Hardware Trojan Detection Acceleration Based on Word-Level Statistical Properties Management

He Li and Qiang Liu

School of Electronic Information Engineering, Tianjin University, Tianjin 300072, China
{heli@tju.edu.cn, qiangliu@tju.edu.cn}

***Abstract*—Hardware Trojan insertion has raised serious concerns to semiconductor industry and government agencies. Hardware Trojan is usually activated under rare conditions associated with low transition bits in a circuit. The damage includes circuit functional failure or important information leakage. Previous research on hardware Trojan detection is mainly based on side-channel analysis and Trojan activation. Long activation time is a major concern during the detection process. In this paper, we propose a novel approach for efficiently accelerating Trojan activation by increasing the transition activity of rare bits. In particular, the proposed approach increases the bit-level transition activity by controlling signal word-level statistical properties, such as changing the variance and autocorrelation of the signal. In addition, by analyzing the signal propagation statistical properties through various digital signal processing (DSP) operators such as adders and multipliers, the proposed approach can control the statistical properties of internal signals and then enhance the internal bit transition activity from the primary input of the circuit. The proposed approach is evaluated on several circuits. The results show that the transition activity of rare bits can be dramatically increased by up to 166.7 times and Trojan activation time can be reduced by up to 121 times.**

***Index Terms*—Hardware Trojan detection; signal statistical properties; transition activity**

I. Introduction

Globalization of semiconductor industry makes governments and enterprises raise serious concerns regarding hardware Trojan (HT) [1]. Potential adversary inserts HTs to a circuit design without affecting the functionality of the design. The attempt is to cause the circuit failure or to steal confidential information.

Detection of HTs is very challenging because they are designed to be inactive most of their life time and be activated under very specific conditions, such as connected to low transition activity nets or sensing a specific signal (*e.g.,* power and temperature [1]). The widely used HT detection methods include side-channel analysis and Trojan activation. The side-channel analysis approaches usually require a long stimulating time, so that HTs' abnormal behavior can be observed by measuring circuits' power [2], transient current [3, 4] and delay [5, 6]. The light weight of HTs also makes them very difficult to be distinguished from the effect of process variations. Trojan activation approaches [7–10] exploit the characteristics of designs to activate HTs in order to reduce the detection time, and have been used with the side-channel analysis.

This work provides a further progress in accelerating the HT detection process. For efficient detection, it is critical to find the rare bits with low transition activity, and then increase their activity. Different from the existing activation approaches [7–10], our approach increases the transition activity of bits by managing the word-level signal characteristics of circuit nodes involving DSP operators. DSP operations are the core functions in wireless communication and multimedia applications, and in recent years the architecture and performance of FPGA have been adapted to the requirement of digital signal processing with embedded DSP blocks. In the proposed approach, the bit-level transition activity is first determined by analyzing the word-level statistical properties of a signal. Then, by changing the signal's variance and temporal autocorrelation, we can effectively increase the transitions of the bits, which have low transition activity in a word. Moreover, the proposed approach analyzes the signal propagation statistical properties through DSP operators such as adders, multipliers, multiplexers and delays. Based on the analysis, the statistical properties of internal signals can be controlled from the primary inputs of the circuit. As a result, the proposed approach can enhance the bit-level transition activity of internal rare bits by input signal regulation. By increasing the transitions of rare bits, HTs can be activated in a short stimulating time.

The contributions of this paper are listed as follows:

- We propose enhancement approaches for bit-level transition activity by signal word-level statistical properties management. Specifically, increasing variance and decreasing temporal correlation could effectively increase the transitions of the bits which have low transition activity in a word.
- We propose a traceback approach to determine the statistical properties of primary input signals to increase the transitions of an internal bit from the primary inputs, based on the propagation statistical properties of DSP operations.
- We evaluate the proposed approach with typical DSP circuits, such as FIR and IIR on an FPGA. The experimental results show that rare bits transition activity can be dramatically increased by up to 166.7 times and the Trojan activation time can be reduced by 121 times.

The rest of the paper is organized as follows. Section II briefly introduces related works on HT detection and on bit transition activity analysis. Section III presents the enhancement approaches for bit-level transition activity and a traceback approach to determine the input signals in order

978-1-4799-6245-7/14/$31.00 ©2014 IEEE

to increase the transition activity of internal bits. Section IV shows experiment results and is followed by a conclusion in Section V.

II. BACKGROUND

A. Related Work on Hardware Trojan Detection

Hardware Trojans are modifications to original circuit by adversaries to gain access to confidential information or destroy the circuit. Trojans are usually very small and embedded deeply in their host design, with negligible impacts on the host design's metrics. Therefore, they are very difficult to control and observe from the primary input and output of a circuit [7].

Some existing literatures involve side-channel analysis to detect HTs, based on circuit physical parameters measurement, such as power [2], current [3, 4] and circuit delay [5, 6]. The approach in [2] used transient power analysis to detect HTs. During the analysis, a circuit was divided into multiple regions and at a time only one region was activated while keeping other parts quiet. This helps magnify the Trojan's impact on the circuit transient power in order to improve the efficiency of the side-channel analysis. In [4], authors presented a Trojan isolation method based on localized current analysis to detect HTs in integrated circuits. A test point insertion method was designed to detect HTs in term of gate-level delay characterization in [5]. In [6], path-delays were collected to construct footprints which represent the characteristics of a genuine design. Chips were then validated by comparing their delay parameters to the footprints.

The side-channel analysis approaches face the pressure of long stimulating time. To reduce the time, Trojan activation approaches have been developed. In [7], a sustained vector technique was applied to both genuine circuit and Trojan circuit with the constant primary inputs for several clock cycles, in order to reduce extraneous toggles within the genuine circuit. The power consumption difference between the actual and the Trojan circuitry was then magnified. In their experiments, the technique can reach a stable state where no further change in the state variables occurs. In [8], an approach was proposed to construct a unique probabilistic signature for every functionally distinct circuit. The idea was to build a probability distribution on the input such that the probability distribution of the output is unique. Then, a technique based on hypothesis testing was used to statistically infer the presence of a Trojan. The approach in [9] first identified nets with transition probability less than a specific threshold. Then dummy scan flip-flops (DSFF) were inserted to these nets to augment their transition probabilities, and thus increase activity of HTs in circuits. This method introduced area overhead due to the inserted dummy flip-flops. The proposed methodology in [10] is conceptually similar to *N*-detect test used in stuck-at ATPG (automatic test pattern generation), where test set is generated to detect each single stuck-at fault in a circuit by at least *N* different patterns.

The previous HT activation approaches introduced above were based on the bit-level logic properties of circuits. In this paper, we present the first approach to increase the probability of Trojan detection with reducing time based on word-level statistical properties of circuits. The word-level signal analysis and management are more convenient than bit-level with reduced complexity. Characteristics of various Trojan activation techniques can be seen in Table I.

B. Preliminaries about Transition Activity

Since HTs are associated with rare bits in a circuit, the rare bits are the key to HT detection. A rare bit in a circuit is defined as a bit with low 0-1 transition activity.

The transition activity of a bit in DSP hardware circuits can be analyzed based on its word-level signal statistical description [11, 12]. In [11] and [12], the transition activity was investigated to estimate power consumption.

Let $x(n)$ be a *B*-bit word signal, and $b_i(n)$ is the value of the ith bit at the time index n. This signal can be modeled by ARMA [11] with mean μ, variance σ^2, and temporal autocorrelation ρ. In the rest of this paper, all signals are from the ARMA model. The transition activity (t_i) of the ith bit of $x(n)$ can be calculated as [11]

$$t_i = 2p_i(1-p_i)(1-\rho_i) \tag{1}$$

where p_i is the probability of the ith bit being 1, and ρ_i is the temporal autocorrelation of the ith bit, *i.e.,* the correlation between $b_i(n)$ and $b_i(n-1)$. Therefore, the transition activity is closely related to p_i and ρ_i. The approach [9] increases the transition activity from the aspect of p_i, while our approach focuses on ρ_i.

In [11] and [12], the bits in a word is partitioned into three regions: LSB (least significant bits), linear, and MSB (most significant bits) in terms of the temporal autocorrelation. The breakpoints BP_0 and BP_1 separating the LSB from the linear region and the linear from MSB could be calculated using the word-level statistical properties (μ, σ, ρ), respectively.

For a specific signal $x(n)$, the BP_0 is calculated as [12]

$$BP_0 = [\log_2\sigma + \log_2(\sqrt{1-\rho^2} + |\rho|/8)] \tag{2}$$

where $[k]$ is the number nearest to k. For BP_1, there are two calculation equations as below in [11] and [12], respectively.

$$BP_1 = [\log_2(6\sigma)] \tag{3}$$

$$BP_1 = [\log_2(|\mu| + 3\sigma)] \tag{4}$$

When $|\mu| \leq 3\sigma$, both (3) and (4) are approximately equal. The bits in the LSB region have zero temporal correlation and the bits in the MSB region have high temporal correlation. According to (1), in a word the bits in the LSB region have high transition activity, while the bits in the MSB region have low transition activity. This is in line with practical observations. Therefore, the bits in the MSB region are considered as rare bits in a word.

The basic idea of our approach is to move the bits from the MSB region to the LSB region, and decrease the temporal correlation, in order to increase transition activity of rare bits. This is a different application of word-level signal statistical properties from [11], where transition activity analysis based

TABLE I
COMPARISON OF VARIOUS TROJAN ACTIVATION TECHNIQUES.

Technique	Description	Operation level
[7]	A sustained vector technique magnifying the power ratio in the Trojan circuitry.	Bit level
[8]	HT detection based on probability distribution of circuit's input and output.	Bit level
[9]	Insertion of dummy flip-flops to increase the transition activity.	Bit level
[10]	Multiple excitation of rare logic conditions at internal nodes to detect HT using test patterns.	Bit level
Our Approach	Transition enhancement of rare bits based on word-level signal statistics management.	Word level

on word-level signal statistical description was used for power consumption estimation. In the next, the proposed approach will be presented in details.

III. Proposed approach for transition activity enhancement

In this section, we will present our proposed approaches to increase transition activity of rare bits. To start with, we present techniques for increasing transition activity of rare bits by use of word-level signal statistical properties. Our final goal is to increase the transition activity of internal rare bits from the primary input of a circuit. This is because in practice the primary input is the only port to access the internal space of a circuit. Therefore, the key to increase the transitions of the internal rare bits is to control the internal signal statistical properties from the primary input. To do so, in the second part of this section, we present the propagation statistical properties of DSP operations and the conditions that the signals must meet in order to guarantee the monotonous propagation of signal statistical properties. Finally, a traceback approach will be introduced, which determines the statistical properties of desired primary input signals. Inputting the new signal to the circuit will increase the number transitions of rate bits.

A. Bit-level transition activity enhancement approaches

Based on the theories in [11] and [12], we further partition the bits in a word into four regions in terms of ρ as below.

$$\rho_i = \begin{cases} 0 & i < BP_0 \\ \frac{(i-BP_0+1)\rho}{BP_1-BP_0} & BP_0 \le i \le BP_1 - 1 \\ \rho & BP_1 - 1 < i < BP_2 \\ 1 & i \ge BP_2 \end{cases} \tag{5}$$

where i is the bit position, ρ is the word-level temporal autocorrelation, BP_0 and BP_1 are calculated according to (2)-(4), and BP_2 can be calculated by computing the common most significant bits in the binary representations of the maximum and minimum numbers of $x(n)$ in its dynamic range [11].

From (5) we can see that there are two ways to increase the transition activity of the rare bit $i, (BP_1 - 1 < i < BP_2)$. First, reducing the word-level temporal correlation ρ will reduce ρ_i of the rare bit i and thus increase t_i. Second, shifting the BP_0 and BP_1 positions towards the BP_2 will move the rare bits originally in the MSB region to the LSB region. In this way, ρ_i is reduced and t_i is increased. From (2) and (3), we find that a larger word-level signal variance σ will shift BP_0 and BP_1 closer to the MSB position. Particularly, if we want a rare bit i to be in the linear region, *i.e.,* $i < BP_1$, then solving (3) with the specific i will give us the needed σ.

$$\frac{1}{6} \cdot 2^{i-0.5} < \sigma \tag{6}$$

Similarly, solving (2) with $i \le BP_0$ will move the ith bit into the LSB region.

$$2^{i-0.5-\log_2(\sqrt{1-\rho^2}+|\rho|/8)} \le \sigma \tag{7}$$

B. Signal propagation property and conditions

From the description in the previous section, we know that increasing σ and decreasing ρ at the word level can increase the transition activity of a rare bit. Therefore, an intuitive way to increase the transitions of an internal rare bit is to increase σ or decrease ρ of the word signal, to which the rare bit belongs, from the primary input. Our approach exploits the word-level propagation properties of signals to control σ monotonous increase and ρ monotonous decrease. In this way, increasing σ or decreasing ρ of the primary input signals will have direct impact on the internal signals. For an operation

$$x_3(n) = \text{op}(x_1(n) \cdot x_2(n)) \tag{8}$$

we have the following two requirements to make sure that the output's σ is not smaller than that of the inputs, or the output's ρ is not greater than that of the inputs.

Requirement I:

$$\sigma_3 \ge \sigma_1 \tag{9}$$

Requirement II:

$$\rho_3 \le \rho_1 \tag{10}$$

where σ_1 and ρ_1 are the variance and temporal correlation of the input $x_1(n)$, σ_3 and ρ_3 are for the output $x_3(n)$. We consider four operators widely used in DSP algorithms [11], which are shown in Fig. 1. To meet Requirements I and II, each operator has its own conditions. We present the conditions separately in the next.

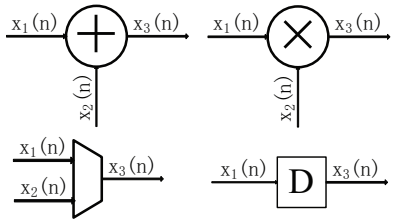

Fig. 1. Four operators considered in this paper: adder, multiplier, multiplexer and delay.

1) Adder: The μ_3, σ_3 and ρ_3 of the output signal $x_3(n)$ of the adder can be computed as the following [13]:

$$\mu_3 = \mu_1 + \mu_2 \tag{11}$$

$${\sigma_3}^2 = {\sigma_1}^2 + 2\rho_{x_1x_2}\sigma_1\sigma_2 + {\sigma_2}^2 \tag{12}$$

$$\rho_3 = \frac{\rho_1{\sigma_1}^2 + \rho_2{\sigma_2}^2 + 2\rho_{x_1x_2}\sigma_1\sigma_2}{{\sigma_1}^2 + 2\rho_{x_1x_2}\sigma_1\sigma_2 + {\sigma_2}^2} \tag{13}$$

where $(\mu_1, \sigma_1, \rho_1)$ and $(\mu_2, \sigma_2, \rho_2)$ are the statistical properties of the inputs $x_1(n)$ and $x_2(n)$, and $\rho_{x_1x_2}$ and $\rho_{x_2x_1}$ are the cross-correlation between $x_1(n)$ and $x_2(n)$.

To satisfy Requirement I, the following condition should be met.

$$\begin{cases} \text{Universal} \quad \text{Approval}, & \text{if} \quad 0 \le \rho_{x_1x_2} \le 1; \\ \sigma_2 \ge 2\sigma_1, & \text{if} \quad -1 \le \rho_{x_1x_2} < 0. \end{cases} \tag{14}$$

If $\rho_{x_1x_2}$ is in between 0 and 1, Requirement I can always be satisfied. If $\rho_{x_1x_2}$ is in between -1 and 0, then the condition is $\sigma_2 \ge 2\sigma_1$. That means if we want to increase σ_3, we should raise the value of σ_1 as well as meet the condition in (14).

Proof: According to (12), if $0 \le \rho_{x_1x_2} \le 1$, then certainly $\sigma_3 \ge \sigma_1$, so Requirement I is satisfied. If $-1 \le \rho_{x_1x_2} \le 0$, then to satisfy Requirement I it requires

$$\sigma_2 \ge -2\rho_{x_1x_2}\sigma_1. \tag{15}$$

Therefore, $\sigma_2 \ge 2\sigma_1$ guarantees Requirement I in this case.

To satisfy Requirement II, the relationship between two input signals should satisfy the following condition:

$$\rho_1 \ge \rho_2 \tag{16}$$

Proof: According to (13),

$$\rho_3 \le \rho_1 \cdot \frac{{\sigma_1}^2 + \rho_2/\rho_1 \cdot {\sigma_2}^2 + 2\rho_{x_1x_2}\sigma_1\sigma_2}{{\sigma_1}^2 + 2\rho_{x_1x_2}\sigma_1\sigma_2 + {\sigma_2}^2} \tag{17}$$

Obviously, if $\rho_1 \ge \rho_2$, then $\rho_3 \le \rho_1$. Requirement II is reached.

2) Multiplier: In this subsection, we will present the conditions to meet Requirement I and Requirement II for multiplier. The proof process is similar to adder and is ignored here. The μ_3, σ_3 and ρ_3 of the output signal $x_3(n)$ of the multiplier are calculated as the following [13]:

$$\mu_3 = \mu_2 \cdot \mu_1 \tag{18}$$

$${\sigma_3}^2 = ({\sigma_1}^2 + {\mu_1}^2) \cdot ({\sigma_2}^2 + {\mu_2}^2) - {\mu_3}^2 \tag{19}$$

$$\rho_3 = \rho_1 \cdot \rho_2 \tag{20}$$

Conditions for the multiplier to satisfy Requirement I are:

$$\begin{cases} \mu_1 \le -1, \mu_2 \ge 1, \quad \text{or} \quad \mu_1 \cdot \mu_2 = 0, \\ \text{if} \quad \rho_{{x_1}^2{x_2}^2} \ge 0 \quad \text{and} \quad \rho_{x_1x_2} \ge 0; \\ \mu_1 \ge 1, \quad \mu_2 \ge 1, \quad \text{or} \quad \mu_1 \cdot \mu_2 = 0, \\ \text{if} \quad \rho_{{x_1}^2{x_2}^2} \ge 0 \quad \text{and} \quad \rho_{x_1x_2} \le 0. \end{cases} \tag{21}$$

Here we only consider the case of $\rho_{{x_1}^2{x_2}^2} \ge 0$, because it is difficult to derive a simple condition for the case of $\rho_{{x_1}^2{x_2}^2} \le 0$. More importantly, $\rho_{{x_1}^2{x_2}^2} \ge 0$ is easy to achieve (*e.g.,* setting one of the inputs as a constant).

For multiplication, Requirement II is always satisfied because of $\rho_3 = \rho_1 \cdot \rho_2 < \rho_1$.

3) Multiplexer: The μ_3, σ_3 and ρ_3 of the output signal $x_3(n)$ of the multiplexer can be calculated as [11]:

$$\mu_3 = (1 - p_c)\mu_1 + p_c\mu_2 \tag{22}$$

$$\begin{aligned} {\sigma_3}^2 = {} & (1 - p_c){\sigma_1}^2 + p_c(1 - p_c){\mu_1}^2 + p_c{\sigma_2}^2 \\ & + p_c(1 - p_c){\mu_2}^2 - 2p_c(1 - p_c)\mu_1\mu_2 \end{aligned} \tag{23}$$

$$\rho_3 = \frac{E[x_3(n)x_3(n-1)] - {\mu_3}^2}{{\sigma_3}^2} \tag{24}$$

where p_c is probability of the control signal being 1, and $E[x_3(n)x_3(n-1)]$ can be computed based on p_c and the mean of the time-delayed product between $x_1(n)$ and $x_2(n)$.

Condition for the multiplexer to meet Requirement I is

$$({\sigma_1}^2 - {\sigma_2}^2) \le (1 - p_c)(\mu_1 - \mu_2)^2 \tag{25}$$

For Requirement II, when a signal propagates through the multiplexer, the output signal is one of the input signals under the probability p_c. Therefore, the output correlation will be smaller than the correlation of the input signals.

4) Delay: A delay module shifts the signal by one clock cycle. Therefore, the signal's statistical properties at the output of a delay module are identical to that in the input.

The derived conditions for above operators guarantee that the statistics of the output signal of the operators meet the requirements in (9) and (10). Then, for each operator, given the expected statistical properties of the output signal, the statistical properties of the input signals can be derived by solving a set of equations, including the property definition equations and the conditions. Taking the adder as an example, we need to solve equations (11)-(14) and (16) to determine the input signal's (μ, σ, ρ).

For a datapath composed of multiple such operators, if we can find the primary input signals which make the conditions for all operators satisfied, the primary input signals can control the transition activities of the circuit internal nodes. In the next, we will propose an approach to find such signals.

C. Traceback approach

Based on the signal propagation properties described in the previous section, we use a traceback approach to determine the primary input signals with $(\mu^*, \sigma^*, \rho^*)$, so that the number of transitions of certain rare bits is increased.

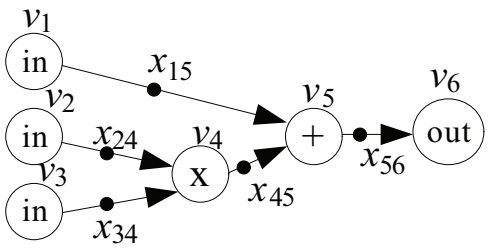

Fig. 2. An example of DFG, where there are six vertices and five word-level signal nodes.

The approach works on a data flow graph (DFG) of a circuit composed of DSP operators. The range of the original primary input signals with the statistical properties (μ, σ, ρ) is given.

The DFG $G(V, E)$ is a set of vertices $v_i \in V$ and a set of edges $e_{ij} \in E$. Vertex v_i represents a primary input/output port or an DSP operator i and e_{ij} shows data flow from v_i to v_j. Define a node x_{ij} on each edge e_{ij} as a word-level signal coming out from v_i and going to v_j. Fig. 2 shows an example of such DFG. In the example, there are three primary inputs, two DSP operators and one primary output. We define five word-level signal nodes, $X = (x_{15}, x_{24}, x_{34}, x_{45}, x_{56})$. Among them, the first three are primary input signals.

The approach first propagates the original primary input signals forward along the DFG, to obtain the statistical properties of the internal signals, *e.g.,* x_{45} and x_{56} in Fig. 2. Then, the four word-level temporal correlation regions can be obtained for each signal node according to (5). Afterwards, the temporal correlation ${\rho_{ij}}^k$ of the kth bit of signal x_{ij} is determined. Sorting all ${\rho_{ij}}^k$ in the non-ascending order will reveal the rare bits. Higher the temporal correlation is, lower the transition activity is. Next, if the kth bit of signal x_{ij} is selected to be the target of transition activity enhancement, the approach will carry out a traceback process as follows.

1) For signal node x_{ij}, determine the expected σ_{ij} or ρ_{ij} through accurate calculation according to (5)-(7), so that ${\rho_{ij}}^k$ will be reduced.

2) Depending on the type of operator v_i which outputs signal x_{ij}, find a set of possible values (μ, σ, ρ) of the input signals of v_i so that σ_{ij} and ρ_{ij} can be realized. This involves solving a set of equations according to the property definitions and conditions described in Section III-B. To reduce the search complexity, we always try to set one of the inputs as a constant and use another input to control the statistical properties of the output signal. For example, in Fig.2 we can set the primary inputs x_{24} and x_{34} to be constants, and find (μ, σ, ρ) of x_{15} to increase the transitions of the rare bits in the word x_{56}.

3) Continue the process backwards along the DFG until the input vertices are reached. In this way, we can obtain $(\mu^*, \sigma^*, \rho^*)$ for the desired primary input signal. Note that the other primary signals may be fixed as constants.

During the process, there could be conflicts if the DFG contains a reconvergent path. In this case, the conditions required by several outgoing signals on the input signal may be contradicted each other. When this kind of conflict occurs, the approach will not force the inputs to be constant, and try different conditions on different paths. Especially, if we cannot meet the conditions for monotonous control of the variance, we switch to control the autocorrelation, because the conditions for the latter are less strict. Furthermore, the worst case complexity of the traceback approach is $O(2^{\frac{L(L+1)}{2}})$, where L is the number of logic levels from the internal bit to the input. The complexity is high for large and complex designs and optimizing the approach to reduce the complexity is our future work. Up to now, we have presented our proposed approach for the transition activity enhancement of internal rare bits. In the next, we will evaluate the approach in terms of its efficiency in increasing the number of transitions and reducing HT activation time.

TABLE II
PRIMARY INPUT SIGNALS FROM THE ARMA MODEL.

Signal	μ	σ	ρ	BP_1position
SIG1	0	10	0.99	6th bit
SIG2	0	10	0.40	6th bit
SIG3	0	10	0.10	6th bit
SIG4	0	1000	0.99	13th bit
SIG5	0	1000	0.70	13th bit
SIG6	0	1000	0.40	13th bit
SIG7	16384	1000	0.99	13th bit
SIG8	-16384	1000	0.99	13th bit
SIG9	0	3000	0.99	14th bit
SIG10	16384	3000	0.99	14th bit

IV. EXPERIMENTAL RESULTS

To evaluate the proposed approach, three typical DSP circuits are used, including FIR, IIR and multiplication-addition (MA). All three circuits are implemented on FPGA and simulated using Modelsim. For the FIR and IIR implementations, the word length is 16 bits. For MA, the implementation uses 32-bit word length. Ten ARMA signals with different statistical properties are used as primary input signals as in [11], and are shown in Table II.

The experimental results presented next include three parts. The first part demonstrates the relationship between the word-level statistical properties and the bit-level transition activity. The second part evaluates the efficiency of the proposed approach to enhance the transition activity of internal rare bits in the three benchmark circuits. The number of transitions shown below is counted during simulation in 1024 clock cycles. Finally, it is shown that the HT activation time can be reduced by the proposed approach.

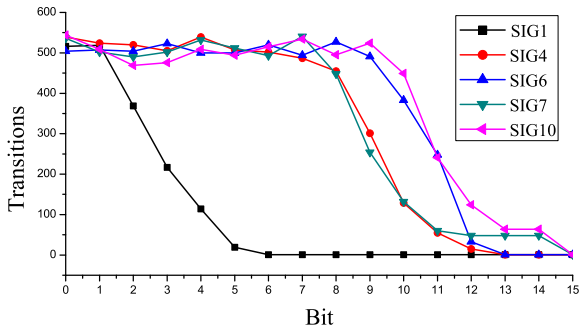

Fig. 3. Relation between the word-level statistical properties and the bit-level transition activity.

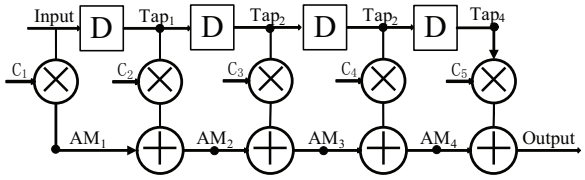

Fig. 4. Five-tap FIR filter architecture.

A. Results of effect of word-level statistical properties on bit-level transition activity

Table II lists ten different signals from the ARMA model. Columns 2-4 show the mean, the variance and the temporal autocorrelation of the signals. The different signals have different values for the three statistical properties. The property difference lead to the different transition activities of the bits in the signal word.

Firstly, the last column of Table II reports the BP_1 position of each signal. Remember that the bits in the position greater than or equal to BP_1 are rare bits, mentioned in Section III-A. Therefore, lower the BP_1 position is, more bits in a word have low transition activity. In Table II, the property which makes the BP_1 position various is σ. As σ increases, the BP_1 position moves closer to MSB.

Secondly, Fig. 3 illustrates the number of transitions of each bit in the signal word. From SIG1 to SIG4, σ increases from 10 to 1000, leading to large transition increases of bits 5-8. The number of transitions of the 5th bit increases sharply from 19 to 507 in 1024 clock cycles. From SIG4 to SIG7, μ increases. However, the number of transitions is not affected much. From SIG4 to SIG6, ρ decreases, leading to remarkable transition increases of bits 9-11. From SIG4 to SIG10, σ further increases from 1000 to 3000, also resulting in similar transition increases of bits 9-11. These observations demonstrate that increasing σ and decreasing ρ can effectively increase the transition activities of some rare bits in signal word.

B. Results of transition activity enhancement

In this section, we use the three benchmark circuits to evaluate the traceback approach, which increases the transition activity of internal bits from the primary input with the determined input signals.

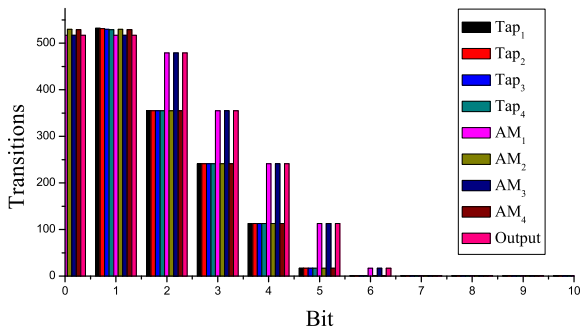

Fig. 5. Bit Transition Activity of nine internal signal nodes in the FIR implementation with the primary input signal SIG1.

To start with, we show the transition activity of the internal bits in the five-tap FIR filter. The FIR filter architecture is shown in Fig. 4, where the coefficients are set as C_1 = C_5 = 2, C_2 = C_4 = -1 and C_3 = 1. Nine internal signal nodes are shown in the figure. Fig. 5 plots the number of transitions of each bit of the nine signal nodes when the primary input signal to the FIR filter is SIG1. From this figure, we can clearly find the rare bits.

Let select the 6th bit of the output signal to enhance its transition activity by use of our approach. With SIG1 as the input to the FIR, the bit has only 19 transitions in 1024 clock cycles. Assume we want to move the bit to the LSB region to increase its transition activity. Then, let $i = 6$ and $\rho = 0.99$ in (7) will obtain the range of the standard deviation of the output signal $\sigma_{out} \geq 169$. Moreover, by using the proposed traceback approach, we can obtain the relation between the output signal properties and the primary input signal's properties of the FIR implementation as $\mu_{out} = 3\mu_{in}$, $\sigma_{out} = 3\sigma_{in}$ and $\rho_{out} = \rho_{in}$. Note that this relation is obtained based the FIR architecture in Fig. 4. From this relation, we can obtain the properties of the primary input signal, which are $\mu_{in} = 0$, $\sigma_{in} \geq 57$, and $\rho_{in} = 0.99$. Hence, SIG4 and SIG9 in Table I satisfy these requirements. We input SIG4 and SIG9 to the FIR and measure the number of transitions of the 6th bit of the output signal. The results are shown in Fig. 6. With SIG4 the number of transitions of the 6th bit (which is bit 5 in the figure) is increased to 489, while with SIG9 the number of transitions is 507. We effectively increase the transition activity of the 6th bit of the output signal by 27 times. Apart from the 6th bit, the number of transitions of the 7th-10th bits are also increased, due to the large σ_{in}.

Moreover, another way to increase the transitions is decreasing ρ of the output signal. As shown in (5), the lower the word-level autocorrelation is, the higher the bit transition activity is. Therefore, we expect the properties of the primary input signal to be $\mu_{in} = 0$, $\sigma_{in} = 10$ and ρ as smaller as better. We choose SIG2 and SIG3 from Table I to input to the FIR. The results are also shown in Fig. 6. The number of transitions of the 6th bit grows from 19 to 93 under SIG2

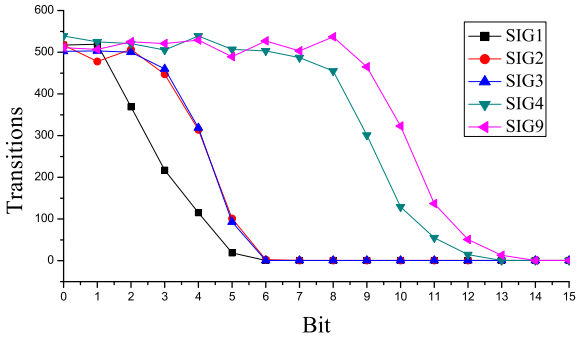

Fig. 6. Transition activity enhancement in the output node of FIR with different input signals. As the primary input signal's standard deviation increases and autocorrelation decreases, the number of transitions of the 6th bit significantly increases.

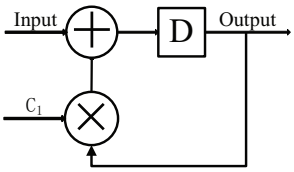

Fig. 7. IIR filter architecture with C_1=0.25.

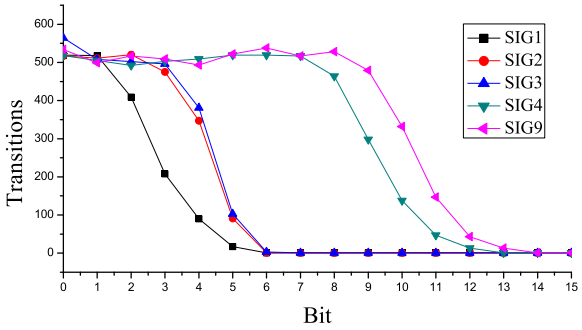

Fig. 8. Transition activity enhancement in the output node of IIR with different input signals. As the primary input signal's standard deviation increases and autocorrelation decreases, the number of transitions of the 6th bit significantly increases.

and to 101 under SIG3, with the decrease of ρ_{in} from 0.99 to 0.1. For the IIR filter shown in Fig. 7, similarly applying the proposed approach to increase the transition activity of the 6th bit of the output signal leads to the results shown in Fig. 8. Inputting SIG9 with increased standard deviation to the IIR increases the number of transitions of the bit from 17 to 531 (31×). With SIG2 as the input, the number of transitions is increased to 103 (6×).

The MA architecture is shown in Fig. 9. It contains an adder, two multipliers and two multiplexers. Firstly, we input SIG4 and SIG5 to $x_1(n)$ and $x_2(n)$, respectively, to calculate the statistical properties of the internal nodes.

For the output signal node, the 25th bit is the rare bit as calculated from (3). To move the 25th bit to the LSB region, the standard deviation of the output node should be $\sigma_{out} \geq 177102$. Then, we use our traceback approach to

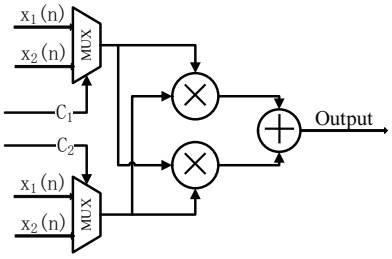

Fig. 9. The MA architecture.

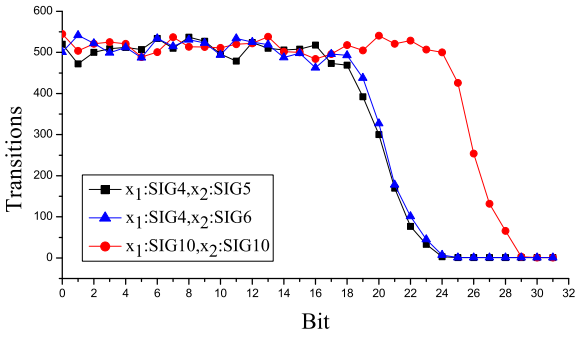

Fig. 10. Transition activity enhancement in the output node of MA with different input signals. As the primary input signals' standard deviation increases and autocorrelation decreases, the number of transitions of the 26th bit significantly increases.

find the appropriate primary input signals. By calculation, we choose $x_1(n)$ =SIG10 and $x_2(n)$ =SIG10 to move the 25th bit into the LSB region. Results can be seen in Fig. 10. The number of transitions of the 25th bit is significantly increased from 3 to 500. To decrease the temporal autocorrelation, signals $x_1(n)$ =SIG4 and $x_2(n)$ =SIG6 are chosen. The increase of the number of transitions is small from 3 to 13.

The above experimental results on the three circuits demonstrate that the proposed approach can effectively increase the transition activity of the internal rare bits by controlling the primary input signals. It shows that the impact of increasing the standard deviation is more significant.

C. Results of trojan activation time reduction

In this subsection, we evaluate the proposed approach for accelerating the HT activation time. It is assumed that an internal rare bit in the three circuits is used to trigger the HT. To keep the results consistent, we also select the 6th bit of the output signal in FIR and IIR and the 26th bit of the output signal in MA as the trigger. We count the number of transitions of the trigger bit in the 1024 clock cycles before and after applying our proposed approach. Also, HT activation time is measured as the number of clock cycles that the trigger bit needs to make 1000 transitions.

The results are reported in Table III. The first column indicates the names of the three benchmark circuits. Columns 2 and 3 show the number of transitions of the trigger bit and the number of activation clock cycles needed before the proposed approach is applied. Columns 4 and 5 list the

TABLE III
RESULTS OF HT ACTIVATION TIME REDUCTION. #TRANS: THE NUMBER TRANSITIONS IN 1024 CLOCK CYCLES. #CYCLES: THE NUMBER OF CYCLES NEEDED FOR A RARE BIT MAKING 1000 TRANSITIONS.

Circuit	Before application of our approach		After application of our approach	
	#Trans	#Cycles	#Trans	#Cycles
FIR	19	56846	541 (28.5×)	1908 (29.8×)
IIR	17	64011	531 (31.2×)	1929 (33.2×)
MA	3	511480	500 (166.7×)	4227 (121×)

numbers after application of our approach. We can see that our approach can significantly increase the transitions up to 166.7 times and effectively reduce Trojan activation time by up to 121 times for 1000 transitions.

V. CONCLUSION

We present a novel approach to accelerate the activation of HTs in hardware circuit with DSP operators. The approach exploits the relation between the word-level signal statistical properties and the bit-level transition activity to increase the transitions of rare bits. A traceback procedure is presented to find the statistical properties of the primary input signals based on a set of conditions, which ensure the monotonous propagation of the signal statistical properties through the circuits. In this way, the internal bits' transition activity can be controlled by the primary input signals. We evaluate the proposed approach on three typical DSP circuits implemented on FPGA. The experimental results demonstrate that the proposed approach can effectively increase the number of transitions of internal circuit bits by up to 166.7 times in 1024 clock cycles and reduce Trojan activation time by up to 121 times.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China under Grand No. 61204022 and the Natural Science Foundation of Tianjin under Grand No. 12JCYBJC30700.

REFERENCES

[1] M. Tehranipoor and F. Koushanfar, "A survey of hardware trojan taxonomy and detection," *Design Test of Computers, IEEE*, vol. 27, no. 1, pp. 10–25, Jan 2010.

[2] H. Salmani, M. Tehranipoor, and J. Plusquellic, "A layout-aware approach for improving localized switching to detect hardware trojans in integrated circuits," in *IEEE International Workshop on Information Forensics and Security (WIFS)*, Dec 2010, pp. 1–6.

[3] Y. Cao, C.-H. Chang, and S. Chen, "Cluster-based distributed active current timer for hardware trojan detection," in *IEEE International Symposium on Circuits and Systems (ISCAS)*, May 2013, pp. 1010–1013.

[4] X. Wang, H. Salmani, M. Tehranipoor, and J. Plusquellic, "Hardware trojan detection and isolation using current integration and localized current analysis," in *IEEE International Symposium on Defect and Fault Tolerance of VLSI Systems(DFTVS)*, Oct 2008, pp. 87–95.

[5] S. Wei, K. Li, F. Koushanfar, and M. Potkonjak, "Provably complete hardware trojan detection using test point insertion," in *IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, Nov 2012, pp. 569–576.

[6] Y. Jin and Y. Makris, "Hardware trojan detection using path delay fingerprint," in *IEEE International Workshop on Hardware-Oriented Security and Trust (HOST)*, June 2008, pp. 51–57.

[7] M. Banga and M. Hsiao, "A novel sustained vector technique for the detection of hardware trojans," in *22nd International Conference on VLSI Design*, Jan 2009, pp. 327–332.

[8] S. Jha and S. Jha, "Randomization based probabilistic approach to detect trojan circuits," in *11th IEEE High Assurance Systems Engineering Symposium (HASE)*, Dec 2008, pp. 117–124.

[9] H. Salmani, M. Tehranipoor, and J. Plusquellic, "A novel technique for improving hardware trojan detection and reducing trojan activation time," *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, vol. 20, no. 1, pp. 112–125, Jan 2012.

[10] R. Chakraborty, F. Wolff, S. Paul, C. Papachristou, and S. Bhunia, "Mero: A statistical approach for hardware trojan detection," in *Cryptographic Hardware and Embedded Systems - CHES 2009*. Springer Berlin Heidelberg, 2009.

[11] S. Ramprasad, N. Shanbha, and I. Hajj, "Analytical estimation of signal transition activity from word-level statistics," *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on*, vol. 16, no. 7, pp. 718–733, Jul 1997.

[12] P. Landman and J. Rabaey, "Architectural power analysis: The dual bit type method," *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, vol. 3, no. 2, pp. 173–187, June 1995.

[13] E. Kyriakis-Bitzaros and S. Nikolaidis, "Estimation of bit-level transition activity in data-paths based on word-level statistics and conditional entropy," *Circuits, Devices and Systems, IEE Proceedings* -, vol. 149, no. 4, pp. 234–240, Aug 2002.