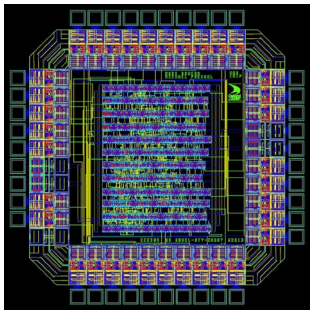
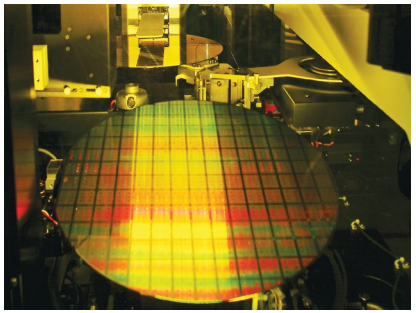
Chapter 5
Hardware Trojan Classification

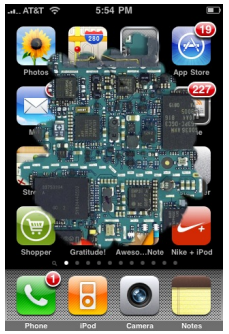
Intentional Trust Issues on ASIC



IC Design



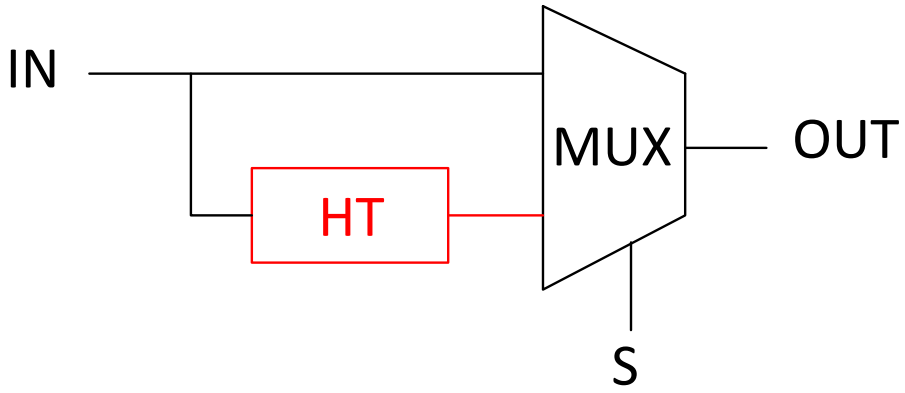
Fabrication



Deployment

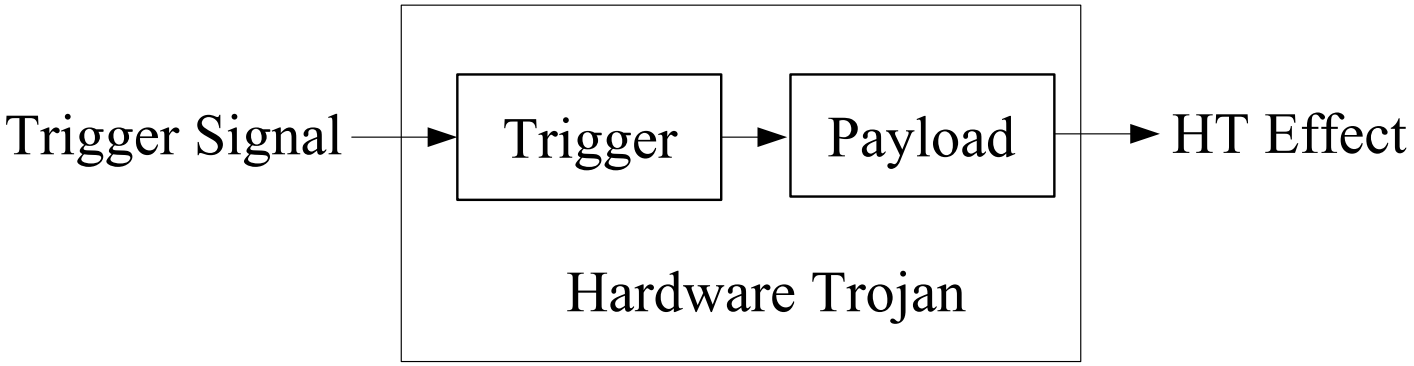
Introduction

- A Hardware Trojan is a malicious and deliberately stealthy modification made to an electronic device such as an IC.



A simple hardware trojan

HT Structure



- The trigger logic monitors HT trigger signal which is a set of signals activating HT circuit thereby with payload effected.
- To maintain stealthy nature of HT in testing mode, the fundamental in this case is that trigger signal has low controllability while HT effect has low observability, that means trigger signal should be relatively very rare case compare to other signals on chip.
- The payload of an HT is the entire activity that the Trojan executes when it is triggered

HT Insertion in IC Life Cycle

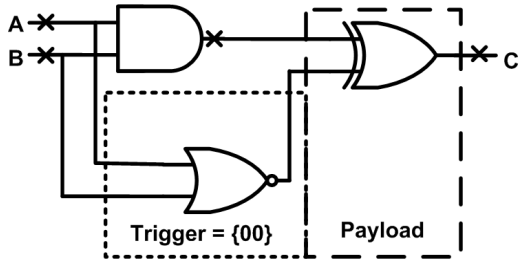
1. Specification
   - Function, size, power, delay, etc.
2. Design
   - Altered by 3rd party CAD tools
     - Functional, logical, timing, physical constraints, etc.
   - 3rd party IP
3. Fabrication
   - Replace wafer mask set
   - Alter chemical composition to increase the electromigration in clock grids → accelerate failures.

HT Insertion in IC Life Cycle

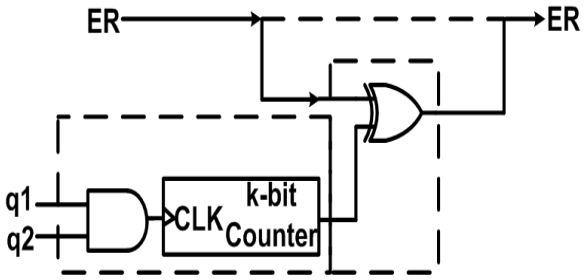
4. Assembly
   - Tested chips & other components are assembled on PCB.
     - Interface → possible trojan insertion
     - E.g. unshielded wires → electromagnetism
5. Testing
   - Not related to Trojan detection
   - Test vectors should be kept secret, the test vectors will be faithfully applied, and the specified actions (accept/reject, binning) will be faithfully followed.

HW Trojan Examples / Models

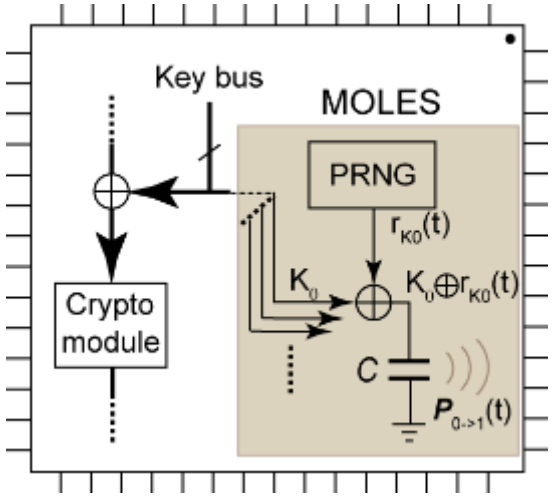
Comb. Trojan Example



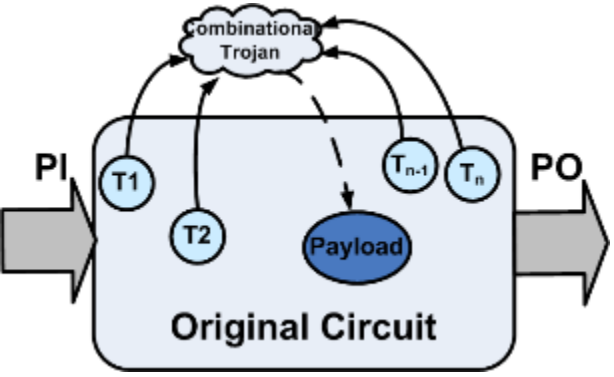
Seq. Trojan Example



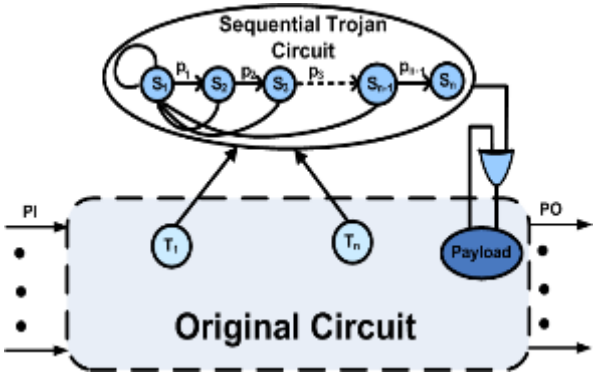
MOLES: Info Leakage Trojan*



Comb. Trojan model



Seq. Trojan Model

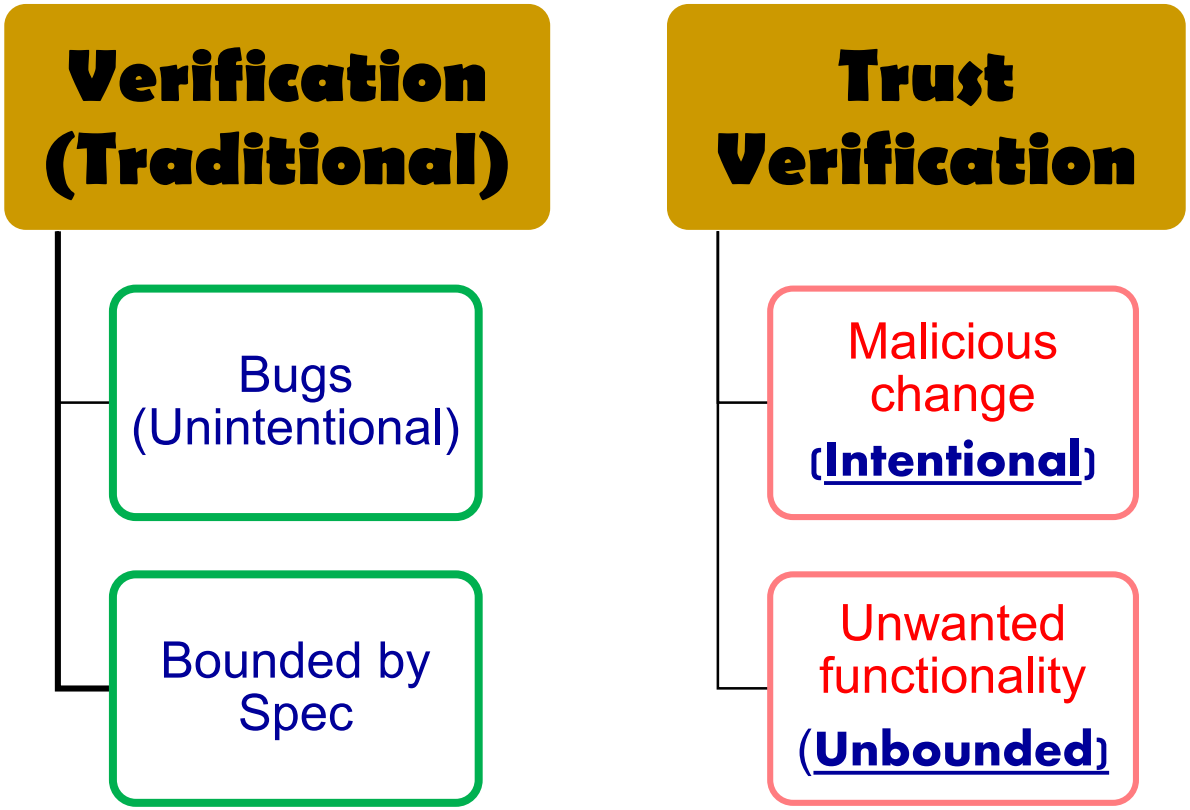


**Lin et al, ICCAD 2009*

HW Trojan evidence!

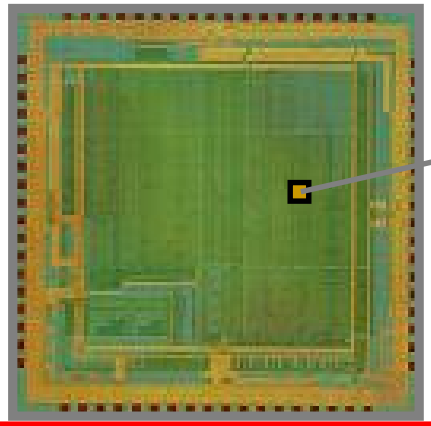
Why is detection of hardware Trojans very difficult?

Bug vs. Malicious Change



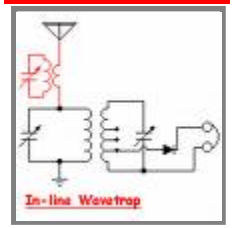
Trojan Attacks → BIGGER verification challenge!

Silicon Back Door



Untrusted Hardware

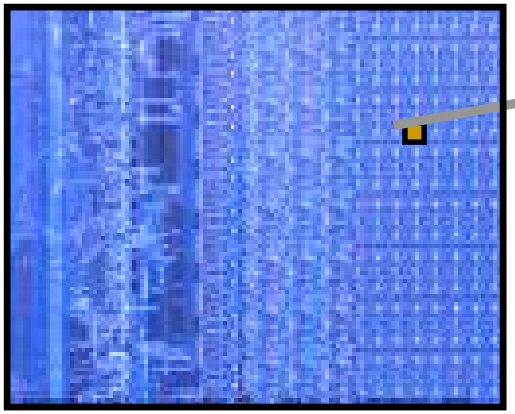
Antenna



- Adversary can send and receive secret information
- Adversary can disable the chip, blowup the chip, send wrong processing data, impact circuit information etc.

- Adversary can place an Antenna on the fabricated chip
- Such Trojan cannot be detected since it does not change the functionality of the circuit.

Silicon Time Bomb



Untrusted Hardware

Counter

Finite state machine (FSM)

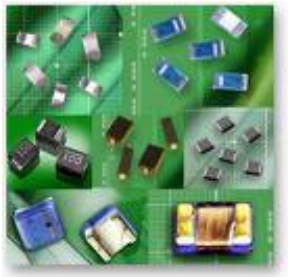
Comparator to monitor key data

Wires/transistors that violate design rules

- **Such Trojan cannot be detected since it does not change the functionality of the circuit.**
- **In some cases, adversary has little control on the exact time of Trojan action**
- **Cause reliability issue**

Applications and Threats

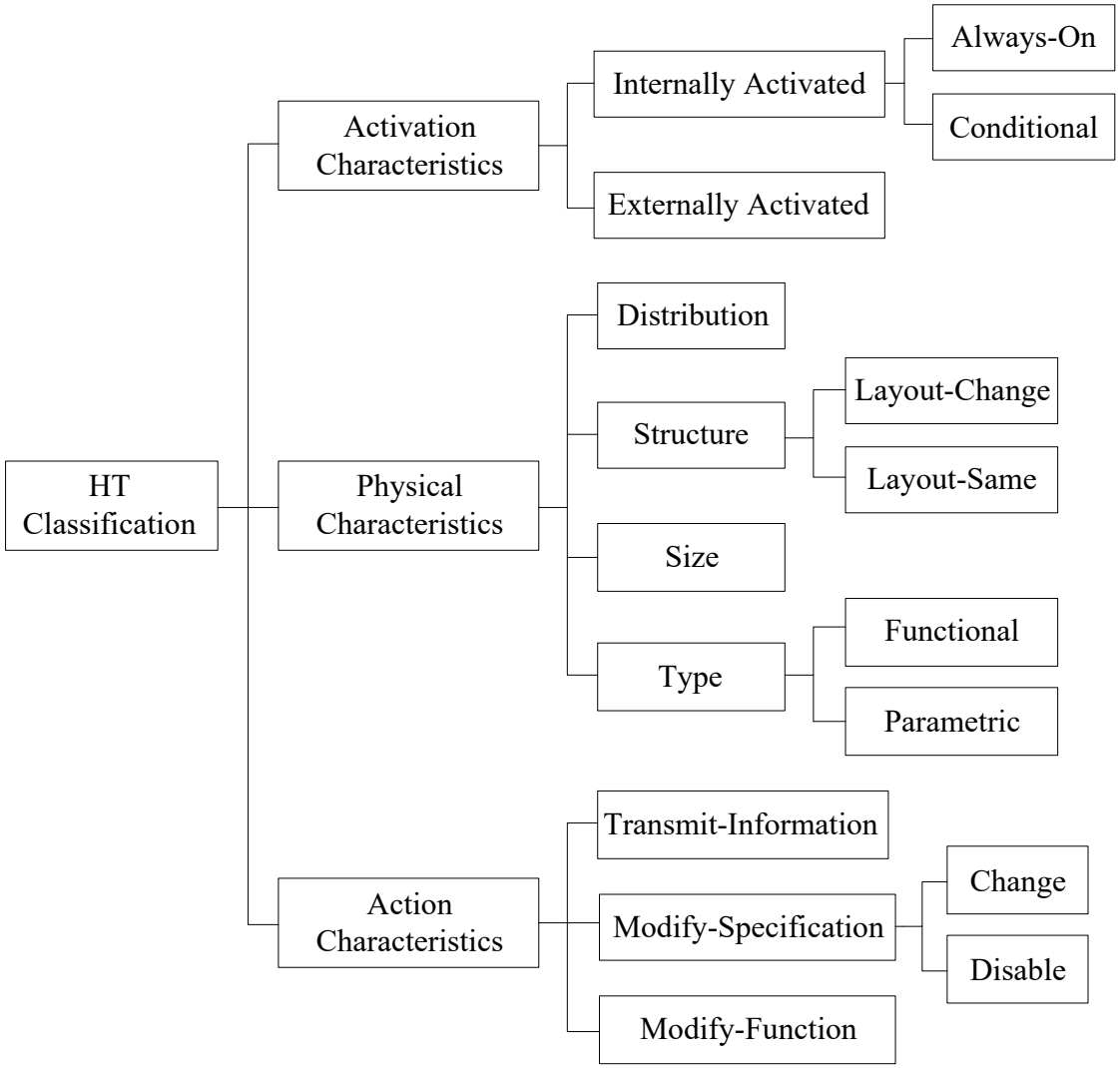
Thousands of chips are being fabricated in untrusted foundries



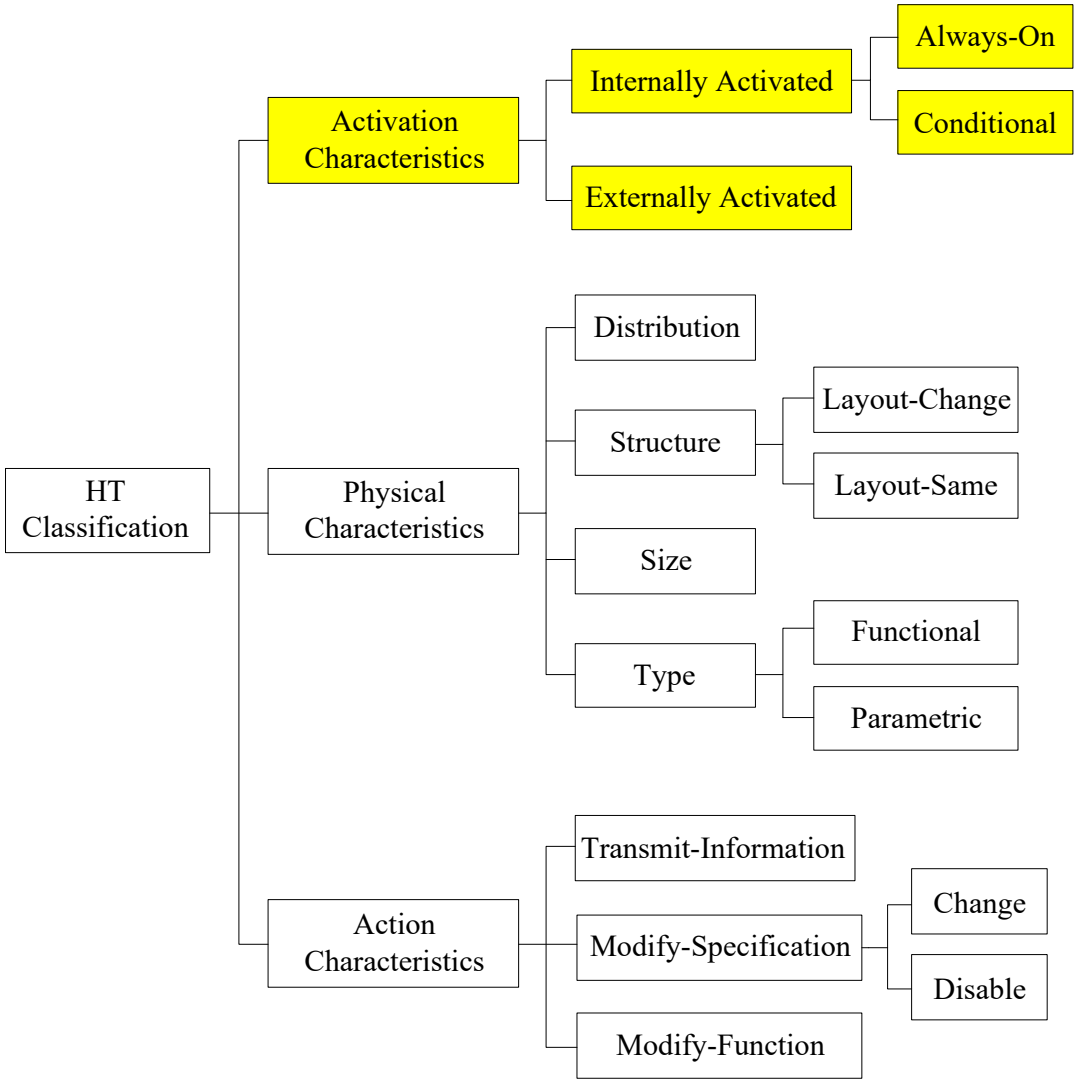
Comprehensive Attack Model

Model	Description	3PIP Vendor	SoC Developer	Foundry
A	Untrusted 3PIP vendor	Untrusted	Trusted	Trusted
B	Untrusted foundry	Trusted	Trusted	Untrusted
C	Untrusted EDA tool or rogue employee	Trusted	Untrusted	Trusted
D	Commercial-off-the-shelf component	Untrusted	Untrusted	Untrusted
E	Untrusted design house	Untrusted	Untrusted	Trusted
F	Fabless SoC design house	Untrusted	Trusted	Untrusted
G	Untrusted SoC developer with trusted IPs	Trusted	Untrusted	Untrusted

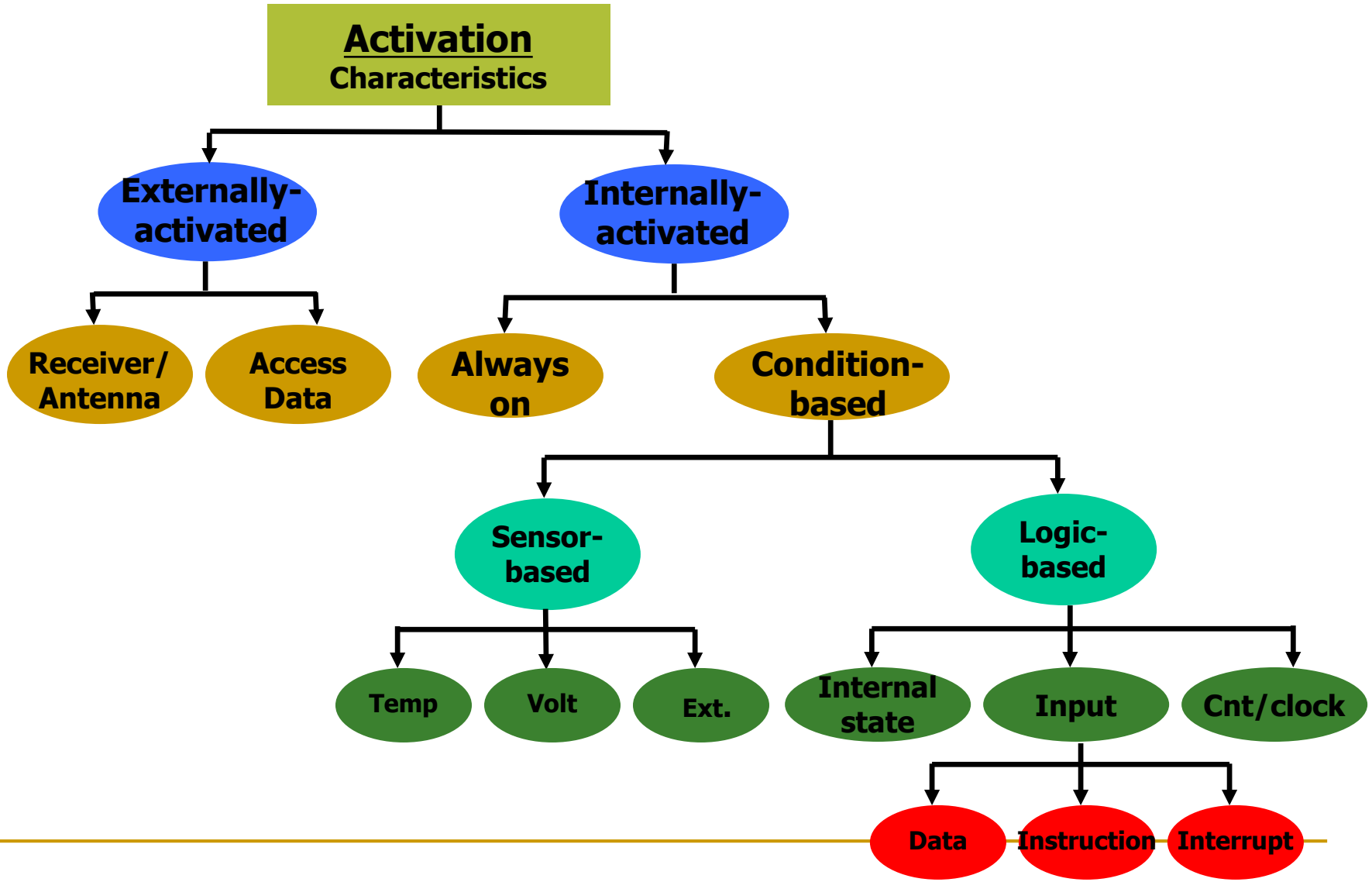
HT Classification



HT Classification

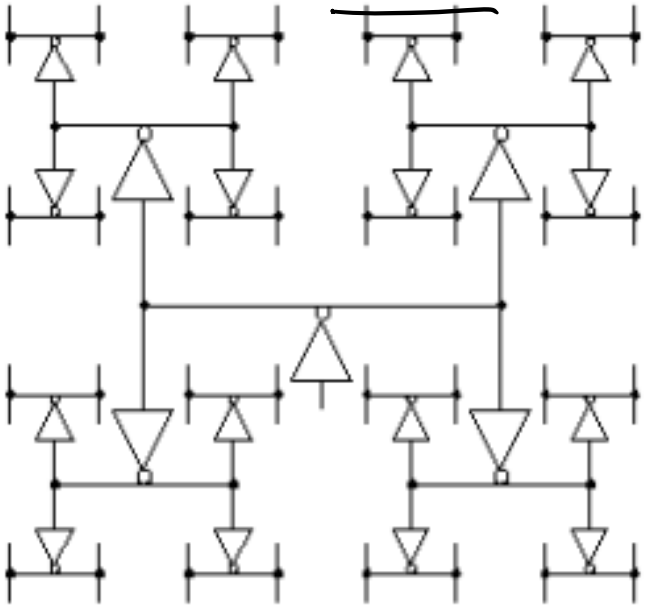


Trojan Taxonomy: Activation



Internal – Always on

- Always-on represents the HT being active at all times.
- HT example: Change the wire width to alter clock phase in different area.



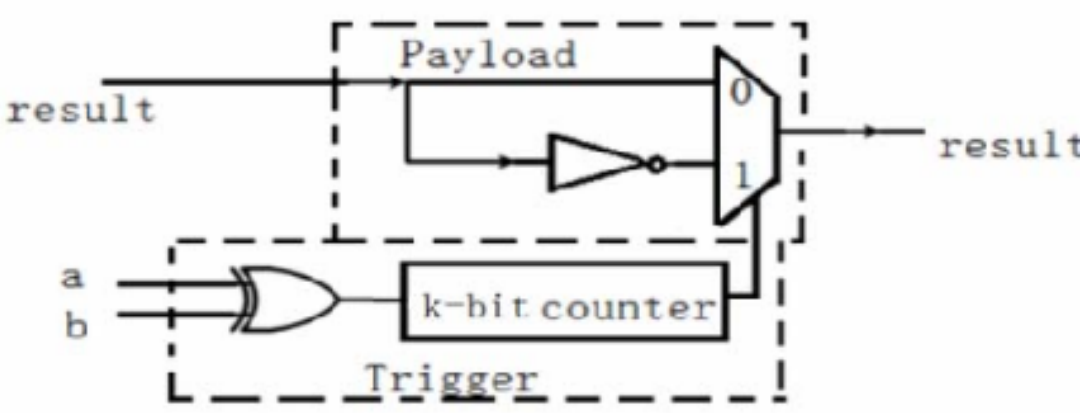
Internal – Conditional

- Condition based activation monitors one or more signals inside IC as trigger.

  Analog signal → e.g., temperature, voltage, timing

  Digital signal → e.g., Boolean logic function, FSM

- HT example: counter-based Trojan circuit.

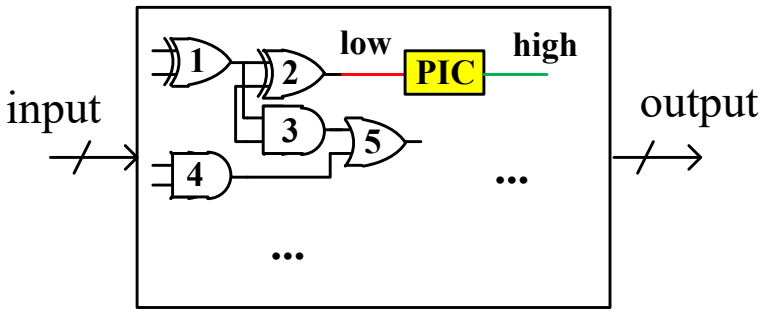


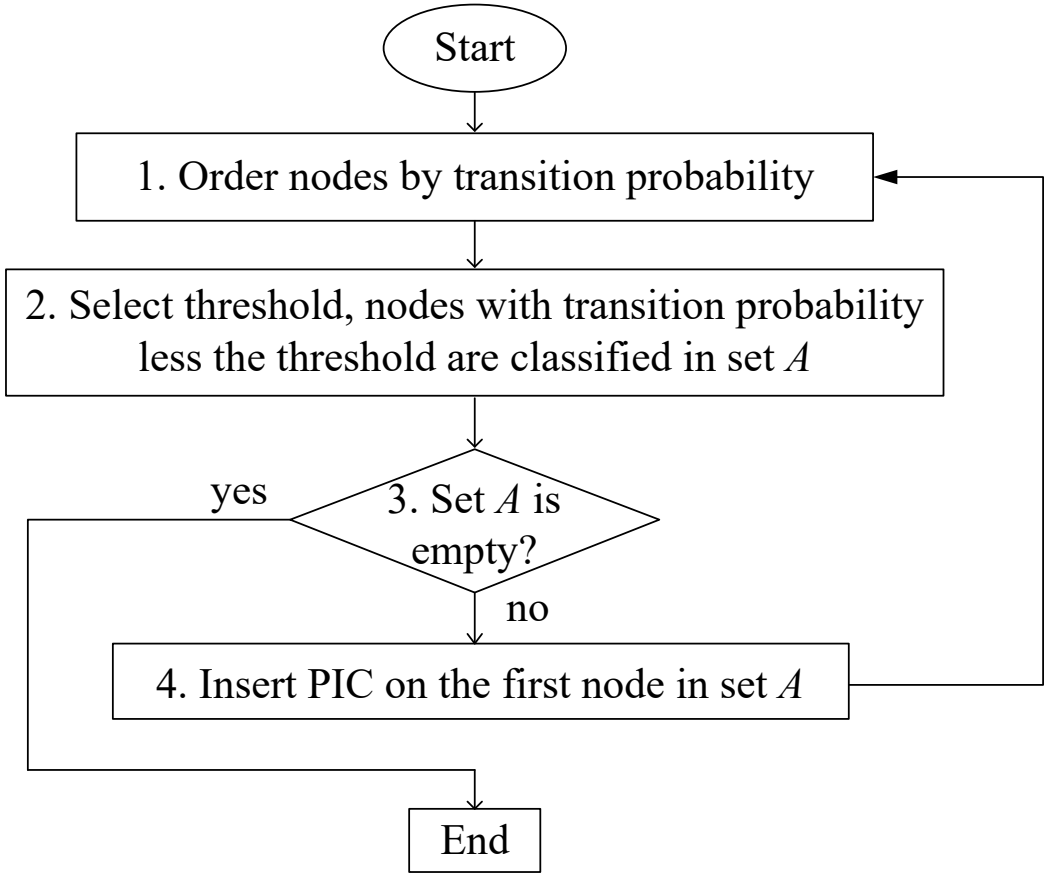
HT Activation

- **Goal** – Increase probability of rare signal which may be used as HT trigger signal to facilitate HT detection efficiency.
- **Method** – Embed a signal *probability-increase-circuit* (PIC) to node with low transition probability, reducing signal transition time.

HT Activation Workflow

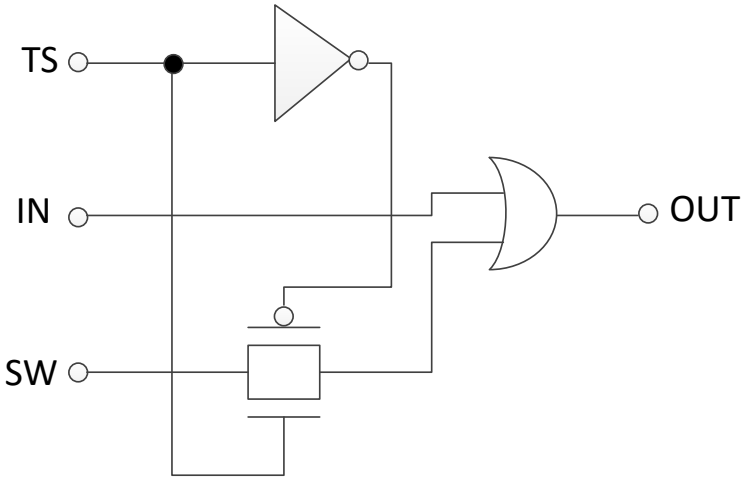
- PIC Insertion Algorithm

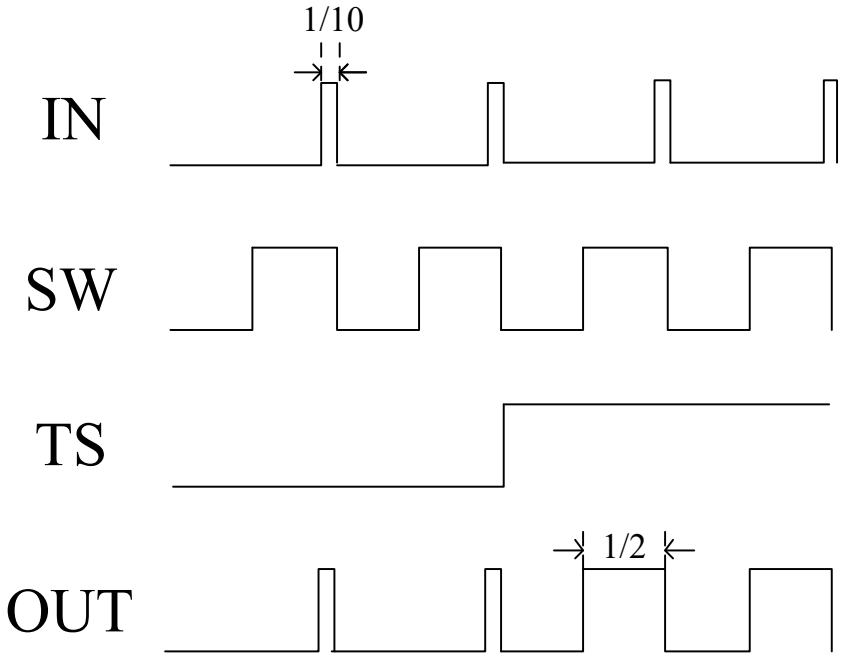




Probability Increase Circuit

- Rare Signal when $P_0 \gg P_1$





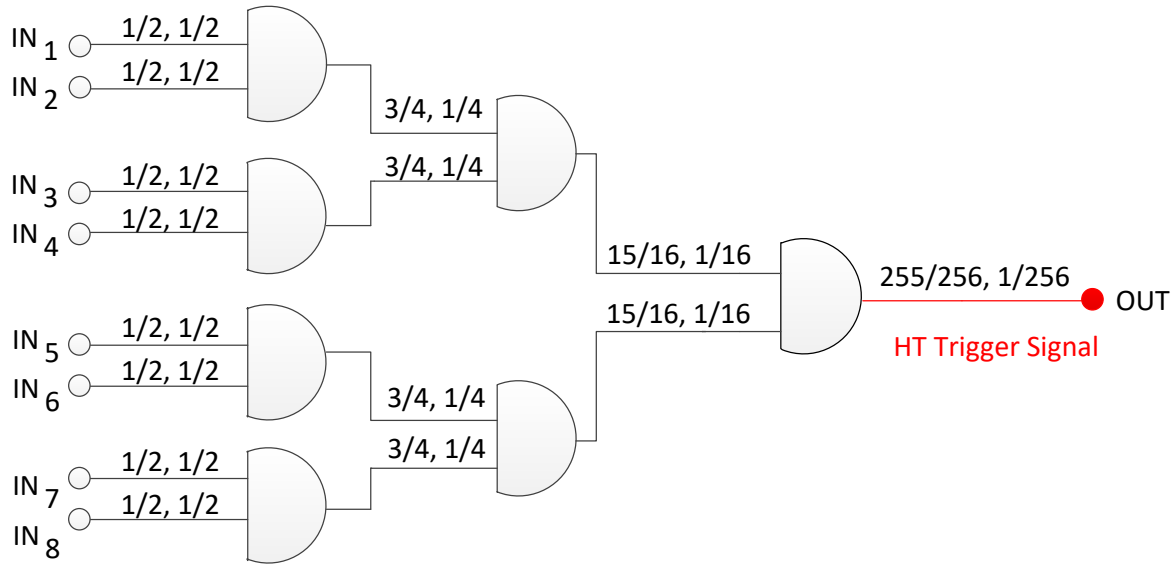
Probability-Increase-Circuit (PIC)

TS: Testing Switch

SW: Square Wave

Rare Signal in 8-bit AND Gate

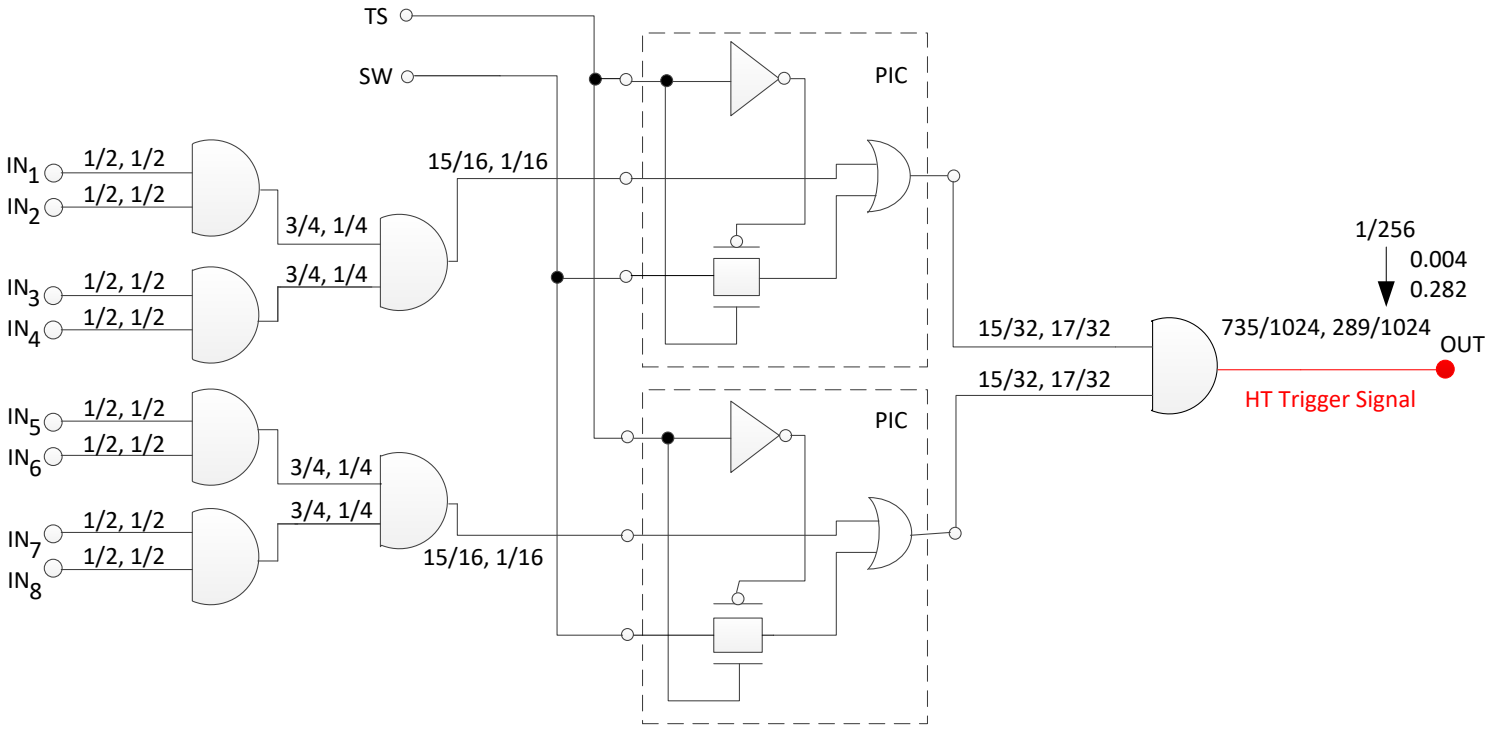
- Rare Signal when $P_0 \gg P_1$



8-Bit AND Gate, Signal Probability Label (P_0, P_1)

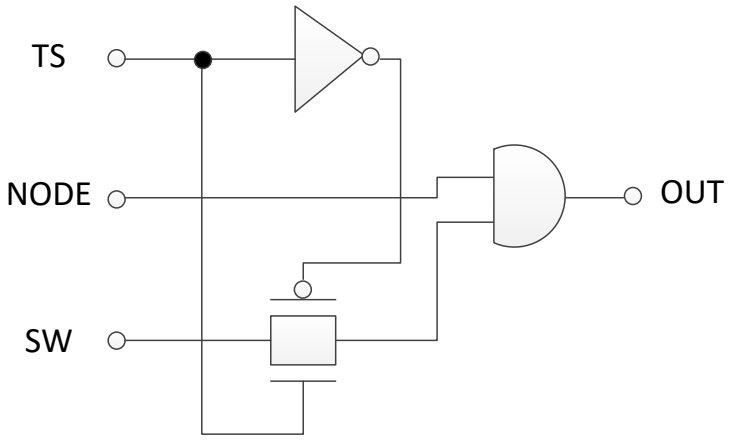
8-bit AND Gate with PIC

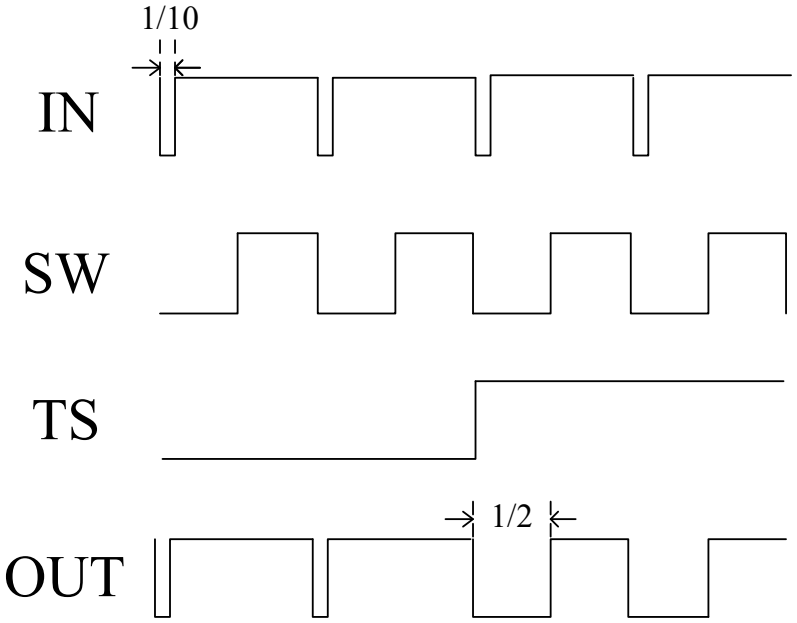
- Rare Signal when $P_0 \gg P_1$



Probability Increase Circuit

- Rare Signal when $P_0 \ll P_1$

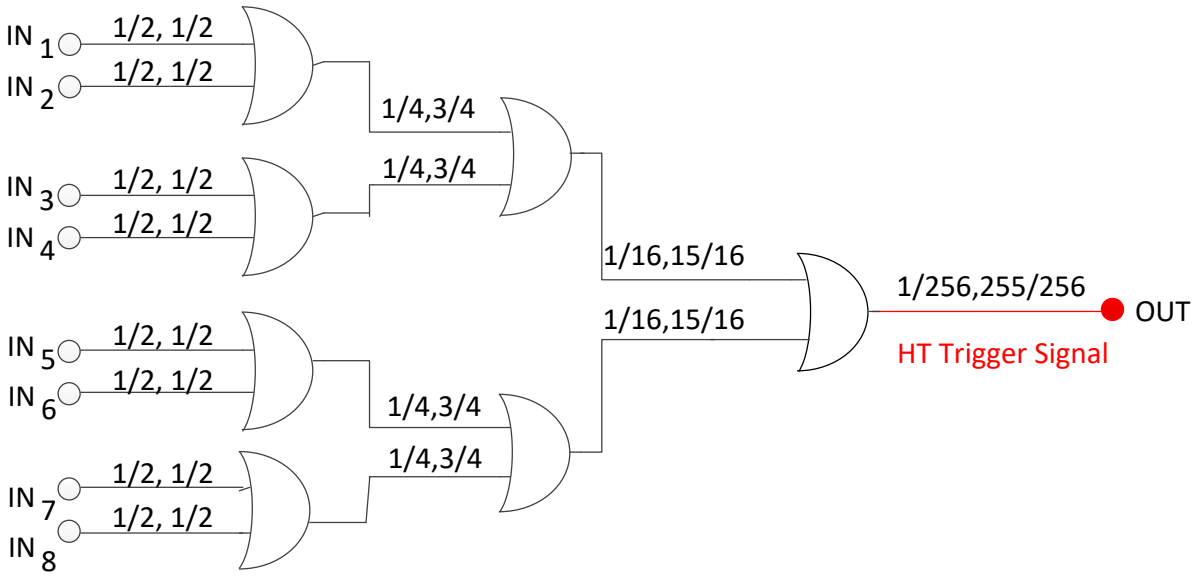




Probability-Increase-Circuit (PIC)
TS: Testing Switch
SW: Square Wave

Rare Signal in 8-bit OR Gate

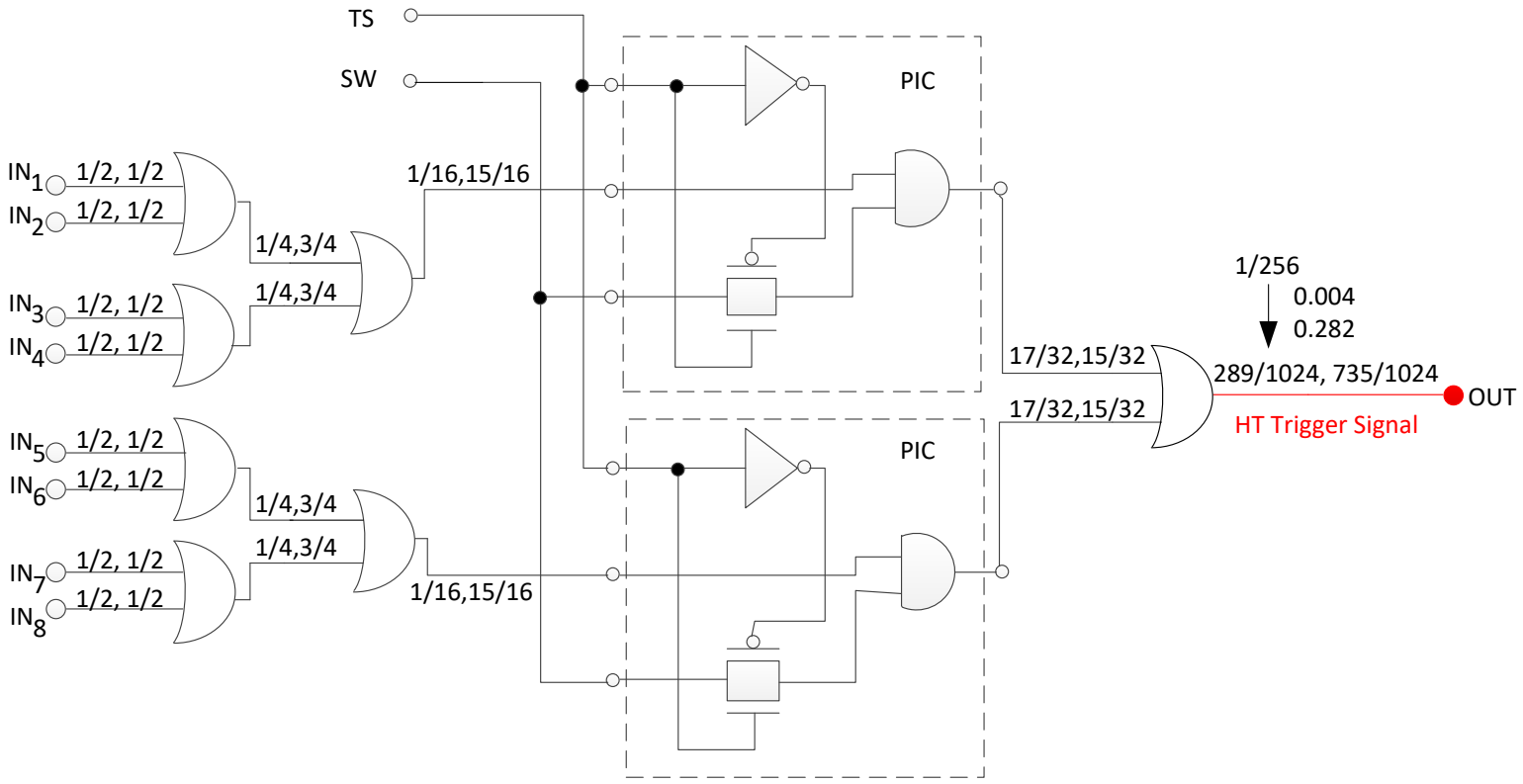
- Rare Signal when $P_0 \ll P_1$



8-Bit OR Gate, Signal Probability Label (P_0, P_1)

8-bit OR Gate with PIC

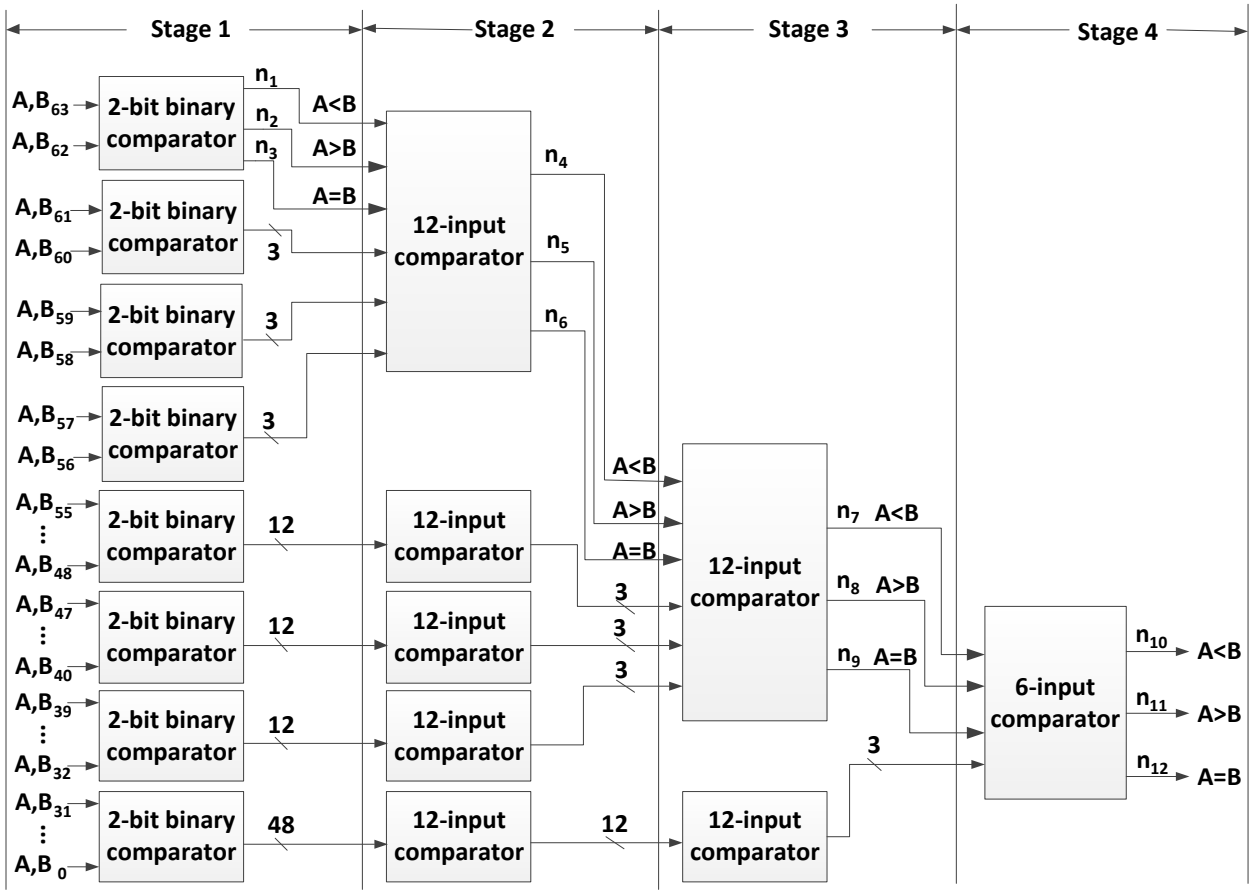
- Rare Signal when $P_0 \ll P_1$



8-Bit OR Gate with PIC

HT Activation Example

- Experimental Circuit



64-bit Binary Comparator

$$P_{n_{12}} = \left(\frac{2^{64}}{2^{2^{64}}}, \frac{2^{2^{64}} - 2^{64}}{2^{2^{64}}}\right)$$

$$TP_{n_{12}} = \frac{2^{64}}{2^{2^{64}}} * \frac{2^{2^{64}} - 2^{64}}{2^{2^{64}}} \approx 0$$

If input signal is at frequency of 1GHz, average transition time of n_{12} is 570 years.

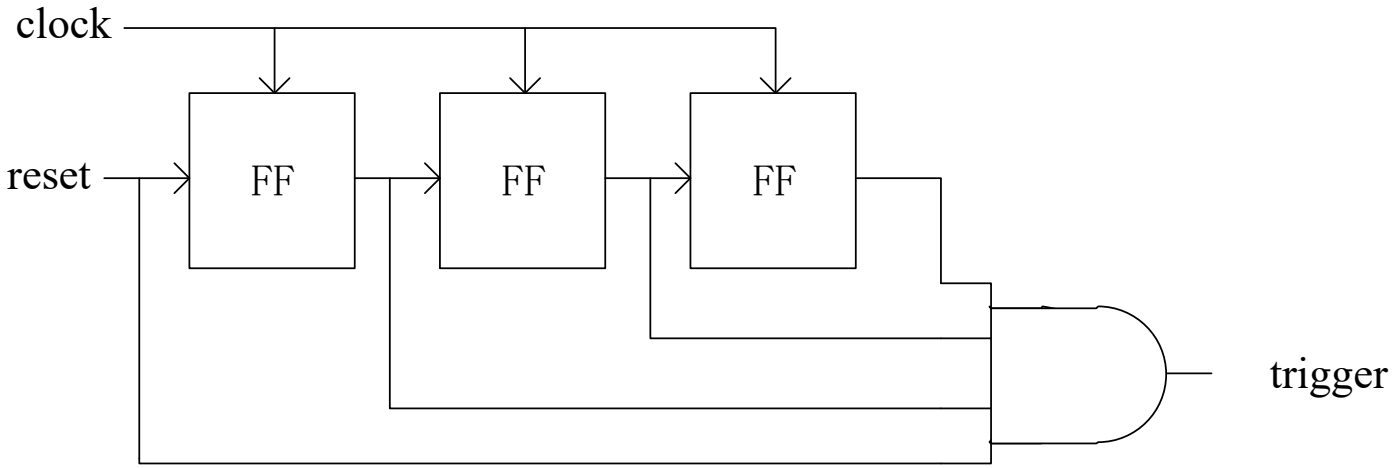
HT Activation Example

- Experimental Result

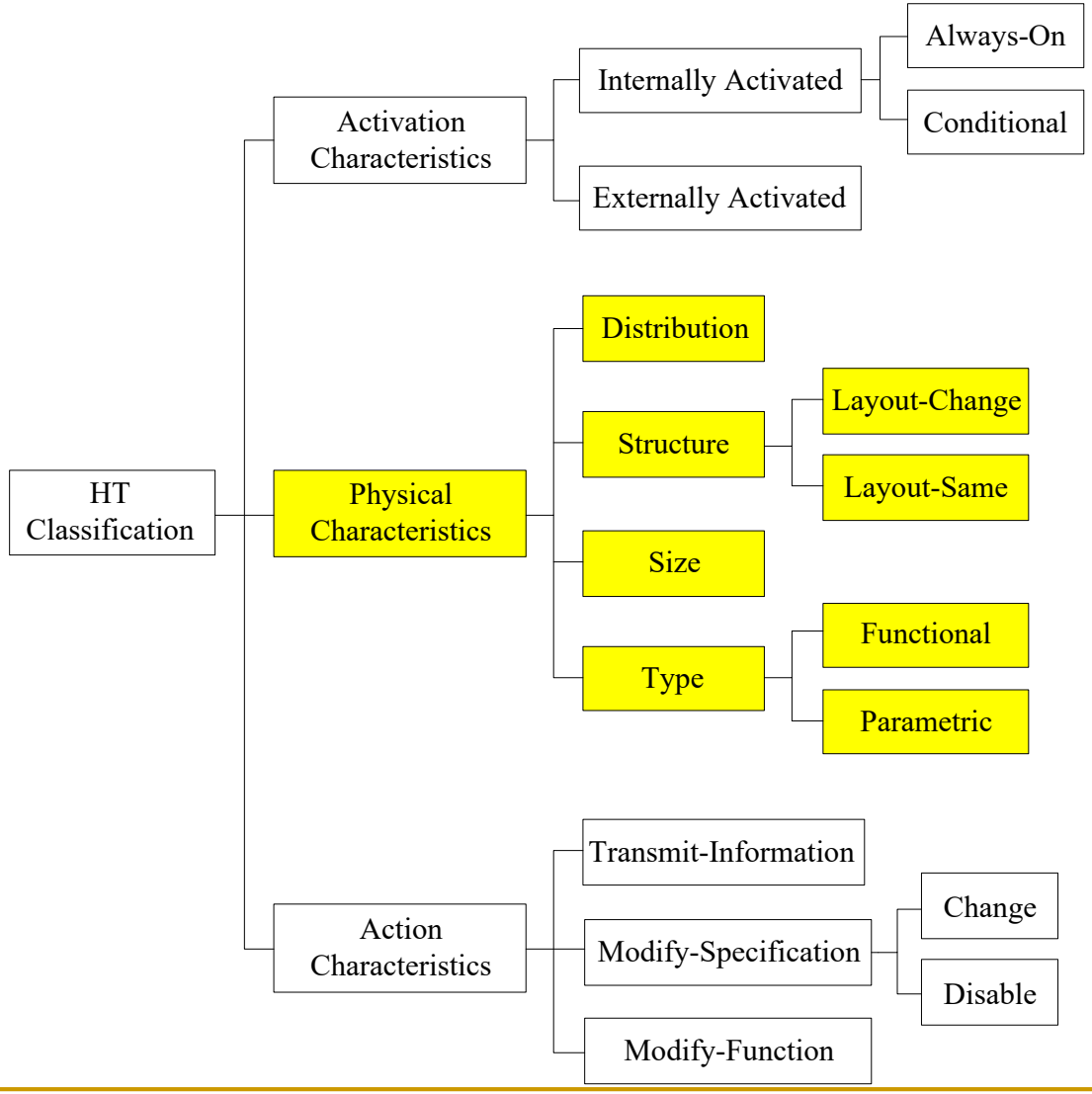
Parameters	CUT	CUT with PIC	
		Parameter	Overhead
Delay (ps)	712	840	18%
Power (mW)	12.78	12.78	0%
Size	1314	1434	1.5%
Transition	570 years	256 ns	100%

External

- Requires an external input or uses an antenna or sensors receiving HT trigger signal from outside of IC.
- HT example: primary input (reset) is used as Trojan trigger.



HT Classification



Distribution

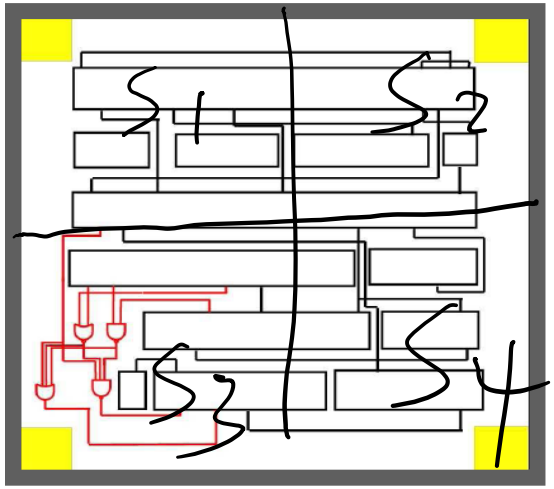
- HT distribution represents the layout density of HT on chip, that consists of tight and loose.
- Reason: 1. Depends on free layout area in original chip.

  Redesign exiting layout → affect operation parameters

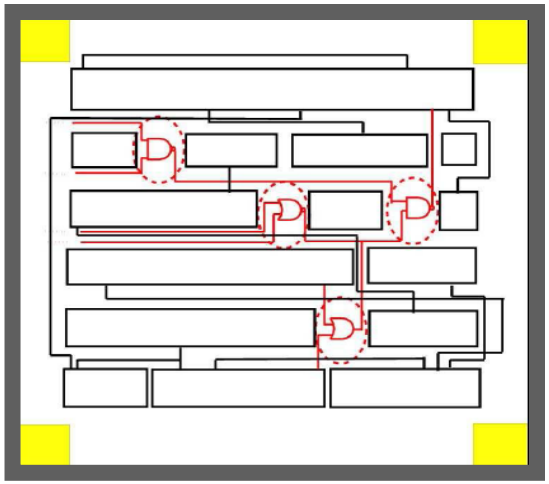
  2. Variable location of HT trigger signal and HT effect on chip to meet HT design specification
- Loose distribution → reduce parameter effect → increase wires → increase timing/power

Example: Distribution

Tight



Loose



- **Tight Distribution**
  - **Trojan components are topologically close in the layout**
- **Loose Distribution**
  - **Trojan components are dispersed across the layout of a chip**

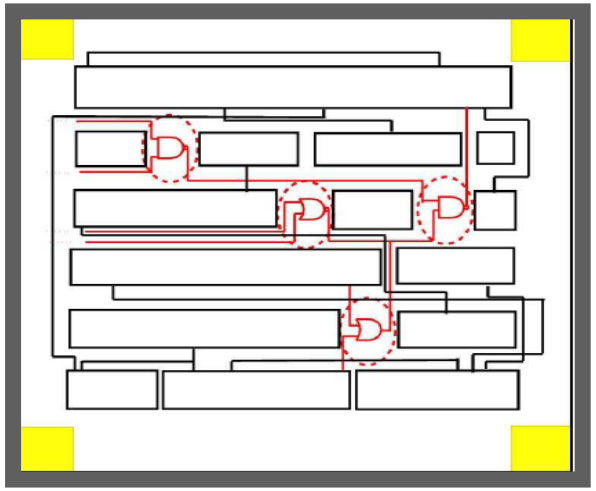
- **Distribution of Trojans depends on the availability of dead spaces on the layout**

Structure – Layout Change

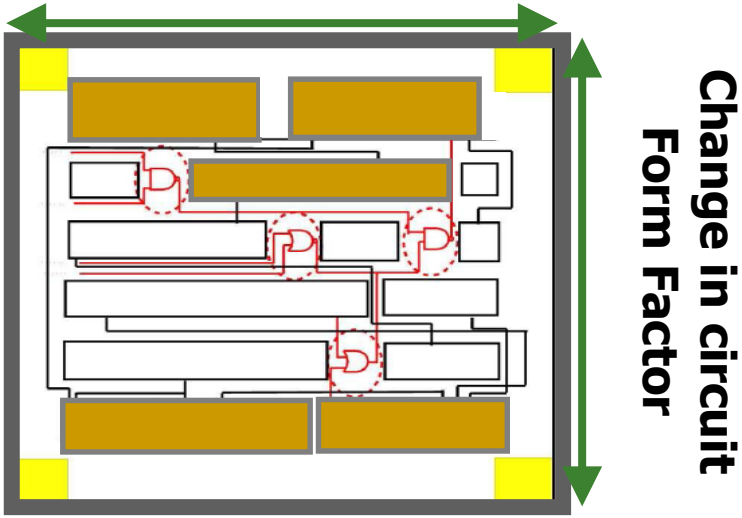
- Disadvantage: increase parameter effect
- Advantage: squeeze into a specific area utilizing the operations in it (e.g., use the signal in this area as trigger or modify the function in this area.)
  - Reduce HT circuit communication wires.
  - Hiding technology (e.g., power gating) may be used to facilitate stealthy nature of HT.

Example: Structure

No-change



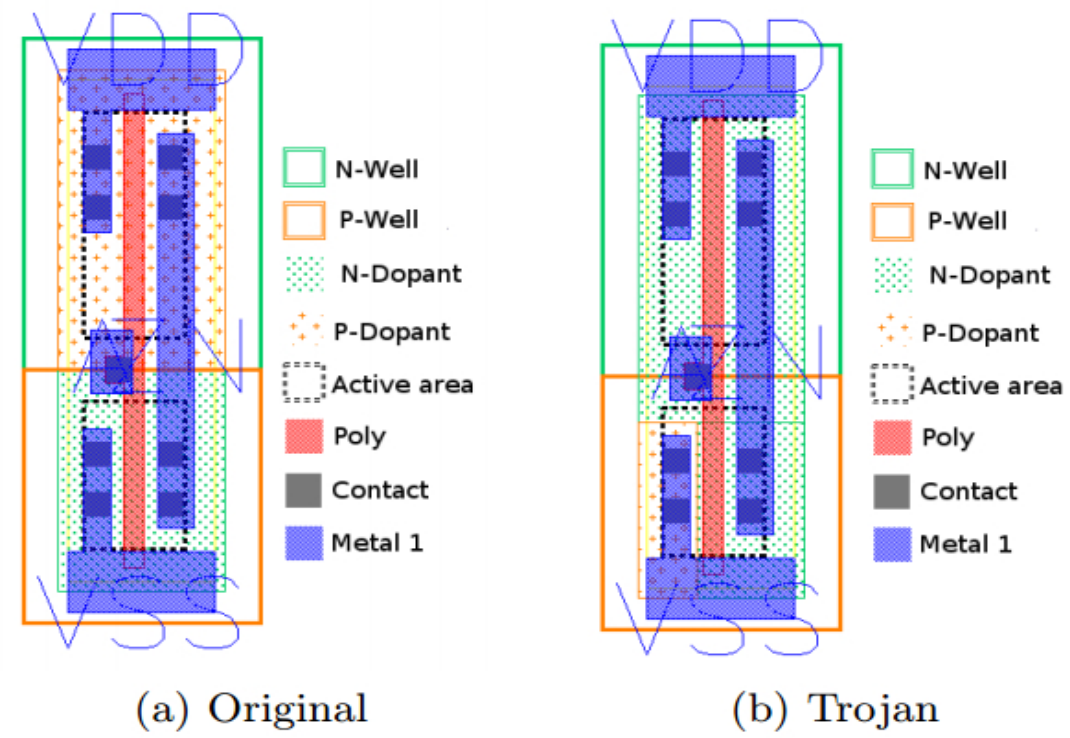
Modified Layout



- The adversary may be forced to regenerate the layout to be able to insert the Trojan, then the chip dimensions change
  - It could result in different placement for some or all the design components
- A change in physical layout can change the delay and power characteristics of chip
  - It is easier to detect the Trojan

Structure – Layout Same

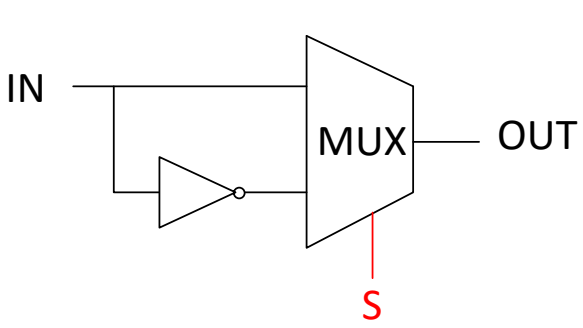
- HT example: n-dopant in PMOS, p-dopant in NMOS.

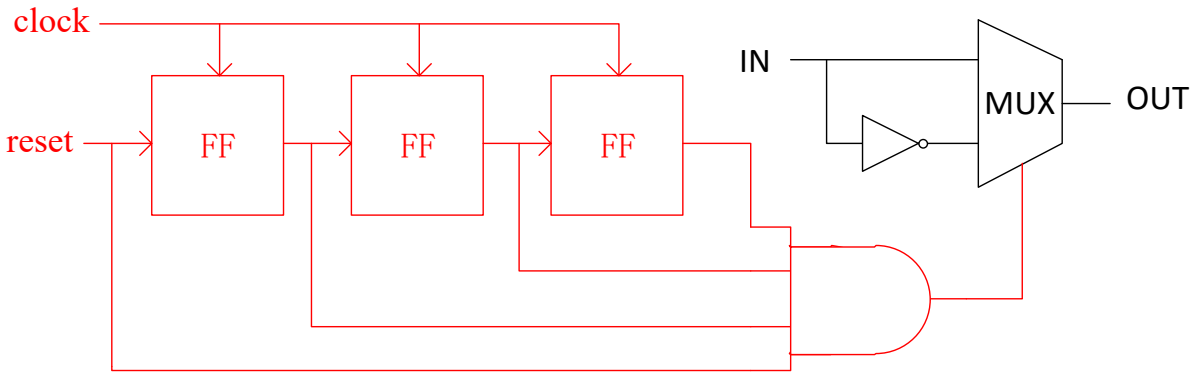


(a) Original (b) Trojan

Size

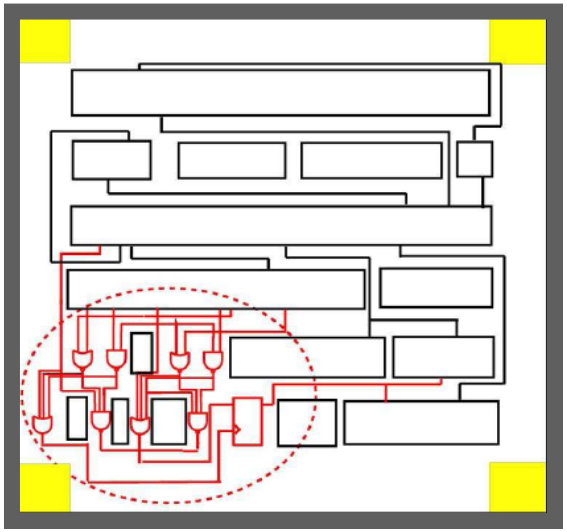
- HT size accounts for the number of component that have been added, deleted or altered.
- HT with large size looks like easier to be detected, however HT with small size may lose ability monitoring complicated rare trigger that increases its activation probability, eventually harming silence of HT.



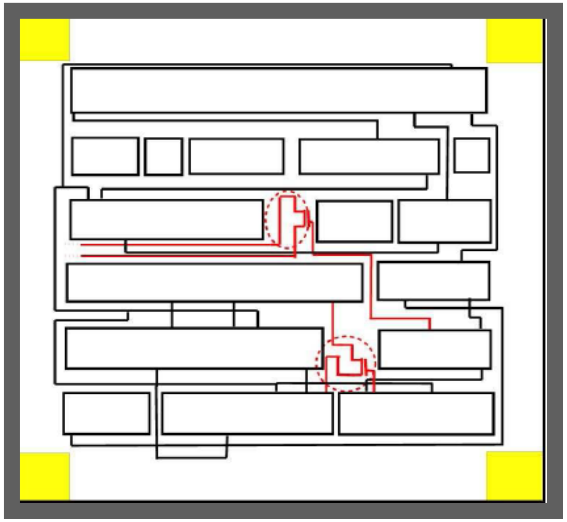


Example: Size

Big



Small

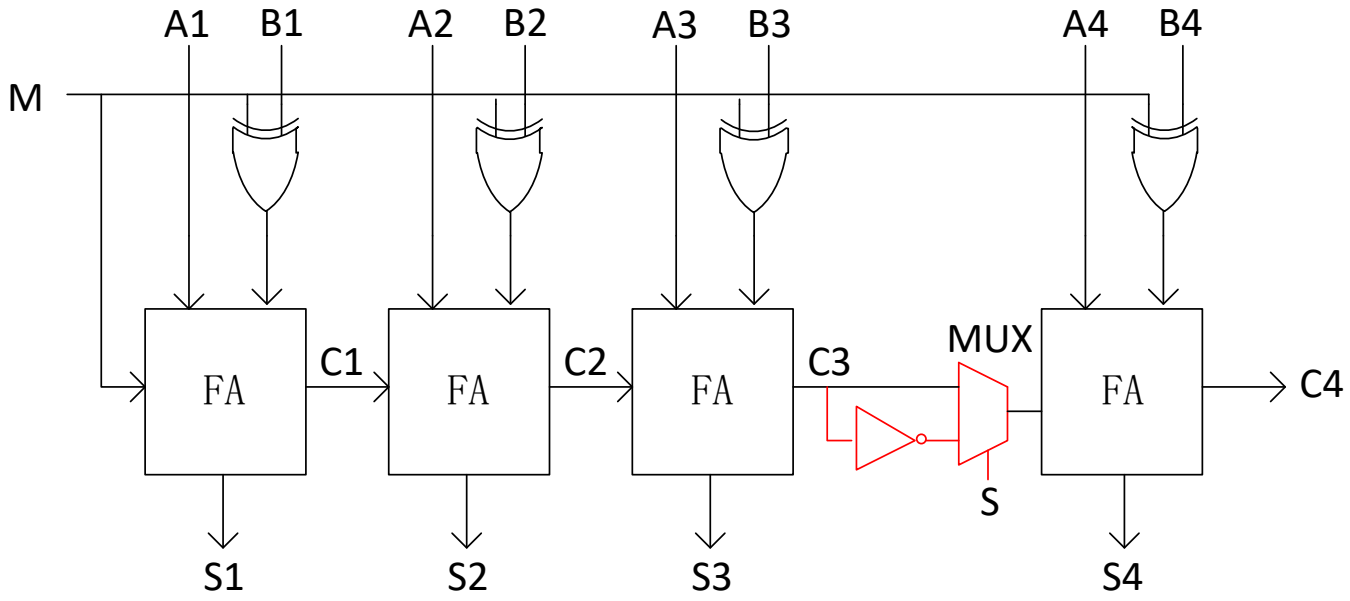


- **Size:**
  - **Number of components added to the circuit**
    - **Small transistors**
    - **Small gates**
    - **Large gates**

- **In case of layout, depends on availability of:**
  - **Dead spaces**
  - **Filler cells**
  - **Decap cells**
  - **Change in the structure**

Type - Functional

- The HT causing addition or deletion of transistors or gates, which is relatively popular due to its various possible functionalities.
- HT example: Alter signal in 4-bit adder/subtractor.

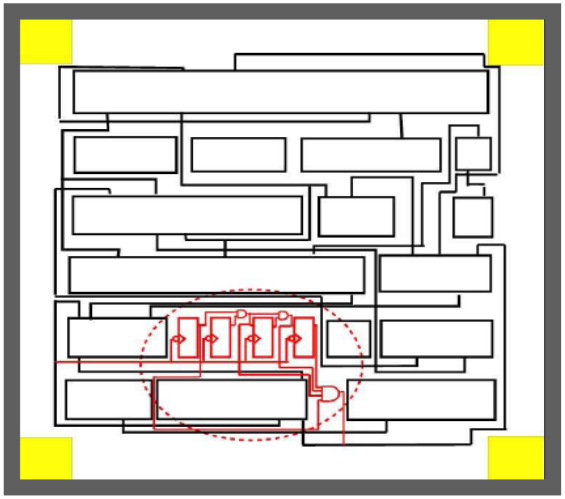


Type - Parametric

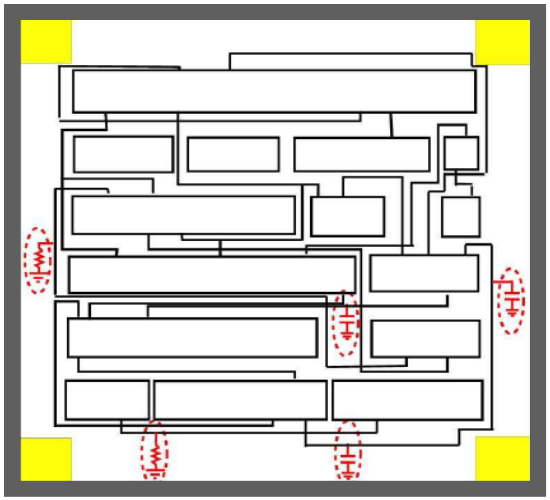
- Modify circuit operation parameters.
- HT example: thin wires, increase wire, increase threshold voltage.

Example: Type

Functional

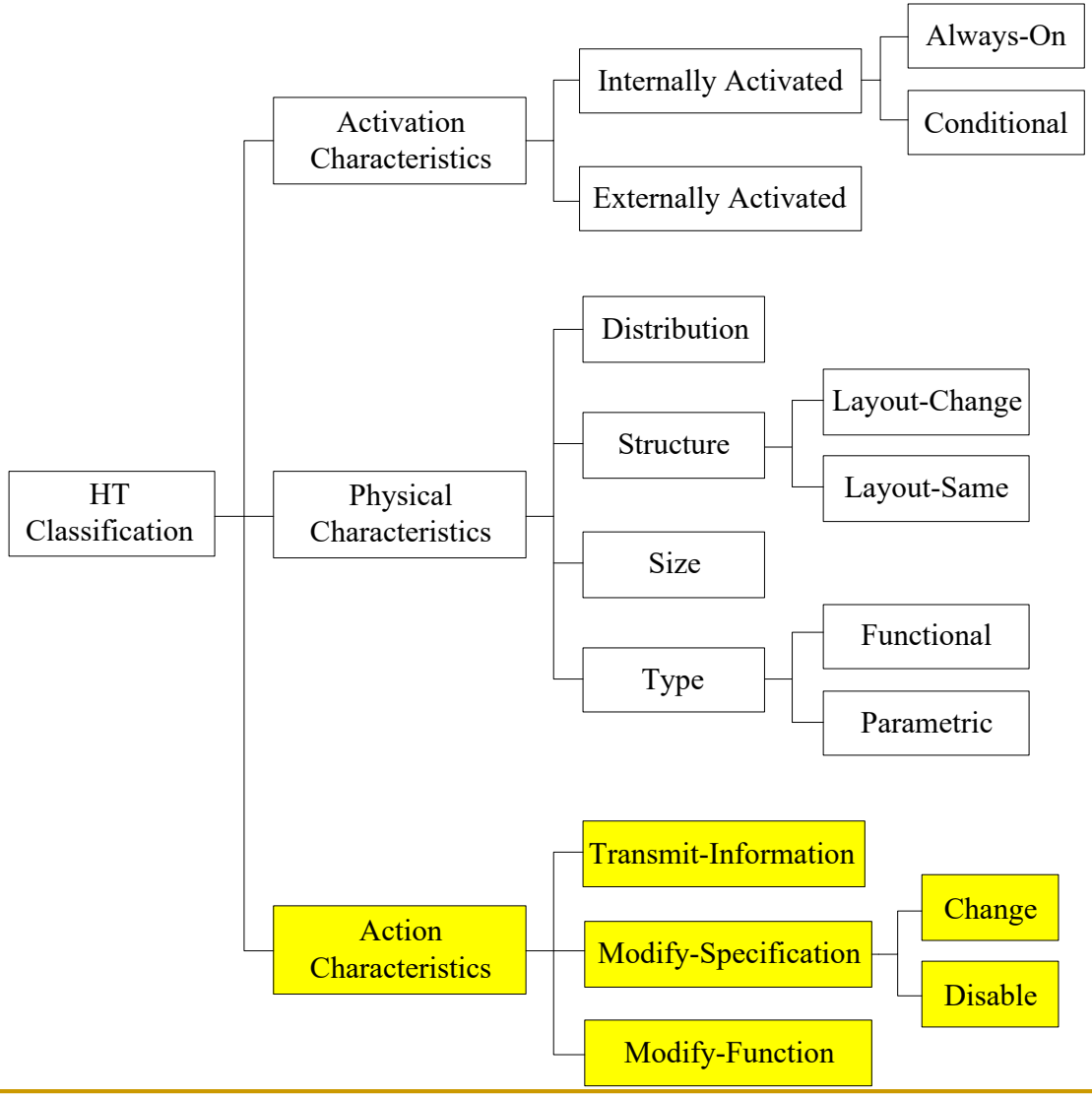


Parametric



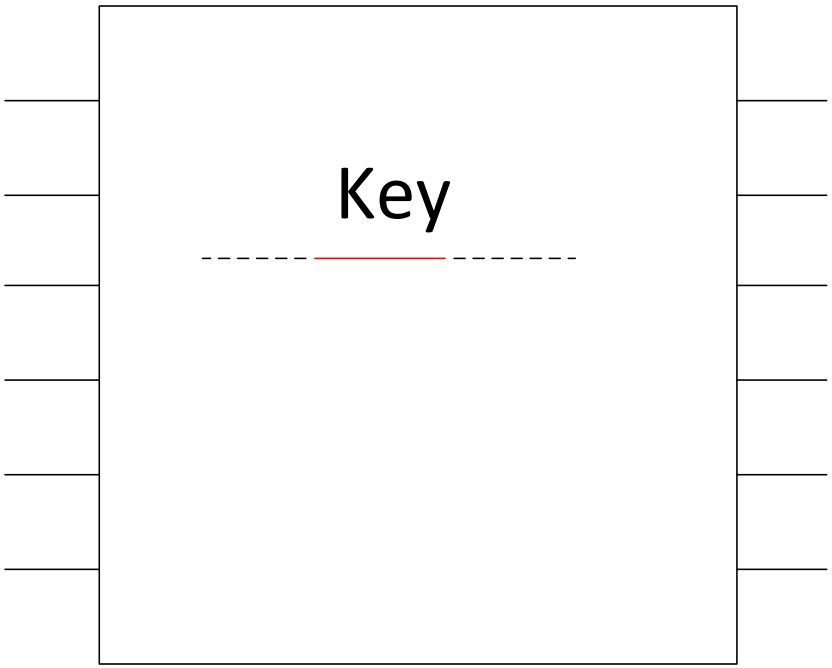
- **Functional**
  - **Addition or deletion of components**
  - **Sequential circuits**
  - **Combinational circuits**
  - **Modification to function or no change**
- **Parametric**
  - **Modifications of existing components**
    - **Wire: e.g. thinning of wires**
    - **Logic: Weakening of a transistor, modification to physical geometry of a gate**
    - **Modification to power distribution network**
  - **Sabotage reliability or increase the likelihood of a functional or performance failure**

HT Classification



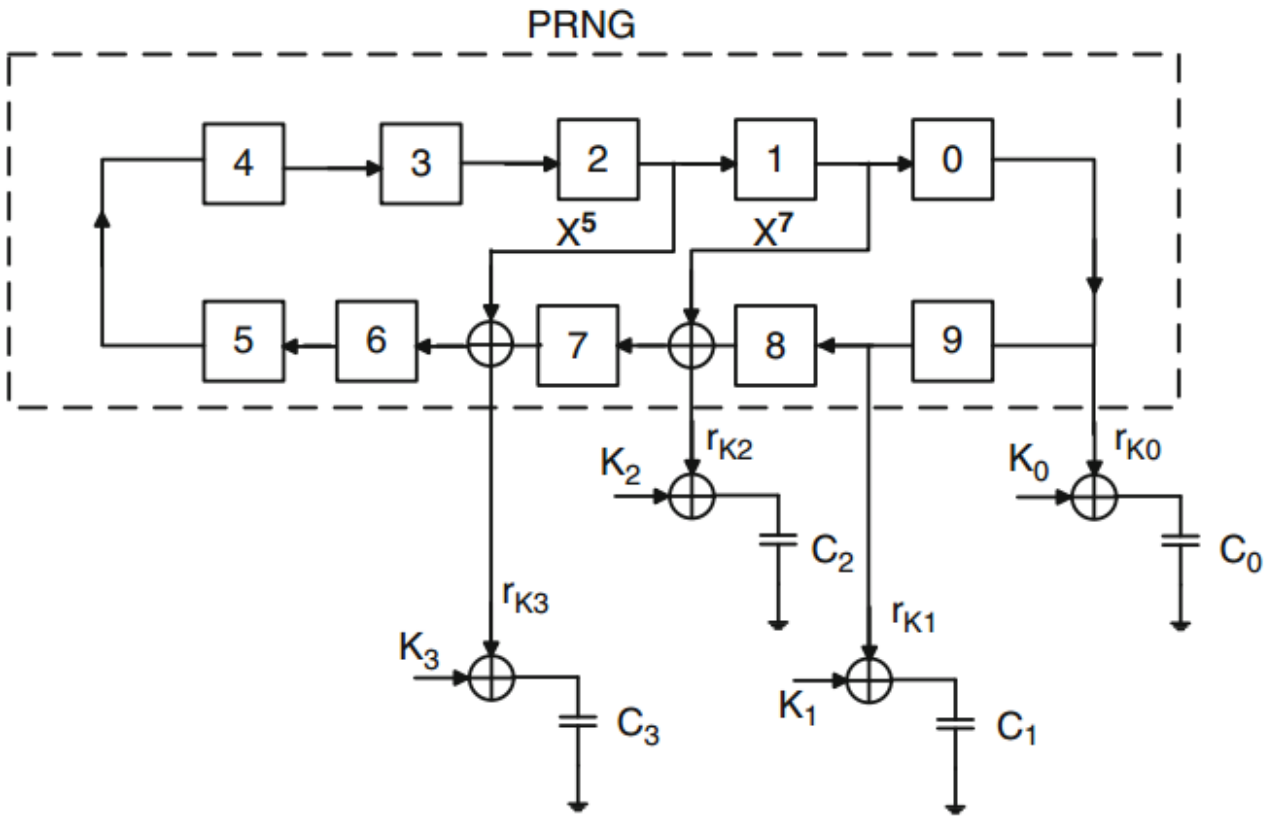
Transmit Information

Have access to IC chip. How to steal the internal key signal?



Transmit Information

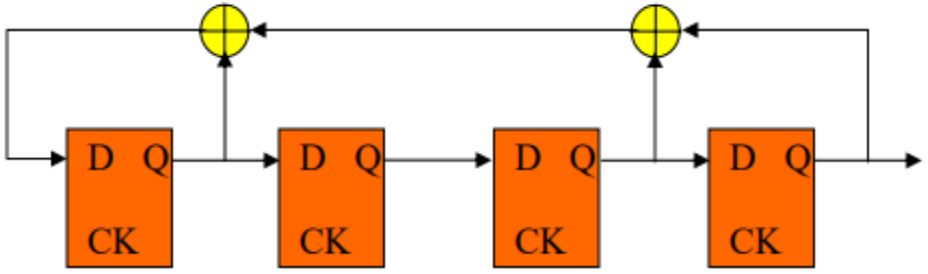
HT example: pseudo random number generator leaks secret key.



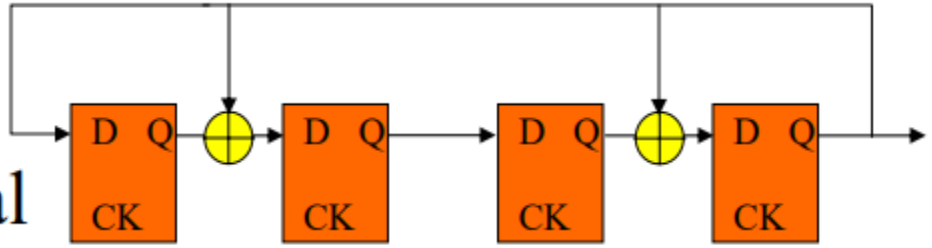
Linear Feedback Shift Registers (LFSRs)

- Efficient design for Test Pattern Generators & Output Response Analyzers (also used in CRC)
  - FFs plus a few XOR gates
  - better than counter
    - fewer gates
    - higher clock frequency
- Two types of LFSRs
  - External Feedback
  - Internal Feedback
    - higher clock frequency
- Characteristic polynomial
  - defined by XOR positions
  - $P(x) = x^4 + x^3 + x + 1$ in both examples

External Feedback LFSR



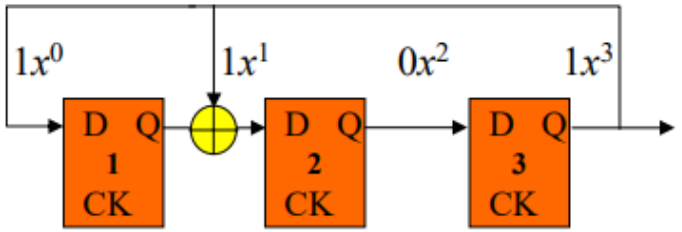
Internal Feedback LFSR



Linear Feedback Shift Registers (LFSRs)

Characteristic polynomial of LFSR

- n = # of FFs = degree of polynomial
- XOR feedback connection to FF i ⇔ coefficient of x^i
  - coefficient = 0 if no connection
  - coefficient = 1 if connection
  - coefficients always included in characteristic polynomial:
    - x^n (degree of polynomial & primary feedback)
    - $x^0 = 1$ (principle input to shift register)
- Note: state of the LFSR ⇔ polynomial of degree n-1
- Example: $P(x) = x^3 + x + 1$

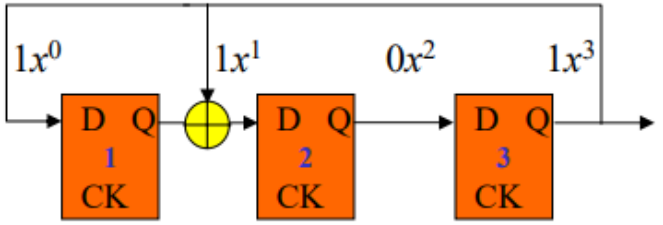


Linear Feedback Shift Registers (LFSRs)

- An LFSR generates periodic sequence
  - must start in a non-zero state,
- The maximum-length of an LFSR sequence is 2^n -1
  - does not generate all 0s pattern (gets stuck in that state)
- The characteristic polynomial of an LFSR generating a maximum-length sequence is a ***primitive polynomial***
- A maximum-length sequence is ***pseudo-random***:
  - number of 1s = number of 0s + 1
  - same number of runs of consectuive 0s and 1s

Linear Feedback Shift Registers (LFSRs)

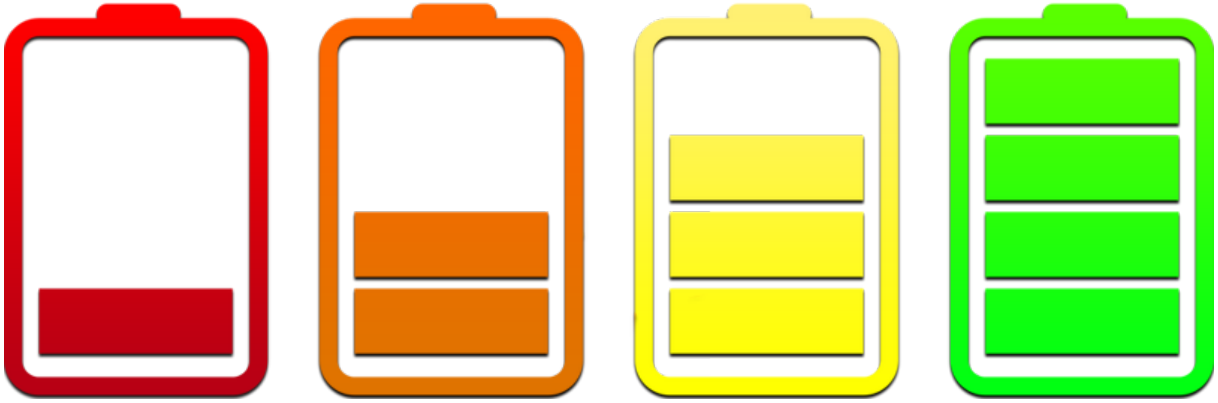
- Example: Characteristic polynomial is $P(x) = x^3 + x + 1$
- Beginning at all 1s state
  - 7 clock cycles to repeat
  - maximal length = 2^n-1
  - polynomial is primitive
- Properties:
  - four 1s and three 0s
  - 4 runs:
    - 2 runs of length 1 (one 0 & one 1)
    - 1 run of length 2 (0s)
    - 1 run of length 3 (1s)
- Note: external & internal LFSRs with same primitive polynomial do not generate same sequence (only same length)



1	1	1	1
1	0	1	2
1	0	0	3
0	1	0	4
0	0	1	5
1	1	0	6
0	1	1	7
1	1	1	

Modify Specification - Change

- HT attacking the parameter properties (e.g., delay) of host-chip.
- HT example: Change the wire width. / exhausting battery power.



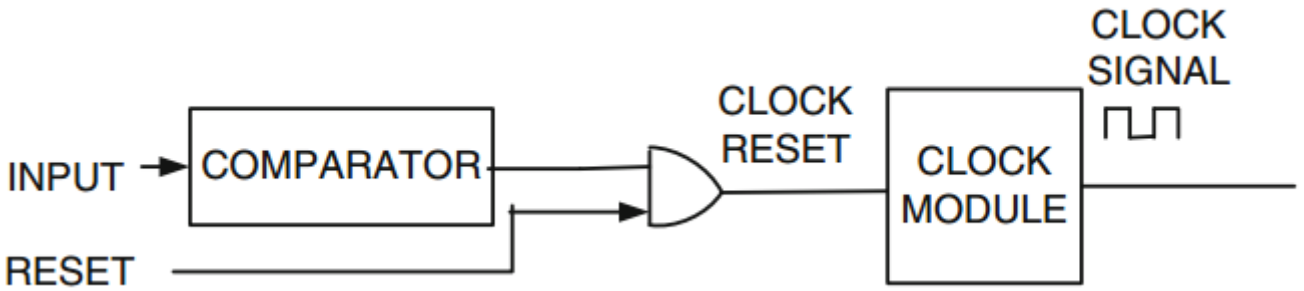
Modify Specification - Disable

- Disable host-circuit operation.
- HT design: How to disable “clock” in an IC chip.

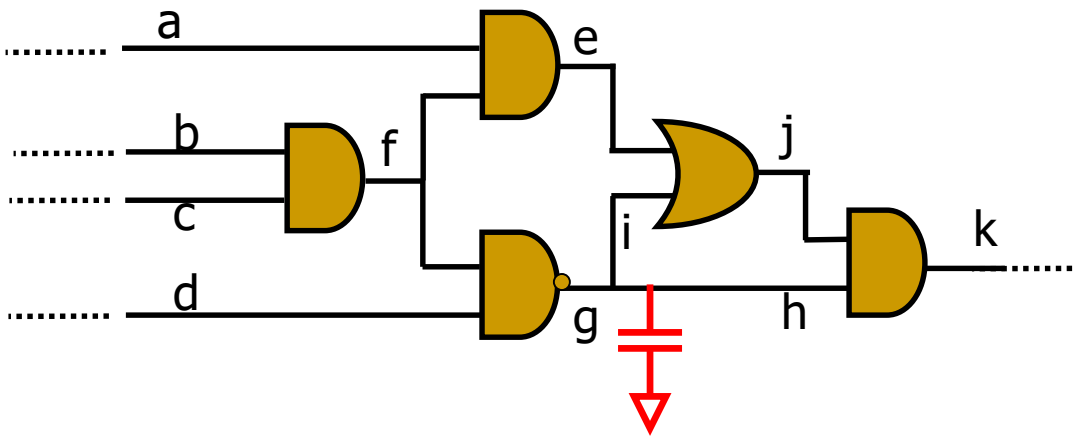
Modify Specification - Disable

- HT example:

1. A series of XOR gates are used to compares the input sequence with a pre-defined binary value. The specific input sequence can freeze clock.
2. simply isolate partial or all power supply from circuits.

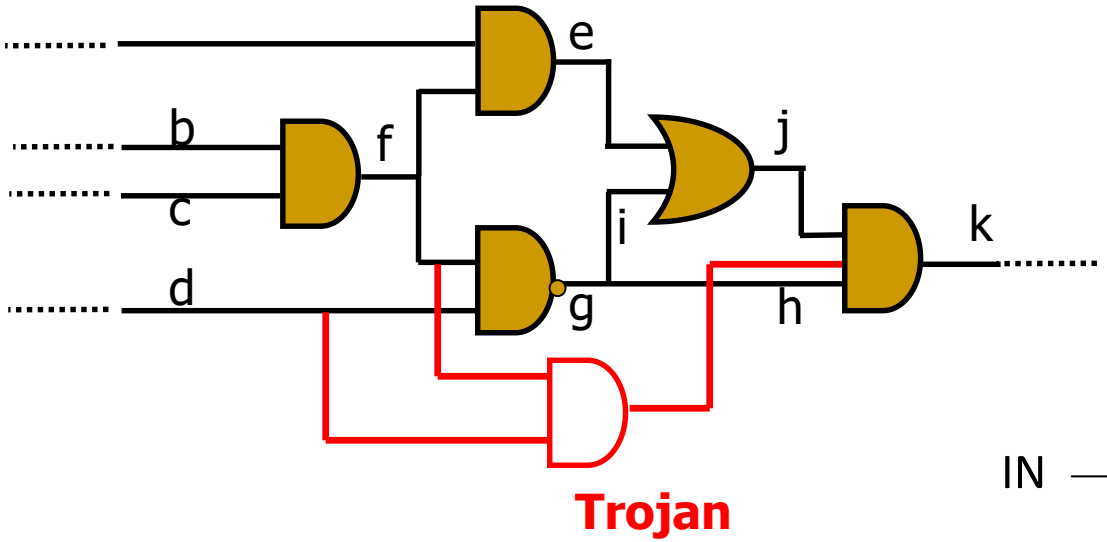


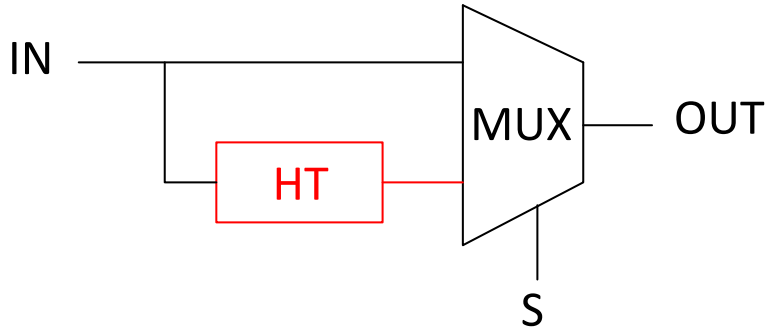
Modify Specification – Noise, Delay and Temperature



Modify Function

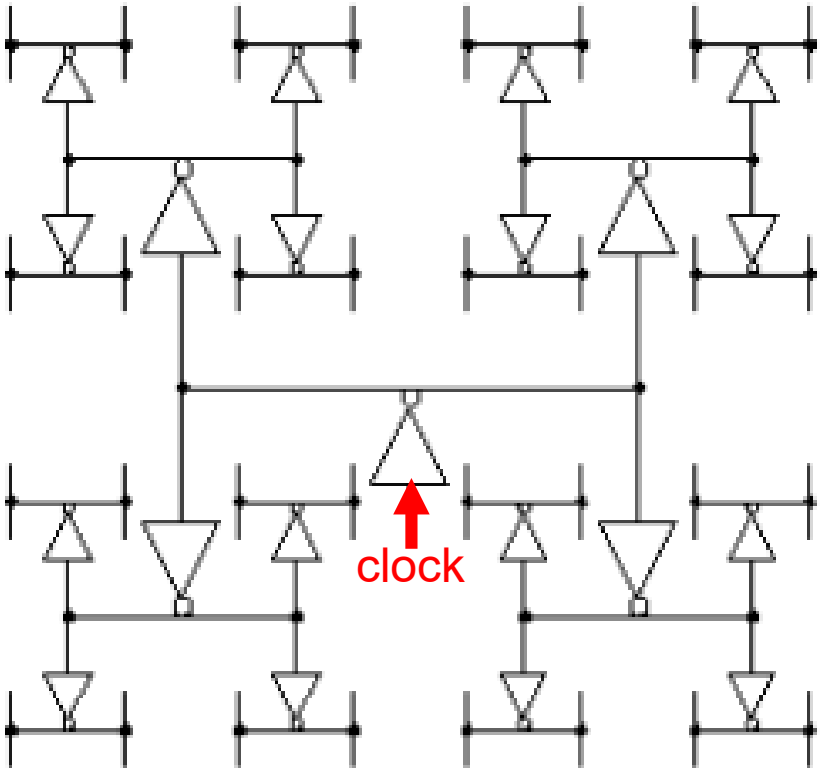
- HT modify designed function of host-chip by adding extra logic components or bypassing original circuits.
- HT design: how to change the original circuit function.





Modify Function

- HT example: Change the wire width to alter clock phase in different area.



Design for Hardware Trust

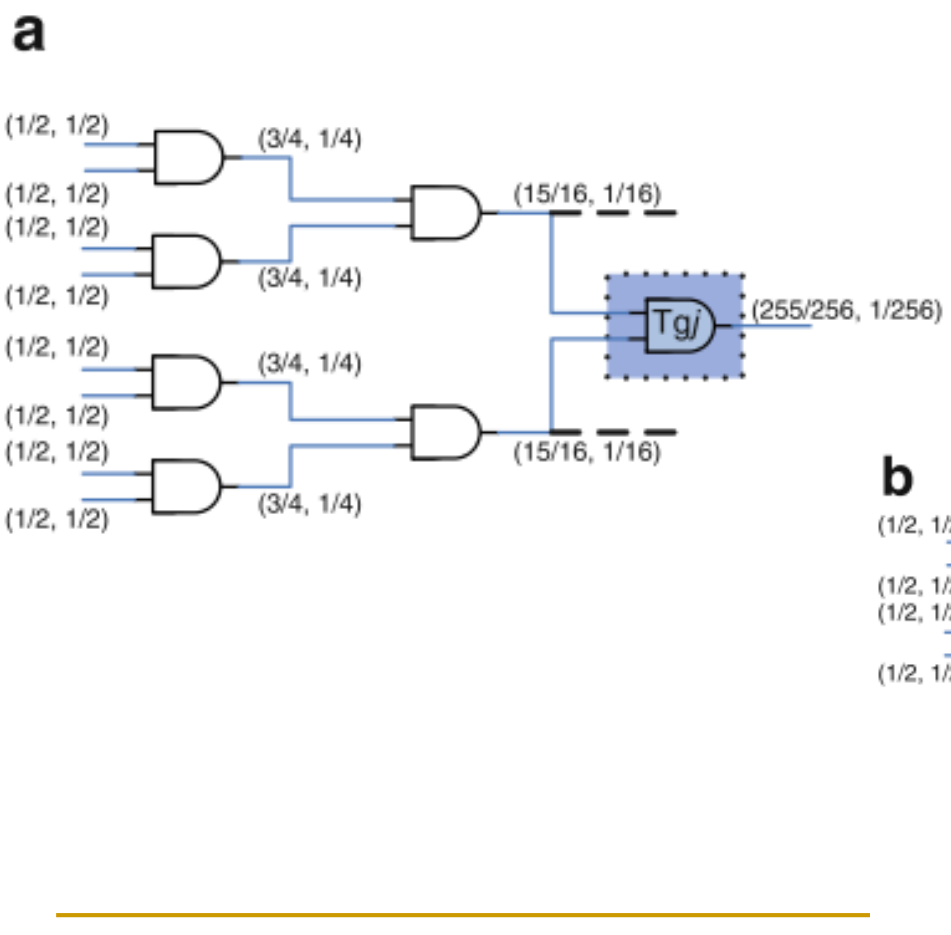
- Since detecting Trojan is extremely challenging, design for hardware trust approaches are proposed to
  - **Improve hardware Trojan detection methods**
    - Improve sensitive to power and delay
    - Rare event removal
  - **Prevent hardware Trojan insertion**
    - Design obfuscation

Rare Event Removal

- Intelligent attackers will choose low-frequency events to trigger the inserted Trojans.
- Improving controllability or observability can make rare events scarce, thereby facilitating detecting Trojans inside the design.
  - Design for Trojan test: inserting probing points
  - Inserting dummy scan flip-flops

Increasing Probability of Partial/Full Activation

- Inserting dummy FFs on path with very low activation probability



Increasing Probability of Partial/Full Activation

- Dummy scan flip-flops are inserted to control hard-to-excite nodes.

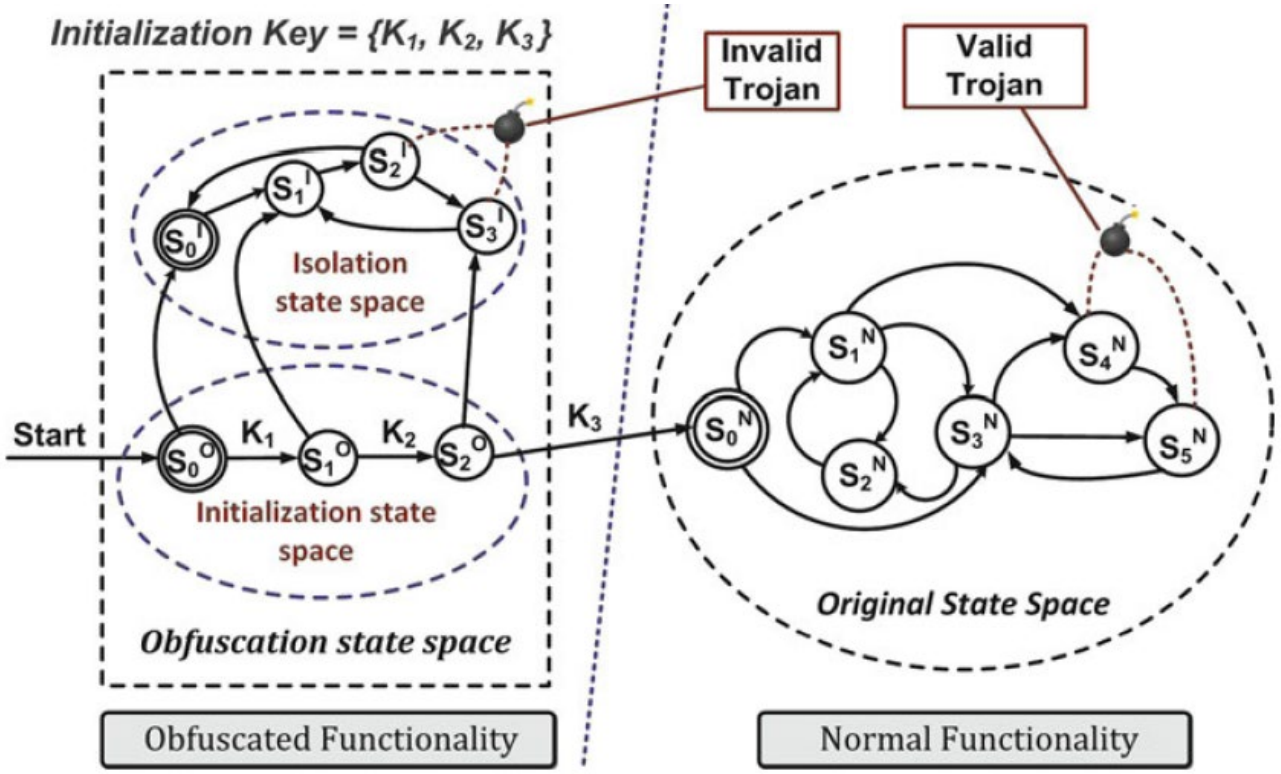
- Usage:
  - **Full activation**: increase controllability
  - **Power-based**: generate switching activities
  - **Delay-based**: activate more paths to improve coverage

Trojan Prevention-Design obfuscation

- The objective is deterring attackers from inserting Trojans inside the design.
- Design obfuscation means that a design will be transformed to another one which is functionally equivalent to the original, but in which it is much harder for attackers to obtain complete understanding of the internal logic, making reverse engineering much more difficult to perform.
- It obfuscates the state transition function to add an obfuscated mode on top of the original functionality (called normal mode).

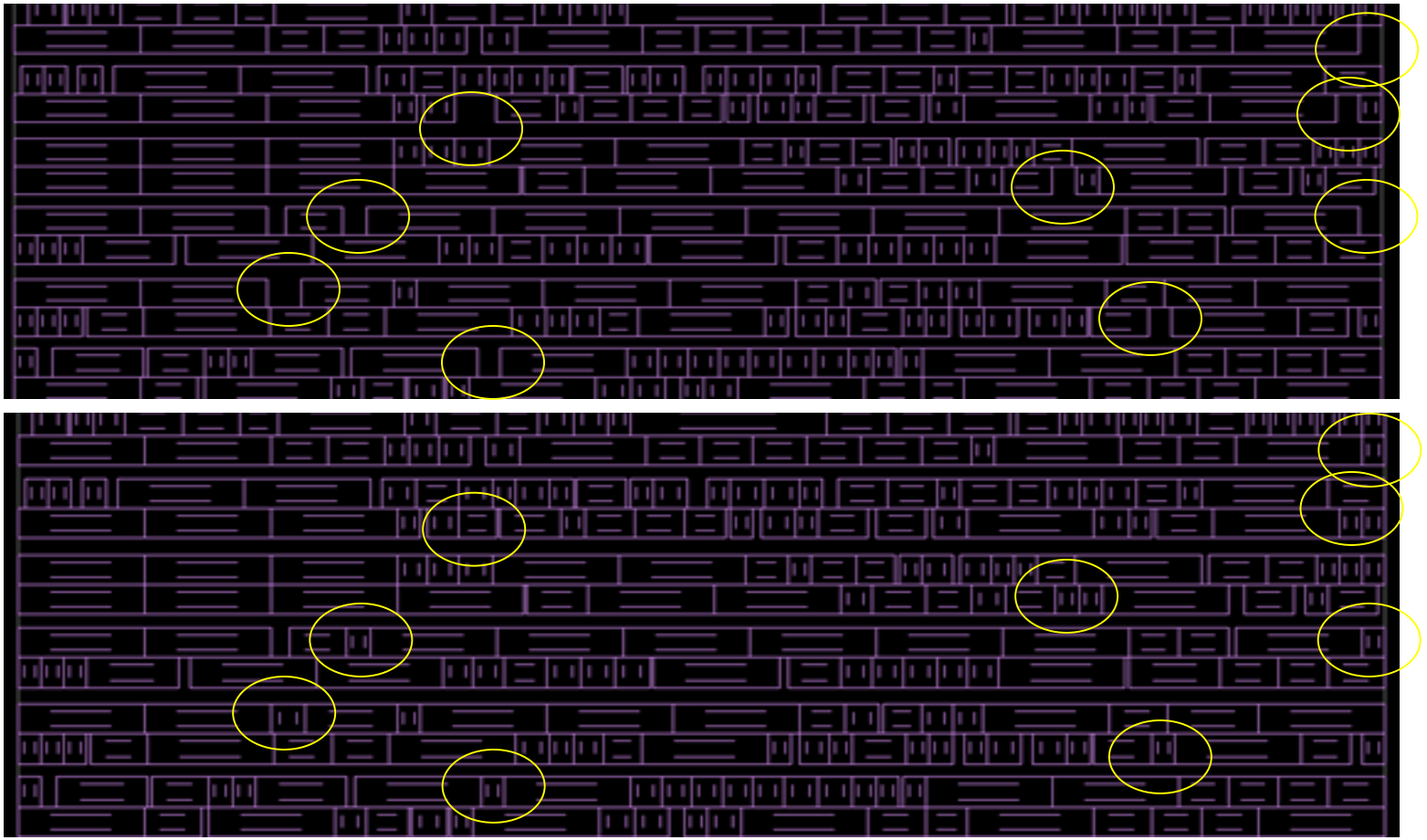
Design obfuscation

- Specified pattern is able to guide the circuit into its normal mode.
- The transition arc K3 is the only way the design can enter normal operation mode from the obfuscated mode.



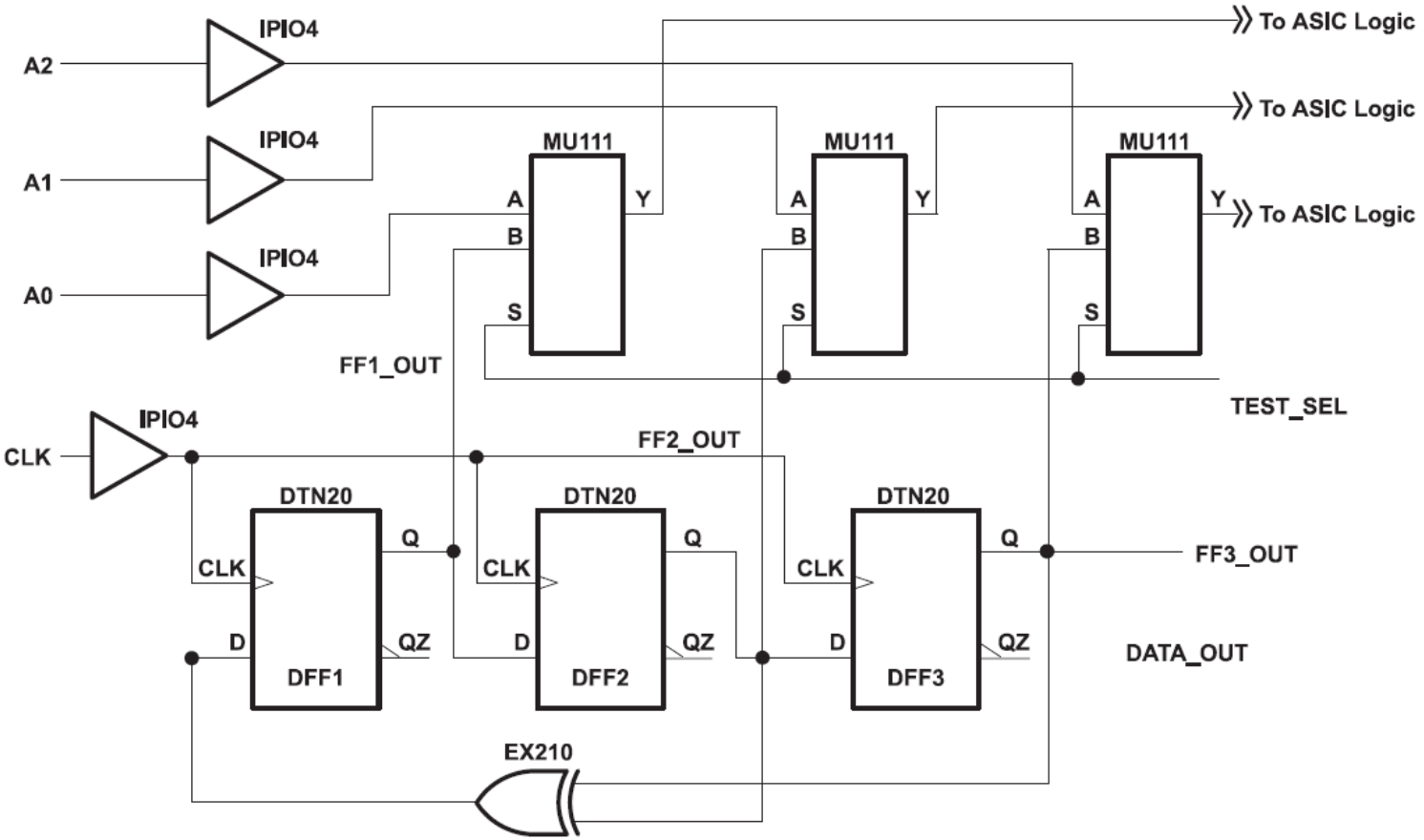
BISA: Built-In Self-Authentication

- Filling all unused spaces with a circuit that can easily test itself

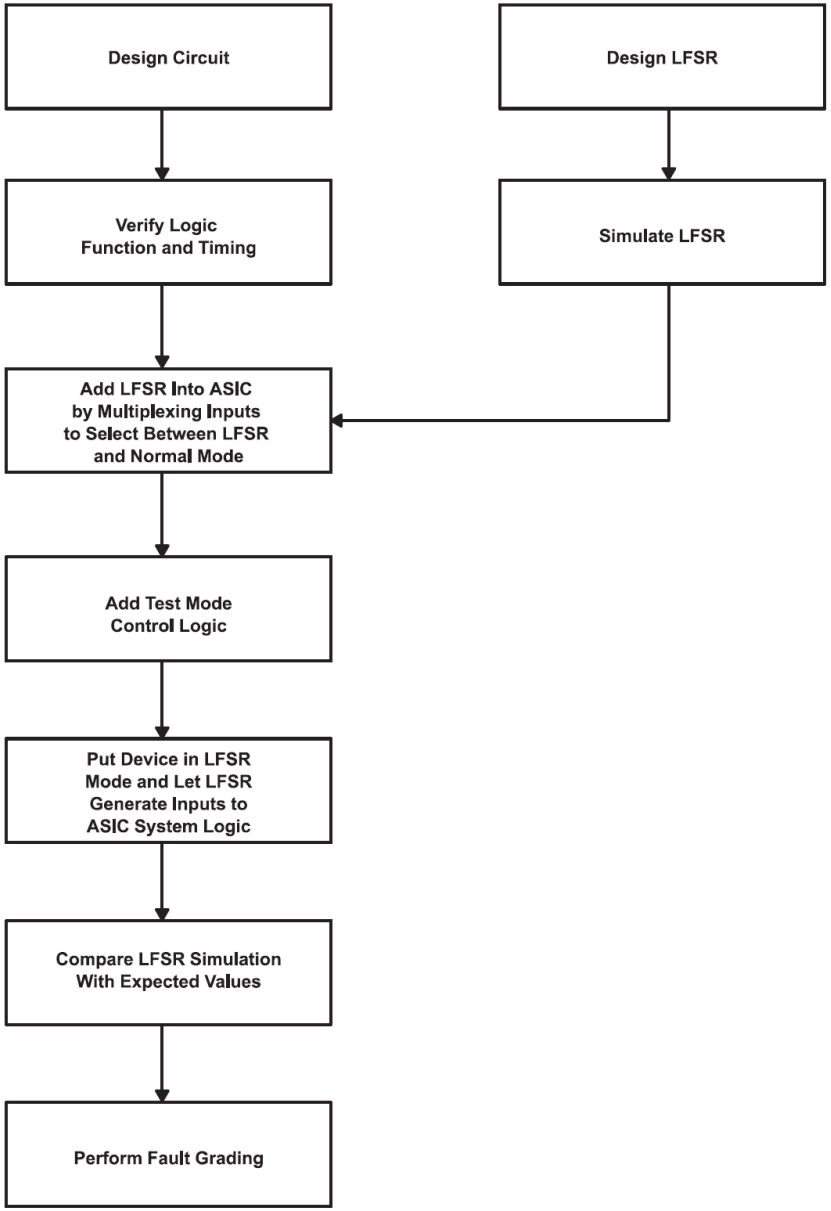


BISA: Built-In Self-Authentication

- Filling all unused spaces with a circuit that can easily test itself
  --LFSR with outputs muxed with ASIC inputs



BISA: Built-In Self-Authentication



- Flowchart for designing an LFSR into an ASIC