Microelectronics Journal 66 (2017) 48–57

Contents lists available at ScienceDirect

Microelectronics Journal

journal homepage: www.elsevier.com/locate/mejo

A fault-analysis oriented re-design and cost-effectiveness evaluation methodology for error tolerant applications

Tong-Yu Hsieh*, Kuan-Hsien Li, Chen-Chia Chung

Department of Electrical Engineering, National Sun Yat-sen University, Kaohsiung, Taiwan

ARTICLE INFO

Keywords:
Design simplification
Effective yield
Error-tolerance
Error-tolerability
Fault-analysis

ABSTRACT

For error-tolerant applications such as multimedia circuits, the circuits can be re-designed by removing the hardware that contains faults inducing only imperceptible errors for human beings. As a result, the area, critical path delay and power consumption of the circuits can become much lower. The chip yield can thus also be improved. However, this method also sacrifices the inherent error-tolerability of the circuits when they suffer from noises such as soft errors in the field. For IoT (Internet of Things) applications, this issue is critical for cost-effectively extending their lifetime, and should be considered, which however is not addressed in the literature. In this paper we will show that performing fault analysis can make the evaluation of the incurred cost in error tolerability easy to be carried out. A fault-analysis oriented methodology is presented in this work to help users develop and evaluate re-design methods for their applications. In the proposed methodology, selection of fault models and implementation of fault injection and evaluation processes is addressed. According to the evaluation results, non-critical hardware parts are identified and removed. Furthermore, yield and error-tolerability evaluation models are presented. We also apply the proposed methodology to the JPEG2000 image encoding process as a case study. The quantitative analysis results conducted in this case study illustrate the importance of evaluating the adverse impact on error-tolerability.

1. Introduction

Yield improvement for nanometer scale chip manufacturing technology is a hot topic for both the academic and the industry. In the past decades, many techniques have been proposed to adjust the semiconductor manufacturing process or re-design chip architectures. In this work we concentrate on the re-design methods.

Well-known re-design methods include fault-tolerance and defect-tolerance [1]. In these techniques some redundant hardware is added to a target design. By doing so, it is guaranteed that no errors would appear at the outputs of the design. Triple Modular Redundancy (TMR) is one representative example of such techniques. Two additional copies of the target design are added in the chip and a voter is used such that as long as at least two of the copies can functionally work, the chip can work correctly. One issue is that the incurred area or performance overhead for these methods may be high. For examples, the TMR method requires at least 200% area overhead. Another well-known method, called Error Correction Code (ECC), may induce additional 50% computation time [2].

In the recent years a new notion called *error-tolerance* has been proposed, which provides an alternative manner for yield improvement [3]. Different from the conventional test notion that focuses on identifying chips as being fault-free or faulty, error-tolerance further examines the "acceptability" of errors. The basis is on the fact that human beings are insensitive to small variations in multimedia signals such as sounds and colors. As a result, a defective chip is likely to still be acceptable and thereby applicable in some lower-end applications. In this way the *effective yield*, i.e., the ratio of the total number of acceptable chips over that of all manufactured ones, can be improved. Note that here the acceptable chips include fault-free chips and faulty ones that produce only minor errors. In the literature error-tolerance has been applied to many multimedia applications such as H.264 video coding, voice recorder and images [3–7]. These investigation results show that high error-tolerability exists in such applications, and significant improvement on effective yield can be achieved if the notion of error-tolerance is well exploited.

Under the notion of error-tolerance, some faults in a certain parts of the target circuit induce only minor performance degradation, and thus are likely to be tolerable. As a result, a reasonable re-design method is to remove the hardware parts where the tolerable faults are located. This makes sense as the faults (value variations) in these parts do not affect the acceptability of the outputs. This means that which logic

* Corresponding author.
E-mail address: tyhsieh@mail.ee.nsysu.edu.tw (T.-Y. Hsieh).

http://dx.doi.org/10.1016/j.mejo.2017.05.018
Received 25 September 2016; Received in revised form 18 April 2017; Accepted 31 May 2017

value these parts have appears to be a “don’t care condition” with respect to its acceptability. Note that this does not mean that these parts are unnecessary over-designs, but these parts only cause minor precision loss and thus are insignificant. Due to this “don’t care” feature, we can remove these non-critical parts. In this way the area, critical path delay and power consumption of the circuit can be reduced, which in turn results in increase in chip yield. This is the fundamental idea of this work.

In the literature a similar notion called *approximate computing* has been proposed, which has received more and more attentions recently [8–10]. Also there have been several similar design simplification methods. In [11] the authors replace logic gates in the typical ripple carry adder and the carry look ahead one by simplified logic gates. Approximate functions of the adders are thus implemented based on a tolerable threshold of errors. For example, the XOR gate is replaced by an OR gate. Under different acceptable thresholds of errors, the corresponding achievable yield improvements are analyzed. In [12] the authors also target adder designs, but analyze the significance of each hardware part of the circuit on the functional correctness and performance of the circuit. Then the authors simplify the non-critical hardware parts, particularly to reduce their signal switching activities for power saving purposes. In [13] the authors develop an analytical framework to evaluate error characteristics of approximate adders. This framework provides a fast and accurate method for users to select proper simplified adders for their applications. A similar simplified design notion is applied to multiplier circuits in [14], which is also for power savings. While the previous work mentioned above focuses on adders and multipliers, in [15] the sequential circuits have been shown to be simplified as well.

Simplification based re-design methods (e.g., approximate computing), however, also bring some *adverse impacts* on error-tolerability and effective yield. Since the non-critical hardware has been removed, when faults or soft errors appear in the re-designed circuit, the produced outputs are very likely to be unacceptable. In contrast, error-tolerance capability exists in the original design. Such degradation in the inherent error tolerability (and thus the effective yield) would be harmful for some applications such as IoT (Internet of Things). For IoT applications, error-tolerance capability will be important for extending the lifetime of the system [16]. As long as the resulting errors are acceptable, the system can still be used without need to repair. As a result, the cost for maintaining the system can be reduced. This issue, however, is not addressed in the literature.

In this work we address the issue mentioned above by proposing a fault-analysis oriented re-design and cost-effectiveness evaluation methodology. The proposed methodology starts with fault model selection, fault injection and simulation, as well as fault analysis. Based on the fault analysis results, circuit re-design methods are then developed and the achievable benefits are evaluated. Moreover, the fault analysis results are further applied to the presented yield and error-tolerability models to evaluate the incurred cost. Our methodology shares the same objective of removing non-critical hardware with that for approximate computing. However, the previous work of approximate computing does not address the evaluation of the incurred adverse impacts on error-tolerability and effective yield due to design simplification. We believe this issue is also critical for users to determine a feasible re-design method, especially for applications where error-tolerance capability is important for extending their lifetime. In this paper we show that fault-analysis can make this evaluation easy to be carried out. The evaluation models are established and the evaluation results illustrate the importance of analyzing the adverse impact on error-tolerability and effective yield.

In our previous work [17], we have done some preliminary work on the fault analysis process. By targeting the JPEG2000 image processing system [18], in [17] we have demonstrated how fault analysis can facilitate the error-tolerability evaluation and non-critical hardware identification processes. In this work we make further extensions, which is described below. To make this paper self-contained, most contents in [17] will also be included.

1.1. Technical contribution

1. A fault-analysis oriented re-design and cost-effectiveness evaluation methodology is proposed. This methodology can be used to evaluate the incurred adverse impacts on error-tolerability and effective yield due to design simplification (e.g., approximate computing). This evaluation, however, is not addressed by the previous work. The proposed evaluation methods can also be integrated with the existing design simplification methods to provide deeper information for developing a proper design approximation scheme.
2. Fault-analysis is conducted to enable design simplification as well as benefit and adverse impact evaluation simultaneously in the proposed methodology. This is also not investigated in the literature. The work of [17] concentrates on only the fault analysis process and the results. How these results are utilized to develop re-design methods and evaluation of these methods is not addressed.
3. Each step of the proposed methodology is qualitatively explained in details, including establishment of the evaluation models and surveys of the related references. A case study on the discrete wavelet transform (DWT) circuit (a core circuit of the JPEG2000 image encoding system) is also provided to illustrate how the proposed methodology can be carried out. In this work we target the DWT circuit because it has great error-tolerability, which is beneficial as a case study to illustrate the ideas and benefits of the proposed methodology. Although this circuit can be easily mapped onto a DSP or a media processor due to its regularity in both control flow and data access, we still find that there are a number of studies on hardware implementation of this circuit in the literature. We believe that this is because that hardware implementation of this circuit can accelerate the JPEG 2000 coding/decoding process. We also believe that applying the proposed method to the whole JPEG2000 system to evaluate the yield is a very good future work. This, however, requires further study on the fault acceptability of the whole system, and is actually one of our on-going research. In this paper our focus and contribution is the development of the methodology. Detailed steps are also described and explained to help users apply the proposed methodology to error tolerable systems/applications. The quantitative analysis results conducted in this case study also illustrate the importance of analyzing the impact on error-tolerability and effective yield due to design implication.

In the rest of this paper we first present the proposed methodology in Section 2. Section 3 then describes the JPEG2000 codec and the DWT circuit as background. The fault analysis process contained in the proposed methodology is then illustrated by using DWT in Section 4. Based on the obtained fault analysis results, Section 5 demonstrates various possible re-design methods. The achieved benefits and the associated costs for the developed re-design methods are then evaluated in Section 6. Finally, Section 7 concludes this paper.

2. Proposed methodology

Fig. 1 gives the flow chart of the proposed methodology. At first the users need to determine the target design that is error tolerant. In the literature, it has been shown that multimedia processing circuits are good candidates. The target design can be a single module, a block or a system. Based on the determined design, a total of six steps are then executed, which is explained as follows.

2.1. Step 1: Select fault model

Common fault models in the literature include single stuck-at faults, transition faults, cell-aware faults, and so on [19,20]. The users may