A Canonical-based NPN Boolean Matching Algorithm Utilizing Boolean Difference and Cofactor Signature

Juling Zhang, Guowu Yang, William N. N. Hung, and Jinzhao Wu

Abstract—This paper presents a new compact canonical-based algorithm to solve the problem of single-output completely specified NPN Boolean matching. We propose a new signature vector Boolean difference and cofactor (DC) signature vector. Our algorithm utilizes the Boolean difference, cofactor signature and symmetry properties to search for canonical transformations. The use of symmetry and Boolean difference notably reduces the search space and speeds up the Boolean matching process compared to the algorithm proposed in [1]. We tested our algorithm on a large number of circuits. The experimental results showed that the average runtime of our algorithm 37% higher and its average search space 67% smaller compared to [1] when tested on general circuits.

Index Terms—Boolean Matching, Boolean difference, NPN equivalent, Symmetry.

I. INTRODUCTION

Judging whether two Boolean functions are equivalent under input negation and/or input permutation and/or output negation (NPN) is an important problem applied in integrated circuit design, logic synthesis, logic verification, and so on. In cryptography, affine equivalence is used to resolve S-box problem [2]. Technology mapping is the process of selecting logic gates from a library to implement a Boolean circuit [1]–[5]. However, every Boolean function has many NPN-equivalent Boolean functions. Technology mapping searches for an optimal combination of logic gates in terms of area, performance and power dissipation. In the cell-library binding process, some of the cells of a library are found to realize some part of a multiple-level representation of a Boolean function [6]. Two Boolean functions \( f \) and \( g \) are NPN equivalent if a transformation \( T \) exists that can transform \( f \) to \( g \). In technology mapping and cell-library binding, Boolean matching is a key step.

In recent years, numerous methods have emerged to solve the NPN Boolean matching problem. The main approaches focus on the canonical-based, pairwise and SAT-based algorithms. In the canonical-based algorithms, two Boolean functions are NPN-equivalent when they have the same canonical form. An NPN-equivalent class has a canonical representative, and all the Boolean functions in an NPN-equivalent class can be transformed to this canonical representative. The canonical representative may have a maximal truth table or a minimal truth table or a maximal signature vector [1], [6]–[12].

A pairwise Boolean matching algorithm searches the correspondence relation of the variables of two Boolean functions by their signatures. In the pairwise Boolean matching process, the search operation terminated when it finds a transformation \( T \) that can transform one Boolean function into the other. The authors of [13]–[16] presented a pairwise Boolean matching algorithm.

The NPN Boolean matching problem is converted to one that involves solving Boolean satisfiability (SAT) problem. To check whether a Boolean function \( f \) is equivalent to a Boolean function \( g \), the SAT-based algorithm first constructs a circuit with the functionality \( f \oplus g \) and then generates a SAT instance (circuit CNF). Finally, it judges the Boolean satisfiability of this circuit CNF [17]. When the circuit CNF is satisfied, \( f \) is not NPN-equivalent to \( g \). In contrast, \( f \) is NPN-equivalent to \( g \) when the circuit CNF is not satisfied.

It is well known that the search space complexity of NPN Boolean matching is \( n!2^{n+1} \) using the exhaustive method. Therefore, the exhaustive method is computationally infeasible when the number of inputs is large, and thus, it is important to reduce the search space regardless of what approach is used. Many properties of Boolean functions are used to reduce the search space, including positive unate, negative unate, cofactor and symmetry, and so on.

This paper studies canonical-based NPN Boolean matching for single-output specified Boolean functions. We exploit a new compact canonical form using a DC signature vector. The canonical representative of an NPN equivalent class has the maximal DC signature vector. The use of the symmetry class, independent class and Boolean difference signature dramatically reduce the search space of the canonical transformation. The experimental results reflect that using the Boolean difference signature is highly conductive for distinguishing variables. The proposed algorithm enhances the NPN Boolean matching speed and reduces the search space significantly.

The remainder of this paper is organized as follows. In Section II, we survey the state of the Boolean matching problem. Section III introduces related preliminaries and definitions used in our algorithm. Section IV presents our canonical-based algorithm in detail. We demonstrate the effectiveness of our algorithm by presenting experimental results in Section V, and
we summarize our work and outline future work in Section VI.

II. RELATED WORKS

Numerous scholars have contributed to this problem in prior Boolean matching studies. The authors of [1]–[12] studied the canonical-based Boolean matching algorithm and proposed many practical methods for computing the canonical representative. The authors of [1] studied the P-equivalence problem and proposed a formal framework that unified the spectral and canonical-based approaches. They found a linear transformation to improve the speed of P-equivalence matching. Chai and Kuehlmann of [8] designed a fast Boolean matching that used satisfied counts to assign the phases of the Boolean function and its variables, the sums of rows or columns to search permutation, and symmetry to refine transformations. In [6], a canonical-based algorithm was presented that used table look-ups and a tree-based breadth-first search to find the canonical representative with the smallest value in the binary representation. Ciric and Sechen [9] proposed the minimum-cost canonical form for P-equivalence matching. In [10], the authors unified many multiple canonical-based approaches and proposed a new P-equivalence matching algorithm. Generally, in the canonical-based Boolean matching algorithms, the canonical representative has the maximal or minimal truth table. Abdollahi and Pedram [1] exploited a new canonical form using signature vector to resolve the NPN Boolean matching problem. The algorithms proposed in [11] and [12] used general symmetry and higher-order symmetry to realize fast NPN Boolean classification. They classified a number of Boolean functions with 6-16 inputs.

Many Boolean matching approaches have emerged based on SAT. In a SAT-based Boolean matching algorithm, Boolean matching is converted to a SAT or UNSAT problem. The authors of [17]–[22] all studied Boolean matching using the SAT technique. The authors of [13] proposed PP-equivalence matching based on graphs, simulation and SAT; their approach can be applied to large-scale circuits. Matsunaga [19] presented a Boolean matching method for LUT-based circuits by using one-hot encoding and the CEGAR technique to speed up Boolean matching [19]. The authors of [20] exploited architectural symmetry in PLB and proposed a SAT-based Boolean matching algorithm to resolve FPGA technology mapping. Wang et al. [17] integrated the simulation and SAT technique, and proposed a P-equivalent Boolean matching approach.

A pairwise Boolean matching algorithm was presented in [13] for multiple-output Boolean functions with don’t care sets and applied Boolean matching to technology mapping to optimize circuits. Abdollahi [14] proposed a signature-based Boolean matching to resolve NPN-equivalent problem for single-output incompletely specified Boolean functions. They tested their Boolean matching approach on 4-10 inputs Boolean functions. The authors of [15] utilized a level-first strategy and a set of filters to reduce the search space in the pairwise Boolean matching process, while [16] unified the structural signature of Boolean functions and Shannon expansion to exploit a pairwise Boolean matching algorithm.

A few other methods exist for solving the Boolean matching problem. In [23], the authors studied NPN Boolean matching using a Walsh Spectral Decision Diagram, and achieved an efficiency higher than that of Luks’ hypergraph method. Lai et al. [24] utilized a conflict-driven learning method to solve the multiple-output Boolean function matching problem.

III. PRELIMINARIES AND PROBLEM STATEMENT

Let \( f(x_1, x_2, \cdots, x_n) \) be a single-output completely specified Boolean function and \( X = \{x_1, \cdots, x_n\} \) be a vector of \( f \). \( |f| \) denotes the number of minterms in \( f \), which is also called the 0th-order signature.

\[
X_{\pi}(x_1, \cdots, x_n) = (x_{\pi(1)}, \cdots, x_{\pi(n)})
\]

is an input permutation and \( X_{\phi}(x_1, \cdots, x_n) = (x_1^{\phi(1)}, \cdots, x_n^{\phi(n)}) \) is an input negation. \( x_i^{\phi(i)} = x_i \) when \( \phi(i) = 1 \), and \( x_i^{\phi(i)} = \overline{x_i} \) when \( \phi(i) = 0 \). NPN equivalence is introduced by \( f \equiv g \iff f = (g \circ X_{\pi} \circ X_{\phi})^\circ \), where \( \phi \in \mathbb{B}^n (\phi \in \mathbb{B}) \).

Definition 1: (NP transformation) An NP transformation, \( T \), is an onto mapping where \( T = \{x_1, \cdots, x_n\} \mapsto \{x_1^{\phi(1)}, \cdots, x_n^{\phi(n)}\} \).

For each Boolean function \( f \) in an NPN-equivalent class with a canonical representative \( F \), there must be an NP transformation \( T \) that can transform \( f \) to \( F \), (i.e., \( f(T X) = F \) or \( T(f X) = F \)). The \( T \) is called a canonical transformation. Boolean function \( f_1 \) is NPN-equivalent to \( f_2 \) if and only if they have the same canonical representative (i.e., \( F_1 = F_2 \)).

A cofactor is a generalized signature of a Boolean function that has been used in many Boolean matching algorithms.

Definition 2: (Cofactor signature): The cofactor signature of \( f \) with respect to \( x_i(\pi_i), |f_{x_i}(\pi_i)(f_{x_i})| \), is the onset size of \( f_{x_i}(\pi_i)(f_{x_i}) \), where \( f_{x_i} = f[x_i \leftarrow 1] \) and \( f_{x_i} = f[x_i \leftarrow 0] \). The cofactor signature of \( f \) with respect to a cube \( b, |f_b| \), is the onset size of \( f_b \), where \( f_b = f[b \leftarrow 1] \). If the cube \( b \) has \( k \) variables, \( |f_b| \) is denoted as the \( k \)-th order signature [1].

A Boolean function may have some independent variables; however, the value of a Boolean function is irrelevant to these independent variables. The negative cofactor and positive cofactor of a variable are used to check its independence.

Definition 3: (Boolean difference signature) The Boolean difference signature of \( f \) with respect to \( x_i, |f_{x_i}|, |f_{\overline{x_i}}| \), is the onset size of \( f_{x_i} \), where \( f_{x_i} = f_{x_i} \) or \( f_{\overline{x_i}} \).

A variable \( x_i \) is said to be an independent variable when it satisfies \( |f_{x_i}| = 0 \). The independence property of an independent variable does not change under NP transformation.

Definition 4: (Variable symmetry): The variables \( x_i \) and \( x_j(\pi_j) \) of a Boolean function \( f \) are symmetric if \( f \) is invariant after swapping \( x_i \) and \( x_j(\pi_j) \).

Two NP transformations are equal when two or more symmetric variables are permuted. Therefore, symmetry is always used to reduce the search space for Boolean matching.

IV. CANONICAL-BASED BOOLEAN MATCHING

The goal of our algorithm is to find the canonical transformation \( T \) of a Boolean function \( f \), which can transform \( f \) to \( F \) that has the maximal DC signature vector. During the search process, we search many candidate canonical transformations by sorting and grouping variables. Then, we compare the DC
signature vectors of these candidate canonical transformations to find the target canonical transformation \( T \).

A. The Proposed Canonical Form

The canonical form presented in [1] has the maximal signature vector. For some Boolean functions, many variables have the same cofactor signature and are not symmetric variables. In this case, the algorithm proposed in [1] will generate multiple splitting ways, which increases the search space of canonical transformation. Therefore, to reduce the search space, further methods to distinguish these variables are required.

Although the variables of many Boolean functions have the same cofactor signature, their Boolean difference signatures are the different. Therefore, we proposes DC signature vector that can resolve the shortcoming in [1] described above.

**Definition 5:** (\( \prec \)) Let “\( \prec \)” denote the lexicographic comparison of two vectors.

**Definition 6:** (1st DC signature value) The 1st Boolean Difference and Cofactor (DC) signature value of a Boolean function \( f \) with respect to \( x_i \) is a two-tuple \((|f_{x_i}|, |f'_{x_i}|)\).

The 1st DC signature values of the two variables \( x_i \) and \( x_j \) of a Boolean function \( f \) are \((|f_{x_i}|, |f'_{x_i}|)\) and \((|f_{x_j}|, |f'_{x_j}|)\), respectively. When they satisfy one of the two following cases, then \((|f_{x_i}|, |f'_{x_i}|) \prec (|f_{x_j}|, |f'_{x_j}|)\) and the relation of \( x_i \) and \( x_j \) is \( x_i < x_j \).

1. \(|f_{x_i}| < |f_{x_j}|\)
2. \(|f_{x_i}| = |f_{x_j}|\) and \(|f'_{x_i}| < |f'_{x_j}|\).

**Definition 7:** (kth-order DC signature value) The kth-order DC signature value of a Boolean function \( f \) with respect to a cube \( b = b_1 \cdots b_{ik} \) is \((|f_{x_1 \cdots x_{ik}}|, |f'_{x_1 \cdots x_{ik}}|)\).

**Definition 8:** (DC signature vector) The DC signature vector of an n-input Boolean function \( f \) is

\[
D_f = \{|f|, |f_{x_1}|, |f'_{x_1}|, \ldots, |f_{x_n}|, |f'_{x_n}|, |f_{x_1x_2}|, |f'_{x_1x_2}|, \ldots, |f_{x_{n-1}x_n}|, |f'_{x_{n-1}x_n}|, |f_{x_1x_2 \cdots x_{n-k}}|, |f'_{x_1x_2 \cdots x_{n-k}}|, |f_{x_1 \cdots x_{n-k}}|, |f'_{x_1 \cdots x_{n-k}}|\},
\]

which is composed of its kth-order signature, 1st DC signature values and higher order DC signature values up to the nth-order signature.

The authors of [1] proved that each Boolean function has a unique signature vector. Therefore, each Boolean function also has a unique DC signature vector. We need not prove that here.

**Definition 9:** (The canonical representative) The canonical representative of an NPN-equivalent class \( EC = \{f_1, \cdots, f_m\} \) is the Boolean function that has the maximal DC signature vector.

Assume that two Boolean functions \( f \) and \( g \) have the DC signature vectors \( D_f \) and \( D_g \). When comparing \( D_f \) and \( D_g \), we compare their 0th-order signature first, then their 1st DC signature values, 2nd-order DC signature values and higher DC signature values until an inequality is encountered. \( D_f \prec D_g \) when the corresponding DC signature value of \( D_f \) is less than that of \( D_g \). If and only if \( D_f \prec D_g \), the order relation of \( f \) and \( g \) is \( f < g \).

**Definition 10:** (The maximal canonical transformation) The transformation \( C_f = (t_1, t_2, \cdots, t_n) \) that has the maximal DC signature vector is the maximal canonical transformation.

In this paper, we use \( T = (t_1, t_2, \cdots, t_n) \) to denote a candidate canonical transformation searched in a canonicalized process where \( t_i = \chi^{(i)}_{\pi(i)} \). For a candidate canonical transformation \( T \), the DC signature vector is as follows:

\[
D_T = \{|f|, |f_{t_1}|, |f'_{t_1}|, \ldots, |f_{t_n}|, |f'_{t_n}|, |f_{t_1t_2}|, |f'_{t_1t_2}|, \ldots, |f_{t_{n-1}t_n}|, |f'_{t_{n-1}t_n}|, \ldots, |f_{t_1 \cdots t_n}|, \ldots, |f'_{t_1 \cdots t_n}|, |f_{t_1 \cdots t_{n-1}}|, |f'_{t_1 \cdots t_{n-1}}|\}
\]

In the canonicalized process, multiple candidate canonical transformations similar to \( T \) exist. The maximal canonical transformation \( C_f \) is the candidate canonical transformation that has the maximal DC signature vector. Consider two candidate canonical transformations \( T_1 \) and \( T_2 \): \( T_1 < T_2 \) when \( D_{T_1} \prec D_{T_2} \).

For an n-input Boolean function \( f \), \( n!2^n \) NP transformations exist. The canonical transformation \( T \) of a Boolean function \( f \) is the transformation that can transform \( f \) to \( F \) when \( F \) has the maximal DC signature vector in its NPN-equivalent class. We use \( C_f \) to express the maximal canonical transformation of Boolean function \( f \) found in the canonical process. The input phase assignment and reindexing transformation is \( X = (x_1, \cdots, x_n) \) to \( C_f = (t_1, \cdots, t_n) \), where \( F(C_f) = f(X) \). The relation between \( f \) and \( F \) can be expressed by an NP transformation \( T \), where \( X = T C_f \). Therefore, the canonical form of a Boolean function \( f \) is \( f(TX) \), and \( f(TX) = f(C_f^{-1}X) \).

The key step of our algorithm is sorting and grouping the variables according to their DC signature values. After each grouping, a result group is obtained. We denote \( G \) as the group result of \( f \), and \( G = \{G_1, \cdots, G_m\} \), where \( m \in \{1, \cdots, n\} \).

Every group \( G_i \) has one or multiple classes, and each class has one or multiple variables. We use \( G_i = \{C_{i1}, \cdots, C_{ik}\} \) denote a group with \( k \) classes.

In the process of grouping variables, each group may have many classes. We categorize these as asymmetric, symmetric and independent classes.

**Definition 11:** (Asymmetric class) Every asymmetric variable in a group is an asymmetric class.

**Definition 12:** (Symmetric class) A variable set \( C_i = \{x_{i1}, \cdots, x_{ik}\} \) and \( k \in \{2, \cdots, n\} \) is a symmetric class if any arbitrary two variables in \( C_i \) are symmetric.

**Definition 13:** (Independent class) A variable set \( C_i = \{x_{i1}, \cdots, x_{ik}\} \) and \( k \in \{1, \cdots, n-1\} \) is an independent class if every variable in \( C_i \) is an independent variable.

B. Compute Canonical Form

The maximal canonical transformation \( C_f \) is found after comparing all the candidate canonical transformations. We first try to use the 1st DC signature value to group the variables. When the 1st DC signature values cannot resolve all the groups, we use the 2nd-order DC signature values and higher DC signature values until all the groups are resolved. Then, our algorithm generates a candidate canonical transformation.

Before search candidate canonical transformation, our algorithm need to do some preparatory works including phase assignment, symmetry checking and independent variable checking. The method of determining the phase of Boolean function and variable is similar to [1].
The first comparison values are $|f|$ and $|\overline{f}|$. If $|f| > |\overline{f}|$, there is no negation to output; otherwise, there is a negation to the output. When $|f| = |\overline{f}|$, our algorithm need to test $f$ and $\overline{f}$.

The phase of a variable is obtained by comparing the size of its positive and negative cofactors. The phase of $x_i$ is positive when $|f_{x_i}| > |\overline{f}_{x_i}|$, and negative when $|f_{x_i}| < |\overline{f}_{x_i}|$. When $|f_{x_i}| = |\overline{f}_{x_i}|$, we need to use a higher signature to determine its phase. If the higher signature cannot determine its phase, we must try both positive and negative.

If variable $x_i$ of a Boolean function $f$ is an independent variable, the phase of $x_i$ is assigned as positive when $|f_{x_i}| = |\overline{f}_{x_i}|$, because $x_i$ is independent to the Boolean function. If $f$ has $k$ independent variables and their phases cannot be determined by comparing the positive and negative cofactors, we may try two phases; which doubles the search space. Therefore, checking independent variables is a necessary operation.

Procedure 1 Compute Canonical Form

**Input:** $f$

**Output:** $F$

**function** CANONICAL($f$)

CT_List=NULL

Create Boolean function $f$ and compute $|f|$

if $|f| < |\overline{f}|$ then

$f = \overline{f}$

end if

Compute the 1st DC signature of $f$

Determine the phase of $x_1, \ldots, x_n$

Check the independent variables

Check the symmetric variable

m=group($f$)

SEARCH($f$, CT_list, group, m)

if $|f| = 2^{n-1}$ then

$f = \overline{f}$

Compute the 1st difference signature of $f$

Determine the phases of $x_1, \ldots, x_n$

Check the independent variables

Check the symmetry variable

m=group($f$)

SEARCH($f$, CT_list, group, m)

end if

$T = C_f^{-1}$

Return $f(TX)$

end function

The purpose of Procedure 1 is to compute the canonical form of a Boolean function $f$. The individual tasks in Procedure 1 are as follows.

1) Compute $|f|$ and determine the phase of $f$.
2) Compute the 1st DC signature values of all variables of $f$.
3) The 1st DC signature value determines the phases of all variables. Variables whose phases are not determined are handled in Procedure 2. The method used is the same as that described in [1].
4) Check the symmetry of all variables.
5) Check the independent variables. If the phases of independent variables are not determined, assign the positive phase to them.
6) Group all variables by comparing their 1st DC signature values.
7) Call Procedure 2 to get the maximal canonical transformation $C_f$.
8) Compute canonical form by the maximal canonical transformation $C_f$.

**Example 1:** Consider a 7-input Boolean function $f(X) = x_3\overline{x}_5x_2 + \overline{x}_1x_2x_4 + \overline{x}_1x_2x_4x_5x_6 + x_3x_4x_5x_6 + x_2x_3x_4x_5x_6 + \overline{x}_2x_3x_4x_5x_6$, whose phase assignment and group variables are as follows.

Procedure 1 computes $|f| = 46$ and assigns a positive phase to $f$. Then, it computes the 1st DC signature value of all the variables of $f$, and the results are $\{(16,28), (16,28), (30,28), (22,44), (24,44), (15,32), (30,28)\}$. The variables $x_3$, $x_5$, and $x_7$ are positive, and the others are negative.

Check the symmetry of every variable. There are two symmetric class $\{x_1, x_2\}$ and $\{x_3, x_7\}$. Group the variables according to the 1st DC signature value. The grouping result is $G = \{G_1, G_2, G_3\}$, and $G_1 = \{x_6\}$, $G_2 = \{x_{21}, x_{22}\} = \{x_1, x_2\}$, $x_3, x_7\}$ and $G_3 = \{C_{31}, C_{32}\} = \{x_4, x_5\}$.

**C. Searching the Canonical Transformations**

After obtaining the initial group results, Procedure 2 begins to search candidate canonical transformations. Procedure 2 first addresses the first group $G_1$, then group $G_2$, and so on until all groups have been resolved. A group $G_i$ is resolved when $G_i$ has only one class. There are three possible cases when group $G_i$ is resolved.

1) Group $G_i$ has only one asymmetric variable and the phase of this variable is determined.
2) Group $G_i$ has multiple variables, an arbitrary two variables of $G_i$ are symmetric, and the phases of all variables of $G_i$ are determined.
3) Group $G_i$ has one or more independent variables. Because the Boolean difference signature of an independent variable is 0 and the Boolean difference signature of dependent variable must not be 0, a group contains only independent variables when any independent variable occurs.

When searching candidate canonical formations, all the candidate canonical transformations are stored in a tree and Procedure 2 uses a depth-first search to find candidate canonical transformations. Each branch of the candidate canonical transformation tree is a candidate canonical transformation. In Procedure 2, CT_list is the candidate canonical transformation tree, which is initialized to NULL. $T$ is the candidate canonical transformation found in the search process, and $C_T$ is the maximal canonical transformation, which is initialized empty.

Let $G = \{G_1, G_2, \ldots, G_m\}$ be the grouping results that Procedure 1 obtained using the 1st DC signature values. The first step in Procedure 2 is to check whether all groups are resolved. Here we use the condition $D_1$ to denote whether all groups are resolved.

When $D_1$ is true, Procedure 2 finds a candidate canonical transformation $T$. When this occurs, Procedure 2 assigns $T$ to $C_T$ if $C_T$ is empty or $D_T \prec D_T$. Then, Procedure 2 continues to search for other candidate canonical transformations until all the candidate canonical transformations have been searched.
Procedure 2 Canonical transformation search process

Input: \( f, \text{map\\_list}, \text{group}, m \)

Output: \( C_f \)

function \( \text{SEARCH}(f, CT\\_\text{list}, \text{group}, m) \)

if \( D_1 \) then

if \( \text{Empty}(C_f) \) then

\( C_f = T \)

else if \( D^{C_f} < D^T \) then

\( C_f = T \)

end if

else

for all \( G_i, i \in (1, \ldots, m) \) do

if \( D_2 \) then

break

end if

end for

if \( D_3 \) then

Add the variables in \( G_i \) to \( \text{CT\\_list} \)

SEARCH\( (f, CT\\_\text{list}, \text{group}, m) \)

else

for all \( x_i, x_j \in G_i \) do

Split \( G_i \)

create \( \text{node\\_list} \)

end for

for all \( node_j \) in \( \text{node\\_list} \) do

UPDATE\_SEQUENCE\( (\text{node}_j, \text{group}) \)

\( m = m + 1 \)

if \( D_4 \) then

if \( \text{Empty}(C_f) \) then

\( C_f = T \)

else if \( \text{Empty}(C_f) \) then

\( C_f = T \)

end if

else

Add the variables in \( G_i \) to \( \text{CT\\_list} \)

Update\_signature\( (f) \)

\( m = m + 1 \)

SEARCH\( (f, CT\\_\text{list}, \text{group}, m) \)

end if

end for

end if

end function

When \( D_1 \) is false, some groups could not be resolved. At this point, Procedure 2 searches the group \( G_i \) which is an unresolved group that has the minimal sequence number. Condition \( D_2 \) becomes true when an unresolved group is found.

When group \( G_i \) (the unresolved group having the minimal sequence number) is found, it is handled in two ways.

1) Group \( G_i \) can be resolved.

The condition \( D_3 \) denotes whether group \( G_i \) can be resolved. When \( D_3 \) is true, Procedure 2 adds all the variables in \( G_i \) to the candidate canonical transformation tree and calls itself recursively.

In Example 1, the condition \( D_1 \) is false. Procedure 2 finds that group \( G_1 \) is the first unresolved group and that it can be resolved in this call. Procedure 2 adds variable \( \pi_6 \) to the candidate canonical transformation tree, giving a new layer with a node. Then, Procedure 2 calls itself recursively.

2) Group \( G_i \) cannot be resolved.

When condition \( D_3 \) is false, Procedure 2 splits group \( G_i \). Group \( G_i \) has \( k \) classes; thus, there are \( k \) approaches for splitting it. Let group \( G_i = \{c_{i1}, \ldots, c_{ik}\} \), select one class as group \( G_j \); the other classes are in group \( G_{i+1} \). The first splitting approach is \( P_1 = \{G_i, G_{i+1}\} = \{\{c_{i1}\}, \{c_{i2}, \ldots, c_{ik}\}\} \); the second splitting approach is \( P_2 = \{G_i, G_{i+1}\} = \{\{c_{i2}\}, \{c_{i1}, c_{i3}, \ldots, c_{ik}\}\} \), and the \( k \)th splitting approach is \( P_k = \{G_i, G_{i+1}\} = \{\{c_{ik}\}, \{c_{i1}, \ldots, c_{ii(k-1)}\}\} \). Note when the phases of the variables of group \( G_i \) are not determined, there are \( 2k \) splitting approaches, because we need to try the positive and the negative.

Suppose there are \( k \) splitting approaches. These \( k \) splitting approaches are stored in the variable \( \text{node\\_list} \). Then, Procedure 2 selects one splitting approach from the variable \( \text{node\\_list} \) in order. Let Procedure 2 selects the \( q \)th splitting approach \( P_q = \{G_i, G_{i+1}\} = \{\{c_{iq}\}, \{c_{iq-1}, c_{iq+1}, \ldots, c_{ik}\}\} \). Procedure 2 calls \( \text{UPDATE\\_SEQUENCE}() \) to update the sequence number of the groups from \( G_i \) to \( G_m \). The group sequence number of the variables of \( C_{iq} \) remains unchanged and the group sequence number of the variables of \( \{c_{iq-1}, c_{iq+1}, \ldots, c_{ik}\} \) is incremented by one. Finally, the group sequence number of the variables of original group \( G_{i+1}, \ldots, G_m \) is incremented by one, and the value of \( m \) is incremented by one.

Then, Procedure 2 checks the condition \( D_1 \). When condition \( D_1 \) is true, Procedure 2 creates a candidate canonical transformation \( T \) and compares the DC signature vector of \( T \) and \( C_f \). When condition \( D_1 \) is false, the \( q \)th branch is added to the \( |C_{iq}| \) layers as one layer with one node. In other words, \( |C_{iq}| \) expresses the number of variables in \( C_{iq} \). Procedure 2 calls \( \text{UPDATE\\_signature}() \) to update the DC signature value of \( f \) and regroups the variables of \( G_{i+1}, \ldots, G_m \), and calls itself recursively.

Procedure 2 starts with the first splitting approach and then processes the next splitting approach until all splitting approaches have been processed.

In Example 1, after handling group \( G_1 \), group \( G_2 \) is the next unresolved group with the minimal sequence number. Group \( G_2 \) has two symmetric classes; therefore, it cannot be resolved. There are two ways to split group \( G_2 \): \( P_1 = \{G_2, G_3\} = \{\{\pi_1, \pi_2\}, \{x_3, \pi_7\}\} \) and \( P_2 = \{G_2, G_3\} = \{\{x_3, \pi_7\}, \{\pi_1, \pi_2\}\} \). The first branch added to the candidate canonical transformation tree has two layers with the nodes \( \pi_1 \) and \( \pi_2 \). The second branch added to the candidate canonical transformation tree has two layers with the nodes \( x_3 \) and \( x_7 \).

Procedure 2 selects one splitting approach in sequence and updates the DC signature values. The method for updating the DC signature is similar to that used in [11] to update signatures. In Example 1, Procedure 2 uses variable \( \pi_6 \) to compute the 2nd-order DC signature value for the other variables. Thus, there are new DC signature values, and Procedure 2 regroups all the unresolved groups and calls itself recursively.

Our algorithm uses three strategies to reduce the search space.

1) A symmetric class has multiple variables. When we group and sort variables, all the variables of a symmetric class are regarded as a single operational object.
2) Take advantage of the independence of independent variable. When the phase of independent variables cannot be determined, we assign positive phase to them. If we do not make use of independent variable, a independent class \( C_{il} \) is a symmetric class. When the phases of the variables in \( C_{il} \) are not determined, the number of candidate canonical transformations is doubled. When there are other symmetric classes having the same signatures to \( C_{il} \), the number of candidate canonical transformations is grow even more.

3) The combination of the cofactor signature and Boolean difference signature can better distinguish variables.

In Example 1, after updating the DC signature using variable \( x_1 \), variables \( x_4 \) and \( x_5 \) are also resolved. There are two candidate canonical transformations \( T1 = \{ x_6, x_1, x_2, x_3, x_7, x_4, x_5 \} \) and \( T2 = \{ x_6, x_3, x_7, x_1, x_2, x_4, x_5 \} \). The candidate canonical transformation tree of Example 1 is shown in Fig. 1.

![Fig. 1. The candidate canonical transformation tree of Example 1](image)

For a 7-input Boolean function, there are \( 7! \times 2^7 \) NP transformations, while only two candidate canonical transformations exist after using our algorithm in Example 1. Through the DC signature vector comparison, the maximal canonical transformation \( C_f \) is \( \{ x_6, x_1, x_2, x_3, x_7, x_4, x_5 \} \).

To demonstrate that our algorithm is superior to the algorithm proposed in [1], we use both algorithms to search the maximal candidate canonical transformations in Example 2.

**Example 2:** Consider a 6-input Boolean function \( f(X) = x_1 x_2 x_3 x_4 + x_1 x_2 x_3 x_4 + x_1 x_2 x_3 x_6 + x_1 x_2 x_3 x_6 + x_1 x_2 x_3 x_6 + x_1 x_2 x_3 x_6 + x_1 x_2 x_3 x_6 + x_1 x_2 x_3 x_6 + x_1 x_2 x_3 x_6 + x_1 x_2 x_3 x_6 \). We use both the signature vector from [1] and the DC signature vector proposed in this paper to search the candidate canonical transformations.

1) Using the algorithm of [1], the candidate canonical transformation search process is as follows.

The algorithm from [1] computes the \( 1^{st} \) signature, checks the symmetry and groups variables. The Boolean function \( f \) has 32 minterms, i.e., \( |f| = 32 \). The results of the \( 1^{st} \) signature are \( |f_{x_1}| = 13 \) and \( |f_{x_2}| = |f_{x_3}| = |f_{x_4}| = |f_{x_5}| = |f_{x_6}| = 16 \). There are no symmetric classes. The phase of variable \( x_1 \) is negative and the phases of the other variables are not determined. According to the \( 1^{st} \) signatures, the grouping result is \( G = \{ G_1, G_2 \} = \{ \{ x_1 \}, \{ x_2, x_3, x_4, x_5, x_6 \} \} \).

The first group \( G_1 \) can be resolved, and the variable \( x_1 \) is used to compute the 2nd-order signature of the variables in \( G_2 \). The results are \( |f_{x_1, x_2}| = 9 \) and \( |f_{x_1, x_3}| = |f_{x_1, x_4}| = |f_{x_1, x_5}| = |f_{x_1, x_6}| = 10 \). We know that \( |f_{x_1} = |f| - |f_{x_1}| = 19 \); therefore, we can determine the phases of these 5 variables. Variable \( x_2 \) is negative and the others are positive. But \( G_2 \) cannot be resolved: these 5 variables have the same 2nd-order signature. Group \( G_2 \) has 5 asymmetric classes, the algorithm of [1] splits \( G_2 \) and there are five ways to split it.

- \( P_1 = \{ G_2, G_3 \} = \{ \{ x_2 \}, \{ x_3, x_4, x_5, x_6 \} \} \)
- \( P_2 = \{ G_2, G_3 \} = \{ \{ x_2 \}, \{ x_4, x_5, x_6 \} \} \)
- \( P_3 = \{ G_2, G_3 \} = \{ \{ x_4 \}, \{ x_2, x_3, x_5, x_6 \} \} \)
- \( P_4 = \{ G_2, G_3 \} = \{ \{ x_5 \}, \{ x_2, x_3, x_4, x_6 \} \} \)
- \( P_5 = \{ G_2, G_3 \} = \{ \{ x_6 \}, \{ x_2, x_3, x_4, x_5 \} \} \)

The algorithm of [1] traverses each possible split shown above. We select the first split approach as an example. The updated groups are \( G_1 = \{ x_1 \}, G_2 = \{ x_2 \} \) and \( G_3 = \{ x_3, x_4, x_5, x_6 \} \).

Group \( G_2 \) can be resolved; therefore, we use variable \( x_2 \) to compute 2nd-order signature and group variables. The 2nd-order signatures of the unresolved variables are \( |f_{x_2, x_3}| = |f_{x_2, x_4}| = |f_{x_2, x_5}| = 8 \). From these results, we also know that group \( G_3 \) also cannot be resolved. There are four ways to split it as follows.

- \( P_1 = \{ G_3, G_4 \} = \{ \{ x_3 \}, \{ x_4, x_5, x_6 \} \} \)
- \( P_2 = \{ G_3, G_4 \} = \{ \{ x_3 \}, \{ x_5, x_6 \} \} \)
- \( P_3 = \{ G_3, G_4 \} = \{ \{ x_4 \}, \{ x_3, x_5, x_6 \} \} \)
- \( P_4 = \{ G_3, G_4 \} = \{ \{ x_5 \}, \{ x_3, x_4, x_6 \} \} \)

Continuing, the algorithm of [1] computes its signature and splits the group. Because variables \( x_2, x_3, x_4, x_5 \) and \( x_6 \) always have the same 2nd-order signature, the following computations will generate a total of \( 5 \times 4 \times 3 \times 2 \) candidate canonical transformations.

2) Using the algorithm proposed by this paper, the candidate canonical transformation search process is as follows.

Our algorithm computes the \( 1^{st} \) DC signature value of each variable, checks the symmetry, and searches the independent variables and groups variables by the \( 1^{st} \) DC signature value.

Boolean function \( f \) has no symmetry classes or independent classes. The \( 1^{st} \) DC signature values are \( |f_{x_1}|, |f_{x_2}| = (13, 64), (|f_{x_2}|, |f_{x_3}|) = (16, 36), (|f_{x_3}|, |f_{x_4}|) = (16, 52), (|f_{x_4}|, |f_{x_5}|) = (16, 20), (|f_{x_5}|, |f_{x_6}|) = (16, 12) \) and \( (|f_{x_6}|, |f_{x_7}|) = (16, 28) \).

Because \( |f| = 32 \), we can determine that the phase of variable \( x_1 \) is negative and the phases of others are not determined. Our algorithm groups the variables and generates 6 groups: \( G_1 = \{ x_1 \}, G_2 = \{ x_3 \}, G_3 = \{ x_2 \}, G_4 = \{ x_6 \}, G_5 = \{ x_4 \} \) and \( G_6 = \{ x_5 \} \).

Group \( G_1 \) can be resolved. The other groups are not resolved because the phases of the variables in these groups have not been determined. Our algorithm uses the variable \( x_1 \) to compute the 2nd-order DC signature values, which are: \( (|f_{x_1, x_2}|, |f_{x_1, x_3}|) = (9, 18), (|f_{x_1, x_4}|, |f_{x_1, x_5}|) = (10, 26), (|f_{x_1, x_6}|, |f_{x_1, x_7}|) = (10, 10), (|f_{x_1, x_8}|, |f_{x_1, x_9}|) = (10, 6) \), and \( (|f_{x_1, x_7}|, |f_{x_1, x_9}|) = (10, 14) \).

From the above 2nd-order DC signature values, the phases of all variables are determined. Therefore, groups \( G_2, G_3, \)
$G_4$, $G_5$ and $G_6$ are all resolved. One candidate canonical transformation is generated: \{\mathcal{F}_1, x_3, \mathcal{F}_2, x_6, x_4, x_5\}.

Because $|f| = |f|$, the algorithm of \[1\] and ours must both perform the same search for $f$. However, the Boolean function $f$ results in 240 candidate canonical transformations when computed by the algorithm of \[1\] but only 2 when computed by the algorithm proposed in this paper.

Using our algorithm, the NPN Boolean matching algorithm is denoted as follows: Given two Boolean function $f(X)$ and $g(X)$, Boolean function $f$ is NPN-equivalent to $g$ if the canonical form $\mathcal{F}$ of $f$ is equal to the canonical form $G$ of $g$.

V. EXPERIMENTAL RESULTS

We reimplemented the algorithm of \[1\] to compare it with the algorithm of this paper and obtain supporting experimental evidence. The two algorithms were tested on a set of randomly generated circuits and a set of MCNC benchmark circuits. The two circuit sets include a number of NPN-equivalent circuits. We tested the NPN matching runtime and the number of candidate canonical transformations generated when searching for the canonical transformation of a Boolean function. The runtime is measured in seconds. The following experimental results were obtained on a computer equipped with 3.3 GHz CPU and 4GB RAM.

Tables I and II show the experimental results on the random circuit and MCNC benchmark circuit sets, respectively. In these two tables, the first column shows the number of inputs (#I), the second and third columns show the average matching runtime (#A.T) and the average number of candidate canonical transformations (#A.C.T) of the algorithm in \[1\], respectively, and the forth and fifth columns show the average matching runtime and the average number of candidate canonical transformations of our algorithm, respectively.

![Fig. 2. The comparison results on the equivalent random circuits](image)

The experimental results listed in Table I show that our algorithm improves the speed of the Boolean matching process and reduces the search space effectively. In the best case, our algorithm’s runtime is 70% faster and its search space is 82% smaller than that of the algorithm from \[1\]. On average, our algorithm improves the runtime by 37% and reduces the search space by 67%. Fig. 2 shows the results on the equivalent random circuits.

| #I | #A.T of \[1\] (Second) | #A.C.N of \[1\] | #A.T of ours (Second) | #A.C.N of ours |
|----|------------------------|----------------|----------------------|----------------|
| 7  | 0.0007                 | 1.32           | 0.0005               | 7.3            |
| 8  | 0.0009                 | 22.1           | 0.0004               | 5.3            |
| 9  | 0.0010                 | 16.8           | 0.0007               | 6.1            |
| 10 | 0.0013                 | 19.6           | 0.0004               | 3.6            |
| 11 | 0.0032                 | 19.1           | 0.0020               | 4.3            |
| 12 | 0.0068                 | 17.2           | 0.0048               | 4.5            |
| 13 | 0.0099                 | 14.8           | 0.0062               | 4.2            |
| 14 | 0.0097                 | 16.5           | 0.0044               | 4.2            |
| 15 | 0.0249                 | 13.9           | 0.0135               | 5.0            |
| 16 | 0.0326                 | 23.4           | 0.0208               | 5.9            |
| 17 | 0.0423                 | 13.3           | 0.0261               | 5.0            |
| 18 | 0.1565                 | 13.8           | 0.1070               | 5.8            |
| 19 | 0.1195                 | 12.5           | 0.0974               | 5.6            |
| 20 | 0.0859                 | 13.3           | 0.0708               | 4.9            |
| 21 | 0.1349                 | 18.3           | 0.1186               | 6.8            |

Form Table II, we can see that the average runtime improves by 83% and the search space is reduced by 96% compared with \[1\] when the number of input variables is 7. However, neither the runtime nor the search space are drastically improved in the other input cases. In majority of the circuits of the MCNC benchmark, using only cofactor signature serves to quickly identify the variables. Therefore, the average runtime of the algorithm of \[1\] when tested on 22 inputs is slightly faster than ours. Despite this, our algorithm improves the average runtime by 6% and reduces the average search space by 19% compared with that of \[1\] when tested on the MCNC benchmark circuits.

1A.T: Average runtime, A.C.N: Average number of candidate canonical transformations
VI. CONCLUSIONS

In this paper, we propose a canonical-based Boolean matching algorithm. The proposed DC signature vector is more effective in computing canonical form than the algorithm of [1]. We test the runtime and the canonical transformation search space of Boolean matching algorithms with 7-22 inputs. The experimental results show that the average runtime of our algorithm is 37% faster and its search space is 67% smaller than those of the algorithm proposed in [1]. The algorithm proposed in this paper is highly effective at reducing the search space and enhancing the Boolean matching speed. In future work, we plan to extend this Boolean matching approach to Boolean matching with don’t care set and multiple-output Boolean matching.

ACKNOWLEDGMENT

We would like to thank the National Natural Science Foundation of China (Grant No. 61572109, No. 11371003) and the Special Fund for Bagui Scholars of Guangxi (Grant No. 113000200230010) for their support for technology.

REFERENCES

[1] A. Abdollahi and M. Pedram, “Symmetry detection and boolean matching utilizing a signature-based canonical form of boolean functions,” IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, vol. 27, no. 6, pp. 1128–1137, 2008.

[2] Y. Zhang, G. Yang, W. N. Hung, and J. Zhang, “Computing affine equivalence classes of boolean functions by group isomorphism,” IEEE Transactions on Computers, vol. 65, no. 12, pp. 3606–3616, 2016.

[3] B. Kapoor, “Improved technology mapping using a new approach to boolean matching,” in European Design and Test Conference, Proceedings, Paris, France, Mar. 1995, pp. 86–90.

[4] F. Mailhot and G. De Micheli, “Technology mapping using boolean matching and don’t care sets,” in Design Automation Conference, 1990., EDAC. Proceedings of the European, Glasgow, UK, 1990, pp. 212–216.

[5] M. Damiani and A. Y. Selchenko, “Boolean technology mapping based on logic decomposition,” in Integrated Circuits and Systems Design, 2003. SBCIC 2003. Proceedings. 16th Symposium on, Sao Paulo, Brazil, Sep. 2003, pp. 35–40.

[6] D. Debnath and T. Sasao, “Efficient computation of canonical form for boolean matching in large libraries,” in Design Automation Conference, 2004. Proceedings of the ASP-DAC 2004. Asia and South Pacific, Yokohama, Japan, Jan. 2004, pp. 591–596.

[7] G. Agosta, F. Bruschi, G. Pelosi, and D. Sciuto, “A transform-parametric approach to boolean matching,” IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, vol. 28, no. 6, pp. 805–817, 2009.

[8] D. Chai and A. Kuehmann, “Building a better boolean matcher and symmetry detector,” in Design, Automation and Test in Europe, 2006. DATE ’06. Proceedings, Munich, Germany, 2006, pp. 1–6.

[9] J. Ciric and C. Sechen, “Efficient canonical form for boolean matching of complex functions in large libraries,” IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, vol. 22, no. 5, pp. 535–544, 2003.

[10] G. Agosta, F. Bruschi, G. Pelosi, and D. Sciuto, “A unified approach to canonical form-based boolean matching,” in Design Automation Conference, 2007. DAC ’07. 44th ACM/IEEE, San Diego, CA, 2007, pp. 841–846.

[11] Z. Huang, L. Wang, and Y. Nasikovskiy, “Fast boolean matching based on npn classification,” in Field-Programmable Technology (FPT), 2013 International Conference on, Kyoto, Japan, Dec. 2013, pp. 310–313.

[12] A. Petkovska, M. Søeken, G. D. Micheli, P. Ienne, and A. Mishchenko, “Fast hierarchical npn classification,” in International Conference on Field Programmable Logic and Applications, Lausanne, Switzerland, Dec. 2016, pp. 1–4.

[13] K. C. Chen and J. C. Y. Yang, “Boolean matching algorithms,” in International Symposium on VLSI Technology, Systems, and Applications, Proceedings, Taipei, Taiwan, May 1993, pp. 44–48.