A Modular Approach for Reliable Nanoelectronic and Very-Deep Submicron Circuit Design Based on Analog Neural Network Principles

Alexandre Schmid and Yusuf Leblebici
Microelectronic Systems Laboratory
Swiss Federal Institute of Technology
Lausanne, Switzerland
alexandre.schmid@epfl.ch, yusuf.leblebici@epfl.ch

Abstract— **Reliability of nanodevices is expected to be a central issue with the advent of very-deep submicon devices and future single-electron transistors. We propose a new approach based on the assumption that a number of circuit-level devices are to be expected to fail. Artificial neural networks can be trained to resists to errors and be used for synthesizing fault-tolerant Boolean functions. The development method is outlined; results based on the feed-forward artificial neural network implementation are presented, while future research directions are discussed with possible applications.**

Fault-tolerance; robust very-deep submicron design; artificial neural networks

I. INTRODUCTION

Prevailing trends indicate that the envisioned novel directions in nano-scale device design and system-level integration could largely benefit from the existing knowledge if these nanoelectronics technologies are developed in ways that are compatible with conventional, very-deep submicron (VDSM) silicon CMOS technologies. Similarly, new system design methodologies and design platforms should be equally applicable to large-scale system design problems using new nano-scale devices as well as using a very large number of conventional CMOS devices. In addition, early studies suggest that many characteristics of novel nano-scale devices such as single-electron transistors (SET) are actually complementary to CMOS devices – offering major advantages in device size, integration density and power dissipation while suffering in some areas where CMOS is traditionally strong, such as voltage gain, drive strength and high speed.

On the other hand, the decreasing reliability of VDSM CMOS devices and the expected reliability deficiencies of novel nanoelectronic devices (such as SETs) call for design approaches that can compensate for these problems. VDSM CMOS transistors already exhibit specific properties, which can be perceived as a degradation of conventional CMOS device characteristics and behavior. Similarly, nanoscale single-electron (or few-electron) transistors are expected to suffer from a relatively large, random and dynamic failure rate, due to background charge affecting their correct behavior, in addition to process related failures. In this paper we propose a modular approach to improving the reliability of nano-devices, which can be applied both to VDSM CMOS and future SET-transistor based systems.

The possible direction that is being proposed for the realization of complex functions using nano-scale devices involves the design of artificial neural network (ANN) architectures. Similar approaches were proposed earlier by several researchers [1-4] and realizations of basic ANN functionality were shown using SET structures. The main functionality of neural building blocks is based on computation of synaptic weighted-sum terms on each node. With their unique characteristics, the SET devices appear to be good candidates for the implementation of such functions. But the main advantage of implementing SET-based neural function blocks would be the inherent hardware redundancy, error-tolerance and self-organization features of such ANN architectures, offering a very effective means to counteract the inherent weaknesses of SET devices, and providing a robust methodology for realization of highly complex systems. Thus, an SET-based neural network architecture would be expected to continue its operation even when some of its components fail to operate properly due to permanent and transient faults.

While offering some significant advantages in terms of system-level robustness, the classical ANN approach also has several difficulties associated with its implementation. The elementary synaptic functions can be realized by using SET devices, but the hardware realization of learning algorithms remains a problematic issue. Also, the interconnect density in general-purpose ANN architectures represents a serious limitation in terms of hardware complexity, which is exacerbated by the introduction of general-purpose learning/adaptation hardware. Finally, the classical ANN paradigms based on learning algorithms and self-adapting hardware are in stark contrast to the conventional, well-established system design methodologies currently being used for CMOS VLSI systems. In the following, we will explore a design approach that borrows some of the fundamental aspects of ANN architectures such as replication of identical hardware blocks, low-level regularity, multiple functional layers, and regular, short-distance interconnections.

0-7803-7977-2/03/$17.00 (C) IEEE

II. CIRCUIT-LEVEL FAULT TOLERANCE USING ARTIFICIAL NEURAL NETWORKS

Fault-tolerance of VLSI circuits has traditionally been addressed at system-level, involving algorithmic adaptation, block-level redundancy and majority voting as the main tools [5]. These techniques have proven to be appropriate for modern VLSI where the defect density is considered to be low. However, correct operation of these techniques cannot be guaranteed in cases where larger defect densities, and a uniform distribution of the failures are observed.

The reliability of ANNs has been tackled by several research groups [6-7]. The algorithmic aspects, as well as the hardware implementation issues have been addressed, mainly with the target of guaranteeing the ANN transfer function under failure, while considering a minimal number of neurons [8-9]. Also the aspects of learning using error-prone analog hardware and/or limited precision digital circuits have been addressed. In our approach, feed-forward ANNs (FFANNs) are used as a tool to implement Boolean functions, where each function is perceived by the designer as a black box. An appropriate training technique, as well as the appropriate neuron redundancy are used to provide enhanced fault-tolerance.

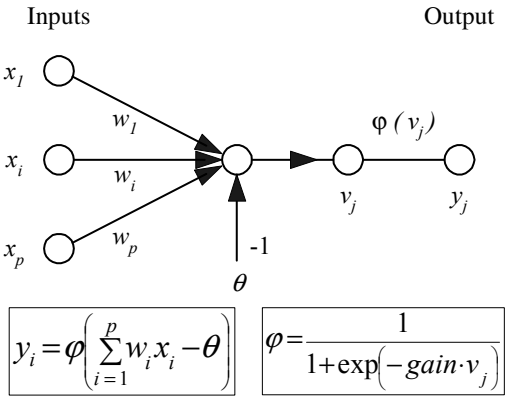

Figure 1. Standard neuron model used in the numerical simulations.

A classical model of neuron as depicted in Figure 1, and the standard error backpropagation learning algorithm have been applied in this paper to demonstrate the correctness of the proposed approach [10].

The internal blocks are composed of a layered arrangement of kernel units as depicted in Figure 2. In its simplest organization, each unit in the first layer is responsible for contributing to a fraction of the expected block response. The second layer units aggregate the first unit answers in an averaging voting scheme to produce the analog block outputs. This arrangement of the so-called atomic units can be understood as an analog artificial neural network. Obviously, any failure of a first layer unit causes a limited error in the global block response, which is at most equal to its weight in the averaging layer.

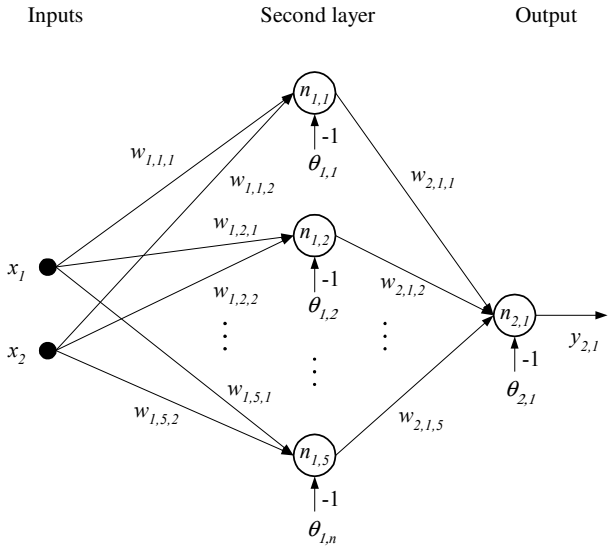

Figure 2. Two layer FFANN implementing a Boolean function.

The circuit in Figure 2 was simulated and proven to operate correctly with 5 neurons in the second layer and one single failing connection, performing a simple Boolean function (NAND). Note that a single neuron would achieve the same result under assumption of full operativness. The artificial neural network was trained, using binary inputs-output pairs and tested with the same data, due to the limited size of the data set.

The main motivation for using multiple neurons in the second layer is clearly the intention to leverage the relatively "cheap" additional devices in return for enhanced robustness of the function. This approach is quite different from the classical focus of ANN design where the goal is usually to identify the minimum number of neurons (or devices) that are needed to perform a function correctly.

III. A NOVEL FEEDFORWARD ANN BOOLEAN FUNCTION SYNTHESIS BLOCK

The key issue is to design each functional block as capable of generating, transmitting and receiving analog levels, and to define "regions of acceptability" that comply with this concept – instead of classical noise margins. The output of any function block may be varying depending on its state, on the possible crosstalk perturbation, and on the actual operation of the basic transistors, which may exhibit unreliable behavior, such as soft and/or unpredictable cut-off or turn-on levels. Rather than handling these issues with noise margins intervals, it is more appropriate to design the input stages to be compliant with these features. Hence, we propose to design each functional module with two complementary analog outputs, where each output is considered as a degree of confidence to be given to logic high and logic low levels respectively. Figure 3 depicts the proposed gate architecture, in the case of a three-layer FFANN with analog, complemented inputs and output designed to perform a simple Boolean operation (NAND, NOR,…)

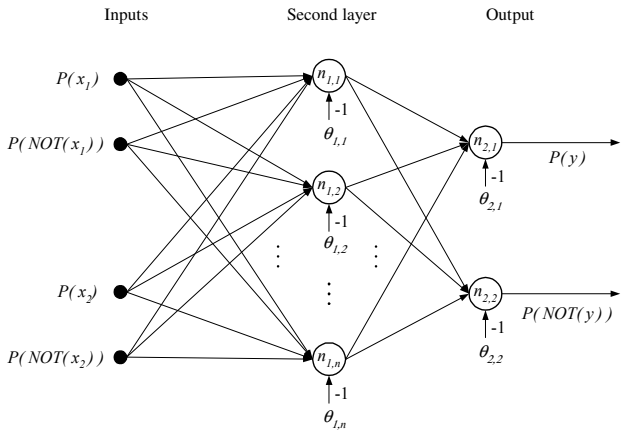

Figure 3. Proposed architecture of a Boolean gate

An appropriate training scheme based on regular FFANN training until a defined mean-square error target is reached, followed by a second training sequence involving dynamic random suppression of connections allows for spreading the global information to the second layer units, thus providing the network with fault-tolerance ability, whereas the redundant analog outputs ensure possible data recovery in case of disconnection.

Figure 4 shows the simulation results of a FFANN composed of 9 second-layer neurons, and performing the NAND Boolean function. The Mean square error computed at the network output is depicted on the upper graph, where a first training phase consisting of regular error backpropagation followed by a second training phase where one random connection is selected for failure can be identified. In both phases, training is stopped when a predefined mean square error target is reached.

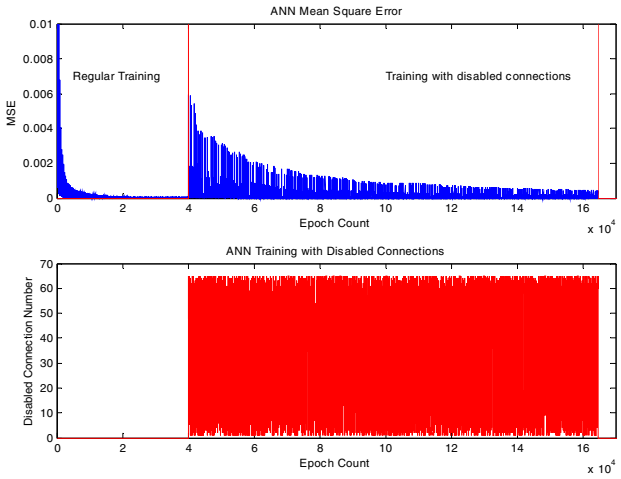

Figure 4. Training results of the proposed architecture of Boolean gate, NAND.

IV. RESEARCH PERSPECTIVES

The target of the work presented in this paper is to provide the designer with a library of fault-tolerant gates that can be practically considered as black boxes, and their mutual connection scheme. The signals are transmitted to the next gate using two complementary lines to improve the robustness of data transmission. Several research directions have to be explored in order to propose a full system with its design methodology.

Conceptual developments and preliminary results show that FFANNs can be successfully applied at circuit-level in order to increase system fault tolerance. Several other ANN models, and learning schemes have been developed and should be considered as candidates to be implemented into the black box gate, along parallel lines to earlier research on strict operator redundancy [11].

The ability to absorb input analog levels as such is gained in the training phase. Noise injection into the training set has been reported in [9] in the case of a digital approach. In our case, using a noisy training set is intended to mimic the imperfect behavior or effect of a damaged or leaky device. Furthermore, the injection of weight noise has been reported to improve the generalization ability and fault-tolerance [12].

The appropriate number of the first layer redundant neurons, as well as the connection scheme between them is to be tackled based on the extensive existing studies on regular artificial neural networks. This approach also confirms the necessity of spatial redundancy, which is believed to be acceptable in the perspective small transistor size, and with respect to the benefit of higher reliability operation. In our case, the gate input stage, i.e. the second layer neurons and their connection scheme, has to be designed very carefully in order to take benefit of the differential signal connections. Moreover, full feedforward connection scheme that has been applied in the numerical simulations so far is unlikely to be the most appropriate connection scheme. Partitioning the network into subnetworks with specific tasks is probably the best way to accommodate and take benefit of the dual nature of transmitted signals.

So far we have addressed the limited case where connections are inoperative. In real circuit operation, faults will mostly appear as stuck-at zero, stuck-at one lines and/or stuck-on, stuck-off transistors. The method proposed in the previous section can be extended to comply with these failure models. The extra cost is to be expected in terms of neuron units to be added in the architecture.

Finally, the hardware implementation of the proposed architecture into actual analog electronics has to be undertaken. Replicating the results shown above and obtained with double-resolution simulation cannot obviously be expected [13-15]. However, using fixed weights, i.e. using the FFANN in forward propagation mode only requires limited precision, typically 6 to 8-bit. Analog ANNs is a mature field for which many circuit solutions have been proposed, and theoretical issues have been tackled.

V. CONCLUSION

The reliability of nano-scale devices has to be considered as a central issue in the successful implementation of future nano-scale devices. 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. Artificial neural networks have proven to be error resistant when they are appropriately trained, and thus offer a very interesting solution to constructing fault-tolerant VLSI at the circuit level. The proposed design is based on a layered, feed-forward arrangement of redundant neurons, where every unit contributes to a portion of the final result. A circuit architecture is proposed, preliminary results presented, and future research directions are discussed.

REFERENCES

[1] R. H. Klunder, J. Hoekstra, "Programmable logic using a SET electron box," IEEE International Conference on Electronics, Circuits and Systems ICECS, 2001.

[2] S. Fölling, Ö. Türel, K. Likharev, "Single-Electron Latching Switches as Nanoscale Synapses," International Joint Conference on Neural Networks IJCNN, 2001.

[3] K. F. Goser, C. Pacha, A. Kanstein, M. L. Rossmann, "Aspects of System and Circuits for Nanoelectronics," Proc. of the IEEE, Vol. 85, No. 4, April 1997, pp. 558-573.

[4] M. J. Goossens, C. J. M. Verhoeven, A. H. M. van Roermund, "Single Electron Tunneling Technology for Neural Networks," Proc. Microneuro'96, 1996.

[5] M. Alderighi, S. D'Angelo, C. Metra, G. R. Sechi, " Achieving Fault-Tolerance by Shifted and Rotated Operands in TMR non-Diverse ALUs," IEEE Int. Symposium on Defect and Fault Tolerance in VLSI Systems, 2000, pp. 155-163.

[6] K. Yamamori, S. Horigushi, "The Efficient Design of Fault-Tolerant Artificial Neural Networks," IEEE Int. Conf. On Neural Networks, 1995, pp. 1487-1491.

[7] P. Kerlirzin, P. Réfrégier, "Theoretical Investigation of the Robustness of Multilayer Perceptrons: Analysis of the Linear Case and Extension to Nonlinear Networks," IEEE TNN, Vol. 6, No. 3, 1995, pp. 560-571.

[8] R. C. Frye, E. A. Rietman, C. C. Wong, "Back-Propagation Learning and Nonidealities in Analog Neural Network Hardware," IEEE TNN, Vol. 2, No. 1, 1991, pp. 110-117.

[9] J. Nijhuis, B. Höfflinger, A. van Schaik, L. Spaanenburg, "Limits to the Fault-Tolerance of a Feedforward Neural Network with Learning," 20th Symposium on Fault-Tolerant Computing, 1990, pp. 228-235.

[10] D. E. Rumelhart, G. E. Hinton, R. J. Williams, "Learning Internal Representations by Error Backpropagation," Parallel Distributed Processing: Exploration in the Microstructures of Cognition, Vol. 1, D. E. Rumelhart, J. L. McClelland, Edts., MIT Press, Cambridge, MA, 1986, pp. 318-362.

[11] A. Schmid, Y. Leblebici, "Robust Circuit and System Design Methodologies for Nanometer-Scale Devices and Single-Electron Principles," IEEE NANO03, San Fransisco, CA, 2003.

[12] A. F. Murray, P. J. Edwards, "Enhanced MLP Performance and Fault Tolerance Resulting from Synaptic Weight Noise During Training," IEEE TNN, Vol. 5, No. 5, 1994, pp. 792-802.

[13] G. Dündar, K. Rose, "The Effects of Quantization on Multilayer Neural Networks," IEEE TNN, Vol. 6, No. 6, 1995, pp. 1446-1451.

[14] K. Asanovic, "Experimental Determination of Precision Requirements for Back-Propagation Training of Artificial Neural Networks," 2nd Int. Conf. On Microelectronic for Neural Networks, Germany, 1991, pp. 9-15.

[15] T. Lundin, P. Moerland, "Quantization and Pruning of Multilayer Perceptrons: Towards Compact Neural Networks," IDIAP Communication, availableat following URL: http://www.idiap.ch