Fault Diagnosis and Tolerance in Cryptography (FDTC2017)

Exploiting Bitflip Detector for Non-Invasive Probing and its Application to Ineffective Fault Analysis

Takeshi Sugawara*, Natsu Shoji*, Kazuo Sakiyama*

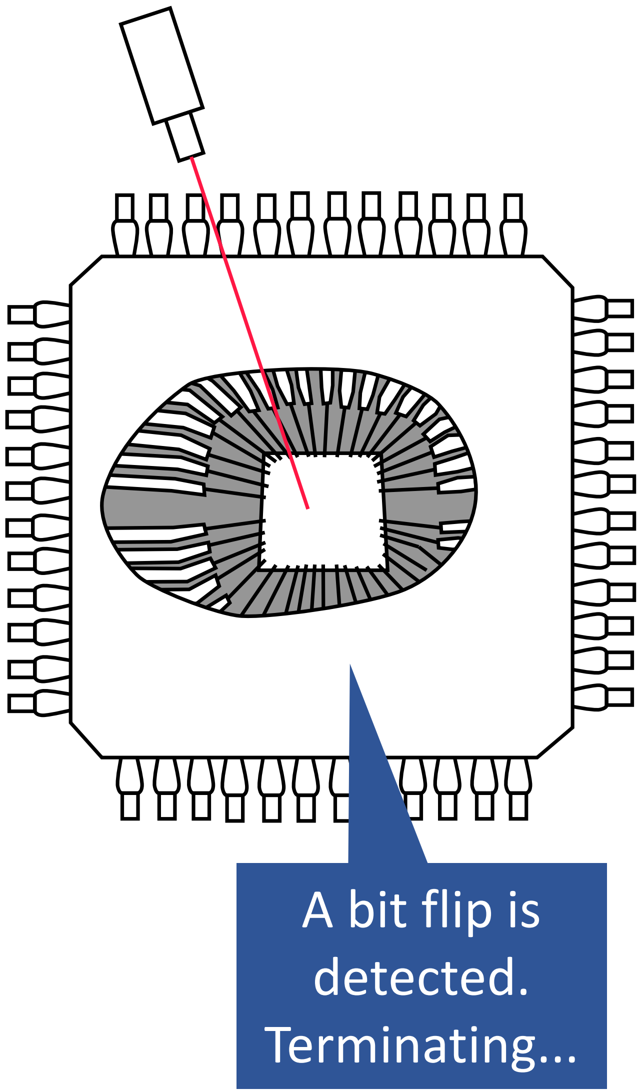
Kohei Matsuda**, Noriyuki Miura**, Makoto Nagata**

*The University of Electro-Communications, Tokyo

**Kobe University, Kobe

Overview

- An attack on a sensor-based countermeasure
- **Fault detector as side channel**
  - Obtaining 1-bit side-channel information by observing how the fault detector reacts to a laser fault injection
- New fault analysis using the above leakage based on **linear cryptanalysis**

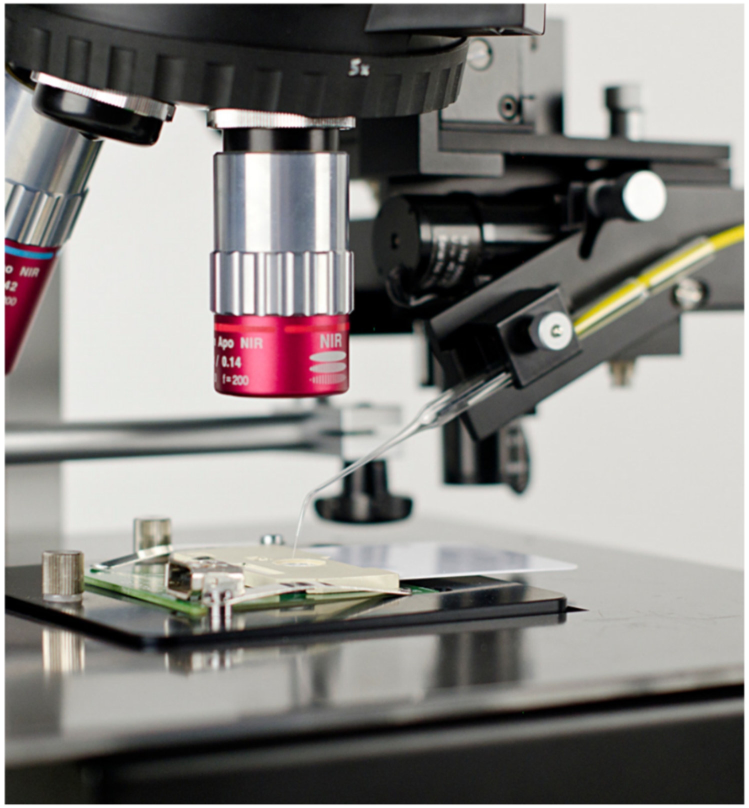


Background:

Sensor-based Countermeasure against Laser Fault Injection

Laser Fault Injection (LFI)

- One of the strongest fault injection technique
  - Instruments are commercially available for testing



The image is taken from https://www.riscure.com

Sensor-based Countermeasure against Laser Fault Injection

- Detects laser fault injection by using sensors and terminates a sensitive operation upon detection

- Photo detectors
  - Disadvantage: limited coverage
    - Laser can be focused in order to avoid photo-sensitive area

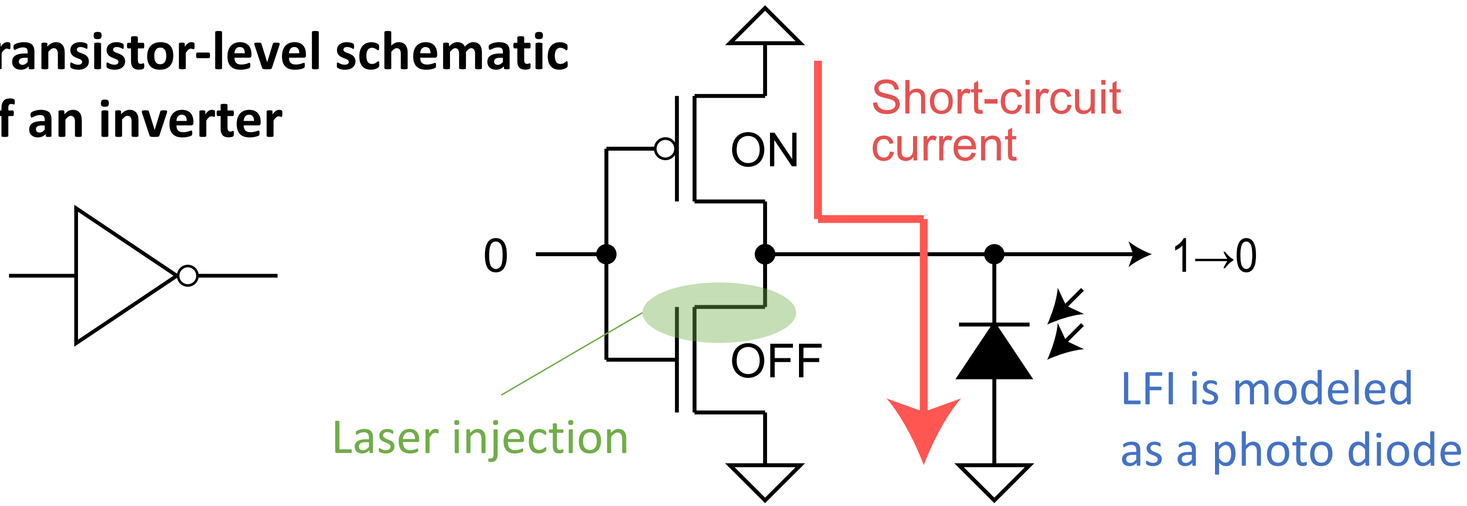
- **Substrate bounce monitoring***
  - Monitors temporal short-circuit current for detecting laser injection

*K. Matsuda, et al., "On-Chip Substrate-Bounce Monitoring for Laser-Fault Countermeasure," AsianHOST 2016.

Mechanism behind bit flip by LFI

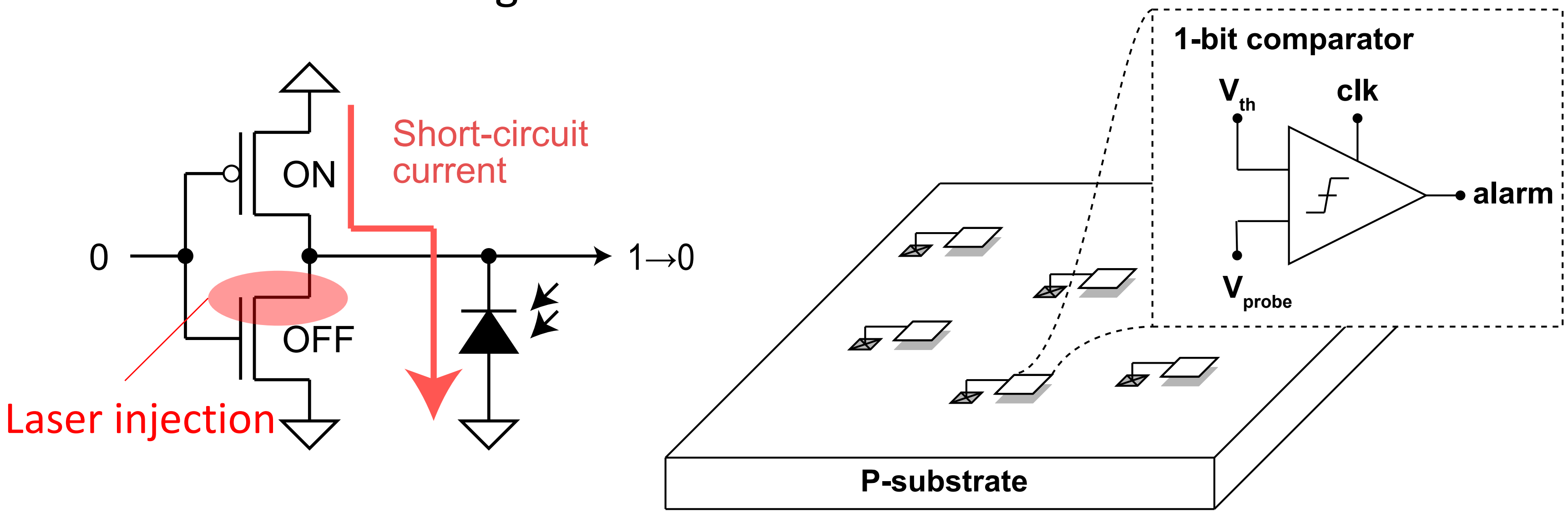
- LFI-induced photo current can short-circuit VDD and GND, which makes voltage drop at the output node, resulting in a bit flip
- **Short circuit is unavoidable for bit flip**

Transistor-level schematic of an inverter



Detecting Laser Fault Injection by Substrate Bounce Monitoring*

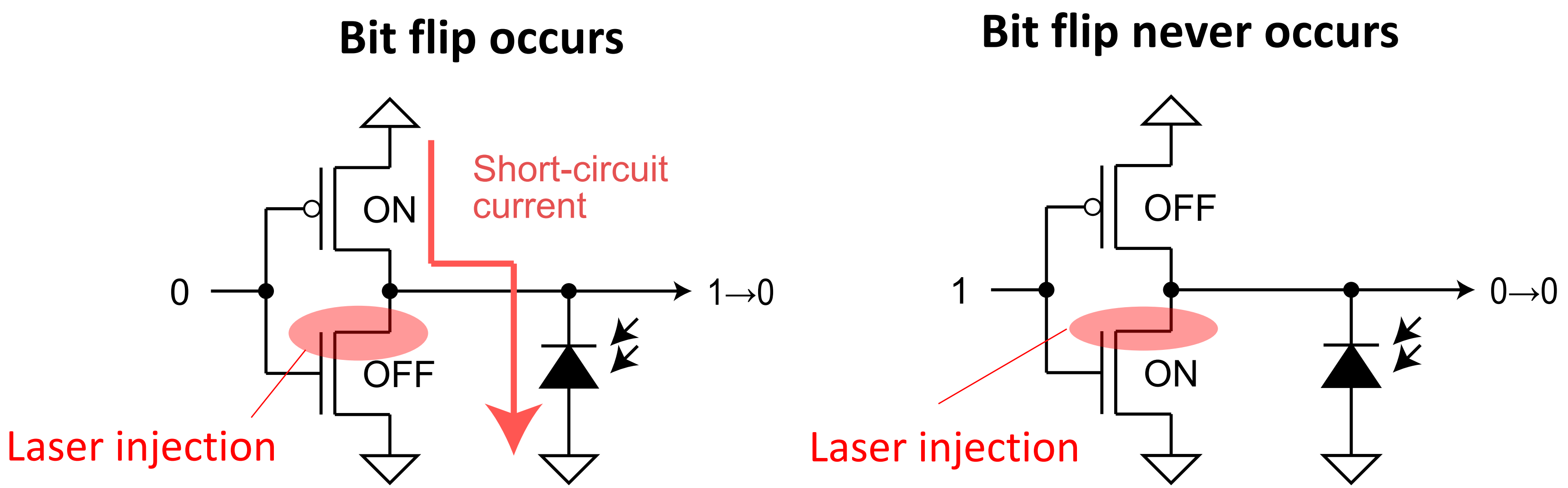
- Using distributed on-chip monitors to detect voltage bounce in silicon substrate caused as a side effect of short circuit
- **Benefit: better coverage**
  - Since the voltage bounce propagates through substrate, a sensor can cover a large area



*K. Matsuda, et al., "On-Chip Substrate-Bounce Monitoring for Laser-Fault Countermeasure," AsianHOST 2016.

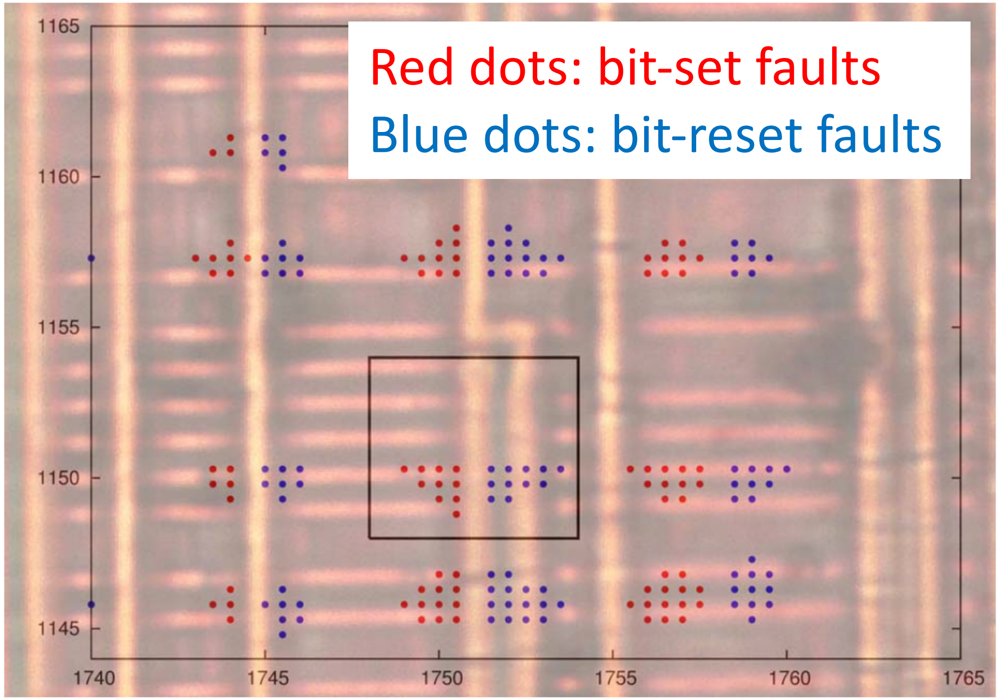
Bit-Set/Reset Faults by LFI

- Unidirectional fault
  - **Bit-set fault**: 0→1 flip only
  - **Bit-reset fault**: 1→0 flip only



Bit-Set/Reset Faults by LFI cont.

- Bit-set/reset faults of SRAM



The image is taken from the paper*

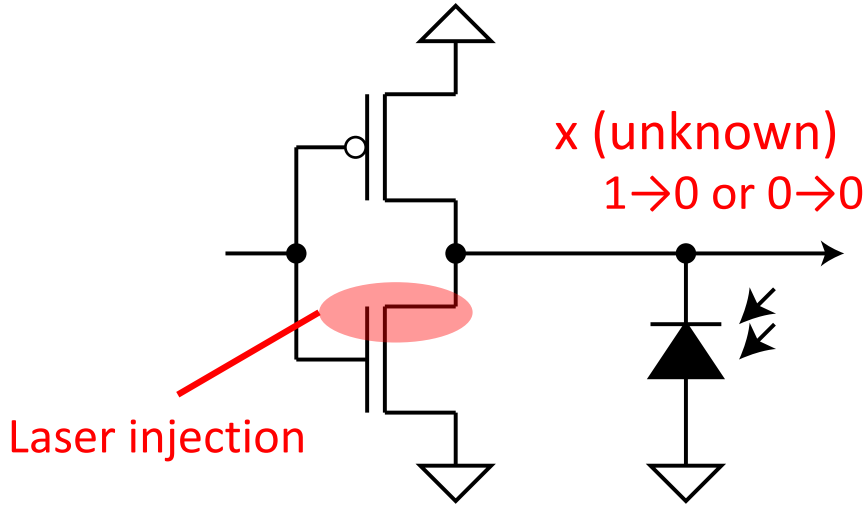
*C. Roscian, A. Sarafianos, J.-M. Dutertre and A. Tria, “Fault Model Analysis of Laser-induced Faults in SRAM Memory Cells”, FDTC 2013

First part:

Exploiting Bitflip Detector for Non-Invasive Probing and its Application to Ineffective Fault Analysis

Idea: Learning an Internal State by Observing Alarm from Sensor

- Prerequisite: position for LFI that causes a bit-set/reset fault



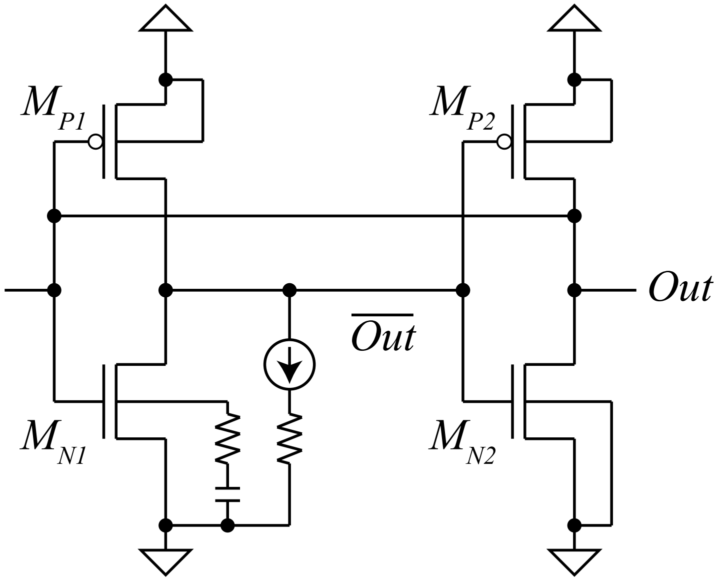
If an alarm is observed:
x = 1
otherwise:
x = 0

- Presence/absence of an alarm (i.e., bit flip) directly corresponds to an internal bit value
  - **Attacker successfully probes 1-bit signal non-invasively**

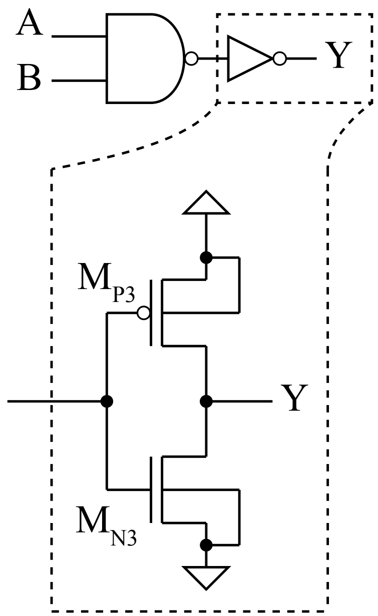
Circuit structures that can be potentially probed

- The discussion on inverter extends to other primitives
  - Cross-coupled inverters in **SRAM** and **flipflop**
  - Inverters and buffers in **logic gates**

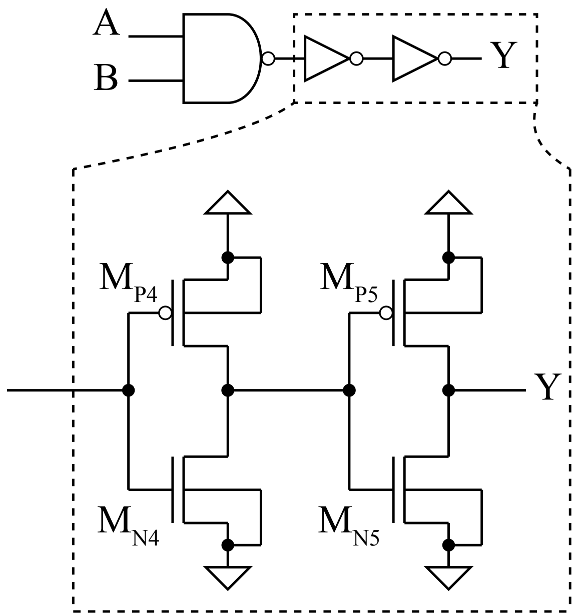
Cross-coupled inverters



AND gate



NAND gate with buffer



Categorization

- **Probing attack**
  - Similarity: attacker has circuit-level resolution and recovers a bit value in a circuit

- **Ineffective fault analysis / safe error attack**
  - Similarity: the attacker retrieves information by presence/absence of a fault

Second part:

Exploiting Bitflip Detector for Non-Invasive Probing and its Application to Ineffective Fault Analysis

(of AES)

Attack using the 1-bit leakage

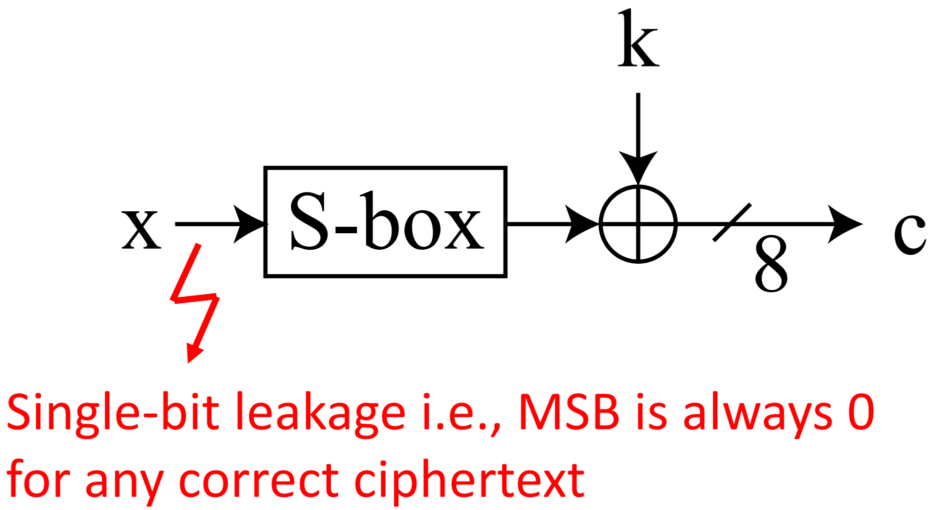
- Known- or chosen-plaintext attacks
  - Conventional **probing attack** on AES by Schmidt and Kim works*

- Ciphertext-only attack
  - **A new setting: correct-ciphertext-only attack**
  - Only correct ciphertext is released. Why?
    - A sensor detects a fault and stops releasing a faulty ciphertext

*J.-M. Schmidt and C. H. Kim, “A Probing Attack on AES,” WISA 2008, LNCS 5379, pp. 256–265, 2008.

Attack on 10th (last) round of AES using Correct Ciphertext Only

- Example: bit-reset fault on the MSB of an Sbox at 10th round



$$MSB(Sbox^{-1}(c \oplus \hat{k})) = 0$$

1. Guess 8-bit key $\hat{k}$
2. Calculate x using $\hat{k}$ and c
3. Check if MSB of x is 0

- The key space is halved for each correct ciphertext
  - Roughly 8 ciphertexts are needed to uniquely determine a key

Extending the attack to 9th round

- Why extend?
  - The previous attack recovers 1-byte key for each LFI position
  - Attacker wants to recover more bytes for each LFI position

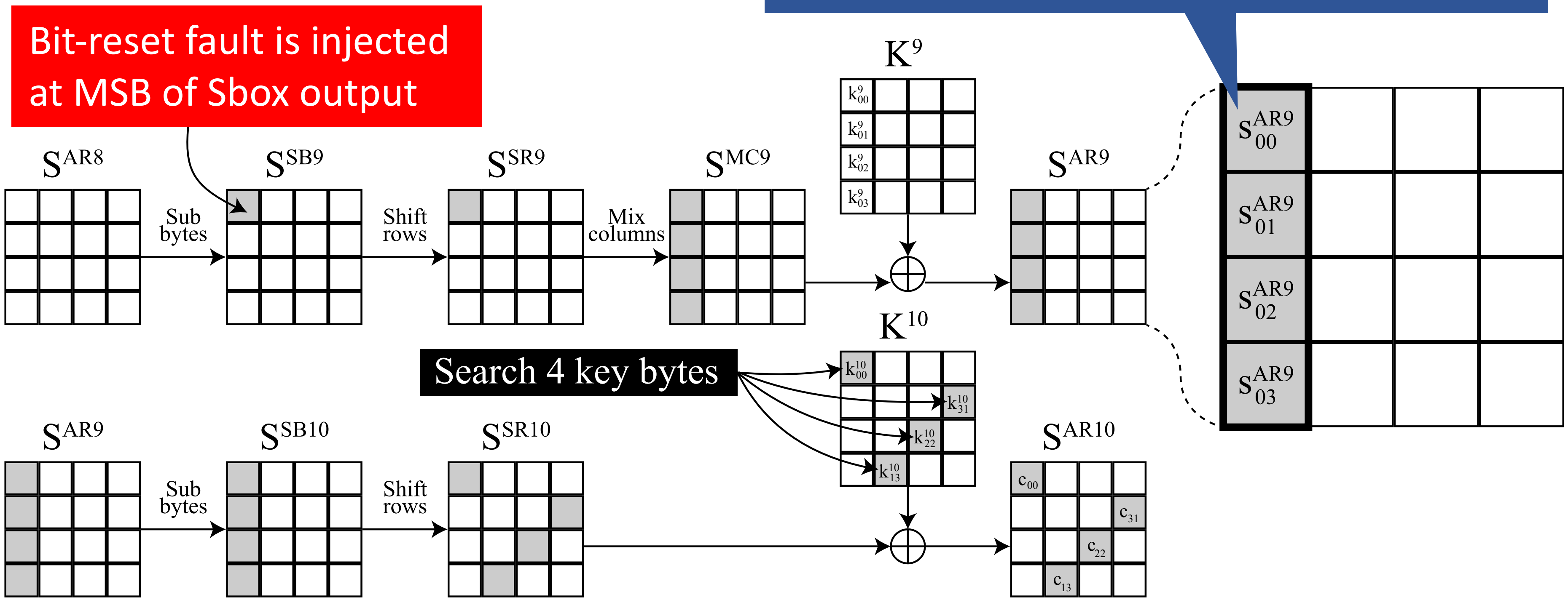
- A common strategy is to induce a **small difference** in either internal state or ciphertext, but, ...
- **Difficulty: no small difference**
  - Only correct ciphertexts are available
  - Output difference is uncontrollable in known-ciphertext setting

- Idea: Using technique from **linear cryptanalysis**

Attack on 9th round of AES using Correct Ciphertext Only (outline)

- A column in S^{AR9} is recovered using a 32-bit key guess
- The key guess is checked using a constraint on S^{AR9}

Check if the columns satisfies an equation



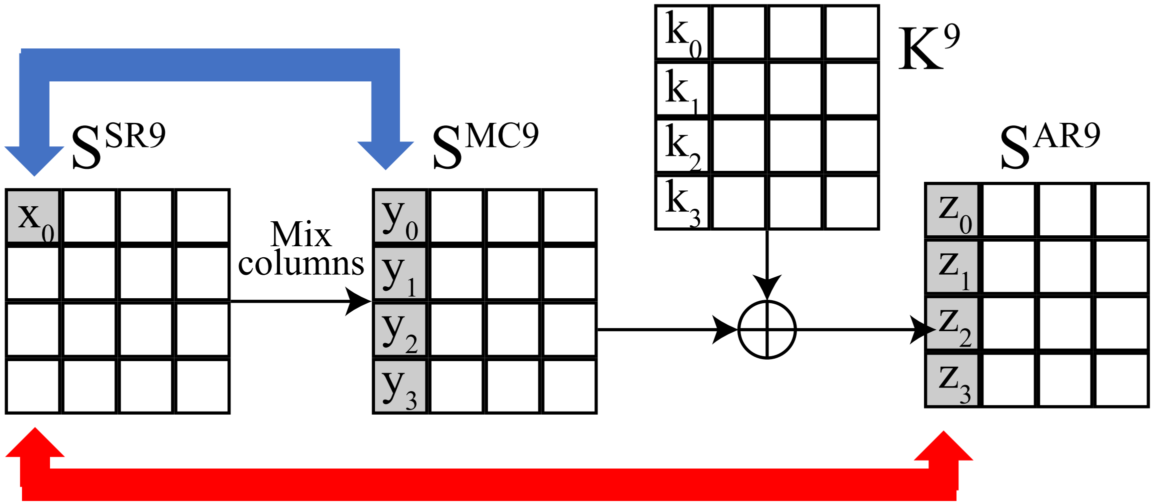
Linear equation

A linear equation determined by MixColumn (focusing on MSB of x_0)

$$x_0[7] = y_0[6] \oplus y_0[5] \oplus y_0[4] \oplus y_1[7] \oplus y_1[6] \oplus y_1[4]$$
$$\oplus y_2[7] \oplus y_2[5] \oplus y_2[4] \oplus y_3[7] \oplus y_3[4]$$

$X[i]$ is i-th bit of a byte X

AES states at 9th round

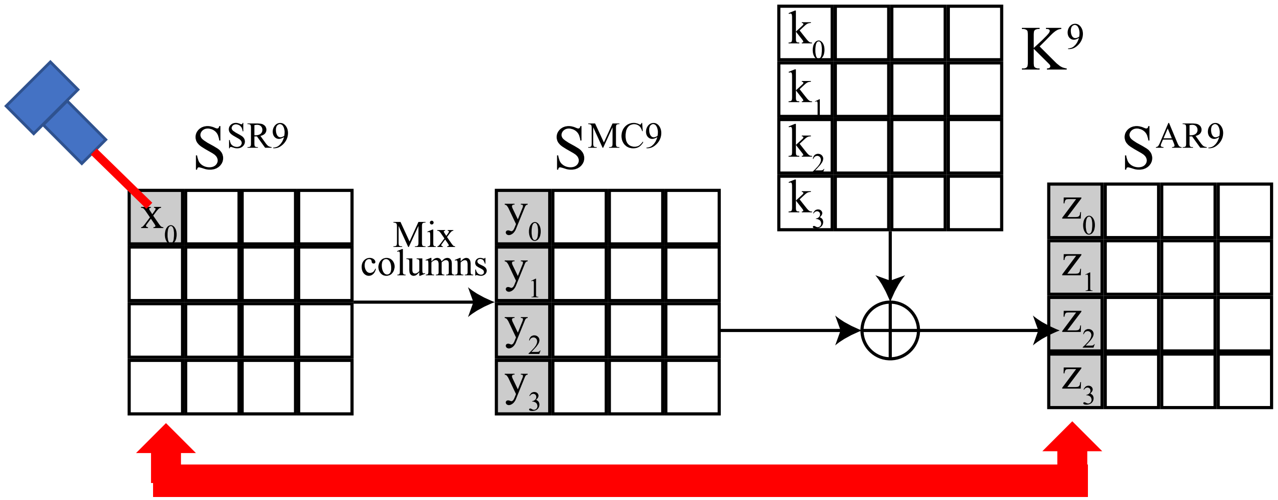


The unknown key $k_0, \ldots, k_3$ degenerates to a 1-bit constant

$$x_0[7] \oplus const = z_0[6] \oplus z_0[5] \oplus z_0[4] \oplus z_1[7] \oplus z_1[6] \oplus z_1[4]$$
$$\oplus z_2[7] \oplus z_2[5] \oplus z_2[4] \oplus z_3[7] \oplus z_3[4]$$

Linear equation cont.

Bit-reset fault on MSB of x_0, then $\boldsymbol{x_0[7] = 0}$ for any correct ciphertext



$$x_0[7] \oplus const = z_0[6] \oplus z_0[5] \oplus z_0[4] \oplus z_1[7] \oplus z_1[6] \oplus z_1[4] \oplus z_2[7] \oplus z_2[5] \oplus z_2[4] \oplus z_3[7] \oplus z_3[4]$$

$$\Downarrow$$

$$const = z_0[6] \oplus z_0[5] \oplus z_0[4] \oplus z_1[7] \oplus z_1[6] \oplus z_1[4] \oplus z_2[7] \oplus z_2[5] \oplus z_2[4] \oplus z_3[7] \oplus z_3[4]$$

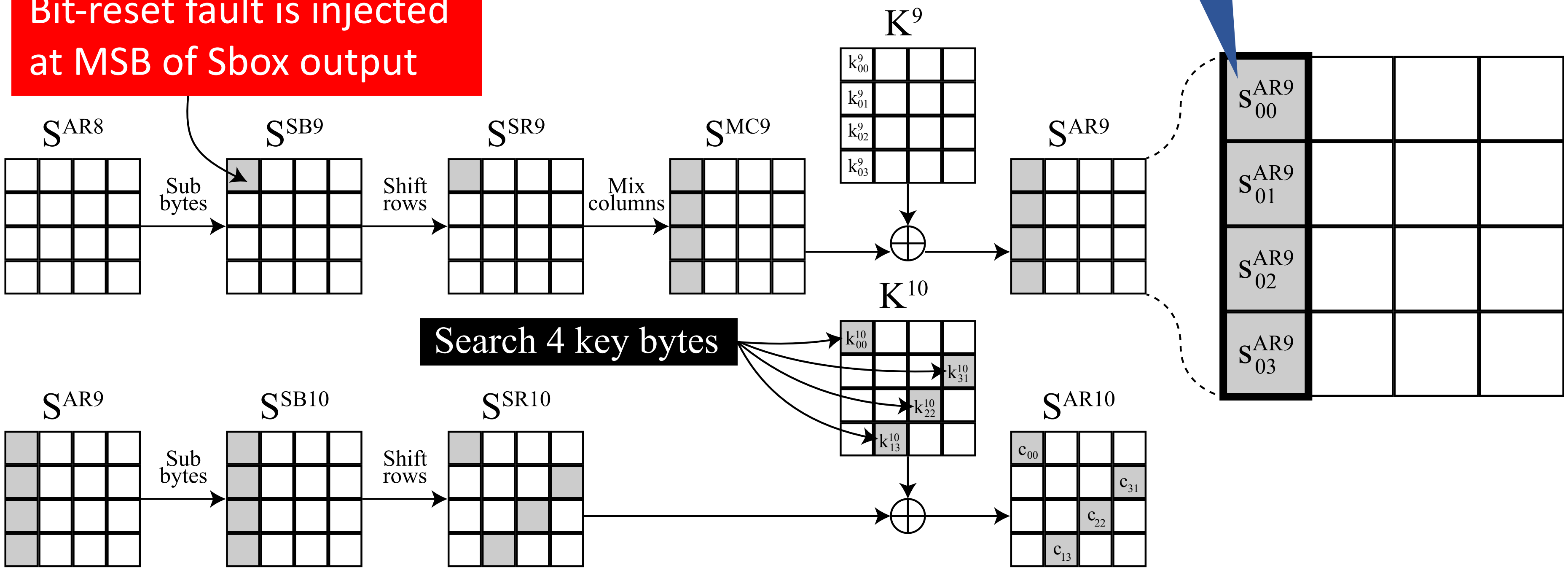
A constraint solely on S^{AR9}

Attack on 9th round of AES using Correct Ciphertext Only

Check if the columns satisfies the equation

$$const = z_0[6] \oplus z_0[5] \oplus z_0[4] \oplus z_1[7] \oplus z_1[6] \oplus z_1[4] \oplus z_2[7] \oplus z_2[5] \oplus z_2[4] \oplus z_3[7] \oplus z_3[4]$$

Bit-reset fault is injected at MSB of Sbox output



Comparison

- Key space is halved for each correct ciphertext
  - For N ciphertexts, key space is reduced to 2^{-N+1}

- The number of LFI positions and ciphertexts needed to recover 16-byte round key:

	The number of LFI positions	**The number of correct ciphertexts**
Attack on 10^{th} round	16	8*16=128
Attack on 9^{th} round	**4**	33*4=132

Conclusion & future work

- Conclusion
  - Bit-flip sensor can be used as a side-channel oracle
  - Ineffective fault analysis on AES using the above leakage

- Future work
  - Experimental verification
  - Further study on probing attack
  - Extension to other sensor-based countermeasures