Artificial Intelligence Hardware Accelerators for Space Applications

Junchao Chen, Milos Krstic

26nd Oct, 2022

innovations for high performance microelectronics

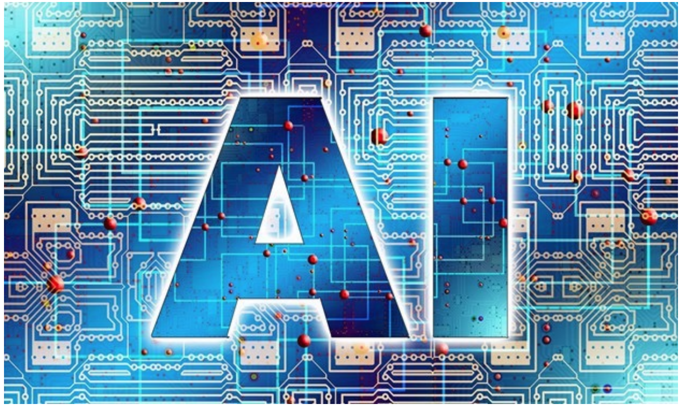
Leibniz Association

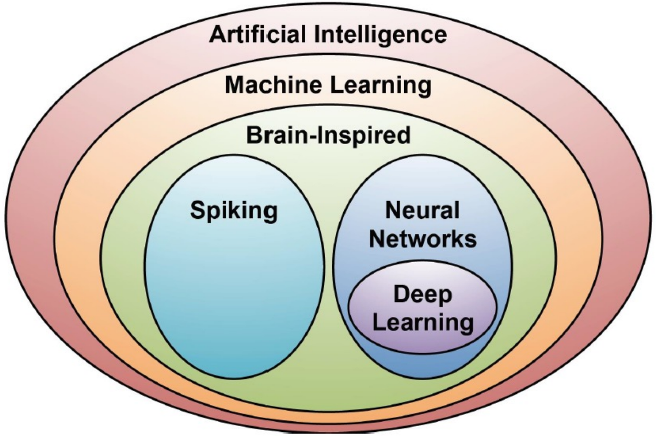
Outline

1. Background (1/3)

❖ **Artificial intelligence (AI) leverages computers and machines to mimic the problem-solving and decision-making capabilities of the human mind**

- ➢ Machine Learning: focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy
- ➢ Neural Network (NN): an interconnected group of nodes, inspired by a simplification of neurons in a brain
- ➢ Deep Learning: uses multiple NN layers to progressively extract higher-level features from the raw input
- ➢ Spiking: artificial networks that more closely mimic natural neural networks



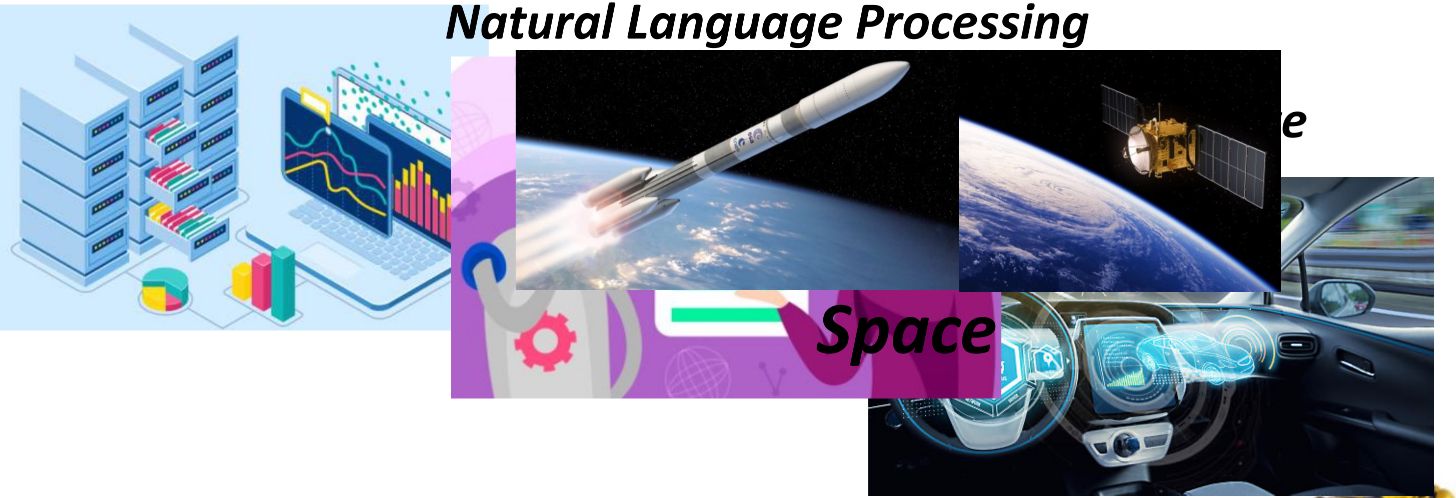


1. Background (2/3)

Artificial intelligence is applied to many fields and contributes to many important applications and research areas, such as …

Data Processing

Natural Language Processing



Space

Robot

1. Motivation (3/3)

❖ **Hostile environment**

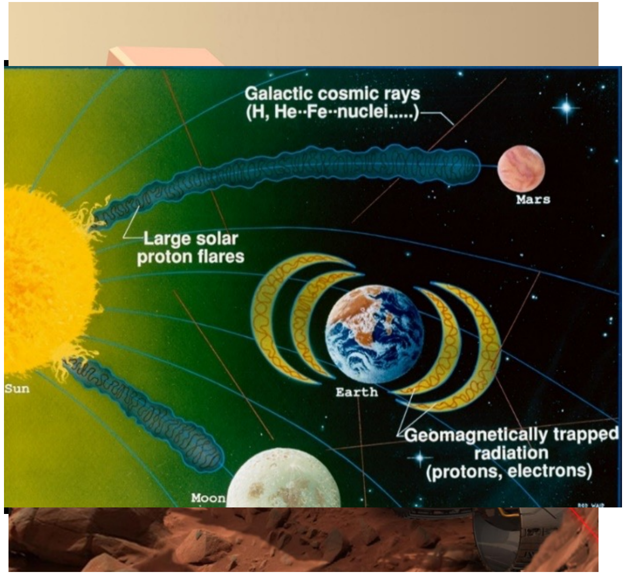
- ➢ Hostile to human life and electronic devices
- ➢ Dynamic space environment (solar events)

❖ **Spacecrafts and missions are incredibly complex**

- ➢ Rely on countless small actions
- ➢ Increase the develop & diagnose efficient

❖ **Communication Issues**

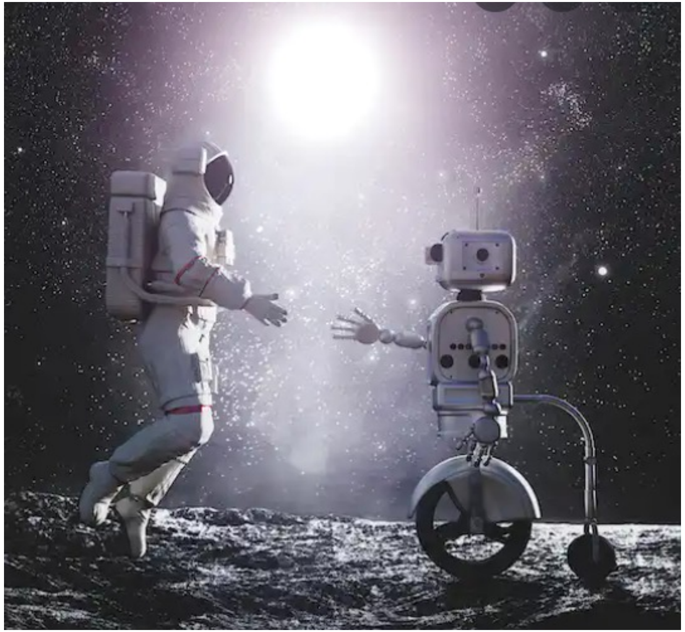
- ➢ Too vast to fast communicate
- ➢ More and more on-board data
- ➢ Blind spots (ground station location, antenna size,…)



AI is a natural fit for space applications

2. State-of-the-art AI use cases in space applications (1/4)

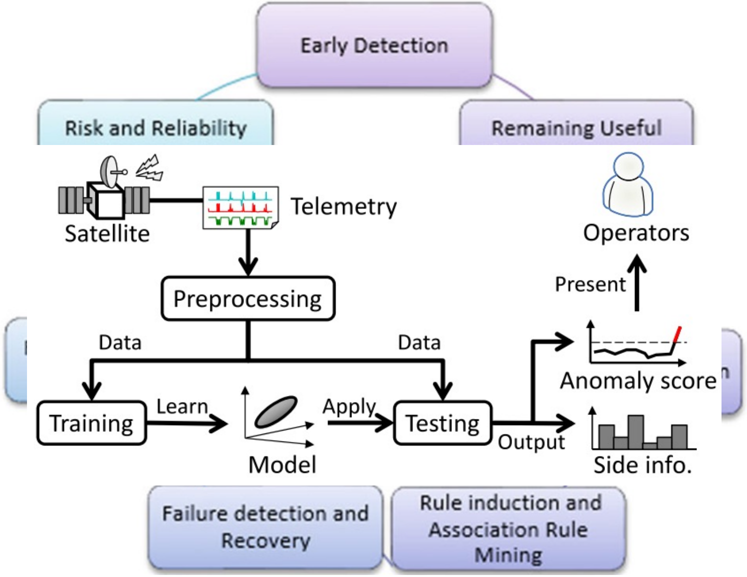
- **Mission Design**
  - Internet of Space Things
  - Design engineering assistant
- **Mission Planning**
  - Satellite health monitoring
  - Astronaut assistant
- **Space Exploration**
  - Deep Exploration, space traffic management
  - Multi-agent system
  - Study of astronomical bodies
- **Earth Observation**
  - Emergency and disaster recovery
  - Climate change, Biodiversity protection
  - Land water, and atmosphere monitoring



Pictures source: https://theconversation.com

2. Example of AI for Satellite Health Monitoring (2/4)

- **Target:**
  - Determines the satellite health state
  - Failure prediction of satellites based on the sensor data
- **Status of satellite can change relative to the various unexpected conditions**
- **Real-time system monitoring, archived data analysis**
- **Expert system is prior but limited due to variety of external environment, performance degradation, etc.**
- **Data mining is suitable to examine telemetry data and extract information to produce advanced system health monitoring**
- **Intelligent approach: detection, diagnosis and prediction based on the historical data**
- **Satellite subsystem generally has several modes with different structures and parameters**



Source: **S. Abdelghafar, et al. "Intelligent Health Monitoring Systems for Space Missions Based on Data Mining Techniques", *In Machine Learning and Data Mining in Aerospace Technology; Springer.***

2. Example of AI for Internet of Space Things (3/4)

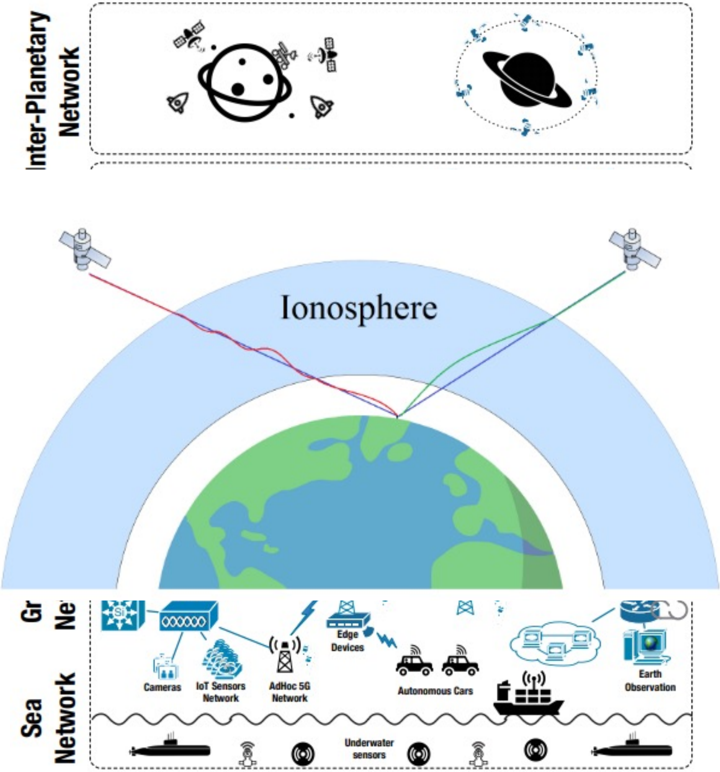
❖ **Target:**

- ➢ Engineering the further internet, reliable machine-to-machine communication, machine-type low-latency communication
- ➢ Edge, fog, cloud computing for satellites
- ➢ Satellite based 5G/6G netwowrk, large-scale data processing
- ➢ Network wearables, apps & robots in space

❖ **Communication and computing techniques are promotion, and critical for automation and ubiquitous connectivity in space**

❖ **AI is critical for dynamically managing, optimizing and addressing various space IoT issues**

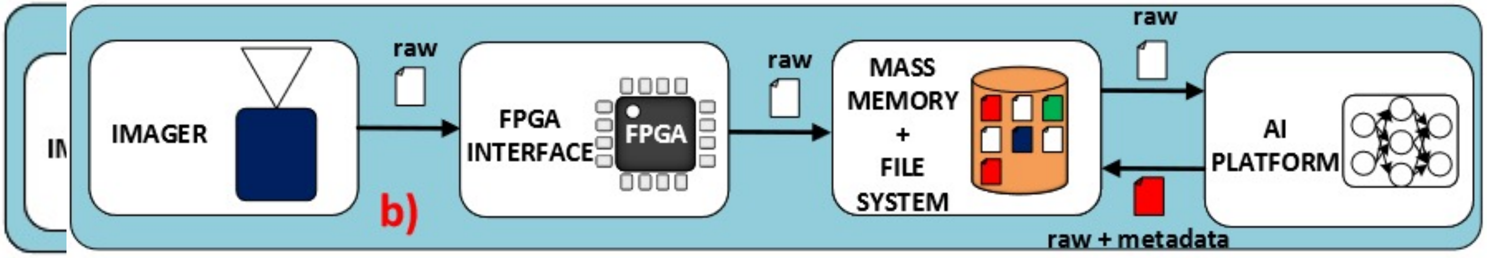
❖ **Example: signal distort from ionospheric scintillation events**



Source: **F. Fourati, et al. "Artificial Intelligence for Satellite Communication: A Review",** ***In Intelligent and Converged Networks***

2. Example of Data Process for Earth Observation (EO) (4/4)

- **Challenge:**
  - Memory budget of satellites, over 10,000 scenes per day (~10TB)
  - Downlink bandwidth, communication limitation and flexibility restrict
  - Increasing resolution with massive data, require on-board processing
- **Deep leaning is used to classify the images depending on their contents, protect & download meaningful files**
- **AI reduces the required number of electronics, saving mass, power, volume and reducing harness complexity**
- **Improve the task execution, increase system flexibility, DNN can be pre-trained through emulated satellite data**
- **Applied in several ESA satellites to improve the EO ability**

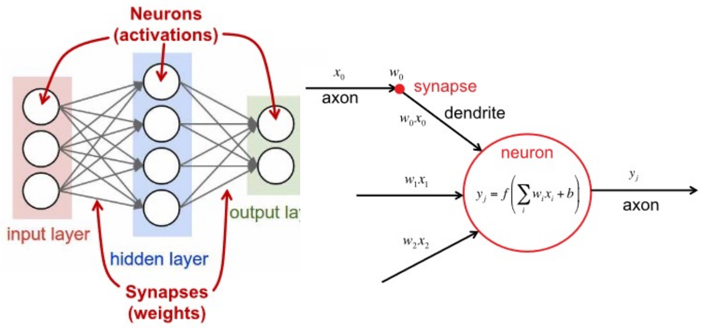


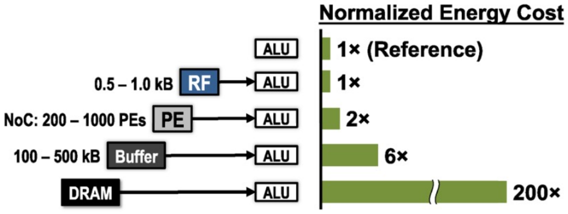
Classical EO satellite architecture
AI based EO satellite architecture

Source: **G. Furano, et al. "AI in space: applications examples and challenges"*, IEEE DFT***

3. Overview of Hardware Accelerator (1/4)

- **Motivation for hardware accelerator:**
  - Various AI applications, especially DNNs, are widely used
  - Severe challenges of data processing speed and scalability on conventional computer systems
  - Domain-specific design offer greater energy efficiency and performance gain than general-purpose processors
- **General Target:** improve throughput and energy efficiency
- **Design guideline:**
  - Use dedicated memories to minimize data movement
  - Invest resources into more arithmetic units or bigger memories
  - Use the easiest form of parallelism that matches the domain
  - Reduce data size and type to the simplest needed for the domain
  - Use a domain-specific programming language

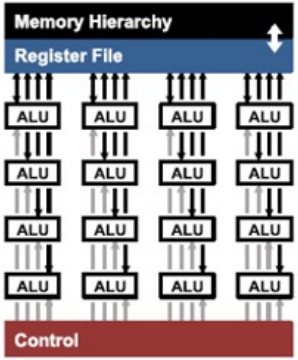




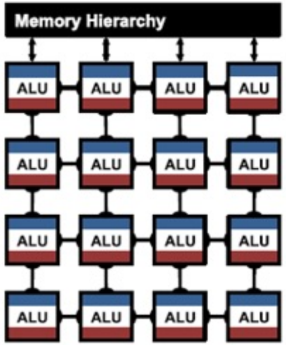
Data Movement Energy

3. Design Aspect of Hardware Accelerator (2/4)

- **Multiply-And-Accumulate (MAC) operations are the fundamental component, can be easily parallelized**
- **Matrix multiplications and convolutions dominate over 90% of the operations**
- **NN layers (CONV, FC) can be mapped to matrix multiplication**
- **High performance & parallel compute: temporal and spatial architectures**
- **Temporal Architecture:**
  - Mostly in CPU & GPU, a centralized control for a large number of ALUs
  - Improve parallelism: SIMD, SIMT, etc.
  - Challenge: reduce the number of multiplications and increase throughput
- **Spatial Architecture:**
  - Mostly for ASIC/FPGA-based design, message passing directly between ALUs
  - ALU can have its own control logic and local memory
  - Reduce the energy cost of data movement by introducing several levels of local memory hierarchy
  - Challenge: increase data reuse from memory to reduce power consumptions



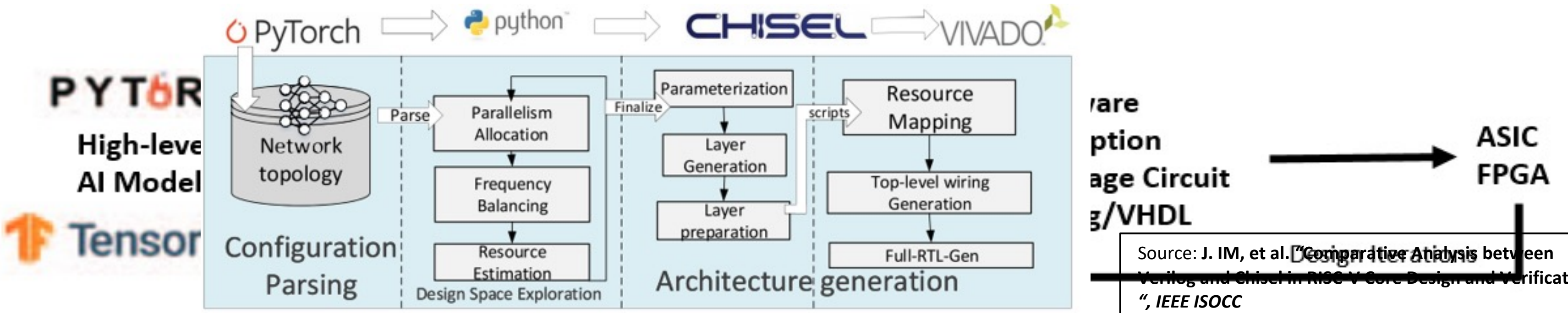
Temporal Arch.



Spatial Arch.

3. Design Aspect of Hardware Accelerator (3/4)

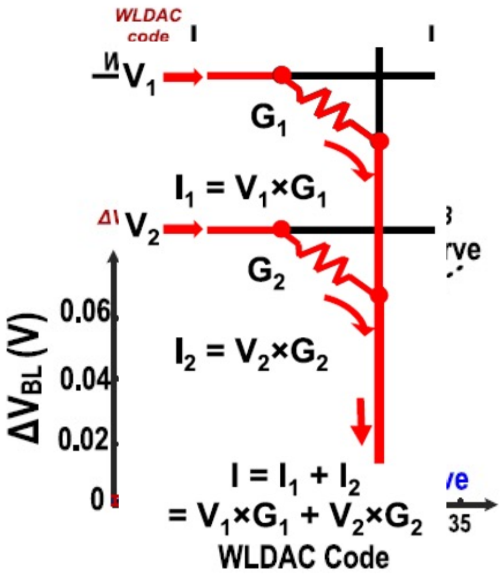
- **AI algorithms generally descript from high-level libraries**
- **Hardware analyze**: parallelism and data reuse opportunities, determine the microarchitecture and dataflow
- **Implement the algorithms:** RTL, High-level Synthesis, Domain-specific Language
- **Example: Agile implementation with Domain-special Language (such as Chisel)：**
  - Improve the design productivity with better source code expressiveness
  - Increased abstraction level in design entry, and improved automation
  - More productive than direct Verilog coding in aspects of both design and verification
  - End-to-end automatic synthesis flow
  - Based on hardware generator fast design, may no- human-in-the-loop



Source: J. IM, et al. "Comparative Analysis between Verilog and Chisel in RISC-V Core Design and Verification", IEEE ISOCC

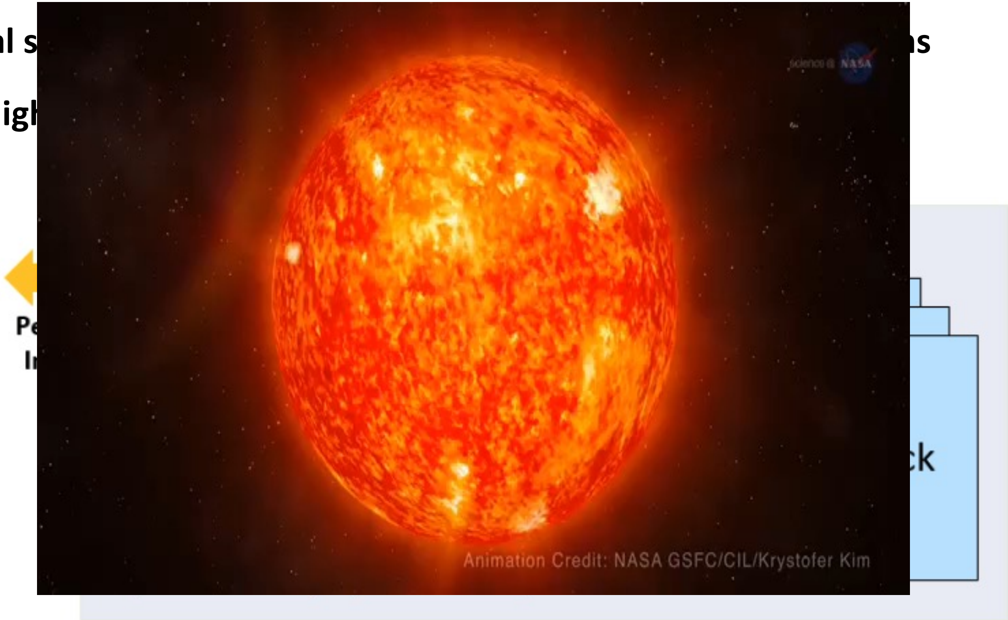
3. Design Aspect of Hardware Accelerator (4/4)

- **Data movement dominates energy consumption -> in-memory computing**
- **Generally use analog processing, compute at reduced precision**
- **Integration of MAC into SRAM Array:**
  - DAC drive the WL to an analog voltage (feature vector)
  - I_{BC} is product of the voltage and value of bit cell
  - 12X energy saving compared to separate Implementation
  - Drawback: density/capacity insufficient, leakage current, limited parallelism
- **Nonvolatile Resistive Memories (RRAM, CBRAM, STT-MRAM, etc.):**
  - Memristors can be used as weights
  - Multiplication is performed with the resistor’s conductance as the weight,
  - Voltage as the input, corresponding current as the output
  - Great immunity to noise and faults and largely tolerated by ML algorithms
  - Drawbacks: precision/ADC/DAC overhead , array size dominated by wires, write energy could be costly, Device-to-Device and cycle-to-cycle variations



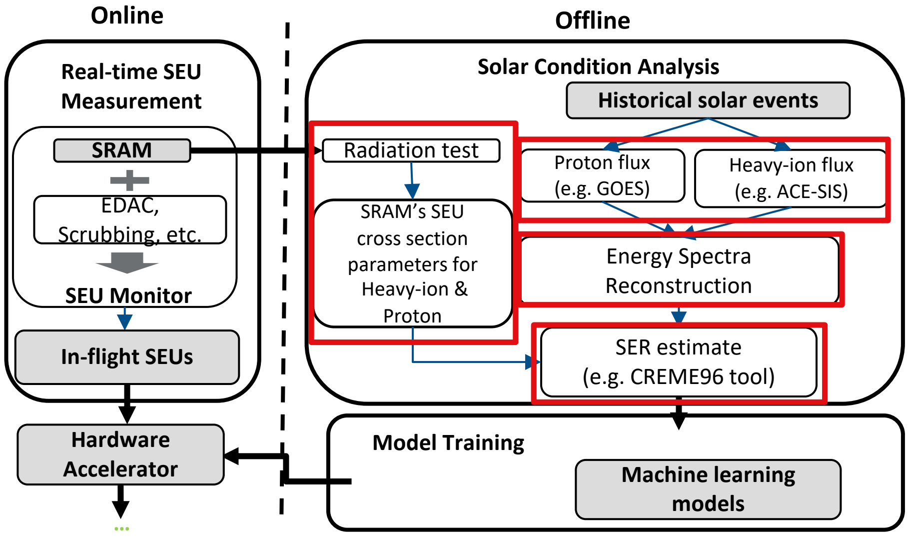
4.1 IHP: Space Environment Prediction with AI (1/4)

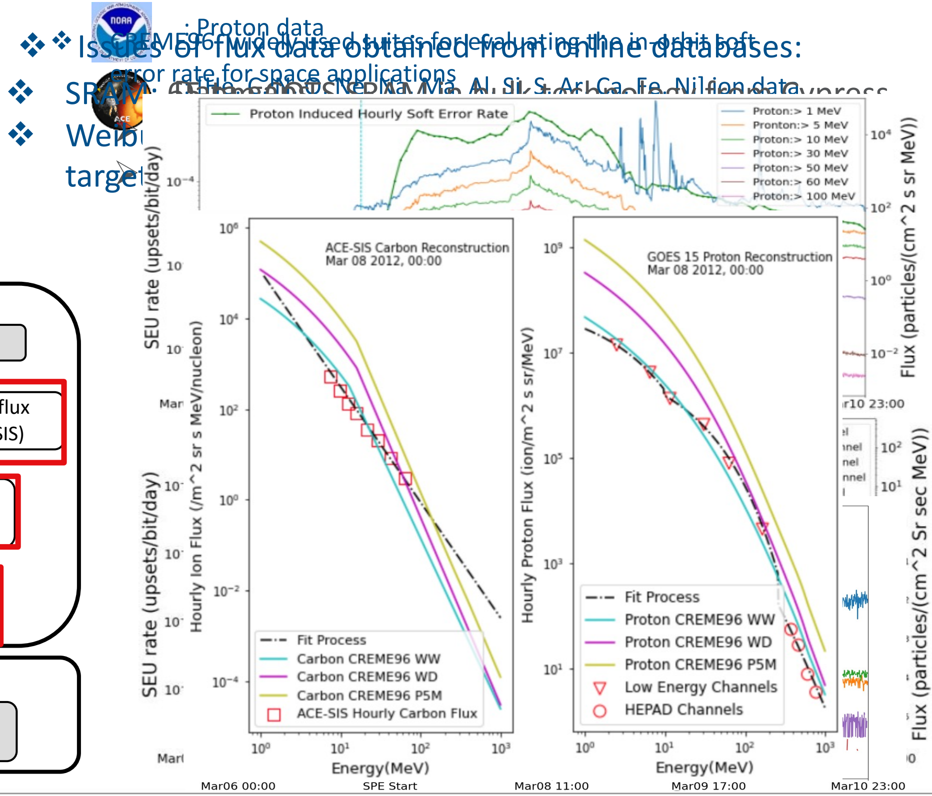
- **Solar Particle Event: can dominate the space radiation environment**
  - particle flux may increase by 2 – 6 orders of magnitude during a period of several hours or days
  - Real-time monitoring & prediction is vital in space applications
- **On-board SRAM-based radiation monitor for in-flight real-time Single Event Upset (SEU) rate**
- **Available public historical s … s**
- **Target: Prediction of in-fligh**



4.1 IHP: Space Environment Prediction with AI (2/4)

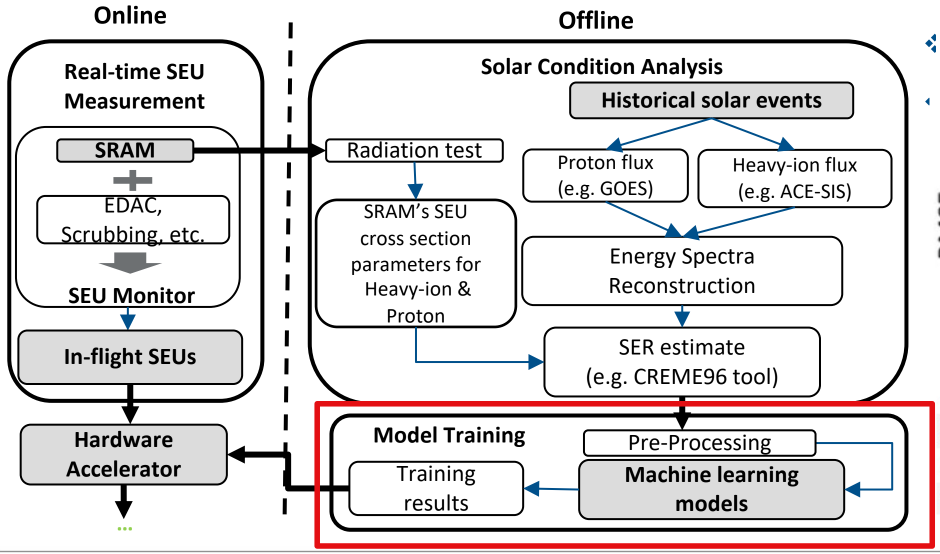
- Target: obtain the historical in-flight target SRAM SEU rates from real historical SPE flux data
- Output: Hourly Soft Error Rate (SEU rate) data of the target SRAM during the selected SPEs
- Four main steps:
  1) Collection of historical solar events flux data;
  2) SPE energy spectra reconstruction;
  3) Target SRAM cross-section parameters;
  4) Hourly SRAM SER estimation.

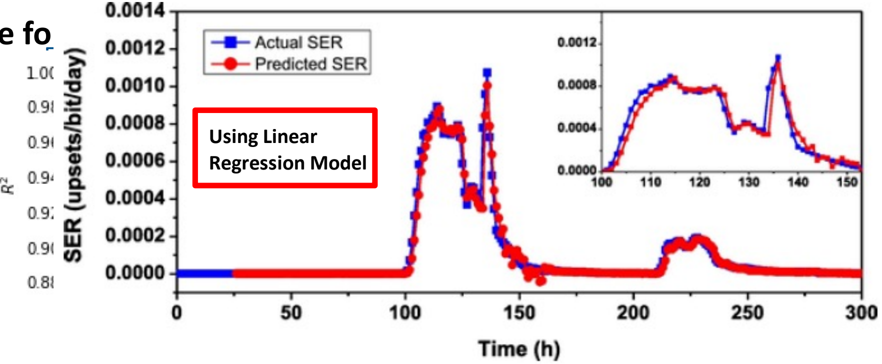


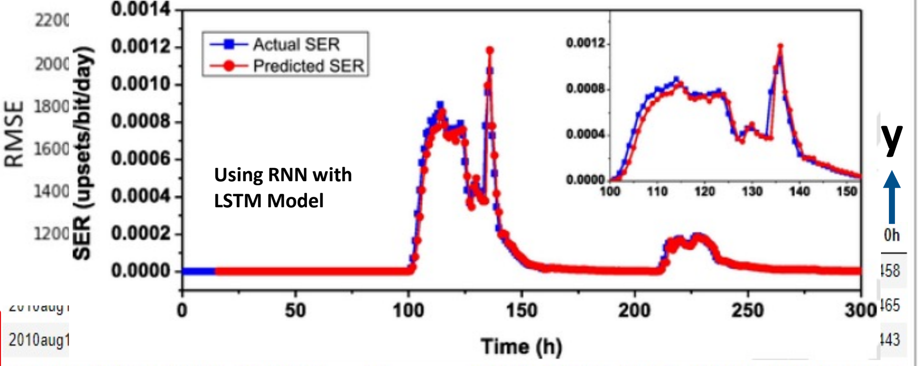


4.1 IHP: Space Environment Prediction with AI (3/4)

- Target: obtain a machine learning model which is able to predict SEUs 1 h in advance
- Using several consecutive past hourly SEU data to predict the fo
- Three main steps:
  1) Pre-processing of the Test and Training Data Set
  2) Model Training
  3) Model evailation and comparison



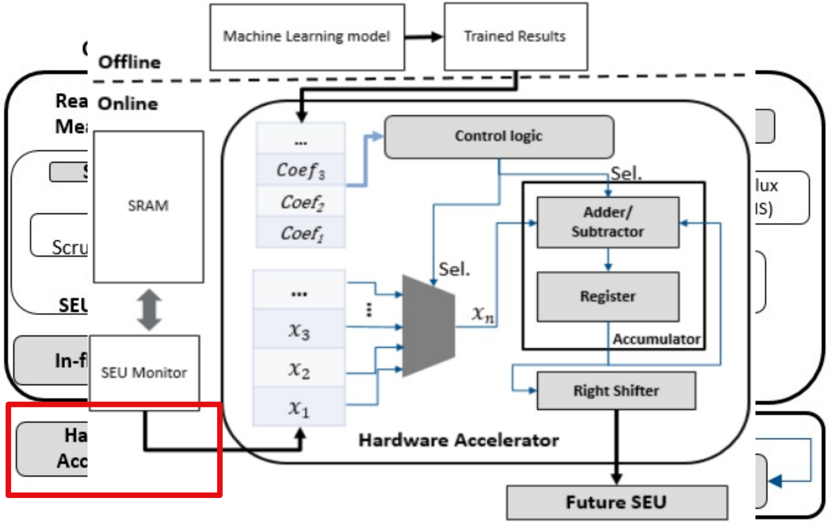




Prediction over a SPE from 2011-01-19 to 2011-02-01

4.1 IHP: Space Environment Prediction with AI (4/4)

- Target: implement the linear regression model with an ultra-low cost
- Hourly SEU data from monitor, coefs from training model
- Accumulator: multiplication operation through successive additions
- Improves area usage and power consumption by slowing down the calculation process



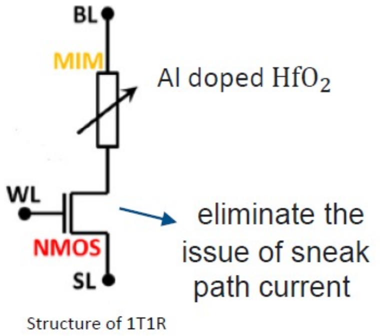
Example: $SEU_{Pred} = 41x1 + 3x2 + (-38)x3 + 33x4 + 3x5 + (-29)x6 + 26x7 + (-1)x8 + (-22)x9 + 21x10 + (-1)x11 + (-15)x12 + 13x13 + (-1)x14 + (-7)x15 + 6x16 + (-2)x17$

- A total of 262 clock cycles is needed for one calculation
- 5.24 us when working frequency is 50 MHz

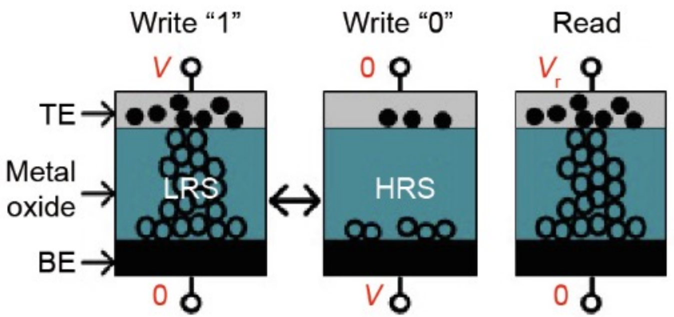
4.2 IHP: RRAM for AI Implementation (1/2)

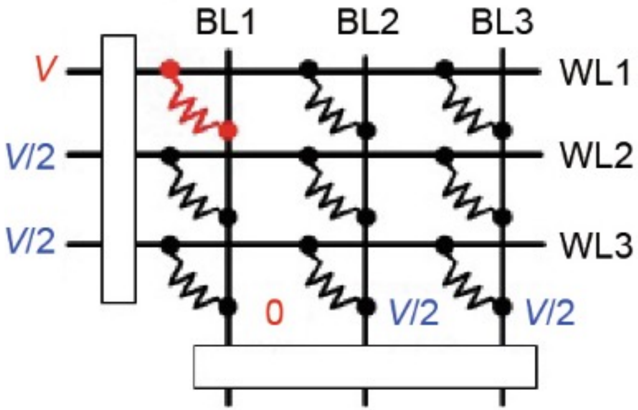
❖ **Memristor (RRAM):**

- Emerging non-volatile memory that stores information using cell resistances
- Two-terminal device, the resistance state is retained based on the history of applied voltages
- 3-4 orders of magnitude higher computation efficiency due to the low-power nature of analog computation.
- Tolerated by ML algorithms, immunity to noise and faults (based on analog signals)
- Perform in situ matrix-vector multiplications in an analog manner



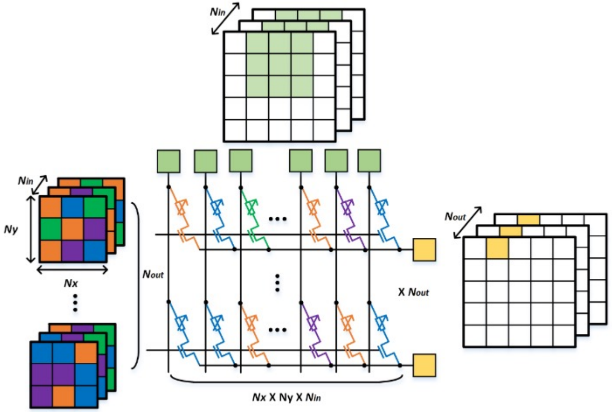
Structure of 1T1R



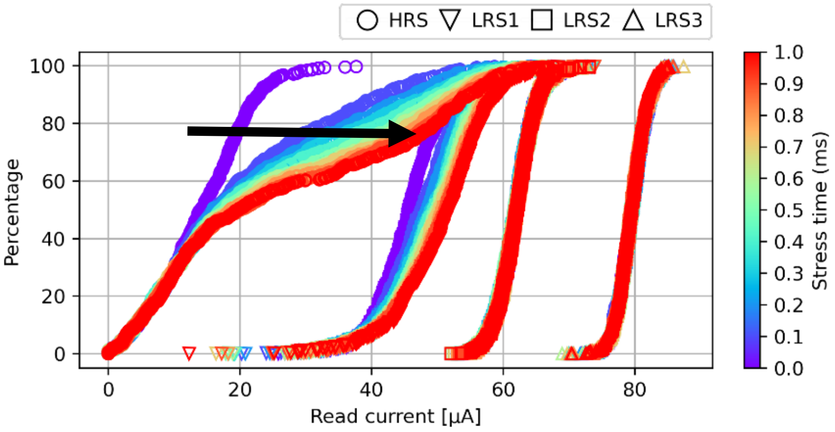


4.2 IHP: RRAM for AI Implementation (2/2)

- Realizing MAC computation inside the RRAM -> Reduce data movement, weight stationary inference
- Active one or more rows -> High parallelism
- Multi-bit computation -> Increased storage density and throughput
- States with low conduactance suffer from read disturb
- Devices trend to become more conductive with the increased number of pulses
- The distribution of devices is wider => increased device-to-device variability



CNN inference with RRAM crossbar



Summary

- **AI is widely used and a natural fit in space applications**
  - AI is addressing 13 main challenges in 4 categories for soace applications
- **Hardware accelerator is important for improving throughput and energy efficiency of various AI applications, especially DNNs**
  - Basic arch: temporal arch for CPU/GPU; spatial arch for ASIC/FPGA
  - Near data processing: SRAM, Nonvolatile Resistive Memories
- **IHP is actively conducting research in related directions**
  - Space environment prediction with on-chip radiation monitor
  - RRAM design for in-memory AI computing

Thank you for your attention!

Junchao Chen

IHP – Innovations for High Performance Microelectronics
Im Technologiepark 25
15236 Frankfurt (Oder)
Germany
Phone: +49 (0) 335 5625 527
Fax: +49 (0) 335 5625 413
Email: andjelkovic@ihp-microelectronics.com

www.ihp-microelectronics.com

innovations for high performance microelectronics

Leibniz Association

A10