SN Computer Science (2020) 1:233
https://doi.org/10.1007/s42979-020-00220-0

ORIGINAL RESEARCH

Comprehensive Analysis on Hardware Trojans in 3D ICs: Characterization and Experimental Impact Assessment

Zhiming Zhang[1] · Jaya Dofe[2] · Pruthvy Yellu[1] · Qiaoyan Yu[1]

Received: 4 April 2020 / Accepted: 8 June 2020

Abstract
Three-dimensional (3D) integration facilitates to integrate an increasing number of transistors into a single package. Despite improved performance and power efficiency, the integration of multiple dies in the same package potentially leads to new security threats, such as 3D hardware Trojans. This work conducts a thorough survey on hardware Trojans reported in 3D integrated circuits (ICs) and systems, and proposes a comprehensive characterization of 3D hardware Trojans. Several case studies are performed to validate the feasibility of 3D hardware Trojan implementation. Our experimental results indicate that 3D ICs indeed provide a better environment for inserting stealthy thermal-based Trojans than 2D ICs. Multiple FPGA boards are utilized to conceptually emulate the stacked 3D ICs that experience multi-tier hardware Trojans. The stealthiness and effectiveness of the proposed multi-tier Trojans are validated in our case studies. The emulation results further show that the existing current-based self-referencing Trojan detection method designed for 2D Trojans will result in a lower detection rate in 3D scenarios.

Keywords Three-dimensional integration · Hardware Trojan · Trojan model · Side-channel analysis attack · Interconnect · Power distribution network (PDN) · Network-on-chip (NoC)

Introduction

Three-dimensional (3D) integration is an emerging technology to ensure further growth in transistor density and performance of future integrated circuits (ICs) [1, 2]. It has been demonstrated that 3D techniques can be leveraged to reduce package size and power consumption while significantly improving bandwidth [3–5]. Unfortunately, 3D techniques also bring in unique and unexplored security threats to 3D ICs [6]. Due to higher integration density and wider process/voltage/temperature (PVT) variation [7, 8], it may be more challenging to address the security threats in 3D ICs than in 2D planar chips [9].

Since 2007, hardware Trojans inserted in 2D ICs have been well studied in the literature [10–14]. To facilitate Trojan detection, researchers categorize hardware Trojans based on their distribution, structure, size, and logic type. Depending on the activation mechanism, a hardware Trojan can be classified as internally or externally triggered. Based on how often hardware Trojans are triggered, the work [15] presents three types of Trojans: always-on, combinational condition triggered, and sequential condition triggered. Once the Trojan trigger condition arrives, the Trojan payload will execute the defined malicious operations, such as transmitting confidential information, modifying function, degrading performance, and consuming extra power.

Thanks to the mature models for 2D Trojans, various functional testing and side-channel analysis approaches have been proposed to detect different kinds of hardware Trojans in 2D ICs [11, 16–18]. However, Trojan detection methods for 3D Trojans have not been widely explored yet. One important reason for that is the lack of a well-established 3D Trojan model. Due to the vertical integration of multiple tiers, 3D Trojans appear with different characteristics than 2D Trojans [19]. Thus, the commonly used Trojan detection

This article is part of the topical collection "Hardware-Assisted Security Solutions for Electronic Systems" guest edited by Himanshu Thapliyal, Saraju P. Mohanty, Wujie Wen and Yiran Chen.

✉ Qiaoyan Yu
qiaoyan.yu@unh.edu

[1] University of New Hampshire, Durham, NH, USA

[2] California State University, Fullerton, CA, USA

Published online: 16 July 2020

methods for 2D Trojans may not be effective to protect chips from 3D Trojans.

The preliminary version of this work introduces four 3D hardware Trojan models. In this work, we highlight the difference between 2D and 3D Trojans using architectural comparison and quantitative assessment with practical implementations. More specifically, the main contributions of this work are summarized as follows.

1. Together with the preliminary version [19], our work does the first thorough survey on hardware Trojans in 3D ICs. Security threats and hardware Trojan models reported in the existing literature are compared in this work.
2. Four representable high-level 3D hardware Trojan cases are characterized. Practical examples for each Trojan model are provided for quantitative analysis. The difference between 2D and 3D Trojans are highlighted in our study.
3. As the thermal issue is prominent in 3D ICs, we designed a thermal-induced 3D hardware Trojan and examined its triggering speed and resilience against Trojan detection in a 3D environment for a pass-code authentication.
4. Multiple FPGA boards were utilized to emulate the multi-tier collaborative hardware Trojans, through which attackers can manipulate the function of the target tier without direct tampering on the victim circuit.
5. We examined the success rate of an existing 2D hardware Trojan detection method in the context of 3D ICs. Our simulation results show that the 2D approach operated in 3D chips is not as effective as it works in the 2D scenario.
6. Comparing to our preliminary work [19], this work provides new simulation and FPGA emulation examples for case 2 and case 3 Trojan models and also examines the Trojan detection rate of an existing 2D-level Trojan detection in the 3D scenario.

The rest of this paper is organized as follows: Sect. "Our Survey on Existing Hardware Trojans in 3D Integrated Circuits and Systemss" summarizes the security threats and hardware Trojan models for 3D ICs discussed in the existing literature. Section "Proposed Comprehensive Characterization of3D Hardware Trojans" proposes comprehensive characterization models for 3D Trojans and their practical implementations. Simulation and emulation results for the 3D Trojans are presented in Sect. "Proposed Comprehensive Characterization of3D Hardware Trojans", too. The effectiveness of a 2D hardware Trojan detection method applied in the scenario of 3D IC is examined in Sect. "Examination of A 2D Trojan DetectionApproach in 3D IC". This paper is concluded in Sect. "Conclusion".

Our Survey on Existing Hardware Trojans in 3D Integrated Circuits and Systems

The increased number of dies in 3D ICs and vertical-dimension integration potentially leaves more attack surfaces open for adversaries to implement hardware Trojans. As multiple dies are vertically integrated into 3D systems, additional manufacturing steps are needed in 3D IC fabrication flow than in their 2D counterparts. Multiple foundries for dies and vertical interconnects will be involved in the 3D integration. In the current semiconductor business model, more and more chip designs are outsourced for fabrication. As a result, neither all single die fabrication foundries nor vertical interconnect manufacturers are trusted [6, 20–23, 26]. The die-to-die bonding may be performed in an untrusted foundry, too. In Fig. 1, we label the possible attack surfaces for 3D Trojan insertion. Trojans can be placed by the single-die manufacturing foundries, independently or cooperatively. Since the bonding foundries have access to all the single dies, they have a more likely-hood to implement a Trojan involving multiple dies.

Based on the existing literature, we categorize the 3D Trojans in Table 1, where we highlight the threat model with special emphasis on threat source and attack target. In addition to Trojan trigger and payload mechanisms, we also identify Trojan locations in 3D ICs. From Table 1, we can see the nature of the 3D IC structure creates new opportunities for hardware Trojan design, for instance, thermal-based Trojans and cross-tier Trojans. In the next three subsections, we discuss the existing literature listed in Table 1 according to their special trigger mechanisms and Trojan locations.

Thermal-Triggered 3D Trojans

The fact of poor heat dissipation in a stacked 3D IC can be exploited to develop Trojan triggers. Although the techniques such as heat sink, liquid cooling, thermal-driven floorplanning and routing, and thermal TSV insertion [27] could address the thermal issue in 3D ICs at certain degree, the heat dissipation along a path could harm the tiers and degrade the chip performance [28]. The heat generated and accumulated in the chip will change the electrical parameters of transistors and the switching speed of logic gates. Thus, the system may have new (and unspecified) transition states. The unexpected transition glitches can be employed to design Trojan triggers. As indicated in [20, 21], thermal-triggered Trojans can be inserted by any malicious foundries with access to the layout of designs. Those Trojans likely congregate near the middle tier, where heat dissipation is harder than in other tiers [21]. The work [6] demonstrates that a thermal triggered Trojan may be hidden in 3D interposers. Thermal Trojans can speed up circuit component

Table 1 Existing work on hardware Trojan in 3D ICs

Work	Threat model		Trojan model		
	Threat source	Attackers' access	Trigger	Payload	Location
[20]	Untrusted die foundries	GDSII files	Thermal effect caused transition glitches	No special requirement	Any tiers in 3D ICs
[21]	Untrusted die foundries	GDSII files	Thermal effect caused transition glitches	No special requirement	Middle tier in 3D ICs
[6]	Untrusted interconnect foundries Untrusted single die manufacturers	GDSII files	Thermal effect, Aging effect	Voids leading to DoS Partially filled TSVs	Interposer TSV
[22]	Untrusted interconnect foundries Untrusted single die manufacturers Untrusted unified foundries	GDSII files	Remote circuits, Distributed circuits	Impacts on target's power Impacts on target's delay	TSV Multiple tiers
[23]	Untrusted single die manufacturers	Lease critical die	Low-activity nets	Leak key from encryption unit	Trojan in different tiers with encryption unit
[24]	Untrusted assemblers	No legitimate dies	No special requirement	Interrupt normal function, Leak information	Extra Trojan die in 3D ICs stack
[25]	Final bonding foundries	Entire layers	Internal nets	No special requirement	Any tiers in 3D ICs
[26]	Untrusted single die manufacturers	GDSII files	No special requirement	No special requirement	Any tiers in 3D ICs

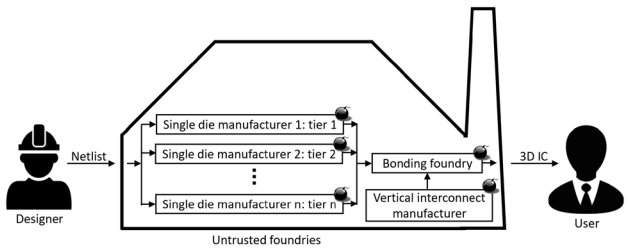

Fig. 1 3D hardware Trojan insertion in untrusted foundries

aging and consequently lead to a Denial-of-Service (DoS) attack [6].

Cross-Tier 3D Trojans

The multiple-die structure of 3D ICs allows attackers to spread the circuit for a Trojan to multiple tiers. This type of Trojans could be inserted by untrusted die manufacturers, interconnect foundries, and unified foundries. The cross-tier concept means that either the trigger and payload circuits of cross-tier Trojans are separated into different tiers, or the trigger circuit split in multiple tiers is activated jointly to enable the payload [22]. The cross-tier Trojans may not be detected by functional testing performed on each individual die since the Trojan trigger condition is extremely rare. The work [23] demonstrates a Trojan located in a different tier than the encryption unit facilitates to leak the secret key. Even if the untrusted foundry only has partial knowledge of the 3D chip, they can launch cross-tier Trojan attacks.

Trojans Exploiting Other 3D Features

The work [24] envisions a new hardware Trojan in stacked 3D ICs: a malicious die is placed between other tiers in the 3D stack. That malicious die, carrying Trojan circuits, may interrupt normal operations in other 3D tiers or store secret information passing through the Trojan tier. Due to the prominent process variation in 3D chips, it is not easy to differentiate the extra delay induced by the 3D hardware Trojan. This type of Trojan can be inserted by untrusted die assemblers. For instance, the work [25] describes that attackers from the bonding foundry could leverage outsourced dies to implement 3D Trojans. In [26], the adversary is an untrusted die manufacturing foundry with access to GDSII files.

Proposed Comprehensive Characterization of 3D Hardware Trojans

The existing literature mentioned in Table 1 showcases diverse 3D Trojans, but they neither have a thorough discussion on the exact Trojan models nor provide quantitative impact assessment. This work fills the gap by characterizing four representable 3D hardware Trojan cases and quantitatively analyzing their practical examples in the following sections.

The major difference between 2D and 3D hardware Trojans is whether or not the Trojan trigger and payload circuits are located in the same tier where the target circuit resides. In 2D chips, the Trojan circuit co-exists with the victim in the same tier. One could perform testing or side-channel analysis to detect the presence of 2D Trojans. In contrast,

conventional testing on 3D chips is typically done in a separate fashion. The die for each tier is tested individually before 3D integration. Once the good dies are stacked vertically, limited testing will be performed to detect the defects between die-to-die connections, rather than extensively examining the correctness of the 3D system's behavior [29].

Based on our survey in Sect. "Our Survey on Existing Hardware Trojans in 3D Integrated Circuits and Systemss", we characterize the 3D hardware Trojan with four cases shown in Fig. 2. To the best of our knowledge, our prior publication [19] and this work are the first efforts that introduce comprehensive characterization for 3D hardware Trojans. The following subsections present four 3D Trojan cases in detail.

Case 1: Cross-tier Trojan Trigger

Characteristics

In case 1, the trigger circuit of the 3D Trojan is placed in tier 1 while the payload circuit is located near the Trojan target. This type of 3D Trojan is similar to the 2D Trojans that are triggered by an external signal [30], but it is more difficult to mitigate compared to the 2D Trojan. In 2D chips, the passive attack from the external trigger signals can be alleviated by adding shielding material or using unit isolation. In contrast, in 3D ICs, the external attack may be originated from the adjacent tiers, which are not removable after the 3D chip fabrication is completed. As heterogeneous 3D integration emerges, varieties of external trigger mechanisms could be implemented in the other 3D tiers, thus challenging the prevention of 3D Trojans. Moreover, since the payload circuit may never or rarely be enabled without the valid cross-tier trigger signal, the symptom of Trojan attacks will not be observed in typical functional testing. Thus, this type of Trojan is stealthy.

We illustrate the case 1 Trojan with an example shown in Fig. 3. The trigger circuit is a heat generator in the top tier. The payload circuit is a temperature-sensitive resistor, which is built in the authentication unit in the middle tier. When the heat from the top tier propagates to the middle tier, the temperature-sensitive resistor could alter the delay of the critical path or cause timing violations, thus resulting in a malfunction of the authentication unit. As reported in [21], the heat from the middle tier of a 3D vertical stacking structure is accumulated easily due to the relatively long dissipation path to the heat sink. Hence, the thermal triggered Trojans will be more likely deployed in 3D integrated circuits and systems than its 2D counterpart.

We performed a transistor-level simulation in Cadence Virtuoso to demonstrate the impact of middle-tier heat dissipation on neighboring tiers. We collected the transient current of the nodes for load connection in the middle tier of our 3D power distribution network (PDN) model [9] to evaluate the thermal effect. Our target module for the thermal effect investigation is an 8-bit S-box module of AES. In the middle tier, we had 30 load nodes arranged as 5 rows by 6 columns and then captured the current of each node for 10 ns. The current collected in the 8th ns is shown in the contour graphs in Fig. 4. Generally, the 3D PDN carries greater currents than the 2D PDN. Although the highest current for both 2D and 3D cases appears in the bottom left area where the S-box is located, the current distribution near the S-box is different in the 3D PDN compared to the 2D PDN. We highlight the difference with red dashed rectangles in Fig. 4a, b. Those observations make sense because any single tier in the 3D chip is not isolated but impacted by its neighboring tiers. Since the thermal dissipation of a circuit is proportional to its current, it is reasonable to believe that the temperature surrounding our target is influenced by its neighboring tiers.

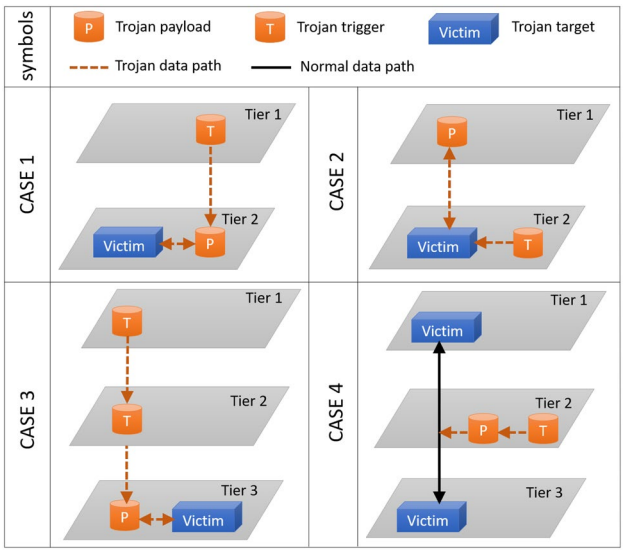

Fig. 2 Proposed characterization of 3D hardware Trojans

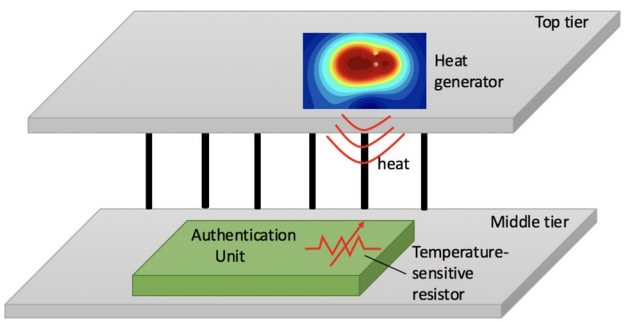

Fig. 3 Thermal-triggered cross-tier Trojan

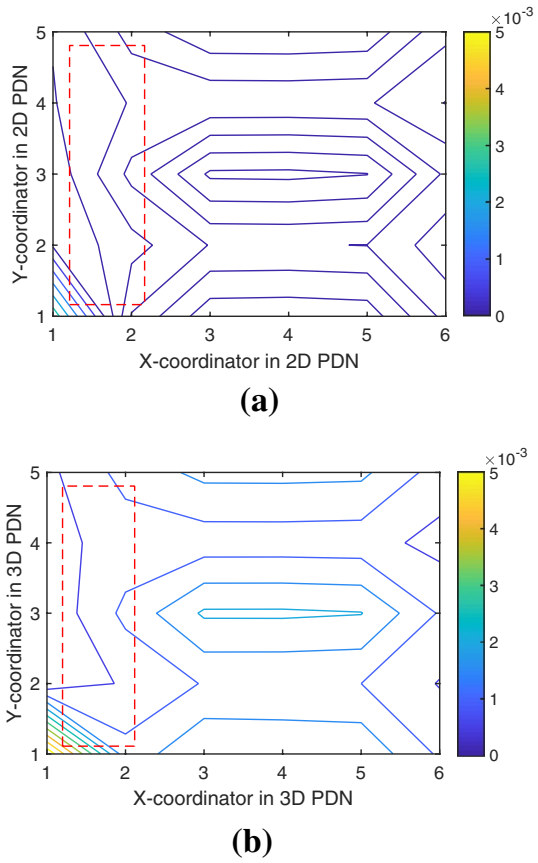

Fig. 4 Current contour maps of **a** 2D and **b** 3D PDNs

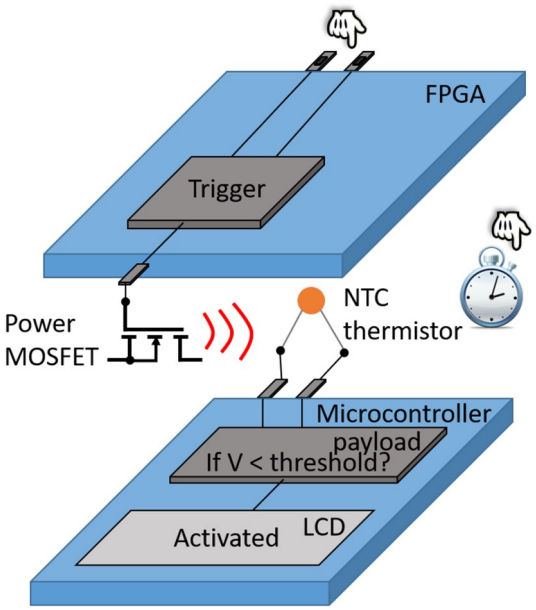

Fig. 5 Experimental setup for the emulation of thermal-triggered hardware Trojan in 3D ICs

Example Analysis

To perform quantitative analysis for the cross-tier 3D hardware Trojan, we conducted a case study on a platform composed of Xilinx Nexys3 Spartan-6 FPGA, TI MSP430FR6989 LaunchPad board, IRF540 MOSFET transistor, and an NTC thermistor. The purpose of this case study is to verify the implementation feasibility of the thermal Trojan (similar to the one shown in Fig. 3) and compare its activation efficiency between the scenarios of 2D and 3D ICs. The overview of our experimental setup is depicted in Fig. 5. The main component of the heat generator circuit is a MOSFET driven by the FPGA board. The MOSFET could burn when its gate voltage exceeds a voltage threshold and the MOSFET temperature can be as high as 175 °C. The sensor circuit composed of an NTC thermistor and multiple resistors in series is powered by the TI microcontroller. When the thermistor senses an increase in the temperature in the surrounding air, its resistance starts to drop. This leads to a reduction in the voltage across the thermistor. To emulate the 2D scenario for comparison, we added a heat sink for the heat generator circuit, to provide a better heat dissipation which is commonly available in 2D ICs.

An authentication system is programmed in the microcontroller to examine the password provided externally. The microcontroller also detects the voltage level of the thermistor. A Trojan trigger logic is programmed in the FPGA to monitor the two input signals controlled by the two switches on the FPGA board. The triggered Trojan turns on the MOSFET (thus it starts to burn) to heat the temperature in the surrounding area. Once the thermistor senses the increased temperature, the microcontroller detects the change on voltage and then drives the authentication system to jump to the password reset status, which is usually only available to legal users. We successfully mimicked a 3D thermal-triggered hardware Trojan and overwrote the authentication password in our hardware demo [31].

Next, we compared the activation speed of the thermal-triggered Trojans for 2D and 3D scenarios. We used the microcontroller to implement a threshold comparator to examine the voltage level of the thermistor. If the voltage of a thermistor exceeds the threshold, the Trojan payload will reset the authentication password. We warmed the air surrounding the thermistor with and without the heat sink to mimic 2D and 3D scenarios, respectively. A timer is used to measure the time that the thermistor takes to drop the voltage below the threshold for each case. The results shown in Table 2 indicate that the Trojan activation time in the 2D scenario is almost twice compared to the 3D case. This means it is easier to implement thermal-triggered Trojans in 3D ICs than in 2D chips. We also measured the speed of temperature changing, which is reflected in the resistance of the thermistor. The dropping trend of the resistance in Fig. 6

Table 2 Trojan activation efficiency

Emulation scenarios	Time to trigger the Trojan (min)
2D	11:12
3D	6:52

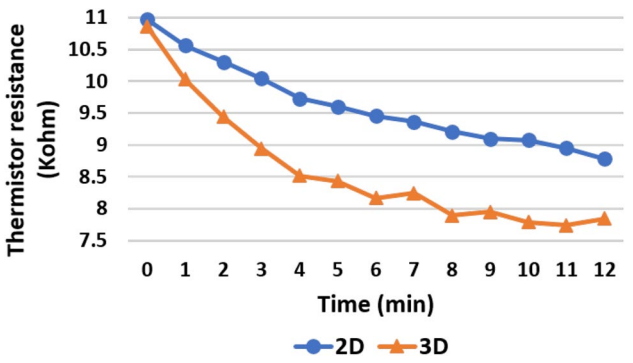

Fig. 6 Resistance dropping of the thermistor used in Fig. 5

implies that the NTC thermistor's resistance for the 3D case drops faster than the 2D. This fact further confirms that heat can be better accumulated in 3D than 2D. Thus, 3D ICs will provide a better environment to facilitate the implementation of thermal-based Trojans than 2D ICs.

Case 2: Cross-tier Trojan Payload

Characteristics

In the Trojan described in case 2, the payload is located in the top tier (tier 1), from where it is relatively easy to probe and measure side-channel signals than from the middle tier. The motivation of this type of 3D Trojan is to steal confidential information from the victim unit. Essentially, the stacked structure of 3D ICs provides a reliable medium for attackers to collect information from the middle and bottom tiers. In addition, as the payload resides in another tier, the effect of this kind of Trojans will not be observable while testing on the individual tiers. Here, we assume that the trigger circuit is small enough to hide its area, delay, and power overhead. This assumption is as reasonable as what we usually have in 2D ICs.

The cross-tier Trojan can facilitate the development of a covert channel to leak information. The victim unit could be an encryption engine, such as the one shown in Fig. 7. The crypto key is loaded from the volatile memory in the top tier. To prevent the leaked key from being visible during the middle tier testing, the pilfered key is first transformed into another format (i.e., obfuscated key), and then the Trojan passes the obfuscated key to the rarely used main memory

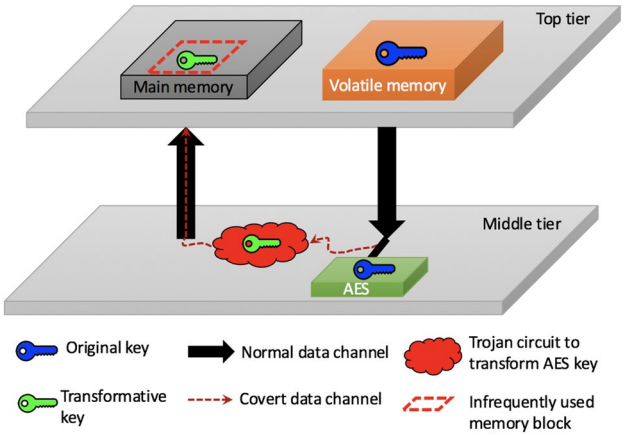

Fig. 7 An example of key leaking via the covert channel formed by a hardware Trojan in a stacked 3D IC

Table 3 Parameters for TSV and wire model

TSV Model (per TSV) [32]					
Diameter	Height	Pitch	Resistance	Inductance	Capacitance
10 μm	60 μm	20 μm	20 mΩ	34.94 pH	283 fF
RC Model for Local Wire Interconnect (per mm) [33]					
Resistance			Capacitance		
3.31 kΩ			170.59 fF		

in the top tier. When we test the top tier, the main memory functions normally. The separated testing on the middle tier will not reveal the presence of the 3D Trojan because the key is obfuscated. However, the key will be leaked by the covert channel built by the cross-tier 3D Trojan since the attacker knows how to de-obfuscate the key.

Example Analysis

In this subsection, we use a combination of transistor-level simulation and FPGA emulation to demonstrate the feasibility of leaking the AES secret key via cross-tier Trojans. We implemented the cross-tier hardware Trojan and the 3D system shown in Fig. 8 in Cadence Virtuoso with a 45 nm NCSU FreePDK technology [32]. The PDN in each tier of the stacked 3D structure is mainly composed of a global power grid and a virtual grid. TSVs connect the global power grids in nearby tiers. The parameters for the TSV and wire model are listed in Table 3. The parameters are verified by [32, 33]. Our transistor-level 3D circuit nearly matches the practical 3D IC. The crypto unit adopted here is a transistor-level AES S-box. To ensure the unipolarity of the channel between key and TSV, a buffer is located in the middle of the channel (not shown in the diagram) so that we can prevent the power data from being transmitted

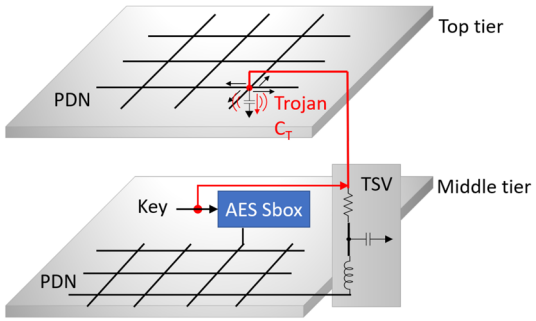

Fig. 8 Experimental setup of key leaking via a cross-tier Trojan

back to the S-box to hinder normal operation. The hardware Trojan shown in Fig. 8 stealthily passes the secret key to a nearby 3D tier. The main component of the Trojan is a capacitor connected with the PDN. Each key is assigned to one Trojan capacitor. The Trojan capacitors are charged or discharged based on the key bits transmitted through TSVs. The charges stored in the Trojan capacitor C_T will facilitate the side-channel analysis for the crypto key retrieval. The capacitor C_T acts like a decoupling capacitor, which can keep the supply power stable. In this way, the normal function of the nearby tier will not be affected so that the stealthiness of the inserted Trojan can be achieved.

In our experiment, we set the key bits to "11111111", and varied C_T from 10fF, 1132fF, to 11320fF. The power consumption of the S-box without Trojan or with different Trojan loads was measured and compared. As shown in Fig. 9a, a smaller Trojan capacitor leads to a smaller power change, but the power difference induced by the Trojan is still less than 2.5% even though we increase C_T to 11320fF. However, the power profiles for different Trojan capacitors are consistent. The slight but consistent variation on the power profile is an important quality to ensure the stealthiness of the cross-tier Trojan. We kept the capacitance of the Trojan as 11320fF but changed the key bits from "11111111", "00000000", "01010101", to "01001011". The power consumption for these four cases is shown in Fig. 9b. It can be observed that the power consumption for each key is unique. Thus, we can correlate the new power profile with the key used in the crypto unit.

Next, we used a SAKURA-G FPGA assessment kit to conduct a side-channel analysis on an AES affected by the cross-tier Trojan. The Trojan model AES-T1000 published on Trust-hub was modified to mimic the 3D Trojan described in Fig. 8. The main difference is, we used FPGA pins to mimic the Trojan capacitors. Each key bit additionally drives eight FPGA pins. Due to the capacitor induced by the Trojan, the total power consumption of the AES module is slightly changed. However, the power difference due to the Trojan accelerates the correlation power analysis (CPA)

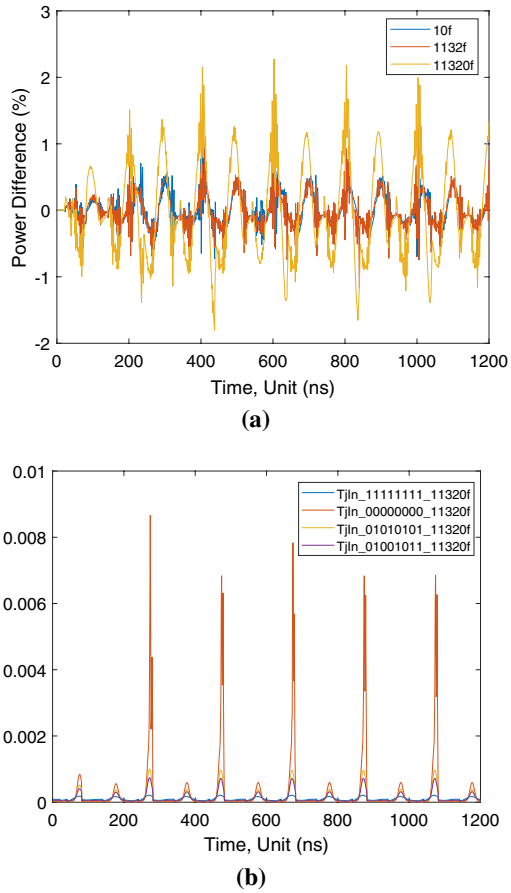

Fig. 9 Impact of cross-tier Trojans on the power consumption of an AES S-box. **a** Power differences caused by the Trojans implemented with different Trojan capacitors, and **b** unique power profiles induced by the same Trojan that snoops the AES S-box with different keys

attack. The key retrieval processes for cases of without Trojan and with Trojan are shown in Fig. 10. The red lines represent the 16 key bytes of AES. As the number of analyzed traces increases, the red lines are getting out of the green zone, which means the key bytes are being retrieved. As a result, the CPA attack on the AES with Trojan is able to retrieve all the key bytes within the use of 6000 power traces. Given the same amount of power traces, the CPA attack without Trojan retrieves only 14 key bytes out of 16 since two lines are still buried in the green zone. This indicates that the Trojan implemented in this example could ease the CPA attack.

Case 3: Multi-tier Collaborative Trojan

Characteristics

There may emerge another kind of 3D Trojan, multi-tier collaborative Trojan, which is more sophisticated than the cross-tier Trojan trigger and payload. The multi-tier Trojan

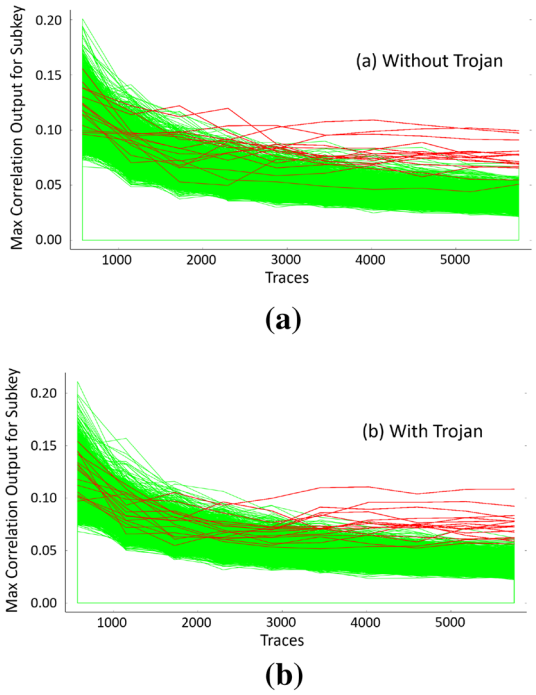

Fig. 10 Correlation power analysis for the AES **a** without Trojan and **b** with Trojan

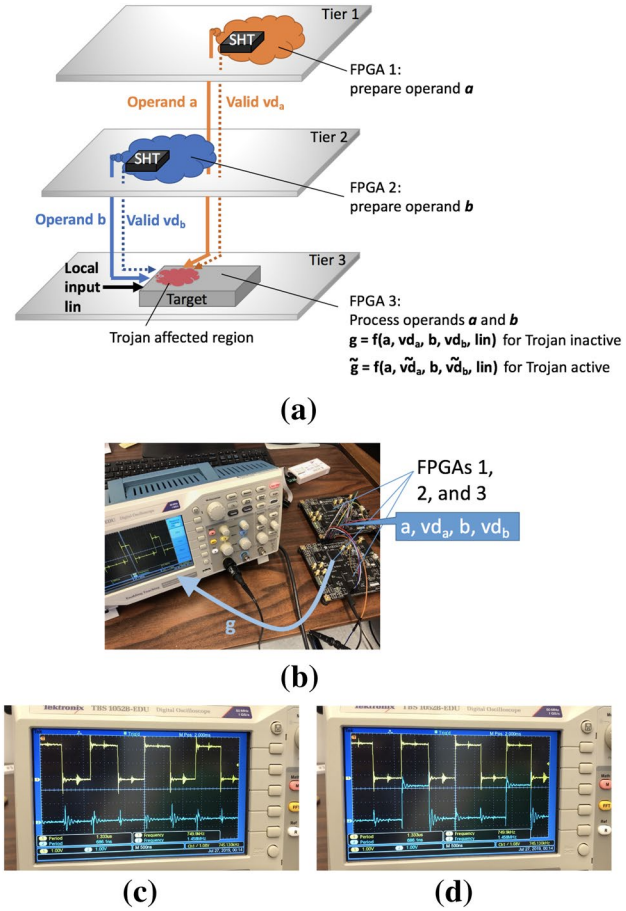

Fig. 11 Multi-tier collaborative hardware Trojan. **a** Conceptual diagram, **b** multi-FPGAs experimental setup, **c** normal output, and **d** Trojan affected output

in case 3 shown in Fig. 2 is activated by the two trigger circuits from tiers 1 and 2, respectively. Compared to hardware Trojans in 2D ICs, the multi-tier Trojan trigger has significantly lower Trojan triggering probability due to a larger pool of trigger signals. Moreover, the collaborative Trojan trigger could be a combination of different trigger mechanisms (e.g., temperature, voltage level, and electromagnetic flux). Multi-tier collaborative Trojans represent the scenario that attackers exploit the security weaknesses of other tiers in the 3D system to breach the target tier with strong security mechanisms, instead of compromising the target tier directly. In terms of cost and effectiveness, multi-tier Trojans are more likely to appear in 3D chips than a single-tier Trojan.

We implemented an example of a multi-tier collaborative Trojan in a 3D system with four tiers. Two FPGA boards, each including two FPGA chips, were utilized to emulate the 3D system. The schematic diagram and FPGA setup are shown in Fig. 11a, b, respectively. Tiers 1 and 2 are weak in the sense of resistance against hardware Trojan insertion. Thus, two hardware Trojan triggers were placed in those two tiers. The 3D Trojan manipulates the signals passing images from tiers 1 and 2 to tier 3. Due to their low trigger probability, sequential hardware Trojan (SHT) triggers were applied in this example. When the SHT trigger is active, the vertical data communication is compromised such that the valid indication signals vd_a and vd_b will allow improper operands a and b to propagate to tier 3. Consequently, the compromised inputs $\widetilde{vd_a}$ and $\widetilde{vd_b}$ lead the Trojan target circuit to behave differently ($\widetilde{g}$) than the normal specification (g). Once the valid signals are compromised by the 3D Trojans, the integrity of the images received by tier 3 will be sabotaged. As a result, image-based authentication will fail.

Example Analysis

In the FPGA platform, we connected those FPGA chips with external wires so that the tier-to-tier communication can be manipulated and observed via the oscilloscope. Figure 11c illustrates that the square-wave signal from tier 1 (the yellow line on the top) is not passed to tier 3 (as the blue signal on the bottom is flat). When the Trojan is triggered, a portion of the yellow line is copied to the blue signal as shown in Fig. 11d. This indicates that the multi-tier collaborative Trojan manipulates the signal filter, which is controlled by the valid signal, and transfers invalid or even malicious data to the target tier. Assume tier 3 in the 3D system examines

whether the images from the top two tiers are highly correlated and then enables the critical mission programmed in tier 3. If the valid signals vd_a and vd_b are tampered by the multi-tier collaborative Trojan, dummy image rows will be dumped to tier 3. Five images shown in Fig. 12 are adopted for correlation analysis in the 3D system mentioned above. Clearly, Fig. 12b–e are different than Fig. 12a, thus the image correlation cannot get close to 0.9. However, when the valid signals for enabling image transfer between tiers are compromised, the image correlation could approach to 0.9 if the hardware Trojan is able to manipulate vd_a and vd_b for a time period long enough to dump 100 dummy image rows.

Case 4: Multi-tier Synergic Trojan Payload

Characteristics

When an IC is expended from planar to vertical dimension, the corresponding Trojan payload will be distributed to multiple tiers as well. In case 4 shown in Fig. 2, the Trojan circuit snoops the data (or even the side-channel signal) available in tier 2. As a result, the confidential information is leaked from tier 2 to other tiers. Often time, both the Trojan trigger and payload are located in the different tiers than the target one. Alternatively, a thin Trojan tier can be integrated into the 3D stack structure to provide flexible and precise control on the snooped information without incurring noticeable delay overhead [24]. We further envision that a 3D Trojan payload could achieve a synergic attack effect in multiple tiers, rather than influencing each tier independently. In summary, a multi-tier synergic Trojan has the potential to impact a bigger area than a 2D Trojan. It will be challenging for module-level testing for a subsystem to identify the underlying security threat in the 3D system. The symptom of a synergic Trojan may seem benign from the viewpoint of a small local area. More importantly, the increased impact area of the synergic Trojan payload will make the technique of isolating malicious hardware ineffective or unrealistic since multiple tiers are involved.

Example Analysis

3D network-on-chip (NoC) [34, 35] has been demonstrated as a promising infrastructure to integrate increasing transistors in multiple tiers. 3D NoC eliminates the need for long global interconnects and reduces the voltage droop and power consumption on long wires. A rogue 2D NoC leads to information leaking and bandwidth depletion [36]. If NoC-based 3D ICs have a synergic Trojan placed in some IP cores or 3D switches, that Trojan leads to a similar consequence, as shown in Fig. 13. The rogue IP core sends an NoC instruction packet to the rogue switch. Next, the rogue switch passes that malicious packet to the victim IP core in the bottom tier. As a result, the multi-tier synergic Trojan

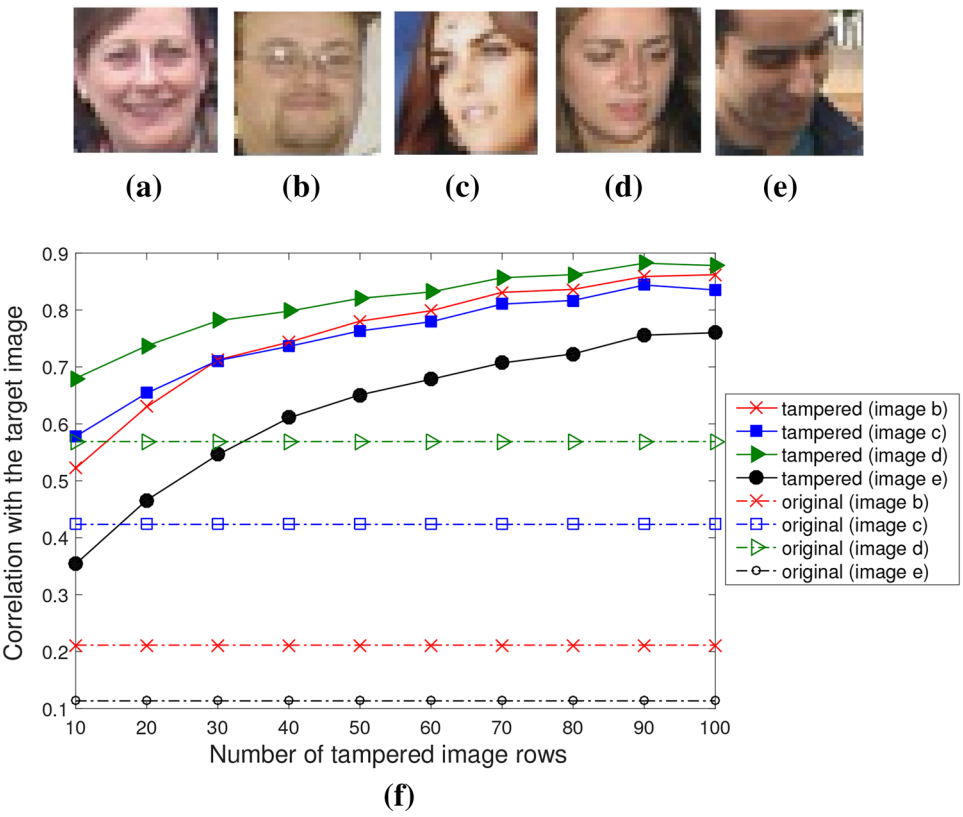

Fig. 12 Impact of multi-tier collaborative hardware Trojans in an image authentication application. **a** Image generated in tier 1, **b**–**e** images for comparison provided by tier 2, and **f** correlation analysis results obtained from tier 3

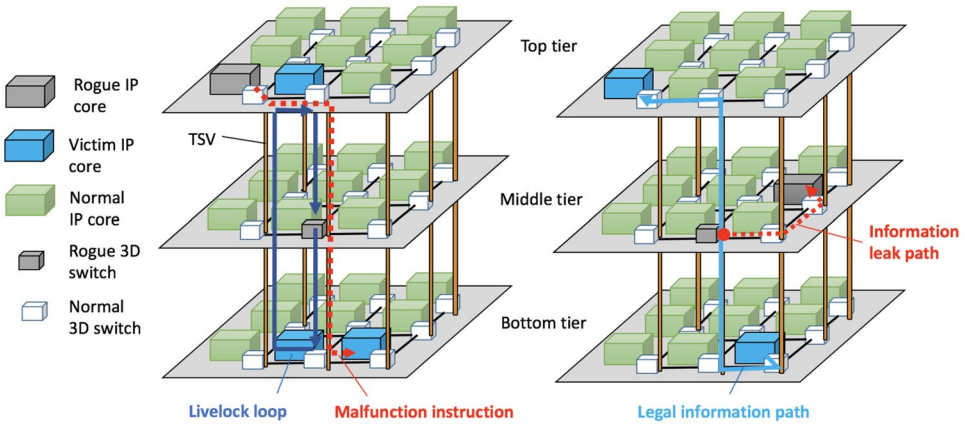

Fig. 13 Multi-tier synergic hardware Trojan payload causing malfunction, communication livelock, and information leaking

eventually causes the victim IP core to have malfunctions. Or, the rogue switch in the middle tier could trigger a livelock between the middle and bottom tiers. The proposed multi-tier synergic Trojan is stealthy because the hardware of the rogue IP core and switch has high similarity with the normal ones and the 'rogue' feature is only visible at the arrival time of special NoC packets. Figure 13 illustrates another practical example of the case 4 Trojan model. The rogue switch and IP core tampered by a hardware Trojan monitor the special packet transferring through the middle tier and the packet of interest in the rogue IP core is stored for future use and analysis. In the case of passing malicious packets in NoCs, the rogue IP core is the Trojan trigger to initialize the attack by issuing the malicious instructions. The rogue 3D switch is the payload, which causes malfunction by delivering malicious instructions to the victim IP cores. The trigger and payload are from different tiers but none of them is in the same tier where the victim locates. In the case of information leaking, the payload formed by a rogue 3D switch is responsible for leaking NoC packets. Although the trigger and payload for this case are in the same tier, they remotely control the victims in other tiers. The Trojan type proposed in this subsection is non-invasive. Moreover, the snooping attack is hidden in the normal data transmission of the middle tier. Side-channel analysis of the entire system may not be able to detect the presence of such hardware Trojans.

Examination of a 2D Trojan Detection Approach in 3D IC

The existing Trojan detection methods are mainly designed for the Trojans in 2D ICs. Due to the unique characteristics of 3D Trojans, as analyzed in Sect. "Proposed Comprehensive Characterization of3D Hardware Trojans", they may not work well in 3D scenarios. Split manufacturing may impact the hardware Trojan insertion in 3D ICs at some level. However, the adversaries in untrusted foundries with partial design details might be able to reverse engineer the whole design. Once the design is recovered, attackers can continue to insert Trojans. On the other hand, split manufacturing is not for securing the stacked 3D ICs in which every single tier is complete. This type of 3D IC is addressed in this work. New countermeasures specifically for 3D Trojans are needed.

In this section, we applied an existing approach [11], originally designed for 2D Trojans, to a 3D system and compared the effectiveness of Trojan detection in 2D and 3D ICs. As 3D chips have severe internal noise, we suspect that Trojan detection using side-channel signals will lose its detection accuracy. Thus, we chose a current based Trojan detection method.

Description of Trojan Detection Method for 2D ICs

The Trojan detection method we examined is Temporal Self-Referencing (TeSR) [11]. In TeSR, a special test vector generator offers the input sequence to ensure the system go through the identical state transitions in a period of time. A Trojan-free system should obtain identical current signatures in two consecutive time windows when it goes through the same state transitions. Any mismatch between the two current signatures will indicate the presence of a hardware Trojan. This method may not work well in 3D scenarios because of the greater internal noise in 3D ICs.

Targeted Hardware Trojan

In the following experiment, we inserted the same MOLES Trojan mentioned in [37] to the 2D and 3D circuits. The MOLES Trojan is composed of a set of registers as a ring generator to generate a series of random numbers, which will be XORed with the key information. The XOR outputs will drive a set of capacitors. Attackers who know the implementation details of the ring generator can decode the

obfuscated key information via power analysis. However, the power consumed in the load capacitors seems like noise if the random sequence is unknown. In the 2D case, MOLES was implemented as an external circuit on the same tier of the target circuit. In the 3D scenario, MOLES and the victim circuit were placed in two different tiers.

Efficiency of TeSR Trojan Detection Method in 2D and 3D ICs

We adopt the metric *point-wise Euclidean distance (PWED)* between the two current signatures to assess Trojan detection efficiency, following the similar process used in the work [11]. The PWED for the Trojan-free case (i.e. TrojanFree) is considered as the noise threshold. If the PWED measured from the Trojan injected case (i.e., $PWED_{TrojanIn}$) is higher than that measured from the Trojan-free case (i.e., $PWED_{TrojanFree}$), the hardware Trojan is detected.

We implemented the TeSR Trojan detection method in the transistor-level 3D IC model built with a 45nm NCSU FreePDK technology [32]. The detailed setting is as same as what described in Sect. "Example Analysis"). One, two, four, six, and eight S-boxes were applied for the purpose of sweeping the size of the victim circuit. The number of registers in the MOLES ring generator was varied to observe the impact of Trojan size on Trojan detection efficiency.

Our simulation results shown in Fig. 14 confirm that the TeSR Trojan detection method is generally less effective in the 3D scenarios than in the 2D cases. The inserted MOLES Trojan can be successfully detected in the 2D environment for all victim sizes tested in the experiment. In contrast, the Trojan in the 3D scenario is not detected in most of the cases because the 3D $PWED_{TrojanIn}$ is lower than $PWED_{TrojanFree}$. We further zoom in the PWEDs for different test cases and define the confidence level of Trojan detection $Confidence_{HTD}$ as the expression shown in Eq. (1).

$$Confidence_{HTD} = \frac{PWED_{TrojanIn} - PWED_{TrojanFree}}{PWED_{TrojanFree}} \quad (1)$$

Table 4 shows $Confidence_{HTD}$ for all the test cases reported in Fig. 14. A positive percentage means that the Trojan is detected. A higher percentage stands for better confidence in the detection result. If the positive percentage is too small, our detection conclusion may be changed by the interruption from some internal noise or process variations. Although TeSR achieves a positive confidence value in the 3D TrojanIn with 2 S-boxes case, the percentage of 12.61% is not as high as that in most of the 2D cases. A negative percentage in Table 4 indicates that the TeSR fails to capture the Trojan. To conclude, the MOLES Trojans in most of the 3D scenarios are not recognized by the TeSR approach.

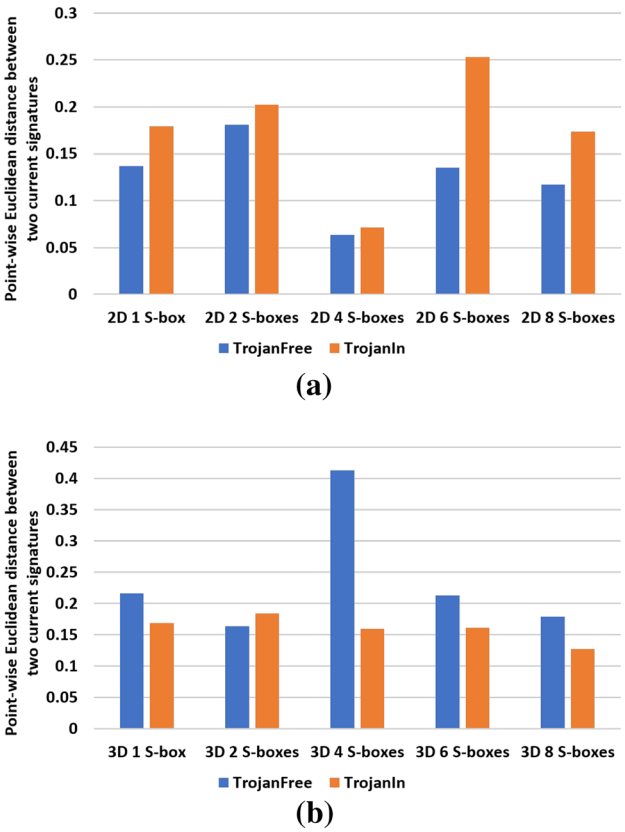

Fig. 14 Trojan detection results achieved by the TeSR approach applied in **a** 2D and **b** 3D ICs with different sizes of victim circuits

Table 4 Trojan detection confidence for different victim sizes

	1 S-box	2 S-boxes	4 S-boxes	6 S-boxes	8 S-boxes
2D	+31.07%	+11.84%	+12.80%	+80.06%	+48.00%
3D	-21.99%	+12.61%	-61.30%	-24.32%	-28.74%

Next, we swept the size of the MOLES Trojans from 20 to 80 registers and obtained the corresponding PWED shown in Fig. 15. As can be seen, the PWED for all 3D TrojanIn cases is less than the TrojanFree case. This indicates that the TeSR approach fails to detect the MOLES Trojans inserted in the 3D circuits even if the Trojan size increases. Another observation we had from our case study is, the PWED does not monotonically increase or decrease with the Trojan size. This is summarized in Table 5.

Conclusion

Three-dimensional integration techniques for integrated circuits leverage vertical-dimension space to increase the chip density and provide better performance than two-dimensional

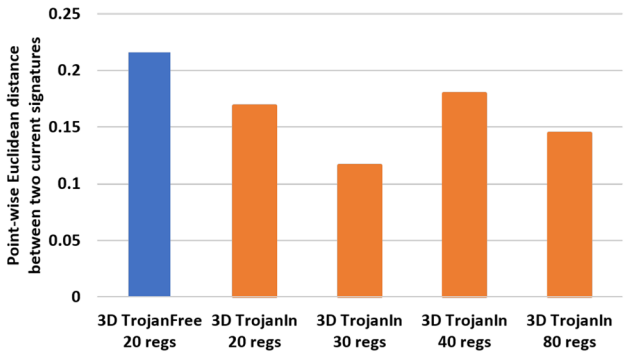

Fig. 15 Trojan detection efficiency of TeSR against 3D MOLES Trojans with different sizes

Table 5 Trojan detection confidence for different Trojan sizes

3D TrojanIn 20 regs	3D TrojanIn 30 regs	3D TrojanIn 40 regs	3D TrojanIn 80 regs
− 21.99%	− 46.34%	− 16.94%	− 33.15%

chips. However, the increased number of transistors in a small footprint leaves more exploration space for attackers to insert stealthy hardware Trojans. Trojans in planar integrated circuits are well modeled and understood, but there is limited work available to investigate hardware Trojans specifically in 3D ICs. This work summarizes the existing effort on 3D hardware Trojans. To improve the awareness of potential attacks that could succeed in 3D ICs, this work characterizes four representable 3D hardware Trojan cases and provides practical simulation/emulation examples for each model. To the best of our knowledge, this is the first comprehensive work that analyzes the 3D Trojan models, especially for cross-tier and multi-tier Trojans, and demonstrates their impact with the quantitative assessment. Our experimental results show that 3D Trojans are feasible to be implemented in 3D integrated circuits and systems. We advocate the research community to investigate unique Trojan detection methods for 3D hardware Trojans.

Acknowledgements This work was supported in part by Semiconductor Research Corporation (SRC) and National Science Foundation award no. 1717130.

Compliance with Ethical Standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Labrak L, O'Connor I. Heterogeneous system design platform and perspectives for 3D integration. In Proceeding of IEEE International Conference on Microelectronics, 2009; pp. 161–164.
2. Xue L, Liu CC, Kim H-S, Kim SK, Tiwari S. Three-dimensional integration: technology, use, and issues for mixed-signal applications. IEEE Trans Electron Devices. 2003;50(3):601–9.
3. Tummala RR. 3d system package architecture as alternative to 3d stacking of ics with tsv at system level. In: *2017 IEEE International Electron Devices Meeting (IEDM)*, 2017; pp. 3.4.1–3.4.3.
4. Tanaka T, 3d-ic technology and reliability challenges. In: *2017 17th International Workshop on Junction Technology (IWJT)*, 2017; pp. 51–53.
5. Li L, Su P, Xue J, Brillhart M, Lau J, Tzeng PJ, Lee CK, Zhan CJ, Dai MJ, Chien HC, Wu ST. Addressing bandwidth challenges in next generation high performance network systems with 3d ic integration. In: 2012 IEEE 62nd electronic components and technology conference, 2012; pp. 1040–1046.
6. Dofe J, Yu Q, Wang H, Salman E. Hardware security threats and potential countermeasures in emerging 3D ICs. In: Proceedings of GLSVLSI'16. ACM, 2016; pp. 69–74.
7. Juan D, Garg S, Marculescu D. Statistical thermal evaluation and mitigation techniques for 3D Chip-Multiprocessors in the presence of process variations. In: Proceedings of DATE'11, 2011; pp. 1–6.
8. Garg S, Marculescu D. System-level process variability analysis and mitigation for 3D MPSoCs. In: *Proceedings of DATE'09*, 2009; pp. 604–609.
9. Dofe J, Yu Q. Exploiting PDN Noise to Thwart Correlation Power Analysis Attacks in 3D ICs. In: 2018 ACM/IEEE International Workshop on System Level Interconnect Prediction (SLIP), 2018; pp. 1–6.
10. Tehranipoor M, Koushanfar F. A survey of hardware trojan taxonomy and detection. IEEE Des Test Comput. 2010;27(1):10–25.
11. Narasimhan S, Wang X, Du D, Chakraborty RS, Bhunia S, TeSR: A robust Temporal Self-Referencing approach for Hardware Trojan detection. In: 2011 IEEE International Symposium on Hardware-Oriented Security and Trust, 2011, pp. 71–74.
12. Bhasin S, Regazzoni F. A survey on hardware trojan detection techniques. In: IEEE International Symposium on Circuits and Systems (ISCAS). 2015;2015:2021–4.
13. Li H, Liu Q, Zhang J. A survey of hardware trojan threat and defense. Integration. 2016;55:426–37.
14. Francq J, Frick F, Introduction to hardware trojan detection methods. In: Design, Automation Test in Europe Conference Exhibition (DATE). 2015;2015:770–5.
15. Inoue T, Hasegawa K, Yanagisawa M, Togawa N, Designing hardware trojans and their detection based on a svm-based approach. In: Proceedings of ASICON'17, 2017; pp. 811–814.
16. Zarrinchian G, Zamani MS. Latch-based structure: a high resolution and self-reference technique for hardware trojan detection. IEEE Trans Comput. 2017;66(1):100–13.
17. Salmani H, Tehranipoor M, Plusquellic J. A novel technique for improving hardware trojan detection and reducing trojan activation time. IEEE Trans Very Large Scale Integr VLSI Syst. 2012;20(1):112–25.
18. Banga M, Hsiao MS. Odette: A non-scan design-for-test methodology for trojan detection in ics. In: IEEE international symposium on hardware-oriented security and trust. 2011;2011:18–23.
19. Zhang Z, Yu Q, Modeling Hardware Trojans in 3D ICs. In: Proceedings of 2019 IEEE Computer Society Annual Symposium on VLSI (ISVLSI), 2019; pp. 483–488.
20. Mossa SF, Hasan SR, Elkeelany O. Hardware trojans in 3-D ICs due to NBTI effects and countermeasure. Integration. 2017;59:64–74.
21. Hasan SR, Mossa SF, Elkeelany OSA, Awwad F. Tenacious hardware trojans due to high temperature in middle tiers of 3-d ics. In:

2015 IEEE 58th International Midwest Symposium on Circuits and Systems (MWSCAS), 2015; pp. 1–4.
22. Dofe J, Gu P, Stow D, Yu Q, Kursun E, Xie Y. Security threats and countermeasures in three-dimensional integrated circuits. In: Proceedings of the on Great Lakes Symposium on VLSI 2017. ACM, 2017; pp. 321–326.
23. Madani S, Bayoumi M. A Security-Aware Pre-partitioning Technique for 3D Integrated Circuits. In: Proceedings of MTV'17, 2017; pp. 57–61.
24. Alhelaly S, Dworak J, Manikas T, Gui P, Nepal K, Crouch AL. Detecting a trojan die in 3D stacked integrated circuits. In: 2017 IEEE North Atlantic Test Workshop (NATW), May 2017; pp. 1–6.
25. Madani S, Madani MR, Dutta IK, Joshi Y, Bayoumi M. A hardware obfuscation technique for manufacturing a secure 3D IC. In: Proceedings of MWSCAS'18, 2018; pp. 318–323.
26. Yang P, Marek-Sadowska M. Making split-fabrication more secure. In: 2016 IEEE/ACM international conference on computer-aided design (ICCAD), 2016; pp. 1–8.
27. Salah K, Survey on 3d-ics thermal modeling, analysis, and management techniques. In: 2017 IEEE 19th electronics packaging technology conference (EPTC), 2017; pp. 1–4.
28. King Jr CR, Thermal management of three-dimensional integrated circuits using inter-layer liquid cooling. Ph.D. dissertation, Georgia Institute of Technology (2012).
29. Marinissen EJ, Challenges and emerging solutions in testing TSV-based 2 1 over 2D- and 3D-stacked ICs. In: Proceedings of DATE '12, 2012; pp. 1277–1282.
30. Ngo XT, Najm Z, Bhasin S, Roy DB, Danger J-L, Guilley S. Integrated sensor: a backdoor for hardware trojan insertions? In: Proceedings of 2015 Euromicro conference on digital system design, 2015; pp. 415–422.
31. Zhang Z, Yu Q. 3D Thermal Triggered Hardware Trojan. Hardware Demo. In: IEEE International Symposium on Hardware Oriented Security and Trust (HOST), 2019. https://easychair.org/smart-program/HOST2019/2019-05-08.html#talk:90210.
32. Satheesh SM, Salman E. Power distribution in TSV-based 3-D processor-memory stacks. IEEE J Emerg Select Top Circ Syst. 2012;2(4):692–703.
33. Nguyen VH, Christie P, Heringa A, Kumar A, Ng R. An analysis of the effect of wire resistance on circuit level performance at the 45-nm technology node. In: Proceedings of IITC'05, 2005; pp. 191–193.
34. Jabbar MH, Houzet D, Hammami O. 3D multiprocessor with 3D NoC architecture based on Tezzaron technology. In: *Proceedings of 3DIC '11*, 2012; pp. 1–5.
35. Jiang L, Xu Q, Fault-Tolerant 3D-NoC architecture and design: recent advances and challenges. In Proceedings of NOCS '15, 2015; pp. 7:1–7:8. https://doi.org/10.1145/2786572.2788709.
36. Frey J, Yu Q, A hardened network-on-chip design using runtime hardware Trojan mitigation methods. Integration, the VLSI Journal, 2017; 56, 15–31. Available: http://www.sciencedirect.com/science/article/pii/S0167926016300311.
37. Lin L, Burleson W, Paar C. MOLES: Malicious off-chip leakage enabled by side-channels. *2009 IEEE/ACM International Conference on Computer-Aided Design - Digest of Technical Papers*, 2009; pp. 117–122.