SAT Based Exact Synthesis using DAG Topology Families

Winston Haaswijk
EPFL
Lausanne, Vaud, Switzerland
winston.haaswijk@epfl.ch

Mathias Soeken
EPFL
Lausanne, Vaud, Switzerland
mathias.soeken@epfl.ch

Alan Mishchenko
University of California, Berkeley
Berkeley, California, United States
alanmi@berkeley.edu

Giovanni De Micheli
EPFL
Lausanne, Vaud, Switzerland
giovanni.demicheli@epfl.ch

ABSTRACT

SAT based exact synthesis is a powerful technique, with applications in logic optimization, technology mapping, and synthesis for emerging technologies. However, its runtime behavior can be unpredictable and slow. In this paper, we propose to add a new type of constraint based on families of DAG topologies. Such families restrict the search space considerably and let us partition the synthesis problem in a natural way. Our approach shows significant reductions in runtime as compared to state-of-the-art implementations, by up to 63.43%. Moreover, our implementation has significantly fewer timeouts compared to baseline and reference implementations, and reduces this number by up to 61%. In fact, our topology based implementation dominates the others with respect to the number of solved instances: given a runtime bound, it solves at least as many instances as any other implementation.

ACM Reference Format:

1 INTRODUCTION

In recent years, there has been a large research effort into SAT based logic synthesis methods. A notable research direction has been the synthesis of optimum Boolean chains, also known as exact synthesis [4, 16, 17]. A Boolean chain is a directed acyclic graph (DAG), in which every vertex corresponds to a 2-input Boolean operator [10]. Boolean chains are compact structures for the representation of multiple-output Boolean functions. They are similar to the concept of unbound logic networks used by the logic synthesis community. Exact synthesis has various applications in logic optimization, technology mapping, and synthesis for emerging technologies [4, 14, 16, 17].

Given a Boolean function, we can solve the problem of finding an optimum Boolean chain for that function by solving sequences of SAT formulae [3, 12]. However, the major drawback of SAT based synthesis is the unpredictable, and potentially long, runtime of SAT invocations. This runtime is perhaps unsurprising if we consider that, in finding optimum Boolean chains, the SAT solver has to simultaneously perform at least two distinct tasks:

(1) finding valid DAG structures for the Boolean chain
(2) assigning Boolean operators to the vertices in these DAGs, such that the entire sequence of the chain corresponds to the specified Boolean function

Another drawback to SAT based synthesis is that, like many logic synthesis and EDA problems, it is difficult to parallelize. Some efforts have been made in parallelizing SAT solvers using techniques such as cube-and-conquer, clause sharing, and so-called portfolio SAT solvers that apply different SAT solvers in a parallel or distributed manner [6, 7]. However, the search space is so large that the impact of using multiple threads is limited. Moreover, parallel SAT solvers based on these methods are largely domain agnostic, and do not take advantage of the structure of specific problem domains.

The main contribution of this paper is the proposal to use DAG topology families to mitigate the runtime problems of exact synthesis. These topology families can be used to provide additional constraints to the SAT solver, and thus to speed up the synthesis process. In other words, topology based synthesis is a proposal to synthesize Boolean chains (and extensions thereof) while providing additional aid to the SAT solver in performing step (1).

The rest of this paper is organized as follows. Section 2 provides some background on Boolean chains, and on finding optimum Boolean chains using SAT based exact synthesis. In Section 3 we introduce a simple generalization of Boolean chains which allows for chains that have n-input operators. Then, in Section 4, we discuss DAG topology families and introduce the related concept of Boolean fences. We discuss some theoretical properties of fences and algorithms for generating them. Next, in Section 5, we show how the conventional SAT based exact synthesis algorithm can be extended to include topological constraints. Then, in Section 6 we perform several experiments with our new synthesis algorithm and compare it to the state-of-the-art. We show that our topology based algorithm is able to obtain significant runtime reductions of up to 63.43%. Perhaps even more importantly, given the runtime budget, it also is able to solve up to 61% more problem instances. Moreover, our topology based algorithm dominates the others in terms of solved instances. Finally, we conclude the paper with a brief discussion in Section 7.
We present here a variant of Knuth's algorithm for the synthesis for
where $\leq$ $≤$ $\leq$ $≤$ $≤$
functions that each compute some values for the same set of
variables. In Knuth's algorithm, such a formula consists of the following variables, for $1 \leq h \leq m$,
$n < i \leq n + r$, and $1 \leq t \leq 2^n$:

\begin{align*}
x_{it} & : t^{th} \text{ bit of } x_i \text{ 's truth table} \\
g_{hi} : f_{h}(x_1, \ldots, x_n) = x_i \\
s_{ijk} : x_i = x_j \circ x_k \text{ for } 1 \leq j < k \leq i \\
f_{ipq} : p \circ q \text{ for } 0 \leq p, q \leq 1
\end{align*}

Here, the $g_{hi}$ variables determine which outputs point to which
steps. The $s_{ijk}$ variables determine the inputs $j$ and $k$, for each step
$i$. These are also known as selection variables. The $f_{ipq}$ encode for
all steps $i$ what the corresponding Boolean operator is.

These variables are then constrained by a set of clauses which
ensure that the chain computes the correct functions. For $0 \leq
a, b, c \leq 1$ and $1 \leq j < k < i$, the main clauses are:

\begin{align*}
(s_{ijk} \lor (x_{it} \circ a) \lor (x_{jt} \circ b) \lor (f_{jbc} \circ \bar{a}))
\end{align*}

Intuitively, these clauses encode the following constraint: if step $i$ has
inputs $j$ and $k$, and the $t^{th}$ bit of $x_i$ is $a$, and the $t^{th}$ bit of $x_j$ is $b$, and the $t^{th}$ bit of $x_k$ is $c$, then it must be the case that $b \circ c = a$.
This can be understood by rewriting the formula as follows:

\begin{align*}
((s_{ijk} \land (x_{it} \circ a) \land (x_{jt} \circ b) \land (x_{kt} \circ c)) \rightarrow (f_{jbc} \circ \bar{a}))
\end{align*}

Note that $a, b, c$ are constants which are used to set the proper
variable polarities.

Next, let $(t_1, \ldots, t_n) = t$ be the binary encoding of $t$. In order
to fix the proper output values, we add the clauses $(\bar{g}_{hi} \lor x_{it})$
or $(\bar{g}_{hi} \lor x_{it})$ depending on the value $f_{h}(t_1, \ldots, t_n)$. We also add
$s_{i} = x_{i+1} \lor x_{i+2}$ and $s_{i} = x_{i-1} \lor x_{i-2}$, so that every output points to a step in the chain and to ensure that every step has two inputs.

These are the only clauses necessary to ensure that a valid chain is found. Other clauses, such as forcing a colexicographic order on the steps, may be added to speed up synthesis. We refer the
interested reader to [11].

One can use these clause to synthesize Boolean chain as follows:

1. Initialize $r = 1$.
2. Generate the clauses for an $r$ step chain.
3. If they are satisfiable, we are done.
4. Otherwise, increment $r$ and goto 2.

The final value for $r$ represents the minimum number of steps
necessary to compute the specified functions.

### 3 GENERALIZED BOOLEAN CHAINS

In Section 2.1 we describe the common definition of Boolean chains. However, that definition includes only binary operators. We can extend the definition to include Boolean operators of arbitrary
numbers of variables. For simplicity, we will impose the restriction that the number of operands of the Boolean operators will be upper bounded by some fixed $n$.

This extension has two main motivations. First, synthesis of
chains with larger operators may be significantly faster. For example,
using Boolean chains with 3-input operators, one can efficiently classify the set of all 5-input functions using SAT based exact syn-
thesis [5], whereas this has not been achieved for 2-input operator
chains. Second, one application of exact synthesis is in technology
mapping, where we are often required to use a diverse set of logic
primitives. For example, we may be given a standard cell library and asked to find an efficient implementation of a function $f$ in

---

1For convenience, we do not consider trivial special cases such as 0-step chains here.
terms of the primitive Boolean operators defined by the standard cells. Generally, we cannot assume that the cell library contains only 2-input operators. Following the convention of [15], we denote such a set of primitives by \( \mathcal{B} \).

We extend the definition as follows. Let \( f = (f_1, \ldots, f_m) \) be a multiple-output Boolean function, such that \( f : \mathbb{B}^n \rightarrow \mathbb{B}^m \). An unbounded Boolean chain is a sequence \((x_{n+1}, \ldots, x_{n+r})\):

\[
x_i = \phi_l(x_{j(i,1)}, \ldots, x_{j(i,i(\phi_l))}) \quad \text{for } n + 1 \leq i \leq n + r
\]

such that \( 1 \leq j(i,v) < i, \) and \( 1 \leq \text{fanin}(\phi_l) \leq n \), and for all \( 1 \leq k \leq m \), either \( f_k(x_1, \ldots, x_n) = x_{l(i,k)} \) or \( f_k(x_1, \ldots, x_n) = x_{l(i,k)} \), where \( 0 \leq l(k) \leq n + r \), and \( x_0 = 0 \). We call \( \text{fanin}(\phi_l) \) the fanin of operator \( \phi_l \). For convenience, in the following we use Boolean chains to refer to this extended definition. Note that for all Boolean chains \( n < i \leq n + r \Rightarrow \phi_l \in \mathcal{B} \). For example, in the common definition of Boolean chains, we have \( \text{fanin}(\phi_l) = 2 \) for all \( i \) and \( \mathcal{B} \) is the set of all binary operators.

### 4 DAG TOPOLOGIES, FENCES, AND FAMILIES

Suppose we are given a DAG \( G = (V, E) \), and a Boolean function \( f : \mathbb{B}^n \rightarrow \mathbb{B}^m \). We may be able to transform the DAG into a Boolean chain for \( f \) by assigning the appropriate operators \( \phi_l \in \mathcal{B} \) to every vertex \( v \in V \). We call such a transformation a labeling of the graph. Recall that finding a labeling corresponds to step (2) of the synthesis process described in the introduction. Finding such a labeling may not be possible, but if it exists, a SAT solver can find it efficiently. For example, consider the 6-input function with truth table 0x9ef7ad9c7193a0f. The smallest known implementation of this function uses 19 2-input gates. When the solution topology is obtained, the labeling may not be possible, but if it exists, a SAT solver can find it efficiently.

The efficiency of labeling may inspire one to think of the following (naive) synthesis algorithm:

```plaintext
function SYNTHESIZE(f)
    while true do
        G ← NextDag()
        if LabelingExists(G, f) then
            Chain ← LabelGraph(G, f)
            return Chain
        end if
    end while
end function
```

Such an algorithm reduces to efficiently finding a DAG with the proper structure for \( f \). However, in general, given \( f \) we do not know a priori which DAG structures have a labeling for \( f \). Given an \( n \)-input function, finding a suitable DAG requires us to search a very large space of DAG structures. Unfortunately, the enumeration of potential DAGs in this space generally outweighs the potential efficiency of graph labeling.

Alternatively, instead of providing the SAT solver with a DAG to label, we can specify a set of clauses which constrain the SAT solver’s search to a particular family of DAG topologies. We then use the SAT solver’s efficient search heuristics to find only those topologies within that family. This approach avoids explicit enumeration of DAGs and provides a middle ground between the unstructured exact synthesis formulation of Section 2.2 and the fully structured labeling of graphs. The idea to explore this middle ground is the main contribution of this paper.

![Figure 2](image.png)
number of nonnegative integer-valued solutions to the equation
\[ x_1 + x_2 + \cdots + x_l = k - l \]
and hence
\[ |\mathcal{F}(k, l)| = \binom{k-1}{l-1}. \tag{1} \]
We can now use Formula 1 to count the total number of fences of \( k \) nodes, \( |\mathcal{F}_k| \) as follows:
\[ |\mathcal{F}_k| = \sum_{i=1}^{k} \binom{k-1}{i-1} = 2^{k-1} \]

The reader may verify that these formulas correctly predicts the numbers of fences in Figure 2.

4.3 Generating Fences

Our synthesis method requires an efficient algorithm for the generation of fences. Suppose we want to generate \( \mathcal{F}_k \). In order to do this, we first observe that the number of fence families in \( \mathcal{F}_k \) closely correspond to different integer partitionings of \( k \). Recall that, given an integer \( k \), an integer partition of \( k \) is a way of writing \( k \) as the sum of positive integers \( k_1 + \cdots + k_l = k \). We can obtain a fence from such a partition by imposing an order on it. Let \( S \) be the multiset of integers corresponding to an integer partition of \( k \), and let \( l = |S| \). Now, we can create a fence \( F \in \mathcal{F}(k, l) \) from this partition by fixing \( F = (k_1, \ldots, k_l) \) where \( k_i \in S \) \((1 \leq i \leq l)\). Note that \( S \) is a unique partition of \( k \). However, \( F \) may not be the only fence corresponding to this partition. To see why, let \( \pi \) be a permutation of \( I \). Then, the fence \( F' = F_{\pi} = (k_{\pi(1)}, \ldots, k_{\pi(n)}) \) is also a fence in \( \mathcal{F}(k, l) \).

Thus, to generate all fences in \( \mathcal{F}_k \), we have to do the following:
- Generate all integer partitions \( S \) of \( k \).
- For all such \( S \), generate all permutations \( \pi_S \).

In practice, we are often not interested in enumerating all \( 2^{k-1} \) fences in \( \mathcal{F}_k \). Instead, we are often satisfied once we obtain a fence that our synthesizer finds a solution with. All of this suggest the lazy fence generating algorithm in Algorithm 1. The algorithm presented here is a coroutine that may be called repeatedly and yields all fences in \( \mathcal{F}_k \) until exhausted. The algorithm is constructed by composing standard integer partitioning and permutation algorithms. In our implementation we use a lazy adaptation of the integer partition algorithm from Knuth [10] (fascicle 3, page 38), which attributes it to Hindenburg [8]. For the permutations we use an algorithm from the C++ standard library.

Algorithm 1 is an efficient procedure that may be function in the inner loop of a synthesis algorithm. For example, we can generate set \( \{\mathcal{F}_k \mid k \leq 10 \} \) in 0.097 seconds. On top of this basic procedure we can also build more sophisticated algorithms, such as algorithms that filter out any fences that are unnecessary for a specific synthesis task. We discuss such methods in Section 5.

5 EXACT SYNTHESIS USING FENCES

We have seen how fences correspond to families of DAG topologies, investigated some of their theoretical properties, and presented a fence generating algorithm. In this section we consider how to use fences to accelerate exact synthesis by using them to provide additional constraints in the SAT formulation. To do so, let us first look at some connections between fences and Boolean chains.

Consider a fence \( F = (\lambda_1, \ldots, \lambda_l) \). Let \( G = (V, E) \) be a DAG, and let \( \tau(v) : V \to \mathbb{N} \) be the function that assigns each vertex from \( G \) to its level. Let \( \tau_1 = \{v \mid \tau(v) = i\} \). We say that \( G \) satisfies \( F \) if and only if \( |\lambda_i| = \tau_1 \). In other words, a DAG satisfies the topological constraints of a fence if its distribution of nodes across levels is the same. We say that a Boolean chain satisfies \( F \) if its underlying DAG structure satisfies \( F \). We consider the primary inputs of the chain to have level 0, and do not consider them in satisfying \( F \).

For example, consider the fence \( F = (\lambda_1, \lambda_2) \in F(4, 2) \) highlighted in Figure 3(a). We have numbered its nodes to make them easier to distinguish. Intuitively, only DAGs with two nodes on the first level and two nodes on the second level satisfy \( F \). For example, Figure 3(b) is a 2-input operator Boolean chain satisfying the constraints from \( F \). Similarly Figure 3(c) is a 3-input Boolean chain that satisfies \( F \). However, Figure 3(d) shows a chain that is invalid for \( F \). It violates the constraint that the step corresponding to fence node 4 be on level 2.

Observe that the topology constraints captured by fences are independent of number of inputs, or operator fanin. This is desirable, as it implies that the same fence generator can be used as the basis for synthesis of generalized Boolean chains and functions of arbitrary input size.

Now consider again the arbitrary fence \( F = (\lambda_1, \ldots, \lambda_l) \in F(k, l) \). Suppose we wish to synthesize a Boolean chain that satisfies \( F \). We know that it must be a \( k \)-step chain. We assign step \( i \) to level \( t \) by setting
\[ \tau(x_i) = t \iff t = \min_{t'} \sum_{j=0}^{l'} |\lambda_{2j}|. \]
where \( |\lambda_0| = n \), the number of primary inputs.

Note that if \( \tau(x_i) = t \), then step \( x_i \) must, by definition, have at least one fanin on level \( t - 1 \). Thus, the fence constrains not only
we can do better. Observe that, since none of the variables outside the way to express the constraints imposed by the fence is by adding, $F$ search. For instance, Figure (b) and Figure (c) satisfy the constraints from $F$ below. This approach is similar to the way that colexicographic clauses analogous to the one described in Section 2.2. However, we can do better. Observe that, since none of the variables outside of $\{s_{ijk} \mid k' \leq k \leq k''\}$ may be true, we do not need to include them in our SAT formula at all. Thus, by using a fence we can significantly reduce the number of variables and clauses in our SAT instances.

To implement exact synthesis with topological constraints we can then proceed as follows: (i) Generate a new fence using the procedure described in Algorithm 1. (ii) Using the constraints implied by the fence, generate a reduced SAT formula. We use a set of clauses analogous to the one described in Section 2.2. However, we exclude any variables or clauses that are rendered unnecessary due to the fence constraints, obtaining a simpler SAT formula. (iii) If the formula is satisfiable, we are done. (iv) Otherwise, go to (i).

The fence generation procedure in Algorithm 1 can be easily modified to yield a stream of fences of increasing size. Thus, we extend the conventional exact synthesis algorithm, while decomposing the search space using increasingly large families of graph topologies. Recall that in Section 4.2 we derived the total number of fences of $k$ nodes. Given an upper bound on the number of nodes to realize a function, we therefore also have an upper bound on the number of decomposed exact synthesis instances we have to solve.

6 EXPERIMENTS

In order to evaluate the performance of our proposed approach, we find exact synthesis for the following collection of Boolean function sets:

- **NPN4**: All 222 4-input NPN classes [9].
- **PDSD8**: 100 partially-DSD decomposable 8-input functions.
- **FDSD8**: 100 fully-DSD decomposable 8-input functions.
- **PDSD6**: 100 partially-DSD decomposable 6-input functions.
- **FDSD6**: 100 fully-DSD decomposable 6-input functions.
- **PDSD8**: 100 partially-DSD decomposable 8-input.

We compare five different SAT based exact synthesis approaches, including a state-of-the-art reference implementation from the well known ABC logic synthesis package [1].

- **SYM**: A baseline implementation of the SAT-based exact synthesis algorithm, which is very similar in architecture and performance to ABC’s state-of-the-art implementation (command exact, [17]).
- **SYM-CEGAR**: Approach SYM with CEGAR loop to add the main clauses incrementally [2].
- **TOP**: Our proposed algorithm based on fence enumeration and the use of additional topological constraints.
- **TOP-CEGAR**: Approach TOP with CEGAR loop to add the main clauses incrementally.
- **ABC**: The state-of-the-art reference implementation in ABC.

Table 1 lists all experimental results. For each approach four values are listed: i) the mean solving time (mean), ii) the standard deviation (dev), both in milliseconds, iii) the number of instances that could not be solved before timing out (# t/o), and the number of instances that were successfully solved (# ok). Note that the number of solved instances is the most important metric here, as it captures in essence how practical an algorithm is. Given a bound on runtime, we obviously prefer the algorithm that can solve the most problems within that bound. A similar metric is commonly used in SAT solver competitions. In our experiments we limit runtime by specifying a conflict limit to each SAT instance. All experiments were performed on an Intel Xeon E5-2680 v3 (Haswell) 2.5 GHz processor with a 30 MB cache and 256GB of RAM.

The results in Table 1 show that using topological structure enumeration can significantly improve the solving time, as well as the number of solved instances. For NPN4, our topology based
All algorithms find the solutions for all problem instances. For $FDS6$, $TOP$-CEGAR is $57.09\%$ faster than $SYM$-CEGAR and $61.63\%$ faster than $ABC$. Again, there are no timeouts. For $FDS6$, $TOP$ is $52.35\%$ faster than $TOP$ and $62.02\%$ faster than $ABC$. We can also solve $50\%$ more instances than both other approaches before timing out. The same observation can be made for the 8-input function sets. For $FDS8$, $TOP$-CEGAR is up to $51.85\%$ faster than $SYM$-CEGAR. For $FDS8$, $TOP$-CEGAR is $63.43\%$ faster than $SYM$-CEGAR and $42.68\%$ faster than $ABC$. Again, our algorithm has fewer timeouts than the others. Note especially the striking difference in timeout percentage with the ABC implementation. Our algorithm has $61\%$ fewer timeouts than ABC. In fact, the table shows that it dominates the other implementations with respect to the number of solved instances.

In summary, we can see that the gains from using topological constraints can be quite significant. They seem to be particularly beneficial as the functions to be synthesized become larger and harder to synthesize.

### 7 DISCUSSION & FUTURE WORK

This paper takes a new look at the difficult problem of SAT-based exact synthesis for Boolean functions. The SAT-based exact synthesis formulation must encode both the structure of the logic network and the functionality of its nodes. These are cleverly encoded using SAT clauses, however, their combination makes the problem very hard to solve. We find that knowing the structure makes the problem significantly easier. Based on this observation, we introduce a SAT-based exact synthesis method based on topological structure enumeration. Since the number of topological structures grows very quickly as the number of gates increases, we collect a set of structures in what we call a Boolean fence. The paper introduces a theory of Boolean fences, illustrates how they are enumerated, and shows how they are used to constrain the SAT-based exact synthesis encoding.

We evaluate our new approach to find optimum logic networks for various sets of practical Boolean functions. Our experimental results show a runtime reduction of up to $63.43\%$. Further, we find many solutions for instances in which the state-of-the-art method, which does not make use of topological constraints, times out. We reduce the number of timeouts by up to $61\%$. Our topology based algorithm dominates the other implementations in the number of solved instances: given a runtime budget it always solves at least as many as any other implementation. This has direct impact on a variety of logic optimization algorithms that use exact synthesis, such as logic rewriting, technology mapping, and synthesis for emerging technologies[4, 14, 16, 17].

In future work, we plan to find enumerate topological structures according to different strategies to reach further speed-ups. Furthermore, DAG topologies provide a natural way of partitioning the synthesis search space: only parts of the space containing potentially valid topologies needs to be examined. An interesting line of investigation would be to use graph topologies as the basis for parallel SAT based exact synthesis algorithms, using the graph topologies as a source of parallelism.

### ACKNOWLEDGMENTS

This work was supported in part by SRC contract 2710.001 “SAT-based methods for scalable synthesis and verification” at UC Berkeley, H2020-ERC-2014-ADG 669354 CyberCare, and the Swiss National Science Foundation (200021-169084 MAJesty).

### REFERENCES


