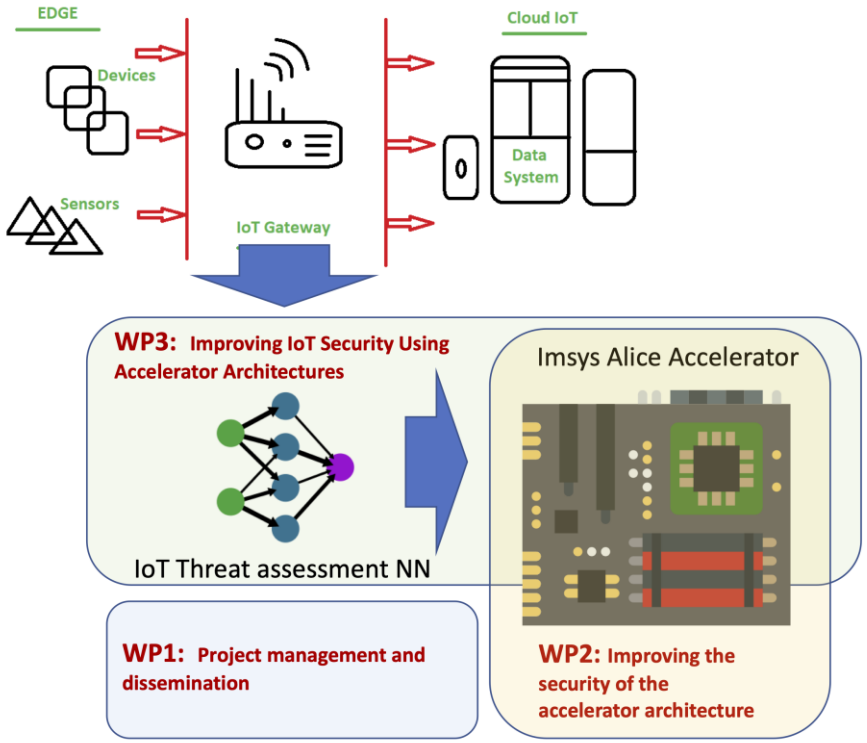
Smart and Secure Gateways for a Secure Internet of Things

The Cybernode Collaboration Conference 2023

Joakim Eriksson, RISE
2023-01-26

Smart and Secure IoT Gatway

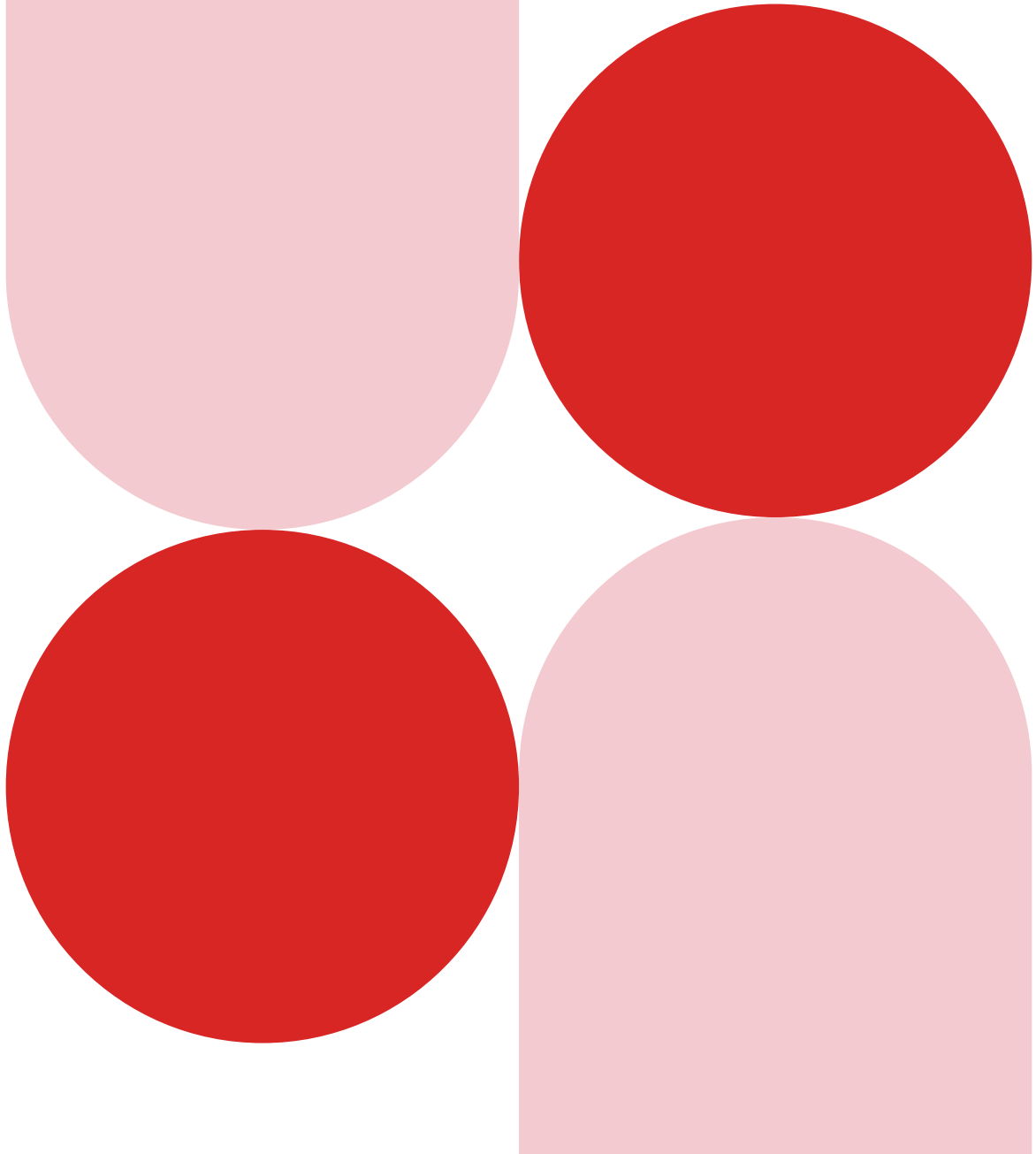
- Partners
  - RISE - project lead and use-case provider:
    IoT – wireless device fingerprinting
  - Uppsala University – research and development of security mechanisms for the AI accelerator
  - Imsys – design and implementation of secure AI accelerator
  - IoT Bridge – IoT company – use-case provider:
    Bridge Safety IoT Application Using AcceleratorArchitecture
  - Wittra – IoT company – use-case provides:
    Secure IoT for asset tracking and asset lock system
- Main contribution
  - Smart and Secure IoT Gateway based on Imsys Alice AI-accelerator
  - Use-cases evaluating Alice



Acceleration in SecureGW

Dag Helmfrid, CTO

Imsys take on AI acceleration with power efficiency



imsys

Inference close to data

- High capacity important for fast response

Contained mission criticality

Resource restricted

Fast response

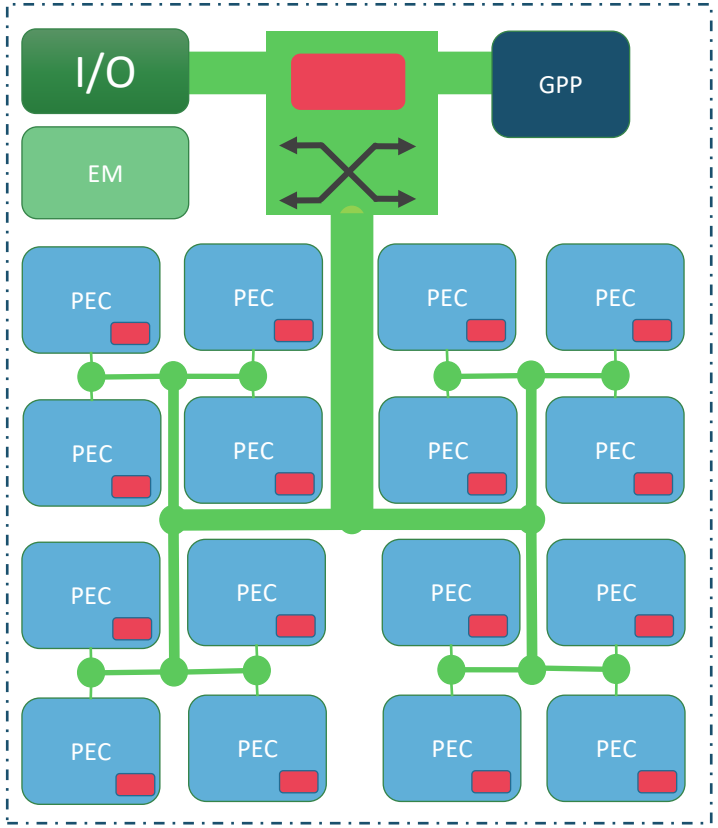
A lot of data

imsys

Alice Accelerator Architecture

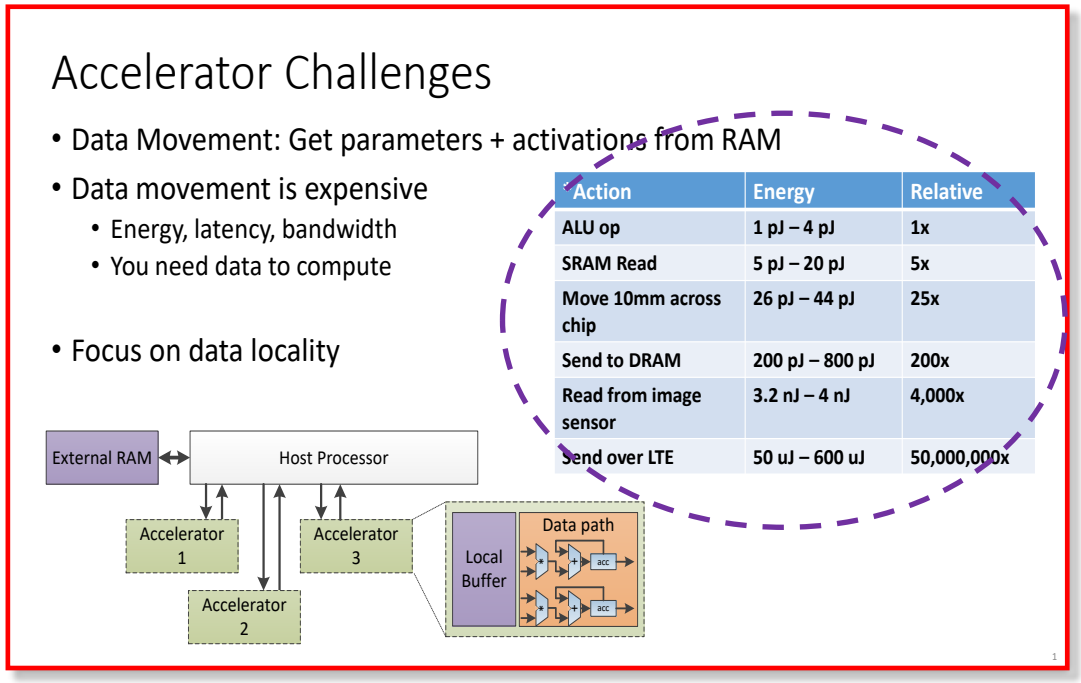
- **Accelerator controller, IM4000 (GPP)**
  - Full software stack
  - High level language support
  - Open and open source based
- **Network on Chip (NoC)**
  - High speed data and control
  - Application controlled peer2peer
- **Processing Element Cluster (PEC)**
  - Multiple Processing Elements with shared memory
  - Processing near memory in PE
- **I/O**
  - External memory, Highspeed i/f, Ethernet …
- **Energy manager (EM)**
  - Sleep modes, Performance, External power source

Flexible
Deep
Neural
Network
Instruction
Set



imsys

The Imsys Accelerator Design for Low energy



DSD 2018 AMDL Keynote, Prof. Dr. Henk Corporaal

Sources of energy consumption challenging the system solution

Don't move data around

- Automated tools for data flow analysis
- Cache less memory access
- Data "reuse"
- Processing near memory

Efficient processing

- Lean data types (int8, BFloat16)
- Low power circuit design matching archite
  and advanced technology nodes for SoC
- Only local data transport

Important project results:
Models have been quantized for energy efficient execution. Same model precision to less than a quarter of the energy
Imsys accelerator's vector engines and tools have been updated.

Automated network optimization and programmed flexibility

- Minimize memory usage & maximize utilization
- Layer fusion, zero pruning, operator fusion …

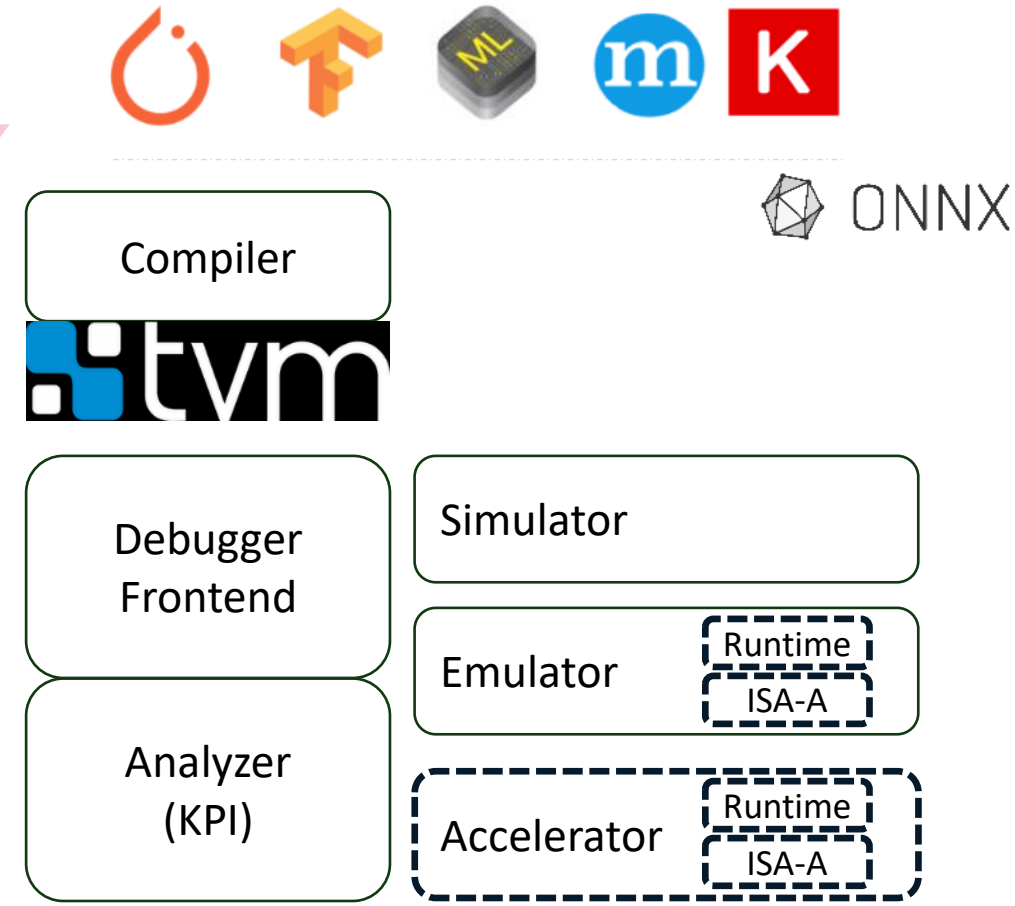
imsys

SecureGW demonstrator platform

imsys

Optimizer, Compiler & Runtime

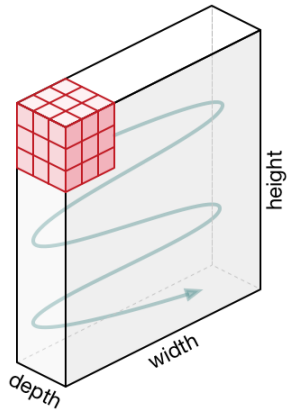
- Supports development flow from inference application graph to optimized target object code
- Quantization support
- Customizable optimization: pipelining, layer fusing …
- Seamlessly integrates with existing AI development frameworks



imsys

Kernel Library (DNN Instruction set, ISA-A) & SDK

- Kernel libraries support
  - Extensive instructions for quantized neural network operations and other kernel-based operations
- Programmable user customization
- Tools for custom kernel development



Microcode Compiler

µDebugger

µSimulator

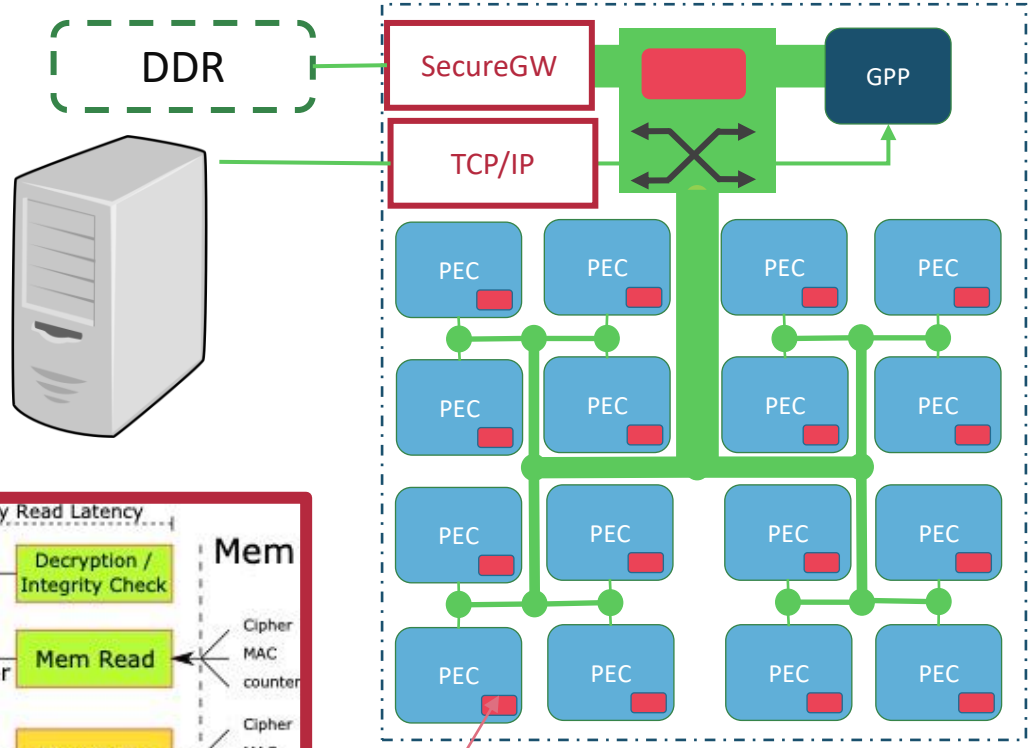
Emulator

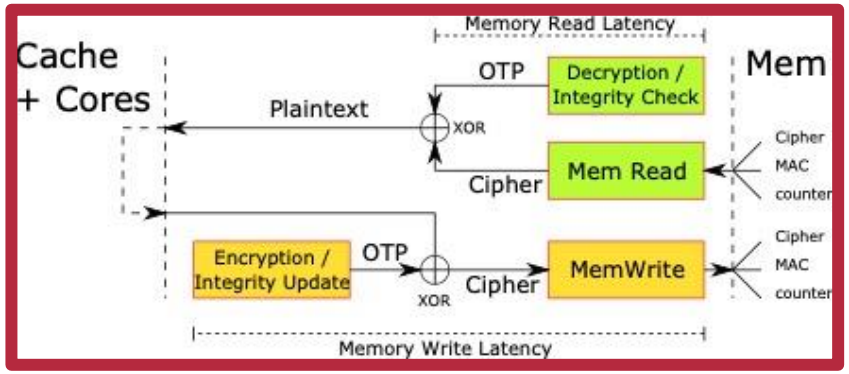
µAnalyzer

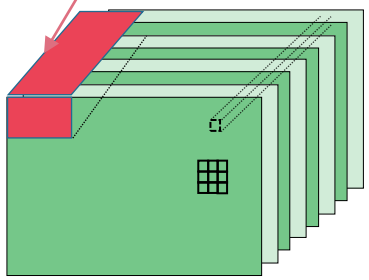
"SecureGW project has triggered new optimizations and kernel library extensions"

imsys

SecureGW application demo

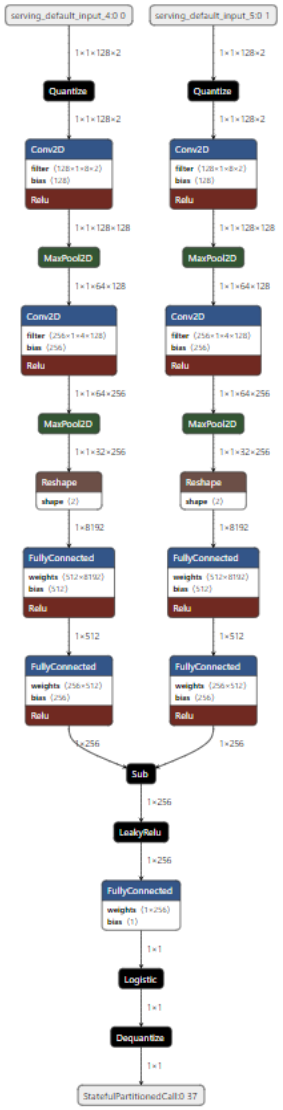






MAC:s dominate computational load

- Point-wise convolution
- Depth-wise convolution

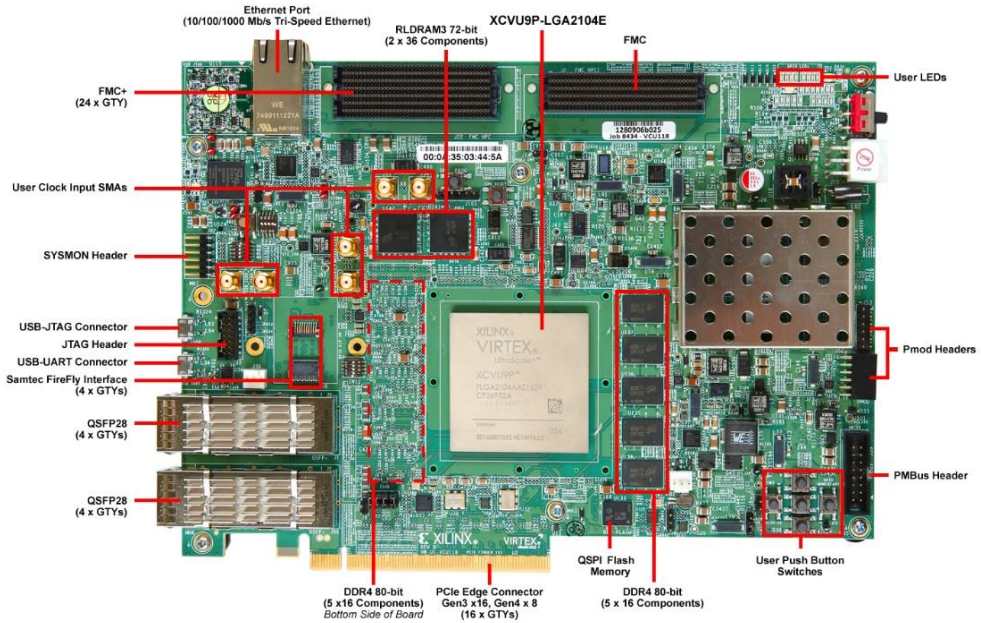


imsys

Accelerator Evaluation Platform (Release R1)

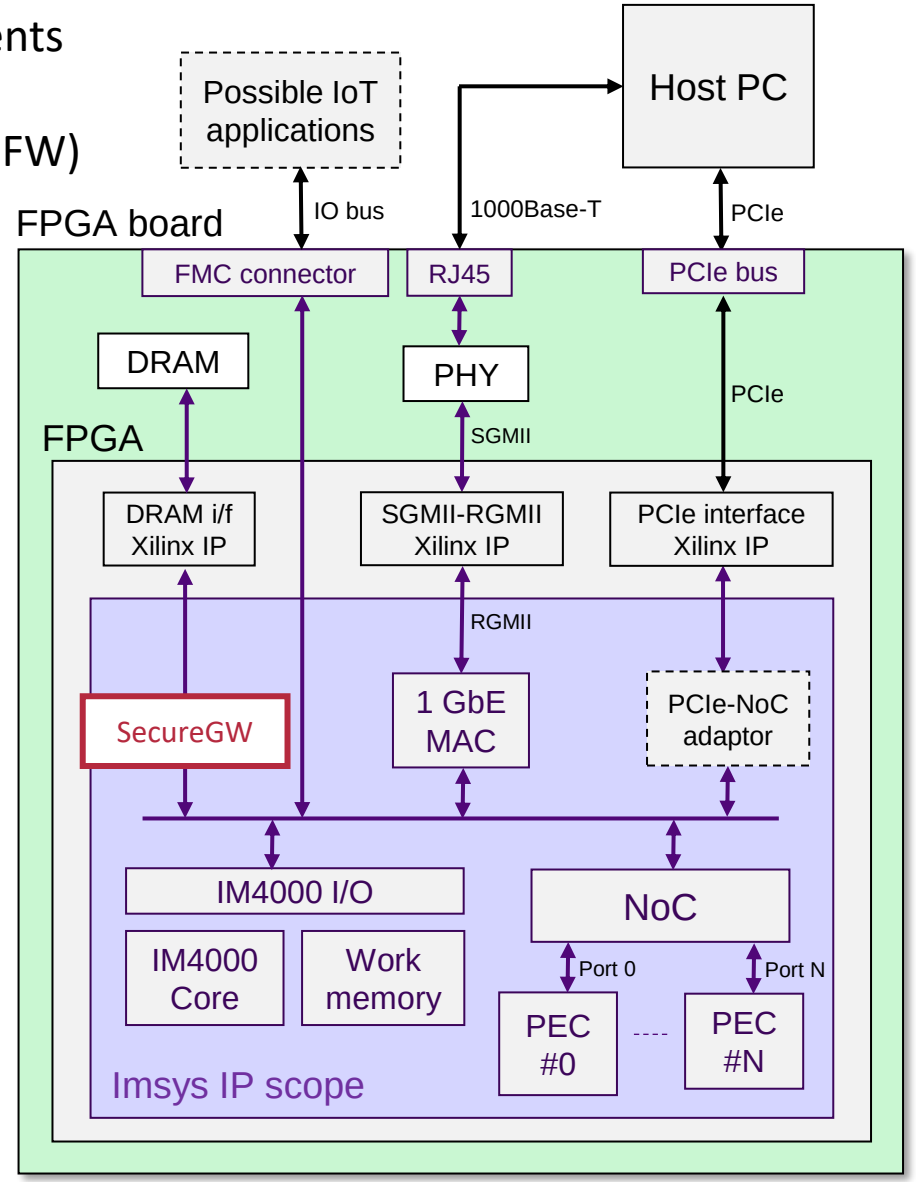
The FPGA board
- ✓ VCU118 Virtex UltraScale+ Evaluation Kit
- ✓ FPGA: UltraScale+ XCVU9P

System Logic Cells (K)	2,586
DSP Slices	6,840
Memory (Mb)	345.9
GTY 32.75 Gb/s Transceivers	120
I/O	832



Acceleration on processing elements
- ✓ 64 PE = 4 PEC (~320 GOP/s)
- ✓ Micro coded DNN operations (FW)

High speed data
- ✓ DRAM
- ✓ Control: Ethernet



imsys

Thank you.