A New Accurate Yield Prediction Method for System-LSI Embedded Memories

Yutaka Shimada and Koichi Sakurai, *Senior Member, IEEE*

***Abstract*—The authors propose a new accurate yield prediction method for system-LSI embedded memories to improve the productivity of chips. Their new method is based on the failure-related yield prediction method in which failure bits in memory are tested to see whether they are repairable or not by using built-in redundancies. The important concept of the new method is called "repairable matrix" (RM). In RM, $rm_{ij} = 1$ means that i row redundancy sets and j column redundancy sets are needed for repair, where rm_{ij} is an element of the matrix. Here, RM can indicate all the candidate combinations of the number of row and column redundancy sets for repair. The new yield prediction method using RM solves two problems, "asymmetric repair" and "link set." These have a significant effect on accurate yield prediction but have not yet been approached by conventional analytical methods. The calculation of yield by the new method is demonstrated in two kinds of advanced memory devices that have different design rules, failure situations, and redundancy designs. The calculated results are consistent with the actual yield. On average, the difference in accuracy between the new method and conventional analytical methods is about 5%.**

***Index Terms*—Embedded memory, failure, redundancy, semiconductor manufacturing, yield prediction.**

NOMENCLATURE

DX	Deviation of "pointer bit" toward X direction.
DY	Deviation of "pointer bit" toward Y direction.
Ef	Existence probability of a failure situation.
Ef_{xy}	Existence probability of a failure situation in MUxy.
E_{RM}	Existence probability of RM.
$E_{\mathrm{RM}}[kxy]$	Existence probability of (kxy)th RM in N_{RM} variations of RM.
fdist	Existence probability function of failure size.
fsize	Existence probability function of failure distribution.
FX	Size of a failure along X axis.
FY	Size of a failure along Y axis.
MU	Minimum unit.
MU_{xy}	MU determined by x and y which are the X and Y coordinates of MU in an analysis unit, respectively.
N_{AU}	Number of analysis units in a chip.
N_{Bit}	Number of bit failures in a MU.
N_{BL}	Number of bit line failures in a MU.
NC	Number of built-in column redundancy sets.
N_{CL}	Number of cross line failures in a MU.
NR	Number of built-in row redundancy sets.
N_{RM}	Number of variations of RM.
N_{WL}	Number of word line failures in a MU.
PF	Pass (repairable) or fail (not repairable).
RM	Repairable matrix.
$rm(1)_{pq}$	Matrix element of the first original RM.
$rm(2)_{(i-p)(j-q)}$	Matrix element of the second original RM.
$rm(\mathrm{comb})_{ij}$	Matrix element of combined RM made from two original RMs.
rm_{ij}	Matrix element of RM.
WC	Bit width of one column redundancy set.
WR	Bit width of one row redundancy set.
X	Number of row redundancy sets for repair.
XMU	Number of MUs in an analysis unit along the X axis.
Y	Number of column redundancy sets for repair.
Y_{AU}	Yield of an analysis unit.
Y_{memory}	Yield of the memory part in a chip.
YMU	Number of MUs in an analysis unit along the Y axis.

Manuscript received March 31, 2003; revised May 13, 2003.

The authors are with the LSI Manufacturing Technology Unit, Renesas Technology Corporation, Hyogo 664-0005, Japan (e-mail: Shimada.Yutaka3@renesas.com; Sakurai.Koichi@renesas.com).

Digital Object Identifier 10.1109/TSM.2003.815636

I. INTRODUCTION

IN THE system-LSI with embedded memories business, a short delivery time is required. In addition, the volume of production is much smaller than that of general-purpose memory chips industry. Therefore, improving productivity, by such means as suitable plans for production and inventory adjustment, is essential to make a profit. Accurate yield prediction is crucial. Inaccurate yield prediction results in a shortage or surplus of lot inputs. For example, an inaccuracy of 10% can result in a ten-lot discrepancy.

In this paper, we focus on the yield prediction method for the memory part of system-LSI, because some multimedia applications require larger embedded memories than 4 Mb. There is continual room for improvement in yield prediction techniques, although the yield of the memory part strongly affects the yield of system-LSI chips [1].

The mainstream of analytical yield prediction methods has been the "failure-related" yield prediction method which judges whether failure bits or defects that cause failure bits in memory can be repaired by built-in row and column redundancies. So far, many researchers have suggested prediction methods for "failure situations" in memory chips of the future [2]–[9]. For example, Sakurai *et al.* successfully predicted the failure situations in future devices by converting the actual failure data of the devices in the manufacturing factories, on the basis of a line

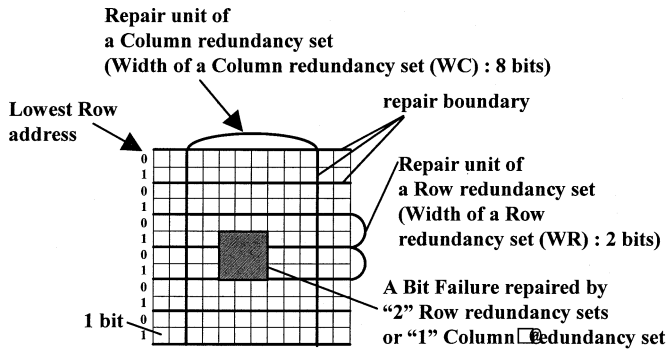


Fig. 1. Example of "asymmetric repair." Number of row and column redundancy sets for repair are different.

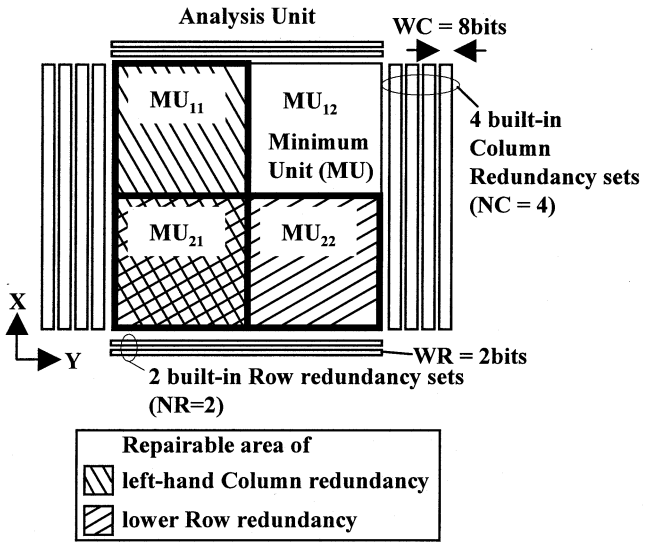


Fig. 2. Example of "link set." Repairable area of row and column redundancies are different.

and space model [7]. Failure situations are determined by three factors. These are the kind of failures, the size of the failures, and the number of failures. In the failure-related yield prediction method, it is possible to make use of the results of these studies on failure situations [2]–[9]. Therefore, the failure-related yield prediction method is superior because a high degree of accuracy for yield prediction can be expected.

However, the conventional analytical methods that have been described so far cannot treat "asymmetric repair" and "link set." These two cases often occur in embedded memories in system-LSIs.

Fig. 1 shows the asymmetric repair where the number of row redundancy sets and the number of column redundancy sets to repair a failure are different. Asymmetric repair is not negligible in system-LSI embedded memories in which the bit width of one column redundancy set is larger than that of the row redundancy set because the repair unit of a column redundancy set is an input/output (I/O) line in most cases, normally a multiple of four bits. In Fig. 1, the width of a row redundancy set is two bits and we assume that a row redundancy set can repair "01" pairs of the lowest row address but cannot repair "10" pairs. The 3×3 bit failure can be repaired by either "2" row redundancy sets or "1" column redundancy set. The number of redundancy sets is determined by the relative location of the failure to the "repair boundary" shown in Fig. 1. It is obvious that the larger the difference between the width of a row redundancy set and that of column redundancy set becomes, the more likely asymmetric repair occurs. Asymmetric repair can occur in bit and cross line (CL) failures; the former failure can be repaired by row or column redundancies and the latter failure is repaired by row and column redundancies. However, some conventional methods do not even treat bit and CL failures [10]–[12], and others that do treat these failures assume that the number of redundancy sets required to repair a failure is equal in both cases as reported by Stapper *et al.* [13]. In the case of the failure in Fig. 1, the conventional methods assume that the failure is repaired by either "1" row redundancy set or "1" column redundancy set, or by either "2" row redundancy sets or "2" column redundancy sets. Ignorance of asymmetric repair results in incorrect yield predictions as demonstrated in Section V.

Fig. 2 shows the second case, link set, where a set of redundancies can cross region boundaries and the repairable areas of row and column redundancies are partially overlapping but not equal. A link set often occurs in large-scale memory chips [16]. In Fig. 2, the repairable areas of row redundancies are two minimum units (MU) along the Y axis and the repairable areas of column redundancies are two MUs along the X axis. It is significant for accurate yield prediction because a failure situation in one MU could affect the repair probability of the neighboring MU. However, there are no conventional analytical methods that take link set into account, and this results in an incorrect yield prediction as described in Section V. Conventional methods are limited because the repairable areas of row and column redundancies are equal [10]–[15].

A Monte Carlo technique is a strong candidate to overcome these two problems. However, it is difficult to verify the relation of cause and effect. On the other hand, the results of analytical methods such as the failure-related method can be easily verified by monitoring the process of the calculation. Therefore, if there are strong alternative methods, we believe that the Monte Carlo technique is not the best solution.

The purpose of this work is to present an accurate analytical yield prediction method based on the failure-related method and to make it possible to treat both asymmetric repair and link set. In Section II of this paper, we analyze the basic failure-related method. In Section III, we propose a new method of yield prediction using the repairable matrix (RM). In Section IV, the process of calculation is described in detail. In Section V, the new method is verified by a comparison between the actual yield and the calculated results; the results calculated by conventional methods are also shown.

II. Analysis of the Basic Failure-Related Yield Prediction Method

A. Basic Failure-Related Yield Prediction Method

The basic failure-related yield prediction method is

$$Y_{\text{memory}} = (Y_{\text{AU}})^{NAU} \tag{1}$$

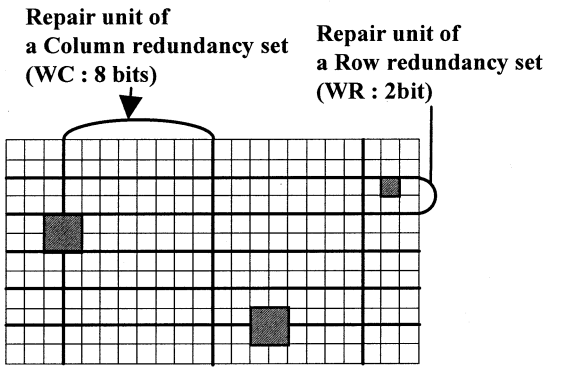


Fig. 3. Example of a failure situation composed of three bit failures. There are five different ways of repair.

where Y_{memory} is the yield of the memory part of a chip, Y_{AU} is the yield of an analysis unit, and N_{AU} is the number of analysis unit in a chip; these abbreviations are shown in the Appendix. An analysis unit is the minimum rectangular area containing the repairable areas of both row and column redundancies as shown in Fig. 2.

Y_{AU} can be shown as follows:

$$Y_{\mathrm{AU}} = \sum_{\text{all repairable failure situations}} \prod_{x=1}^{\mathrm{XMU}} \prod_{y=1}^{\mathrm{YMU}} Ef_{xy} \quad (2)$$

where XMU and YMU are the number of MUs along the X axis and Y axis, respectively (in Fig. 2, $XMU = YMU = 2$), Ef_{xy} is the existence probability of a failure situation in MU$_{xy}$ and x and y are the coordinates of MU in the analysis unit.

Equation (2) is an extension of conventional failure-related yield prediction methods [13], [15] and link set is taken into account. In (2), if a given failure situation is repairable, then the product of Ef_{xy} is added to Y_{AU}. To judge whether it is repairable or not, the number of redundancy sets needed for repair and the number of built-in redundancy sets must be compared. If the former is less than or equal to the latter, the analysis unit is repairable. Therefore, the key point in calculating (2) is counting the number of redundancy sets for repair under an arbitrary failure situation.

B. Point at Issue of Failure-Related Yield Prediction Method

However, there are two problems that make the calculation of (2) complicated. These are explained by the two examples shown in Figs. 3 and 4.

Fig. 3 shows an example of a failure situation composed of two 2×2-bit failures and one 1×1-bit failure in a MU. To repair this failure situation, there are eight candidate combinations as follows: (4,0) / (3,2) / (3,1) / (2,3) / (2,1) / (1,3) / (1,2) / (0,4), where (i, j) means i row redundancy sets and j column redundancy sets are needed for repair. If the number of built-in row redundancy sets (NR) is two and that of column redundancy sets (NC) is four as shown in Fig. 2, the number of candidates becomes five, as follows: (2,3) / (2,1) / (1,3) / (1,2) / (0,4), because the candidates in which the number of row redundancy sets exceeds the number of built-in row redundancy sets are excluded. To deal with link set in which one redundancy repair failure bits in plural MUs, all five candidates must be memorized because some of these five candidates may not be used according to the failure situation in the neighboring MU. Without a method in which plural candidates can be treated simply, it is obvious that the calculation of (2) becomes complicated.

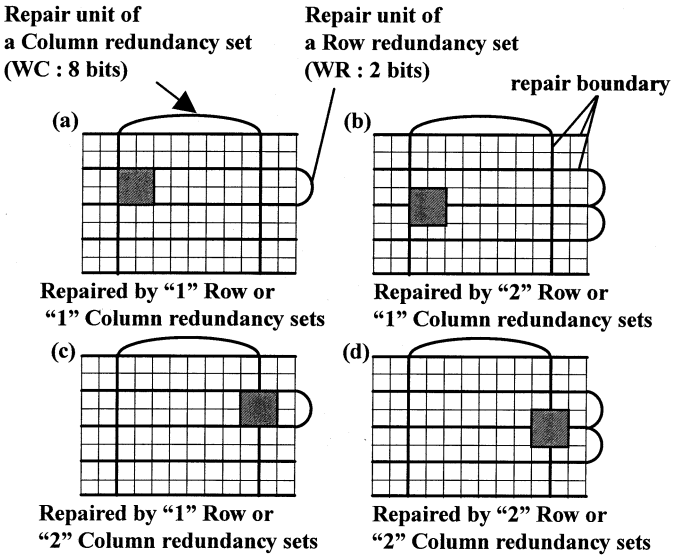


Fig. 4. Example of a 2×2-bit failure. There are four different ways of repair according to the failure's location.

Fig. 4 shows another example of the ways of repairing a failure situation composed of one 2×2-bit failure. In Fig. 4, it is notable that the bit failure can be repaired in four different ways, including the two asymmetric repairs shown in Fig. 4(b) and (c), according to its relative location to the repair boundary of the redundancies in memory. In this case, it is required to judge whether this failure situation is repairable or not at least four times. In the case of a failure situation composed of plural failures, the number of combinations of redundancy sets becomes larger. Moreover, taking link set into consideration, the number of combinations becomes much larger.

Thus, it seems to be tough work to tackle (2) squarely. Therefore, we need a calculation method in which the two conditions described below are met.

1) The several candidates for repair, as mentioned in Fig. 3, should be described in a simple way.

2) The number of combinations for calculation should be smaller.

III. New Yield Prediction Method by Repairable Matrix

To find a way of meeting the first condition described in the last section, a new concept called repairable matrix (RM) is suggested in Section II-A. In Section II-B, we propose a new yield prediction equation by RM. This equation implies two steps for calculation to meet the second condition described in the last section.

A. New Concept—RM

We propose a new criterion, the RM. RM shows the combinations of the number of row redundancy sets and the number

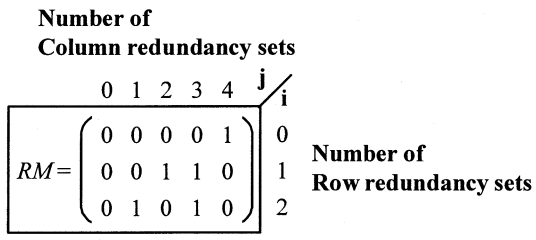


Fig. 5. RM corresponding to the failure situation shown in Fig. 3. Five ways of repair can be expressed by the five "1"s in RM.

of column redundancy sets for repair. Fig. 5 shows the RM corresponding to the failures shown in Fig. 3 which has five candidates for repair. The minimum number of rows and columns in the matrix is zero and the maximum number of rows and columns is *NR* and *NC*, respectively, where *NR* is the number of built-in row redundancy sets and *NC* is that of column redundancy sets. Fig. 5 shows a case where *NR* is two and *NC* is four as shown in Fig. 2. Here, "$rm_{ij} = 1$" indicates that the failures can be repaired by "i" row redundancy sets and "j." Column redundancy sets, where rm_{ij}, is the matrix element. All five candidates described in Fig. 3 can be successfully expressed in RM as shown in Fig. 5, and the first condition shown in the former section is satisfied. It is also obvious that an asymmetric repair can be treated successfully.

N_{RM}, the number of variations of RM, can be calculated as follows:

$$N_{\mathrm{RM}} = 2^{(NR+1)(NC+1)} - 1. \tag{3}$$

Here "2" means the number of cases at RM (0 or 1). "-1" means the exception when all elements are "0." In the case of $NR = 2$ and $NC = 4$, as shown in Fig. 2, N_{RM} becomes 32 767.

B. New Yield Prediction Method by Repairable Matrix

Equation (2) can be rewritten by a new yield prediction method with RM

$$Y_{\mathrm{AU}} = \sum_{k_{11}=1}^{N_{\mathrm{RM}}} \sum_{k_{12}=1}^{N_{\mathrm{RM}}} \cdots \sum_{k_{XMUYMU}=1}^{N_{\mathrm{RM}}} \cdot \{E_{\mathrm{RM}}[k_{11}]\ E_{\mathrm{RM}}[k_{12}] \cdots E_{\mathrm{RM}}[k_{XMUYMU}]\ PF\} \tag{4}$$

where k_{xy} is a variable whose value changes from 1 to N_{RM}, E_{RM} is the existence probability of RM, and $E_{\mathrm{RM}}\,[k_{xy}]$ means the existence probability of (k_{xy})th RM in N_{RM} variations of RM. Also, *PF* stands for pass (repairable) or fail (not repairable) and the value of *PF* becomes as follows:

$$\left.\begin{array}{ll} PF = 1 & (\text{repairable}) \\ PF = 0 & (\text{not repairable}) \end{array}\right\}. \tag{5}$$

Equation (4) means that when a failure situation in MU$_{xy}$ can be repaired by combinations of row and column redundancy sets shown by (k_{xy})th RM, the product of E_{RM} [k_{xy}] of all MU is added to Y_{AU}, if $PF = 1$.

Equation (4) implies two steps for calculation. The first step is converting "all repairable failure situations" in an MU into N_{RM} variations of RM and finding their existence probability (E_{RM}). The second step follows after the first step ends. The second step is to judge whether $PF = 1$ or $= 0$ when each MU has some RMs.

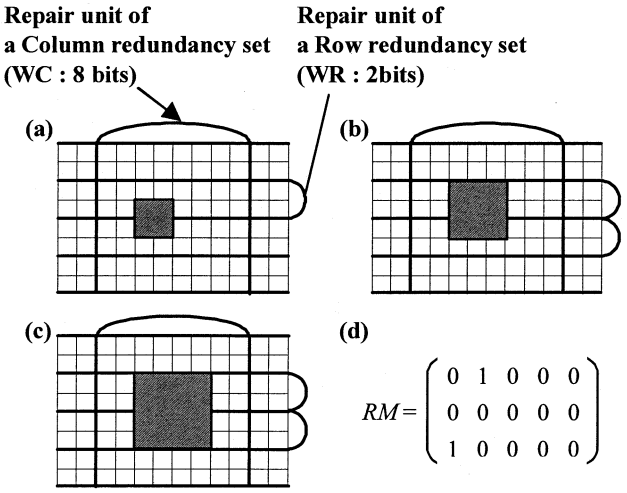


Fig. 6. Example of different failure situations categorized in the same RM. RM helps to reduce the number of combinations required for calculation.

We define all repairable failure situations by the number of failures and the size of failures. First, for the number of failures, we decide all repairable failure situations as follows:

$$\left.\begin{array}{l} N_{\mathrm{WL}} + N_{\mathrm{CL}} \le NR \\ N_{\mathrm{BL}} + N_{\mathrm{CL}} \le NC \\ N_{\mathrm{Bit}} + N_{\mathrm{WL}} + N_{\mathrm{BL}} + 2 \times N_{\mathrm{CL}} \le NR + NC \end{array}\right\} \tag{6}$$

where N_{Bit}, N_{WL}, N_{BL}, and N_{CL} are the number of Bit, WL, BL, and CL failures in the MU, respectively. When $NR = 2$ and $NC = 4$, the number of combinations of N_{Bit}, N_{WL}, N_{BL}, and N_{CL} is 90. Second, for the size of failures, we limit the maximum size of four kinds of failure as follows:

$$\left.\begin{array}{ll} \max(NR\,WR,\ NC\,WC) & : \text{Bit failure} \\ NR\,WR & : \text{WL failure} \\ NC\,WC & : \text{BL failure} \\ \min(NR\,WR,\ NC\,WC) & : \text{CL failure} \end{array}\right\} \tag{7}$$

where WR and WC are the bit width of a row redundancy set and a column redundancy set, respectively, and $\max(NR\,\mathrm{WR}, NC\,\mathrm{WC})$ and $\min(NR\,\mathrm{WR}, NC\,\mathrm{WC})$ are the larger and the smaller numbers of ($NR\,\mathrm{WR}$) and ($NC\,\mathrm{WC}$), respectively. By changing the number of failures and the size of failures according to (6) and (7), respectively, all repairable failure situations can be obtained. When $NR = 2$, $NC = 4$, $WR = 2$, and $WC = 8$ as in Fig. 2, the number of all repairable failure situations is about 6 billion.

The key point of (4) is to categorize all 6 billion repairable failure situations in an MU into N_{RM} variations of RM. Fig. 6 shows examples of three different bit failures that are categorized into one RM. In (2), these failures are treated separately. Another key point is to treat each MU independently at first, then judge whether repairable or not by putting all the MUs in an analysis unit together. If the analysis unit is treated altogether from the beginning, the number of combinations for calculation becomes 47 billion times larger than the proposed method. As

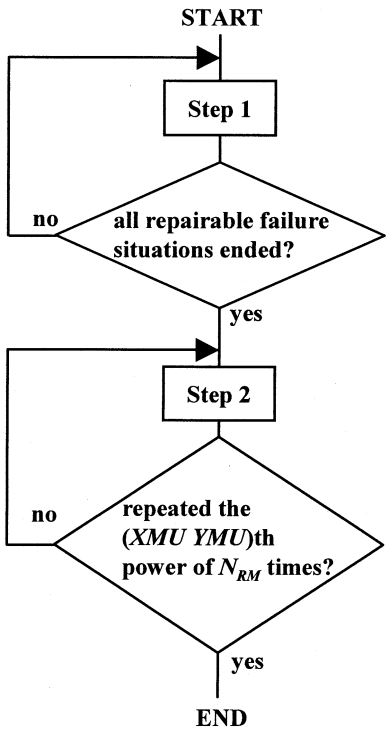


Fig. 7. Outline of calculation of Y_{AU}. Steps 1 and 2 are repeated.

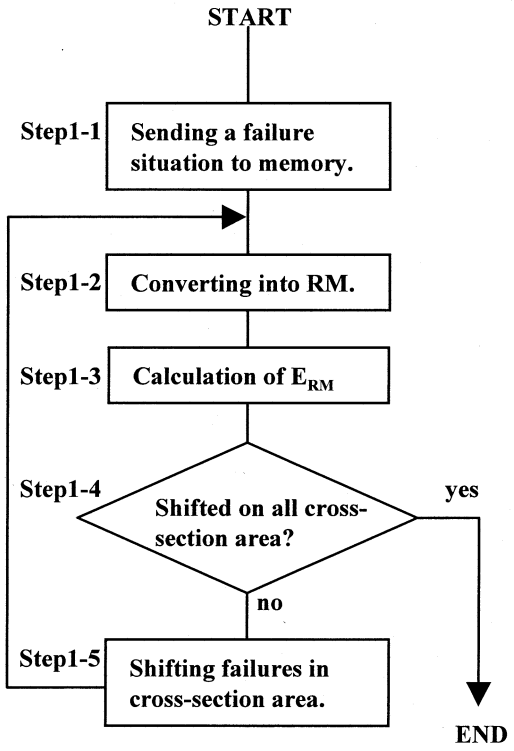


Fig. 8. Outline of Step 1. Failure situation is converted into RM.

a result, (4) reduces the number of combinations required for calculation.

IV. Calculation Method of Y_{AU}

To calculate Y_{AU} in (4), two steps are needed as described in the previous section. This section describes the method of calculating these two steps in detail. As shown in Fig. 7, the first step is repeated under all repairable failure situations indicated in (6) and (7), and the second step is repeated the (XMU YMU)th power of N_{RM} times.

In the last part of this section, we discuss how to apply this calculation method to other redundancy schemes such as flexible redundancy or block redundancy.

A. Step 1—Conversion Into Repairable Matrix

The main process of Step 1 is to convert a failure situation into RM. Fig. 8 shows an outline of Step 1 and Fig. 9 shows an example of the process of Step 1.

In Step 1-1, the failure situation is sent to memory. In Fig. 9, there is one 2 × 2-bit failure. We call the upper left corner of the failure the "pointer bit."

From Steps 1–2 to 1–5, the conversion of failures into RM and a shift of failures are repeated. The failures are shifted in the "cross-sectional area" of the repair unit of a row redundancy set and that of a column redundancy set. In Fig. 9, the cross-sectional area is composed of 16 bits because the width of a row redundancy set is 2 and that of a column redundancy set is 8. Therefore, Steps 1-2–1-5 are repeated 16 times. The 2 × 2-bit failure can take four variations of RM in 16 locations as shown in Fig. 9(b), and the E_{RM}, the existence probability of RM, is shown in the lower right corner of each RM. For example, the

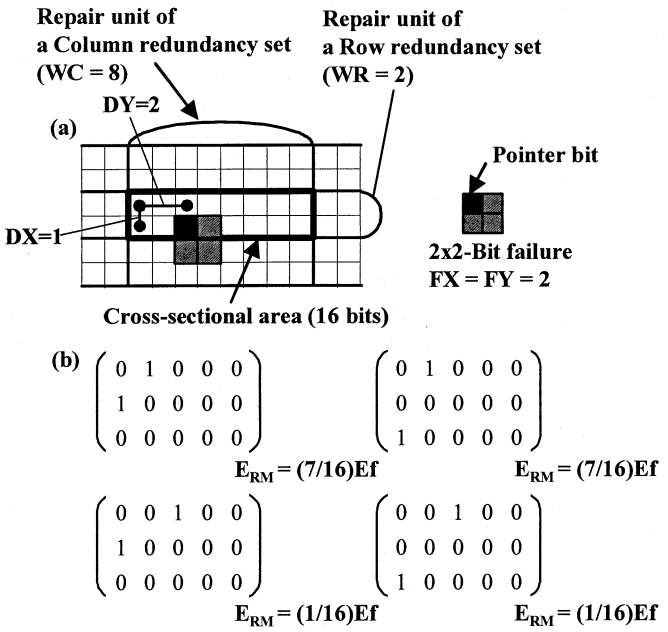


Fig. 9. Example of Step 1. 2 × 2-bit failure is converted into four RMs.

E_{RM} of the upper left RM shown in Fig. 9 can be expressed by $(7/16)Ef$, the existence probability of a failure situation, as follows:

$$\begin{aligned} E_{\mathrm{RM}} &= (7/16)Ef \\ &= (7/16)\{\mathrm{fdist}(N_{\mathrm{Bit}}=1)\ \mathrm{fsize}(2\times 2\text{−}\mathrm{Bit})\} \\ &\quad \cdot \mathrm{fdist}(N_{\mathrm{WL}}=0)\mathrm{fdist}(N_{\mathrm{BL}}=0)\mathrm{fdist}(N_{\mathrm{CL}}=0) \end{aligned} \tag{8}$$

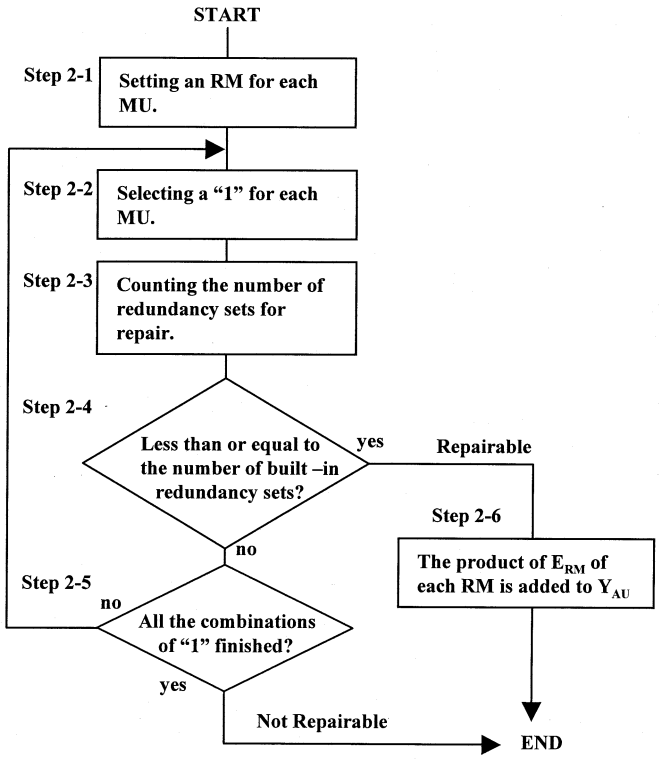


Fig. 10. Outline of Step 2, which judges whether the analysis unit is repairable or not.

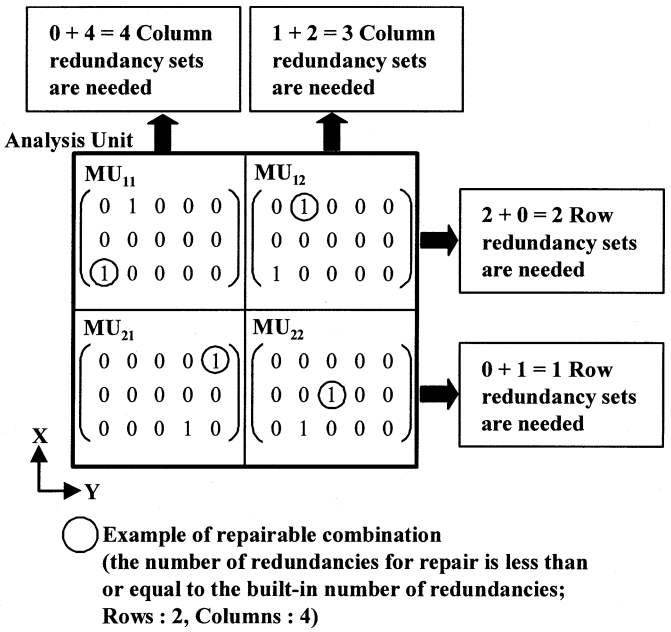


Fig. 11. Example of Step 2. If the number of redundancy sets for repair is less than or equal to that of built-in redundancy sets, the analysis unit is repairable.

where $\mathrm{fdist}(N_{\mathrm{Bit}}=1)$ is the probability of a failure distribution where there is one bit failure in the MU, and so on. Here, $\mathrm{fsize}(2\times 2\text{−Bit})$ is the probability that the bit failure's size is 2×2. These values of fdist and fsize can be obtained on the basis of the results of previous works on the prediction methods of failure situations [2]–[9].

In general, the RM of a single failure is determined by the following:

$$\left.\begin{array}{ll} rm_{X0}=1 & \text{if } X\le NR \\ rm_{0Y}=1 & \text{if } Y\le NC \end{array}\right\} \quad \text{[Bit failure]} \tag{9}$$

$$\text{[WL failure]}\quad \mathrm{RM}_{X0}=1 \quad \text{if } X\le NR \tag{10}$$

$$\text{[BL failure]}\quad \mathrm{RM}_{0Y}=1 \quad \text{if } Y\le NC \tag{11}$$

$$\text{[CL failure]}\quad \mathrm{RM}_{XY}=1 \quad \text{if } X\le NR \text{ and } Y\le NC \tag{12}$$

where X and Y in these equations are the number of row redundancy sets and column redundancy sets for repair, respectively. If X or Y does not meet the conditional equations that are shown on the right side of each equation, the failure becomes unrepairable because the number of redundancy sets for repair exceeds the number of built-in redundancy sets. Here, X and Y are calculated by following:

$$X=\mathrm{Int}\{(DX+FX)/WR\}+1 \tag{13}$$

$$Y=\mathrm{Int}\{(DY+FY)/WC\}+1 \tag{14}$$

where $\mathrm{Int}(a/b)$ is the quotient of (a/b), FX and FY are the size of the failure along X axis and Y axis, respectively, DX and DY are the deviation of the pointer bit from the upper left corner bit of the cross-sectional area, and WR and WC are the bit width of a row redundancy set and a column redundancy set, respectively, as shown in Fig. 9.

The RM for a failure situation composed of two failures can be calculated by making a combined RM from two RMs. Equation (4) shows how to obtain a combined RM from two RMs

$$\left.\begin{array}{ll} rm(\mathrm{comb})_{ij}=1 & \text{if } \sum_{p=0}^{i}\sum_{q=0}^{j}(rm(1)_{pq}rm(2)_{(i-p)(j-q)})\ge 1 \\ rm(\mathrm{comb})_{ij}=0 & \text{if } \sum_{p=0}^{i}\sum_{q=0}^{j}(rm(1)_{pq}rm(2)_{(i-p)(j-q)})=0 \end{array}\right\} \tag{15}$$

where $rm(\mathrm{comb})_{ij}$ is the element of the combined RM, and $rm(1)_{pq}$ and $rm(2)_{(i-p)(j-q)}$ are the elements of the two original RMs. For a failure situation composed of more than three failures, the combined RM can be calculated by repeating (15).

Step 1 is repeated under all repairable failure situations, and they are categorized in N_{RM} variations of RM, and the E_{RM} of each RM is calculated by adding up the existence probability of failures that have the same RM.

B. Step 2—Judgment of Pass or Fail

Step 2 judges whether the analysis unit is repairable or not. Fig. 10 shows an outline of Step 2 and Fig. 11 shows an example of the process of Step 2.

In Step 2-1, an RM is set for each MU. In Fig. 11, four RMs are set in four MUs.

In Step 2-2, a "1" in each MU is selected. As was described in the third section, the number of "1"s in an RM means the number of candidates for repair. Therefore, the product of the number of "1"s of all the RMs in an analysis unit indicates the number of candidates for repair. In Fig. 11, the number of candidates is 16 $(=2\times 2\times 2\times 2)$ because each RM has two "1"s. The circles in Fig. 11 show an example of a set of selected "1" s.

In Steps 2-3 and 2-4, it is judged whether the set of selected "1"s is repairable or not by comparing the number of redundancy sets for repair and that of built-in redundancy sets. The number of row redundancy sets for repair is calculated by adding the row coordinates of the "1"s in RM along the Y axis, and that of the column redundancy sets is calculated by adding the column coordinates of the "1"s in RM along the X axis. In Fig. 11, the candidates shown by circles are repairable because the number of redundancy sets for repair is less than or equal to the number of built-in redundancy sets.

If repairable, as shown in Step 2-6, the product of E_{RM} of each RM is added to Y_{AU}.

On the other hand, if there are no sets of "1"s in which the number of redundancy sets for repair is less than or equal to the number of built-in redundancy sets, the analysis unit becomes unrepairable.

Step 2 is repeated by changing the combination of RMs. The number of combinations becomes the (XMU YMU)th power of N_{RM}.

C. Application to Other Redundancy Schemes

Application of the new yield prediction method to flexible redundancy [17]–[20] is possible by setting the proper parameters of redundancy design. However, the application to block redundancy [21] seems difficult. In the case of flexible redundancy schemes in which all the redundancies can replace any failures without any regulation of a block division, there is no need to divide a chip into plural MUs such as Fig. 2. Therefore, the calculation itself can be conducted properly by our method without "Step 2" described in this chapter.

On the other hand, in the case of block redundancy schemes in which a block with defective bits is replaced by a spare block, our method cannot deal with it. However, we consider that the yield calculation in the case of spare block is simpler than in the case of normal spare row and column redundancy schemes, because only the number of failures should be taken into account for the calculation in the block redundancy. Other failure information such as the kinds of failures or the size of failures does not affect the calculated result. The calculation technique of this case has already been established [11].

V. Verification of Accuracy

In this section, we compare the actual yield and calculated yield in two kinds of system-LSI embedded memory devices, embedded memory A and embedded memory B whose design rules are different. The size of the failure distributions of embedded memory A and embedded memory B are shown in Fig. 12. Fig. 12 shows the average data from 24 wafers of embedded memory A and that from 21 wafers of embedded memory B, taken from fail bit map (FBM) data. Fig. 13 shows the redundancy design of embedded memory A and B. The redundancy design of embedded memory A is $NR = 2$, $WR = 1$, $NC = 1$, $WC = 8$ and the repairable area of row redundancies is two blocks along the Y axis and that of column redundancies is four blocks along the X axis. In the same way, the redundancy design of embedded memory B is $NR = 8$, $WR = 4$, $NC = 2$, $WC = 8$ and repairable area of row redundancies is the whole chip and that of column redundancies is half a chip.

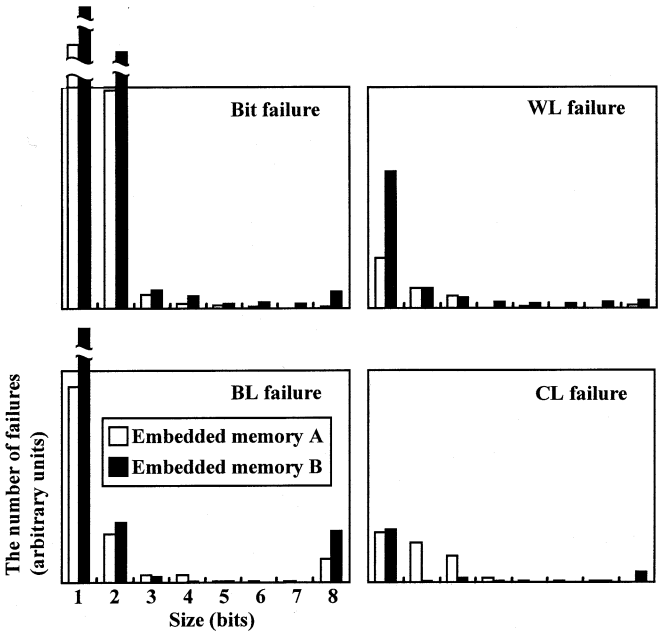


Fig. 12. Size distribution of failures in embedded memory A and B.

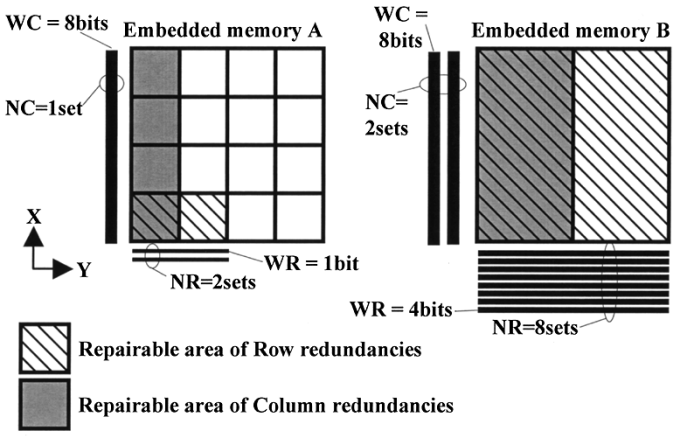


Fig. 13. Redundancy design of embedded memory A and B.

A. Embedded Memory A

Fig. 14 shows a comparison between the actual yield of embedded memory A and three yield prediction methods, the new method and two conventional methods. In Fig. 14, the two conventional methods do not consider asymmetric repair and assume that the number of row redundancy sets and that of column redundancy sets for repairing a bit or a CL failure are equal. In conventional method 1 the number of row redundancy sets for repair is equalized to that of column redundancy sets and *vice versa* in conventional method 2. For example, the 2×2 bit failure shown in Fig. 1 is repaired by one row or one column redundancy set in conventional method 1 and repaired by two Row or two Column redundancy sets in conventional method 2. These assumptions are taken into account when RMs are generated. The 2×2 bit failure shown in Fig. 1 must be converted

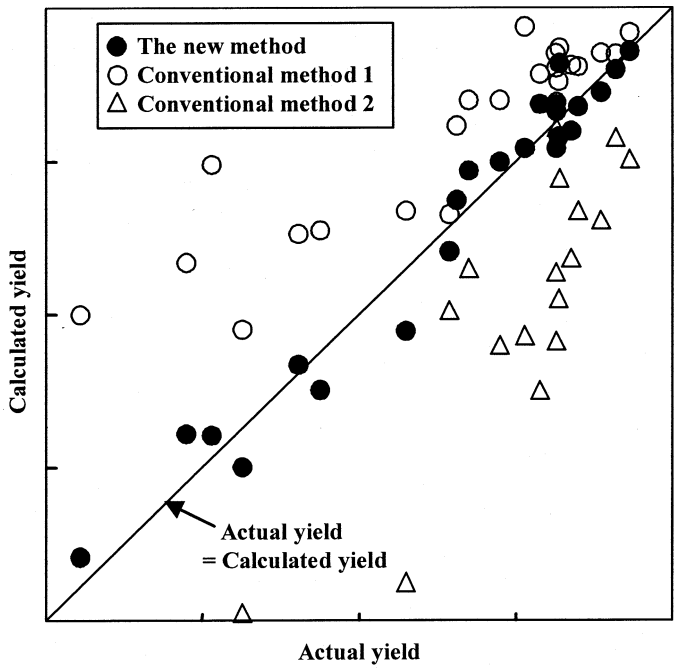


Fig. 14. Comparison of calculated yield with actual yield of embedded memory A. Interval of graduations is 5%.

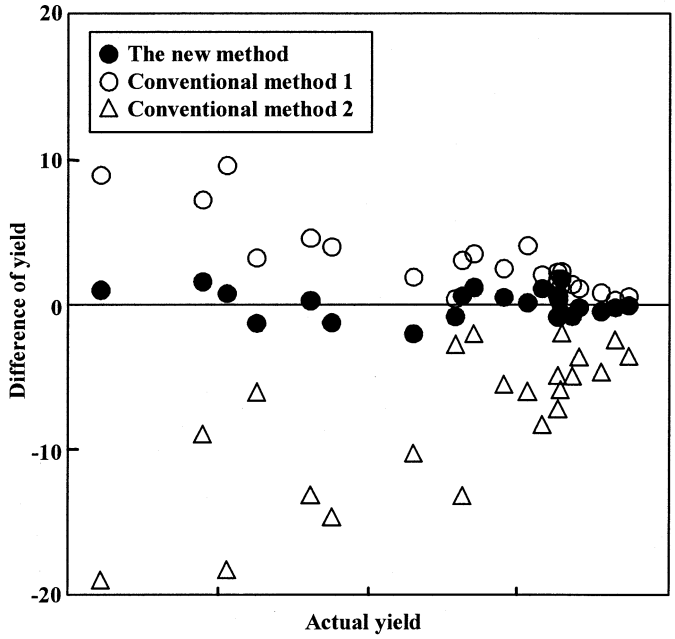


Fig. 15. Difference between calculated yield and actual yield of embedded memory A.

to "$rm_{21} = 1$;" however, in conventional method 1 it is converted to "$rm_{11} = 1$" and in conventional method 2 it is converted to "$rm_{22} = 1$." The yields of these conventional methods are calculated by the same method as described earlier. Fig. 14 shows that the new method is more accurate than the conventional methods.

Fig. 15 shows the difference between the actual yield and the calculated yield. The difference between the actual yield and the yield of the two conventional methods becomes larger as the actual yield becomes lower. The reason is that the number of bit and CL failures that cause asymmetric repair increases. On the other hand, the difference between the actual yield and yield by new method is small regardless of the value of the actual yield. On average, the difference in accuracy between the new method and conventional method 1 is about 3% and that of conventional method 2 is about 8%.

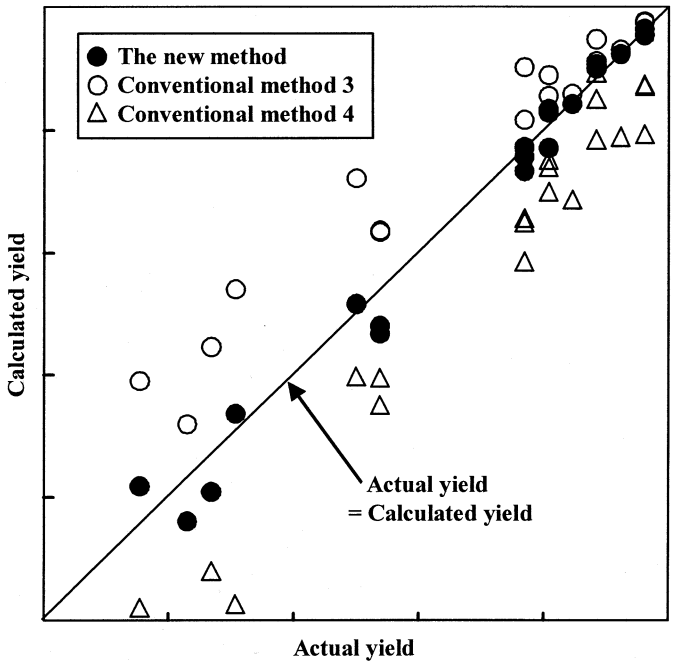


Fig. 16. Comparison of calculated yield with actual yield of embedded memory B. Interval of graduations is 20%.

It is often the case with system-LSI embedded memory that the width of one column redundancy set is much larger than that of one row redundancy. Therefore, to repair a bit or a CL failure, the number of column redundancy sets for repair is somewhat smaller than that of row redundancy sets. As a result, the yield of conventional method 1 is higher and conventional method 2 is lower than the actual yield. In general, the larger the difference of WR, the width of a row redundancy set, and WC, the width of a column redundancy set, becomes, the larger the inaccuracy of conventional methods becomes.

Compared with embedded memory B, the effect of asymmetric repair in embedded memory A was larger. The reason is that WR in embedded memory A is smaller than that of embedded memory B, while WC in both embedded memories is equal. The difference between WR and WC becomes larger in embedded memory A and that results in a larger effect for asymmetric repair. On the other hand, the effect of link set is not large in embedded memory A, because the number of failures is so small that the repair probability is not affected by the difference in the repairable area.

B. Embedded Memory B

Fig. 16 shows a comparison between the actual yield of embedded memory B and three yield prediction methods, the new method and two conventional methods. The two conventional methods do not consider link set. In conventional method 3, it is assumed that the repairable area of row redundancies is equal to that of column redundancies (repairable area is half a chip) and

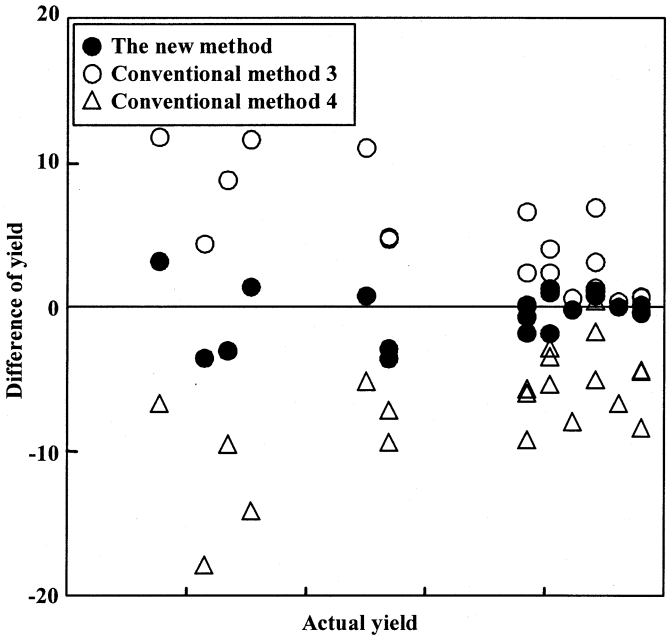


Fig. 17. Difference between calculated yield and actual yield of embedded memory B.

vice versa in conventional method 4 redundancies (repairable area is a chip).

Fig. 16 shows that the new method is more accurate than the conventional methods. Fig. 17 shows the difference between the actual yield and calculated yield. On average, the difference in accuracy between the new method and conventional method 3 is about 4% and that of conventional method 4 is about 7%.

In embedded memory B, the repairable area of row redundancies is larger than that of column redundancies. In other words, conventional method 3 assumes that the number of row redundancy sets in a chip is larger than the actual number, and conventional method 4 assumes that the number of column redundancy sets in a chip is smaller than the actual number. As a result, the yield of conventional method 3 is too optimistic and conventional method 4 is too pessimistic.

The effect of the link set is large and that of asymmetric repair is negligible in embedded memory B, contrary to embedded memory A. In this embedded memory B, the main mode of fatal failures was a large number of small failures in a chip. Therefore, the possible repair of such failures is strongly affected by the difference in repairable area. On the other hand, the effect of asymmetric repair is small because the difference between WR and WC in embedded memory B is smaller and the effect of asymmetric repair becomes small.

VI. Conclusion

The new yield prediction method using the RM successfully predicted more accurate yields for system-LSI embedded memories with various kinds of design rule, failure situation, and redundancy design. In our method, we take both asymmetric repair and link set into consideration. On average, the difference in accuracy between the new method and conventional methods is about 5% in advanced embedded SRAMs in system-LSIs. As a result, the shortage or surplus of lot inputs that amounts to two lots per 1000 wafers can be reduced. Thus, the new yield prediction method can be utilized not only to optimize the redundancy design but also to improve the productivity of chips in such ways as suitable plans for mass production and inventory adjustment.

Acknowledgment

The authors would like to thank to M. Ikeno of the ULSI Development Center for technical discussions. Also, they would like to thank A. Onoyama for technical advice.

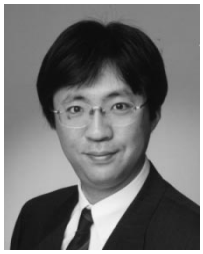
References

[1] A. Shubat, "Moving the market to embedded memory," *IEEE Design Test Comput.*, vol. 18, pp. 5–6, May–June 2001.
[2] C. H. Stapper, "Defect density distributions for LSI yield calculations," *IEEE Trans. Electron Devices*, vol. ED-20, pp. 655–657, July 1973.
[3] A. V. Ferris-Prabhu, "Role of defect size distribution in yield modeling," *IEEE Trans. Electron Devices*, vol. ED-32, pp. 1727–1736, Sept. 1985.
[4] J. A. Cunningham, "The use and evaluation of yield models in integrated circuit manufacturing," *IEEE Trans. Semiconduct. Manufact.*, vol. 3, pp. 60–71, May 1990.
[5] R. Glang, "Defect size distribution in VLSI chips," *IEEE Trans. Semiconduct. Manufact.*, vol. 4, pp. 265–269, Nov. 1991.
[6] C. N. Berglund, "A unified yield model incorporating both defect and parametric effects," *IEEE Trans. Semiconduct. Manufact.*, vol. 9, pp. 447–454, Aug. 1996.
[7] K. Sakurai, Y. Shimada, and K. Yamanishi, "A defect-situation forecasting technology to optimize future DRAM-redundancy design," in *Int. Symp. Semiconductor Manufacturing Proc. (ISSM)*, 1997, pp. E43–E46.
[8] H. Sato, M. Ikota, A. Sugimoto, and H. Masuda, "A new defect distribution metrology with a consistent discrete exponential formula and its applications," *IEEE Trans. Semiconduct. Manuf.*, vol. 12, pp. 409–418, Nov. 1999.
[9] C. Hess, D. Stashower, B. E. Stine, L. H. Weiland, G. Verma, K. Miyamato, and K. Inoue, "Fast extraction of defect size distribution using a single layer short flow NEST structure," *IEEE Trans. Semiconduct. Manufact.*, vol. 14, pp. 330–337, Nov. 2001.
[10] J. Khare, D. B. I. Feltham, and W. Maly, "Accurate estimation of defect-related yield loss in reconfigurable VLSI circuits," *IEEE J. Solid-State Circuits*, vol. 28, pp. 146–156, Feb. 1993.
[11] I. Koren and Z. Koren, "Defect tolerance in VLSI circuits: Techniques and yield analysis," *Proc. IEEE*, vol. 86, pp. 1819–1838, Sept. 1998.
[12] K. Mitsutake and Y. Ushiku, "Yield model with redundancy based on critical area calculations," in *Int. Symp. Semiconductor Manufacturing Proc. (ISSM)*, 2000, pp. 187–190.
[13] 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.
[14] S. Kikuda, H. Miyamoto, S. Mori, M. Niiro, and M. Yamada, "Optimized redundancy selection based on failure-related yield model for 64-Mb DRAM and beyond," *IEEE J. Solid-State Circuits*, vol. 26, pp. 1550–1555, Nov. 1991.
[15] C. H. Stapper, "Improved yield models for fault-tolerant memory chips," *IEEE Trans. Comput.*, vol. 42, pp. 872–881, July 1993.
[16] J. R. Day, "A fault-driven, comprehensive redundancy algorithm," *IEEE Design Test Comput.*, vol. 2, pp. 35–44, June 1985.
[17] T. Sugibayashi, I. Naritake, S. Utsugi, K. Shibahara, R. Oikawa, H. Mori, S. Iwao, T. Murotani, K. Koyama, S. Fukuzawa, T. Itani, K. Kasama, T. Okuda, S. Ohya, and M. Ogawa, "A 1-Gb DRAM for file applications," *IEEE J. Solid-State Circuits*, vol. 30, pp. 1277–1280, Nov. 1995.
[18] T. Yamagata, H. Sato, K. Fujita, Y. Nishimura, and K. Anami, "A distributed globally replaceable redundancy scheme for sub-half-micron ULSI memories and beyond," *IEEE J. Solid-State Circuits*, vol. 31, pp. 195–201, Feb. 1996.
[19] T. Kirihata, Y. Watanabe, H. Wong, J. K. DeBrosse, M. Yoshida, D. Kato, S. Fujii, M. R. Wordeman, P. Poechmueller, S. A. Parke, and Y. Asao, "Fault-tolerant designs for 256 Mb DRAM," *IEEE J. Solid-State Circuits*, vol. 31, pp. 558–566, Apr. 1996.
[20] I. Koren and Z. Koren, "Yield analysis of a novel scheme for defect-tolerant memories," in *Proc. 1996 IEEE Int. Conf. Innovative System in Silicon*, Austin, TX, Oct. 1996, pp. 269–278.

[21] J. H. Yoo, C. H. Kim, K. C. Lee, K. H. Kyung, S. M. Yoo, J. H. Lee, M. H. Son, J. M. Han, B. M. Kang, E. Haq, S. B. Lee, J. H. Sim, J. H. Kim, B. S. Moon, K. Y. Kim, J. G. Park, K. P. Lee, K. Y. Lee, K. N. Kim, S. I. Cho, J. W. Park, and H. K. Lim, "A 32-bank 1 Gb self-strobing synchronous DRAM with 1 Gbyte/s bandwidth," *IEEE J. Solid-State Circuits*, vol. 31, pp. 1635–1644, Nov. 1996.

Yutaka Shimada received the B.S. and M.S. degrees in materials science and engineering from Kyoto University, Japan, in 1994 and 1996, respectively.

He joined Mitsubishi Electric Corp. in April 1996 and worked in the productivity enhancement of CMOS devices including redundancy optimization, BIST architecture design, and defect analysis. He is currently working at Renesas Technology Corp.



Koichi Sakurai (M'96–SM'01) received the B.S. and M.S. degrees in electrical engineering from Keio University, Japan, in 1983 and 1985, respectively, and a D.E. degree from Kobe University, Japan, in 2002.

He joined Mitsubishi Electric Corporation in April 1985. He studied electron beam applications at the University of Southern California as a Visiting Researcher under a Mitsubishi Electric Scholarship from 1992 to 1993. Since then, he has been engaged in the yield enhancement of CMOS devices including defect inspection and reduction, redundancy optimization, and statistical analysis. He is currently working at Renesas Technology Corp.