

Cancellation-Free Circuits in Unbounded and Bounded Depth [☆]

Magnus Gausdal Find¹, Joan Boyar²

*Department of Mathematics and Computer Science, University of Southern Denmark,
Denmark*

Abstract

We study the notion of “cancellation-free” circuits. This is a restriction of XOR circuits, but can be considered as being equivalent to previously studied models of computation. The notion was coined by Boyar and Peralta in a study of heuristics for a particular circuit minimization problem. They asked how large a gap there can be between the smallest cancellation-free circuit and the smallest XOR circuit. We present a new proof showing that the difference can be a factor $\Omega(n/\log^2 n)$. Furthermore, our proof holds for circuits of constant depth. We also study the complexity of computing the Sierpinski matrix using cancellation-free circuits and give a tight $\Omega(n \log(n))$ lower bound.

Keywords: Circuit Complexity, Cancellation-free, Linear Circuits

1. Introduction

Let \mathbb{F}_2 be the field of order 2, and let \mathbb{F}_2^n be the n -dimensional vector space over \mathbb{F}_2 . For $n \in \mathbb{N}$, we let $[n] = \{1, \dots, n\}$. A Boolean function $f: \mathbb{F}_2^n \rightarrow \mathbb{F}_2^m$ is said to be linear if there exists a Boolean $m \times n$ matrix A such that $f(\mathbf{x}) = A\mathbf{x}$ for every $\mathbf{x} \in \mathbb{F}_2^n$. This is equivalent to saying that f can be computed using only XOR gates.

An *XOR circuit* (or a *linear circuit*) C is a directed acyclic graph. There are n nodes with in-degree 0, called the *inputs*. All other nodes have in-

[☆]A preliminary version of this paper appears in [1]. Both authors are partially supported by the Danish Council for Independent Research, Natural Sciences.

Email addresses: `magnusgf@imada.sdu.dk` (Magnus Gausdal Find),
`joan@imada.sdu.dk` (Joan Boyar)

¹Part of this work was done while visiting the University of Toronto.

²Part of this work was done while visiting the University of Waterloo.

degree 2 and are called *gates*. There are m nodes which are called the *outputs*; these are labeled y_1, \dots, y_m . The value of a gate is the sum of its two children (addition in \mathbb{F}_2 , denoted \oplus). The circuit C , with inputs $\mathbf{x} = (x_1, \dots, x_n)$, computes the $m \times n$ matrix A if the output vector computed by C , $\mathbf{y} = (y_1, \dots, y_m)$, satisfies $\mathbf{y} = A\mathbf{x}$. In other words, output y_i is defined by the i th row of the matrix. The *size* of a circuit C , is the number of gates in C . The *depth* is the number of gates on a longest directed path from an input to an output. For simplicity, we will let $m = n$ unless otherwise explicitly stated. For a matrix A , let $|A|$ be the number of nonzero entries in A .

Our contributions: In this paper we deal with a restriction of XOR circuits called *cancellation-free* circuits, coined in [2], where the authors noticed that many heuristics for finding small XOR circuits always produce cancellation-free XOR circuits. They asked the question of how large a separation there can be between these two models. Recently, Gashkov and Sergeev [3] showed that the work of Grinchuk and Sergeev [4] implied a separation of $\Omega\left(\frac{n}{\log^6 n \log \log n}\right)$. An improved separation of $\Omega\left(\frac{n}{\log^2 n}\right)$ follows from Lemma 4.1 and Lemma 4.2 in [5], although this implied separation was not published until recently [6]. We present an alternative proof of the same separation. Our proof is based on a different construction and uses communication complexity in a novel way that might have independent interest. Like the separation implied in the work [6], but unlike the separations demonstrated in [3, 7], our separation holds even in the case of constant depth circuits. We conclude that many heuristics for finding XOR circuits do not approximate better than a factor of $\Theta\left(\frac{n}{\log^2 n}\right)$ of the optimal. We also study the complexity of computing the Sierpinski matrix using cancellation-free circuits. We show that the complexity is exactly $\frac{1}{2}n \log n$. Furthermore, our proof holds for OR circuits. As a corollary to this we obtain an explicit matrix where the smallest OR circuit is a factor $\Theta(\log n)$ larger than the smallest OR circuit for its complement.

We also study the complexity of computing the *Sierpinski matrix* (described later), and show a tight $\frac{1}{2}n \log n$ lower bound for OR circuits and cancellation-free circuits. This results follows implicitly from the work of Kennes [8], however our proof is simpler and more direct. Also we hope that our proof can be strengthened to give an $\omega(n)$ lower bound for XOR circuits for the Sierpinski matrix. A similar lower bound was shown independently by Selezneva in [9, 10].

2. Cancellation-Free XOR Circuits

For XOR circuits, the value computed by every gate is the parity of a subset of the n variables. That is, the output of every gate u can be considered as a vector $\kappa(u)$ in the vector space \mathbb{F}_2^n , where $\kappa(u)_i = 1$ if and only if x_i is a term in the parity function computed by the gate u . We call $\kappa(u)$ the *value vector* of u , and for input variables define $\kappa(x_i) = e^{(i)}$, the unit vector having the i th coordinate 1 and all others 0. It is clear by definition that if a gate u has the two children w, t , then $\kappa(u) = \kappa(w) \oplus \kappa(t)$, where \oplus denotes coordinate-wise addition in \mathbb{F}_2 . We say that an XOR circuit is *cancellation-free* if for every pair of gates u, w where u is an ancestor of w , then $\kappa(u) \geq \kappa(w)$, where \geq denotes the usual coordinate-wise partial order. These are also called SUM circuits in [7, 6].

If this is satisfied, the circuit never exploits the \mathbb{F}_2 -identity, $a \oplus a = 0$, so things do not “cancel out” in the circuit.

Although it is not hard to see that cancellation-free circuits is equivalent to addition chains [11, 12] and “ensemble computations” [13], we stick to the term “cancellation-free”, since we will think of it as a special case of XOR circuits.

For a simple example demonstrating that cancellation-free circuits indeed are less powerful than general XOR circuits, consider the matrix

$$A = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \end{pmatrix}.$$

In Figure 1, two circuits computing the matrix A are shown, the circuit on the right uses cancellations, and the circuit on the left is cancellation-free, and has one gate more. For this particular matrix, any cancellation-free circuit must use at least 5 gates.

A different, but tightly related kind of circuits is OR circuits. The definition is exactly the same as for XOR circuits, but with \vee (logical OR) instead of \oplus , see [14, 6, 13]. Cancellation-free circuits is a special case of OR circuits and every cancellation-free circuit can be interpreted as an OR circuit for the same matrix, as well as an XOR circuit.

For a matrix A , we will let $C_{\oplus}(A)$, $C_{CF}(A)$, $C_{\vee}(A)$ denote the smallest XOR circuit, the smallest cancellation-free circuit and the smallest OR circuit computing the matrix A .

By the discussion above, the following is immediate:

Proposition 1. *For every matrix, A , $C_{\vee}(A) \leq C_{CF}(A)$.*

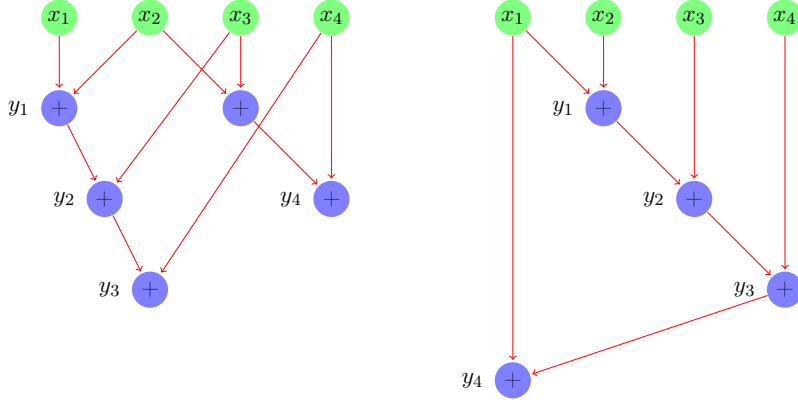


Figure 1: Two circuits computing the matrix A . The circuit on the left is cancellation-free, and has size 5 - one more more than the circuit on to the right.

This means in particular that any lower bound for OR circuits carries over to a lower bound for cancellation-free circuits. However, the converse does not hold in general [7]. A simple example showing this is the matrix

$$B = \begin{pmatrix} 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}.$$

For this matrix, there exist an OR circuit with 6 gates, however any cancellation-free circuit must have at least 7 gates.

Every matrix admits a cancellation-free circuit of size at most $n(n-1)$. This can be obtained simply by computing each row independently. It was shown by Nechiporuk [14] and Pippenger [11] (see also [6]) that this upper bound can be improved to $(1 + o(1)) \frac{n^2}{2 \log n}$.

A Shannon-style counting argument gives that this is tight up to low order terms. A proof of this can be found in [11]. Combining these results, we get that for most matrices, cancellation does not help much:

Theorem 1. *For every $0 < \epsilon < 1$, for sufficiently large n , a random $n \times n$ matrix has $\frac{C_{CF}(A)}{C_{\oplus}(A)} \leq 1 + \epsilon$ with probability at least $1 - \epsilon$.*

We also use the following upper bound, which also holds for cancellation-free circuits, and hence also for OR circuits and XOR circuits.

Theorem 2 (Lupanov [15]). *Any $m \times n$ matrix, admits a cancellation-free XOR circuit of size $O\left(\min\left\{\frac{mn}{\log n}, \frac{mn}{\log m}\right\} + n + m\right)$.*

The theorem follows directly from Lupanov’s result and an application of the “transposition principle” (see e.g. [16]).

A matrix A is k -free if it does not have an all one submatrix of size $(k+1) \times (k+1)$. The following lemma will be used later. According to Jukna and Sergeev [6], it was independently due to Nechiporuk [17], Mehlhorn [18], Pippenger [19], and Wegener [20].

Lemma 1 (Nechiporuk, Mehlhorn, Pippenger, Wegener). *For k -free A , $C_{\vee}(A) \in \Omega\left(\frac{|A|}{k^2}\right)$.*

3. Relationship Between Cancellation-Free XOR Circuits and General XOR Circuits

In [2], Boyar and Peralta exhibited an infinite family of matrices where the sizes of the cancellation-free circuits computing them are at least $\frac{3}{2} - o(1)$ times the corresponding sizes for smallest XOR circuits for them. We call this ratio the *cancellation ratio*, $\rho(n)$, defined as

$$\rho(n) = \max_{A \in \mathbb{F}_2^{n \times n}} \frac{C_{CF}(A)}{C_{\oplus}(A)}.$$

The following proposition on the Boolean Sylvester-Hadamard matrix was pointed out by Edward Hirsch and Olga Melanich [21]. The $n \times n$ Boolean Sylvester-Hadamard matrix H_n , is defined recursively:

$$H_1 = (1), H_{2n} = \begin{pmatrix} H_n & H_n \\ H_n & \overline{H_n} \end{pmatrix}$$

Where \overline{A} means the Boolean complement of the matrix A . It is known that $C_{\oplus}(H_n) \in O(n)$, but that in depth 2 it requires circuits of size $\Omega(n \log n)$ [22].

Proposition 2. *The $n \times n$ Boolean Sylvester-Hadamard matrix requires cancellation-free circuits of size $C_{CF}(H_n) \in \Omega(n \log n)$.*

Since $\log |\det(H_n)| \in \Omega(n \log n)$, this proposition follows from following theorem due to Morgenstern, ([23], see also [24, Thm. 13.14]).

Theorem 3 (Morgenstern). *For a Boolean matrix M ,*

$$C_{CF} \in \Omega(\log |\det(M)|).$$

The statement holds more generally, namely for circuits with addition over the complex numbers and scalar multiplication by any constant $c \in \mathbb{C}$ with $|c| \leq 2$. Cancellation-free circuits can be seen as a special case of this.

Using the recursive structure of H_n , it is not hard to show that $C_{\oplus}(H_n) \in O(n)$, so this demonstrates that $\rho(n) \in \Omega(\log n)$. It should be noted that no $n \times n$ Boolean matrix can have determinant larger than $n!$, so this technique cannot give lower a bound on $\rho(n)$ stronger than $O(\log n)$.

As mentioned in the introduction, the ratio

$$\lambda(n) = \max_{A \in \mathbb{F}_2^{n \times n}} \frac{C_{\vee}(A)}{C_{\oplus}(A)}$$

has been studied, (see [3, 6]). Using the techniques of [5], it can be derived (as is done in [6]) that $\lambda(n) \in \Omega(n/\log^2 n)$.

We present a different construction exhibiting the same gap. The construction is different, and in some sense simpler. Furthermore our proof is quite different. More concretely we use communication complexity for the analysis to show that certain conditional random variables are almost uniformly distributed in a way that might have independent interest. Also our construction gives a similar separation for circuits of constant depth (see Section 5).

Theorem 4. $\lambda(n) \in \Omega\left(\frac{n}{\log^2 n}\right)$.

The proof uses the probabilistic method. We construct randomly two matrices, and let A be their product. In order to use Lemma 1 on A , we need to show that with high probability, the matrix A will be $2 \log n$ -free. We do this via Lemma 2 by showing that the marginal distribution of any entry in a fixed $2 \log n \times 2 \log n$ submatrix is almost uniformly random.

In the following, for a matrix M , we let M_i (M^i) denote its i th row (column). And for $I \subseteq [n]$, we let M_I (M^I) denote the submatrix consisting of the rows (columns) with indices in I .

Lemma 2 might seem somewhat technical. However, there is a very simple intuition behind it: Suppose M is obtained at random as in the

statement of the lemma. Informally we want to say that the entries do not “depend” too much on each other. More formally we want to show that given all but one entry in M it is not possible to guess the last entry with significant advantage over random guessing. The proof idea is to transform any good guess into a deterministic communication protocol for computation of the inner product, and to use a well known limitation on how well this can be done [25, 26].

We will say that two (partially) defined matrices are *consistent* if they agree on all their defined entries.

Lemma 2. *Let M be an $m \times m$ partially defined matrix, where all entries except M_p^q are defined. Let B, C be matrices over \mathbb{F}_2 with dimensions $m \times 7m$ and $7m \times m$ respectively, be uniformly random among all possible pairs (B, C) such that BC is consistent with M .*

Then for sufficiently large m , the conditional probability that M_p^q is 1, given all other entries, is contained in the interval $(\frac{1}{2} - \frac{1}{m}, \frac{1}{2} + \frac{1}{m})$, where the probability is over the choices of B and C .

Before proving the lemma, we will first recall a fact from communication complexity, due to Chor and Goldreich [27], see also [26].

Theorem 5 (Chor, Goldreich). *Let \mathbf{x} and \mathbf{y} be independent and uniformly random vectors, each of n bits. Suppose a deterministic communication protocol is used to compute the inner product of \mathbf{x} and \mathbf{y} , and the protocol is correct with probability at least $\frac{1}{2} + p$. Then on some inputs, the protocol uses $\frac{n}{2} - \log(1/p)$ bits of communication.*

PROOF (OF LEMMA 2). Suppose for the sake of contradiction that there exists a partially defined matrix M , such that when all entries but one are revealed, the conditional probability of the last entry being a is at least $\frac{1}{2} + \frac{1}{m}$ for some $a \in \{0, 1\}$.

Assuming this, we will first present a randomized communication protocol computing the inner product of two independent and uniformly random $7m$ bit vectors \mathbf{x} and \mathbf{y} that always uses m bits of communication and is correct with probability at least $\frac{1}{2} + \frac{2^{-2m}}{4m}$. We will then argue that this protocol can be completely derandomized. This results in a deterministic communication protocol that violates Theorem 5. From this we conclude that such a partially defined matrix, with this large probability of the last entry being a , does not exist.

Let Alice and Bob have as input vectors \mathbf{x} and \mathbf{y} , respectively, each of length $7m$. Before getting their inputs, they use their shared random bits

to agree on a random choice of the two matrices B and C distributed as stated in the Lemma. To compute the inner product of \mathbf{x} and \mathbf{y} , Alice replaces the row B_p with \mathbf{x} and Bob replaces the column C^q with \mathbf{y} , let the resulting matrices be B' and C' . Let $M' = B'C'$. Notice that M and M' are consistent, except possibly on row p and column q . Alice can compute the entire p th row of M' (except $(M')_p^q$). Similarly Bob can compute the entire q th column (except $(M')_p^q$). The communication in the protocol consists of first letting Alice send the $m - 1$ bits in the part of the p th row she can compute to Bob. Bob now knows all the entries in M' , except the entry M_p^q .

In order for M' and M to be consistent, it is only necessary that the $m - 1$ defined entries in row p and the $m - 1$ defined entries in column q are equal in the two matrices, since B' and C' were defined such that all other entries were equal. This occurs with probability at least 2^{-2m-2} .

In this case, the value Alice and Bob want to compute is exactly the only unknown entry M_p^q . By assumption, this last entry is a with probability at least $\frac{1}{2} + \frac{1}{m}$, so Bob outputs a . If the known entries in M' are not consistent with the known entries in M , Bob outputs a uniformly random bit. This is correct with probability $\frac{1}{2}$. Thus, the probability of this protocol being correct is at least:

$$\begin{aligned} & 2^{-2m-2} \left(\frac{1}{2} + \frac{1}{m} \right) + (1 - 2^{-2m-2}) \frac{1}{2} \\ = & 1/2 + \frac{2^{-2m}}{4m} \end{aligned}$$

So when the inputs are uniformly distributed, the randomized protocol computes the inner product of two $7m$ bits vectors with m bits communication, and it is correct with probability at least $\frac{1}{2} + \frac{2^{-2m}}{4m}$. By an averaging argument it follows that there exist a deterministic communication protocol with the same success probability. According to Theorem 5, any deterministic algorithm for computing the inner product with this success probability must communicate at least

$$\frac{7m}{2} - \log(1/p) = \frac{7}{2}m - \log \left(\frac{4m}{2^{-2m}} \right) = \frac{3}{2}m - \log m - 2$$

Which is larger than m for sufficiently large values of m ($m \geq 16$ suffices), and we arrive at the desired contradiction. \square

We now use this to prove Theorem 4. We will use following result on the ‘‘Zarankiewicz problem’’ [28], see also [29].

Theorem 6 (Kovári, Sós, Turán). *Let M be an $(a-1)$ -free $n \times n$ matrix. Then the number of ones in M is at most $(a-1)^{1/a} n^{2-1/a} + (a-1)n$.*

PROOF (OF THEOREM 4). We will probabilistically construct two matrices B, C of dimensions $n \times 14 \log n$, $14 \log n \times n$. Each entry in B and C will be chosen independently and uniformly at random on \mathbb{F}_2 . We let $A = BC$. First notice that it follows directly from Theorem 2 that B and C can be computed with XOR circuits, both of size $O(n)$. Now we can let the outputs of the circuit computing C be the inputs of the circuit computing B . Notice that this composed circuit will have many cancellations. The resulting circuit computes the matrix A and has size $O(n)$. We will argue that with probability $1 - o(1)$ this matrix will not have a $2 \log n \times 2 \log n$ submatrix of all ones, while $|A| \in \Omega(n^2)$. By Lemma 1 the results follows.

We show that for large enough n , with high probability neither of the following two events will happen:

1. BC has a submatrix of dimension $2 \log n \times 2 \log n$ consisting of all ones or all zeros
2. $|BC| \leq 0.3n^2$

1). Fix a submatrix M of BC with dimensions $2 \log n \times 2 \log n$. That is, some subset I of the rows of B , and a subset J of the columns in C so $M = B_I C^J$. We now want to show that the probability of this matrix having only ones (or zeros) is so small that a union bound over all choices of $2 \log n \times 2 \log n$ submatrices gives that the probability that there exists such a submatrix goes to 0. Notice that this would be easy if all the entries in M were mutually independent and uniformly distributed.

Although this is not case, Lemma 2 for $m = 2 \log n$ states, that this is almost the case. More precisely, the conditional probability that a given entry is 1 (or 0) is at most $\frac{1}{2} + \frac{1}{2 \log n}$. We can now use the union bound to estimate the probability that A has a submatrix of dimension $2 \log n \times 2 \log n$ with all the entries being either 0 or 1:

$$\begin{aligned} 2 \binom{n}{2 \log n}^2 \left(\frac{1}{2} + \frac{1}{2 \log n} \right)^{4 \log^2 n} &\leq 2 \frac{n^{4 \log n}}{(2 \log n)!} \left(\frac{1 + \frac{1}{\log n}}{2} \right)^{4 \log^2 n} \\ &\leq 2 \left(\frac{\left(1 + \frac{1}{\log n} \right)^{4 \log^2 n}}{(2 \log n)!} \right) \end{aligned}$$

This tends to 0, so we arrive at the desired result.

2). Note that if one wants to show that with positive probability the number of ones is $\Omega(n^2)$, a straightforward application of Markov's inequality suffices. Here we will show the stronger statement that with probability $1 - o(1)$, the number of ones is at least $\frac{n^2}{2} - o(n^2)$. By the proof above, we may assume that the Boolean complement of A , \bar{A} , does not have a $2 \log n$ submatrix of all ones. By Theorem 6, the number of ones in \bar{A} is at most

$$(2 \log n - 1)^{1/2 \log n} n^{2-1/2 \log n} + (2 \log n - 1)n$$

One can verify that

$$\lim_{n \rightarrow \infty} \frac{(2 \log n - 1)^{1/2 \log n} n^{2-1/2 \log n} + (2 \log n - 1)n}{n^2} = \frac{1}{\sqrt{2}}$$

So if there is not a $2 \log n \times 2 \log n$ matrix of all zeros in A , the number of zeros in A is at most $n^2(1 - \frac{1}{\sqrt{2}}) < 0.3n^2$. Hence the probability of $|A|$ being less than $0.3n^2$ tends to 0.

□

Remark 1: It has been pointed out by Avishay Tal that in order to show that the matrix is $O(\log n)$ -free, a significantly simpler argument suffices. We present it here: Let B, C be random matrices as in the construction of Theorem 4 but with dimensions $n \times 5 \log n$ and $5 \log n \times n$, respectively, and let $A = BC$. Now any $5 \log n \times 5 \log n$ submatrix of A is a product of two $5 \log n \times 5 \log n$ dimensional matrices, one being a submatrix of B and one being a submatrix of C . Now recall the theorem from linear algebra:

Theorem 7 (Sylvester's Rank Inequality). *For two $m \times m$ matrices B, C*

$$\text{rank}(BC) \geq \text{rank}(B) + \text{rank}(C) - m$$

The probability that a random $k \times k$ matrix has rank less than d is at most $2^{k-(k-d)^2}$ (see e.g. the proof of Lemma 5.4 in [30]). Now a union bound shows that the probability that there is a $5 \log n \times 5 \log n$ submatrix of B or C with rank smaller than $0.51 \cdot 5 \log n$ tends to 0. So for large enough n , with high probability, every $5 \log n \times 5 \log n$ of A will have rank at least $0.02 \cdot 5 \log n$. A submatrix consisting of all ones or all zeros has rank 0 or 1, which is less than $0.1 \log n$ for large enough n . Thus, the probability of this occurring tends to zero.

In the matrix constructed in [6, Theorem 5.8], they highlight the property that the matrix is *t-Ramsey*, meaning that both the matrix and its

complement are $(t - 1)$ -free, and it is a somewhat interesting fact that such matrices admit small XOR circuits. It follows immediately from the proof of Theorem 4 that this holds as well for the matrix constructed, and we state this a separate corollary.

Corollary 1. *For large enough n , with high probability, the bipartite graph with adjacency matrix A from Theorem 4 is t -Ramsey for $t = 2 \log n$.*

Notice that by Theorem 2, the obtained separation is at most a factor of $O(\log n)$ from being optimal. Also, except for lower bounds based on counting, all strong lower bounds we know of are essentially based on Lemma 1. Following that line of thought, one might hope to improve the separation above by coming up with a better choice of A that does not have a $O(\log^{1-\epsilon} n) \times O(\log^{1-\epsilon} n)$ all 1 submatrix to get a stronger lower bound on $C_\vee(A)$, or perhaps hope that a tighter analysis than the above would give a stronger separation. However, this direction does not seem promising. To see this, it follows from Theorem 6 that a matrix without a $\log^{1-\epsilon} n \times \log^{1-\epsilon} n$ all 1 submatrix, the lower bound obtained using Lemma 1 would be of order $O\left(\frac{n^{2-\frac{1}{\log^{1-\epsilon} n}}}{(\log^{1-\epsilon} n)^2}\right)$, which is $o\left(\frac{n^2}{\log^2 n}\right)$.

4. Smallest XOR Circuit Problem

As mentioned earlier, the notion cancellation-free was introduced by Bonyar and Peralta in [2]. The paper concerns shortest straight line programs for computing linear forms, which is equivalent to the model studied in this paper. In [13], it is shown that the Ensemble Computation Problem (recall that this is equivalent to cancellation-free) is **NP**-complete. For general XOR circuits, the problem remains **NP**-complete [2]. It was observed in [2] that several researchers have used heuristics that will always produce cancellation-free circuits, see [31, 32, 33]. By definition, any heuristic which only produces cancellation-free circuits cannot achieve an approximation ratio better than $\rho(n)$. By Proposition 1, $\rho(n) \geq \lambda(n)$. Thus, Theorem 4 implies that techniques which only produce cancellation-free circuits are not guaranteed to be very close to optimal.

Corollary 2. *The algorithms in [31, 32, 33] do not guarantee approximation ratios better than $\Theta\left(\frac{n}{\log^2 n}\right)$.*

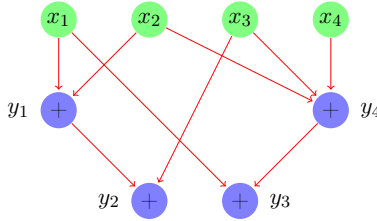


Figure 2: An example of a depth 2 circuit, computing the same matrix as the circuits in Figure 1. Notice that some gates have fan-in larger than 2. This circuit has size 9.

5. Constant Depth

For unbounded depth, there is no known family of (polynomial time computable) matrices known to require XOR circuits of superlinear size. However, if one puts restrictions on the depth, superlinear lower bounds are known [16]. In this case, we allow each gate to have unbounded fan-in, and instead of counting the number of gates we count the number of wires in the circuit. See Figure 2 for an example of a depth two circuit.

In particular, the circuit model where the depth is bounded to be at most 2 is well studied (see e.g. [16]). Similarly to previously, an XOR circuit in depth 2 is a circuit where each gate computes the XOR of its inputs. When considering matrices computed by XOR circuits, the general situation in the two circuit models is very similar. The following two results are due to Lupanov [15], see also [16].

Theorem 8 (Lupanov). *For every $n \times n$ matrix A , there exists a depth 2 cancellation-free circuit with at most $O\left(\frac{n^2}{\log n}\right)$ wires computing A . Furthermore, almost every such matrix requires $\Omega\left(\frac{n^2}{\log n}\right)$ wires.*

Let $\lambda^d(n)$ denote $\lambda(n)$ for circuits restricted to depth d (recall that now size is defined as the number of wires). Neither the separation in [3] nor that

in [7] seems to carry over to bounded depth circuits in any obvious way. The separation presented in [6, Theorem 5.8] holds for any depth $d \geq 2$.

By inspecting the proof of Theorem 4, the upper bound on the size of the XOR circuit worked as follows: First construct a circuit to compute C , and then construct a circuit for B with the outputs of C as inputs, that is, a circuit for B that comes topologically after C . To get to an upper bound of $O(n)$ wires, we use Theorem 2. By using Theorem 8 twice, we get a depth 4 circuit of that size.

For depths $d = 2$ and $d = 3$, one can use arguments similar to those in given in the proof of [6, Theorem 5.8]) to show that the separation still holds in these two cases. We summarize this in the following theorem.

Theorem 9. *Let $d \geq 2$. $\lambda^d(n) \in \Omega\left(\frac{n}{\log^2 n}\right)$.*

6. Computing the Sierpinski Matrix

In this section we prove that the $n \times n$ Sierpinski matrix, S_n , needs $\frac{1}{2}n \log n$ gates when computed by a cancellation-free circuit, and that this suffices. The proof strategy is surprisingly simple, it is essentially gate elimination where more than one gate is eliminated in each step. Neither Theorem 3 nor Lemma 1 gives anything nontrivial for this matrix.

As mentioned previously, there is no known (polynomial time computable) family of matrices requiring XOR circuits of superlinear size. However there are simple matrices that are conjectured to require circuits of size $\Omega(n \log n)$. One such matrix is the Sierpinski matrix, (Aaronson, personal communication and [34]). The $n \times n$ Sierpinski (also called *set disjointness*) matrix, S_n , is defined inductively

$$S_2 = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}, S_{2n} = \begin{pmatrix} S_n & 0 \\ S_n & S_n \end{pmatrix}$$

Independently of this conjecture, Jukna and Sergeev [6, Problem 7.11] have very recently asked if the “set intersection matrix”, K_n , has $C_{\oplus}(K_n) \in \omega(n)$. The motivation for this is that $C_V(K_n) \in O(n)$, so if true this would give a counterpart to Theorem 4.

If n is a power of two, the $n \times n$ set intersection matrix K_n can be defined by associating each row and column with a subset of $[\log n]$, and letting an entry be 1 if and only if the corresponding row and column sets have non-empty intersection. One can also define K_n inductively:

$$K_2 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, K_{2n} = \begin{pmatrix} K_n & K_n \\ K_n & J \end{pmatrix},$$

where J is the $n \times n$ matrix with 1 in each entry. It is easy to see that up to a reordering of the columns, the complement of K_n contains exactly the same rows as S_n . Thus, $C_{\oplus}(K_n)$ is superlinear if and only if $C_{\oplus}(S_n)$ is, since either matrix can be computed from the other with at most $2n - 1$ extra XOR gates, using cancellation heavily.

To see that the set intersection matrix can be computed with OR circuits of linear size observe that over the Boolean semiring, K_n decomposes into $K_n = B \cdot B^T$, where the i th row in B is the binary representation of i . Now apply Theorem 2 to the $n \times \log n$ matrix B and its transpose and perform the composition.

Any lower bound against XOR circuits must hold for cancellation-free circuits, so a first step in proving superlinear lower bounds for the set intersection matrix is to prove superlinear cancellation-free lower bounds for the Sierpinski matrix. Below we show that $C_{CF}(S_n) = \frac{1}{2}n \log n$. Our technique also holds for OR circuits. This provides a simple example of a matrix family where the complements are significantly easier to compute with OR circuits than the matrices themselves.

Gate Elimination. Suppose some subset of the input variables are restricted to the value 0. Now look at the resulting circuit. Some of the gates will now compute the value $z = 0 \oplus w$. In this case, we say that the gate is eliminated since it no longer does any computation. The situation can be more extreme, some gate might “compute” $z = 0 \oplus 0$. In both cases, we can remove the gate from the circuit, and forward the input if necessary (if z is an output gate, w now outputs the result). In the second case, the parent of z will get eliminated, so the effect might cascade. For any subset of the variables, there is a unique set of gates that become eliminated when setting these variables to 0.

In all of the following let n be a power of 2, and let S_n be the $n \times n$ Sierpinski matrix. The following proposition is easily established.

Proposition 3. *For every n , the Sierpinski matrix S_n has full rank, over both \mathbb{R} and \mathbb{F}_2 .*

We now proceed to the proof of the lower bound of the Sierpinski matrix for cancellation-free circuits. It is our hope that this might be a step towards proving an $\omega(n)$ lower bound for XOR circuits.

Theorem 10. *For every $n \geq 2$, any cancellation-free circuit that computes the $n \times n$ Sierpinski matrix has size at least $\frac{1}{2}n \log n$.*

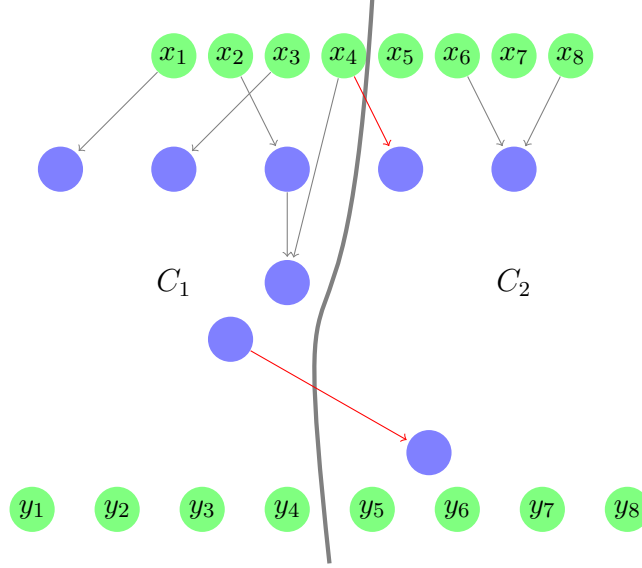


Figure 3: Figure illustrating the inductive step. Due to monotonicity there is no wire crossing from right to left. The gates on the left hand side are in C_1 . Notice that wires crossing the cut are red, and that these wires become constant when x_1, \dots, x_n are set to 0, so the gates with one such input wire are in C_3 . The rest are in C_2 .

PROOF. The proof is by induction on n . For the base case, look at the 2×2 matrix S_2 . This clearly needs at least $\frac{1}{2}2 \log 2 = 1$ gate.

Suppose the statement is true for some n and consider the $2n \times 2n$ matrix S_{2n} . Denote the output gates y_1, \dots, y_{2n} and the inputs x_1, \dots, x_{2n} . Partition the gates of C into three disjoint sets, C_1, C_2 and C_3 (Figure 3 illustrates the situation), defined as follows:

- C_1 : The gates having only inputs from x_1, \dots, x_n and C_1 . Equivalently the gates not reachable from inputs x_{n+1}, \dots, x_{2n} .
- C_2 : The gates in $C - C_1$ that are not eliminated when inputs x_1, \dots, x_n are set to 0.
- C_3 : $C - (C_1 \cup C_2)$. That is, the gates in $C - C_1$ that do become eliminated when inputs x_1, \dots, x_n are set to 0.

Obviously $|C| = |C_1| + |C_2| + |C_3|$. We will now give lower bounds on the sizes of C_1 , C_2 , and C_3 .

C_1 :. Since the circuit is cancellation-free, the outputs y_1, \dots, y_n and all their predecessors are in C_1 . By the induction hypothesis, $|C_1| \geq \frac{1}{2}n \log n$.

C_2 :. Since the gates in C_2 are not eliminated, they compute S_n on the inputs x_{n+1}, \dots, x_{2n} . By the induction hypothesis, $|C_2| \geq \frac{1}{2}n \log n$.

C_3 :. The goal is to prove that this set has size at least n . Let $\delta(C_1)$ be the set of wires from $C_1 \cup \{x_1, \dots, x_n\}$ to $C_2 \cup C_3$. We first prove that $|C_3| \geq |\delta(C_1)|$.

By definition, all gates in C_1 attain the value 0 when x_1, \dots, x_n are set to 0. Let $(v, w) \in \delta(C_1)$ be arbitrary. Since $v \in C_1 \cup \{x_1, \dots, x_n\}$, w becomes eliminated, so $w \in C_3$. By definition, every $u \in C_3$ can only have one child in C_1 . So $|C_3| \geq |\delta(C_1)|$.

We now show that $|\delta(C_1)| \geq n$. Let the endpoints of $\delta(C_1)$ in C_1 be e_1, \dots, e_p and let their corresponding value vectors be v_1, \dots, v_p .

The circuit is cancellation-free, so coordinatewise addition corresponds to addition in \mathbb{R} . Now look at the value vectors of the output gates y_{n+1}, \dots, y_{2n} . For each of these, the vector consisting of the first n coordinates must be in $\text{span}_{\mathbb{R}}(v_1, \dots, v_p)$, but the dimension of S_n is n , so $p \geq n$. We have that $|C_3| \geq |\delta(C_1)| \geq n$, so

$$|C| = |C_1| + |C_2| + |C_3| \geq \frac{1}{2}n \log n + \frac{1}{2}n \log n + n = \frac{1}{2}(2n) \log(2n).$$

□

This is tight:

Proposition 4. *The Sierpinski matrix can be computed by a cancellation-free circuit using $\frac{1}{2}n \log n$ gates.*

PROOF. This is clearly true for S_2 . Assume that S_n can be computed using $\frac{1}{2}n \log n$ gates. Consider the matrix S_{2n} . Construct the circuit in a divide and conquer manner by constructing recursively on the variables x_1, \dots, x_n and x_{n+1}, \dots, x_{2n} . This gives outputs y_1, \dots, y_n . After this use n operations to finish the outputs y_{n+1}, \dots, y_{2n} . This adds up to exactly $\frac{1}{2}(2n) \log(2n)$. □

Circuits With Cancellation. In the proof of Theorem 10, we used the cancellation-free property when estimating the sizes of both C_1 and C_3 . However, since S_n has full rank over \mathbb{F}_2 , a similar dimensionality argument to that used when estimating C_3 holds even if the circuits use cancellation. Therefore we might replace the cancellation-free assumption with the assumption that for the $2n \times 2n$ Sierpinski matrix, there is no path from x_{n+i} to y_j for $i \geq 1$, $j \leq n$. We have not been able to show whether or not this is the case for minimum sized circuits, although we have experimentally verified that even for circuits where cancellation is allowed, the matrices S_2, S_4, S_8 do not admit circuits smaller than the lower bound from Theorem 10.

OR circuits. In the proof of Theorem 10, the estimates for C_1 and C_2 hold for OR circuits too, but when estimating C_3 , it does not suffice to appeal to rank over \mathbb{F}_2 or \mathbb{R} . However, it is not hard to see that any set of row vectors that “spans” S_n (with the operation being coordinate-wise OR) must have size at least n .

Theorem 11. *Theorem 10 holds for OR circuits as well.*

This proof strategy for Theorem 10 has recently been used by Sergeev to prove similar lower bounds for another family of Boolean matrices in the OR model [35]. As mentioned in the introduction, Theorem 10 can be shown using another strategy. In [8], Kennes gives a lower bound on the additive complexity for computing the Möbius transformation of a Boolean lattice. It is not hard to verify that the Sierpinski matrix corresponds to the Möbius transformation induced by the subset lattice. Combining this observation with Kennes’ result gives the same lower bound.

Since $C_V(K_n) \in O(n)$ and K_n contains the same rows as \bar{S}_n , the complement of S_n , the Sierpinski matrix is harder to compute than its complement.

Corollary 3. $C_V(S_n) = \Theta(\log n)C_V(\bar{S}_n)$.

Until very recently, this was the largest gap between the OR complexity of A and \bar{A} for an explicit matrix. See [36] for a very recent manuscript describing a construction greatly improving on this. [6]).

7. Conclusions and Open Problems

We show the existence of matrices, for which OR circuits and cancellation-free XOR circuits are both a factor of $O\left(\frac{n}{\log^2 n}\right)$ larger than the smallest XOR circuit. This separation holds in unbounded depth and any constant depth of at least 2.

This means that when designing XOR (sub)circuits, it can be important that the methods employed can produce circuits which have cancellation.

If a cancellation-free or an OR circuit computes the Sierpinski matrix correctly, it has size at least $\frac{1}{2}n \log n$. For this particular family of matrices, it is not obvious to what extent cancellation can help. It would be very interesting to determine this, since it would automatically provide a converse to Theorem 4.

Acknowledgments

The authors would like to thank Elad Verbin for an idea which eventually led to the proof of Theorem 4, Igor Sergeev and Stasys Jukna for references to related papers, Janne Korhonen for pointing out the result of Kennes, Avishay Tal for pointing out an alternative proof of a slightly weaker version of Theorem 4, and Mika Göös for helpful discussions.

We would also like to thank the anonymous referees for many valuable suggestions.

References

- [1] J. Boyar, M. G. Find, Cancellation-free circuits in unbounded and bounded depth, in: L. Gasieniec, F. Wolter (Eds.), FCT, volume 8070 of *Lecture Notes in Computer Science*, Springer, 2013, pp. 159–170.
- [2] J. Boyar, P. Matthews, R. Peralta, Logic minimization techniques with applications to cryptology, *J. Cryptology* 26 (2013) 280–312.
- [3] S. Gashkov, I. Sergeev, On the complexity of linear Boolean operators with thin matrices, *Journal of Applied and Industrial Mathematics* 5 (2011) 202–211.
- [4] M. Grinchuk, I. Sergeev, Thin circulant matrices and lower bounds on the complexity of some boolean operators, *Discret. Anal. & Issl. Oper.* 18 (2011) 38–53.
- [5] S. Jukna, Disproving the single level conjecture, *SIAM J. Comput.* 36 (2006) 83–98.
- [6] S. Jukna, I. Sergeev, Complexity of linear boolean operators, *Foundations and Trends in Theoretical Computer Science* 9 (2013) 1–123.
- [7] M. Find, M. Göös, P. Kaski, J. Korhonen, Separating Or, Sum and XOR Circuits, *arXiv preprint* (2013).
- [8] R. Kennes, Computational aspects of the mobius transformation of graphs, *IEEE Transactions on Systems, Man, and Cybernetics* 22 (1992) 201–223.
- [9] S. 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, 2012.

- [10] S. Selezneva, Lower bound on the complexity of finding polynomials of boolean functions in the class of circuits with separated variables, *Computational Mathematics and Modeling* 24 (2013) 146–152.
- [11] N. Pippenger, On the evaluation of powers and related problems (preliminary version), in: *FOCS*, IEEE Computer Society, 1976, pp. 258–263.
- [12] N. Pippenger, On the evaluation of powers and monomials, *SIAM J. Comput.* 9 (1980) 230–250.
- [13] M. R. Garey, D. S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*, W. H. Freeman, 1979.
- [14] E. Nechiporuk, Rectifier networks, in: *Soviet Physics Doklady*, volume 8, 1963, p. 5.
- [15] O. Lupanov, On rectifier and switching-and-rectifier schemes, *Dokl. Akad. Nauk SSSR* 111, 1171–1174. (1956). English translation available at <http://www.thi.informatik.uni-frankfurt.de/~jukna/lupanov56.pdf>.
- [16] S. Jukna, *Boolean Function Complexity: Advances and Frontiers*, Springer Berlin Heidelberg, 2012.
- [17] E. Nechiporuk, On the topological principles of self-correction, *Problemy Kibernetika* (1969) 5–102. (In Russian).
- [18] K. Mehlhorn, Some remarks on Boolean sums, *Acta Informatica* 12 (1979) 371–375.
- [19] N. Pippenger, On another boolean matrix, *Theor. Comput. Sci.* 11 (1980) 49–56.
- [20] I. Wegener, A new lower bound on the monotone network complexity of boolean sums, *Acta Inf.* 13 (1980) 109–114.
- [21] E. Hirsch, O. Melanich, Personal communication, 2012.
- [22] N. Alon, M. Karchmer, A. Wigderson, Linear circuits over $\text{GF}(2)$, *SIAM J. Comput.* 19 (1990) 1064–1067.
- [23] J. Morgenstern, Note on a lower bound on the linear complexity of the fast Fourier transform, *J. ACM* 20 (1973) 305–306.

- [24] P. Bürgisser, M. Clausen, M. A. Shokrollahi, Algebraic complexity theory, volume 315 of *Grundlehren der mathematischen Wissenschaften*, Springer, Heidelberg, 1997.
- [25] B. Chor, O. Goldreich, Unbiased bits from sources of weak randomness and probabilistic communication complexity, *SIAM J. Comput.* 17 (1988) 230–261.
- [26] E. Kushilevitz, N. Nisan, Communication Complexity, Cambridge University Press, 1997.
- [27] B. Chor, O. Goldreich, Unbiased bits from sources of weak randomness and probabilistic communication complexity (extended abstract), in: FOCS, IEEE Computer Society, 1985, pp. 429–442.
- [28] T. Kovári, V. Sós, P. Turán, On a problem of K. Zarankiewicz, in: Colloquium Math, volume 3, 1954, pp. 50–57.
- [29] S. Jukna, Extremal Combinatorics - With Applications in Computer Science, Texts in Theoretical Computer Science, Springer, Heidelberg, 2001.
- [30] I. Komargodski, R. Raz, A. Tal, Improved average-case lower bounds for DeMorgan formula size, in: FOCS, IEEE Computer Society, 2013, pp. 588–597.
- [31] D. Canright, A very compact s-box for AES, in: CHES, volume 6049 of *LNCS*, Springer, Heidelberg, 2010.
- [32] C. Paar, Some remarks on efficient inversion in finite fields, in: B. Whistler (Ed.), IEEE International Symposium on Information Theory, volume 5162 of *LNCS*, Springer, Heidelberg, 1995, p. 58.
- [33] A. Satoh, S. Morioka, K. Takano, S. Munetoh, A compact Rijndael hardware architecture with S-box optimization, in: C. Boyd (Ed.), ASIACRYPT, volume 2248 of *LNCS*, Springer, Heidelberg, 2001, pp. 239–254.
- [34] S. Aaronson, Thread on cstheory.stackexchange.com, <http://cstheory.stackexchange.com/questions/1794/circuitlower-boundsoverarbitrarysetsofgates>, 2012.
- [35] I. Sergeev, On additive complexity of a sequence of matrices, arXiv preprint (2012).

- [36] I. Sergeev, On the or complexity of matrices and their complements, <http://lovelace.thi.informatik.uni-frankfurt.de/~jukna/Knizka/comment1.html>, 2014.