Microelectronics Reliability 80 (2018) 124–133

Contents lists available at ScienceDirect

Microelectronics Reliability

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

Design for Small Delay Test - A Simulation Study☆

Matthias Kampmann*, Sybille Hellebrand

Paderborn University, Warburger Straße 100, Paderborn 33098, Germany

ARTICLE INFO

Keywords:
Faster-than-at-speed-test
Small Delay Faults
BIST
X-tolerant compaction
X-masking

ABSTRACT

Early Life Failures (ELFs) are becoming an important reliability issue in state-of-the-art technologies. ELFs can be indicated by Small Delay Faults (SDFs), however, some SDFs may not be detectable even with modern at-speed tests. For these hidden SDFs, Faster-than-at-Speed Test (FAST) provides a solution. However, FAST imposes new challenges on the test method. Unknown logic values (X-values) are a major challenge in FAST, due to the increased frequencies. A special Design-for-FAST architecture relying on an accordingly adjusted scan configuration and a simple, but efficient X-masking scheme can support X-tolerant compaction in the context of FAST. This work analyzes the trade-offs of this concept within the framework of a standard industrial workflow and presents a comprehensive case study. Simulation results indicate that for some designs, the conventional synthesis workflow does not produce optimal circuit behavior under FAST. In these cases, the Design-for-FAST approach can increase the fault efficiency, while at the same time reducing the amount of X-values in the test responses considerably.

1. Introduction

Modern technology nodes allow for both tight integration densities and reduced supply voltages and leakage powers, thereby enabling an optimized performance. However, recent technologies are also getting more susceptible to manufacturing challenges like process variations, which may, in extreme cases, lead to marginal hardware. Shortly after manufacturing, the marginal hardware is functional and passes the test. However, it can quickly degrade and cause an *Early Life Failure (ELF)* of the system, which is becoming an increasingly important issue for applications with high reliability demands. While traditional burn-in tests try to avoid ELFs by expedited aging, the prediction of ELFs already during manufacturing test can help to reduce the overall test cost considerably. As experiments at the Stanford University have shown, *Small Delay Faults (SDFs)* are possible indicators for ELFs [3].

To detect an SDF, a pattern pair must propagate a transition along a path through the fault site and cause a timing violation at the output. If the additional delay introduced by the fault is very small, or the slacks of the possible paths in the circuit are very large, SDFs may be undetectable or *hidden* even with state-of-the-art timing-aware at-speed tests [4–7]. Consequently, such faults are called *Hidden Delay Faults (HDFs)*.

Faster-than-at-Speed Test (FAST) [8–14] samples the outputs of the circuit at a frequency higher than the nominal operating frequency of the device. It can make even very small HDFs detectable, as increasing the frequency corresponds to reducing path slacks. Typically, multiple frequencies are used, which can be up to three times higher than the operational frequency in practical applications [15,16]. In [17] and [18], a silicon evaluation has confirmed the effectiveness of FAST.

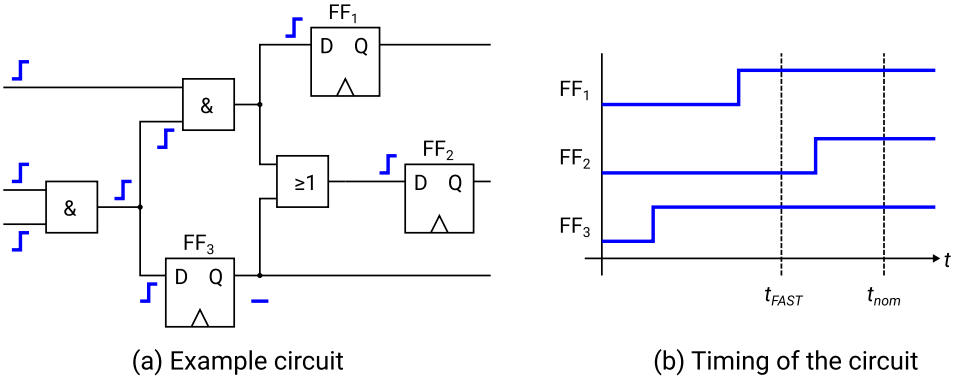
Since low-cost Automatic Test Equipment (ATE) is often not able to generate the frequencies required for FAST, on-chip clock generation is preferred. Using clock generation schemes like the programmable capture generator [19] or launch and capture clock generation [20], it is possible to generate the desired frequencies. Although in scan-based test, only the launch and capture for the second pattern are overclocked, increased IR-drop and overheating are challenges associated with FAST. They have been addressed in [11,12], where the authors performed an extended simulation to determine the IR-drop occurring during FAST. FAST IR-drop increases the switching delays of the elements in a circuit, causing chips to fail the test even though they are defect-free. Thus, IR-drop can be treated as a sort of frequency-dependent delay offset that needs to be considered during frequency selection. Further steps towards a complete Built-in Self-Test (BIST) solution have been presented in [21], where the appropriate selection of FAST frequencies has been mapped to an optimization problem and solved in [22].

A major challenge associated with FAST is a high and potentially unbounded number of unknown logic values (X-values) at the outputs, as not all paths might have finished their computation at the time the outputs are sampled (cf. Fig. 1).

☆ This work is based on the preliminary results presented in [1] and [2].

* Corresponding author.
E-mail addresses: matkam@mail.uni-paderborn.de (M. Kampmann), sybille.hellebrand@uni-paderborn.de (S. Hellebrand).

https://doi.org/10.1016/j.microrel.2017.11.019
Received 31 July 2017; Received in revised form 31 October 2017; Accepted 15 November 2017



(a) Example circuit (b) Timing of the circuit

Fig. 1. An example circuit and its timing during FAST.

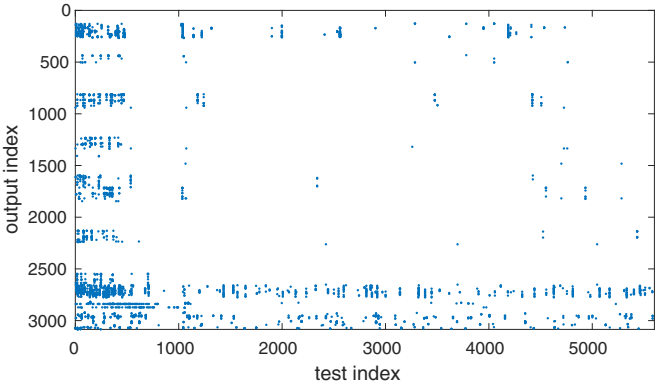
In Fig. 1a, an example circuit with three Flip-Flops (FFs) is shown, with the corresponding signal waveforms of the data inputs given in Fig. 1b, when rising transitions are applied at the inputs and FF_3 contains a 0. At the nominal observation time t_{nom}, (corresponding to the nominal frequency f_{nom}), all signals are stable. However at the earlier observation time t_{FAST}, FF_2 captures a non-stable signal value. Pessimistically, such values are treated as X-values during simulation.

The amount of X-values captured during FAST is much higher than during a conventional or at-speed test, since the X-values caused by timing violations add up to the already existing X-sources present in the Circuit Under Test (CUT). Furthermore, the amount also increases with the frequency, since more aggressive timings cause more paths to fail and the IR-drop induced by FAST rises [11,12]. As a result, X-tolerant compaction is required to handle high and varying rates of X-values. Although test response compaction in the presence of unknowns has been in the focus of research for many years (cf. e.g. [23–25]), the specific requirements of FAST still need further investigations.

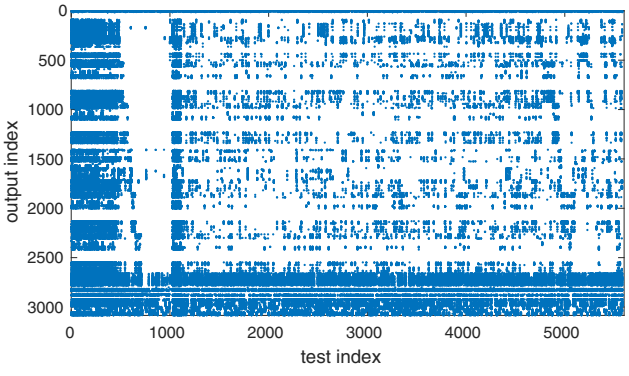
A first approach specifically tailored for FAST [26] uses multiplexers to select X-free outputs during each shift cycle. However, this approach suffers from large pattern sets needed to obtain desired fault efficiencies. In [21], the X-canceling Multiple Input Signature Register (MISR) [24,27] is combined with an on-chip memory for storing intermediate signatures that can be downloaded and processed offline after the test. The number of intermediate signatures, and thus the memory size, depends on the X-rates at the MISR inputs. For more efficient solutions, the typical scenario of FAST must be better exploited. In [1] it was observed that the X-values during FAST are not equally distributed over the outputs of the circuit, but rather follow a distinct distribution, as shown in Fig. 2.

This figure shows the distribution of X-values over the outputs of the ITC'99 benchmark circuit b18_1 [28]. The same pattern set was simulated for two different observation times: 50% of t_{nom} (cf. Fig. 2a) and 33% of t_{nom} (cf. Fig. 2b). In each diagram, the X-axis shows the test pattern index, and the Y-axis shows the output index. A dot in the diagram indicates that the output with given index captures an X-value when the corresponding test pattern is applied. It can be seen that there are distinct patterns in the distribution of X-values. Multiple outputs capture X-values at the same patterns, whereas others are (almost) X-free during the complete test. In the following, the characterization of an output with respect to X-values is also called the *X-Profile* of that output.

Knowledge about the X-profiles of the outputs is very beneficial for X-tolerant compaction. For example, if it is known that a certain output will capture a large number of X-values at a target frequency, that output can simply be masked before reaching the compactor. This has been exploited in the past during scan-chain stitching as well. For instance, in [29], a small subset of scan-chains has been marked as specific "X-chains", which contain only Scan-Flip-Flops (SFFs) capturing a lot of X-values during test. Masking one of these "X-chains" can greatly reduce the number of X-values going into the compaction hardware. Its



(a) 50% of t_{nom}



(b) 33% of t_{nom}

Fig. 2. X-Distribution of the b18_1 circuit for two different observation times.

effectiveness has been confirmed in combination with an X-canceling MISR in [30].

In [1], this approach has been extended for FAST. The X-Profile for each SFF is explicitly determined and used to create clusters of SFFs that have similar or equal X-Profiles for all provided FAST frequencies. This allows very fine-grained control over the distribution of X-values over all scan-chains. The clustering approach does not require a reordering of existing scan-chains. Instead, the implemented algorithm provides constraints to the synthesis tool guiding the scan stitching. This way, the approach can be easily integrated into existing design flows. The "Design-for-FAST" approach is completed by specially tuned algorithms generating the control data for a very simple and efficient masking scheme [2].

In this work, the techniques developed in [1] and [2] are evaluated in the context of a standard design flow. For this, the complete Design-for-FAST workflow has been integrated into an industrial tool chain, ultimately showing a more practical approach to utilize the Design-for-