

Create reduced binary decision tree

Table of Contents

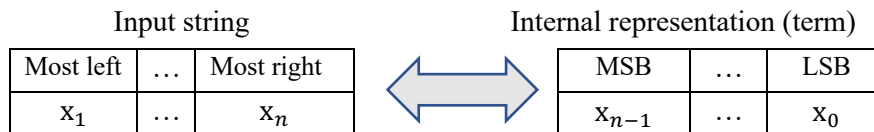
Introduction.....	2
Description of Algorithm.....	2
Buildcompactbdt.....	3
Evalcompactbdt.....	6
Internal representation of minterm in C++	7
Implemten of Buildcompactbdt in C++	8
Generate minterm.....	8
Find prime implicant.....	9
generate the binary decision tree.....	11
Implemten of Evalcompactbdt 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^{1,2} 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)³ Using chart(heuristic) or Petrick's method⁴ 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.

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

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

³ 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 1st part of Experimental Evaluation for the experiment.

⁴Donald Krambeck (2016) *Prime Implicant Simplification Using Petrick's Method*
<https://www.allaboutcircuits.com/technical-articles/prime-implicant-simplification-using-petricks-method/> [29th/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”⁵ 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		
$x_1x_2x_3x_4$ ⁶	$x_4x_3x_2x_1$	mask	Minterm
1-10	01- ⁷ 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) ⁸	1000	0000	01-(1)	0010	0101
--1(0)	1100	0010	-00(1)	1000	0000

⁵Donald Krambeck (2016) *Everything About the Quine-McCluskey Method*

<https://www.allaboutcircuits.com/technical-articles/everything-about-the-quine-mccluskey-method/>
[29th/04/2018]

⁶ $x_1x_2x_3x_4$ 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).

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

⁸ 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
$x_4x_3x_2x_1$	$x_4x_3x_2x_1$	$x_4x_3x_2x_1$	$x_4x_3x_2x_1$
-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		
$x_4x_3x_2x_1$	mask	Minterm	$x_4x_3x_2x_1$	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 $\overline{x_3} \overline{x_2} x_1$ 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 $x_3 \overline{x_2} x_1$ evaluate to 0.

The binary decision tree produced by this method is shown below

<pre> 0 x1 1 x2 x3 2 x3 1 x2 x4 3 1 0 1 0 1 0 4 </pre>	
*output from my print tree function	*graphical enhanced by hand

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 “ x_{num} ”, 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 wdlength 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 implicant⁹

Quine–McCluskey algorithm is used in this step, also by taking inspiration from (Jadhav Vitthal, 2012)¹⁰, 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.

⁹ 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.

¹⁰ Corporate author (2012) *Modified Quine-McCluskey Method*
<https://pdfs.semanticscholar.org/650a/ae0c3e59bdd12e7b22158162166e4437949e.pdf> [29th/04/2018]

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

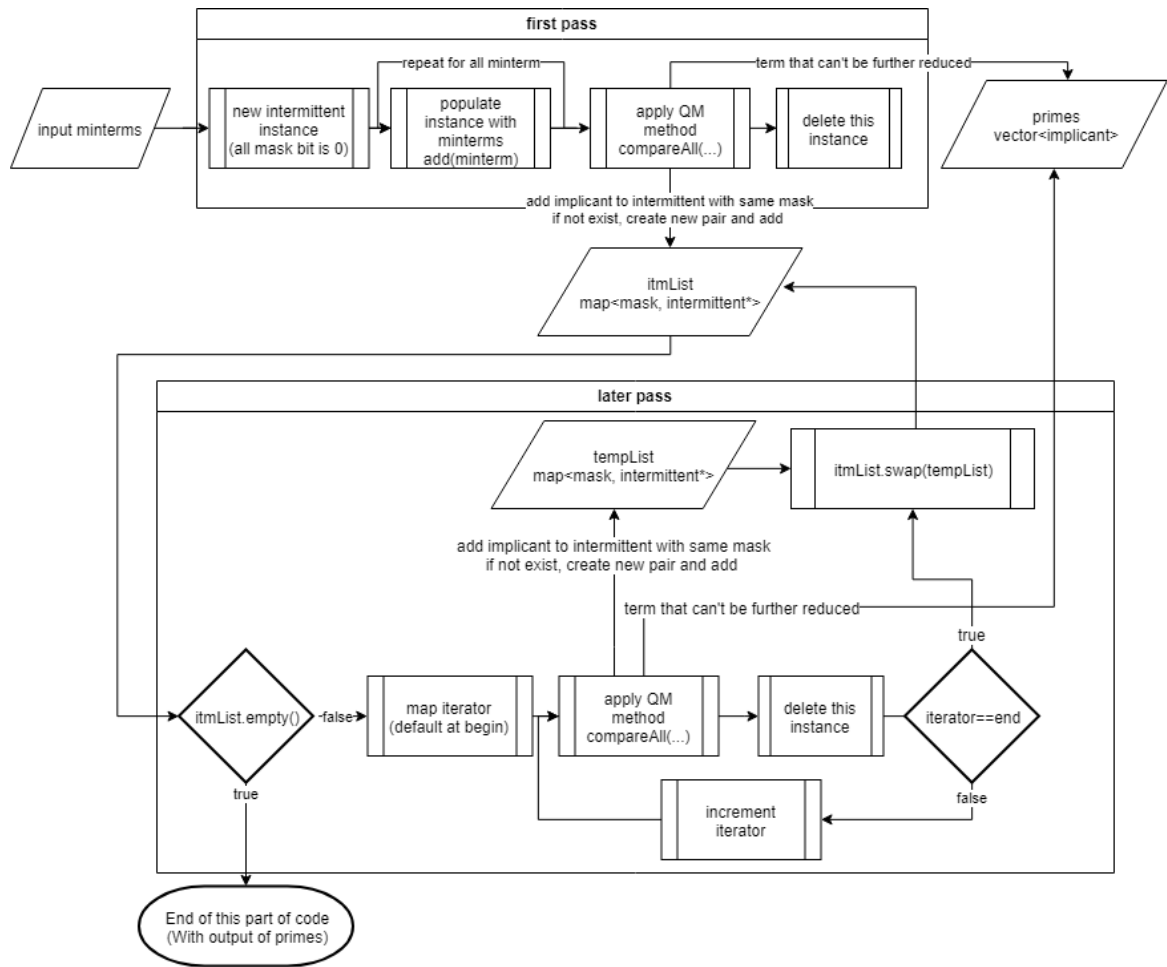


Figure 1

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 tempList, 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 automatedly 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))
}

```

¹¹ 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 ith 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 set¹² 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 resources¹³ 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	2	x3 1 x2 x4
3	1 0 1 x4 0 x4	3	1 0 1 0 1 0
4	1 0 1 0	4	
17 nodes		13 nodes	

The second set I tested is with minterm¹⁴ 40,51,37,38,27,50,14,5,42,53,61,20,21,16,15,22,63,30,25,47,6,19,17,23,13,56,26,9,52,45,24,12,1,7,62,41,28,32,58,8,0,29,39,35,43,60,57,59,11,31,3,44,54,46,36,4,48,2,34,18, 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.

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

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

¹⁴ 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 optimised	After optimised
<pre>the tree has 1 nodes compare to tree in ass1 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, 34,356,724 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 ass1 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) /usr/Desktop-61A49FE /usr/include/C3</pre>
<pre>minterm=minterm&baseNot; int pcount=popcount(minterm); minterms[pcount].push_back(minterm); used[pcount].push_back(false);</pre>	<pre>minterm=minterm&baseNot; int pcount=popcount(minterm); if(notRepeated¹⁵(minterms[pcount],minterm)){ used[pcount].push_back(false); }</pre>

¹⁵ 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 ¹⁶	All 3 condition i,ii,iii
00001 00010	<pre> 0 1 2 3 4 5 x1 x2 0 x3 0 x4 0 x5 x5 0 1 1 0 </pre> <p>13 nodes in total</p>	<pre> 0 1 2 3 4 5 x1 x2 0 x3 0 x4 0 x5 x5 0 1 1 0 </pre> <p>13 nodes in total</p>
10000 01000	<pre> 0 1 2 3 4 5 x1 x2 x2 0 x3 x3 0 x4 0 x4 0 x5 0 x5 0 1 0 1 0 </pre> <p>19 nodes in total</p>	<pre> 0 1 2 3 4 5 x3 x4 0 x5 0 x1 0 x2 x2 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 10,000¹⁷ terms of fvalue. From in the range of 0000000000000000 (minterm 0) to 1111111111111111 (minterm $2^{15}-1$).

Minterm Set reference number ¹⁸	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
Average Nodes	16828.6	16814.6	65535

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.

¹⁶ As condition described in step 4 of section Buildcompactbdt

¹⁷ Max number of integers allowed to generate by random.org

¹⁸ 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.

[illegible]

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 000000000000 (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 compare to tree in ass1 require 16383 node reduced by 99.9939%		the tree has 1 nodes compare to tree in ass1 require 16383 node 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	$\gg 0 \& 100100$ ¹⁹	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)	Digit 1	010 (2)
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 $1*2=2$.

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

All the term in the list below are in form of minterm representation of fvaules, all from range 0 to $2^{15}-1$, generated by integer set from random.org.

Set 2

Set 3

1000 1536 1548 1560 1574 1584 1594 1596 1598 1600 1602 1604 1606 1608 1610 1612 1614 1616 1618 1620 1622 1624 1626 1628 1630 1632 1634 1636 1638 1640 1642 1644 1646 1648 1650 1652 1654 1656 1658 1660 1662 1664 1666 1668 1670 1672 1674 1676 1678 1680 1682 1684 1686 1688 1690 1692 1694 1696 1698 1700 1702 1704 1706 1708 1710 1712 1714 1716 1718 1720 1722 1724 1726 1728 1730 1732 1734 1736 1738 1740 1742 1744 1746 1748 1750 1752 1754 1756 1758 1760 1762 1764 1766 1768 1770 1772 1774 1776 1778 1780 1782 1784 1786 1788 1790 1792 1794 1796 1798 1800 1802 1804 1806 1808 1810 1812 1814 1816 1818 1820 1822 1824 1826 1828 1830 1832 1834 1836 1838 1840 1842 1844 1846 1848 1850 1852 1854 1856 1858 1860 1862 1864 1866 1868 1870 1872 1874 1876 1878 1880 1882 1884 1886 1888 1890 1892 1894 1896 1898 1900 1902 1904 1906 1908 1910 1912 1914 1916 1918 1920 1922 1924 1926 1928 1930 1932 1934 1936 1938 1940 1942 1944 1946 1948 1950 1952 1954 1956 1958 1960 1962 1964 1966 1968 1970 1972 1974 1976 1978 1980 1982 1984 1986 1988 1990 1992 1994 1996 1998 2000 2002 2004 2006 2008 2010 2012 2014 2016 2018 2020 2022 2024 2026 2028 2030 2032 2034 2036 2038 2040 2042 2044 2046 2048 2050 2052 2054 2056 2058 2060 2062 2064 2066 2068 2070 2072 2074 2076 2078 2080 2082 2084 2086 2088 2090 2092 2094 2096 2098 2100 2102 2104 2106 2108 2110 2112 2114 2116 2118 2120 2122 2124 2126 2128 2130 2132 2134 2136 2138 2140 2142 2144 2146 2148 2150 2152 2154 2156 2158 2160 2162 2164 2166 2168 2170 2172 2174 2176 2178 2180 2182 2184 2186 2188 2190 2192 2194 2196 2198 2200 2202 2204 2206 2208 2210 2212 2214 2216 2218 2220 2222 2224 2226 2228 2230 2232 2234 2236 2238 2240 2242 2244 2246 2248 2250 2252 2254 2256 2258 2260 2262 2264 2266 2268 2270 2272 2274 2276 2278 2280 2282 2284 2286 2288 2290 2292 2294 2296 2298 2300 2302 2304 2306 2308 2310 2312 2314 2316 2318 2320 2322 2324 2326 2328 2330 2332 2334 2336 2338 2340 2342 2344 2346 2348 2350 2352 2354 2356 2358 2360 2362 2364 2366 2368 2370 2372 2374 2376 2378 2380 2382 2384 2386 2388 2390 2392 2394 2396 2398 2400 2402 2404 2406 2408 2410 2412 2414 2416 2418 2420 2422 2424 2426 2428 2430 2432 2434 2436 2438 2440 2442 2444 2446 2448 2450 2452 2454 2456 2458 2460 2462 2464 2466 2468 2470 2472 2474 2476 2478 2480 2482 2484 2486 2488 2490 2492 2494 2496 2498 2500 2502 2504 2506 2508 2510 2512 2514 2516 2518 2520 2522 2524 2526 2528 2530 2532 2534 2536 2538 2540 2542 2544 2546 2548 2550 2552 2554 2556 2558 2560 2562 2564 2566 2568 2570 2572 2574 2576 2578 2580 2582 2584 2586 2588 2590 2592 2594 2596 2598 2600 2602 2604 2606 2608 2610 2612 2614 2616 2618 2620 2622 2624 2626 2628 2630 2632 2634 2636 2638 2640 2642 2644 2646 2648 2650 2652 2654 2656 2658 2660 2662 2664 2666 2668 2670 2672 2674 2676 2678 2680 2682 2684 2686 2688 2690 2692 2694 2696 2698 2700 2702 2704 2706 2708 2710 2712 2714 2716 2718 2720 2722 2724 2726 2728 2730 2732 2734 2736 2738 2740 2742 2744 2746 2748 2750 2752 2754 2756 2758 2760 2762 2764 2766 2768 2770 2772 2774 2776 2778 2780 2782 2784 2786 2788 2790 2792 2794 2796 2798 2800 2802 2804 2806 2808 2810 2812 2814 2816 2818 2820 2822 2824 2826 2828 2830 2832 2834 2836 2838 2840 2842 2844 2846 2848 2850 2852 2854 2856 2858 2860 2862 2864 2866 2868 2870 2872 2874 2876 2878 2880 2882 2884 2886 2888 2890 2892 2894 2896 2898 2900 2902 2904 2906 2908 2910 2912 2914 2916 2918 2920 2922 2924 2926 2928 2930 2932 2934 2936 2938 2940 2942 2944 2946 2948 2950 2952 2954 2956 2958 2960 2962 2964 2966 2968 2970 2972 2974 2976 2978 2980 2982 2984 2986 2988 2990 2992 2994 2996 2998 3000 3002 3004 3006 3008 3010 3012 3014 3016 3018 3020 3022 3024 3026 3028 3030 3032 3034 3036 3038 3040 3042 3044 3046 3048 3050 3052 3054 3056 3058 3060 3062 3064 3066 3068 3070 3072 3074 3076 3078 3080 3082 3084 3086 3088 3090 3092 3094 3096 3098 3100 3102 3104 3106 3108 3110 3112 3114 3116 3118 3120 3122 3124 3126 3128 3130 3132 3134 3136 3138 3140 3142 3144 3146 3148 3150 3152 3154 3156 3158 3160 3162 3164 3166 3168 3170 3172 3174 3176 3178 3180 3182 3184 3186 3188 3190 3192 3194 3196 3198 3200 3202 3204 3206 3208 3210 3212 3214 3216 3218

Set 5

12842, 6629, 2897, 14669, 14725, 3197, 12123, 4323, 14752, 4304, 12288, 14471, 14505, 12472, 1978, 465, 4554, 2711, 5691, 2955, 12783, 15944, 3449, 12553, 2969, 1750, 10499, 10380, 7532, 237, 7347, 14224, 1214, 11329, 13571, 3802, 12211, 901, 10590, 14610, 4971, 7846, 7793, 10774, 1856, 5302, 12399, 3256, 14991, 12775, 13627, 4980, 10923, 15265, 13733, 13211, 1056, 5538, 2262, 7194, 3689, 15194, 13760, 1439, 6186, 5589, 14714, 3015, 2149, 1863, 10962, 12504, 3820, 1516, 4935, 6644, 14880, 5094, 13785, 7465, 13066, 11933, 13716, 5019, 11877, 4139, 13509, 1389, 9970, 15975, 2925, 9733, 10336, 2610, 5595, 14054, 2004, 8648, 11407, 1970, 6307, 14240.

