Cancellation-Free Circuits in Unbounded and Bounded Depth

Joan Boyar* and Magnus Gausdal Find**
{joan, magnusgf}@imada.sdu.dk
Department of Mathematics and Computer Science
University of Southern Denmark

Abstract. We study the notion of “cancellation-free” circuits. This is a restriction of linear Boolean circuits (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 linear circuit. We show that the difference can be a factor \( \Omega\left(\frac{n}{\log^2 n}\right) \). This improves on a recent result by Sergeev and Gashkov who have studied a similar problem. 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.

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 \( \mathbb{F}_2 \). A Boolean function \( f : \mathbb{F}_2^n \to \mathbb{F}_2^m \) is said to be linear if there exists a Boolean \( m \times n \) matrix \( A \) such that \( f(x) = Ax \) for every \( x \in \mathbb{F}_2^n \). This is equivalent to saying that \( f \) can be computed using only XOR gates.

A linear circuit (or XOR-circuit) \( C \) is a directed acyclic graph. There are \( n \) nodes with in-degree 0, called the inputs. All other nodes have in-degree 2 and are called gates. There are \( m \) nodes which are called the outputs; these are labeled \( y_1, \ldots, y_m \). The value of a gate is the sum of its two children (addition in \( \mathbb{F}_2 \), denoted \( \oplus \)). The circuit \( C \), with inputs \( x = (x_1, \ldots, x_n) \), computes the \( m \times n \) matrix \( A \) if the output vector computed by \( C \), \( y = (y_1, \ldots, y_m) \), satisfies \( y = Ax \). 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 \( C \). For simplicity, we will let \( m = n \) unless otherwise

* Partially supported by the Danish Council for Independent Research, Natural Sciences. Part of this work was done while visiting the University of Waterloo.
** Part of this work was done while visiting the University of Toronto.
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 linear circuits called cancellation-free circuits, coined in [3], where the authors noticed that many heuristics for finding small linear circuits would always produce cancellation-free linear circuits. They asked the question of how large a separation there can be between these two models. Recently, a separation of $\Omega\left(\frac{n}{\log n \log \log n}\right)$ due to Gashkov and Sergeev was given in [9]. We improve on this result by giving a slightly stronger separation, namely $\Omega\left(\frac{n}{\log^2 n}\right)$. Furthermore, our proof gives a similar separation for the case of linear circuits with constant depth. We conclude that many heuristics for finding linear 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.

## 2 Cancellation-Free Linear Circuits

For linear circuits, the value computed by every gate is the parity of some 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 a linear 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. If this is satisfied, the circuit never exploits the fact that in $\mathbb{F}_2$, $a \oplus a = 0$, so things do not “cancel out” in the circuit.

Although it is not hard to see that the model is equivalent to addition chains [21][23] and “ensemble computations”, [8], in this paper we will stick to the term “cancellation-free”, since we will think of it as a special case of linear circuits. A different, but tightly related kind of circuits are OR circuits. The definition is exactly the same as for linear circuits, but with $\lor$ instead of $\oplus$, see [19][13][8]. In particular every cancellation-free circuit gives an OR-circuit for the same matrix, so lower bounds in the OR-model carry over to lower bounds on cancellation-free circuits. However, the converse does not hold in general [7].
For a matrix $A$, we will let $C_{\oplus}(A)$, $C_{CF}(A)$, $C_{\lor}(A)$ denote the smallest linear circuit, the smallest cancellation-free circuit and the smallest $OR$ circuit computing the matrix $A$. By the discussion above, the following is immediate:

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

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 [19] and Pippenger [21] (see also [13]) that this upper bound can be improved to $(1 + o(1)) \frac{n^2}{\log n}$. The proof due to Nechiporuk is for the OR model, but the proof also holds for cancellation-free circuits.

A Shannon-style counting argument gives that this is tight up to low order terms. A proof of this can be found in [21]. Combining these results, we get that for most matrices, cancellation does not help much.

**Theorem 1.** For every $\epsilon > 0$, almost every $n \times n$ matrix has $\frac{C_{CF}(A)}{C_{\oplus}(A)} \leq 1 + \epsilon$.

For $m \times n$ matrices, the following upper bound also holds.

**Theorem 2 (Lupanov [16]).** Any $m \times n$ matrix, admits a cancellation-free linear circuit of size $O\left(\min\{\frac{mn}{\log n}, \frac{mn}{\log m}\} + n + m\right)$.

The theorem follows directly from Lupanov’s result and the application of the “transposition principle” (see e.g. [13]).

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 due to Mehlhorn [17] and Pippenger [22] will be used later.

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

### 3 Relationship Between Cancellation-Free Linear Circuits and General Linear Circuits

In [3], Boyar and Peralta exhibited an infinite family of matrices where the sizes of cancellation-free circuits computing them are at least $\frac{3}{4} - o(1)$ times larger than the optimum. 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 [10]. The $n \times n$ Boolean Sylvester-Hadamard matrix $H_n$, is defined recursively:

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

It is known that $C_\oplus(H_n) \in O(n)$, but that in depth 2 it requires circuits of size $\Omega(n \log n)$ [2].

**Proposition 1.** The $n \times n$ Boolean Sylvester-Hadamard matrix requires cancellation-free circuits of size $C_{CF}(H_n) \in \Omega(n \log n)$.

It is based on the following theorem due to Morgenstern, ([18], see also [1 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 of any constant $c \in \mathbb{C}$ with $|c| \leq 2$. Cancellation-free can be seen as a special case of this.

The determinant of $H_n$ is $n^{\Omega(n)}$, so the lower bound follows. 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 $n!$ (since the permanent is at most $n!$), so this technique cannot give lower a bound on $\rho(n)$ stronger than $\Omega(\log n)$.

In [9], Gashkov and Sergeev studied the ratio $\lambda(n) = \max_{A \in \mathbb{F}_2^{n \times n}} \frac{C_\lor(A)}{C_\oplus(A)}$. They showed that $\lambda(n) \in \Omega\left(\frac{n^2}{\log^4 n \log \log n}\right)$. We improve this by showing that $\lambda(n) \in \Omega\left(\frac{n^2}{\log^2 n \log \log n}\right)$. Alternative proofs of this are given in more recent work [7,11]. We include the proof below because the construction is different, and the technique we use to analyze the matrix is as well. More concretely we use communication complexity for the analysis 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.** $\lambda(n) \in \Omega\left(\frac{n^2}{\log^2 n \log \log n}\right)$.

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 a technical lemma stated below, which will be useful in showing that with high probability, our matrix will be $3\log n$-free.

In the following, for a matrix $M$, we let $M_i$ ($M^i$) denote the $i$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 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 [6][15].

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 $M^q_p$ are defined. Let $B, C$ be matrices over $\mathbb{F}_2$ with dimensions $m \times 8m$ and $8m \times m$ 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 $M^q_p$ 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 [6], see also [15].

**Theorem 5 (Chor, Goldreich).** Let $x$ and $y$ be independent and uniformly random vectors, each of $n$ bits. Suppose a deterministic communication protocol is used to compute the inner product of $x$ and $y$, and the protocol is correct with probability at least $\frac{1}{2} + \rho$. Then on some inputs, the protocol uses $\frac{n}{2} - \log(1/\rho)$ 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 $8m$ bit vectors $x$ and $y$ 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 \( x \) and \( y \), respectively, each of length \( 8m \). 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 \( x \) and \( y \), Alice replaces the row \( B_p \) with \( x \) and Bob replaces the column \( C_q \) with \( y \), 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 \( (M'_q)_p \)). 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 \( (M'_q)_p \).

In order for \( M' \) and \( M \) to be consistent, it is only necessary that the \( m-1 \) defined entries in row \( p \) and \( m-1 \) defined entries in column \( q \) be 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 \( (M'_q)_p \). 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 0. This will be the correct value for the inner product on at least half of all inputs. Thus, the probability of this protocol being correct is at least:

\[
2^{-2m-2} \left( \frac{1}{2} + \frac{1}{m} \right) + (1 - 2^{-2m-2}) \frac{1}{2}
\]

\[
= \frac{1}{2} + 2^{-2m} \frac{4}{4m}
\]

So when the inputs are uniformly distributed, the randomized protocol computes the inner product of two \( 8m \) bits vectors with \( m \) bits communication, and it is correct with probability at least \( \frac{1}{2} + 2^{-2m} \frac{4}{4m} \). The only use of randomness was by choosing the matrix pair \( (B, C) \). By an averaging argument it follows that instead of choosing \( (B, C) \) randomly there is a fixed choice of \( (B, C) \) that gives at least the same success probability. Hence we arrive at a deterministic communication protocol violating Theorem 5.

We now use this to prove Theorem 4. \( \square \)

We now use this to prove Theorem 4.
Proof (of Theorem 4). We will probabilistically construct two matrices \( B, C \) of dimensions \( n \times 24 \log(n) \), \( 24 \log(n) \times n \). 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 linear circuits, both of size \( O(n) \). Now we can let the outputs of the circuit computing \( C \) be the inputs of the circuit computing \( B \). The resulting circuit computes the matrix \( A \) and has size \( O(n) \). We will argue that with nonzero probability this matrix will not have a \( 3 \log(n) \times 3 \log(n) \) all 1 submatrix, while \( |A| \in \Omega(n^2) \). By Lemma 1 the results follows. We show that with nonzero probability neither of the following two events will happen:

1. \(|BC| < \frac{n^2}{9}\)
2. \(BC\) has an all one submatrix of dimension \( 3 \log(n) \times 3 \log(n) \)

(\textit{1}) Let \((BC)^{ij}_k\) be the random variable indicating the value of entry \( i, j \) in the matrix \( BC \). By definition, the number of ones in \( BC \) is a random variable, \(|BC| = \sum_{i,j} (BC)^{ij}_i\). \((BC)^{ij}_i\) is computed as the inner product of the row \( i \) in \( B \) with column \( j \) in \( C \). The probability that column \( j \) in \( C \) is all zeros is \( 2^{-24 \log(n)} \). If it is not zero, the probability of the inner product having the result one is exactly \( \frac{1}{2} \) (for any \( 1 \) in \( C^j \) the probability of it being added into the sum in the inner product is exactly \( \frac{1}{2} \)). Thus, the expected value of \((BC)^{ij}_i\) is \( (1 - 2^{-24 \log(n)}) \frac{1}{2} \), and, by the linearity of expectations, \( \mathbb{E}[|BC|] = n^2 \left( \frac{1}{2} - \frac{1}{2 \log(n) + 1} \right) \). We want to show that \( \mathbb{P}[|BC| < \frac{n^2}{9}] \leq \frac{3}{4} \). Notice that for large \( n \), \( \mathbb{E}[|BC|] \geq \frac{n^2}{3} \). By applying Markov’s inequality to the non-negative random variable \( n^2 - |BC| \), we get

\[
\mathbb{P}\left[|BC| < \frac{n^2}{9}\right] = \mathbb{P}\left[n^2 - |BC| \geq \frac{8n^2}{9}\right] \leq \frac{n^2 - \frac{n^2}{9}}{8n^2} = \frac{3}{4}.
\]

(\textit{2}) Fix a submatrix \( M \) of \( BC \) with dimensions \( 3 \log(n) \times 3 \log(n) \). That is, some subset \( I \) of the rows of \( B \), and a subset \( J \) of the columns in \( C \) so \( M = B_J C^J \). We now want to show that the probability of this matrix having only 1’s is so small that a union bound over all choices of \( 3 \log(n) \times 3 \log(n) \) submatrices gives that the probability that there exists such a submatrix is less than \( \frac{1}{4} \). Notice that this would be easy if all the entries in \( M \) were mutually independent and uniformly distributed.

Though this is not case, Lemma (2) for \( m = 3 \log(n) \) states, that this is almost the case. More precisely, the conditional probability that a given entry is 1 is at most \( \frac{1}{2} + \frac{1}{3 \log(n)} \). We can now use the union bound to
estimate the probability that $A$ has an all one submatrix of dimension $3\log n \times 3\log n$:

\[
\left( \frac{n}{3\log n} \right)^2 \left( \frac{1}{2} + \frac{1}{3\log n} \right) ^ {9\log^2 n} \leq \frac{n^{6\log n}}{(3\log n)!} \left( \frac{1 + \frac{2}{3\log n}}{2} \right)^{9\log^2 n} \leq \frac{(1 + \frac{2}{3\log n})^{9\log^2 n}}{(3\log n)!2^{3\log^2 n}}
\]

This tends to 0, so for sufficiently large $n$ this probability is strictly smaller than $\frac{1}{4}$. By the union bound we conclude that with nonzero probability, neither of the two events occur. Thus, with nonzero probability, $A$ is $3\log n$-free with $|A| \in \Omega(n^2)$ and $C_{\oplus}(A) \in O(n)$. By Lemma 1 this proves the theorem.

Remark: Originally, this result was slightly weaker, obtaining a separation of $\Omega \left( \frac{n}{\log^{2+\epsilon} n} \right)$. Motivated by this result, the $\Omega \left( \frac{n}{\log^2 n} \right)$ was obtained in [7]. Referring to that result, Stasys Jukna and Igor Sergeev pointed out that Theorem [4] can be proved using a slightly different method [11]. Roughly speaking, it is shown that certain “$t$-Ramsey” graphs admit small circuits. A bipartite graph is said to be $t$-Ramsey, if it neither contains a (bi)clique with $t$ vertices in each vertex set nor an independent set with $t$ vertices in each vertex set. Though the proofs and constructions are different, they are along the same general lines. The proof of Theorem [3] is mainly concerned with the asymptotic values. Therefore, we only showed that with high probability, the matrix is $3\log n$-free. However, it is not hard to see that in fact it is also $2\log n$-free with high probability. Applying the same argument to estimate the probability of the existence of a $2\log n \times 2\log n$ all zero matrix, we conclude that with high probability $A$ is $2\log n$-Ramsey.

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

Notice that by Theorem 4 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 $O(\log^{1-\epsilon}) \times O(\log^{1-\epsilon})$ all 1 submatrix to get a stronger lower bound on $C_{\oplus}(A)$, or perhaps hope that a tighter analysis than the above would give
a stronger separation. However, this direction does not seem promising. To make this precise, we will use following result on “Zarankiewicz problem” [14], see also [12].

**Theorem 6 (Kovári, Sós, Turán).** Let $M$ be an $n \times n$ matrix without an $a \times a$ submatrix of all ones. Then the number of ones in $M$ is at most 
\[(a - 1)^{1/a}n^{2-1/a} + (a - 1)n.\]

This implies that for 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 $O \left( \frac{n^2}{\log^{1-\epsilon}(n)^2} \right)$, which is $o \left( \frac{n^2}{\log n} \right)$.

### 4 Smallest linear circuit problem

As mentioned earlier, the notion cancellation-free was introduced by Boyar and Peralta in [3]. The paper concerns shortest straight line programs for computing linear forms, which is equivalent to the model studied in this paper. In [8], it is shown that the Ensemble Computation Problem (recall that this is equivalent to cancellation-free) is $\mathsf{NP}$-complete. For general linear circuits, the problem remains $\mathsf{NP}$-complete ([3]). It was observed in [3] that several researchers have used heuristics that will always produce cancellation-free circuits, see [5,20,24]. By definition, any heuristic which only produces cancellation-free circuits cannot achieve an approximation ratio better than $\rho(n)$. By Remark 1, $\rho(n) \geq \lambda(n)$. By Theorem 4, we get that techniques which only produce cancellation-free circuits are not guaranteed to be very close to optimal.

**Corollary 2.** The algorithms in [5,20,24] do not guarantee approximation ratios better than $\Theta \left( \frac{n}{\log^{2} n} \right)$.

### 5 Constant Depth

For unbounded depth, there is no known family of (polynomial time computable) matrices known to require linear circuits of superlinear size. However, if one puts restriction on the depth, superlinear lower bounds are known [13]. 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. In particular, the circuit model where the depth is bounded to be at most 2 is well studied (see e.g. [13]). Similarly to previously, we say that a circuit $C$ is *linear* if every gate computes the XOR or its inputs.
When considering matrices computed by linear circuits, the general situation in the two circuit models are very similar. The upper bound comes from Lupanov [16], and the lower bound is folklore. See also [13].

**Theorem 7 (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 of the separations in [9] and [7] seem to carry over to bounded depth circuits in any obvious way. By inspecting the proof of Theorem 4, the upper bound on the size of the linear 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 $O(n)$, we used Theorem 2. By using Theorem 7 twice, we get a depth 4 circuit of that size.

**Theorem 8.** Let $d \geq 4$. $\lambda^d(n) \in \Omega\left(\frac{n}{\log^2 n}\right)$.

For depths 2 and 3, we can construct a depth 1 circuit for each of $B$ and $C$. This results in a depth 2 circuit for $A$ with $O(n \log n)$ wires.

**Theorem 9.** Let $d \in \{2, 3\}$ $\lambda^d(n) \in \Omega\left(\frac{n}{\log^2 n}\right)$.

### 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 in this case.

As mentioned previously, there is no known (polynomial time computable) family of matrices requiring linear 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 [1]). 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 [11] has very recently asked if the “set intersection matrix” (also called the Kneser matrix in [11]), $K_n$, has $C_{\oplus}(K_n) \in \omega(n)$. The motivation for this is that $C_{\lor}(K_n) \in O(n)$, so if true this would give a counterpart to Theorem 4. 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 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 ith row in $B$ is the binary representation of $i$. Now apply Theorem 2 to the $n \times \log n$ matrix $B$ and its transpose.

Any lower bound against linear 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. 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 2.** For every $n$, the Sierpinski matrix 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 a $\omega(n)$ lower bound for linear circuits.

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

**Proof.** The proof is by induction on $n$. For the base case, look at the $2 \times 2$ matrix $S_2$. This clearly needs at least $\frac{1}{2}2 \log 2 = 1$ gate. (Figure 1 illustrates the situation).

![Figure 1](image-url)

**Fig. 1.** Figure illustrating the inductive step. Due to monotinicity there is no wire crossing from right to left. The gates on the left hand side are in $C_1$. Wires crossing the cut are red. Notice that these wires become constant when $x_1, \ldots, x_n$ are set to 0, so the gates with one such input wire are in $C_3$. The rest are in $C_2$.

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

- $C_1$: The gates having only inputs from $x_1, \ldots, x_n$ and $C_1$. Equivalently the gates not reachable from inputs $x_{n+1}, \ldots, x_{2n}$.
\[ C \begin{align*}
&\text{C}2: \text{The gates in } C - C_1 \text{ that are not eliminated when inputs } x_1, \ldots, x_n \\
&\text{are set to 0.}
&\text{C}3: \text{C} - (C_1 \cup C_2). \text{ That is, the gates in } C - C_1 \text{ that do become}
\end{align*}
\]

eliminated when inputs } x_1, \ldots, x_n \text{ is set to 0.}

Obviously } |C| = |C_1| + |C_2| + |C_3|. \text{ We will now give lower bounds on the}

sizes of } C_1, C_2, \text{ and } C_3.

\[ C_1: \text{Since the circuit is cancellation-free, the outputs } y_1, \ldots, y_n \text{ and all their predecessors are in}
\]

\[ C \begin{align*}
&\text{C}1 \text{ union } \{x_1, \ldots, x_n\} \text{ to } C_2 \cup C_3. \text{ We first prove that } |C_3| \geq |\delta(C_1)|. \text{ By definition, all gates in } C_1 \text{ attain the value 0 when } x_1, \ldots, x_n \\
&\text{are set to 0. Let } (v, w) \in \delta(C_1) \text{ be arbitrary. Since } v \in C_1 \cup \{x_1, \ldots, x_n\}, \text{ } w \text{ becomes eliminated, so } w \in C_3. \text{ By definition, every } u \in C_3 \text{ can only}
\end{align*}
\]

\[ \text{have one child in } C_1. \text{ So } |C_3| \geq |\delta(C_1)|.
\]

\[ \text{We now show that } |\delta(C_1)| \geq n. \text{ Let the endpoints of } \delta(C_1) \text{ in } C_1 \text{ be}
\]

\[ e_1, \ldots, e_p \text{ and let their corresponding value vectors be } v_1, \ldots, v_p.
\]

\[ \text{The circuit is cancellation-free, so coordinate-wise addition corresponds}
\]

\[ \text{to addition in } \mathbb{R}. \text{ Now look at the value vectors of the output gates}
\]

\[ y_{n+1}, \ldots, y_{2n}. \text{ For each of these, the vector consisting of the first } n \text{ co-}
\]

\[ \text{ordinates must be in } \text{span}_{\mathbb{R}}(v_1, \ldots, v_p), \text{ but the dimension of } S_n \text{ is } n, \text{ so}
\]

\[ p \geq n. \text{ We have that } |C_3| \geq |\delta(C_1)| \geq n, \text{ so}
\]

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

\[ \square
\]

This is tight:

\[ \text{Proposition 3. The Sierpinski matrix can be computed by a cancellation-}
\]

\[ \text{free circuit using } \frac{1}{2}n \log n \text{ gates.}
\]

\[ \text{Proof. This is clearly true for } S_2. \text{ Assume that } S_n \text{ can be computed using}
\]

\[ \frac{1}{2}n \log n \text{ gates. Consider the matrix } S_{2n}. \text{ Construct the circuit in a}
\]

\[ \text{divide and conquer manner by constructing recursively on the variables}
\]

\[ x_1, \ldots, x_n \text{ and } x_{n+1}, \ldots, x_{2n}. \text{ This gives outputs } y_1, \ldots, y_n. \text{ After this use}
\]

\[ n \text{ operations to finish the outputs } y_{n+1}, \ldots y_{2n}. \text{ This adds up to exactly}
\]

\[ \frac{1}{2}(2n) \log 2n. \]

\[ \square
\]
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.

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

Corollary 3. \( C_\lor(S_n) = \Theta(\log n)C_\lor(\overline{S}_n) \).

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 [25].

7 Conclusions and Open Problems

For circuits of unbounded depth, we show the existence of matrices, for which OR circuits and cancellation-free linear circuits are both a factor of \( \Omega\left(\frac{n}{\log^2 n}\right) \) larger than the smallest linear circuit. For circuits of constant depth we give a separation of \( \Omega\left(\frac{n}{\log^D n}\right) \) where \( D = 3 \) for circuits of depth 2 or 3 and \( D = 2 \) for any larger depth.

This means that when designing linear (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 \( \Omega(n \log n) \). 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 also solve Jukna’s conjecture concerning set intersection matrices.

Acknowledgments

The authors would like to thank Elad Verbin for an idea which eventually led to the proof of Theorem 4. They would also like to thank Igor Sergeev and Stasys Jukna for references to related papers and Mika Göös for helpful discussions.

References

1. Aaronson, S.: Thread on cstheory.stackexchange.com. http://cstheory.stackexchange.com/questions/1794/circuit-lower-bounds-over-arbitrary-sets-of-gates
2. Alon, N., Karchmer, M., Wigderson, A.: Linear circuits over GF(2). SIAM J. Comput. 19(6), 1064–1067 (1990)
3. Boyar, J., Matthews, P., Peralta, R.: Logic minimization techniques with applications to cryptology. Journal of Cryptology 26, 280–312 (2013)
4. Bürgisser, P., Clausen, M., Shokrollahi, M.A.: Algebraic complexity theory, Grundlehren der mathematischen Wissenschaften, vol. 315. Springer (1997)
5. Canright, D.: A very compact S-box for AES. Cryptographic Hardware and Embedded Systems–CHES 2005 pp. 441–455 (2005)
6. Chor, B., Goldreich, O.: Unbiased bits from sources of weak randomness and probabilistic communication complexity (extended abstract). In: FOCS. pp. 429–442. IEEE Computer Society (1985)
7. Find, M., Göös, M., Kaski, P., Korhonen, J.: Separating Or, Sum and XOR Circuits. arXiv preprint (1304.0513) (April 2013)
8. Garey, M.R., Johnson, D.S.: Computers and Intractability: A Guide to the Theory of NP-Completeness. W. H. Freeman (1979)
9. Gashkov, S., Sergeev, I.: On the complexity of linear Boolean operators with thin matrices. Journal of Applied and Industrial Mathematics 5(2), 202–211 (2011)
10. Hirsch, E., Melanich, O.: Personal communication (2012)
11. Jukna, S.: Xor versus or circuits. http://www.thi.informatik.uni-frankfurt.de/~jukna/boolean/comment9.html
12. Jukna, S.: Extremal Combinatorics - With Applications in Computer Science. Texts in Theoretical Computer Science, Springer (2001)
13. Jukna, S.: Boolean Function Complexity: Advances and Frontiers. Springer Berlin Heidelberg (2012)
14. Kovári, T., Sós, V., Turán, P.: On a problem of K. Zarankiewicz. In: Colloquium Math. vol. 3, pp. 50–57 (1954)
15. Kushilevitz, E., Nisan, N.: Communication Complexity. Cambridge University Press (1997)
16. Lupanov, O.: On rectifier and switching-and-rectifier schemes. Dokl. Akad. 30 Nauk SSSR 111, 1171-1174. (1956)
17. Mehlhorn, K.: Some remarks on Boolean sums. Acta Informatica 12, 371–375 (1979)
18. Morgenstern, J.: Note on a lower bound on the linear complexity of the fast Fourier transform. J. ACM 20(2), 305–306 (1973)
19. Nechiporuk, É.: Rectifier networks. In: Soviet Physics Doklady, vol. 8, p. 5 (1963)
20. Paar, C.: Some remarks on efficient inversion in finite fields. In: Whistler, B. (ed.) IEEE Internatiol Symposium on Information Theory. Lecture Notes in Computer Science, vol. 5162, p. 58. Springer (1995)
21. Pippenger, N.: On the evaluation of powers and related problems (preliminary version). In: FOCS. pp. 258–263. IEEE Computer Society (1976)
22. Pippenger, N.: On another Boolean matrix. Theor. Comput. Sci. 11, 49–56 (1980)
23. Pippenger, N.: On the evaluation of powers and monomials. SIAM J. Comput. 9(2), 230–250 (1980)
24. Satoh, A., Morioka, S., Takano, K., Munetoh, S.: A compact Rijndael hardware architecture with S-box optimization. In: Boyd, C. (ed.) ASIACRYPT. Lecture Notes in Computer Science, vol. 2248, pp. 239–254. Springer (2001)
25. Sergeev, I.: On additive complexity of a sequence of matrices. arXiv preprint (1209.1645) (2012)