

DenseQMC: an efficient bit-slice implementation of the Quine-McCluskey algorithm*

Aleksei Udovenko

SnT, University of Luxembourg
aleksei@affine.group

Abstract. This note describes a new efficient bit-slice implementation DenseQMC of the Quine-McCluskey algorithm for finding all prime implicants of a Boolean function in the dense case. It is practically feasible for $n \leq 23$ when run on a common laptop or for $n \leq 27$ when run on a server with 1 TiB RAM.

This note also outlines a very common mistake in the implementations of the Quine-McCluskey algorithm, leading to a quadratic slowdown. An optimized corrected implementation of the classic approach is also given (called SparseQMC).

The implementation is freely available at github.com/hellman/Quine-McCluskey.

Keywords: Boolean minimization · Two-level minimization · Quine-McCluskey method · Implementation · Bit-slice

1 Introduction

Boolean/logic minimization is a broad topic with numerous applications, including logic synthesis, expressing problems for SAT solvers (e.g., in symmetric-key cryptanalysis [24]), performing Qualitative Comparative Analysis (QCA) in social sciences. The simplest minimization setting is the *two-level minimization*, in which the target function has to be expressed in one of the two following types of formulas: *sum-of-products* (SOP) or *product-of-sums* (POS), also known as *disjunctive normal form* (DNF) and *conjunctive normal form* (CNF) respectively. Here, a “sum” refers to the logic OR operator and a “product” refers to the logic AND operator. The two formula types are dual to each other, as they are related by the De Morgan’s laws. Therefore, it is sufficient to study any one of the two, and this work describes algorithms for DNF minimization.

A DNF/SOP formula consists of *clauses* connected by the OR operator, and where each clause consists of variables or their negations connected by the AND operator. According to the sum/product interpretation, the OR operator is often denoted by “+” and the AND operator is omitted. Furthermore, the logic negation (NOT) is denoted by the prime mark (for example, x'). For example, the 3-bit majority function Maj can be expressed in DNF as

$$\text{Maj}(x, y, z) = xy + xz + yz.$$

The DNF minimization problem asks, given an expression of a Boolean function, to find a DNF formula with the smallest possible number of logic AND/OR operations (sometimes, only OR operations are counted, leading to minimization of the number of clauses). This work considers the most general case when the input function is given by its truth table.

The Quine-McCluskey method In a series of works [15, 17, 18], Quine and McCluskey developed an algorithm for two-level minimization. It consists of two steps:

* The work was supported by the Luxembourg National Research Fund’s (FNR) and the German Research Foundation’s (DFG) joint project APLICA (C19/IS/13641232).

1. Find all minimal products that can be consistently included in the DNF (also called *prime implicants*). Minimality here means that no proper divisor of the product is consistent with the target function.
2. Choose a subset of minimal products according to the problem’s goal: minimizing the number of products or minimizing the total number of AND/OR operations.

The second step is an instance of the SetCover problem, which is NP-hard. Furthermore, NP-hard instances were proven to occur in this setup. Therefore, one has to resort to heuristic methods. Classically, the Petrick’s method [16] was used. However, a modern integer optimization suite may perform better (as noted already in [6]). In addition, there are several dedicated exact or approximate heuristic solvers for the SetCover problem in general [7, 12, 23], as well as methods tailored to the Boolean minimization problem [11].

New method This work focuses on the first step of the Quine-McCluskey algorithm, namely, finding all prime implicants of the function. Although the second step is overwhelmingly dominating the full procedure, an efficient solution to the first step opens doors to fast heuristic approximate methods for the second step. In particular, the set of prime implicants itself provides a DNF expression of the function (or CNF of the function’s complement) which can be sufficiently compact (even if containing redundant terms).

For the first step of the Quine-McCluskey algorithm, we propose a new implementation based on multidimensional ternary transforms and on the bit-slicing technique, called DenseQMC. To the best of our knowledge, the new approach significantly outperforms all available previous implementations and expands the feasible number n of function’s inputs to $n = 23$ when run on a laptop and to $n = 27$ when run on a server equipped with 1 TiB of RAM (see Table 1). It is in particular useful for very dense problems (computing a DNF of a dense function or a CNF of a sparse function), which are the worst-case for the Quine-McCluskey algorithm and where sparse methods are too slow. For a fair comparison and due to absence of competitive classic (sparse) implementations, we also developed an optimized implementation of the classic Quine-McCluskey method, called SparseQMC. Surprisingly, it also outperforms existing reports and even the new DenseQMC when density is at most 50%-60%.

A brief benchmark is given in Table 1. We do not provide explicit comparison with existing implementations, since we could not find any competitive implementation and/or sufficient performance information. For a rough comparison, the work [10] reports 34 seconds for a function with $n = 11$, [5] reports 5 seconds for a very sparse function with $n = 15$, all of which are done instantly by any of our implementations; [14] gives mixed CPU/GPU timings such as 1000 seconds and 10 GiB RAM for the dense case of $n = 20$, 10 minutes for an $n = 24$ -bit function of density 70%, 2000 seconds for an $n = 28$ -bit function of density 30%, close to 10^6 seconds on an $n = 32$ -bit function of density 42% using disk storage (due to ambiguous reports and absence of available implementation, it is difficult to provide a clear comparison). Note that the dense case is often occurring in practice when optimizing a CNF formula of a sparse function, for example, [2] report 2 hours of work for the case of $n = 16$ and 82%-dense function, appearing in cryptographic applications. Note that this work does not intend to compete with sparse or approximate methods such as ESPRESSO-Exact [3] or more recent “Consistency Cubes” method [4].

The source code is publicly available at

github.com/hellman/Quine-McCluskey

2 Definitions and notation

The Boolean AND,OR,NOT operations are denoted by \wedge, \vee, \neg respective. They can be operate on single bits or bitwise on bit-vectors. The left and right shift operations on bit-vectors are denoted by \lll and \ggg respectively.

Table 1: Time/memory benchmark of the optimized implementations of new DenseQMC (Section 4, Algorithm 4) and classic SparseQMC (Section 5) algorithms on random n -bit Boolean functions with a given density. Ran on a single core of an AMD EPYC 3.2 GHz processor with 1 TiB of RAM available.

n	DenseQMC density: any		SparseQMC density: 25%		SparseQMC density: 50%		SparseQMC density: 99%	
	RAM	Time	RAM	Time	RAM	Time	RAM	Time
16	5 MiB	0.01 s	2.0 MiB	0.01 s	4.0 MiB	0.07 s	0.2 GiB	18 s
17	16 MiB	0.04 s	4.0 MiB	0.03 s	16 MiB	0.2 s	1.0 GiB	62 s
18	49 MiB	0.2 s	8 MiB	0.05 s	32 MiB	0.7 s	2.0 GiB	3.2 m
19	0.1 GiB	0.5 s	16 MiB	0.2 s	64 MiB	1.9 s	6 GiB	11.2 m
20	0.4 GiB	1.5 s	32 MiB	0.6 s	0.1 GiB	5 s	16 GiB	0.6 h
21	1.3 GiB	5 s	64 MiB	1.6 s	0.2 GiB	10 s	64 GiB	1.6 h
22	3.8 GiB	13 s	0.1 GiB	3 s	0.5 GiB	22 s	-	-
23	12 GiB	39 s	0.2 GiB	7 s	1.5 GiB	49 s	-	-
24	35 GiB	2.0 m	0.5 GiB	18 s	3.0 GiB	1.9 m	-	-
25	0.1 TiB	6.0 m	1.0 GiB	35 s	6 GiB	5.0 m	-	-
26	0.3 TiB	0.3 h	2.0 GiB	78 s	12 GiB	13.1 m	-	-
27	0.9 TiB	1.1 h	4.0 GiB	3.2 m	24 GiB	0.5 h	-	-
28	-	-	8 GiB	8.2 m	64 GiB	1.2 h	-	-
29	-	-	24 GiB	0.3 h	0.1 TiB	3.1 h	-	-
30	-	-	48 GiB	0.8 h	-	-	-	-
31	-	-	96 GiB	1.8 h	-	-	-	-

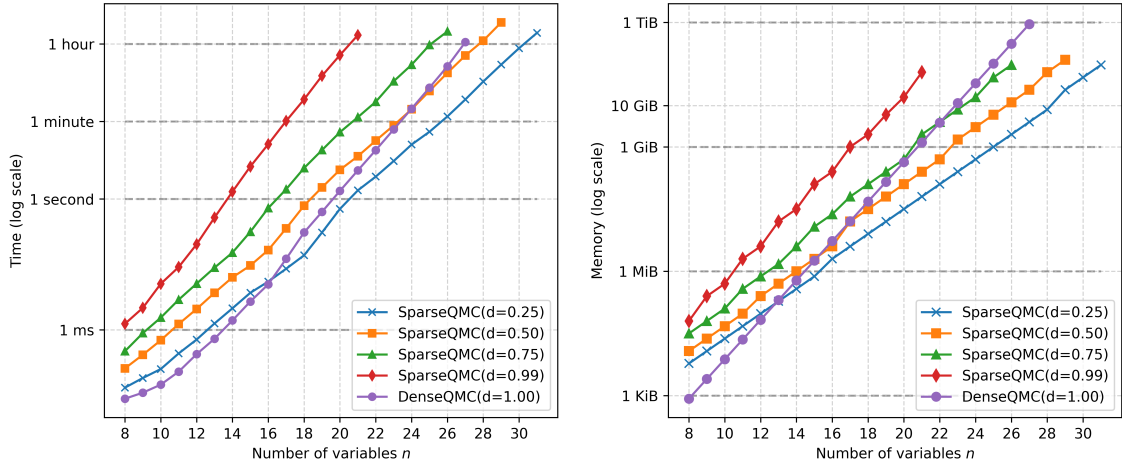


Fig. 1: Comparison of DenseQMC and SparseQMC implementations (time and memory usage) on random n -bit Boolean functions with a given density d . Evaluated on a single core of a 3.2GHz CPU.

Let $\Sigma = \{0, 1, *\}$ be the alphabet. The symbol $*$ is called a *wildcard*. Throughout the work, the number of input variables to the considered function is denoted by n . A *literal* α is either an input variable x_i or its negation $\neg x_i$.

Definition 1. A minterm is a product $\alpha_1 \dots \alpha_m$ of literals $\alpha_j \in \{x_{t_j}, x'_{t_j}\}$, $t_j \in \{1, \dots, n\}$ such that $t_i \neq t_j$ for all $i \neq j$. In other words, every variable occurs at most once in the minterm. The minterm $\alpha_1 \dots \alpha_m$ is equivalently described by the string $s = (s_1, \dots, s_n) \in \Sigma^n$ such that

$$\begin{cases} s_k = 0, & \text{if } x'_{t_k} \text{ is present in the product,} \\ s_k = 1, & \text{if } x_{t_k} \text{ is present in the product,} \\ s_k = *, & \text{otherwise.} \end{cases}$$

The weight of a minterm is defined as the number of wildcards in the string representation, equal to n minus the minterm's degree.

Example 1. Let $n = 5$. Then, the minterm $x_1 x'_3$ corresponds to the string $1*0**$.

Definition 2. A minterm $\alpha_1 \dots \alpha_m$ is called an *implicant* of a Boolean function $f : \{0, 1\}^n \rightarrow \{0, 1\}$, if $\alpha_1 \dots \alpha_m(x) = 1$ implies $f(x) = 1$ for all $x \in \{0, 1\}^n$.

Definition 3. An implicant $\alpha_1 \dots \alpha_m$ of a Boolean function f is said to be *prime*, if no proper divisor of $\alpha_1 \dots \alpha_m$ is an implicant of f . Otherwise, it is said to be *redundant*.

3 Classic Quine-McCluskey algorithm

The Quine-McCluskey algorithm constructs all implicants of the function, removing redundant implicants along the way, so that the final remaining implicants are prime. The construction relies on the following observation about combining minterms.

Observation 1. Let $\alpha_1 \dots \alpha_m$ and $\beta_1 \dots \beta_m$ be two minterms such that, for some $i \in \{1, \dots, m\}$, it is $\alpha_i = \beta'_i$ and $\alpha_j = \beta_j$ for all $j \neq i$. Then, the sum of $\alpha_1 \dots \alpha_m$ and $\beta_1 \dots \beta_m$ is functionally equivalent to the minterm $\prod_{j \neq i} \alpha_j$.

In terms of strings, two minterms can be combined if they differ exactly at one position, and the new minterm is obtained by setting the string at this position to $*$.

Example 2. The sum of minterms $0011*$ and $0111*$ is $0*11*$.

The observation works in the other way too: any implicant with at least one wildcard can be constructed as a sum of two implicants each with one wildcard less. This means that all implicants can be constructed bottom up from implicants with no wildcards.

The idea of the algorithm is to start with implicants of degree n (having no wildcards in their strings). Then, at w -th level, $w \in \{1, \dots, n\}$, all implicants with w wildcards are constructed by combining implicants from the previous level in pairs. When a new implicant is constructed, the two used implicants are marked redundant (and can be removed from memory after the level w is fully finished). The pseudocode for the high-level procedure is given in [Algorithm 1](#).

Implementation caveat So far, we have not considered how to find pairs of implicants that differ in exactly one position (as strings). This crucial step is performed in line 5 of [Algorithm 1](#). Unfortunately, there is a widespread misconception about this step, stemming from a heuristic used in the initial works of Quine and McCluskey, who did not focus on formal algorithmic complexity of the approach. In fact, most works from 19xx describe various heuristic variants in a “pen-and-paper” style. Yet, even modern implementations follow the inefficient approach, see for example [\[1, 5, 9, 10, 25\]](#). A few works that follow the efficient approach are [\[21\]](#). In the following, we first discuss the inefficient approach and then describe the efficient variants.

Algorithm 1 Quine-McCluskey algorithm for finding all prime implicants

Input: truth table of a Boolean $f : \{0, 1\}^n \rightarrow \{0, 1\}$

Output: set $S \subseteq \Sigma^n$ of prime implicants of f

```
1:  $L_0 \leftarrow \{x \in \{0, 1\}^n : f(x) = 1\} \subseteq \Sigma^n$ 
2: for  $w \in \{1, \dots, n\}$  do
3:    $R \leftarrow \emptyset$ 
4:    $L_w \leftarrow \emptyset$ 
5:   for  $s, t \in L_{w-1} : \exists i \in \{1, \dots, n\} s_i = 0, t_i = 1$  and  $s_j = t_j$  for all  $j \neq i$  do
6:      $R \leftarrow R \cup \{s, t\}$ 
7:      $L_w \leftarrow L_w \cup \{s + t\}$ 
8:    $L'_{w-1} \leftarrow L'_{w-1} \setminus R$ 
9:  $L'_n \leftarrow L_n$ 
10:  $S \leftarrow \bigcup_{w=0}^n L'_w$ 
```

Finding compatible pairs of implicants (inefficient) Since the strings s, t differ only at one position, at which they are equal to 0 and 1 respectively, it follows that the number of 1s in s is by one less than that in t . It is thus commonly suggested to sort and group each L_{w-1} by the number of 1s in the increasing order. Let $L_{w-1,u}$ denote the subset of strings from L_{w-1} which have exactly u symbols 1, $0 \leq u \leq n - w + 1$. Then, for each compatible pair s, t considered by the algorithm, it must be $s \in L_{w-1,u}$ and $t \in L_{w-1,u+1}$ for some u . It follows that, at step w , it is sufficient to only consider pairs from the set

$$(L_{w-1,0} \times L_{w-1,1}) \cup \dots \cup (L_{w-1,n-w} \times L_{w-1,n-w+1})$$

The problem is that this observation is often interpreted as the direct method to find the compatible pairs, by enumerating all pairs from $L_{w-1,u} \times L_{w-1,u+1}$, for each fitting u . This approach however leads to a quadratic complexity blowup.

Proposition 1. *There exists an infinite family of functions f such that Algorithm 1 implemented using the described heuristic has*

$$|L_{w-1,u} \times L_{w-1,u+1}| = \Omega\left(\frac{6^n}{n^3}\right).$$

for some index u .

Proof. Consider the worst case for the Quine-McCluskey algorithm (in general), the constant-1 function $f = 1$, for which all possible minterms are implicants (although only the minterm 1 is prime). For simplicity, we assume that $n = 3m$ for some integer m . Then, $L_{m,m}$ consists of strings from Σ^n with exactly m zeroes, m ones, and m wildcards, the number of which is

$$\frac{(3m)!}{m!m!m!} = \Theta\left(\sqrt{m} \left(\frac{3m}{e}\right)^{3m} \cdot \sqrt{m}^{-3} \left(\frac{m}{e}\right)^{-3m}\right) = \Theta\left(\frac{3^n}{n}\right)$$

for some constant c (using Stirling's approximation). Since $L_{m,m}$ can be obtained by replacing one 0 by 1 in each position in each string from $L_{m,m-1}$, we have $L_{m,m-1} \geq \frac{L_{m,m}}{m+1}$. It follows that for $w = m + 1$, $u = m - 1$, the set

$$L_{w-1,u} \times L_{w-1,u+1}$$

has size $\Omega\left(\frac{6^n}{n^3}\right)$. □

Finding compatible pairs of implicants (efficient) A correct implementation was described by Wegener [26, Section 2.2], who suggested to iterate through the list $L_{w-1,u}$ and, for each string s in it, check all strings that can be obtained by replacing exactly one 1 with 0 inside the set $L_{w-1,u-1}$. Note that there are at most n such strings to check. As a membership test, Wegener proposed to sort $L_{w-1,u-1}$ in advance and use binary search for each query. This leads to complexity $\mathcal{O}(3^n n^2)$ (3^n maximum minterms, at most n neighbors per a string, logarithmic cost of the binary search in $L_{w-1,u-1}$ of size at most 3^n). In practice, the membership test can be performed using a hash table in amortized constant time, leading to a probabilistic algorithm with expected time complexity of $\mathcal{O}(3^n n)$ operations on minterms (n -bit strings) and memory accesses.

Note that *there is no benefit* for the asymptotic complexity in grouping elements of L_{w-1} by the number of 1s: the cost of a membership test is asymptotically the same for sets of size 3^n and $3^n/n$ (both for binary search and hash-based methods). Although such grouping may improve locality of memory operations (and thus, improve the constant behind the asymptotic time complexity), it also makes the implementation more complex.

4 DenseQMC: New bit-slice algorithm for the dense case

This section describes a new algorithm for finding prime implicants, which exploits the ideas of Quine and McCluskey in a more careful implementation. It has complexity $\mathcal{O}(3^n n)$ of bit operations on *any* function of n variables, which is further reduced by a bit-slice style implementation allowing to maximally utilize the CPU registers. Its high-level structure is based on iterating over *dimensions* (variables), not over *weights*. A similar idea was proposed already in [19], but not sufficiently developed and detailed. Furthermore, we also employ and develop the idea of [5] to consider the full possible state of size 3^n packed in a dense bit-vector similar to implementations in [14, 21]. In addition, the core of the algorithm essentially describes a Boolean circuit, reminiscent of hardware accelerators proposed in [13, 14].

4.1 Multidimensional ternary transforms

State representation The problem of finding prime implicants takes as input a subset of $\{0, 1\}^n$ and outputs a subset of $\Sigma^n = \{0, 1, *\}^n$. The input set can be also naturally embedded in Σ^n . This allows to represent the problem as a map from 2^{Σ^n} to itself. In the general (dense) case, it is efficient to represent a subset of Σ^n as a bit-vector S of size 3^n (i.e., $S \in \{0, 1\}^{3^n}$). For now, we assume that bits in the memory are indexed by strings from Σ^n ; concrete implementation is described in [Subsection 4.2](#). For example, by $S(s)$ we denote the single bit indicating whether s belongs to the set represented by S or not. As we shall see, it is possible to implement the required mapping by manipulating such a state in-place by a fixed Boolean circuit.

The Quine-McCluskey algorithm can be directly reinterpreted to work on such a state. Essentially, we only change the data structure behind the sets L_w , which is a further evolution of the observation about efficient implementation in the previous section.

Definition 4. *Define two Boolean operations*

$$\begin{aligned} \text{Merge} : \mathbb{F}_2^3 &\rightarrow \mathbb{F}_2^3 : (\chi_0, \chi_1, \chi_*) \mapsto (\chi_0, \chi_1, \chi_* \vee (\chi_0 \wedge \chi_1)), \\ \text{Reduce} : \mathbb{F}_2^3 &\rightarrow \mathbb{F}_2^3 : (\chi_0, \chi_1, \chi_*) \mapsto (\chi_0 \wedge \neg \chi_*, \chi_1 \wedge \neg \chi_*, \chi_*). \end{aligned}$$

These operators can be viewed as maps from 2^Σ to itself. The Merge operator will be used to combine two compatible minterms; the Reduce operator will be used to remove redundant minterms.

Theorem 1. *The Quine-McCluskey algorithm can be implemented as a Boolean circuit mapping the set $\{0, 1\}^{3^n}$ to itself, interpreted as 2^{Σ^n} . The circuit uses $3^{n-1}n$ Boolean OR and NOT gates, and $3^n n$ AND gates.*

Proof. See [Algorithm 2](#). Complexity: the algorithm considers all triples having $n-1$ positions equal and the other position taking all three possible values (in the fixed order $(0, 1, *)$). There are $n3^{n-1}$ such triples. The Merge and Reduce operations are applied exactly once per such a triple.

Correctness: it is easy to see that I represents the same state as the set $L_0 \cup L_1 \dots \cup L_w$ from [Algorithm 1](#) at matching steps. More precisely, each iteration of the first inner loop in [Algorithm 2](#) has the same effect as an iteration of the inner loop in [Algorithm 1](#); the second inner loop in [Algorithm 2](#) has the same effect as the line $L_w \leftarrow L_w \setminus R$ in [Algorithm 1](#). \square

Algorithm 2 Boolean circuit implementation of the Quine-McCluskey algorithm for finding all prime implicants

Input: indicator set $I \in \{0, 1\}^{3^n}$ of the support of a Boolean function $f : \{0, 1\}^n \rightarrow \{0, 1\}$

Output: indicator set $O \in \{0, 1\}^{3^n}$ of the set $O \subseteq \Sigma^n$ of prime implicants of f

```

1:  $S \leftarrow I$  ▷ working state
2: for  $w \in \{1, \dots, n\}$  do
3:   for  $s, t, u \in \Sigma^n$  :
        $u$  has  $w$  wildcards,
        $\exists i \in \{1, \dots, n\} s_i = 0, t_i = 1, u_i = *$  and  $s_j = t_j = u_j$  for all  $j \neq i$  do
4:      $(S(s), S(t), S(v)) \leftarrow \text{Merge}(S(s), S(t), S(v))$ 
5:   for  $s, t, u \in \Sigma^n$  :
        $u$  has  $w$  wildcards,
        $\exists i \in \{1, \dots, n\} s_i = 0, t_i = 1, u_i = *$  and  $s_j = t_j = u_j$  for all  $j \neq i$  do
6:      $(S(s), S(t), S(v)) \leftarrow \text{Reduce}(S(s), S(t), S(v))$ 
7:  $O \leftarrow S$ 

```

Remark 1. Lines 2 and 4 of [Algorithm 2](#) define wirings in the Boolean circuit, and so do not contribute to its complexity. Software implementation of it would require careful enumeration of minterms with w wildcards for all values of w . It is not included here as the alternative implementation below is superior.

The Quine-McCluskey algorithm (and its Boolean circuit implementation from [Algorithm 2](#)) traverses the space of minterms in the order of increasing number of wildcards w in the associated strings (equal to the decreasing order of minterms' degree). The correctness of the processing order is crucial, since Merge transfers information from lower to higher w , while Reduce goes in the other direction. A crucial observation is that these two processes can be completely separated. More precisely, second inner loop in [Algorithm 2](#) (Lines 4-5) can be performed in a separate loop on w following the first loop on w . However, the second loop has to iterate w in the decreasing order. In terms of implicants, the interpretation is that we can first compute all the implicants of the function, and only then remove all the redundant implicants.

Lemma 1. Let $\text{MergeAll}, \text{ReduceAll} : \{0, 1\}^{3^n} \rightarrow \{0, 1\}^{3^n}$ be maps such that, for all $I \in \{0, 1\}^{3^n}$, $M = \text{MergeAll}(I)$, $O = \text{ReduceAll}(M)$, it holds:

$$M(u) = \begin{cases} I(u) & \text{if } u \in \{0, 1\}^n, \\ \bigvee_{s, t \in \Sigma^n: s+t=u} M(s) \wedge M(t) & \text{otherwise,} \end{cases}$$

$$O(s) = M(s) \wedge \neg \bigwedge_{u, t \in \Sigma^n: s+t=u} M(u).$$

Then:

1. For all $u \in \Sigma^n$ with w wildcards, $M(u) = 1$ if and only if $u \in L_w$ computed in [Algorithm 1](#) if and only if $I(u) = 1$ at the iteration w before the second inner loop in [Algorithm 2](#).
2. For all $u \in \Sigma^n$ with w wildcards, $O(u) = 1$ if and only if $u \in L'_w$ computed in [Algorithm 1](#) if and only if $I(u) = 1$ at the iteration w after the second inner loop in [Algorithm 2](#).

Proof. Follows from the dataflow in the algorithms. □

We are now ready to explain the idea of multidimensional transforms, which essentially consists in switching the traversal order of the minterms from monotonic in the number of wildcards to traversal of each coordinate (“dimension”) at a time. In other words, the same operations

$$\begin{aligned} (S(s), S(t), S(v)) &\leftarrow \text{Merge}(S(s), S(t), S(v)), \\ (S(s), S(t), S(v)) &\leftarrow \text{Reduce}(S(s), S(t), S(v)) \end{aligned}$$

are performed, but on a differently ordered sequence of the triples (s, t, v) . The idea is to first iterate over the position i where $s_i = 0, t_i = 1, v_i = *$. Note that all such triples are independent as any such pair of triples would differ in at least one other position. All such triples can be enumerated for example by going through the whole string space Σ^n filtered by i -th coordinate equal to 0, and obtaining the respective t, v by changing this coordinate appropriately.

The Boolean circuit utilizing the alternative traversal order is given by [Algorithm 3](#).

Algorithm 3 Boolean circuit implementation of the Quine-McCluskey algorithm for finding all prime implicants (multidimensional transforms)

Input: indicator set $I \in \{0, 1\}^{3^n}$ of the support of a Boolean function $f : \{0, 1\}^n \rightarrow \{0, 1\}$

Output: indicator set $O \in \{0, 1\}^{3^n}$ of the set $O \subseteq \Sigma^n$ of prime implicants of f

```

MergeAll :
1:  $S \leftarrow I$ 
2: for  $i \in \{1, \dots, n\}$  do
3:   for  $s \in \Sigma^n : s_i = 0$  do
4:      $t \leftarrow s$  with  $i$ -th position set to 1
5:      $u \leftarrow s$  with  $i$ -th position set to *
6:      $(S(s), S(t), S(v)) \leftarrow \text{Merge}(S(s), S(t), S(v))$ 
7:  $M \leftarrow S$ 
ReduceAll :
8:  $S \leftarrow M$ 
9: for  $i \in \{1, \dots, n\}$  do
10:  for  $s \in \Sigma^n : s_i = 0$  do
11:     $t \leftarrow s$  with  $i$ -th position set to 1
12:     $u \leftarrow s$  with  $i$ -th position set to *
13:     $(S(s), S(t), S(v)) \leftarrow \text{Reduce}(S(s), S(t), S(v))$ 
14:  $O \leftarrow S$ 

```

Proposition 2. [Algorithm 2](#) and [Algorithm 3](#) describe the same circuits (up to reordering gates).

Although the circuits produced by the two algorithms are equivalent, the new traversal order is more regular and easier to implement in software. More importantly, it allows efficient bit-slice implementation described below.

4.2 Bit-slicing the transforms

While the Boolean circuit structure may be useful for theoretic purposes, for practical purposes it favors hardware implementations, which does not make sense for this kind of a problem. Therefore,

it is crucial to optimize a software implementation of the algorithm. For this purpose, we employ the bit-slicing technique, which is (in particular) commonly used in implementations and designs of symmetric-key cryptographic primitives such as block ciphers [8].

The idea is that CPU instructions perform operations on full registers, not on single bits. The most common general register size is $\omega = 64$ bits, while there are vector extensions processing even more bits at once. Essentially, a bitwise operation such as OR, AND, NOT can be performed on ω bits in parallel by a single instruction. A straightforward implementation of a Boolean circuit in software would waste this potential. Furthermore, a Boolean circuit needs to have a very regular structure to be efficiently bit-sliceable. On the other hand, it is easy to perform *batch* executions for any Boolean circuit: the i -th bit in each register is simply associated to the i -th parallel execution. Besides loading the inputs and reconstructing the outputs (which require single bit manipulations), the algorithm itself can be executed by simply translating the circuit into a straight-line program (using arbitrary topological order). In the following, we will use both the batch method and a custom bit-slice optimization relying on circuit's regular structure.

Data structure If the working state $S \in \{0,1\}^{3^N}$ is represented by a single bit-stream in the memory.

We represent the working state $S \in \{0,1\}^{3^n}$ in a two-layer approach. In both layers, the strings over Σ are treated as base-3 numbers (with $*$ representing the digit 2). For both layers, we employ 0-based indexes, as they are more natural for algorithmic computations.

Notation 1. For a string $s \in \Sigma^n$, we associate the integer $\rho(s) = \sum_{i=1}^n 3^{i-1} \rho(s_i)$, where $\rho(0) = 0, \rho(1) = 1, \rho(*) = 2$.

The bottom layer covers h dimensions, $1 \leq h \leq n$, while the top layer covers the remaining $n-h$ dimensions. The whole working state consists of 3^{n-h} independent subsets of Σ^h , each represented by a contiguous bit-stream of length 3^h closely fitting a single CPU register or a small number of registers. All such bit-streams are always aligned to the register's least significant bit and padded to the register size. When such bit-streams are stored in memory, we will call them *blocks*. Unused bits in blocks are not employed (i.e., are wasted). Therefore, it is desirable to choose a block size equal to a small multiple of the register size close to the maximum power of 3 it fits. Note that a large multiple implies bigger overhead on bitwise shifts.

Example 3. The 32-bit blocks are well fitted by $h = 3$: 27/32 bits are used (15% of used memory is wasted). The 64-bit blocks do not add any value over 32-bit blocks, while 128-bit blocks waste about 37% of memory. The best fit on practice are 256-bit blocks: with $h = 5$ only 5% of memory is wasted, it is a small multiple of 64-bit register size provided by modern high-performance CPUs and also matches the bit-size of the AVX2 vector extension.

Notation 2. The block in a bottom layer will be denoted by $\text{Block} \in \{0,1\}^{3^h}$. We recall that it represents a subset of Σ^{3^h} by its indicator bit-string of length 3^h , while requiring memory of 2^ℓ bits, where ℓ is smallest integer such that $2^\ell \geq 3^h$.

A working state S consists of 3^{n-h} blocks of size 3^h bits. For $a \in \Sigma^{n-h}$ we let $S[a]$ denote the Block indexed by $\rho(a)$, and for $b \in \Sigma^h$ we let $S[a][b]$ denote the bit indexed by $\rho(b)$ in the block $S[a]$. It corresponds to the indicator bit for the minterm string $(a||b)$, where $||$ denotes the concatenation.

Processing the top layer The top layer is straightforward to process as it can be seen as batch Boolean circuit evaluation, described above. To process i -th dimension, $1 \leq i \leq n-h$, we consider all triples of bits at distance 3^{i-1} , non-intersecting and properly aligned. This can be done by considering groups (of blocks) of size 3^i . Inside each block at offset b , we consider first 3^{i-1} offsets c inside the block, which correspond to minterms-strings having 0 at position i . The respective strings having 1 and $*$ at position i are obtained by adding 3^{i-1} to the index. Thus, the 3 indexes

corresponding to these strings are given by

$$(i_s, i_t, i_u) \leftarrow (b + c, b + c + 3^{i-1}, b + c + 2 \cdot 3^{i-1}),$$

where b takes values $0, 3^i, 2 \cdot 3^i, \dots, 3^{n-h} - 3^i$; c takes values $0, 1, \dots, 3^{i-1} - 1$.

Processing the bottom layer In the bottom layer, the same procedure is applied to each block independently. Recall that a block is essentially a register $v \in \{0, 1\}^{3^h}$ storing the indicator vector of a subset of Σ^h having 3^h bits, and extra unused bits which we can ignore. When processing a dimension i covered by the bottom layer, we can extract the subset of relevant indicator bits by a computing AND with a fixed bitmask. For example, the 0-valued positions in the lowest dimension correspond to the mask

$$\mu_1 = (0, \dots, 0, 0, 0, 1, 0, 0, 1, \dots, 0, 0, 1)$$

(most-significant bits first), where the first group of zeroes corresponds to the padding (unused memory). The 1-valued positions can then be extracted by the same mask shifted by 1 position to the left:

$$(\mu_1 \lll 1) = (0, \dots, 0, 0, 1, 0, 0, 1, 0, \dots, 0, 1, 0),$$

and the *-valued positions are given by one more shift:

$$(\mu_1 \lll 2) = (0, \dots, 0, 1, 0, 0, 1, 0, 0, \dots, 1, 0, 0),$$

Since the extracted bits need to be aligned to allow computations, it is more convenient to shift the register before applying the mask. Then, we can apply the Merge/Reduce operations as:

$$\begin{aligned} s &\leftarrow (v \ggg 0) \wedge \mu_1, \\ t &\leftarrow (v \ggg 1) \wedge \mu_1, \\ u &\leftarrow (v \ggg 2) \wedge \mu_1, \\ (s, t, u) &\leftarrow \text{Merge}(s, t, u), \\ v &\leftarrow (s \lll 0) \vee (t \lll 1) \vee (u \lll 2). \end{aligned}$$

In this way, the operation is applied to 3^{h-1} bit triples at once. This is possible due to the regular structure of the algorithm, allowing aligning bits by simple shift operation. This approach is easily generalized to any dimension i , by computing the mask μ_i via summing bits at indexes given by i_s from (4), and replacing the shift amount by 3^{i-1} .

Notation 3. The i -th mask $\mu_i \in \{0, 1\}^{3^h}$ is defined as

$$\mu_i = \sum_{b=0}^{3^h - i \cdot 3^{i-1}} \sum_{c=0}^{3^{i-1}-1} e_{3^i b + c},$$

where e_j denotes the unit vector with j -th rightmost bit equal to 1.

Note that, compared to the batch processing of the top layer, which almost fully exploits the available register (excluding only the unused ‘padding’ bits), an operation in the bottom layer is performed on approximately only 1/3 of the register’s bits in parallel. In addition, the shift and mask operations slow down the process even further. Yet, all this operations are performed locally on a single block, where as the top-layer processing requires more expensive memory accesses.

The full algorithm processing the two layers is given in [Algorithm 4](#). The bit-slicing optimization can be summarized as full exploitation of a machine’s word size, and generalized to the following theoretical result.

Theorem 2. In the RAM computational model with ω -bit words, all prime implicants of an explicitly given Boolean set over n -bits can be computed in $\mathcal{O}(3^n n / \omega)$ bitwise word operations and memory accesses over a storage of $\mathcal{O}(3^n / \omega)$ words.

Algorithm 4 Bit-slice implementation of the algorithm for finding all prime implicants

Input: the support $I \subseteq \{0, 1\}^n$ of a Boolean function $f : \{0, 1\}^n \rightarrow \{0, 1\}$

Output: set $O \subseteq \Sigma^n$ of prime implicants of f (in the two-layer data structure)

Load input:

- 1: $S \leftarrow$ Two-layer data structure
- 2: **for** $x \in I$ **do**
- 3: $(a, b) \in \Sigma^{n-h} \times \Sigma^h \leftarrow x$ as a Σ -string
- 4: $S[\rho(a)][\rho(b)] \leftarrow 1$

MergeAll :

Δ Processing top layer (batch processing)

- 5: **for** $i \in \{1, \dots, n-h\}$ **do**
- 6: **for** $b \in 3^i \cdot \{0, \dots, 3^{n-h-i} - 1\}$ **do**
- 7: **for** $c \in \{0, \dots, 3^{i-1} - 1\}$ **do**
- 8: $(i_s, i_t, i_u) \leftarrow (b+c, b+c+3^{i-1}, b+c+2 \cdot 3^{i-1})$
- 9: $(S[i_s], S[i_t], S[i_u]) \leftarrow \text{Merge}(S[i_s], S[i_t], S[i_u])$ ▷ bit-slice operation on blocks

Δ Processing bottom layer (bit-masks manipulation on single blocks)

- 10: **for** $a \in \{0, \dots, 3^{n-h} - 1\}$ **do** ▷ iterate over all blocks
- 11: **for** $i \in \{1, \dots, h\}$ **do**
- 12: $s \leftarrow (S[a] \ggg 0 \cdot 3^{i-1}) \wedge \mu_i$
- 13: $t \leftarrow (S[a] \ggg 1 \cdot 3^{i-1}) \wedge \mu_i$
- 14: $u \leftarrow (S[a] \ggg 2 \cdot 3^{i-1}) \wedge \mu_i$
- 15: $(s, t, u) \leftarrow \text{Merge}(s, t, u)$
- 16: $S[a] \leftarrow (s \lll 0 \cdot 3^{i-1}) \vee (t \lll 1 \cdot 3^{i-1}) \vee (u \lll 2 \cdot 3^{i-1})$

ReduceAll :

- 17: Same operations as in the MergeAll procedure above,
 but with Reduce instead of Merge.

Extracting output:

- 18: $O \leftarrow \emptyset$
 - 19: **for** $a \in \Sigma^{n-h}, b \in \Sigma^h$ **do**
 - 20: **if** $S[\rho(a)][\rho(b)] = 1$ **then**
 - 21: $O \leftarrow O \cup (a||b)$
-

5 SparseQMC: Optimized hash-based implementation for the sparse case

For a fair comparison, we implemented the classic Quine-McCluskey algorithm using hash-based neighbor search and various optimizations, which we describe in this section. This implementation is particularly useful in the sparse case, including the case of random functions with density of 50%.

There also exist a few other heuristic algorithms for the sparse case [20, 22], which however are not sufficiently detailed to be directly and efficiently implemented. For example, these methods often involve manipulation of formulas defining the function per each prime implicant

Bit-based representation of minterms Each minterm is represented by a $2n$ -bit value (fitting a 64-bit machine word when $n \leq 31$, assuming 2 bits are reserved for control flags). Each 2-bit chunk represents a symbol from Σ by correspondences $00_2 \mapsto 0$, $01_2 \mapsto 1$, $10_2 \mapsto *$. This allows efficient minterm manipulation and hashing by bitwise operations. A similar idea was proposed in [10, 19].

Avoiding repeated implicant generation Each implicant containing w wildcards can be constructed in w possible pairs of implicants with $w - 1$ wildcards (per each wildcard position). To avoid duplicates, it is sufficient to merge minterms on positions that don't have preceding wildcards. For example, $1**$ can be merged from $(10*, 11*)$ and from $(1*0, 1*1)$. The latter one can be skipped as there is a wildcard preceding the merging position 3. However, such skipped pairs might not be marked as redundant as they should. Therefore, an extra pass over all new implicants is required to mark all w possible merge pairs as redundant.

This optimization allows, at step w , to first collect all new implicants in a list (which is more efficient than a hash-table due to sequential memory access), and then to convert it to the hash-based set L_w (requiring 1 random memory access per element if linear probing is used, see below). Then, all w merge pairs of each new implicant has to be marked redundant (removed from L_{w-1}).

Custom hash-based set As described above, the step w of Algorithm 1 can be processed in two separate stages: generating the full set of implicants L_w , and removing redundant implicants from L_{w-1} . A hash-table with linear probing reduces the amount of random memory accesses, which is crucial for large memory structures. While generally linear probing becomes complicated when elements are removed, the two-stage process simplifies the implementation: when elements are removed, they can be simply marked in the hash-space (e.g., by using a control bit) without performing the expensive shrinking procedure requiring to rehash the whole set. At the end of the step, non-redundant (prime) implicants are collected and the set can be cleared to free the memory.

6 Experimental evaluation

We implemented the new dense algorithm and the optimized sparse variant in the C++ programming language. In addition, for the dense algorithm, we added obvious optimizations, such as unrolling the bottom layer dimension loop and merging the bottom layers of Merge and Reduce (using the fact that the two Reduce layers are independent and can be swapped). The optimized version performs about 3 times faster than the basic one (measured on $n = 22$).

Experiments were performed on a single core of an AMD EPYC 3.2 GHz CPU, run on the Ubuntu Server OS ¹ inside a QEMU ² virtual machine, with 1 TiB of RAM available. The code was compiled using the GNU g++ compiler (version 9.4.0). Time measurements and memory requirements are listed in Table 1 and illustrated in Figure 1. Measurements were done on single runs due to time constraints.

The source code is publicly available at

¹ <https://ubuntu.com/download/server>

² <https://www.qemu.org/>

References

1. Banerji, S.: Computer simulation codes for the Quine-McCluskey method of logic minimization. *CoRR abs/1404.3349* (2014) [4](#)
2. Boura, C., Coggia, D.: Efficient MILP modelings for sboxes and linear layers of SPN ciphers. *IACR Trans. Symm. Cryptol.* **2020**(3), 327–361 (2020). <https://doi.org/10.13154/tosc.v2020.i3.327-361> [2](#)
3. Brayton, R.K., Hachtel, G.D., McMullen, C.T., Sangiovanni-Vincentelli, A.L.: *Logic Minimization Algorithms for VLSI Synthesis*, The Kluwer International Series in Engineering and Computer Science, vol. 2. Springer US, Boston, MA (1984). <https://doi.org/10.1007/978-1-4613-2821-6> [2](#)
4. Duşa, A.: Consistency cubes: a fast, efficient method for exact Boolean minimization. *R J.* **10**, 357 (2018) [2](#)
5. Duşa, A., Thiem, A.: Enhancing the minimization of Boolean and multivalued output functions with eQMC. *The Journal of Mathematical Sociology* **39**, 108 – 92 (2015) [2](#), [4](#), [6](#)
6. Fridshal, R., Cobham, A., North, J.H.: An application of linear programming to the minimization of Boolean functions. In: 2013 IEEE 54th Annual Symposium on Foundations of Computer Science. pp. 3–9. IEEE Computer Society, Los Alamitos, CA, USA (oct 1961). <https://doi.org/10.1109/FOCS.1961.5> [2](#)
7. Gao, C., Yao, X., Weise, T., Li, J.: An efficient local search heuristic with row weighting for the unicost set covering problem. *Eur. J. Oper. Res.* **246**, 750–761 (2015) [2](#)
8. Grosso, V., Leurent, G., Standaert, F.X., Varici, K.: LS-designs: Bitslice encryption for efficient masked software implementations. In: Cid, C., Rechberger, C. (eds.) *FSE 2014*. LNCS, vol. 8540, pp. 18–37. Springer, Heidelberg (Mar 2015). https://doi.org/10.1007/978-3-662-46706-0_2 [9](#)
9. Jadhav, V., Buchade, A.: Modified Quine-McCluskey method. *CoRR abs/1203.2289* (2012) [4](#)
10. Jain, T.K., Kushwaha, D.S., Misra, A.K.: Optimization of the Quine-McCluskey method for the minimization of the Boolean expressions. In: *Fourth International Conference on Autonomic and Autonomous Systems (ICAS'08)*. pp. 165–168 (2008). <https://doi.org/10.1109/ICAS.2008.11> [2](#), [4](#), [12](#)
11. Kagiwal, A., Balachandran, S.: Set-cover heuristics for two-level logic minimization. 2012 25th International Conference on VLSI Design pp. 197–202 (2012) [2](#)
12. Lin, W., Ma, F., Su, Z., Zhang, Q., Li, C., Lü, Z.: Weighting-based parallel local search for optimal camera placement and unicost set covering. In: *Proceedings of the 2020 Genetic and Evolutionary Computation Conference Companion*. p. 3–4. GECCO '20, Association for Computing Machinery, New York, NY, USA (2020). <https://doi.org/10.1145/3377929.3398184> [2](#)
13. Madoš, B., Bilanová, Z., Chovancová, E., Ádám, N.: Field Programmable Gate Array hardware accelerator of prime implicants generation for single-output Boolean functions minimization. In: 2019 17th International Conference on Emerging eLearning Technologies and Applications (ICETA). pp. 493–498 (2019). <https://doi.org/10.1109/ICETA48886.2019.9040020> [6](#)
14. Madoš, B., Ádám, N., Bilanová, Z., Chovanec, M.: FPGA HW accelerator of the first step of systematic two-level minimization of single-output Boolean function. *Acta Polytechnica Hungarica* **17**, 25–43 (2020) [2](#), [6](#)
15. McCluskey, E.J.: Minimization of boolean functions. *The Bell System Technical Journal* **35**(6), 1417–1444 (1956), <https://doi.org/10.1002/j.1538-7305.1956.tb03835.x> [1](#)
16. Petrick, S.R.: On the minimisation of Boolean functions. In: *Proc. Symp. Switching Theory*. ICIP, Paris (1959) [2](#)
17. Quine, W.V.: The problem of simplifying truth functions. *The American Mathematical Monthly* **59**(8), 521–531 (1952), <https://doi.org/10.1080/00029890.1952.11988183> [1](#)
18. Quine, W.V.: A way to simplify truth functions. *The American Mathematical Monthly* **62**(9), 627–631 (1955), <https://doi.org/10.1080/00029890.1955.11988710> [1](#)
19. Scheinman, A.H.: A method for simplifying Boolean functions. *Bell System Technical Journal* **41**, 1337–1346 (1962) [6](#), [12](#)
20. Sen, M.K.: Minimization of Boolean functions of any number of variables using decimal labels. *Inf. Sci.* **30**, 37–45 (1983) [12](#)
21. Siládi, V., Povinský, M., Trajtel', L., Satymbekov, M.: Adapted parallel Quine-McCluskey algorithm using GPGPU. 2017 IEEE 14th International Scientific Conference on Informatics pp. 327–331 (2017) [4](#), [6](#)

22. Slagle, J., Chang, C.L., Lee, R.: A new algorithm for generating prime implicants. *IEEE Transactions on Computers* **C-19**(4), 304–310 (1970). <https://doi.org/10.1109/T-C.1970.222917> 12
23. Su, Z., Zhang, Q., Lü, Z., Li, C.M., Lin, W., Ma, F.: Weighting-based variable neighborhood search for optimal camera placement. In: *AAAI Conference on Artificial Intelligence* (2021) 2
24. Sun, L., Wang, W., Wang, M.: Accelerating the search of differential and linear characteristics with the SAT method. *IACR Trans. Symm. Cryptol.* **2021**(1), 269–315 (2021). <https://doi.org/10.46586/tosc.v2021.i1.269-315> 1
25. Tomaszewski, S., Celik, I.U., Antoniou, G.E.: WWW-based Boolean function minimization. *International Journal of Applied Mathematics and Computer Science* **13**, 577–583 (2003) 4
26. Wegener, I.: *The Complexity of Boolean Functions*. Wiley (1987) 6