A high efficiency Hardware Trojan detection technique based on fast SEM imaging

Franck Courbon*†, Philippe Loubet-Moundi*, Jacques J.A. Fournier‡ and Assia Tria‡

*GEMALTO, Security Labs, La Ciotat, France
†Ecole des Mines de Saint-Etienne, CMP-GC/LSAS, Gardanne, France
‡CEA, CEA Tech Region, DPACA/LSAS, Gardanne, France

***Abstract*—In the semiconductor market where more and more companies become fabless, malicious integrated circuits' modifications are seen as possible threats. Those Hardware Trojans can have various effects and can be implemented by different entities with different means. This article includes the integration of an almost automatic Hardware Trojan detection. The latter is based on a visual inspection implemented within the integrated circuit life cycle. The proposed detection methodology is quite efficient regarding tools, user experience and time needed. A single layer of the chip is accessed and then imaged with a Scanning Electron Microscope (SEM). The acquisition of several hundred images at high magnification is automated as does the images registration. Then depending on the reference availability, one can check if any supplementary gates have been inserted in the design using a golden reference or a graphic/text design file. Depending on the reference, either basic image processing is used to compare the chip extracted image with a golden model or some pattern recognition can be used to retrieve the number of occurrences of each standard cell. The depicted methodology aims to detect any gate modification, substitution, removal or addition and so far require an invasive approach and a reference.**

INTRODUCTION

In Hardware Trojan detection research activities, several families of techniques have been studied and proposed [1]. However, the choice of the detection technique can be done by performing a vulnerability analysis regarding HT insertion risk within the entire production chain. Once the associated chain of trust is known, the locations where it is possible and interesting to apply detections steps are identified. The goals of the detection technique can vary depending on the final application of the circuit.

In our study, the real case of secure circuits development and supply chain like smart cards is used as example. The role of a software developer that receives the circuit from a chip maker is taken as hypothesis. The selection of the HT detection technique is driven by the following parameters that are relevant for the given application: detection time available, kind of HT detected, detection success rate, size of the HT detected in digital circuit, robustness regarding process variation, golden model availability and cost of detection.

To the best of our knowledge, the only method achieving a '100%' efficiency is a full reverse engineering which is cost and time expensive [2]. Whereas other detection methods such as Side Channel Analysis (SCA), affected by process and measure variations reach a lower detection rate, and may not cover any a HTs size or activity. For instance, *Aarestad et al* [3] do not detect small Hardware Trojans yet, do not take count of process variations and give only simulations results. Delay-based detection are only simulations so far. Both SCA and delay-based techniques may also require large data sets and may be sensitive to process and setup measure variations. Closer to our approach, only simulations have been done on a partial reverse engineering hardware trojan detection [4] and only top metal layer hardware trojan detection tests have been performed on FPGA design tool images [5]. The main contribution of the paper is to introduce a methodology based on standard techniques, reaching a '100%' efficiency for any HT size or activation type, taking count of the reference type availability and tested on an integrated circuit.

I. PROCESS FLOW VS CHAIN OF TRUST

A. Dedicated process flow

Integrated life-cycle steps with their respective level of trust has been previously depicted [6]. For a given study, it must be adapted to reach the real production flow. The confidence level can also be different depending from who is looking the process flow. e.g. the chip designer could trust the chip design steps whereas the final user could have a lower confidence in design steps. That's why the role of the person who is doing the process flow vulnerability analysis must also be considered.

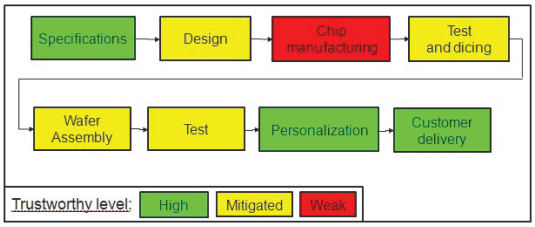


Fig. 1. Integrated circuit standard life cycle

The Fig. 1 represents the view of integrated circuit life cycle. In the industry where designer becomes fabless, we first assume that the wafers don't return especially to the designer, thus the integrity verification can directly be done by the client. The manufacturing flow is not impacted by this detection method. Thus, each client could decide to use or not this detection method depending on his security needs.

B. Choice of the detection technique

According to the modified process flow, possible detection steps could be introduced during wafers incoming tests or

978-3-9815370-4-8/DATE15/©2015 EDAA

during the tests performed after chip packaging. As the test time constrain after assembly is a key element in production cost it was not easy to add additional tests for HT detection in that step. However, delays can be observed between wafer receptions and production start. Inserting detection step at this stage seems relevant and gives sufficient time to perform deeper investigations. The duration of the detection is not the only important parameter; the other points are described here after. The technique must be able to detect digital HT inserted after a trusted chip design, the detection rate must be close to '100%', the detection success rate must detect HT's of any size -smaller or larger Trojans can be detected in the same way-, the process variation must not impact the detection rate, the related cost of the detection must be in line with the risk analysis performed, a physical golden model is not always available.

As naked dice - wafer delivery - are available, the idea is to use physical chip observation. As non functional dice (generally identified by an ink spot) are also present on the wafer frame, they can be used for destructive investigations without impacting the yield. Depending on the steppers capability, lithographic steps are made on several chips at once. A correct sampling could be put in place to cover several wafer locations or several wafers on a batch and detect potential HT that could be inserted only on some wafers or some wafer locations.

II. Methodology

A. Assumptions on HT insertion

One of the assumptions is that the chip designer is not fully trusted. It potentially be possible for a very few number of people to have access to the design database. Modifications for the HT insertion can be, VHDL modification, gate insertions during place and route, gate insertion after place and route. In the current use case, modifying the circuit during manufacturing seems less relevant has many organizational and environmental procedures are used for secure chip manufacturing, it is also considered that reverse-engineering the GDSII and performing few changes to add the required HT is not realistic due to manufacturing lead times.

So, the result of the infection will increase or decrease the overall gates number, will add or remove few gates before or after place and route.

B. Invasive approach

The methodology is based on the hardware chip structure itself. Performing the full physical chip reverse engineering to detect a netlist modification is not possible due to timing constraints. However, a quick and easy sample preparation associated with an efficient observation technique could be considered. According to the assumptions on the insertion effect, a unique hardware layer of the chip could be sufficient to check the presence of a Hardware Trojan. At the layer of interest, only the transistors' active regions are implemented, thus any gates modification, add, removal is detected. To reach this layer, the sample preparation is fast and low-cost. Then, a Scanning Electron Microscope is used to image the component and the multiple image acquisition and stitching is fully automated. Then, image processing allows finding the number and the location of each standard cell. Finally, the tool compares these previous informations with a design file. The Fig. 2 illustrates the impact over the integrated life cycle.

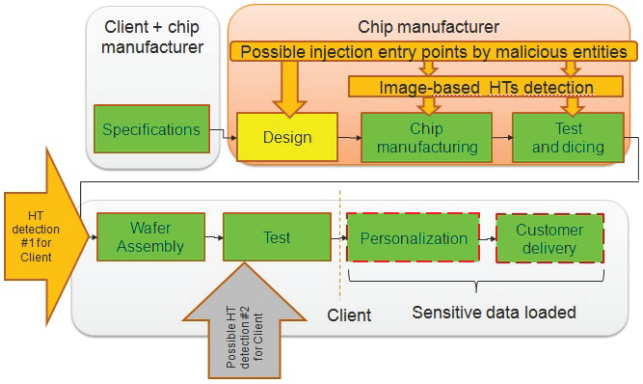


Fig. 2. Adapted integrated circuit flow with SEM-based detection

In our case the role of a final smart card vendor is chosen. Tested circuits are delivered on sawn wafers from the chip manufacturer. The chain of trust is modified to be adapted accordingly. Then, only after this work the locations where it could be possible and useful to insert detection steps are identified. In the example chosen, the global threat identified is the leakage of sensitive data on the field. The last step where the detection is possible is during assembly test before any sensitive data loading.

Moreover, our image-based Hardware Trojans detection, comparing a post-design file with data gained at the wafer reception make sure that intermediary steps are trusted. This technique enables a '100%' detection.

C. Methodology overview

The Fig. 3 illustrates the proposed methodology composed of three different steps; each step is impacted by the previous one. For Step $\#S3$, we give three different detection schemes depending on the reference file type availability.

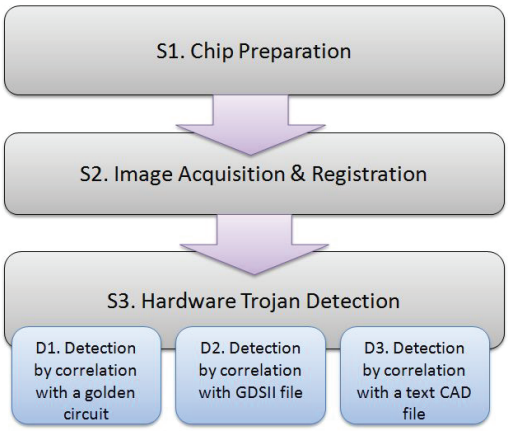


Fig. 3. Global methodology flowchart

D. Methodology pros and cons

The proposed methodology answers to the concerns described at the beginning of the paper. Main advantages of the technique are the following ones: low cost, easy preparation, do not affect the yield, full chip covering, valid for any CMOS technology node, practiced technique.

On the other hand, our technique needs a reference, is strongly dependant on the preparation step and also depends on image processing used.

These pros and cons are non exhaustive but allows having a global overview. The methodology being introduced, we first briefly describe the circuit under test before applying this methodology on the latter.

E. Device Under Test

The device used is a hardware cryptographic circuit [7]. The latter is composed of 15000 standard cells covering an area about 700 by $700\mu m$. The device is made of 5 Metal layers in a $130nm$ technology node. Only synthesized logic is present over this circuit, any circuit modification would affect the circuit logic. So far our investigations are based on simulated infected circuit, the infection brings 4 additional standard gates within the chip that are added in the bottom left area. This is representative of an insertion after the place and route step that uses free spaces.

III. Preparation/Acquisition

A. Chip preparation

The first step in our methodology is to prepare the chip in order to access the physical layer of interest. We use different preparation baths, an HydroFluoric (HF) acid bath is performed to access the source and drain implementations of each single transistor. Only p-well and n-well remain on the substrate [8]. Afterwards, to remove residues, a cleaning ultra-sonic bath is used. Thus, the complete preparation step does not require a specialized operator, it is done in several minutes and requires few euros but also personal protective equipment.

B. Image acquisition & registration

The second step of the methodology is the acquisition at high magnification of the implementation of each standard cell. We choose working with a SEM. The main advantage of SEM is to be able to acquire images with no resolution limit as it gets rid of light resolution limit unlike optical microscopy. Moreover, no careful tilt adjustment has to be done, however some minutes are required at the machine start up to keep the vacuum in the SEM chamber. The sample is placed under the Scanning Electron Microscope column, and several parameters of the electron gun and the detector can be controlled. Some of them are the contrast, luminosity, detector type, detector current, accelerating voltage, working distance, magnification, acquisition area size and the acquired zone. In our case, we use a $2,2k$ times magnification as it allows to differentiate shapes at this magnification (Fig. 4).

Once the image is good enough according to the operator, no more detector, electron gun or image parameters have

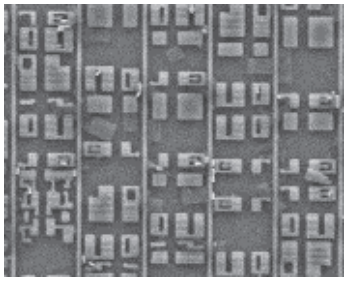


Fig. 4. Image part of a $2,2kX$ magnification

to be changed. We then define the area to scan, 700 by $700\mu m$, the wished acquisition magnification, $2,2kX$, and the overlap between each images, '10%'. This overlap will facilitate image registration based on similar points over two successive images. A routine is created under the SEM to automatically scan the area. The routine is a loop done until the full matrix is recovered consisting in freezing the image once the electron beam covered all the acquired zone line by line, saving the displayed image and moving the stage. In our case, 64 images are required to reconstruct the full image at a magnification which is sufficient for our investigations. The matrix of images size depends on the scanned area size and the magnification needed. Whereas the acquisition speed mainly depends on the electron gun scanning speed affecting the image noise as seen in Fig. 5.

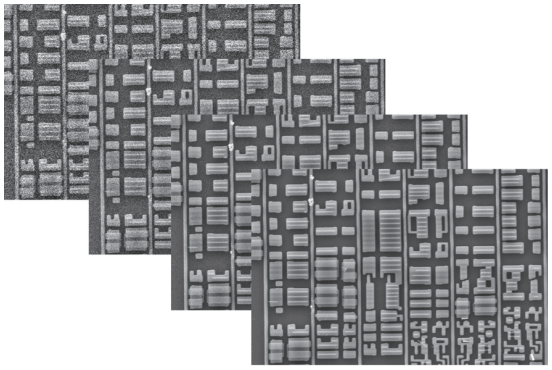


Fig. 5. From left to right: Fast to low scanning speed acquisitions

At the end, it results in a set of 8 by 8 images, captured one by one that has to be registered together to reconstruct the whole chip. In the next section, this image will be called 'physical extracted view'. Different open source softwares [9] or libraries to perform this task are available online. They can be based on invariant features over each successive images and requires a couple of minutes. The result does not contain any image registration artifacts. It has to be noted that automatic image acquisition and registration are in particular included in FEI [10] or ZEISS [11] softwares (several tens of thousands euros). For this example, Steps $\#S1$ and $\#S2$ of Fig. 3 only requires 20 minutes. It depends on the chip size and the chosen SEM parameters.

IV. Hardware Trojans Detection

Once the full chip is reconstructed the idea is to correlate with a reference the 'physical extracted view' obtained with

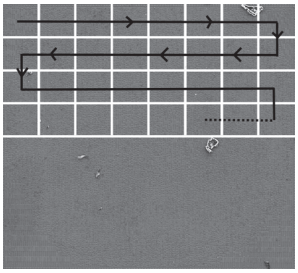


Fig. 6. Full chip reconstructed with a $2,2kX$ magnification

our methodology. It requires either a golden circuit or a design file as reference. The golden circuit is prepared on the same basics and a direct image comparison can be made. Whereas comparing with a design file require either a graphical file, such as a GDSII, or a text file such as a DEF file. The first part needs to modify the GDSII polygon view with the real physical shapes. The second technique identifies cells from the 'physical extracted view', saves their location and correlates each recognized occurrence with the text design file. Thus, we developed some image processing responding to the need.

A. Detection by correlation with a golden circuit

In order to simulate a Hardware Trojans infected circuit, the same chip is imaged twice with different SEM parameters(this can be pointed out in Fig. 7) and on the left image, 4 gates are manually added. Images features might vary between golden model and device under test acquisitions.

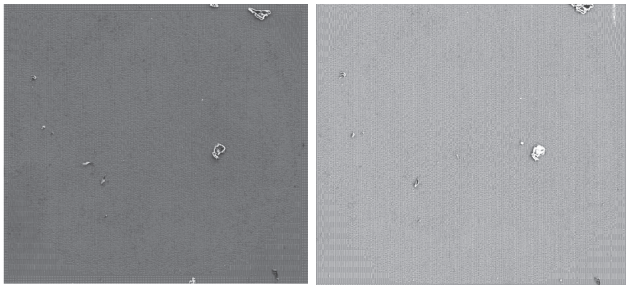


Fig. 7. Same chip, different acquisitions

A direct image subtraction can not be realized to check if any circuit modification has been inserted. The chip images don't have the same orientation, the same scanned area and features. Therefore, some image processing has to be first used in order to register an image to the other one. Afterwards, some image processing tools have to be applied to point differences between both images out. Applying histogram equalization and image subtraction highlights the four additional gates in Fig 8.

Back to a real case insertion, this technique could be performed over two different circuits, a golden circuit and a circuit to test. Mainly due to the preparation type, acquired images under the microscope can show some local differences, such as presence of remaining superior layers at some locations. Moreover, preparing the golden circuit and then each new selected chip may result in some matters to overcome. Indeed, first the application of acids has to be a little bit controlled in time and concentration.

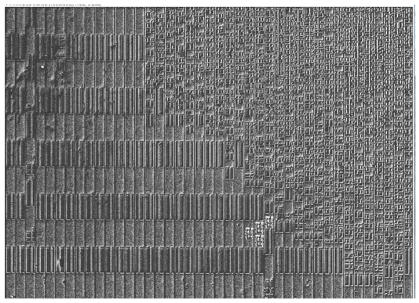


Fig. 8. Detecting Hardware Trojans with golden circuit

B. Detection by correlation with GDSII file

The Golden model circuits may not be available - e.g. if no full trusted manufacturing process accessible, starting from the same image we describe how to compare the chip extracted image reference with a graphical CAD file. We got the layout of a flip-flop standard cell and only kept the n-wells and p-wells implanted zone by layer selection under the CAD tool. It gives the left part of Fig. 9. This reference is then modified to facilitate the correlation with the hardware structure retrieved with our preparation methodology. It gives the pattern on the right. Standard closing morphology image processing and dilation are used.

Fig. 9. GDSII layout and modified GDSII

From our methodology, the operator can obtain the physically extracted shapes seen in Fig. 10. A correlation test is made between the modified GDSII shape and the physically extracted view. If at a specific location, the correlation significantly dropped, a circuit modification is detected.

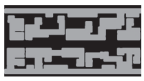
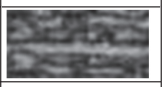
	Localization 1	Localization 2
Modified GDSII		
Physically Extracted		
HTs presence	NO	YES

Fig. 10. Detecting Hardware Trojans with (modified) GDSII

C. Detection by correlation with a text CAD file

A CAD text output file such as a DEF file can be used as a reference as it may be a CAD file easier (less sensitive)

to transmit compared to a GDSII file -it does not contain connection between standard cells. In this type of file, each standard cell name is written as does its localization in μm compared to an origin. The idea is to first class all the instances by standard cells and add each type and size similar standard cell to obtain a number of similar shapes to retrieve over the circuit under test. In Fig. 11, taking a flip-flop cell instance and using libraries based on normalized cross correlation algorithm [13], we obtain localizations of the given flip-flop over the complete circuit. A correlation threshold has to be set. Using a quite high correlation threshold, despite non retrieving all flip-flop cells instances, it makes sure that no false recognitions are in the result. In the end, one can spot the 4 additional gates that are not present within the DEF file at the retrieved location. Thus, a circuit modification has been inserted.

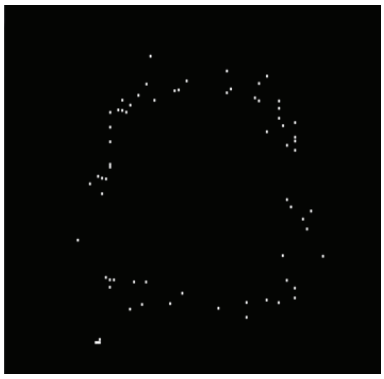


Fig. 11. Recognized flip-flop localizations

The pattern recognition tool has to be enhanced to improve detection rate. The following sum up the final reasoning with an efficient chip preparation and pattern recognition. If the number of gates occurrences present in the DEF file, N_{occ}, differs from the number of recognized occurrences, N_{rec}, a circuit modification has been introduced within the device. Localizations of gates should also correlate if the device under test is trojan-free.

TABLE I. DETECTION EXAMPLE WITH A TEXT CAD FILE

Name	Localization in DEF file in microns	Total number of occurrences	Extracted view	Number of recognized occurrences	Localization of instances recognized (in pixel)
FF1	DX_1 , DY_1 DX_2 , DY_2 DX_n , DY_n	N_{occ}		N_{rec}	RX_1 , RY_1 RX_2 , RY_2 RX_n , RY_n

CONCLUSION

The adaptation of the chain of trust for a given application allows deciding where it is possible to insert a dedicated Hardware Trojan detection step. The results obtained with an invasive approach have shown very good results. The detection rate can reach '100%' with efficient image processing as used in other fields, and it is not affected by the process variations. Small HT of few gates can be caught efficiently even on large circuits which is more complex to achieve with side-channel techniques for example. The main limitation will apply to the sample preparation technique that must be upgrade for smaller technologies. Whereas the observation technique based on SEM imaging is applicable to any CMOS technology node. An other interesting advantage to the developed solution is to use several possible inputs -derived from design database files- that can be substitute to a missing golden model. On the use case studied, the focus was on digital logic HT insertion but an analogue part modification will also be detected in the same way.

PERSPECTIVES

Real case detection

The next steps of our study will be to use reference and infected circuit. This detection process steps that were based on simulated infected circuit will be applied on real infected circuit provided by academic partners. Pertinent image processing will be tested to obtain perfect detection rate.

Non destructive visualization

A strong improvement of our methodology could be to remove the sample preparation step. An infrared camera can be used to see through the Silicon substrate of the chip. So far, it is possible to detect the length of a standard cell position at a given position. However unlike SEM imaging it is not possible by now to get some details information within the standard cell itself. So the next step will be to evaluate how SEM and IR imaging can be coupled to distinguish synthesized logic modification.

Sub active region detection

The idea is to add a step in our proposed methodology in order to be able to distinguish possible HTs present under the transistor's well implementation [14]. $Suguwara\ et\ al$ [15] proposed a method requiring a prepared chip with an access above the chip contact layer. From our side, the idea is to characterize the dopant type from the layer obtained by our fast and low cost preparation. Adding a supplementary step based on a KOH aqueous solution [8] could be an interesting solution in order to have etching profile dependant on the dopant type. Those differences would be visible with a Scanning Electron Microscope.

ACKNOWLEDGEMENT

This collaborative research work has been done within the HOMERE project funded by the French Government (BPI-OSEO) under grant FUI#14. The research work of Franck Courbon has been funded by the ANRT CIFRE funding #2012-2008. We would also like to thank people involved in the design of the test chip, and more generally both first author teams, the Secure Architectures and Systems laboratory and the Gemalto Security Hardware Labs.

REFERENCES

[1] J. Aarestad, D. Acharyya, R. Rad, and J. Plusquellic, "A Survey of Hardware Trojan Taxonomy and Detection," Design Test of Computers, IEEE, January 2010, p. 10-25.

[2] R. Torrance and D. James, "The state-of-the-art in IC reverse engineering," in Cryptographic Hardware and Embedded Systems-CHES 2009. Springer, 2009, p. 363381.

[3] J. Aarestad, D. Acharyya, R. Rad, and J. Plusquellic, "Detecting trojans through leakage current analysis using multiple supply pad s," Information Forensics and Security, IEEE Transactions on, vol. 5, no. 4, pp. 893–904, Dec 2010.

[4] C. Bao, D. Forte, and A. Srivastava, "On application of one-class svm to reverse engineering-based hardware trojan detection," in Quality Electronic Design (ISQED), 2014 15th International Symposium on, March 2014, pp. 47–54.

[5] S. Bhasin, J.-L. Danger, S. Guilley, X. Ngo, and L. Sauvage, "Hardware trojan horses in cryptographic ip cores," in Fault Diagnosis and Tolerance in Cryptography (FDTC), 2013 Workshop on, Aug 2013, pp. 15–29.

[6] R. Chakraborty, S. Narasimhan, and S. Bhunia, "Hardware trojan: Threats and emerging solutions," in High Level Design Validation and Test Workshop, 2009. HLDVT 2009. IEEE International, Nov 2009, pp. 166–171.

[7] J. Fournier, J.-B. Rigaud, S. Bouquet, B. Robisson, A. Tria, J.-M. Dutertre, and M. Agoyan, "Design and characterisation of an aes chip embedding countermeasures," IJIEI, vol. 1, no. 3/4, pp. 328–347, 2011.

[8] F. Beck, Integrated circuit failure analysis : a guide to preparation techniques, 1st ed. New York, NY, USA: Cambridge University Press, 1997.

[9] "http://hugin.sourceforge.net/."

[10] "http://www.fei.com."

[11] "http://www.zeiss.com."

[12] A. Poonawala, "Mask design for single and double exposure optical microlithography: An inverse imaging approach."

[13] J. Lewis, "Fast normalized cross-correlation," Vision interface, vol. 10, no. 1, pp. 120–123, 1995.

[14] G. T. Becker, F. Regazzoni, C. Paar, and W. P. Burleson, "Stealthy dopant-level hardware trojans: extended version," J. Cryptographic Engineering, vol. 4, no. 1, pp. 19–31, 2014. [Online]. Available: http://dx.doi.org/10.1007/s13389-013-0068-0

[15] T. Sugawara, D. Suzuki, R. Fujii, S. Tawa, R. Hori, M. Shiozaki, and T. Fujino, "Reversing stealthy dopant-level circuits," in Cryptographic Hardware and Embedded Systems CHES 2014, ser. Lecture Notes in Computer Science, L. Batina and M. Robshaw, Eds. Springer Berlin Heidelberg, 2014, vol. 8731, pp. 112–126.