

Published in IET Computers & Digital Techniques
 Received on 31st October 2008
 Revised on 24th August 2009
 doi: 10.1049/iet-cdt.2008.0146



Fault-tolerance techniques for hybrid CMOS/nanoarchitecture

A. Melouki S. Srivastava B.M. Al-Hashimi

School of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, UK
E-mail: ss3@ecs.soton.ac.uk

Abstract: The authors propose two fault-tolerance techniques for hybrid CMOS/nanoarchitecture implementing logic functions as look-up tables. The authors compare the efficiency of the proposed techniques with recently reported methods that use single coding schemes in tolerating high fault rates in nanoscale fabrics. Both proposed techniques are based on error correcting codes to tackle different fault rates. In the first technique, the authors implement a combined two-dimensional coding scheme using Hamming and Bose–Chaudhuri–Hocquenghem (BCH) codes to address fault rates greater than 5%. In the second technique, Hamming coding is complemented with bad line exclusion technique to tolerate fault rates higher than the first proposed technique (up to 20%). The authors have also estimated the improvement that can be achieved in the circuit reliability in the presence of Don't Care Conditions. The area, latency and energy costs of the proposed techniques were also estimated in the CMOS domain.

1 Introduction

Molecular electronics holds the promise to overcome the physical limitation of lithography-based VLSI technology and offer the possibility of significantly denser circuits. However, tremendous growth in device density will be accompanied by a substantial increase in hard [1–3] and soft [4–6] faults. To achieve acceptable levels of manufacturing yield and computational reliability, fault tolerance must be integrated into the design flow of nanoscale circuits.

Much work has already been done in the area of fault tolerance/avoidance for nanotechnology to increase circuit reliability in the presence of increased hard and soft error rates. One of the proposed techniques, triple-modular-redundancy (TMR), is based on the use of three copies of the same module and an arbitration unit [7, 8]. TMR technique fails when there are faults in more than one module. The reliability of TMR is also limited by that of the final arbitration unit making this approach insufficient in the presence of high defect rates [7]. Reconfiguration is another technique that can circumvent physical defects by first mapping defects on reconfigurable fabrics then synthesising a feasible configuration to realise an

application for each nanofabric instance [9–11]. However, defect mapping and reconfiguration is performed on a per chip basis which poses a scalability challenge. The prohibitively low reliability of these new nanodevices dictate that they must be interfaced with complementary metal-oxide semiconductor (CMOS) circuits to tolerate the inevitable high fault rates. This leads to a new paradigm of hybrid CMOS/nanoarchitecture [12–15] to perform reliable computing using unreliable components (nanodevices). In this architecture, nanoscale devices offer a highly dense fabric for data storage and computation, whereas CMOS components are utilised for interfacing and for highly critical circuit operations. Both CMOS circuits and high fault rates will reduce the net density delivered by these nanodevices.

Recently, error correcting codes (ECCs) have been proposed as a promising approach to improve the reliability and yield of heterogeneous CMOS/nanodevices systems. In [6, 12], ECCs were mainly used for the suppression of soft errors rather than physical defects that is maintaining the fault-tolerance level rather than enhancing defect tolerance. In [12], authors suggested a hybrid fault-tolerance technique based on Hamming code and reconfiguration. In [6], the authors proposed an implementation of ECCs based on the theory of Markov

random fields (MRF) to combat soft faults thus increasing the reliability of hybrid systems. In [13], two nanoelectronic memory fault-tolerant system design approaches based on Bose–Chaudhuri–Hocquenghem (BCH) codes were suggested. Previously, single ECCs such as Hamming and BCH have been used in the context of reliable memory designs [12, 16]. In [16], the authors explored combining error correction codes with various repair techniques to combat the high defect rates in hybrid CMOS/nanofabrics with particular focus on memory architectures. The previous works have only addressed fault tolerance in memory architectures. ECC-based techniques can also be applied to memory-based implementation of logic circuits [i.e. look-up tables (LUTs)] which includes Don't Care Conditions (DCCs). The presence of DCCs in Boolean logic functions presents a strong case to apply these techniques to circuits implemented as LUTs on CMOS/nanodevice fabrics. As we will demonstrate in this work, the existence of DCCs can be exploited in this type of architecture since it helps in masking of erroneous bits which is not possible in memory design.

Fig. 1 gives an overview of the targeted hybrid CMOS/nano architecture. The proposed architecture is technology-independent that is the nanoscale fabric is built using any of the recently proposed nanodevices including carbon nanotubes (CNT) or silicon nano-wires (NW). The techniques proposed in this work target CMOS/nano computational architecture incorporating a LUT implementation of logic functions, as outlined in [17]. LUT implementation is an effective functional-coding approach that provides low-level protection of individual Boolean functions [18, 19]. In our experiments, the LUTs under test are represented by randomly generated symmetric matrices of sizes ranging from $2^3 \times 3$ to $2^6 \times 6$ where the probability of 0 and 1 are equal. The errors are injected randomly in the nanofabric causing the corresponding bits to change their values (i.e. $1 \rightarrow 0$ or $0 \rightarrow 1$). We have assumed random distribution of errors to simulate the worst-case scenario as correlated errors are technology specific. The proposed fault-tolerant techniques are based on ECC and partial redundancy to address the permanent and transient faults in nanoscale LUTs. In the first technique, we implement a two-dimensional coding scheme using Hamming and BCH codes to address both hard and soft errors in the presence of high fault rates. In the second technique, we target the high physical defect

rates in the nanofabric by integrating ECCs with bad line exclusion technique. In this technique, the high bit density offered by nanodevices is exploited to provide the necessary spare rows to compensate for the defective lines. Although the exact manufacturing defect rate and transient error rate are not yet pinpointed, it is believed that they will easily exceed 10% [2]. The authors in [17] assume small fault rates (less than 10%) in nanofabrics for small LUT sizes with 50% of LUT entries set as DCCs. We first investigate the effectiveness of our proposed techniques over the methods [12, 17] in high defect rate scenario.

2 Primitives

We first examine the ineffectiveness of using single ECCs such as Hamming and BCH in the presence of high error rates for different LUT sizes.

2.1 Hamming

Hamming is a single-error-correcting and double-error-detecting code that is the code is capable of correcting one error and detecting two errors in a codeword. A typical Hamming code is $(2^m - 1, 2^m - m - 1)$, in other words, for $2^m - m - 1$ data bits, m parity bits need to be added for full protection.

To demonstrate the reliability improvement that can be achieved from the techniques discussed in this work, we have performed experiments on randomly generated symmetric LUTs where the probability of 0 and 1 is equal. The LUTs are of sizes ranging from $2^3 \times 3$ (three inputs, three outputs) to $2^6 \times 6$ (six inputs, six outputs). The circuit failure probability F_m , resulting from randomly injecting m errors, is obtained by calculating the ratio of defective LUTs after decoding to the total number simulation iterations $I = 5000$. In this work, we assume that a nanodevice is subject to both stuck-open and stuck-short defects with equal probabilities. We also assume that errors are uniformly distributed across the fabric where both physical and transient errors are random and statistically independent. For comparison purposes, we use the simple Hamming code [17] as a reference point for the evaluation of our proposed techniques.

Theoretically, the probability of a row of length l having m defective bits is given by the following binomial equation

$$P(m) = \binom{l}{m} P^m (1 - P)^{l-m} \quad (1)$$

where P is the error rate of the fabric. The probability that the Hamming decoder fails to correctly decode an erroneous codeword is equal to the probability of having more than one error per row. Using (1), this is given by the

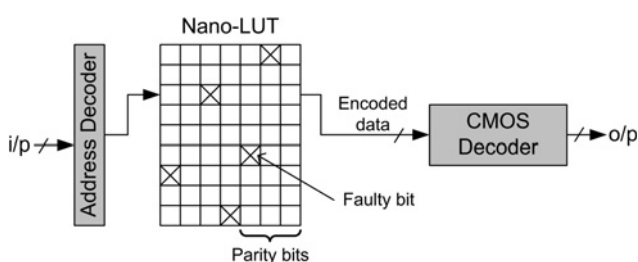


Figure 1 Hybrid Nano/CMOS architecture overview

following equation

$$P_{\text{row}} = \sum_{k=2}^{r+r_{\text{par}}} \binom{r+r_{\text{par}}}{k} P^k (1-P)^{r+r_{\text{par}}-k} \quad (2)$$

where r and r_{par} are the number of bits and number of parity bits in a row, respectively. The failure rate of a LUT with c columns is equal to the probability that at least one row is defective and it can be computed as

$$P_{\text{failure}} = \sum_{k=1}^c \binom{c}{k} P_{\text{row}}^k (1 - P_{\text{row}})^{c-k} \quad (3)$$

In the case of $2^4 \times 4$ LUT, (2) and (3) can be rewritten as

$$P_{\text{row}} = \sum_{k=2}^{4+3} \binom{4+3}{k} P^k (1-P)^{4+3-k}$$

$$P_{\text{failure}} = \sum_{k=1}^{16} \binom{16}{k} P_{\text{row}}^k (1 - P_{\text{row}})^{16-k}$$

Fig. 2 illustrates the failure rate obtained both theoretically [using (2) and (3)] and through the above simulation procedure. As can be seen, the two graphs are almost identical, validating the derived theoretical equations.

Fig. 3 shows the variation of failure rate with respect to several percentages of injected error rates for different LUT sizes. For defect rates as small as 1%, the Hamming code perfectly detects and corrects all faults for LUTs of sizes smaller or equal to $2^4 \times 4$ as reported in [17]. However, it can be seen that even for a small LUT size of $2^3 \times 3$, and error rate greater than 5%, more than 10% of circuits fail. We can also observe that as the LUT size increases, the fault tolerance of the circuit falls more rapidly indicating the inefficiency of this scheme.

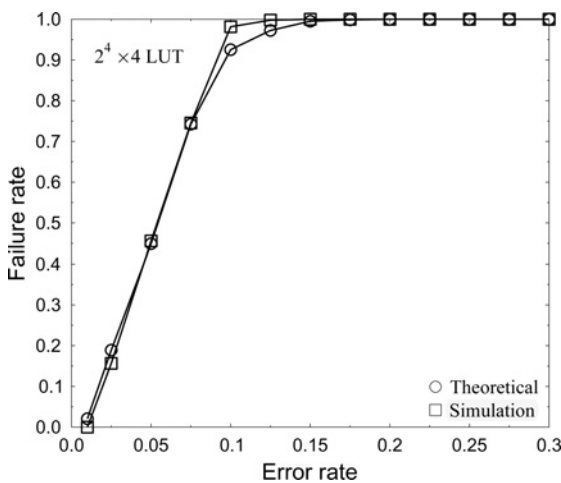


Figure 2 Hamming – failure rate obtained through simulation and theory

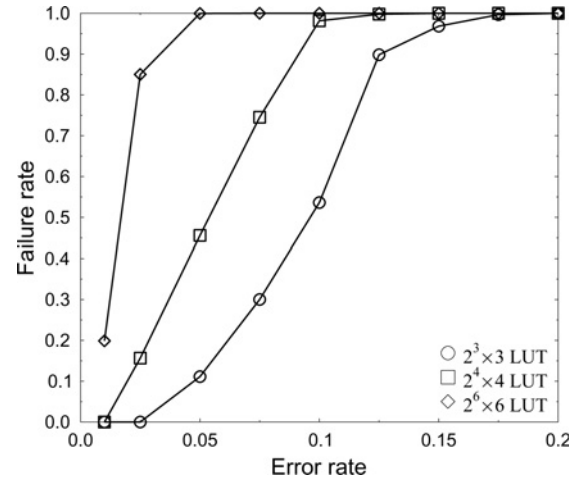


Figure 3 Hamming – effect of varying LUT size on failure rate

2.2 Bose–Chaudhuri–Hocquenghem

In this section, we examine fault-tolerant approach using stronger BCH coding and evaluate its efficiency in dealing with the high defect probabilities in nanoscale LUTs. BCH is a multilevel, variable length and easy to decode ECC used to correct multiple random errors in a codeword. The simplest form of BCH codes is the single error-correcting BCH(7,4,1) which is equivalent to Hamming code. We first examine BCH [13] with 0% DCCs using BCH(15,7,2), which adds eight parity bits in order to detect and correct two errors in the codeword. The required word length is 7 bits; however, the size of each entry in our LUT is only 4-bits in the case of $2^4 \times 4$ LUT. Therefore we need to pad the data bits with the necessary bits so that it is equal to the required data word size, as shown in Fig. 4.

The graph shown in Fig. 5 represents the failure rate achieved using the 2-bit error-correction BCH and in the absence of DCCs. The fault tolerance obtained from the simulation results revealed a performance very similar to Hamming. The reason behind this is that even though the BCH(15,7,2) code can tolerate more errors than single error correction techniques, there is a higher probability of errors per codeword in LUTs. The padded bits and the redundant bits added to the data word doubles the probability of faults in each entry of the LUT. The codeword is 15-bit long for the 2-bit error-correction BCH which is twice longer than the 7-bit long codeword for the single error correcting Hamming code. Hence, strong ECCs have the capability of tolerating more errors at the cost of more parity bits added to the codeword, which in turn makes them more vulnerable to higher fault rates; and hence a rapid drop in their efficiency as the fault rate increases in the LUTs.

Instead of coding row entries in LUTs, we use stronger ECCs such as BCH codes to encode columns.

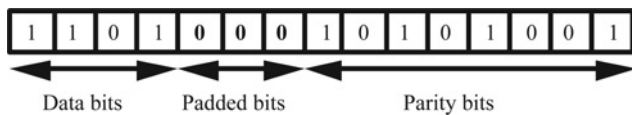


Figure 4 BCH – padding

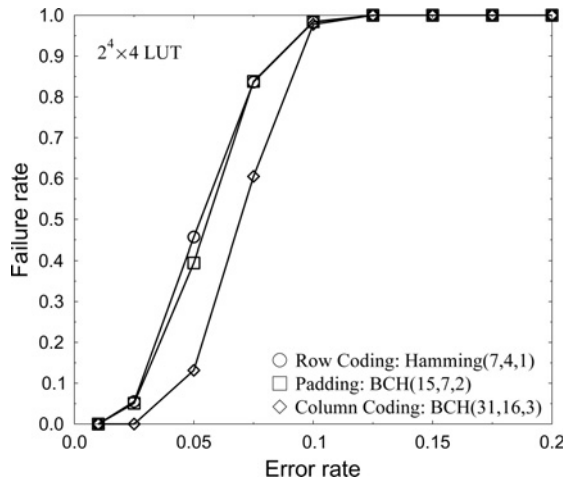


Figure 5 Failure rate comparison between different 1-D coding techniques

BCH(31,16,3) for example can detect and correct three errors per column, but at the cost of adding 15 parity bits. The results obtained from simulations are shown in Fig. 5. For low error rates, BCH exhibits a better performance than Hamming. At 5% of errors, we can notice a 70% improvement in failure rate over Hamming. However, when errors exceed 10%, this coding technique completely fails.

3 Proposed fault-tolerance techniques

In this section, we present two different hybrid fault-tolerant techniques to address the high fault rates in nano-LUTs. As we have seen earlier, single error correction schemes prove to be inefficient in dealing with such high error rates. The first proposed approach combines Hamming and BCH to target higher error rates (greater than 5%), whereas the second technique combines Hamming with bad line exclusion to address error rates as high as 20%.

3.1 Combined two-dimensional coding: technique 1

To reduce the failure rate for fault rates exceeding 5%, we implemented a two-dimensional coding technique based on Hamming and BCH codes (also known as product code [20, 21]). The idea is to encode both rows and columns in LUTs as shown in Fig. 6a. In [21], the authors proposed a Hamming-based two-dimensional coding scheme to tolerate the occurrence of random and burst errors in on-chip interconnects. The single error correction Hamming

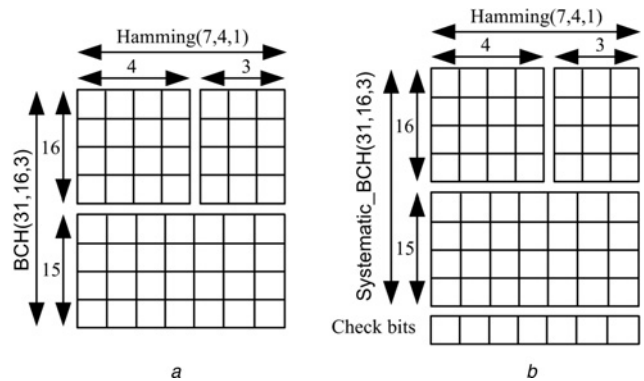


Figure 6 2D coding for $2^4 \times 4$ LUT

a Hamming & BCH

b Hamming and systematic BCH with check bits

In the row encoding, Hamming(7,4,1) is used to detect and correct one error per row by adding three parity bits. BCH(31,16,3) is used to encode columns can detect and correct three errors per column at a cost of adding 15 parity bits

code is used for both row and column encoding. In our proposed technique Hamming code is used to encode data bits in each row of the LUT, and then a stronger BCH code is used to encode each column. The choice of BCH for column encoding is due to its ability to tolerate more errors in a codeword and given the size of columns which is 2^N , this choice seems appropriate. Retrieving data from the encoded LUT comprises of two decoding steps. In the first step, columns are first decoded using the BCH decoder. This step will allow the detection and correction of the biggest portion of errors because of the capability of the BCH decoder to correct more errors in the codeword than the Hamming decoder. Then, in the second decoding step, the Hamming decoder is used to remove the remaining faults.

We can further improve the fault tolerance of this technique by using systematic BCH code along with check bits as illustrated in Fig. 6b. In systematic block codes, data bits remain unchanged in the codeword, and the parity bits are attached to the end of the data bit sequence. We exploit the fact that the number of 1's in any wrong decoded word will most probably be different from the number of 1's in the expected correct word. Therefore we use check bits to store the number of 1's of each column after all row entries of LUT are encoded using Hamming. If the number of faults per column exceeds the error correcting capability limit of systematic BCH, the BCH decoder will generate the wrong output and hence cause the entire technique to fail. Therefore to avoid failure the check bits are always compared with the number of 1's of the output of BCH decoder, if they are not equal, the codeword remains unchanged and all the faults in the first 16 bits of the corresponding column are left to the second iteration of decoding to be corrected using the Hamming decoder. The flowchart in Fig. 7 illustrates the decoding process to retrieve row M from a LUT of size $2^N \times N$.

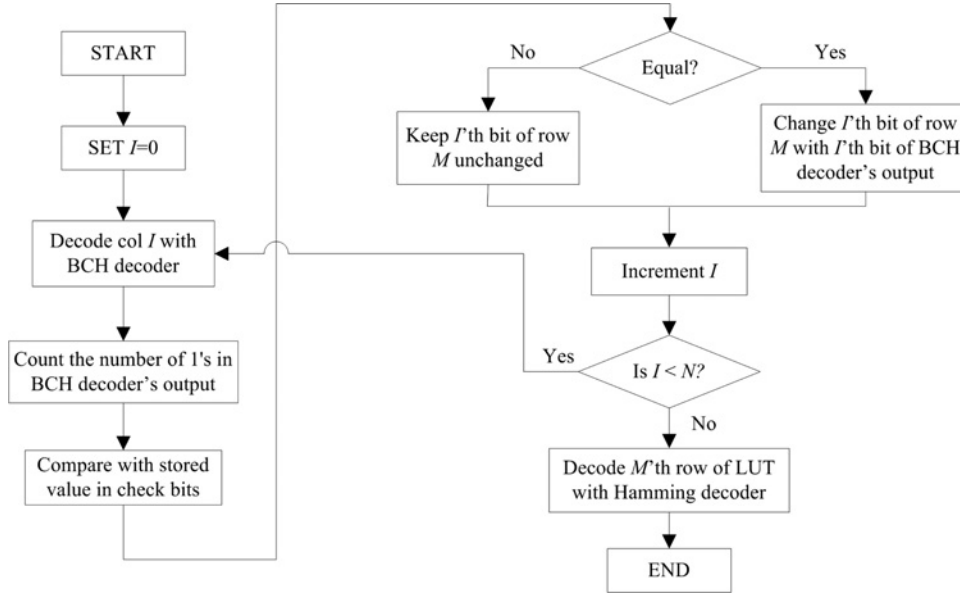


Figure 7 Flowchart to illustrate the multiple-step decoding process of the proposed 2D coding technique

Assuming the same error probability P for each bit, the probability of having m defective bits in a column of length $(c + c_{\text{par}})$ follows the binomial distribution given in (1). Therefore the probability that the BCH decoder fails to correct a column because the number of faults exceeds its correction capability bch_err is given by

$$P_{\text{col}} = \sum_{k=\text{bch_err}+1}^{c+c_{\text{par}}} \binom{c+c_{\text{par}}}{k} P^k (1-P)^{c+c_{\text{par}}-k} \quad (4)$$

where c and c_{par} are the number of bits and number of parity bits in a column, respectively.

The BCH correction of columns reduces the probability of a bit being erroneous by a factor of P_{col} . Therefore the remaining faults that are randomly distributed over the rows will have a new error probability P_{new} that is given by the following equation

$$P_{\text{new}} = P \times P_{\text{col}} \quad (5)$$

Using this new error rate, the failure rate for each row after Hamming decoding is obtained using (2), as follows

$$P_{\text{row}} = \sum_{k=2}^{r+r_{\text{par}}} \binom{r+r_{\text{par}}}{k} P_{\text{new}}^k (1-P_{\text{new}})^{r+r_{\text{par}}-k} \quad (6)$$

Hence, the final failure probability of our combined 2D coding technique is computed as

$$P_{\text{failure}} = 1 - (1 - P_{\text{row}})^r \quad (7)$$

For the example used in Fig. 6, (4), (6) and (7) reduce to

$$P_{\text{col}} = \sum_{k=3+1}^{16+15} \binom{16+15}{k} P^k (1-P)^{16+15-k}$$

$$P_{\text{row}} = \sum_{k=2}^{4+3} \binom{4+3}{k} P_{\text{new}}^k (1-P_{\text{new}})^{4+3-k}$$

$$P_{\text{failure}} = 1 - (1 - P_{\text{row}})^{16}$$

Fig. 8 shows the failure rate obtained both theoretically (P_{failure}) and through simulation in the case of $2^4 \times 4$ LUT. As we can see, there is an excellent correlation between experimental and theoretical results.

The plots for simulation results of the circuit failure rate for Hamming(7,4,1), BCH(31,16,3) and 2D coding

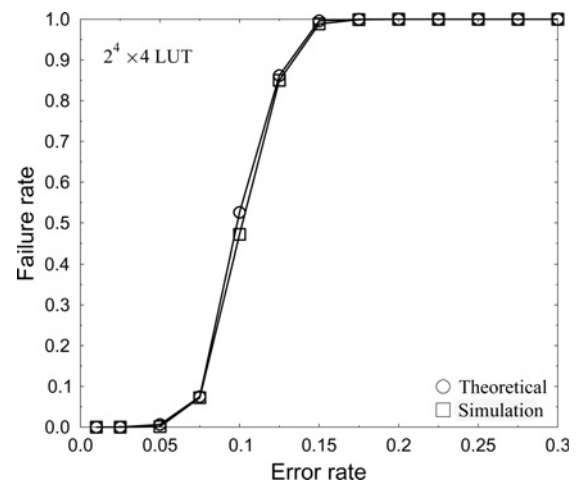


Figure 8 2D coding – failure rate obtained through simulation and theory

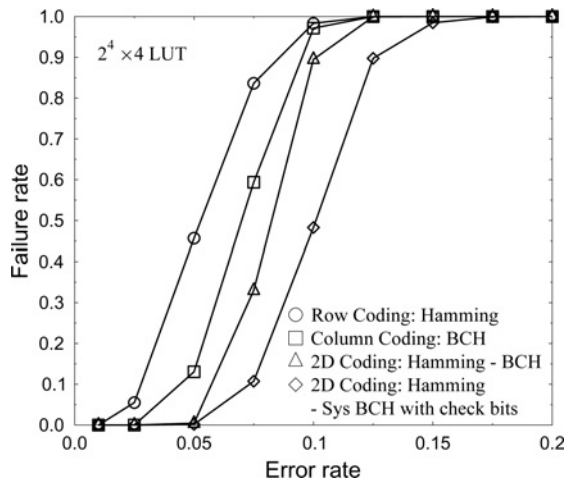


Figure 9 Failure rate comparison between 1D and 2D coding techniques

techniques are shown in Fig. 9. It can be seen that for error rates smaller than 15% 2D coding technique (without check bits) exhibit better tolerance than both Hamming and BCH. For example, when the percentage of injected faults is 5%, 2D coding perfectly detects and corrects all injected faults, whereas Hamming code achieves a failure rate of approximately 45%. However, as the fault rate increases beyond 15%, this technique completely fails. This improvement in reliability is achieved at the cost of a higher number of parity bits which will result in additional area and energy overhead.

Fig. 9 also illustrates the enhancement achieved in fault tolerance by incorporating the check bits into the technique. 2D coding with check bits achieves significantly lower failure rates for error rates greater than 5% and up to 10% as compared to basic 2D coding technique resulting in an improvement of 37% in fault tolerance. As can be seen, using check bits will improve the fault tolerance of our combined 2D coding technique. However, the need to store the number of 1's in a highly reliable memory (i.e. store at most approximately $\lceil \log_2(\text{num}_{\text{rows}}) \times \text{num}_{\text{columns}} \rceil$ bits in a CMOS memory) will incur an extra area and delay overhead that need to be evaluated based on practical IC designs as explained in Section 4.

Next we examine the effect of varying LUT size on circuit reliability. As can be seen in Fig. 10, the failure rate increases rapidly for bigger LUTs. For a fault density of 10%, the failure probability to successfully instantiate a LUT on the defective nanofabric increases from 5% in the case of $2^3 \times 3$ LUT to complete failure for $2^6 \times 6$ LUT. Comparing the results of Fig. 10 with Fig. 3, we can observe that combined 2D coding technique outperforms one-dimensional coding in terms of fault tolerance despite its insufficiency in coping with fault rates higher than 10% and bigger LUT sizes. In the case of $2^6 \times 6$ LUT, we can observe a nearly 0% failure rate at 5% error rate, whereas Hamming completely fails.

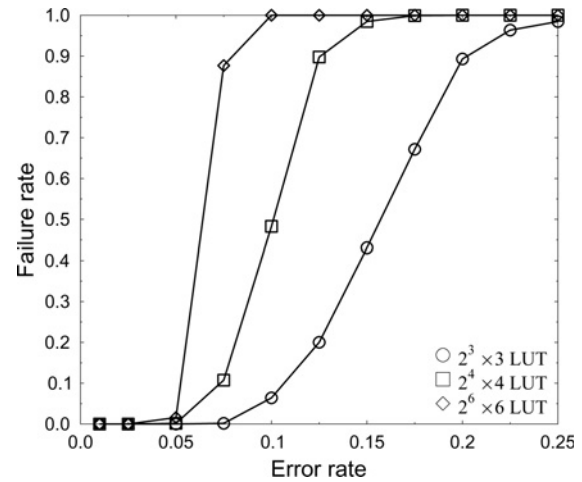


Figure 10 2D coding – effect of varying LUT size on failure rate

Boolean functions are defined by their On-set, OFF-set and their DCC-set [17]. If an entry in a LUT is a DCC, the output can be either 0 or 1. We examine the impact of DCCs on our 2D coding technique. In order to theoretically calculate the circuit failure rate given the percentage of DCCs in the LUT, we first need to calculate the failure probability of the BCH decoder and the new error rate after decoding which are given by (4) and (5). After column decoding, the probability that the Hamming decoder fails to correctly detect and correct all errors does not only depend on the number of faults per each row, but also on the number of erroneous bits in the output of the decoder. Therefore the probability that a given number of bits are erroneous in the output of the decoder, denoted as $P_{(n)\text{bit}}$, assuming a number of errors in the codeword has to be estimated where n is smaller or equal to the number of bits of the decoded word. For a $2^4 \times 4$ LUT, the values of $P_{(n)\text{bit}}$ are shown in Table 1.

The failure rate of correctly decoding a row using the Hamming decoder is obtained using the following equation

$$P'_{\text{row}} = \sum_{k=2}^{r+r_{\text{par}}} \left[\binom{r+r_{\text{par}}}{k} P_{\text{new}}^k (1 - P_{\text{new}})^{r+r_{\text{par}}-k} \right] \times \sum_{n=1}^r \left[P_{(n)\text{bit}} \sum_{m=1}^n \binom{n}{m} (1 - P_{\text{DCC}})^m P_{\text{DCC}}^{n-m} \right] \quad (8)$$

Table 1 Hamming decoder: distribution of erroneous bits in the output word – $2^4 \times 4$ LUT

$P_{(n)\text{bit}}$	Number of errors in a codeword					
	2	3	4	5	6	7
$P_{1\text{bit}}$	0.4306	0.1952	0.3639	0.1415	0	0
$P_{2\text{bit}}$	0.4287	0.4274	0.4335	0.4271	0	0
$P_{3\text{bit}}$	0.1407	0.3774	0.2026	0.4314	0	0
$P_{4\text{bit}}$	0	0	0	0	1	1

The total failure probability is computed as

$$P_{\text{failure}} = 1 - (1 - P'_{\text{row}})^r \quad (9)$$

Fig. 11 presents the failure rate obtained theoretically, based on the previously outlined equations, and using the simulation procedure in the presence of 50% DCCs. The two graphs perfectly match each other which validates our theoretical predictions.

Fig. 12 shows the results obtained before and after injecting 50% of Don't Cares in $2^4 \times 4$ LUTs. As can be observed, the optimum improvement is recorded at 10% error rate where the failure rate is reduced from nearly 50 to 37%.

3.2 Hamming with bad line exclusion: technique 2

ECCs are usually preserved for the suppression of transient faults to enhance computational reliability. However, the high defect densities in future nano-circuits [2] dictate involving coding techniques at the initial repair of hard errors to minimise the required amount of redundant resources. As can be seen from Fig. 9, ECCs alone are not able to address the issue of high defect rates induced during manufacturing. Hence it is imperative to use ECCs in conjunction with other techniques in order to detect and correct larger portions of physical defects (up to 20%).

To deal with higher defect rates, we have combined Hamming code with bad line exclusion technique. This technique requires allocating enough redundant rows for each LUT to be repaired. The use of redundant wires to tolerate physical defects was presented in [16, 22]. As will be shown, the amount of spare rows depends on two main factors: the defect rate of the fabric and the size of the LUT. Repair consists of two phases: a must-repair phase and a final-repair phase. In the must-repair phase, line

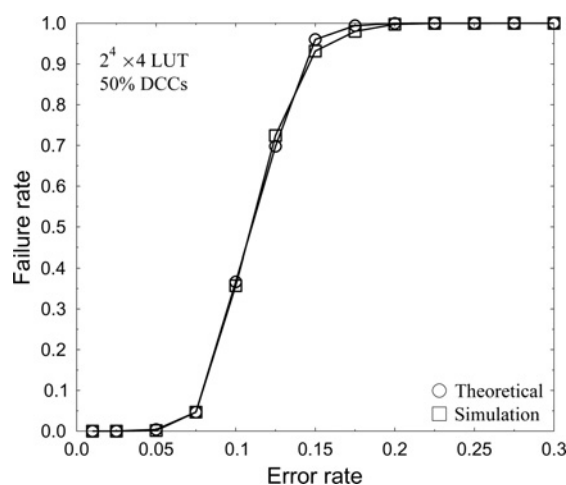


Figure 11 2D coding – failure rate obtained through theory and simulation in the presence of 50% DCCs

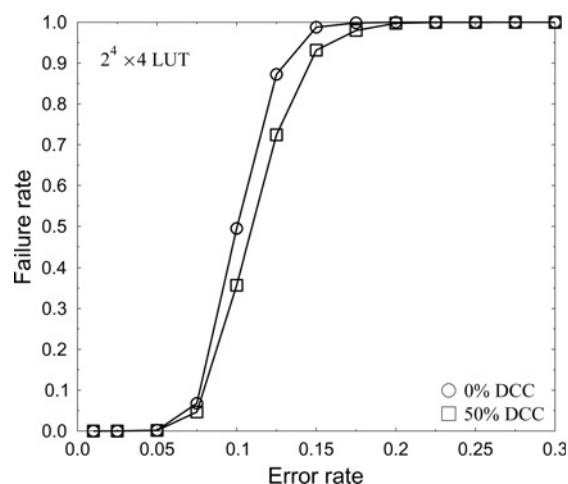


Figure 12 2D coding – effect of Don't Cares on failure rate

exclusion is applied only in one dimension where the defective rows are excluded and replaced with spare rows if the number of defects per row exceeds the correction capability of Hamming. Therefore defect counters for the faulty rows are required during the initial testing process of the nanofabric. It is worth mentioning that the configuration process is performed only once during the manufacturing phase (i.e. off-line configuration) and hence the area of the configuration logic is not added to the CMOS area overhead of this technique. In the final-repair phase, the Hamming decoder detects and corrects the remaining defects as shown in Fig. 13. During the initial analysis of the fabric, the bad rows are detected and their physical address is used to create a special table to map the continuous logical address to the actual physical location of defect-free rows. Such a mapping table has to be stored in a highly reliable memory implemented in CMOS. The physical implementation of this logical-to-physical mapping table is beyond the scope of this work and is not included in the area overhead estimation of this technique in the next section.

To obtain the probability of failure, we first calculate the probability P_{row} of a row of length l being excluded. P_{row} is

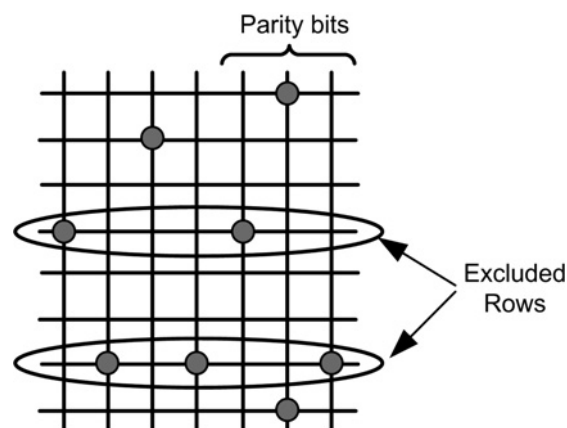


Figure 13 Hamming with bad line exclusion

equal to the probability of having more than one bad bit in that row given by

$$P_{\text{row}} = \sum_{k=2}^l \binom{l}{k} P^k (1-P)^{l-k} \quad (10)$$

where l is the number of bits in a row including the parity bits and P is defect probability of each bit. Therefore the probability of failure to instantiate a LUT on the fabric given the original number of rows r and upper limit of spare rows r_{sp} can be computed as

$$P_{\text{failure}} = \sum_{k=r_{\text{sp}}+1}^{r+r_{\text{sp}}} \binom{r+r_{\text{sp}}}{k} P_{\text{row}}^k (1-P_{\text{row}})^{r+r_{\text{sp}}-k} \quad (11)$$

In the case of $2^4 \times 4$ LUT and 25% spare rows, (8) and (9) can be rewritten as

$$P_{\text{row}} = \sum_{k=2}^7 \binom{7}{k} P^k (1-P)^{7-k}$$

$$P_{\text{failure}} = \sum_{k=4+1}^{16+4} \binom{16+4}{k} P_{\text{row}}^k (1-P_{\text{row}})^{16+4-k}$$

Fig. 14 shows the failure rates for a $2^4 \times 4$ LUT for different error rates in the presence of various amounts of spare rows. As can be seen, this technique is capable of tolerating an unprecedented percentage of defects when compared to the Hamming alone and combined 2D coding technique shown in Fig. 9. This is demonstrated by a failure rate of nearly 0% for up to 20% of injected faults into randomly generated $2^4 \times 4$ LUT.

To further our analysis, we examine the effect of variation in number of spare rows on the failure rate for different LUT sizes. Fig. 15 demonstrates that as the error rate increases,

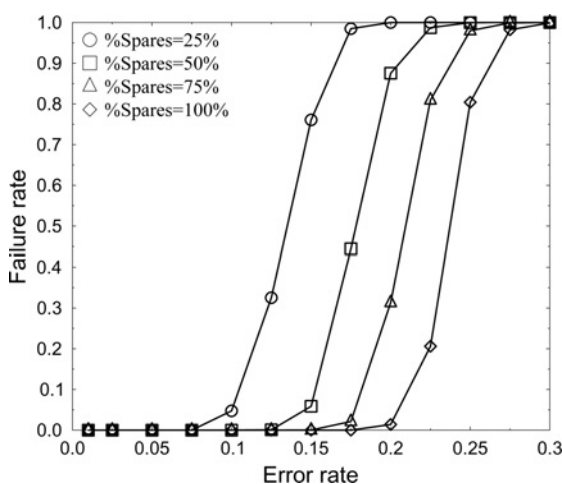


Figure 14 Hamming and bad line exclusion – failure rate obtained for different percentages of spare rows – $2^4 \times 4$ LUT

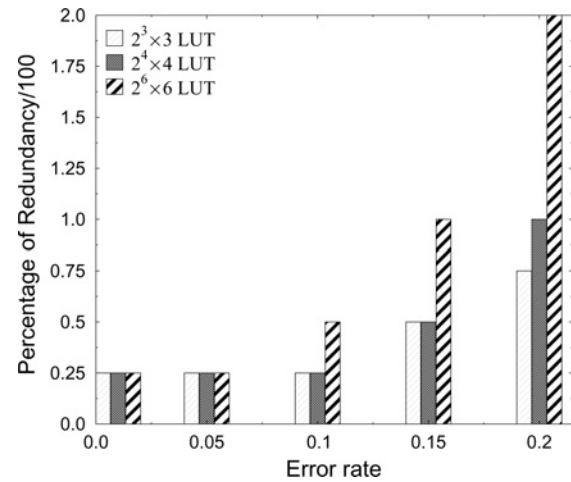


Figure 15 Percentage of redundancy needed to achieve 0% failure rate for different LUT sizes

more spares are needed to keep the level of fault tolerance close to 0%. In the case of $2^4 \times 4$ LUT for instance, only 25% of spare rows are needed that is four more rows, to completely tolerate up to 10% of faults rates. And as more errors are injected into the LUT, more spares should be allocated and hence decreasing the useful bit density of the fabric. It can also be observed that as the LUT size increases, the percentage of spares also increases to achieve low failure rates. For example, $2^6 \times 6$ LUT requires twice its original size to tolerate 20% defect rate. However, we can minimise such high redundancy by adopting a more powerful ECC such as BCH instead of Hamming.

While the authors in [17] assumed 50% DCCs, our results have shown remarkable improvement in fault tolerance even with 0% DCCs in the LUT implementation. However, it can be seen from Fig. 16 that the fault tolerance of this technique is significantly improved when we assume some

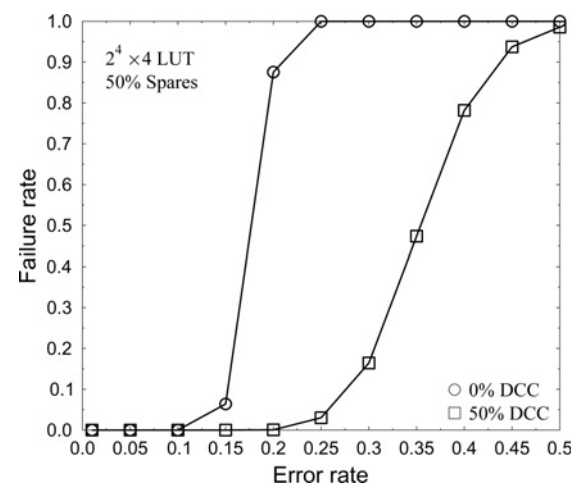


Figure 16 Hamming and bad line exclusion – variation of failure rate in the presence of Don't Care entries

Inclusion of 50% DCCs improves fault tolerance by almost two times

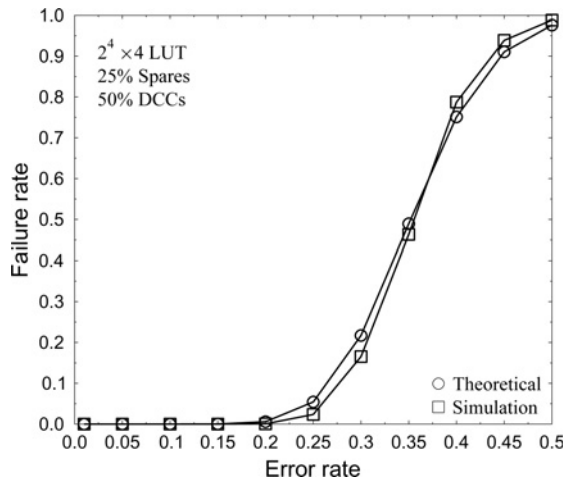


Figure 17 Hamming and bad line exclusion – failure rate obtained through theory and simulation in the presence of 25% spares and 50% DCCs

of the entries in the implementation as DCCs as compared to our results in Fig. 14.

The existence of DCCs (P_{DCC}) in LUTs significantly reduces the bit failure rate as outlined in the following equation

$$P' = P \times (1 - P_{DCC}) \quad (12)$$

The new probability of a row being excluded after injecting Don't Cares is obtained by replacing the fault rate P with the new fault rate P' in (8) as follows

$$P'_{\text{row}} = \sum_{k=2}^l \binom{l}{k} P'^k (1 - P')^{l-k} \quad (13)$$

Fig. 17 illustrates the failure rate obtained both theoretically using (11) and (13) and through simulations. As can be observed, there is an identical match between the two graphs indicating the correctness of our derived mathematical equations.

4 Implementation

The realisation of fault tolerance in nano/CMOS nanoelectronic architecture will incur area, energy and operational latency overhead in CMOS domain [15]. Such overhead must be taken into account when investigating and evaluating hybrid CMOS/nanodevice fault-tolerant architectures. The high reliability decoders are implemented in CMOS and therefore incur an increase in the area and energy consumption compared to the denser and low-energy nano-LUTs. Additional clock cycles are also lost in decoding and correcting codewords which cause latency overhead.

In order to obtain an estimate of the area, latency and energy overheads per codeword, the corresponding decoders were designed in VHDL and thoroughly tested through simulation using the appropriate test benches. Both Hamming and BCH

decoders are serial that is they receive 1-bit input and generate a 1-bit output per clock cycle, therefore the decoding latency is proportional to the codeword length. Using the 0.12 μm CMOS standard cell library, the area overhead of the Hamming decoder is 906 μm^2 and decoding one 7-bit long codeword requires 13 clock cycles and an energy overhead of approximately 9 pJ/MHz. The BCH decoder incurs higher overhead because of its high complexity: an area overhead of 9132 μm^2 , a latency of 69 clock cycles and an energy overhead of 286 pJ/MHz. However, this needs to be further studied in the context of optimum design strategy where we negotiate the advantage of higher density and low-power dissipation against increased delay overhead.

The CMOS components as well as the redundant parity bits and spare rows in our techniques will reduce the useful bit density offered by nanodevices. Therefore another design parameter called area per useful bit ratio is used to compare the efficiency of the various techniques. Area per useful bit (a) reflects the area necessary to achieve certain useful bit capacity and is obtained by dividing the total area of the fabric by the number of useful bits in the LUT.

The total area of the fabric comprises of the area of nanodevices and that of CMOS subsystems. We adopt a model presented in [14] to estimate the area of nanoscale memory. Each bank in the memory is composed of a set of crossed nanoscale wires supported by a set of interface microscale wires. For a nanocircuit of 2^n inputs and m outputs, the area can be estimated as shown in [17]

$$A = (W_{\text{lith}}(n + \log_2 m) + W_{\text{nano}}2^n) \cdot (W_{\text{lith}}n + mW_{\text{nano}}2^n) \quad (14)$$

The main parameters in the model are the number of rows 2^n and columns m . The area of the nanomemory is dominated by the address lines which are microwires. $W_{\text{lith}} = 105$ nm is the wire pitch of the lithographic address wires and $W_{\text{nano}} = 10$ nm is the pitch for the nanoscale wires. For instance, in the case of a $2^4 \times 4$ LUT, the area of the nano-LUT before encoding is 0.84 μm^2 .

Table 2 compares the area overhead of the proposed technique with the earlier approaches. Although we have seen that Hamming in combination with bad line exclusion achieves much better failure rates as compared to error correcting schemes such as Hamming or BCH, this improvement in failure rate is achieved with little or no increase in area overhead when compared with the simplest correction technique proposed in [17]. It should be noted that BCH has the highest area overhead because of the complexity of its decoding circuitry. Table 2 also shows the area per useful bit ratio for a $2^4 \times 4$ LUT implemented using the three techniques.

Although significant area improvement can be achieved over current CMOS for high density fabrics using hybrid

Table 2 Area/useful bit of the proposed techniques

Technique	Total area (μm^2)	Area/useful bit ($\mu\text{m}^2/\text{bit}$)
Hamming alone [17]	907	14
BCH alone [13]	9133	142
Hamming and bad line excl. (proposed)	908	14

nano/CMOS architecture [13], it can be shown that further improvement in terms of useful bit density can be achieved by sharing the decoders by multiple LUTs using time multiplexing strategy as outlined in [17]. Another way to minimise the CMOS area overhead is to synthesise logical circuits into smaller LUTs because the size of the decoder increases proportionally with the size of the LUT. Moreover, as shown in the experimental results (Figs. 10 and 15), using smaller LUTs allows achieving higher levels of fault tolerance at the cost of low area overhead.

5 Conclusion

In this paper we investigated a promising LUT-based implementation of Boolean logic functions in heterogeneous CMOS/nanodevice architectures. Our studies showed that single ECCs such as Hamming or BCH prove their insufficiency in tolerating high error rates. We presented two hybrid fault-tolerance techniques that address faults caused because of physical defects and transient faults. In the first technique, we encoded both rows and columns of LUTs (using Hamming and BCH codes, respectively) to target higher number of faults. This technique significantly improves the fault tolerance with respect to single error correction schemes for error rates greater than 5%. In the second technique, we complemented Hamming with bad line exclusion. This technique results in remarkable improvement of failure rate against a substantial fraction of bad nanodevices (up to 20%). This is achieved at the cost of minimal increase in area overhead compared with Hamming, yet with much higher efficiency in tolerating errors. Based on our studies this technique is very effective for LUT-based Boolean logic architectures. We have also shown that the presence of DCCs in LUTs can significantly improve circuit failure rates when we combine coding with bad line exclusion. Finally, we investigated the impact of these techniques in terms of area, latency and energy overheads and showed that improved fault tolerance can be achieved using the proposed techniques with little overheads compared to previous coding techniques.

6 Acknowledgments

The authors would like to acknowledge the EPSRC (UK) for funding this project in part under grant EP/E035965/1 as well as the Algerian Ministry of Higher Education and Scientific Research.

7 References

- [1] BROWN J.G., BLANTON R.D.: 'A built-in self-test and diagnosis strategy for chemically assembled electronic nanotechnology', *J. Electron. Test.*, 2007, **23**, (2–3), pp. 131–144
- [2] WANG Z., CHAKRABARTY K.: 'Built-in self-test and defect tolerance in molecular electronics-based nanofabrics', *J. Electron. Test.*, 2007, **23**, (2–3), pp. 145–161
- [3] BAHAR R.I.: 'Trends and future directions in nano structure based computing and fabrication'. Int. Conf. on Computer Design, October 2006, pp. 522–527
- [4] JACORNE M., HE C., DE VECIANA G., BIJANSKY S.: 'Defect tolerant probabilistic design paradigm for nanotechnologies'. Proc. 41st, DAC'04, 2004, pp. 596–601
- [5] ZHAO C., DEY S., BAI X.: 'Soft-spot analysis: targeting compound noise effects in nanometer circuits', *IEEE Des. Test Comput.*, 2005, **22**, (4), pp. 362–375
- [6] NEPAL K., BAHAR R.I., MUNDY J., PATTERSON W., ZASLAVSKY A.: 'MRF reinforcer: a probabilistic element for space redundancy in nanoscale circuits', *IEEE Micro.*, 2006, **26**, (5), pp. 19–27
- [7] NIKOLICK, SADEKA, FORSHAW M.: 'Fault-tolerant techniques for nanocomputers', *Nanotechniques*, 2002, **13**, (3), pp. 357–362
- [8] THAKER D., AMIRTHARAJAH R., IMPENS F., CHUANG I., CHONG F.: 'Recursive TMR: scaling fault tolerance in the nanoscale era', *IEEE Des. Test Comput.*, 2005, **22**, (4), pp. 298–305
- [9] MISHRA M., GOLDSTEIN S.: 'Defect tolerance at the end of the roadmap', *ITC*, 2003, **1**, pp. 1201–1210
- [10] HE C., JACOME M., DE VECIANA G.: 'A reconfiguration-based defect-tolerant design paradigm for nanotechnologies', *IEEE Des. Test.*, 2005, **22**, (4), pp. 316–326
- [11] COPEN GOLDSTEIN S., BUDIU M.: 'NanoFabrics: spatial computing using molecular electronics'. IEEE ISCA, 2001, pp. 178–189
- [12] JEFFERY C., BASAGALAR A., FIGUEIREDO R.: 'Dynamic sparing and error correction techniques for fault tolerance in nanoscale memory structures'. Fourth IEEE Conf. on Nanotechnology, 2004, August 2004, pp. 168–170
- [13] SUN F., ZHANG T.: 'Defect and transient fault-tolerant system design for hybrid CMOS/nanodevice digital memories', *Nanotechniques*, 2007, **6**, (3), pp. 341–351
- [14] DEHON A., GOLDSTEIN S., KUEKES P., LINCOLN P.: 'Nonphotolithographic nanoscale memory density prospects', *IEEE Trans. Nanotechnol.*, 2005, **4**, (2), pp. 215–228

- [15] BAHAR R.I., HAMMERSTROM D., HARLOW J., ET AL.: 'Architectures for silicon nanoelectronics and beyond', *Computer*, 2007, **40**, (1), pp. 25–33
- [16] STRUKOV D.B., LIKHAREV K.K.: 'Prospects for terabit-scale nanoelectronic memories', *Nanotechniques*, 2005, **16**, (1), pp. 137–148
- [17] SINGH A., ZEINEDDINE H., AZIZ A., VISHWANATH S., ORSHANSKY M.: 'A heterogeneous CMOS-CNT architecture utilizing novel coding of boolean functions'. NANOARCH 07, October 2007, pp. 15–20
- [18] SHANBHAG N.R., MITRA S., DE VECIANA G., ET AL.: 'The search for alternative computational paradigms', *IEEE Des. Test Comput.*, 2008, **25**, (4), pp. 334–343
- [19] ZIEGLER M., STAN M.: 'CMOS/nano co-design for crossbar-based molecular electronic systems', *IEEE Trans. Nanotechnol.*, 2003, **2**, (4), pp. 217–230
- [20] MIZUOCHI T., MIYATA Y., KOBAYASHI T., ET AL.: 'Forward error correction based on block turbo code with 3-bit soft decision for 10-Gb/s optical communication systems', *IEEE J. Sel. Topics Quantum Electron.*, 2004, **10**, (2), pp. 376–386
- [21] FU B., AMPADU P.: 'An energy-efficient multiwire error control scheme for reliable on-chip interconnects using Hamming product codes', *VLSI Des.*, 2008, **14**, pp. 1–14
- [22] LEHTONEN T., LILJEBERG P., PLOSILA J.: 'Online reconfigurable self-timed links for fault tolerant NoC', *VLSI Des.*, 2007, **13**, pp. 1–13