

# Create reduced binary decision tree

Leyang Shen 01354299

## Table of Contents

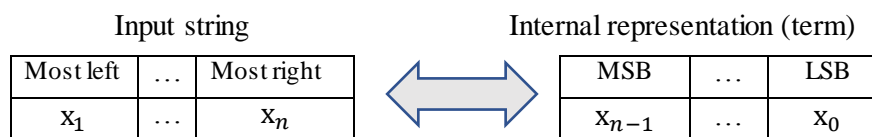
Introduction .....	2
Description of Algorithm .....	2
Buildcompactbdt.....	3
Evalcompactbdt .....	6
Internal representation of minterm in C++ .....	7
Impletion of Buildcompactbdt in C++ .....	8
Generate minterm .....	8
Find prime implicant .....	9
generate the binary decision tree .....	11
Impletion of Evalcompactbdt in C++ .....	13
Experimental Evaluation .....	14
Is it worth to implement functions to find essential prime implicants? .....	14
Memory and time evalution and optimisation.....	15
Testing on case where can't be merged.....	16
Testing on large fvalues.....	16
Manual test on fvalue with length greater than 64.....	17
Effect of speed on how Internal representation of minterm is implemented .....	18
Appendix .....	19
Binary trick .....	19
Tricks for bitcount .....	20
List of testing case for large fvalues: .....	21
Set 1.....	21
Set 2.....	22
Set 3.....	23
Set 4.....	24
Set 5.....	25

## Introduction

The specification for this assignment is to produce a function (buildcompactbdt) that generate a reduced binary tree with minimum nodes, and a function (evalcompactbdt) will use this tree to evaluate input string to produce true "1", or false "0".

## Description of Algorithm

The input for buildcompactbdt is a list of string, where each entry in string corresponds to a row in truth table which evaluate to 1. This input can be considered as an POS (Product of Sum) expression, where each string corresponds to a minterm, with the left most side as  $x_1$ , which corresponds to LSB (Least significant bit) of the minterm, as illustrated by the diagram below.



To simplify the binary decision tree, instead of approach where seems most forward, to generate a tree then rotate and remove repetitive node.

I decide to per-process the string in term of simplify SOP expression, remove "don't cares" and get (essential) prime implicants, which is the simplest form possible, if done as binary decision graph, there will be using least node possible.

But the requirement of this assignment is binary decision tree, so I need to turn the graph into tree by creating duplicate node, with the least amount of duplication possible.

## Buildcompactbdt

This function takes one argument, which is input strings (fvalues) and the return with root node of the decision tree. Below are the descriptions of the algorithm

1. Convert vector of string to vector of minterm
2. Using Quine–McCluskey<sup>1,2</sup> algorithm to find prime implicants.  
By using this method, up to this step, the output is optimal, there are no possible simplification to further reduce the term.
3. (optional)<sup>3</sup> Using chart(heuristic) or Petrick's method<sup>4</sup> to find essential prime implicants  
By using Petrick's method will ensure the number of term is optimal, but a much faster approached is to use heuristic method to approximate the essential prime implicants. However, as shown in later experimental section, this step does not produce a much different for the node of binary decision tree.
4. Using heuristic algorithm to generate the binary decision tree, with aim to make duplicated node appear as deeper node as possible.  
By using 3 conditions in order of propriety to choice the node, firstly choice node with most don't care, secondly choice node with most unbalanced 1 and 0, finally choice by numerical order from node 1 to node n.

---

<sup>1</sup>Donald Krambeck (2016) *Everything About the Quine-McCluskey Method*  
<https://www.allaboutcircuits.com/technical-articles/everything-about-the-quine-mccluskey-method/>  
[ 29<sup>th</sup>/04/2018]

<sup>2</sup> corporate author (2018) *Quine–McCluskey algorithm*  
[https://en.wikipedia.org/wiki/Quine-McCluskey\\_algorithm](https://en.wikipedia.org/wiki/Quine-McCluskey_algorithm) [ 29<sup>th</sup>/04/2018]

<sup>3</sup> This step was considered at the start, but I had chosen not to implement due to the level of optimisation and time constraint. Please refer to 1<sup>st</sup> part of Experimental Evaluation for the experiment.

<sup>4</sup>Donald Krambeck (2016) *Prime Implicant Simplification Using Petrick's Method*  
<https://www.allaboutcircuits.com/technical-articles/prime-implicant-simplification-using-petricks-method/> [ 29<sup>th</sup>/04/2018]

Below is an example of work through, for input fvalues of 0000(0), 1000(1), 0100(2), 1010(5), 0110(6), 1110(7), 0001(8), 1001(9), 0101(10) and 0111(14).

The website “allaboutcircuits”<sup>5</sup> showed how to simplify the fvalues above with Quine–McCluskey algorithm to obtain essential prime implicants so I don’t repeat it here. At start of step 4, the table below shows the representation of prime implicants.

essential prime implicants	Internal representation in struct implicant		
$x_1x_2x_3x_4$ <sup>6</sup>	$x_4x_3x_2x_1$	mask	Minterm
1-10	01- <sup>7</sup> 1	0010	0101
-00-	-00-	1001	0000
01--	--10	1100	0010

In step4, 3 condition are used for this heuristics algorithm, listed below in the order of how decision is taken:

- choice the  $x_n$  with least don’t care, to make duplicated node appear as deeper node as possible,
- If there are multiple  $x_n$  with same number of don’t care, additional screening is used to choose the node with most unbalanced 1 and 0. Such that for example with internal representation ( $x_4x_3x_2x_1$ ) of 0001 and 0010, node  $x_3$  then  $x_4$  will be chosen as it is most unbalanced. Reference the full example from “Testing on case where can’t be merged” in Evaluation section.
- If there are multiple  $x_n$  have same ranking from condition i and ii, then numerical order is choice.

Continue with the example, reference the table above,  $x_3, x_2, x_1$  have the least don’t care (“-” represented as 1 in mask), and same unbalance of 1 and 0 (all have 1 ones and 2 zeros). Which means numerical order will be used, which choice  $x_1$  as root of the tree.

When  $x_1$  is chosen as root node, the table above can be split into 2.

When $x_1$ is 0, left child			When $x_1$ is 1, right child		
$x_4x_3x_2x_1$	mask	Minterm	$x_4x_3x_2x_1$	mask	Minterm
-00(0) <sup>8</sup>	1000	0000	01-(1)	0010	0101
--1(0)	1100	0010	-00(1)	1000	0000

<sup>5</sup>Donald Krambeck (2016) *Everything About the Quine-McCluskey Method*  
<https://www.allaboutcircuits.com/technical-articles/everything-about-the-quine-mccluskey-method/>  
 [ 29<sup>th</sup>/04/2018]

<sup>6</sup> $x_1x_2x_3x_4$  is the order of the input string fvaules, but internally is stored as  $x_4x_3x_2x_1$ . Later on will always using internal representation unless when evaluate against fvalue(s).

<sup>7</sup>“-“ here means don’t care, 01-1 is same as  $\overline{x_4}x_3x_1$

<sup>8</sup> Bracket means this bit ( $x_1$  in this case) had been envaulted, the mask and minterm at this bit will be normalised to 0

Further consider node (bit) not been evaluated, using same heuristic strategy,  $x_2$  will be chosen, as  $x_4, x_3, x_2$  have don't care 2,1,0 respectively, for the internal node for left child of root node  $x_1$ . By the same method,  $x_3$  will be chosen as internal node for the right node. Which produces a table as shown below.

When $x_1$ is 0, left child		When $x_1$ is 1, right child	
When $x_2$ is 0	When $x_2$ is 1	When $x_3$ is 0	When $x_3$ is 1
$x_4 x_3 x_2 x_1$	$x_4 x_3 x_2 x_1$	$x_4 x_3 x_2 x_1$	$x_4 x_3 x_2 x_1$
-0 (0) (0)	--(1)(0)	-(0)0(1)	0(1)-(1)

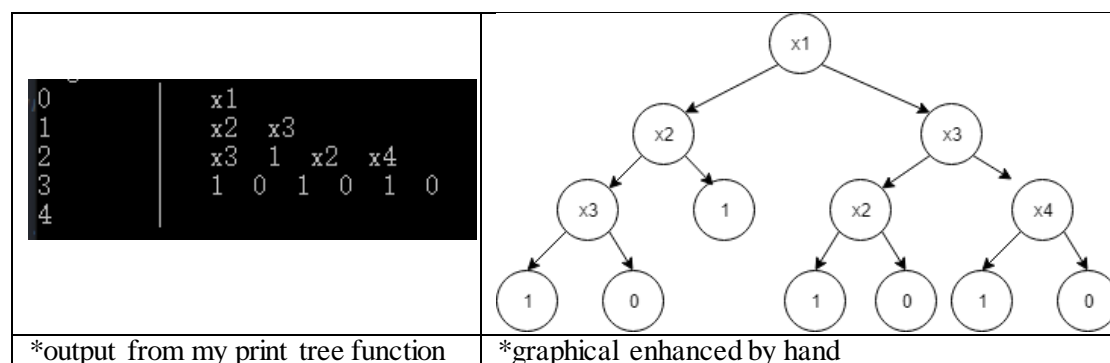
By repeat steps above, it will reach 2 kind of base case. First base case is when there is no implicant in the table, second base case is when there is *one* implicant with non-evaluated bit are all don't care (""). For example, when  $x_1$  is 1,  $x_3$  is 0.

When $x_1$ is 1 and $x_2$ is 0					
When $x_2$ is 0			When $x_2$ is 1		
$x_4 x_3 x_2 x_1$	mask	Minterm	$x_4 x_3 x_2 x_1$	mask	Minterm
-(0)(0)(1)	1000	0000			

As shown on table above, on left side when  $x_2$  is 0, the only term  $x_4$  is not evaluated which is don't care. This node has label 1 and no child node, represented in Boolean is  $\bar{x}_3 \bar{x}_2 x_1$  evaluate to 1.

On the other hand, on left side when  $x_2$  is 1, there are no term as the table content. This node has label 0 and no child node, represented in Boolean is  $x_3 \bar{x}_2 x_1$  evaluate to 0.

The binary decision tree produced by this method is shown below



By using this method, only 13 nodes is required, compare to a full binary decision tree as did in assignment 1, requires 31 nodes.

## Evalcompactbdt

This function takes two arguments, the first is the root of the tree, second is a string for input to evaluate.

In the binary decision tree, except for the leaf which will either be “1” or “0”, the rest of node is in the form of “xnum”, where *num* is a index for a char at position (num-1). When evaluate the tree, if the string[num-1] is “0”, it goes to left branch, otherwise, goes to right branch.

For example, continues example above, when evaluate string is “0100”( $x_1x_2x_3x_4$ ), correspond to internal minterm 0010 ( $x_4x_3x_2x_1$ ). Check it is not leaf node(“1” or “0”), remove first char of string “x1”, get 1, evaluate char of input string at index 0, which is 0, evaluate left node.

Check it is not leaf node, remove first char of string “x2”, get 2, evaluate char of input string at index 1, which is 1, evaluate right node.

Check it is leaf node, output the content, returns “1”.

## Internal representation of minterm in C++

For flexibility of the code, all the algorithm and function implemented below uses same abstract data structure called “term”, which is treated as a vector consist of multiple word (integer) of constant length. All mask, minterm, implicant are implemented by using term.

The reason for use word (64 bit integer) is because of efficiency. Another way to implement such data structure is using `std::vector<bool>`, which although occupies same space (stores multiple bool in same word), but it requires mask when read and write and it does not support read and write multiple bits at a time, such as XOR of two 64-bit word.

Bitwise operator (`~`, `|`, `^`, `&`) had been overloaded, to allow readable and more maintainable code.

```
term operator~(const term& v1);  
term operator&(const term& v1, const term& v2);  
term operator|(const term& v1, const term& v2);  
term operator^(const term& v1, const term& v2);
```

Below are how current implementation of term is defined. Each word is a 64 bit unsigned integer (unsigned long long), to form a vector of any length required by the fvalues.

```
#define wlength 64  
#define wdMax 0xFFFFFFFFFFFFFFFF  
typedef uint64_t wd;  
typedef std::vector<wd> term;
```

## Impletion of Buildcompactbdt in C++

### Generate minterm

This step is performed by function call `genMinterm(fvalues,minterms)` ,where `fvalues` is a vector of string of same length and `minterms` is vector of minterm, both pass by reference.

The algorithm can be represented by C++ like pseudo code below:

```
int numwd=fvalues[0].size() / wlength;
int lastlen=fvalues[0].size() % wlength;
For( str : fvalues){
    int p=fvalues[0].size()-1;
    term minterm(numwd+1,0); //init minterm to (numwd+1) word consist of 0

    //Most significant word of the term
    wd word=0
    Repeat lastlen times{
        if (str[p]==1){ word = word *2+1 } else { word = word *2};
        p--;
    }
    minterm[numwd]=word

    //All other less significant word of the term
    While(p>0){
        word=0;
        Repeat wlength times{
            if (str[p]==1){ word = word *2+1 } else { word = word *2};
            p--;
        }
        minterm[(p+1) / wlength]=word
    }
}
Minterms.push_back(minterm)
```



## Find prime implicant<sup>9</sup>

Quine–McCluskey algorithm is used in this step, also by taking inspiration from (Jadhav Vitthal, 2012)<sup>10</sup>, a structure is made to recursively compute this. The structure is called intermittent.

At the start the structure intermittent is populated by minterms with same mask, normalised by doing a bitwise AND with the mask and then separated into smaller group according to number of 1's to contain, the mask has same length as minterm. For example,  $X_1\overline{X_4}$  and  $X_1X_4$  will be populated to same instance of the structure, as they both have mask of 0110, but they will be speared to 2 group with minterm represented by 0001(one 1's) and 1001(two 1's) respectively.

The structure intermittent had a function called compareAll, which will compare neighbour groups of 1s iteratively by Quine–McCluskey algorithm, and create new instances of the structure by using “new “ on HEAP and append it to a map with the common mask as key. For the part that did not use to create new instance, it is push backed to a vector of another structure implicant, as it is prime implicant.

Structure implicant only consist of two data field with no functions, first is the mask and the second is minterm, for example  $X_3\overline{X_4}$  will have mask 0011 represent for “don't care” of term  $X_1X_2$  and minterm 1000, where least significant 2 bit is forced to 0 due to the mask.

Pseudo code is not provided for this section, as it involved with multiple struct with their public data and function. A flowchart is used to given a overview of how these struct linked together to execute efficient Quine–McCluskey algorithm.

---

<sup>9</sup> ALL pseudo code from this sub-section onward, unless otherwise stated, treat term as a single integer, a simplification of how the real code is implemented.

<sup>10</sup> Corporate author(2012) *Modified Quine-McCluskey Method*

<https://pdfs.semanticscholar.org/650a/ae0c3e59bdd12e7b22158162166e4437949e.pdf> [29<sup>th</sup>/04/2018]

The chart below shows how these structures linked together to find prime implicant.

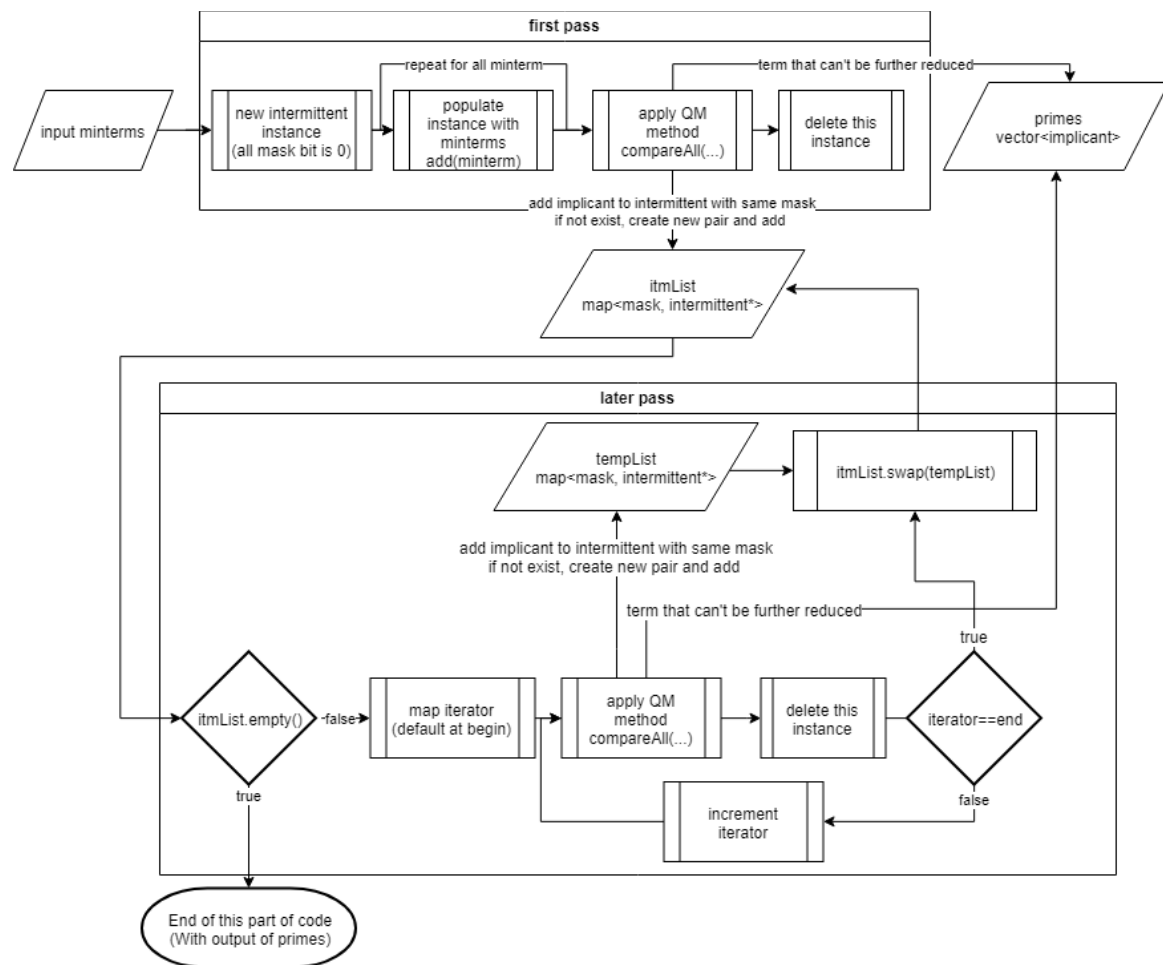


Figure 1

As shown from the chart, in the first pass, an intermittent is newed and constructed, then all the minterm produced by previous step is added to the struct, compareAll(...) function is called, produce an output of map itmList and primes for primes implicant. The current struct is then deleted to save HEAP space and avoid memory leak.

In all the later pass, itmList is firstly checked if it is empty. If true, means no more term need be further simplified and exist to next part, if false, iterate through the map for all the intermittent struct. For each intermittent struct, call compareAll(...) function but put output map to itmList, and delete the current instance and iterate to next struct in the itmList. After iteration finish, swap between tempList and itmList, ready for next pass. The pointer in itmList which swapped to tempList is no longer needed and delete automatically as went out of scope.

The reason for the swap is feasible is due to implicant can only merge when mask is the same and its only different by one bit, but mask in tempList and itmList will never be, as tempList is like children of itmList, have 1 more bit of mask.

## generate the binary decision tree

By using the algorithm introduced in Description of Buildcompactbdt from previous section. A recursive function can be used to implement this algorithm, with base cases of all don't cares and empty content.

Function `recTreeConstructor(node, primes, nodeRemains)` is used to implement the function. Where `node` is the root of the sub tree that is currently constructing, `primes` are the (essential) prime implicant produced by previous part, `nodeRemains` are the bit indicator of either bit had been used.

For example, for `fvalue` consist of 4 chars, if bit  $x_3, x_2$  had been used, but  $x_4, x_1$  haven't, the `nodeRemains` will be 1001 (by internal representation  $x_4x_3x_2x_1$ ).

Structure `bitCount` is used, it has function of `construct(...)`, which takes vector of prime implicant (`primes`), count number of set for each bit of the don't care (`primes.mask`) and count number of ones for each bit of the minterm (`primes.minterm`).

It has a function called `getMask (digit)`, which returns the sum of set bit for that digit. And function called `getTermLeast(digit)`, which returns the minimum between sum of ones and sum of zeros for that digit, such that the smaller the number returns, the more unbalanced ones and zeros it has.

Below is C++ like pseudo code used to describe this function.

```
If(primes.size()==0){
    //2 base case
    Node.lable="0" //first base case
    return;
}else{
    For(prime :primes){
        maxMaskCount=popcount(nodeRemains)
        If( popcount(prime.mask)== maxMaskCount){
            Node.lable="1" //second base case
            return;
        }
    }
}
```

```

// heuristic method to decide which root to choose
bitCount cnt;
cnt.constrct(primes, ...);
unsigned int mincount= (int)-1 //use type casting to set mincount to highest possible

for(digit in nodeRemains){
    if(cnt.getMask(digit)< mincount){
        // condition i
        mincount= cnt.getMask(digit)
        optimalDigit = digit
    }elseif(cnt.getMask(digit)==mincount && cnt.getTermLeast(digit)<mostUnbalanced){
        // condition ii
        mostUnbalanced= cnt.getTermLeast(digit)
        optimalDigit = digit
    }
}
Node.labe="x"+ optimalNode

// prepare for the recursive call
For(prime :primes){
    If(prime.minterm[optimalDigit]==0){
        // as described in algorithm description, have 2 possible cases
        If(prime.mask[digit]!=0){
            tempPrime=prime
            tempPrime.mask= prime.mask & (~optimalMask11)
            leftPrimes.push_back(prime)
            rightPrimes.push_back(prime)

        }else{
            rightPrimes.push_back(prime)
        }
    }else{
        leftPrimes.push_back(primes[i]);
    }
}

// start recursive call
recTreeConstructor(node->left, leftPrimes, nodeRemains& (~optimalMask))
recTreeConstructor(node->right, rightPrimes, nodeRemains& (~optimalMask))
}

```

---

<sup>11</sup> optimalMask is a mask where only optimalDigit bit is set to 1, rest is 0. ~ means bitwise NOT, & means bitwise AND, it is used to normalise the optimalDigit bit on mask to 0

## Impletion of Evalcompactbdt in C++

From algorithm introduced in Description of Evalcompactbdt, below is the simple implementation.

```
While(node is not 0 or 1){  
    If(str[node.val]!=0){  
        node=node.right  
    }else{  
        node=node.left  
    }  
}  
Return node.val
```

Where node.val is the content in the node, str[i] is the i<sup>th</sup> char in string str indexing from 0.

## Experimental Evaluation

All the test is performed on Win10 Ubuntu, with average  $13.5\% \pm 0.5\%$  CPU (i7-7700K) with DDR3 RAM. The time shown is the time taken to execute `buildcompactbdt()` function.

In all of test below, unless stated otherwise, is done by taking integer set<sup>12</sup> as input, treated as minterm, generate fvalues and run `buildcompactbdt()`, and finally testing correctness by iterate all the possible combination and compare if “1” only get returned when it is a member of fvalues.

### Is it worth to implement functions to find essential prime implicants?

I used my code without step 3 mentioned in Description of Buildcompactbdt algorithm to shown number of node produced by tree with only prime implicants as control group. Used online resources<sup>13</sup> to generate the essential prime implicant, inject to step 4, to see the performance with essential prime implicants, as experimental group.

The first set I tested is with minterm used in previous work through 0,1,2,5,6,7,8,9,10,14 which forms a fvalues with size 10 and each fvalue string is made by 4 chars.

The result from control group, without reduce term to find essential prime implicants is 17 nodes. There is only one form of essential prime implicant, which is  $\bar{x}_2 \bar{x}_3 + \bar{x}_1 x_2 + \bar{x}_4 x_3 x_1$ , by inject to step 4, creating a tree with 13 node. Both tree are drawn in the table below:

Use prime implicants						Use essential prime implicants					
0			x1			0			x1		
1			x2	x2		1			x2	x3	
2			x3	1	x3	x3			x3	1	x2
3		1	0	1	x4	0	x4		1	0	1
4		1	0	1	0						
17 nodes						13 nodes					

The second set I tested is with minterm<sup>14</sup> 40,51,37,38,27,50,14,5,42,53,61,20,21,16,15,22, 63,30,25,47,6,19,17,23,13,56,26,9,52,45,24,12,1,7,62,41,28,32,58,8,0,29,39,35,43,60,57,59,1 1,31,3,44,54,46,36,4,48,2,34,18, which forms a fvalues with size 59 and each fvalue string is made by 6 chars.

The result from control group, without reduce term to find essential prime implicants is 29 nodes, the result from experimental group with all 12 different form of simplification measured, it had max of 35 nodes, minimum of 31 nodes, median of 35 nodes.

Interestingly for this case, by find essential prime implicants increases the number of node required.

The last set I tested is with minterm 0,2,4,8,9,10,11,12,16,21,22,23,24,25,26,27,28,29,30, 31,32,39,40,44,46,49,51,52,61,63 which forms a fvalues with size 29 and each fvalue string is made by 6 chars.

The result from control group, without reduce term to find essential prime implicants is 61 nodes, the result from experimental group with all 2 different form of simplification is 57 nodes and 63 nodes.

As shown, from 3 set of data above, step 3 does not always lead to reduced to node and due to time constraint, it is not implemented.

<sup>12</sup> All testing data are generated randomly from random.org by using integer set generator.

<sup>13</sup> Frédéric Carpon (2012) *quine-mccluskey-frederic-carpon-implementation* <http://quine-mccluskey-frederic-carpon-implementation.e-geii.eu/> [29<sup>th</sup>/04/2018]

<sup>14</sup> Randomly generated from random.org <https://www.random.org/integer-sets/>

## Memory and time evaluation and optimisation

Before I optimise the step 2 of Buildcompactbdt (Quine–McCluskey), the runtime of step 3 is heavily depends on the size of the input fvaules, it will take significant part of HEAP after fvalues size reaches 64 and above, also the runtime is often greater than 1 minutes. Hence, I decide to debug the find out how can I improve it.

Here is the performance before optimise, with input of minterms from 0 to 31, generate a fvaules of size 32 and each fvalue string have 5 chars.

Before the optimisation, the Buildcompactbdt function will take 15.34 seconds to run, uses 34.4MB on HEAP maximum, but there is no memory leak, as all struct created by new are deleted before pointer go out of scope.

After the optimisation, with the same input, Buildcompactbdt functions only takes 104.1 ms to construct the tree, with 101kB on HEAP at maximum, with 99.3% reduction on time and 99.7% reduction on HEAP.

See second row of table below for original output, first 5 lines are my debug code in main(), after that are information provided by valgrind.

I start this evaluation by identity the part where takes most of time and memory happened at “later pass” described in flowchart Figure 1. By using “cout” to show intermediate process in this section, I found the problem is due to duplicate entry to each of intermittent struct tempList.

To solve this problem, I changed the code in add function of struct intermittent to perform additional check of duplicate, shown on third row of table below.

Before optised	After optised
<pre> the tree has 1 nodes compare to tree in assl require 63 node reduced by 98%  the buildcompactbdt function takes 15360.3ms  ==414== ==414== HEAP SUMMARY: ==414==   in use at exit: 72,704 bytes in 1 blocks ==414==   total heap usage: 933 allocs, 932 frees, 34,856,724 bytes allocated ==414== ==414== LEAK SUMMARY: ==414==   definitely lost: 0 bytes in 0 blocks ==414==   indirectly lost: 0 bytes in 0 blocks ==414==   possibly lost: 0 bytes in 0 blocks ==414==   still reachable: 72,704 bytes in 1 blocks ==414==   suppressed: 0 bytes in 0 blocks ==414== Rerun with --leak-check=full to see details of leaked memory ==414== ==414== For counts of detected and suppressed errors, rerun with: -v ==414== ERROR SUMMARY: 0 errors from 0 contexts (suppressed: 0 from 0) </pre>	<pre> the tree has 1 nodes compare to tree in assl require 63 node reduced by 98%  the buildcompactbdt function takes 103.015ms  ==420== ==420== HEAP SUMMARY: ==420==   in use at exit: 72,704 bytes in 1 blocks ==420==   total heap usage: 597 allocs, 596 frees, 100,756 bytes allocated ==420== ==420== LEAK SUMMARY: ==420==   definitely lost: 0 bytes in 0 blocks ==420==   indirectly lost: 0 bytes in 0 blocks ==420==   possibly lost: 0 bytes in 0 blocks ==420==   still reachable: 72,704 bytes in 1 blocks ==420==   suppressed: 0 bytes in 0 blocks ==420== Rerun with --leak-check=full to see details of leaked memory ==420== ==420== For counts of detected and suppressed errors, rerun with: -v ==420== ERROR SUMMARY: 0 errors from 0 contexts (suppressed: 0 from 0) ==420== MEMSCKTOP: 61440KB: /tmp/algos/3 </pre>
<pre> minterm=minterm&amp;baseNot; int pcount=popcount(minterm);  minterms[pcount].push_back(minterm); used[pcount].push_back(false); </pre>	<pre> minterm=minterm&amp;baseNot; int pcount=popcount(minterm);  if(notRepeated<sup>15</sup>(minterms[pcount],minterm)){     used[pcount].push_back(false); } </pre>

<sup>15</sup> notRepeated(A,B) function will return Boolean whether B is already in A, and push back B to A if not exist

## Testing on case where can't be merged

One special case that I have experimented is when there are not term to merge to produce don't care, such that the only factor is going to make a difference is the order of the node chosen.

Below is the example of how node will be chosen for 00001, 00010 and 10000, 01000.

Term $x_1x_2x_3x_4x_5$	with condition i and ii <sup>16</sup>	All 3 condition i,ii,iii
00001 00010	<pre> 0      x1 1      x2  0 2      x3  0 3      x4  0 4      x5  x5 5      0  1  1  0           </pre> <p>13 nodes in total</p>	<pre> 0      x1 1      x2  0 2      x3  0 3      x4  0 4      x5  x5 5      0  1  1  0           </pre> <p>13 nodes in total</p>
10000 01000	<pre> 0      x1 1      x2  x2 2      0  x3  x3  0 3      x4  0  x4  0 4      x5  0  x5  0 5      1  0  1  0           </pre> <p>19 nodes in total</p>	<pre> 0      x3 1      x4  0 2      x5  0 3      x1  0 4      x2  x2 5      0  1  1  0           </pre> <p>13 nodes in total</p>

As shown from the table above, with condition ii, choice the node with most unbalanced 1 and 0 applies, it create minimum node consistently. Whereas if after condition i just choice numerically by condition iii, the number of nodes is heavily depend on the order of the input, leading to 46% increasement on number of nodes in this case from 13 nodes to 19 nodes.

## Testing on large fvalues

Below is the result for 10,000<sup>17</sup> terms of fvalue. From in the range of 0000000000000000 (minterm 0) to 11111111111111 (minterm 2<sup>15</sup>-1).

Minterm Set reference number <sup>18</sup>	Number of nodes with condition i ii	Number of nodes with condition i ii iii	Number of nodes without simplification
Set 1	16861	16857	65535
Set 2	16659	16639	65535
Set 3	16815	16875	65535
Set 4	16991	16915	65535
Set 5	16817	16787	65535
<b>Average Nodes</b>	<b>16828.6</b>	<b>16814.6</b>	<b>65535</b>

As shown, there is a average of 74.3% of reduction by using my method compare to the node without simplification. By using condition ii, before choosing node numerically will reduce number of node further by 14 on average, although there are times where by using only condition i ii performs, but it is just luck.

<sup>16</sup> As condition described in step 4 of section Buildcompactbdt

<sup>17</sup> Max number of integers allowed to generate by random.org

<sup>18</sup> The Minterm Set used in the test bench can be find from end of appendix



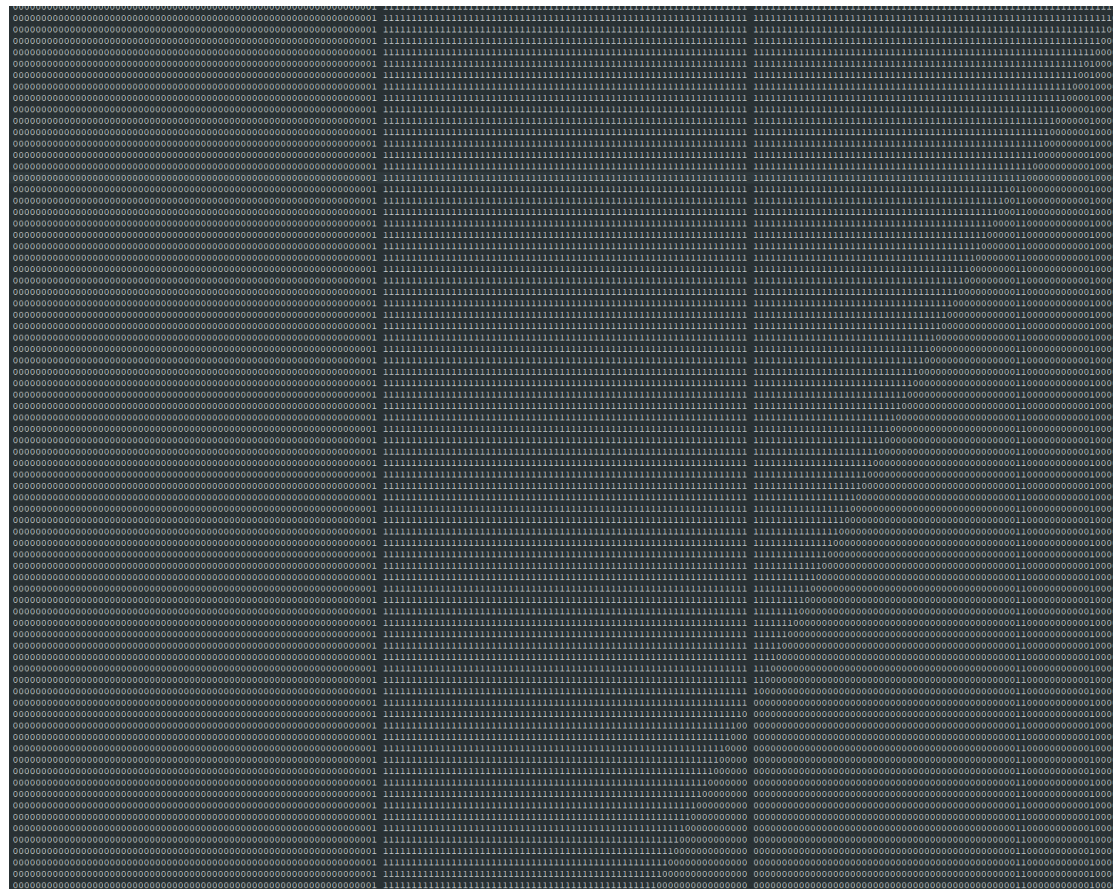
## Manual test on fvalue with length greater than 64

Due to longest integer supported by c++11 is 64bit, for fvalue longer than 64 bit, automated test can't be easily performed. Hence manual test is done to confirm that it still functioning correctly.

The test contain 5 input fvalues each consist of 129 chars as shown below

[illegible]

To check if the tree is correct, all 5 string in fvalues are evaluated and confirmed to return 1. To check all other string will return 0, tree is printed on screen and verified the tree is the same as I expected.



The picture shown above is the terminal output of how node is considered, just as expected, in general node is picked from left to right (from  $x_0$ ), where bit turn from 1 to 0 meaning it was in consideration. It's interesting to note that bit 5 and 17,18 had stayed 1, it is due to condition ii, due to  $x_5$  and  $x_{17}, x_{18}$  can be merged while calculating the implicant.

Effect of speed on how Internal representation of minterm is implemented

For the cause of general purpose, the minterm is internal represented as vector in the end, to allow simplification of trees with each of fvalue string longer than 64 bit, which can not otherwise performed with a single word.

However, it creates overhead for strings that is shorter or equal to 64 bit, which leading to longer run time. To understanding the effect, a test is done, by generating fvalues string from 00000000000 (minterm 0) all to 111111111111 (minterm  $2^{13}-1$ ), total of  $2^{13}$  fvaules. With the result shown in the table below, is about 10 times slower, but consider the number of term access during the MQ algorithm growth exponentially (NP-hard) its affect will become more significant if number of term increase, such as  $2^{32}$  terms.

Term as vector		Term as single 64-bit word	
0	1	0	1
1		1	
2		2	
3		3	
4		4	
5		5	
6		6	
7		7	
8		8	
9		9	
10		10	
11		11	
12		12	
13		13	
the tree has 1 nodes		the tree has 1 nodes	
compare to tree in assl require 16383 node		compare to tree in assl require 16383 node	
reduced by 99.9939%		reduced by 99%	
the buildcompactbdt function takes 62391ms		the buildcompactbdt function takes 5560.45ms	

## Appendix

### Binary trick

As some of the binary comparison will be performed many times, and the time complexity of the problem trying to solve is NP-hard. It is essential to have an efficient binary operation, as it will reduce the base of the exponential time function.

But to improve readability of the code, most of the binary tricks are warped inside structure/functions with self-explanatory names.

Please refer to the start of “Find prime implicant” for the explanation of the most important and complicated structure “intermittent” .

Structure “bitcount” is explained in the next page, to explain the concept of this optimisation. If have further question about how these function are implemented, please refer to the source file which have comment explains the why such a binary bitwise operation is performed.

## Tricks for bitcount

Tricks for bit count is quite complicated, so here given a overview of how does it count all the digit efficiently.

### Counting each digit

At start assume term are single integer. For example to count the following term vector: 001111 000001 001000 001111 000000. As shown, it the vector has size 5 (0b101), so the maximum count for any digit will max by 3 bit. Hence, leave space 3 for each count.

Start by creating the mask, which will be 100100 in this case. Do shift and mask and save to a vector by using method show below, for the 001111.

Term been counting	Mask and shift	Value stored in vector	Vector index
001111	$\gg 0 \& 100100$ <sup>19</sup>	001001	0
001111	$\gg 1 \& 100100$	000001	1
001111	$\gg 2 \& 100100$	000001	2

Do the same for next term 000001, get {001010, 000001, 000001} and repeat for all term, finally get {011011, 000010, 000010}

### Reading each digit

To read count for digit  $x$ , simply to let column be the quotient and row be the remainder of  $x/3$ . As table below, for example for  $x=2$ , it have count of 2. (bolded)

Vector index	Higher 3 bit		Lower 3 bit	
	Count digit	value	Count digit	value
0	Digit 3	011 (3)	Digit 0	011 (3)
1	Digit 4	000 (0)	<b>Digit 1</b>	<b>010 (2)</b>
2	Digit 5	000 (0)	Digit 2	010 (2)

### For term as vector

The same method can be applied for vector of integers.

By splitting vector to word of integers called zones and count them separately. For example, with terms being  $\{\{a_1, a_0\}, \{b_1, b_0\}, \{c_1, c_0\}\}$ , the zone 0 will be count for  $a_0, b_0, c_0$  and zone 1 will count for  $a_1, b_1, c_1$ .

And store the count vector as 1D, by setting the start index for each zone being the zone number multiplied by space, for start index zone 1 in this case is  $1*2=2$ .

---

<sup>19</sup>  $\gg 0 \& 100100$  means means right shift 001111 by 0, bitwise AND with 100100

11356, 5930, 3424, 12819, 3327, 8367, 13177, 15911, 903, 16154, 10573, 11752, 4063, 4299, 1759, 6850, 13273, 13223, 1901, 14457, 3518, 16095, 15926, 6711, 7812, 13295, 3623, 6358, 6236, 14460, 11147, 10894, 8867, 836, 16253, 12257, 10710, 5581, 5881, 14011, 11103, 2010, 11693, 8345, 7523, 10122, 278, 15702, 4301, 8285, 8799,





## Set 3

1209 1208 1207 1206 1205 1204 1203 1202 1201 1200 1199 1198 1197 1196 1195 1194 1193 1192 1191 1190 1189 1188 1187 1186 1185 1184 1183 1182 1181 1180 1179 1178 1177 1176 1175 1174 1173 1172 1171 1170 1169 1168 1167 1166 1165 1164 1163 1162 1161 1160 1159 1158 1157 1156 1155 1154 1153 1152 1151 1150 1149 1148 1147 1146 1145 1144 1143 1142 1141 1140 1139 1138 1137 1136 1135 1134 1133 1132 1131 1130 1129 1128 1127 1126 1125 1124 1123 1122 1121 1120 1119 1118 1117 1116 1115 1114 1113 1112 1111 1110 1109 1108 1107 1106 1105 1104 1103 1102 1101 1100 1099 1098 1097 1096 1095 1094 1093 1092 1091 1090 1089 1088 1087 1086 1085 1084 1083 1082 1081 1080 1079 1078 1077 1076 1075 1074 1073 1072 1071 1070 1069 1068 1067 1066 1065 1064 1063 1062 1061 1060 1059 1058 1057 1056 1055 1054 1053 1052 1051 1050 1049 1048 1047 1046 1045 1044 1043 1042 1041 1040 1039 1038 1037 1036 1035 1034 1033 1032 1031 1030 1029 1028 1027 1026 1025 1024 1023 1022 1021 1020 1019 1018 1017 1016 1015 1014 1013 1012 1011 1010 1009 1008 1007 1006 1005 1004 1003 1002 1001 1000 999 998 997 996 995 994 993 992 991 990 989 988 987 986 985 984 983 982 981 980 979 978 977 976 975 974 973 972 971 970 969 968 967 966 965 964 963 962 961 960 959 958 957 956 955 954 953 952 951 950 949 948 947 946 945 944 943 942 941 940 939 938 937 936 935 934 933 932 931 930 929 928 927 926 925 924 923 922 921 920 919 918 917 916 915 914 913 912 911 910 909 908 907 906 905 904 903 902 901 900 899 898 897 896 895 894 893 892 891 890 889 888 887 886 885 884 883 882 881 880 879 878 877 876 875 874 873 872 871 870 869 868 867 866 865 864 863 862 861 860 859 858 857 856 855 854 853 852 851 850 849 848 847 846 845 844 843 842 841 840 839 838 837 836 835 834 833 832 831 830 829 828 827 826 825 824 823 822 821 820 819 818 817 816 815 814 813 812 811 810 809 808 807 806 805 804 803 802 801 800 799 798 797 796 795 794 793 792 791 790 789 788 787 786 785 784 783 782 781 780 779 778 777 776 775 774 773 772 771 770 769 768 767 766 765 764 763 762 761 760 759 758 757 756 755 754 753 752 751 750 749 748 747 746 745 744 743 742 741 740 739 738 737 736 735 734 733 732 731 730 729 728 727 726 725 724 723 722 721 720 719 718 717 716 715 714 713 712 711 710 709 708 707 706 705 704 703 702 701 700 699 698 697 696 695 694 693 692 691 690 689 688 687 686 685 684 683 682 681 680 679 678 677 676 675 674 673 672 671 670 669 668 667 666 665 664 663 662 661 660 659 658 657 656 655 654 653 652 651 650 649 648 647 646 645 644 643 642 641 640 639 638 637 636 635 634 633 632 631 630 629 628 627 626 625 624 623 622 621 620 619 618 617 616 615 614 613 612 611 610 609 608 607 606 605 604 603 602 601 600 599 598 597 596 595 594 593 592 591 590 589 588 587 586 585 584 583 582 581 580 579 578 577 576 575 574 573 572 571 570 569 568 567 566 565 564 563 562 561 560 559 558 557 556 555 554 553 552 551 550 549 548 547 546 545 544 543 542 541 540 539 538 537 536 535 534 533 532 531 530 529 528 527 526 525 524 523 522 521 520 519 518 517 516 515 514 513 512 511 510 509 508 507 506 505 504 503 502 501 500 499 498 497 496 495 494 493 492 491 490 489 488 487 486 485 484 483 482 481 480 479 478 477 476 475 474 473 472 471 470 469 468 467 466 465 464 463 462 461 460 459 458 457 456 455 454 453 452 451 450 449 448 447 446 445 444 443 442 441 440 439 438 437 436 435 434 433 432 431 430 429 428 427 426 425 424 423 422 421 420 419 418 417 416 415 414 413 412 411 410 409 408 407 406 405 404 403 402 401 400 399 398 397 396 395 394 393 392 391 390 389 388 387 386 385 384 383 382 381 380 379 378 377 376 375 374 373 372 371 370 369 368 367 366 365 364 363 362 361 360 359 358 357 356 355 354 353 352 351 350 349 348 347 346 345 344 343 342 341 340 339 338 337 336 335 334 333 332 331 330 329 328 327 326 325 324 323 322 321 320 319 318 317 316 315 314 313 312 311 310 309 308 307 306 305 304 303 302 301 300 299 298 297 296 295 294 293 292 291 290 289 288 287 286 285 284 283 282 281 280 279 278 277 276 275 274 273 272 271 270 269 268 267 266 265 264 263 262 261 260 259 258 257 256 255 254 253 252 251 250 249 248 247 246 245 244 243 242 241 240 239

## Set 4

[illegible]



Set 5

[illegible]