A Combined Fault Detection and Discrimination Strategy for Resource-Sensitive Platforms

Richard McWilliam, Philipp Schiefer and Alan Purvis
School of Engineering and Computing Sciences, Science Laboratories
Durham University, Durham, DH1 3LE, United Kingdom
Email: r.p.mcwilliam@durham.ac.uk

***Abstract*—This paper presents a combined fault detection and discrimination strategy for CMOS logic incorporating active resource mitigation and monitoring. The approach is demonstrated for a NOR gate using a dual redundant gate design with selective mitigation and analogue or digital detection. The potential benefits of the approach are discussed with respect to resource awareness and management within fine-grained logic.**

I. Introduction

Fault detection and mitigation within CMOS logic structures is a long-standing challenge that is seeing new a emphasis for nanoscale and printable electronics. The possibility for intrinsic resource awareness and management without obfuscating management at higher design levels is an attractive proposition but requires new gate and transistor level strategies. This paper presents ongoing work into a combined fine-grained redundancy and active mitigation approach with minimal resource overhead that enables selective fault detection, masking and discrimination close to the point of fault manifestation.

A. Existing Methods

On-line fault strategies have been discussed at length for future nanoscale electronics where massive redundancy concepts become feasible [1]. However, resource-sensitive platforms typically involve more conservative duplicate gate and/or interconnect structures combined with majority signal generation in order to mask faults and prevent their manifestation at critical outputs. An practical example of this is seen in [2], in which combined logic interleaving and quad-transistor structures are employed. While the use of regular cell structures is attractive, the typical area overheads range between 3x and 8x without explicit fault detection or discrimination. It could be argued that fault detection triggers may be generated within quad-transistor majority logic but determination of the specific fault location and type becomes abstracted by the internal process of converting critical faults to sub-critical faults.

Fine-grained fault tolerant strategies for future nanoscale CMOS logic have been proposed to combat anticipated manufacturing defects. An example is reported in [3] wherein a defect present in either the N-type or P-type networks invokes switched, active pull up/down loads. In this case however defect detection is not part of the repair method and would typically be provided by additional built-in self-test (BIST) logic and/or external test equipment. Hard-fault mitigation approaches have been explored based on active switching matrices [4] however self-detection is not included in the strategy.

FPGA platforms provide fixed cellular architectures and full or partial configuration, though the total resource utilisation rarely approaches 100%. Efficient resource allocation is a difficult task in cellular architectures based on arrays of regular logic cells, prompting the use of partial reconfiguration. Even so, it is not yet clear how spare resources may be reallocated online without resorting to external supervisory hardware/software as typified in [5]. While solutions based on custom programmable architectures have been proposed that aim to address this limitation [6] by enabling dynamic resource allocation, fault detection is still achieved through data coding and error detection and correction (EDC) hardware that is abstracted from the exact nature and location of the fault.

II. Proposed Method

The proposed strategy relies upon an alternative method referred to here as *Stuck-At Fault Resilient* (SAFR) design, wherein fixed dual redundancy is combined with a fault triggering mechanism [7]. An example logic gate implemented by the SAFR approach is shown in Fig. 1a, where dual redundancy is employed within the P- and N-type networks.

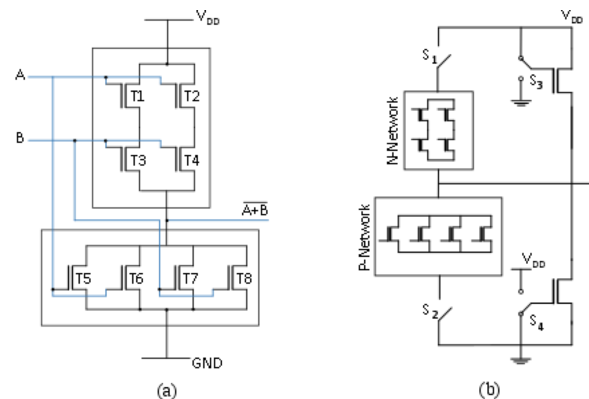

Fig. 1. Gate design strategy. (a) Example of redundancy scheme for NOR gate employing P- and N-type networks. (b) Potential implementation for active fault mitigation according to [3].

TABLE I
STUCK-HIGH FAULT RESPONSE OF CMOS NETWORK

Input	Stuck-on fault location[a]							
AB	T1	T2	T3	T4	T5	T6	T7	T8
00	1	1	1	1	X	X	X	X
01	0	0	X	X	0	0	0	0
10	X	X	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0

[a]Output error denoted by 'X'

A. Detection Strategy

The dual redundancy strategy permits masking of any single stuck-off fault and selective fault triggers for stuck-on faults depending upon the state of the inputs. Of particular note is the fact that fault discrimination is not retained when higher redundancy factors are used i.e., triple- and quad-transistors. Hence, a resource trade-off between fault masking capacity and fault identification is present in this approach.

B. Discrimination and Mitigation

Selective fault masking allows for detection of stuck-on faults considered to be critical due to potential high current flow between VDD and GND. Examination of the gate output response under fault condition, summarised in Table I, shows that at there is at least one input combination that generates current flow between VDD and VSS for every single stuck-on fault. This may be exploited to achieve discrimination of fault type by monitoring current imbalance in the CMOS network or else periodic exercising of the gate inputs via digital test. The P- and N-networks are combined with the switching network shown in Fig. 1b, which includes weak active pull-up/down loads typically used for defect repair [3], but which are used here for selective online fault discrimination.

III. RESOURCE AWARENESS AND MANAGEMENT

Resource considerations will be important for emerging printable and nanoscale electronics due to their differing densities and scope for building redundancy structures based upon multi-gate and/or sub-gate nano-structures. Resource management extending to the fine-grained levels should be explored for both defect tolerance and hard-fault mitigation.

Combining the above approach with weak active pull-up/down loads creates an efficient active mitigation mechanism that, when further combined with dual redundancy within the P- and N-networks, opens up further resource awareness options in the presence of faults. Once a fault has been detected, partial isolation proceeds by switching to pseudo-NMOS or PMOS mode wherein the nature of the fault may be further characterised. For example, assuming a stuck-at high fault occurring within the P-network (Transistors T1-T4 in Fig. 1a), the location of the fault is not known *a-priori*. The circuit may first be switched to pseudo-NMOS mode (setting switches S1 and S3 in Fig. 1b) and, due to the complimentary nature of the design, a second analogue/digital test will would reveal the same fault behaviour summarised in Table I. However, depending on the value of the weak pull-down resistance of transistor T9, the digital test may pass without error and the adapted circuit may continue to be used in a degraded state. Alternatively, the circuit may be switched into pseudo-PMOS mode (switches S2 and S4) whereupon the error no longer persists. Hence the state of the P- and N-networks may be individually ascertained. The reverse situation of a fault occurring within the N-network would proceed in identical fashion as described above. At all times stuck-at low fault events are intrinsically masked.

An further extension of resource awareness concerns continual resource monitoring in the presence of intermittent faults. For the above case of the pseudo-PMOS configuration being activated in response to a stuck-high fault within the P-network, a further option would be to periodically switch to the pseudo-NMOS configuration and check the P-network response to determine whether the fault persists. This serves two functions: first, intermittent faults may be handled in a graceful manner and with specific knowledge of their locality. Second, disappearance of the fault allows for restoration of the full CMOS network and non-degraded performance.

IV. CONCLUSION

Detection remains a fundamental challenge for resource management across multiple system levels. The proposed dual-redundancy method achieves discrimination between stuck-high/stuck-low faults and selective masking, thus reserving active mitigation for stuck-high faults. Fine-grained resource mitigation proceeds by combining redundancy with weak pull-up/down networks. Ongoing work is investigating further logic gate configurations and functional logic built from such gates.

ACKNOWLEDGEMENT

This work was supported by the UK EPSRC Centre for Innovative Manufacturing in Through-life Engineering Services (EP/I033246/1).

REFERENCES

[1] J. Von Neumann, "Probabilistic logics and the synthesis of reliable organisms from unreliable components," *Automata studies*, vol. 34, pp. 43–98, 1956.

[2] J. Han, E. Leung, L. Liu, and F. Lombardi, "A Fault-Tolerant Technique Using Quadded Logic and Quadded Transistors," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. PP, no. 99, pp. 1–1, 2014.

[3] M. Ashouei, A. Singh, and A. Chatterjee, "Reconfiguring CMOS as Pseudo N/PMOS for Defect Tolerance in Nano-Scale CMOS," in *21st International Conference on VLSI Design, 2008. VLSID 2008*, 2008, pp. 27–32.

[4] R. Kothe, H. Vierhaus, T. Coym, W. Vermeiren, and B. Straube, "Embedded Self Repair by Transistor and Gate Level Reconfiguration," in *Design and Diagnostics of Electronic Circuits and systems, 2006 IEEE*, 2006, pp. 208 –213.

[5] J. Emmert, C. Stroud, and M. Abramovici, "Online Fault Tolerance for FPGA Logic Blocks," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 15, no. 2, pp. 216–226, Feb. 2007.

[6] P. Bremner, et al. "SABRE: a bio-inspired fault-tolerant electronic architecture," *Bioinspir. Biomim.*, vol. 8, no. 1, p. 016003, Mar. 2013.

[7] P. Schiefer, R. McWilliam, and A. Purvis, "Fault Tolerant Quadded Logic Cell Structure with Built-in Adaptive Time Redundancy," *Procedia CIRP*, vol. 22, pp. 127–131, 2014.