Embedded Method of Soc Memory Repairing

V. Hahanov, W. Gharibi, K. Mostovaya

Computer Engineering Faculty, Kharkov National University of Radioelectronics, Lenin Ave. 14, Kharkov, Ukraine, 61166, phone: (057) 70-21-421, (057) 70-21-326, e-mails: hahanov@kture.kharkov.ua, wgharibi@kkwu.edu.sa, kit@kture.kharkov.ua

SOC Infrastructure Intellectual Property Technologies

The memory diagnosis and repair problem [1-12] is related to the tendency of continuous reduction of chip area, which is allocated to original and standardized logic, and simultaneous growth of embedded memory. As it is represented in Fig. 1, increase of the memory specific weight on a chip reduces to its complete dominance for data and program storage, which will reach 94% by the year 2014 [2]. It will provide not only fast response of carrying out of functions, but also flexibility that is appropriate to software concerning design error correction.

The memory element feature is the fact that some cells can fail under fault influence in the process of production and operation. This fact not always brings a memory matrix into a critical state, when the repair is not possible. So, such technical memory state, when the total quantity of faulty cells is not greater than spare capacities of a device (intended for repair), is considered below.

The research aim is to develop an algebra-logical method of embedded matrix memory diagnosis and repair in real time.


Fig. 1. Memory specific weight on a chip SoC

Modern design technologies of digital systems on chips propose along with creation of functional blocks F-IP development of service modules I-IP, which are based on complex solving of the project quality problem and yield increasing in manufacturing that is determined by implementation of the following services into a chip [9]:

1) Diagnosis of failures and faults by analysis of information, which is obtained on the stage of testing and use of special methods of embedded fault lookup on the basis of the standard IEEE 1500 [12];
2) Repair of functional modules and memory after fixation of a negative testing result, fault location and identification of a fault type in carrying out of the diagnosis phase;
3) Measurement of the general characteristics and parameters of a device operation on the basis of on-chip facilities, which enables to make time and volt-ampere measurements;
4) Reliability and fault tolerance of a device operation in working that is obtained by diversification of functional blocks, redundancy of them and repair of SoC in real time.

Memory diagnosis and repair method

It is the representation of an exact method of memory elements diagnosis and repair by spares that enables to cover a set of fault cells by the minimally possible quantity of spares. The method is oriented on implementation of the Infrastructure Intellectual Property for SoC functionality. The structure solutions for realization of the method of diagnosis and repair of memory matrix fault cells are proposed [5-7, 10].

In the process of operation and repair any kinds of memory it is necessary the guarantee of its technical compliance. In this regard, three procedures are carried out as given below: 1) Memory testing that consists of test patterns input, which oriented on identification of specific kinds of faults [1, 7]; 2) In the case of fault appearance, it is necessary an additional diagnosis procedure that enables to determine location, cause and kind of fault; 3) After fault detection, which blocks carrying out of the memory function, it is necessary to activate the repair process – replacement of faulty elements by spares, which initially are on a chip [5, 6]. Thereby, aforementioned actions are oriented on the growth of yield without significant
additional time and material costs. To repair, it is necessary to apply a special mechanism for memory repair, by means of replacement of faulty components by faultless ones from the chip reserve.

As a rule the testing procedure is realized by BIST-block (Built-In Self Test), which is hardware fast-acting generator of test patterns, as well as an analyzer (signature) of reactions of memory outputs on test patterns. Repair analysis consists of definition of covering possibility of faulty memory elements by available reserve components. Memory module has two parts: 1) functional cells, which are used for data and program storage, when a module is used in SoC; 2) reserve or spare cells, which are designed for memory repair in case of functional cells failure. Functional and reserve cells are joined together in the form of columns and rows. When a fault is detected, a row (a column), which includes a faulty element, is disconnected from the functional structure of memory cells and a row (a column) from chip reserve is connected on its place. The number of reserve components is limited, so it is necessary to apply a special mechanism of effective allocation of repair resource, for support of faulty memory elements covering by the minimally possible quantity of redundant rows and columns.

The search procedure of faulty cells covering by the minimal quantity of reserve rows and columns described above, can be realized as on-chip repair module or external. In the second case data about errors is received from external modules; they are processed and pass to the controller that provides memory repair. It results in considerable time loss. So, the preferable solution is on-chip module realization, when data about errors is passed from BIST directly. Such mechanism is called as BIRA [5, 8] – Built-In Repair Analysis.

Memory repair is realized by disconnection of faulty elements (rows and columns of a matrix) by means of electrical fusion of metal links and connection of reserve ones. The fuse process can be electrical or laser. Electrical fuse equipment has smaller dimensions than laser one and it is used more frequently. Fuse is carried out by means of an instruction set, which can be stored in permanent memory inside chip (hard repair) or in random-access memory (soft repair) [5–7]. Soft repair has several advantages: when a defect appears, a new corrected instruction can be recorded to memory easily; there provide economic use of chip area and sufficient reliability [3]. Hard repair enables to use a simplified manufacturing test and provides detection of errors, which can not be fixed by soft repair under certain circumstances (for instance, overheat).

The structure of on-chip memory analysis and soft repair processes (BISR) [5–7] is represented in Fig. 2. 1) Chip activation, filling of the BISR register by zero values. 2) Run the BIST controller. Memory testing and accumulation of information about faulty cells in the BIRA register. 3) Transfer of information about faulty cells to the BISR register for subsequent fusion. 4) Scanning the BIRA registers, which contain the repair status, by the BIST controller for obtaining of faults information 5) Run the fuse controller in record mode and transfer the repair instructions from the BISR. 6) Chip restart. Recording the fuse information to the BISR register, replacement of faulty rows and columns by reserve components is fulfilled. 7) Run the BIST controller for repeated memory testing and verification of the repair result correctness.

The objective function $Z$ of the given research can be defined on the basis of modern progress in the field of online memory repair in the following way: minimization of the repair cost (hardware costs) of a memory module $M = M_{ij}$ in the process of SoC operation by means of applying algebra-logical method of minimization of faulty-cells-set covering by a system of reserve elements subject to the constraints $N$ on quantity of ones:

$$Z = \min \{ Q_i(F) \} \text{ s.t. } Q_i(F) \leq N_{1 \text{ max}} = N_{1} + N_{c}, \quad (1)$$

where $Q_i(F)$ – the cost of $i$-th solution variant of the memory module $M = M_{ij}$ repair by means of the minimal subset of rows and columns $R = \{R_i, R_c\}$ of chip reserve that covers the set $F$ of faulty memory cells

$$R \cap F = F, \quad Z^* = \max | F_i |, F_i \in F \leftrightarrow \forall R_i. \quad (2)$$

Method of minimal covering obtainment on an example, a memory matrix with five faulty cells [8], two reserve rows and a reserve column (Fig. 3) is considered below. Every reserve component (a row or a column) can repair from one up to $n$ faulty cells, which belong to a row or a column.

The idea behind this method is to optimize replacement of faulty memory matrix elements by means of solving covering problem of faulty columns by row reserve. To illustrate the method, it is proposed originally to use the covering matrix of given faults $F$ by some quantity of rows (it can be test patterns or reserve rows) $X$ and

$$|F| \geq |X| = \{F_1, F_2, F_3, F_4, F_5, F_6, F_7, F_8\} \geq \{X_1, X_2, X_3, X_4, X_5, X_6\}. \quad (3)$$

Let the matrix $Y$ is specified:

$$(F_j \cap X_i = \varnothing \Rightarrow Y_{ij} = 1) \& (F_j \cap X_i = \neq \Rightarrow Y_{ij} = 0): \quad (4)$$

$$\begin{array}{cccccc}
X & X_1 & X_2 & X_3 & X_4 & X_5 & X_6 \\
X_1 & 0 & 1 & 0 & 1 & 1 & 0 \\
X_2 & 1 & 1 & 1 & 0 & 0 & 0 \\
X_3 & 1 & 1 & 1 & 0 & 0 & 0 \\
X_4 & 1 & 1 & 1 & 0 & 0 & 0 \\
X_5 & 1 & 1 & 1 & 0 & 0 & 0 \\
X_6 & 1 & 1 & 1 & 0 & 0 & 0 \\
\end{array}$$
The exact solution of the covering problem of faults by minimal quantity of reserve memory rows is based on synthesis of the Boolean function which is written as product of sums, written by constituents of unities, corresponding to columns of the matrix:

\[ Y = (X_1 \lor X_4) \& (X_2) \& (X_2 \lor X_3) \& (X_4 \lor X_5) \& (X_3 \lor X_6) \& (X_1 \lor X_6) \& (X_2) \& (X_2 \lor X_3). \]  

(5)

In the given case, an analytic notation in the Boolean function form, represented in conjunctive normal form (CNF), is the initial model that contains a full set of covering problem solutions which is solved by finding disjunctive normal form (DNF). The transformation procedure of CNF to DNF by means of all multiplication terms is performed. As a result of equivalent transformations, performed in compliance with the algebra of logic rules, it is came out the Boolean function that contains all possible fault covers, defined by four variants of row combinations:

\[ Y = (X_1 X_2 \lor X_2 X_4)(X_2 X_4 \lor X_4 X_5 \lor X_5 X_6 \lor X_2 X_6) \& (X_1 X_3 \lor X_6 X_6 \lor X_1 X_6)(X_2 X_2 \lor X_2 X_3) =
(X_1 X_2 X_4 \lor X_2 X_4 X_6 \lor X_4 X_2 X_6 \lor X_2 X_2 X_6). \]  

(6)

The minimal solution of covering problem contains three reserve rows, which can cover 8 faults in the memory matrix:

\[ Y = X_2 X_4 X_6. \]  

(7)

For the use of proposed memory-repair-method, it is necessary to remember that every fault \( F_i \) in a memory matrix belongs to a row and a column simultaneously. So, transformation of the topological fault model to the covering matrix consists of assignment of row and column numbers, which are distorted by given fault, to every fault. For instance (Fig.3), where there are 5 faulty cells, which are covered by three columns and 4 rows, the transformation turns a memory matrix into the covering table, where left column specifies one-to-one correspondence between fault coordinates (row and column numbers of a memory matrix) and rows of a fault covering:

\[
\begin{array}{|c|c|c|c|c|c|c|c|c|c|c|}
\hline
\text{C} & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\
\hline
1 & \\
2 & \\
3 & \\
4 & \\
5 & \\
6 & \\
7 & \\
8 & \\
9 & \\
10 & \\
\hline
\end{array}
\]

Fig. 3. Memory matrix with fault cells

In other words, the memory matrix topology is transformed from two-dimensional metrics to one-dimensional row structure, which have defined covering features concerning about fault columns.

The following Boolean function forms logical product of disjunctions, written by constituents of unities, corresponding to columns of the matrix

\[ (F_j \cap X_i \neq \emptyset \rightarrow Y_{ij} = 1): \]

\[
Y = (X_3 \lor X_4)(X_1 \lor X_4)(X_2 \lor X_5)(X_1 \lor X_6)(X_2 \lor X_7) =
(X_1 X_3 \lor X_1 X_4 \lor X_3 X_4 \lor X_4)X_1 X_2 X_5 \lor X_2 X_6 \lor X_2 X_7
\]

(8)

The equivalent transformations enable to simplify the complex construction – conjunctive normal form – and to obtain the minimal set of all solutions, a number of that is equal to six in this case. Subset of minimal solutions is defined by three conjunctive terms, every of which contains 3 reserve elements for memory matrix repair:

\[ Y = X_1 X_2 X_3 \lor X_1 X_2 X_4 \lor X_2 X_4 X_6. \]  

(9)

Formalization of optimal algebra-logical memory repair method

The objective function is defined as minimization of reserve components of the memory matrix (S – spare), which are needed for its repair in the process of SoC operation by means of synthesis of disjunctive normal form of faulty elements, covering and subsequent choice of the minimal conjunctive term \( X^l \in Y \) that satisfies the limitations on quantity of the reserve rows and columns \( S^r_{\text{max}}, S^c_{\text{max}} \), which enter into the logical product:
\[
Z = \min_{i=1}^{m}(X^i) = \min_{i=1}^{m} \left| X^i \right| = \left| S_{\text{max}} \right| \left| S_{\text{max}} \right| \left| S_{\text{max}} \right| \left| S_{\text{max}} \right|, \quad \text{(10)}
\]

\[
X^i \in Y = \{ X^1, X^2, \ldots, X^i, \ldots, X^n \}, \quad \text{(11)}
\]

\[
X^i = (X^i_1, X^i_2, \ldots, X^i_{m_1}, \ldots, X^i_1, X^i_m), \quad \text{(12)}
\]

where every resulting conjunctive term of the function \( Y \) is made from the row and column identifiers \( X^i = (R^1, C^1) \), which cover all faults in the memory matrix. The best solution is a term of minimal length at Quine mark, in which there are rows and columns, covering all faults. In a particular case, a solution can contain none rows (columns), when existing columns (rows) from memory matrix reserve are sufficient for memory repair. The model of definition process of minimal quantity of spares, which cover all detected faults in a memory matrix, comes to the following items:

1. Transformation of two-dimensional model of a memory matrix faults to the fault covering table by reserve rows and columns. To achieve the goal of topological memory model in the form of matrix, identifying detected faults, is considered:

\[
M = M_{ij} \quad M_{ij} = \begin{cases} 1 & \text{if } \Omega(f) = 1; \\ 0 & \text{if } \Omega(f) = 0. \end{cases} \quad \text{(13)}
\]

Here a matrix coordinate is equal to 1, if the fault-free behaviour function of a cell gives unit value on a test, the coordinate is identified as faulty. After fixation of all faults construction of the fault covering table \( Y = [Y_{ij}], i = 1, n; j = 1, m \) is carried out, where columns correspond to the set of detected faults \( m \) and rows are numbers of columns and rows of a memory matrix, which have faults:

\[
Y = [Y_{ij}] \quad Y_{ij} = \begin{cases} 1 & \text{if } C_i(R_j) \cap F_j \neq \emptyset; \\ 0 & \text{if } C_i(R_j) \cap F_j = \emptyset. \end{cases} \quad \text{(14)}
\]

Instead of the two-dimensional metrics components \( C \) and \( R \) the one-dimensional vector is used, it is concatenated from two sequences \( C \) and \( R \), the power of which is equal to \( n = p + q \):

\[
X = C^*R = (C_1, C_2, \ldots, C_p, R_1, R_2, \ldots, R_q) = X^c \times X^r = (X^c_1, X^c_2, \ldots, X^c_n, X^r_1, X^r_2, \ldots, X^r_q). \quad \text{(15)}
\]

At that there exists one-to-one correspondence between the initial set elements \( C \) and \( R \), the resulting vector \( X \), which is defined in the first column of the matrix \( Y \). It is necessary to say that transformation \( X = C^*R \) is carried out, for ease of consideration and subsequent forming of disjunctive normal form (uniformity of variables, forming the Boolean function). If the procedure is not carried out, the function is defined by two kinds of variables, containing rows and columns of a memory matrix.

2. Construction of conjunctive normal form for analytic, complete and exact solution of the covering problem. After generation of the covering matrix that contains zero and unit coordinates the synthesis of analytic covering form is carried out by writing of CNF by columns. Here, a number of conjunctive terms is equal to quantity of table columns and every disjunction is written by unit values of a current column:

\[
Y = \bigwedge_{j=1}^{m} (Y_{pj} \lor Y_{qj}) \quad (Y_{pj}, Y_{qj}) = 1 = \bigwedge_{j=1}^{m} (X_{pj} \lor X_{qj}). \quad \text{(16)}
\]

From the last expression, it is obvious that every column has two coordinates only, which has unit value, and the number of logical products is equal to the total quantity of faults \( m \), detected in a memory matrix.

3. Transformation of CNF to DNF that enables to find out all solutions of the covering problem. It is necessary to apply an operation of logical multiplication and the minimization (absorption) rules to conjunctive normal form to obtain the following disjunctive normal form:

\[
Y = \bigvee_{j=1}^{w} (k_1 \lor k_2 \lor \ldots \lor k_n), k_i = \{0, 1\}. \quad \text{(17)}
\]

It is the generalized DNF notation, where the limit number of terms is equal to \( w = 2^n \), where \( n \) is the quantity of rows in the generalized set \( (C, R) \) or quantity of the variables \( X \) in the matrix \( Y \), on the set of which all solutions are formed (fault covering by reserve components); if \( k_i \) at \( X_i \) is equal to zero, the variable \( X_i \) is nonessential.

4. Choice of minimal and exact solutions of the covering problem. It is related to the determination of minimal length conjunctions in the achieved DNF. The following transformation executing rows and columns of a memory matrix on the basis of the above-mentioned correspondence which enables to write a minimal covering or set of ones in two-dimensional metrics of rows and columns, which satisfies the conditions (limitations) of the objective function on the quantity of reserve components.

The computational complexity of algebra-logical memory repair method in the part of solving of the covering problem is determined by the following expression:

\[
Q = 2^{[F]} + |C + R| \times 2^{[F]}, \quad \text{(18)}
\]

where \( 2^{[F]} \) is costs related to DNF synthesis by logical multiplication of two-component disjunctions (fault coordinate is defined by row and column numbers), where their quantity is equal to the quantity of faulty cells; \( |C + R| \times 2^{[F]} \) is upper limit of computational costs, which are needed for minimization of the obtained DNF on maximum set of variables which is equal to the total quantity of rows and columns \( |C + R| \).

In the worst case, when coordinates of all faulty cells are not correlated by rows and columns (they are unique), for instance, diagonal faults, the computational complexity of the matrix method is dependent only on the quantity of faulty cells and its analytic notation is transformed to the following view:

\[
Q = 2^{[F]} + 2^{[F]} \times 2^{[F]} \times 2^{[F]} = 2^{[F]} + 2^{[F]} \times 2^{[F]} \times (1 + 2^{[F]}). \quad \text{(19)}
\]
If instead of fault set power to use quantity $m$ of them, the previous expression can be represented in more simplified form:

$$Q = 2^m \times (1 + 2 \times m) = 2^m (2m + 1).$$

(20)

According to the SoC Functional Intellectual Property Infrastructure, the matrix repair method on the basis of solving covering problem is implemented into a chip as one of IP components, designed for the functional support of SoC matrix memory.

**Conclusion**

Scientific novelty is to follow. SoC memory in the future will occupy more than 90% of the chip area which is oriented on use flexible software. Development of models and methods of quick and exact diagnosis, as well as technologies for repairing faulty cells by on-chip facility in real time and on all life cycle stages of a product are urgent problems. It enables to decrease quantity of chip pins, to raise yield, to decrease time-to-market, to reduce service costs, as well as to remove output diagnosis and repair facility.

The algebra-logical memory repair method is based on solving the faulty cells covering problem by spares by means of the Boolean algebra apparatus. The method has quadratic computational complexity and can have hardware or software realization that is service module of fault correction, which enables to carry out memory elements repair in the process of operation.

The classical covering problem use two one-dimensional vectors $(X, F)$, where the covering operator $P$ enables to find minimal subset of the components $X$, which cover all elements from $F$:

$$X_{\text{min}} \subseteq P(X, F) = X \cap F = X_{\text{min}}$$

by its aggregate functionality. The statement of covering problem of one-dimensional vector $F$ features by two-dimensional matrix $M = (C \times R)$ is needed in the reduction of both components to a single matrix (such coordinate system that is common denominator for both structures). Such metrics for the matrix $M = (C \times R)$ and the vector $F$ is one-dimensional structure. So, in this case it is necessary to carry out transformation of two-dimensional structure (memory fault matrix) $M = (C \times R)$ to one-dimensional one by means of concatenation operation $X = (C \times *)$ for subsequent solving of the classical covering problem by application of formal actions, which are defined by the operator $X_{\text{min}} = P(X, F)$.

The proposed method of optimal memory fault repair differs from analogs by application of algebra-logical technology of fault covering by two-dimensional memory matrix topology that enables to obtain minimal and full solutions for subsequent repair in real time, which is based on utilization of spares in the form of memory rows and columns.

Practical significance of the research consists of implementation of the method to SoC Functional Intellectual Property Infrastructure. It enables to raise yield essentially (5-10%) on the electronic technology market by means of faulty chip repair in the process of production and operation, as well as to increase the life cycle duration of memory matrices by repairing them in real time.

On-chip repair is based on all objects, which have an address: memory, multiplexers and matrix processors. If it is necessary to repair other structures, they must be designed with an allowance for component addressability. The addressability and regularity of components turn a system into reliable, robust, repairable and durable.

Further research is oriented on development of testability structure of the system and hardware, BIRA module for embedded memory repair in appearance of faults on production and operating stages.

Yervant Zorian (EWDTS’ 2007, Yerevan): “Now the main problem of system on a chip repair is development of embedded technologies and methods of the logic repair that occupies no more than 10% of chip area”.

**References**


Received 2008 10 15
It is proposed SoC Functional Intellectual Property Infrastructure that differ by minimal set of the embedded diagnosis processes in real time and enables to realize such services: functional testing on basis of generable input patterns and analysis of output reactions; SoC diagnosis with given resolution of fault location; fault simulation to carry out the first two procedures on basis of the fault detection table. Structural-algebraic method of embedded fault diagnosis of SoC functional blocks is proposed. The method uses preliminary analysis of the fault detection table to reduction its size and subsequent DNF construction computation, which forms all solutions of SoC functional diagnosis in real time. Ill. 3, bibl. 12 (in English; summaries in English, Russian and Lithuanian).


Предложен алгебро-логический метод оптимального восстановления работоспособности памяти, основанный на решении задачи покрытия дефектных ячеек резервными элементами путем использования аппарата булевой алгебры. Метод позволяет автоматически выполнять восстановление работоспособности элементов памяти в процессе функционирования и может иметь аппаратную или программную встроенную реализацию, представляющую собой сервисный модуль исправления дефектов. Ил. 3, библ. 12 (на английском языке; рефераты на английском, русском и литовском яз.).


It is proposed SoC Functional Intellectual Property Infrastructure that differ by minimal set of the embedded diagnosis processes in real time and enables to realize such services: functional testing on basis of generable input patterns and analysis of output reactions; SoC diagnosis with given resolution of fault location; fault simulation to carry out the first two procedures on basis of the fault detection table. Structural-algebraic method of embedded fault diagnosis of SoC functional blocks is proposed. The method uses preliminary analysis of the fault detection table to reduction its size and subsequent DNF construction computation, which forms all solutions of SoC functional diagnosis in real time. Ill. 3, bibl. 12 (in English; summaries in English, Russian and Lithuanian).