Average and Worst Case Number of Nodes in Decision Diagrams of Symmetric Multiple-Valued Functions

Jon T. Butler, *Fellow, IEEE,* David S. Herscovici,
Tsutomu Sasao, *Fellow, IEEE,* and
Robert J. Barton III, *Member, IEEE*

Abstract—We derive the average and worst case number of nodes in decision diagrams of r-valued symmetric functions of n variables. We show that, for large n, both numbers approach $\frac{n^r}{r!}$. For binary decision diagrams ($r = 2$), we compute the distribution of the number of functions on n variables with a specified number of nodes. Subclasses of symmetric functions appear as features in this distribution. For example, voting functions are noted as having an average of $\frac{n^2}{6}$ nodes, for large n, compared to $\frac{n^2}{2}$, for general binary symmetric functions.

Index Terms—Decision diagrams, BDD, symmetric functions, multiple-valued functions, complexity, asymptotic approximation, average case.

1 Introduction

DURING the past ten years, significant effort has been devoted to the study of the (ordered) binary decision diagram (BDD). With origins that predate VLSI (Akers [1]), BDDs were set on a firm mathematical basis when Bryant [2] proved that, for each switching function, there exists a canonical BDD. A natural extension is the multiple-valued decision diagram (MDD), used to represent a multiple-valued function, $f: R^n \to R$, where $R = \{0, 1, 2, \ldots, r-1\}$.

To construct the MDD for a given function $f(x_1, x_2, \ldots, x_n)$, we use a root vertex to represent the function itself, and attach r children to represent $f(0, x_2, \ldots, x_n)$, $f(1, x_2, \ldots, x_n)$, and so on up to $f(r-1, x_2, \ldots, x_n)$. To each of these children, attach r children to represent the assignments to x_2, and continue until all variables are assigned. The leaf nodes represent the functions 0, 1, and so on up to $r-1$. Whenever some function appears more than once in the graph, we merge all instances of that function into a single node. We also delete a node that has identical children.

An important measure of MDD complexity is the number of nodes. For example, the first two MDDs in Fig. 1 are BDDs of functions on nine variables. For a complete description of this figure, the reader is referred to Section 2. Fig. 1a represents the AND function, while Fig. 1b represents the deBruijn function, so called because it corresponds to a Hamiltonian cycle in the deBruijn graph B_3 (Fredricksen [5] and Wegener [10]). A deBruijn function is a function on $n = 2^k + k - 2$ variables, represented by a string σ that contains one (and only one) copy of every possible substring of length k. Nodes in this BDD correspond to every distinct substring of σ. The number of such nodes is asymptotic to $\frac{n^2}{2}$ for large n, as is suggested by Fig. 1b in which the arrangement of nodes is approximately one half a square of size n by n. This represents the worst case.

In this paper, we enumerate nodes in BDDs of totally symmetric functions. This is an important subset of all functions, in which the function is unchanged by any permutation of variables. From now on, we will use the term "symmetric" to denote totally symmetric functions. There are at least four papers that discuss the worst case number of nodes in BDDs of symmetric functions. Bryant [2] noted, in passing, that the worst case complexity of BDDs for symmetric functions is $O(n^2)$. Two complete papers have been devoted to the calculation of the worst case number of nodes, the first of which is Ross et al. [8]. In the second, Heap [6] presented a correction to the results in [8]. Independently, Sasao [9] derived a similar expression. The last paper is an example of the growing number of papers on MDDs, e.g., Miller [7]. Within binary, we know of only one other paper devoted to the average number of nodes in a class of switching functions. Butler and Sasao [3] consider a special class of threshold functions called Fibonacci functions.

An interesting and important question is whether the typical BDD of a symmetric function has a complexity more representative of the BDD in Fig. 1a or of the BDD in Fig. 1b. We answer this question, showing that the typical BDD is more like that of Fig. 1b. Our results, however, go beyond this. We derive the average number of nodes in MDDs of r-valued functions of n variables, where n is large. Since the worst case number nodes for general MDDs has not been previously demonstrated, we do this too, showing that, in the limit as n approaches infinity, both numbers are asymptotically the same. Another contribution of this paper is the experimental results shown in Section 4. Here, we show the actual distribution of BDDs with respect to the number of nodes and the number of functions. Important subclasses of the symmetric functions, e.g., AND, OR, and other voting functions, are seen as ridges in this distribution. Finally, we compare the general symmetric function to voting functions, showing that the worst case number of nodes in BDDs of voting functions is one-half that of general binary functions, while the average number of nodes in BDDs of voting functions is one-third that of general binary symmetric functions.

2 MDDs of Symmetric Functions

A symmetric function $f(X)$ on n *binary* variables can be represented by a binary string of length $n + 1$, $b_0 b_1 \ldots b_i \ldots b_n$, where b_i is the value of $f(X)$ when i of the n variables are 1. For example, binary string representations 001, 010, 101, and 011 correspond to the AND, Exclusive OR, Exclusive NOR, and OR function, respectively, on two variables. With $f(X)$ represented as a binary $(n + 1)$-tuple, $B_f = b_0 b_1 \ldots b_i \ldots b_n$, any *substring*, $b_s b_{s+1} \ldots b_t$ of B_f represents a function that is a node in the BDD of B_f. This can be seen in Fig. 1a. The root node is labeled by 0000000001, which represents the AND function on nine variables; i.e., the function that is 1 iff all nine variables are 1. One of the two successor nodes is labeled 000000001, the AND of eight variables. Indeed, there is a succession of nodes representing the AND function from the root down to the terminal nodes.

We can use triangles to represent three-valued symmetric functions. For example,

• *J.T. Butler is with the Department of Electrical and Computer Engineering, Naval Postgraduate School, Code EC/Bu, Monterey, CA 93943-5121. E-mail: butler@cs.nps.navy.mil.*
• *D.S. Herscovici is with the Department of Mathematical Sciences, St. Mary's College, P.O. Box 3517, Moraga, CA 94575-3517. E-mail: hersco@stmarys-ca.edu.*
• *T. Sasao is with the Department of Computer Science and Electronics, Kyushu Institute of Technology, Iizuka 820, Japan. E-mail: sasao@aries65.cse.kyutech.ac.jp.*
• *R.J. Barton III is with Fraunhofer Center for Research in Computer Graphics, 321 South Main Street, Providence, RI 02906. E-mail: rjbarton@crcg.org.*

Manuscript received May 17, 1995; revised Jan. 7, 1996.
For information on obtaining reprints of this article, please send e-mail to: transcom@computer.org, and reference IEEECS Log Number C96099.

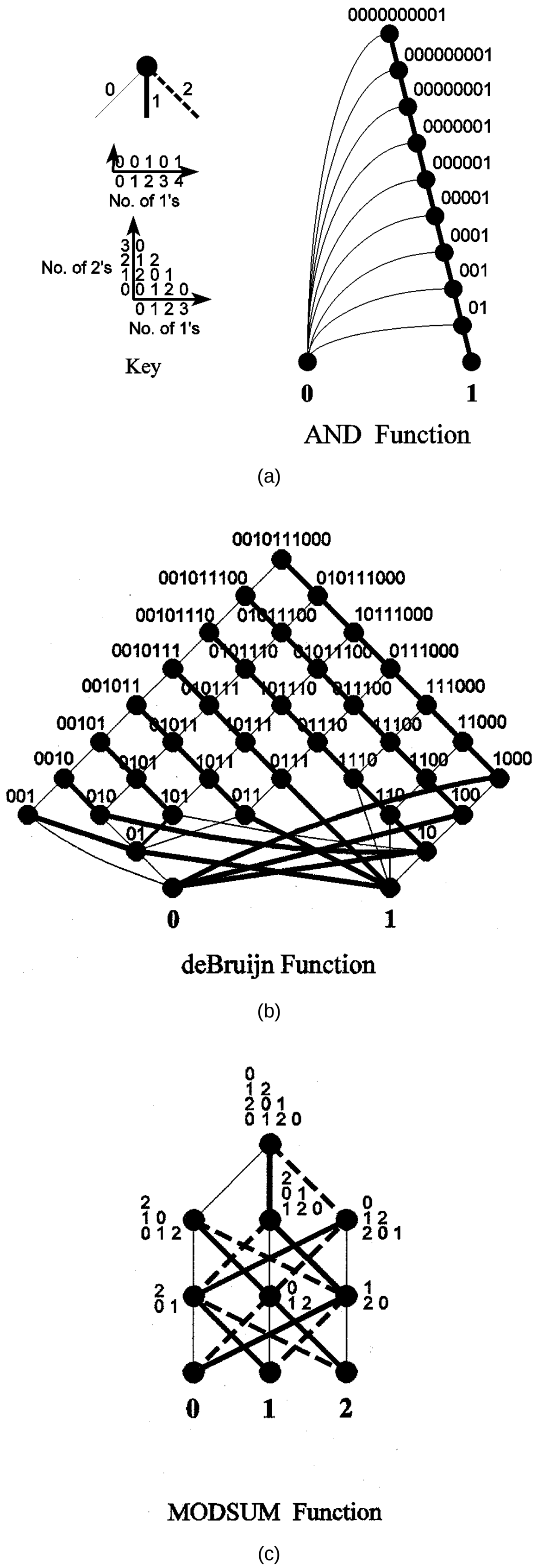
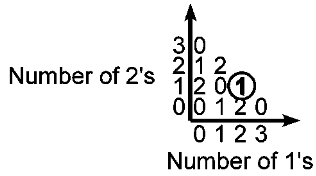


Fig. 1. Examples of decision diagrams of symmetric functions.



shows the use of a triangle to represent the MODSUM function on three three-valued variables. The circled **1** means that the function evaluates to 1 when there are zero 0s, two 1s, and one 2 among the variables. Fig. 1c shows the MDD of the MODSUM function of three three-valued variables. In the case of this example, the root node (the MODSUM function itself) is represented by the 4×4 triangle shown above, while all nodes at the next level down correspond to distinct 3×3 subtriangles, nodes in the next level down by distinct 2×2 subtriangles, and nodes in the last level by distinct 1×1 subtriangles; i.e., by the three logic values. Extending this to four values results in a symmetric function representation that is a tetrahedral.

A general representation can be obtained as follows. Let P be the set of *ordered* partitions of n into r nonnegative parts. That is, P consists of r-tuples $(n_0, n_1, ..., n_{r-1})$, where $n_i \geq 0$ represents the number of variables that take on logic value i, and $n_0 + n_1 + ... + n_{r-1} = n$. The number of such partitions is the number of ways of choosing n objects from r with repetition or $\binom{n+r-1}{r-1}$. A symmetric function f on n r-valued variables can be represented as a mapping $F_f : P \to R$, where $R = \{0, 1, ..., r-1\}$ is the set of logic values of f. The number of such mappings is $r^{\binom{n+r-1}{r-1}}$.

If η is a node in the MDD of a given symmetric multiple-valued function f, then the function f_η associated with η is a symmetric function obtained from f by assigning values to certain variables. Because f is symmetric, it makes no difference which variables are assigned, and f_η depends only on the number of variables to which each logic value is assigned. Let v_i be the number of variables assigned the logic value i, for $0 \leq i \leq r-1$. Then, $F_{f_\eta}\left(\left(n_0 - v_0\right), \left(n_1 - v_1\right), \cdots, \left(n_{r-1} - v_{r-1}\right)\right) = F_f\left(n_0, n_1, \cdots, n_{r-1}\right)$, whenever $((n_0 - v_0), (n_1 - v_1), \cdots, (n_{r-1} - v_{r-1}))$ is an ordered partition, i.e., whenever $n_i - v_i \geq 0$, for all $0 \leq i \leq r-1$.

3 Average and Worst Case Number of Nodes

We have

THEOREM 1. *Both the average and worst case number of nodes in the MDDs of r-valued symmetric functions are asymptotic to* $\frac{n^r}{r!}$ *for large n, where n is the number of variables.*

We calculate the average number of nodes in MDDs of n-variable, r-valued symmetric functions by calculating the expected number of nodes at any level. Specifically, we show that, as n approaches infinity, this number is close to the largest number possible number, for all levels except those near the bottom.

Let k index the levels of the MDD, where $k = 0$ corresponds to the root node and the given function. k represents the number of variables that have been assigned values so far in the MDD. Therefore, at level $k = 1$, there are at most r nodes corresponding to the assignment of a value to the first variable, and at level $k = n$, there are at most r nodes, corresponding to the r possible logic values in the function. In general, there are at most

$N = \binom{k+r-1}{r-1}$ nodes at level k, since there can be no more nodes than there are ways to choose k logic values from r with repetition. Each of these N nodes can represent one of $M = r^{\binom{n-k+r-1}{r-1}}$ different symmetric functions. That is, each symmetric function is specified by a mapping from a partition to the set $\{0, 1, \ldots, r-1\}$. The number of elements in the domain of this mapping is $\binom{n-k+r-1}{r-1}$, since at level k, $n - k$ variables remain to be assigned values, and this corresponds to choosing $n - k$ values from r with repetition. We can now state

LEMMA 2. *The worst case number of nodes in the MDDs of symmetric functions on n r-alued variables is asymptotic to* $\frac{n^r}{r!}$, *for large n.*

PROOF. Because there are at most $N = \binom{k+r-1}{r-1}$ nodes at level k, the total number of nodes in the MDD is at most

$$\sum_{k=0}^{n}\binom{k+r-1}{r-1}.$$

This sum has a simple form. We can choose r of $n + r$ numbered balls by identifying the largest number. That is, if we choose the ball with the largest number to be $k + r$, there are $\binom{k+r-1}{r-1}$ ways to choose the remaining balls. Summing over $0 \le k \le n$ yields the sum shown above. However, $\binom{n+r}{r}$ is also the number of ways to choose r of $n + r$ balls. Thus, the above expression is equivalent to

$$\binom{n+r}{r}.$$

But, for large n, $\binom{n+r}{r}$ is asymptotic to $\frac{n^r}{r!}$. □

The likelihood that any given pair of nodes at level k represents the same function is $\frac{1}{M}$. Since there are $\binom{N}{2}$ pairs of nodes, the expected number of matches is

$$\frac{\binom{N}{2}}{M} \le \frac{N^2}{M},$$

and the expected number of distinct nodes (functions) is at least

If we choose

$$k \le k_{max} = n + r - 1 - \left(2r!\log_r n\right)^{\frac{1}{r-1}},$$

then we have

$$M = r^{\binom{n-k+r-1}{r-1}} \ge n^{2r},$$

and

$$N = \binom{k+r-1}{r-1} \le \frac{(k+r-1)^{(r-1)}}{(r-1)!} \le \frac{(n+r-1)^{(r-1)}}{(r-1)!}.$$

When $n > r$ (so $n + r < 2n$), we have

$$\frac{N}{M} \le \frac{(n+r-1)^{(r-1)}}{(r-1)!\,n^{2r}} \le \frac{2^{(r-1)}}{(r-1)!\,n^{r+1}}.$$

Therefore, the likelihood that no two functions at level k are the same is at least

$$1 - \frac{2^{(r-1)}}{(r-1)!\,n^{r+1}},$$

which approaches 1 as n approaches infinity. Thus, at level k, the expected number of different symmetric functions is at least

$$\binom{k+r-1}{r-1}\left(1 - \frac{2^{(r-1)}}{(r-1)!\,n^{r+1}}\right).$$

Over the complete MDD of a random symmetric function on n r-valued variables, the expected number of nodes is at least

$$\left(1 - \frac{2^{(r-1)}}{(r-1)!\,n^{r+1}}\right)\sum_{k=0}^{k_{max}}\binom{k+r-1}{r-1} = \left(1 - \frac{1}{(r-1)!\,n^{r+1}}\right)\binom{k_{max}+r}{r}.$$

Here, the sum in the left term can be replaced by $\binom{k_{max}+r}{r}$ by Lemma 2. Since $\binom{n+r}{r}$ is asymptotic to $\frac{n^r}{r!}$ for large n, the average number of nodes in r-valued symmetric functions on n variables is asymptotic to $\frac{n^r}{r!}$.

4 EXPERIMENTAL RESULTS FOR BINARY DECISION DIAGRAMS

Fig. 2 shows the distribution of BDDs with respect to the number of nodes, as computed by a program that counts the exact number of BDDs (shown along the vertical axis) of n-variable symmetric functions (with n plotted along one axis in the horizontal plane) that have a given number of nodes (shown along the other axis in the horizontal plane). This is an exact distribution, not a sampled one. One horizontal axis corresponds to n, the number of variables, while the other corresponds to the number of nodes in a BDD with n variables. The vertical axis shows the number of n-variable functions with the number of nodes shown. The worst case can be clearly seen as a limit (which follows an $\frac{n^2}{2}$ curve) beyond which there are no functions. Also, the proximity of the average case to the worst case is seen as large, vertical extensions near the limit.

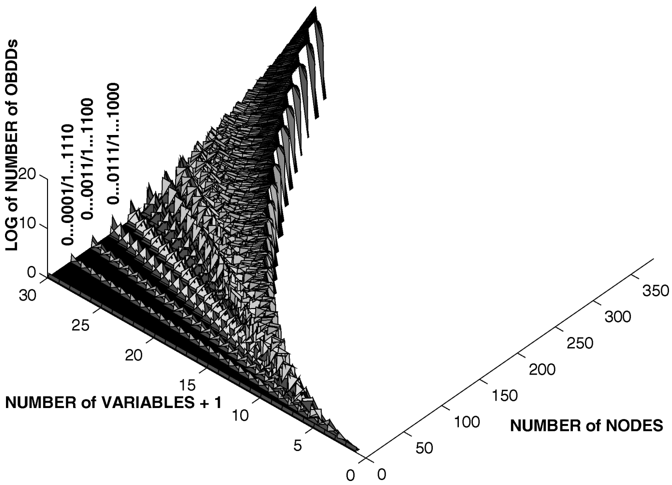


Fig. 2. Distributions of n-variable symmetric functions with respect to the number of nodes.

Note that the vertical axis plots the log of the number of cases to show detail of the function over a wide range of n. This also allows patterns to be discerned among the data. Specifically, ridges extending from the origin have various slopes and characterize certain classes of functions. These functions generate BDDs of minimal size, and the exact number of nodes is easily calculated. For example, the ridge labeled 0 … 0001/1 … 1110 in Fig. 2 corresponds to the AND function and its complement. These

TABLE 1
NUMBER OF NODES IN SAMPLE BDDS OF BINARY FUNCTIONS

Function	String representation	BDD node count
Constant 0	00 … 0	1
AND (n of n voting function)	00 … 01	$n + 2$
$n - 1$ of n voting function	00 … 011	$2n$
$n - 2$ of n voting function	00 … 0111	$3n - 4$
0^v1^w (w of n voting function)	00 … 011 …1 (v 0s, w 1s)	$vw + 2$
OR (1 of n voting function)	011 … 1	$n + 2$
Constant 1	11 … 1	1
Parity	0101 … 01	$2n + 1$

functions are represented by BDDs with $n + 2$ nodes. Other ridges in the figure correspond to symmetric functions with the string representation 0^v1^w (v 0s followed by w 1s, where $v + w = n + 1$). These are the voting functions, which are 1 if and only if at least w of n variables are 1. In the BDD of such a function, a node can be labeled by i 0s followed by j 1s, where $1 \le i \le v$ and $1 \le j \le w$, or with a 0 or a 1, for a total of $vw + 2$ nodes. Table 1 gives the size of the BDDs of voting functions and of the parity function on n variables.

It is interesting to compare our results above with the special case of binary voting functions. It can be shown that the average number of nodes in n-variable voting functions is $\frac{n^2}{6}$ (see Butler et al. [4]), for large n, or one-third that of n-variable symmetric functions. Further, the worst case number of nodes in n-variable voting functions is $\frac{n^2}{4}$, for large n, or one-half that of n-variable symmetric functions.

ACKNOWLEDGMENTS

Partial funding for this research was provided by a grant from the Tateishi Science and Engineering Foundation.

REFERENCES

[1] S.B. Akers, “Binary Decision Diagrams,” *IEEE Trans. Computers*, vol. 27, no. 6, pp. 509-516, June 1978.
[2] R.E. Bryant, “Graph-Based Algorithms for Boolean Function Manipulation,” *IEEE Trans. Computers*, vol. 35, no. 8, pp. 677-691, Aug. 1986.
[3] J.T. Butler and T. Sasao, “Average Number of Nodes in Binary Decision Diagrams of Fibonacci Functions,” *The Fibonacci Quarterly*, vol. 34.5, pp. 413-422, Nov. 1996.
[4] J.T. Butler, J.L. Nowlin, and T. Sasao, “Planarity in ROMDD’s of Multiple-Valued Symmetric Functions,” *Proc. 26th Int’l Symp. Multiple-Valued Logic*, pp. 236-241, May 1996.
[5] H. Fredricksen, “A Survey of Full Length Nonlinear Shift Register Cycle Algorithms, “ *SIAM Review*, vol. 24, no. 2, pp. 195-220, Feb. 1982.
[6] M. Heap, “On the Exact Ordered Binary Decision Diagram Size of Totally Symmetric Functions,” *J. Electronic Testing: Theory, Application*, vol. 4, no. 2, pp. 191-195, May 1993.
[7] D.M. Miller, “Multiple-Valued Logic Design Tools,” *Proc. 23rd Int’l Symp. Multiple-Valued Logic*, pp. 2-11, May 1993.
[8] D.E. Ross, K.M. Butler, and M.R. Mercer, “Exact Ordered Binary Decision Diagram Size When Representing Classes of Symmetric Functions,” *J. Electronic Testing: Theory, Application*, vol. 2, pp. 243-259, 1991.
[9] T. Sasao, “Optimization of Multiple-Valued AND-EXOR Expressions Using Multiple-Place Decision Diagrams,” *Proc. 22nd Int’l Symp. Multiple-Valued Logic*, pp. 451-458, May 1992.
[10] I. Wegener, “Optimal Decision Tree and One-Time-Only Branching Programs for Symmetric Boolean Functions,” *Information and Control*, pp. 129-143, Aug./Sept. 1984.