

TRUSTED ICs Proposers Day Metrics Discussion

Dan Wilt
240-228-5332
daniel.wilt@jhuapl.edu

The logo for Applied Physics Laboratory (APL) at Johns Hopkins University, consisting of the letters 'APL' in a large, bold, serif font.

The Johns Hopkins University
APPLIED PHYSICS LABORATORY

How to Measure TRUST

Conventional security metrics are not useful for TRUST (threat characterization, mitigation, and risk assessment)

- The IC fabrication process is inherently *untrusted*
- Securing this process from unauthorized access is *not possible*

We have chosen to focus on a detection-based approach, measuring the ability to detect alterations (Trojan Horses) in the IC's intended design

Since Trojans are difficult to categorize and characterize at high-level, the TRUST program focuses on feature (transistor) level metrics

- The probability of correctly detecting altered transistors, P_{d}^t
- The probability of falsely detecting unaltered transistors, P_{fa}^t

Proposers are required to state their program goals in these terms

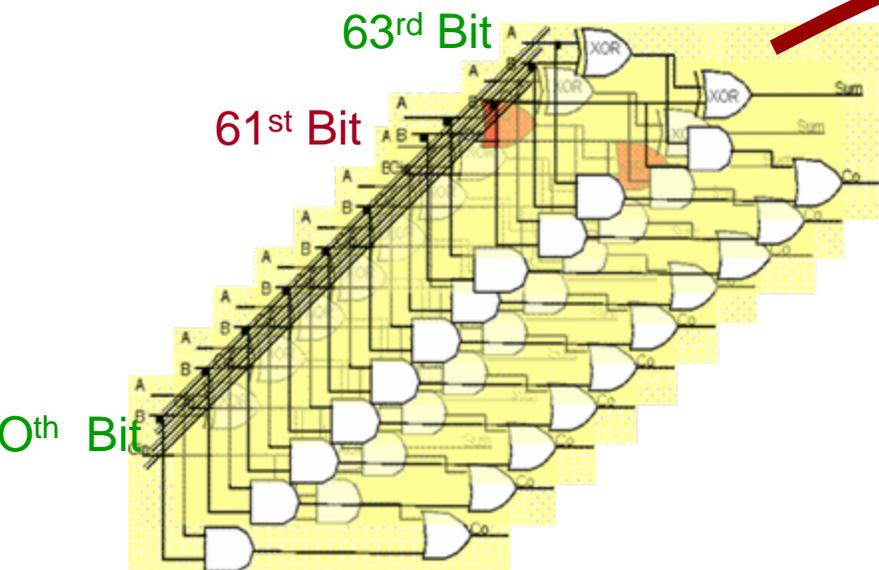
- Go/No-Go milestones will be based upon them

TRUST Program Goals

Process	Area 1—Hardware Validation Case 1 Trusted Design and Untrusted FAB			Area 2—Design Validation Case 2 Untrusted Design ASIC			Area 2—Design Validation Case 3 Untrusted Design FPGA		
	Phase 1	Phase 2	Phase 3	Phase 1	Phase 2	Phase 3	Phase 1	Phase 2	Phase 3
P_D	90.0%	99.0%	99.9%	80.0%	90.0%	99.0%	90.0%	99.0%	99.9%
P_{FA}	10^{-3}	10^{-5}	10^{-7}	10^{-3}	10^{-4}	10^{-6}	10^{-3}	10^{-5}	10^{-6}
Transistors Evaluated (n_t)	10^5	10^6	10^8	10^5	10^6	10^8	10^5	10^6	10^7
Time to Evaluate*	480 H	240 H	120 H	480 H	240 H	120 H	480 H	240 H	120 H

*Combined man hours plus wall clock time.

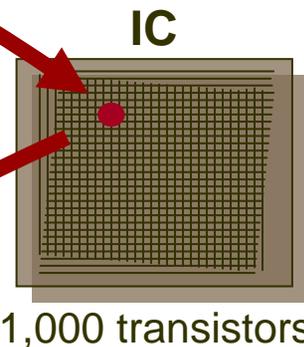
Example Pd/Pfa Calculations (at the transistor level)



64 Bit Adder

- 256 gates
- 2048 transistors
- 2 transistors mis-designed to cause arithmetic errors in the 61st bit of the adder

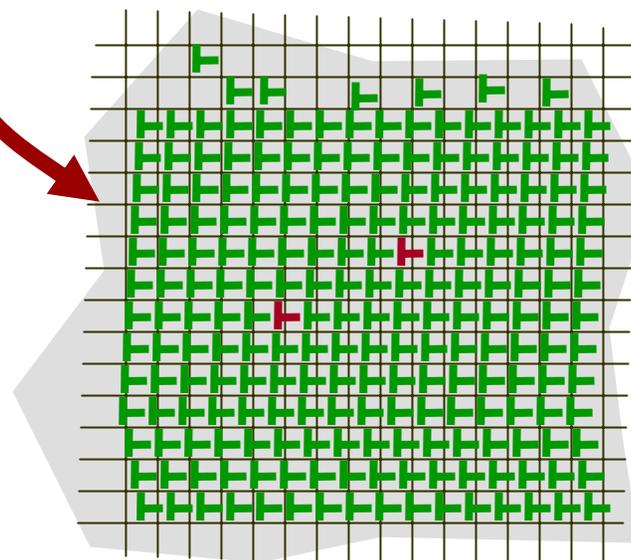
Adder region



1,000 transistors

10⁶ total transistors

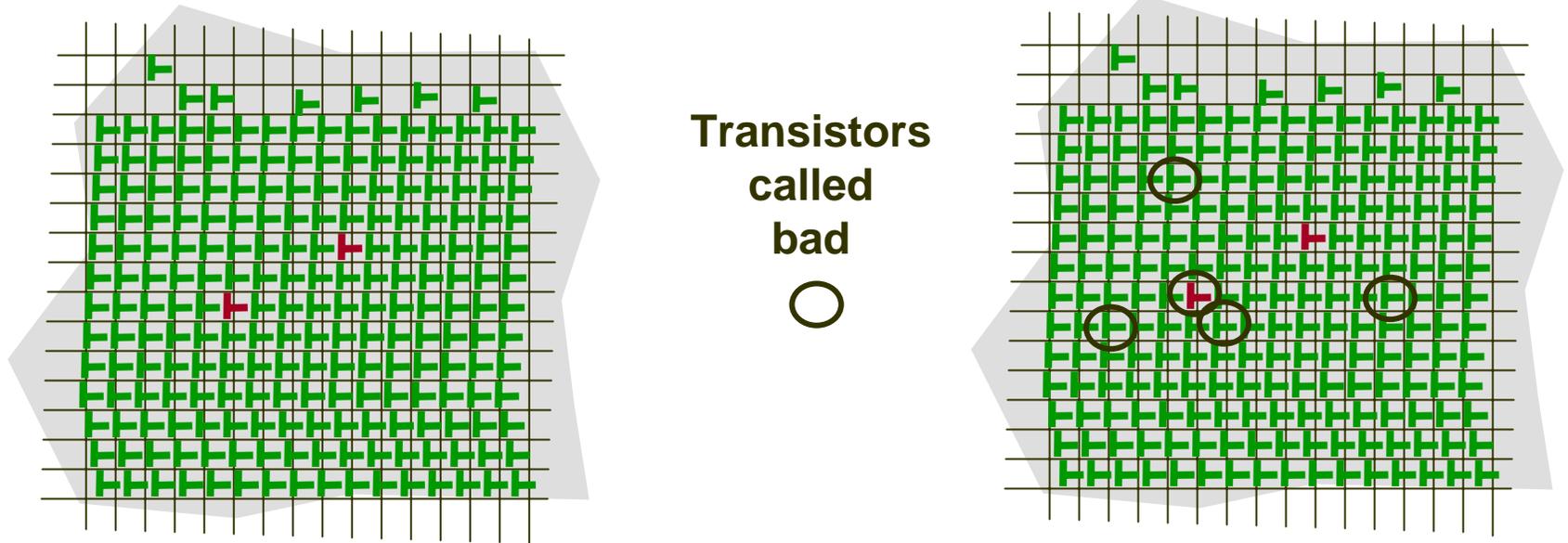
Adder region



Good Transistor

Trojan Transistor

Case 1 – Each Transistor Can Be Tested To Determine If The Entire Circuit Can Be Trusted



Case 1 – Test at Transistor Level

$$P_d^t = 1/2 = 50.0\%$$

$$P_{fa}^t = 4/10^6 = 4 \cdot 10^{-6}$$

Maximum likelihood result

Relating P_{d}^t and P_{fa}^t to Other Measures

For techniques not directly related to measuring transistors, how to capture the appropriate P_{d}^t and P_{fa}^t ?

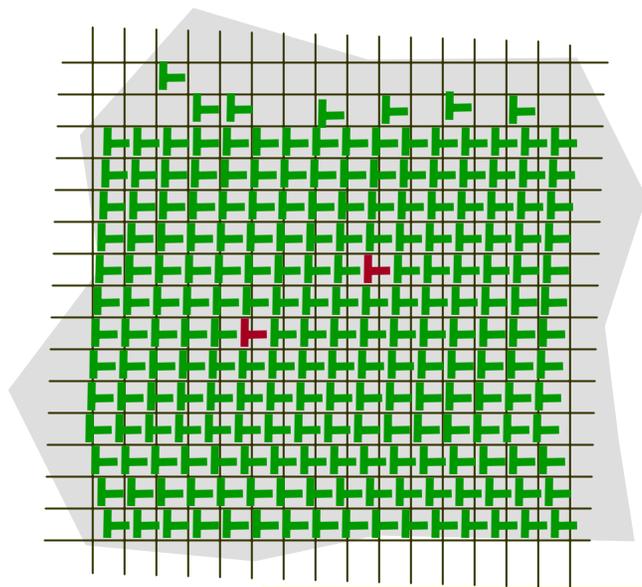
- A model is required to relate these techniques to transistor level
- There are many possible models – the BAA provides an example
- Different techniques are likely to have different models

One of the Metrics Team roles is to work with performers on this problem

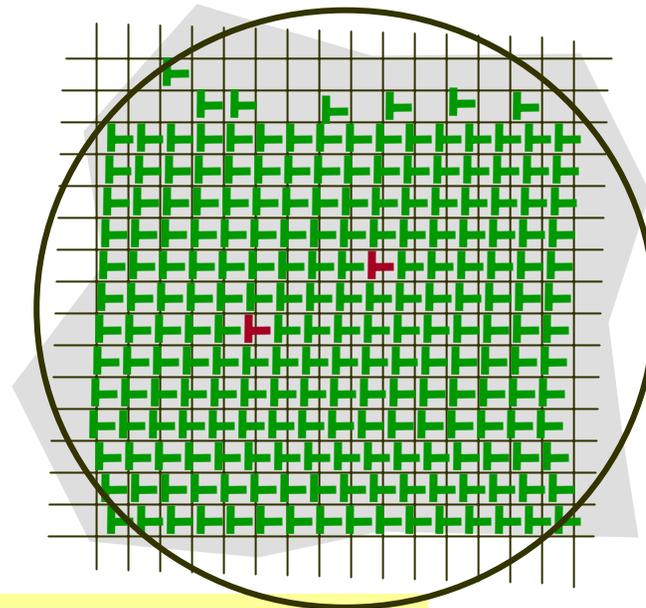
- Insure mathematical & statistical consistency and validity
- Insure compatibility with overall program goals

Case 2 – Cannot Test Each Transistor

(but can check to determine if the adder is working properly)



Adder called bad



Case 2 – Test at Functional Level
“2048 transistor adder” does not function properly

$$P_d^t = 2/2 = 100.0\%$$

$$P_{fa}^t = (2048-2)/10^6 = 2.046 \cdot 10^{-3}$$

} Implied model

IC-Level Decision Problem

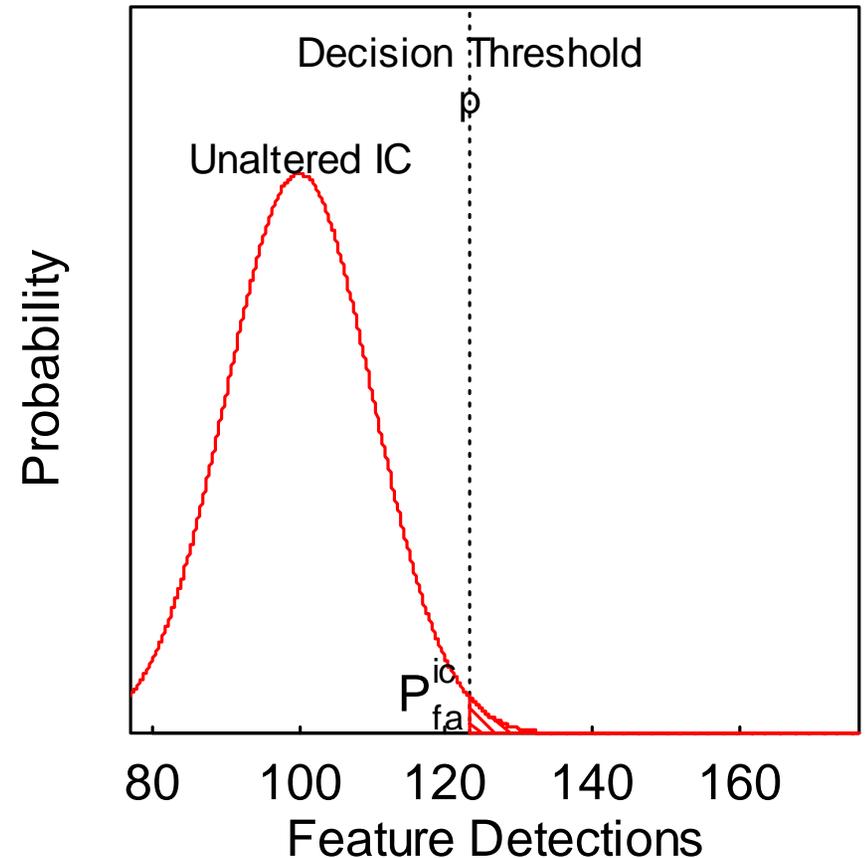
Ultimately TRUST is a decision made at IC level – to accept or reject an IC

In the later phases of the program, performers will relate P_d^t and P_{fa}^t metrics to IC-level decision probabilities P_d^{ic} and P_{fa}^{ic}

True Situation	IC-Level Decision	
	Reject	Accept
Trojan	Correct P_d^{ic}	Wrong $1 - P_d^{ic}$
No Trojan	Wrong P_{fa}^{ic}	Correct $1 - P_{fa}^{ic}$

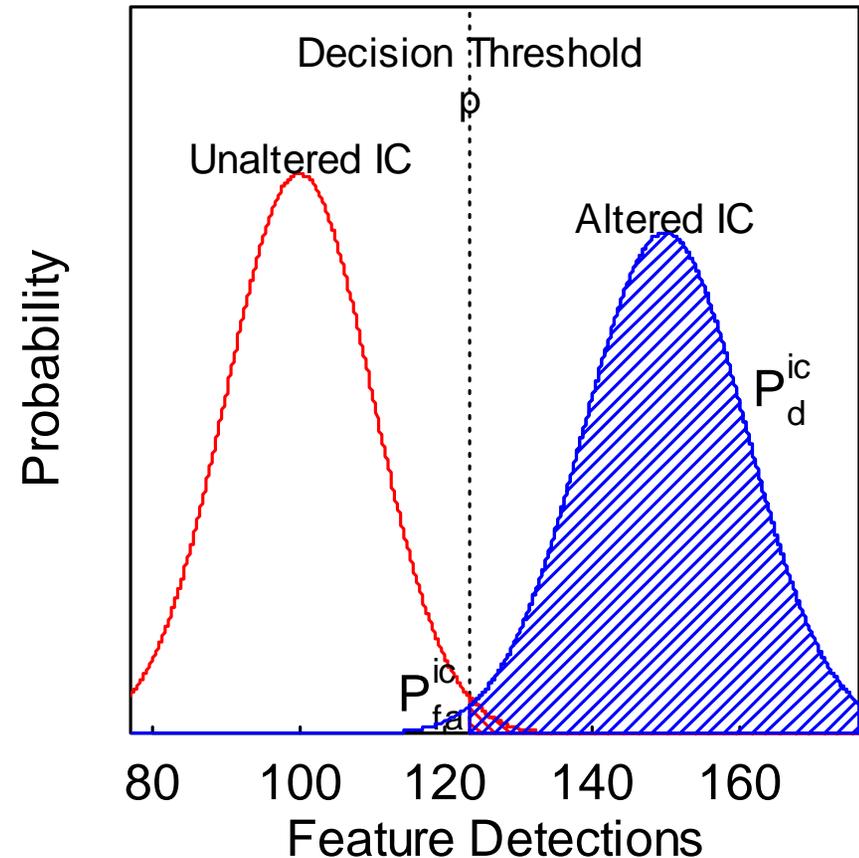
ROC-Curve Relationships

- Case 1 example: Compare a trusted design to an IC, and count mismatched circuit features (transistors)
- For an unaltered IC there is a distribution of falsely detected transistors (noise) with a mean of $n_t P_{fa}^t$
- At IC level, choose a decision threshold p to achieve a desired IC-level false alarm probability P_{fa}^{ic} given by the red shaded tail area under the distribution



ROC-Curve Relationships

- For an altered IC containing a Trojan, there is a shifted distribution of detected transistors containing both signal and noise
- The IC-level detection probability is determined by:
 - Transistor-level false alarms (noise)
 - Number of inserted rogue transistors
 - Probability of detecting those rogue transistors P_d^t
- The IC-level detection probability P_d^{ic} is given by the blue shaded tail area under the shifted distribution



Metrics Team Interaction with TRUST Performers

Our role in the efforts:

- Work with DARPA and government teams to define a metrics framework for TRUST:
 - For performers
 - For overall TRUST program evaluation
- Work with performers to identify TRUST metrics
- Advise performers on evaluating their metrics

Each Metrics Team participant signed an individual nondisclosure agreement with DARPA (which remains active)

- When desired, JHUAPL has also signed mutual NDAs with individual performers
- As the BAA indicates, any other JHUAPL effort will be firewalled from the Metrics team

Proposed Model for Metrics Team Interactions in the TRUST Program

The Metrics team will be working with all TRUST performers to:

- Help performers to define the metrics for their approach
- Help performers to design experiments that demonstrate that their approach can meet TRUST program goals
- Insure that the analysis techniques that are applied to the data are valid

The Metrics team members will sign mutual nondisclosure agreements (if needed)

In addition, the Metrics team will be working with the Test Article Team and the Red Team to insure that the test articles and red-team activities support the metrics goals of the TRUST program

Finally, the Metrics team will assist DARPA in doing overall metrics evaluations for TRUST at the programmatic level



Thank You!
Questions?