Robust Circuit and System Design Methodologies for Nanometer-Scale Devices and Single-Electron Transistors

Alexandre Schmid and Yusuf Leblebici

Microelectronic Systems Laboratory
Swiss Federal Institute of Technology
Lausanne, Switzerland
alexandre.schmid@epfl.ch, yusuf.leblebici@epfl.ch

Abstract—**This paper addresses the functional robustness and fault-tolerance capability of very-deep submicron CMOS and single-electron transistor (SET) circuits. A set of guidelines is identified for the design of very high-density digital systems using inherently unreliable and error-prone devices. Empirical results based on SPICE simulations show that the proposed design method improves fault immunity at transistor level. Graceful degradation of circuit performance allows recovery of information, where classical circuits would fail.**

Fault-tolerance; robust very-deep submicron design; CMOS-SET design

I. INTRODUCTION

The phenomenal advancement of semiconductors and microelectronics over the last few decades was made possible mainly by a unique combination of factors, namely: (i) the availability of reliable, mass-producible basic building blocks, i.e. MOS transistors, (ii) the capability of the manufacturing technologies to reduce the device dimensions and thereby increase the integration density on a continuous basis, (iii) the existence of well-defined levels of abstraction that enabled circuit and system designers to develop very complex integrated systems in a hierarchical manner, and (iv) the availability of design tools that allowed the designers to increase their productivity in step with the increasing complexity of the systems. Yet as we explore the limits of very deep-submicron scaling, we are approaching a point where one or more of these factors will become questionable. As a result, many fundamental aspects related to digital circuit and system design will have to be re-considered, especially taking into account the radically different properties of future nanometer-scale devices.

A very wide array of nano-scale quantum device architectures and related technologies are currently under investigation for future nano-scale computation, such as solid state nanoelectronic devices (RSFQ, RTD, SETs, spin transistor, etc.) and molecular electronics (architectures based on small conductive molecules, carbon nano-tubes and others). Single Electron Transistors (SETs) could be among the most interesting and promising candidates for future nano-

electronics because of their particular functionality and complementary characteristics with respect to CMOS.

One key question for the future success of SET and other nano-scale logic circuit applications is about the efficiency of mimicking principles and functionality of CMOS (logic) circuits. Many successful logic applications have been reported by mimicking CMOS, but real competing performance with CMOS still remains to be demonstrated. Background charge sensitivity [1,2] and room temperature operation [3] are among major issues to be solved for single electronics.

Hence, the necessity to cope with intrinsic errors at the device and circuit level must be recognized as a key aspect of nano-scale systems design. To implement such robustness and fault tolerance, new circuit design approaches will need to be considered at the low level. The path to high-level (top-down) synthesis, however, should still follow the well-established route being used for CMOS logic architectures. This observation also underlines the need to create fundamentally novel elementary function blocks on the one hand, and well-defined interfaces to the CMOS-world on the other, so that the inherent advantages of both technologies can be exploited.

In the following, the fundamental principles of a highly regular, redundant, and scalable design approach based on fixed-weight neural networks and multiple-valued logic are presented. It is demonstrated that the proposed design technique offers significantly improved immunity to permanent and transient faults occurring at the transistor level, and that it results in graceful degradation of circuit performance in response to device failures. As such, the proposed design approach is equally suitable for systems consisting of fundamentally novel devices such as SETs, as well as for ultra high-density systems consisting of nanometer-scale "classical" CMOS devices.

II. FUNDAMENTAL CHARACTERISTICS OF SINGLE-ELECTRON TRANSISTORS

The structure of the typical SET consists of an isolated conductive island (single-electron box) that is separated from the two external electrodes (source and drain) by tunneling junctions. The transfer of individual electrons between the

electrodes and the isolated island can be controlled by the voltage that is applied to the gate electrode, based on the fundamental principle of Coulomb blockade. The dimensions of the conductive island and the tunneling junctions need to be in the order of a few nanometers to a few tens of nanometers. While larger device dimensions allow observable device operation at very low temperatures, the dimensions may need to be reduced to sub-nanometer levels in order to achieve Coulomb blockade near room temperature [2]. Still, it has already been demonstrated that many fundamental device operations can be performed with this structure.

The simplified structure of an SET is compared with that of a MOSFET in Figure 1. Indeed, the device is reminiscent of a usual MOSFET, but with a small conducting island embedded between two tunnel barriers, instead of the usual inversion channel. At small drain-to-source voltages, there is no current since the tunneling rate is between the electrodes and the island is very low. This suppression of DC current at low voltage levels is known as the Coulomb blockade [2]. At a certain threshold voltage, the Coulomb blockade is overcome, and for higher drain-to-source voltages, the current approaches one of its linear asymptotes. It is evident that the device can be operated as a switch controlled by the gate electrode, capable of performing a number of tasks.
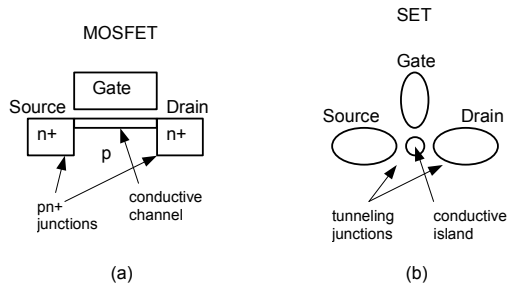

Figure 1. Simplified structure of a MOSFET (a), compared with that of an SET (b).

One of the most significant difficulties of designing complex functions using SETs will be the inherent sensitivity of their characteristics to background charge fluctuations. This effect is the result of (permanent or transient) random variations in local charge due to fabrication irregularities, leakage, or external perturbations such as noise. Background charge effects may permanently or temporarily disrupt device function, rendering one or more SETs inoperative within a functional block in a random manner. To ensure reliable operation and to reduce the sensitivity of devices to background charge effects (especially at room temperature), the device dimensions must be reduced to sub-nanometer levels, which is not very feasible in the foreseeable future. A more likely scenario is that the functional blocks be designed with a certain degree of fine-grained, built-in immunity to such permanent and transient faults, such that they are capable of absorbing a number of errors and still be able to perform their functions. This type of pervasive fault tolerance, which is based on the implicit acceptance that a certain percentage of devices in the system will fail in a random fashion, may require a novel approach to robust design that is quite different from classical techniques of fault-tolerance and redundancy.

III. FAULT-TOLERANT CIRCUIT ARCHITECTURE

The proposed fault-tolerant architecture consists of four layers in which the data is strictly processed in a feed-forward manner (Figure 2). The first layer is denoted as the input layer, accepting conventional Boolean (binary) signal levels. The core operation is performed in the second layer, which consists of a number of identical, redundant units implementing the desired logic function. It will be seen that the fault immunity increases with the number of redundant units, yet the operation is quite different from the classical majority-based redundancy. The third layer receives the outputs of the redundant logic units in the second layer, creating a weighted average with re-scaling. Note that the output of the third layer becomes a multiple-valued logic level. Finally, the fourth layer is the decision layer where a binary output value is extracted using a simple threshold function.
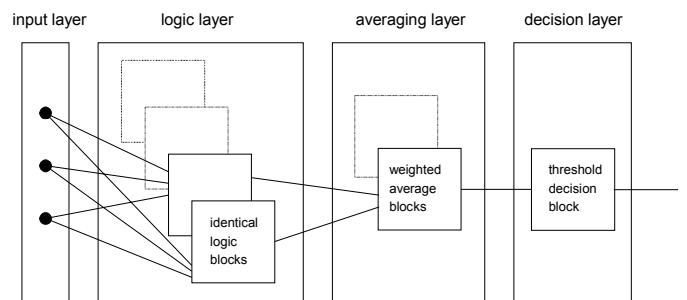

Figure 2. The proposed fault-tolerant architecture based on multiple layers.

Figure 3 shows the typical weighted-averaging and re-scaling function that is performed by one of the third-layer blocks. It was already shown in the literature that this particular type of weighted-sum functions can be implemented quite easily with SET devices [4]. Similarly, proposals have been made to exploit the particular characteristics of SETs for the implementation of multiple-valued logic functions [5,6]. Note that in our approach, the multiple-valued logic levels are not used to increase the information content of a signal, rather, they serve as a mechanism that allows reducing the impact of random device failures on the global transfer function surface. Hence, the multiple-valued logic representation of the function provides an effective means for absorbing faults that would otherwise manifest themselves as catastrophic errors. In the simplest possible implementation, a single averaging block with equal weights already provides significant fault immunity. The benefits of this circuit architecture can be further increased by using multiple averaging units and/or different input weighting schemes.
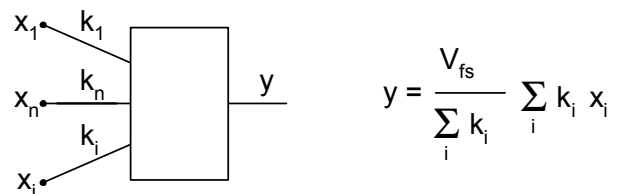

$$y = \frac{V_{fs}}{\sum_i k_i} \sum_i k_i x_i$$

Figure 3. The general weighted averaging and re-scaling function used in the third layer.

IV. SIMULATION EXAMPLES

The first example consists of two identical logic blocks in the second layer, and one averaging block in the third layer. It is assumed that the NOR function blocks in the second layer are realized using the straightforward construction approach based on dual (series-parallel) switch networks, similar to classical CMOS logic gates. Each NOR block in the second layer receives two binary inputs, and produces one binary output. The outputs of the second layer are processed further in the averaging block to produce the multiple-valued output. As long as all devices operate correctly, three of the four possible input combinations will produce a logic-zero output in both of the second-layer logic blocks, and only one input combination, "00" will produce a logic-one output. Figure 4 shows the output level of the averaging block, for correct operation. It can be seen that the transfer function surface generated at the output of the averaging block clearly reproduces the expected two-input NOR function, with the fourth-layer decision threshold set as shown. Note that the input-output behavior presented here is not particularly influenced by the circuit implementation of the logic function blocks in the second layer, nor by the realization of the averaging function in the third layer. The fundamental characteristics of the proposed architecture are largely independent of the specific implementation of the logic function blocks and the averaging block.
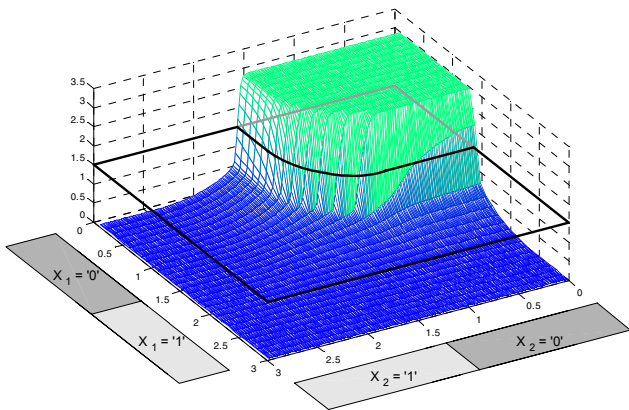

Figure 4.   Output transfer function generated by the averaging layer of the two-input NOR circuit, showing correct operation (no device failures). The fourth-layer decision threshold is also indicated.

Considering random device failures, it can be shown that the proposed architecture can successfully absorb all single-faults occurring anywhere in the second layer, as long as there are two or more identical logic units in the second layer. This is a property that can only be achieved using three or more redundant units in the conventional approach based on majority decisions. Furthermore, the new circuit architecture is capable of producing correct output behavior even when some devices in the third layer (averaging block) are faulty. This is in stark contrast to the limited fault immunity of conventional redundant systems where even a single-fault in the majority decision block cannot be tolerated. The significant benefits of the proposed design approach become evident especially when

considering multiple device failures. Figure 5 shows the output transfer function surface of the circuit described above (two identical NOR blocks in the second layer, one averaging block in the third layer) where a total of four devices are assumed to be faulty. It can be seen that the correct output behavior can be extracted by setting the decision threshold level as shown. A fixed decision threshold level appears to be sufficient in most cases, while dynamically adjustable decision threshold levels may further increase the flexibility of the proposed approach.
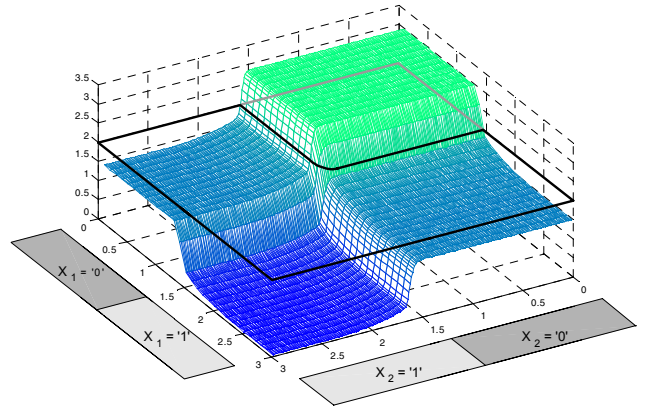

Figure 5.   Output transfer function generated by the averaging layer of the two-input NOR circuit, with a total of four device failures in both of the second-layer logic blocks. It is seen that the correct output can still be obtained with the proper threshold decision in the fourth layer.

Figure 6 shows the third-layer output transfer function surfaces of the NOR circuit with two redundant logic units, under the assumption of several different (single or multiple) device failures. It is seen that the multiple-valued output level of the averaging block is capable of preserving the essential function behavior in most of the cases, and that the correct Boolean output can be extracted by applying a simple decision threshold.
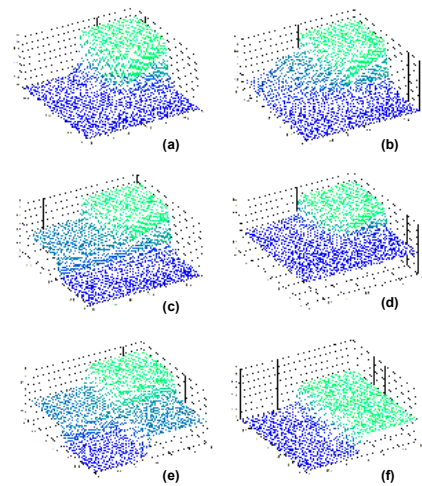

Figure 6.   Output transfer function of the two-input NOR circuit, with two redundant units: (a) no device failure, (b) one failure in the 2nd layer, (c) two failures in the 2nd layer, (d) one failure in the 3rd layer, (e) four failures in the 2nd layer, (f) four failures in the 2nd layer.

The classical fault-tolerant design methods based on triple-redundancy, for example, rest on the inherent assumption that only one of the three identical function blocks will be subject to any single-device failure at any time. If the defect density is high, on the other hand, ensuring this proposition is not very straightforward. The ability of the proposed approach to withstand multiple device failures and still produce the correct output function is the strongest advantage of our design method in the presence of very high defect densities (Figure 7).
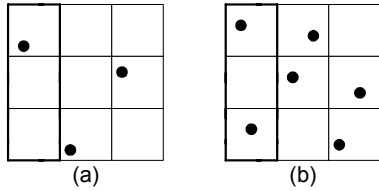

Figure 7. (a) Low defect density distribution where classical triple redundancy would be sufficient. (b) High defect density distribution where triple redundancy is not sufficient.
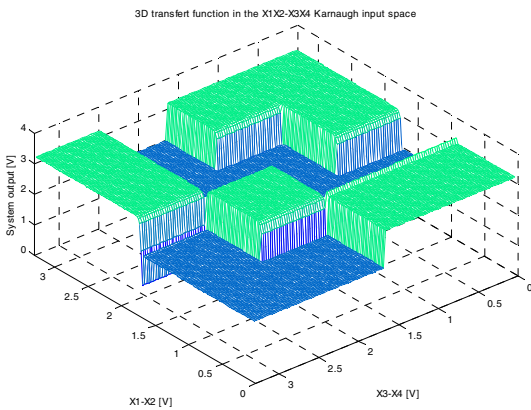

Figure 8. Transfer function surface of a 4-input mapping onto a 1-output, with a large number of distributed errors.

The fault-tolerant design approach is not limited to the mappings of two input variables onto one output variables, and can be extended to the cases where several input and/or output variables are involved. These cases require adaptive systems, where the output thresholds should be multiple, and should be adjusted in real time. Figure 8 shows the 4-input variables onto 1-output variable mapping of the following Boolean function:

$$f(x_1, x_2, x_3, x_4) = x_1\overline{x_4} + \overline{x_2}\,\overline{x_3} + x_1\overline{x_2}\,\overline{x_3} + \overline{x_1}x_2x_3x_4,$$ where three identical function blocks are used in the second layer, to ensure robust operation. The minimal number of transistors to synthesize $f(x_1, x_2, x_3, x_4)$, i.e. the number of transistors in each function block, is equal to 40. The entire circuit with three identical units in the second layer has a total of 125 transistors. In the specific example above, 44 out of these 125 devices are allowed to fail (either stuck-on or stuck-off). It can be seen that the correct output function surface can still be reconstructed, even with a failure rate of close to 30%.

Based on the empirical simulation results described in the previous, the proposed fault-tolerant design technique can be recognized to significantly increase circuits' robustness, and allowing them to absorb a large number of transistor-level errors, i.e. without propagation of the error to the gate output node. Very clearly, this technique deserves closer attention, and a thorough theoretical approach to define its properties and limits.

V.   CONCLUSION

In this paper, various circuit and system level design challenges for nanometer-scale devices and single-electron transistors (SET) are discussed, especially concentrating on the functional robustness and fault tolerance point-of-view. The main issues are identified for the design of very high-density digital systems using inherently unreliable and error-prone devices. The fundamental principles of a highly regular, redundant, and scalable design approach based on fixed-weight neural networks and multiple-valued logic are presented. The proposed circuit architecture, based on redundant layers arranged in a feed-forward manner, is capable of absorbing the effects of multiple simultaneous device failures and still offering a high probability of correct operation. This approach is fundamentally different from the conventional logic redundancy strategies that rely on multiple identical units and majority decisions on their outputs. It is demonstrated that the proposed design technique offers significantly improved immunity to permanent and transient faults occurring at the transistor level, and that it results in graceful degradation of circuit performance in response to device failures. We believe that the design technique presented in this paper can be adapted to a wide variety of cases to address the potential reliability limitations of novel nanometer-scale devices.

REFERENCES

[1] C. Wasshuber, Computational Single–Electronics, Springer Ver., Wien, 2001.

[2] K.K. Likharev, "Single-electron devices and their applications," Proceedings of IEEE, vol. 87, Issue 4, pp. 606 -632, 1999.

[3] M. Kirihara, N. Kuwamura, K. Taniguchi, and C. Hamaguchi, "Monte Carlo study of single-electronic devices," in Ext. Abst. Int. Conf. on Solid State Devices and Materials, Yokohama, Japan, 1994, pp. 328–330.

[4] C. Lageweg, S. Cotofana, S. Vassiliadis, "A Linear Threshold Gate Implementation in Single Electron Technology," in Proc. IEEE Comp. Soc. Workshop on VLSI, 2001, pp. 93-98.

[5] H. Inokawa, A. Fujiwara, Y. Takahashi, "A Multiple-Valued SRAM with Combined Single-Electron and MOS Transistors," Device Research Conference, 2001.

[6] H. Inokawa, A. Fujiwara, Y. Takahashi, "A Multiple-Valued Logic with Merged Single-Electron and MOS Transistors," Electron Device Meeting IEDM, 2001.