Resilient AI System for Edge Applications

Marko Andjelkovic[1)], Fabian Vargas[1)], Artur Jutman[2)], Maksim Jenihhin[2)], Milos Krstic[1)]

[1)] IHP - Leibniz-Institut für innovative Mikroelektronik, Frankfurt (Oder), Germany

[2)] TalTech - Tallinn University of Technology, Tallinn, Estonia

MOTIVATION

- Transient and permanent faults may degrade the performance of AI systems.
- Future AI chips employed in safety critical applications need to be resilient, i.e., supporting self-awareness and self-healing, with a capability to anticipate possible fault scenarios.

GOALS

- Design an architecture of a resilient AI system with a processor and multiple AI accelerators.
- Explore techniques for efficient and low-cost on-chip sensing, and on-chip assessment of AI system health based on data collected from sensors.

DESIGN CONCEPT OF A RESILIENT AI SYSTEM

Multi-core AI accelerator

- Enables to optimize the performance, power consumption and reliability at runtime.

RISC-V processor

- Controls the AI accelerator and executes the application program.

On-chip sensors

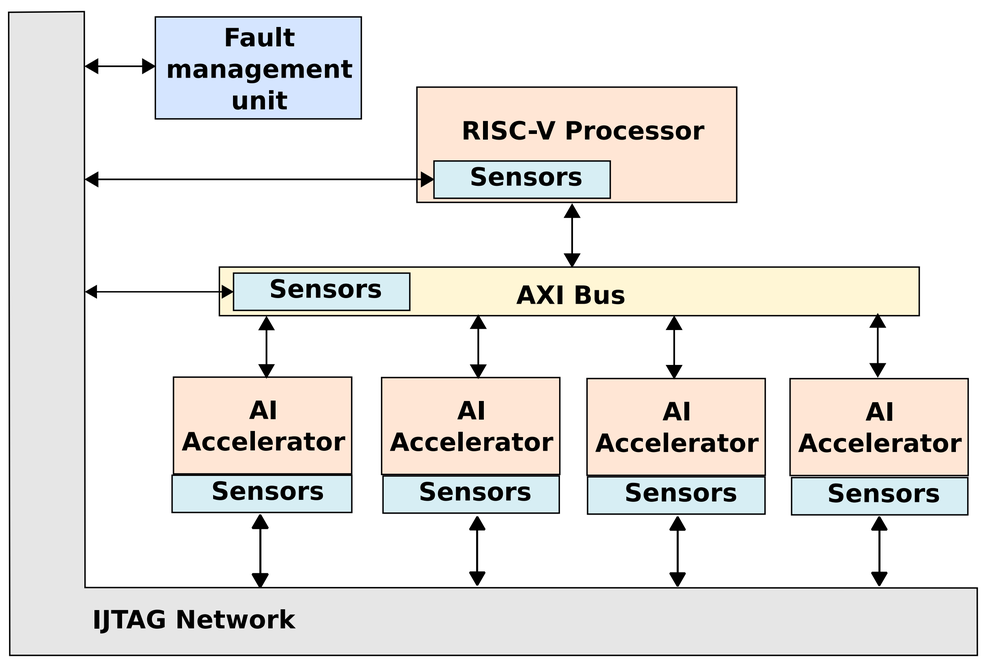
- Sensors for detection of faults and monitoring of status parameters (e.g., supply voltage, temperature).

IJTAG network

- Collects data from sensors, which is then used for further processing to assess the system state/health.

Fault management unit

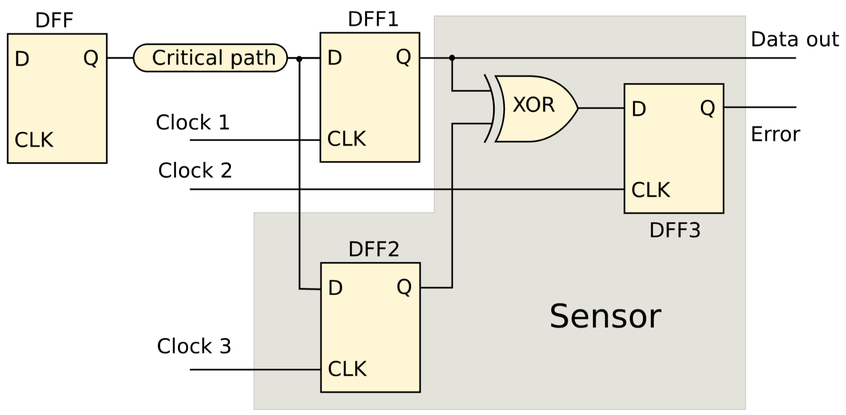
- Evaluates system health state and performs reconfiguration.



TYPICAL ON-CHIP SENSORS FOR A RESILIENT AI SYSTEM

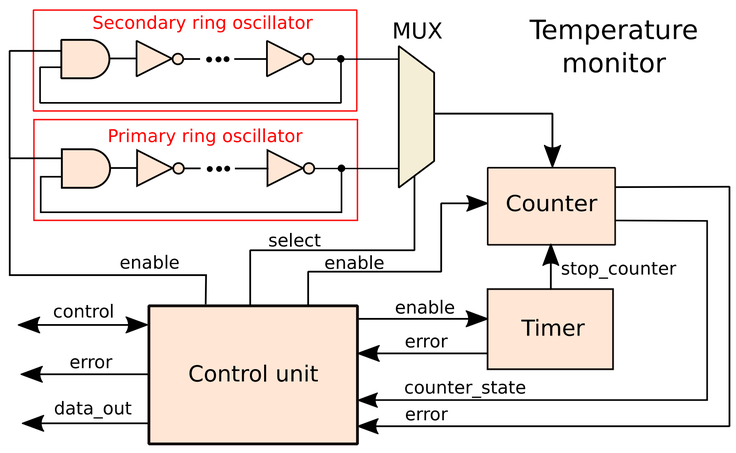
Multipurpose sensor for monitoring aging and soft errors

- Monitors aging evolution across critical paths and detects soft errors in respective flip-flops.



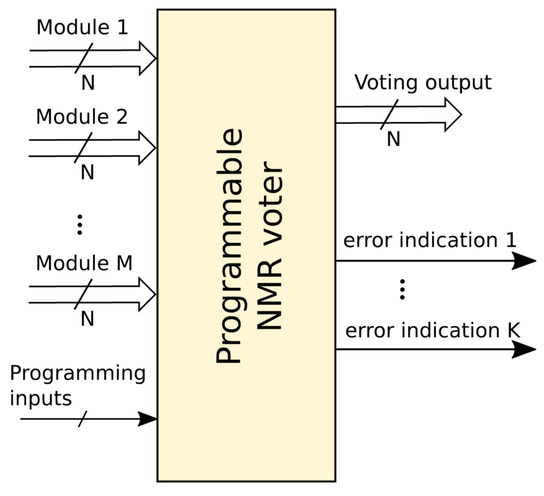
Digital temperature and supply voltage monitor

- Multiple ring oscillators with different lenghts are used to detect temperature and supply voltage changes.



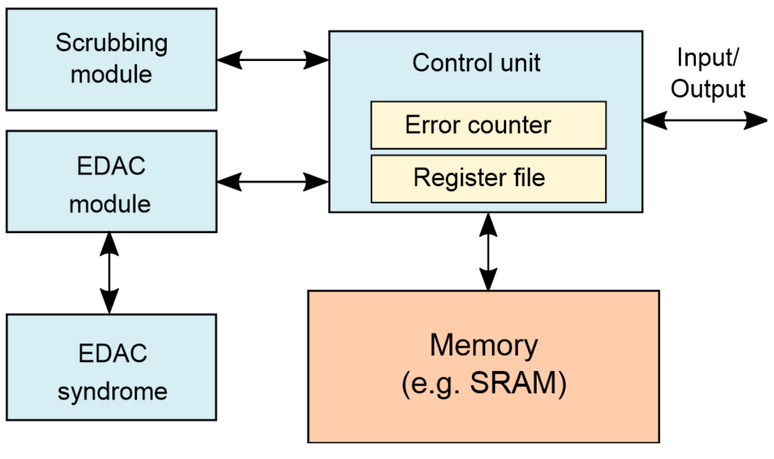
Voter for detection of transient and permanent faults

- Compares the outputs of two or more identical functional units and stores the number of detected faults.
- Generates fault indication whenever a mismatch between input signals is detected.



Error detection and correction (EDAC) for memory

- Standard EDAC and scrubbing techniques are used to detect and correct bit flips in memory cells.
- Dedicated error counter is added to collect the information on detected bit flips.



Funding: This work was done in the framework of EU-funded Horizon Europe Twinning project TAICHIP (Grant No. 101160182).