Chapter 6
Design for Hardware Trust: Ring Oscillator Network

In this chapter, a hardware Trojan detection method is presented, which is performed by combining measurements from an on-chip structure with external dynamic current measurements [1]. This method monitors power fluctuations and differentiates fluctuations due to hardware Trojans from fluctuations due to measurement noise and process variations. This method considers Trojan impact on the power consumption of neighboring cells and the entire IC. A number of ring oscillators (ROs) acting as *power monitors* are inserted into a circuit, forming a ring oscillator network (RON). Each row of the circuit under authentication contains at least one inverter of an RO in the RON. Thus any malicious inclusions in each row would be captured by one of these ring oscillators. Off-chip test equipment will measure the transient current of the IC, which will be combined with the ROs' cycle counts to generate a power signature for the entire IC. The signature of the CUT is then compared against the Trojan-free signatures.

6.1 Analyzing Impact of Power Supply Noise on Ring Oscillators

Two simple five-stage ring oscillators are shown in Fig. 6.1: the ring oscillator in Fig. 6.1a consists of inverters and the ring oscillator in Fig. 6.1b is composed of NAND gates. The second ring oscillator has a higher sensitivity to supply noise since one of its inputs is connected to the power supply but this ring oscillator offers larger area overhead. The first ring oscillator is used as a power monitor since its behavior can easily be described analytically. The frequency of this ring oscillator is determined by the total delay of all the inverters in the presence of supply voltage variations and process variations. Assume that each stage in the ring oscillator provides a delay of t_d. The delay of the n-stage ring oscillator, then, is approximately $2 \times n \times t_d$ and the oscillation frequency will be $f = 1/(2 \times n \times t_d)$. The delay of each inverter is given by

M. Tehranipoor et al., *Integrated Circuit Authentication: Hardware Trojans and Counterfeit Detection*, DOI 10.1007/978-3-319-00816-5_6,

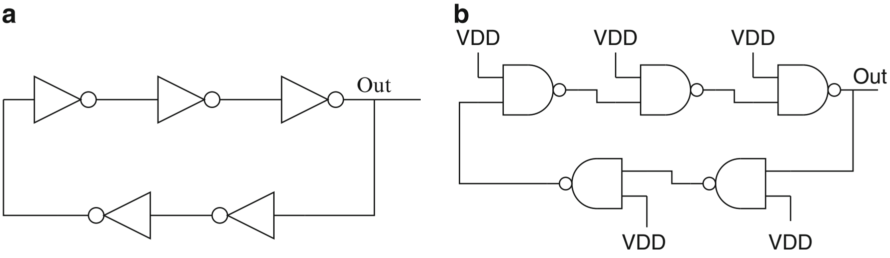


Fig. 6.1 Five-stage ring oscillators

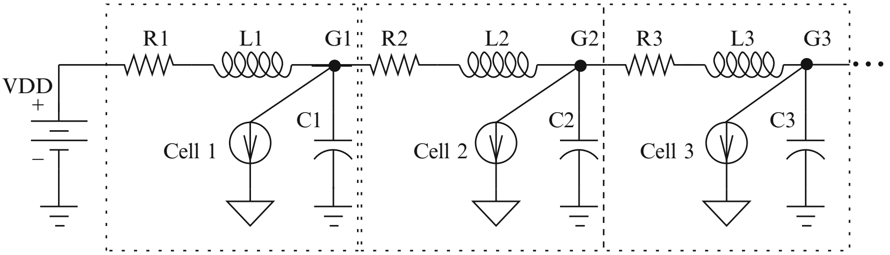


Fig. 6.2 The RLC model of a simple power line in a power distribution network

$$t_d = 0.52 \times \frac{C_L V_{dd}}{(W/L)k' V_{DSAT}(V_{dd} - V_{th} - V_{DSAT}/2)} \tag{6.1}$$

where C_L is load capacitance, $k^{'}$ represents transconductance, and W/L denotes the transistor channel width to channel length ratio. V_{DSAT}, V_{th}, and V_{dd} represent saturation voltage, threshold voltage, and power supply voltage, respectively [2].

Power supply noise (also known as voltage drop) increases the delays of logic gates in an IC. Thus, a change in the supply voltage of any inverter in a ring oscillator impacts the delay of all associated gates, and therefore impacts the oscillation frequency.

Concerning today's tightly designed power supply distribution networks, transitions in one gate can impact the power supply of other gates within close proximity [3]. Figure 6.2 shows a simple power line model in which VDD line supplies one row in the standard cell design. The indicated VDD line represents the point where a via connects the power rail to the upper metal layer in a power distribution network. Nodes G1, G2, and G3 connect to adjacent cells represented as current sources in Cell 1, Cell 2, and Cell 3. Here, for the sake of simplicity, the power via is assumed to have zero impedance and each interconnect is modeled by a resistance, inductance, and capacitance (RLC) network. The contribution of each current source to the overall noise is described in (6.2) where $V1$, $V2$, and $V3$ (voltage at nodes G1, G2, and G3) are the power supply noise spectrum, $Vii = Z_{ii} \times I_{ii} (i = 1, 2, 3)$ (Z_{ii} is the impedance of node i and I_{ii} is the current) is the power noise,

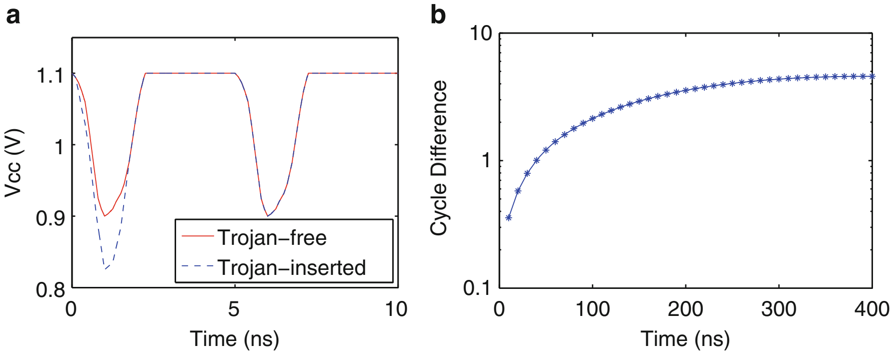


Fig. 6.3 (**a**) Power supply variations for Trojan-free and Trojan-inserted circuits; (**b**) Cycle difference caused by Trojan gates' switching

$\rho_{ij}(i, j = 1, 2, 3)$ is the voltage division coefficient, and ω is the frequency of the circuit. As can be seen from the equation, $V1$, $V2$, and $V3$ are related to the power noise of neighboring gates, demonstrating that a transitioning gate has an effect on neighboring gates connected to the same VDD line.

$$\begin{aligned} V1 &= V11 + \rho_{21}(\omega) \times V22 + \rho_{31}(\omega) \times V33 \\ V2 &= \rho_{12}(\omega) \times V11 + V22 + \rho_{32}(\omega) \times V33 \\ V3 &= \rho_{13}(\omega) \times V11 + \rho_{23}(\omega) \times V22 + V33 \end{aligned} \tag{6.2}$$

Therefore, with the same input patterns, the power supply noise affecting the Trojan-free IC and Trojan-inserted IC will differ due to the switching gates within a Trojan. In order to verify the impact of the Trojan on the frequency of the ring oscillator, a 5-stage ring oscillator (shown in Fig. 6.1a) is implemented in the 90 nm technology for simulation.

In Fig. 6.3a, the dashed line denotes the dynamic power in the presence of a Trojan and the solid line denotes the Trojan-free power (assuming $VDD = 1.2\,\text{V}$). The two supply voltages only differ during the first 2 ns. These two power waveforms are applied to the ring oscillator for 400 ns. Figure 6.3b shows the cycle count difference between the Trojan-free and Trojan-inserted ICs. At time 0, the two ring oscillators denoting *with* and *without* an inserted Trojan have the same period. However, due to the Trojan-induced power supply noise, the cycle count difference grows steadily as the measurement duration increases.

A Trojan, composed of 20 combinational gates, is also simulated to demonstrate its effect on the frequency of a 5-stage ring oscillator at 25°C. The simulation time is 10 µs. Figure 6.4a shows the locations of the ring oscillator and the Trojan: the ring oscillator is placed at the left corner of a standard cell row, while the Trojan is located at some position between locations *1* and *11*. There is one flip-flop (FF) between every two possible Trojan locations. CC_f denotes the cycle count of the

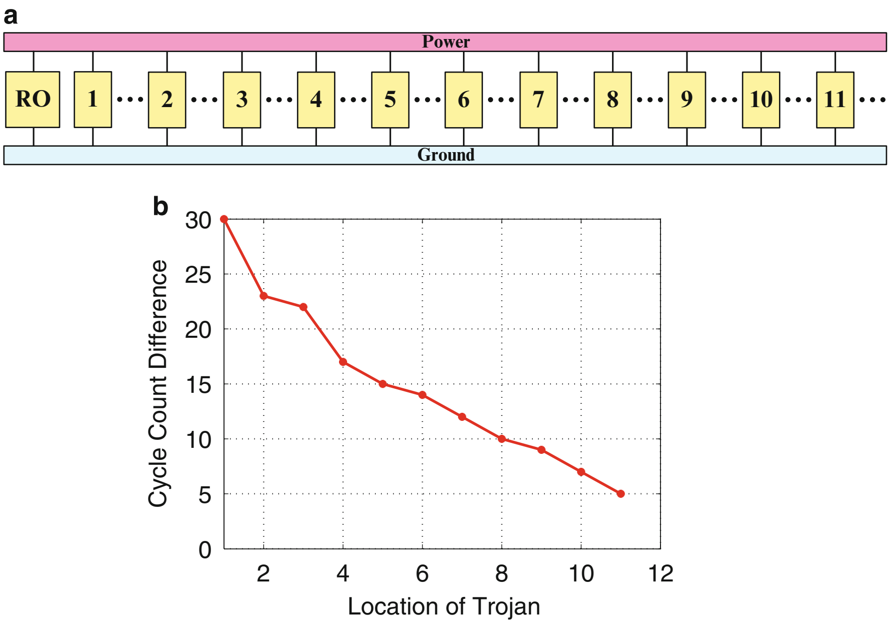


Fig. 6.4 (**a**) RO and Trojan's location; (**b**) Cycle count difference caused by Trojan gates-induced voltage drop

ring oscillator in a Trojan-free IC, CC_t denotes the cycle count of the ring oscillator in an IC with a Trojan, and $\Delta CC_{ft} = CC_f - CC_t$. Figure 6.4b illustrates the relationship between ΔCC_{ft} and the Trojan's location. As can be seen in the figure, the Trojan gates' switching will reduce the ring oscillator's frequency by an amount related to the distance between the Trojan and the ring oscillator. The farther the Trojan is placed from the ring oscillator, the less impact it has on the ring oscillator's frequency.

6.2 The Relationship Between RO Frequency and Localized and Total Dynamic Current

The delay of each inverter in the ring oscillator can also be expressed as $t_d = k_g/I_g$ where k_g is a gate-dependent constant, and I_g is the dynamic current of the inverter [6]. Based on the Alpha-Power Model mentioned in [7], the dynamic current of a switching gate is

$$I = \mu_g \times (V_{dd} - V_{th})^{\alpha} \tag{6.3}$$

where α is the velocity saturation index. Thus the frequency of the n-stage ring oscillator can be expressed as:

$$f = \frac{\mu_g \times (V_{dd} - V_{th})^{\alpha}}{2n \times k_g} \tag{6.4}$$

In the presence of a Trojan, the ring oscillator frequency is modeled by (6.5) rather than (6.4) where the voltage-drop ΔV_t represents the Trojan's contribution to the ring oscillator frequency. As the equation shows, the frequency of the ring oscillator f_t is more sensitive to the voltage-drop ΔV_t when the stage of the ring oscillator n is smaller. However, if n is too small, the frequency of the ring oscillator will be too high to be measured by an on-chip counter. Using (i) an operating frequency of f=1 GHz, (ii) $V_{dd} = 1.2\,\text{V}$, and (iii) Synopsys 90nm technology in a Nanosim simulation, a 5-stage RO would be the smallest allowable RO. Thus, 5-stage ring oscillators will be used.

$$f_t = \frac{\mu_g \times (V_{dd} - \Delta V_t - V_{th})^{\alpha}}{2n \times k_g} \tag{6.5}$$

The dynamic current of the entire Trojan-free chip is:

$$I_{total} = \sum_{i=0}^{i=N} \lambda_i \times N \times \mu_g \times (V_{dd} - V_{th})^{\alpha} \tag{6.6}$$

where N is the total number of switching gates in the IC, and λ_i denotes the gate-dependent constant of the i_{th} gate. The constant, λ_i, depends only on the type of gate specified, not the particular instance of such a gate. The relationship between the frequency of the n-stage ring oscillator embedded into the chip and the dynamic current of entire chip will be:

$$\frac{I_{total}}{f} = \sum_{i=0}^{i=N} \lambda_i \times N \times 2n \times k_g \tag{6.7}$$

For ICs with n_t Trojan gates inserted, (6.7) becomes:

$$\frac{I_{total,t}}{f_t} \approx \sum_{i=0}^{i=N+n_t} \lambda_i \times (N + n_t) \times 2n \times k_g (1 + \alpha \times \frac{\Delta V_t}{V_{dd} - \Delta V_t - V_{th}}) \tag{6.8}$$

when $\Delta V_t << V_{dd} - \Delta V_t - V_{th}$ [9]. By comparing (6.8) with (6.5), it can be concluded that combining ring oscillator frequency measurements with current measurements will achieve greater sensitivity to Trojans than either measurement alone.

The above analysis is based on ring oscillators made with standard threshold voltage (SVT) transistors. However, ring oscillators with high threshold voltage (HVT) transistors are more sensitive to power supply noise, as shown by the simulation results in Fig. 6.5, performed using the same technology with a 5-stage

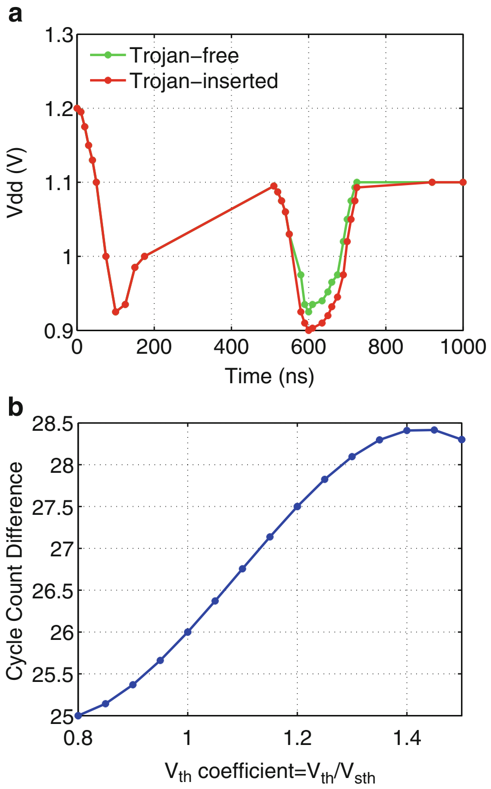


Fig. 6.5 (**a**) Power supply variations for Trojan-free and Trojan-inserted circuits; (**b**) Cycle count difference increases as threshold voltage increases

ring oscillator. The green line in Fig. 6.5a denotes the power supply voltage of Trojan-free ICs during the 1,000 ns simulation period while the red line represents the power supply voltage of Trojan-inserted ICs. Figure 6.5b shows that for a particular ring oscillator, the cycle count difference between Trojan-free ICs and Trojan-inserted ICs will increase as the threshold voltage of the transistors increases until a maximum is reached. Once this maximum has been reached, increasing the threshold adversely affects the cycle-count difference (and thus the sensitivity to inserted Trojans). The X axis in Fig. 6.5b represents the threshold voltage coefficient, V_{th}/V_{sth}, where V_{sth} is the SVT of the MOS transistors. In the Synopsys 90nm technology library, the threshold voltage coefficient of the HVT transistors is 1.2. With HVT ring oscillators, (6.8) becomes:

$$\frac{I_{total,t}}{f_t} = \sum_{i=0}^{i=N+n_t} \lambda_i \times (N + n_t) \times 2n \times k_g (1 + \alpha \times \frac{\Delta V_t + V_{hth} - V_{sth}}{V_{dd} - \Delta V_t - V_{hth}}) \quad (6.9)$$

where V_{hth} is the high threshold voltage of the transistors in the ring oscillators. As seen in (6.9), the relationship between the IC's dynamic current and the frequency of a ring oscillator in the circuit will be more sensitive using HVT transistors. In addition, Trojans with larger (n_t) and more IR-drop (ΔV_t) are easier to detect.

Some of the parameters in (6.9) will change with process and environmental variations. Since ICs are tested under the same temperature condition in a production test environment, only small environmental variations will be considered. All remaining parameters are susceptible to process variations and statistical analysis will help to separate the contributions of process variations and Trojans to the transient power.

6.3 Ring Oscillator Network Structure

As previously discussed, Trojan gates' switching impacts both the frequency of nearby ring oscillators and the IC's dynamic current. Since a Trojan's effects may be localized (i.e. tightly distributed), and the impact of a Trojan on a ring oscillator is dependent upon the distance between them, one ring oscillator may not be sensitive enough to distinguish the effects of Trojans from process variations throughout the entire IC. An improved RON, however, can improve sensitivity to Trojan noise.

Figure 6.6 shows a circuit into which the proposed on-chip structure with N_{RO} n-stage ring oscillators is inserted. These n-stage ring oscillators are each composed of one NAND gate and $n - 1$ inverters with one component located in each of the n rows of the standard cell design. The ring oscillators are more sensitive to the voltage drop caused by a Trojan if they share the same power strap. Therefore, it is highly advantageous to ensure complete coverage of the power distribution network by placing at least one ring oscillator component in each row of the standard cell (and thus near each power strap). One set of n-stage ring oscillators will be inserted between two vertical straps. If there are M vertical power straps and R rows in the design, $N_{RO} = (M + 1)\lceil R/n \rceil$. However, the number of ring oscillators can be adjusted according to the required Trojan detection sensitivity and the minimum sensitivity to Trojan activity.

The on-chip structure also includes a linear feedback shift register (LFSR), one decoder, one multiplexer, and one counter. The LFSR will supply random functional patterns for the entire IC during the signature generation and authentication processes; the same seed must be used for each golden IC and each IC under authentication. The decoder and the multiplexer are used to select which ring oscillator is measured. When a ring oscillator is selected, the decoder enables that particular RO and the multiplexer transmits the output of that RO to the counter. The counter measures the cycle count of the selected ring oscillator over a specified duration. The number of stages in a ring oscillator is limited by the operating speed of the counter, which is determined by the technology node. For example, using Synopsys 90nm technology, a 16-bit counter can operate at a maximum frequency of $1GHz$ according to HSPICE simulation.

Since the ring oscillators are only enabled during the production test and the authentication phase, their power overhead in the field is negligible. The proposed architecture has a small area overhead, due mainly to the ring oscillators. For larger

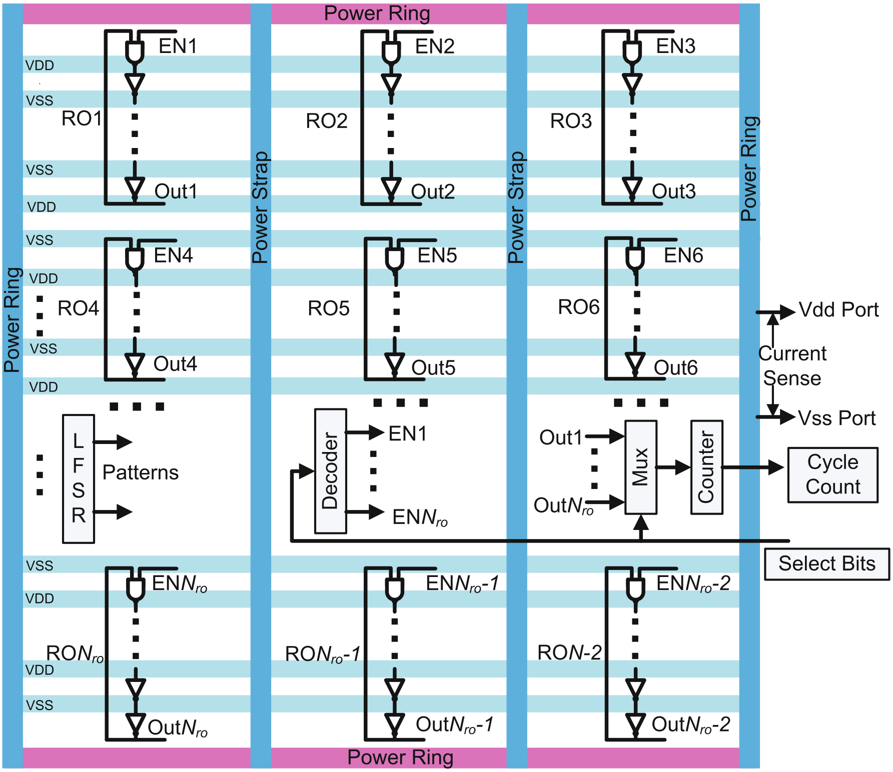


Fig. 6.6 The proposed on-chip structure with each gate of the ring oscillators placed in a standard cell row

circuits, assuming there is one vertical power strap for every 20 FFs or 80 gates, the area overhead of the ring oscillators will be approximately $1/(20 \times 4) = 1.25\,\%$. The total area overhead will be approximately 2.5% if there is only one vertical strap for every 10 FFs or 40 gates in the design. For a small circuit, the counter may play a significant part in the area overhead, but the counter size does not increase linearly with the size of the circuit. Since LFSRs are commonly used for built-in self-tests in modern designs, it can be ignored when analyzing the area overhead. However, even with LFSRs, the area overhead of a RON in large designs is still quiet small, since the area of the LFSRs does not increase significantly with the size of the circuit either. Transient current will be measured externally (i.e. with no area cost). In summary, the area overhead will be less than 3% for a large circuit and would be slightly larger for a smaller circuit. For instance, the overhead of RONs with LFSRs is 5.58% for the ISCAS'89 benchmark circuit s38584 (which contains four vertical power straps), 2.47% for an AES circuit (with six vertical straps), and 1.99% for a DES circuit (with six vertical power straps). The AES and DES circuits are provided in [10].

Since the ring oscillators are distributed across the entire IC, it is inherently difficult for an adversary to remove or tamper with one. If one of the ring oscillators reports data outside of a certain range or does not report data, it must have been attacked. In addition, this proposed on-chip structure is resilient to modeling. Some attackers may build up a lookup table to repeatedly generate the same cycle count for each ring oscillator, which would attempt to replace the Trojan-effected counter values with known good values. However, the current consumed by the lookup tables may be captured by the external current measurement and the power signature generated by the outlier analysis would also be changed. On the other hand, if the ROs are replaced with lookup tables embedded in the design, the frequency of the same ring oscillator at the same location, but on a different chip would stay at the same value in different ICs. However, unlike the value stored in a lookup table, the measured frequency of an RO in different ICs should be slightly different due to different process variations in Trojan-free circuits. If one ring oscillator in all CUTs has exactly the same frequency, designers will easily know that the IC was tampered with using embedded lookup tables.

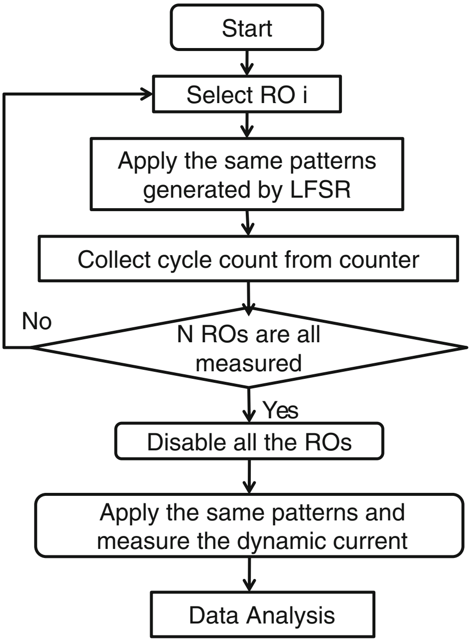
6.4 Measurement Flow and Statistical Analysis

The measurement flow for each IC is shown in Fig. 6.7. To measure the frequency of N_{RO} ring oscillators, the LSFR patterns with the same seed will be applied N_{RO} times. The transient current is measured externally. A signature is generated by recording the cycle count of each ring oscillator and the transient current from a large number of ICs of the same design. Since the ICs will all be subject to different process variations, this signature can be statistically more tolerant to similar variations in chips under authentication. In order to separate the effect of process variations and Trojans, a data analysis flow is suggested which includes the following three methods: (i) Simple Outlier Analysis, (ii) Principal Component Analysis (PCA), and (iii) Advanced Outlier Analysis.

Simple outlier analysis is based on the statistical distribution of the oscillation cycle count for each ring oscillator in the RON. For each ring oscillator, the oscillation count is within a certain range for Trojan-free ICs. If the oscillation count of even one ring oscillator in the IC under authentication is outside of the range, this IC is considered suspicious and might contain a Trojan. This method uses the information from individual ring oscillators but not the relationships among ring oscillators in the network, nor the dynamic current of the entire IC. This method can often identify a small number of Trojan-inserted ICs but may fail to detect most Trojan-inserted ICs. If the oscillation cycle count of all ring oscillators in an IC under authentication is within each Trojan-free IC's signature, the data collected from this IC will be processed by PCA and advanced outlier analysis.

The principal component analysis method [4] is used to account for the $N_{RO} + 1$ variables. With one variable representing one ring oscillator, there are N_{RO} variables, and the $N_{RO} + 1$th variable represents the dynamic current. The relationship

Fig. 6.7 Measurement flow of proposed method


Start
Select RO i
Apply the same patterns generated by LFSR
Collect cycle count from counter
N ROs are all measured
No
Yes
Disable all the ROs
Apply the same patterns and measure the dynamic current
Data Analysis

between the data from the N_{RO} ring oscillators and the dynamic current is considered by PCA when it transforms the $N_{RO} + 1$ variables into uncorrelated variables. The $N_{RO} + 1$ variables are transformed by PCA and the first three of the resulting components in Trojan-free ICs are used to construct a convex hull [5]. If the output of the CUT is beyond the convex, a Trojan must exist in the IC under authentication. However, if the output is inside the convex, an advanced outlier will be used for further analysis and validation.

An advanced outlier analysis has been developed to identify Trojan-inserted ICs that cannot be detected by simple outlier analysis and PCA. It considers the relationships among ROs in the RON and the dynamic current of the entire chip. The pseudo-code is shown in Fig. 6.8. For each Trojan-free IC, two out of N_{RO} ring oscillators will be selected along with the dynamic current information to generate a power signature (shown in Fig. 6.8a). For a particular Trojan-free IC, the total oscillation cycle count from the RON is $CC_{RON} = \sum_{m=1}^{N_{RO}} CC_m$. Then, the data from the RO_i (CC_i) and RO_j $(j \neq i)$ (CC_j) are selected to calculate $x_i = (CC_{RON} - CC_i)/C_i$ and $y_j = (CC_{RON} - CC_i)/CC_j$. Finally, (x_i, y_j, I) from all the Trojan-free ICs would be used to generate the power signature, PS_{ij}. There will be $N_{RO} \times (N_{RO} - 1)$ unique power signatures in total. The same process will be applied to the CUT (shown in Fig. 6.8b). If the CUT lies within the signature, it may be assumed that the circuit is Trojan-free. Otherwise, if one of the $N_{RO} \times (N_{RO} - 1)$ signatures does not match the Trojan-free signature, it will be treated as a suspicious part, i.e., Trojan-inserted.

a

```
Power Signature Generation
Collect data from $N_{TF}$ Trojan-free ICs with $N_{RO}$ ring oscillators
for ($i = 1, i <= N_{RO}, i++$) {                         //select $i^{th}$ $RO$
  for ($j = 1, j <= N_{RO}, j++$) ($j \neq i$) {          //select $j^{th}$ $RO$
    for ($k = 1, k <= N_{TF}, k++$) {           //select $k^{th}$ Trojan-free IC
      $x_{ki}$=($\sum_{m=1}^{N_{RO}} CC_{km} - CC_{ki})/CC_{ki}$;
      $y_{kj}$=($\sum_{m=1}^{N_{RO}} CC_{km} - CC_{ki})/CC_{kj}$;
      plot($x_{ki}$, $y_{kj}$, $I_K$);
    }
    The power signature, $PS_{ij}$, is created from all $N_{TF}$ ICs.
  }     //$x_{ki}$ is named as the first vector
}     //$y_{kj}$ is named as the second vector
Note: $CC_{km}$, $CC_{ki}$, $CC_{kj}$: Oscillation cycle count of $RO_m$, $RO_i$, $RO_j$ in $k^{th}$ IC
      $I_k$: The dynamic current of $k^{th}$ IC
```

b

```
Authentication
For each IC under authentication:
Collect data from $N_{RO}$ ring oscillators ($CC_{RON} = \sum_{i=1}^{N_{RO}} CC_i$)
for ($i = 1, i <= N_{RO}, i++$) {
  for ($j = 1, j <= N_{RO}, j++$) ($j \neq i$) {
    $x = (CC_{RON} - CC_i)/CC_i$;
    $y = (CC_{RON} - CC_i)/CC_j$;
    plot($x$,$y$,$I$);
    if (($x$,$y$) is outside of the power signature $PS_{ij}$
      {The IC is Trojan-inserted; Break; }
    else go on;
  }
}
Note: $CC_i$, $CC_j$: Oscillation cycle count of $RO_i$, $RO_j$
      $I$: Dynamic current of the under test IC
```

Fig. 6.8 Advanced outlier analysis procedure

6.5 Simulation Results and FPGA Implementation Analysis

The proposed approach is implemented on a small s9234 benchmark using Synopsys 90nm technology and a larger circuit, the AES benchmark, on Xilinx Spartan-6 FPGAs (45nm technology). For IC simulation, the s9234 benchmark was designed with two vertical power straps and 35 rows, with $N_{RO} = 15$ ring oscillators constituting the on-chip structure. Twenty Trojans (T_1 to T_{20}) with different sizes, gates types, and physical distributions were inserted into s9234. Table 6.1 shows these twenty Trojans where FF represents a flip-flop, $Cen.$ indicates that the Trojan is centrally located, and $Dis.$ indicates that the Trojan is physically distributed (shown in Fig. 6.9). Ten combinational Trojans (T_1–T_{10}) tap internal signals working as comparators, and the sequential Trojans (T_{11}–T_{20}) act as shift registers. None of

Table 6.1 Twenty Trojans inserted in s9234 circuit

	Combinational Trojans									
	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	T_9	T_{10}
Sizes	2 gates	3 gates	4 gates	5 gates	7 gates	8 gates	10 gates	12 gates	16 gates	20 gates
Area overhead (%)	0.09	0.16	0.2	0.25	0.37	0.43	0.5	0.66	0.82	0.92
Distribution	Cen.	Cen.	Dis.	Dis.	Cen.	Dis.	Dis.	Cen.	Dis.	Dis.

	Sequential Trojans									
	T_{11}	T_{12}	T_{13}	T_{14}	T_{15}	T_{16}	T_{17}	T_{18}	T_{19}	T_{20}
Sizes	2 FFs	3 FFs	4 FFs	5 FFs	6 FFs	7 FFs	8 FFs	10 FFs	12 FFs	16 FFs
Area overhead (%)	0.41	0.62	0.81	0.98	1.18	1.4	1.61	1.83	2	2.21
Distribution	Cen.	Cen.	Dis.	Dis.	Cen.	Dis.	Dis.	Dis.	Dis.	Dis.

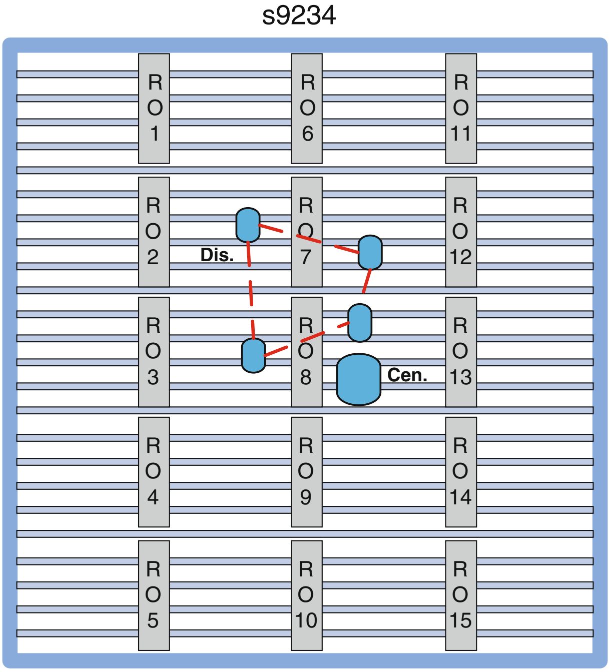


Fig. 6.9 s9234 with 15 ROs and 20 Trojans. One Trojan at a time is inserted into the circuit

these Trojans were detected by a test suite made up of 80,000 random functional patterns and 206 structural patterns (created by ATPG tools) for detecting stuck-at and transition delay faults. StarRC was used to extract parasitic parameters from the layout of benchmarks and to generate SPICE files. A Monte Carlo simulation (performed with Synopsys Nanosim) was used to emulate the effects of process variations that impact the frequencies of the ring oscillators and the dynamic current. The simulation temperature was 25°C with ±5°C variations. For hardware

Table 6.2 The oscillation cycle count of some of the ring oscillators and the circuit dynamic current in the presence of hardware Trojans without process variations

		T_1			T_3			T_6			T_{10}		
		TF	TI	ΔT	TF	TI	ΔT	TF	TI	ΔT	TF	TI	ΔT
Average dynamic current (μA)		29.8	29.84	0.04	25.56	25.65	0.09	24.94	25.08	0.14	24.94	26.1	1.16
RO (CC)	RO8	2790	2,787	−3	3,396	3,392	−5	3,064	3,054	−10	3,064	3,024	−40
	RO7	3,021	3,021	0	3,528	3,528	−2	3,008	3,005	−3	3,008	2,998	−10
	RO1	2,952	2,952	0	3,377	3,377	0	2,985	2,984	−1	2,985	2,982	−3
	RO15	3,103	3,103	0	3,406	3,406	0	2,803	2,803	0	2,803	2,801	−2

		T_{11}			T_{16}			T_{20}		
		TF	TI	ΔT	TF	TI	ΔT	TF	TI	ΔT
Average dynamic current (μA)		27.48	27.6	0.12	23.14	23.77	0.63	26.85	29.05	2.25
RO (CC)	RO8	3,150	3,141	−9	3,120	3,084	−36	3,031	2,972	−59
	RO7	3,117	3,117	−0	3,158	3,150	−8	2,925	2,914	−11
	RO1	3,042	3,042	0	3,198	3,198	0	2,980	2,977	−3
	RO15	3,132	3,132	0	3,210	3,210	0	3,012	3,011	−1

validation, eight Trojans (T_{21}–T_{28}) with different gates and distributions were inserted into an AES benchmark. Trojan-inserted and Trojan-free versions of the AES benchmark were both implemented on multiple FPGAs at room temperature; multiple FPGAs were used to analyze the effects of both inter-die and intra-die process variations.

6.5.1 Effectiveness Demonstration

Trojan Size and Distribution Analysis: Using a simulation without variations, the detailed cycle count and dynamic current results of T_1–T_3, T_6, T_7, and T_{12} with four ring oscillators (RO1, RO7, RO8, and RO15) during a 1,000-clock cycle LFSR test are shown in Table 6.2. Since the IC's dynamic current varies with the test pattern applied, the waveform of the dynamic current is recorded during the simulation. The average dynamic current in the measurement time window is shown in the table as well. In Table 6.2, TF indicates that the data in this column was collected from Trojan-free ICs while TI denotes data from Trojan-inserted ICs. ΔT represents the difference between the Trojan-inserted ICs and the Trojan-free ICs. As can be seen from Table 6.2, the Trojans consume extra power, increase the dynamic current, and decrease the cycle count of the ring oscillators.

Table 6.2 shows that T_1, T_3, and T_{11} have a larger impact on the oscillation frequency of RO8 than the other ring oscillators. Similarly, for T_6, T_{10}, T_{16} and T_{20}, there is a larger impact on RO8 and RO7 than on RO1 and RO15. This phenomenon is explained by the power supply voltage's dependence on the voltage division coefficient which is partially determined by the distance (resistive path) between two gates; a smaller distance implies a greater Trojan impact on ring oscillators. The remaining Trojans not shown in Table 6.2 exhibit similar behavior upon ring oscillators. However, the total dynamic current does not vary with the distributions of Trojans.

As Table 6.2 shows, in these seven Trojans, the oscillation cycle count difference ΔCC_{ft} of RO8 increased with Trojan size from −3 (for T_1) to −59 (for T_{20}). This occurred due to the greater power supply noise imparted from the Trojan with more gates. The dynamic current difference between a Trojan-free IC and a Trojan-inserted IC varies from 0.04 to 2.25 μA. Larger Trojans consume more power. Similar results can be observed for the Trojans not shown in the table. In general, larger Trojans have a greater impact on the power supply network, and consequently have a greater impact on the ring oscillators and dynamic current measurements.

Process Variations Analysis: Random process variations, consisting of $3\sigma = 10\%$ voltage threshold (5% inter-die and 5% intra-die), $3\sigma = 3\%$ oxide thickness (2% inter-die and 1% intra-die), and $3\sigma = 10\%$ channel length (5% inter-die and 5% intra-die) are used in the following simulations to analyze their impact on the method. 200 Trojan-free ICs and 100 Trojan-inserted ICs for each Trojan are generated by Monte Carlo simulations. The statistical data analysis flow was used to process the data collected from these ICs. T_{10}, composed of 20 combinational gates, is used to show the detailed results of the data analysis flow.

A simple outlier analysis is first applied to distinguish the effects of Trojans and process variations. Histograms obtained from RO1, RO7, RO8, and RO15 in Fig. 6.10 show the distribution of oscillation cycle counts in the presence of process variations with T_{10}. Figure 6.10a displays a histogram of the cycle counts reported by RO8 with the inserted-Trojan, and Fig. 6.10b shows the same result without (w/o) the Trojan. The distributions of the two sets of oscillation cycle counts are plotted in Fig. 6.10c. The remaining figures (Fig. 6.10d–l) show the data distributions collected from RO7, RO1, and RO15, respectively. There is no significant change in RO7, RO1, and RO15. However, due to the presence of T_{10}, which is proximal to RO8, the RO8's distribution shifts leftward considerably. For RO8, the oscillation cycle range is 2756–3090 in Trojan-free ICs. 3 ICs out of the 100 ICs under authentication fell outside of the range, which are identified as containing a Trojan.

PCA is performed to analyze the data for the remaining 97 ICs. Figure 6.11a shows the power signature comparison using PCA with N_{RO} ring oscillators and the dynamic current for Trojan detection. The convex is drawn from the first three principal components with 200 Trojan-free ICs. The asterisks denote data obtained from ICs with Trojans that are shown to be separate from the convex hull. Thus, with the RON architecture and statistical analysis, T_{10} can be detected with 100% accuracy. However, with limited statistical analysis, or if the RON is subjected to the increasingly large variations of nano-scale technologies, *smaller Trojans* may not necessarily be detected with such accuracy, which was the case for T_1 to T_8 and T_{11} to T_{17}.

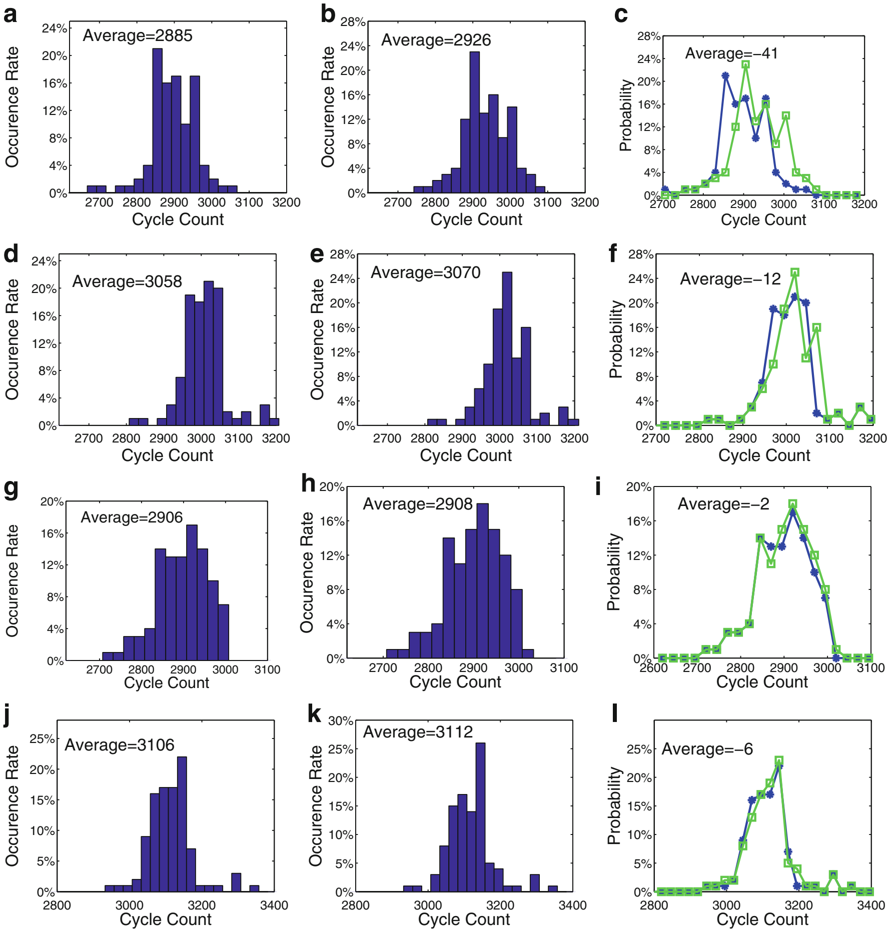


Fig. 6.10 Oscillation cycle distribution of RON with Monte Carlo simulations when T_{10} is inserted in s9234. (**a**) RO8 with Trojan; (**b**) RO8 w/o Trojan; (**c**) Cycle count distribution of RO8; (**d**) RO7 with Trojan; (**e**) RO7 w/o Trojan; (**f**) Cycle count distribution of RO7; (**g**) RO1 with Trojan; (**h**) RO1 w/o Trojan; (**i**) Cycle count distribution of RO1; (**j**) RO15 with Trojan; (**k**) RO15 w/o Trojan; (**l**) Cycle count distribution of RO15

The advanced outlier analysis shown in Fig. 6.8 is also used to identify Trojan-inserted ICs. There are a total of $15 \times 14 = 210$ power signatures generated by the Trojan-free ICs. Some power signatures could identify more Trojan-inserted ICs than others. In the following advanced outlier analysis results, only the power signature that can detect the most Trojan-inserted ICs is shown. Figure 6.11b shows the advanced outlier analysis results with Trojan T_{10}. The ring oscillator that was selected as x in Fig. 6.8 is defined as the first vector and y as the second vector. The

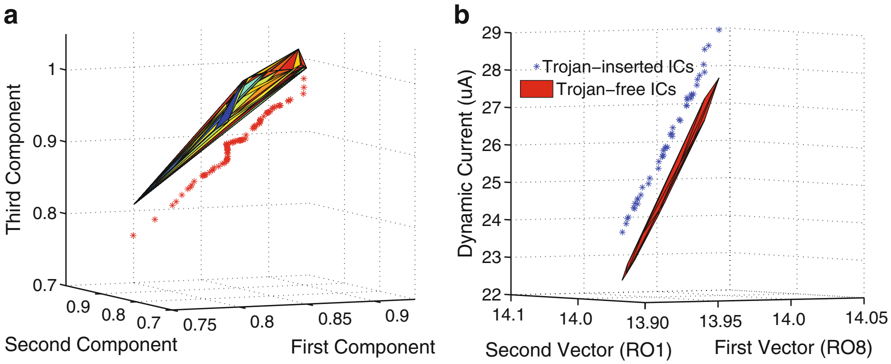


Fig. 6.11 Power signature for Trojan-free ICs and Trojan-inserted ICs with T_{10} using (**a**) PCA and (**b**) advanced outlier analysis

Table 6.3 Trojan detection rates with process variations

	Combinational Trojans									
	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	T_9	T_{10}
Trojan detection rate (%)	75	80	86	100	100	100	100	100	100	100

	Sequential Trojans									
	T_{11}	T_{12}	T_{13}	T_{14}	T_{15}	T_{16}	T_{17}	T_{18}	T_{19}	T_{20}
Trojan detection rate (%)	100	100	100	100	100	100	100	100	100	100

blue dots represent Trojan-free ICs and the red asterisks denote Trojan-inserted ICs. As can be seen in the figure, all of the Trojan-inserted ICs are outside of the Trojan-free signature. Thus, the detection rate with T_{10} using advanced outlier analysis is 100%.

Similarly, the remaining 19 Trojans with 200 Trojan-free ICs and 100 Trojan-inserted ICs are also simulated and the data analysis flow is applied for every Trojan. The Trojan-inserted ICs with T_1, T_4, T_{11}, and T_{20} are selected to present detailed results using advanced outlier analysis, shown in Fig. 6.12a–d. The detection rates of Trojans T_{11} and T_{20} shown in Fig. 6.12 are 100% with only one signature. For T_4, 98% of the Trojan-inserted ICs are detected using one signature, shown in Fig. 6.12b. When all 210 power signatures are used, the detection rate for Trojan T_4 is 100%. Complete results for all Trojans using all of the power signatures are shown in Table 6.3. From Table 6.3 and Fig. 6.12, it can be concluded that for Trojan T_4–T_{20}, the detection rates are all 100%. The power signatures of the Trojan-free ICs are completely separate from the Trojan-inserted ICs. However, the Trojan-inserted ICs with T_4 are close to the Trojan-free ICs. For the very small Trojans T_1–T_3, the detection rates are less than 100% because of their diminished impact on the power supply lines.

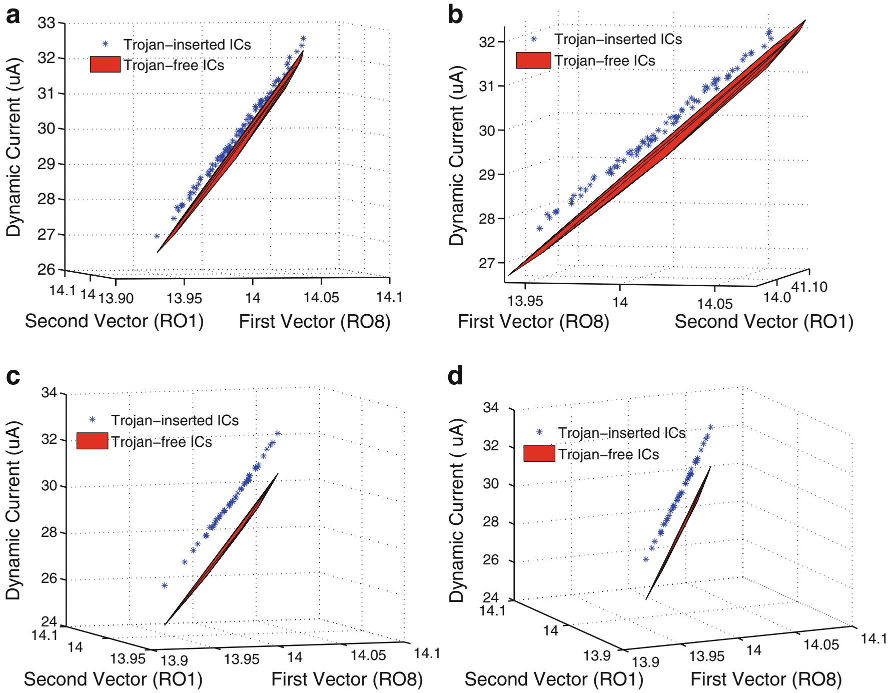


Fig. 6.12 Signatures with outlier data analysis from IC simulation for Trojan (**a**) T1, (**b**) T4, (**c**) T11 and (**d**) T20

6.5.2 Sensitivity Analysis

Ring Oscillator Number Analysis: Trojans T_1, T_2, and T_3 were chosen for ring oscillator number analysis since their detection rates are less than 100% with N_{RO}=15 ring oscillators. RONs with N_{RO}=10, 20, and 25 ring oscillators were implemented with Monte Carlo simulation. The location of the inserted Trojans is fixed throughout this analysis. For RONs with different quantities of ring oscillators, the layout is similar to Fig. 6.9. The three columns of ring oscillators were replaced by 2, 4, and 5 columns of ring oscillators.

Figure 6.13 shows Trojan detection rates using advanced outlier analysis with a different number of ring oscillators in RON for Trojans T_1, T_2, and T_3. With 10 ring oscillators, the detection rates for T_1, T_2, and T_3 are 40%, 48%, and 53%, respectively. With 25 ring oscillators, the detection rates increase to 95%, 100%, and 100%. These results imply that increasing the number of ring oscillators in the circuit improves detection rates. This is because a Trojan will likely be closer to a ring oscillator (or perhaps several) with more of them embedded in a design.

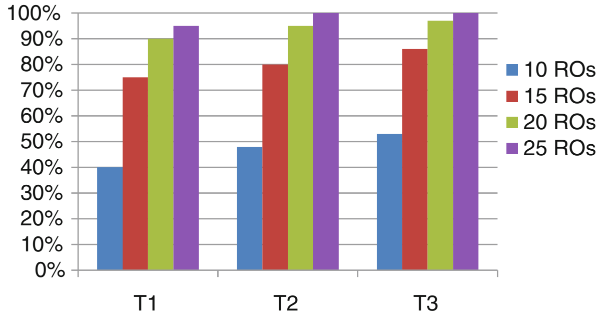


Fig. 6.13 Ring oscillator number (N_{RO}) analysis with Trojans T_1, T_2, and T_3

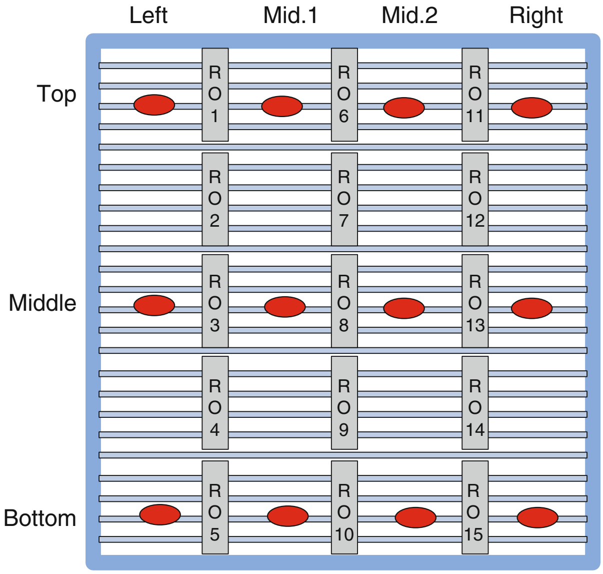


Fig. 6.14 Placing T_2 at different location in the s9234 circuit

When the number of ring oscillators in the RON is increased, the power consumption will remain unchanged while the circuit is under normal operation; the RON is only on for a short time during testing and remains off during functional operation. The area overhead would increase slowly with the number of ring oscillators.

For the simulation, the area overheads are 2.5%, 3.75%, 5.0%, and 6.25% with 10, 15, 20, and 25 ring oscillators in the RON inserted in s9234. However, the increase in area overhead is small in comparison to the increase in Trojan detection rates. Thus, the RON structure may be adjusted to meet desired area overhead and detection resolution values.

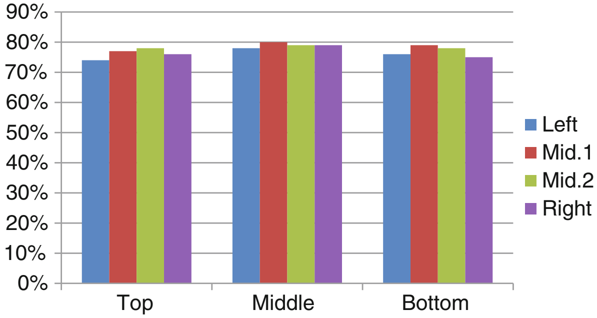


Fig. 6.15 Trojan location analysis with T_2

Trojan Location Analysis: In order to verify the impact of a Trojan's location on its detection rate, Trojan T_2 was placed in twelve locations (shown in Fig. 6.14). For each location, 200 Trojan-free ICs and 100 Trojan-inserted ICs were generated by a Monte Carlo simulation. A RON of 15 ROs was embedded into the benchmark. The detection rates using advanced outlier analysis are shown in Fig. 6.15. As the figure shows, when Trojan T_2 was placed around boundary corners, fewer Trojan-inserted ICs would be detected than if it was placed centrally. This occurs because the Trojan is closer to a greater number of ring oscillators when placed towards the center. However, the Trojan detection rate varies by less than 8% for the twelve locations. This can likely be alleviated with greater design coverage; placing ring oscillators in columns adjacent to the outermost edges of the IC will limit the maximum distance between a Trojan and an RO.

Pattern Analysis: Since different inputs could cause different switching activities within an IC, the pattern generated by the LFSR during testing can impact Trojan detection resolution in two ways: (1) Trojan switching activity (and thus the Trojan contribution to changes in dynamic power) depends on circuit inputs and thus the pattern selected, and (2) the total switching activity in the circuit may be altered by the patterns. Increased switching among Trojan gates implies a greater Trojan contribution to side-channel information. Decreased total switching in the circuit under authentication implies reduced background noise and a greater chance that Trojan activity will not be obfuscated. It is crucial to note that Trojan switching activity does not refer to the event of actually activating a Trojan to launch its malicious function, but rather refers to any amount of switching in the gates which comprise the Trojan. For example, for Trojan T_3, which is composed of four gates, if only one gate transitions, there is switching activity in the Trojan, whether or not the Trojan was completely activated. The LFSR was simulated to verify the impact of pattern selection on the combined ring oscillator network and the dynamic current method. Different seeds are used in the LFSR to generate different patterns; 1,000 patterns are generated by one seed.

Figure 6.16 shows the detection rates with four different seeds ($S1$, $S2$, $S3$, and $S4$) in the LFSR. The four different seeds were randomly generated by MATLAB.

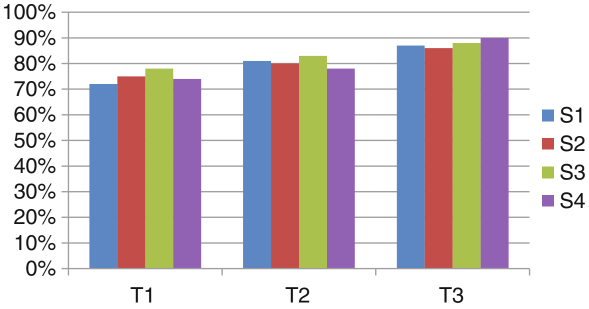


Fig. 6.16 Pattern analysis with Trojans T_1, T_2, and T_3

Trojans T_1, T_2, and T_3 were selected to show the results. All these Trojans were fixed at locations shown in Fig. 6.9. As seen in Fig. 6.16, the Trojan detection method gives different detection rates using different random patterns. Generally, the detection rate will be higher if the Trojan switching activity is greater. However, the Trojan detection rated does not vary significantly with random patterns. If special patterns are generated, such as ones that could cause more switching at the nets that rarely activate in the design, the Trojan detection method would be more effective.

6.5.3 Experimental Results from Spartan-6 FPGA

Xilinx Spartan-6 FPGA boards (shown in Fig. 6.17a) were used for the hardware validation of the proposed method and 24 ring oscillators were inserted into an AES benchmark circuit (shown in Fig. 6.17b). An Atmel Atmega328P microcontroller is connected to the FPGA to facilitate the collection of ring oscillator cycle count data from the counter. Transient current waveforms (shown in Fig. 6.18a) are collected using Digilent Adept software [8]. 28 Trojan-inserted FPGAs and 60 Trojan-free FPGAs were used to verify the impact of process variations. Several measurements were done for each ring oscillator in each FPGA in order to eliminate measurement noise, and the average oscillation count was used to perform data analysis. The Trojans implemented in the following analysis are composed of arbitrary combinational gates of varied sizes. The malicious function to be carried out by the Trojans will not be important since this analysis is intended to demonstrate the ability of the method to detect arbitrarily added yet difficult to detect malicious gates.

Eight different Trojans T_{21}–T_{28} with different sizes were inserted into the AES benchmark. As seen in Table 6.4, some Trojans are extremely small and switch rarely during functional operation. For example, the switching probability of Trojan T_{21} is 0.0016%. These Trojans were found at location L_3 (shown in Fig. 6.17b). The area overhead and detection rates of these Trojans are shown in Table 6.4. T_{26} was used to show the detailed results of the advanced outlier analysis in Fig. 6.18b. As can be seen from the figure, the Trojan detection rate for T_{26} is 100%. With all the Trojan detection rates (shown in Table 6.4), it can be concluded that most of

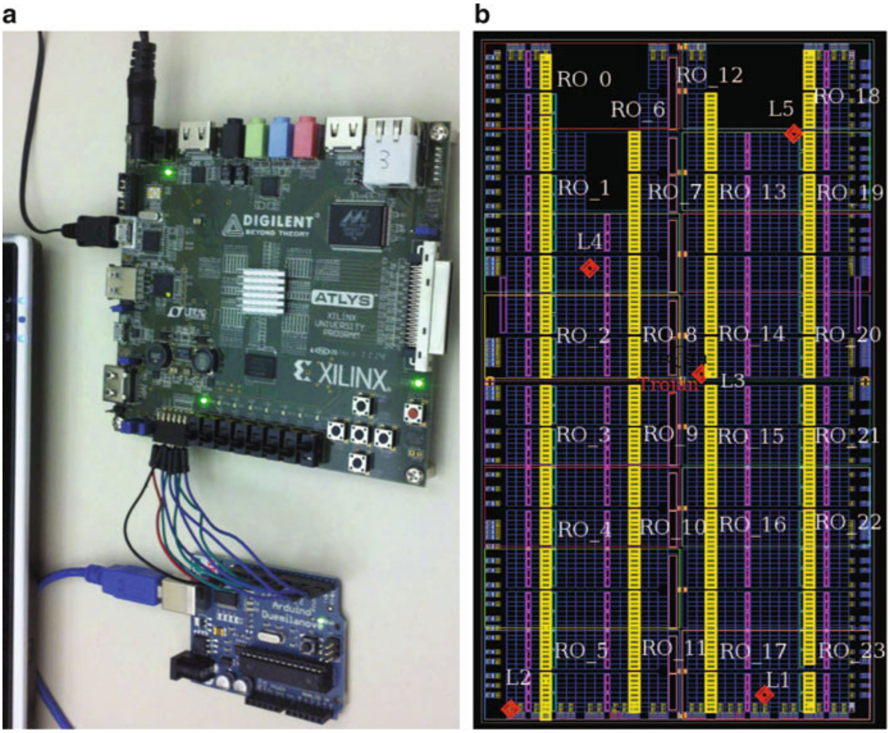
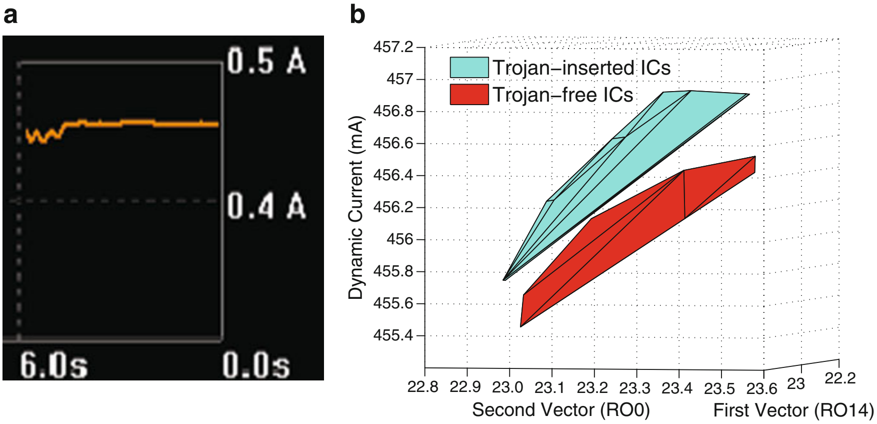


Fig. 6.17 (**a**) Xilinx Spartan-6 FPGA board (45nm technology) and (**b**) AES layout after placement



T_{26} T_{26}

Fig. 6.18 (**a**) Transient current waveform and (**b**) Outlier analysis results with Trojan T_{26} from FPGA implementation

Table 6.4 Trojans inserted in FPGAs and their detection rate when $N_{RO} = 24$

	T_{21}	T_{22}	T_{23}	T_{24}	T_{25}	T_{26}	T_{27}	T_{28}
Area overhead (%)	0.0016	0.012	0.025	0.05	0.08	0.1	0.15	0.2
Trojan detection rate (%)	73	86	100	100	100	100	100	100

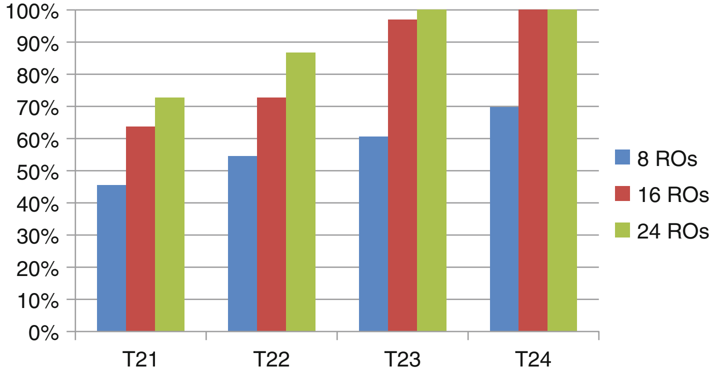


Fig. 6.19 Ring oscillator number analysis with Trojans T_{21}, T_{22}, T_{23} and T_{24} in FPGAs

Trojans were detected with a 100% detection rate. However, for very small Trojans, the detection rates were lower.

The impact of the number of ring oscillators on detection rates was analyzed on Xilinx Spartan-6 FPGAs, in addition to the simulation results presented earlier. Here, the number of oscillators in the network is varied and Trojans of diverse sizes are inserted into the circuit. The Trojans are placed in the same location, and the same LFSR seed is applied to each part of this experiment. RONs, composed of 8, 16, and 24 ring oscillators, were implemented in the AES benchmark circuit. Figure 6.17b shows the RON with 24 ring oscillators and RONs with 8 ring oscillators and 16 ring oscillators similarly implemented. With 60 Trojan-free FPGAs and 28 Trojan-inserted FPGAs, Fig. 6.19 shows detection rates with different RONs for Trojans T_{21}, T_{22}, T_{23}, and T_{24}. The figure shows that the number of ring oscillators in the RON plays a considerable role in the effectiveness of the method. For T_{23} and T_{24}, a detection rate of 100% is achieved by increasing the size of the network from 8 to 24 ring oscillators.

Also, a significant improvement is achieved by increasing the number of ROs from 8 to 16, but a smaller improvement is seen when the number of ROs is increased from 16 to 24. This suggests that detection resolution is not linear with the number of ring oscillators in RON.

To analyze the sensitivity of the method to the location of Trojans, T_{22} was placed in different locations, from L_1 to L_5. Figure 6.20 shows results using the data analysis flow. The detection rate vacillates between 88.3% and 79.3% with changes in the Trojan's location. When the Trojan was placed in locations L_4 and L_3, the detection rate was relatively higher, since it impacted more ring oscillators.

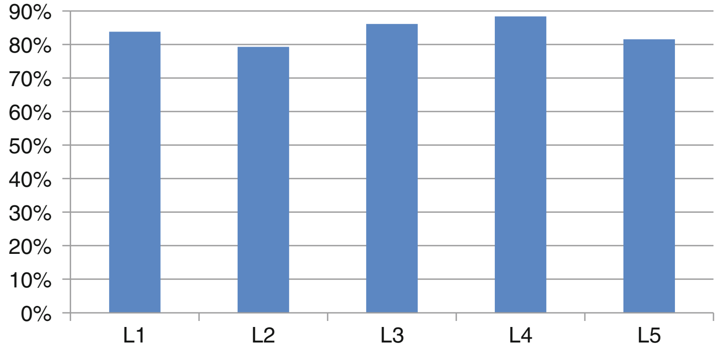


Fig. 6.20 Trojan location analysis with Trojans T_{22} in FPGAs

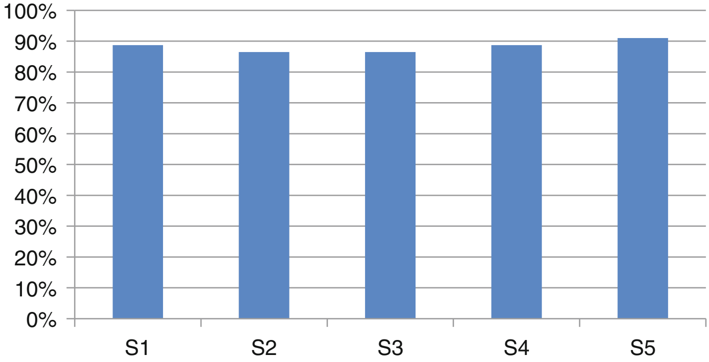


Fig. 6.21 Patterns analysis with Trojans T_{22} in FPGAs

When the Trojan was located in L_2, at a corner of the FPGA, the Trojan detection rate is at its lowest.

To analyze the impact of these patterns, Trojan T_{22} was located in L_3. Six randomly selected seeds were applied to the LFSR. The ring oscillator cycle counts and transient current waveforms were collected and analyzed. Figure 6.21 shows the data analysis results. The figure shows that random patterns do not have a significant impact on the Trojan detection rate. However, if a designer were to intelligently select a set of patterns that control background noise and net coverage, additional improvements in detection resolution are possible.

6.6 ASIC Evaluation

6.6.1 Test Chip Design

In order to analyze the effectiveness of the RON structure, 40 test chips were designed and fabricated using IBM 90nm technology through MOSIS. All chips

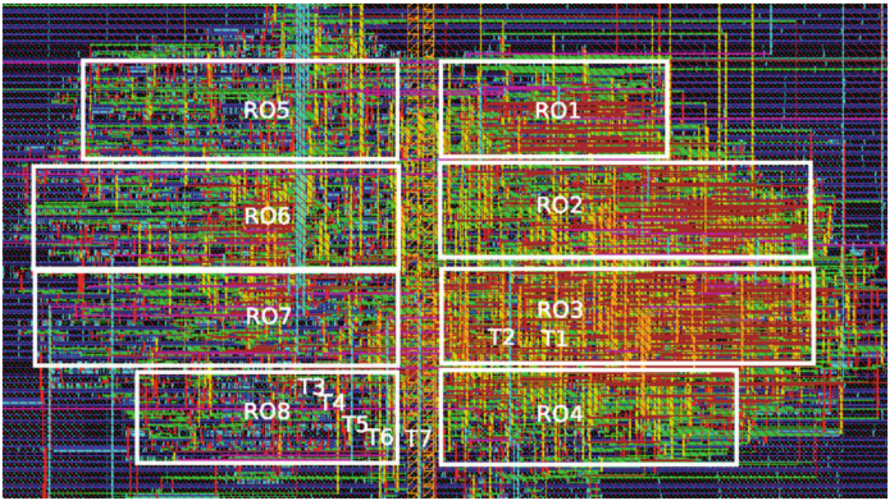


Fig. 6.22 Layout for the test chip design

were fabricated on the same wafer. The RON architecture is inserted into the ISCAS s9234 benchmark, which represents the design to be protected in the test chip.

Figure 6.22 shows the layout of the test chips with the RON structure composed of $N_{ro} = 8$ $n = 61$-stage ROs (RO_j, where $1 \leq j \leq 8$) with one NAND gate and 60 inverters, each distributed across the chip. It is important to note that the areas labeled RO_1 to RO_8 show the broad area in which that RO is confined rather than the total area occupied by that RO. Ring oscillator stages are placed in each standard cell row in an intentionally loosely distributed fashion that improves its coverage of the power distribution network. Therefore, these areas are also occupied by background circuit and control structure components and the area overhead of the oscillators is substantially lower than the labeled areas. The approximate locations of the seven Trojan stages (T_i where $1 \leq i \leq 7$) are labeled as well. The number of RO stages was selected so that the maximum observed frequency would not exceed the 400 MHz operating frequency of the 90 nm counters used in this design. The distance between the two adjacent RO components is limited to 10 times the width of the flip-flops. Given this design rule and the area of the chip, 8 ROs were used.

The feedback polynomial of the LFSR used in the test chip is

$$X^7 + X^3 + 1 \tag{6.10}$$

To conserve area, this design uses an LFSR with only 8-bits to generate patterns for the 36 input s9234 benchmark. A broadcasting technique is used to assign this 8-bit output to the 36 inputs. An 8-bit decoder and 8-bit multiplexer are used for RO selection. A 16-bit counter is used to measure the number of oscillations observed in the test duration, which is controlled by a timer. In this design, the test duration of 500 clock cycles was selected based on the technology node and test area overhead.

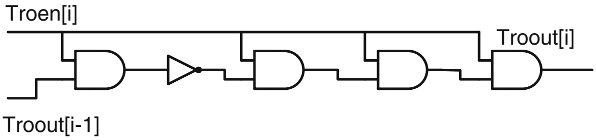


Fig. 6.23 Design of a hardware Trojan stage T_i

Table 6.5 An estimation of the area occupied by s9234 in terms of the number of transistors

Component	Quantity	Total transistors
D Flip-Flops	211	7,174
Inverters	3,570	7,140
Gates	2,027	8,108
Total	5,808	22,422

6.6.2 Hardware Trojan Design

Each IC contains seven combinational hardware Trojan designs that may be completely deactivated. Since this design is implemented in 90nm CMOS technology, the static power dissipation and thus the side-channel contribution is negligible when the Trojans are deactivated. By using a single-IC multiple-Trojan design we are able to not only carry out a more extensive set of Trojan impact tests, but we are also able to isolate the effect of process variations from the effect of inserted Trojans on RO characteristic frequencies. Further, since the static power is present in the Trojan-free case, it is neglected and the detection results provide a lower-bound.

The gate-level implementation of a Trojan stage is shown in Fig. 6.23 where $troout[i]$ is the output of the ith Trojan stage, $troout[i-1]$ is the output of the previous Trojan stage, and $troen[i]$ is the enable signal for the ith stage which also asserts all prior enable signals when enabled.

Trojan T_i contains i stages consisting of $i \times (4AND + 1INV)$ gates where each stage $i-1$ is also enabled if stage i is enabled. The first Trojan, T_1 is driven by the 200 MHz clock signal at the location of signal $troout[0]$. Note that the Trojan T_i is not derived of the trigger-payload Trojan design used in [11–13]. Here, each Trojan gate transitions once per clock cycle; therefore, the partial activity of each of these Trojans is simply $5i$ partial activations per clock cycle. The average ratio of Trojan partial activation to background circuit activity is estimated in the fourth column of Table 6.6.

The s9234 benchmark consists of 211 D flip-flops, 3570 inverters, and 2027 other gates. The number of transistors used in the s9234 benchmark is estimated in Table 6.5 by assuming that each flip-flop consists of 8 NAND or NOR gates and 2 inverters. As previously mentioned, there are a total of seven Trojans (T_1–T_7) in this design. The area overhead of each Trojan is summarized in Table 6.6.

Table 6.6 An estimation of Trojan area overheads and noise

Trojan number	Transistors	Percent area (%)	Trojan to background circuit switching ratio (%)
T1	26	0.12	0.11
T2	52	0.23	0.22
T3	78	0.35	0.33
T4	104	0.47	0.45
T5	130	0.58	0.56
T6	156	0.70	0.67
T7	182	0.81	0.78

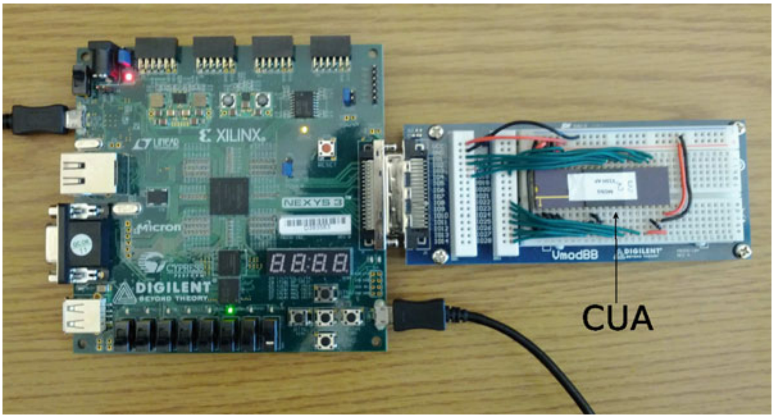


Fig. 6.24 A data collection setup including a Spartan 6 FPGA connected to a prototyping board through a serial connector

6.6.3 Experimental Setup

During data collection, the IC is mounted on and wired to a prototyping board that includes a high-density serial connector. The serial connector allows the prototyping board to interface with a Xilinx Spartan-6 FPGA on a Digilent Nexys 3 board. The FPGA is programmed to control the test sequence supplied to the IC and to transmit the outputs of the IC to a computer using an on-board USB-UART module. The complete setup is shown in Fig. 6.24.

The nominal supply voltage of the IC pins is 2.5V. This is converted internally to the nominal core voltage of 1.2V using a voltage divider. Since the s9234 benchmark circuit used in this design is small compared to a modern IC, in order to emulate the tight power design of a modern circuit, an external voltage divider is used to supply the IC with 1.875V and the core with 0.9V, which is greater than the 0.80V minimum core voltage. Reducing the power supply voltage will reduce the background circuit switching activity and improve Trojan detection rates. Therefore, it is desirable to reduce the supply voltage during measurement.

The FPGA includes a state machine which sequences through each ring oscillator, begins a data collection trial, selects each 4-bit window of the counter output for the current ring oscillator, and transmits each 4-bit window as a hex digit over the USB-UART connection. The process is repeated for 10 trials on each ring oscillator of each IC. The IC is supplied with 1.875V using a voltage divider and the board's 2.5V peripheral power supply over the serial connection along with a 200 MHz clock signal. Each trial lasts 500 clock cycles.

As shown in Fig. 6.22, each of the 40 ICs contains $N_T = 7$ pre-inserted hardware Trojan designs. During Trojan-free data collection, each hardware Trojan circuit is disabled, as is any Trojan not being analyzed. Since the designs are implemented with CMOS circuits, the static dissipation is negligibly low. Furthermore, since all Trojan measurements are compared to the Trojan-free results (which include static dissipation), the detection results provide a conservative lower bound.

6.6.4 Experimental Results and Analysis

The frequency of a single ring oscillator on a single IC was measured 10 times. The measurement noise is then calculated with

$$\frac{Max\{f_{Trial1}, \ldots, f_{Trial10}\} - Min\{f_{Trial1}, \ldots, f_{Trial10}\}}{0.1 \sum_{m=1}^{10} f_{Trialm}} \tag{6.11}$$

for a single IC and a single ring oscillator where f_{Trialm} is the mth repeated measurement of frequency for that RO. This is repeated for all ICs and all ROs and averaged, resulting in a measurement noise of 0.23%.

The impact of intra-die variation on an RO's frequencies was analyzed by comparing a single RO on an IC with other ROs on that same IC. For a single IC, intra-die variation is calculated with

$$\frac{Max\{f_{RO_1}, \ldots, f_{RO_8}\} - Min\{f_{RO_1}, \ldots, f_{RO_8}\}}{0.125 \sum_{j=1}^{8} f_{RO_j}} \tag{6.12}$$

where f_{ROj} is the frequency of the jth RO. This calculation is repeated for all ICs and averaged, resulting in a mean intra-die variation impact on frequency of 8.05%.

Of the 40 fabricated ICs, 38 functioned correctly and the remaining faulty ICs were omitted. The impact of inter-die variation on the frequency of a ring oscillator was determined by selecting a single RO and comparing the frequency of this RO across each IC. For a single RO, the inter-die variation is calculated with

$$\frac{Max\{f_{IC1}, \ldots, f_{IC38}\} - Min\{IC_1, \ldots, f_{IC38}\}}{(1/38) \sum_{k=1}^{38} f_{ICk}} \tag{6.13}$$

Table 6.7 A summary of validation data

Measurement noise	0.23%
Intra-die variation	8.05%
Inter-die variation	16.67%
Mean RO frequency	291 MHz

where f_{ICk} is the frequency of the individual RO of interest on the kth integrated circuit. This calculation is repeated for all ROs and averaged, resulting in a mean inter-die variation impact on frequency of 16.67%. The average RO frequency of all ROs on all ICs was 291 MHz. The maximum recorded frequency was 315 MHz, which was less than the 400 MHz frequency the counter was timing closed at. These results are summarized in Table 6.7.

Trojan Impact Analysis: The direct impact of hardware Trojan induced power supply noise on ring oscillator frequencies is analyzed by measuring the frequency of each RO on each IC for the Trojan-free case, as well as for each Trojan. The mean impact of a particular Trojan on a particular RO is then computed by comparing the frequency of that RO on a particular IC with the frequency of that RO on the same IC when the Trojan is disabled. The computation is thus

$$TROI_{ROj,Ti} = (1/38) \sum_{k=1}^{k=38} \frac{|RO_{j,k,Tfree} - RO_{j,k,Ti}|}{RO_{j,k,Tfree}} \times 100\% \tag{6.14}$$

where $TROI_{ROj,Ti}$ is the mean impact of the ith Trojan on the jth RO across all ICs compared to the Trojan-free case. $RO_{j,k,Tfree}$ is the Trojan-free frequency for the jth RO on the kth IC, and similarly, $RO_{j,k,Tj}$ is the frequency of the jth RO on the kth IC with the ith Trojan activated.

It is with this calculation that the value of the single-IC multiple-Trojan design is best demonstrated. By comparing measurements made with a Trojan enabled against measurements made on the same IC with the Trojan disabled, inter-die variation is eliminated from the analysis. Had separate ICs been fabricated with Trojans inserted and Trojans removed, only comparisons between different ICs would be possible, and the computation would include inter-die process variation. By restricting comparisons to the same RO, intra-die process variations are eliminated from the computation as well.

The results for Trojan impact are presented in Fig. 6.25. It is immediately clear that Trojans of greater area and those that more frequently partially activate induce a greater change in the frequencies of nearby ROs since they consume more power. The maximum induced change for the largest Trojans in this experiment is representative of one of the core issues in the IC trust problem. The Trojan induces at most a change of 2.5% to frequencies, yet as Table 6.7 reports, intra-die variation and inter-die variation induce far greater changes, suggesting these Trojans would be completely obfuscated in a test where these variations are not isolated. However, Trojan detection is still possible with this technique. The manner in which Trojan

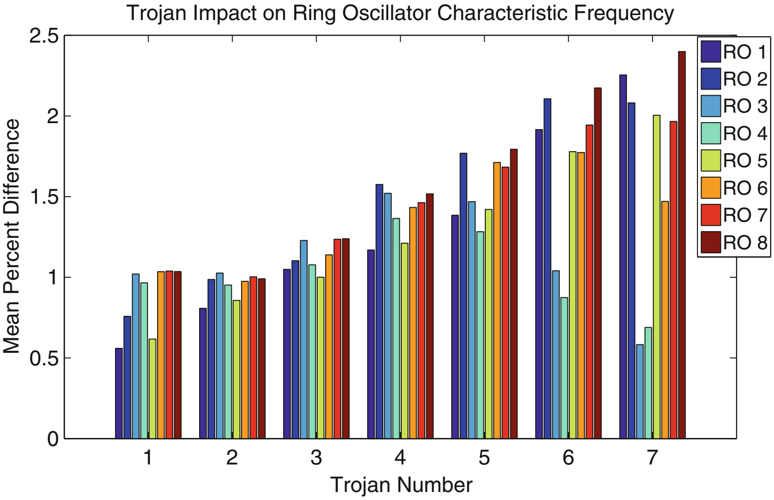


Fig. 6.25 The impact of inserted hardware Trojans on RO frequencies isolated from process variations

impact is distributed across ROs, including the decrease in impact on RO_3 and RO_4 for larger Trojans.

Spatial Locality Analysis: To analyze the effect of Trojan location, the ring oscillator that experiences the greatest Trojan impact calculated with (6.14) is determined for each IC with a particular Trojan. A histogram showing the frequency with which each ring oscillator was the most impacted on an IC is shown in Fig. 6.26. The location of Trojan gates relative to the gates of the ROs and the vertical power line is shown in Fig. 6.22

Notably, RO_8 is impacted most frequently of all Trojans since several of its gates are closest to the vertical power strap, thereby causing a portion of the overall power supply noise to affect this RO. For T_1 and T_2, a substantial portion of the Trojan impact is distributed on RO_2 and RO_3, since these Trojans are located close to these ROs and likely share power lines.

Since the majority of the gates in subsequent Trojans are closest to RO_8, more of the Trojan impact is distributed on this RO. Perhaps counter-intuitively, the distribution becomes more focused on a single RO as the Trojan expands in size. Had the Trojan expanded vertically and towards multiple ROs it is likely the distribution would have become less focused. However, for these Trojans that extend primarily horizontally, the increase in area and activity further increases the Trojan impact without expanding into other regions of the power network.

For T_7, the Trojan becomes less localized on RO_8 since T_7 is particularly close to the vertical power strap. For this reason, the Trojan impact is more evenly distributed across ROs since the vertical power strap supplies power to the entire circuit. Finally, the reduced impact on RO_3 and RO_4 for T_6 and T_7 shown in Fig. 6.25 is due to the

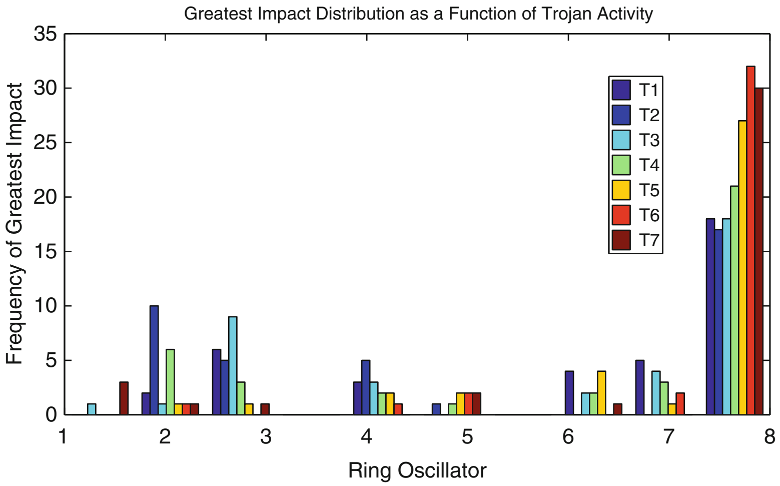


Fig. 6.26 The number of instances of each RO being most impacted by a Trojan

loosely distributed nature of these ROs away from the vertical power line and the placement of these Trojans close to the vertical power line.

IC Classification and False-Positive Analysis: In the previous section, it was shown that all Trojans used in this study impacted the RO frequencies substantially less than inter-die and intra-die process variations. However, using the principal component analysis (PCA) [4] based classification scheme presented below, it is still possible to detect these Trojans. In order to verify that this data is adequately represented in fewer than 8 principal components, the percent of the total variance in each PCA representation is computed by dividing the cumulative sum of the latent of the PCA representation by the total sum. The percent variance for each representation is shown in Table 6.8. The results imply that any representation of at least 2 components should adequately represent this data.

Table 6.8 The percent variation contained in a representation of *h* principal components

Components	Percent variation (%)
1	89.4
2	99.39
3	99.59
4	99.79
5	99.87
6	99.93
7	99.97
8	100

To succeed, a classification scheme must perform two functions: (1) it must correctly label Trojan-inserted circuits as tampered and (2) it must correctly label

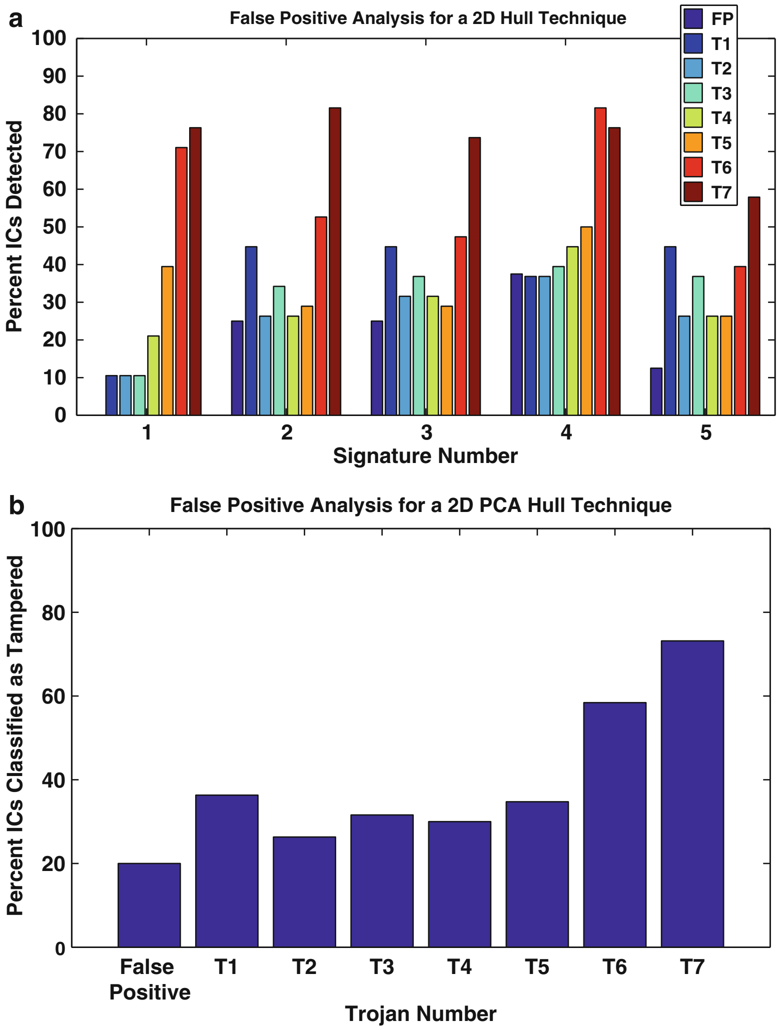


Fig. 6.27 Classification using the presented scheme and 2 dimensions. (**a**) All cases. (**b**) Mean rates

Trojan-free circuits as uncompromised. The steps for the presented classification scheme are:

1. Form a matrix from golden (Trojan-free) data in which each row is a verified Trojan-free IC and each column is a ring oscillator. Append a similar row containing the data from the chip under authentication (CUA) to the matrix.
2. Obtain a representation of this matrix using the first h principal components.
3. Render an h-dimensional convex hull [5] with all data, except that of the CUA.

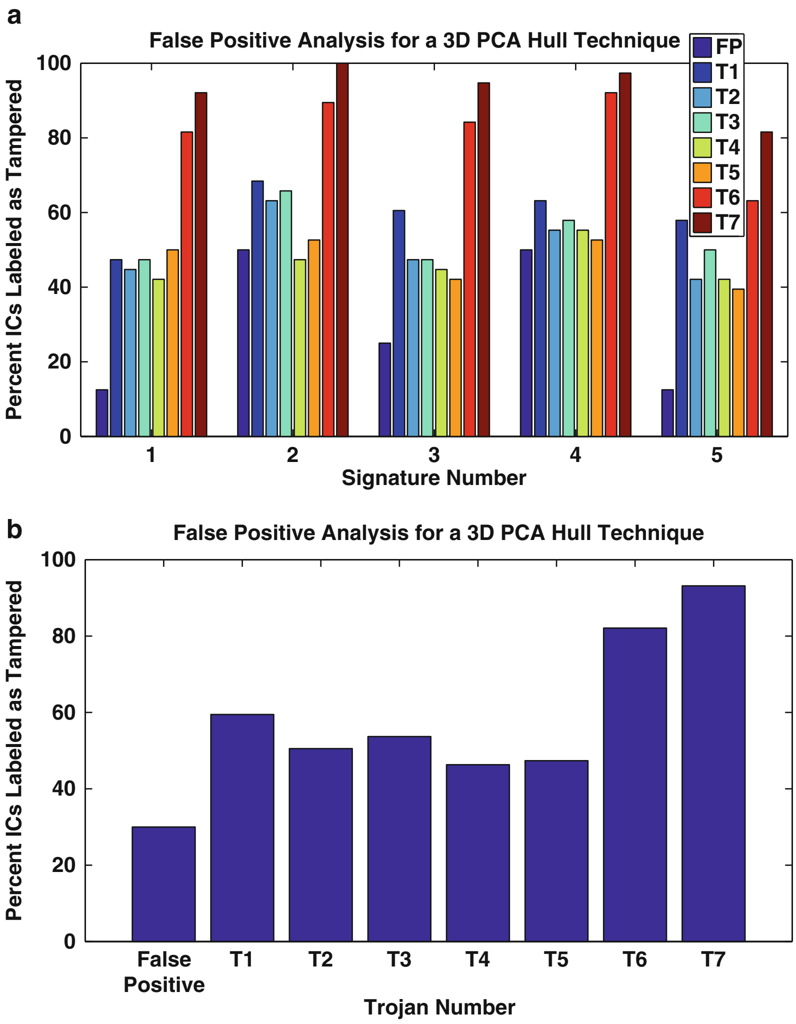


Fig. 6.28 Classification using the presented scheme and 3 dimensions. (**a**) All cases. (**b**) Mean rates

4. Determine if the CUA point falls within the hull. If it is within the boundaries of the hull it is considered Trojan-free.

To examine the performance of this classification scheme, the data is organized into five cases in which 8 of the 38 functioning ICs are randomly selected to represent Trojan-free chips to be authenticated and the remaining ICs are used to build the golden signature. All 38 ICs are used as Trojan-inserted chips under authentication.

The classification scheme was tested using both 2 and 3 dimensional hulls using the same subset cases for both hull types. The percent chips labeled as Trojan-inserted are shown for each case using both 2 and 3 dimensions in Figs. 6.27a and 6.28a respectively. "FP" indicates the number of Trojan-free chips that were incorrectly classified. For both 2 and 3 dimensions, the behavior varies among the randomly selected cases. Thus, for clarity, the average rates among all cases are shown in Figs. 6.27b and 6.28b. For both the 2 and 3 dimensional schemes, the false positive rates are lower than the detection rates for even the smallest Trojans in the experiment. For Trojans T1–T5 the detection rates are under 50%. This is unsurprising since these Trojans consisting of fewer than 130 transistors were intentionally designed to determine and emphasize the limitations of this technique.

For the larger Trojans, the detection rates are as high as 60–70% for the 2 dimensional case and 80–90% for the 3 dimensional case. Notably, the percent ICs labeled Trojan-inserted tends to be higher for the 3 dimensional case, indicating sensitivity is related to the number of dimensions used. However, the three-dimensional case also achieves a higher ratio of detection rate to false positive rate for some cases.

These results demonstrate that the ring oscillator network scheme and the presented classification scheme can adequately separate Trojan-inserted designs from the Trojan-free designs despite the presence of obfuscating process variations. Although intra-die and inter-die variations introduce roughly 8% and 17% variations in RO frequencies, compared to the 1–3% change induced by the inserted Trojans, this technique successfully classifies ICs by exploiting the spatially correlated nature of process variations.

6.7 Summary

In this chapter, an effective Trojan detection framework is presented, which combines an on-chip structure with off-chip current measurements. This technique has the capability of detecting very small Trojans with very little contribution to circuit transient current. Statistical analysis distinguishes the effects of hardware Trojans from process variations. The experimental results on 45nm FPGAs and on 90nm test chips demonstrated that this approach is very effective at identifying Trojan-inserted ICs.

References

1. Xuehui Zhang, Andrew Ferraiuolo, and Mohammad Tehranipoor, "Detection of Trojans using a Combined Ring Oscillator Network and Off-chip Transient-Power Analysis," *ACM Journal on Emerging Technologies in Computing Systems (JETC)*, 2012.
2. J. M. Rabaey, A. Chandrakasan, and B. Nikolic, "Digital Integrated Circuits: A Design Perspective (2nd Edition)," Prentice Hall, ISBN: 0-13-090996-3, 2003.

3. S. Zhao, K. Roy, and C. Koh, "Frequency Domain Analysis of Switching Noise on Power Supply Network," *Tecnical Reports*, 2000.
4. I. T. Jolliffe, "Principal Component Analysis (2ed Edition)," Springer, pp. 150–165, 2002.
5. F. P. Preparata and S. J. Hong, "Convex Hulls of Finite Sets of Points in Two and Three Dimensions," *Commun. ACM*, vol. 20, no. 2, pp. 87–93, 1977.
6. S.H.K. Embabi. "Digital BiCMOS Integrated Circuit Design." Kluwer, 1993.
7. T. Sakurai and R. Newton, "Alpha-power law MOSFET model and its applications to CMOS inverter delay and other formulas," *IEEE J. Solid-State Circuits*, vol. 25, no. 2, pp. 584–594, Apr. 1990.
8. *http://digilentinc.com/Products/Detail.cfm?NavPath=2,66,828&Prod=ADEPT2*.
9. S. Narasimhan, D. Dongdong, R. Chakraborty, S. S. Paul, F. Wolff, C. Papachristou, K. Roy, and S. Bhunia, "Multiple-parameter Side-channel Analysis: A non-invasive Hardware Trojan Detection Approach," in Proc. *IEEE HOST*, pp. 13–18, 2010.
10. *http://trust-hub.org/resources/benchmarks*.
11. D. Agrawal, S. Baktir, D. Karakoyunlu, P. Rohatgi, and B. Sunar, "Trojan Detection using IC Fingerprinting," in Proc. *IEEE Symposium on Security and Privacy (SP)*, pp. 296–310, 2007.
12. M. Banga and M. Hsiao, "A Region based Approach for the Identification of Hardware Trojans," in Proc. *IEEE HOST*, pp. 40–47, 2008.
13. F. Wolff, C. Papachristou, S. Bhunia, and R. S. Chakraborty, "Towards Trojan-free Trusted ICs: Problem Analysis and Detection Scheme" in Proc. *Design, Automation and Test in Europe (DATE)*, pp. 1362–1365, 2008.