Determining the Multiplicative Complexity of Boolean Functions using SAT

Mathias Soeken
Microsoft, Switzerland

Abstract—We present a constructive SAT-based algorithm to determine the multiplicative complexity of a Boolean function, i.e., the smallest number of AND gates in any logic network that consists of 2-input AND gates, 2-input XOR gates, and inverters. In order to speed-up solving time, we make use of several symmetry breaking constraints; these exploit properties of XAGs that may be useful beyond the proposed SAT-based algorithm. We further propose a heuristic post-optimization algorithm to reduce the number of XOR gates once the optimum number of AND gates has been obtained, which also makes use of SAT solvers. Our algorithm is capable to find all optimum XAGs for representatives of all 5-input affine-equivalent classes, and for a set of frequently occurring 6-input functions.

I. INTRODUCTION

We are considering the minimization of AND gates in XOR-AND graphs (XAGs), which are logic networks that can have 2-input AND gates and 2-input XOR gates. XAGs can represent all normal Boolean functions, also called 0-preserving functions, which are all Boolean functions can represent all normal Boolean functions, also called 0-preserving functions, which are all Boolean functions.

The multiplicative complexity of a Boolean function is the smallest number of AND gates in any XAG that represents the function [11]. Determining the multiplicative complexity is intractable. Find has shown that if one-way functions exist [2], no algorithm can compute the multiplicative complexity of an n-variable Boolean function in time $2^{O(n)}$ given as input the truth table of $f$ [3]. Therefore, heuristic optimization techniques that minimize the number of AND gates in XAGs have been used as a tool to assess the multiplicative complexity of the function [4], [5], [6], [7], [8], since it provides an upper bound of the actual complexity.

The concept of multiplicative complexity and the optimization of AND gates in XAGs play an important role in cryptography and fault-tolerant quantum computing. In cryptography, the number of AND gates correlates to the degree of vulnerability of a circuit [9]. Further, the multiplicative complexity of a function directly correlates to the resistance of the function against algebraic attacks [10]. The number of AND gates also plays an important role in high-level cryptography protocols such as zero-knowledge protocols, fully homomorphic encryption (FHE), and secure multi-party computation (MPC) [11], [12], [6]. For example, the size of the signature in post-quantum zero-knowledge signatures based on “MPC-in-the-head" [13] depends on the multiplicative complexity in the underlying block cipher [12]. Moreover, the number of computations in MPC protocols based on Yao’s garbled circuits [14] with the free XOR technique [15] is proportional to the number of AND gates. Regarding FHE, XOR gates are considered cheaper and less noisy compared to AND gates. In fault-tolerant quantum computing, the multiplicative complexity directly corresponds to the number of qubits and expensive operations ($T$ gates), and therefore optimizing AND gates in an XAG can lead to more efficient quantum circuits [16].

In this paper, we present a SAT-based algorithm to determine the multiplicative complexity of a Boolean function. The algorithm is constructive and returns an XAG, in which the number of AND gates is minimum. The algorithm can only be applied to either small functions or to functions which have a small multiplicative complexity. It is inspired by SAT-based exact logic synthesis techniques [17], [18], [19]. While XAGs with minimum number of AND gates are known for all 6-input functions [20], our algorithms can also be used to enumerate multiple structurally different solutions, in order to minimize the number of XOR gates, while keeping the number of AND gates unchanged. For this purpose, we exploit an existing SAT-based algorithm to minimize the number of XOR gates in logic networks for linear functions [21]. Our algorithm is capable to find all optimum XAGs for representatives of all 5-input affine-equivalent classes, and for a set of frequently occurring 6-input functions. We further used the algorithm to find a new collection of optimum XAGs for 5-input functions, categorized with respect to representatives of a recently proposed affine equivalence classification algorithm [22].

II. PRELIMINARIES

As general notation, we are using $[n] = \{1, \ldots, n\}$. We model an XAG for a Boolean function over variables $x_1, \ldots, x_n$ as a sequence of $r$ steps, where each step has one of the two following forms:

$$x_i = x_{j_1} \oplus x_{j_2}, \quad \text{or} \quad x_i = x_{j_1} \land x_{j_2},$$

for $n < i \leq n + r$. The values $1 \leq j_1 < j_2 < i$ point to primary inputs or previous steps in the network. The function
value is computed by the last step \( f = x_{n+r} \). We assume that all logic functions represent Boolean functions that depend on all primary input variables. To represent the constant function, we define \( x_0 = 0 \).

**Example 1:** The if-then-else function \( x_1 \land x_2 : x_3 = x_1 x_2 \oplus \bar{x}_1 x_3 = x_1 x_2 \oplus x_1 x_3 \oplus x_3 \) can be computed by the 4-step XAG

\[
x_4 = x_1 \land x_2, \quad x_5 = x_1 \land x_3, \quad x_6 = x_4 \oplus x_5, \quad x_7 = x_3 \oplus x_6.
\]

The multiplicative complexity of an XAG is the number of AND gates it contains. It is an upper bound to the multiplicative complexity of the Boolean function, which is the smallest number of AND gates in any XAG that realizes the function.

**Example 2:** The multiplicative complexity of the XAG to realize the if-then-else function in Example 1 is 2, however, of steps only have a single input, in case the fanin of an AND gate has been defined in a similar way by the authors in [20]. In abstract XAGs from the previous examples can be represented by the abstract

\[
\begin{align*}
x_4 &= L_1 \land L_2, \\
x_5 &= L_1 \land L_3, \\
f &= L_{3,4,5},
\end{align*}
\]

and

\[
\begin{align*}
x_4 &= L_1 \land L_{2,3}, \\
f &= L_{3,5},
\end{align*}
\]

respectively.

It is straightforward to translate an XAG into an abstract XAG by merging all the XOR gates between AND gates into single linear functions. The inverse direction, however, is more complicated. A naïve translation of an abstract XAG into an XAG may lead to a large number of XOR gates, if gate sharing and cancellations are not taken into account.

### III. Finding Optimum Abstract XAGs

We solve the optimization problem of finding the multiplicative complexity of a normal Boolean function \( f \), by solving a series of decision problems that ask whether there exists an abstract XAG for \( f \) with \( r \) gates. In that series of decision problems, we can either increment \( r \), starting from some lower bound, or decrementing \( r \), starting from an upper bound. A good lower bound is \( d-1 \), where \( d \) is the algebraic degree of \( f \), which is the size of the largest monomial in its algebraic normal form [1]. A good upper bound is the number of AND-like gates in a logic network for \( f \) over binary gates.

#### A. SAT Encoding

We encode the decision problem as a SAT problem over variables \( s_{cij} \) for \( c \in \{1,2\} \), \( n \leq i \leq n+r \) and \( 1 \leq j \leq i \) that are true, whenever \( j \in S_{cij} \), and similarly, variables \( s_j \) that are true, whenever \( j \in S \) for all \( 1 \leq j \leq n+r \). Further variables \( f_{ix} \) for \( n \leq i \leq n+r \) and \( 0 < x < 2^n \) encode the function computed by AND gate \( i \) for input assignment \( b_1, \ldots, b_n \) when \( x=(b_n \ldots b_1)_2 \). It is not necessary to consider the case \( x=0 \), since \( f \) is normal. As notation we also introduce variables \( \hat{s}_{ij} \) which correspond to variables \( s_{cij} \) and \( s_j \) with respect to the definition of the sets \( S_j \) above. Note that these are not additional variables for the SAT problem.

Clauses are added to relate the variables that encode the structure of the abstract XAG with the variables that encode the function of the abstract XAG. For each \( x=(b_n \ldots b_1)_2 \), we add the gate clauses

\[
f_{ix} \leftrightarrow \bigwedge_{c \in \{1,2\}} \bigoplus_{j=1}^{i-1} (s_{cij} \land f_{jx}), \tag{5}
\]

for all \( n \leq i \leq n+r \). Here, \( f_{jx} = b_j \) whenever \( j \leq n \). We also add output clauses

\[
f(b_1, \ldots, b_n) \leftrightarrow \bigoplus_{j=1}^{n+r} (s_j \land f_{jx}). \tag{6}
\]

Note that \( f(b_1, \ldots, b_n) \) evaluates to a constant value. All clauses are added to the SAT solver by making use of the Tseytin encoding [23].

#### B. Additional constraints and symmetry breaking

The clauses in the previous section are sufficient to find abstract XAGs with a maximum number of AND gates \( r \) for some given function \( f \). In this section, we introduce additional constraints that help the SAT solver, e.g., by using redundant clauses, or by filtering out some solutions using symmetry breaking.
Non-constant linear fan-in: Unless \( f \) is the constant-0 function, all linear functions involved in the abstract XAG should contain at least one variable, i.e., \( S_l \neq \emptyset \) for all \( l \), since they would otherwise evaluate to the 0 function. We can help the SAT solver by adding the clauses

\[
\left( \hat{s}_{i1} \lor \cdots \lor \hat{s}_{il} \right)_{l \in S_l} 
\]

for all \( 1 \leq l \leq 2r + 1 \).

**Commutativity:** Since the AND operation is commutative, we can interchange \( S_{11} \) and \( S_{12} \) in any step without changing the function. We can force the SAT solver to break this symmetry by asserting that \( S_{1i} \) must be lexicographically smaller than \( S_{2i} \), i.e., \( \max\{S_{1i} \setminus \Delta S_{2i}\} \in S_{2i} \), where ‘\( \Delta \)’ denotes the symmetric difference. This constraint can be enforced by adding the clauses

\[
\bigwedge_{j=1}^{i-2} \left( (\hat{s}_{1ij} \lor \hat{s}_{2ij} \lor \hat{a}_{i(j-1)}) (\hat{s}_{1ij} \lor \hat{a}_{ij} \lor \hat{a}_{i(j-1)}) (\hat{s}_{2ij} \lor \hat{a}_{ij} \lor \hat{a}_{i(j-1)}) \right)
\]

\[
(\hat{s}_{1(i-1)} \lor \hat{a}_{i(i-2)}) (\hat{s}_{2(i-1)} \lor \hat{a}_{i(i-2)})
\]

for \( n < i \leq n + r \), and \( i = 2 \) auxiliary variables \( a_{ij} \), where all literals \( \bar{a}_{i0} \) are omitted [13].

**Symmetric variables:** A function \( f(x_1, \ldots, x_n) \) is symmetric in two variables \( x_j \) and \( x_k \), if swapping \( x_j \) with \( x_k \) in \( f \) does not change the function. In that case, we can enforce that \( x_j \) “is used before” \( x_k \), by adding the clauses

\[ \hat{s}_{lk} \rightarrow \bigvee_{1 \leq l' \leq l} \hat{s}_{lj} \]

for \( 1 \leq l \leq 2r + 1 \).

**All gates and all essential variables must be used:** A variable \( x_i \) is essential in \( f \) if \( f_{\bar{x}_i} \neq f_{x_i} \), where \( f_{\bar{x}_i} \) and \( f_{x_i} \) are the negative and positive co-factors of \( f \), which are obtained by setting \( x_i \) to 0 and 1, respectively. In an optimum solution each essential variable and each AND gate must be present in at least one of the index sets \( \hat{S}_l \). This can be enforced by the clauses

\[
\bigvee_{l \text{ s.t. } |S_l| \geq j} \hat{s}_{li}
\]

for all \( i \in [n] \) such that \( x_i \) is an essential variable, and for all \( n < i \leq n + r \).

**Linear fan-ins are not subsets:** It can be shown that there always exists a minimum abstract XAG in which \( S_{1i} \not\subseteq S_{2i} \) and \( S_{2i} \not\subseteq S_{1i} \) for all steps \( i \). This property has also been used by the authors in [20] to reduce the number of topologies in the enumeration of minimum abstract XAGs for all 6-variable Boolean functions. The proof makes use of the following lemma:

**Lemma 1:** Let \( S_1 \subseteq S_2 \), then

\[
L_{S_1} \land L_{S_2} = L_{S_1} \oplus (L_{S_1} \land L_{S_2 \setminus S_1})
\]

**Proof:** We start by expanding the left-hand side of the equality using the distributivity law:

\[
L_{S_1} \land L_{S_2} = \bigoplus_{(i,j) \in S_1 \times S_2} x_i x_j
\]

Splitting \( S_2 \) into the disjoint union \( S_1 \cup (S_2 \setminus S_1) \) leads to

\[
\bigoplus_{(i,j) \in S_1 \times S_1} x_i x_j \quad \bigoplus_{(i,j) \in S_1 \times (S_2 \setminus S_1)} x_i x_j.
\]

Since for each pair \((i, j) \in S_1 \times S_1 \) with \( i \neq j \), there also exists \((j, i) \in S_1 \times S_1 \), all monomials of degree 2 on the left-hand side cancel, simplifying the expression to

\[
\bigoplus_{i \in S_1} x_i \oplus \bigoplus_{(i,j) \in S_1 \times (S_2 \setminus S_1)} x_i x_j.
\]

Now, it is easy to see that the term on the left-hand side is the linear function \( L_{S_1} \), and the term on the right-hand side equals \( L_{S_1} \land L_{S_2 \setminus S_1} \) by laws of distributivity.

We can now formulate and prove the following theorem.

**Theorem 1:** Any abstract XAG can be expressed using steps in which \( S_{1i} \not\subseteq S_{2i} \) and \( S_{2i} \not\subseteq S_{1i} \) without changing the number of steps.

**Proof:** Assume we have an abstract XAG with \( r \) steps that does not fulfill the stated property. That is, there exists some step \( i \) for which \( S_{1i} \subseteq S_{2i} \) or \( S_{2i} \subseteq S_{1i} \). Without loss of generality, let us assume that \( S_{1i} \subseteq S_{2i} \). Then due to Lemma 1 we can rewrite that step as

\[
x_i = L_{S_{1i}} \land S_{2i} \setminus S_{1i},
\]

and replace all \( S' \in \{S\} \cup \{S_{1j}, S_{2j} \mid j > i\} \) by \( S' \Delta S_{1i} \), whenever \( i \in S' \). We repeat this process for all gates that do not fulfill the property, and since a change only affects gates with a larger index, this procedure eventually terminates.

Based on this result, we add the clauses

\[
\bigwedge_{j=1}^{i-1} (s_{1ij} \lor s_{2ij}) \land \bigwedge_{j=1}^{i-1} (\hat{s}_{1ij} \lor s_{2ij})
\]

for all \( n < i \leq n + r \) to the SAT formula, thereby ruling out abstract XAGs with the property in Theorem 1.

**Multi-level subset relation:** More symmetry breaking can be taken into account when considering subset relations among linear functions that are not from the same AND gate. For this purpose, we make use of the following lemma, which is proven in Appendix A.

**Lemma 2:** Let \( S, T, \) and \( U \) such that \( S \subseteq T, \) and \( S \subseteq U \). Then

\[
L_S \oplus (L_T \land L_U) = L_T \setminus S \oplus (L_T \land L_T \Delta U).
\]

On the left-hand side, by definition \( S \subseteq T \) and \( S \subseteq U \). However, on the right-hand side, \( T \setminus S \subseteq T, \) but \( T \setminus S \subseteq T \Delta U, \) if and only if \( T \cap U = S \). Therefore, we can restrict the number of solutions, by constraining that \( S \subseteq T \) and \( S \subseteq U \) implies that \( T \cap U = S \), by adding the clauses

\[
\bigwedge_{i=n+1}^{\mid \hat{S}_i \mid} \left( \hat{s}_{li} \land \bigwedge_{j=1}^{i-1} (\hat{s}_{ij} \leftrightarrow s_{1ij} \land s_{2ij}) \right)
\]

for all \( 1 \leq l \leq 2r + 1 \). In that case, the SAT solver will rule out the left-hand side expression in [11], unless \( T \cap U = S \).
In this section we describe different solving strategies to solve the optimization problem of finding an XAG with the minimum number of AND gates for an \( n \)-variable Boolean function \( f \). For this purpose, we make use of the following notation. Function \( F_r(x) \) describes the conjunction of the main clauses \( \{3\} \) and \( \{6\} \) for some given number of steps \( r \) and input assignment \( x \). The function \( A_r \) is a conjunction of additional and symmetry breaking constraints described in the previous section. Note that \( A_r \) does not necessarily need to contain all of these constraints. The value \( r_{\text{low}} \) is a lower bound for \( r \), and is either set to \( d - 1 \), where \( d \) is the algebraic degree of \( f \), or an application specific value. The function \( \text{solve} \) solves a SAT formula using a SAT solver and returns either \( \text{unsat} \) and is either set to \( r \) of these constraints. The value \( r \) is a conjunction of additional and symmetry breaking constraints described in the previous section. Note that \( A_r \) does not necessarily need to contain all of these constraints. The value \( r_{\text{low}} \) is a lower bound for \( r \), and is either set to \( d - 1 \), where \( d \) is the algebraic degree of \( f \), or an application specific value. The function \( \text{solve} \) solves a SAT formula using a SAT solver and returns either \( \text{unsat} \) or \( \text{sat} \), and is either set to \( r \) of these constraints. The value \( r_{\text{low}} \) is a lower bound for \( r \), and is either set to \( d - 1 \), where \( d \) is the algebraic degree of \( f \), or an application specific value. The function \( \text{solve} \) solves a SAT formula using a SAT solver and returns either \( \text{unsat} \) or \( \text{sat} \), and is either set to \( r \)

**Direct method:** The direct method solves the optimization problem as follows:

```plaintext
set \( r \leftarrow r_{\text{low}} \);
while solve(\( \bigwedge_{r=1}^{2^n-1} F_r(x) \land A_r \)) = \text{unsat} do
    set \( r \leftarrow r + 1 \);
end
return extract_xag();
```

Starting from a lower bound, it constrains all input assignments (except \( x = 0 \)), and returns one XAG for the first satisfying solution.

**Counter-example guided abstraction refinement:** In this method inspired by \[24\] we do not constrain all input assignments, but call the SAT solver incrementally, constraining new input assignments that are derived from counter-examples of wrong solutions:

```plaintext
set \( r \leftarrow r_{\text{low}} \);
while true do
    set \( F \leftarrow A_r \);
    while solve(\( F \)) = \text{sat} do
        set \( N \leftarrow \text{extract}_x\text{ag}() \);
        set \( f' \leftarrow \text{simulate}(N) \);
        if \( f' = f \) then
            return \( N \);
        else
            set \( x \leftarrow \min_x \{ f(x) \neq f'(x) \} \);
            set \( F \leftarrow F \land F_r(x) \);
        end
    end
    set \( r \leftarrow r + 1 \);
end
```

After each satisfying SAT call, we extract a candidate XAG \( N \) and simulate it to extract its function \( f' \). If \( f' = f \), then \( N \) is a minimum XAG, otherwise we refine \( F \) by constraining the smallest input assignment in which \( f \) and \( f' \) differ.

### Multiple solutions:
Next, we assume that we have found some \( r \), for which there exists an XAG that implements \( f \).

```plaintext
set \( N \leftarrow \emptyset \);
set \( F \leftarrow \bigwedge_{r=1}^{2^n-1} F_r(x) \land A_r \);
while solve(\( F \)) = \text{sat} do
    set \( N \leftarrow N \cup \{ \text{extract}_x\text{ag}() \} \);
    set \( F \leftarrow F \land \text{block}() \);
end
return \( N \);
```

While \( F \) is satisfiable and solutions can be extracted, we iteratively block solutions, by adding a single large clause that contains all variables \( s_{ij} \) in a polarity opposite to their value in the last satisfying solution.

### IV. Optimizing the Number of XOR Gates

There exist several differently structured abstract XAGs, all having the same number of steps, and all realizing the same Boolean function \( f \). Further, for each of these abstract XAGs there exists several possible ways to transform it into an XAG without changing its structure, i.e., without changing the linear functions that fan-in to the AND gates. Finding the XAG with the smallest number of XOR gates for a given number of AND gates essentially requires to solve the SAT problem on the XAG structure rather than on the abstract XAG, which makes the problem much more complicated. Instead, in this section, we propose a method that (i) heuristically determines an abstract XAG that potentially leads to a small number of XOR gates in the XAG, and (ii) finds an XAG with the smallest number of XOR gates for that particular abstract XAG exactly using a SAT-based method.

#### A. Heuristic to find good abstract XAG as starting point

The number of XOR gates is possibly small, if the linear functions in the abstract XAG have small index sets. In other words, we aim to minimize the sum \( |S_1| + \cdots + |S_{2^r+1}| \). For this purpose, we add the constraint

\[
\sum_{l=1}^{2r+1} \sum_{i=1}^{m} s_{ij} < p \tag{13}
\]

to an otherwise satisfying instance to find an abstract XAG with \( r \) gates, where \( r \) is already known to be optimum. An initial value for \( p \) can be extracted from a known solution, and then incrementally made smaller until no more satisfying solution can be found. It is a common approach to implement such a search using a sorter network, which has as inputs all variables \( s_{ij} \). We can then constrain \( \sum_{l=1}^{2r+1} \sum_{i=1}^{m} s_{ij} \) by forcing the \( p^{th} \) most significant output of the sorter network to be 0 \[25\].

#### B. Finding the shortest linear network with SAT

Next we describe a SAT-based approach that finds the smallest XAG representation in terms of XOR gates for a given abstract XAG. For this purpose, we introduce multi-output linear networks. Let \( L_{S_1}, \ldots, L_{S_m} \) be \( m \) linear functions over \( n \) variables. We can represent all functions by a single \( m \times n \) Boolean matrix \( A = (a_{ij}) \) with \( a_{ij} = [j \in S_i] \). A
linear network (or linear straight-line program) for \( m \) linear functions over \( n \) variables is a sequence of steps

\[
x_i = x_{j_1} \oplus x_{j_2},
\]

for \( n < i \leq n + r \) and a mapping \( f_l \in [n+r] \) for \( 1 \leq l \leq m \).

Given an abstract XAG over \( n \) variables and \( r \) steps, we can extract a linear network over \( n + r \) variables and \( 2r + 1 \) linear functions, where variables \( x_1, \ldots, x_{n+r} \) correspond to the original inputs and steps in the abstract XAG and the \( 2r + 1 \) linear functions are \( S_1(n+1), S_2(n+1), \ldots, S_1(n+r), S_2(n+r+1) \), \( S \), in that order. From the linear network we can then construct an XAG, where the inputs \( x_{n+1}, \ldots, x_{n+r} \) correspond to the outputs of an AND gate, and where all but the last linear function correspond to the inputs of an AND gate. To ensure that the XAG corresponds to a directed acyclic graph, we must control how steps are computed in the linear network. For example, in order to compute a function that corresponds to an input of AND gate \( x_i \), we cannot use any variable with index \( i \) or higher.

For a single linear function \( L_S \), where \( S \neq \emptyset \), the shortest linear network requires \( r = |S| - 1 \) steps. Determining the smallest linear network for a set of linear functions is more complicated, since one needs to take step sharing and step cancellations into account. Boyar, Matthews, and Peralta have shown that finding a linear network with the smallest number of steps for some set of linear functions is MaxSNP-complete [25].

Fuhs and Schneider-Kamp have presented a SAT-based approach to find the shortest linear network for a given Boolean matrix \( A \) [21]. Similarly to the approach presented in this paper, they are solving a sequence of decision problems that ask whether there exists a linear network with \( r \) steps. Their SAT encoding for this decision problem consists of variables \( b_{ij} \) that encode whether step \( x_i \) uses input or step \( x_j \) for \( n < i \leq n + r \) and \( 1 \leq j < i \), and variables \( f_{li} \) that encode whether step \( x_i \) computes output \( f_i \) for \( n < i \leq n + r \) and \( 1 \leq l \leq m \). For simplicity, we assume that each row in \( A \) has at least two 1s, i.e., no output is a constant-0 or projection function. Then, the constraints

\[
\sum_{j=1}^{i-1} b_{ij} = 2 \quad \text{for } n < i \leq n + r \tag{15}
\]

ensure that each step has two inputs and the constraints

\[
\sum_{i=n+1}^{n+r} f_{li} = 1 \quad \text{for } 1 \leq l \leq m \tag{16}
\]

ensure that each output is computed by exactly one step. The main clauses

\[
\bigwedge_{i=n+1}^{n+r} \left( f_{li} \rightarrow \bigwedge_{j=1}^{n} (\psi(j, i) \leftrightarrow a_{ij}) \right) \quad \text{for } 1 \leq j \leq m \tag{17}
\]

ensure that if the function \( f_i \) is computed by step \( x_i \), the linear function at that step matches the entries in row \( l \) of \( A \). For this purpose, the recursive function

\[
\psi(j, i) = b_{ij} \oplus \bigoplus_{i'=n+1}^{i-1} b_{ii'} \land \psi(j, i') \tag{18}
\]

encodes whether variable \( x_j \) is in the linear function computed by step \( x_i \) for \( 1 \leq j \leq n \) and \( n < i \leq n + r \). Further constraints to remove redundant solutions or break symmetries can speed up the SAT solver and are discussed in [21].

As described above, a linear network extracted from an abstract XAG must obey additional constraints to ensure that linear functions that correspond to input of gate \( x_i \) do not make use of variables that correspond to the output of a gate \( x_j \) with \( j > i \). One possibility to ensure these constraints is to force the linear network to be cancellation-free [26], however, this is unnecessarily restrictive. A better way is to forbid using some inputs in the computation of an output. For this purpose, we extend the original SAT encoding to find the shortest linear network by the recursive function

\[
\varphi(j, i) = b_{ij} \lor \bigvee_{i'=n+1}^{i-1} b_{ii'} \land \varphi(j, i'), \tag{19}
\]

for \( 1 \leq j \leq n \) and \( n < i \leq n + r \), which is similar to \( \psi(j, i) \) in (18), but captures whether variable \( x_j \) has been used in any intermediate step to compute \( x_i \). If we wish to enforce cancellation-free logic networks, we can add the constraints \( \psi(j, i) \leftrightarrow \varphi(j, i) \), however, for our application it is sufficient to simply forbid that some outputs cannot use some inputs by constraining

\[
f_{li} \rightarrow \overline{b_{ij}} \tag{20}
\]

for all \( n < i \leq n + r \), \( 1 \leq l \leq m \) and \( n + \left\lceil \frac{n}{2} \right\rceil \leq j \leq n + r \).

V. Experimental Results

We have implemented the algorithm using the EPFL logic synthesis libraries [27]. In our experiments, we used the exact synthesis algorithms to verify known multiplicative complexities for Boolean functions with up to 6 variables [9], [20]. More precisely, we found optimum XAGs for affine-equivalent classes, since the multiplicative complexity is invariant for all functions in a class. As representative function, we chose the function determined by the classification algorithm in [22]. The experiment \textit{affine}(n) finds optimum XMGs for all affine equivalent class representatives for \( n \)-variable Boolean functions, starting from a lower bound based on the function’s algebraic degree. There are 8 and 48 classes for all 4- and 5-variable Boolean functions, respectively [28]. The experiment \textit{practical6}(k) finds optimum XAGs for the \( k \) most occurring 6-variable Boolean functions found by enumerating cuts in all arithmetic and random control benchmarks of the EPFL logic synthesis benchmark suite [29]. Appendix B shows how these functions are obtained using ABC [80].

We report the total number of 2-input XOR gates, the total number of 2-input AND gates, and the overall run-time in
We presented a SAT-based algorithm to determine the multiplicative complexity of a Boolean function. The algorithm is only applicable to small Boolean functions. The multiplicative complexity is known for all Boolean functions up to 6 variables, however, the approaches that were used to determine these numbers are based on the exhaustive enumeration of all 150,357 equivalence classes. Instead, our approach can be used to find one or multiple solutions for one particular Boolean function of interest. Therefore, for functions with up to 6 variables, it is of interest when the goal is to explore different structures for the same function. For functions with more than 6 variables, it can find solutions, if the multiplicative complexity is low, or can prove that no solution with a small number of AND gates can exist.

Future insight in the structure of XAGs can be exploited as additional symmetry breaking rules in order to speed up solving time. Similarly, advances in SAT solvers have a positive impact on our proposed algorithm.

Acknowledgments: We like to thank Thomas Häner, Robin Kothari, René Peralta, Michael Miller, Bruno Schmitt, and Eleonora Testa for valuable feedback in the preparation of this manuscript.

VI. CONCLUSIONS

We presented a SAT-based algorithm to determine the multiplicative complexity of a Boolean function. The algorithm is only applicable to small Boolean functions. The multiplicative complexity is known for all Boolean functions up to 6 variables, however, the approaches that were used to determine these numbers are based on the exhaustive enumeration of all 150,357 equivalence classes. Instead, our approach can be used to find one or multiple solutions for one particular Boolean function of interest. Therefore, for functions with up to 6 variables, it is of interest when the goal is to explore different structures for the same function. For functions with more than 6 variables, it can find solutions, if the multiplicative complexity is low, or can prove that no solution with a small number of AND gates can exist.

Future insight in the structure of XAGs can be exploited as additional symmetry breaking rules in order to speed up solving time. Similarly, advances in SAT solvers have a positive impact on our proposed algorithm.

We like to thank Thomas Häner, Robin Kothari, René Peralta, Michael Miller, Bruno Schmitt, and Eleonora Testa for valuable feedback in the preparation of this manuscript.

REFERENCES

1. C.-P. Schnorr, “The multiplicative complexity of Boolean functions,” in Int’l Conf. on Applied Algebra, Algebraic Algorithms and Error-Correcting Codes, 1988, pp. 45–58.
2. L. A. Levin, “The tale of one-way functions,” Problems of Information Transmission, vol. 39, no. 1, pp. 92–103, 2003.
3. M. G. Find, “On the complexity of computing two nonlinearity measures,” in Int’l Computer Science Symposium in Russia, 2014, pp. 167–175.
4. J. Boyar, P. Matthews, and R. Peralta, “Logic minimization techniques with applications to cryptography,” Journal of Cryptology, vol. 26, no. 2, pp. 280–312, 2013.
5. E. Testa, M. Soeken, L. G. Amarù, and G. De Micheli, “Reducing the multiplicative complexity in logic networks for cryptography and security applications,” in Design Automation Conference, 2019, p. 74.
6. M. S. Turan and R. Peralta, “The multiplicative complexity of Boolean functions on four and five variables,” in Int’l Workshop on Lightweight Cryptography for Security and Privacy, ser. Lecture Notes in Computer Science, T. Eisenbarth and E. Oztürk, Eds., vol. 11738. Springer, 2019, pp. 158–167.
7. D. H. Nicolás T. Courtois and T. Mauw, “Search for the knowledge for Boolean circuits,” in Int’l Workshop on Security and Trust Management, ser. Lecture Notes in Computer Science, S. Mauw and M. Conti, Eds., vol. 11738. Springer, 2019, pp. 21–33.
8. E. Testa, M. Soeken, H. Rienen, and G. Luca Gaetano Amarù, “A logic synthesis toolbox for reducing the multiplicative complexity in logic networks,” in Design, Automation and Test in Europe, 2020.
9. M. S. Turan and R. Peralta, “The multiplicative complexity of Boolean functions on four and five variables,” in Int’l Workshop on Lightweight Cryptography for Security and Privacy, ser. Lecture Notes in Computer Science, T. Eisenbarth and E. Oztürk, Eds., vol. 8898. Springer, 2014, pp. 21–33.
10. D. H. Nicolás T. Courtois and T. Mauw, “Search for the knowledge for Boolean circuits,” in Int’l Workshop on Security and Trust Management, ser. Lecture Notes in Computer Science, S. Mauw and M. Conti, Eds., vol. 11738. Springer, 2019, pp. 158–167.
APPENDIX

A. Proof of Lemma 2

We introduce a notation and identities for a subset of quadratic forms, which will be useful for the proof. Quadratic forms are algebraic normal forms, in which all monomials have degree 2. A subset of quadratic forms can be constructed for two disjoint sets $S$ and $T$, i.e., $S \cap T = \emptyset$. We define

$$Q_{ST} = L_S \land L_T = \bigoplus_{(i,j) \in S \times T} x_i x_j.$$  \hspace{1cm} (21)

Note that $Q_{ST}$ contains of $|S| \times |T|$ monomials of degree 2, since no cancellation can take place. Also note that $Q_{ST} = Q_{TS}$. Let $S_1, S_2$, and $T$ such that all three sets are pairwise disjoint. Then

$$Q_{S_1T} \lor Q_{S_2T} = Q_{(S_1 \cup S_2)T}.$$ \hspace{1cm} (22)

Further, for any two sets $S$ and $T$, we have

$$L_S \lor L_T = L_{S \Delta T}$$ \hspace{1cm} (23)

and

$$L_S \land L_T = L_{S \lor T} \lor Q_{(S \lor T)(S \land T)} \lor Q_{(S \lor T)(T \land S)}.$$ \hspace{1cm} (24)

We prove one non-trivial lemma, based on these identities:

**Lemma 3:** Given two sets $T$ and $U$, we have $L_T \land L_{T \land U} = L_T \lor (L_T \land U).$

**Proof:** Since $T \cap (T \land U) = T \setminus U$, $T \land (T \land U) = U$, $T \setminus (T \land U) \setminus U$, and $(T \land U) \setminus T = U \setminus T$, by expanding the left-hand side of the lemma using (24), we get

$$L_{T \land U} \lor Q_{(T \setminus U)(T \land U)} \lor Q_{(T \setminus U)(U \land T)}.$$ \hspace{1cm} (25)

Applying (23) to the first two linear forms, and (22) to the first and third quadratic form gives

$$L_T \lor L_{T \land U} \lor Q_{(T \setminus U)(T \land U)} \lor Q_{(T \setminus U)(U \land T)},$$

which simplifies to $L_T \lor (L_T \land U)$ after applying (24). \hspace{1cm} ■

We can now prove Lemma 2 and show that $L_{T \land S} \lor (L_T \land L_U)$, if $S \subseteq T$ and $S \subseteq U$.

**Proof:**

$$L_{T \land S} \lor (L_T \land L_U) \overset{\text{Lemma 0}}{=} L_{T \land S} \lor L_T \lor (L_T \land L_U)$$

Since $S \subseteq T$, we have $(T \setminus S) \Delta T = S$, and applying (23) yields $L_S \lor (L_T \land L_U).$ \hspace{1cm} ■

B. Functions in practical6(n, k)

The following procedure generates the $k$ practical 6-variable functions in $\text{practical6}(n, k)$ using ABC \cite{ABC} and the arithmetic and random control instances of the EPFL logic synthesis benchmarks \cite{EPFL}.

```
$ cat functions | grep -e "6\$" | head -nk
| cut -d ' ' -f1
```

| Class Function Logic network MC |
|-------------------------------|-------------------------------|
| 0  f(x)                     | 0  |
| 1  $f(x) = x \land x \land x$ | 4  |
| 2  $f(x) = x \land x \land x \land x$ | 3  |
| 3  $f(x) = x \land x \land x \land x \land x$ | 4  |
| 4  $f(x) = x \land x \land x \land x \land x \land x$ | 3  |
| 5  $f(x) = x \land x \land x \land x \land x \land x \land x$ | 4  |
| 6  $f(x) = x \land x \land x \land x \land x \land x \land x \land x$ | 3  |
| 7  $f(x) = x \land x \land x \land x \land x \land x \land x \land x \land x$ | 4  |
| 8  $f(x) = x \land x \land x \land x \land x \land x \land x \land x \land x \land x$ | 3  |
| 9  $f(x) = x \land x \land x \land x \land x \land x \land x \land x \land x \land x \land x$ | 4  |
| 10  $f(x) = x \land x \land x \land x \land x \land x \land x \land x \land x \land x \land x \land x$ | 3  |

*TABLE II*

Optimum XAGs for all 48 affine equivalent class for 5-variable Boolean functions.