A Markov Chain-Based Yield Formula for VLSI Fault-Tolerant Chips

Bruno Ciciani and Giuseppe Iazeolla

Abstract—**A new yield calculation method for the yield formula of fault-tolerant VLSI chips is introduced, which improves existing methods and puts together generalities, ease of computation, and predictability in approximation levels.**

The innovative part of the method is concerned with the evaluation of the probability that the chip is acceptable given n defects. This is done by introduction of a Markov chain model, in which each state represents an operating chip configuration and the state transitions take place by the presence of manufacturing defects.

Comparisons are made with three existing yield calculation methods to prove the "ease of use," the "accuracy," and the "representativeness" characteristics of the new proposed one.

I. Introduction

INCREASED chip functionality and performance can be achieved by reducing the device dimensions and by increasing the die size. However, as the number of devices increases, the chance of having device failures also increases. The major problems that must be solved are then the yield and the reliability of large chips. The yield is the probability of chip acceptability at manufacturing time. The most promising approach for maximizing the yield is by using on-chip redundancies and restructuring techniques [1], [2].

The yield evaluation formula essentially consists of two terms [3]–[5]: the random defect (fault) statistics term and the term providing the probability of chip acceptability given n defects.

In this paper (as suggested in [3] and followed in [5]–[8]) the generalized negative binomial statistics defect model is taken as the first term of the formula. This model, taking into account the clustering phenomena, has been proven to be one of the statistics that best fits experimental data [9].

As far as the formula's second term is concerned, in the literature there is a general lack of yield formulas featuring both accuracy and ease of use. This is because yield formula accuracy requires a knowledge of the practically infinite series of all possible chip fault patterns [3]. On the other hand, ease of use has been achieved only by introducing gross simplifying assumptions which reduce the formula's representativeness. In particular are the following.

i) Some yield formulas do not take into consideration that there are cases in which replacement of defective components with redundant ones could not take place because of the "connectiveness constraints" imposed by logic circuitry. An example of such is found in [6] where it is assumed that, given K faulty components out of N, the system could still be reconfigured using the remaining $N - K$ components, without information about their effective connections.

ii) Some other yield formulas do not allow for a comparison of different reconfiguration strategies, because they fail to consider the "functional relationships" existing between basic and redundant circuits. Examples of such may be found in [1], [4], [5], [8], [10], and [11].

The aim of this paper is, therefore, to propose a yield formula which features ease of use, though without sacrificing representativeness and accuracy. As in the majority of the cited papers we do not consider the "gross" defect presence. Ease of use and representativeness are obtained by a double aggregation technique of the chip state space which replaces the infinite space of fault patterns (used in [3]) with a finite state space of an easily defined Markov chain. Accuracy is preserved by control of the approximation bounds.

This paper takes the yield formula for memory chip yield evaluation introduced in [3] (which is considered to be one of the most realistic and complete yield evaluation methods) as the "baseline yield formula" and starting from this, introduces the "new yield formula." The initial definitions are thus borrowed from the memory chip terminology used in [3]. However, as application examples will prove, the method extends to general fault tolerant VLSI chips.

The paper is organized in the following manner. In Section II the "baseline yield formula" is discussed. In Section III the new yield formula is introduced. In Section IV the physics for the formula is analyzed and the calculation algorithm introduced. In Section V a threefold evaluation of the achieved "ease of use," "accuracy," and "representativeness" features is introduced by comparing results with three existing different approaches. In order to discuss the accuracy feature in particular, a comparison is made with the results obtained in [3] and in [6], while the representativeness and ease of use features are illustrated by a comparison with the method proposed in [8].

II. The Baseline Yield Formula

As shown in [3] and [9], adequate evaluation of chip manufacturing yield requires knowledge of the following.

1) The "fault types" $(1, 2, \cdots, F)$ originating from manufacturing defects. In Table I the fault types of a conventional memory chip are reported. Quantity F denotes the number of such types. The average number of type "i" faults per chip will be denoted by λ_i $(i = 1, 2, \cdots, F)$. Evaluation of λ_i may be done by conventional methods as in [3], [9] or [12], [13].

2) The "circuit types" $(1, 2, \cdots, L)$ are subject to faults. In Table II the circuit types of a conventional memory chip are shown. Quantity L denotes the number of such types.

Manuscript received December 20, 1988; revised July 3, 1989 and January 23, 1990. This work was supported in part by the CNR Project "Materials and Devices for Solid State Electronics" and by MPI National Research Grants. This paper was recommended by Editor M. R. Lightner.

The authors are with the Electronics Engineering Department, University of Rome II, 00173 Roma, Italy.

IEEE Log Number 9040260.

0278-0070/91/0200-0252$01.00 © 1991 IEEE

TABLE I
FAULT TYPE EXAMPLE

	1. Single cell	(SC)
	2. Double cell on a word line	(DCW)
	3. Double cell on a bit line	(DCB)
	4. Single word line	(SWL)
	5. Double word line	(DWL)
	6. Single bit line	(SBL)
	7. Double bit line	(DBL)
$F =$	8. Chip kill	(CK)

TABLE II
CIRCUIT TYPE EXAMPLE

	1. Array Cell	(AC)
	2. Word Line	(WL)
	3. Bit Line	(BL)
$L =$	4. Support Circuit	(SC)

Fault and circuit types will be referred to a $M \times M \times 1$ b fault-tolerant VLSI dynamic RAM, i.e., with M bitlines (BL), and M wordlines (WL) of 1 b. It is also assumed there are r spare BL's and r spare WL's. Quantity r is also called the chip "redundancy level." We assume that replacement of defective components (e.g., WL's or BL's) can be performed, for example, by fuse firing techniques such as those used in [3].

For a chip with F possible fault types, the baseline formula [3] bases yield calculations on the concept of fault pattern (FP), defined by the vector:

$$\mathrm{FP} = (i_1, i_2, \cdots, i_F)$$

where i_k is the number of type-k faults ($k = 1, \cdots, F$).

Based on this, the chip manufacturing yield is defined by:

$$Y = \sum_{\text{all fixable FP}_i} \text{prob}(\mathrm{FP}_i) \tag{1}$$

where FP_i is the chip's ith fault pattern.

In order to clarify the example, let us consider the four FP's in Table III for a chip memory with "redundancy level" $r = 1$. FP_1 is characterized by two faults of the SC-type. FP_2 is characterized by 1 fault of the SWL-type and 1 of the SBL-type, while FP_3 and FP_4 are characterized by 3 and 7 faults, respectively, all of the same type (SC). In [3], FP_1 and FP_2 are considered to be fixable, while FP_3 and FP_4 are considered not fixable (see [3, table 4]). The decision concerning FP fixability is made by the analyst on the basis of experience and available redundancies.

In [3] it is also shown that (without considering "gross" defect presence) by using generalized negative binomial statistics, it can be stated

$$\begin{aligned}\text{Prob}(i_1, i_2, \cdots, i_F) &= \frac{\Gamma(i_1 + i_2 + \cdots + i_F + \alpha)}{\Gamma(\alpha)} \\ &\cdot \frac{(1/\alpha)^{i_1 + i_2 + \cdots + i_F}}{(1 + \lambda/\alpha)^{n+\alpha}} \cdot \frac{\lambda_1^{i_1} \cdot \cdots \cdot \lambda_F^{i_F}}{i_1! \cdot \cdots \cdot i_F!}\end{aligned} \tag{2}$$

where

$\Gamma(\cdot)$ gamma function,

α "fault clustering" parameter ($0 \le \alpha \le \infty$): as shown in [14] $\alpha = 0$ denotes maximum clustering and $\alpha = \infty$ minimum clustering,

λ_i average number of faults of type i (for $i = 1, \cdots n$),

$\lambda = \lambda_1 + \lambda_2 + \cdots + \lambda_F$ total average number of faults per chip.

TABLE III
EXAMPLE FP'S

Fault Types	SC	DCW	DCB	SWL	DWL	SBL	DBL	CK
Fault pattern								
FP_1	2	0	0	0	0	0	0	0
FP_2	0	0	0	1	0	1	0	0
FP_3	3	0	0	0	0	0	0	0
FP_4	7	0	0	0	0	0	0	0

The generalized negative binomial is the statistic that best fits the experimental results [9]. As pointed out in [15] its appeal stems from the fact that it does not assume that all defects (faults) are evenly distributed throughout the wafer but rather allows defects to cluster. References [5], [6], and [8] also definitely choose this statistic to model defect distribution throughout the wafer.

If, at a given time, there are n faults on the chip, then it may be stated that

$$\sum_{k=1}^{F} i_k = n$$

since faults can be of F different types.

If one simply assumes that faults are functionally and stochastically independent, then these can be seen as "n" objects to be placed into F different sites. Therefore, for any given "n" one can have a number of

$$\binom{n + F - 1}{F - 1} \tag{3}$$

different FP's and, since $n = 0, 1, 2, \cdots, \infty$, the FP set is a countable infinite set denoted as

$$\left\{(i_1, i_2, \cdots, i_F) \Big| \sum_{K=1}^{F} i_K = n, \quad n = 0, 1, 2, \cdots, \infty\right\} \tag{4}$$

It should be noted that within this infinite set there are fixable and unfixable FP's. In order to overcome the problem of having to decide (in the formula application) between fixable and unfixable FP's, two new placement notions are introduced in Section III-3.1.

III. THE NEW YIELD FORMULA

3.1. Preliminary Definitions

In order to present the new yield formula, a pair of further FP notions must be introduced, in addition to the FP notion used above. These are the notions of "fault distribution" (FD) and chip "operating configuration" (OC).

The chip FD is an assignment of faults to the chip circuits. In Table II, for example, we have 4 circuit types. We shall assume circuits of type 1, 2, and 3 (denoted AC, WL, BL) to

be the reconfigurable ones due to the presence of spare BL's and spare WL's.

In general, one can find a one-to-many fold mapping between chip FP's and chip FD's. The chip fault pattern $FP_3 = (3, 0, \cdots, 0)$ in Table III, for example, may give rise to a larger number of chip fault distributions, some of which are the fault distributions introduced below and denoted FD_{31}, FD_{32}, FD_{33}, FD_{34} and FD_{35}, each with the following characteristics:

FD_{31} = 3 SC-faults on three AC's belonging to three different WL's and three different BL's;
FD_{32} = 2 SC-faults on the same WL and 1 SC-fault on a different WL;
FD_{33} = 3 SC-faults on the same WL;
FD_{34} = 2 SC-faults on the same BL and 1 SC-fault on a different BL;
FD_{35} = 3 SC-faults on the same BL.

On the other hand, the chip fault pattern $FP_2 = (0, 0, 0, 1, 0, 1, 0, 0)$ gives rise to only one fault distribution, denoted FD_{21} and defined as follows:

FD_{21} = 1 SWL fault on a given WL and 1 SBL on a given BL.

It is important to note that, for a given redundancy level ($r = 1$ in the example) some of the above fault distributions give rise to reconfigurable situations (FD_{32}, FD_{33}, FD_{34}, FD_{35} and FD_{21}) while some others do not (FD_{31}). From this, the necessity of introducing the second fault placement notion, that of the chip OC (or "chip state") arises. On the basis of the chip restructuring policy, FD's can be mapped by a many-to-one fold relationship into chip configurations. Chip OC is the chip state expressed in terms of the number of existing error-free reconfigurable circuit types.

Denoting the chip state with $M + s$ error free WL's and $M + t$ error-free BL's $(s, t \leq r)$ as (s, t), Table IV shows the relation between chip FD's and chip OC's or states.

In other words, state (r, r) is the chip's completely error-free state, and state $(0, 0)$ is the state the chip reaches after exhaustion of all spare components. State-to-state transitions are governed by the chip "reconfiguration strategy." For any given reconfiguration strategy a transition diagram can be obtained. Fig. 1 is the transition diagram for the reconfiguration strategy which reorganizes the residual fault-free cells in a square. Circles denote chip OC's. So called chip "unacceptable" OC's are aggregated into a single "absorbing state." This is a faulty state in which the chip has to be discarded. It is reached either when a fault hits some unreconfigurable circuit (e.g., timing and control circuit or data and address buffers), or when the number of operating WL's or BL's becomes less than M. The state-to-state transitions are governed by the following events (assuming that the system is in state (s, t) (i.e., there are $(M + s) \times (M + t)$ active cells)).

i) A SC type of fault hits a single AC out of $(M + s) \times (M + t)$ active ones. Then, if $s \geq t$ the BL on which the faulty cell resides is replaced by a spare line. Otherwise the corresponding WL is replaced.

ii) A double-cell type of fault hits a pair of AC's out of $(M + s) \times (M + t)$ active ones. These cells reside on a common word (bit) line, but on a pair of different bit (word) lines. Then, if $s \geq 1$ $(t \geq 1)$, the interested word (bit) line is replaced by a spare one. Otherwise, if $s = 0$ and $t \geq 2$ $(t = 0, s \geq 2)$, the pair of interested bit (word) lines is replaced by spare ones.

TABLE IV
ILLUSTRATION OF SOME CHIP CONFIGURATIONS

Chip FD	No. of Error-Free (WL's) × (BL's)	Chip OC or State
none initially	$(M + r) \times (M + r)$	$(r, r) = OC_m$
FD_{31}	$(M + r - 2) \times (M + r - 1)$	$((r - 2), (r - 1))$
FD_{32}	$(M + r - 1) \times (M + r - 1)$	$((r - 1), (r - 1))$
FD_{33}	$(M + r - 1) \times (M + r)$	$((r - 1), r)$
FD_{34}	$(M + r - 1) \times (M + r - 1)$	$((r - 1), (r - 1))$
FD_{35}	$(M + r) \times (M + r - 1)$	$(r, (r - 1))$
FD_{21}	$(M + r - 1) \times (M + r - 1)$	$((r - 1), (r - 1))$
in general	$(M + s) \times (M + t)$	(s, t)
after exhaustion of spares	$M \times M$	$(0, 0) = OC_1$

iii) A SBL (SWL) type of fault hits a single bit (word) line out of $M + t$ $(M + s)$ active ones. Then the faulty line is replaced by a spare line.

iv) A DBL (DWL) type of fault hits a pair of bit (word) lines out of $M + t$ $(M + s)$ active ones. Then the faulty pair is replaced by the two spare lines.

v) A DBL (DWL) type of fault hits a single bit (word) line out of $M + t$ $(M + s)$ active ones. Then the active faulty one is replaced by a spare line.

The set of chip states can be denoted by $S = \{OC_0, OC_1, \cdots, OC_m\}$. State OC_0 denotes the absorbing state, i.e., the set of chip's unacceptable OC's. State OC_m denotes the fault-free configuration, and states OC_j $(j \neq 0, m)$ are the intermediate acceptable configurations.

Set S can be interpreted as the state space of a Markov chain (see Fig. 1, for example) with an $m \times m$ transition matrix $\boldsymbol{T} = \|t(k, j)\|$. The transition probability $t(k, j)$ gives the probability that departing from state OC_k, the chip has to be reconfigured to state OC_j due to the occurrence of an additional fault.

Based on the FD and OC definitions, the new yield formula can now be introduced. As we shall see, the new formula embeds the infinite series of all possible FP's into the paths of a finite Markov chain transition diagram.

3.2. *Mathematics of the New Yield Formula*

Here it shall be demonstrated that, through simple arguments formula (1) can be changed into one which is simpler, and which works on just a finite state space only.

According to the notation presented in Section III-3.1, denoting the chip's ith fault pattern by FP_i, (1) can be rewritten as follows:

$$Y = \sum_{\text{all fixable FP}_i} \text{prob}(\text{FP}_i) = \sum_{\forall i} \text{prob}(\text{FP}_i)\,\text{prob}(\text{FP}_i \text{ is fixable}).$$

Since OC_j is the chip's jth operating configuration, through simple conditional probability arguments, it can also be stated that

$$\begin{aligned} Y &= \sum_{\forall i} \text{prob}(\text{FP}_i) \sum_{\forall j} \text{prob}\big[(\text{FP}_i \text{ leads to } \text{OC}_j) \\ &\qquad \cap\ (\text{OC}_j \text{ is acceptable})\big] \\ &= \sum_{\forall i,j} \text{prob}\big[\text{FP}_i \cap (\text{FP}_i \text{ leads to } \text{OC}_j) \\ &\qquad \cap\ (\text{OC}_j \text{ is acceptable})\big] \\ &= \sum_{\text{all acceptable OC}_j} \sum_{\forall i} \text{prob}\big[\text{FP}_i \cap (\text{FP}_i \text{ leads to } \text{OC}_j)\big] \end{aligned}$$

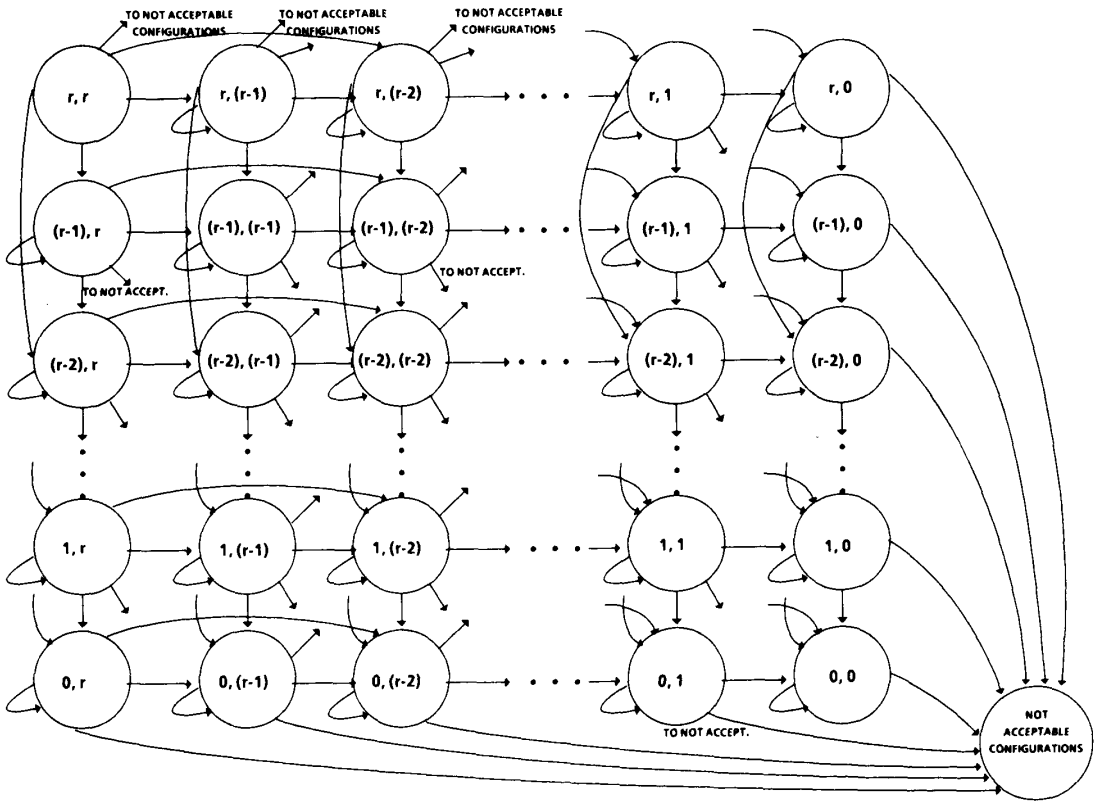


Fig. 1. Markov chain state transition diagram for the $(M + r) \times (M + r) \times$ 1-b dynamic RAM.

and by conditioning on the total number of n faults:

$$Y = \sum_{n=0}^{\infty} \sum_{\text{all acceptable OC}_j} \sum_{\forall i} \text{prob}\left[\text{FP}_i \cap (\text{FP}_i \text{ leads to OC}_j) \middle| n\right] \cdot p(n) \quad (5)$$

where $p(n)$ is the probability that n faults are found on the chip at the end of the manufacturing process, which can be expressed [9] by the generalized negative binomial statistics

$$p(n) = \frac{\Gamma(n + \alpha)}{n!\Gamma(\alpha)} \frac{(\lambda/\alpha)^n}{(1 + \lambda/\alpha)^{n+\alpha}}. \quad (6)$$

At this point it is opportune to review the subexpression

$$\sum_{\forall i} \text{prob}\left[\text{FP}_i \cap (\text{FP}_i \text{ leads to OC}_j) \middle| n\right] = \text{prob}(\text{OC}_j | n) \quad (7)$$

that we define as $\text{prob}(\text{OC}_j | n)$.

This gives the aggregated probability that, if n faults hit the chip, then any one of the FP_i patterns may take place and lead to the specific chip state OC_j. We call this the first level aggregation.

Using (7) we now turn to a further subexpression derived from (5), that is

$$\sum_{\text{all acceptable OC}_j} \text{prob}(\text{OC}_j | n) = \text{prob}(C | n) \quad (8)$$

which we denote as $\text{prob}(C | n)$, and which gives the aggregated probability that the chip is acceptable given n faults. We call this second level aggregation.

Using (7) and (8), the yield formula (5) can be rewritten as follows:

$$Y = \sum_{n=0}^{\infty} p(n) * \text{prob}(C | n). \quad (9)$$

As already stated in Section I, this formula consists of two terms, and we are concerned with the $\text{prob}(C | n)$ term. This term is the result of two probability aggregations of the chip state space: the first level aggregation aggregates all possible fault patterns and the second level aggregation all possible acceptable operating configurations.

The resulting formula (9) is much simpler than the conventional one (1), since the infinite space of the fault patterns has been aggregated and replaced by the finite state space of the operating configurations (e.g., see Fig. 1). No information has been lost, since the infinite set of fault patterns is now given by the infinite number of paths one can randomly walk by state-to-state transitions in the finite state space of the Markov chain.

Based on this chain, the evaluation of term $\text{prob}(C | n)$ in (9) is a matter of elementary stochastic process algebra. In fact, it can first be stated that

$$\text{prob}(C | n) = \text{prob}(\text{the chain transits by } n \text{ steps from } \text{OC}_m \text{ to } \text{OC}_{j \neq 0})$$

and then [16]:

$$\text{prob}(C|n) = \sum_{j=1}^{m} t^n(m, j) \tag{10}$$

where $t^n(m, j)$ is the (m, j)th entry of the n-step transition matrix $\boldsymbol{T}^n$.

By using (6) and (10), then the evaluation of the new yield formula (9) is quite straightforward. Starting from different assumptions, formulas similar to (9) have been obtained by other authors [4], [5], [10], [15]. These formulas will be discussed in Section IV.

3.3. Formula Approximation Bounds

The truncation of the infinite sum in (9) can be affected in a controlled manner. For example let us assume that (9) is truncated at $n = n^*$. The resulting error ϵ would then be

$$\epsilon = \sum_{n=n^*+1}^{\infty} p(n) * \text{prob}(C|n)$$

and, since $\text{prob}(C|n) < 1$, it can then be stated that

$$\epsilon < \sum_{n=n^*+1}^{\infty} p(n) = 1 - \sum_{n=0}^{n^*} p(n). \tag{11}$$

For example, by using (7) with $\alpha = 1$, (11) yields

$$\epsilon < 1 - \sum_{n=0}^{n^*} \frac{\lambda^n}{(1+\lambda)^{n+1}} = \left(\lambda/(1+\lambda)\right)^{n^*} \tag{12}$$

and for $\lambda = 3$, truncating the sum in (9) for $n^* = 11$ gives rise to an error of only $\epsilon < 0.01$. In other words, the computation can be stopped after the first few powers of matrix $\boldsymbol{T}$ have been derived. Computationally speaking, it is worth noting that only the product of mth row of $\boldsymbol{T}$ by $\boldsymbol{T}$ itself has to be performed each time.

IV. Yield Formula Calculation Algorithm and Physics

The calculation algorithm for the yield calculation, according to (9), can be schematized as follows.

1) Selection of the defect statistics model, $p(n)$, e.g., generalized negative binomial [3], Poisson [9].

2) Definition of the Markov chain state-diagram for the chip under consideration (e.g., Fig. 1, or Figs. 2 and 3 in the following).

3) Derivation of the state-to-state transition probabilities, $t(m, j)$. These are functions of the redundancy level and the reconfiguration strategy (e.g., see Section V-5.1, or V-5.2 in the following).

4) Selection of the approximation level ϵ and derivation, by (11), of the truncation level n^*.

5) Computation of the yield by (9) truncating its sum at n^*.

As far as the physics of the formula-term $\text{prob}(C|n)$ is concerned, it is easy to see that the Markov chain approach relates chip operating configurations with fault distributions. In other words, each Markov state (excluding the absorbing one) represents a possible acceptable operating configuration. The Markov chain walks connecting the fault-free operating configuration (state OC_m, i.e., the (r, r) state in Fig. 1) to any acceptable operating configuration (state OC_j, with $j \neq 0$), physically generate the set of fixable FP's. The walks connecting OC_m to any unacceptable configuration (aggregated into the state OC_0), physically generate the set of unfixable chip fault patterns.

Formulas like that designed in (9) have already been introduced in the literature, though starting from a different assumption. These formulas are based on the Bose–Einstein or Maxwell–Boltzmann statistics to obtain the $\text{prob}(C|n)$ term. However, in their on-paper form ([4], [5], [10], [15])) these formulas can only efficiently handle the so called M-out-of-N yield problem (that is given H faulty components out of N, the system could always be reconfigured using the remaining $M = N - H$ components without information about their effective connections). For topologically limited chip architecture problems, they have to be altered in an unspecified way.

An example is given by the formula presented in [4] and in [10], based on Bose–Einstein statistics for the scopes of the $\text{prob}(C|n)$. The formula evaluates the yield of a two-dimensional arrangement of N modules, of which $N - R$ are required to be functional, with intermodule connections running vertically and horizontally between the modules. The relation between the reconfiguration strategy, the faulty modules and faulty connections is taken into account (see [10, formula 1]) by the "bypass coverage factor," defined as the conditional probability that a module will be bypassed, given that it is faulty. However, unfortunately the authors do not explicate how this factor can be expressed in terms of architecture complexity, interconnection density, amount of redundancies, and switching mechanisms. However, it should be noted that this formula gives rise to some critical comments, found in [17].

Similarly, in [5] and [15] the yield evaluation of redundant chips is studied, in which s defective elements can be tolerated. The generalized binomial defect statistics and the Maxwell–Boltzmann (or alternatively the Bose–Einstein) statistics are used for the scopes of the $p(n)$ and the $\text{prob}(C|n)$ terms, respectively. Yet neither the authors deal with reconfiguration strategy nor with the relation between different types of defects. They assert (see [15, p. 709]) only that in case of defect type interactions the formula has to be changed accordingly, without specifying in what way. In [8] a solution is given for a particular architecture, but, as demonstrated in Section V-5.2, it is too simplistic.

V. Numerical Results

In the following sections a three part evaluation of the accuracy, representativeness, and ease of use features achieved by the formula proposed here will be presented by comparing results with three different existing approaches. To demonstrate its accuracy a comparison is drawn with the results obtained in [3] and in [6]; to emphasize its representativeness and ease of use, it is contrasted with the method proposed in [8].

5.1. Formula Accuracy

In this subsection the accuracy of the proposed method is illustrated by a comparison with the exact solution provided in [6] for the M-out-of-N yield problem. We shall show that the same problem can be solved by the use of a very simple finite Markov chain with $N - M + 1$ states, and shall then compare results with the [6] exact solution. It should be noted that the closed form solution for the simple M-out-of-N yield problem proposed by Harden and Straden has been obtained starting with (1) (proposed in [3]) and is used in this paper as the baseline formula.

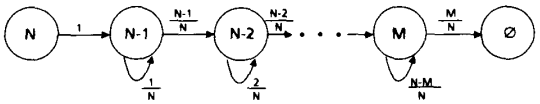


Fig. 2. M-out-of-N cells problem state diagram.

In Fig. 2 the state diagram for the M-out-of-N cells problem is given. A chip with N functional units (cells) is considered. The assumption is that, at the end of the manufacturing time, the chip is acceptable if M-out-of-N cells are fault-free.

State N is the completely fault-free state of the chip and M is the state the chip reaches after $N - M$ cells are faulty. State 0 is the absorbing state, which aggregates all unacceptable operating configurations. State-to-state transitions are governed by the events of cell faults. For example, transition from state $N - 1$ to state $N - 2$ takes place when a fault hits any one of the $N - 1$ fault-free cells, while transition from state $N - 1$ to itself takes place when the fault hits the already existing faulty cell. The first transition takes place with probability $(N - 1)/N$, and the latter with probability $1/N$.

In Table V the case $N = 5$, $\lambda_i = 0.2$ (average number of faults per cell), $\alpha = 0.2$ (clustering parameter) is presented, for varying M values. A comparison of results provided by this paper's formula (9) and those obtained using the formula proposed in [6] is given. The results are shown for three different truncation levels (n^*). The resulting error (ϵ) ranges between 10^{-2} and 10^{-5} for $M = 1$ and vanishes for increasing values of M.

In other words, the number n^* of iteration can be limited to $n^* = 5$ if a 10^{-2} error can be tolerated, while by placing n^* to 20, a 10^{-5} error is obtained even for very unfavorable values of M.

Further, to show the new yield formula accuracy feature, a comparison with the approach proposed in [3] to evaluate a redundant memory chip (with $(M + r) \times (M + r)$ cells, in which r is the redundancy level) is presented.

To compare the two approaches we need to know all fixable fault patterns. While in our approach these can be computer evaluated (they are the walks connecting the fault-free state of the chip, state (r, r) in Fig. 1, with all the acceptable operating configurations), in the cited paper, instead, they are subjectively chosen by the analyst. In particular, by examining Table V in [3] the reader can see that a number of 1316 fixable FP's, have to be chosen for the yield calculation of an $(M + r) \times (M + r)$ memory chip with $r = 5$. However, in the cited paper only the fixable FP's (chosen by the authors) are listed for a redundancy level $r = 1$, so we can only compare our results to this particular case, even if, given the redundancy level, the comparison is of academic interest only.

The necessary Markov chain for our approach is schematized in Fig. 1, while the fault parameters employed are listed in Table VI. The values represent the average number of faults per chip and reproduce the experimental data used in [3], while the defect clustering parameter is fixed at 1.

In Table VII the yield values obtained with our method are compared with those obtained by the method proposed in [3]. The higher predicability in approximation levels featured by the proposed formula may be clearly seen. The data precision which results, obtained using (9), is 10^{-6}.

To sum up, the new formula allows for the selection from among different memory dimensions on the basis of the yield parameter, while the baseline formula did not. The discrepancy arises from the fact that the new formula permits situations considered unfixable in [3] to be considered as fixable. For example, FP_3 (see [3, tables 3 and 5]) which belongs to the set of all possible FP's with $n = 3$, is subjectively considered unfixable in the cited paper, and thus it is not entered into (1). In the case of FP_3, however, based on the available redundancy level ($r = 1$), the situation of 2 (out of the 3) faults occuring in the same row or the same column (see fault distributions FD_{32}, FD_{33}, FD_{34} and FD_{35} in Table IV) is a fixable situation. It is also true that such a situation is less likely to occur in larger memories, which consequently show a lower yield value than the corresponding smaller ones in Table VII.

Observating Table VII it can also be seen that the accuracy of the conventional yield formula (1) depends on M. This proves that the subjective choice that has to be made in (1) for only a limited number of FP's can have negative and unpredictable effects on the yield estimation accuracy. These effects appear to be negligible for chips with M larger than 64. A more visible effect may be obtained by placing the redundancy level r at a value larger than 1. However, this has not been done in this paper since $r = 1$ is the only redundancy level for which it is possible to use the fixable FP's listed in [3].

TABLE V
YIELD VALUES FROM THE PROPOSED METHOD (9) AND THE METHOD IN [6]

M		$N = 5$ $\alpha = 0.2$ $\lambda_i = 0.2$ $n^* = 5$	$n^* = 10$	$n^* = 20$
1	Y based on (9)	0.95943	0.97679	0.97882
	Y based on [6]	0.97885	0.97885	0.97885
2	Y based on (9)	0.94251	0.94654	0.94663
	Y based on [6]	0.94664	0.94664	0.94664
3	Y based on (9)	0.89964	0.89985	0.89985
	Y based on [6]	0.89985	0.89985	0.89985
4	Y based on (9)	0.82859	0.82859	0.82859
	Y based on [6]	0.82859	0.82859	0.82859
5	Y based on (9)	0.69883	0.69883	0.69883
	Y based on [6]	0.69883	0.69883	0.69883

TABLE VI
MEMORY FAULT TYPES AND PARAMETERS FOR TABLE VII

Fault-Type	Average Number of Faults
Single cell	$\lambda_{SC} = 0.58$
Double cell on a word line	$\lambda_{DCW} = 0.07$
Double cell on a bit line	$\lambda_{DCB} = 0.07$
Single word line	$\lambda_{SWL} = 0.56$
Double word line	$\lambda_{DW} = 0.21$
Single bit line	$\lambda_{SBL} = 0.44$
Double bit line	$\lambda_{DBL} = 0.21$
Chip kill	$\lambda_{CK} = 0.02$

TABLE VII
YIELD VALUES FROM THE PROPOSED METHOD (9) AND THE METHOD IN [3]

Memory Dimension $(M + 1) \times (M + 1)$	$M = 4$	$M = 16$	$M = 64$	$M = 256$	$M = 1024$
Y based on (9)	0.587	0.563	0.562	0.561	0.56
Y based on (1)	0.56	0.56	0.56	0.56	0.56

5.2. *Formula Representativeness and Ease of Use*

In this subsection, the representativeness of the new yield formula and its ease of use is illustrated by proving that certain proposals (e.g., [8]) fail to consider functional relationships existing between basic and redundant circuits. Consequently these do not allow for a comparison of different reconfiguration strategies.

In [8], the case of a square grid chip with $(15 + s) \times (15 + s)$ processor elements (PE's) is considered, where s is the redundancy level. When a fault in a processor or a connecting bus (between processors) occurs, the following strategy is suggested (see [8, p. 23]). If the fault occurs within a processor, then all the processors in the corresponding row and column are declared connecting elements (CE's) and do not participate in later processing. If a connecting bus is faulty, only the processors in the appropriate row or column become connecting elements.

According to [8] the probability that the chip is acceptable is (see [8, formula (21)]):

$$\sum_{i=0}^{s} \sum_{j=0}^{2(s-i)} \Pr\{I = i, J = j\} \tag{13}$$

where $\Pr\{I = i, J = j\}$ is the probability that i PE's and j communication buses are faulty.

Formula (13) provides a very simple formula which is unable to take into consideration the necessary functional relationship between faulty and fault-free circuits. For example, according to (13) a configuration with $(15 + s + i) \times (15 + s - i - j)$ PE's would be accepted whenever j faults hit communication buses on j distinct horizontal lines. In this case, if $i + j > s$ (where for example $i = 0$ and $j > s$) also holds, then the resulting configuration could not be reconfigured as a 15 by 15 square grid. Moreover, the yield formula (13) is unable to take into account the simultaneous presence of faults on the (l, m)th PE as well as on its lth row and its mth column. More precisely, if two faults hit the chip, with one fault being a PE fault (say the (l, m)th PE) and the second a bus fault, then (13) is unable to distinguish from among the following cases:

i) the bus-fault hits the kth row, with $k \neq l$;
ii) the bus-fault hits the hth column, with $h \neq m$;
iii) the bus-fault hits either the lth row or the mth column.

As a consequence, cases i), ii), and iii) are considered as unfixable cases when $s = 1$. On the contrary, situation iii) is actually a fixable one, when $s = 1$.

With new yield formula, on the contrary, all such situations are correctly taken into account. This is possible by use of a very simple finite state Markov chain with $(s + 1) \times (s + 1) + 1$ states. Such a chain is schematized in Fig. 3. The state-to-state transitions are governed by the events of PE or connection bus faults, and relative transition probability values can be obtained simply considering that they are functions of the number and type of the components. For example, the transition from state (s, s) to state $((s - 1), (s - 1))$ takes place when a fault hits any one of the $(15 + s) \times (15 + s)$ PE's, while transition from state $((s - 1), (s - 1))$ to state $((s - 2), (s - 2))$ takes place when a fault hits anyone of the $((15 + s - 1) \times (15 + s - 1))$ fault-free PE's not belonging to the same row and column of the faulty PE that caused the transition from (s, s) to $((s - 1), (s - 1))$.

The numerical consequences of the roughness of (13) are shown in Tables VIII–X obtained for a $(15 + s) \times (15 + s)$ square grid for different values of the redundancy level s and different fault parameters. Table VIII is obtained by employing the same parameters used in [8] to obtain one of their Fig. 6

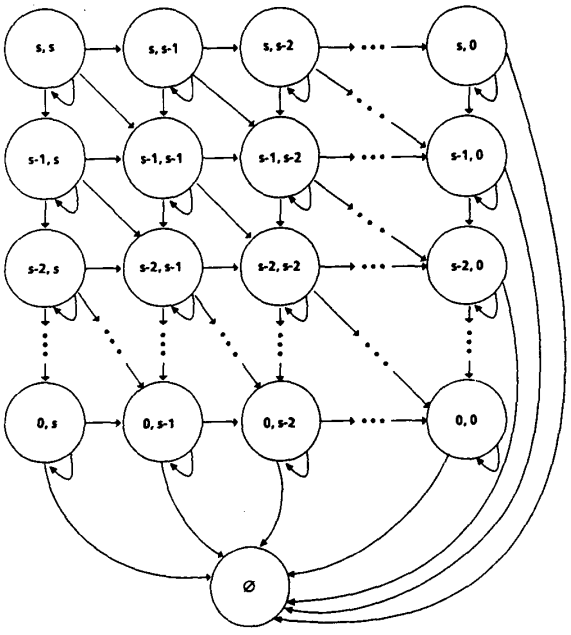


Fig. 3. Markov chain state transition diagram for the $(15 + s) \times (15 + s)$ square grid.

TABLE VIII
YIELD VALUES FROM THE PROPOSED METHOD (9) AND THE METHOD IN [8]

s	$\lambda = 3 \quad \alpha = 2 \quad \beta = 2 \quad \delta = 0.3$ Yield Evaluation Based on (13)	Yield Evaluation Based on (9)
0	0.03125	0.03125
1	0.06111	0.06063
2	0.08143	0.08197
3	0.09285	0.09550
4	0.09776	0.10289
5	0.09839	0.10591
6	0.09635	0.10595
7	0.09274	0.10406
8	0.08827	0.10096
9	0.08340	0.09713
10	0.07840	0.09293

TABLE IX
YIELD VALUES FROM THE PROPOSED METHOD (9) AND THE METHOD IN [8]

s	$\lambda = 3 \quad \alpha = 6 \quad \beta = 2 \quad \delta = 0.3$ Yield Evaluation Based on (13)	Yield Evaluation Based on (9)
0	0.00781	0.00781
1	0.02177	0.02147
2	0.03633	0.03690
3	0.04814	0.05102
4	0.05625	0.06232
5	0.06094	0.07047
6	0.06292	0.07575
7	0.06290	0.07866
8	0.06152	0.07973
9	0.05924	0.07944
10	0.05642	0.07816

TABLE X
YIELD VALUES FROM THE PROPOSED METHOD (9) AND THE METHOD IN [8]

$\lambda = 6$ $\alpha = 2$ $\beta = 2$ $\delta = 0.3$		
s	Yield Evaluation Based on (13)	Yield Evaluation Based on (9)
0	0.01020	0.01020
1	0.02116	0.02097
2	0.02950	0.02976
3	0.02950	0.03610
4	0.03761	0.04023
5	0.03860	0.04259
6	0.03835	0.04363
7	0.03730	0.04371
8	0.03577	0.04312
9	0.03396	0.04207
10	0.03202	0.04073

curves. In particular the square grid considered here is for duplicated PE's ($\beta = 2$), with a percentage of communication buses out of all chip defects $\delta = 0.3$, an average number of defects $\lambda = 3$, and a defect clustering parameter $\alpha = 2$. Table IX is obtained by changing only the defect clustering parameter to $\alpha = 6$, while Table X is obtained by changing the average number of defects to $\lambda = 6$. From the results of the tables one can see there are differences both in the absolute yield value (that increases with the redundancy level) and in the optimal value of the redundancy level. Indeed, in Table VIII the yield optimal value is obtained using (13) for a redundancy value $s = 5$, while the yield optimal value is obtained using (9) for $s = 6$; moreover, the difference between the yield values for the optimal redundancy level is about 10%. This discrepancy increases both in Tables IX and X, where the optimal redundancy level changes from 6 to 8 and from 5 to 7 with an error of about 23 and 13%, respectively.

VI. CONCLUSIONS

In this paper a new calculation method has been introduced for the yield calculation in VLSI fault-tolerant chips, which features ease of use without sacrificing representativeness and accuracy.

Ease of use and representativeness has been obtained by a double aggregation technique of the chip state space which replaces the infinite space of FP's (used in [3]) by a finite state space. In particular, the paper proves that this space can be computer-generated by randomly walking the finite state space of an easily defined Markov chain. Formula accuracy is preserved by control of the formula approximation bounds.

In order to demonstrate the powerfulness of the new yield calculation method, a comparison has been made between the proposed one and three known methods existing in literature: the [3] method for memory chip yield evaluation, the [6] method for the M-out-of-N yield problem evaluation, and the [8] method for the square grid chip yield evaluation.

ACKNOWLEDGMENT

The authors wish to thank the anonymous referees for their insightful observations which substantially improved the quality of this paper.

REFERENCES

[1] T. Mano, M. Wada, N. Ieda, and M. Tanimoto, "A redundancy circuit for a fault-tolerant 256K MOS RAM," *IEEE J. Solid-State Circuits*, vol. 17, pp. 726–731, Aug. 1982.

[2] W. R. Moore, "A review of fault-tolerant techniques for the enhancement of integrated circuit yield," *Proc. IEEE*, vol. 74, pp. 684–698, May 1986.

[3] C. H. Stapper, A. N. McLaren, and M. Dreckmann, "Yield model for productivity optimization of VLSI memory chips with redundancy and partially good product," *IBM J. Res. Develop.*, vol. 24, no. 3, pp. 398–409, May 1980.

[4] T. E. Mangir, "Sources of failures and yield improvement and restructurable interconnects for RVLSI and WSI: Part I—Sources of failures and yield improvement for VLSI," *Proc. IEEE*, vol. 72, pp. 690–708, June 1984.

[5] I. Koren and D. K. Pradhan, "Modeling the effect of redundancy on yield and performance of VLSI systems," *IEEE Trans. Comput.*, vol. C-36, pp. 344–355, Mar. 1987.

[6] J. C. Harden and N. R. Strader, "Architectural yield optimization for WSI," *IEEE Trans. Comput.*, vol. 37, pp. 88–110, Jan. 1988.

[7] N. Jarwala and D. K. Pradhan, "TRAM: A design methodology for high performance, easily testable, multimegabit RAMs," *IEEE Trans. Comput.*, vol. 37, pp. 1235–1250, Oct. 1988.

[8] I. Koren and M. A. Breuer, "On area and yield considerations for fault-tolerant VLSI processor array," *IEEE Trans. Comput.*, vol. 33, pp. 21–27, Jan. 1984.

[9] C. H. Stapper, F. M. Armstrong, and K. Saji, "Integrated circuit yield statistics," *Proc. IEEE*, vol. 71, pp. 453–469, Apr. 1983.

[10] T. E. Mangir and A. Avizienis, "Fault-tolerant design for VLSI: Effect of interconnect requirements on yield improvement of VLSI design," *IEEE Trans. Comput.*, vol. 31, pp. 609–615, July 1982.

[11] C.-L. Wey, "On yield consideration for the design of redundant programmable logic array," *IEEE Trans. Computer-Aided Design*, vol. 7, pp. 528–535, Apr. 1988.

[12] B. Ciciani, V. Grassi, and G. Iazeolla, "Yield and performability evaluation of VLSI reconfigurable multiprocessor structures," presented at 2nd ACM Int. Workshop on Applied Mathematics and Performance/Reliability Models of Computer/Communications Systems," Rome, Italy, May 1987.

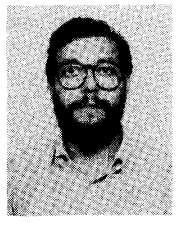
[13] B. Ciciani and G. Iazeolla, "A straightforward yield model for fault-tolerant VLSI memory chips," presented at IFIP TC-10 Conf. on Design Methodology in VLSI and Computer Architecture," Pisa, Italy, Sept. 1988.

[14] C. H. Stapper, "On yield, fault distributions, and clustering of particles," *IBM J. Res. Develop.*, vol. 30, no. 3, pp. 326–338, May 1986.

[15] I. Koren and D. K. Pradhan, "Yield and performance enhancement through redundancy in VLSI and WSI multiprocessor systems," *Proc. IEEE*, vol. 74, pp. 699–711, May 1986.

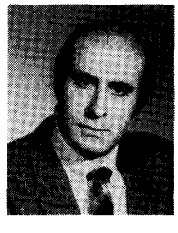
[16] E. Cinlar, *Introduction to Stochastic Processes.* Englewood Cliffs, NJ: Prentice-Hall, 1975.

[17] J. C. Harden, "Comments on 'Sources of failures and yield improvement and restructurable interconnects for RVLSI and WSI: Part I—Sources of failures and yield improvement for VLSI'," *Proc. IEEE*, vol. 74, p. 515, Mar. 1986.



Bruno Ciciani received the Ph.D. degree in electronics engineering from the University of Rome in 1980.

Since 1984, he has been a Research Associate of Computer Science at the University of Rome II. His research includes distributed computer systems, languages for parallel processing, fault-tolerant computing, and computer system performance and reliability evaluation.



Giuseppe Iazeolla received the Ph.D. degree in electronics engineering from the University of Naples in 1963.

Since 1972, he has been with the University of Pisa, and since 1984, he has been a Professor of Computer Science at the University of Rome II. His research is in computer system performance and reliability evaluation.