Shielding and Securing Integrated Circuits with Sensors

Davood Shahrjerdi†, Jeyavijayan Rajendran†, Siddharth Garg†, Farinaz Koushanfar‡, and Ramesh Karri†
†Electrical and Computer Engineering Department, New York University, Brooklyn, NY, USA, 11201
‡Electrical and Computer Engineering Department, Rice University, Houston, TX, USA,77005
Email: davood@nyu.edu, jv.ece@nyu.edu, siddharth.j.garg@gmail.com, farinaz@rice.edu, rkarri@nyu.edu

***Abstract*—An integrated circuit (IC) Supply Chain Hardware Integrity for Electronics Defense (SHIELD) is envisioned to enable advanced supply chain hardware authentication and tracing capabilities. The suggested SHIELD is expected to be a ultra-lower power, minuscule electronic component that is physically attached to the host IC. This paper focuses on two important adversarial acts on SHIELD: physical reverse engineering and physical side-channel analysis. These attacks can be launched through mechanical or optical means and they can reveal and/or modify the confidential on-chip data or enable reverse-engineering of the design. For detection of these attacks and subsequent erasing of the sensitive data, sensors, erasure devices, and the relevant control circuitry need to be added to the SHIELD. We describe the device-level operation of the optical (photodetectors) and mechanical (nano- or micro-electromechanical switches) sensors and how they can be integrated within an IC to detect physical attacks. The operation of these micro/nano-scale sensors is unreliable due to environmental, operational, and structural fluctuations and noise. We outline system-level approaches to design a reliable countermeasure against physical attacks using unreliable sensors.**

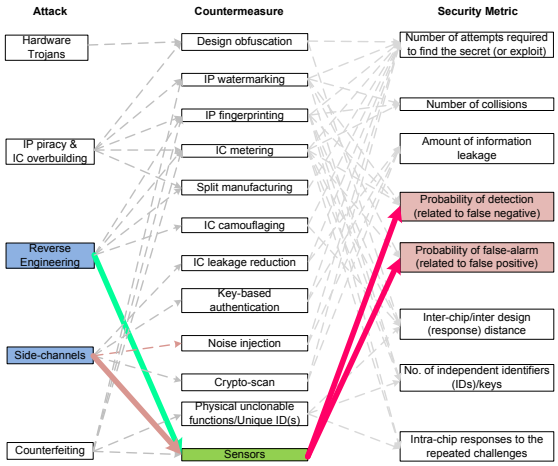


Fig. 1. Five classes of hardware attacks in [2] (left column), suggested countermeasures (middle column) and evaluation metrics (right column.) In this work, we discuss the hardware attacks, countermeasures, and metrics that are shown as colored (shaded) blocks and solid edges in the context of SHIELD.

I. Introduction

A. Motivation

The Defense Advanced Research Project Agency (DARPA) Supply Chain Hardware Integrity for Electronics Defense (SHIELD) program proposes to build trust into the Integrated Circuit (IC) supply chain of Design→ Manufacturing → Testing → Integration → Packaging → Distribution [1]. SHIELD is envisioned to be the hardware root-of-trust which when packaged into any IC equips that IC with a permanent, unique, and unclonable identifier that is infeasible to alter or copy. The SHIELD root-of-trust can verify the provenance of an IC as it goes through the IC supply chain and can continuously authenticate the sensitive ICs in a system throughout its lifetime.

The SHIELD root-of-trust is expected to be miniscule and have no on-board power source. This is realized by powering it from outside and by using a simple authentication protocol. [1]. SHIELD is expected to offer strong security by supporting a 256-bit cryptographicon-SHIELD encryption engine. Further, any sensitive material such as the keys and unique identifiers should be stored within the SHIELD in a way that is prohibitively expensive to reverse engineer. Moreover, SHIELD shall be resilient to non-invasive attacks and self destruct upon invasive attacks. Finally, SHIELD shall be electronically robust to last a lifetime.

B. Security of SHIELD

Threats, countermeasures, and metrics for ICs in general have been systematized in [2]. The systematization, shown in Figure 1, summarizes the attacks (left column), the applicable countermeasures (the middle column), and the metrics (the right column), respectively. The SHIELD designers can use this framework to clearly state the targeted threats, develop countermeasures, and use metrics to evaluate security. As shown in Figure 1, the threat models relevant to SHIELD include [2]: (1) hardware trojans, (2) IC piracy and intellectual property (IP) overbuilding, (3) reverse engineering, (4) side-channel analysis, and (5) counterfeiting.

The confidential information being protected within the SHIELD can be cryptographic keys, unique identifiers, or the design itself. Unique identifiers can be generated using physical unclonable functions (PUFs). PUFs use process variations to generate unique identification codes per IC [3].

C. Thwarting physical attacks on SHIELD with sensors

This paper focuses on physical reverse engineering and physical side-channel attacks on SHIELD. These attacks can reveal and/or modify confidential data and reveal the IP. These attacks can be launched through mechanical or optical means and are described in Section II.

The defense mechanism against these attacks should detect the attacks and subsequently destroy the confidential information on the IC. This therefore calls for the integration of sensors, erasure devices, and the associated circuits on CMOS ICs. Photodetectors can detect optical attacks that use light source [4], [5], [6], and Nano- or Micro-electromechanical switches (NEMS/MEMS) can detect mechanical attacks such as delayering or inserting microprobes [7], [8]. The erasure mechanism is determined by the properties of CMOS devices to be destroyed. For example, nano-fluidic chambers can be used for chemical destruction of the devices [9]. Section III

describes the device-level operation of these sensors and how they can be integrated within an IC to thwart physical attacks.

The operation of nano-scale sensors is unreliable due to process variations, noise, and environmental effects. Such unreliable operation may destroy the IC in the absence of an attack or may not destroy the IC during an attack. Section IV describes system-level approaches to design a reliable countermeasure against physical attacks using unreliable sensors.

D. Related work

According to U.S. Federal Information Processing Standard (FIPS) 140-2, a cryptographic module with the highest level of security should have the capability to erase confidential data on detecting an unauthorized physical access to the module [10]. In order to achieve this standard, [11] states that a module should have tamper resistant, tamper evident, and tamper response capabilities. While different approaches for achieving tamper resistant and tamper evident capabilities are discussed in that paper, it did not address how to achieve the tamper response capability. IBM 4758 meets the FIPS standard [12]. However, this IC can erase confidential information only when there is alteration in voltage, temperature or radiation.

In the context of using sensors for security, MEMS devices have been used to generate seeds for random number generators [13]. The general classes of attacks, countermeasures and metrics that are applicable to SHIELD have been outlined in [14].

II. Physical attacks on SHIELD

Reverse engineering and physical side-channel analysis can use mechanical and optical modalities. Physical attacks on SHIELD can be either invasive or semi-invasive.

A. Invasive attacks

One can reverse engineer an IC by de-packaging, delayering, imaging the layers, and extracting the netlist [15], [16]. Shrinking device dimensions have not hampered reverse engineering. For instance, Intel's 22nm Xeon processor has been successfully reverse engineered [17]. Picosecond imaging-based circuit analysis (PICA) can invasively monitor the state of the signals by monitoring the luminescence of hot-carriers [18]. Although PICA was used for failure analysis of ICs, one can use PICA to recover the secrets stored in the IC. An attacker can insert probes into the IC and monitor the signals while it is functioning. For example, one can probe the bus signals in an IC and readout the contents of the memory [19]. Besides such passive monitoring of the internal signals, one can use focused ion beams (FIB) to alter the state of the signals in a functioning IC. For instance, one can cut off the trigger signal from alerting the CPU of an attack [19]. A practical FIB attack has been demonstrated by probing the backside of an IC [20].

B. Semi-invasive attacks

In this class of attacks, the attacker does not delayer the IC. Instead, he disrupts its functionality using light and other sources of radiation [21]. The target IC is mounted under a radiation (laser) source. The injected light source is magnified to an extent where its photon energy is greater than the bandgap energy of the targeted CMOS devices. The energized photon, on hitting a CMOS device, will ionize the surrounding semiconductor region creating a transient bit-flip, for example in a SRAM memory cell. An optical radiation attack on a PIC microcontroller using a $30 light source has been able to modify the contents of the SRAM [21]. Furthermore, such attacks have been shown to be successful using radiation sources ranging from flashbulbs to lasers [21]. Since the amount of photon energy required to ionize the semiconductor is directly proportional to its bandgap energy, emerging technology nodes are even more sensitive to optical radiation.

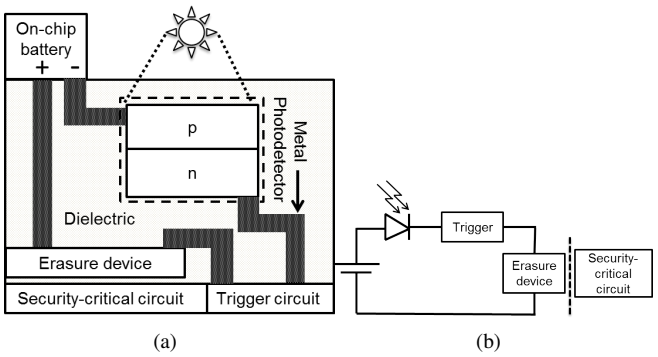


Fig. 2. (a) Side-view of an IC using photodetectors to detect optical attacks [4]. (b) Equivalent circuit diagram.

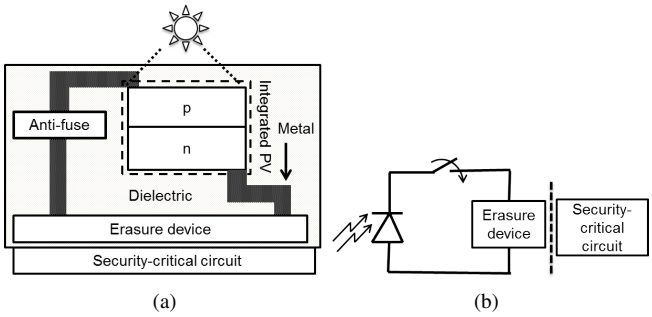
III. Securing ICs with Sensors

The types of adverse attempts to tamper with ICs appear to be diverse and distinct. Since most of the invasive and semi-invasive attacks use mechanical force [15], [17], [20] or light source [21], one can classify the underlying attack mechanisms into two general categories: mechanical and optical. This simple classification, from the defense standpoint, promotes the design of overarching security mechanisms that will potentially provide higher level of security against various attacks of similar nature. Some of the design considerations of such systems include robustness, durability, cost-effectiveness, and compatibility with CMOS integration processes.

A. Key idea

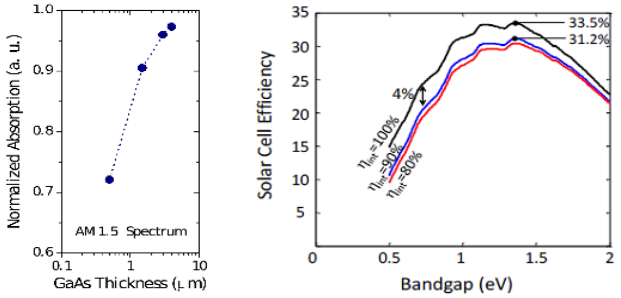
One can embed sensors within an IC to detect attacks, specifically mechanical and optical attacks [4]. Figure 2(a) schematically illustrates a hardware configuration that can detect physical attacks and destroy the confidential information. This IC consists of sensors, trigger circuitry, battery and erasure devices. In case of optical attacks, the sensor, i.e. photodetector, triggers the erasure device. This erasure device erases the confidential information stored in the associated circuitry. Figure 2(b) shows the equivalent circuit diagram.

From the device standpoint, the ability to integrate various sensors such as PV devices [4], [5], [6] and MEMS-based sensors [7], [8] offers flexibility in designing an IC to thwart the considered attacks. However, the reliance on the on-chip battery as a single energy source for powering up the defense mechanism could undermine the robustness of the security scheme. For example, an attacker can turn the battery off



(a) (b)

Fig. 3. (a) Side-view of an IC using energy-harvesting PV cells to detect optical attacks [4]. (b) Equivalent circuit diagram.



(a) (b)

Fig. 4. (a) Calculated normalized absorption in GaAs as a function of thickness, illustrating its strong absorption properties. (b) Theoretical calculation of the conversion efficiency as a function of the absorber bandgap [22], illustrating the decline in efficiency as the bandgap is reduced.

while launching the attacks, rendering any battery-powered defense useless. Furthermore, although batteries are a reliable source of energy, their longevity, size, and form factor could limit their effectiveness in applications such as distributed sensor networks and wearable electronics. In these applications, the relatively small size, arbitrary shape of the devices and conceivably the remote location of the electronic system can tremendously curtail the use of batteries.

To overcome these problems, one can use energy harvesting devices to replace or supplement on-chip batteries. One can use the energy harvester as a sensor to detect the attack and to simultaneously generate power for setting off the erasure device. In this section we review the potential and technological challenges of two prominent energy-harvesting schemes – (i) photo-voltaic (PV) devices and (ii) Nano- and micro-electromechanical systems (NEMS- and MEMS-) based piezoelectric cantilevers – and describe how they can be used to thwart physical attacks.

B. Thwarting optical attacks with photo-voltaic (PV) devices [4]

One can use PV cells in an IC to thwart imaging attacks (both invasive [17] and semi-invasive [21]). These PV cells will detect the light/radiation used for imaging the device and will trigger the erasure device, which then destroys the confidential information.

1) PV device operation: PV devices absorb the incident light/radiation and generate photo-voltaic energy. The choice of an absorber material is an important design consideration in order to result in the favorable response of the PV device, i.e. strong absorption of the light and its subsequent efficient conversion for producing adequate energy to operate the erasure device. Furthermore, it is notable that some of the imaging techniques used for reverse engineering ICs are largely based on long wavelength photons. This therefore calls for the integration of efficient PV devices, in which the bandgap energy of the absorber layer is direct and small. As a result, III-V materials owing to their unique electronic and optical properties appear to be the most viable candidates.

2) Structure and circuit operation: Figure 3(a) conceptually illustrates an IC that employs integrated PV devices. The PV device is connected to the erasure device through an anti-fuse. The erasure device is coupled with the security-critical CMOS circuit, which stores the confidential information. Figure 3(b) shows the equivalent circuit diagram.

During fabrication and testing, the anti-fuse is disabled. This prevents the PV device from accidentally triggering the erasure device. This anti-fuse is enabled only when the IC is deployed in the field. When an attacker tries to image the IC, the PV device will detect the light and activate the erasure device. The erasure device then destroys the underlying security-critical circuit, preventing an attacker from stealing the confidential information.

3) Challenges: Figure 4(a) illustrates the theoretical calculations for the normalized absorption of gallium arsenide (GaAs) absorber layer versus its thickness, where GaAs was chosen as a typical example of a direct gap material for making high-efficiency PV devices. It is important to highlight that the reasonably high absorption level of the spectrum by direct gap materials at very small thicknesses potentially enables the integration of high-efficiency cells in the CMOS stack. However, the use of narrow gap materials, for enhancing the sensitivity of the PV device at long wavelengths, degrades the conversion efficiency. This primarily stems from the reduction in the open circuit voltage of the device. Figure 4(b) illustrates the theoretical calculations of the conversion efficiency under one-sun illumination as a function of the bandgap of the absorber material [22]. Another challenge associated with the integration of the III-V PV devices is the epitaxial growth of these materials on silicon substrates. However, some promising attempts such as aspect-ratio-trapping [23] are viable options for the integration of III-V PV devices in CMOS.

C. Thwarting mechanical attacks using NEMS/MEMS cantilevers

Layer-by-layer mechanical deconstruction of an IC is another approach to glean electrical and structural information about the IC [17]. Deconstruction of circuits may involve mechanical polishing or grinding of the IC. Therefore, harnessing the low-frequency mechanical force that is exerted by the grinding disk during the delayering process appears to be a viable option for triggering and powering the erasure device.

1) NEMS/MEMS cantilever device operation: NEMS/MEMS cantilevers convert an external mechanical vibration into electrical energy through the piezoelectric effect, i.e. the occurrence of electric dipole moments. These devices are typically modeled as a mass-spring-damper

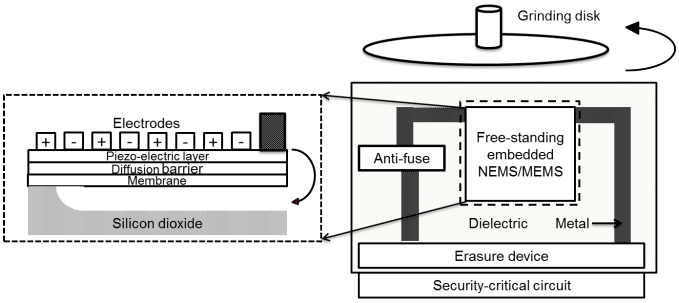


Fig. 5. Side-view of an IC using NEMS/MEMS-based devices to detect mechanical attacks. The inset shows the structure of a NEMS/MEMS cantilever.

system. The maximum power in these devices is achieved when the frequency of the external mechanical vibration is equal to the natural frequency ω_n of the system [7], [8].

2) Structure and circuit operation: Figure 5 conceptually illustrates the schematic of an IC using a NEMS/MEMS-based energy harvester to thwart delayering attacks. The inset shows an exemplary structure of a typical vibration-based energy harvester. The NEMS/MEMS device is connected to the erasure device through an anti-fuse, which is enabled after fabrication. The erasure device is located on top of the security-critical CMOS circuit, which needs to be protected from reverse engineering.

3) Challenges: NEMS/MEMS-based cantilevers have a relatively small output power density [24]. Thus, the power generated by NEMS/MEMS devices might not be sufficient for the erasure device to operate. Furthemore, these devices have a narrow operating bandwidth [24]. Consequently, the devices may be less sensitive to the mechanical force exerted by the grinding disk. Innovative approaches for realizing more efficient piezoelectric materials and beam structures are required in order to achieve higher power densities and wider bandwidth.

IV. System-level challenges for shielding ICs

An important aspect of a sensor-based solution to detect and react to semi-invasive and invasive attacks is a mechanism to determine if the signal picked up by the sensor is indeed the consequence of malicious activity or the regular operation of an IC. This is particularly relevant given that several defense mechanisms, for instance nano-fludic chamber [9] and laser auto-destruct system [25] can, in fact, permanently disable the chip when an attack is detected.

Certain sensors that detect an invasive attack might also be sensitive to signals generated during regular IC operation. For instance, the NEMS/MEMS cantilever sensor in Section III is intended to detect mechanical stress on metal wires as a consequence of grinding during delayering an IC. However, metal wires also undergo thermo-mechanical stress because of the difference in the coefficient of thermal expansion of metal and the surrounding insulator. This is exacerbated by increasing chip temperature of high-performance ICs and fast temporal thermal variations.

In addition, if an attacker is aware that a sensor has been deployed to detect a certain attack modality, the attacker might try to circumvent the sensor by, for instance, applying the minimum possible mechanical force on the metal wires, or using a low intensity light source for a longer duration of time (for the sensors described in Section III). This sets up a trade-off — making the sensor more sensitive is more likely to detect malicious attacks, but also more likely to trigger due to noise during regular operation.

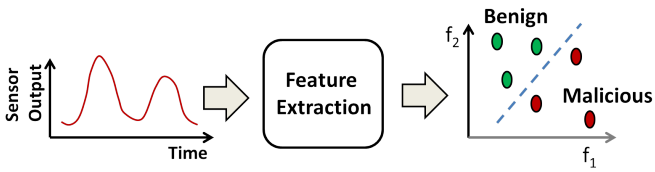


Fig. 6. Structure of a basic classification module including a feature extraction block. The input is a time domain signal that is output by the sensor.

To address these issues, the photodetector in Figure 2(a) and NEMS/MEMS cantilever in Figure 5 will likely need to be augmented with a classification algorithm implemented as custom logic to minimize power dissipation and area.

A. Structure and operation of a classifier

The classifier takes as input a time domain analog signal from the sensor and converts it into a set of relevant features. For instance, for the cantilever sensor, the relevant features might be the magnitude of the Fourier coefficients for different frequency values. The feature extraction block outputs one or more binary, integer or real valued features. Now, based on training data that includes both potential malicious signals and those from regular IC operation, one can design a classifier to distinguish between these two classes. The classifier shown in Figure 6 is, for instance, a simple linear classifier.

Once a sensor output signal has been classified as malicious, the defense mechanism can be triggered. Of course, it is critical to ensure that the feature extraction and classification circuits have a low footprint in terms of area and power consumption.

B. Challenges

While feature extraction and classification might be essential for certain types of sensors, several research challenges need to be addressed in deploying such solutions. One needs to determine the types of features and classifiers which are the most accurate for a given defense mechanism. Furthermore, the feature extractors and classifiers can be implemented using digital or analog circuits. Hence, one has to chose an implementation of a classifier depending up on power, performance, area, and accuracy trade-offs. In addition, one has to train these classifiers with data from normal and malicious operations.

V. Conclusion

To thwart different types of attacks, one needs to embed different type of sensors – photodetectors to detect optical attacks that use light source, and NEMS/MEMS cantilevers to detect attacks that exert mechanical force. However, using different types of sensors has two issues. First, even though the individual sensors are CMOS-compatible, the manufacturing process of one type of sensor may not be compatible with

that of another type of sensor. Therefore, one needs to design and use sensors whose manufacturing processes are mutually compatible. Second, using multiple types of sensors may lead to placement issues. Consider the case where a certain type of sensors are placed within the IC such that its ability to detect an attack is maximized. However, this placement scheme may inhibit the detection ability of another type of sensor. Hence, one needs to place these sensors in an optimized way. To conclude, though shielding ICs with sensors for security is a promising approach, it presents several device, circuit, system, security, and computer-aided design challenges.

REFERENCES

[1] Defense Advanced Research Projects Agency (DARPA), Microsystems Technology Office/MTO Broad Agency Announcement, "Supply chain hardware integrity for electronics defense (SHIELD)," 2014.

[2] M. Rostami, F. Koushanfar, and R. Karri, "A primer on hardware security: Threat models, metrics, and remedies," *Proceedings of the IEEE*, vol. 102, no. 8, pp. 1283–1295, Aug 2014.

[3] B. Gassend, D. Clarke, M. Van Dijk, and S. Devadas, "Silicon physical random functions," *ACM Conference on Computer and Communications Security*, pp. 148–160, 2002.

[4] J. Chu, G. Fritz, H. Hovel, Y. Kim, D. Pfeiffer, and K. Rodbell, "Integrated circuit tamper detection and response," http://www.google.com/patents/US20140103286, 2014, US Patent App. 13/654,078.

[5] L. Hsu, D. Kruger, J. Mason, and R. Oldrey, "Implementing tamper resistant integrated circuit chips," https://www.google.com/patents/US8089285, 2012, US Patent 8,089,285.

[6] W. Abadeer, "CMOS imaging sensor having a third FET device with a gate terminal coupled to a second diffusion region of a first FET device and a first terminal coupled to a row select signal," https://www.google.com/patents/US7692130, 2010, US Patent 7,692,130.

[7] E. Fuentes-Fernandez, P. Shah, W. Mechtaly-Debray, and B. Gnade, "Mems-based cantilever energy harvester," https://www.google.com/patents/US20140087509, 2014, US Patent App. 14/094,590.

[8] W. Choi, Y. Jeon, J.-H. Jeong, R. Sood, and S. Kim, "Energy harvesting MEMS device based on thin film piezoelectric cantilevers," *Journal of Electroceramics*, vol. 17, no. 2-4, 2006.

[9] R. Das, V. Markovich, J. McNamara, and M. Poliks, "Anti-tamper microchip package based on thermal nanofluids or fluids," http://www.google.com.ar/patents/US8288857, 2012, US Patent 8,288,857.

[10] Federal Information Processing Standard Publication (FIPS PUB 140-2), "Security requirements for cryptographic modules," http://csrc.nist.gov/publications/fips/fips140-2/fips1402.pdf, 2002.

[11] K. Lemke, "Embedded Security: Physical Protection against Tampering Attacks," http://dx.doi.org/10.1007/3-540-28428-1_12, pp. 207–217, 2006.

[12] J. G. Dyer, M. Lindemann, R. Perez, R. Sailer, L. van Doorn, S. W. Smith, and S. Weingart, "Building the IBM 4758 Secure Coprocessor," *Computer*, vol. 34, no. 10, pp. 57–66, 2001.

[13] J. English, D. Coe, R. Gaede, D. Hyde, and J. Kulick, "MEMS-Assisted Cryptography for CPI Protection," *IEEE Security and Privacy*, vol. 5, no. 4, pp. 14–21, 2007.

[14] F. Koushanfar and R. Karri, "Can the shield protect our integrated circuits?" *Proceedings of IEEE MWSCAS*, 2014.

[15] Chipworks, "Reverse engineering software," http://www.chipworks.com/en/technical-competitive-analysis/resources/reverse-engineering-software.

[16] Degate, http://www.degate.org/documentation/.

[17] R. Torrance and D. James, "The State-of-the-Art in IC Reverse Engineering," *Cryptographic Hardware and Embedded Systems*, pp. 363–381, 2009.

[18] J. C. Tsang, J. A. Kash, and D. P. Vallett, "Picosecond Imaging Circuit Analysis," *IBM Journal of Research and Development*, vol. 44, no. 4, pp. 583–603, 2000.

[19] R. J. Anderson, "Security Engineering. In A Guide to Building Dependable Distributed Systems," *Wiley*, vol. 44, 2010.

[20] C. Helfmeier, D. Nedospasov, C. Tarnovsky, J. S. Krissler, C. Boit, and J.-P. Seifert, "Breaking and Entering Through the Silicon," *Proceedings of the ACM Conference on Computer and Communications Security*, pp. 733–744, 2013.

[21] S. Skorobogatov and R. Anderson, "Optical fault induction attacks," *Cryptographic Hardware and Embedded Systems*, pp. 31–48, 2002.

[22] O. D. Miller, E. Yablonovitch, and S. R. Kurtz, "Strong Internal and External Luminescence as Solar Cells Approach the Shockley–x Queisser Limit," *IEEE Journal of Photovoltaics*, vol. 2, no. 3, pp. 303–311, 2012.

[23] J.-S. Fiorenzaa, James G.and Parkb, J. Hydrickb, J. Lib, J. Lic, M. Curtinc, M. Carrollc, and A. Lochtefeldc, "Aspect Ratio Trapping: A Unique Technology for Integrating Ge and III-Vs with Silicon CMOS," *ECS Transactions*, vol. 33, pp. 963–976, 2010.

[24] S.-G. Kim, S. Priya, and I. Kanno, "Piezoelectric MEMS for Energy Harvesting," *MRS Bulletin*, vol. 37, pp. 1039–1050, 2012.

[25] G. J. Knowles and L. Quarrie, "Integrated laser auto-destruct system for electronic components," https://www.google.com.ar/patents/US20090070887, 2008, US Patent App. 12/205,693.