Choice of granularity for reliable circuit design using dynamic reconfiguration

Atin Mukherjee *, Anindya Sundar Dhar

Department of Electronics and Electrical Communication Engineering, Indian Institute of Technology Kharagpur, India

A B S T R A C T

While designing fault tolerant systems using dynamic reconfiguration, choice regarding the size of the granule influences the area, the power and the delay overheads. In this paper, attempt has been made to determine the optimum granule size that would incur minimum overhead vis-à-vis other design parameters such as the number of faults to be tolerated etc. In order to facilitate the design process, mathematical expressions have been provided showing the relationships among the area of single granule, the number of the external connections, the area of the reconfiguration multiplexers and the probability of failure of the system. Optimum granule-sizes in designing various fault tolerant circuits from ripple carry adder to CORDIC as well as Viterbi decoder have been derived.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

The ongoing miniaturization through shrinking the device dimensions not only increases the packing density, but also elevates the probability of failure of the system due to internal faults such as gate dielectric breakdown and electromigration or for external influences, such as radiation induced defects in space applications. Due to the high density of transistors and complex topology used in modern technologies, failure rates have been increased and hence incorporation of fault tolerance has become essential in the design of various systems catering to different applications for ensuring reliable and fault-free operations.

For resource-constrained systems where the amount of hardware devoted for active computing must be maximized, dynamic reconfiguration is the preferred fault tolerant technique [1–3] that uses a fault detection and reconfiguration unit to identify the faulty modules and replaces them with fault-free spares. For static redundancy methods, all the spares are active along with the normally working modules and hence more power is consumed compared to the dynamic reconfiguration method, where the backup units become operational only upon detection of faults in some active modules. Popular static methods like triple modular redundancy (TMR) [1], multiplexing technique [4], quadded logic (QL) [5], quadded transistor (QT) [6] etc. require at least three times of the original hardware needed in case of the non-redundant ones. Hence for critical applications such as in satellites and avionics where increase in payload is a major concern, dynamic reconfiguration technique gets priority over the static methods. The generalized idea for self-repair including testing and reconfiguration in case of dynamic reconfiguration has been briefed in [7].

The major limitation of the dynamic recovery method is the associated delay due to the time required for testing and reconfiguration. But this impact can be reduced easily if the reconfiguration is performed during idle time or on an idle hardware portion of the system, if any [8]. But such arrangement cannot provide a real-time protection of the system as testing and reconfiguration are performed only when the resources become idle. For real time operation, we consider a topology that incorporates the *hot-standby* feature in dynamic recovery, where testing and reconfiguration are carried out simultaneously without stopping the normal operation of the system [9]. In this topology, spare modules are tested for faults and if found to be non-faulty, operation of some normal active modules is transferred to them making the operative modules act as spares and tested for errors. If any module is identified as faulty, further activation of that module in the circuit is prohibited and hence no extra time is required for reconfiguration.

A detailed study on generalized modular redundancy scheme enhancing the fault tolerance for combinational circuits has been carried out in [10]. Most of the sequential circuits can be thought of as an integration of some combinational logic and registers (e.g. up-counter = incrementer + registers). In general scan chain technique [11] is used to locate faulty registers/flip-flops and dismantle them from the system. Registers can also be designed fault tolerant using triple redundant storage [12]. Hence a sequential circuit can be made capable of tolerating faults by combining fault tolerant registers with fault tolerant combinational blocks of its parts. Proper choice of the spare module-size in case of dynamic reconfiguration plays an important role in minimizing the area and delay overheads of the system. Granularity is the minimum module-size in which a system is broken down and the minimum sized replaceable module of the system is called as granule. Proper

* Corresponding author.
  E-mail addresses: mukherjeeatin@ece.iitkgp.ernet.in (A. Mukherjee),
asd@ece.iitkgp.ernet.in (A.S. Dhar).

choice of granularity helps in minimizing the overall cost for designing of the fault tolerant circuit. But most of the recent literatures that deal with the design of reliable architectures using dynamic reconfiguration consider some specific granularity [9–10,13–14]: fine or coarse [15], and to the best of our knowledge, analysis on choice of the size of the spare modules has not been previously presented.

In this paper, we make an attempt for proper selection of the granule-size that would help the designers to minimize the area and delay overheads of a system for a given reliability depending on the particular requirements. For implementing fault tolerance using dynamic reconfiguration, our first objective is to identify the structural regularity within a circuit and then divide the circuit into symmetrical modules and make the approach systematic. Array of such symmetrical modules is also known as iterative logic array (ILA), where identifying the smaller blocks and defining the proper granule-size for the circuit to make it module-wise fault tolerant not only increase the understandability of the approach, but also make the system easy to debug in future. We compute the specific value of the granule-size for which the total area overhead for fault tolerant design of the given system is minimum, which optimizes the reliability as well the delay overhead. We also formulate an analytical expression for granularity that optimizes the trade-off between the area overhead of a system with the number of faults tolerated maximizing the overall fault coverage.

Major contributions of our work in this paper are as follows:

- We analyze how the area and delay overheads change with different factors such as the granule-size chosen, number of inputs, number of outputs and the interconnections among the intermediate modules and find out the optimal value of the granule. We show that if single-bit granule is chosen instead of the optimal size of 4-bit granule, design of fault tolerant 64-bit ripple carry adder (RCA) requires 20% higher area overhead and that becomes >50% when the chosen granule-size = 64.
- We extend the design approach making it capable of tolerating multiple faults. Keeping in mind about the trade-offs among the granule-size chosen, total area overhead and the number of faults tolerated, we also provide an analytical discussion that helps us to choose the granule-size precisely such that area overhead is minimized for maximum fault coverage. We also prove that for particular selection of granularity, a circuit can tolerate multiple faults instead of just a single fault at the same hardware cost.
- We incorporate *hot-standby* topology that makes the fault tolerant mechanism online, i.e., we do not need any extra time for testing and reconfiguration, and any module identified as faulty is immediately disconnected from the system prohibiting it from further participation in the normal operation.

The rest of the paper is organized as follows. Section 2 describes the theoretical background required for the present work. Considering single fault cases, how the area and delay overheads vary with the granule-size is discussed in Section 3. Section 4 describes the methodology of testing and reconfiguration proposed in the present work. In Section 5, the optimization analysis is done showing how proper selection of the module size plays a significant role in any fault tolerant ILA design. Section 6 considers multiple fault tolerance. In Section 7, the concept for choice of granularity is applied on some real-life digital functional units like RCA, conditional sum adder (CSA), comparator, incrementer, multiplier, Viterbi decoder and COordinate Rotation DIgital Computer (CORDIC) for their optimal fault tolerant designs. The paper is concluded in Section 8.

2. Theoretical background

A fault tolerant approach enabling the autonomous restoration of the defective module in a system, avoiding fault accumulation and re-establishing the correct circuit state in real-time has been presented in

[16]. Self-repairing procedure for permanent faults and in-the-field self-testing for specific applications has been presented in [17]. But these methods believe on offline testing only limiting their usages for real time applications. A detailed study on self-healing approach and its optimization for asynchronous circuits have been developed in [18]. The authors have also highlighted the efficiencies of their method in terms of resource occupation, fault tolerance, reconfiguration speed and capability to tolerate permanent as well as transient faults. But these works have not discussed anything on the granularity, i.e. at which level the extra circuitry for testing and reconfiguration should be added.

Selection of granularity for achieving different trade-offs among cost, performance and recovery time for fault tolerant designs using TMR has been discussed in [19]. In case of dynamic reconfiguration, there are two levels of adding redundancy: One is *coarse-grain redundancy* (CGR) approach [20] that uses spare rows and columns to an array tolerating clustered defects. But it has limitations in tolerating multiple, distributed random defects. The other one is *fine-grain redundancy* (FGR) approach [21] that uses spare wires eliminating the need for rerouting and minimizing timing variance due to correction. At high defect levels, it requires lower area overhead than CGR, but at lower defect rate, CGR requires less area overhead than FGR [15]. Combination of CGR and FGR has been efficiently used in many systems to tolerate random distributed defects as well as clustered and bridging faults. In case of CGR, proper choice of granularity plays an important role in achieving higher reliability at lower area and delay overheads [22]. Some works are available in literature that shows the trade-off among reliability, redundancy and performance of the system [15, 23–25] for changing the granularity. But their discussions are completely system specific. Here we have derived a generalized formula to calculate the optimum granularity for any system having structural regularity.

The outputs of the currently operative modules are monitored by a fault detection and reconfiguration unit that activates the spare module in place of a working module upon identifying faults in the later. Area overhead increases for the spare module as well as for the two levels of multiplexers (MUXes) those are needed for proper routing of the inputs to and the outputs from the non-faulty active modules bypassing the faulty one. If we vary the size of the granule, the area and delay overheads differ due to the change in the number of extra MUXes required for input and output signal selection depending on the granularity chosen. Hence we need to optimize these overheads with proper choice of the granularity. In most of the practical cases, the modules of a circuit are interconnected and extra MUXes are required at the interconnections among the modules for proper routing of the signals through the non-faulty ones. Here, the number of MUXes used for selection of signals at intermediate connections among the modules decreases with increase in the size of the granule and hence proper selection of granularity is very important to minimize the overall cost in designing a fault tolerant circuit. In this paper, we analyze how the choice of the granule-size influences the area and delay overheads of the fault tolerant circuit and also determine the optimal size of the granule for a given design for which minimum hardware cost is achieved.

For ease of understanding, we call the minimum possible sized module of a circuit as a 1-bit granule and hence the club of k minimum-sized module as k-bit granule. Representative example of 1-bit granule with a number of primary inputs, b number of primary outputs and c number of interconnected inputs and outputs fed from and to the neighboring granules for some digital module of an arbitrary system suitable for making fault tolerant using dynamic reconfiguration is shown in Fig. 1. The fault tolerant design is also cascadable in nature so that the number of bits to be handled by it can be increased as required by connecting similar circuit blocks.

Our fault tolerant structure can tolerate almost all types of faults like transistor stuck-open, stuck-close faults, input–output stuck-at faults and bridging faults occurring within a single granule. To incorporate