Quantification of Power Based Hardware Trojan Detection Under Realistic Variations and Separation of Power Supplies

Mehmet Kocabas, Alp Arslan Bayrakci

Department of Computer Engineering
Gebze Technical University
mkocabas@gtu.edu.tr, abayrakci@gtu.edu.tr

Abstract

Advances in technology and the increasing economic burden of manufacturing fabs resulted in integrated circuit (IC) manufacturing outside the homeland. The availability of intentional insertion of malicious circuitry (hardware Trojan) drew concerns towards hardware Trojan detection. Side channel analysis based on power is one of the proposed detection methods. However, the main enemy of the method is unavoidable process variations which can easily hide the effect of Trojan. In this paper, using precise Spice simulations, we quantify the difficulty of Trojan detection based on power under realistic variations. Also, a method is devised to decrease the number of additional pins when separating the circuit into many grids, which is shown to be a necessity for Trojan detection in actual integrated circuits.

1. Introduction

Hardware Trojan (HT) issue has become a hot research topic for especially the last decade. Due to mostly economical reasons, for most of the countries including the developed ones, it is impossible to manufacture all operation critical chips in home country. This makes especially the manufacturing stage vulnerable to any intentional attacks. Any adversary can insert a malicious circuitry to alter the circuit characteristics and functionality as all layout is handled to the manufacturing fab.

There is a lot amount of research in both academia and defence industry [1] as well as articles [2] in IEEE Spectrum magazine focusing on the topic. There is a large variety of HT shown to be possible and proposed in the literature [3].

The first approach is to break up the manufactured circuit by stripping layer by layer and using electron microscope and reverse engineering to check whether there is any inserted HT inside. This is called destructive method but as the insertion can be done selectively to only a fraction of chips, this does not guarantee the authenticity of the chip. The method is also too expensive and time consuming. The functional test can be used to reveal Trojan existence but the Trojans can be designed to be passive until a specific or a series of specific rare events occur. This makes it impossible to detect HT by conventional functional tests. Yet, side channel analysis (SCA) can be used to reveal Trojan existence. By looking at observable circuit signals like delay and power, the effect of HT can become evident. However because of manufacturing variations on parameters like transistor gate length and threshold voltage, each copies of the same chip comes with different delay and power characteristics which can easily hide the effect of the inserted Trojan. This makes parameter variations the main source of failure for SCA based Trojan detection.

Regional, localized power measurement for HT detection is devised to reveal the Trojan effect on power consumption [4, 5] but most research on SCA based Trojan detection methods either lack a realistic, accurate variation model considering all inter and intra die variations with spatial correlations or use large Trojans based on counters to be able to detect. Yet, tiny sized but smart Trojans can harm the circuit impressively, which complicate SCA based detection. Utilization of simulation models less precise than Spice may result in deviations from the actual case. Need for golden model is another shortcoming for most SCA based HT detection methods.

In this paper, we first apply a realistic variation model explained in Section 4.2. For all results in the paper, the precise transistor level Spice simulations in a Monte Carlo fashion are performed. This paper aims to quantify the effect of Trojan on power, the effect of parameter variations on HT detection, the detection performance using single vs. multiple power supplies with layout separation to grids, the required level of grid separation for a successful detection without a need for golden model, the effect of circuit inputs on detection. In Section 3.3 we also devise a circuit that decreases the number of additional input pins to make the grid separation applicable to actual circuits. Results on our test circuit synthesized with 45nm cell library, reveals the importance of the separation level of the circuit layout and the negative effect of variations on avoiding the power based HT detection.

Section 2 gives the preliminaries to clarify power based HT detection. Section 3 explains the utilized variation model, separation the circuit to grids and the proposed circuitry to decrease the number of additional inputs. Section 4 gives all results to quantify the effect of variations, grid size, inputs and circuit size on power based HT detection. Last section concludes the paper.

2. Preliminaries

2.1. Instantaneous power

Side channel analysis methods are based on the analysis of observable and measurable quantities like path delay, temperature and total power consumption of the circuit. Instantaneous power shown in (1) is observable as the supply voltage (V_{supp}) is known and current from the voltage supply ($i(t)$) can be measured.

$$P(t) = i(t) \times V_{supp} \tag{1}$$

The current in this equation depends on the transistor gate length and threshold parameters which are highly affected from manufacturing process variations [6]. This constitutes the main bottleneck of power based Trojan detection as the resultant variation on the instantaneous current can surely hide the effect of any potential malicious insertion to the circuit.

In conventional power based HT detection the whole circuit is connected to single power supply. Measuring the current from supply and computing power and energy, the different samples (chips) can be investigated to reveal any Trojan existence. It is expected that circuits with Trojan consume much power due to the additional malicious circuitry and can be detected as they deviate from the ones without any Trojan. However, the detection probability is highly related with the applied inputs and the parameter variations as will be clarified later in Section 4.

2.2. Hardware Trojan

There are lots of different types of hardware Trojans so that the taxonomy of Trojans has been worked on [3]. In general, HT circuitry includes two parts: payload and trigger. The trigger part is responsible from the activation of the Trojan, whereas the payload part alters the signal in the circuit when it is triggered. Without loss of generality, we assume a hardware Trojan with a payload of one xor gate and a trigger of one nand gate similar to devised in [7]. Such a two gate Trojan is enough to both alter the circuit and stay passive until activated by a specific input. This makes it impossible to be detected by functional tests. Whenever the output of nand connected to side input of xor becomes logic-1, it gets activated and inverses the circuit signal. Yet, SCA based methods can detect the Trojan without activating it.

3. Power Based HT Detection Using Separation to Grids

3.1. Modeling Parameter Variations

The main issue of SCA based hardware Trojan detection is non-negligible and unavoidable manufacturing process variations. Each manufactured copy of the same chip behaves differently due to these variations. It is very difficult to discriminate the effect of Trojan on power consumption from the effect of the variations. Therefore, it is very important to model variations accurately. Otherwise the HT detection results would be very optimistic and misleading.

In order to model process variations accurately one must model the properties of the variations. The parameter variations occur both between different chips (inter-die) and inside the same chip (intra-die) [8]. The amount of variations for the corresponding technology node is also important. Underestimating that amount results in misleading optimistic HT detection results. One other property of variations is the spatial correlations, i.e. correlation due to location. Spatial correlations occur due to the inherent nature of the manufacturing defects and cause the transistors located in a close proximity have similar random parameter values. Any model with smaller variation amounts or ignoring intra die variations or ignoring spatial correlations may cause misleading results. In this paper, considering the size of the circuit, we employ a 4 level version of the quad tree model proposed in [9], which can model both inter and intra die variations as well as the spatial correlations where the actual variation amounts are taken from ITRS report [6].

3.2. Separating Power Supplies

One of the main weakness of power based HT detection is that all consumption in the circuit is through a single supply voltage. When this is combined with the effect of variations, it is very difficult to discriminate the effect of the Trojan from the whole circuit. One approach [5] to isolate Trojan in the circuit is separating the circuit into regions and using multiple power supplies to collect regional power consumption. For that purpose, we divide the circuit layout into $n \times n$ grids during the design phase. Each grid must have its own power supply pin so that, after the manufacturing, the current for that grid can be measured separately. The total power consumption can be written as the sum of the power consumption at each grid as shown by (2).

$$P(t) = \sum_{x,y} P(x, y, t) \tag{2}$$

where x is the x-index and y is the y-index of the corresponding grid in the circuit. Both can take values from 0 to $n - 1$.

As the number of grids, i.e. n increases, the Trojan gets isolated more and the possibility to discriminate any Trojan effect on power increases. However, this also increases the number of input supply pins as each supply voltage requires a new pin connection. For instance, when n is 8, a total of 64 supply input pins are required. In this paper, for a test circuit, we also search for the required n value for number of additional inputs to be less whereas the grid size to be small enough to capture any Trojan effect.

For a grid (x, y) and a sample (chip), power and energy per transition are defined as shown in (3) and (4). t in these equations is time and tr_i is the time period required for dynamic power consumption during the transition for input i.

$$P(x, y, t) = i(x, y, t) \times V_{supp}(x, y) \tag{3}$$

$$E(x, y, tr_i) = \int_{tr_i} P(x, y, t)dt \tag{4}$$

3.3. I/O Issue of Separated Grids Approach

The main problem with separating power supplies is the need for additional supply pins for the circuit. The number of grids for accurate HT detection may get very large for larger circuits with plenty of transistors. In most real cases, providing each grid with a different supply may not even be possible other than being complicated. In this paper, we propose a decoder based circuit shown in Fig. 1 for exponentially decreasing the required number of additional pins.

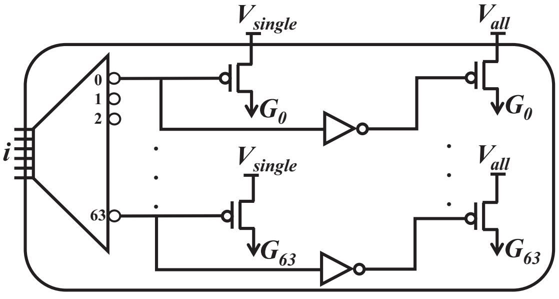

Figure 1. Circuit to provide individual power supply

This circuit is made up of a 6-to-64 line inverting decoder, inverters and PMOS transistors. It illustrates the case for 64 (8x8) grids. Note that each PMOS turns on and links supply to a grid only if the gate of the PMOS is connected to low voltage. The single voltage supply (V_{single}) is used to provide voltage only for one selected grid. The grid selection is done by input i. For instance, if input i is 24 then only grid 24 (G_{24}) is supplied by V_{single} because the PMOS connected to 24^{th} decoder output turns on. All remaining grids are supplied by V_{all}. Therefore, by this setup, the power consumption of each grid can be measured individually by utilizing only $log_2(N) + 1(V_{single})$ additional pins where N is the number of total grids in the circuit. For instance, for 16×16 grid case, instead of 256 supply pins only 9 additional pins are enough.

The number of additional pins can even be decreased to 2, independent of the number of grids in the circuit. For that purpose, a shift register can be utilized to hold i inside. Yet, this would increase the experiment duration due to the fact that i is entered to shift register serially instead of parallel in Fig. 1. In such a case, for instance for 256 grid case, setting i would require 8 clock cycles. But additional pins would be 2 instead of 9. Therefore, the choice of shift register is more meaningful when plenty of grids are required in extra large circuits or when there is no room for additional pins in the chip packaging.

4. Results

4.1. Experimental Setup

Variation model explained in 3.1 applied to the test circuit. The inter and intra die variations are assumed to have equal portion of variation [8] and the spatial correlations are applied by quad-tree model [9]. Two most effective circuit parameters, transistor gate length (L_{eff}) and threshold voltage (V_{th}) are assumed to be random as devised in [10]. The variation amounts are defined according to ITRS report [6], which states $3\sigma/\mu$ ratio of transistor gate length and threshold voltage are 12% and 35% respectively and for 45nm process node.

The test circuits are 64-bit and 16-bit integer adders with a size of 576 and 144 two input nand gates respectively. Without loss of generality the Trojan is taken to be as a two input xor gate as payload and a two input nand gate as trigger part inserted at the carry out of bit-46 for 64-bit adder and bit-15 for 16-bit adder.

The circuit is synthesized using 45nm Nangate open cell library [11] and Spice decks are constructed for 1000 samples according to our variation model in a Monte Carlo fashion. 500 samples are Trojan-free and the other 500 samples include the Trojan. Each sample circuit is tested with 64 randomly generated 128-bit and 128 randomly generated 32-bit input vectors respectively for 64-bit and 16-bit adder and given by 1ns time intervals. Precise Spice power simulations with increased sensitivity (as small as 1e-6 reltol value) are performed on both test circuits and for each sample and each of 1000 inputs. Simulation of one sample requires about half an hour to finalize.

In Section 4.2, we demonstrate the effect of realistic process variations on HT detection. Section 4.3 investigates the separation of power supplies and the corresponding impact on Trojan detection. The dependence of HT detection on the given input, on the number of grids and on the circuit size are explained in Sections 4.4, 4.5 and 4.6.

4.2. Effect of Variations on Conventional Power Test

In this subsection the effect of variations on hiding the Trojan and the inadequacy of the conventional power tests are demonstrated. By conventional test, we mean the case where there is only one or a couple of power supplies for the whole circuit. Before showing the results please note that detection of the Trojan looking at power consumption highly depends on the input. This is empirically shown in next subsection. In our tests on 64-bit adder, the Trojan effect can best be observed when 48^{th} random input is given. Yet, even 48^{th} input does not reveal Trojan when conventional test is preferred. Under the realistic variation model and when the conventional power test using single power supply is applied, the resultant power consumption for 48^{th} input is shown by (a) and the total energy consumption during the transition for the 48^{th} input is shown by (b) of Fig. 2.

Throughout the paper, the green circles represent the samples (chips) without Trojan and the red asterisks represent the samples with inserted Trojan. In the instantaneous power ($P(t)$) plots, each asterisk or circle belongs to an instantaneous power for a specific sample and specific time instance where the asterisks or circles for the same sample are connected with lines. On the other hand, each asterisk or circle in the energy ($E(tr)$) plots belongs to the total energy consumed for a specific sample. Thus, the x-axis of instantaneous power plots is time showing the whole transition period for an input and the x-axis of the energy plots shows the sample index that the asterisk or circle belongs to.

In Fig. 2 (a), instantaneous power for each sample and time instance are plotted over the transition time period of 48^{th} input which is between 48ns and 48.3ns and in (b) energy consumed during the same transition period is shown for each sample. The x axis in (b) shows the sample index from 1 to 500 (500 Trojan-free and 500 Trojan-inserted samples). Results clearly show that the samples with and without Trojan are totally intermixed for both instantaneous power and energy, which shows that it is not possible to discriminate any Trojan solely looking at power from single source even if you give the right input pair to the circuit.

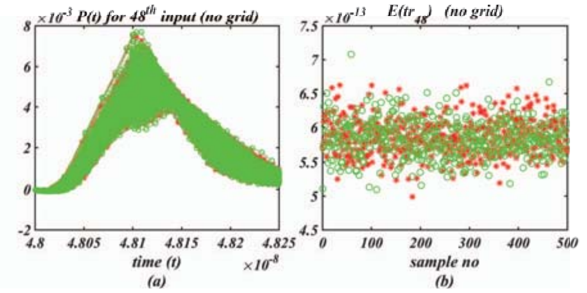

Figure 2. Conventional power based HT detection with single power supply for 48^{th} input (64-bit adder): (a) Power (in Watt) vs. time (b) Energy (in Joule) vs. sample no

4.3. Results for Separated Power Supplies

Instead of one supply in the previous section, we divide the 64-bit adder circuit layout into n-by-n grids where n is 2, 4 or 8. Again for 48^{th} input, the power (a1), (a2), (a3) and energy (b1), (b2), (b3) graphs are plotted in Fig. 3 respectively for 2×2, 4×4 and 8×8 grid divisions. The figures are plotted for the grid where Trojan is located, which is (2, 2), (3, 4) and (6, 7) respectively for 2×2, 4×4 and 8×8 grid. The red asterisks correspond to samples with Trojan and the green circles correspond to Trojan-free samples. Please be aware that there would be no red or green plots for the tester but all would be the same from the perspective of the tester as the tester does not know the

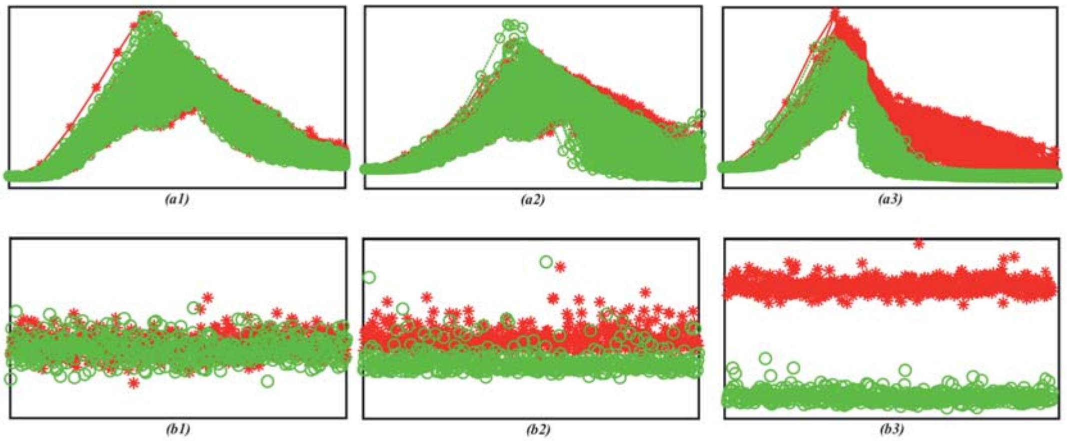

Figure 3. Power vs. time and energy vs. sample no plots for 64-bit adder (same x, y axis and scale as in Fig. 2): P(t) for (a1) 2 × 2 grid (a2) 4 × 4 grid (a3) 8 × 8 grid and E(tr_{48}) for (b1) 2 × 2 grid (b2) 4 × 4 grid (b3) 8 × 8 grid

Trojan inserted samples a priori. Fig. 3 reveals the following facts:

- It is better to use energy per transition ((b1), (b2) and (b3)) instead of instantaneous power ((a1), (a2) and (a3)) to reveal the Trojan existence.
- Observing the results for 2×2 grid ((a1) and (b1)) shows that separating the layout into 4 grids doesn't reveal HT existence for our test circuit. Even separating to 16 grids ((a2) and (b2)) cannot reveal HT existence.
- Separating the circuit into 64 grids and observing the energy plot in Fig. 3 (b3) can reveal the Trojan inserted circuits. The good point is that for that separation a golden model is not required, because all the samples located above are Trojan-inserted whereas all samples at below are Trojan-free.

4.4. Dependence on the Input

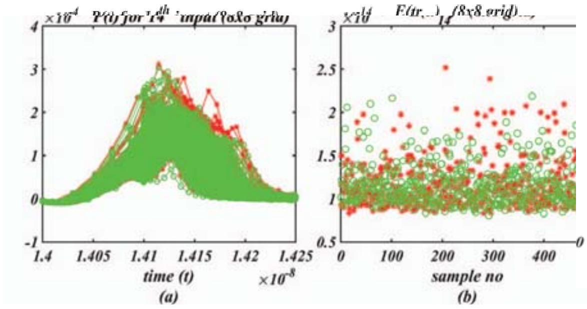

Figure 4. Power vs. time and energy vs. sample plots for 14^{th} input for grid (6, 7) (64-bit adder): (a) P(t) (in Watt), 8 × 8 grid (b) E(tr_{14}) (in Joule), 8 × 8 grid

Until now, we used the 48^{th} input as it is revealed in our experiments to be the best candidate for HT detection among the 64 inputs simulated by Spice for each sample. To demonstrate the effect of a different input on detection, Fig. 4 shows the results for another random input (14^{th}). Although the 8 × 8 grid is used neither the instantaneous power nor the energy can reveal any Trojan. The importance of using right inputs can be seen by comparing the energy plot in Fig. 4 with Fig. 3 (b3).

4.5. Dependence on the Grids

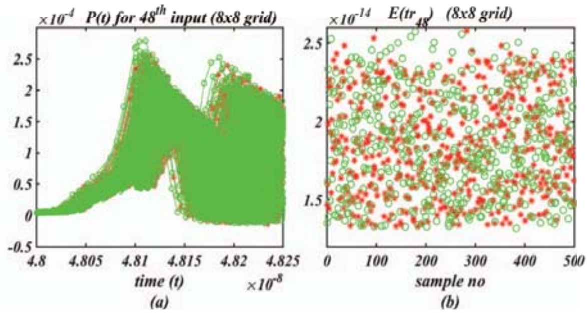

Figure 5. Power vs. time and energy vs. sample no plots for 48^{th} input for grid (5, 6) (64-bit adder): (a) P(t) (in Watt), 8 × 8 grid (b) E(tr_{48}) (in Joule), 8 × 8 grid

As the circuit is divided into grids, it is important to use the measurements from the right grid where the right grid include any part of the Trojan to observe its effect on energy or power. In our case, the Trojan is inserted at grid (6, 7) in 8 × 8 grid. Therefore Fig. 3 (a3) and (b3) belong to that grid. In Fig. 5 the power and energy for grid (5, 6) are plotted as (a) and (b) respectively. Although it is a neighboring grid for the grid containing the Trojan and even the best input (48^{th}) is used, the intermixing at the resultant plot in Fig. 5 clearly shows the necessity to use the power consumption for the grid on which the Trojan is located.

4.6. Dependence on the Circuit Size

Lastly it is important to show the relationship with the separation to grids and the circuit size. For that purpose we change the test circuit and apply 1000 random sample test for 16-bit adder circuit. Until now, all plots belong to the 64-bit adder. Fig. 6 is same as Fig. 3 except it belongs to 16-bit adder cir-

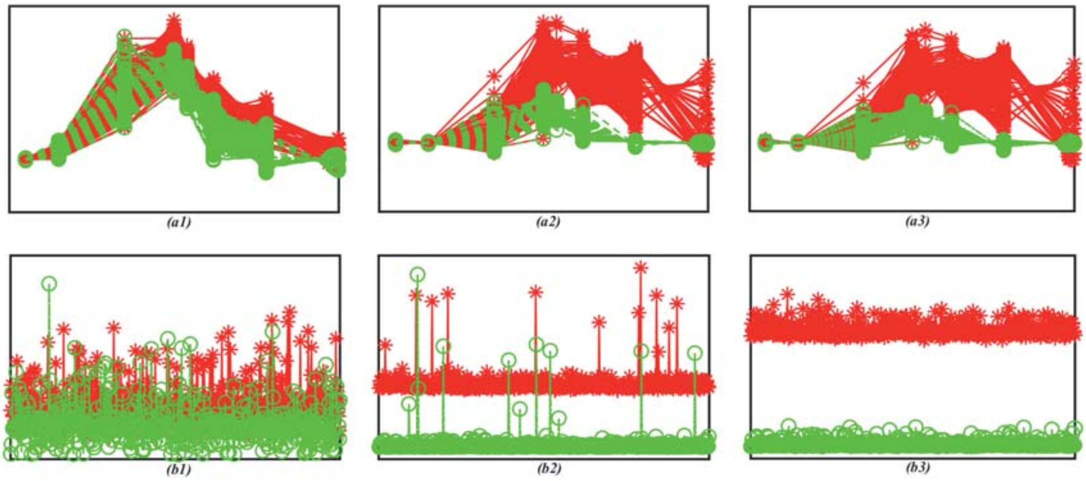

Figure 6. Power vs. time and energy vs. sample no plots for 16-bit adder (same x, y axis and scale as in Fig. 2): P(t) for (a1) 2 × 2 grid (a2) 4 × 4 grid (a3) 8 × 8 grid and E(tr_{82}) for (b1) 2 × 2 grid (b2) 4 × 4 grid (b3) 8 × 8 grid

cuit. This time, the 82^{nd} random input reveals the Trojan best and it is used in this figure. When we compare two figures, it is observed that the Trojan in 16-bit adder can be revealed even by 4 × 4 grid separation, which is expected as the 16-bit adder circuit has fewer number of transistors.

5. Conclusion

It is shown that under realistic variations it is impossible to reveal small Trojans when only one power supply is used (Fig. 2), even when enough number of circuit separation is not performed (Fig. 3). It is also not possible if the power consumption of a grid without the Trojan is measured (Fig. 5) or an input not activating the Trojan is given to the circuit (Fig. 4). It is possible to reveal the Trojan in the circuit by only separating the circuit into enough number of grids and measuring the energy per transition for a right input and for each grid. But a loose grid separation or any error in choosing the grid or input substantially decrease the Trojan detection probability. This shows the difficulty of HT detection under realistic variations.

Considering that the actual circuits are larger than the 64-bit adder circuit used for these plots, the separation may require too many grids and distinct power supplies. Observing Fig. 3, it can be deduced that only separating the circuit into 64 grids could reveal the Trojan existence for 64-bit adder circuit with 2304 transistors. By a rough estimation, 4096 grid separation is required for a circuit having about 150,000 transistors. Employing such a large number of additional supply pins is surely impossible but using the decoder circuit proposed in Fig. 1 can resolve that problem. Even the number of additional pins can be decreased to two by using shift register as devised in this paper, but with the additional cost of measurement duration.

6. References

[1] A darpa apprach to trusted microelectronics. https://www.darpa.mil/about-us/darpa-approach-to-trusted-microelectronics. Accessed January 2019.

[2] Stopping hardware trojans in their tracks. https://spectrum.ieee.org/semiconductors/design/stopping-hardware-trojans-in-their-tracks, Jan 2015. Accessed January 2019.

[3] Mohammad Tehranipoor and Farinaz Koushanfar. A survey of hardware trojan taxonomy and detection. *IEEE design & test of computers*, 27(1), 2010.

[4] Mainak Banga and Michael S Hsiao. A region based approach for the identification of hardware trojans. In *2008 IEEE International Workshop on Hardware-Oriented Security and Trust*, pages 40–47. IEEE, 2008.

[5] Xiaoxiao Wang, Hassan Salmani, Mohammad Tehranipoor, and Jim Plusquellic. Hardware trojan detection and isolation using current integration and localized current analysis. In *2008 IEEE international symposium on defect and fault tolerance of VLSI systems*, pages 87–95. IEEE, 2008.

[6] http://www.itrs2.net/. International Technology Roadmap for Semiconductors (ITRS) Report 2011 Edition.

[7] F. N. Esirci and A. A. Bayrakci. Hardware trojan detection based on correlated path delays in defiance of variations with spatial correlations. In *Design, Automation Test in Europe Conference Exhibition (DATE), 2017*, pages 163–168, March 2017.

[8] J. Cong, P. Gupta, and J. Lee. Evaluating statistical power optimization. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 29(11):1750–1762, Nov 2010.

[9] A. Agarwal, D. Blaauw, and V. Zolotov. Statistical timing analysis for intra-die process variations with spatial correlations. In *ACM/IEEE International Conference on Computer Aided Design (ICCAD)*, 2003.

[10] Dennis Sylvester, Kanak Agarwal, and Saumil Shah. Invited paper: Variability in nanometer cmos: Impact, analysis, and minimization. *Integration, the VLSI Journal*, 41(3):319–339, May 2008.

[11] http://www.nangate.com/. NanGate 45nm Open Cell Library.