{F}_2$; but are they the "best" way? To address this question, we consider *general* circuits that allow *arbitrary* boolean functions at its gates. Despite this model's crazy power, we still don't know if there is any example where it computes an $\mathbb{F}_2$-linear operator more efficiently than XOR circuits do. Moreover, some of our lower bound techniques apply also to this stronger model, and we describe some of this work in Chapter 6.

For general circuits computing *linear* $\mathbb{F}_2$-operators, the strongest known explicit lower bounds have the form $\Theta(n(\ln n/\ln\ln n)^2)$ in depth 2, and the form $\Theta(n\ln\ln n)$ in depth 3; these bounds are tight and are achievable even by XOR circuits (see Chapter 6). This highlights the power of XOR circuits and difficulties of dealing with them. In larger depths, the known lower bounds for XOR circuits are only barely superlinear.

If we consider *non-linear* operators in the arbitrary gates model, then we have explicit $\Omega(n^{3/2})$ bounds in depth 2, and $\Omega(n\ln n)$ in depth 3. These bounds were proved by Cherukhin [18, 19] and Jukna [48] using entropy arguments which do not work for linear operators. In larger depths, the known bounds are only barely better than those known for linear operators.

Though organized as a survey, the text also contains some new, previously unpublished results. These include:

1. Hansel–Krichevski type lower bound (Theorem 3.5).
2. Rectangle-area lower bounds (Theorem 3.12).
3. Depth-2 lower bound for block matrices (Theorem 3.18(iii)).
4. Lower bound for Kronecker products (Theorem 3.20(ii)).
5. Bounds for the Kneser–Sierpinski matrix (Lemma 4.2).
6. Upper bounds for the Sylvester matrix (Theorem 4.3).
7. Balanced decomposition of the triangular matrix (Lemma 5.3).
8. Coverings vs. decompositions in depth 2 (Theorem 5.4).
9. An XOR/OR gap in depth 2 (Theorem 5.12).
10. Matrix/complement gaps (Theorem 5.13, items (i) and (iii)).
11. Linearization of half-linear depth-2 circuits (Lemma 7.17).

Most of the remaining (known) results are given with proofs—in most

cases, substantially simplified—or at least with detailed proof sketches. The subject of this survey previously found an only fragmentary exposition in the books by Wegener [108], Dunne [24], Jukna [50], and in an earlier very short survey by Lupanov [63].

What we do not cover To compute linear operators over fields $(S, +, \cdot)$, and in particular over infinite fields, it is natural to allow multiplication by arbitrary field elements as a basic circuit operation. Such circuits are called *linear circuits*. If $S = \{0, 1\}$, then these are just the addition circuits considered in this survey. However, the ability to use "for free" arbitrarily complex coefficients of arbitrary magnitude is one of the central "mysteries" in arithmetic circuit complexity over infinite fields.

Research in this direction also has long history, starting with the seminal works of Morgenstern [67, 68], Grigoriev [37] and Valiant [106]. In this case, gates may compute arbitrary linear combinations of their inputs, not just 0/1 combinations. It is still an open problem to prove more than linear lower bounds on circuits computing a linear form Ax defined by an explicit 0/1 matrix A—such bounds are only known when either the matrix A has very "complicated" entries (say, square roots of the first n^2 distinct primes) or when the circuit is not allowed to use large coefficients; see, for example, the book by Bürgisser, Clausen, and Shokrollahi [13], or the more recent survey by Lokam [61].

1.1 Concepts used

We first recall some (mostly basic) concepts concerning boolean matrices which we will use later. A matrix is *boolean* if it only has entries 0 and 1. If not otherwise stated,

by a "matrix" we will always mean a "boolean matrix".

For such a matrix A, $|A|$ denotes the number of 1-entries in A. A *rectangle* in a matrix is an all-1 submatrix. If this is an $a \times b$ rectangle, then we define its *weight* as $a + b$, its *area* as $a \cdot b$, and its *density* as $a \cdot b/(a+b)$. For a positive integer r, $[r] = \{1, \ldots, r\}$ will always denote the set of the first r positive integers.

The *Kronecker product* $A \otimes B$ of a $p \times q$ matrix $A = (a_{i,j})$ and an $n \times m$ matrix B is an $np \times mq$ block-matrix obtained by replacing 1-entries of A by copies of B. The *direct sum* of matrices A and B is the matrix $A \boxplus B = \left[\begin{smallmatrix} A & 0 \\ 0 & B \end{smallmatrix}\right]$.

We can view a rectangle R in an $n \times n$ matrix A as an $n \times n$ matrix $\hat{R}$ with all entries outside R filled by 0s. A set $R_1, \ldots, R_s$ of rectangles in a matrix A is a:

- SUM covering (or a *decomposition*) of A if $A = \sum_{i=1}^{s} \hat{R}_i$;
- OR covering (or just a *covering*) of A if $A = \bigvee_{i=1}^{s} \hat{R}_i$;
- XOR covering of A if $A = \bigoplus_{i=1}^{s} \hat{R}_i$.

The *weight* of a covering is the sum of weights of its rectangles. For

$$\mathsf{L} \in \{\text{SUM,OR,XOR}\}\,,$$

the L-*rank* of A is the smallest number of rectangles in an L-covering of A. The L-*product* of two matrices is their product over the corresponding semiring. Thus, the L-rank of A is the smallest number r such that A can be written as an L-product $A = PQ^{\top}$, where P and Q are $n \times r$ matrices. To visually better distinguish the three ranks, we will use $\mathrm{rk}_{+}(A)$, $\mathrm{rk}_{\vee}(A)$ and $\mathrm{rk}(A)$ to denote, respectively, the SUM-, OR- and XOR-rank of A. In communication complexity,[2] $\log \mathrm{rk}_{\vee}(A)$ is exactly the *nondeterministic* communication complexity of A (see, e.g. [50, § 4.2]).

The *term rank*, $\mathrm{tr}(B)$, of a boolean matrix B is the largest number of its 1s, no two of which lie in the same row or column. By the König–Egerváry theorem, this is exactly the smallest number of rows and columns covering all 1s of B. It is easy to see that

$$\mathrm{tr}(B) \geqslant \mathrm{rk}_{+}(B) \geqslant \mathrm{rk}_{\vee}(B)\,.$$

Indeed, $\mathrm{tr}(B)$ is the smallest number $a+b$ such that, after some permutation of rows and columns, the matrix B can be written in the form $B = \left[\begin{smallmatrix} C & D \\ F & 0 \end{smallmatrix}\right]$, where C is an $a \times b$ matrix. We can therefore write B as a sum of $a + b$ pairwise disjoint rectangles, each corresponding to one row or column of B.

[2] If not specified otherwise, $\log n$ will always stand for $\log_2 n$.

A matrix A is (k,l)-*free* ($k,l \geqslant 1$) if it does not contain a $k \times l$ rectangle; being k-*free* means being (k,k)-free. Known upper bounds for the Zarankiewicz problem (see, for example, [58] or the book [8]) state that, if A is a (k,l)-*free* matrix of dimension $m \times n$, then

$$|A| \leqslant (k-1)^{1/l}(n-l+1)m^{1-1/l} + (l-1)m\,. \tag{1.2}$$

A matrix is (k,l)-*Ramsey matrix* if both the matrix and its complement are (k,l)-free.

We will often use the arithmetic-geometric mean inequality

$$\frac{1}{n}\sum_{i=1}^{n} x_i \geqslant \Big(\prod_{i=1}^{n} x_i\Big)^{1/n}, \tag{1.3}$$

as well as a special version of the Jensen inequality for a convex function f:

$$\sum_{i=1}^{n} f(x_i) \geqslant n \cdot f\Big(\frac{X}{n}\Big)\,, \tag{1.4}$$

where $X = \sum_{i=1}^{n} x_i$ and all $x_i \geqslant 0$. In particular, by taking $f(x) = x\log x$, we obtain

$$\sum_{i=1}^{n} x_i \log x_i \geqslant X \log \frac{X}{n}\,, \tag{1.5}$$

In some estimates we will also use the binary entropy function

$$H(\alpha) = \alpha \log \frac{1}{\alpha} + (1-\alpha)\log\frac{1}{1-\alpha}\,.$$

Asymptotic notation To spare parenthesis (in larger expressions), we will occasionally write $f \succcurlyeq g$ instead of $f = \Omega(g)$, $f \preccurlyeq g$ instead of $f = O(g)$, and $f \asymp g$ instead of $f = \Theta(g)$. Also, $f \ll g$ stands for $f = o(g)$. Notation $f \sim g$ means the usual (tight) asymptotic $f/g \to 1$. By saying "the $n \times n$ matrix A has complexity $\succcurlyeq g(n)$" we will actually mean that we have an infinite sequence $\{A_n\}$ of $n \times n$ matrices ($n = 1, 2, \ldots$) for which there exists a constant $\epsilon > 0$ such that "complexity of A_n is $\geqslant \epsilon g(n)$" holds for infinitely many n. By writing "A has complexity $\geqslant g(n)$", we will mean that this holds for all large enough dimensions n.

1.2 Simple observations

We denote the minimum number of edges in an OR, XOR and SUM circuit implementing a given matrix A by $\mathsf{OR}(A)$, $\mathsf{XOR}(A)$ and $\mathsf{SUM}(A)$. If we speak only about circuits of depth $\leqslant d$, then the corresponding measures are denoted by $\mathsf{OR}_d(A)$, $\mathsf{XOR}_d(A)$ and $\mathsf{SUM}_d(A)$.

As we noted above, SUM circuits constitute the weakest model: each such circuit can be turned into an OR circuit or an XOR circuit just by replacing the operations computed at their nodes. So, for every matrix A, we have that

$$\mathsf{OR}(A) \leqslant \mathsf{SUM}(A) \quad \text{and} \quad \mathsf{XOR}(A) \leqslant \mathsf{SUM}(A)\,.$$

In the case of depth-d circuits, we will assume that the underlying graph is "leveled" in the following sense. We have $d+1$ levels of nodes. The first level consists of input nodes, the last consists of output nodes, and each edge goes from one level to the next one. Thus, if A_i is the boolean adjacency matrix of the bipartite graph between the $(i+1)$-th and i-th levels, then these measures give the smallest weight $\sum_{i=1}^{d} |A_i|$ of the presentation of A as a product $A = A_d \cdot A_{d-1} \cdots A_1$ of boolean matrices over the corresponding semirings, where $|A_i|$ is the number of 1s in A_i. That is,

$$\mathsf{L}\text{-complexity of } A = \text{smallest weight of an } \mathsf{L}\text{-factorization of } A\,.$$

Observation 1.1 (Transposition principle)**.** The complexities of a matrix A and its transpose $A^\top$ are the same.

Proof. Given any circuit for A, one may reverse the direction of all edges to obtain a circuit for $A^\top$. □

Observation 1.2. The complexity of a submatrix is at most the complexity of the entire matrix.

Proof. Given a circuit for a matrix, we can remove all input and output nodes that are not in the submatrix. □

For counting reasons, it is sometimes convenient to transform the circuit so that every inner node (non-input node) has fanin at most 2,

and then count the nodes in a new circuit rather than the edges in the original one.

Observation 1.3. An unbounded fanin circuit with e edges and v non-input nodes can be turned into an equivalent fanin-2 circuit with $e - v$ nodes.

Proof. Just replace every node of fanin $d > 2$ by a binary tree with $d - 1$ inner nodes. The difference $e' - v'$ in the new circuit equals $e - v$ in the original circuit. See [50, Section 1.8] for more details. □

Depth-1 complexity is a trivial measure: we have $\mathsf{SUM}(A) \leqslant \mathsf{SUM}_1(A) = |A| \leqslant n^2$ for every $n \times n$ matrix A. Depth-2 circuits constitute the first non-trivial model. We already know that $\mathsf{L}_2(A) = \min\{|B| + |C| \colon A = B \cdot C\}$. Here and in what follows, $\mathsf{L}(A)$ stands for the SUM, OR or XOR complexity, and the matrix product is over the corresponding semiring. On the other hand, depth-2 circuits have also a *combinatorial* description in terms of coverings.

Observation 1.4. For every matrix A, $\mathsf{L}_2(A)$ is the minimum weight of an L-covering of A.

Proof. The paths going through one node on the middle level of a circuit for A define a rectangle in A. □

Let again $\mathsf{L}(A)$ stand for the SUM, OR or XOR complexity, and let $A + B$ and $A \cdot B$ denote the matrix sum and the matrix product over the corresponding semiring. Then we have:

1. $\mathsf{L}(A + B) \leqslant \mathsf{L}(A) + \mathsf{L}(B)$, if the matrices can be added;
2. $\mathsf{L}(A \cdot B) \leqslant \mathsf{L}(A) + \mathsf{L}(B)$, if the matrices can be multiplied;
3. $\mathsf{L}(A \boxplus B) \leqslant \mathsf{L}(A) + \mathsf{L}(B)$;
4. $\mathsf{L}(A \otimes B) \leqslant a \cdot \mathsf{L}(B) + b \cdot \mathsf{L}(A)$, if A has a rows, and B has b columns.

Only (4) needs a proof. First, we rewrite the Kronecker product as $A \otimes B = (I_a \otimes B)(A \otimes I_b)$, and observe that $A \otimes I_b = P(I_b \otimes A)Q$ for particular permutation matrices P and Q. Since, $I_a \otimes B = B \boxplus B \boxplus \cdots \boxplus B$ is a direct sum (a times), the desired inequality (4) follows from (2) and (3).

1.3 Some basic matrices

Let us recall the definitions of some basic matrices whose complexities we will investigate later. These matrices are well-suited to demonstrate known lower bound techniques. This section is just for later reference, so that the reader can safely skip it, and proceed with the next section.

Full triangular matrix The *full triangular* matrix T_n, known also as the *prefix matrix*, is an $n \times n$ matrix with 1s on the main diagonal and below it, and zeroes elsewhere. For $n = 2^r$, these matrices can be defined recursively as follows:

$$T_2 = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, \quad T_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix}, \quad T_{2n} = \begin{bmatrix} T_n & 0 \\ 1 & T_n \end{bmatrix}.$$

This gives the recursion $\mathsf{SUM}_2(T_n) \leqslant 2 \cdot \mathsf{SUM}_2(T_{n/2}) + n$, which results to

$$\mathsf{SUM}_2(T_n) \leqslant n \log n + n \,. \tag{1.6}$$

Complement of identity matrix To demonstrate some bounds, we will also use the complement $\overline{I}_n = T_n \oplus T_n^\top$ of the identity matrix I_n for $n = 2^r$. For this matrix, we have that

$$\mathrm{rk}_\vee(\overline{I}_n) \leqslant 2r = 2 \log n \quad \text{and} \quad \mathsf{OR}_2(\overline{I}_n) \leqslant 2r 2^r = 2n \log n \,. \tag{1.7}$$

To see this, label the rows and columns of $\overline{I}_n$ by vectors $u \in \{0,1\}^r$. For each position $i \in \{1, \ldots, r\}$, we have two rectangles: one consists of all pairs (u, v) such that $u_i = 0$ and $v_i = 1$, and the other consists of all pairs (u, v) such that $u_i = 1$ and $v_i = 0$. This way, we obtain a covering of $\overline{I}_n$ by $2r$ rectangles of total weight $4r2^{r-1} = 2n \log n$.

A general construction of some important $n \times n$ matrices, for $n = 2^r$ being a power of two, is the following. Label the rows and columns by distinct subsets u of $[r]$. The $n \times n$ matrix M_f induced by a function $f : \{0, 1, \ldots, r\} \to \{0, 1\}$ is then defined by: $M_f[u, v] := f(|u \cap v|)$.

Kneser–Sierpinski (disjointness) matrix In graph theory, the Kneser graph is the graph whose nodes correspond to the k-element subsets of a set of r elements, and where two nodes are adjacent if and only if the two corresponding sets are disjoint. Kneser graphs are named after Martin Kneser, who first investigated them in 1955.

By analogy, the *Kneser-Sierpinski $n \times n$ matrix* (known also as the *disjointness matrix*) $D = D_n$ is the f-intersection matrix induced by the function $f(x) = 1$ if and only if $x = 0$. That is, the rows and columns of $D = D_n$ with $n = 2^r$ are labeled by distinct subsets u of $[r]$, and $D[u, v] = 1$ if and only if $u \cap v = \emptyset$. These matrices can be defined inductively as follows:

$$D_2 = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, \quad D_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix}, \quad D_{2n} = \begin{bmatrix} D_n & 0 \\ D_n & D_n \end{bmatrix}. \tag{1.8}$$

The Kneser–Sierpinski matrix D is an important object. This matrix is also sometimes called the Sierpinski matrix, since it resembles the well-known "Sierpinski gasket". In particular, it gives a linear transformation between the vector of the values of a boolean function f and the vector of coefficients of its unique representation as multilinear polynomial over the 2-element field. This polynomial is also known as the *Zhegalkin polynomial* for f.

To see this, consider a boolean function $f : 2^{[r]} \to \{0, 1\}$ and its XOR-polynomial $f(X) = \bigoplus_{u \subseteq [r]} g(u) X_u$ with boolean coefficients $g(u)$, and $X_u = \prod_{i \in u} x_i$. Then the $2^r \times 2^r$ matrix D induces a linear mapping from the vector $(g(u) \colon u \subseteq [r])$ to the vector $(f(v) \colon v \subseteq [r])$: just note that $X_u(v) = 1$ if and only if $u \subseteq v$, or, in other words, if and only if $u \cap \overline{v} = \emptyset$. Moreover, the inverse map is also given by the matrix D, since $D = D^{-1}$ (easy to check).

Intersection matrix The *intersection $n \times n$ matrix* is the f-intersection matrix induced by the function $f(x) = 1$ if and only if $x > 0$. That is, the intersection matrix is just the complement $Đ_n = \overline{D_n}$ of the Kneser–Sierpinski matrix with $n = 2^r$. The rows and columns are labeled by distinct subsets u of $[r]$, and $Đ[u, v] = 1$ if and only if

$u \cap v \neq \emptyset$. These matrices also have a recursive definition:

$$\mathcal{D}_2 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad \mathcal{D}_4 = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \mathcal{D}_{2n} = \begin{bmatrix} \mathcal{D}_n & 1 \\ \mathcal{D}_n & \mathcal{D}_n \end{bmatrix}.$$

By identifying subsets u with their characteristic vectors, we see that $\mathcal{D}[u,v] = \bigvee_{i=1}^{r} u_i \wedge v_i$. Thus, over the boolean semiring, we have that $\mathcal{D} = B \cdot B^\top$ for the $n \times r$ matrix B whose rows are all vectors of length r. This yields

$$\mathsf{OR}_2(\mathcal{D}_n) \leqslant 2r2^{r-1} = n \log n\,. \tag{1.9}$$

In the *unique intersection matrix* $\mathcal{D}^u$ we have a stronger condition for 1s: $\mathcal{D}^u[u,v] = 1$ if and only if $|u \cap v| = 1$.

Sylvester matrices The Sylvester $n \times n$ matrix H_n for $n = 2^r$ is the $n \times n$ f-intersection matrix induced by the function $f(x) = x \mod 2$. That is, the rows and columns of H are labeled by distinct subsets u of $[r]$, and $H[u,v] = 1$ if and only if $|u \cap v|$ is odd. Sylvester matrices can be defined inductively as follows:

$$H_2 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \quad H_4 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix}, \quad H_{2n} = \begin{bmatrix} H_n & H_n \\ H_n & \overline{H}_n \end{bmatrix}. \tag{1.10}$$

By identifying subsets of $[r]$ with their characteristic vectors $u \in \{0,1\}^r$, we see that $H[u,v] = \langle u, v \rangle = u_1 v_1 \oplus u_2 v_2 \oplus \cdots \oplus u_r v_r$ is the scalar product of u and v over $\mathbb{F}_2$. Thus, H is just a "counting version" of the intersection matrix, and we have $H = B \cdot B^\top$ over $\mathbb{F}_2$ for the $n \times r$ matrix B whose rows are all n binary vectors of length r. This yields

$$\mathsf{XOR}_2(H_n) \leqslant 2r2^{r-1} = n \log n\,. \tag{1.11}$$

A basic property of Sylvester matrices is expressed by the following lemma, whose simple proof can be found, say, in [28, p. 88].

Lindsey's Lemma. The Sylvester $n \times n$ matrix contains no monochromatic $a \times b$ submatrices, unless $ab \leqslant n$.

We will show in § 3.3 that $\mathsf{OR}(A) \geqslant |A|/k^2$ holds for every k-free matrix A. Here we give some examples of such matrices.

Random matrices A random $n \times n$ matrix A, where each entry is drawn uniformly and independently from $\{0,1\}$, has $\Omega(n^2)$ ones, and is k-free for relatively small k. This holds because A fails to be k-free with probability at most $\binom{n}{k}^2 2^{-k^2} \ll \mathrm{e}^{2k\ln n - k^2}$: there are $\binom{n}{k}^2$ $k \times k$ submatrices, and the probability that all k^2 entries of a given $k \times k$ submatrix are 0s is 2^{-k^2}. For $k \geqslant 2\ln n$, this probability tends to 0 as $n \to \infty$. Thus, k-free $n \times n$ matrices A with $k = O(\log n)$ and $|A| = \Omega(n^2)$ exist.

Singer matrix [100] The upper bound (1.2) for the Zarankiewicz problem implies that no 2-free $n \times n$ matrix can have more than $n^{3/2}+n$ ones. On the other hand, there are several *explicit* constructions of 2-matrices with almost this number of 1s. One of the oldest construction is due to Singer [100].

For a prime power q, a projective plane $PG(2,q)$ has $n = q^2+q+1$ points and n subsets of points (called lines). Every point lies in $q+1$ lines, every line has $q+1$ points, any two points lie on a unique line, and any two lines meet in the unique point.

The point-line incidence matrix of a finite projective plane P was introduced by Singer [100]. Label rows by points x, columns by lines L, and let $P[x,L] = 1$ if and only if $x \in L$, then the obtained matrix is 2-free. The number of 1s is $|P| = (q+1)n > n^{3/2}$.

A 2-free matrix similar in spirit to Singer's was constructed by Kövari–Sós—Turán [58] and Nechiporuk [74]. This matrix is related to the point-line incidences in a finite *affine* plane. Here rows and columns correspond to pairs of numbers in $\mathbb{F}_q$, and each row (a,b) has 1s in positions $(x, ax-b)$ with $x \in \mathbb{F}_q$. Thus, $|A| = nq = q^3 = n^{3/2}$. The matrix is 2-free because every system of two equations $ax = b+y$ and $cx = d+y$ has at most one solution.

Circulant matrices A matrix is *circulant* if each its row is a cyclic shift (by one position to the left) of the previous one. Singer [100]

proved that his $n \times n$ matrices P with $n = q^2 + q + 1$ and q a prime power are circulant: there exists a subset $S \subseteq \{0, 1, \ldots, n-1\}$ of size $|S| = q+1$ such that (after permutation of rows and columns) we have that $P[x, y] = 1$ if and only if $y = x + a \mod n$ for some $a \in S$. The circulant property is significant for us because such matrices have small XOR complexity (see § 5.4).

We can define a circulant matrix by giving a subset $S = \{s_1, \ldots, s_k\}$ of $\mathbb{Z}_n = \{0, 1, \ldots, n-1\}$: these are the positions of 1s in the first row. The resulting circulant matrix A has $|A| = kn$ ones. Such a set is called a *Sidon set* if all differences modulo n of two its elements are distinct. It is well known (and not difficult to show) that, if the support S of a circulant matrix A is a Sidon set, then the matrix A is 2-free: the 1s of A stay on $|S|$ diagonals determined by position in S. It is known that no Sidon set can have more than $\sqrt{n} + 1$ elements. Explicit examples of Sidon sets S with $|S| \sim \sqrt{n}$ were given by Alexeev [1], Bose [9], Ruzsa [94], and other authors; see a survey by O'Bryant [77].

Norm matrices Let q be a prime-power, $t \geqslant 2$ an integer, and consider the field $\mathbb{F}_{q^t}$ with q^t elements. The norm of an element a of this field is defined as the element $\|a\| := a \cdot a^q \cdots a^{q^{t-1}} = a^{(q^t-1)/(q-1)}$ of this field. Now let $n = q^t$, and construct an $n \times n$ matrix $N = N_{n,t}$ whose rows and columns are labeled by elements of $\mathbb{F}_{q^t}$. The entries are defined by letting $N[a, b] = 1$ if and only if $\|a + b\| = 1$.

It is known that the number of solutions in $\mathbb{F}_{q^t}$ of the equation $\|x\| = 1$ is $(q^t - 1)/(q - 1)$; see e. g., the book by Lidl and Niederreiter [60]. Hence, each row of N has $r = (q^t - 1)/(q - 1)$ ones, implying that the total number of ones is $|N| = rq^t \geqslant q^{2t-1} = n^{2-1/t}$.

Kollár, Rónyai and Szabó [57] proved that, for every t distinct elements $a_1, \ldots, a_t$ of $\mathbb{F}_{q^t}$, the system of equations $\|a_1 + x\| = 1, \|a_2 + x\| = 1, \ldots, \|a_t + x\| = 1$ has at most $t!$ solutions $x \in \mathbb{F}_{q^t}$. This implies that the constructed matrix N has no $t \times (t! + 1)$ all-1 submatrix. Hence, the constructed matrix A is $(t, t! + 1)$-free. Explicit matrices with slightly worse parameters were constructed earlier by Andreev [7].

2

General Upper Bounds

The upper bounds presented in this chapter hold for SUM circuits, and hence, also for OR and XOR circuits. So, let

$$\mathsf{L} \in \{\mathsf{SUM}, \mathsf{OR}, \mathsf{XOR}\}$$

stand for any of these measures, and let L_d be the depth-d version of L. Let also $\mathsf{L}(n,m)$ and $\mathsf{L}_d(n,m)$ denote the maximum of these measures over all $n \times m$ matrices. These functions are symmetric: by the Transposition Principle (Observation 1.1), we have $\mathsf{L}(m,n) = \mathsf{L}(n,m)$ and $\mathsf{L}_d(m,n) = \mathsf{L}_d(n,m)$.

2.1 Lupanov's decomposition bound

We start with two observations made by Lupanov and Nechiporuk more than 50 years ago. They are amazingly simple, but turned out to be very useful later. In particular, they have led to asymptotically tight bounds on the Shannon function for addition circuits.

If we decompose a given $p \times q$ matrix A row-wise or column-wise, then we get a decomposition of weight $|A| + \min\{p,q\}$. Observe, however, that the same row (or column) may appear many times in this decomposition. In this case, it is better to join them into one rectangle.

Lemma 2.1 (Lupanov [62])**.** Every boolean $p \times q$ matrix has a decomposition of weight at most $p + q2^{q-1}$.

Proof. Given a $p \times q$ matrix A, split the rows of into groups, where the rows in one group all have the same values. This gives us a decomposition of A into $t \leqslant 2^q$ rectangles. For the i-th of these submatrices, let r_i be the number of its nonzero rows, and c_i the number of its nonzero columns; hence, the submatrix is an $r_i \times c_i$ rectangle. Since each nonzero row of A lies in exactly one of the these matrices, the weight of the decomposition is

$$\sum_{i=1}^{t}(r_i + c_i) \leqslant p + \sum_{j=0}^{q}\sum_{i:c_i=j} j \leqslant p + \sum_{j=0}^{q}\binom{q}{j}\cdot j = p + q2^{q-1}\,. \qquad \square$$

This gives us an asymptotically tight upper bound in depth 2.

Theorem 2.2 (Lupanov [62])**.** If $\log n \ll m \leqslant n$ then

$$\mathsf{L}_2(n,m) \sim \frac{nm}{\log n}\,.$$

Proof. To show the upper bound, let A be an $n \times m$ matrix. Take a positive integer q, and split A column-wise into $t \leqslant m/q$ submatrices of dimension $n \times q$; the last submatrix may have fewer than q columns. By applying Lemma 2.1 to each of them, we obtain a decomposition of A of weight at most $nm/q + m2^{q-1}$. It remains to take q about $\log n - 2\log\log n$.

The lower bound follows by counting arguments. Suppose we have r nodes on the middle level, and L edges in total. The number of ways to assign degrees to these nodes is then at most the number $\binom{L+r-1}{r-1} < 2^{L+r}$ of ways to decompose L into a sum of r non-negative numbers. After the degrees are fixed, there are at most $(m+n)^L$ ways to specify with what nodes on the input or output level the edges are actually incident. Thus, the number of depth-2 circuits with n inputs, m outputs, r middle-level nodes and L edges is at most $2^{L+r}(m+n)^L$, implying that the number of minimal circuits with n inputs, m outputs and L edges is at most

$$\sum_{r\leqslant L/2} 2^{L+r}(m+n)^L < 2^{3L/2+1}(m+n)^L < 2^{L\log n+3L}\,.$$

For this to be at least the number 2^{nm} of all $n \times m$ matrices, we need $L \geqslant (1 - o(1))nm/\log n$. □

This result was later re-discovered by Tuza [104] and Bublitz [12]. A version of it (decomposition of graphs into bipartite complete subgraphs) was proved by Chung, Erdős and Graham [21]. These results were obtained by repeatedly applying the counting argument to show the existence of a large complete bipartite graph and removing its edges. Let us note that, besides being asymptotically tight, Lupanov's decomposition is much simpler.

The case when m is only logarithmic in n was considered by Orlov [78] who proved that, if $m = k \log n$ then $\mathsf{L}_2(n, m) \sim (k+1)n$.

To get asymptotics for depth 3 and for unbounded depth, Nechiporuk used the following version of Lemma 2.1.

Lemma 2.3 (Nechiporuk [71, 75])**.** Every boolean $p \times q$ matrix A has a decomposition of weight at most $\frac{1}{2}|A| + q + p^2$.

Recall that $|A| + p$ and $|A| + q$ are trivial upper bounds.

Proof. By Observation 1.4, it is enough to construct a SUM circuit of depth 2 with at most $\frac{1}{2}|A| + q + p^2$ edges. In each column of A, group all its 1s into pairs; at most one 1 remains non-paired. Let $P(i, j)$ be the sum of all input variables x such that the x-th column has paired 1s in rows i and j. Let also $N(i)$ be the sum of variables corresponding to non-paired 1s of the i-th row. Note that, the sum $P(i, j)$ is disjoint from (shares no variable in common with) $N(i)$ as well as from $P(i, j')$ for all $j' \neq j$. (We need this disjointness to get a SUM circuit.) Compute all the sums $P(i, j)$ and $N(i)$ at the first layer. Since the 1s are paired, and since at most one 1 in each column remains non-paired, the total number of edges in this layer is at most $|A|/2 + q$. Then compute the sum y_i corresponding to the i-th row of A as

$$y_i = P(i, 1) + \ldots + P(i, i-1) + P(i, i+1) + \cdots + P(i, p) + N(i)\,.$$

This results in at most p^2 edges on the second layer. □

Using Lemma 2.3 one can prove the asymptotically tight bounds for square matrices.

Theorem 2.4 (Nechiporuk [71, 75])**.**

$$\mathsf{L}(n,n) \sim \mathsf{L}_3(n,n) \sim \frac{n^2}{2\log n}\,.$$

Proof. The lower bound follows again by counting arguments. We give an elegant one due to Pippenger [83]. Let us first count the number of graphs having r labeled nodes (some of which may be isolated) and ℓ edges. As in the proof of Theorem 2.2, the number of ways to assign out-degrees is at most $\binom{\ell+r-1}{r-1} \leqslant 2^{\ell+r-1}$. After the out-degrees are fixed, there are at most $r^\ell = 2^{\ell \log r}$ ways to specify which nodes the outgoing edges do actually enter. Since a graph with no isolated nodes and at most L edges can have at most $2L$ nodes, the number of such graphs is therefore at most

$$\sum_{\ell=0}^{L}\sum_{r=1}^{2L} 2^{\ell\log r+\ell+r-1} \leqslant \sum_{\ell=0}^{L} 2^{\ell\log L+2\ell+2L} \leqslant 2^{L\log L+4L+1}\,.$$

In each such graph, there are at most $(2L)^{2n}$ ways to choose n input and n output nodes. Thus, we have at most $2^{L\log L+O(L+n\log L)}$ minimal circuits with at most L edges. For this to be at least the number 2^{n^2} of all $n\times n$ matrices, we need $L \geqslant (1-o(1))n^2/\log n^2$.

To show the upper bound, let A be an $n\times n$ matrix. We first apply Lupanov's decomposition: split the matrix column-wise into groups of $k < \log n$ columns and compute all possible (distinct) sums in each group in depth 1; for simplicity, we assume that k divides n. Note that we have at most $q := n2^k/k$ such sums, and all they can be computed by a trivial depth-1 circuit with at most $n2^{k-1}$ edges. Treat these sums as new variables.

Consider the $n\times q$ matrix M such that $M[i,z]=1$ if and only if z is the (new) variable corresponding to a sum of some of the n/k groups in the i-th row of A. Hence, the matrix M has $|M| \leqslant n\cdot(n/k) = n^2/k$ ones, and we have that $A\vec{x} = M\vec{z}$. Now split the matrix M horizontally into submatrices of height p (also dividing n), and apply Lemma 2.3 to each of them. This way we implement M by a depth-2 SUM circuit of size at most $|M|/2 + (n/p)q + (n/p)p^2$. The size of the entire depth-3 circuit is thus at most

$$n2^{k-1} + \frac{|M|}{2} + \frac{nq}{p} + np \leqslant n2^{k-1} + \frac{n^2}{2k} + \frac{n^2 2^k}{pk} + np\,.$$

The desired upper bound is then obtained by taking $p \sim n/\log^2 n$ and k about $\log n - 3\log\log n$. □

In fact, Nechiporuk proved a more general result providing a tight depth-3 asymptotics for implementation of $m \times n$ matrices with various ratios m/n. Using a relaxation of the bounded-depth condition and a generalization of Nechiporuk's method, Pippenger [81, 83] constructed a SUM circuit of asymptotically optimal size $(1+o(1))mn/\log(mn)$ for an arbitrary $m \times n$ matrix and any $\log n \ll m \leqslant n$.

Nechiporuk [72, 75] has also proved the following upper bound for every $n \times n$ matrix A:

$$\mathsf{L}_2(A) \preccurlyeq \frac{|A|}{\log n} \log \frac{n^2}{|A|}$$

This bound was re-discovered by Feder and Motwani [29]. Actually, for the maximum $\mathsf{L}_d(n,\alpha)$ of $\mathsf{L}_d(A)$ over all $n \times n$ matrices with $|A| = \alpha n^2$ ones, Nechiporuk [72, 73, 75] proved asymptotically tight estimates

$$\mathsf{L}_2(n,\alpha) \sim H(\alpha)\frac{n^2}{\log n} \quad \text{and} \quad \mathsf{L}(n,\alpha) \sim \mathsf{L}_3(n,\alpha) \sim H(\alpha)\frac{n^2}{2\log n}\,,$$

where $H(\alpha)$ is the binary entropy function; these estimates hold as long as $\log n = o(nH(\alpha))$ and $-\log\min\{\alpha, 1-\alpha\} = o(\log n)$. The upper bounds here also hold for SUM circuits, and are obtained by applying a greedy-type algorithm.

An OR circuit may be viewed as an encoding of the *adjacency* relation of a given bipartite $n \times m$ graph G (represented by its adjacency matrix) by the *connectivity* relation in another graph (circuit) H with more nodes. The circuit H in this case has $n+m$ "poles", n input poles and m output poles. The fewer edges H has, the better is the encoding. Note that $\mathsf{L}(n,m)$ is exactly the maximum, over all bipartite $n \times m$ graphs G, of the smallest number of edges in a code H of G.

In the case when G is not bipartite, its circuit H has only n poles corresponding to the vertices of G. The circuit encodes G if there is a path from vertex i to vertex j in G precisely when there is a path from the pole i to the pole j in H. That is, in this case we are trying to encode the *connectivity* (not just adjacency) relation in the graph

G by the connectivity relation in another graph H with more nodes. For the corresponding Shannon function $\mathsf{L}(n)$, Andreev [6] proved the asymptotic bound

$$\mathsf{L}(n) \sim \frac{n^2}{8 \log n}\,.$$

2.2 The low-rank bound

If an $n \times n$ matrix A has L-rank r, then $\mathsf{L}_2(A) \leqslant 2rn$, just because A can be written as an L-product $A = P \cdot Q^\top$ of two $n \times r$ matrices. If $r \preccurlyeq \log n$, then $\mathsf{L}_2(A) \preccurlyeq n \log n$. In depth 3 we can do better.

Lemma 2.5 (Pudlák and Rödl [88]). If a boolean $n \times n$ matrix A has L-rank r, then

$$\mathsf{L}_3(A) \preccurlyeq \frac{rn}{\log n}\,.$$

Proof. The lemma was proved in [88, Proposition 10.1] by showing that A can be written as a product of three matrices, each with $O(n)$ ones. Here we give a direct construction of the circuit. Let us first do this for $\mathsf{L} = \mathsf{XOR}$.

Since the matrix A has an XOR-rank r, it can be written as a product $A = B \cdot C$ over $\mathbb{F}_2$ of an $n \times r$ matrix B and an $r \times n$ matrix C. Set $t = \lceil 2r/\log(2n) \rceil$, and split the matrix B "vertically" into t submatrices B_i with equal number of columns, and split matrix C "horizontally" into t submatrices C_i with equal number of rows. To implement a product $B_i \cdot C_i$, take the trivial depth-1 circuit F implementing the $n \times n$ Sylvester matrix $H = H_n$ with $n = 2^{r/t}$. Add new output nodes corresponding to the rows of B_i, and new input nodes corresponding to the columns of C_i. For each input node x, if the corresponding column of C_i is u, then connect x with the u-th input node of F. For each output node y, if the corresponding row of B_i is v, then connect the v-th output of F with y. Clearly, the described circuit implements $B_i \cdot C_i$ and has at most $2n + 2^{2r/t}$ edges. Since $A = \sum_i B_i \cdot C_i$, we can get the circuit for A as a union of t circuits for $B_i \cdot C_i$ utilizing $2nt + t2^{2r/t} \preccurlyeq rn/\log n$ edges in total.

The same upper bound holds also for $\mathsf{OR}_3(A \cdot B)$ and $\mathsf{SUM}_3(A \cdot B)$, when the product is taken over the corresponding semiring: just use the

intersection matrix for OR circuits, and the unique intersection matrix for SUM circuits. □

2.3 Recursively defined matrices

Some of the $n \times n$ matrices A_n with $n = 2^r$ a power of two, which we will consider below, can be defined recursively as

$$A_{2n} = \begin{bmatrix} f_1(A_n) & f_2(A_n) \\ f_3(A_n) & f_4(A_n) \end{bmatrix}$$

where each $f_i(A_n)$ is either the matrix A_n, or its complement $\overline{A}_n$, or the all-0 matrix, or the all-1 matrix. In particular, Kneser-Sierpinski matrices and Sylvester matrices, as well as their complements, have this form. For such matrices we have a simple upper bound.

Lemma 2.6. For every recursively defined boolean $n \times n$ matrix A, we have that

$$\mathsf{L}(A) \leqslant 4n \log n + 2n \,.$$

Proof. We compute recursively triples t_n containing matrix A_n, its complement $\overline{A}_n$ and all-1 row b_n of length n. To obtain the desired implementation for t_{2n} we need circuits implementing t_n on each half of inputs and two additional edges to complete each row of A_{2n}, $\overline{A}_{2n}$ and b_{2n} ($8n+2$ additional edges in total). Then, the required bound follows from the recursion $\mathsf{L}(t_{2n}) \leqslant 2 \cdot \mathsf{L}(t_n) + 8n + 2$. □

Note that the constructed circuit is of depth $\log n$.

2.4 Upper bounds for Kronecker products

We have mentioned in § 1.2 that

$$\mathsf{L}(A \otimes B) \leqslant a \cdot \mathsf{L}(B) + b \cdot \mathsf{L}(A)$$

holds, where a is the number of rows in A, and b is the number of rows in B. In depth-2 we have the following.

Lemma 2.7. $\mathsf{L}_2(A \otimes B) \leqslant \mathsf{L}_2(A) \cdot \mathsf{L}_2(B)$.

Proof. Let $\{R_i\}$ be an optimal L-covering of B by rectangles. Hence, $\mathsf{L}_2(B) = \sum_i (a_i + b_i)$ where $a_i \times b_i$ is the dimension of R_i. One $s \times t$ rectangle in A gives rise to an $a_i s \times b_i t$ rectangle in $A \otimes R_i$. Thus, if we take an optimal L-covering of A by $s_j \times t_j$ rectangles, then

$$\begin{aligned}\mathsf{L}_2(A \otimes B) &\leqslant \sum_i \sum_j (a_i s_j + b_i t_j) \leqslant \sum_i (a_i + b_i) \sum_j (s_j + t_j) \\ &= \mathsf{L}_2(B) \sum_i (a_i + b_i) = \mathsf{L}_2(B) \cdot \mathsf{L}_2(A)\,. \quad \square\end{aligned}$$

Lemma 2.8. Let A be an $m \times m$ matrix, and B a matrix having an L-covering by s rectangles of total weight W. Then

$$\mathsf{L}(A \otimes B) \leqslant s \cdot \mathsf{L}(A) + mW\,.$$

Moreover, if the circuit for A has depth d, then the resulting circuit for $A \otimes B$ has depth $d + 2$.

Proof. Let R be an $a \times b$ rectangle in B. We want first to show that then

$$\mathsf{L}(A \otimes R) \leqslant \mathsf{L}(A) + (a + b)m\,. \tag{2.1}$$

The input string of variables $\vec{x} = (\vec{x}_1, \ldots, \vec{x}_m)$ consists of m consecutive blocks, each of length k. Fix a nonzero row of R, and take a trivial depth-1 circuit of size bm computing the scalar products $z_1, \ldots, z_m$ of this row with all of these m blocks. Now let F be a circuit for the matrix A of size $L(A)$. Then $F(z_1, \ldots, z_m)$ computes all possible distinct sums of $(A \otimes R)\vec{x}$. It remains to add an additional depth-1 circuit with at most am edges to make copies of equal nonzero sums. This completes the proof of (2.1).

Now suppose we have a covering $\{R_i\}$ of B by s rectangles of dimensions $a_i \times b_i$; hence, $W = \sum_{i=1}^{s} (a_i + b_i)$ is the total weight of the covering. By (2.1),

$$\mathsf{L}(A \otimes B) \leqslant \sum_{i=1}^{s} \mathsf{L}(A \otimes R_i) \leqslant s \cdot \mathsf{L}(A) + m \sum_{i=1}^{s} (a_i + b_i) = s \cdot \mathsf{L}(A) + mW\,. \quad \square$$

For example, the $m \times m$ all-1 matrix J_m has a trivial covering with $s = 1$ rectangle of weight $W = 2m$. Hence, for any matrix A, we have

that $\mathsf{L}_3(A \otimes J_m) \leqslant |A| + 2m^2$. We will use this observation later in § 3.3 to show the optimality of Nechiporuk's rectangle bound.

Given a covering of a matrix by $a_i \times b_i$ rectangles, its *smaller weight* is the sum $s = \sum_i \min\{a_i, b_i\}$ of the lengths of the shorter sides, and its *larger weight* is the sum $\ell = \sum_i \max\{a_i, b_i\}$ of the lengths of the longer sides. Hence, $s + \ell$ is the (total) weight of the covering. Note that every $m \times m$ matrix A has a covering with parameters $s \leqslant m$ and $\ell \leqslant |A| \leqslant m^2$: just take rows as rectangles.

The following lemma extends Lupanov's upper bound, given in Theorem 2.2, to Kronecker products.

Lemma 2.9. Let B be an $m \times m$ matrix, and A an arbitrary matrix admitting an L-covering of smaller weight s and larger weight $\ell \leqslant m^2/2\log m$. Then

$$\mathsf{L}_2(A \otimes B) \preccurlyeq \frac{m^3}{\log m} + \frac{sm^2}{\log(m^2/\ell)}\,.$$

Note that, if A is also an $m \times m$ matrix, then a direct application of Theorem 2.2 would only give an upper bound of about $m^4/\log m$.

Proof. Take an L-covering of A of smaller weight s and larger weight ℓ. By Lemma 2.1, for every integer $t \geqslant 1$, both B and $B^\top$ have coverings with lengths of their first sides summing to m^2/t, and lengths of their first sides summing to $m2^t$. If R is an $a \times b$ rectangle in our covering, then $R \otimes B$ has a covering of weight

$$\min\{a, b\}m^2/t + \max\{a, b\}m2^t\,.$$

By taking the union of these coverings over all rectangles R in the covering of A, we obtain a covering of $A \otimes B$ of weight at most $sm^2/t + \ell m2^t$. It remains to take t about $\log(m^2/\ell) - \log\log m$. □

Lemma 2.10 (Find et al. [30])**.** Let A and B be $m \times m$ matrices. If r is the L-rank of A, then

$$\mathsf{L}_3(A \otimes B) \leqslant 3rm^2 \quad \text{and} \quad \mathsf{L}_6(A \otimes B) \leqslant 6rm^2/\log m\,.$$

Proof. The matrix A can be written as a product $A = PQ^\top$ where P and Q are $m \times r$ matrices. Using the mixed-product property

$$A \cdot B \otimes C \cdot D = (A \otimes C) \cdot (B \otimes D)$$

of the Kronecker product, we can represent $A \otimes B$ as

$$\begin{aligned} A \otimes B &= P \cdot Q^\top \otimes B \\ &= P \cdot I_r \cdot Q^\top \otimes I_m \cdot B \cdot I_m \\ &= (P \otimes I_m)(I_r \otimes B)(Q^\top \otimes I_m)\,. \end{aligned}$$

As each of the three matrices has at most rm^2 ones, we immediately derive the first bound $\mathsf{L}_3(A \otimes B) \leqslant 3rm^2$.

To prove the second bound, observe that Theorem 2.2 gives

$$\mathsf{L}_2(P \otimes I_m) = m \cdot \mathsf{L}_2(P) \leqslant 2rm^2/\log m$$

and similarly

$$\mathsf{L}_2(I_r \otimes B) \leqslant 2rm^2/\log m, \qquad \mathsf{L}_2(Q^\top \otimes I_m) \leqslant 2rm^2/\log m$$

leading to the second claim of the lemma. □

The proof actually shows that, if A is an $m \times m$ matrix which can be represented as an L-product $A = P \cdot Q^\top$, where P and Q are $m \times r$ matrices, then for every matrix B,

$$\mathsf{L}(A \otimes B) \leqslant r \cdot \mathsf{L}(B) + m \cdot \mathsf{L}(P) + m \cdot \mathsf{L}(Q)\,.$$

3

General Lower Bounds

Here we collect some basic lower-bound arguments, in increasing order of technical difficulty: bounds for SUM circuits, then bounds for OR circuits, and finally, those for XOR circuits.

3.1 Determinant lower bounds

When restricted to SUM circuits, one of the first general lower bounds on the size of linear circuits over the field of real numbers is the following one.

Theorem 3.1 (Morgenstern [67]). For every square matrix A,

$$\mathsf{SUM}(A) \geqslant \log |\det(A)| .$$

Although Morgenstern counts the number of nodes in fanin-2 circuit, the bound is also applicable to the number of edges (see Observation 1.3). But if we count edges directly, then the same proof technique leads to a sharper bound.

Theorem 3.2 (Kochergin [56]). For every square matrix A,

$$\mathsf{SUM}(A) \geqslant 3 \log_3 |\det(A)| .$$

Proof. We prove the desired inequality

$$3^{\mathsf{SUM}(A)} \geqslant |\det(A)|^3$$

by induction on $\mathsf{SUM}(A)$. The base case $\mathsf{SUM}(A) = 0$ (that is, $A = I_n$) is straightforward. Next we prove the induction step.

In the minimal circuit computing A, consider an output node v with no outgoing edges. Suppose this node implements the i-th row of the matrix A. Let r be the number of edges entering v, and A_j be the matrix obtained from A by replacing its i-th row by the row computed[1] at the j-th node ($1 \leqslant j \leqslant r$) incident to v; the i-th row is the sum of these r rows.

Clearly, $\mathsf{SUM}(A_j) \leqslant \mathsf{SUM}(A) - r$ holds for every j: we can remove all r edges incident to v, and the resulting circuit will still compute A_j. By induction hypothesis, we also have $|\det(A_j)| \leqslant 3^{\mathsf{SUM}(A_j)/3}$. Using well-known properties of the determinant, we obtain:

$$\begin{aligned} |\det(A)| = \Big|\sum_{j=1}^{r} \det(A_j)\Big| &\leqslant \sum_{j=1}^{r} |\det(A_j)| \leqslant \sum_{j=1}^{r} 3^{\mathsf{SUM}(A_j)/3} \\ &\leqslant r3^{(\mathsf{SUM}(A)-r)/3} = r3^{-r/3}3^{\mathsf{SUM}(A)/3} \leqslant 3^{\mathsf{SUM}(A)/3}, \end{aligned}$$

since the function $r3^{-r/3}$, for natural numbers r, achieves its maximum when $r = 3$. □

A classical inequality of Hadamard for complex valued $n \times n$ matrices $A = (a_{ij})$ states that

$$\det(A) \leqslant \prod_{j=1}^{n} \sqrt{\sum_{i=1}^{n} |a_{ij}|}\,.$$

Using the arithmetic-geometric mean inequality (1.3), this yields the following inequality for every boolean matrix A:

$$\det(A) \leqslant \Big(\frac{|A|}{n}\Big)^{n/2}. \tag{3.1}$$

[1] The row computed at a node is the coefficient vector of the sum computed at that node.

Theorem 3.3 (Pudlák [87])**.** For every boolean $n \times n$ matrix A, and every integer $d \geqslant 1$,

$$\mathsf{SUM}_d(A) \geqslant dn|\det(A)|^{\frac{2}{dn}} .$$

Proof. As we mentioned above, the computation of linear forms $x \mapsto Ax$ associated to a matrix A by depth-d circuits corresponds to a factorization into the product of d rectangular matrices $A_1, \ldots, A_d$ (each A_i being the adjacency matrix of one level in the circuit). The number of edges in the circuit is $S = \sum_{i=1}^{d} |A_i|$. Using (3.1) and the geometric-arithmetic mean inequality (1.3), one can derive

$$|\det(A)| \leqslant \left(\frac{\sum_{i=1}^{d} |A_i|}{dn}\right)^{\frac{dn}{2}} = \left(\frac{S}{dn}\right)^{\frac{dn}{2}} ,$$

from which the desired inequality $S \geqslant dn|\det(A)|^{2/dn}$ follows. □

Natural candidates to apply Theorem 3.3 are Sylvester matrices because they have essentially the largest determinant among all boolean matrices.

Corollary 3.4. Let $H = H_n$ be the $n \times n$ Sylvester matrix. Then for every $d \geqslant 1$,

$$\mathsf{SUM}(H) \succcurlyeq n \log n \quad \text{and} \quad \mathsf{SUM}_d(H) \geqslant dn\Big(\frac{n}{4}\Big)^{1/d} .$$

Proof. Let H' be the $(n-1)\times(n-1)$ matrix obtained from H by removing its all-0 row and all-0 column. By Theorem 3.1 and Theorem 3.3, it is enough to show that

$$|\det(H')| = 2(n/4)^{n/2} .$$

For this, consider the ± 1-version M of H_n with each entry $a \in \{0, 1\}$ replaced by $2a - 1$ (that is, $0 \mapsto -1$ and $1 \mapsto +1$). The resulting ± 1 matrix M satisfies $MM^{\top} = nI_n$ over the reals (is a Hadamard matrix), implying that $|\det(M)| = n^{n/2}$. The first row of M contains only -1s. Divide by 2 all rows, except the first one. (This multiplies the determinant by 2^{1-n}.) Then subtract the first row divided by 2 from

each other row. This way we obtain a matrix

$$\begin{bmatrix} -1 & -1 & \dots & -1 \\ 0 & & & \\ \vdots & & H' & \\ 0 & & & \end{bmatrix}.$$

The absolute value of its determinant is

$$|\det(H')| = 2^{1-n} \cdot |\det(M)| = 2(n/4)^{n/2}. \qquad \square$$

3.2 Hansel–Krichevski type bounds

For a square matrix A, let $\alpha(A)$ denote the largest number d such that A has a $d \times d$ all-0 submatrix on the diagonal, that is, a principal $d \times d$ minor consisting of 0s. Note that if A is the adjacency matrix of a graph, then $\alpha(A)$ is the largest size of an independent set in this graph.

Theorem 3.5. If A is an $n \times n$ matrix with zero main diagonal, then

$$\mathsf{OR}_2(A) \geqslant n \log \frac{n}{\alpha(A)}.$$

Theorem 3.5 is a direct consequence of the following two lemmas.

A *bipartite covering* of an undirected graph $G = (V, E)$ is a set of pairs (S_i, T_i), $i = 1, \dots, m$ of disjoint subsets of V such that $E \subseteq \cup_{i=1}^{m} S_i \times T_i$. The *weight* of pairs (S_i, T_i) is $|S_i| + |T_i|$. The weight of a covering is the sum of the weights of all pairs in the covering. In the *biclique covering* we have an additional restriction that $S_i \times T_i \subseteq E$ must hold for all i. The *biclique covering number*, $\mathsf{bc}(G)$, of a graph G is the minimum weight of any biclique covering of G.

The relation with OR circuits is given in the following lemma. If A is a square 0/1 matrix with zero diagonal, then its graph is the graph G_A whose adjacency matrix is $A \vee A^{\top}$.

Lemma 3.6. For every square matrix A with zeroes on the main diagonal, $\mathsf{OR}_2(A) \geqslant \mathsf{bc}(G_A)$.

Proof. Let A be an $n \times n$ matrix with zero main diagonal. Take a depth-2 OR circuit implementing A, and consider the graph $G = ([n], E)$

where two nodes i and j are adjacent if and only if either there is a path from the i-th input to the j-th output, or from the j-th input to the i-th output in the circuit. Note that the adjacency matrix of the resulting graph is $A \vee A^\top$; hence, $G = G_A$.

Let V be the set of nodes on the middle level of our circuit for A. For $v \in V$, let $S_v \subseteq [n]$ be the set of (indices of) inputs having edges to v, and $T_v \subseteq [n]$ the set of (indices of) outputs having edges from v in the circuit. Since A has only zeroes on the main diagonal, we have that $S_v \cap T_v = \emptyset$ for each $v \in V$. Thus, we obtain a covering of the graph G by the bicliques $S_v \times T_v$, $v \in V$, and the weight $\sum_{v \in V}(|S_v| + |T_v|)$ of this covering is the total number of edges in our circuit. □

The following classical result of Hansel [44] and Krichevski [59] gives a general lower bound on the weight of bipartite coverings of graphs. Independently, a similar result was also proved by Katona and Szemerédi [52]. For an undirected graph G, let $\alpha(G)$ denote the maximum number of nodes in its independent set.

Lemma 3.7 (Hansel–Krichevski)**.** Every bipartite covering of an n-node graph G must have weight at least $n \log \frac{n}{\alpha(G)}$. In particular,

$$\mathsf{bc}(G) \geqslant n \log \frac{n}{\alpha(G)}\,.$$

Proof. We give an elegant probabilistic proof due to Radhakrishnan [91]. Fix a graph G on n nodes, and take an arbitrary bipartite covering (S_i, T_i), $i = 1, \ldots, m$. For a node v, let $m_v = |\{i\colon v \in S_i \cup T_i\}|$ be the number of pairs in our covering containing v. By double-counting, the sum $\sum_v m_v$ is exactly the weight $\sum_i(|S_i| + |T_i|)$ of the covering.

Now, for each i, randomly choose one of S_i and T_i, and delete all its nodes from the graph. Since every edge of G must belong to at least one $S_i \times T_i$, at most $\alpha(G)$ nodes can survive at the end. On the other hand, $\Pr[v \text{ survives}] = 2^{-m_v}$. By the linearity of expectation, $\sum_v 2^{-m_v} \leqslant \alpha(G)$. When applied with $x_i = 2^{-m_i}$, the arithmetic-geometric mean inequality (1.3) yields

$$\frac{\alpha(G)}{n} \geqslant \frac{1}{n}\sum_v 2^{-m_v} \geqslant \Big(\prod_v 2^{-m_v}\Big)^{1/n} = 2^{-\frac{1}{n}\sum_v m_v}\,,$$

from which $2^{\frac{1}{n}\sum_v m_v} \geqslant n/\alpha(G)$, and hence, also $\sum_v m_v \geqslant n \log \frac{n}{\alpha(G)}$ follows. □

In fact, as shown by Newman and Wigderson [76], the extra factor can be improved to $\log H(G)$, where $H(G)$ is the entropy of G; see also [50, Section 6.13].

To illustrate Theorem 3.5 "at work", consider the $n \times n$ intersection matrix $\mathcal{D}_n$, the full triangular matrix T_n, and the complement $\overline{I}_n$ of the identity matrix I_n. By (1.9), (1.7) and (1.6), we know that $\mathsf{OR}_2(\mathcal{D}_n)$, $\mathsf{OR}_2(\overline{I}_n) \leqslant n \log n$ and $\mathsf{OR}(T_n) \leqslant n \log n + n$. Theorem 3.5 implies that these trivial upper bounds are also optimal.

Corollary 3.8.

$$\mathsf{OR}_2(\overline{I}_n), \mathsf{OR}_2(\mathcal{D}_n), \mathsf{OR}_2(T_n) \geqslant n \log n\,.$$

The lower bounds for T_n and $\overline{I}_n$ were proved by Tarjan [103].

Proof. The first two lower bounds follow directly from Theorem 3.5, because the matrices $\overline{I}_n$ and $\mathcal{D}_n$ have zero main diagonals, and $\alpha(\overline{I}_n) = \alpha(\mathcal{D}_n) = 1$. To show the last lower bound, extend T_n to an $(n+1) \times (n+1)$ matrix T'_n by adding the all-0 row on the top, and the all-0 column on the right. Since $\alpha(T'_n) = 1$ and $\mathsf{OR}_2(T_n) = \mathsf{OR}_2(T'_n)$, the desired lower bound follows. □

Note that $\mathsf{XOR}_2(\overline{I}_n) \leqslant 4n$. Just take one node on the middle level connected to all n inputs and outputs and add, for each $1 \leqslant i \leqslant n$, a new length-2 path from the i-th input to the i-th output. Thus, the Hansel–Krichevski approach does not yield a superlinear lower bound for XOR circuits.

3.3 Nechiporuk's bounds

To make our arguments in what follows more intuitive, let us say that a node v *is seen* by node w (and that w *can see* v) if there is a path from v to w. Our convention is that a node can see itself and is seen by itself. We also say that an edge $e = (u, v)$ *is seen* by a node w if its

second endpoint v is seen by w, and that the edge *can see* w if its first endpoint u can see w.

Most of the lower bounds for SUM and OR circuits are based on their "rectangle property" (not shared by XOR circuits):

> every edge in a SUM or OR circuit for a matrix A defines a rectangle in A.

This rectangle consists of all entries (i, j) such that the j-th input node is seen by the edge, and the edge is seen by the i-th output.

A *cut* in a circuit is a set of its edges such that every input-output path contains at least one edge in this set. A cut is *minimal* if no two of its edges lie on the same path. Thus, every minimal cut gives us a covering of A by rectangles; in the case of SUM circuits, we have a decomposition. Note also that the rectangle property alone can only give a lower bound $\mathsf{OR}(A) \geqslant |A|/r(A)$, where $r(A)$ is the maximal area of (number of 1-entries in) a rectangle of A. This holds because *every* cut in a circuit for A must have at least $|A|/r(A)$ edges.

To obtain stronger lower bounds on the number of edges, one tries to find a large minimal cut in a circuit for A, whose induced covering of A is "legal" in some sense, say, either the dimensions or the areas of its rectangles do not exceed some given threshold. In order to show that $\mathsf{OR}(A)$ must be large, one then argues in two steps:

1. Find a minimal cut inducing a "legal" covering of A.
2. Show that every "legal" covering must have a large number of rectangles.

Note that in (2) we are dealing with the *number* of rectangles in coverings, not with their *weights*.

The following lower bound was proved by Nechiporuk [75, Theorems 1.2 and 1.3]. This result remained unknown, and was later independently (and almost simultaneously) re-discovered by Mehlhorn [64], Pippenger [84] and Wegener [107].

Theorem 3.9 (Nechiporuk [75, 73]). If A is $(k+1, l+1)$-free, then

$$\mathsf{OR}(A) \geqslant \frac{|A|}{k \cdot l} \quad \text{and} \quad \mathsf{OR}_2(A) \geqslant \frac{|A|}{\max\{k, l\}}\,.$$

In particular, the theorem implies that 2-free matrices A have no better OR circuits than the trivial one: $\mathsf{OR}(A) = |A| = \mathsf{OR}_1(A)$.

Proof. We use an elegant argument due to Pippenger [84]. Take a minimal OR circuit implementing A. Say that an edge e is *legal* if the dimension $a \times b$ of the rectangle induced by e satisfies $a \leqslant k$ and $b \leqslant l$. Since each such rectangle can cover at most $k \cdot l$ 1-entries of A, it is enough to show that a cut consisting of only legal edges exists. To show this, let l_u be the number of input nodes seen by u, and k_u the number of output nodes seeing the node u.

Take an arbitrary input-output path P. Let P_1 be the initial part of P consisting of edges (u, v) with $l_u \leqslant l$, and P_2 the final part of P consisting of edges (u, v) with $k_v \leqslant k$. Let $e = (u, v)$ be the last edge of P_1. If e is the last edge of the entire path P, then $l_v \leqslant l$, implying that e is a legal edge (because $k_v = 1$). Otherwise, we have that $l_u \leqslant l$ but $l_v > l$. Suppose that $e \notin P_2$. Then $k_v > k$. This means that the node v can see $> l$ inputs and is seen by $> k$ outputs, contradicting the $(k+1, l+1)$-freeness of A. Thus, P_1 and P_2 must have some edge $e = (u, v)$ in common, meaning that $l_u \leqslant l$ and $k_v \leqslant k$, that is, the edge e is legal, as desired.

In the depth-2 case, every edge $e = (u, v)$ (be it legal or not) has either $k_u = 1$ or $l_v = 1$. Thus, if e is legal, then the induced $a \times b$ rectangle must satisfy $ab \leqslant \max\{k, l\}$. □

Observation 3.10 (Optimality of Theorem 3.9)**.** For every $k \geqslant 2$, there is a $(k+1)$-free $n \times n$ matrix A such that

$$|A|/k^2 \leqslant \mathsf{OR}(A) \leqslant |A|/k^2 + 2n\,.$$

Thus, for all $k = o(n^{1/3})$, the lower bound given by Theorem 3.9 is slack by only a lower-order additive term.

Proof. Take a 2-free $m \times m$ matrix M with $|M| = m^{3/2}$ ones. Such matrices can be easily constructed (see § 1.3). Take also a $k \times k$ all-1 matrix J, and consider the Kronecker product $A = M \otimes J$. That is, A is an $n \times n$ block-matrix with $n = km$ obtained from M by replacing its 1s by copies of J; hence, A has $|A| = k^2 m^{3/2}$ ones. Since the matrix A

is $(k+1)$-free, Theorem 3.9 gives $\mathsf{OR}(A) \geqslant |A|/k^2$. On the other hand, Lemma 2.8 yields $\mathsf{OR}_3(A) \leqslant m^{3/2} + 2km = |A|/k^2 + 2n$. □

3.4 Rectangle-area based bounds

Theorem 3.9 gives strong lower bounds on the OR- complexity for matrices that are dense (have may 1s) and are k-free for a small k. In particular, it yields an almost maximal lower bound $\mathsf{OR}(N_{n,t}) \succcurlyeq n^{2-1/t}$ for the norm $n \times n$ matrix with $n = q^t$ for a prime power q and any integer constant $t \geqslant 1$. But this theorem fails on the matrices that have at least one $\sqrt{n} \times \sqrt{n}$ all-1 submatrix: the resulting lower bound is then at most n. In such cases, the following rectangle-area based bound can still yield superlinear lower bounds.

Recall that the *area* of an $a \times b$ rectangle is the number ab of its 1-entries, and its *density* is the fraction $ab/(a+b)$. Let $r(A)$ be the maximal area, and $\delta(A)$ the maximal density of a rectangle in A. From the arithmetic-geometric mean inequality $(a+b)/2 \geqslant \sqrt{ab}$, we obtain that

$$\delta(A) \leqslant \frac{1}{2}\sqrt{r(A)}\,.$$

Lemma 3.11 (Nechiporuk [75])**.** Every covering of a matrix A by rectangles of density $\leqslant \delta$ has weight $\geqslant |A|/\delta$.

Proof. Take a covering of A by $a_i \times b_i$ rectangles R_i, $i = 1, \ldots, t$. Define the weight $w(e)$ of a 1-entry e of A by

$$w(e) = \sum_{i\colon e \in R_i} \frac{a_i + b_i}{a_i b_i}\,.$$

Then the total weight of the covering is[2]

$$\sum_{i=1}^{t}(a_i + b_i) = \sum_{i=1}^{t}\sum_{e\in R_i}\frac{a_i+b_i}{a_ib_i} = \sum_{e\in A}\sum_{i\colon e\in R_i}\frac{a_i+b_i}{a_ib_i} = \sum_{e\in A} w(e)\,.$$

Since $w(e) \geqslant 1/\delta$ for every 1-entry e of A, the desired lower bound $|A|/\delta$ on the weight of the covering follows. □

[2]Here $e \in A$ means that e is a 1-entry of A.

Together with Observation 1.4, this gives

$$\mathsf{OR}_2(A) \geqslant \frac{|A|}{\delta(A)} \geqslant \frac{2|A|}{\sqrt{r(A)}}\,.$$

For circuits of larger depth, we have the following general area-based lower bounds. Recall that $\mathsf{OR}(A) \geqslant |A|/r(A)$ is a trivial lower bound, and this bound is $\leqslant n$ for every $n \times n$ matrix, because $r(A)$ is at least the maximum number of 1s in a row.

Theorem 3.12. For every $n \times m$ boolean matrix A, and every integer $d \geqslant 1$, we have

$$\mathsf{OR}(A) \geqslant \frac{3|A|}{r(A)} \log_3 \frac{|A|}{n} \quad \text{and} \quad \mathsf{OR}_d(A) \geqslant \frac{d|A|}{r(A)} \left(\frac{|A|}{n}\right)^{1/d}.$$

A similar lower bound was implicit in Grigoriev's paper [38], where he proved that every OR circuit for the Sylvester matrix $H = H_n$ requires at least $\frac{1}{2} n \log n$ fanin-2 gates.

Theorem 3.12 itself is an easy consequence of the following two graph-theoretic lemmas, versions of which are well-known and widely used facts.

Let $G = (V, E)$ be a directed acyclic graph. An *r-regular* weighting of G is an assignment of non-negative weights l_e to the edges e such that no edge can be seen by more than r/l_e output nodes. That is, at most r/l_e output nodes are reachable from each edge e. The *weight* l_u of a node $u \in V$ is the sum of weights of all edges seen by this node.

Lemma 3.13. Let $G = (V, E)$ be a directed acyclic graph, and $Y \subset V$ be the set of its output nodes. Then for every r-regular weighting $l : E \to \mathbb{R}_+$,

$$|E| \geqslant \frac{1}{r} \sum_{u \in Y} l_u\,.$$

Proof. Let $L = \{l_e : e \in E\}$ be the set of all distinct weights assigned to the edges of G. For every node $u \in V$ and every weight $l \in L$, let $E_u \subseteq E$ be the set of all edges seen by the node u, and $W_l \subseteq E$ be the set of all edges of weight l. Thus, $\sum_{l \in L} |W_l|$ is the total number of edges. Our goal is to lower-bound this sum.

Fix one weight $l \in L$. Since the weighting is r-regular, no edge in W_l can be seen by more than r/l output nodes $u \in Y$. Thus,[3] the sum $\sum_{u \in Y} |E_u \cap W_l|$ cannot exceed $\frac{r}{l}|W_l|$. This gives us the lower bound

$$|W_l| \geqslant \frac{l}{r} \sum_{u \in Y} |E_u \cap W_l| \tag{3.2}$$

for every $l \in L$. On the other hand, by the definition of the weight of a node, for every node u, we have that

$$l_u = \sum_{l \in L} l \cdot |E_u \cap W_l| \,. \tag{3.3}$$

Thus,

$$\begin{aligned} |E| = \sum_{l \in L} |W_l| &\geqslant \sum_{l \in L} \frac{l}{r} \sum_{u \in Y} |E_u \cap W_l| \\ &= \frac{1}{r} \sum_{u \in Y} \sum_{l \in L} l \cdot |E_u \cap W_l| = \frac{1}{r} \sum_{u \in Y} l_u \,. \quad \square \end{aligned}$$

Now, if $G = (V, E)$ is an OR circuit implementing a matrix A, the *standard weighting*, which assigns to each edge $e = (w, v)$ the number l_e of input nodes seen by its terminal endpoint v, is r-regular for every $r \geqslant r(A)$. This holds because at most $r(A)/l_e$ output nodes can see e. Thus, by Lemma 3.13, we only need to lower-bound the weights l_u of output nodes under this weighting.

Recall that l_u is the sum of weights of all edges seen by u. We can remove some edges seen by u until we obtain a (directed) tree T_u rooted at u, whose leaves (fanin-0 nodes) are exactly the input nodes seen by u in the original graph. Since the weights are non-negative, l_u is at least the weight of the tree T_u, defined as the sum of weights of all its edges.

Lemma 3.14. Under the standard weighting, every tree with m leaves has weight at least $3m \log_3 m$. If the tree has depth d, then its weight is at least $dm^{1+1/d}$.

Proof. Let $f(m)$ denote the minimum weight of a tree with m leaves. Our goal is to show that $f(m) \geqslant 3m \log_3 m$. We argue by induction on

[3] Here we use the trivial fact that, if no element belongs to more than d of the sets $S_1, \ldots, S_t$, then $\sum_i |S_i| \leqslant d|\cup_i S_i|$.

m. Take a tree with m leaves and minimal weight. Let k be the number of edges entering the root, and let m_i be the number of leaves in the subtree of its i-th predecessor; hence, $\sum_{i=1}^{k} m_i = m$. Then

$$\begin{aligned} f(m) &= km + \sum_{i=1}^{k} f(m_i) \geqslant km + 3\sum_{i=1}^{k} m_i \log_3 m_i \\ &\geqslant km + 3\Big(\sum_i m_i\Big) \log_3 \frac{\sum_i m_i}{k} = km + 3m \log_3 \frac{m}{k} \\ &= 3m \log_3 m + m(k - 3\log_3 k) \geqslant 3m \log_3 m\,. \end{aligned}$$

The first inequality here follows by the induction hypothesis, and the second from Jensen's inequality (1.4), since the function $x \log x$ is convex.

To prove the second claim, let $f_d(m)$ denote the minimum weight of a tree of depth d with m leaves. We will prove $f_d(m) \geqslant dm^{1+1/d}$ by induction on d. The case $d = 1$ is obvious. For an arbitrary depth, take a depth-$(d+1)$ tree with m leaves and minimal weight. Let k be the number of edges entering the root, and let m_i be the number of leaves in the subtree of its i-th predecessor. Then

$$\begin{aligned} f_{d+1}(m) &= km + \sum_{i=1}^{k} f_d(m_i) \geqslant km + d\sum_{i=1}^{k} m_i^{1+1/d} \\ &\geqslant km + kd\Big(\frac{m}{k}\Big)^{1+1/d} = m\Big[k + d\Big(\frac{m}{k}\Big)^{1/d}\Big] \\ &\geqslant m\Big[(d+1)m^{1/(d+1)}\Big]\,. \end{aligned}$$

The first inequality here follows by the induction hypothesis, and the second follows from Jensen's inequality (1.4), since the function $x^{1+1/d}$ is convex. The last inequality follows from the arithmetic–geometric mean inequality (1.3) applied to the summand k and d summands $(m/k)^{1/d}$. □

Proof of Theorem 3.12. Take a minimal OR circuit $G = (V, E)$ for A, and its standard weighting $l : E \to \mathbb{R}_+$. As we observed above, the weighting is r-regular for $r = r(A)$, because A cannot have a rectangle

of area larger than r. By Lemma 3.13, the circuit must have at least

$$\frac{1}{r}\sum_{i=1}^{n} l_i$$

edges, where l_i is the sum of weights of edges seen by the i-th output node. On the other hand, Lemma 3.14 yields $l_i \geqslant 3m_i \log_3 m_i$, where m_i is the number of 1s in the i-th row of A. Since $m_1 + \cdots + m_n = |A|$, Lemma 3.13 and Jensen's inequality (1.4) yield

$$\mathsf{OR}(A) \geqslant \frac{3}{r}\sum_{i=1}^{n} m_i \log_3 m_i \geqslant \frac{3|A|}{r}\log_3 \frac{|A|}{n}\,.$$

In the case of depth-d circuits, Lemma 3.14 implies that $l_i \geqslant d m_i^{1+1/d}$ for all $i = 1, \ldots, n$. Thus, Lemma 3.13 and Jensen's inequality (1.4) yield

$$\mathsf{OR}_d(A) \geqslant \frac{d}{r}\sum_{i=1}^{n} m_i^{1+1/d} \geqslant \frac{d}{r}\cdot n\left(\frac{|A|}{n}\right)^{1+1/d}. \qquad \square$$

If a matrix A has one large rectangle, then the fraction $|A|/r(A)$ is automatically small, even though the remaining 1-entries of A might be hard to cover. To capture such situations, one can take some set X of potentially "hard-to-cover" 1-entries of A such that the largest possible number $r_A(X)$ of entries in X lying in a rectangle of A is much smaller than $|X|$. Then we have:

$$\mathsf{OR}(A) \geqslant \frac{3|X|}{r_A(X)}\log_3 \frac{|X|}{n} \quad \text{and} \quad \mathsf{OR}_d(A) \geqslant \frac{d|X|}{r_A(X)}\left(\frac{|X|}{n}\right)^{1/d}.$$

To show this, it is enough in the proof of Theorem 3.12 to let m_i be the number of input nodes corresponding to the 1-entries of X in the i-th row of A, and to ignore output nodes seeing no entry in X.

In some cases, the bounds resulting from Theorem 3.12 can be slightly improved by using more subtle definitions of tree-weights. The resulting improvements are not substantial, but could help when trying to find asymptotically tight bounds. To demonstrate this, let us consider the $n \times n$ Sylvester matrix H.

By Lindsey's Lemma, H contains no monochromatic $a \times b$ submatrix with $ab > n$. In fact, the area of rectangles (1-monochromatic submatrices) in H is two times smaller.

Observation 3.15. If H_n contains an $s \times t$ or $t \times s$ rectangle, then $s \leqslant n/2^{1+\lceil \log t \rceil}$.

In particular, $r(H_n) = n/2$.

Proof. Recall that $n = 2^r$. Let $a_1, \ldots, a_t \in \mathbb{F}_2^r$ be distinct labels of a set of rows of $H = H_n$. All these rows can have a 1 in the x-th column only if x satisfies the system of t linear equations: $\langle a_1, x\rangle = 1, \ldots, \langle a_t, x\rangle = 1$. Let m be the rank of this system. Then, the number of solutions is at most 2^{r-m}. On the other hand, in a subspace of dimension m we can choose at most $t \leqslant 2^{m-1}$ vectors a_i such that the system above has a solution. These are some m linearly independent vectors and sums of odd numbers of these vectors. Hence, the number of solutions is at most $2^{r-1-\lceil \log t \rceil} = n/2^{1+\lceil \log t \rceil}$, as claimed. □

Since H has $|H| = \binom{n}{2}$ ones, the area lower bound (Theorem 3.12) implies that $\mathsf{OR}(H) \geqslant (3 - o(1))n \log_3 n$. We now show how to improve this to $\mathsf{OR}(H) \geqslant (2 - o(1))n \log n$.

To prove this, we need an analogue of Lemma 3.14 for another definition of the edge weight. Once again consider a rooted tree. If ℓ is the number of leaves in the subtree T_v rooted at v, then we define the *rounded weight* of an edge (u, v) to be 2^q where q is the unique integer such that $2^{q-1} < \ell \leqslant 2^q$. As before, the *rounded weight* of the tree is the sum of rounded weights of all its edges.

Lemma 3.16. The rounded weight of every tree with n leaves is at least $2n\lceil \log n \rceil$.

Proof. Let $f(n)$ denote the minimum rounded weight of a tree with n leaves. Our goal is to show that $f(n) \geqslant 2n\lceil \log n \rceil$. For this, note first that for all integers $p \leqslant q$ the following inequality holds:

$$2^q \geqslant 2^{p+1}(q - p)\,.$$

Hence, for all $m \leqslant 2^q$:

$$2^q \geqslant 2m(q - \lceil \log m \rceil). \tag{3.4}$$

Now, we prove the statement of lemma by induction on n. Take a tree with n leaves and minimal rounded weight. Let k be the number

of edges entering the root, and let n_i be the number of leaves in the subtree of the i-th child; hence, $\sum_{i=1}^{k} n_i = n$. Let $q = \lceil \log n \rceil$. Then

$$f(n) = k2^q + \sum_{i=1}^{k} f(n_i) \geqslant \sum_{i=1}^{k} (2n_i \lceil \log n_i \rceil + 2^q) \geqslant \sum_{i=1}^{k} 2n_i q = 2nq\,.$$

The second inequality here follows by the induction hypothesis, and the third follows from (3.4). □

Lemma 3.17. For the Sylvester $n \times n$ matrix H, we have

$$\mathsf{OR}(H) \geqslant (2 - o(1))n \log n\,.$$

Proof. We proceed along the lines of the proof of Theorem 3.12. First, instead of (3.2), Observation 3.15 gives us the inequality

$$|W_l| \geqslant \frac{2^{\lceil \log l \rceil + 1}}{n} \sum_{u \in Y} |E_u \cap W_l|\,. \tag{3.5}$$

Further, instead of using (3.3), we now have, by Lemma 3.16,

$$l_u = \sum_{l \in L} 2^{\lceil \log l \rceil} |E_u \cap W_l| \geqslant 2m_u \log m_u\,, \tag{3.6}$$

where l_u stands for the rounded weight of the subtree rooted at u, and m_u is the number of 1s in the u-th row of A.

From (3.5) and (3.6) we now obtain:

$$\begin{aligned}\mathsf{OR}(H) = \sum_l |W_l| &\geqslant \sum_l \frac{2^{\lceil \log l \rceil + 1}}{n} \sum_u |E_u \cap W_l| = \\ &= \frac{2}{n} \sum_u \sum_l 2^{\lceil \log l \rceil} |E_u \cap W_l| \geqslant \frac{4}{n} \sum_u m_u \log m_u \\ &\geqslant \frac{4}{n} |H| \log \frac{|H|}{n} = (2 - o(1))n \log n\,. \quad \square\end{aligned}$$

There is still a gap between the lower bound given by Lemma 3.17 and the upper bound $\mathsf{OR}(H) \leqslant (4 + o(1))n \log n$ given by Lemma 2.6.

3.5 Bounds for block-matrices

If a matrix M has the form

$$M = \begin{bmatrix} A & 0 \\ B & C \end{bmatrix} \tag{3.7}$$

then it is clear that $\mathsf{L}(M) \geqslant \mathsf{L}(A) + \mathsf{L}(C)$, where $\mathsf{L}(M)$ is the SUM or OR complexity of M. In fact, we have stronger lower bounds in terms of the corresponding rank of B.

Theorem 3.18. If a matrix M has the form (3.7), then

$$\mathsf{SUM}(M) \geqslant \mathsf{SUM}(A) + \mathsf{SUM}(C) + \mathrm{rk}_{+}(B)\,, \tag{i}$$
$$\mathsf{OR}(M) \geqslant \mathsf{OR}(A) + \mathsf{OR}(C) + \mathrm{rk}_{\vee}(B)\,, \tag{ii}$$
$$\mathsf{OR}_2(M) \geqslant \mathsf{OR}_2(A) + \mathsf{OR}_2(C) + \mathrm{tr}(B)\,. \tag{iii}$$

A version of (i) for so-called "triangular" circuits was first proved by Grigoriev [39]. For Kneser–Sierpinski matrices, the bound (i) was independently proved by Boyar and Find [10], and Selezneva, and extended to (ii) by Boyar and Find in [11].

Proof. We first prove item (ii). For simplicity, we assume that the matrix C contains no zero rows; by an easy modification of the definition of set of edges E_3 below, the proof works also without this requirement.

Take an OR circuit computing $M\vec{x}$. It must compute $\vec{y}_1 = A\vec{x}_1$ and $\vec{y}_2 = B\vec{x}_1 + C\vec{x}_2$ where $(\vec{x}_1, \vec{x}_2) = \vec{x}$ and $(\vec{y}_1, \vec{y}_2) = \vec{y}$:

$$\begin{array}{c} \\ \vec{y}_1 \\ \vec{y}_2 \end{array} \begin{array}{c} \begin{array}{cc} \vec{x}_1 & \vec{x}_2 \end{array} \\ \begin{bmatrix} A & 0 \\ B & C \end{bmatrix} \end{array}$$

Since the circuit is optimal, every its edge must see at least one input, and must be seen by at least one output node. Consider the following three sets of edges:

$E_1 =$ edges seen by some outputs in $\vec{y}_1$; it is clear that $|E_1| \geqslant \mathsf{OR}(A)$.

$E_2 =$ edges seeing some inputs in $\vec{x}_2$; it is clear that $|E_2| \geqslant \mathsf{OR}(C)$.

$E_3 =$ edges seeing only inputs $\vec{x}_1$ and entering nodes seeing some inputs in $\vec{x}_2$.

It is not difficult to see, that these sets are disjoint. The sets E_2 and E_3 are disjoint by their definition. If E_1 and $E_2 \cup E_3$ shared an edge, then we could set $\vec{x}_1 = \vec{0}$, $\vec{x}_2 \neq \vec{0}$, and the circuit would wrongly output $\vec{y}_1 \neq \vec{0}$.

Observe that every path from $\vec{x}_1$ to $\vec{y}_2$ must contain exactly one edge in E_3 (thus, E_3 forms a minimal $\vec{x}_1 \to \vec{y}_2$ cut). This follows because otherwise, no node along these paths—including the last node—could see any input from $\vec{x}_2$. Thus, $|E_3| \geqslant \mathrm{rk}_\vee(B)$. This gives item (ii).

In the case of SUM circuits, the set E_3 gives a decomposition of B into disjoint rectangles, implying that $|E_3| \geqslant \mathrm{rk}_+(B)$. Together with $|E_1| \geqslant \mathsf{SUM}(A)$ and $|E_2| \geqslant \mathsf{SUM}(C)$, this yields item (i).

If the circuit has depth 2, then each rectangle corresponding to an edge in E_3 has either only one nonzero row or only one nonzero column. Thus, in this case the 1s of B can be covered by at most $|E_3|$ lines (rows and columns). By the König–Egerváry theorem, $|E_3| \geqslant \mathrm{tr}(B)$. This yields item (iii). □

The proof works for every commutative semiring $(S, +)$ with identity 0, in which $x + y = 0$ implies $x = y = 0$. This holds because then setting to 0 one of the inputs of a gate (in the fanin-2 version of the circuit) eliminates the need of that gate. This is no longer the case in, say, the XOR group $(S, \oplus)$, and no analogue of Theorem 3.18 is known for XOR circuits, even in depth 2.

3.6 Bounds for Kronecker products

Recall that the *Kronecker product* $A \otimes B$ of an $a \times b$ matrix A and an $n \times m$ matrix B is an $an \times bm$ block-matrix obtained by replacing 1-entries of A by copies of B. Let $\mathsf{L} \in \{\mathsf{SUM}, \mathsf{OR}, \mathsf{XOR}\}$.

Theorem 3.19 (Gál [32])**.** For every boolean matrices A and B,

$$\mathsf{L}_2(A \otimes B) \geqslant \mathrm{tr}(A) \cdot \mathsf{L}_2(B)\,.$$

Proof. The proof is based on a fact that, in depth 2, we have

$$\mathsf{L}_2 \begin{bmatrix} X & C \\ D & Y \end{bmatrix} \geqslant \mathsf{L}_2 \begin{bmatrix} X & 0 \\ 0 & Y \end{bmatrix} = \mathsf{L}_2(X) + \mathsf{L}_2(Y)\,. \tag{3.8}$$

To verify this, just take for each node on the middle level two its copies, and appropriately split the edges among these copies. Now take $r = \mathrm{tr}(A)$ 1-entries of A, with no two on the same row or column, replace

the remaining 1s by 0s, and let A' be the resulting matrix. By (3.8), $\mathsf{L}_2(A \otimes B) \geqslant \mathsf{L}_2(A' \otimes B) \geqslant r \cdot \mathsf{L}_2(B)$, as desired. □

In depths larger than 2, the fact (3.8) no longer holds. For example, if A is an $n \times n$ matrix, then

$$\mathsf{OR}\begin{bmatrix} A & A \\ A & A \end{bmatrix} \leqslant \mathsf{OR}(A) + 4n \quad \text{but} \quad \mathsf{OR}\begin{bmatrix} A & 0 \\ 0 & A \end{bmatrix} = 2 \cdot \mathsf{OR}(A)\,.$$

To see the upper bound, take a circuit F for A, draw $2n$ edges from the first and the second half of variables to input nodes of F, and draw $2n$ edges from the output nodes to the first and the second part of output variables.

The following lower bounds extend those in Nechiporuk's Theorem 3.9.

Theorem 3.20. Let A be a square matrix, and B a $(k+1, l+1)$-free matrix. Then

$$\mathsf{SUM}(A \otimes B) \geqslant \mathrm{rk}_{+}(A) \cdot \frac{|B|}{k \cdot l}\,, \tag{i}$$

$$\mathsf{OR}(A \otimes B) \geqslant \mathrm{rk}_{\vee}(A) \cdot \frac{|B|}{k \cdot l}\,. \tag{ii}$$

The lower bound (i) was recently proved by Find et al. [30] by using ideas of Mehlhorn's proof of Theorem 3.9 in [64]. Below we give a different proof following the ideas of the Nechiporuk–Pippenger type proof of this theorem given above. This proof is much simpler and, more importantly, it directly applies to OR circuits as well.

Proof. Take a minimal SUM circuit for $A \otimes B$. If A is an $n \times n$ matrix, then by the *type* of the m-th row (column) of $A \otimes B$ we will mean the number m mod n. Note that the rows of type s and columns of type t correspond to one and the same entry $b_{s,t}$ of $B = (b_{ij})$ (see Figure 3.1).

Let τ_v denote the number of different types among the input nodes seen by the node v. Let also E be the set of edges $e = (u, v)$ in our circuit such that

(a) either v is an output node and $\tau_v \leqslant l$,
(b) or $\tau_u \leqslant l$ and $\tau_v \geqslant l + 1$.

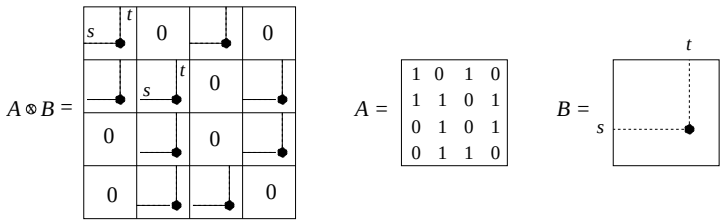


Figure 3.1: The rows of type s and columns of type t in $A \otimes B$ correspond to one and the same entry of B.

Let $E_{s,t} \subseteq E$ consist of all edges $e \in E$ such that e can see at least one input of type t, and is seen by at least one output of type s.

Due to the $(k+1, l+1)$-freeness of B, no edge $e = (u, v)$ in E can be seen by outputs of more than k different types, because $\tau_v \geqslant l+1$ (or the edge enters an output node). Moreover, no such edge e can see inputs of more than l different types. Thus, no edge $e \in E$ can belong to more than $k \cdot l$ of the sets $E_{s,t}$, implying that

$$|E| \geqslant \frac{1}{k \cdot l} \sum_{s,t} |E_{s,t}| = \frac{1}{k \cdot l} \sum_{(s,t)\colon\, b_{s,t}=1} |E_{s,t}| \,.$$

It remains, therefore, to show that, for every s, t,

$$|E_{s,t}| \geqslant \mathrm{rk}_+(A) \,.$$

This follows by observing that every path P from an input x_j of type t to an output y_i of type s must contain an edge in $E_{s,t}$. If y_i is seen by inputs of fewer than l types ($\tau_{y_i} \leqslant l$), then already the last edge of P entering y_i belongs to $E_{s,t}$ by (a). Otherwise, there must be an edge $e = (u, v)$ in P such that $\tau_u \leqslant l$ and $\tau_v \geqslant l+1$. In this case, e is in $E_{s,t}$ by (b).

So, $E_{s,t}$ forms a cut in a subcircuit connecting inputs of type t with outputs of type s (and hence implementing matrix A) . In fact, this is a minimal cut, because no two edges in E can lie on the same path. Since we have a SUM circuit, the rectangles R_e induced by the edges $e \in E_{s,t}$ form a decomposition of A, implying that the number $|E_{s,t}|$ of rectangles in this decomposition must be at least $\mathrm{rk}_+(A)$.

The proof in the case of OR circuit is the same with the exception that now the rectangles R_e need not be disjoint. □

3.7 Graph-theoretic bounds

SUM and OR circuits are *monotone* models: the presence of any input-output path forces the corresponding entry of the matrix to be non-zero. This is why we can use various rectangle arguments to lower-bound their size. The model of XOR circuits is non-monotone: an additional path can switch the entry from 1 to 0. This model is much harder to deal with, and only few general lower bound arguments are known for them. Most of them use the following "graph-theoretic" approach proposed by Valiant [106] and independently by Grigoriev [37].

(i) For some property $\mathcal{P}$ of graphs, show that every XOR circuit computing Ax must have this property.
(ii) Show that every graph with property $\mathcal{P}$ must have many edges.

As $\mathcal{P}$ one usually takes some connectivity property like "it is not possible to disconnect all (or almost all) input-output pairs by removing a small number of nodes". This has led to notions of a "concentrator", "superconcentrator", "grate", "meander", etc. The basic observation here is that already the rank of A determines the size of node-cuts in any XOR circuit computing Ax.

A *node-cut* in a circuit is a set of nodes, whose removal destroys all input-output paths. By Menger's theorem, the minimal size of such a cut is equal to the largest number of node-disjoint input-output paths.

Observation 3.21. If a circuit F computes Ax over $\mathbb{F}_2$, then every node-cut of F must have at least $\mathrm{rk}(A)$ nodes.

Proof. If the circuit F has a node-cut of size t, then the operator $x \mapsto Ax$ can take at most 2^t distinct values, since the output is determined by the values of nodes in the cut. On the other hand, if A has rank r, then the operator takes at least 2^r distinct values, implying that $t \geqslant r$. □

The same also holds for the submatrices of A: if a submatrix has rank r, then there must be at least r node-disjoint paths between the

corresponding subsets of input and output nodes. Note that the observation holds for *general* circuits where *arbitrary* boolean functions, not just XORs, are allowed as gates.

Suppose now that every $m \times m$ submatrix of A has rank at least r. Then the Observation 3.21 implies that for every circuit F computing Ax over $\mathbb{F}_2$, the associated digraph of F must have the following property:

> for every subset S of input nodes, and for every subset T of output nodes such that $|S|, |T| \geqslant m$, there must be at least r node-disjoint paths from S to T.

Such a digraph is called an (m, r)-*superconcentrator*.

In light of this relation between the rank of submatrices of A and the purely graph-theoretic property of circuits computing Ax, several variants of the superconcentrator property were investigated. The strongest version, called the *superconcentrator property*, requires the digraph to be an (m, m)-superconcentrator for all m. Unfortunately, in the boolean case, we do not know of any matrix all of whose large enough square submatrices have full rank over $\mathbb{F}_2$. Moreover, Valiant [105] has shown that superconcentrators of size $O(n)$ *exist*. Pippenger [82] has then shown that such small superconcentrators exist already in depth $O(\log n)$. Dolev et al. [22] substantially decreased the depth to the inverse of any primitive-recursive function. Still one can use superconcentrators to prove superlinear lower bounds for *constant*-depth circuits.[4]

Let $f : \mathbb{N} \to \mathbb{R}^+$ be a function, $f(n) \leqslant \log n$. An f-*superconcentrator* is an (m, r)-superconcentrator for $m = n/2^{f(n)}$ and $r = f(n)$. It is clear that depth-2 f-superconcentrators with $O(n \cdot f(n))$ edges exist: just take $\lceil f(n) \rceil$ nodes on the middle level, and join them with all input and all output nodes. Using a Ramsey-type reasoning, Alon and Maass have shown that this trivial upper bound is almost optimal.

Theorem 3.22 (Alon and Maass [4])**.** Every f-superconcentrator of depth 2 has $\Omega(n \cdot f(n))$ edges.

We postpone the proof of this theorem to the end of this section, and turn to its applications.

[4] We will give yet another application of superconcentrators in § 6.2.

Take an XOR circuit of depth 2 computing some linear form $y = Ax$ over $\mathbb{F}_2$. Suppose that every $m \times m$ submatrix of A with $m = \lceil n/2^{f(n)} \rceil$ has rank at least $f(n)$. By Observation 3.21, the circuit must be an f-superconcentrator, and hence, must have $\Omega(n \cdot f(n))$ edges.

Alon, Karchmer and Wigderson [3] combined this observation with Theorem 3.22 to prove the first superlinear lower bound on the XOR complexity (in depth 2). They have proved this bound for the Sylvester matrix, but a closer look at their argument shows that it works for many other matrices, as well.

Recall that a matrix A is (k,l)-*Ramsey matrix* if both the matrix and its complement are (k,l)-free. Say that A is *weakly Ramsey* if it is (n^{1-2c}, n^{1-c})-Ramsey for some constant $c > 0$. Note that only relatively large monochromatic submatrices are forbidden.

Theorem 3.23 (Alon, Karchmer and Wigderson [3])**.** For every weakly Ramsey $n \times n$ matrix A, we have $\mathsf{XOR}_2(A) = \Omega(n \log n)$.

Proof. We are going to apply Theorem 3.22 with $f(n) = c \log n - 1$, where the constant $c > 0$ comes from the definition of A being weakly Ramsey. Set $m = n/2^{f(n)} = 2n^{1-c}$. It is enough to verify that every $m \times m$ submatrix M of A has rank $r \geqslant f(n)$.

Since the span of any r rows of M can have at most 2^r distinct vectors, some row must appear in M at least $m/2^r$ times, implying that M must have a monochromatic $(m/2^r) \times (m/2)$ submatrix. Since A is weakly Ramsey and $m/2 = n^{1-c}$, we have that $m/2^r < n^{1-2c}$ must hold, from which $2^r > m/n^{1-2c} = 2n^c$, and hence, $r \geqslant f(n)$ follows. □

By Lindsey's Lemma, the $n \times n$ Sylvester matrix $H = H_n$ has no monochromatic $a \times b$ rectangle with $ab > n$. Thus, this matrix is weakly Ramsey (where we may take any $c < 1/3$ in the definition). So, we have that

$$\mathsf{XOR}_2(H) = \Omega(n \log n)\,. \tag{3.9}$$

In § 6.1 we will give a simpler and direct proof of a weaker lower bound $\mathsf{XOR}_2(H) \succcurlyeq n \ln n / \ln \ln n$, also due to [3]. This latter bound holds for any matrix whose every two columns differ in $\Omega(n)$ positions.

We already mentioned explicit constructions of maximal 2-free matrices, that is, $n \times n$ matrices A that are 2-free and still have at least

$\sqrt{n}$ ones in each row (see § 1.3).

Corollary 3.24. For every maximal 2-free $n \times n$ matrix A, we have

$$\mathsf{XOR}_2(A) = \Omega(n \log n)\,.$$

Proof. By Theorem 3.23, it is enough to show that every such matrix A is weakly Ramsey.[5] To see this, suppose that A has an $a \times b$ all-0 submatrix. Then A must have an $a \times (n-b)$ submatrix M with at least $\sqrt{n}$ ones in each row. By the upper bound (1.2) for the Zarankiewicz problem, we have that

$$a \cdot \sqrt{n} \leqslant |M| \leqslant a(n-b)^{1/2} + n - b < a(n-b)^{1/2} + n$$

from which $b \leqslant 2n^{3/2}/a$ follows. Now, if $a = n^{1-2c}$, then b cannot exceed $2n^{1/2+2c}$, which is smaller than n^{1-c}, when $c < 1/6$. □

Proof of Theorem 3.22 The theorem is an easy consequence of a more general result about so-called "meanders".

Let X be some alphabet, and $A, B \subseteq X$ be two disjoint subsets of size $|A|, |B| \geqslant n$. Given a string x over X, and subsets $S \subseteq A$ and $T \subseteq B$, a *link* between S and T in x is a substring of x whose first and last symbols belong to different sets S and T, and such that none of the remaining symbols lying in-between belongs to $S \cup T$. Let $x(S,T)$ denote the number of links between S and T in x.

Note that $x(S,T)$ is just the number of alternations between 0 and 1 in the string obtained from x as follows: replace by 0 each occurrence of a symbol from S, replace by 1 each occurrence of a symbol from T, and remove all remaining symbols. Consider, for example, the case when $X = [6]$, $S = \{1,3\}$ and $T = \{2,4\}$. Then the string

x	=	1	6	2	3	2	6	4	1	5	3	1	2	4	6	5	1	3	2
		0		1	0	1		1	0		0	0	1	1			0	0	1

has length 18, and the derived string has $x(S,T) = 7$ alternations between 0 and 1.

[5] This observation was inspired by Alon's result in [2] that the Singer $n \times n$ matrix (see § 1.3 for its definition) does not have all-0 submatrices of area larger than $n^{3/2}$.

A string x is an (n, s)-*meander* if $x(S, T) \geqslant s$ holds for all subsets $S \subseteq A$ and $T \subseteq B$ of size $|S| = |T| \geqslant n/2^s$. The question we are interested is: how short can an (n, s)-meander be?

Let us first show that every depth-2 superconcentrator gives a meander. Recall that a circuit is an f-*superconcentrator* if for every subset S of $|S| \geqslant n/2^{f(n)}$ input nodes, and for every subset T of $|T| \geqslant n/2^{f(n)}$ output nodes, there are at least $f(n)$ node-disjoint paths from S to T.

Observation 3.25. Every f-superconcentrator of depth-2 with n input nodes, n output nodes, and m edges gives an $(n, f(n))$-meander of length m.

Proof. As our alphabet we take $A \cup B$ where A is the set of all input nodes, and B is the set of all output nodes. Associate with each middle node a string consisting of all input nodes seen by this node, followed by all output nodes seeing that node. Let x be the concatenation of these strings. Clearly, the string x has length m. We claim that x is an (n, s)-meander with $s = f(n)$.

To show this, take any two subsets $S \subseteq A$ and $T \subseteq B$ of size $|S| = |T| \geqslant n/2^s$. Since we have an s-superconcentrator, there must be s node-disjoint paths between S and T. The segments of x, corresponding to the middle nodes of these s paths, give us s distinct links between S and T in x, implying that $x(S, T) \geqslant s$, as desired. □

By Observation 3.25, Theorem 3.22 is a direct consequence of the following more general result.

Theorem 3.26 (Alon and Maass [4])**.** Every (n, s)-meander has length $\Omega(n \cdot s)$.

Proof. Our first goal is to show that every meander must contain many symbols that appear often enough.

Claim 3.27. Let x be a string over X in which each $a \in A$ appears at most k_A times and each $b \in B$ appears at most k_B times. Let $k = k_A + k_B$. Then there exist subsets $S \subseteq A$ and $T \subseteq B$ of size $|S|, |T| \geqslant n/2^k$ such that $x(S, T) < k$.

Proof. We apply induction on k. If $k = 1$, then either k_A or k_B is 0, and we have $x(A, B) = 0$. For the induction step, assume w.l.o.g. that each symbol appears in x at least once (otherwise, extend x with the missing symbols in an arbitrary way).

Now examine the symbols of x one by one until we reach a location where we already have seen $n/2$ symbols of one of A and B but fewer than $n/2$ of the other; such a location must exist since $A \cap B = \emptyset$. Denote the prefix by y and the rest of x by z; hence, $x = yz$. Assume that $n/2$ symbols of A appear in y (the other case is handled identically). Let $A' \subseteq A$ be the set of symbols of A that appear in y, and $B' \subseteq B$ the set of symbols of B that *do not* appear in y. It follows that $|A'|, |B'| \geqslant n/2$.

Since every symbol of A' appeared in y, it can appear in z at most $k_A - 1$ times. So, by the induction hypothesis, there exist subsets $S \subseteq A'$ and $T \subseteq B'$ of size $|S|, |T| \geqslant (n/2)/2^{k-1} = n/2^k$ such that $z(S, T) < k - 1$. Since the prefix y of x can only contain symbols of A' but none of B', the entire string $x = yz$ can have at most one more link between S and T, implying that $x(S, T) < k$, as desired. □

Now we can finish the proof of Theorem 3.26 as follows. Let A, B be disjoint alphabets of size n, and let x be an (n, s)-meander of length m over $A \cup B$. Set $p = \lceil m/n \rceil$. If $4p + 1 \geqslant s$, then $m \geqslant n(s - 5)/4$, and we are done. So, assume that $4p + 1 < s$.

Let $A' \subseteq A$ be the set of all symbols $a \in A$ that appear at most $2p$ times in x, and let $B' \subseteq B$ be the set of all symbols $b \in B$ that appear at most $2p$ times in x. Clearly, $|A'|, |B'| \geqslant n/2$. Remove from x all symbols that are not in $A' \cup B'$, and consider the resulting string x'. When applied with $k_{A'} = k_{B'} = 2p$, Claim 3.27 gives us two subsets $S \subseteq A'$ and $T \subseteq B'$ of size $|S|, |T| \geqslant (n/2)/2^{4p} = n/2^{4p+1} \geqslant n/2^s$ such that the number $x'(S, T)$ of links between S and T in x' (and hence, also in x) is at most $4p < s$. But this is impossible since x is an (n, s)-meander. □

3.8 Rigidity lower bounds

Superconcentrator type arguments aim to show that any XOR circuit for a matrix A must have many edges, if submatrices of A have large

rank. One can, however, take a different route, and ask: how many edges must one remove from a circuit implementing A in order to reduce the rank of the computed transformation down to some given threshold r? This approach leads to the following notion of "matrix rigidity".

The *rigidity* of a matrix A over some field $\mathbb{F}$ is defined as the function $R_A(r)$ which, for each r, gives the minimum number of entries of A that one has to change in order to reduce the rank of A over $\mathbb{F}$ to r or less. We will consider the rigidity of boolean matrices over the field $\mathbb{F} = \mathbb{F}_2$; hence,

$$R_A(r) = \min\{|B| : \mathrm{rk}(A \oplus B) \leqslant r\}.$$

It is easy to show that $R_A(r) \leqslant (n-r)^2$ holds for every $n \times n$ matrix A. To see this, let B be the bottom-left $r \times r$ submatrix of A of full rank r (up to permutation of rows and columns, there must be one, if A has rank larger than r); hence, $A = \left[\begin{smallmatrix} B & C \\ D & E \end{smallmatrix}\right]$ where $\mathrm{rk}(B) = r$. The i-th column of C is a linear combination Bx of the columns of B for some vector x. Replace the i-th column of E by Dx (using the *same* vector x). This way, every column of the obtained matrix A' is a linear combination of the first r columns. Since we have only changed the entries of E, the upper bound $(n-r)^2$ on the rigidity follows.

Valiant [106] has shown that boolean $n \times n$ matrices with

$$R_A(r) \geqslant \frac{(n-r)^2 - 2n - \log n}{\log(2n^2)}$$

for all $r < n - \sqrt{2n + \log n}$, exist.

The concept of the rigidity of matrices itself was proposed by Valiant [106]. A similar notion of "separability" was independently proposed by Grigoriev [37]; as explained in [40, Section 15], the separability property is similar to being a "grate" in a sense of [106]. A comprehensive survey about the rigidity over *large* fields, its variants and applications can be found in the book by Lokam [61]. Here we restrict ourselves with applications of the rigidity over $\mathbb{F}_2$ for XOR circuits.

A directed acyclic graph is an $f(r)$-*grate* if no matter which r nodes are removed, at least $f(r)$ distinct input-output pairs will remain connected by paths. The faster $f(r)$ grows, the better the grate is.

Lemma 3.28 (Valiant [106])**.** For every matrix A, every XOR circuit computing Ax is an $R_A(r)$-grate.

Proof. Assume for the sake of contradiction that for some $1 \leqslant r \leqslant n$ it is possible to remove r nodes so that fewer than $R_A(r)$ distinct input-output pairs remain connected. This means that the matrix B implemented by the resulting circuit has $|B| < R_A(r)$ ones. However, the rows of B differ from the corresponding rows of A only by linear combinations of linear forms computed by the original circuit at the removed nodes. It follows that $A = B \oplus C$ for some matrix C of rank $\leqslant r$. Since $C = A \oplus B$, by the definition of rigidity, we have that $R_A(r) \leqslant |B| < R_A(r)$, a contradiction. □

The next task is to show that "good" grates (those with $f(r)$ growing fast enough) must have a superlinear number of edges. That $f(r)$ must grow indeed fast was shown by Schnitger in [96], where he constructed a sequence of $f(r)$-grates with a linear number of edges such that $f(r) \geqslant cn^{1/3}$ for all $r \leqslant cn$ and an absolute constant $c > 0$. Klawe [54] showed the existence of graphs with similar properties and additionally having only logarithmic depth.

On the other hand, Valiant [106] has earlier shown that $f(r)$-grates with $f(n/2) \geqslant n^{1+c}$ for a constant $c > 0$ must already have superlinear number of edges. He obtains this as a consequence of the following "depth-reduction" lemma, which itself generalizes and simplifies an analogous result of Erdős, Graham and Szemerédi [26].

A *labeling* of a directed acyclic graph G is an assignment of a positive integer to each its non-input node. Such a labeling is *legal* if, for each edge, the label of its second node is strictly greater than the label of the first mode. The *canonical labeling* is to assign each node the length of a longest directed path that terminates at that node.

Observation 3.29. The depth of G is the smallest number of distinct labels used by a legal labeling of G.

Proof. If the graph has depth d then the canonical labeling gives us a labeling using only d labels $1, \ldots, d$. On the other hand, since in any legal labeling, all labels along a directed path must be distinct, we have

that the depth of a graph does not exceed the number of labels used by any legal labeling. □

Lemma 3.30 (Valiant [106]). Let $D = 2^p$ and $1 \leqslant q \leqslant p$ be integers. In any directed graph with M edges and depth D it is possible to remove at most $(q/p)M$ edges so that the depth of the resulting graph does not exceed $D/2^q$.

Schnitger in [95] has shown that this lemma cannot be substantially improved. Namely, he presented a sequence of amazingly simple n-node graphs with $n \log n$ edges such that, for every constant $0 \leqslant c < 1$, a constant fraction of edges must be removed in order to reduce the depth till n^c. In [96] he proved the existence of graphs with a *linear* number of edges and with the same property.

Proof of Lemma 3.30. Consider any directed graph with M edges and depth D, and consider the canonical labeling using labels $1, \ldots, D$. For $i = 1, \ldots, p$, let E_i be the set of all edges, the binary representations of labels of whose endpoints differ in the i-th position (from the left) for the *first time*.

If edges in E_i are removed from the graph, then we can relabel the nodes using integers $1, \ldots, D/2$ by simply deleting the i-th bit in the binary representations of labels. It is not difficult to see that this is a legal labeling (of the new graph): if an edge (u, v) survived, then the first difference between the binary representations of the old labels of u and v was *not* in the i-th position; hence, the new label of u remains strictly smaller than that of v. Consequently, if any $q \leqslant p$ of the *smallest* sets E_i are removed, then at most qM/p edges are removed, and a graph of depth at most $D/2^q$ remains. □

Valiant used his depth-reduction to relate the XOR complexity of matrices with their rigidity.

Theorem 3.31 (Valiant [106]). Let $\epsilon, c, k > 0$ be constants, and A be a boolean $n \times n$ matrix with $R_A(n/2) > \epsilon n^{1+c}$. Then every XOR circuit of fanin two and depth $k \log n$ computing Ax has at least $np/3 \log p$ gates, where $p = \log \log n$.

Proof. Assume we have fewer than $np/3\log p$ nodes. By applying Lemma 3.30 with q about $\log p$, we obtain that some set of $n/2$ nodes can be removed so as to leave no paths longer than $d = k\log n/\log\log n$. Hence, each of the n output nodes will be connected to at most $2^d = n^{k/\ln\ln n} = o(n^c)$ inputs after the deletion, implying that at most $o(n^{1+c})$ input-output pairs will remain connected. This implies that the circuit is not an $R_A(n/2)$-grate, contradicting Lemma 3.28. □

For small-depth circuits, we have the following relation with the rigidity.

Theorem 3.32 (Pudlák [86])**.** For every $n \times m$ matrix A, every integer $d \geqslant 1$, and every r,

$$\mathsf{XOR}_d(A) \geqslant r \cdot \left(\frac{R_A(r)}{n}\right)^{1/d}.$$

Proof. Take an XOR circuit of depth d for A, and let S be the number of edges in it. A node is *large* if its out-degree is $\geqslant S/r$; otherwise, the node is *small*. Note that there cannot be more than r large nodes. Let L be the $n \times m$ matrix where $L[i,j]$ is the parity of the number of paths from i to j that hit at least one large node. Thus $\mathrm{rk}(L) \leqslant r$. The matrix $B := A \oplus L$ is computed by a subcircuit formed by paths that go through only small nodes. As there are at most $n(S/r)^d$ such paths, we have $|B| \leqslant n(S/r)^d$. On the other hand, $|B| \geqslant R_A(r)$, and the desired lower bound on S follows. □

Unfortunately, the largest known lower bound on the rigidity of *explicit* boolean matrices are only of the form $R_A(r) \succcurlyeq n^2/r$; some of them have an additional $\ln(n/r)$ factor. However, Pudlák [86] has shown that even already existing lower bounds on the rigidity *can* yield superlinear lower bounds for depth-2 circuits.

When dealing with circuits of depth 2, the following particular case of the *Karamata inequality* (see e. g., [45, p. 89]) turned out to be very useful: If $p_1 \geqslant \ldots \geqslant p_m \geqslant 0$ and $q_1 \geqslant \ldots \geqslant q_m \geqslant 0$ satisfy

$$p_r^2 + \cdots + p_m^2 \geqslant q_r^2 + \cdots + q_m^2 \quad \text{for all } r = 1, 2, \ldots, m,$$

then

$$p_1 + \cdots + p_m \geqslant q_1 + \cdots + q_m\,.$$

This, in particular, yields the following

Lemma 3.33. If a sequence of numbers $p_1 \geqslant \ldots \geqslant p_m \geqslant 0$ satisfies

$$\sum_{i=r+1}^{m} p_i^2 \geqslant \frac{1}{r} \cdot \Delta^2$$

for all $s \leqslant r \leqslant t < m$, then

$$\sum_{i=1}^{m} p_i \geqslant \Delta \cdot \ln \frac{t}{s}\,.$$

Proof. Set $q_i = \Delta/i$ for $i = s, s+1, \ldots, t$, and $q_i = 0$ for $i > t$. For every r between s and t, we have

$$\sum_{i=r+1}^{m} p_i^2 \geqslant \frac{1}{r} \cdot \Delta^2 \geqslant \Delta^2 \sum_{i=r+1}^{\infty} \frac{1}{i^2} \geqslant \sum_{i=r+1}^{m} q_i^2\,.$$

The Karamata inequality yields

$$\sum_{i=1}^{m} p_i \geqslant \sum_{i=s}^{t} q_i = \Delta \sum_{i=s}^{t} \frac{1}{i} \geqslant \Delta \cdot \ln \frac{t}{s}\,. \qquad \square$$

Theorem 3.34 (Pudlák [86])**.** Let A be an $n \times n$ matrix, $f : \mathbb{N} \to \mathbb{R}^+$ a non-decreasing function, $1 \leqslant s \leqslant t \leqslant n$ be integers. If

$$R_A(r) \geqslant \frac{1}{r} \cdot f(n)^2$$

holds for all for all r between s and t, then

$$\mathsf{XOR}_2(A) \geqslant 2 f(n) \ln \frac{t}{s}\,.$$

Proof. For arbitrary factorization $A = BC$ over $\mathbb{F}_2$, we will prove the desired lower bound on $|B| + |C|$. Let B be an $n \times m$ matrix and C be an $m \times n$ matrix. Notice that $m \geqslant t$, since $\mathrm{rk}(A) \geqslant t$ by the condition on rigidity.

Let b_i be the number of 1s in the i-th column of B and c_i be the number of 1s in the i-th row of C. Thus, by setting the i-th column of B (alternatively, the i-th row of C) to zero we affect at most $b_i c_i$ entries of A.

We may assume that $b_1 \geqslant \ldots \geqslant b_m$ and $c_1 \geqslant \ldots \geqslant c_m$. If not, consider the corresponding permutation of columns of B and rows of C, and note that it does not affect rigidity properties of A.

By setting any $m-r$ columns of the matrix B to zero we decrease its rank, and hence, the rank of the product BC to at most r. With the choice of the $m-r$ rightmost columns and by the condition on rigidity we have

$$\sum_{i=r+1}^{m} b_i c_i \geqslant f(n)^2/r$$

for all $r \leqslant t$. Together with the arithmetic-geometric mean inequality $(x+y)^2/4 \geqslant xy$, this yields

$$\sum_{i=r+1}^{m} \left(\frac{b_i + c_i}{2f(n)}\right)^2 \geqslant \sum_{i=r+1}^{m} \frac{b_i c_i}{f(n)^2} \geqslant \frac{1}{r}.$$

Lemma 3.33 applied with $\Delta = f(n)$ and $p_i = (b_i + c_i)/2f(n)$ yields

$$|B| + |C| = \sum_{i=1}^{m} (b_i + c_i) \geqslant 2f(n) \ln \frac{t}{s}\,. \qquad \square$$

Pudlák and Vavrín [90] have shown that the full triangular matrix T_n satisfies the condition of Theorem 3.34 with both $f(n)$ and t/s at least cn for a constant $c > 0$. In fact, they have even found an *exact* expression for $R_{T_n}(r)$. Actually, when combined with Theorem 3.34, even non-exact bound yields almost asymptotically tight bound on $\mathsf{XOR}_2(T_n)$.

By (1.6), we have that $\mathsf{XOR}_2(T_n) \leqslant \mathsf{SUM}_2(T_n) \leqslant n \log n + n$.

Lemma 3.35 (Pudlák and Vavrín [90], Pudlák [86])**.** For every sufficiently large n,

$$\mathsf{XOR}_2(T_n) \geqslant (1 - o(1)) n \ln n\,.$$

Proof. We first prove a lower bound on the rigidity of T_n, and then apply Theorem 3.34. Let B be any matrix such that $\mathrm{rk}(T_n \oplus B) \leqslant r$. We will prove that $|B| \geqslant (1 - o(1))n^2/4r$ holds for all values of r in a sufficiently large interval. For this, let the rows of T_n be numbered by the number of 1s in them. Consider some s rows of T_n with numbers l_i

satisfying $l_1 < \ldots < l_s$. It is easy to verify that the sum of these rows has at least

$$(l_s - l_{s-1}) + (l_{s-2} - l_{s-3}) + \cdots \geqslant \frac{s}{2} \min_i \{l_{i+1} - l_i\} \tag{3.10}$$

ones. (Here and below, under a "sum" we understand a sum modulo 2.)

Our first goal is to show that B must contain a row with at least $\lfloor n/(2r+2) \rfloor$ ones. Set $k = \lfloor n/(r+1) \rfloor$, and consider rows of matrices T_n, B and $T_n \oplus B$ with numbers $k, 2k, \ldots, (r+1)k$. Since $\mathrm{rk}(T_n \oplus B) \leqslant r$ some s of these rows have zero sum in $T_n \oplus B$. By (3.10), the sum of these rows in T_n as well as in B has weight at least $sk/2$, so B contains a row with at least $k/2$ ones, as desired.

After removing the latter row and a column with the same number from all three matrices, we still have $\mathrm{rk}(T_{n-1} \oplus B') \leqslant r$, where B' is a matrix obtained from B. We continue the above procedure of reducing matrices until possible. At the end, the total number of 1s in removed rows of B is at least

$$\sum_{i=1}^{n} \left\lfloor \frac{i}{2r+2} \right\rfloor \geqslant \frac{n(n+1)}{4(r+1)} - n \geqslant \frac{1}{r} \cdot f(n)^2$$

with $f(n)^2 = (1 - o(1))n^2/4$ when $\omega(1) \leqslant r \leqslant o(n)$, and the desired lower bound on $\mathsf{XOR}_2(T_n)$ follows from Theorem 3.34. □

Slightly larger than $n \log n$ lower bounds were proved for generator matrices of asymptotically good linear codes. Every $m \times n$ matrix A generates a linear code $\{Ax \colon x \in \mathbb{F}_2^n\} \subseteq \mathbb{F}_2^m$. The distance d of the code is the smallest Hamming distance $|y \oplus z|$ between any two codewords, that is,

$$d = \min\{|Ax| \colon x \neq \vec{0}\}\,.$$

The relative distance is $\delta = d/m$. The smaller the length m of codewords is, and the larger the relative distance is, the better is the code. A code is called *asymptotically good*, if $\delta > 0$ is a constant and m is at most constant times larger than the message length n. Such linear codes can be explicitly constructed; for instance, Spielman [101] constructed explicit good code matrices A with $\mathsf{XOR}(A) \preccurlyeq n$.

Theorem 3.36 (Pudlák and Rödl [88]). Let A be an $m \times n$ generator matrix of a code with the relative distance $\delta > 0$. Then for $r \leqslant n/16$, the rigidity of A satisfies

$$R_A(r) \geqslant \frac{\delta}{8} \cdot \frac{mn}{r + \log(n/r)} \cdot \log \frac{n}{r}\,.$$

A similar result was proved earlier by Friedman [31].

Proof. Let $r \leqslant n/16$, and fix an even integer k such that

$$\frac{2r}{\log(n/r)} \leqslant k < \frac{2r}{\log(n/r)} + 2\,.$$

With this choice of k, it is enough to show that

$$R_A(r) \geqslant \frac{\delta mn}{4k}\,. \tag{3.11}$$

Suppose we change fewer than $\delta mn/4k$ entries in the matrix. Then there will be at least $n/2$ columns with $t \leqslant \delta m/2k$ entries changed in each of them. Let M be the corresponding $n \times (n/2)$ submatrix of A. The choice of k implies that

$$r \leqslant \frac{k}{2} \log \frac{n}{k} < \log \binom{n/2}{k/2}\,.$$

On the other hand, the condition on A implies that $|Mx| \geqslant \delta m - k \cdot t \geqslant \delta m/2 \neq 0$ must hold for every nonzero vector $x \in \{0,1\}^{n/2}$ with at most $k/4$ ones. This, in particular, means that all $\binom{n/2}{k/2}$ sums over all $k/2$-tuples of columns of M must be distinct, implying that

$$\operatorname{rk}(M) \geqslant \log \binom{n/2}{k/2} > r\,.$$

Thus, at least $\delta mn/4k$ entries of the matrix A should have been changed to reduce its rank until r or fewer. This gives the desired lower bound (3.11) on the rigidity, proving the theorem. □

By adding zero columns to matrices from Theorem 3.36, we obtain $n \times n$ matrices A with rigidity

$$R_A(r) \succcurlyeq \frac{n^2}{r} \log \frac{n}{r}\,,$$

Table 3.1: Summary of general lower bounds.

Bound	**Property of A**
$\mathsf{SUM}(A) \geqslant 3\log_3 \lvert\det(A)\rvert$	
$\mathsf{SUM}_d(A) \geqslant dn\lvert\det(A)\rvert^{2/dn}$	
$\mathsf{OR}_2(A) \geqslant n\log\frac{n}{k-1}$	zero diagonal and no $k \times k$ all-0 principal minor
$\mathsf{OR}(A) \geqslant \frac{\lvert A\rvert}{k^2}$	k-free
$\mathsf{OR}_2(A) \geqslant \frac{\lvert A\rvert}{k}$	k-free
$\mathsf{OR}(A) \geqslant \frac{3\lvert A\rvert}{r}\log_3\frac{\lvert A\rvert}{n}$	no $a \times b$ rectangle with $ab > r$
$\mathsf{OR}_d(A) \geqslant \frac{d\lvert A\rvert}{r}\left(\frac{\lvert A\rvert}{n}\right)^{1/d}$	no $a \times b$ rectangle with $ab > r$
$\mathsf{XOR}_2(A) \geqslant \epsilon n\log n$	weakly Ramsey
$\mathsf{XOR}_2(A) \geqslant 2f(n)\ln\frac{t}{s}$	$R_A(r) \geqslant \frac{1}{r}f(n)^2$ for all $s \leqslant r \leqslant t$
$\mathsf{XOR}_2(A) \geqslant \epsilon k \cdot \frac{\ln n}{\ln\ln n}$	$(2, n-k)$-Ramsey (Theorem 6.1)

for all $\ln n \ll r \leqslant n/c$ and a constant $c > 1$. By taking $f(n) = \Theta(n\sqrt{\ln n})$ in Theorem 3.34, we obtain that

$$\mathsf{XOR}_2(A) \succcurlyeq n\ln^{3/2} n$$

holds for every such matrix. This remained the strongest known lower bound for XOR circuits of depth 2 for many years. Recently, Gál et al. [33] improved this to $\Omega(n(\ln n/\ln\ln n)^2)$, and showed that no larger lower bound can be obtained by only using the goodness of the code. We will describe their argument in § 6.2.

4

Complexity of Some Basic Matrices

Here we give applications of the general lower bound techniques described in the previous chapter to some basic matrices defined in § 1.3. Besides being important objects, these matrices demonstrate the limitations of these techniques.

4.1 Full triangular matrices

Complexity of full triangular matrices T_n was considered by many authors. In particular, tight (up to constant factors) bounds on the depth-d OR complexity and XOR complexity of these matrices for all $d \geqslant 1$ are already known. For depth-2 OR circuits, Corollary 3.8 and (1.6) yield an asymptotically tight estimate

$$\mathsf{OR}_2(T_n) \sim n \log n \,. \tag{4.1}$$

For depth-2 XOR circuits, Lemma 3.35 and (1.6) also yield an estimate

$$\mathsf{XOR}_2(T_n) \asymp n \log n \,.$$

The bound $\mathsf{OR}_2(T_n) \asymp n \log n$ was proved earlier by Tarjan [103]. To state known results about the complexity of T_n for *larger* depths, we

introduce a sequence of nondecreasing functions $\lambda_d(n)$ taking natural values as follows: $\lambda_1(n) := n$, $\lambda_2(n) := \lfloor \log n \rfloor$, $\lambda_3(n) := 1 + \lfloor \log \log n \rfloor$, and for $d > 3$, $\lambda_d(n)$ is the smallest k such that the k-fold superposition of $\lambda_{d-2}(n)$ is $\leqslant 1$.

Theorem 4.1. For every integer $d \geqslant 1$,

$$\mathsf{OR}_d(T_n) \asymp n \cdot \lambda_d(n) \quad \text{and} \quad \mathsf{XOR}_d(T_n) \asymp n \cdot \lambda_d(n)\,.$$

The multiplicative constants in these estimates depend only on the depth.In particular, we have that

$$\mathsf{OR}_3(T_n), \mathsf{XOR}_3(T_n) \asymp n \ln \ln n\,.$$

Bounds on the OR complexity of T_n were independently proved by Chandra, Fortune and Lipton [15, 14], and Grinchuk [41]. Bounds on the XOR complexity of T_n were proved by Pudlák [86] using some techniques for superconcentrators developed by Dolev, Dwork, Pippenger, and Wigderson [22], and by Alon and Pudlák [5].

4.2 Intersection matrices

By (1.9), we have that $\mathsf{OR}_2(Ð) \leqslant n \log n$. Moreover, Lemma 2.5 yields $\mathsf{OR}_3(Ð) \preccurlyeq n$. On the other hand, Corollary 3.8 yields $\mathsf{OR}_2(Ð) \geqslant n \log n$. Thus,

$$\mathsf{OR}_2(Ð) = n \log n \quad \text{and} \quad \mathsf{OR}_3(Ð) \asymp n\,. \tag{4.2}$$

4.3 Kneser–Sierpinski matrices

The Kneser–Sierpinski $n \times n$ matrix (or the disjointness matrix) $D = D_n$ is the complement of the intersection matrix. For this matrix, we have

$$\mathsf{OR}(D) \sim \tfrac{1}{2} n \log n\,. \tag{4.3}$$

The upper bound $\mathsf{OR}(D) \leqslant \mathsf{SUM}(D) \preccurlyeq n \log n$ follows from Lemma 2.6. Actually, when applied to this specific matrix, Lemma 2.6 is a bit too slack to provide best possible bounds. Next we give a simple construction showing that $\mathsf{OR}(D) \leqslant \frac{1}{2} n \log n + 2n$.

Consider a directed graph G on $n = 2^r$ nodes v_S corresponding to distinct subsets $S \subseteq [r]$. The graph contains an edge (v_S, v_T) if and only if $S \subset T$ and $|T| = |S| + 1$. (It is easy to see that G is a boolean cube.) By definition, G contains a path from v_S to v_T if and only if $S \subset T$. To complete G to an OR circuit implementing D we add n input nodes x_S and n output nodes y_S and also edges (x_S, v_S) and $(v_S, y_{\overline{S}})$ for each subset S, where $\overline{S} = [r] \setminus S$. Clearly, an input x_S is connected with an output y_T via oriented path if and only if S is a subset of $\overline{T}$, that is, if S and T are disjoint, as required. By the construction, the circuit has $\frac{1}{2}n \log n + 2n$ edges.

On the other hand, since D is a full triangular matrix with some 1s below the diagonal switched to 0ss, we have $\mathrm{rk}_\vee(D) = n$. Theorem 3.18, together with the recursive definition (1.8) of D_n, yields the recursion $\mathsf{OR}(D_{2n}) \geqslant 2 \cdot \mathsf{OR}(D_n) + n$, which results in $\mathsf{OR}(D) \geqslant \frac{1}{2}n \log n + n$. The lower bound $\mathsf{SUM}(D) \geqslant \frac{1}{2}n \log n$ was proved independently by Boyar and Find [10], and Selezneva [98]. The bound was extended to OR circuits in [11] (though it was implicit already in [10, 98]).

In depth 2, we have the following bounds.

Lemma 4.2.

$$\mathsf{OR}_2(D) \succcurlyeq n^{1.16} \quad \text{and} \quad \mathsf{SUM}_2(D) \preccurlyeq n^{\log_2(1+\sqrt{2})} < n^{1.28}.$$

Proof. To show the lower bound, consider the submatrix D^α of D whose rows and columns correspond to subsets $u \subseteq [r]$ of size $|u| = \alpha r$. The submatrix has $|D^\alpha| = \binom{r}{\alpha r}\binom{r-\alpha r}{\alpha r}$ ones, and is obviously k-free for $k = \binom{r/2}{\alpha r} + 1$. Using the bound $\binom{n}{\alpha n} = \frac{1}{\Theta(\sqrt{n})} 2^{nH(\alpha)}$ where $H(\alpha) = \log \left(\frac{1}{\alpha}\right)^\alpha \left(\frac{1}{1-\alpha}\right)^{1-\alpha}$ is the binary entropy function, and using Theorem 3.9, we obtain:

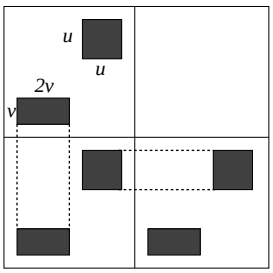
$$\mathsf{OR}_2(D) \geqslant \mathsf{OR}_2(D^\alpha) \geqslant \frac{|D^\alpha|}{k} \geqslant 2^{r \cdot h(\alpha) - o(r)},$$

where $h(\alpha) = H(\alpha) + (1-\alpha)H(\frac{\alpha}{1-\alpha}) - \frac{1}{2}H(2\alpha)$. By taking $\alpha = 0.4$, we have $\mathsf{OR}_2(D) \succcurlyeq n^{1.16}$.

To show the upper bound, we use the recursive definition of the disjointness $n \times n$ matrices D_n with $n = 2^r$, $r = 1, 2, \ldots$

$$D_{2n} = \begin{bmatrix} D_n & 0 \\ D_n & D_n \end{bmatrix} \quad \text{with} \quad D_2 = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}.$$

The desired decomposition of D_n will consist of "squares" (all-1 submatrices with equal side lengths) and "rectangles" whose side lengths have the ratio $1:2$. To decompose the matrix D_{2n}, we use the decompositions of its three submatrices D_n, as shown above. In every triple of rectangles, we merge two of them along their longer side as depicted below (of course, squares and rectangles need not to consist of adjacent columns):



In this way, a $u \times u$ square from the decomposition of D_n generates one square of the same dimension, and one $u \times 2u$ rectangle in the decomposition of D_{2n}. One $v \times 2v$ rectangle generates one rectangle of the same dimensions, and a $2v \times 2v$ square. Thus, if we let u_n be the sum of lengths of the sides of squares, and v_n the sum of lengths of the shorter sides of the rectangles in the decomposition of D_n, then we obtain the recursion

$$\begin{bmatrix} u_{2n} \\ v_{2n} \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} u_n \\ v_n \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix}^r \cdot \begin{bmatrix} u_2 \\ v_2 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix}^r \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

The eigenvalues of the matrix $A = \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix}$ are $1 \pm \sqrt{2}$. So,

$$\begin{bmatrix} u_n \\ v_n \end{bmatrix} = P \cdot \begin{bmatrix} 1+\sqrt{2} & 0 \\ 0 & 1-\sqrt{2} \end{bmatrix}^r \cdot P^{-1} \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

for an invertible 2×2 matrix P. Thus, both u_r and v_r are at most a constant times $(1+\sqrt{2})^r$, as desired. □

Since no rectangle in D^α can have area larger than $\binom{r/2}{\alpha r}^2$, Theorem 3.12 yields that

$$\mathsf{OR}_d(D) \gtrcurlyeq n^{1+c_d}$$

for a constant $c_d > 0$ depending only on the depth d. Routine computations give, for example, $\mathsf{OR}_3(D) \succcurlyeq n^{1.03}$, $\mathsf{OR}_4(D) \succcurlyeq n^{1.01}$, etc.

4.4 Sylvester matrices

For the Sylvester $n \times n$ matrices $H = H_n$ with $n = 2^r$, (1.11) and (3.9) yield

$$n \log n \preccurlyeq \mathsf{XOR}_2(H) \leqslant n \log n\,.$$

For circuits of larger depths, Lemma 2.5 gives $\mathsf{XOR}_3(H) \preccurlyeq n$. For OR circuits, Lemma 3.17 and Lemma 2.6 yield

$$(2 - o(1))n \log n \leqslant \mathsf{OR}(H) \leqslant (4 + o(1))n \log n\,.$$

In the class of bounded-depth OR circuits, we have lower bounds

$$\mathsf{OR}_d(H) \geqslant (c_d - o(1))n^{1+1/d} \tag{4.4}$$

with $c_d = d/2^{1/d}$. This follows from the area lower bound (Theorem 3.12) because H has $|H| = \binom{n}{2}$ ones, and since the maximum area of a rectangle in H is $n/2$ (see Observation 3.15). These bounds extend the lower bounds of Corollary 3.4 to OR circuits. The lower bound $\mathsf{OR}_2(H) \succcurlyeq n^{1.5}$ was also proved by Tarjan [103].

Let us now show that the order of magnitude of the lower bound (4.4) is correct.

Theorem 4.3. Let $d \geqslant 2$, and n be a power of 2^d. Then

$$\mathsf{SUM}_d(H) \leqslant 2(d-1)n^{1+1/d}\,.$$

In particular, in depth 2 we have

$$(\sqrt{2} - o(1))n^{1.5} \leqslant \mathsf{OR}_2(H) \leqslant \mathsf{SUM}_2(H) \leqslant 2n^{1.5}\,.$$

Proof. For simplicity, we give a full proof only for the case $d = 2$. The proof for larger depths goes along the same lines. The idea is to simultaneously implement *both* the matrix $H = H_n$ and its complement $\overline{H}$ using the recursive definition (1.10) of H. To visualize the construction, we will treat each edge leaving an input node as either *positive* or *negative*; the remaining edges are ordinary. We then say that a circuit implements a boolean matrix $A = (a_{ij})$ if:

1. for any i and j there exists unique path connecting j-th input with the i-th output;
2. if $a_{ij} = 1$ then this path goes through a positive edge;
3. if $a_{ij} = 0$ then this path goes through a negative edge.

Note that we can obtain an ordinary SUM circuit for matrix A by just removing all negative edges, as well a SUM circuit for $\overline{A}$ by removing all positive edges. If we invert the "polarities" of the edges (that is, make positive edges negative and negative edges positive) in a circuit for A, then we obtain a circuit for the $\overline{A}$.

Let $\widehat{H}_n = \left[\begin{smallmatrix} H_n \\ \overline{H}_n \end{smallmatrix}\right]$ stand for a $(2n, n)$-matrix composed of matrices H_n and $\overline{H}_n$. To prove the theorem we will show that depth-2 circuit for matrix $\widehat{H}_{4n}$ can be derived as a union of eight circuits for matrix $\widehat{H}_n$ (some of the circuits may be inverted).

Let us rewrite the recursive definition of $\widehat{H}_n$. For convenience we split rows and columns of the matrix into groups denoted by Y_i's and X_j's:

$$\widehat{H}_1 = \begin{array}{c} \\ y_1 \\ y_2 \end{array} \begin{array}{c} x \\ \left[\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right] \end{array} \qquad \widehat{H}_{4n} = \begin{array}{c} \\ Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \\ Y_6 \\ Y_7 \\ Y_8 \end{array} \begin{array}{c} \begin{array}{cccc} X_1 & X_2 & X_3 & X_4 \end{array} \\ \left[\begin{array}{cccc} H & H & H & H \\ H & \overline{H} & H & \overline{H} \\ H & H & \overline{H} & \overline{H} \\ H & \overline{H} & \overline{H} & H \\ \overline{H} & \overline{H} & \overline{H} & \overline{H} \\ \overline{H} & H & \overline{H} & H \\ \overline{H} & \overline{H} & H & H \\ \overline{H} & H & H & \overline{H} \end{array}\right] \end{array} \quad \text{with} \quad H = H_n\,.$$

Circuits for $n = 1$ and $n = 4$ are shown in Figure 4.1. Thick lines on the first level are used to represent positive edges and thin dashed lines represent negative edges. The second circuit will also serve as a circuit for the general case (for $n = 4$, groups X_i and Y_j are single inputs and outputs).

To jump from n to $4n$ we introduce a procedure of merging of circuits. Figure 4.2(a) shows symbolically a circuit for $\widehat{H}_n$. We denote by h a matrix implemented at the nodes of the middle level in a circuit.

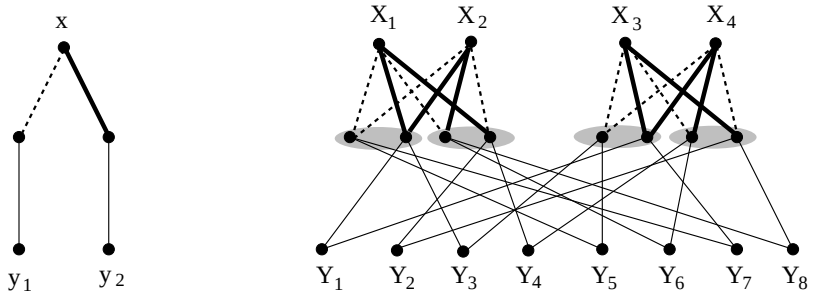


Figure 4.1: Circuits for $\widehat{H}_1$ and $\widehat{H}_4$.

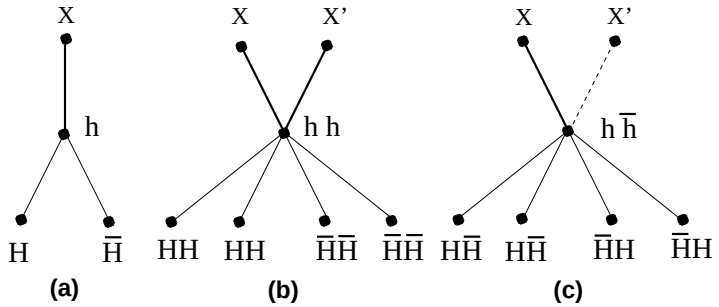


Figure 4.2: Superposition of circuits.

Consider two groups of inputs X and X' and combine the circuits implementing the matrix $\widehat{H}_n$ for each of the groups via merging of corresponding nodes on the middle level. The obtained circuit is shown in Figure 4.2(b). Its middle level implements the matrix $[h|h]$, hence the circuit itself implements two pairs of matrices $[H|H]$ and $[\overline{H}|\overline{H}]$, where $H = H_n$.

If in the circuit from Figure 4.2(b) we invert the (polarities of) edges connected with inputs in X', then we obtain a circuit with middle level implementing the matrix $[h|\overline{h}]$, and two pairs of matrices $[H|\overline{H}]$ and $[\overline{H}|H]$ being implemented at the outputs (Figure 4.2(c)).

Finally, a circuit for $\widehat{H}_{4n}$ can be derived as a union of two circuits from Figure 4.2(b) and two circuits from Figure 4.2(c). Inputs of these circuits are connected to groups X_i and outputs are connected to

Table 4.1: Matrices witnessing the optimality of lower bounds.

Result	Witness	Reference
Thms 3.1, 3.2 and 3.3	Sylvester matrix H_n	Thm 4.3
Thms 3.34 and 3.5	full triangular matrix T_n	Sect. 4.1
Thm 3.9	Obs. 3.10, random matrix	
Thm 3.12	Sylvester matrix H_n	Thm 4.3
Thm 3.18(i),(ii)	Kneser–Sierpinski matrix D_n	Eq. (4.3)
Thm 3.18(iii)	full triangular matrix T_n	Eq. (4.1)
Thms 3.19 and 3.20	$I_n \otimes A$, A dense k-free	Obs. 3.10

groups Y_j in a way shown in Figure 4.1, where middle levels of merged subcircuits are depicted.

It is not difficult to verify (by induction) that the circuit for $\widehat{H}_n$ contains $n^{1.5}$ positive and $n^{1.5}$ negative edges at the first level, $n^{1.5}$ edges connected with outputs implementing the matrix H_n and $n^{1.5}$ edges connected with outputs implementing its complement $\overline{H}_n$. Thus, the claim of the theorem in the case of depth $d = 2$ follows.

In the general case, we start from the trivial depth-d circuit (with polarities) for $\widehat{H}_{2^d}$ provided by Lemma 2.6. Next, in the proof of induction step from n to $2^d n$ we use the former circuit as a guide for merging subcircuits. Finally, we have an implementation of $\widehat{H}_n$ via depth-d circuit with $2n^{1+1/d}$ edges on each level: half of edges on the first and on the last levels are to be eliminated after reduction to the SUM circuit implementing the matrix H_n alone. □

Table 4.1 shows that, with two exceptions, all the general lower bounds proved in § 3 are provably optimal (or almost optimal). The only exceptions are Theorems 3.31 and 3.32 whose optimality we do not know.

5

Complexity Gaps

For the maximum $\mathsf{OR}(n)$, $\mathsf{XOR}(n)$ and $\mathsf{SUM}(n)$ over all $n \times n$ matrices, Theorem 2.4 yields

$$\mathsf{SUM}(n) \sim \mathsf{OR}(n) \sim \mathsf{XOR}(n) \sim \frac{n^2}{2 \log n} .$$

Thus the "worst-case" behavior of all three measures is essentially the same. But the situation changes drastically, if we compare the complexities of a fixed matrix A: here we may have growing gaps. Note that the largest possible gap cannot exceed $n/2 \log n$.

5.1 SUM/OR gaps

Although non-trivial OR/XOR-gaps were long known, the SUM/OR-gaps were apparently not known until recently. Lemma 2.10 and Theorem 3.20(i) suggest that a possible gap could appear on Kronecker products $B \otimes A$ of particular $m \times m$ matrices: Lemma 2.10 gives an upper bound

$$\mathsf{OR}(B \otimes A) \preccurlyeq \mathrm{rk}_\vee(B) \cdot \frac{m^2}{\log m} ,$$

and Theorem 3.20(i) gives a lower bound

$$\mathsf{SUM}(B \otimes A) \geqslant \mathrm{rk}_+(B) \cdot \frac{|A|}{k^2}\,,$$

if A is k-free. So, we only need to take a matrix B whose SUM-rank is much larger than the OR-rank, and a matrix A which has $\Omega(m^2)$ ones but is still k-free for some small enough k.

A standard example of a matrix whose SUM-rank is exponentially larger than the OR-rank is the complement $\overline{I}_m$ of the identity $m \times m$ matrix I_m. For A, we can take a random $m \times m$ matrix: this matrix is k-free for $k \preccurlyeq \log m$ and has $|A| \succcurlyeq m^2$ ones. So, fix such a matrix.

Theorem 5.1 (Find et al. [30])**.** Let $n = m^2$. For the $n \times n$ matrix $F = \overline{I}_m \otimes A$, we have

$$\mathsf{SUM}(F)/\mathsf{OR}_6(F) \succcurlyeq \frac{\sqrt{n}}{\log^2 n} \quad \text{and} \quad \mathsf{SUM}(F)/\mathsf{OR}_3(F) \succcurlyeq \frac{\sqrt{n}}{\log^3 n}\,.$$

Proof. By (1.7), we know that $\mathrm{rk}_\vee(\overline{I}_m) \preccurlyeq \log m$. Lemma 2.10 gives $\mathsf{OR}_6(\overline{I}_m \otimes A) \preccurlyeq m^2 = n$.

On the other hand, the matrix A is k-free for $k \preccurlyeq \log m$ and has $|A| \succcurlyeq m^2$ ones. Since the integer (and even real) rank $r = \mathrm{rk}_+(\overline{I}_m)$ of $\overline{I}_m$ is $\Omega(m)$, Theorem 3.20 implies that

$$\mathsf{SUM}(F) \geqslant \frac{r|A|}{k^2} \succcurlyeq \frac{m^3}{\log^2 m} \succcurlyeq \frac{n^{3/2}}{\log^2 n}\,,$$

and the desired gap follows. The gap in depth 3 also follows, since $\mathsf{OR}_3(F) \preccurlyeq n \log n$ by Lemma 2.10. □

Instead of using a random matrix A in Theorem 5.1, one can also take an explicit dense matrix without large rectangles, for example, the norm-matrix (see § 1.3). The resulting gap will be almost the same.

5.2 SUM/OR gap in depth two

In the proof of Theorem 5.1, it was important that OR circuits of depth $d > 2$ were allowed. We first show that particular matrices F analyzed in this theorem cannot exhibit any non-constant SUM/OR

gap in depth 2. We then will show that a particular submatrix of the intersection matrix exhibits a logarithmic gap.

Theorem 5.2. For F as in Theorem 5.1,

$$\mathsf{SUM}_2(F) \asymp \mathsf{OR}_2(F) \asymp \frac{n^{3/2}}{\log n}\,.$$

Proof. Since A has $|A| \succcurlyeq m^2$ ones, and is k-free for $k \preccurlyeq \log m$, Theorem 3.9 yields $\mathsf{OR}_2(A) \succcurlyeq m^2/\log m$. Since the term-rank of $\overline{I}_m$ is m, Theorem 3.19 yields

$$\mathsf{OR}_2(F) \geqslant \mathrm{tr}(\overline{I}_m) \cdot \mathsf{OR}_2(A) \succcurlyeq \frac{m^3}{\log m}\,.$$

On the other hand, since $\mathsf{SUM}_2(T_m) \leqslant m \log m + m$ holds for the full triangular $m \times m$ matrix T_m (see (1.6)), we have that $\mathsf{SUM}_2(\overline{I}_m) \preccurlyeq m \log m$. Since $\mathsf{SUM}_2(A) \preccurlyeq m^2/\log m$ (by Theorem 2.2), Lemma 2.7 implies that $\mathsf{SUM}_2(F) \preccurlyeq m^3$. To improve this upper bound by a logarithmic factor, we will use Lemma 2.9 together with the following fact.

Recall that the smaller weight s of a covering is the sum of the lengths of the shorter sides, and its larger weight ℓ is the sum of the length of the longer sides of its rectangles.

Lemma 5.3. For every $1 < k \leqslant n$, the full lower triangular matrix T_m admits a decomposition with parameters $s \preccurlyeq m \log m/\log k$ and $\ell \leqslant s \cdot k$.

By taking $k = \sqrt{m}$, we have that $\overline{I}_m$ admits a decomposition with parameters $s \preccurlyeq m$ and $\ell \preccurlyeq m^{3/2}$. Lemma 2.9 gives

$$\mathsf{SUM}_2(\overline{I}_m \otimes A) \preccurlyeq \frac{m^3}{\log m} + \frac{sm^2}{\log(m^2/\ell)} \preccurlyeq \frac{m^3}{\log m}\,.$$

So, it remains to prove Lemma 5.3.

To avoid floors and ceilings, let us assume that k divides m. Call an $a \times b$ rectangle *k-balanced* if $\max\{a,b\} \leqslant k \cdot \min\{a,b\}$. Let $f(m) = f_k(m)$ be the minimum possible sum of the shorter sides in a k-balanced decomposition of T_m. It is enough to show that $f(m) \leqslant 2m \log m/\log k$ for $m \geqslant k$. We do this by induction on m.

For the induction basis, assume that $m \leqslant k^2$. In this case we take m/k vertical "stripes" of width k each. The i-th of these stripes, for $i = 1, \ldots, m/k$, is an $(m - ik) \times k$ submatrix consisting of consecutive columns. Then cover the remaining 1s by stripes of width 1. The resulting covering is k-balanced, and we obtain $f(m) \leqslant k \cdot (m/k) + m = 2m \leqslant 2m \log m / \log k$.

For the induction step, take k vertical stripes of width m/k each. The covering is k-balanced, since the height of such a rectangle is at most $m - m/k < m = k \cdot (m/k)$. This gives the recurrence

$$f(m) \leqslant k(m/k) + k \cdot f(m/k)\,.$$

Using the induction hypothesis $f(p) \leqslant 2p \log p / \log k$, this yields

$$\begin{aligned} f(m) &\leqslant m + k \cdot 2(m/k) \log(m/k) / \log k \\ &= m + 2m \log m / \log k - 2m \leqslant 2m \log m / \log k\,. \end{aligned}$$

This concludes the proof of Lemma 5.3, and hence, the proof of the theorem. □

To exhibit a growing SUM/OR gap in depth 2, we use another matrix. Namely, let $M = M_n$ be an $n \times n$ matrix with $n = \binom{m}{2}$ whose rows and columns are labeled by 2-element subsets of $[m]$, and $M[a, b] = 1$ if and only if $a \cap b \neq \emptyset$. That is, M is a submatrix of the intersection $2^m \times 2^m$ matrix $\mathcal{D}$ formed by rows and columns whose label-sets have exactly two elements.

Theorem 5.4.

$$\mathsf{OR}_2(M_n) \leqslant 4n \quad \text{but} \quad \mathsf{SUM}_2(M_n) \succcurlyeq n \log n\,.$$

This theorem (as well as its proof) was inspired by a recent result of Pinto [80] concerning the "replication number" of coverings and partitions of the edges of graphs by bicliques; this is the maximal number of bicliques in the covering/partition sharing a vertex in common. He constructed an n-vertex graph which admits a covering with replication 2 such that every partition must have replication $\Omega(\log n)$.

Proof. The bound $\mathsf{OR}_2(M_n) \leqslant 4n$ is obvious, since the matrix has a trivial covering by m rectangles Q_x, where each Q_x consists of all rows

and columns whose labels contain element $x \in [m]$. The weight of each Q_x is $2m - 2$.

To prove the lower bound on $\mathsf{SUM}_2(M_n)$, fix an optimal decomposition $\mathcal{R}$ of $M = M_n$ into pairwise disjoint rectangles. Based on this decomposition, our strategy is to select submatrices of M such that: (i) the submatrices are row- and column-disjoint, that is, no two of them share a row or a column in common, and (ii) the weight of the induced decomposition of each submatrix is large.

When doing this, we concentrate on the "diagonal" entries of M. An entry is *diagonal* if its row and column are labeled by the same 2-element subset of $[m]$. To show (ii), we will use the Hansel–Krichevski argument for matrices with the zero diagonal (Theorem 3.5).

An $s \times t$ rectangle is *thin* if either $s = t = 2$ or $s = 1$ or $t = 1$.

Claim 5.5. Let R be a rectangle covering a given diagonal entry corresponding to a subset $\{x, y\}$. Then at least one of the sub-rectangles $R \cap Q_x$ and $R \cap Q_y$ is thin.

Proof. If at least one of the sub-rectangles $R \cap Q_x$ and $R \cap Q_y$ contains only one row or only one column, then we are done. So, assume that both sub-rectangles have at least two rows an at least two columns.

Since $R \cap Q_y$ must contain at least two rows and at least two columns, R must contain at least one row and at least one column without x. Since the labels must intersect, this implies that at most two rows and at most two columns of R can contain element x. Therefore, $R \cap Q_x$ (and, by symmetry, $R \cap Q_y$) is a 2×2 rectangle. □

Now fix in each 2-element subset $\{x, y\}$ of $[m]$ any one of its elements x or y for which the conclusion of Claim 5.5 holds, and call this element the *type* of the corresponding to $\{x, y\}$ diagonal entry of our matrix M. This is the only place where the disjointness of the rectangles in $\mathcal{R}$ is used: the type x of a diagonal entry alone ensures that the restriction $R \cap Q_x$ of the unique rectangle R covering this entry has small density.

Since we have $n = \binom{m}{2}$ diagonal entries and only m possible types, some set of $k := n/m = (m-1)/2$ diagonal entries must have the same type $x_1 \in [m]$. Consider the submatrix A_1 of M consisting of rows and columns covering exactly these entries. Note that A_1 is a

$k \times k$ all-1 matrix contained in the rectangle Q_{x_1}. Consider the induced decomposition of A_1 by sub-rectangles $R \cap Q_{x_1}$ with $R \in \mathcal{R}$. Let $w(A_1)$ be the weight of this decomposition, and D_1 be the set of diagonal entries of A_1.

By Claim 5.5, we know that every rectangle $R \cap Q_{x_1}$ intersecting D_1 is thin. Let A'_1 be the $k \times k$ matrix obtained from A_1 by switching to 0 all diagonal entries of A_1. Let also $w(A'_1)$ be the weight of the induced decomposition of A'_1, where we remove rectangles covering switched entries and cover remaining 1s optimally. It is easy to see that $w(A'_1) \leqslant w(A_1) + k$: only 2×2 rectangles covering only one entry of D_1 contribute $+1$ to $w(A'_1)$. On the other hand, since $\alpha(A'_1) = 1$, the Hansel–Krichevski argument (Theorem 3.5) implies that $w(A'_1) \geqslant \mathsf{OR}_2(A'_1) \geqslant k \log k$. Thus,

$$w(A_1) \geqslant w(A'_1) - k = k \log \frac{k}{2} = \frac{m-1}{2} \log \frac{m-1}{4} .$$

Next, remove from M all rows and columns containing the element x_1 among its labels, take an element $x_2 \in [m]$ such that at least $(m-2)/2$ of the remaining diagonal entries are of type x_2, build the next submatrix A_2, and argue in the same way to show that

$$w(A_2) \geqslant \frac{m-2}{2} \log \frac{m-2}{4} .$$

Proceeding in this way, we construct row- and column-disjoint submatrices $A_1, \ldots, A_{m-1}$ of M such that

$$\begin{aligned} \mathsf{SUM}_2(M) &\geqslant \sum_{i=1}^{m-1} w(A_i) \geqslant \sum_{i=1}^{m-1} \frac{i}{2} \log \frac{i}{4} \\ &\geqslant \frac{m(m-1)}{4} \log \frac{m}{8} \succcurlyeq n \log n , \end{aligned}$$

where the first inequality follows because the submatrices A_i do not share common rows or columns, the third inequality follows from (1.5), and the last inequality follows because $n = \binom{m}{2}$. □

Note that the lower bound also holds for the weight of the *coverings* of M under the restriction that no diagonal entry can be covered more than once.

5.3 OR/XOR gaps

Known bounds for the Sylvester $n \times n$ matrix $H = H_n$ (see § 4.4) give the gap

$$\mathsf{OR}_2(H)/\mathsf{XOR}_2(H) \asymp \frac{\sqrt{n}}{\log n}\,.$$

This remains the strongest known explicit OR/XOR gap in depth 2. (As before, by "explicit gaps" we mean gaps established on *explicit* matrices.) For larger depths, we have larger explicit gaps; we will mention them in the next section. Here we show that random submatrices of H exhibit almost maximal *non-explicit* gaps.

For this, we exploit some Ramseyean properties of submatrices of the Sylvester matrix. A boolean matrix A is *k-Ramsey matrix* if it contains no monochromatic $k \times k$ submatrices, that is, if both A and its complement $\overline{A}$ are k-free.

Let H_{n^2} be the $n^2 \times n^2$ Sylvester matrix with $n = 2^r$. Each subset $S \subseteq \{0,1\}^{2r}$ gives us an $|S| \times |S|$ submatrix H_S of H whose rows and columns correspond to vectors in S.

Lemma 5.6 (Pudlák and Rödl [88]). If $|S \cap V| < t$ holds for every subspace $V \subseteq \mathbb{F}_2^{2r}$ of dimension $\leqslant r$, then H_S is a t-Ramsey matrix.

Proof. Suppose that H_S contains a monochromatic $t \times t$ submatrix T. Our goal is to show that then there is a subspace $V \subseteq \mathbb{F}_2^{2r}$ of dimension r such that $|S \cap V| \geqslant t$. The submatrix T corresponds to two subsets of vectors $X, Y \subseteq S$ such that $\langle u, v\rangle = a$ for some $a \in \{0,1\}$ and all $u \in X$ and $v \in Y$. Viewing the vectors in X as the rows of the coefficient matrix and the vectors in Y as (columns of) unknowns, we obtain that the sum $\dim(X') + \dim(Y')$ of the dimensions of vector spaces X' and Y', spanned by X and by Y, cannot exceed $2r + a \leqslant 2r + 1$. Hence, at least one of these dimensions, say $\dim(X')$ must be $\leqslant r$. So, we can take $V = X'$, and obtain that $|S \cap V| \geqslant |X| = t$, as claimed. □

It is difficult to explicitly construct large sets S satisfying the condition of Lemma 5.6, but one can show their existence by a simple probabilistic argument, as one used by Jukna in [47].

Lemma 5.7. For almost all subsets $S \subseteq \{0,1\}^{2r}$ of size $|S| = 2^r$, the submatrix H_S is a $2r$-Ramsey matrix.

Proof. We will use the following versions of Chernoff's inequality (see, e.g., [69], § 4.1): if X is the sum of n independent Bernoulli random variables with the success probability p, and $\mu = \mathrm{E}[X]$, then

$$\Pr[X \leqslant a] \leqslant \mathrm{e}^{-(\mu-a)^2/2n} \tag{5.1}$$

for $0 \leqslant a < \mu$, and $\Pr[X \geqslant c\mu] \leqslant 2^{-c\mu}$ for $c > 2\mathrm{e}$.

Now pick a subset $S \subseteq \mathbb{F}_2^{2r}$ at random, by including each vector in S independently with probability $p = 2^{1-r} = 2/n$. By Chernoff's inequality, $|S| \geqslant pn/2 = 2^r$ holds with probability at least $1 - o(1)$.

Let $V \subseteq \mathbb{F}_2^{2r}$ be a fixed subspace of dimension r. Then $|V| = 2^r$, hence, $cp|V| = 2c$. By Chernoff's inequality, $\Pr[|S \cap V| \geqslant 2c] \leqslant 2^{-2c}$ holds for any $c > 2e$. The number of vector spaces in $\mathbb{F}_2^{2r}$ of dimension r does not exceed $\binom{2r}{r} \leqslant 2^{2r}/\sqrt{2r}$. We can therefore take $c = r$ and conclude that the set S intersects some r-dimensional vector space V in $2c = 2r$ or more elements with probability at most $\binom{2r}{r}2^{-2c} \leqslant 1/\sqrt{2r} = o(1)$. Hence, with probability $1 - o(1)$, both $|S| \geqslant pn/2 = 2^r = n$ and $|S \cap V| < 2r$ hold for every r-dimensional subspace V. Lemma 5.6 implies that almost all $n \times n$ submatrices H_S of H_{n^2} are t-Ramsey for $t \leqslant 2r = 2\log n$.  □

Theorem 5.8. For almost all $n \times n$ submatrices A of H_{n^2}, we have that $\mathsf{XOR}_2(A) \preccurlyeq n \log n$ and $\mathsf{XOR}_3(A) \preccurlyeq n$, but

$$\mathsf{OR}_2(A) \succcurlyeq \frac{n^2}{\log n} \quad \text{and} \quad \mathsf{OR}(A) \succcurlyeq \frac{n^2}{\log^2 n}\,.$$

Thus,

$$\mathsf{OR}_2(A)/\mathsf{XOR}_2(A) \succcurlyeq \frac{n}{\log^2 n} \quad \text{and} \quad \mathsf{OR}(A)/\mathsf{XOR}_3(A) \succcurlyeq \frac{n}{\log^2 n}\,.$$

Proof. Every $n \times n$ submatrix A of H_{n^2} can be written as the product $A = P \cdot Q^\top$ over $\mathbb{F}_2$ of two $n \times 2r$ matrices P and Q. Hence, $\mathsf{XOR}_2(A) \leqslant n \log n$, and Lemma 2.5 gives $\mathsf{XOR}_3(A) \preccurlyeq n$.

To show the lower bounds, take any of the submatrices H_S guaranteed by Lemma 5.7. Let A be H_S or its complement (depending on

which of the matrices has more 1s). Thus, A has $|A| \succcurlyeq n^2$ ones, and is k-free for $k = 2\log n$. Theorem 3.9 implies that $\mathsf{OR}_2(A) \geqslant |A|/k \succcurlyeq n^2/\log n$ and $\mathsf{OR}(A) \geqslant |A|/k^2 \succcurlyeq n^2/\log^2 n$, as desired. □

The same OR/XOR-gap was recently established by Find et al. [30]. They showed that the gap is achieved on matrices of the form $A = P^\top BP$, where B is a random $m \times m$ matrix with $m \asymp \sqrt{n}$, and P is an $m \times m$ matrix whose each k-columns are linearly independent, for an appropriate value of k. Their proof uses the concept of "local independence" in random matrices, as well as some non-trivial results in the spirit of Theorem 6.9 below.

Yet another proof of the same gap was recently given by Boyar and Find [11]. They show this for a matrix A of the form $A = P \cdot Q^\top$, where P and Q are random $n \times 24\log n$ matrices. Using a result of Chor and Goldreich [20] about the randomized communication complexity of the inner product function, they show that such a matrix is $(3\log n)$-free.

5.4 Explicit gaps

As observed already by Mitiagin and Sadovskiy [66], circulant matrices are easy for XOR circuits. To see why this is true, consider a product of two binary polynomials of degree $< n$ modulo $(x^n + 1)$:

$$\sum_{i=0}^{n-1} c_i x^i = \left(\sum_{i=0}^{n-1} a_i x^i\right)\left(\sum_{i=0}^{n-1} b_i x^i\right) \mod (x^n + 1)\,.$$

It is easy to see that coefficients c_k satisfy

$$c_k = \bigoplus_{i+j \,\equiv\, k \bmod n} a_i b_j\,. \tag{5.2}$$

Vector $(c_0, \ldots, c_{n-1})$ is called a *cyclic convolution of order* n of vectors $(a_0, \ldots, a_{n-1})$ and $(b_0, \ldots, b_{n-1})$.

Eq. (5.2) implies that the operator of cyclic convolution of constant vector a and a vector x of variables is a linear operator with circulant matrix having first column a. On the other hand, any circulant matrix is a matrix of a cyclic convolution operator.

Small bilinear (XOR, AND)-circuits computing cyclic convolution are known. Bilinear means that the circuit is linear in any of the two

input vectors. Thus, if we set one vector to be a constant vector, then an (XOR, AND)-circuit reduces to an XOR circuit. So, we can reformulate the best known complexity result for convolution as follows.

Theorem 5.9 (Schönhage [97]). For any circulant $n \times n$ matrix A:

$$\mathsf{XOR}(A) \preccurlyeq n \ln n \ln\ln n\,.$$

In fact, there is also a bounded-depth analogue:

Theorem 5.10 (Sergeev [99]). For every integer constant $t \geqslant 1$ there is a constant $c = c(t)$ such that, for every circulant $n \times n$ matrix A:

(i) $\mathsf{XOR}_{2t-1}(A) \leqslant cn^{1+1/t}$,
(ii) $\mathsf{XOR}_{2t}(A) \leqslant cn^{1+1/t}/\log^{1/t} n$.

Some explicit circulant matrices that are hard for OR circuits are long known. One example is the point-line incidence matrix P of a finite projective plane which is 2-free matrix with $|P| > n^{3/2}$ ones. By Theorem 3.9, we have $\mathsf{OR}(P) = \mathsf{OR}_1(P) > n^{3/2}$, and hence,

$$\mathsf{OR}(P)/\mathsf{XOR}(P) \succcurlyeq \frac{\sqrt{n}}{\ln n \ln\ln n}\,. \tag{5.3}$$

In the case of bounded-depth circuits we have

$$\mathsf{OR}(P)/\mathsf{XOR}_4(P) \succcurlyeq \sqrt{\log n} \quad \text{and} \quad \mathsf{OR}(P)/\mathsf{XOR}_{2t-1}(P) \succcurlyeq n^{1/2-1/t}$$

for any $t \geqslant 3$. The bound (5.3) was proved by Gashkov and Sergeev [34] and is the strongest known gap for 2-free matrices (in either explicit and non-explicit cases). An obvious upper bound is $O(\sqrt{n})$. To the best of our knowledge, $d = 4$ is the minimal depth known to separate XOR- and OR-complexities of 2-free matrices.

As observed in [34], under appropriate choice of parameters, the $n \times n$ norm-matrix N constructed in § 1.3 is Δ-free and has about n^2/Δ ones, where

$$\Delta = 2^{-\Theta(\sqrt{\ln n \ln\ln n})}.$$

Thus, Theorem 3.9 implies that $\mathsf{OR}(N) \succcurlyeq n^2/\Delta$. Using fast multi-dimensional convolution, the gap

$$\mathsf{OR}(N)/\mathsf{XOR}(N) \succcurlyeq \frac{n}{\Delta} \tag{5.4}$$

was shown by Gashkov and Sergeev in [34]. Alternatively, as they explained in [35], the norm-matrix can be converted into a circulant matrix which is also Δ-free and has about n^2/Δ ones. Then (5.4) follows from Schönhage's theorem.

Concerning *bounded-depth* circuits, the circulant analogue of the norm-matrix N provides separation

$$\mathsf{OR}(N)/\mathsf{XOR}_{2t-1}(N) \succcurlyeq \frac{n^{1-1/t}}{\Delta}$$

for any $t \geqslant 1$.

For completeness, we mention that, as shown by Grinchuk and Sergeev in [43], a certain randomized construction of an $\times n$ matrix Z is with high probability $O(\log n)$-free and has $|Z| \succcurlyeq n^2/\log^3 n$ ones. Thus, for such a matrix we have the gaps

$$\mathsf{OR}(Z)/\mathsf{XOR}(Z) \succcurlyeq \frac{n}{\ln^6 n \ln\ln n} \quad \text{and} \quad \mathsf{OR}(Z)/\mathsf{XOR}_{2t-1}(Z) \succcurlyeq \frac{n^{1-1/t}}{\ln^5 n}$$

for any $t \geq 1$. It is a slight improvement over the result due to Grinchuk [42] who showed the existence of $\log^2 n$-free circulant matrices with about $n^2/\log^8 n$ ones.

It is worth to mention that OR/XOR gaps may be much larger than SUM/OR gaps. Namely, Theorem 3.9 implies that $\mathsf{SUM}(A)/\mathsf{OR}(A) \leqslant k^2$ holds for *every* k-free matrix. On the other hand, as we mentioned above, there are explicit k-free matrices A such that the gap $\mathsf{OR}(A)/\mathsf{XOR}(A)$ is at least $n^{1/2-\epsilon}$, for constant k, and even $n^{1-o(1)}$ for $k \preccurlyeq \log n$.

5.5 XOR/OR gap in depth two

Unlike for SUM/OR and OR/XOR gaps, no non-trivial XOR/OR gaps were known. In this section we present the first such gap. Our starting object is the k-intersection matrix $\mathcal{D}_{n,k}$ which is defined as follows. Fix an n-element set X, and label the rows and columns by distinct non-empty subsets $S \subseteq X$ of size $|S| \leqslant k$. Thus, the matrix has

$$N(n,k) := \sum_{i=1}^{k} \binom{n}{i}$$

rows and $N(n,k)$ columns. The (S,S')-th entry is 1 if and only if S and S' have a non-empty intersection.

To show a growing gap, we use Kronecker product matrices of the form $A = Đ_{n,k} \otimes Đ_{m,p}$. That is, we now have two fixed disjoint sets X and Y of sizes $|X| = n$ and $|Y| = m$. Rows and columns of A are labeled by pairs (S,T) of subsets $S \subset X$ and $T \subset Y$ of sizes $|S| \leqslant k$ and $|T| \leqslant p$, and

$$A[(S,T),(S',T')] = 1 \quad \text{iff} \quad S \cap S' \neq \emptyset \text{ and } T \cap T' \neq \emptyset.$$

Important property of such matrices is that they have full rank over $\mathbb{F}_2$. The following lemma is an extension of a well-known fact that $Đ_{n,k}$ itself has full rank (see e.g. [50, Lemma 4.11]).

Lemma 5.11. The matrix $A = Đ_{n,k} \otimes Đ_{m,p}$ has full rank over $\mathbb{F}_2$.

Proof. As in [50, Lemma 4.11], we follow Razborov's short argument for $Đ_{n,k}$ given in [93]. The matrix A has NM rows and as many columns, where $N = N(n,k)$ and $M = N(m,p)$. We have to show that the rows of A are linearly independent over $\mathbb{F}_2$, i.e., that for any nonzero vector

$$\lambda = (\lambda_{I_1,J_1}, \lambda_{I_1,J_2}, \ldots, \lambda_{I_N,J_M})$$

in $\mathbb{F}_2^{NM}$ indexed by pairs of subsets $I \subset [n]$ and $J \subset [m]$ of sizes $1 \leqslant |I| \leqslant k$ and $1 \leqslant |J| \leqslant p$, we have $\lambda^\top A \neq 0$; the sets I and J are sets of indices of the corresponding subsets $S \subset X$ and $T \subset Y$. For this, consider the following boolean function

$$f(x_1,\ldots,x_n,y_1,\ldots,y_m) := \bigoplus_{\substack{1\leq|I|\leqslant k\\ 1\leq|J|\leqslant p}} \lambda_{I,J}\Big(\bigvee_{i\in I} x_i\Big)\cdot\Big(\bigvee_{j\in J} y_j\Big)$$

on $n+m$ boolean variables corresponding to the elements of X and Y. Since $\lambda \neq 0$, at least one of the coefficients $\lambda_{I,J}$ is nonzero, and we can find some pair (I_0,J_0) such that $\lambda_{I_0,J_0} \neq 0$ and $\lambda_{I,J} = 0$ for all other pairs (I,J) satisfying $I \supset I_0$ and $J \supset J_0$. So, the pair (I_0,J_0) is maximal. Assume w.l.o.g. that $I_0 = \{1,\ldots,t\}$ and $J_0 = \{1,\ldots,u\}$, and make in the function f the substitution $x_i = 0$ for all $i \notin I_0$ and $y_j = 0$ for all $j \notin J_0$.

After this substitution has been made, a *nonzero polynomial* over the variables $x_1, \dots, x_t, y_1, \dots, y_u$ remains with the leading-degree term $x_1 \cdots x_t \cdot y_1 \cdots y_u$; here we use the maximality of (I_0, J_0) and the fact that the polynomial for the OR function has maximal degree. Hence, after the substitution we obtain a polynomial which is 1 for some assignment $(a_1, \dots, a_t, b_1, \dots, b_u)$ to its variables. But this means that the polynomial f itself takes the value 1 on the assignment

$$c = (a_1, \dots, a_t, 0, \dots, 0, b_1, \dots, b_u, 0, \dots, 0)\,.$$

Hence,

$$1 = f(c) = \bigoplus_{I,J} \lambda_{I,J} \Big(\bigvee_{i \in I} c_i \Big) \cdot \Big(\bigvee_{j \in J} c_{j+n} \Big)$$

Let $I' := \{i \colon a_i = 1\}$ and $J' := \{j \colon b_j = 1\}$. Then $|I'| \leqslant k$ and $|J'| \leqslant p$. Moreover, $\big(\bigvee_{i \in I} c_i\big) \cdot \big(\bigvee_{j \in J} c_{j+n}\big) = 1$ if and only if $I \cap I' \neq \emptyset$ and $J \cap J' \neq \emptyset$, which is equivalent to $A[(I, J), (I', J')] = 1$. Thus,

$$\bigoplus_{I,J} \lambda_{I,J} A[(I, J), (I', J')] = 1\,,$$

meaning that the (I', J')-th coordinate of the vector $\lambda^\top A$ is nonzero. □

Next, consider the $n \times n$ matrix $K_n = Đ_{k^2,k} \otimes Đ_{p^2,p}$ with

$$k^2 \asymp \frac{1}{p} N(p^2, p) \quad \text{and} \quad n = N(p^2, p) \cdot N(k^2, k)\,. \tag{5.5}$$

Hence,

$$k \asymp \frac{\ln n}{\ln \ln n} \quad \text{and} \quad p \asymp \frac{\ln \ln n}{\ln \ln \ln n}\,.$$

That is, we have two fixed disjoint sets X and Y of sizes $|X| = k^2$ and $|Y| = p^2$. Rows and columns of the matrix $K = K_n$ are labeled by pairs (S, T) of non-empty subsets $S \subset X$ and $T \subset Y$ of sizes $|S| \leqslant k$ and $|T| \leqslant p$, and

$$K[(S, T), (S', T')] = 1 \quad \text{iff} \quad S \cap S' \neq \emptyset \text{ and } T \cap T' \neq \emptyset.$$

Theorem 5.12.

$$\mathsf{OR}_2(K_n) \preccurlyeq n \cdot \frac{\ln n}{\ln \ln n} \quad \text{but} \quad \mathsf{XOR}_2(K_n) \succcurlyeq n \cdot \frac{\ln n \cdot \ln \ln \ln n}{\ln \ln n}.$$

Proof. The upper bound for OR_2 complexity is straightforward. It is derived from the representation of the matrix K_n as a product $B \cdot B^\top$ (over the boolean semiring), where B is an $n \times (|X|+|Y|)$ matrix defined by

$$B[(S,T),q] = 1 \text{ iff } q \in S \cup T$$

(cf. representation of D_n in § 1.3). Since $|S| \leqslant k$ and $|T| \leqslant p$, each row of B contains at most $k+p$ 1s, hence, the weight of B does not exceed

$$(k+p)n \asymp n \cdot \frac{\ln n}{\ln \ln n}.$$

To prove the desired lower bound on the XOR_2 complexity of K_n, we exploit the rigidity argument, as given by Pudlák's theorem (Theorem 3.34). Recall that according to this theorem, $\mathsf{XOR}_2(A) \geqslant 2f(n)\ln(b/a)$ holds as long as $R_A(r) \geqslant f(n)^2/r$ holds for *all* integers r between a and b.

First, we will prove that there exist constants $c_1, c_2 > 0$ such that

$$R_{K_n}(c_1 r) \geqslant \frac{c_2 n^2}{r} \tag{5.6}$$

holds for a sequence of integers r of the form

$$r = r(t,u) := N(k^2,t) \cdot N(p^2,u)$$

with $t \in [k]$ and $u \in [p]$. However, to fulfil conditions of Theorem 3.34, we need bounds not just for "rank thresholds" of the form $r = c_1 r(t,u)$ but also for all other intermediate[1] r's between $c_1 r(1,1)$ and $c_1 r(k,p)$. For this, we use simple estimates

$$\frac{r(t,u+1)}{r(t,u)} = \frac{N(p^2,u+1)}{N(p^2,u)} \leqslant 1 + \frac{\binom{p^2}{u+1}}{\binom{p^2}{u}} \leqslant p^2$$

and, by (5.5),

$$\frac{r(t+1,1)}{r(t,p)} = \frac{N(k^2,t+1) \cdot N(p^2,1)}{N(k^2,t) \cdot N(p^2,p)} \leqslant \frac{k^2 p^2}{N(p^2,p)} \asymp p\,.$$

[1] To "fill in" these intervals was the only reason to consider a more complicated case of Kronecker products instead of intersection matrices themselves.

Now, these estimates ensure that, for every integer r between $c_1 r(1,1)$ and $c_1 r(k,p)$, there must be t and u such that $r \leqslant c_1 r(t,u) \leqslant p^2 r$. Therefore,

$$R_{K_n}(r) \geqslant R_{K_n}(c_1 r(t,u)) \geqslant \frac{c_2 n^2}{r(t,u)} \geqslant \frac{c_2 (n/p)^2}{r}\,.$$

When applied with $f(n) = \sqrt{c_2}\, n/p$, $a = c_1 r(1,1)$ and $b = c_1 r(k,p)$, Theorem 3.34 yields

$$\mathsf{XOR}_2(K_n) \succcurlyeq (n/p)\ln n \asymp n \cdot \frac{\ln n \cdot \ln\ln\ln n}{\ln\ln n}\,.$$

So, it remains to prove (5.6). In what follows, we will use inequalities

$$\binom{k^2}{t} \leqslant N(k^2,t) \leqslant \left(1 + t \cdot \frac{t}{k^2 - t}\right)\binom{k^2}{t} \leqslant 6 \cdot \binom{k^2}{t}$$

holding for all $t \leqslant 2k$ and $k \geqslant 10$, as well as

$$\binom{k^2}{t} \geqslant \binom{k^2 - 2k}{t} \geqslant e^{-32}\binom{k^2}{t}$$

holding for all $t \leqslant 2k$ and $k \geqslant 4$. Hence,

$$N(k^2,t) \geqslant N(k^2 - 2k, t) \geqslant e^{-32} N(k^2,t)\,.$$

Fix some $t \in [k]$, $u \in [p]$, and mark a minimal set of entries in K_n one has to change in order to reduce the rank until $c_1 r(t,u)$.

We define a (t,u)-*submatrix* of K_n by fixing a pair I, I' of disjoint $(k-t)$-element subsets of X, a pair J, J' of disjoint $(p-u)$-element subsets of Y. The submatrix then consists of all rows (S,T) and all columns (S',T') such that $I \subset S$, $I' \subset S'$, $J \subset T$, and $J' \subset T'$ (proper inclusions). A *restriction* of such a (t,u)-submatrix A is the submatrix B of A whose labels satisfy the following additional disjointness condition $I \cap S' = I' \cap S = J \cap T' = J' \cap T = \emptyset$.

By the construction, the submatrix B (up to a permutation of rows and columns) is the intersection matrix

$$Đ_{k^2-2(k-t),t} \otimes Đ_{p^2-2(p-u),u}\,.$$

This is because $B[(S,S'),(T,T')] = 1$ if and only if

$$(S \setminus I) \cap (S' \setminus I') \neq \emptyset \text{ and } (T \setminus J) \cap (T' \setminus J') \neq \emptyset.$$

By Lemma 5.11, each such submatrix has full rank

$$\operatorname{rk}(B) = N(k^2 - 2(k-t), t) \cdot N(p^2 - 2(p-u), u)\,.$$

Thus, $\operatorname{rk}(B) \geqslant c_3 r(t,u)$ for a constant $c_3 > 0$. Now set $c_1 := c_3/2$. Since the rank of B cannot be halved by flipping fewer than $\operatorname{rk}(B)/2$ entries of B, the matrix B (and hence the (t,u)-submatrix A) must have at least $c_1 r(t,u)$ marked entries.

Now, we will bound the total number of marked entries in K_n. We have at least $c_1 r(t,u)$ such entries in each (t,u)-submatrix A. On the other hand, each single entry of K_n can belong to at most $\binom{k}{t}^2 \cdot \binom{p}{u}^2$ such submatrices. To see this, fix an arbitrary entry $[(S,T),(S',T')]$ of K_n; hence, $S, S' \subseteq X$, $T, T' \subseteq Y$, $|S|, |S'| \leqslant k$, and $|T|, |T'| \leqslant p$. If a (t,u)-submatrix A is defined by pairs (I,J) and (I',J'), then this entry can belong to A only if $I \subset S$, $I' \subset S'$, $J \subset T$, and $J' \subset T'$. Since $|I| = |I'| = k - t$ and $|J| = |J'| = p - u$, we have at most $\binom{|S|}{k-t} \leqslant \binom{k}{t}$ choices for I (and for I'), and at most $\binom{|T|}{p-u} \leqslant \binom{p}{u}$ choices for J (and for J').

Since there are $\binom{k^2}{k-t} \cdot \binom{k^2-(k-t)}{k-t}$ disjoint pairs (I,J) and $\binom{p^2}{p-u} \cdot \binom{p^2-(p-u)}{p-u}$ disjoint pairs (I',J'), the multiplied by $r(t,u)$ total number of marked entries is bounded from below as

$$\begin{aligned}
&c_1 r^2(t,u) \frac{\binom{k^2}{k-t}\binom{k^2-(k-t)}{k-t}\binom{p^2}{p-u}\binom{p^2-(p-u)}{p-u}}{\binom{k}{t}^2\binom{p}{u}^2} \\
&\geqslant c_4 \left[\frac{\binom{k^2}{t}\binom{k^2}{k-t}\binom{p^2}{u}\binom{p^2}{p-u}}{\binom{k}{t}\binom{p}{u}}\right]^2 \\
&\geqslant c_4 \left[\frac{(k^2-k)^k (p^2-p)^p}{k!\,p!}\right]^2 \\
&\geqslant c_5 n^2\,,
\end{aligned}$$

where c_4 and c_5 are some positive constants. Therefore, the desired inequality $R_{K_n}(c_1 r) \geqslant c_2 n^2/r$ holds with $c_2 := c_5$ for all $r = r(t,u)$. This completes the proof of (5.6), and thus, the proof of the theorem. □

5.6 Gaps for matrices and their complements

If a matrix A has a small circuit, can then its complement $\overline{A}$ require large circuits? In the case of XOR circuits, the gap $\mathsf{XOR}(\overline{A})/\mathsf{XOR}(A)$ cannot be $\omega(1)$. This is because we always have that $\mathsf{XOR}(\overline{A}) \leqslant \mathsf{XOR}(A) + 2n$: given a circuit for A, just add one new node connected with all inputs and all outputs to get a circuit for $\overline{A}$. In the case of OR circuits, however, the situation is completely different: here the gaps $\mathsf{OR}(\overline{A})/\mathsf{OR}(A)$ may be large.

Let $D = D_n$ be the $n \times n$ Kneser–Sierpinski matrix. Since D is the complement of the intersection matrix, (4.2) and Lemma 4.2 yields the following trade-offs between the complexities of D and $\overline{D}$ (see also Section 4.3):

$$\mathsf{OR}(\overline{D}) = n \text{ but } \mathsf{OR}(D) \sim \tfrac{1}{2} n \log n,$$

and

$$\mathsf{OR}_2(\overline{D}) = n \log n \text{ but } \mathsf{OR}_2(D) \succcurlyeq n^{1+c} \text{ for } c > 0.16.$$

This yields the gaps

$$\mathsf{OR}(\overline{D})/\mathsf{OR}(D) = \tfrac{1}{2} \log n \quad \text{and} \quad \mathsf{OR}_2(\overline{D})/\mathsf{OR}_2(D) \succcurlyeq n^{0.16}.$$

In fact, submatrices of D give even larger gaps. Say, for $\alpha = 0.27$, the submatrix D^α of D (as defined in the proof of Lemma 4.2) gives the depth-2 gap of about $n^{0.23}$.

We are now going to show that there exist matrices A almost achieving the maximal possible gap of $n/\ln n$.

To show this, we will use probabilistic arguments. In particular, we will use following simple consequence of the Chebyshev inequality for sums of weakly dependent random 0/1 variables.

Recall that the Chebyshev inequality states that, for every random variable X and every real number $a > 0$, $\Pr[|X - \mathrm{E}[X]| \geqslant a]$ is at most $\mathrm{Var}[X]/a^2$. Now let $X = \sum_{i=1}^n x_i$ be the sum of random 0/1 variables, and $\mu = \mathrm{E}[X]$. For an index i, let $J(i)$ be the set of indices j such that x_i and x_j are *not* independent. If $|J(i)| \leqslant K$ holds for all i, then for every real number $\alpha > 0$,

$$\Pr[|X - \mu| \geqslant \alpha\mu] \leqslant \frac{K}{\alpha^2 \mu}. \tag{5.7}$$

Indeed, one can easily show that, the variance of X cannot exceed μK: by the linearity of expectation, we have

$$\operatorname{Var}[X] = \mathrm{E}\left[X^2\right] - \mu^2 = \sum_{i,j} \mathrm{E}[x_i x_j] - \mu^2 .$$

Using the inequality $\mathrm{E}[x_i x_j] \leqslant \mathrm{E}[x_i]$ holding for all pairs of 0/1 variables, and $\mathrm{E}[x_i x_j] = \mathrm{E}[x_i]\,\mathrm{E}[x_j]$ holding for independent pairs x_i and x_j, we obtain

$$\sum_{i,j} \mathrm{E}[x_i x_j] \leqslant \sum_{i} \mathrm{E}[x_i] \left(K + \sum_{j \notin J(i)} \mathrm{E}[x_j]\right) \leqslant \mu K + \mu^2 .$$

Thus, $\operatorname{Var}[X] \leqslant \mu K$, and (5.7) follows from the Chebyshev inequality.

Theorem 5.13. For n sufficiently large, there exist $n \times n$ matrices A (not necessarily the same for all three items) such that:

(i) $\overline{A}$ is 2-free and has $\Omega(n^{5/4})$ ones, but $\mathsf{OR}_2(A) = O(n \ln^2 n)$.
(ii) $\overline{A}$ is 2-free and has $\Omega(n^{1.1})$ ones, but $\mathrm{rk}_\vee(A) = O(\ln n)$.
(iii) $\overline{A}$ is $(\ln n)$-free and has $\Theta(n^2)$ ones, but $\mathsf{OR}_3(A) = O(n \ln n)$.

Moreover, the matrix from item (iii) also has $\mathsf{OR}_2(A) \preccurlyeq n \ln^2 n$. The second claim (ii) was earlier proved by Katz [53] (using slightly different arguments) and inspired the extension (i) and (iii) above.

Proof. Consider a random depth-2 OR circuit F with n inputs, n outputs, and $t = (3/4)\ln^3 n$ nodes on the middle level; t is assumed to be integer for simplicity. Connect each input and each output node with each node on the middle level independently with probability $p = 1/\ln n$. Let B be the random $n \times n$ matrix implemented by the entire circuit F. Let X be the number of 0s, Y the number 0-squares (that is, 2×2 all-0 submatrices) in B, and Z the number of edges in the circuit.

Since $\mathrm{E}[Z] = 2ptn = (3/2)n \ln^2 n$, Markov's inequality $\Pr[X > a] < 1/a$ implies that

$$\Pr[Z > 6n \ln^2 n] < 1/4 . \tag{5.8}$$

A given length-2 path is present in F with probability p^2. So, the probability that a given entry of B is 0 is the probability $(1-p^2)^t$

that none of the t length-2 paths between the corresponding input and output nodes is present. Since $p^2t = (3/4)\ln n$ and $(1-x) > \mathrm{e}^{-x-x^2}$ holds for $0 < x < 1/2$, this yields

$$\mathrm{E}[X] = n^2(1-p^2)^t \geqslant n^2\mathrm{e}^{-p^2t-p^4t} \geqslant an^{5/4}\,,$$

for a constant $a \geqslant 1/\mathrm{e}$. The number X is the sum of n^2 not necessarily independent but identically distributed boolean variables (entries in our matrix B). Two entries can be dependent only if they lie in the same row or in the same column. Applying (5.7) with $K = 2n$ and $\alpha = n^{-1/8}$, we obtain that $\Pr[|X - an^{5/4}| > an^{9/8}] = O(n^{-1/2})$. Thus,

$$\Pr[X < (a/2)n^{5/4}] < 1/4\,. \tag{5.9}$$

To upper bound $\mathrm{E}[Y]$, fix a 2-element subset I of input nodes, and a 2-element subset J of output nodes. For each middle node v, the probability that I and J are connected by a length-2 path going through v is the probability $1-(1-p)^2 = p(2-p)$ that v is connected to I times the same probability that v is connected to J. Thus, the probability that the corresponding to I and J submatrix of B has no 1s is

$$(1-p^2(2-p)^2)^t \leqslant \mathrm{e}^{-p^2(2-p)^2t} \leqslant \mathrm{e}^{-4p^2t+4p^3t} < (3/n)^3\,,$$

because $p^2t = (3/4)\ln n$ and $p = 1/\ln n$. Since there are only $\binom{n}{2}^2 < n^4$ squares in B, this gives $\mathrm{E}[Y] \leqslant bn$ for a constant $b \leqslant 27$. Markov's inequality yields

$$\Pr[Y > 4bn] < 1/4\,. \tag{5.10}$$

Now fix a circuit avoiding all three events (5.8)-(5.10). The circuit has $Z \leqslant 6n\ln^2 n$ edges, while the implemented by it matrix B has $X \geqslant (a/2)n^{5/4}$ zeroes and only $Y \leqslant 4bn$ zero-squares. For each of these squares, pick one its entry, add a node on the middle level and join the corresponding input and output nodes by a length-2 going through this node. The resulting circuit has $Z + 2Y = O(n\ln^2 n)$ edges, while the complement $\overline{A}$ of the matrix A implemented by this circuit is 2-free and still has at least $X - Y = \Omega(n^{5/4})$ ones. This completes the proof of item (i).

To prove item (ii), we argue similarly. In this case, set p to be a very small positive constant, say $p \leqslant 1/37$, and set $t = \frac{3.5}{p^2(2-p)^2} \ln n$. Now we expect

$$\mathrm{E}[X] = n^2(1-p^2)^t = n^{2-3.5\ln(1-p^2)/p^2(2-p)^2} \geqslant n^{1.1}$$

zeros in B. On the other hand, we expect only

$$\mathrm{E}[Y] \leqslant n^4(1-p^2(2-p)^2)^t \leqslant n^4 \mathrm{e}^{-p^2(2-p)^2 t} = \sqrt{n}$$

zero-squares in B. We can kill all these squares by picking one entry in each of them, and joining the corresponding input and output node by depth-2 paths going through just *one* new node on the middle level. This reduces the total number of zeros in B by at most $Y^2 \leqslant n$, and the boolean rank of the resulting matrix is $t+1 = O(\ln n)$, as desired.

To prove item (iii), let k be an integer nearest to $\ln n$. Consider a random depth-3 OR circuit F with n inputs, n outputs and $t = 48k^2$ nodes on each of the two intermediate levels. As before, we connect each input and each output node with every node on the neighboring level independently with probability $p = 1/k$. We also connect nodes of the two intermediate levels independently with probability $q = 1/k^2$. Hence, the middle layer of the circuit implements a random (though non-uniform) boolean $t \times t$ matrix U. Let B be the random $n \times n$ matrix implemented by the entire circuit F. As before, let X be the number of 0s, Y the number of $k \times k$ all-0 submatrices in B, and Z the number of edges in the circuit.

Since $\mathrm{E}[Z] = 2ptn + qt^2 = O(kn) = O(n \ln n)$, Markov's inequality, $\Pr[X > a] < 1/a$ implies that $Z = O(n \ln n)$ holds with an arbitrarily large constant probability. Our goal is to show that also the event "$Y = O(1)$" holds with an arbitrarily large constant probability, and that the event "$X = \Omega(n^2)$" holds with some constant probability $c > 0$. This will imply that a circuit F achieving these values of X, Y and Z exist.

By adding a constant number of edges to the circuit F we can then switch one entry in each $k \times k$ all-0 submatrix of the matrix A implemented by F to make the complement $\overline{A}$ of the resulting matrix A k-free. Since the number of edges in the resulting circuit for A remains $O(n \ln n)$, we will be done. Moreover the construction also implies that

$\mathrm{rk}_\vee(A) \leqslant t + O(1) = O(\ln^2 n)$, which gives $\mathsf{OR}_2(A) = O(n \ln^2 n)$ in the depth-2 case.

Let us now show that $X = \Omega(n^2)$ holds with a constant probability $c > 0$. The event that F has no paths connecting a given input node and a given output node is that the rows and columns of U, corresponding to the neighbors of these two nodes in F, form an all-0 submatrix of U. Clearly, an input (or output) of the circuit F has degree $\leqslant 2pt$ with probability at least $1/2$. So, the probability that a given entry of the matrix B is zero is at least the probability $1/4$ that the input and the output both have such small degrees times the probability $(1-q)^{4p^2t^2}$ that the given $2pt \times 2pt$ submatrix of the matrix U is an all-0 submatrix. Since $qp^2t^2 = \Omega(1)$, this probability is also $\Omega(1)$. Thus, there is a constant $c > 0$ such that $\mathrm{E}[X] \geqslant 2cn^2$. Since $X \leqslant n^2$, this implies $\Pr[X \geqslant cn^2] \geqslant c$, for otherwise we would have that

$$\mathrm{E}[X] < c \cdot n^2 + 1 \cdot cn^2 = 2cn^2 .$$

It remains to show that $\mathrm{E}[Y] = O(1)$, i.e., that we can expect at most a constant number of zero $k \times k$ submatrices in B. To show this, set $r = 10k^2$. Consider a fixed pair I, J of subsets of input and output node of F, each of size k, and let $B_{I,J}$ be the corresponding submatrix of B. Let P_I (and P_J) be the probability that the nodes in I (in J) have fever than r neighbors, and let P_U be the probability that the matrix U contains some $r \times r$ all-0 submatrix.

If both I and J have at least r neighbors, and if U *does not* contain any $r \times r$ all-0 submatrix, then there will be a path from I to J in the circuit. Thus, $\Pr[B_{I,J}$ is an all-0 submatrix$] \leqslant P_U + P_I + P_J$. Since U has only $\binom{t}{r}^2$ $r \times r$ submatrices, each of which is an all-0 submatrix with probability $(1-q)^{r^2}$, we have that

$$P_U \leqslant \binom{t}{r}^2 (1-q)^{r^2} \leqslant \mathrm{e}^{2t-r^2} = \mathrm{e}^{-4k^2} .$$

To upper-bound P_I, we can view the number of neighbors of the input nodes in I as the sum $S = s_1 + \ldots + s_t$ of independent random 0/1 variables, where $s_i = 1$ if and only if the i-th inner node is connected to some of the nodes in I. Hence,

$$\mathrm{E}[s_i] = 1 - (1-p)^k \geqslant 1 - \mathrm{e}^{-1} ,$$

implying that $\mu := \mathrm{E}[S] > t/2$. By the Chernoff inequality (5.1),

$$P_I = \Pr[S < r] \leqslant \mathrm{e}^{-(\mu-r)^2/2t} \leqslant \mathrm{e}^{-2k^2}\,.$$

This implies that one fixed $k \times k$ submatrix $B_{I,J}$ of B can be an all-0 submatrix with probability at most $P_U + P_I + P_J \leqslant 3\mathrm{e}^{-2k^2}$. Since the total number of $k \times k$ submatrices in B is only $\binom{n}{k}^2 < n^{2k} = \mathrm{e}^{2k^2}$, the expected number of all-0 $k \times k$ submatrices is constant, as desired. By Markov's inequality, $Y = O(1)$ holds with an arbitrarily large constant probability. □

By Lemma 2.5, every matrix satisfying (i) has $\mathsf{OR}_3(A) \asymp n$, though its complement is a rather dense 2-free matrix. For matrices from the claim (iii) we have almost maximal gaps

$$\mathsf{OR}_2(\overline{A})/\mathsf{OR}_2(A) \succcurlyeq \frac{n}{\ln^3 n} \quad \text{and} \quad \mathsf{OR}(\overline{A})/\mathsf{OR}_3(A) \succcurlyeq \frac{n}{\ln^3 n}\,. \tag{5.11}$$

Finally, let us note that one can achieve also *explicit* gaps via derandomization of Theorem 5.13. To demonstrate how this can be done, we show an explicit square root gap.

Let $N = \binom{n}{k}$ and $n = 2^r$ for an even r. Consider a depth-3 circuit with N inputs and outputs labeled by distinct k-element subsets of $[n]$. The circuit also has n nodes associated with elements of $[n]$ on each of two middle levels. Connect each input and each output node $a \in \binom{[n]}{k}$ with all k neighbor-level nodes corresponding to the elements of a. Let the nodes on the middle levels be connected in a way to implement the Sylvester matrix H_n on the middle layer; thus, we view these nodes as vectors in $\mathbb{F}_2^r$, and two nodes are connected if and only if their scalar product over $\mathbb{F}_2$ is 1. Let A be an $N \times N$ matrix implemented by the circuit.

Since $\mathrm{rk}_\vee(A) \leqslant n$ by the construction, we have $\mathsf{OR}_2(A) \preccurlyeq nN$. On the other hand, the number of zeroes in A is the number of $k \times k$ all-0 submatrices in H_n. The later number is bounded from below by the number $\binom{n}{k} = N$ of possibilities to choose a k-tuple $\{a_1, \ldots, a_k\}$ of rows times the number $\binom{n2^{-k}}{k}$ of possibilities to choose k of the at least $2^{r-k} = n2^{-k}$ solutions x of the system of linear equations $\langle a_1, x\rangle = \ldots = \langle a_k, x\rangle = 0$ over $\mathbb{F}_2$. Thus, $|\overline{A}| \geqslant N\binom{n2^{-k}}{k}$.

Table 5.1: Best known results for gaps.

Gap	**Explicit**	**Non-explicit**
$\mathsf{OR}(A)/\mathsf{XOR}(A)$	$n2^{-\Theta(\sqrt{\ln n \ln\ln n})}$	$n/\ln^2 n$
$\mathsf{OR}_2(A)/\mathsf{XOR}_2(A)$	$\sqrt{n}/\ln n$	$n/\ln^2 n$
$\mathsf{SUM}(A)/\mathsf{OR}(A)$	$\sqrt{n}\,2^{-\Theta(\sqrt{\ln n \ln\ln n})}$	$\sqrt{n}/\ln^2 n$
$\mathsf{SUM}_2(A)/\mathsf{OR}_2(A)$	$\ln n$	$\ln n$
$\mathsf{OR}(\overline{A})/\mathsf{OR}(A)$	$\ln n$	$n/\ln^3 n$
$\mathsf{OR}_2(\overline{A})/\mathsf{OR}_2(A)$	$\sqrt{n}\,2^{-\Theta(\ln^{2/3} n)}$	$n/\ln^3 n$
$\mathsf{XOR}_2(A)/\mathsf{OR}_2(A)$	$\ln\ln\ln n$	$\ln\ln\ln n$

Our next goal is to show that $\overline{A}$ is K-free for $K = \binom{\sqrt{n}+1}{k}$. For this, observe that, given a set S of $m \geqslant k$ nodes on the first (or second) middle level (corresponding to rows/columns of H_n), exactly $\binom{m}{k}$ of input (resp., output) nodes are connected to none of the middle nodes outside S: these are precisely the input/output nodes whose label-sets are contained in S. Were $\overline{A}$ not K-free, this would mean that H_n contains an all-0 $m \times m$ submatrix for $m = \sqrt{n}+1$, contradicting the Lindsey Lemma.

By Theorem 3.9, we have that

$$\mathsf{OR}_2(\overline{A}) \geqslant \frac{|\overline{A}|}{K} \geqslant N\frac{\binom{n2^{-k}}{k}}{\binom{\sqrt{n}+1}{k}}\,.$$

By taking k about $\ln^{1/3} N$, and hence, $n = 2^{\Theta(\ln^{2/3} N)}$, we derive

$$\mathsf{OR}_2(\overline{A})/\mathsf{OR}_2(A) \succcurlyeq \sqrt{N}2^{-\Theta(\ln^{2/3} N)}\,.$$

6

Bounds for General Circuits

We now consider computation of linear operators $y = Ax$ over $\mathbb{F}_2$ in the class of *general circuits*, where not only XOR but arbitrary(!) boolean functions are allowed as gates.

Let $\mathsf{GEN}_d(A)$ denote the smallest number of edges in a depth-d circuit with arbitrary boolean functions as gates computing the linear operator Ax over $\mathbb{F}_2$.

6.1 Column-distance lower bounds

The following "column-distance" lower bound for depth-2 XOR circuits was first proved by Alon, Karchmer and Wigderson [3] via a very elegant argument, which we give below. Cherukhin [18] used different arguments to prove a similar lower bound for general circuits. Jukna [49] has shown that the argument of [3] also works for general circuits. Moreover, it yields such a lower bound even when the circuit is only required to correctly compute Ax on the unit vectors $x \in \{\vec{e}_1, \ldots, \vec{e}_n\}$; it is also shown there that $O(n \log n)$ edges are already enough to compute *any* linear operator in this weak sense.

Theorem 6.1. Let A be an $m \times n$ matrix each two columns of which

differ in at least k positions. Then

$$\mathsf{GEN}_2(A) \succcurlyeq k \cdot \frac{\ln n}{\ln \ln n}$$

even if the circuit is only required to correctly compute Ax on n unit vectors x.

Note that, like Theorem 3.23, this lower bound is also based of Ramseyan properties of A: the distance condition on A is equivalent to A being a $(2, m-k+1)$-Ramsey matrix.

Proof. We will need the well-known "Sunflower Lemma" of Erdős and Rado [27]. A *sunflower with k petals* is a family of k sets, all pairwise intersections of which give the same set; this common subset of the intersection is the *core* of the sunflower. The *Sunflower Lemma* states that every family of more than $s!(k-1)^s$ sets of size at most s must contain a sunflower with k petals. This can be proved by an easy induction on s.

Now take a depth-2 circuit with arbitrary boolean functions as gates computing $x \mapsto Ax$. For simplicity, we assume that all gates on the middle level are *symmetric* boolean functions; the case of non-symmetric gates is the same with a bit more subtle reasoning at the point (6.1) below (see [49]). For $i \in [n]$ and $j \in [m]$, let S_i be the set of intermediate nodes (on the second level) that are connected to the i-th input node, and T_j the set of intermediate nodes connected to the j-th output node. We may assume that $k > 0$ (since for $k = 0$ there is nothing to prove). Hence, all sets $S_1, \ldots, S_n$ must be distinct. Set

$$t = c \cdot \frac{\ln n}{\ln \ln n}$$

for a sufficiently small constant $c > 0$; for simplicity, assume that t is integer. If $\sum_{i=1}^n |S_i| > nt$, then we are done. So, assume that $\sum_{i=1}^n |S_i| \leqslant nt$. Our goal is to show that then $\sum_{j=1}^m |T_j| \geqslant kt$.

Since $\sum_{i=1}^n |S_i| \leqslant nt$, at least $n/2$ of the sets S_i must be of size at most $2t$. By the Sunflower Lemma, these sets must contain a sunflower with $2t$ petals. Having such a sunflower with a core C, we can pair its members arbitrarily, $(S_{p_1}, S_{q_1}), \ldots, (S_{p_t}, S_{q_t})$. Important for us is that

all t symmetric differences $D_l = S_{p_l} \oplus S_{q_l} = (S_{p_l} \cup S_{q_l}) \setminus C$ are mutually disjoint.

Let (S_p, S_q) be any of these pairs, and let f_j be the function computed at the j-th output gate. On input vector x, $f_j(x)$ must be the j-th position of the vector Ax. Suppose that the p-th and the q-th columns of A differ in the j-th row. Then $f_j(\vec{e}_p) \neq f_j(\vec{e}_q)$, where $\vec{e}_p$ is the p-th unit vector. This implies that

$$(S_p \oplus S_q) \cap T_j \neq \emptyset\,. \tag{6.1}$$

To show this, let g_v be a gate at a middle node v. If $v \notin S_p \oplus S_q$, then on *both* inputs $\vec{e}_p$ and $\vec{e}_q$, the gate g_v will receive either only 0s (if $v \notin S_p \cup S_q$) or exactly one 1 (if $v \in S_p \cap S_q$). Since the gate g_v is symmetric, it must behave in the same manner on both inputs. But if (6.1) does not hold, then the j-th output gate can see no other middle gates, implying that $f_j(\vec{e}_p) = f_j(\vec{e}_q)$, a contradiction.

Now, the distance property of A implies that, for every pair (S_{p_l}, S_{q_l}), $l = 1, \ldots, t$, there is a set $J_l \subseteq [m]$ of $|J_l| \geqslant k$ rows on which the p_l-th and the q_l-th columns of A differ. By (6.1), we have that

$$|D_l \cap T_j| \geqslant 1 \quad \text{for all } 1 \leqslant l \leqslant t \text{ and all } j \in J_l.$$

Since the sets $D_1, \ldots, D_t$ are pairwise disjoint, the desired lower bound follows:

$$\sum_{j=1}^{m} |T_j| \geqslant \sum_{j=1}^{m} \sum_{l=1}^{t} |D_l \cap T_j| = \sum_{l=1}^{t} \sum_{j=1}^{m} |D_l \cap T_j| \geqslant \sum_{l=1}^{t} |J_l| \geqslant kt\,. \qquad \square$$

Since every two columns of the Sylvester $n \times n$ matrix $H = H_n$ differ in at least $n/2$ positions, Theorem 6.1 implies that

$$\mathsf{GEN}_2(H) \gtrsim n \cdot \frac{\ln n}{\ln \ln n}\,,$$

even if the circuit is only required to correctly compute Hx on n unit vectors x.

On the other hand, large Hamming distance between columns *alone* cannot lead to larger lower bounds. That is, the lower bound given by Theorem 6.1 is, in fact, optimal.

Theorem 6.2 (Drucker [23])**.** There are explicit $n \times n$ matrices A whose every two columns differ in at least $n/8$ positions, but

$$\mathsf{SUM}_2(A) \preccurlyeq n \cdot \frac{\ln n}{\ln \ln n}\,.$$

Proof. Let $n = p^s$ where p is a prime power and $1 \leqslant s \leqslant p$ an integer. Set $m := n/p$, and fix a boolean $p \times m$ matrix M whose rows are labeled by elements $a \in \mathbb{F}_p$, columns by numbers $i \in \{1, \ldots, m\}$, and every two rows in M differ in at least $1/4$ of their positions.[1] For $i \in [m]$, we identify the i-th column of M with the set $S_i \subseteq \mathbb{F}_p$ of its 1-positions. Thus, we have m sets S_i in $\mathbb{F}_p$ such that, for every two elements $a \neq b \in \mathbb{F}_p$, at least $m/4$ of the sets satisfy $|S_i \cap \{a, b\}| = 1$.

We now define the desired $n \times n$ matrix A as follows. The rows of A are labeled by the pairs (a, i) with $a \in \mathbb{F}_p$ and $i \in [m]$ (recall that $pm = n$), and columns are labeled by polynomials $f(z)$ of degree at most $s - 1$ over $\mathbb{F}_p$. The matrix A has a 1 in an entry $((a, i), f)$ if and only if $f(a) \in S_i$.

Claim 6.3. For every prime power p and every integer $1 \leqslant s \leqslant p$, every two columns of A differ in at least $n(p - s)/4p$ positions.

Proof. Let $N(f) = \{(a, i) \colon f(a) \in S_i\}$ be the set of 1-entries in the f-column of A. Our goal is to show that $|N(f) \oplus N(g)| \geqslant n(p - s)/4p$ holds for every two columns f and g of A. For $a \in \mathbb{F}_p$, let

$$\Delta_a := \{(a, i) : f(a) \in S_i \text{ iff } g(a) \notin S_i\}\,.$$

Note that $(a, i) \in \Delta_a$ if and only if the i-th column of our "ambient" matrix M has different values in the $f(a)$-th and $g(a)$-th rows. Since every two distinct rows of M differ in at least $1/4$ of their $m = n/p$ positions, we have that $|\Delta_a| \geqslant m/4 = n/4p$ holds for every $a \in D := \{a \in \mathbb{F}_p \colon f(a) \neq g(a)\}$. On the other hand, since any polynomial of degree s can have at most s roots, the set D has $|D| \geqslant p - s$ elements. Thus,

$$|N(f) \oplus N(g)| = \sum_{a \in \mathbb{F}_p} |\Delta_a| \geqslant \sum_{a \in D} |\Delta_a| \geqslant (p - s)\frac{n}{4p}\,. \qquad \square$$

[1] One can, for example, take p rows of a Sylvester $N \times N$ matrix, where N is a smallest power of 2 such that $N \geqslant p^s$.

Claim 6.4.

$$\mathsf{SUM}_2(A) \leqslant 2pn\,.$$

Proof. Take a depth-2 circuit with p^2 nodes on the middle level, indexed by the pairs $(a,b) \in \mathbb{F}_p^2$. Input nodes (columns of A) correspond to polynomials $f(z)$, whereas output nodes correspond to pairs (a,i) with $a \in \mathbb{F}_p$ and $i \in [m]$. Associate with each input node f and each output node (a,i) the following subsets of nodes on the middle level:

$$V_f := \{(a,b) \in \mathbb{F}_p^2 \colon f(a) = b\} \quad \text{and} \quad W_{(a,i)} := a \times S_i\,.$$

Join each input node f to all nodes in V_f, and each output node (a,i) to all nodes in $W_{(a,i)}$. Since f is a (single-valued) function, the intersection

$$V_f \cap W_{a,i} = V_f \cap (a \times S_i)$$

can have at most one element: the element $(a, f(a))$ if $f(a) \in S_i$, and no elements otherwise. Thus, $|V_f \cap W_{a,i}| = 1$ if the entry $(f,(a,i))$ of A is 1, and $|V_f \cap W_{a,i}| = 0$ otherwise. In other words, we have exactly one input-output path for each 1-entry of A, and no paths for 0-entries. Thus, we have a depth-2 SUM circuit for A. Since $|V_f| = p$ for every polynomial f, and $|S_i| \leqslant p$ for every $i \in [m]$, the total number of edges in this circuit is at most

$$\sum_f |V_f| + \sum_{(a,i)} |S_x| \leqslant p^s \cdot p + p \cdot m \cdot p = 2pn\,,$$

as desired. □

By taking $n = p^s$ with $s = \lfloor p/2 \rfloor$ in Claims 6.3 and 6.4, we obtain an explicit matrix A such that $\mathsf{SUM}_2(A) \leqslant 2pn \preccurlyeq n \ln n / \ln\ln n$ and every two columns of A differ in $\geqslant n/8$ positions, completing the proof of the theorem. □

Note that the column-distance argument cannot yield superlinear lower bounds for circuits of depth $d > 2$ because, say, every two columns of the Sylvester $n \times n$ matrix differ in at least $n/2$ positions, but Lemma 2.5 implies that $\mathsf{XOR}_3(H) \preccurlyeq n$.

6.2 Lower bounds for code matrices

The *distance* of A is the smallest Hamming distance between the images of the linear operator $x \mapsto Ax$:

$$\operatorname{dist}(A) = \min\{|Ax \oplus Ay| \colon x \neq y\}\,.$$

A *good n-code matrix* is an $n \times m$ matrix A such that $n \preccurlyeq m$ and $\operatorname{dist}(A) \succcurlyeq n$. Such a matrix encodes 0/1 messages x of length m into codewords $y = Ax$ of length n such that codewords of any two distinct messages differ in a constant fraction of positions. Thus, good code matrices are the generator matrices of a linear self-correcting codes with very good parameters: they have constant rate (codewords are only constant times longer than messages) and nevertheless can correct a constant fraction of errors.

We have proved in § 3.8 that $\mathsf{XOR}_2(A) \succcurlyeq n \ln^{3/2} n$ holds for every good n-code matrix A. Recently, Gál et. al. [33] improved this lower bound to $\mathsf{GEN}_2(A) \succcurlyeq n(\ln n/\ln\ln n)^2$, and showed that this cannot be further improved: there are good code matrices A such that $\mathsf{XOR}_2(A) \preccurlyeq n(\ln n/\ln\ln n)^2$. Currently, their lower bound is the strongest known bound for depth-2 general, and even XOR circuits. Below we sketch their proof ideas. Their proof goes by first showing that any general circuit for a good code matrix must be a kind of a superconcentrator, and then showing that every such superconcentrator must have many edges.

Bounds on superconcentrators

A circuit with m inputs and n outputs is called a *δ-superconcentrator* if, for every integer $0 < k \leqslant m$, for every k-element subset X of inputs, and for a random k-element subset Y of output nodes, the expected number of node-disjoint paths from X to Y is $\geqslant \delta k$. This is a weakening of the property of superconcentrators by letting Y be random, and not requiring to have exactly k node-disjoint paths. Yet another weakening of the superconcentrator property was earlier considered by Dolev et al. [22], and Pudlák [86]: here both X and Y are random k-element subsets.

For some time, it was a hope that superconcentrators must have superlinear number of edges. However, as we already mentioned in Section 3.7, Valiant [105] refuted this hope: superconcentrators with $O(n)$ edges exist. Pippenger [82] has shown that such (surprisingly small) superconcentrators exist already in logarithmic depth. Dolev et al. [22] proved the existence of linear-size superconcentrators already in depth d, where $d = d(n)$ is a function growing slower than the inverse of any primitive-recursive function. The next question was: what happens if the depth is constant?

A lower bound $\Omega(n \ln n)$ and an upper bound $O(n \ln^2 n)$ in depth-2 were proved by Pippenger [85]. For larger depths d, Dolev et al. [22], and Pudlák [86] proved matching bounds of the form $n\lambda_d(n)$ for depth-d (weak) superconcentrators, where $\lambda_2(n) \asymp \ln n$, $\lambda_3(n) \asymp \ln \ln n$, and $\lambda_d(n)$ is extremely slowly growing function for larger depths d (see § 4.1).

These bounds hold even for the weak versions of superconcentrators. For the size of the strong version of superconcentrators, where every set of k inputs must be connected with every subset of k outputs by node-disjoint paths, better lower bounds were obtained in depth 2. First, Alon and Pudlák [5] proved the lower bound of $n \ln^{3/2} n$. By using different arguments, Radhakrishnan and Ta-Shma [92] were then able to prove the optimal bound of $n \ln^2 n / \ln \ln n$.

Codes and superconcentrators

A connection between superconcentrators and circuits computing error-correcting codes was already observed by Spielman [101]. His construction of linear-size encoding circuits was inspired by known constructions of linear-size superconcentrators. Spielman also observed that some similarity to superconcentrators is necessary. He proved that circuits with m inputs and n outputs computing codes with minimum distance δn have δm node-disjoint paths from any set X of $|X| = \delta m$ inputs to any set Y of $|Y| = (1 - \delta)n$ outputs.

Gál et. al. [33] proved the following stronger connection between circuits encoding error-correcting codes and superconcentrators, which may be of independent interest.

Lemma 6.5 (Gál et. al. [33])**.** Let A be an $n \times m$ matrix. If $\text{dist}(A) \geqslant \delta n$, then every general circuit for an $n \times m$ matrix A is a δ-superconcentrator.

Proof. Take a general circuit for A, and let X be a fixed subset of its $|X| = k$ inputs. For a subset Y of output nodes, let $f(Y)$ denote the maximal number of node-disjoint paths from inputs in X to outputs in Y. Call an output node $v \notin Y$ *bad*, if $f(Y \cup \{v\}) = f(Y)$. Let W be the set of all $|W| = n$ output nodes.

By a result of Perfect [79] in matroid theory (see also [109, Chapter 13]), the subsets of W formed by the sets of endpoints of node-disjoint paths from X to W are independent sets of a matroid over W. Thus, if B is the set of all output nodes that are bad for Y, then $f(Y \cup B) = f(Y)$.

Claim 6.6. For any Y with $|Y| < k$, at least δn output nodes are not bad for Y.

Proof. Let B be the set of all output nodes that are bad for Y, and let $\ell < k$ be the number of node-disjoint paths from X to Y. Hence, $f(Y \cup B) = f(Y) = \ell$. By Menger's theorem, the smallest cut separating X and $Y \cup B$ is of size ℓ. If we set all input bits except for X to 0, then by varying inputs to X we have $2^{|X|} = 2^k$ different inputs. However, over these 2^k inputs, the outputs belonging to $Y \cup B$ will take on at most 2^ℓ different settings, as these output bits will be determined by the values at the gates of the cut separating X and $Y \cup B$, which is of size $\ell < k$.

Thus there exist two different inputs x and x' such that the outputs of our circuit on these two inputs agree on the $Y \cup B$ part. So the Hamming distance between the outputs of the circuit on x and x' is at most the number of output nodes outside of $Y \cup B$. However, since G computes a code with minimum distance δn, the Hamming distance between the encodings of any two different inputs has to be at least δn. Thus, the number of output nodes outside of $Y \cup B$ is at least δn. □

Now pick a random k-element subset Y of output nodes by picking at random one element at a time. By the claim, we have that as long as

$|Y| < k$, with probability at least δ the next randomly chosen output will increase the number of node-disjoint paths from X to the current Y by one. By linearity of expectation we get at least δk node-disjoint paths on average. □

Theorem 6.7 (Gál et. al. [33])**.** For every good n-code matrix A we have

$$\mathsf{GEN}_2(A) \succcurlyeq n \cdot \left(\frac{\ln n}{\ln \ln n}\right)^2 .$$

Proof. We only sketch the proof without calculating the specific constants arising along the way. The proof uses some ideas invented by Radhakrishnan and Ta-Shma [92] in their proof of an optimal bound of $\Theta(n \ln^2 n / \ln \ln n)$ on the size of depth-2 superconcentrators; see also Dutta and Radhakrishnan [25] for a simplification and generalization of these ideas.

Take a general depth-2 circuit for A, and let E be set of edges in it. We assume that n is sufficiently large. For a node u and a set of nodes V, let $\deg(u)$ be the degree of u (number of all incoming and outgoing edges), and $\deg_V(u)$ the number of nodes in V incident with u. For the sake of contradiction, assume that $|E| < cnt^2$ holds for every constant $c > 0$, where

$$t := \frac{\ln n}{\ln \ln n} .$$

For a parameter k, split the set V of nodes on the second level into three classes:

$$S = \left\{ v \in V : \deg(v) < \frac{n}{k \ln^2 n} \right\} \qquad \text{(small-degree nodes)}$$

$$M = \left\{ v \in V : \frac{n}{k \ln^2 n} \leqslant \deg(v) < \frac{n \ln^2 n}{k} \right\} \qquad \text{(medium-degree nodes)}$$

$$L = \left\{ v \in V : \deg(v) \geqslant \frac{n \ln^2 n}{k} \right\} \qquad \text{(large-degree nodes)}$$

Our first goal is to choose a value of k for which at most $c_1 nt$ edges[2] are incident with nodes in M. For this, let $k_i := \ln^{4i} n$, and let M_i be the

[2]In what follows, $c_1, c_2, \ldots$ stand for appropriately chosen constants; these are small constants depending only on the constants ρ and δ in the definition of a good code-matrix.

set M of medium degree nodes when $k := k_i$. Since $(n/k_{i+1}) \ln^2 n \leqslant n/k_i \ln^2 n$, the sets M_i are disjoint. Consider the integers i between $t/16$ and $t/8$. Were M_i be incident with at least $16cnt$ edges for all these integers i, then the total number of edges would be at least cnt^2, contradicting our assumption $|E| < cnt^2$.

Thus, we can take $c_1 = 16c$ and fix a k between $n^{1/4}$ and $n^{1/2}$ for which at most $c_1 nt$ edges are incident with nodes in M. Then at least $c_2 n$ output nodes w are "good" in the sense that $\deg_M(w) \leqslant c_2 t$, and at least $c_3 n$ input nodes u are "good" in the sense that $\deg_M(u) \leqslant c_4 t$ and $\deg(u) \leqslant c_5 t^2$.

After all these technical preparations, we now come to the crux of the argument. Let $p = 1/\ln n$, and consider the following random process: [3] for each node $v \in M$, with probability p remove all the edges leaving v, and with the remaining probability $1-p$ remove all the edges entering v. An input or output node *survives* if no edge incident to it was removed during this process. Let U be the set of surviving good input nodes, and W the set of surviving good output nodes.

Using the consequence (5.7) of Chebyshev's inequality for sums of weakly dependent random 0/1 variables, and making some computations as in the proof of Theorem 5.13, one can show that $|W| \geqslant (1-c_6)n$ and $|U| \geqslant c_7 n^{9/10}$ both hold with probability > 0. In particular, $|U| \geqslant k$ if n is sufficiently large. Fix U and W with these properties. We know that:

> there were no paths from U to W in the original circuit going through M.

Now let $X \subseteq U$ be a random k-element subset of U, and Y be a random k-element subset of output nodes. The proof is finished by proving the following two contradictory claims:

(i) The expected number of node-disjoint paths from X to Y through S is $\geqslant c_8 k$.
(ii) The expected number of all paths from X to Y through S is $o(k)$.

[3] This idea goes back to Hansel and Krichevski; see the proof of Lemma 3.7. The same idea was also used in [92, 25].

Claim (i) follows since the size of L is $o(k)$, the expected number of node-disjoint paths from X to Y is $\Omega(k)$ (by Lemma 6.5), and U is not connected through M with almost all outputs (since $|W| \geqslant (1 - c_6)n$). Claim (ii) follows from the bounds on the degree of good inputs and nodes in S. □

The lower bound of Theorem 6.7 is optimal as well.

Theorem 6.8 (Gál et. al. [33])**.** There exist good n-code matrices A such that

$$\mathsf{XOR}_2(A) \preccurlyeq n \cdot \Big(\frac{\ln n}{\ln \ln n}\Big)^2 .$$

Proof. We will only give a rough idea of the proof of a weaker upper bound $\mathsf{XOR}_2(A) \preccurlyeq n \ln^2 n$ (the proof of the stronger bound uses existence of good expander graphs). First, note that

$$\mathrm{dist}(A) = \min\{|Ax| \colon x \neq \vec{0}\} ,$$

where $|Ax|$ is the weight of (number of 1s in) the vector Ax. Thus, it is enough to show that there exists a depth-2 XOR circuit with $\preccurlyeq n \ln^2 n$ edges which, on every nonzero input, outputs a vector with $\succcurlyeq n$ ones. (The task is non-trivial because the circuit must output $\vec{0}$ on input $x = \vec{0}$.) That is, the circuit must be a "magnifier" in the sense that it maps all nonzero vectors to vectors of very large weight, while mapping $\vec{0}$ to $\vec{0}$.

The idea is first to consider $r = \ln n$ slices of $\{0,1\}^n$, the i-th of which consists of all vectors of weight between $w_i = n/\mathrm{e}^i$ and w_i/e, and to design depth-1 magnifiers of size about $L_i = \frac{n}{w_i} \ln \binom{n}{w_i}$ for each of these slices. Then one puts these depth-1 magnifiers in parallel, and constructs a depth-2 circuit as follows: at each gate on the last (output) level, choose at random one output in each of r depth-1 magnifiers, and take the XOR of them.

The number of edges on the second level is $nr = n \log n$. Since

$$\frac{n}{w_i} \ln \binom{n}{w_i} \leqslant \frac{n}{w_i} \ln \Big(\frac{en}{w_i}\Big)^{w_i} = n \ln \mathrm{e}^i = n \cdot i ,$$

the number of edges on the first level also does not exceed the desired upper bound:

$$\sum_{i=1}^{r} L_i \leqslant n \sum_{i=1}^{r} i \leqslant nr^2 = n \ln^2 n$$

A depth-1 magnifier for the slice of vectors of weight w is also constructed by probabilistic arguments. Roughly, at each output node, one joins it with a subset T of about $\frac{1}{w} \ln \binom{n}{w} \asymp \ln(n/w)$ inputs. Then one shows that, for every w-element subset S of inputs, $|S \cap T|$ is odd with at least some constant probability $p > 0$; cf. (6.2) below. □

In fact, Gál et al. [33] prove matching bounds for generator matrices of good codes in *all* depths. For example, in depth 3, the bound is $\Theta(n \ln \ln n)$. The *lower* bounds for depths $d > 2$ are almost direct consequences of the lower bounds for depth-d weak superconcentrators proved earlier by Dolev et al. [22], and Pudlák [86].

6.3 Hashing is easy for XOR circuits

An operator $f : S \to \{0,1\}^m$ cannot be injective, if $m < \log |S|$. But if we allow the dimension of the range be just twice larger (than this "counting barrier"), then an even *linear* operator of *linear* XOR complexity can do the job!

Theorem 6.9 (Miltersen [65]). For any constant $c > 0$ and any subset $S \subseteq \{0,1\}^n$, there exists a boolean $m \times n$ matrix A with $m \leqslant (2 + c) \log |S|$ rows such that $\mathsf{XOR}(A) \preccurlyeq n$ and $Ax \neq Ay$ for all distinct $x, y \in S$.

Goldreich and Wigderson [36] proved an earlier, somewhat weaker upper bound using universal hashing.

Proof. We present a simplified proof due to Chashkin [17] (see also [16]). Let us first sketch the proof idea. The desired matrix A has the form $A = L \cdot G$, where G is $4n \times n$ matrix which maps every nonzero vector to a vector of weight at least γn for a constant $\gamma > 0$. The matrix L has the property that it maps the G-images of $|S|^2$ differences of vectors in S into nonzero vectors.

That the mapping $x \mapsto Ax$ is injective follows from the following considerations:

1. The mapping Ax is injective on S if and only if it does not map any difference of two distinct vectors of S into zero vector.
2. The mapping Gx is injective on $\{0,1\}^n$ (simply follows from definition). Hence, it maps all differences of vectors of S into distinct vectors of large weight.
3. The mapping Lx maps all images of differences into nonzero vectors.

The matrix L is taken to be a certain type of random matrix with $O(n)$ ones; each its entry is 1 with probability $\Theta(1/n)$. To construct a matrix G of complexity $O(n)$, a recursion is used, exploiting special building blocks also provided by probabilistic arguments. When constructing G, Chashkin's proof essentially follows the exposition due to Sudan [102].

Now we turn to the actual proof.

Expanding operator G Fix a constant $\delta < 2^{-8}$, and let n be sufficiently large. Say that a $n \times 2n$ matrix is *dispersed* if

(i) every column has 7 ones, and
(ii) every $k \leqslant 2\delta n$ columns have a 1 in more than $4k$ rows.

Claim 6.10. Dispersed $n \times 2n$ matrices $M_{n,2n}$ exist.

Proof. We will show this by counting. Consider the set of all $n \times 2n$ matrices with 7 ones in each column. Let p be the fraction of these matrices that are non-dispersed. It is easy to see that the number of possibilities to choose k columns with 1s in $4k$ rows is $\binom{2n}{k}$, and there are $\binom{n}{4k}$ possibilities to choose these $4k$ rows. The 1s in these k columns can be displaced in at most $\binom{4k}{7}^k$ ways, and in the remaining columns this can be done in $\binom{n}{7}^{2n-k}$ ways. Thus, p is at most the sum over all $1 \leqslant k \leqslant 2\delta n$ of

$$\binom{2n}{k}\binom{n}{4k}\binom{4k}{7}\binom{n}{7}^{2n-k}\binom{n}{7}^{-2n} = \binom{2n}{k}\binom{n}{4k}\binom{4k}{7}\binom{n}{7}^{-k}$$

$$\leqslant \Big(\frac{3\cdot 2n}{k}\Big)^k \Big(\frac{3\cdot n}{4k}\Big)^{4k} \Big[\frac{4k(4k-1)\cdots(4k-6)}{n(n-1)\cdots(n-6)}\Big]^k$$
$$\leqslant \Big(\frac{3\cdot 2n}{k}\Big)^k \Big(\frac{3n}{4k}\Big)^{4k} \Big(\frac{4k}{n}\Big)^{7k} = 3^{5k}2^{7k}\Big(\frac{k}{n}\Big)^{2k} < (3^5 2^7 \delta^2)^k \leqslant 2^{-k}\,.$$

Thus, the fraction of not dispersed matrices is $p \leqslant \sum_{k=1}^{2\delta n} 2^{-k} = 1 - 2^{-2\delta n} < 1$, as desired. □

Claim 6.11. Let $M = M_{n,2n}$ be a dispersed $n\times 2n$ matrix. Then $|Mx| > |x|$ holds for every vector $x \in \mathbb{F}_2^{2n}$ of weight $|x| \leqslant 2\delta n$.

Proof. It is enough to show that, for every $k \leqslant 2\delta n$ columns, there are more than k rows, each having exactly one 1 in these columns. To show this, take any $k \leqslant 2\delta n$ rows of M, and let M' be the corresponding $n \times k$ submatrix of M. Let a be the number of rows having exactly one 1, and b the number of rows having at least two 1s in M'. Since the matrix M is dispersed, we have that $a+b > 4k$ and $a+2b \leqslant 7k$, which yields $a > 4k - b \geqslant 4k - (7k-a)/2 = k/2 + a/2$, that is, $a > k$, as desired. □

Claim 6.12. There is a constant $0 < \gamma < 1$, and a $4n \times n$ matrix $G = G_{4n,n}$ such that $\mathsf{XOR}(G) \preccurlyeq n$ and $|Gx| \geqslant 4\gamma n$ for every $x \in \mathbb{F}_2^n$, $x \neq \vec{0}$.

Proof. Induction on $\log n$. Let m be the smallest m for which there exists a $4m \times 2m$ matrix from Claim 6.11. Let $G = G_{4m,m}$ be the matrix such that $Gx = (x,x,x,x)$. That is, G consists of 4 identity matrices I_m. It is clear that then $|Gx| \geqslant 4\gamma m$ holds for all $x \neq \vec{0}$ with $\gamma = 1/m$.

Now assume a matrix $G_{4n,n}$ exists for some $n \geqslant m$. Using this matrix and dispersed matrices $M_{n,2n}$ guaranteed by Claim 6.11, we define the $8n \times 2n$ matrix $G = G_{8n,2n}$ as follows:

$$G_{8n,2n} = \begin{bmatrix} I_{2n} \\ G_{4n,n}\cdot M_{n,2n} \\ M_{2n,4n}\cdot G_{4n,n}\cdot M_{n,2n} \end{bmatrix}.$$

Set $\gamma = \min\{1/m, \delta/4\}$, and let $x \in \mathbb{F}_2^{2n}$ be an arbitrary nonzero vector. Our goal is to show that $|Gx| \geqslant 8\gamma n$. This clearly holds if

$|x| \geqslant 8\gamma n$. If $|x| < 8\gamma n$ then, by Claim 6.11, the vector $x' = M_{n,2n}x$ has nonzero weight, and the induction hypothesis implies that the vector $y = G_{4n,n}x'$ has weight $|y| \geqslant 4\gamma n$. If we have an even stronger inequality $|y| \geqslant 8\gamma n$, then we are done. If $4\gamma n \leqslant |y| < 8\gamma n$ then, by Claim 6.11, we have that the vector $z = M_{2n,4n}y$ has weight $|z| \geqslant |y|$, and hence, $|Gx| \geqslant |y| + |z| \geqslant 8\gamma n$.

It remains to show that $\mathsf{XOR}(G_{4n,n})$ is linear in n. At this point, it is convenient to allow that output nodes may have nonzero fanout: it is clear that by adding a linear number of edges, we can obtain a standard circuit. Under this proviso, we have that $\mathsf{XOR}(G_{4n,n}) \leqslant 42n$. Indeed, since the matrix $M_{n,2n}$ has only $14n$ ones, we obtain that

$$\begin{aligned}\mathsf{XOR}(G_{8n,2n}) &\leqslant \mathsf{XOR}(M_{2n,n}) + \mathsf{XOR}(G_{4n,n}) + \mathsf{XOR}(M_{4n,2n}) \\ &\leqslant 14n + 42n + 28n = 42 \cdot 2n\,. \quad \square\end{aligned}$$

Contracting operator L Now let $t > 0$ be an integer parameter, and $0 < \gamma < 1$ a constant.

Claim 6.13. For every set $D \subseteq \mathbb{F}_2^n$ of $0 < |D| < 2^n$ vectors such that $|x| \geqslant \gamma n$ for all $x \in D$, there exists an $m \times n$ matrix L with $m = \lceil (1 + 2^{-t}) \log 2|D| \rceil$ such that $\mathsf{XOR}(L) \preccurlyeq mt$ and $Lx \neq \vec{0}$ for all $x \in D$.

In the proof of this last claim, we will use the following simple fact: if $X = X_1 + \cdots + X_t$ is a sum of independent 0-1 Bernoulli random variables, each with success probability α, then

$$\Pr[X \text{ is odd}] = \frac{1 - (1 - 2\alpha)^t}{2}\,. \tag{6.2}$$

To verify this, it is enough to consider the product Y of t ± 1 random variables $Y_i = 1 - 2X_i$. Hence, the sum of X_i is odd if and only if $Y = -1$. Since the Y_i's are independent, we have $\mathrm{E}[Y] = \prod_i \mathrm{E}[Y_i] = \prod_i (1 - 2\alpha)$. It remains to observe that $\mathrm{E}[Y] = \Pr[Y = 1] - \Pr[Y = -1] = 1 - 2 \cdot \Pr[Y = -1]$.

Proof of Claim 6.13. Take a random $m \times n$ matrix L, where each entry is set to 1 independently and with equal probability α. Let $x \in \mathbb{F}_2^n$ be

a fixed vector of weight $|x| \geqslant d$ where $d \geqslant \gamma n$. By (6.2),

$$\Pr[Lx = \vec{0}] \leqslant 2^{-m}[1 + (1-2\alpha)^d]^m \leqslant 2^{-m}(1 + 2^{-2\alpha d})^m \leqslant 2^{-m} 4^{m 2^{-2\alpha d}}.$$

If we take $\alpha = (t+2)/2d$, then the right-hand side is at most $2^{-m(1-2^{-t-1})}$. For our choice of m, we have that $\Pr[Lx = \vec{0}] < 1/(2|D|)$, from which

$$\Pr[Lx = \vec{0} \text{ for some } x \in D] < |D|/(2|D|) = 1/2$$

follows. On the other hand, the expected number of 1s in L is $\alpha n m \leqslant (t+1)\gamma m$. By Chebyshev's inequality, L will have more than $2(t+1)\gamma m$ ones with probability $< 1/2$. Thus, the desired matrix L exists. □

Proof of Theorem 6.9 Now we can finish the proof of the theorem as follows. Let $c > 0$ be a given constant, and take $t = \lceil -\log c + 2 \rceil$. Let $S \subseteq \mathbb{F}_2^n$, and consider the set $D = \{x \oplus y \colon x \neq y \in S\}$ of all differences (modulo 2) between the vectors in S. Claim 6.12 gives us a $4n \times n$ matrix $G = G_{4n,n}$ such that $\mathsf{XOR}(G) \preccurlyeq n$ and $|Gx| \geqslant 4\gamma n$ for every $x \in D$. On the other hand, Claim 6.13 gives us an $m \times 4n$ matrix L such that $\mathsf{XOR}(L) \preccurlyeq mt \preccurlyeq n$, $Lx \neq \vec{0}$ for all $x \in D$ and

$$m = \lceil (1 + 2^{-t}) \log 2|D| \rceil \leqslant 2(1 + c/4) \log |S| + 1 \leqslant (2 + c) \log |S|.$$

Thus, $A = L \cdot G$ is the desired $m \times n$ matrix with $Ax \neq Ay$ for all $x \neq y \in S$. This completes the proof of Theorem 6.9. □

By taking S to be the set of all vectors with Hamming weight $\leqslant k$, we obtain the following consequence.

Corollary 6.14. For every $1 \leqslant k \leqslant n/2$, there is an $m \times n$ matrix A with $m \preccurlyeq \log \binom{n}{k}$ rows such that $\mathsf{XOR}(A) \preccurlyeq n$ and any k columns of A are linearly independent.

Note that the matrix in this corollary is a *parity-check* matrix of a linear self-correcting code with very good parameters. As we mentioned in § 6.2, Gál et al. [33] also proved surprisingly small upper bounds for *generator matrices* A of good self-correcting codes. These bounds show why lower bounds on the size of XOR circuits are so difficult to prove: these circuits may have unexpected power!

7

Conclusion and Open Problems

We described known and new results concerning the computational complexity of linear operators over different semirings. Unlike the XOR complexity, the SUM and the OR complexities are relatively well understood. Still, even there some questions remain. In particular, we already know (Theorem 5.1) that the SUM/OR gap is at least $n^{1/2-o(1)}$, if we allow OR circuits of depth at least 3. But what about depth 2? We know that the gap $\mathsf{SUM}_2(A)/\mathsf{OR}_2(A)$ may be at least logarithmic (Theorem 5.4).

Problem 7.1. How large the SUM/OR gaps can be?

We know that the intersection matrices $Ð_n$ (complements of Kneser–Sierpinski matrices) have small OR complexity; see (4.2).

Problem 7.2. What is the SUM complexity of $Ð_n$?

For the $n \times n$ Kneser–Sierpinski matrix D_n, we know that $\mathsf{OR}_2(D_n) \asymp n^{1+c}$ for some constant c lying somewhere between 0.16 and 0.28 (see Lemma 4.2).

Problem 7.3. What is the right order of magnitude of $\mathsf{OR}_2(D_n)$?

As we mentioned in § 2.1, Pippenger [81, 83] achieved the asymptotics $\mathsf{L}(m,n) \sim mn/\log(mn)$ for the complexity of the hardest boolean $n \times m$ matrix for all $\log n \ll m \leqslant n$ via constructing circuits of growing depth. Several years before, Nechiporuk conjectured that this can be achieved in constant depth, and even in depth 4.

Problem 7.4. Does $\mathsf{L}_d(m,n) \sim mn/\log(mn)$ hold for all $\log n \ll m \leqslant n$ and a constant d?

We used Kronecker products of matrices to show the SUM/OR gap (Theorem 5.1) as well as the optimality of Nechiporuk's theorem (Theorem 3.9). We have also shown (Theorem 3.20) that $\mathsf{L}(A \otimes B) \geqslant r \cdot |A|/k^2$ holds for every $(k+1)$-free matrix B, where r is the L-rank of A, and $\mathsf{L} \in \{\mathsf{SUM}, \mathsf{OR}\}$. In depth 2, we have a stronger bound $\mathsf{L}_2(A \otimes B) \geqslant \mathrm{tr}(A) \cdot \mathsf{L}_2(B)$ (Theorem 3.19).

Problem 7.5 (Find et al. [30])**.** Does $\mathsf{L}(A \otimes B) \geqslant r \cdot \mathsf{L}(B)$ hold?

The next problem is to better understand the effect of the depth.

Problem 7.6. How much can the restriction to depth 2 increase the L-complexity?

For the Sylvester $n \times n$ matrix $H = H_n$, we have that $\mathsf{OR}_2(H) \asymp n^{3/2}$ (Theorem 4.3) but $\mathsf{OR}(H) \preccurlyeq n \log n$. Thus, $\mathsf{OR}_2(H)/\mathsf{OR}(H) \succcurlyeq \sqrt{n}/\log n$. Theorems 5.1 and 5.2 show that the same gap is achievable on a matrix A with $\mathsf{OR}(A) \preccurlyeq n$. By taking the $n \times n$ matrix $M = H_m \otimes J_k$, where $n = km$ and J_k is the $k \times k$ all-1 matrix with $k = \log n$, one can slightly improve the gap until $\sqrt{n/\log n}$. Since the matrix M is t-free for $t = k\sqrt{m}$, Theorem 3.9 yields $\mathsf{OR}_2(M) \geqslant |M|/t \succcurlyeq k^2m^2/t = km^{3/2}$. On the other hand, Lemma 2.8 and Lemma 2.6 yield $\mathsf{OR}(M) \leqslant L(H_m) + 2km \preccurlyeq km$.

It would be interesting to beat this "square root" gap. Yet fewer is known about what happens with XOR circuits. Spielman [101] constructed explicit good code matrices A with $\mathsf{XOR}(A) \preccurlyeq n$. Together with Theorem 6.7, this yields an explicit gap of about $(\frac{\ln n}{\ln \ln n})^2$.

Problem 7.7. Do matrices A with $\mathsf{XOR}_2(A)/\mathsf{XOR}(A)$ growing faster than polylog(n) exist?

In Theorem 5.13 we have shown that matrices A achieving the gaps $\mathsf{OR}(\overline{A})/\mathsf{OR}(A) \succcurlyeq n^{1-o(1)}$ exist.

Problem 7.8. Exhibit *explicit* matrices achieving such a gap.

We have shown that almost all submatrices of the Sylvester matrix exhibit almost maximal OR/XOR gaps (Theorem 5.8). Explicit gaps up to n/Δ with $\Delta = 2^{O(\sqrt{\ln n \ln\ln n})}$ are also known (see § 5.4).

Problem 7.9. Improve this explicit gap.

A related problem (cf. § 5.4) is

Problem 7.10. What are the largest possible gaps $\mathsf{OR}_2(A)/\mathsf{XOR}_2(A)$ and $\mathsf{OR}_3(A)/\mathsf{XOR}_3(A)$ for a 2-free matrix A?

Unlike for OR/XOR gaps, much less is known about XOR/OR gaps. The gap of $\Omega(\ln\ln\ln n)$ in depth 2 was shown in Theorem 5.12.

Problem 7.11. Do matrices A with $\mathsf{XOR}(A)/\mathsf{OR}(A) \to \infty$ exist?

Intersection matrix Đ_n could be again a natural candidate to try.

One can also consider an analogue of one-wayness problem for linear operators: how large the gap $\mathsf{XOR}(A)/\mathsf{XOR}(A^{-1})$ may be for an invertible matrix A over $\mathbb{F}_2$. One can easily see that in constant depth the gap grows unboundedly: just consider the full triangular matrix T_n and note that T_n^{-1} is a bidiagonal matrix. Thus,

$$\mathsf{XOR}_2(T_n)/\mathsf{XOR}_1(T_n^{-1}) \asymp \log n \quad \text{and} \quad \mathsf{XOR}_d(T_n)/\mathsf{XOR}_1(T_n^{-1}) \to \infty$$

for any constant d. In unbounded depth the following problem is open and deserves investigation.

Problem 7.12. Do matrices A with $\mathsf{XOR}(A)/\mathsf{XOR}(A^{-1}) \to \infty$ exist?

Hiltgen [46] has (implicitly) proved the following non-trivial upper bound

$$\mathsf{XOR}(A)/\mathsf{XOR}(A^{-1}) \preccurlyeq (n/\log n)^{1/2}\,.$$

for any triangular matrix A, that is, any matrix obtained from T_n by flipping to 0 some its 1s outside the diagonal.

The situation with *explicit* lower bounds for XOR circuits is even worse. The strongest known lower bounds are due to Gál et al. [33]. In depth 2 these bounds are of the form $n(\ln n/\ln\ln n)^2$, and are of the form $n\ln\ln n$ in depth 3. On the other hand, as noted already by Valiant [106], dense 2-free matrices "should" require large XOR circuits, at least in depth 2.

Problem 7.13. Does any of the known dense 2-free $n\times n$ matrices A require $\mathsf{XOR}_2(A)\succcurlyeq n^{1+\epsilon}$ for a constant $\epsilon>0$?

A related problem of Pudlák, Rödl and Savický [89] asked whether the complements $\overline{A}$ of dense 2-free matrices A must have large boolean rank $\mathrm{rk}_\vee(A)$. If true, this together with Valiant's reduction [106], would imply a superlinear lower bound for fanin-2 circuits over $\{\wedge,\vee,\neg\}$ of logarithmic depth (see [50, Chapter 11] for how this happens).

Recently, Katz [53] almost refuted this belief via probabilistic arguments: there exist 2-free $n\times n$ matrices A with $|A|\succcurlyeq n^{1+\epsilon}$ ones, for a constant $\epsilon>0$, such that $\mathrm{rk}_\vee(\overline{A})\preccurlyeq\log n$; see Theorem 5.13(ii) for a simpler proof. We write "almost refuted", because his matrices are not dense enough. In § 1.3, we have seen explicit 2-free matrices with $|A|\succcurlyeq n^{3/2}$ ones. To have the desired consequences for fanin-2 circuits of logarithmic depth, it would be enough to show that there is an arbitrary small constant $c>0$ such that the complement of a matrix A', obtained by removing all but $n^{3/2-c}$ ones from A, has boolean rank at least n^c.

The following two problems deal with the effect of circuit fanin/fanout and memory size. Although we do not touched these aspects, they deserve an attention.

Problem 7.14. Find an asymptotic of the complexity of $n\times n$ boolean matrices in the class of fanin-2 and fanout-2 XOR circuits.

By Theorem 2.4 and Observations 1.3 and 1.1, the asymptotic value has the form $cn^2/\log n$ for some constant c between 1 and 2. The problem is to determine this constant. Important here is that *both* the fanin and the fanout are bounded. If we leave one of them unbounded, then similar arguments as in the proof of Theorem 2.4 give $n^2/2\log n$ nodes, and hence, $n^2/\log n$ edges.

One can also consider memory (or space) restriction which is naturally applied to the straight-line version of XOR circuits. Assume we are given an $n \times n$ matrix A and $m \geqslant n$ bit registers. Initially, n of the registers are filled with the bits of a given input vector x, and the value of any other register is zero. A computation proceeds step by step: at each step, the sum of values in some registers is computed, and the result is written into one of the registers; the old value of this register disappears. Finally, some n of registers output the required vector $y = Ax$. The complexity of a computation is the total number of summands in computed sums, which is the number of edges in the corresponding XOR circuit. In the case $m = n$ the computation (program or corresponding circuit) is usually called *in-place*. Note that this model is even more restricted than that of leveled circuits with at most n nodes on each level: in an in-place circuit, the registers cannot be changed in parallel.

Unlike for the case of boolean functions and boolean circuits, each boolean matrix can be computed by an in-place XOR circuit. This can be done, for example, by Gaussian elimination: registers $x_1, \ldots, x_n$ correspond to columns, and the addition of the i-th column to the j-th column means to replace the current content of x_j by $x_i \oplus x_j$. In fact, if the matrix has full rank, then in-place circuit just *is* a Gaussian elimination procedure.

Problem 7.15 (Wigderson [110])**.** Prove an explicit nonlinear lower bound on the in-place XOR complexity.

A related and apparently easier question is to show that matrices A whose in-place complexity is larger than $\mathsf{XOR}(A)$ exist.

We already know that $n \times n$ requiring XOR circuits of size about $n^2/\log n$ exist; see Theorem 2.4. Also, all known superlinear lower bounds for depth-2 XOR circuits (except those proved using rigidity arguments) actually hold for *general* circuits where arbitrary boolean functions are allowed as gates.

Problem 7.16. Do $n \times n$ matrices with $\mathsf{GEN}_2(A) \succcurlyeq n^2/\log n$ exist?

We only know an affirmative answer for "half-linear" circuits, where either all middle gates or all output gates are required to be linear. Ac-

tually, every such circuit can even be "linearized", that is, transformed into an XOR circuit of depth 2 computing the same operator without increasing the complexity.

Lemma 7.17. Half-linear depth-2 circuits can be linearized.

Proof. First, consider the case when XOR-gates are on the output level. Imagine each function $f(x)$ computed at a gate on the middle level as its multilinear XOR (Zhegalkin polynomial) representation. The linear part of a polynomial is a sum of degree-1 monomials. Clearly, each sum computed at an output gate is a sum of linear parts of its inputs (nonlinear parts must be canceled to get a linear function). So, all middle-level functions can be replaced by their linear parts.

Next, consider the case when XOR-gates are on the middle level. Take one output gate, and let V be the linear span of sums computed in the middle-level inputs of this gate. Suppose that the function f computed at this gate is not an XOR of some of its inputs. Since the function f must be linear, this means that f is of the form $f = g \oplus h$ with $g \in V$ and $h \notin V$. Let $g_1, \ldots, g_s$ be a basis of V. Our assumption $h \notin V$ implies that for every boolean $\alpha_1, \ldots, \alpha_s, \beta$, the system $g_1(\vec{x}) = \alpha_1, \ldots, g_s(\vec{x}) = \alpha_s, h(\vec{x}) = \beta$ has a solution.

Now consider solutions $\vec{x}_0$ and $\vec{x}_1$ of two such systems differing only by values 0 and 1 of β. Then all functions in V take the same values on $\vec{x}_0$ and $\vec{x}_1$. Hence, the function f must take the same values as well. But this is impossible, because $h(\vec{x}_0) = 0$ and $h(\vec{x}_1) = 1$. This shows that f must be a linear combination of its inputs, as desired. □

For a measure $\mathsf{L}(A)$ with $\mathsf{L} \in \{\mathsf{XOR}, \mathsf{GEN}\}$, let $\mathsf{L}[A]$ denote its "relaxed" version, where it is only required that a circuit correctly computes the operator $y = Ax$ over $\mathbb{F}_2$ on n unit vectors $x \in \{\vec{e}_1, \ldots, \vec{e}_n\}$; on other inputs x it may output arbitrary values. It is easy to see that $\mathsf{XOR}[A] = \mathsf{XOR}(A)$, that is, in the case of XOR circuits, this is no relaxation at all. Thus, $n \times n$ matrices A with $\mathsf{XOR}[A] \succcurlyeq n^2/\log n$ exist. On the other hand, Jukna [49] has shown that $\mathsf{GEN}_2[A] \preccurlyeq n \log n$ holds for all matrices.[1] Thus, $\mathsf{XOR}[A]/\mathsf{GEN}_2[A] \succcurlyeq n/\log^2 n$. That is, under

[1] Moreover, the constructed circuits are half-linear with polynomials of only logarithmic degree as output gates.

the relaxation (be correct only on basis vectors), usage of superlinear gates *can* help to substantially reduce the circuit size. But what about the gap $\mathsf{XOR}(A)/\mathsf{GEN}(A)$?

Problem 7.18. Do non-linear gates help to compute $\mathbb{F}_2$-linear operators?

It is known that the answer is "no" for circuits over *infinite* fields; see, for example, [13, Theorem 13.1].

When trying to approach this question over the field $\mathbb{F}_2$, one faces the following "min-rank conjecture". A completion of a partially defined 0/1 matrix A is a 0/1 matrix obtained by setting undefined entries to 0 and 1. Let $R(A)$ be the smallest possible rank over $\mathbb{F}_2$ of a completion of A. A system of semi-linear equations for a partial matrix A has the form $A^0\vec{x} = f(\vec{x})$, where A^0 is obtained from A by setting all undefined entries to 0, and $f : \mathbb{F}_2^n \to \mathbb{F}_2^n$ is an operator, the i-th coordinate of which can only depend on variables corresponding to the undefined entries in the i-th row of A. Let $s(A)$ be the maximum, over all possible operators f, of the number of solutions of such a system.

Problem 7.19 (Jukna and Schnitger [51])**.** Is $s(A) \leqslant 2^{n-\epsilon\cdot R(A)}$ for some constant $\epsilon > 0$?

If true, this would give a negative solution for Problem 7.18: then $\mathsf{XOR}(A)/\mathsf{GEN}(A) = O(1)$.

Note that Problem 7.18 concerns the power of non-linear gates when computing linear operators $y = Ax$ over the XOR group. One can also ask a similar question for the OR semigroup: Do non-linear gates can help to simultaneously compute boolean sums? That is, if $\mathsf{GENOR}(A)$ denotes the smallest size of a *general* (not just OR) circuit simultaneously computing all the n ORs $y_i = \bigvee_{a_{ij}=1} x_j$, how large can the gap $\mathsf{OR}(A)/\mathsf{GENOR}(A)$ be?

Results above imply that this gap can be almost maximally possible, can be as large as $\Omega(n/\log^3 n)$. This follows from the gaps (5.11) between the OR complexities of matrices and their complements, together with a simple observation that $\mathsf{GENOR}(\overline{A}) \leqslant \mathsf{GENOR}(A) + 2n$. To see this upper bound, take a circuit for A (computing all ORs given

by A). To obtain a circuit for $\overline{A}$, just add one gate computing the OR h of all n variables, and replace each output gate y_i by $y_i \wedge \neg h$. Important in this construction was that we allow NOT gates: Nechiporuk [74], Mehlhorn [64], and Pippenger [84] have shown that every circuit with only OR and AND gates for a $(k+1)$-free matrix A must gave at least $|A|/k^3$ wires.

Finally, let us mention that the problems above are chosen from the perspective of seeking *incremental* improvements in this research area. The really important longer-term goals are much more ambitious:

- Can we find explicit rigid boolean matrices?
- Can we prove $\mathsf{XOR}_2(A) \geqslant n^{1+\Omega(1)}$ in depth 2?
- Can we prove a super-linear lower bound on $\mathsf{XOR}(A)$, at least when the depth is restricted to be logarithmic?

Acknowledgements

We would like to thank Alexander Chashkin, Andrew Drucker, Sergey Gashkov, Dmitri Grigoriev, and Avi Wigderson for useful comments and suggestions. Detailed comments of the anonymous referee greatly helped to improve the presentation.

Research of Igor Sergeev was supported by the Russian Foundation for Basic Research (grants no. 11–01–00508 and 11–01–00792), OMN RAN "Algebraic and combinatorial methods of mathematical cybernetics" program (project "Problems of optimal synthesis of control systems"). Research of Stasys Jukna was supported by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) grant SCHN 503/6-1.

References

[1] V.E. Alexeev. Two constructions of difference sets. *Problemy Kibernetiki*, 38:259–262, 1981 (in Russian). 15

[2] N. Alon. Eigenvalues, geometric expanders, sorting in rounds, and Ramsey theory. *Combinatorica*, 6(3):207–219, 1986. 48

[3] N. Alon, M. Karchmer, and A. Wigderson. Linear circuits over GF(2). *SIAM J. Comput.*, 19(6):1064–1067, 1990. 47, 47, 47, 91, 91

[4] N. Alon and W. Maass. Meanders and their applications in lower bounds arguments. *J. Comput. Syst. Sci.*, 37(2):118–129, 1988. 46, 49

[5] N. Alon and P. Pudlák. Superconcentrators of depths 2 and 3; odd levels help (rarely). *J. Comput. Syst. Sci.*, 48(1):194–202, 1994. 61, 97

[6] A.E. Andreev. On the complexity of realization of transitivity relations by rectifier circuits. In *Physical and mathematical modeling of discrete systems*, volume 56, pages 11–21. MEI, Moscow, 1985 (in Russian). 21

[7] A.E. Andreev. On a family of Boolean matrices. *Vestnik Moskow Univ.*, (2):97–100, 1986. Engl. transl. in: Moscow. Univ. Math. Bull. 1986. 41, 79–82. 15

[8] B. Bollobás. *Extremal graph theory.* Academic Press, 1978. 8

[9] R.C. Bose. An affine analogue of Singer's theorem. *J. Indian Math. Soc.*, 6:1–15, 1942. 15

[10] J. Boyar and M. Find. Cancellation-free circuits: An approach for proving superlinear lower bounds for linear boolean operators. Technical report, arXiv:1207.5321, 2012. 41, 62, 62

[11] J. Boyar and M. Find. Cancellation-free circuits in unbounded and bounded depth. Technical report, arXiv:1305:3041, 2013. 41, 62, 76

[12] S. Bublitz. Decomposition of graphs and monotone formula size of homogeneous functions. *Acta Inf.*, 23(6):689–696, 1986. 18

[13] P. Bürgisser, M. Clausen, and M.A. Shokrollahi. *Algebraic complexity theory*. Springer, 1997. 6, 113

[14] A.K. Chandra, S. Fortune, and R.J. Lipton. Unbounded fan-in circuits and associative functions. *J. Comput. Syst. Sci.*, 30(2):222–234, 1985. 61

[15] A.K. Chandra, S. Fortune, and R.L. Lipton. Lower bounds for constant depth circuits for prefix problems. In *Proc. in 10th Int. Colloq. on Automata, Languages and Programming (ICALP)*, volume 154 of *Springer Lect. Notes in Comput. Sci*, pages 109–117, 1983. 61

[16] A.V. Chashkin. Perfect linear hashing in boolean cube. In *Transactions on Discrete Mathematics and its Applications*, volume 5, pages 56–67. Keldysh Institute of Applied Mathematics, 2009 (in Russian). 102

[17] A.V. Chashkin. On linear operators injective on subsets of the space $\mathrm{GF}^n(p)$. *Discrete Mathematics and Applications*, 2013 (to appear). 102

[18] D.Yu. Cherukhin. On complexity of linear operators on the class of circuits of depth 2. *Diskretnaya Matematika*, 20(1):109–119, 2008. Engl. transl. in: Discrete Mathematics and Applications. 2008. 18(2), 143–154. 5, 91

[19] D.Yu. Cherukhin. Lower bounds for complexity of boolean circuits of finite depth with arbitrary elements. *Discrete Mathematics and Applications*, 21(4):499–508, 2011. 5

[20] B. Chor and O. Goldreich. Unbiased bits from sources of weak randomness and probabilistic communication complexity. *SIAM J. Comput.*, 17(2):230–261, 1988. 76

[21] F.R.K. Chung, P. Erdős, and J. Spencer. On the decomposition of graphs into complete bipartite graphs. In *Studies in Pure Mathematics, To the Memory of Paul Turán*, pages 95–101. Akadémiai Kiadó, 1983. 18

[22] D. Dolev, C. Dwork, N. Pippenger, and A. Wigderson. Superconcentrators, generalizers and generalized connectors with limited depth (preliminary version). In *Proc. of 15th Ann. ACM Symp. on Theory of Computing*, pages 42–51, 1983. 46, 61, 96, 97, 97, 102

[23] A. Drucker. Limitations of lower-bound methods for the wire complexity of boolean operators. In *IEEE Conference on Computational Complexity*, pages 170–180, 2012. Full version in ECCC Report Nr. 125, 2011. 94

[24] P.E. Dunne. *The complexity of Boolean networks*. Academic Press Professional, Inc., San Diego, CA, 1988. 6

[25] C. Dutta and J. Radhakrishnan. More on a problem of Zarankiewicz. In *Proc. of 23rd Int. Symp. on Algorithms and Computation, ISAAC 2012*, volume 7676 of *Springer Lect. Notes in Comput. Sci.*, pages 257–266, 2012. arXiv.1201.1377. 99, 100

[26] P. Erdős, R.L. Graham, and E. Szemerédi. On sparse graphs with dense long paths. In *Computers and Math. with Appl.*, pages 365–369. Pergamon, Oxford, 1976. 52

[27] P. Erdős and R. Rado. Intersection theorems for systems of sets. *J. London Math. Soc*, 35:85–90, 1960. 92

[28] P. Erdős and J. Spencer. *Probabilistic methods in combinatorics*. Academic Press, 1974. 13

[29] T. Feder and R. Motwani. Clique partitions, graph compression and speeding-up algorithms. *J. Comput. Syst. Sci.*, 51(2):261–272, 1995. 20

[30] M. Find, M. Göös, P. Kaski, J. Korhonen, M. Koivisto, and J.H. Korhonen. Separating OR, SUM, and XOR circuits. Technical report, arXiv.1304.0513, 2013. 24, 43, 69, 76, 108

[31] J. Friedman. A note on matrix rigidity. *Combinatorica*, 13(2):235–239, 1993. 58

[32] A. Gál. On the complexity of realization of some classes of matrices by rectifier networks. *Matematicheskije Voprosy Kibernetiki*, 1:234–235, 1988 (in Russian). 42

[33] A. Gál, K.A. Hansen, M. Koucký, P. Pudlák, and E. Viola. Tight bounds on computing error-correcting codes by bounded-depth circuits with arbitrary gates. In *STOC 2012*, pages 479–494, 2012. Full preliminary version in ECCC Report Nr. 150, 2011. 59, 96, 97, 98, 99, 101, 102, 106, 110

[34] S.B. Gashkov and I.S. Sergeev. On the complexity of linear boolean operators with thin matrices. *Diskretn. Anal. Issled. Oper.*, 17(3):3–18, 2010. Engl. transl.: J. Applied and Industrial Math. 2011. 5(2), 202–211. 77, 77, 78

[35] S.B. Gashkov and I.S. Sergeev. A method for deriving lower bounds for the complexity of monotone arithmetic circuits computing real polynomials. *Matematicheskii Sbornik*, 203(10):33–70, 2012. Engl. transl. in: Sbornik: Mathematics. 2012. 203(10), 1411–1447. 78

[36] O. Goldreich and A. Wigderson. On the circuit complexity of perfect hashing. In *Studies in Complexity and Cryptography*, pages 26–29. 2011. Preliminary version in ECCC Report Nr. 41, 1996. 102

[37] D.Yu. Grigoriev. An application of separability and independence notions for proving lower bounds of circuit complexity. In *Notes of the Scientific Seminar Leningrad branch of the Steklov Institute*, volume 60, pages 38–48. 1976. English translation in J. Soviet Math. 14:5 (1980), 1450–1456. 6, 45, 51

[38] D.Yu. Grigoriev. On a nonlinear lower bound for circuit complexity of a set of disjunctions in monotone boolean basis. In *Notes of the Scientific Seminar Leningrad branch of the Steklov Institute*, volume 68, pages 19–25. 1977. English translation in J. Soviet Math. 15:1 (1981), 11–13. 35

[39] D.Yu. Grigoriev. Additive complexity in directed computations. *Theoret. Comput Sci.*, 19:39–87, 1982. 41

[40] D.Yu. Grigoriev. Lower bounds in algebraic complexity. In *Notes of the Scientific Seminar Leningrad branch of the Steklov Institute*, volume 118, pages 25–82. 1982. English translation in J. Soviet Math. 29 (1985), 1388–1425. 51

[41] M.I. Grinchuk. On the complexity of realization of boolean triangular matrices by rectifier schemes of various depths. *Metody Diskretnogo Analiza*, 4:3–23, 1986 (in Russian). 61

[42] M.I. Grinchuk. Complexity of the realization of cyclic boolean matrices by gate circuits. *Izvestija VUZov. Matematika*, 7:39–44, 1988. English translation in Soviet Math. 32:7 (1988), 65–72. 78

[43] M.I. Grinchuk and I.S. Sergeev. Thin circulant matrices and lower bounds on the complexity of some boolean operators. *Diskretn. Anal. Issled. Oper.*, 18(5):38–53, 2011 (in Russian). 78

[44] G. Hansel. Nombre minimal de contacts de fermeture nécessaires pour réaliser une fonction booléenne symétrique de n variables. *C. R. Acad. Sci.*, 258(25):6037–6040, 1964 (in French). 30

[45] G.H. Hardy, J.E. Littlewood, and G. Polya. *Inequalities*. University Press Cambridge, 1934. 54

[46] A.P. Hiltgen. Towards a better understanding of one-wayness: facing linear permutations. In *EUROCRYPT*, volume 1403 of *Springer Lect. Notes in Comput. Sci.*, pages 319–333, 1998. 109

[47] S. Jukna. Disproving the single level conjecture. *SIAM J. Comput.*, 36(1):83–98, 2006. 74

[48] S. Jukna. Entropy of operators or why matrix multiplication is hard for depth-two circuits. *Theory of Computing Systems*, 46(2):301–310, 2010. 5

[49] S. Jukna. Representing (0,1)-matrices by depth-2 circuits with arbitrary gates. *Discrete Mathematics*, 310:184–187, 2010. 91, 92, 112

[50] S. Jukna. *Boolean Function Complexity.* Springer, 2012. 4, 6, 7, 10, 31, 79, 79, 110

[51] S. Jukna and G. Schnitger. Min-rank conjecture for log-depth circuits. *J. Comput. Syst. Sci.*, 77(6):1023–1038, 2011. 113

[52] G. Katona and E. Szemerédy. On a problem of graph theory. *Studia Scientiarum Mathematicarum Hungarica*, 2:23–28, 1967. 30

[53] N.H. Katz. On the CNF-complexity of bipartite graphs containing no squares. *Lithuanian Math. Journal*, 52(4):385–389, 2012. 85, 110

[54] M.M. Klawe. Shallow grates. *Theor. Comput. Sci.*, 123(2):389–395, 1994. 52

[55] D.E. Knuth. *Art of programming: seminumerical algorithms*, volume 2. Reading, Massachusetts: Addison-Wesley, 1997. Third Edition. 3

[56] V.V. Kochergin. On the complexity of rectifier networks with multiple number of paths. In *Materials of the 18-th International School on Synthesis and Complexity of Control Systems (Penza, 2009)*, pages 51–56, 2009 (in Russian). 26

[57] J. Kollár, L. Rónyai, and T. Szabó. Norm-graphs and bipartite Turán numbers. *Combinatorica*, 16(3):399–406, 1996. 15

[58] T. Kövari, V.T. Sós, and P. Turán. On a problem of K. Zarankiewicz. *Colloq. Math.*, 3:50–57, 1954. 8, 14

[59] R.E. Krichevski. Complexity of contact circuits realizing a function of logical algebra. *Doklady Akad. Nauk SSSR*, 151(4):803–806, 1963. English translation in Soviet Physics Doklady 8 (1963), 770–772. 30

[60] R. Lidl and H. Niederreiter. *Introduction to Finite Fields and their Applications.* Cambridge University Press, 1986. 15

[61] S.V. Lokam. Complexity lower bounds using linear algebra. *Foundations and Trends in Theoretical Computer Science*, 4(1-2):1–155, 2009. 6, 51

[62] O.B. Lupanov. On rectifier and switching-and-rectifier schemes. *Doklady Akad. Nauk SSSR*, 111:1171–1174, 1956 (in Russian). 3, 17, 17

[63] O.B. Lupanov. On rectifier schemes. *Acta Cybernetica*, 4(4):311—315, 1980 (in Russian). 6

[64] K. Mehlhorn. Some remarks on Boolean sums. *Acta Informatica*, 12:371–375, 1979. 32, 43, 114

[65] P. Miltersen. Error correcting codes, perfect hashing circuits, and deterministic dynamic dictionaries. In *SODA*, pages 556–563, 1998. 102

[66] B.S. Mitiagin and B.N. Sadovskiy. On linear Boolean operators. *Doklady Akad. Nauk SSSR*, 165(4):773–776, 1965 (in Russian). 76

[67] J. Morgenstern. Note on a lower bound of the linear complexity of the Fast Fourier Transform. *J. of the ACM*, 20(2):305–306, 1973. 6, 26

[68] J. Morgenstern. The linear complexity of computation. *J. of the ACM*, 22(2):184–194, 1974. 6

[69] R. Motwani and P. Raghavan. *Randomized Algorithms*. Cambridge University Press, 1995. 75

[70] E.I. Nechiporuk. On multipolar switching circuits realizing functions of multi-valued logics. *Problemy Kybernetiki*, 5:49–60, 1961 (in Russian). 3

[71] E.I. Nechiporuk. Rectifier networks. *Doklady Akad. Nauk SSSR*, 148(1):50–53, 1963. English translation in: Soviet Physics Doklady 8 (1963), 5–7. 18, 19

[72] E.I. Nechiporuk. On the synthesis of rectifier networks. *Problemy Kibernetiki*, 9:37–44, 1963 (in Russian). 20, 20

[73] E.I. Nechiporuk. Self-correcting diode networks. *Doklady Akad. Nauk SSSR*, 156(5):1045–1048, 1964. English translation in: Soviet Physics Doklady 9(6) (1964), 422–425. 20, 32

[74] E.I. Nechiporuk. On a boolean matrix. *Problemy Kibernetiki*, 21:237—240, 1969. English translation in: Systems Theory Res. 21 (1970), 236–239. 14, 114

[75] E.I. Nechiporuk. On the topological principles of self-correction. *Problemy Kibernetiki*, 21:5–102, 1969. English translation in: Systems Theory Res. 21 (1970), 1–99. 18, 19, 20, 20, 32, 32, 34

[76] I. Newman and A. Wigderson. Lower bounds on formula size of boolean functions using hypergraph entropy. *SIAM J. Discrete Math.*, 8(4):536–542, 1995. 31

[77] K. O'Bryant. A complete annotated bibliography of work related to Sidon sequences. *Electronic Journal of Combinatorics*, 11:1–39, 2004. 15

[78] V.A. Orlov. Realization of "narrow" matrices by rectifier networks. *Problemy Kybernetiki*, 22:45–52, 1970. English translation in: *Systems Theory Research*, 22, 42–50, 1972. 18

[79] H. Perfect. Applications of Menger's graph theorem. *Journal of Mathematical Analysis and Applications*, 22:96–111, 1968. 98

[80] T. Pinto. Biclique covers and partitions. Technical report, arXiv:1307.6363, 2013. 71

[81] N. Pippenger. On the evaluation of powers and related problems. In *FOCS*, pages 258–263, 1976. 3, 20, 108

[82] N. Pippenger. Superconcentrators. *SIAM J. Comput.*, 6(2):298–304, 1977. 46, 97

[83] N. Pippenger. The minimum number of edges in graphs with prescribed paths. *Math. Syst. Theory*, 12(1):325–346, 1979. 19, 20, 108

[84] N. Pippenger. On another Boolean matrix. *Theor. Comput. Sci.*, 11:49–56, 1980. 32, 33, 114

[85] N. Pippenger. Superconcentrators of depth 2. *J. Comput. Syst. Sci.*, 24(1):82–90, 1982. 97

[86] P. Pudlák. Communication in bounded depth circuits. *Combinatorica*, 14(2):203–216, 1994. 54, 54, 55, 56, 61, 96, 97, 102

[87] P. Pudlák. A note on the use of determinant for proving lower bounds on the size of linear circuits. *Inf. Process. Letters*, 74:197–201, 2000. 28

[88] P. Pudlák and V. Rödl. Some combinatorial-algebraic problems from complexity theory. *Discrete Math.*, 136(1-3):253–279, 1994. 21, 21, 58, 74

[89] P. Pudlák, V. Rödl, and P. Savický. Graph complexity. *Acta Inf.*, 25(5):515–535, 1988. 110

[90] P. Pudlák and Z. Vavrín. Computation of rigidity of order n^2/r for one simple matrix. *Comm. Math. Univ. Carol.*, 32(2):213–218, 1991. 56, 56

[91] J. Radhakrishnan. Entropy and counting. Manuscript, available at http://www.tcs.tifr.res.in/∼jaikumar/mypage.html, 2001. 30

[92] J. Radhakrishnan and A. Ta-Shma. Bounds for dispersers, extractors, and depth-two superconcentrators. *SIAM J. Discrete Math.*, 13(1):2–24, 2000. 97, 99, 100

[93] A.A. Razborov. Lower bounds on the size of bounded-depth networks over a complete basis with logical addition. *Mat. Zametki*, 41(4):598–607, 1987. Engl. transl. in: Math. Notes of the Acad. of Sci. of the USSR 41 (1987), 333–338. 79

[94] I.Z. Ruzsa. Solving a linear equation in a set of integers. I. *Acta Arith.*, 65:259–282, 1993. 15

[95] G. Schnitger. A family of graphs with expensive depth-reduction. *Theor. Comput. Sci.*, 18:89–93, 1982. 53

[96] G. Schnitger. On depth-reduction and grates. In *24th IEEE Ann. Symp. on Foundations of Comput. Sci.*, pages 323–328, 1983. 52, 53

[97] A. Schönhage. Schnelle multiplikation von Polynomen über Körpern der Charakteristik 2. *Acta Inf.*, 7:395–398, 1977 (in German). 77

[98] S.N. Selezneva. Lower bound on the complexity of finding polynomials of Boolean functions in the class of circuits with separated variables. In *Proc. of 11-th Int. Seminar on Discrete Math. and Its Appl. (Moscow, June 2012)*, pages 216–218, 2012. Journal version in: Computational Mathematics and Modeling, Consultants Bureau (United States), 24(1), 146–152, 2013. 62, 62

[99] I.S. Sergeev. Implementation of linear maps with circulant matrices via modulo 2 rectifier circuits of bounded depth. Technical report, arXiv.1305.4389, 2013. 77

[100] J. Singer. A theorem in finite projective geometry and some applications to number theory. *Trans. Amer. Math. Soc.*, 43(3):377–385, 1938. 14, 14, 14, 14

[101] D.A. Spielman. Linear-time encodable and decodable error-correcting codes. *IEEE Transactions on Information Theory*, 42(6):1723–1731, 1996. 57, 97, 108

[102] M. Sudan. Essential coding theory, 2002. Manuscript, available at http://research.microsoft.com/en-us/um/people/madhu/. 103

[103] T.G. Tarjan. Complexity of lattice-configurations. *Stud. Sci. Math. Hung.*, 10:203–211, 1975. 31, 60, 64

[104] Z. Tuza. Covering of graphs by complete bipartite subgraphs; complexity of 0-1 matrices. *Combinatorica*, 4(1):111–116, 1984. 18

[105] L.G. Valiant. Graph-theoretic properties in computational complexity. *J. Comput. Syst. Sci.*, 13(3):278–285, 1976. 46, 97

[106] L.G. Valiant. Graph-theoretic methods in low-level complexity. In *Springer Lect. Notes in Comput. Sci.*, volume 53, pages 162–176, 1977. 6, 45, 51, 51, 51, 52, 52, 53, 53, 110, 110

[107] I. Wegener. A new lower bound on the monotone network complexity of Boolean sums. *Acta Inform.*, 15:147–152, 1980. 32

[108] I. Wegener. *The complexity of Boolean functions*. Wiley-Teubner, 1987. 6

[109] D.J. Welsh. *Matroid theory*. Academic Press, London, 1976. 98

[110] A. Wigderson. P, NP and mathematics - a computational complexity perspective. In *Proceedings of the ICM 06 (Madrid)*, volume 1, pages 665–712. EMS Publishing House, Zurich, 2007. 111