



Hardware Trojan Detection Leveraging a Novel Golden Layout Model Towards Practical Applications

Yanjiang Liu¹ · Jiaji He¹ · Haocheng Ma¹ · Yiqiang Zhao¹

Received: 27 January 2019 / Accepted: 26 June 2019 / Published online: 20 July 2019
© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract

Globalization trend in integrated circuit design and manufacturing process has increased the vulnerability of integrated circuit. These vulnerabilities mainly caused by hardware Trojan have a serious impact on the security of integrated circuits. Although side-channel analysis approach is the most promising Trojan detection approach, nearly all side-channel analysis approaches rely heavily on the availability of golden chips, which are extremely difficult to obtain. In this paper, a golden layout model instead of fabricated golden chips is introduced for the practical application of hardware Trojan detection approaches. The simulated voltage variations generated from the golden layout model at different process corners serve as golden reference, thus fabricated golden chips are not required during detection. Further, silicon measurements are performed to obtain the voltage variations of fabricated chips, and a model calibration algorithm is utilized to calibrate the golden model in the presence of process variations and random noise. Finally, the Trojan detection is formulated as a two-class classification problem, and the Trojan is identified using the partitioning around medoids algorithm. Experimental results demonstrate that the similarities between the simulated traces and measured traces are greater than 98.81%, and the proposed approach distinguishes the Trojan chips correctly even under $\pm 15\%$ process variation.

Keywords Integrated circuit · Hardware trojan · Side-channel analysis · Golden layout model · Model calibration algorithm · Partitioning around medoids algorithm

1 Introduction

With an ever growing of economic globalization, fabless has become a main trend in the integrated circuits (ICs) market. Although such trend dramatically lowers the fabrication cost and reduces the time to market, it also provides an opportunity for adversary to implant hardware Trojan into the original design. Once the hardware Trojan is triggered, it can cause functional changes, information leakage, privilege escalation or even system failure. As a well-designed malicious modification, hardware Trojan has become a serious threat to the security of critical

applications spanning on the domains of space, military, communication, finance, and so on [45].

Given the catastrophic consequences caused by hardware Trojan, various detection approaches have been explored over the past decades. It can be broadly classified into: 1) destructive (e.g. reverse engineering) and 2) non-destructive (e.g. logic testing, run-time monitoring and side-channel analysis) [17]. Among all existing Trojan detection approaches, side-channel analysis approaches have been widely recognized by researchers. Starting with the global power consumption-based method presented in [1], numerous side-channel parameters, including the leakage current [10, 40], transient current [29], path delay [24, 26], electromagnetic emanation [4, 27, 43], temperature [5], thermal signal [37, 46], as well as multiple parameters combinations [33, 36], have been explored to strengthen the side-channel analysis approaches. The primary part of side-channel analysis approaches is golden chips, which are exploited as golden reference to identify the existence of hardware Trojan by comparing with chips under test. However, the real case is that the golden chips are extremely

Responsible Editor: M. Hsiao

✉ Yiqiang Zhao
yq.zhao@tju.edu.cn

¹ School of Microelectronics, Tianjin University, Tianjin, 300072, China

difficult to achieve, thus, side-channel analysis approaches have always been criticized for its applicability.

To address this issue, several golden chip-free Trojan detection approaches are proposed over the past few years, which shift the need of fabricated golden chips to golden model and identify the Trojan by comparing with the simulated traces of golden model [11, 13, 20–22, 30, 35, 47–49, 51, 53]. Majority of existing golden chip-free Trojan detection approaches concentrate on the gate-level simulation, however, several problems introduced by place and route, such as the timing imbalances, parasitic effects and coupling effects, are not taken into consideration during the golden model establishment process. Only in [30], the post-layout simulation data serves as the golden reference, but the intra-die variations and random noise are not taken into account during the model calibration process.

In this paper, a golden model instead of fabricated golden chips is proposed for practical application of hardware Trojan detection approaches. For the golden modeling, layout data is used to perform the post-layout simulation, and the simulated voltage variations at different process corners are regarded as golden reference. Further, a model calibration algorithm, including the data denoising process and model compensation process, is proposed to calibrate the golden model with the measured traces. During the model calibration, a data denoising algorithm is exploited to reduce the random noise mixed into the measured traces, and then a model compensation algorithm is proposed to compensate the golden model with the denoised traces. Finally, the partitioning around medoids algorithm (PAM) is used to classify the calibrated traces and denoised traces, and the hardware Trojan is identified when it falls outside the class of golden model. The main contributions are listed as follows.

- A golden layout model is established and the simulated voltage variations generated from the golden layout model serve as golden reference. There is no need for fabricated golden chips during detection.
- A model calibration algorithm is exploited to calibrate the golden model, which makes the golden model matches well with the actual silicon measurements.
- Fabricated chips instead of Field Programmable Gate Arrays (FPGAs) are utilized to validate the proposed approach, and the Trojan chips are identified correctly using the PAM algorithm.

The remainder of this paper is organized as follows. Section 2 investigates the attack model, related works about the golden chip-free Trojan detection approaches and power simulation methods. Section 3 presents the theoretical analysis of golden layout modeling methodology, and Section 4 gives the Trojan detection framework, model calibration algorithm and partitioning around medoids

algorithm. Section 5 analyzes the experimental results and Section 6 concludes this paper.

2 Background

2.1 Attack Model

Through outsourcing the manufacturing process to the third party foundry, a malicious adversary in foundry may insert hardware Trojans into the fabricated chips. Therefore, the hardware Trojan attack scenario is mainly at the third party foundry, and the synthesized gate-level netlist and layout are assumed to be trusted. The supply voltage variation is chosen as the side-channel parameter instead of other parameters like electromagnetic emanation, delay, thermal and so on, because there exist several matured commercial simulation tools (e.g. HSPICE, Nanosim [23] and Spectre [8]) to obtain the voltage variations accurately.

2.2 Golden Chip-Free Trojan Detection Approaches

Most existing side-channel analysis approaches require several golden chips for reference, however, fabricated golden chips are not always available during detection. To address this problem, several golden chip-free Trojan detection approaches have been proposed over the past decades, which replace the requirement of golden chips with golden model. The method presented in [20] proposes golden electromagnetic model. In [30], a golden current model is built using the HSPICE. Besides, authors regard the hardware Trojan detection as a two-class classification problem in [48] and train the classification algorithms using the simulation data. The Trojan chips are identified when it fall outside the classification boundary. Due to the randomness over the manufacturing process, it is extremely difficult to obtain an accurate model of process variations and thus there exist obvious differences between the simulated traces of golden model and the measured traces of fabricated chips in real applications. Therefore, such approaches suffer from low detection sensitivity with ever-decreasing Trojan size because the influences of well-designed Trojans are minute and carefully hidden within the design margins allowed for process variations.

Taking the side-channel traces as its own reference, self-referencing method eliminates the influences of inter-die variations and avoids the requirement of fabricated golden chips at the same time, which has become a popular research direction. A regional self-referencing approach is proposed to divide a large design into several small sub-regions, and then the relationship of transient current among the different sub-regions is analyzed and the hardware Trojan is identified when the relationship is changed [11,

22, 47]. This idea is also extended to detect the recycled chips [53]. However, some types of Trojans, like always-on and parametric changes type, this approach is unable to detect it. Besides, a temporal self-referencing approach is presented in [35], the transient current with its own at two different time windows is compared to detect the Trojan and improvements are made in [21]. Such approaches are effective in identifying the rarely-activated sequential Trojans, but it is only applied to the case that fully aware of the implementation of Trojan during detection. Moreover, a spatial self-referencing approach is proposed in [13, 49, 51], which identifies the path of implanted hardware Trojan when the delay differences among the symmetry paths exceed the threshold value. Indeed, there exist millions of paths that possibly inserted the hardware Trojan and the verification effort has increased exponentially for achieving a high detection sensitivity.

2.3 Power Simulation Methods

Concerning the power simulation of circuit, several papers have been explored over the past several decades. In general, some circuit simulators, like the HSPICE, Nanosim and Spectre, are exploited to carry out the power simulation. Four models, including the Hamming distance [9], Hamming weight [7], Hamming distance zero to one and Hamming distance one to zero [32, 39], are proposed to obtain the power traces of cryptographic circuit. Moreover, Daisuke et al. [14] introduce a time series divided parasitic capacitance model to acquire the power traces quickly for performing the side-channel attacks. From [34], the ON current of a switching gate is presented and the total transient current is obtained by adding up all the switching gates. In [19], a failure physics model is proposed, which simplifies the digital circuit as a second-order band pass filter.

While there are some works giving similar solutions already, they mainly concentrate on the gate-level simulation [7, 9, 14, 20, 30, 32, 34, 39, 48]. However, some

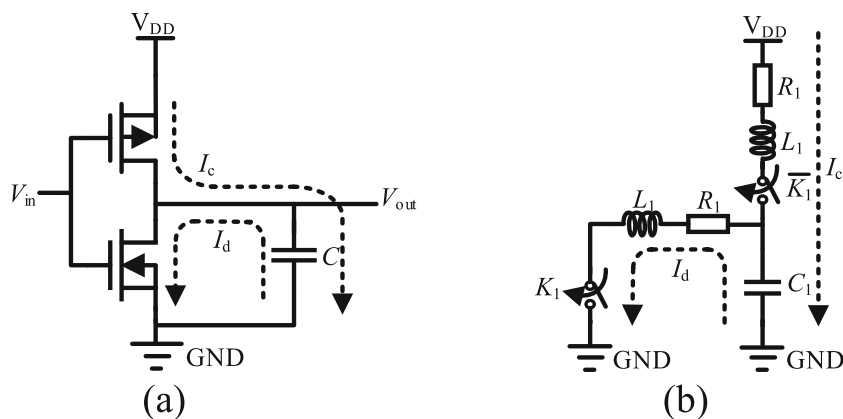
problems, like the timing imbalances, parasitic effects and coupling effects and so on, introduced by the place and route have not taken into account in the gate-level simulation. As the technology today has shrunk to nanometer dimensions, these problems become more and more evident which affect the side-channel traces seriously. Addressing this issue, the post-layout simulation has been explored over the past decades, including the side-channel attack evaluation [6, 44], circuit performance prediction [25, 31, 52], Trojan detection [30] and so on. Accordingly, the post-layout simulation is performed and regarded as golden reference in this paper, and this golden reference matches well with the actual silicon measurements by comparing with the [30]. This is mainly because the inter-die variations, intra-die variations and random noise are all considered during the model calibration process.

3 Golden Layout Modeling Methodology

In this section, the switching characteristics of a basic logic gate, i.e. inverter, are analyzed firstly, and then the simplified switch-based model of an inverter is given. Further, the simplified switch-based model of chip is deduced by summing over all the switching gates, and the Trojan detection based on the simplified switch-based model is presented finally.

Figure 1 illustrates the traditional structure and simplified switch-based model of an inverter. When the input V_{in} switches from logic “1” to logic “0”, the load capacitance C is charged and the charging current I_c flows from the power port V_{DD} to the ground port GND. Contrarily, C is discharged and there exist a discharging current I_d between the C and NMOS. The charge-discharge process of an inverter is illustrated in Fig. 1a. The I - V characteristics of an inverter can be modelled as charging and discharging process of a series RLC circuit depicted as the Fig. 1 (b). Where R_1 , L_1 and C_1 respectively are the lumped resistance, inductance and capacitance, which can be

Fig. 1 Traditional structure and simplified switch-based model of an inverter. **a** Traditional structure and charge-discharge process of an inverter. **b** Simplified switch-based model and charge-discharge process of an inverter



extracted from the layout using the Calibre xRC [8]. When the V_{in} switches from logic “1” to logic “0”, the switch K_1 and \bar{K}_1 are turn on and off respectively, and C_1 is charged. On the contrary, K_1 and \bar{K}_1 are turn off and on respectively, and C_1 is discharged. According to the engineering circuit analysis theory [18], the transient current $I_g(t)$ flows through the ground port is presented in Eq. 1. Where the voltage $V_{C_1}(0)$ across the C_1 is equal to V_{DD} and 0 respectively before the C_1 is discharged and charged.

$$L_1 \frac{dI_g(t)}{dt} + R_1 I_g(t) + \frac{1}{C_1} \int_0^t I_g(t)dt - V_{C_1}(0) = 0 \quad (1)$$

The Laplace transformation of Eq. 1 is presented in Eq. 2.

$$L_1 C_1 s^2 I_g(s) + R_1 C_1 s I_g(s) + I_g(s) = 0 \quad (2)$$

By using the inverse Laplace transformation to Eq. 2, $I_g(t)$ is obtained and depended on the R_1 , L_1 and C_1 .

For a large-scale digital chip which consists of millions or billions of gates, the total transient current flows to the ground plane is determined by adding up the transient current of all switching gates of chip. Thus, the simplified switch-based model of a gate described in Fig. 1b is parallel with each other to form the switch-based model of chip which is depicted in Fig. 2.

Summing over all the switching gates in the chip, the total transient current $I_{tot.g}(t)$ of golden chip is described in Eq. 3. Where n is the total number of switching gates of chip. In addition, several correlations between the switching characteristics of the different CMOS gates are also taken into account. For example, the output of a CMOS gate feeds to the input of another gate would certainly cause their switching times to cascade, overlap, and so on. Where Δt_g is the switching time difference of $I_g(t)$ caused by correlations between the switching gates.

$$I_{tot.g}(t) = \sum_{g=1}^n I_g(t - \Delta t_g) \quad (3)$$

In the presence of a hardware Trojan in the chip, the total transient current $I_{tot.t}(t)$ of Trojan chip is shown in Eq. 4.

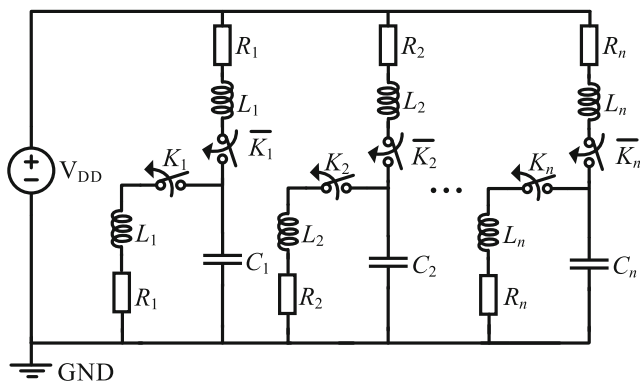


Fig. 2 The simplified switch-based model of chip

Where $I_{ht}(t)$ is the total transient current incurred by h switching gates of hardware Trojan. Although hardware Trojan keeps silence in the majority of life, the Trojan’s trigger part keeps active and monitors the internal logic value or state all the time. Therefore, h is always greater than 0 and the $I_{ht}(t)$ makes it possible to detect the Trojan. Moreover, the structural changes of chip introduced by the hardware Trojan, may change the value of parasitic parameters of m switching gates near the hardware Trojan, and the $\Delta I_{tot.g}(t)$ is the total current fluctuations of these switching gates.

$$I_{tot.t}(t) = I_{tot.g}(t) + I_{ht}(t) + \Delta I_{tot.g}(t) \quad (4)$$

Comparing the Eqs. 3 and 4, hardware Trojan distorts the total transient current of chip, and these distortions ($\Delta I_{tot.g}(t)$ and $I_{ht}(t)$) make it possible to detect the Trojan.

4 Golden Chip-Free Statistics Side-channel Analysis and Hardware Trojan Detection

4.1 Trojan Detection Framework

The overall framework of golden chip-free Trojan detection is shown in Fig. 3, including the pre-silicon simulation phase, model calibration phase, and Trojan detection phase. In the pre-silicon simulation phase, the spice netlist, parasitic parameters, and stimuli are feed into the power simulator and the simulated traces of golden model at different process corners are obtained, which is $I_S(t)$. In the model calibration phase, the measured traces $I_F(t)$ of fabricated chips are acquired using the ASIC-platform, and a data denoising algorithm is utilized to eliminate the random noise mixed into the $I_F(t)$, and a model compensation algorithm is exploited to calibrate the $I_S(t)$ with the $I_{DG}(t)$. The denoised traces of golden chips and chips under test respectively are $I_{DG}(t)$ and $I_{DC}(t)$, while the calibrated traces of golden model are $I_M(t)$. Finally, the partitioning around medoids algorithm (PAM) is exploited to classify the $I_M(t)$ and $I_{DC}(t)$, and the chips under test are classified as golden chips or Trojan chips during the Trojan detection phase.

4.2 Golden Model Calibration

The principal basis of golden chip-free Trojan detection methodology is to find the differences between the simulated traces generated from golden model and the measured traces of chips under test. To ensure a better Trojan detection sensitivity, golden model should be designed for matching well with the actual silicon measurements. However, random noise and process variations exist in the real experiment, and affect the side-channel traces of fabricated chips

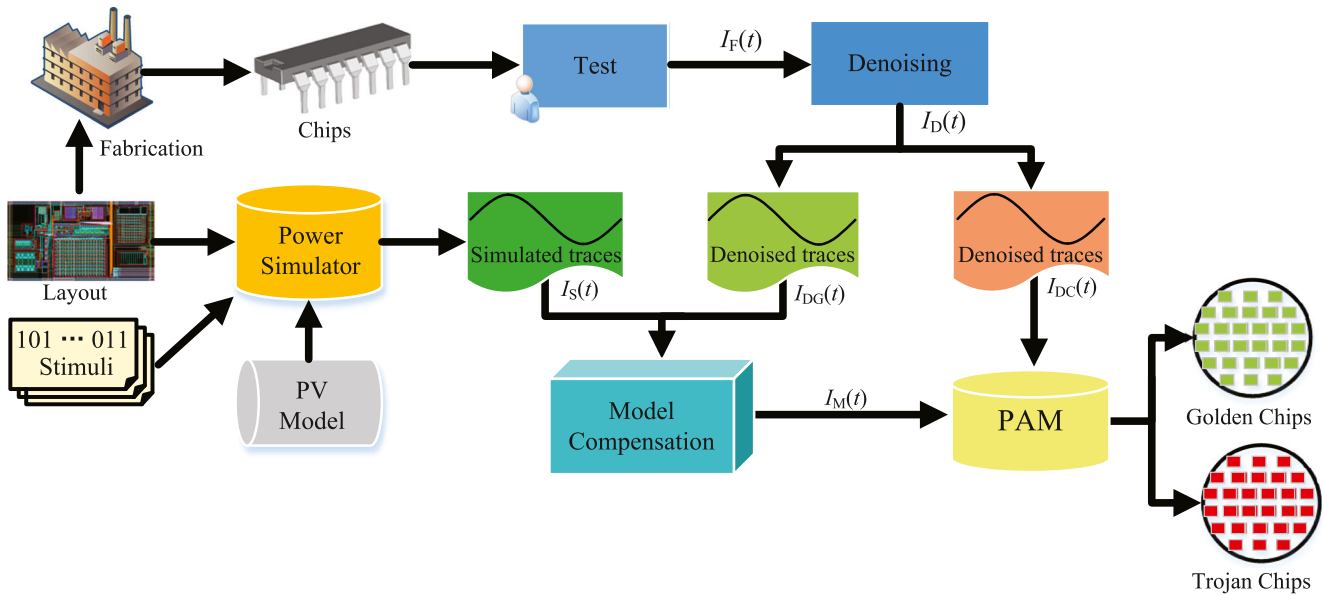


Fig. 3 Overall framework of Trojan detection

seriously. For a well-designed Trojan, the influences of Trojan are minor and easily hidden within these differences, thus, the detection sensitivity is reduced with decreasing Trojan sizes and increasing circuit sizes [20, 21]. Therefore, process variations and random noise should be taken into consideration during the golden model establishment process, and a model calibration technique is necessary in order to make the golden model matches well with the actual silicon measurements.

4.2.1 Denoising Analysis

As for the random noise, a data denoising algorithm combines empirical mode decomposition (EMD) with the Savitzky-Golay filter is exploited to reduce the influences of random noise mixed into the measured traces. Compared with the other traditional decomposition methods, EMD decomposes the signal into a set of intrinsic mode functions (IMFs) without any predefined basic functions [3]. After the decomposition process, the random noise mixed into each IMF is filtered out using the Savitzky-Golay filter [42], and several denoised IMFs are summed to obtain the denoised signal.

The first trace I_{F_1} of $I_F(t)$ is used to explain the whole denoising process, and the decomposed signal IMF_a after the decomposition process is presented in Eq. 5. Where c is the number of IMFs, r is the final residue, and IMF_i is the i -th IMF.

$$I_{F_1} = IMF_a + r = \sum_{i=1}^c IMF_i + r \tag{5}$$

The decomposition process is completed until the current residue r is monotonic. The decomposition process is described in Eq. 6. Where IMF_0 is the measured trace I_{F_1} , and m_{i-1} is the mean value of the upper and lower envelope of IMF_{i-1} .

$$IMF_i = IMF_{i-1} - m_{i-1} \tag{6}$$

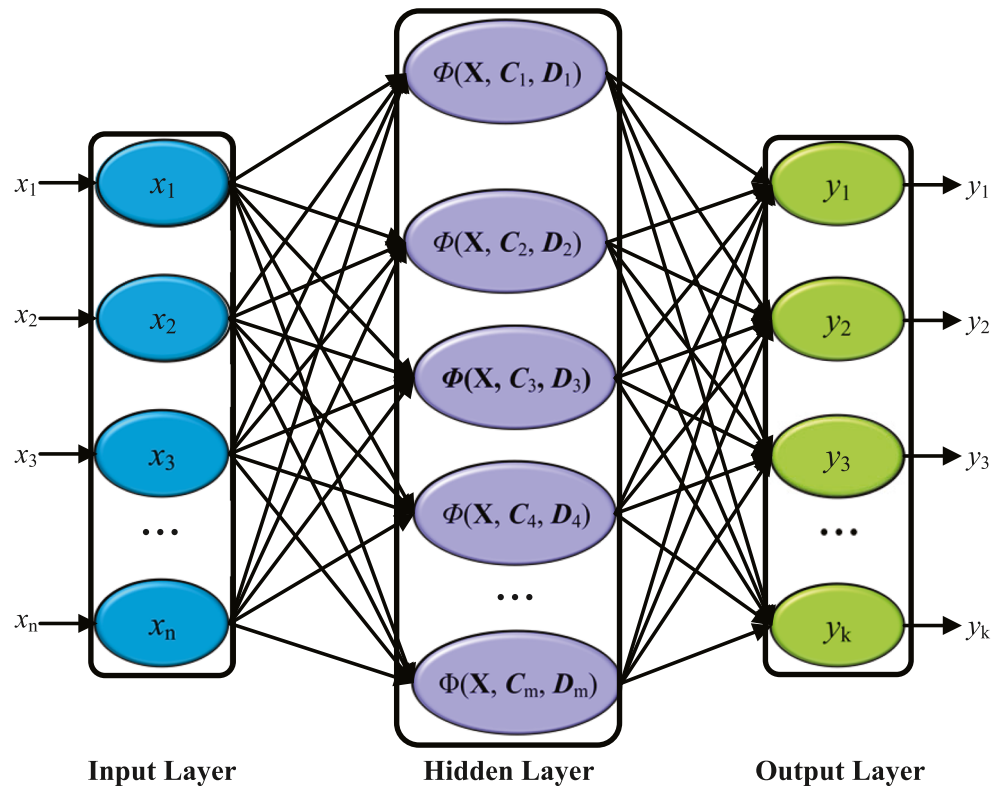
After decomposed by the EMD, all IMFs are denoised using the Savitzky-Golay filter, and the denoised trace I_{D_1} is obtained as described in Eq. 7. Where IMF'_i is the denoised IMF_i .

$$I_{D_1} = \sum_{i=1}^c IMF'_i \tag{7}$$

4.2.2 Golden Model Compensation

Radial basis function neural network (RBFNN) is a feed-forward neural network, which is consisted of three layers shown in Fig. 4: input layer, pattern layer and output layer. For the RBFNN, the multilayer feed-forward network structure not only avoids the cumbersome calculations in back-propagation network, but also overcomes the problem of local minimum in the gradient descent algorithm. Moreover, RBFNN involves a hybrid learning process and the weights between hidden layer and output layer are adaptive adjusted dynamically. In summary, the simple structure, fast learning process and high approximation accuracy have made RBFNN an ideal choice to be used in the nonlinear regression applications. For the influences introduced by process variations, there exist some mismatches between the simulated traces and

Fig. 4 Traditional structure of RBFNN



measured traces, however, these mismatches between the $I_S(t)$ and $I_{DG}(t)$ can be determined with any given accuracy using the radial basis function neural network (RBFNN) [2, 16, 28, 41]. The $I_S(t)$ and $I_{DG}(t)$ are used to explain the nonlinear regression process of RBFNN, and the relationship f between $I_S(t)$ and $I_{DG}(t)$ is presented in Eq. 8.

$$I_S(t) \xrightarrow{f} I_{DG}(t) \Rightarrow I_{DG}(t) = f(I_S(t)) \quad (8)$$

The input layer with n neurons (e.g. x_1, x_2, \dots, x_n) receives n -dimensional input data set $I_S(t) = \{I_{S1}, I_{S1}, \dots, I_{S1}\}$. The hidden layer calculates the radial basis function value $\Phi(X, C, D)$ of each hidden neuron, which is presented in Eq. 9. To ensure the local approximation ability, the Gaussian function is used as the radial basis function in this paper. Where the $C(i) = \{c_{i1}, c_{i2}, \dots, c_{in}\}$ and $D(i) = \{d_{i1}, d_{i2}, \dots, d_{in}\}$ are the center and spread of i -th hidden neuron respectively. The value of output layer is the sum of the multiplications of the weights and the output values of hidden neurons and shown in Eq. 10. Where y_j is the j -th output neuron and w_{ij} is the weight between the j -th output neuron and i -th hidden neuron.

$$\Phi(I_S(t), C_i, D_i) = \exp\left(-\left\|\frac{I_S(t) - C_i}{D_i}\right\|^2\right) \quad (9)$$

$$y_j = \sum_{i=1}^m w_{ij} \Phi(I_S(t), C_i, D_i) \quad (10)$$

Finally, in the following Eq. 11, the root mean square error RMSE between the output layer and expect output (denoted as $I_{DG}(t)$) is calculated to evaluate the fitting performance. Where k is the number of output neurons and $I_{DG}(t) = \{I_{DG,1}, I_{DG,2}, \dots, I_{DG,k}\}$. The nonlinear regression process is completed when the RMSE is less than the fitting error ϵ , otherwise, the weight w_{ij} , center C_i and spread D_i are adjusted iteratively using the gradient descent algorithm.

$$RMSE = \sqrt{\frac{\sum_{j=1}^k (y_j - I_{DG,j})^2}{k}} \quad (11)$$

For n fabricated golden chips, the side-channel traces are always differed from the pre-designed one due to the effect of process variations, which are denoted as $I_{DG_1}(t), I_{DG_2}(t), \dots$, and $I_{DG_n}(t)$, thus, the corresponding relationship f is not consistent with each other. The Eq. 8 can be re-written as the Eq. 12. Where f_i is the mapping relationship between the $I_S(t)$ and $I_{DG_i}(t)$.

$$\begin{Bmatrix} I_{DG_1}(t) \\ I_{DG_2}(t) \\ \vdots \\ I_{DG_n}(t) \end{Bmatrix} = \begin{Bmatrix} f_1(I_S(t)) \\ f_2(I_S(t)) \\ \vdots \\ f_n(I_S(t)) \end{Bmatrix} \quad (12)$$

In the real scenario, only a small amount of fabricated golden chips required from the model calibration process can be verified using reverse engineering, which incurs

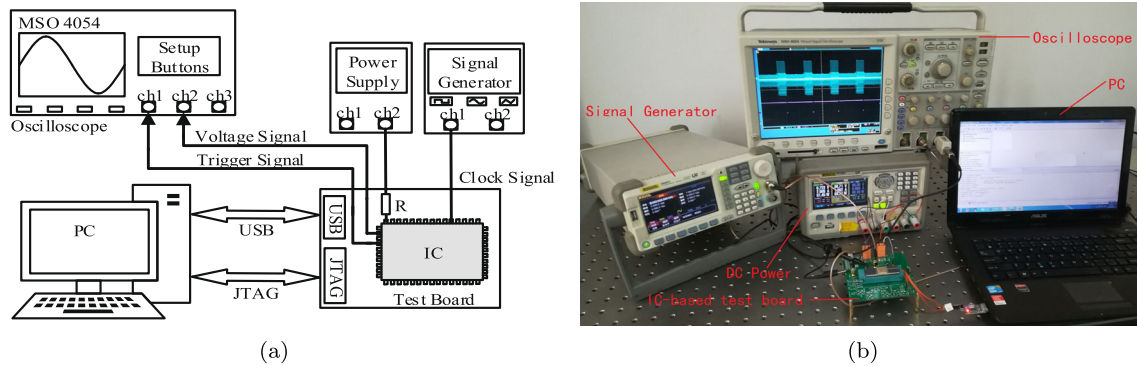


Fig. 5 Experiment setup. **a** Block diagram of ASIC-Platform. **b** Photograph of ASIC-Platform

are fabricated in Chartered 180 nm technology [15], and this library provided by foundry is utilized to carry out the post-layout simulation.

5.2 Golden Reference Generation

Monte Carlo simulations under $\pm 5\%$ PV, $\pm 10\%$ PV and $\pm 15\%$ PV are performed in HSPICE, and 20 simulated traces under $\pm 5\%$ PV randomly selected are illustrated in Fig. 6. Each trace represents the voltage variations of circuit during an encryption process. From Fig. 6, it can be observed that all traces are similar and there exist several differences at some points actually. Analysis results show that process variations can vary the maximum difference within 44.8 mV, and more than 90% differences in magnitude fall in 8.7 mV. Euclidean distance is an effective similarity measurement method, which can be used to differentiate the hardware Trojan when the differences exceed the threshold value. The Euclidean distances between the voltage variations of golden circuit and Trojan circuit are calculated and the results are shown

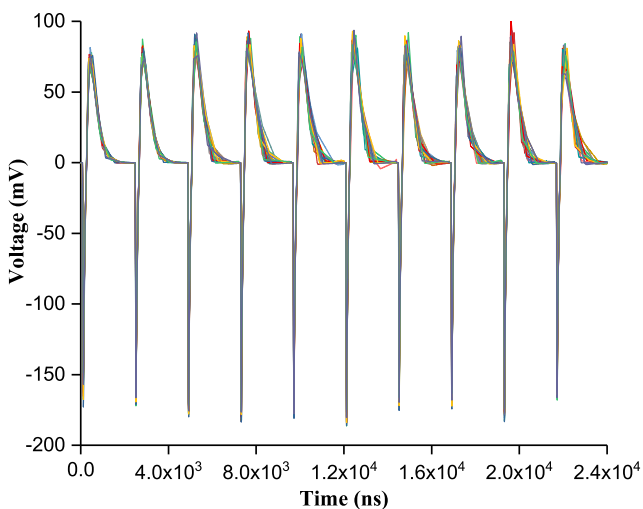


Fig. 6 20 simulated traces of golden model

in Fig. 7. Regarding Fig. 7, the Euclidean distances are ranged from 0.1 to 0.22, and the Euclidean distances of Golden, Trojan1, Trojan2 and Trojan3 overlap with each other. This results indicate that the side-channel differences caused by hardware Trojan are minor and easily masked by the process variations. Therefore, the hardware Trojan cannot be differentiated with simple measurement.

Besides, the voltage variations of fabricated chips are exposed to all kinds of random noise in the real experiment and thus the differences between the simulated traces and measured traces are further increased. Therefore, a model calibration technique is exploited to make the golden model matches well with the real case.

5.3 Golden Model Validation

The calibrated results are shown in Fig. 8. The denoised traces $I_{DG}(t)$ are denoted as Golden_actual, while the traces denoted as Golden_sim5, Golden_sim10, and Golden_sim15 respectively represent the $I_M(t)$ under $\pm 5\%$ PV, $\pm 10\%$ PV and $\pm 15\%$ PV. From results in the Fig. 8a, Golden_actual is an over-damped response, and all traces are too similar to distinguish the obvious differences from each other. Actually, there exist a few differences at some points as shown in Fig. 8b.

Figure 9 shows the distribution histogram of these differences. The differences between the Golden_actual and Golden_sim5, Golden_sim10, Golden_sim15 respectively are denoted as Error_5, Error_10 and Error_15. More specifically, the maximum difference is only 29.65 mV and more than 90% differences in magnitude fall 10.86 mV. This kind of differences mainly caused by the process variations, the influences of ultra_small Trojan are easily masked by these differences. In this paper, Trojans are set as always activated for validating the proposed approach better and a test generation technique is considered to activate the hardware Trojan effectively in the future.

To further demonstrate the whole similarity between the $I_M(t)$ and $I_{DG}(t)$, the correlation analysis is performed

Fig. 7 Euclidean distances between the golden circuit and Trojan circuit

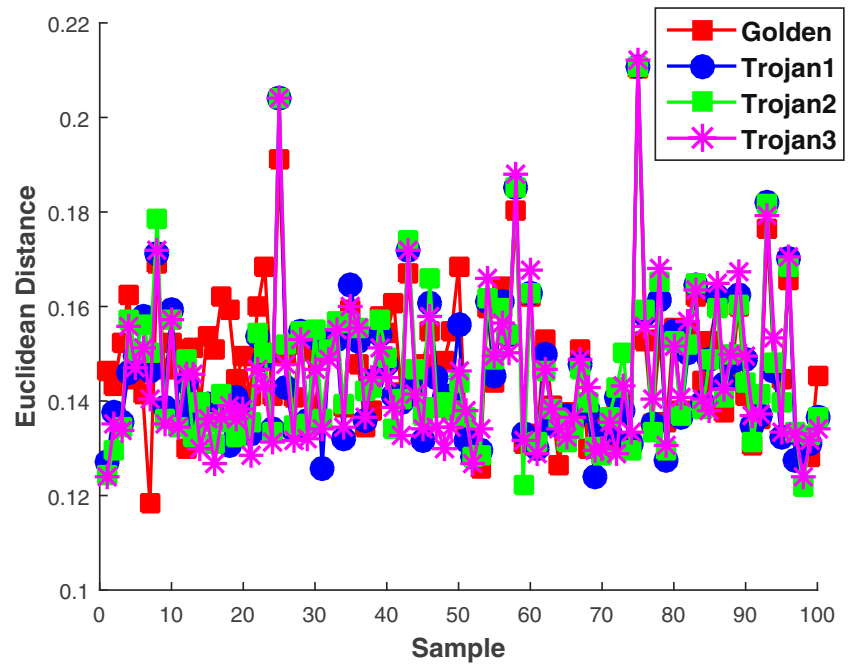
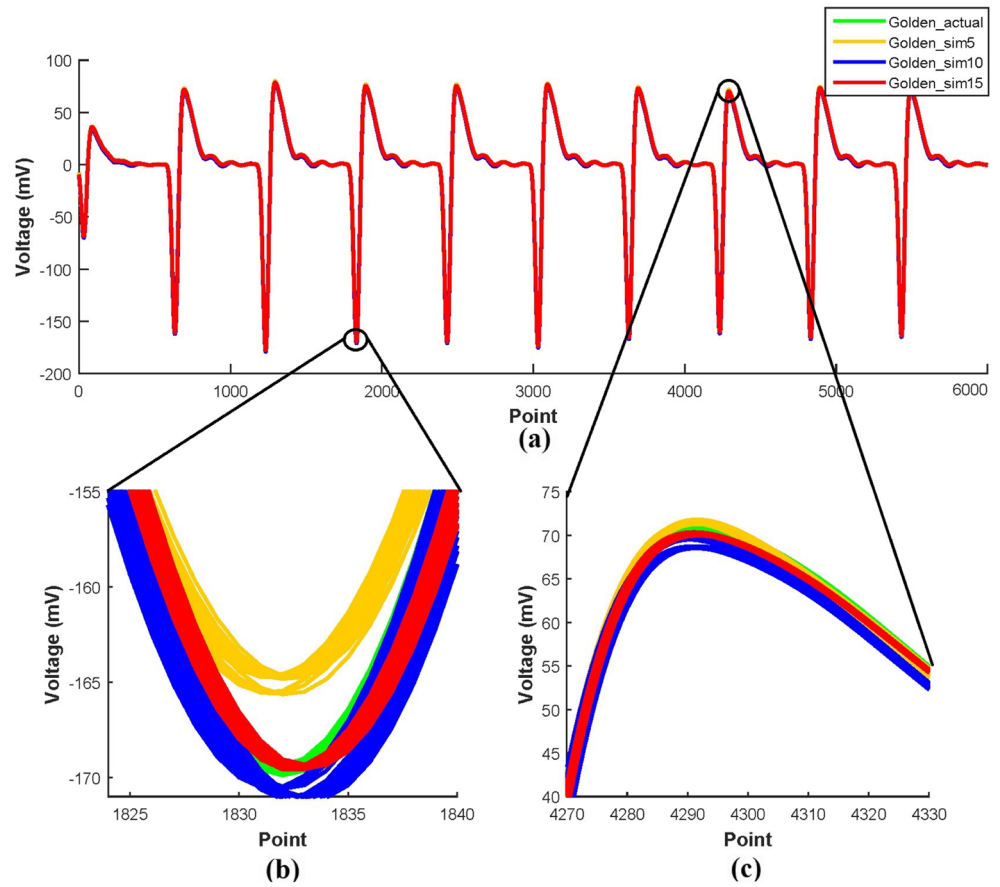


Fig. 8 Model calibration results at different process corners



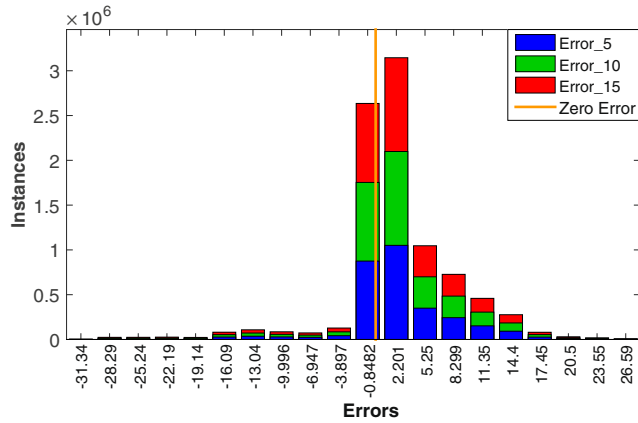


Fig. 9 Distribution histogram of the differences between the $I_M(t)$ and $I_{DG}(t)$

after the model calibration, and the distribution histogram of correlation coefficients between the Golden_actual and Golden_sim5 is shown in Fig. 10. It is clear from this figure, the correlation coefficients are greater than 98.85%. Similarly, the correlation coefficients under $\pm 10\%$ PV and $\pm 15\%$ PV are greater than 98.81%. On the whole, the calibrated golden model matches well with the actual silicon measurements. Therefore, the simulated data of calibrated golden model can substitute for the measured data of fabricated chips even in the presences of process variations and random noise.

5.4 Trojan Detection

After the model calibration, the calibrated traces $I_M(t)$ and denoised traces ($I_{DG}(t)$ and $I_{DC}(t)$) are clustered to two classes using the PAM. The centroids of two classes are determined when the cost function \mathcal{F} reaches the minimum value, and the Euclidean distance between each sample and

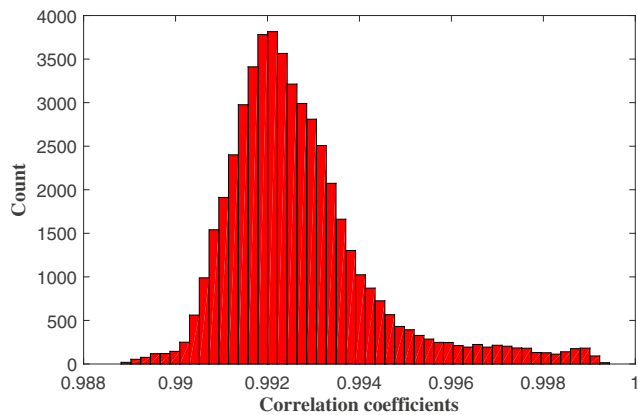


Fig. 10 Distribution histogram of correlation coefficient results

the centroid (denoted as Centroid_G) of golden class is calculated to evaluate the Trojan detection results.

As mentioned in Section 4.2, the random noise can not be eliminated in the real experiment completely, and the process variations are extremely difficult to emulate accurately, and the Euclidean distances vary accordingly. Fig. 11 shows the fluctuation of Euclidean distances under $\pm 5\%$ PV. The Euclidean distances between the Centroid_G and $I_M(t)$ are denoted as S_Dist, while the Euclidean distances between the Centroid_G and the $I_{DC}(t)$ of AES_G, AES_Trojan1, AES_Trojan2 and AES_Trojan3 respectively are denoted as G_Dist, T1_Dist, T2_Dist and T3_Dist. Regarding Fig. 11, S_Dist and G_Dist are more similar than the others, and T1_Dist, T2_Dist and T3_Dist respectively are all greater than the S_Dist or G_Dist. More specifically, the maximum value of G_Dist is 177.25, and the minimum value of T1_Dist is 378.09, and thus the margin is 200.84. Thus, any Euclidean distance that exceeds the 177.25 can be considered as the hardware Trojan, otherwise, there does not exist the hardware Trojan. This margin reduces with the increasing process variations, and the margin is 135 under $\pm 15\%$ PV, however, it is large enough to identify the hardware Trojan in the real experiment. Therefore, the PAM is capable of detecting the hardware Trojan correctly at different process corners.

Moreover, the averaged Euclidean distance of 100 traces is used to represent the robustness of proposed approach in the real experiment. Figure 12 illustrates the averaged Euclidean distance at different process corners. From the Fig. 12, G_Dist is greater than the S_Dist, and T1_Dist, T2_Dist, T3_Dist respectively are much greater than the S_Dist at different process corners. The differences between

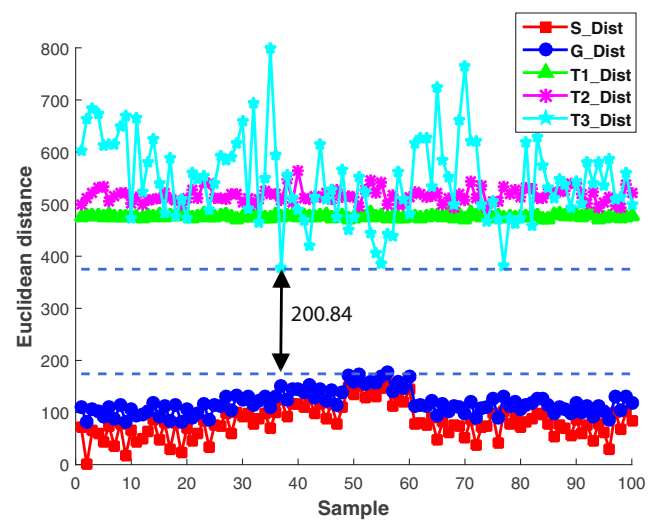


Fig. 11 Fluctuation of Euclidean distances under $\pm 5\%$ PV

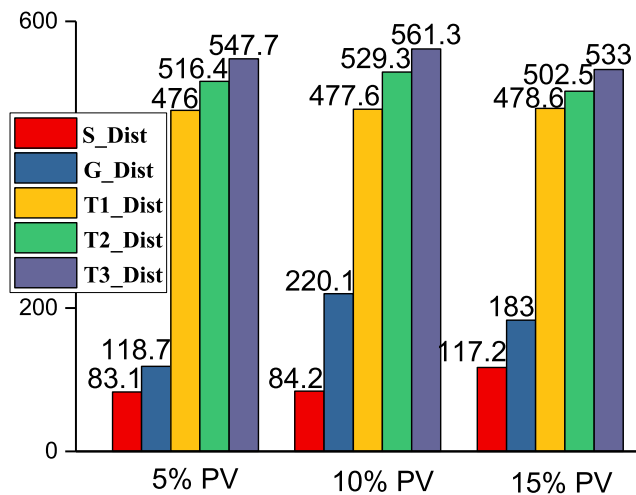


Fig. 12 The averaged Euclidean distance at different process corners

S_Dist and G_Dist mainly caused by process variations, however, the differences Trojan introduced can be identified even under the process variations. Therefore, the proposed approach is robust to the Trojan detection in the presence of process variations and random noise.

6 Conclusion and Future Work

In this paper, a hardware Trojan detection approach based on the golden layout model is presented, which eliminates the requirement of fabricated golden chips during detection. The simulated voltage variations at different process corners match well with the measured traces that can be used as the golden reference, and the hardware Trojan is identified using the PAM correctly in the presence of process variations.

While the proposed approach is success, but it is not limited to detect the Trojans in the cryptographic chips and the scalability of large designs (e.g. microprocessor and system on chip) may be explored in the future. Simultaneously, efficient test generation techniques are considered to activate the hardware Trojan and magnify the influences of Trojan effectively, and the limitation of the proposed approach against ultra-small Trojans is also further analyzed. Furthermore, there exist numerous types and insertions of hardware Trojan, and the side-channel influences of Trojans may be varied, and the proposed method against various Trojans are further studied.

Acknowledgments This work is funded partially by the National Natural Science Foundation of China (61832018) and China Postdoctoral Science Foundation (2019TQ0167).

References

1. Agrawal D, Baktir S, Karakoyunlu D, Rohatgi P, Sunar B (2007) Trojan detection using ic fingerprinting. In: Proceedings of the 2007 IEEE Symposium on Security and Privacy (SP'07), IEEE, pp 296–310
2. Alexandridis A, Chondrodima E, Giannopoulos N, Sarimveis H (2016) A fast and efficient method for training categorical radial basis function networks. *IEEE Transactions on Neural Networks and Learning Systems* 28(11):2831–2836
3. Bai Y, Li H, Zhang Y (2016) A hybrid model for congestion prediction in hf spectrum based on complete ensemble empirical mode decomposition. In: Proceedings of the 2016 CIE International Conference on Radar (RADAR), IEEE, pp 1–5
4. Balasch J, Gierlichs B, Verbauwhede I (2015) Electromagnetic circuit fingerprints for hardware trojan detection. In: Proceedings of the 2015 IEEE International Symposium on Electromagnetic Compatibility (EMC), IEEE, pp 246–251
5. Bao C, Forte D, Srivastava A (2015) Temperature tracking: toward robust run-time detection of hardware trojans. *IEEE Trans Comput Aided Des Integr Circuits Syst* 34(10):1577–1585
6. Bhasin S, Danger JL, Graba T, Mathieu Y, Fujimoto D, Nagata M (2014) Physical security evaluation at an early design-phase: A side-channel aware simulation methodology. In: Proceedings of the International Workshop on Engineering Simulations for Cyber-Physical Systems (ES4CPS'14), ACM, p 13
7. Brier E, Clavier C, Olivier F (2004) Correlation power analysis with a leakage model. In: Proceedings of the International Workshop on Cryptographic Hardware and Embedded Systems (CHES), Springer, pp 16–29
8. Calibre. <https://www.cadence.com/>. Accessed December 27, 2018
9. Chakraborty A, Mazumdar B, Mukhopadhyay D (2017) A combined power and fault analysis attack on protected grain family of stream ciphers. *IEEE Trans Comput Aided Des Integr Circuits Syst* 36(12):1968–1977
10. Du D, Narasimhan S, Chakraborty RS, Bhunia S (2010) Self-referencing: a scalable side-channel approach for hardware trojan detection. In: Proceedings of the International Workshop on Cryptographic Hardware and Embedded Systems (CHES), Springer, pp 173–187
11. Du D, Narasimhan S, Chakraborty RS, Bhunia S (2010) Self-referencing: a scalable side-channel approach for hardware trojan detection. In: Proceedings of the International Workshop on Cryptographic Hardware and Embedded Systems (CHES), Springer, pp 173–187
12. Elnaggar R, Chakraborty K (2018) Machine learning for hardware security: opportunities and risks. *J Electron Test* 34(2):183–201
13. Exurville I, Zussa L, Rigaud JB, Robisson B (2015) Resilient hardware trojans detection based on path delay measurements. In: Proceedings of the 2015 IEEE International Symposium on Hardware Oriented Security and Trust (HOST), IEEE, pp 151–156
14. Fujimoto D, Nagata M, Katashita T, Sasaki A, Hori Y, Satoh A (2011) A fast power current analysis methodology using capacitor charging model for side channel attack evaluation. In: Proceedings of the IEEE International Symposium on Hardware-Oriented Security and Trust (HOST), IEEE, pp 87–92
15. Global Foundries. <https://www.globalfoundries.com/>. Accessed December 27, 2018
16. Han HG, Lu W, Hou Y, Qiao JF (2018) An adaptive-pso-based self-organizing rbf neural network. *IEEE transactions on Neural Networks and Learning Systems* 29(1):104–117

17. Hasegawa K, Yanagisawa M, Togawa N (2017) Trojan-net feature extraction and its application to hardware-trojan detection for gate-level netlists using random forest. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences* 100(12):2857–2868
18. Hayt WH, Kemmerly JE, Durbin SM (2002) *Engineering Circuit Analysis*. McGraw-Hill, New York
19. He C, Hou B, Wang L, En Y, Xie S (2015) A failure physics model for hardware trojan detection based on frequency spectrum analysis. In: *Proceedings of the 2015 IEEE International Reliability Physics Symposium (IRPS)*, pages PR.1.1–PR.1.4
20. He J, Zhao Y, Guo X, Jin Y (2017) Hardware trojan detection through chip-free electromagnetic side-channel statistical analysis. *IEEE Trans Very Large Scale Integr VLSI Syst* 25(10):2939–2948
21. Hoque T, Narasimhan S, Wang X, Mal-Sarkar S, Bhunia S (2017) Golden-free hardware trojan detection with high sensitivity under process noise. *J Electron Test* 33(1):107–124
22. Hossain FS, Yoneda T, Inoue M, Orailoglu A (2017) Detecting hardware trojans without a golden ic through clock-tree defined circuit partitions. In: *Proceedings of the 2017 22nd IEEE European Test Symposium (ETS)*, pp 1–6
23. HSPICE. <https://www.synopsys.com/>. Accessed December 27, 2018
24. Jin Y, Makris Y (2008) Hardware trojan detection using path delay fingerprint. In: *Proceedings of the 2008 IEEE International workshop on hardware-oriented security and trust (HOST)*, IEEE, pp 51–57
25. Kim M, Chung E, Yoon S (2009) High-speed post-layout logic simulation using quasi-static clock event evaluation. *IEEE Trans Comput Aided Des Integr Circuits Syst* 28(8):1274–1278
26. Lamech C, Plusquellic J (2012) Trojan detection based on delay variations measured using a high-precision, low-overhead embedded test structure. In: *Proceedings of the 2012 IEEE International Symposium on Hardware-Oriented Security and Trust (HOST)*, pp 75–82
27. Lecomte M, Fournier J, Maurine P (2017) An on-chip technique to detect hardware trojans and assist counterfeit identification. *IEEE Trans Very Large Scale Integr VLSI Syst* 25(12):3317–3330
28. Leung CS, Wan WY, Feng R (2017) A regularizer approach for rbf networks under the concurrent weight failure situation. *IEEE Transactions on Neural Networks and Learning Systems* 28(6):1360–1372
29. Lin L, Burleson W, Paar C (2009) Moles: malicious off-chip leakage enabled by side-channels. In: *Proceedings of the 2009 international conference on computer-aided design (ICCAD)*, ACM, pp 117–122
30. Liu Y, Huang Y, Makris Y (2014) Hardware trojan detection through golden chip-free statistical side-channel fingerprinting. In: *Proceedings of the 2014 51st ACM/EDAC/IEEE Design Automation Conference (DAC)*, pp 1–6
31. Liu L, Pokharel R (2014) Post-layout simulation time reduction for phase-locked loop frequency synthesizer using system identification techniques. *IEEE Trans Comput Aided Des Integr Circuits Syst* 33(11):1751–1755
32. Menichelli F, Menicocci R, Olivieri M, Trifiletti A (2008) High-level side-channel attack modeling and simulation for security-critical systems on chips. *IEEE Trans Dependable Secure Comput* 5(3):164–176
33. Narasimhan S, Du D, Chakraborty RS, Paul S, Wolff F, Papachristou C, Roy K, Bhunia S (2010) Multiple-parameter side-channel analysis: a non-invasive hardware trojan detection approach. In: *Proceedings of the 2010 IEEE International Symposium on Hardware-Oriented Security and Trust (HOST)*, pp 13–18
34. Narasimhan S, Du D, Chakraborty RS, Paul S, Wolff FG, Papachristou CA, Roy K, Bhunia S (2013) Hardware trojan detection by multiple-parameter side-channel analysis. *IEEE Trans Comput* 62(11):2183–2195
35. Narasimhan S, Wang X, Du D, Chakraborty RS, Bhunia S (2011) Tesr: a robust temporal self-referencing approach for hardware trojan detection. In: *Proceedings of the 2011 IEEE International Symposium on Hardware-Oriented Security and Trust (HOST)*, pp 71–74
36. Ngo XT, Exurville I, Bhasin S, Danger JL, Guilley S, Najm Z, Rigaud JB, Robisson B (2015) Hardware trojan detection by delay and electromagnetic measurements. In: *Proceedings of the 2015 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, EDA Consortium, pp 782–787
37. Nowroz AN, Hu K, Koushanfar F, Reda S (2014) Novel techniques for high-sensitivity hardware trojan detection using thermal and power maps. *IEEE Trans Comput Aided Des Integr Circuits Syst* 33(12):1792–1805
38. Park HS, Jun CH (2009) A simple and fast algorithm for k-medoids clustering. *Expert Systems with Applications* 36(2):3336–3341
39. Peeters E, Standaert FX, Quisquater JJ (2007) Power and electromagnetic analysis: improved model, consequences and comparisons. *Integr VLSI J* 40(1):52–60
40. Rad R, Plusquellic J, Tehranipoor M (2008) Sensitivity analysis to hardware trojans using power supply transient signals. In: *Proceedings of the 2008 IEEE International Workshop on Hardware-Oriented Security and Trust (HOST)*, pp 3–7
41. Raitoharju J, Kiranyaz S, Gabbouj M (2015) Training radial basis function neural networks for classification via class-specific clustering. *IEEE Transactions on Neural Networks and Learning Systems* 27(12):2458–2471
42. Seo J, Ma H, Saha T (2018) On savitzkygolay filtering for online condition monitoring of transformer on-load tap changer. *IEEE Trans Power Delivery* 33(4):1689–1698
43. Söll O, Korak T, Muehlberghuber M (2014) Em-based detection of hardware trojans on fpgas. In: *Proceedings of the 2014 IEEE International Symposium on Hardware-Oriented Security and Trust (HOST)*, pp 84–87
44. Tiri K, Verbauwhe I (2005) A vlsi design flow for secure side-channel attack resistant ics. In: *Proceedings of the Design, Automation and Test in Europe (DATE)*, Vol 3. pp 58–63
45. Wu TF, Ganesan K, Hu YA, Wong HP, Wong S, Mitra S (2016) Tpad: hardware trojan prevention and detection for trusted integrated circuits. *IEEE Trans Comput Aided Des Integr Circuits Syst* 35(4):521–534
46. Wei S, Meguerdichian S, Potkonjak M (2011) Malicious circuitry detection using thermal conditioning. *IEEE Trans Inf Forensics Secur* 6(3):1136–1145
47. Xue H, Ren S (2018) Self-reference-based hardware trojan detection. *IEEE Trans Semicond Manuf* 31(1):2–11
48. Xue M, Wang J, Hux A (2016) An enhanced classification-based golden chips-free hardware trojan detection technique. In: *Proceedings of the 2016 IEEE Asian Hardware-Oriented Security and Trust (AsianHOST)*, pp 1–6
49. Yoshimizu N (2014) Hardware trojan detection by symmetry breaking in path delays. In: *Proceedings of the 2014 IEEE International Symposium on Hardware-Oriented Security and Trust (HOST)*, pp 107–111

50. Yu D, Liu G, Guo M, Liu X (2018) An improved k-medoids algorithm based on step increasing and optimizing medoids. *Expert Syst Appl* 92:464–473
51. Zarrinchian G, Zamani MS (2017) Latch-based structure: a high resolution and self-reference technique for hardware trojan detection. *IEEE Trans Comput* 66(1):100–113
52. Zhao X, Han L, Feng Z (2015) A performance-guided graph sparsification approach to scalable and robust spice-accurate integrated circuit simulations. *IEEE Trans Comput Aided Des Integr Circuits Syst* 34(10):1639–1651
53. Zheng Y, Yang S, Bhunia S (2016) Semia: self-similarity-based ic integrity analysis. *IEEE Trans Comput Aided Des Integr Circuits Syst* 35(1):37–48

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Yanjiang Liu received the B.S. degree in Electronic and Information Engineering from the Zhoukou Normal University, Zhoukou, China, in 2013, and the M.S. degree in Circuits and Systems from Guangdong University of Technology, Guangzhou, China, in 2016. He is currently pursuing the Ph.D. degree at the School of Microelectronics at the Tianjin University. His current research interests include digital circuit design and hardware security.

Jiayi He received the B.S. degree in electronic science and technology and the M.S. degree in microelectronics from Tianjin University, Tianjin, China, in 2013 and 2015, respectively, where he is currently pursuing the Ph.D. degree at the School of Microelectronics. He used to be a Visiting Scholar with the University of Central Florida, Orlando, FL, USA, under the guidance of Y. Jin, from 2016 to 2017. His current research interests include digital circuit design and hardware security.

Haocheng Ma received the B.S. degree in electronic science and technology from Tianjin University, Tianjin, China, in 2017, where he is currently pursuing the M.S. degree at the School of Microelectronics. His current research interests include digital circuit design and hardware security.

Yiqiang Zhao received the B.S. degree in semiconductor physics and device, the M.S. degree in microelectronics, and the Ph.D. degree in microelectronics and solid-state electronics from Tianjin University, Tianjin, China, in 1988, 1991, and 2006, respectively. In 1991, he joined the Jinhang Technical Physics Institute, Tianjin, where he was responsible for analog and mixed signal circuit design. Since 2001, he has been with the School of Electronic Information Engineering and the School of Microelectronics, Tianjin University, where he is currently a Professor. His research interests include mixed-signal integrated circuits, security chips, and hardware security.