Combined methods of tolerance increasing for embedded SRAM

I A Shchigorev and I I Shagurin
National Research Nuclear University MEPhI (Moscow Engineering Physics Institute), Kashirskoe shosse 31, 115409 Moscow, Russian Federation
Corresponding author e-mail address: l.shchigorev@module.ru

Abstract. The abilities of combined use of different methods of fault tolerance increasing for SRAM such as error detection and correction codes, parity bits, and redundant elements are considered. Area penalties due to using combinations of these methods are investigated. Estimation is made for different configurations of 4K × 128 RAM memory block for 28 nm manufacturing process. Evaluation of the effectiveness of the proposed combinations is also reported. The results of these investigations can be useful for designing fault-tolerant “system on chips”.

1. Introduction
One of the most important challenges in the design of electronic systems is ensuring the validity and reliability of storage and transfer of information (instructions, addresses, data). This problem is topical for using chips containing embedded memory cores, as it is the main source of failures especially with the influence of ionizing radiation. Most modern chips are manufactured on commercial sub-100-nm processes, which are often unique and extremely specialized. Memory cells of these devices have increased sensitivity to the effects of external factors, such as a strike of heavy charged particles, which cause the occurrence of failures [1].

The increase of radiation tolerance of chips produced on commercial CMOS technology is achieved by using special circuit design and algorithmic techniques – Radiation Hardening by Design (RHBD) [2].

Detection and error correction during data storage and transfer are made by special error detection and correction codes (EDAC) [3-4] and redundancy. EDAC can detect and/or correct the errors associated with the emergence of faults under the influence of heavy charged particles or electromagnetic noise, but in the case of failed memory elements, their corrective abilities are greatly reduced. Thus, single error correcting code (SEC) will not be able to correct failures under the influence of heavy charged particles, or electromagnetic noise if one faulty element already exists.

Double or triple redundancy of memory blocks requires considerable hardware costs, which increases the area of VLSI and reduces its technical and economic characteristics. More economical is the use of redundant elements (column and/or rows) in the matrix of memory elements that is connected instead of failed ones. This method is used at the final stage of production of memory chips to increase the yield rate of the chip [5].

This article is devoted to the investigation of the possibility of combined use of EDAC and the redundant elements of SRAM for memory blocks failure tolerance improvement.
2. Used methods
To identify single-bit errors occurring in the transmission of information, the parity bits of the transmitted data words are used. Error correction is implemented using a Hamming code that provides double detection and correction of single errors. EDAC encoders and decoders are introduced into the composition of VLSI.

To improve the yield of chips the redundant memory elements (columns and/or rows) are used. To implement this method, built-in self-test (BIST) and built-in self-repair (BISR) units are included. The BIST can conduct unit testing and determine the position of the damaged elements; then the BISR unit replaces the faulty element by the redundant one. Redundant rows and columns that are not used for production control can be used during the operation of the VLSI by replacing the failing memory elements.

To estimate the number of redundant elements required for manufacturing repairing, we define the probability of obtaining faulty memory elements, depending on the area \( A \) , occupied by the memory block on the chip, and the density of defects \( D \). Calculation of the probability of \( k \) faulty memory elements is made according to the equation (1) obtained for the distribution of defects in accordance with the Poisson distribution [6].

\[
P(k) = \frac{e^{-\lambda} \lambda^k}{k!}.
\]

The value of \( \lambda \) is calculated as the product of chip area \( A \) and the density of defects per unit area \( D \):

\[
\lambda = A \times D.
\]

Table 1 shows the probability of \( k \) (0, 1, 2 and 3) number of error in the memory block for different areas: 1 mm\(^2\), 10 mm\(^2\) and 100 mm\(^2\). The value \( D \) is \( 1.395 \times 10^{-3} \) faults/mm\(^2\) according to the experts ITRS for 2016 [7].

| A (mm\(^2\)) | P(0)   | P(1)   | P(2)   | P(3)   |
|------------|--------|--------|--------|--------|
| 1          | 0.9986 | 0.0014 | 0      | 0      |
| 10         | 0.9861 | 0.0138 | 0.0001 | 0      |
| 100        | 0.8697 | 0.1213 | 0.0085 | 0.0004 |

The results show that even for large memory blocks (\( A = 100 \) mm\(^2\)), the probability of failure due to crystal defects does not exceed 12%. With an area of 10 mm\(^2\) and less than this probability does not exceed 1.4%. Thus, a significant part of the redundant elements remains unused for production control and can be used to perform self-repairing during the further operation of the chip.

To evaluate the possibility of using redundant elements as the factors of fault tolerance increasing SRAM 4K \( \times \) 128 was designed. The results are shown for different ways of memory block dividing: 1 bank 4K \( \times \) 128 bit, 2 banks 4K \( \times \) 64 bit, 4 banks 4K \( \times \) 32 bit, 8 banks 4K \( \times \) 16 bit and 16 banks 4K \( \times \) 8 bit. Each of the banks contains 2 redundant columns.

To evaluate the proposed method of reserving two columns, the simulation of repair rate of redundant elements was performed. It is defined as the ratio of repaired memory elements to the total number of faulty elements. The simulation was performed for 10000 SRAM blocks using programs in the System Verilog language. According to Poisson probability distribution \( k \) faulty elements errors occurring repair ratio for manufacturing repairing is calculated for different memory dividing. The evaluation was performed for 100mm\(^2\) area. The result is shown in Figure 1.

These obtained results allow making the conclusion that the repair ratio for all memory dividing is more than 99%.
The next step is to evaluate a number of redundant elements which will be available after yield improvement. For this, the estimation of the percentage of unused redundant elements (URR: unused redundancy ratio) was made. This value is calculated by the equation (3), where MRE (made redundancy) means redundant elements that were designed and URE (used redundancy) means redundant elements which were used in the process of yield improvement at the semiconductor factory.

$$URR = \frac{MRE - URE}{MRE} \times 100\%.$$  \hspace{1cm} (3)

Result is shown in Figure 2.

To evaluate the effectiveness of the method of reserving two columns the simulation of repair ratio for different numbers of faulty elements per block of memory was made. In Figure 3 the dependence of repair ratio from the number of faulty elements per block of memory is shown.
These obtained results allow that the implementation of memory blocks in multiple banks provide a high probability of correction (90% or more) for both single and double failures occurrence during the operation.

3. Combined methods of fault tolerance increasing

The use of redundant elements is proposed as one of the methods of increasing the fault tolerance for SRAM. This way can be applied alone or in combination with EDAC. The proposed variants are combinations of the three methods: byte parity checking (BPC), single error correction (SEC) and 2 redundant columns per bank (RC). Different combinations can contribute to the total fault tolerance effect differently. For example, the combination of methods SEC and BPC allows to correct the read data words in the case of failures of storage elements and to detect errors occurring on the READ and WRITE buses. The combination of RC and SEC, in its turn, allows to activate the redundant columns with addition of saving the error symptom and address after the READ operation for this address (without interrupting access to the memory and preservation of the available information), followed by periodic sessions of self-repair after detection of SEC the damaged items (with the temporary termination of access to the memory and loss of information).

4. Area penalties estimation

However, all these methods require additional hardware and costs. So EDAC requires additional area for additional bits, encoders, and decoders. Redundant elements require additional area not only for themselves but also for the built-in self-test (BIST) and built-in self-repair (BISR) units. Thus, the challenge of increasing the fault tolerance leads to finding the balance between the additional costs of hardware and the achieved level of data protection.

For estimation of hardware penalties different configurations of 4Kx128 RAM memory block with additional units such as EDAC encoders-decoders, BIST and BISR for 28 nm manufacturing process were produced.

Area estimation was provided in the following situations:
- without fault tolerance increasing (WOFT);
- with per byte parity bits (PB);
- with single error correction (SEC);
the combination of PB and SEC (PB+SEC);
the combination of SEC and redundant columns (SEC+RC);
the combination of SEC, RC and PB (SEC+RC+PB).

Result is shown in Figure 4.

Figure 4. Area for different ways of fault tolerance increasing as a function of different arrays dividing.

These obtained results allow to make the following conclusions:
- The combination of all three considered methods (SEC+PB+RC) leads to increasing required area from 30 to 75% when compared variants have the same dividing;
- The dividing of SRAM block into the banks increases the total area up to 1.6 – 2.1 times depending on the number of banks and method of increasing the failure tolerance;
- The most protected case is obtained by dividing a block into 16 8-bit banks and the use of SEC. In this embodiment, the number of detected errors increases by 16 times. Also the number of redundant elements increases by 16 times (with 2 to 32).

Conclusion
The results of the investigation show that the use of combined techniques for increasing the fault tolerance of the embedded SRAM blocks requires 1.6 – 2.1 times area penalties.

The efficiency of array dividing is shown by evaluating the repair ratio. The number of detected errors increases in proportion to the number of used banks. The probability of correction single or double errors is more than 90% for the case of all combined methods using.

The obtained results can be used in the development of embedded memory blocks included in the design of "systems on a chip" and the evaluation of their characteristics.

References
[1] Uznanski S, Gaisot G, Roche P and Autran J 2010 IEEE T. Nucl. Sci. 57 1876-83
[2] Cannon E, Tostenrude J, Cabanas-Holmen M, Meaker B, Neathery C, Carson M and Brees R 2013 IEEE T. Nucl. Sci. 60 4207-13
[3] Krasnyuk A and Petrov K 2013 Russian Microelectronics 42 53-8
[4] Hamming R 1950 Bell Syst. Tech. J. 29 147-60
[5] Chang D Li J and Huang Y 2008, J. of electronic testing-theory and applications 24 181-92
[6] Horiguchi M, Itoh K 2011 Nanoscale Memory Repair (New York: Springer) p 215
[7] International Technology Roadmap for Semiconductors 2007 Yield Enhancement Available at: http://www.itrs2.net/itrs-reports.html.