Global Software Technology Summit (GSTS) 2024

Reliability challenges and solutions in the new age of artificial intelligence-powered devices

Annachiara Ruospo

Assistant Professor (RTD-A)

CAD & Reliability Group, DAUIN

Politecnico di Torino, Italy

July 3rd, 2024

Politecnico di Torino

1859

Acknowledgements

Prof. Matteo Sonza Reorda, CAD & Reliability Group

Prof. Riccardo Cantoro, CAD & Reliability Group

Prof. Ernesto Sanchez, CAD & Reliability Group

Outline

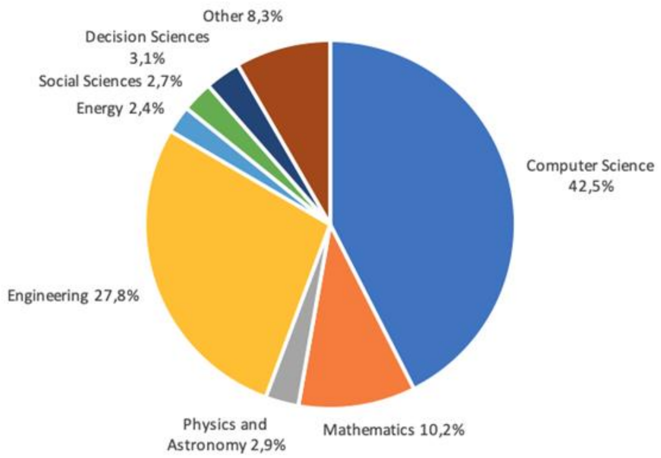
1. Introduction
   - General Introduction
   - Long-term expertise
2. Need for Reliability in AI systems
3. Major Challenges and Solutions
   - Complexity of the reliability assessment
     - Statistical Fault Injections
   - Need for effective in-field fault detection solutions
     - Image Test Libraries
4. Conclusions and Future Directions

Politecnico di Torino

- Turin, Italy
- Founded in 1859 as a school for the army officers
- About 40,000 students
- Consistently ranked among the top technical universities in Europe
- Strongly connected with the surrounding economical environment, rich in companies especially active in the automotive, aerospace and telecommunication areas
- 11 departments managing research, teaching and technology transfer.

Department of Control and Computer Engineering (DAUIN)

- Core Research and Teaching Activities on Computer Sciences and Engineering
    - **84** Faculty members
    - **30** Administrative and Technical staff members
    - **198** PhD students and Research Assistants
- Deep involvement in Research Centers and Industrial Collaborations
- **DAUIN HPC:** Academic Computing Center
    - More than **520** hosted projects
    - More than **50** teaching courses.



The CAD & Reliability group

- Major group within the Dept. of Control and Computer Engineering
- 9 faculty members and about 20 young researchers (post-doc, PhD students, visitors)
- Long expertise in relationships with industry, university, research centers worldwide
- Strong involvement in European projects
- International reputation: average Scholar h-index > 20
- www.cad.polito.it

The CAD & Reliability group

Matteo Sonza Reorda is the leader of the CAD & Reliability Group.

- Full Professor with the Department of Control and Computer Engineering, Politecnico di Torino, Italy
- He published more than 400 papers in the area of test and fault-tolerant design of reliable circuits and systems
- He received several best paper awards at major international conferences
- IEEE Fellow
- h-index=52

matteo.sonzareorda@polito.it

Matteo Sonza Reorda
Full Professor
Politecnico di Torino, Italy

Long-term Expertise

- Reliability-oriented *hardware* and *software* hardening solutions
  - DWC / TMR / ABFT / Code hardening
- In-field software-based fault detection mechanisms for automotive and space SoCs
  - Software Test Libraries
- GPU Reliability
  - Software-Based Self Test mechanisms for different GPU units (e.g., functional units, warp scheduler, pipeline registers)
- Reliability and functional safety assessment solutions
- AI Reliability and Safety
  - Analysis of the impact of permanent and transient faults on AI algorithms
  - Artificial Neural Networks.

Outline

1. Introduction
2. **Need for Reliability in AI systems**
3. Major Challenges and Solutions
    - Complexity of the reliability assessment
        - Statistical Fault Injections
    - Need for effective in-field fault detection solutions
        - Image Test Libraries
4. Conclusions and Future Directions

Are AI models inherently resilient and able to tolerate hardware-induced failures?

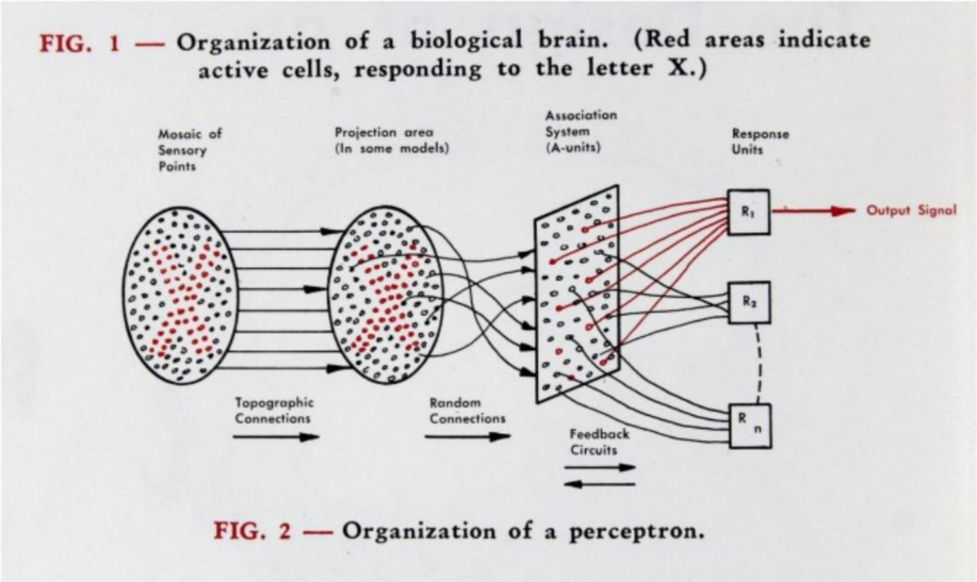
ANNs and resilience to faults

Rosenblatt's perceptron: a «**perfect mimic** of the human brain».

Can Artificial Neural Networks be considered inherently resilient ?

- Biological networks
- Distributed processing
- Overprovisioning

Plastic ability to self-repair from damage.



Division of Rare and Manuscript Collections

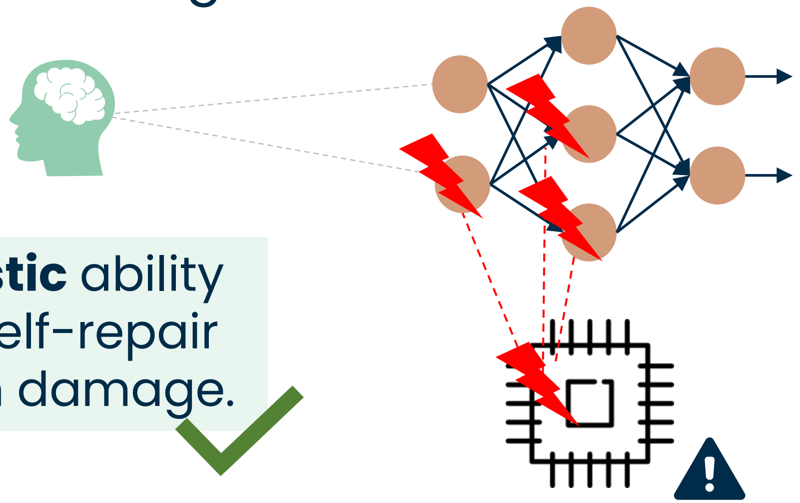
An image of the perceptron from Rosenblatt's "The Design of an Intelligent Automaton," Summer 1958.

Image from: Professor's perceptron paved the way for AI – 60 years too soon | Cornell Chronicle

ANNs and resilience to faults

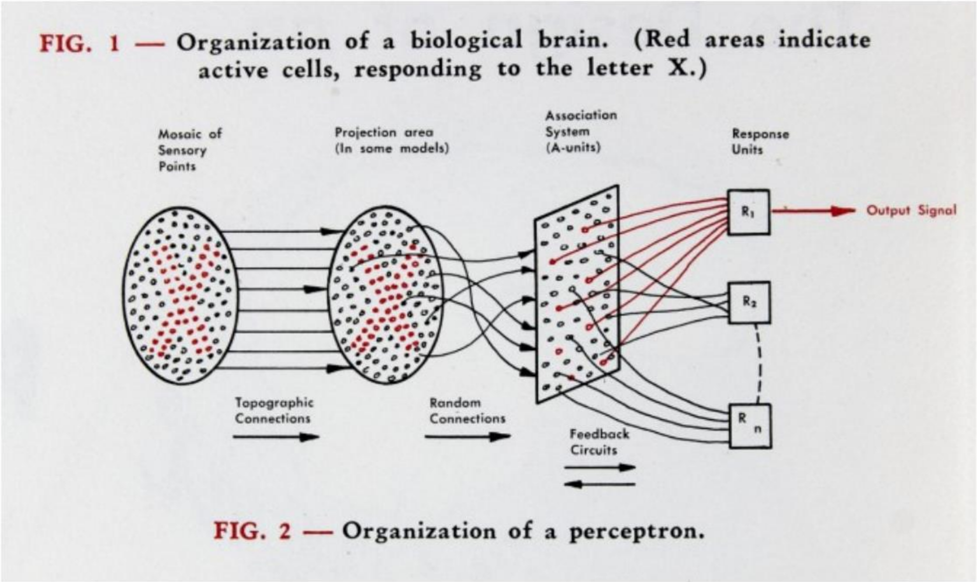
Can Artificial Neural Networks be considered inherently resilient ?

- Biological networks
- Distributed processing
- Overprovisioning



Plastic ability to self-repair from damage.

Rosenblatt's perceptron: a «**perfect mimic** of the human brain».



Division of Rare and Manuscript Collections

An image of the perceptron from Rosenblatt's "The Design of an Intelligent Automaton," Summer 1958.

Artificial Neural Networks:

- No self-repair from errors
- **Single** physical faults – **many** artificial neurons affected.

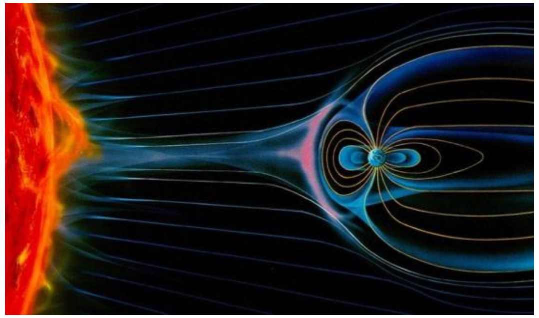
Image from: Professor's perceptron paved the way for AI – 60 years too soon | Cornell Chronicle

Advance in semiconductor technology

Current devices are manufactured with the most advanced semiconductor technologies (to achieve performance and reduce power consumption)

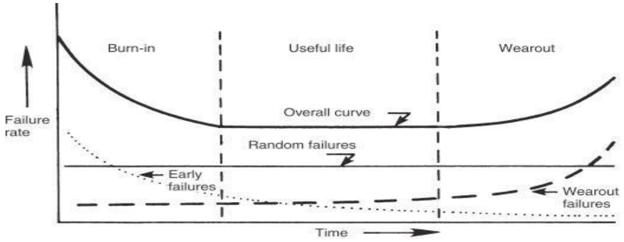
- **New technologies are more critical in terms of reliability**
  - New defects
  - Faster aging
  - More susceptible to radiations

Random Hardware Faults



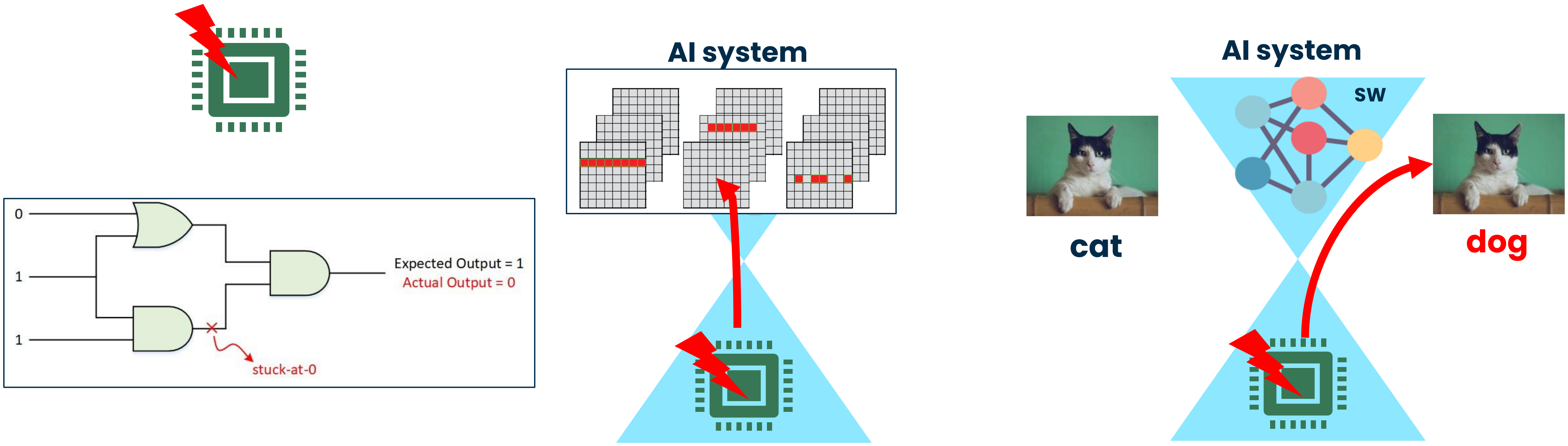
Transistors Do Age … Like Humans

- Transistors slow down over time
- Rate of aging very related to stress



Fault-Error-Failure in AI systems

Fault —activation→ Error —propagation→ Failure



How frequently?

2021 – Call to action - Facebook

Silent Data Corruptions at Scale

Sneha Pendharkar
Facebook, Inc.
spendharkar@fb.com

Matt Beadon
Facebook, Inc.
mbeadon@fb.com

Chris Mason
Facebook, Inc.
clm@fb.com

hakravarthy
book, Inc.
@fb.com

Bharath Muthiah
Facebook, Inc.
bharathm@fb.com

Sriram Sankar
Facebook Inc.
sriramsankar@fb.com

n have negative impact on large-
Cs are not captured by error re-
ntral Processing Unit (CPU) and

machine learning inferences, ranking and recommendation systems. However, it is our observation that computations are not always *accurate*. In some cases, the CPU can perform computations incorrectly. For example, when you perform 2x3, the CPU may give

2021 – Call to action - Google

Cores that don't count

Peter H. Hochschild
Paul Turner
Jeffrey C. Mogul
Google
Sunnyvale, CA, US

Rama Govindaraju
Parthasarathy
Ranganathan
Google
Sunnyvale, CA, US

David E. Culler
Amin Vahdat
Google
Sunnyvale, CA, US

Abstract

We are accustomed to thinking of computers as fail-stop, especially the cores that execute instructions, and most system software implicitly relies on that assumption. During most of the VLSI era, processors that passed manufacturing tests and were operated within specifications have insulated us from this fiction. As fabrication pushes towards smaller feature sizes and more elaborate computational structures, and as increasingly specialized instruction-silicon pairings are introduced to improve performance, we have observed ephemeral

MI, USA. ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3458336.3465297

1 Introduction

Imagine you are running a massive-scale data-analysis pipeline in production, and one day it starts to give you wrong answers – somewhere in the pipeline, a class of computations are yielding corrupt results. Investigation fingers a surprising cause: an innocuous change to a low-level library. The change itself was correct, but it caused servers to make heavier use of otherwise

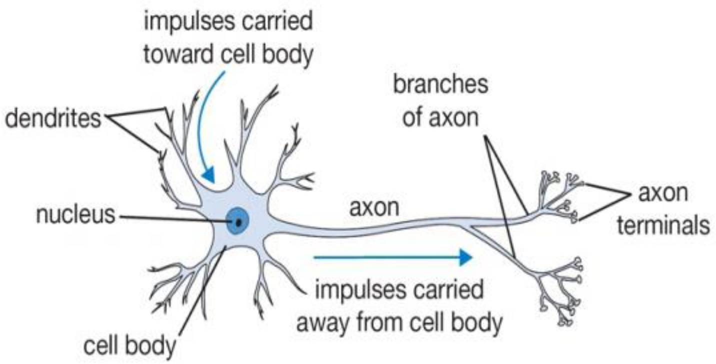
Silent Data Corruption (SDC)
Errors in data (e.g., main memory, disk, or in other storage) that occur without any indication or warning, typically at the **hardware level**

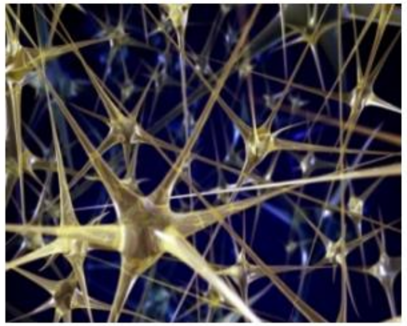
- random causes such as manufacturing defects, alpha particles, cosmic rays.

Outline

1. Introduction
2. Need for Reliability in AI systems
3. **Major Challenges and Solutions**
    - **Complexity of the reliability assessment**
        - Statistical Fault Injections
    - Need for effective in-field fault detection solutions
        - Image Test Libraries
4. Conclusions and Future Directions

Complexity of modern AI systems (I)



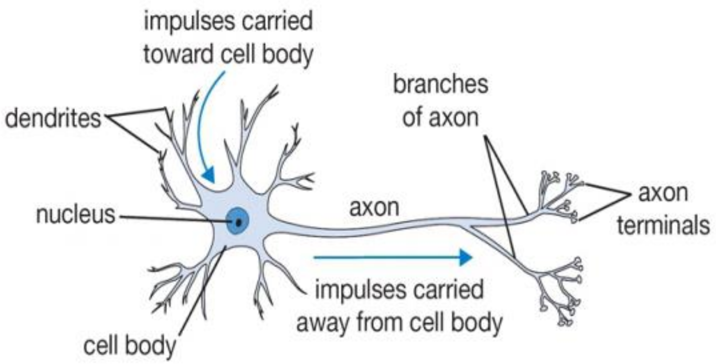


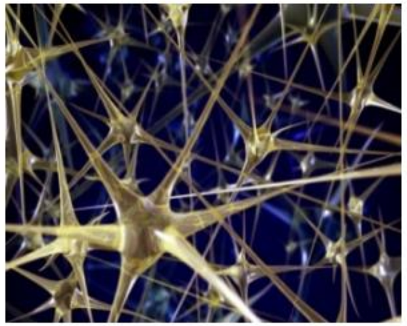
The basic computational unit of the brain is a **neuron**

- 86B neurons in the brain (1 billion = 10^9)
- connected with nearly 10^{14}– 10^{15} synapses

Source Image: https://www.science.org/content/article/brain-cells-chat-even-without-synapse

Complexity of modern AI systems (I)





Source Image: https://www.science.org/content/article/brain-cells-chat-even-without-synapse

The basic computational unit of the brain is a **neuron**

- 86B neurons in the brain (1 billion = 10^9)
- connected with nearly $10^{14} - 10^{15}$ synapses

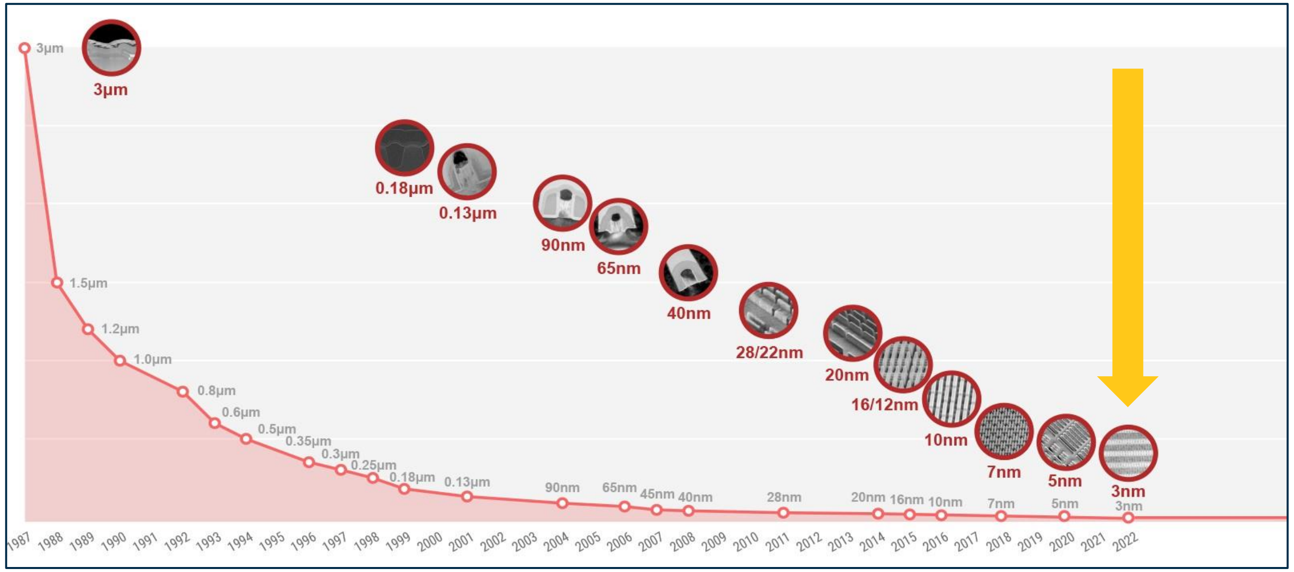
GPT-3: one of the largest neural network ever created has 1.5 trillion neurons (1 trillion = 10^{12})

Complexity of modern AI systems (II)

***Trillion**-parameters DNNs running on **3 nm** silicon chips*

1998

DL Models	Number of parameters
LeNet-5	~60 * 10^3
AlexNet	~62.3 * 10^6
VGG-16	~138 * 10^6
ResNet-18	~11 * 10^6
ResNet-152	~60 * 10^6
OpenAI's GPT-3	~175 * 10^9
Switch Transformer	~1.6 * 10^{12}
GPT-4	~100*10^{12}



Source image:
https://www.tsmc.com/english/dedicatedFoundry/technology/logic/l_3nm

The complexity makes it increasingly difficult for **human minds** to form a comprehensive picture of all relevant elements and behaviors of the system and its environment.

Smarter techniques for reliability assessment need to be found

Fault Injections

Complexity of today's FI-based reliability assessments

How many faults do we have to inject to get a comprehensive reliability assessment?

Exhaustive Fault Injections

+ Comprehensive results

− Prohibitive as the complexity of newer AI models grows

Random Fault Injections

+ Reduced FI complexity

− Inaccurate results

Statistical Fault Injections

+ Realistic simulation times

+ Statistically significant results

+ *Statistical hypothesis and constraints must be met.*

Deep Learning Models	Number of parameters
LeNet-5	$\sim 60 * \mathbf{10^3}$
AlexNet	$\sim 62.3 * 10^6$
VGG-16	$\sim 138 * 10^6$
ResNet-20	$\sim 11 * 10^6$
ResNet-152	$\sim 60 * 10^6$
OpenAI's GPT-3	$\sim 175 * \mathbf{10^9}$
Switch Transformer architecture by Google Research	$\sim 1.6 * \mathbf{10^{12}}$

1998

2022

If 1 injection takes 1 ms, an exhaustive FI would take **2,029 years**

Outline

1. Introduction
2. Need for Reliability in AI systems
3. Major Challenges and Solutions
   - Complexity of the reliability assessment
     - **Statistical Fault Injections**
   - Need for effective in-field fault detection solutions
     - Image Test Libraries
4. Conclusions and Future Directions

Statistical Fault Injections

Total number of possible **faults** in a system.

Population (**N**)

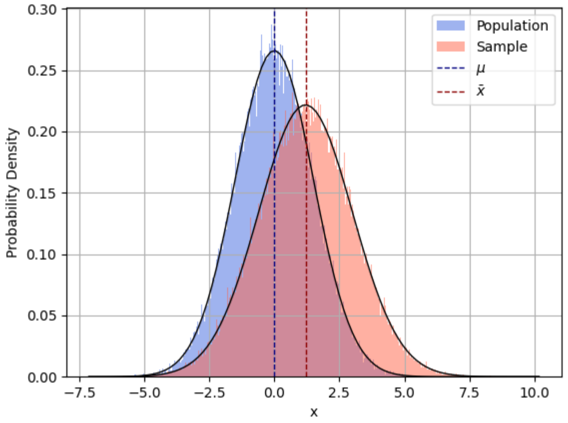
Sample (**n**)

Subset of random faults representative of the entire population.

We are interested in computing **n,** given a **margin of error**

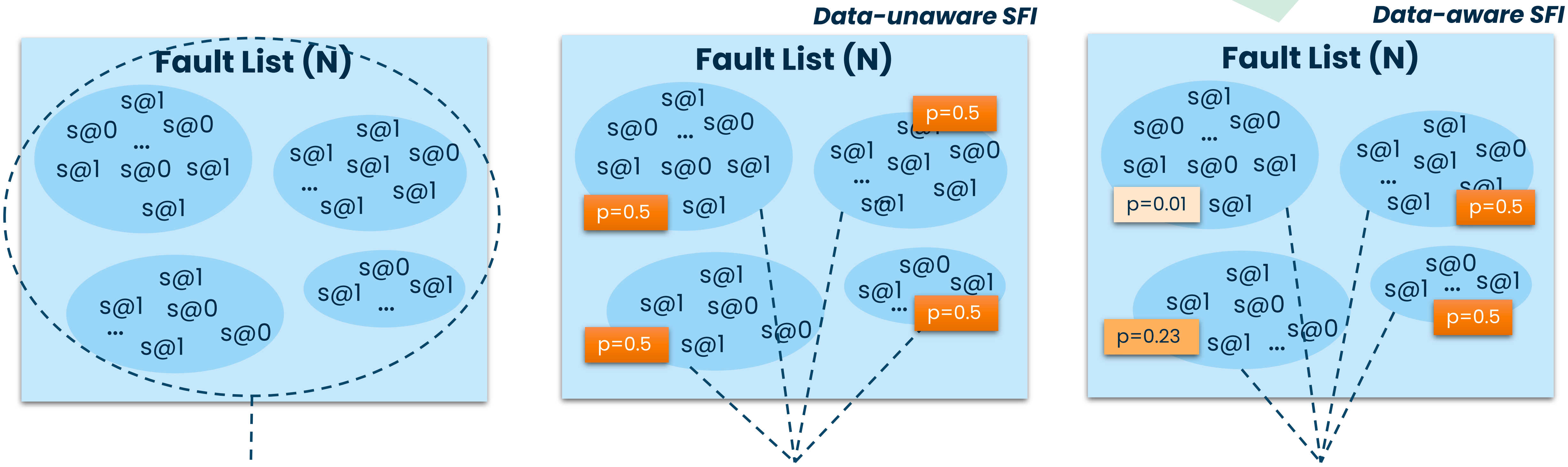
Error Margin

- Observing a statistical sample will always mean introducing an error



SFIs: proposed approaches

Domain-specific feature
to adjust the **p** probability



- Overall percentage of critical faults
- No vulnerability analysis on structural internal units ❌

- Vulnerability analysis on structural **internal units** ✓
- Data-independent SFI approach

- Vulnerability analysis on structural internal units ✓
- Data-dependent SFI approach
- Reduced number of FIs

A. Ruospo et al., "Assessing Convolutional Neural Networks Reliability through Statistical Fault Injections," 2023 Design, Automation & Test in Europe Conference & Exhibition (DATE), Antwerp, Belgium, 2023, pp. 1-6, doi: 10.23919/DATE56975.2023.10136998.

Main achievement

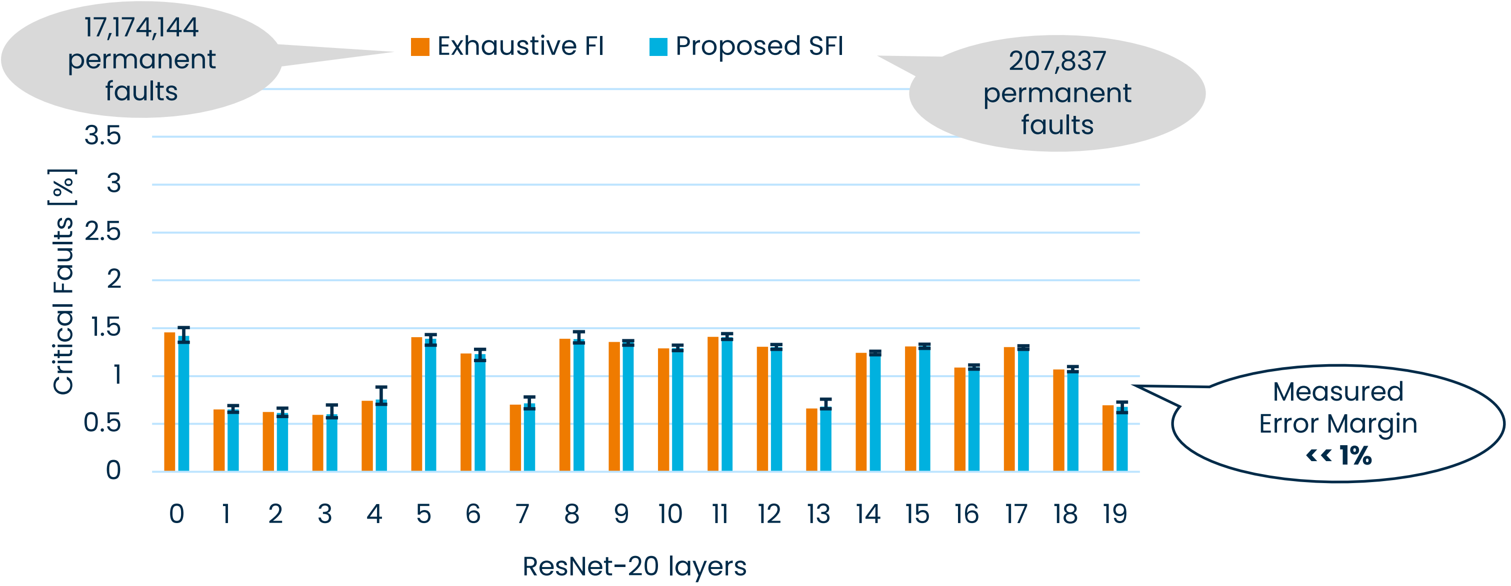
A. Ruospo, G. Gavarini, C. De Sio, J. Guerrero, L. Sterpone, M. Sonza Reorda, E. Sanchez, R. Mariani, J. Aribido, J. Athavale "*Assessing Convolutional Neural Networks Reliability through Statistical Fault Injections*» 2023 Design Automation and Test in Europe Conference (**DATE**), 2023

- The paper got nominated as **Best Paper Award Candidate**

Methodology to correctly perform Statistical Fault Injections (SFI)

- by reducing the sample size to only **1.21%** (ResNet-20) or **0.55%** (MobileNetV2) of the full fault list, the proposed SFI techniques achieve an estimate close to the exhaustive result with an error margin always lower than the target one (1%).

Statistically valid!



Outline

1. Introduction
2. Need for Reliability in AI systems
3. Major Challenges and Solutions
    - Complexity of the reliability assessment
        - Statistical Fault Injections
    - **Need for effective in-field fault detection solutions**
        - Image Test Libraries
4. Conclusions and Future Directions

In-field Testing

In-field test is performed when the device is in the operational phase

- This test is mandatory for devices deployed in safety-critical applications to detect permanent faults occurring during the operational phase (e.g., due to aging)
- Necessary to meet industry safety standards (e.g., ISO 26262 in automotive area).

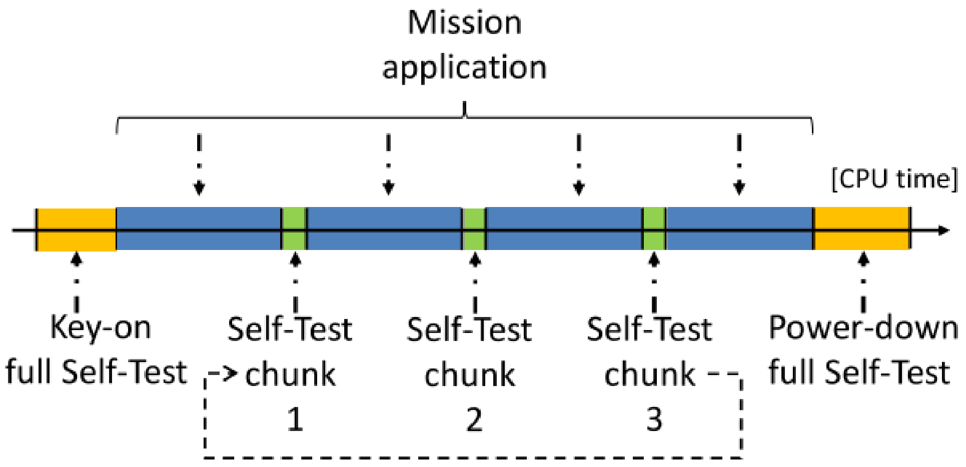


Image from: P. Bernardi, R. Cantoro, S. De Luca, E. Sánchez and A. Sansonetti, "Development Flow for On-Line Core Self-Test of Automotive Microcontrollers," in IEEE Transactions on Computers, vol. 65, no. 3, pp. 744-754, 1 March 2016, doi: 10.1109/TC.2015.2498546.

In-field Testing Techniques

1. **Hardware-based** solutions
2. **Software-based** or functional approaches.

   Software Test Library (STL): a set of **Software-Based Self-Test (SBST)** routines

   - They are based on functional stimuli, often corresponding to suitably developed *test programs*
   - The test program is stored in a memory accessible by the processor/GPU
   - The processor/GPU executes the test program
   - Results are gathered and compared with the expected ones.

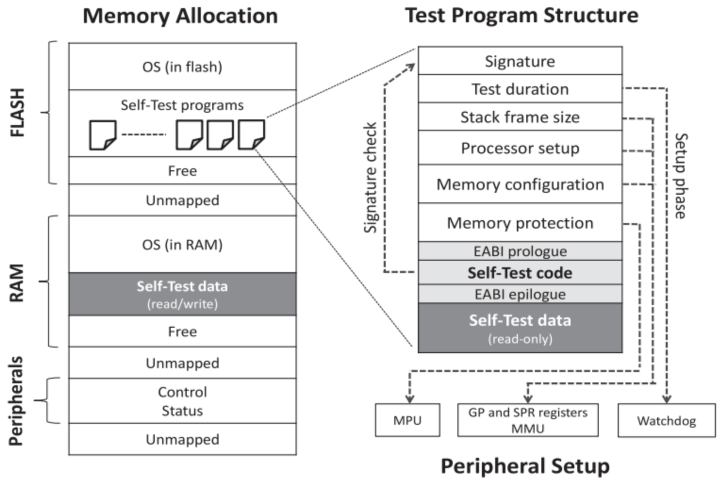


Image from: P. Bernardi, R. Cantoro, S. De Luca, E. Sánchez and A. Sansonetti, "Development Flow for On-Line Core Self-Test of Automotive Microcontrollers," in IEEE Transactions on Computers, vol. 65, no. 3, pp. 744-754, 1 March 2016, doi: 10.1109/TC.2015.2498546.

Software Test Libraries (STLs)

Companies which provide STLs include:

- ARM, STMicroelectronics, Infineon, Renesas, Synopsys.

Advantages

- ✓ No additional HW requirements
- ✓ No HW or performance overhead
- ✓ At-speed testing
- ✓ Suitable for on-line testing
- ✓ Flexible w.r.t. new requirements.

Disadvantages

- × Suitable and compact test programs are required
- × No EDA tools for their generation
- × Manual generation can be very expensive
- × Grading test programs requires huge efforts.

Our experience with STLs

- Almost **10 years** of collaboration with STMicroelectronics
  - More than 9 automotive System-on-Chip devices (ranging from ASIL-B to ASIL-D)
  - > 700,000 faults (stuck-at faults for each CPU)
  - > 80% Fault Coverage (ASIL decomposition).
- More than 30 Conference and 10 Journals papers
- New methods to develop STLs were demonstrated for a wide range of **open-source devices**, ranging from ARM0 to RISC-V and NVIDIA-like GPU model (FlexGripPlus)
- We developed automated and semi-automated techniques to identify **Safe faults**.

Image Test Libraries (ITLs)

New technique to generate test images for the in-field self-test of computational units of hardware devices running CNNs

- to develop high-quality test stimuli in the form of test images to be used during the normal CNN inference process.

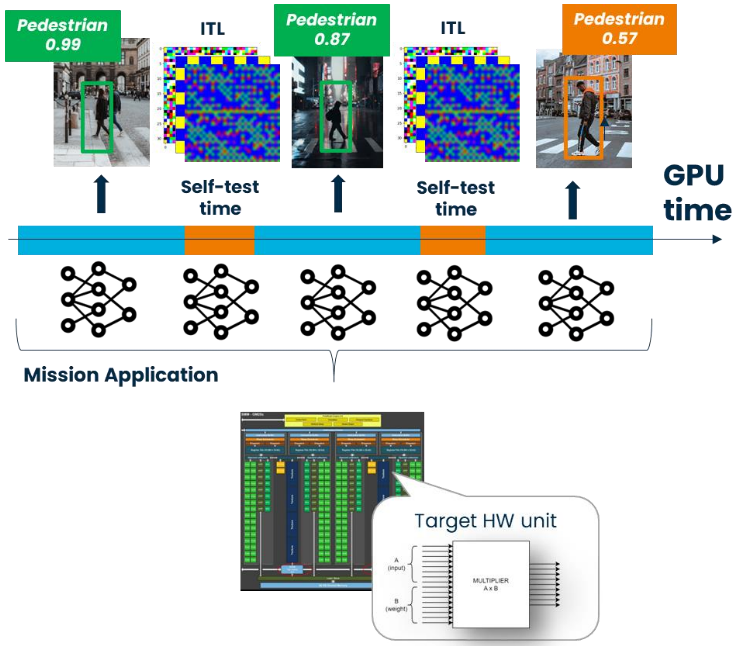


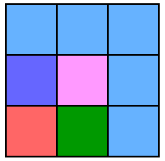
Image Test Libraries (ITLs)

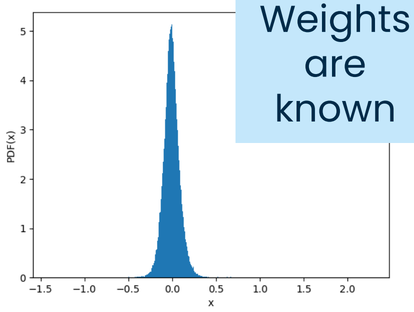
Overall idea

A. Ruospo, G. Gavarini, A. Porsia, M. Sonza Reorda, E. Sanchez, R. Mariani, J. Aribido, J. Athavale "Image Test Libraries for the on-line self-test of functional units in GPUs running CNN" 2023 IEEE European Test Symposium (**ETS**), **2023**

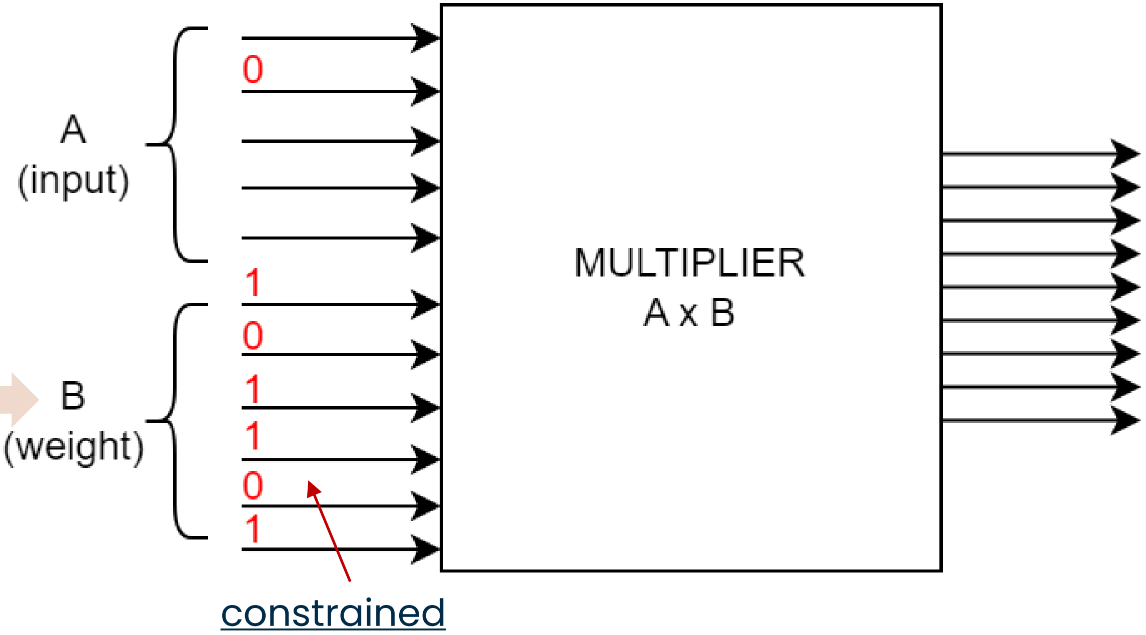
- The paper got nominated as **Best Paper Award Candidate**

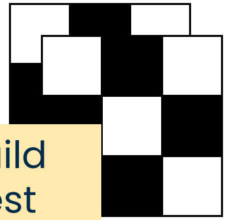
ATPG-based images





Weights are known





Build test images

ATPG patterns
Set of input patterns that **maximise** the Fault Coverage (FC) of the multiplier

ITLs for GPUs

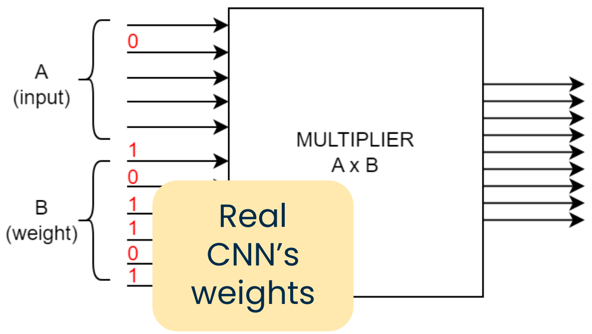
A. Ruospo, G. Gavarini, A. Porsia, M. Sonza Reorda, E. Sanchez, R. Mariani, J. Aribido, J. Athavale "Image Test Libraries for the on-line self-test of functional units in GPUs running CNN" 2023 IEEE European Test Symposium (**ETS**), **2023**

- The paper got nominated as **Best Paper Award Candidate**

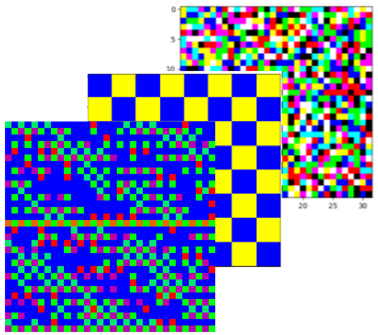
Case Study: CUTLASS GEMM Convolutional Algorithm, NVIDIA Maxwell GPU, ResNet-20, open-source FP32 multiplier from OpenCores.

Results: A total of **6 test images** covering about the **95%** of stuck-at faults.

Proposed ITL	Num. of Images	Avg. TC [%]	Self-test time [ms]	Memory space for storing the ITL [KB]
ResNet-20	6	94.75	0.35	467



ITL type	#img	Avg. Test Coverage [%] on FP32 mul			
		Warp0	Warp1	Warp2	Warp3
Proposed ITL	6	94.72	94.76	94.76	94.72
Checkerboard ITL	6	81.36	81.12	81.12	81.46
Random ITL	6	85.23	84.73	85.01	85.13
CIFAR10 ITL	6	84.42	84.85	84.56	84.61



Outline

1. Introduction
2. Need for Reliability in AI systems
3. Major Challenges
   - Complexity of the reliability assessment
     - Statistical Fault Injections
   - Need for effective in-field fault detection solutions
     - Image Test Libraries
4. **Conclusions and Future Directions**

Conclusions and future directions

As the complexity of AI models increases, the problem of reducing the costs of reliability assessment procedures assumes great significance.

Proposed Statistical FIs:

- To obtain statistically significant results
- To further reduce the costs of the assessment while achieving accurate results
- The method does not depend on the adopted fault models
- The method will be extended to hardware resilience assessments

Different activities are ongoing to support the in-field test of devices used in safety-critical applications

- ITL development for GPUs, AI accelerators, low-cost ASIC devices.

Thank you

Politecnico
di Torino

annachiara.ruospo@polito.it