Abstract—Three-dimensional (3D) integration facilitates to integrate increasing number of transistors into a single package. Despite of improved performance and power efficiency, the integration of multiple dies into the same package potentially leads to new security threats, such as 3D hardware Trojans. In this work, we first provide a thorough survey of reported hardware Trojans in 3D integrated circuits and systems, and then propose comprehensive 3D hardware Trojan models. A case study is performed to verify the implementation feasibility of thermal-triggered 3D Trojan. The activation speed of the 3D Trojan is compared to its 2D counterpart to confirm that 3D IC provides a better environment to hide thermal Trojans.

Index Terms—3D IC, hardware Trojan, side-channel analysis attack, interconnect, through-silicon via (TSV)

I. INTRODUCTION

Three-dimensional (3D) integration is an emerging technology to ensure the 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. Unfortunately, 3D techniques also bring in unique and unexplored security threats to 3D ICs [3]. It is more challenging to address the security threats in 3D ICs than in 2D planar chips.

A larger number of transistors integrated into the single package makes the exhaustive functional testing more sophisticated and time consuming. Limited probing capacity does not allow us to simultaneously access all die-to-die vertical communication channels for thorough functional testing. Although each die and Through Silicon Via (TSV) can be examined during the pre-bond and mid-bond stages, the testing probe may damage some TSVs and thus harm downstream integration [4]. Due to these challenges on traditional testing, malicious component detection via functional testing on 3D ICs is not optimistic [5].

Larger variation on temperature, process, and voltage in 3D ICs may lead to a higher false positive rate if we adopt side-channel signals based hardware Trojan detection methods for 2D circuits. The work [6] provides a temperature distribution comparison among 2D, 2-tier 3D and 4-tier 3D Chip-Multiprocessors (CMPs). That work indicates that the variation on temperature increases with the increasing number of 3D tiers. The standard derivation on temperature for the 4-tier 3D chip is approximately 40 times higher than that for the 1-tier 2D chip. Using an analytical system-level variability model, the work [7] predicts that the performance degradation due to the process variation in a 3D MPSoC is larger than that in the 2D counterpart. The work [7] also predicts that the performance degradation will get even worse when the number of 3D tiers increases, as the process variation in 2D, 2-tier 3D and 4-tier 3D will degrade perform by 8%, 14% and 17%, respectively.

As discussed above, more thorough testing approaches and harmless facility are needed to perform hardware Trojan detection in 3D ICs. To assure the integrity and security of 3D chips, it is imperative to improve our knowledge on potential hardware Trojans in the context of 3D integration. However, the existing work considering 3D-Trojan threats mostly focuses on the Trojan insertion in malicious foundries. Limited work addresses the exact Trojan models which could be implemented in 3D ICs. High-level modeling for 3D Trojans and quantitatively assessment on side-channel signals will benefit the hardware security community to propose effective countermeasures against 3D hardware Trojan insertion. More specifically, our main contributions are as follows.

1) This is the first work that does a thorough survey on hardware Trojans in 3D ICs. Threat models and Trojan models reported in the existing literature are compared in details.
2) Four high-level 3D hardware Trojan models are proposed in this work. Practical examples for each Trojan model is provided, as well.
3) Thermal-induced 3D hardware Trojan is emulated in a platform composed of FPGA, microcontroller, heat generator and heat sink to assess the Trojan trigger/detection speed in a passcode based authentication application.

The rest of this paper is organized as follows: Section 2 briefly introduces the preliminary knowledge on hardware Trojan in 2D ICs. Section 3 summarizes the threat model and hardware Trojan model for 3D Trojans in existing literature. Section 4 proposes comprehensive abstract model of cross-tier 3D Trojan and practical Trojan implementation. Section 5 provides simulation and emulation results for the 3D Trojan implementations. This paper is concluded in Section 6.

II. PRELIMINARIES TO HARDWARE TROJANS

Hardware Trojans are malicious modification made on hardware to fulfill attackers’ intentions such as sabotaging the original function carried by the target hardware, causing hardware
performance degradation, and leaking confidential information embedded in the hardware. Hardware Trojan insertion could occur in several stages of the IC supply chain, including functional design stage [8], netlist synthesis stage [9], and fabrication stage [10], [11]. The survey paper [12] provides a comprehensive classification of hardware Trojans in terms of Trojan insertion stages, trigger conditions, and payload impact.

A. Trojan Triggering Mechanisms

Depending on how often the hardware Trojan is triggered, the work [13] presents three types of Trojan, always-on, combinational condition and sequential condition. The commonly observed combinational and sequential hardware Trojans [13] typically rely on a group of internal signals to form the rarely triggered condition. While the combinational Trojan enables the payload circuit as soon as the trigger condition is satisfied, the sequential Trojan waits for multiple arrivals of the same trigger condition. An example of always-on Trojan is realized with a ring oscillator, which is composed of eight inverters and causes extra power consumption but not harming the system function. Another example is parametric Trojan, which slightly alters the geometry defined by the sign-off GDSII, doping concentration, dopant polarity, or doping area [14], [15]. As a result, the new voltage transfer characteristic of a CMOS transistor will lead that transistor to have a shifted switching threshold.

B. Trojan Payload Mechanisms

After the arrival of Trojan trigger condition, the operation defined in Trojan payload will be executed. Trojan payload is used to implement attackers’ intention, such as modifying the logic value of the original function, or executing malicious operations in hardware. Among different payload designs, covert channel is one of the most challenging payloads for detection. A covert channel generated with the assistance from hardware Trojans could help adversary to extract confidential information, such as the encryption key, without disturbing normal operations of the victim system. A hardware Trojan introduced in [16] monitors the subkey access in the process of encryption and copies the subkeys to an internal memory. After being XORED with the original signal, the encoded subkeys are transmitted via the universal asynchronous receiver transmitter (UART) line. By probing the UART channel, attackers can retrieve the original encryption key. Some hardware Trojans also leverage the internal memory unit to build the covert data channel [17], [18].

III. OUR SURVEY ON EXISTING HARDWARE TROJANS IN 3D ICS

A. 3D-Trojan Insertion Scenarios

The increased number of transistors and the vertical dimension integration in 3D ICs potentially leave adversary more exploration space to implement hardware Trojans. As a general trend, more and more chip designs will be outsourced for fabrication. Not all single die fabrication foundries and vertical interconnect (e.g. TSV) manufacturers are trusted. The die-to-die bonding may be performed in an untrusted foundry, too. Figure 1 shows a simplified process of 3D IC manufacturing. We highlight the possible attack surfaces for Trojan insertion. Trojans can be implemented during the single die manufacturing foundries, independently or cooperatively. Since the bonding foundries have access to all the single dies, they have a good chance to implement a Trojan involving multiple dies.

Based on the existing literature, we categorize the 3D Trojans in Table I. We highlight the threat model with special emphasis on threat source and attack target. In addition to Trojan trigger and payload mechanisms, we also point out Trojan locations in the 3D ICs. From the table, we can see that the nature of 3D IC structure creates new opportunities for hardware Trojan design, such as thermal-based Trojans and cross-tier Trojans.

In the next three subsections, we discuss the existing literature listed in Table I according to their special trigger mechanisms and Trojan locations.

B. Thermal Trojans

Thermal issue due to poor heat dissipation in a 3D stack can be exploited to develop Trojan triggering mechanisms. 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 exploited as Trojan triggers. This type of hardware Trojans does not need explicit trigger circuits. As indicated in [19], [20], thermal-triggered Trojans can be inserted by any malicious foundries with the access to the layout of design. Those Trojans likely locate near the middle tier, where the heat is harder to dissipate than in other tiers [20]. The work [3] demonstrates that the thermal triggered Trojan may also be hidden in interposer to cause short circuit or these thermal Trojans can be utilized to speedup circuit component aging and consequently lead to Deny-of-Service (DoS).

C. Cross-Tier Trojans

The multiple-die structure of 3D ICs allows attackers to spread the Trojan circuit to multiple tiers. Either the trigger circuit and payload circuit are separated into different tiers, or the trigger circuit being split and located in multiple tiers jointly activates the payload [21]. The Trojans inserted in these ways may bypass the functional testings on individual dies, because those Trojans are not activated by just running tests on individual tiers. The Trojans could be inserted by
TABLE I
EXISTING WORK ON HARDWARE TROJAN IN 3D ICs

<table>
<thead>
<tr>
<th>Existing Work</th>
<th>Threat Model</th>
<th>Trojan Model</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Threat source</td>
<td>Attacker access</td>
<td>Trigger</td>
</tr>
<tr>
<td>[19]</td>
<td>Untrusted die foundries</td>
<td>GDSII files</td>
<td>Thermal effect caused transition glitches</td>
</tr>
<tr>
<td>[20]</td>
<td>Untrusted die foundries</td>
<td>GDSII files</td>
<td>Thermal effect caused transition glitches</td>
</tr>
<tr>
<td>[21]</td>
<td>Untrusted interconnect foundries</td>
<td>GDSII files</td>
<td>Remote circuit</td>
</tr>
<tr>
<td></td>
<td>Untrusted single die manufacturers</td>
<td>GDSII files</td>
<td>Aging effect</td>
</tr>
<tr>
<td></td>
<td>Untrusted unified foundries</td>
<td>GDSII files</td>
<td>Thermal effect</td>
</tr>
<tr>
<td>[22]</td>
<td>Untrusted single die manufacturers</td>
<td>Least critical die</td>
<td>Low-activity nets</td>
</tr>
<tr>
<td>[23]</td>
<td>Untrusted assemblers</td>
<td>No legitimate dies</td>
<td>No special requirement</td>
</tr>
<tr>
<td>[24]</td>
<td>Final bounding foundries</td>
<td>Entire layers</td>
<td>Internal nets</td>
</tr>
<tr>
<td>[25]</td>
<td>Untrusted single die manufacturers</td>
<td>GDSII files</td>
<td>No special requirement</td>
</tr>
</tbody>
</table>

untrusted die manufacturing foundries, interconnect foundries and unified foundries. The work [22] also introduces a Trojan, which takes advantage of this particular structure of 3D ICs to steal encryption keys. The Trojan and the target encryption unit are located in different tiers and the Trojan is triggered with low-activity nets. Even the untrusted foundry with partial knowledge of the chip can launch this kind Trojan attack.

D. Trojans Exploiting Other 3D Features

The work [23] 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, the extra delay induced by the 3D hardware Trojan is difficult to be differentiated from process variation. This type of Trojans can be inserted by untrusted assemblers even without the access to legitimate dies. The work [24] introduces the scenarios that attackers are in the foundry who bonds all the outsourced dies. In [25], the adversaries are untrusted die manufacturing foundries with the access to GDSII files. These two works [24], [25] do not have thorough discussions on exact Trojan models.

IV. PROPOSED COMPREHENSIVE 3D-TROJAN MODELS

The major difference between 2D and 3D hardware Trojans is whether or not the Trojan trigger and payload circuits is 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 done in a separated fashion. The die for each tier is tested individually before 3D integration. Once the good dies are stacked vertically, very limited testing will be performed to detect the defects between die-to-die connections, rather than extensively examining the correctness of the 3D system behaviors.

Based on our survey in Section III, we propose four 3D Trojan models shown in Fig. 2. To the best of our knowledge, this is the first work that introduces comprehensive high-level 3D-Trojan models. In cases 1, 2, and 3, the Trojan trigger circuit and payload circuit are located in different 3D tiers. The case 4 describes the always-on Trojan, in which the payload circuit is on the tier that does not carry the victim circuit. The following subsections will discuss the four Trojan cases.

A. Cross-Tier Trojan Trigger

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 Trojans is similar with the 2D Trojans that are triggered by an external signal. For instance, the work [26] demonstrates an external probe sensor that initiates the Trojan by manipulating the ring oscillator and LC coil. As the emerging of heterogeneous 3D integration, the external trigger can be originated from the other tiers. Since the payload circuit is never or rarely enabled without the valid cross-tier
trigger signal, the symptom of the target under Trojan attack will not be observed. Thus, this type of Trojans 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 is passed 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.

B. Cross-Tier Trojan Payload

In the case 2 shown in Fig. 2, the payload of a stealthy 3D Trojan is located in the top tier (tier 1), from where it is relatively easier to probe and measure side-channel signals than from the middle tier(s). The motivation of this type of 3D Trojans is to snoop the confidential information from the victim unit (e.g., crypto engine). As the payload resides in another tier, the effect of this kind of Trojan will not be observable when we do tier-level testing. 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 victim unit in the example shown in Fig. 4 is an AES encryption module. The crypto key is loaded from the volatile memory in tier 1. The purpose of this Trojan is to leak the crypto key. To prevent the key leakage from being captured during the tier 2 testing, the snooped key is transformed into another format (i.e., obfuscated key), and then the Trojan passes the obfuscated key to the rarely used main memory in tier 1. When we test tier 1, the main memory function is normal. As a result, the separated testing on tiers 1 and 2 will not reveal the presence of the 3D Trojan. However, the key will be leaked by the covert channel built by the cross-tier 3D Trojan.

C. Multi-Tier Collaborative Trojan Trigger

The cross-tier hardware Trojan 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, this type of cross-tier hardware Trojan may have significantly lower Trojan triggering probability due to a larger pool of trigger signals. Similar with the example mentioned in Section IV-A, the collaborative Trojan trigger could be a combination of different trigger mechanisms (e.g., temperature, voltage level, and electromagnetic flux). Alternatively, the trigger circuits are composed of multiple portions, which are distributed in multiple tiers.

3D network-on-chip (NoC) [27], [28] 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 [29]. If NoC based 3D ICs have a collaborative Trojan placed in the IP core and 3D switch, that Trojan leads to the similar consequence, as shown in Fig. 5. The rogue IP core sends a 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 collaborative Trojan eventually causes the victim IP core having malfunctions. Or, the rogue switch in the middle tier could trigger a livelock between the middle and bottom tiers. The proposed multi-tier collaborative 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.

D. Information Leaking in Passive Layer

In the case 4 shown in Fig. 2, the Trojan circuit snoops the data (or even the side-channel signal) available in the middle tier. Thanks to the vertical integration of heterogeneous tiers, it is much easier to implement the snooping attack in a malicious tier. Compared with 2D chips, a thin malicious tier provides better flexibility and control on the snooped
Fig. 6. Cross-tier collaborative hardware Trojan causing information leaking. As envisioned in the work [23], a Trojan tier could observe the information passing between tiers without leading to noticeable delay overhead. Figure 6 illustrates a practical example for the case 4 Trojan model. As shown, the rogue switch and IP core monitor the special packet transversing through the middle tier and the packet of interest in the rogue IP core is stored for future use and analysis. The Trojan type proposed in this subsection is non-invasive, as the Trojan does not alter the normal operation and communication of the system. Moreover, the snooping attack is hidden in the normal data transmission of the middle tier. Side-channel analysis on the entire system may not be able to detect the presence of such hardware Trojan.

V. EXPERIMENTAL VERIFICATION ON THERMAL TROJAN

A case study on the thermal-triggered Trojan shown in Fig. 3 is performed in this section. We performed a case study on a platform composed of Xilinx Nexys3 Spartan-6 FPGA, TI MSP430FR6989 LaunchPad board, IRF540 MOSFET transistor, and NTC thermistor. The purpose of this case study is to verify the implementation feasibility of the thermal Trojan and compare its activation efficiency between the scenarios of 2D and 3D ICs.

A. Experimental Setup

In the experiment below, we demonstrate how an attacker uses thermal-triggered Trojan to compromise the authentication system. The overview of our experimental setup is shown in Fig. 7. The two isolated blocks in the breadboard are used to mimic two adjacent tiers in a 3D IC. A heat generator circuit and a thermal sensing circuit are implemented in the two blocks, respectively. The main component of the heat generator circuit is a MOSFET driven by the FPGA board shown Fig. 7. The sensor circuit composed of a NTC thermistor connected with multiple resistors in series is powered by the TI microcontroller shown in Fig. 7. When the thermistor senses temperature rising, its resistance starts to drop, which leads to a decrease on its voltage level. To simulate the 2D scenario for comparison, we add a heat sink for the heat generator circuit, as shown in Fig. 8, to provide a better heat dissipation which is commonly provided in 2D IC.

A Trojan trigger logic is programmed in the FPGA to monitor two input signals, which are controlled by the two switches. An authentication system is programmed in the microcontroller to examine the password provided externally. The microcontroller also detects the voltage level of the thermistor. 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 temperature increasing, then the microcontroller detects the change on voltage and drives the authentication system jump to the password reset status. We successfully mimicked a 3D thermal-triggered hardware Trojan and overwrote the authentication password.

B. Trojan Activation Efficiency

In this subsection, we compare the activation speed of the thermal-triggered Trojans emulated 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 thermistor exceeds the threshold, the Trojan payload will reset the authentication password. The trigger of our Trojan is thermal effect. We warmed the air surrounding the thermistor with and without the heat sink to mimic 2D and 3D scenarios, respectively. Then, we measured the time that the thermistor takes to drop the voltage below the threshold for each case. The results are shown in Table II. In the 2D scenario, the Trojan needs almost twice of the time to be activated compared to the 3D scenario. Due to the poor heat dissipation, it is easier to implement thermal-triggered Trojans in 3D ICs. We also measured the temperature changing within 12 minutes for each scenario. The changing speed is reflected by the resistance of thermistor. The dropping trend of the resistance is plotted in Fig. 9. As can be seen, the NTC thermistor’s resistance for the 3D case drops faster than that for the 2D case. This fact further confirms that heat can be better accumulated in 3D than 2D and 3D ICs will facilitate the implementation of thermal-based Trojans.
TABLE II

<table>
<thead>
<tr>
<th>Emulation scenarios</th>
<th>Time to trigger Trojan (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2D</td>
<td>11:12</td>
</tr>
<tr>
<td>3D</td>
<td>6:52</td>
</tr>
</tbody>
</table>

Fig. 9. Speed of thermistor resistance dropping.

VI. CONCLUSION

Vertical integration of multiple dies into a single packet potentially leaves more exploration space for attackers to insert stealthy hardware Trojans. Due to the limited testing techniques for 3D ICs, the hardware Trojans in 3D ICs are not easy to detect. In this work, we propose four 3D hardware Trojan models, which are unique for 3D ICs. We provide a practical implementation for each proposed Trojan model and analyze its stealthiness. FPGA and microcontroller based platform is developed in this work to emulate the thermal-triggered hardware Trojan in 3D ICs. Our experimental results show that 3D Trojan can be successfully triggered with a faster speed than 2D ones.

ACKNOWLEDGEMENT

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

REFERENCES