SS-CPS&IoT'2022

3rd Summer School on Cyber-Physical Systems and Internet-of-Things

Budva, Montenegro, June 07-11, 2022

Proceedings of the 3rd Summer School on Cyber-Physical Systems and Internet-of-Things

Vol. III

Editors:

Lech Jóźwiak
Eindhoven University of Technology, The Netherlands

Radovan Stojanović
University of Montenegro, MECOnet, Montenegro

Nikolaos Voros
University of Peloponnese, Greece

MECOnet, MANT, SMART4ALL
Montenegro, June 2022

Disclaimer

The responsibility for content in presentations and other contributions in this publication rests solely with their authors.

Official website of the event:

https://mecoconference.me/ss-cpsiot2022/

Proceedings of the 3rd Summer School on Cyber-Physical Systems and Internet-of-Things, Vol. III, June 2022

Edited by Lech Jozwiak and Radovan Stojanovic

Technical editors:

Prof. dr Budimir Lutovac, University of Montenegro

Stevan Djuraskovic, University of Florida

Contributors:

Abdelhakim Baouya, Alessandro Capotondi. Andrea Marongiu, Anish Bhobe, Axel Jantsch, Christoph Schmittner, Daniel Madroñal Quintin, Dimitrios Serphanos, Dominique Blouin, Eugenio Villar, Filippo Cugini, Francesca Palumbo, Gianluca Bellocchi, Hector Posadas, Jose María Gandara, Josip Knezović, Lech Jozwiak, Letizia Jaccheri, Luca Benini, Mario Kovač, Martin Ron, Milica Orlandić, Morayo Adedjouma, Muhammad Abdullah Hanif, Muhammad Shafique, Nikolaos Voros, Pavel Burget, Peter Mörtl, Radovan Stojanović, Ramiro Samano Robles, Raul Gomez, Reda Nouacer, Roberto Giorgi, Rupert Schlick, Samir Ouchani, Stavros Koubias, Thomas Bauer, Zhonghai Lu

DOI: https://doi.org/10.5281/zenodo.6698645

Cite as:

Author/s, “Title of contribution-presentation”, in Proceedings of the 3rd Summer School on Cyber-Physical Systems and Internet-of Things, Editors: Lech Jozwiak, Radovan Stojanovic and Nikolaos Voros, Vol. III, June 2022, pp. xx-yy, DOI: https://doi.org/10.5281/zenodo.6698645

Published by Consortium of MECOnet d.o.o. , MANT Association and SMART4ALL Project

For publisher: MECOnet, www.meconet.me

Filipa Lainovica 19, Podgorica, Montenegro

Message from the Chairs,

The Summer School on Cyber-Physical Systems and Internet of Things (SS-CPS&IoT'2022) is the third school in a series, organized in Charming Budva, a Mediterranean pearl.

This year we adapted to still actual COVID19 situation and managed the event on two tracks, on-line and in-venue.

SS-CPS&IoT'2022 aims at serving the following main purposes:

- **advanced training** of industrial and academic researchers, developers, engineers and decision-makers; academic teachers, Ph.D. and M.Sc. students; entrepreneurs, investors, research funding agents, and policy makers; and other participants who want to learn about CPS and IoT engineering;
- **dissemination, exchange and discussion** of advanced knowledge and project results from numerous European R&D projects in CPS and IoT;
- **promotion and facilitation of international contacts and collaboration** among people working or interested in the Embedded Computing, CPS and IoT areas.

The School is open to everybody, but previous knowledge or equivalent practical experience at least at the Bachelor level in engineering (e.g. system, computer, electronic, electrical, automotive, aviation, mechanical, or industrial engineering), computer science, informatics, applied physics or similar is recommended. Industry participation is encouraged.

SS-CPS&IoT is not only to follow courses and learn new knowledge on Embedded Systems, CPS and IoT from top professionals, but to meet people, interact and discuss with outstanding researchers, developers, academic lecturers, advanced students, and other participants, collaborate or start collaborations, and meet many talented people who may become employees of your companies as well.

Distinguishing features of this advanced traditional Summer School are that its lectures, demonstrations, and practical hands-on sessions are given by top European and Worldwide specialists in particular CPS and IoT fields from industry and academia, delivering very fresh advanced

knowledge. They are based on results from numerous currently running or recently finished European R&D projects in CPS and IoT, what gives an excellent opportunity to get acquainted with issues and challenges of CPS and IoT development; actual industrial problems, designs and case studies; and new concepts, advanced knowledge and modern design methods and tools created in the European R&D projects.

This year, we had the honor to invite outstanding lecturers from and outside Europe. Part of the students and lecturers came from the **H2020 project SMART4ALL,** "Self-sustained customized cyber physical system experiments for capacity building among European stakeholders", so it can be said that it was a Joint School of our community with this significant project.

SS-CPS&IoT'2022 is collocated with CPSIoT'2022, 10th International Conference on Cyber-Physical Systems and Internet-of-Things and MECO'2022, 11th Mediterranean Conference on Embedded Computing. The Summer School participants were encouraged to submit their papers to CPSIoT'2022 and MECO'2022, and thus gain additional experience of presenting work in one of the TOP conference in computing.

The CPS&IoT'2022 Summer School Program was composed of four days of lectures, demonstrations, practical hands-on sessions, and discussions, as well as free participation in MECO'2022 and CPSIoT'2022 sessions. The topics of the lectures, demonstrations, and practical hands-on sessions cover major CPS fields and applications. We had about 80 lecturers and students, coming from over 25 countries, WorldWide. We worked for four days in a 32-hour capacity, that is equivalent to an academic workload of **3 ECTS credits**. Detailed list of the presentations including the names of their authors and presenters is provided in this Proceedings. This Proceedings, in addition to its research and educational component, serves as a supplement to the diplomas awarded to School's participants, after testing their activities and knowledge.

SS-CPS&IoT'2022 taken place in Hotel Budva, Montenegro. Budva is a 3500years old town located at the Adriatic Sea coast of Montenegro. It is a popular touristic destination, with its charming Old Town, beautiful natural environment, 35 clean sandy beaches, and proximity to many famous touristic attractions as Kotor, Boka Kotorska, Sveti Stefan, Dubrovnik, and

several national parks. It is an excellent place to have a summer school in a relaxed and friendly atmosphere.

The Chairs of the SS-CPS&IoT’2022 express their thanks to all authors and presenters as well as, to all other people who contributed to the success of the Summer School. We are especially proud on 3rd generation of students who successfully finished School and showed an enviable level of knowledge and interest.

We are very grateful to **Professor Budimur Lutovac**, Publication Chair of CPSIoT’2022 and MECO’2022 and student moderator **Stevan Djurašković** helping us in logistics and in composing these Proceedings.

We hope to see you again next year, mostly, in-venue, in good health and friendly atmosphere.

Yours,

Lech Jóźwiak Eindhoven University of Technology, The Nederland

Radovan Stojanović University of Montenegro, Montenegro

Nikolaos Voros, University of Peloponnesus, Coordinator of SMART4ALL Project, Greece

Contents

Disclaimer

Lech Jóźwiak, Radovan Stojanović
Introduction to the CPS&IoT'2022 Summer School 1

Luca Benini
PULP: Extreme Energy Efficiency for Extreme Edge AI Acceleration 4

Lech J· Ŧwiak
Green CPS and IoT for Green World 56

Mario Kovač, Josip Knezović
European Processor Initiative Technology for Exascale Era 158

Gianluca Bellocchi, Alessandro Capotondi, Andrea Marongiu, Francesca Palumbo, Daniel Madroñal Quintin
Accelerator-Rich FPGA Architecture Exploration via a Programmable and Reconfigurable Overlay 189

Reda Nouacer, Morayo Adedjouma
From Embedded-Systems towards swarms: opportunities and challenges 287

Letizia Jaccheri
Software for a Better Society. 323

Roberto Giorgi
Extending Performance and Reliability via Modular FPGA Clusters. 371

Filippo Cugini, Pavel Burget, Martin Ron
Edge computing: the BRAINE solution. 410

Axel Jantsch, Zhonghai Lu
Embedded Machine Learning. 471

Muhammad Shafique, Muhammad Abdullah Hanif
Embedded Machine Learning for the Edge: From Algorithms to Architectures. 558

Eugenio Villar, Hector Posadas, Raul Gomez, Jose María Gandara
Modeling, design and Implementation of drone-based services 592

Dimitrios Serphanos, U. Patras and CTI, Stavros Koubias, U. Patras and ISI
Synthesis of Runtime Monitors for Safe and Secure Industrial Systems. 637

Dominique Blouin, Anish Bhobe
Embedded systems modeling, analysis and automatic code generation with AADL and RAMSES. 667

Rupert Schlick, Thomas Bauer
How to design and tailer a perfect fitting verification and validation process for your CPS&IoT project?. 719

Peter Mörtl, Nikolai Ebinger
Framework to facilitate Trustworthiness of Smart Systems for End Users……. 777

Ramiro Samano Robles
Reference architecture for trusted AIoT systems: certification, standardization, and regulation. 806

Christoph Schmittner
Cybersecurity Engineering and Management 875

Samir Ouchani
Secure and Reliable Smart Cyber Physical Systems 946
Abdelhakim Baouya
Artificial Intelligence meets Formal Methods: Generation and verification of learned stochastic automata 1013
Radovan Stojanovic
Principles of performance effective nodes design for smart systems.....1050
Milica Orlandić
Data Processing Pipelines on small satellites and drones: challenges and solutions 1076
Nikolaos Voros
The achievements of SMART4ALL project in Customized Low-Energy Computing for CPS 1109
Abeer Akkad, Gary Wills, Abdolbaghi Rezazadeh
An IoT-enabled Smart Grid: Definitions, Characteristics, Challenges, and Future Directions 1155

Schedule-CPS&IoT'2022 Summer School on Cyber-Physical Systems and Internet-of- Things 1164

3rd Summer School on Cyber Physical Systems and Internet of Things - SS-CPSIoT'2022 3rd Generation (Students and Teachers) 1165

Certificate of Attendance 1166

Author Index 1167

Photo gallery 1169

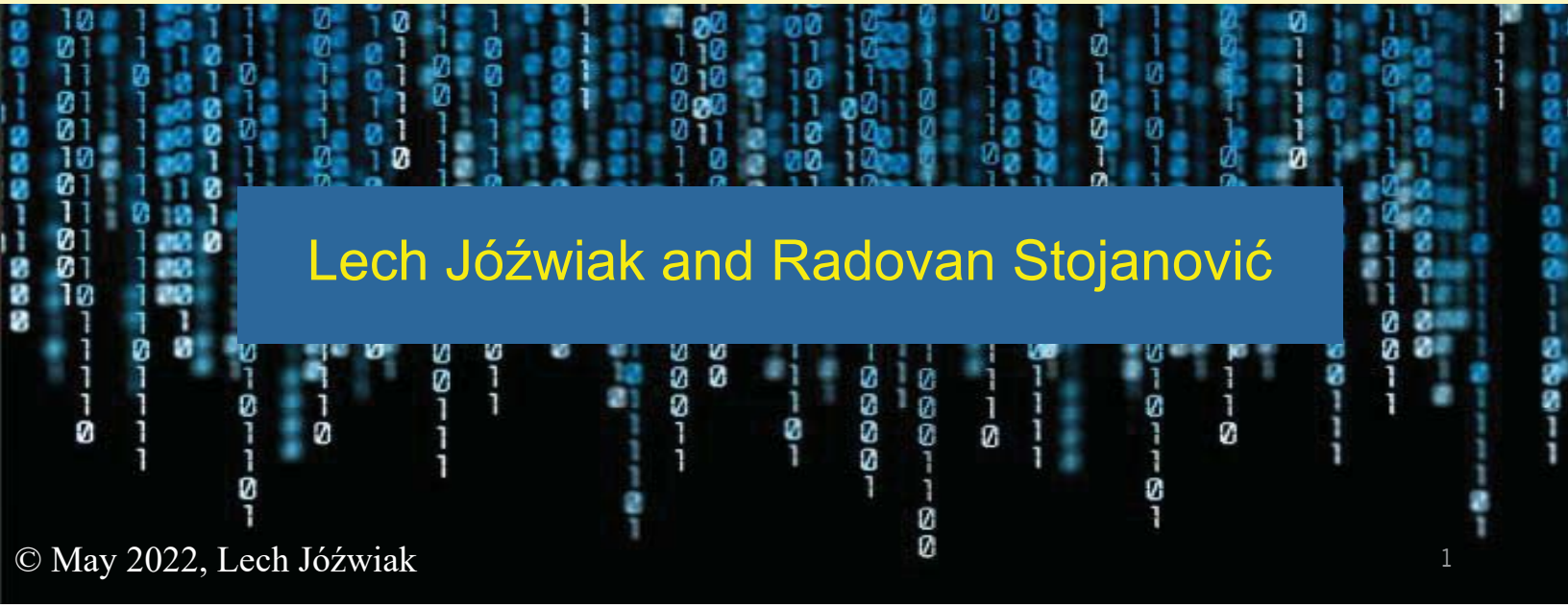
CPS&IoT

CPS&IoT'2022 Summer School

SMART4ALL

Budva, Montenegro
June 7-10, 2022

Introduction



Lech Jóźwiak and Radovan Stojanović

Introduction

- Systemic drawbacks of the traditional economy and cumulation of bad decisions driven by the short-term profit and made without adequately accounting for long-term consequences resulted in the **huge global environmental disaster**
- Innovations exploiting modern CPS and IoT technologies have a high potential to significantly improve systems used by us or that we are part of
- To recover from the environmental disaster and further develop:
  - *a model of a well regulated and controlled effective and efficient system should be applied to all kinds of systems, collaboration chains and related flows*
  - *modern CPS and IoT technologies should be used to much better control and optimize the social, physical and life systems than till now*
  - *methodologies of circular regenerative economy and quality-driven design should be used to design the systems*
- In this CPS&IoT Summer School you will have a unique occasion to be informed on and to discuss the most recent European R&D developments in CPS and IoT

2

Outline of the CPS&IoT'2021 Summer School

1. Introduction to CPS and IoT
2. Green CPS and IoT
3. Computing and communication technologies for CPS and IoT
4. Machine Learning and Edge Computing
5. Modeling, design and implementation of CPS and IoT
6. Trustworthy CPS and IoT: reliability, security and safety
7. Energy-efficient computing for CPS and IoT
8. Closing of the CPS&IoT2022 Summer School

PULP PLATFORM
Open Source Hardware, the way it should be!

PULP: Extreme Energy Efficiency for Extreme Edge AI Acceleration

Luca Benini
<lbenini@iis.ee.ethz.ch,luca.Benini@unibo.it>

Prof. of Digital Circuit and Systems @ ETHZ and UNIBO. h-index=114, 58'000+ citations, 1'000+ publications, fellow IEEE, ACM, Chief Architect in STMicroelectronics (2009-2012) Group of 100+ people

European Research Council

EuroHPC Joint Undertaking

Horizon Europe

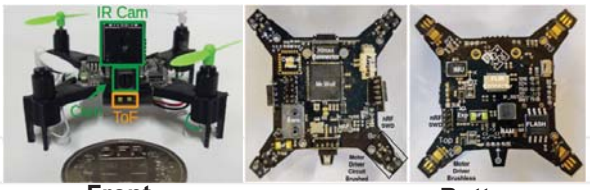
FONDS NATIONAL SUISSE
SCHWEIZERISCHER NATIONALFONDS
FONDO NAZIONALE SVIZZERO
SWISS NATIONAL SCIENCE FOUNDATION

http://pulp-platform.org @pulp_platform

ETH zürich

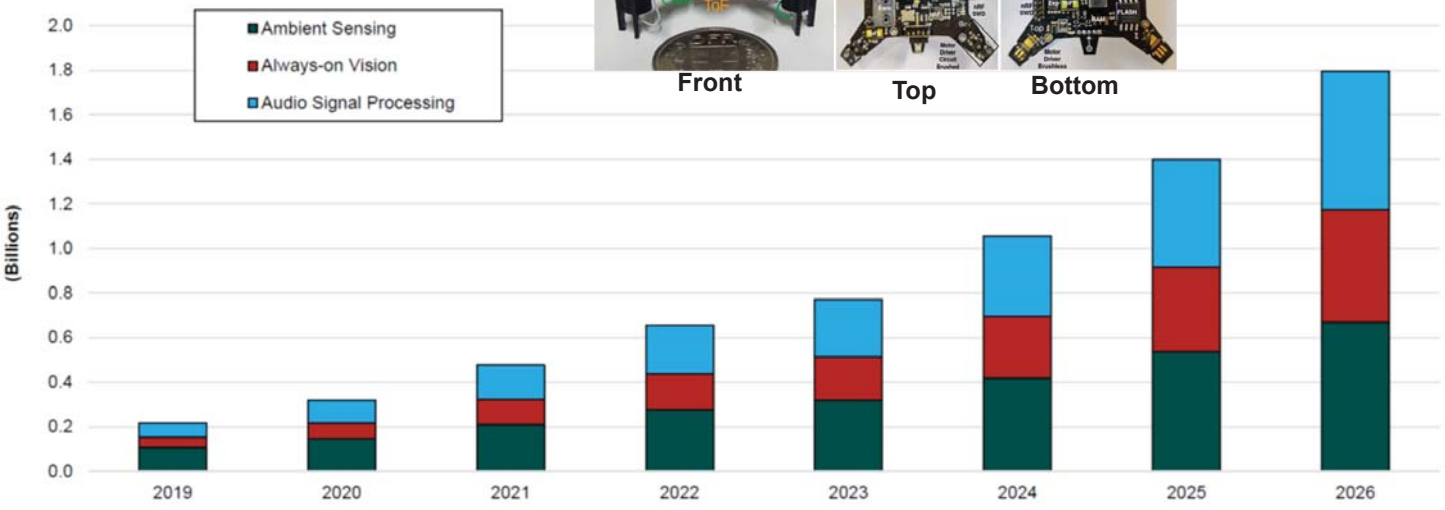
CloudML → TinyML

TinyML Opportunity



Front Top Bottom

[ABI research 21]

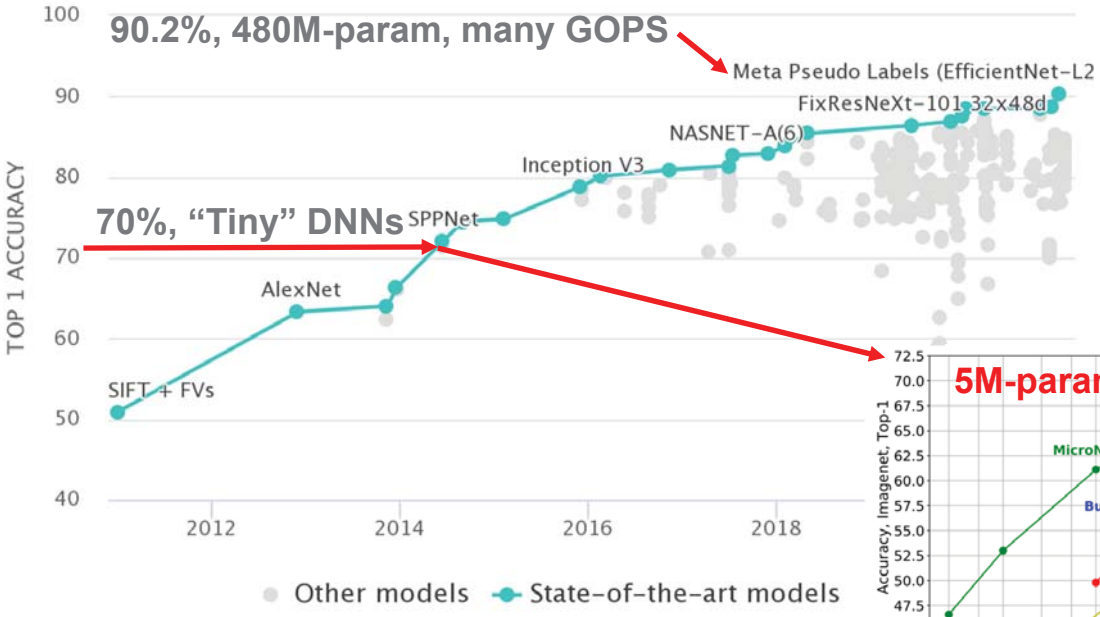


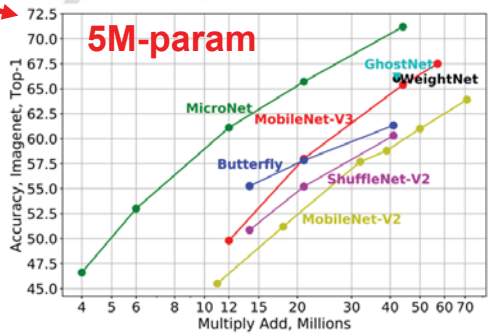
TinyML challenge

AI capabilities in the power envelope of an MCU: **10-mW peak (1mW avg)**

AI Workloads - DNNs

H Pham 2021(Google) arXiv:2003.10580v3

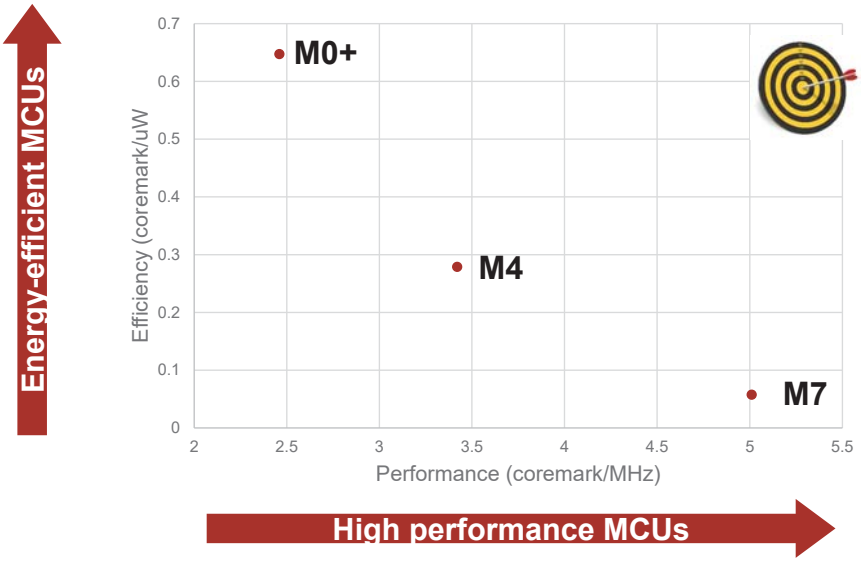




High OP/B ratio
Massive Parallelism
MAC-dominated
Low precision OK

Energy efficiency @ GOPS is the Challenge

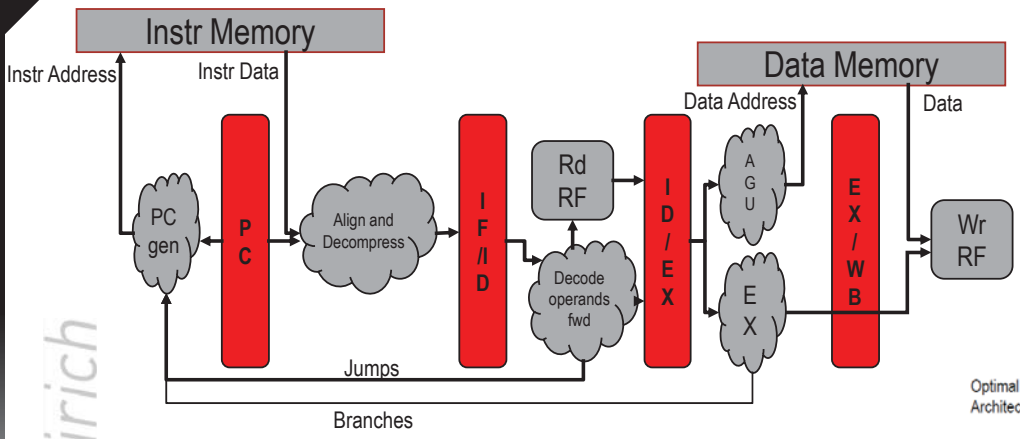
ARM Cortex-M MCUs: M0+, M4, M7 (40LP, typ, 1.1V)*



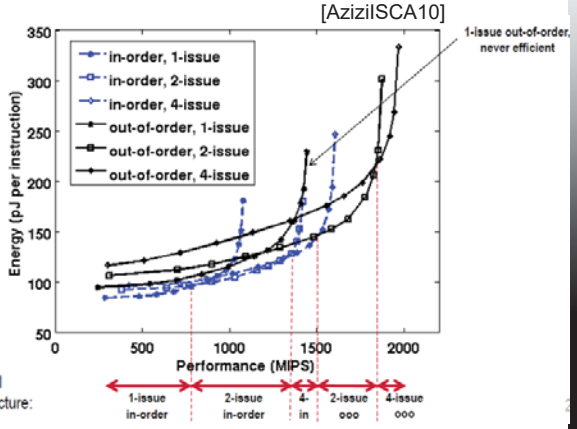
1pJ/MAC=1GMAC/mW

*data from ARMs web

The Tunnel: High-Performance vs. Energy-Efficient



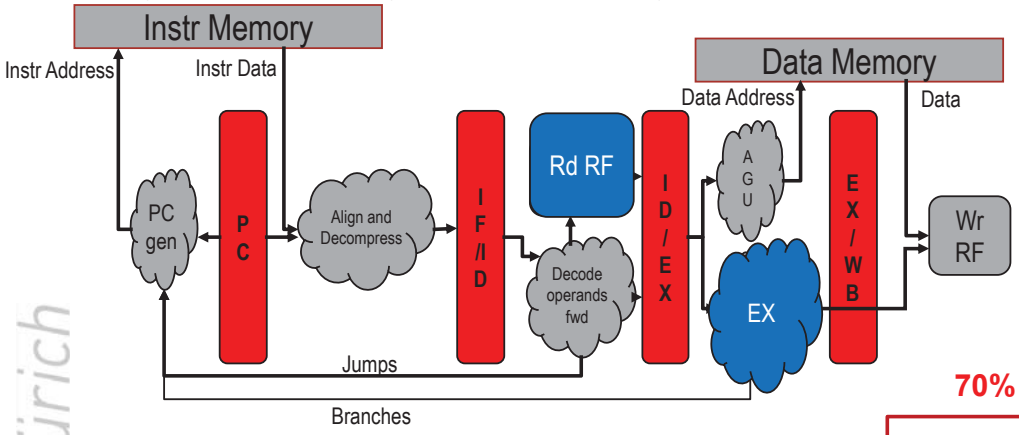
[AziziISCA10]



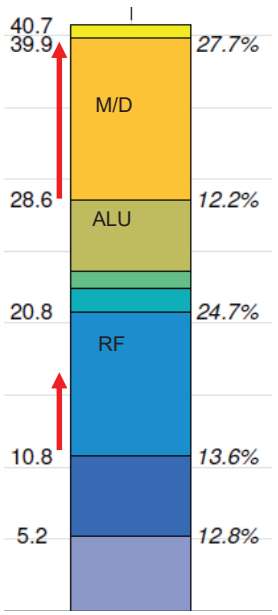
- **"Classical" core performance scaling trajectory**
  - Faster CLK → deeper pipeline → IPC drops
  - Recover IPC → superscalar → ILP bottleneck (dependencies)
  - Mitigate ILP bottlenecks → OOO → huge power, area cost!

A way Out: Processor Specialization

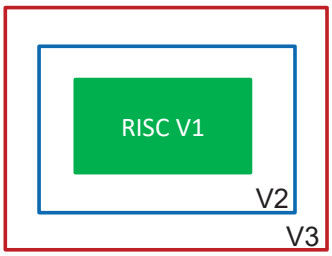
3-cycle ALU-OP, 4-cyle MEM-OP→ only IPC loss: LD-use, Branch



[Gautschi et al. TVLSI 2017]



70% RF+DP



V1 Baseline RISC **(not good for ML)**

V2
Extensions for Data Processing
Data motion (e.g. auto-increment)
Data processing (e.g. MAC)

V3
Domain specific data processing
Narrow bitwidth
HW support for special arithmetic

ISA extension cost 25 kGE → 40 kGE (1.6x), energy efficient if 0.6T_{exec}

PULP-NN: Xpulp ISA exploitation

8-bit Convolution

HW Loop

LD/ST with post increment

8-bit SIMD sdotp

8-bit sdotp + LD

RV32IMC

N

```
addi  a0,a0,1
addi  t1,t1,1
addi  t3,t3,1
addi  t4,t4,1
lbu   a7,-1(a0)
lbu   a6,-1(t4)
lbu   a5,-1(t3)
lbu   t5,-1(t1)
mul   s1,a7,a6
mul   a7,a7,a5
add   s0,s0,s1
mul   a6,a6,t5
add   t0,t0,a7
mul   a5,a5,t5
add   t2,t2,a6
add   t6,t6,a5
bne   s5,a0,1c000bc
```

RV32IMCXpulp

N/4

```
lp.setup
p.lw  w1, 4(a0!)
p.lw  w2, 4(a1!)
p.lw  x1, 4(a2!)
p.lw  x2, 4(a3!)
pv.sdotsp.b  s1, w1, x1
pv.sdotsp.b  s2, w1, x2
pv.sdotsp.b  s3, w2, x1
pv.sdotsp.b  s4, w2, x2
end
```

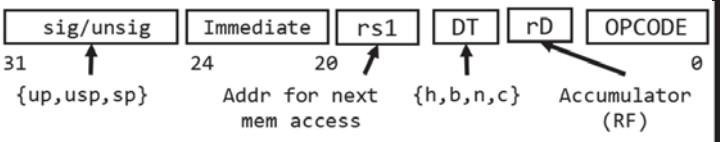
can we remove?

Yes! DOTP+LW

N/4

```
Init NN-RF (outside of the loop)
lp.setup
pv.nnsdotup.h  s0,ax1,9
pv.nnsdotsp.b  s1, aw2, 0
pv.nnsdotsp.b  s2, aw4, 2
pv.nnsdotsp.b  s3, aw3, 4
pv.nnsdotsp.b  s4, ax1, 14
end
```

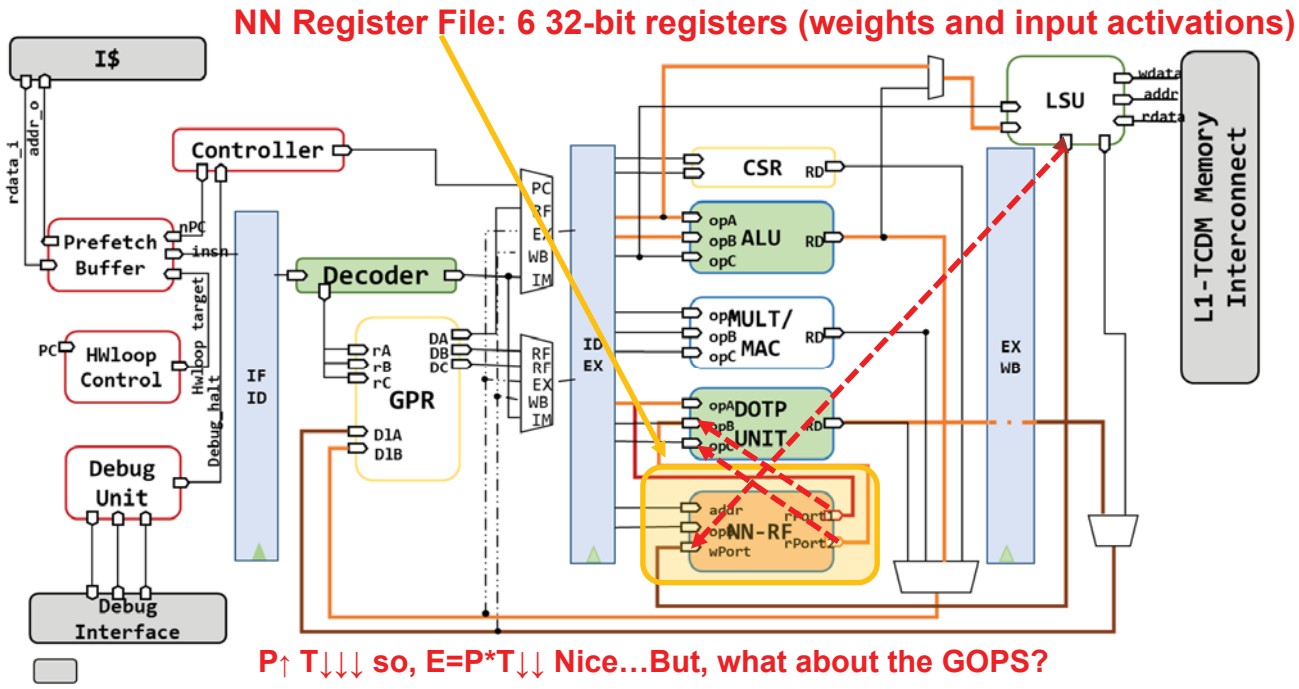
```
pv.nnsdot{up,usp,sp}.{h,b,n,c} rD, rs1, Imm
```



9x less instructions than RV32IMC

14.5x less instructions at an extra 3% area cost (~600GEs)

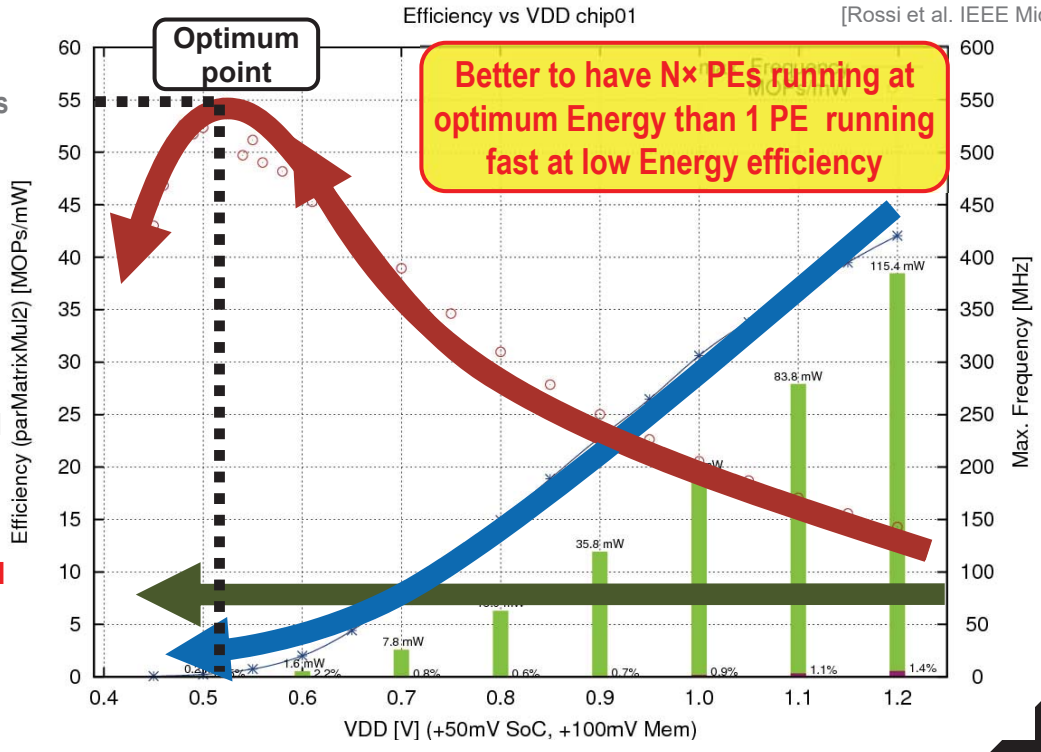
Supporting dotp+ld



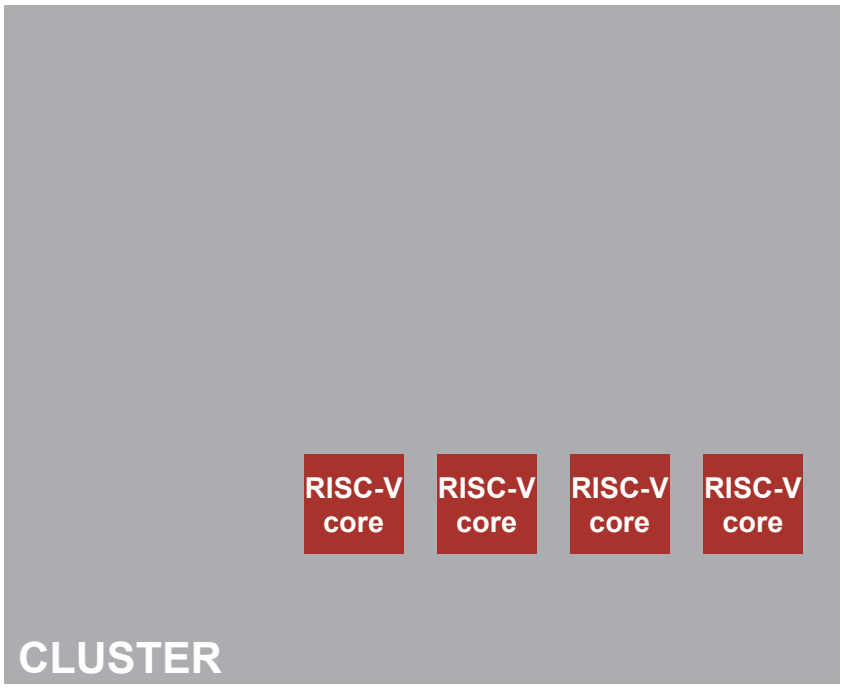
ML & Parallel, Near-threshold: a Marriage Made in Heaven

- As VDD decreases, operating speed decreases
- However efficiency increases→ more work done per Joule
- Until leakage effects start to dominate
- Put more units in parallel to get performance up and keep them busy with a parallel workload

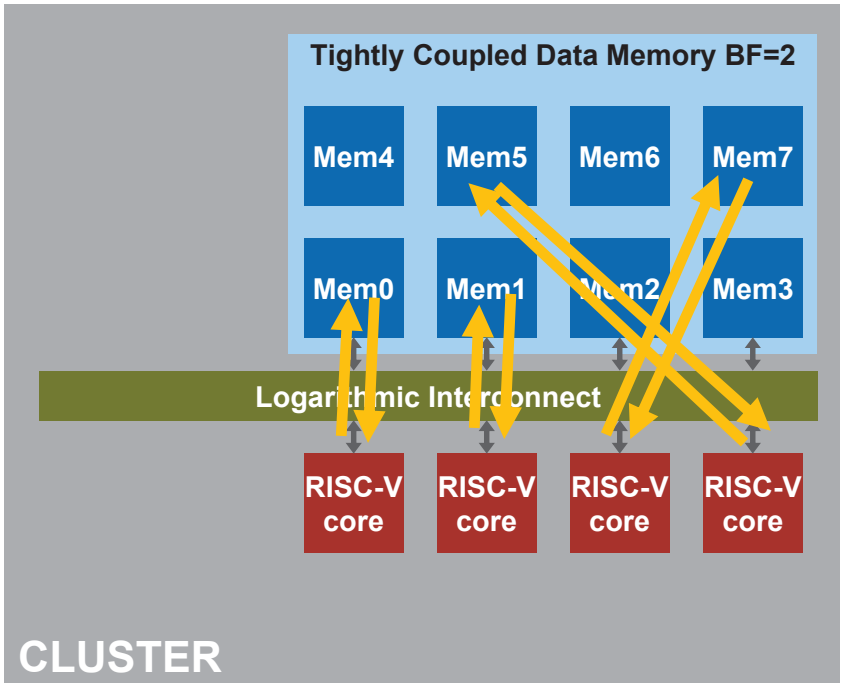
ML is massively parallel and scales well (P/S ↑ with NN size)



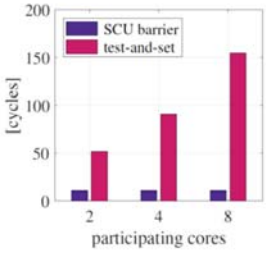
Multiple RI5CY Cores (1-16)

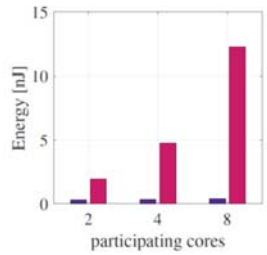


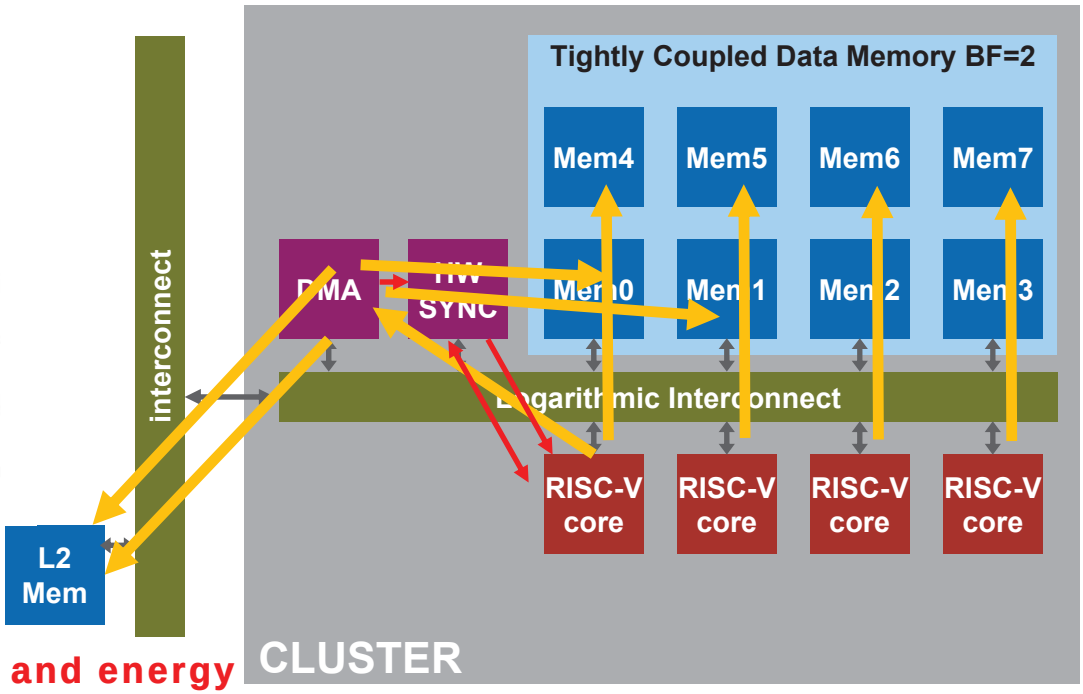
Low-Latency Shared TCDM



DMA for data transfers from/to L2 + Fast synchro



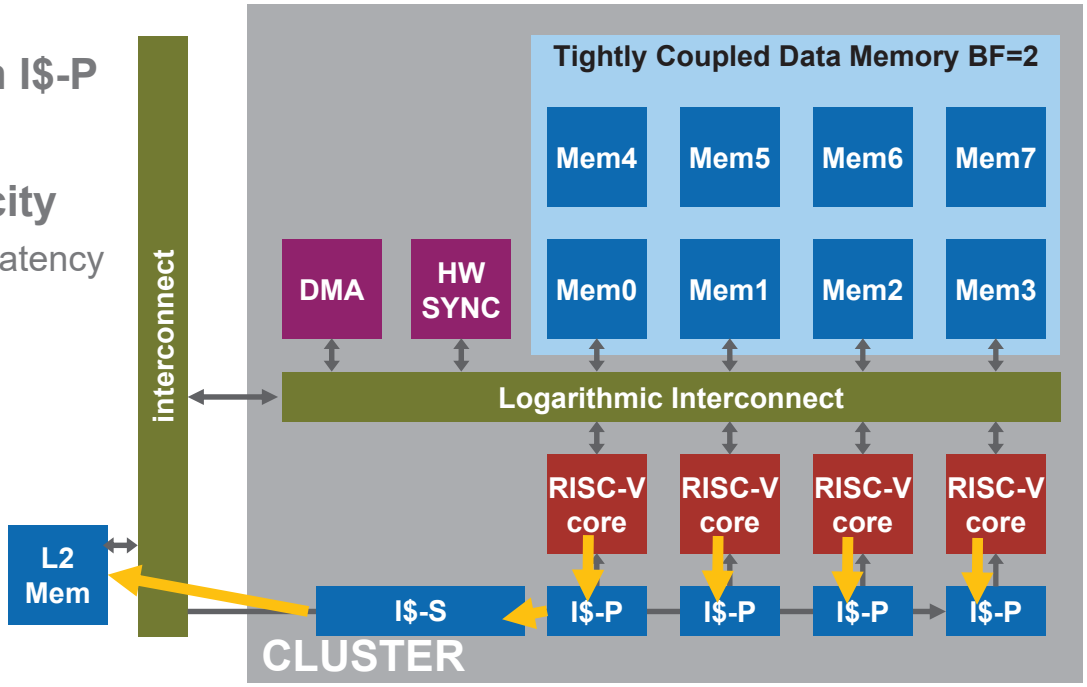




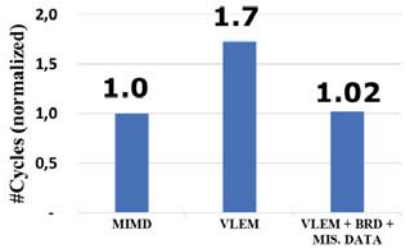
~15x latency and energy reduction for a barrier [Glaser TPDS20]

Shared instruction cache with private "loop buffer"

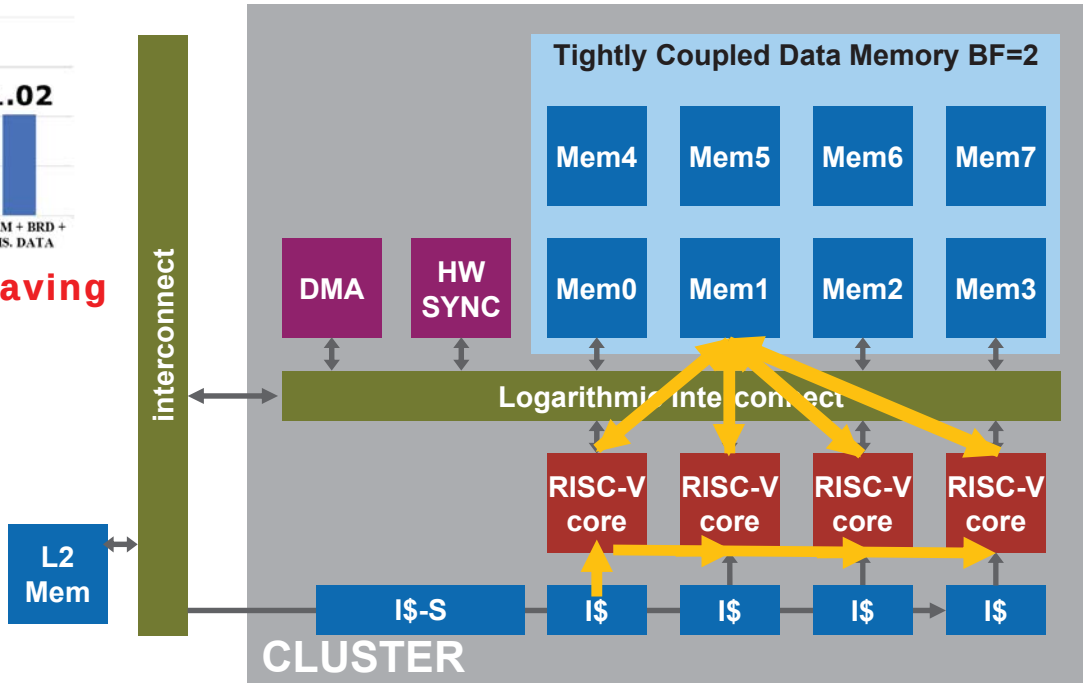
- **Most IFs from I$-P**
  - Low IF energy
- **I$-S for capacity**
  - Reduces miss latency



Vector Lockstep Execution Mode (VLEM)



~45% energy saving

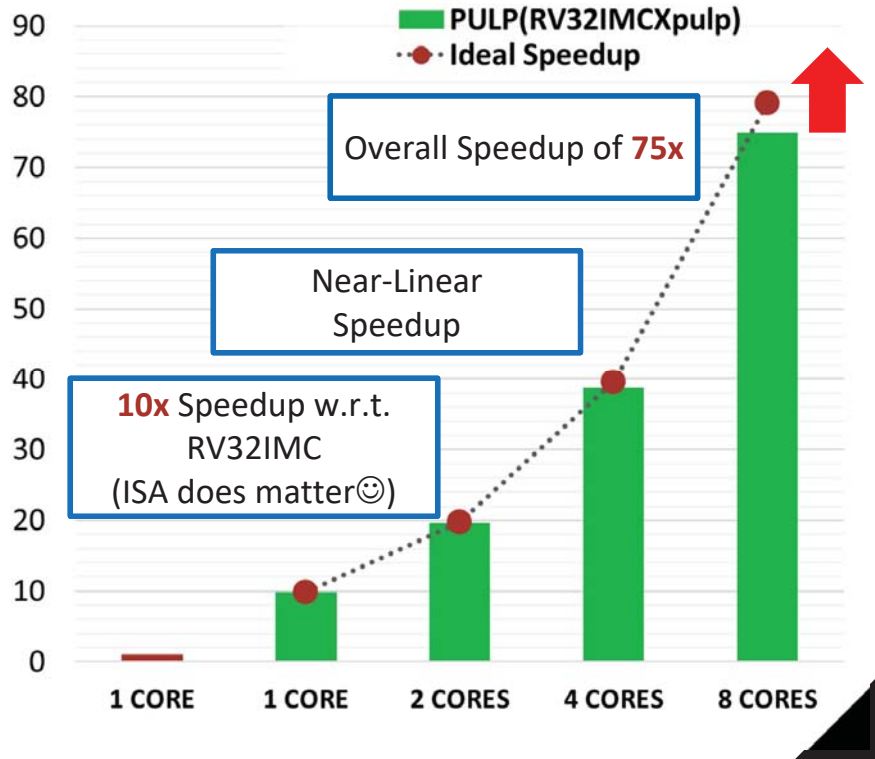


[Garofalo et al. ESSCIR21]

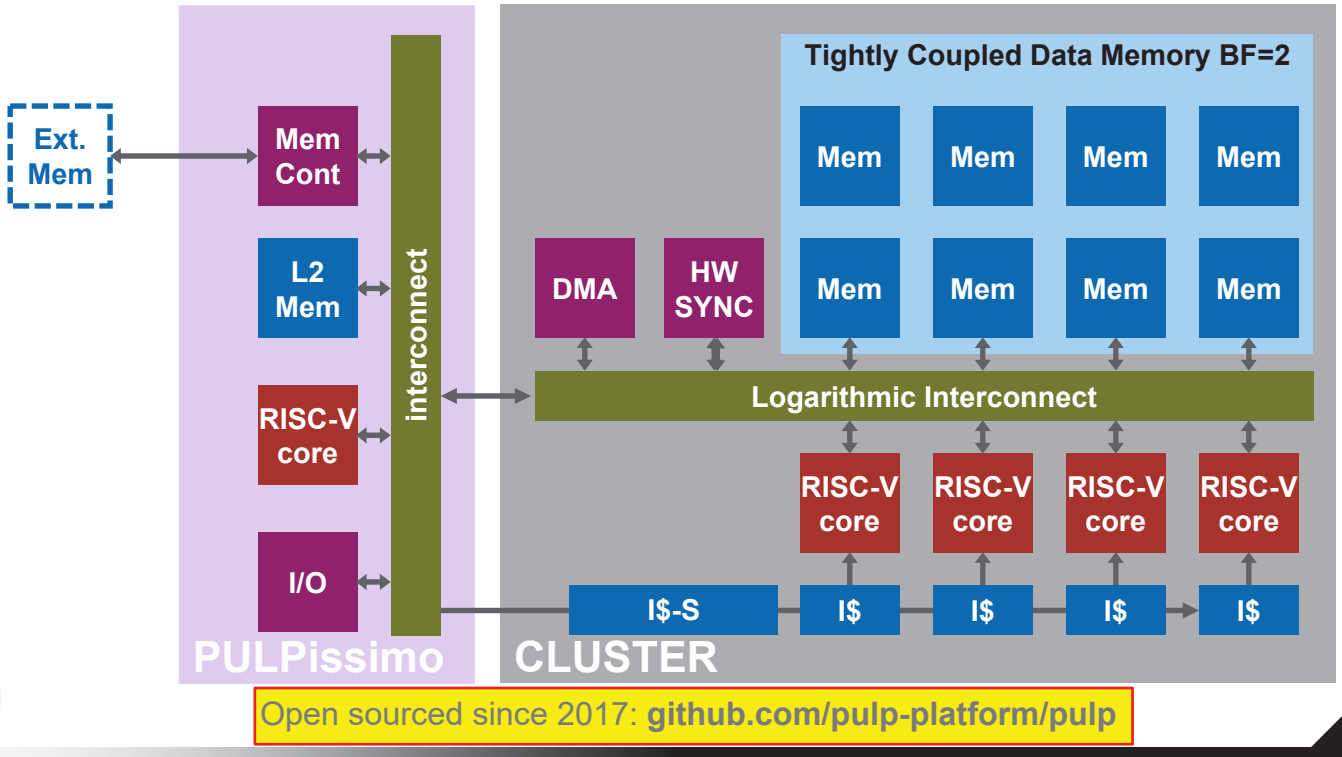
Results: RV32IMCXpulp vs RV32IMC

[Garofalo et al. Philos. Trans. R. Soc 20]

- **8-bit convolution**
  - Open source DNN library
- **10x through xPULP**
  - Extensions bring real speedup
- **Near-linear speedup**
  - Scales well for regular workloads
- **75x overall gain**
  - 2 orders of magnitude with DOTP+LW (**122x**)
  - Sub-byte (nibble, crumb) supported (**537x**, **939x**)



An additional I/O controller is used for IO & PM



Open sourced since 2017: **github.com/pulp-platform/pulp**

Deploying DNNs on PULP

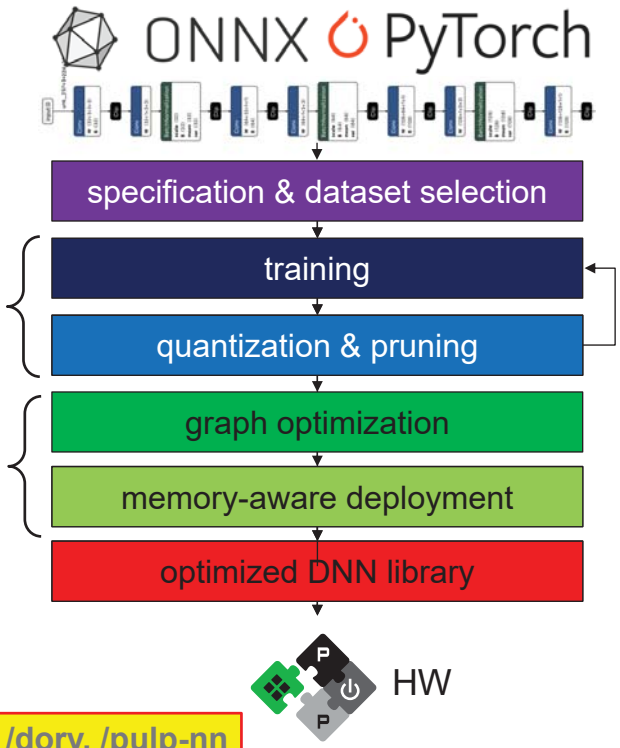
[Burrello et al. TCOMP21]

QuantLab
Quantization Laboratory

NEMO
NEural **M**inimization for pyt**O**rch

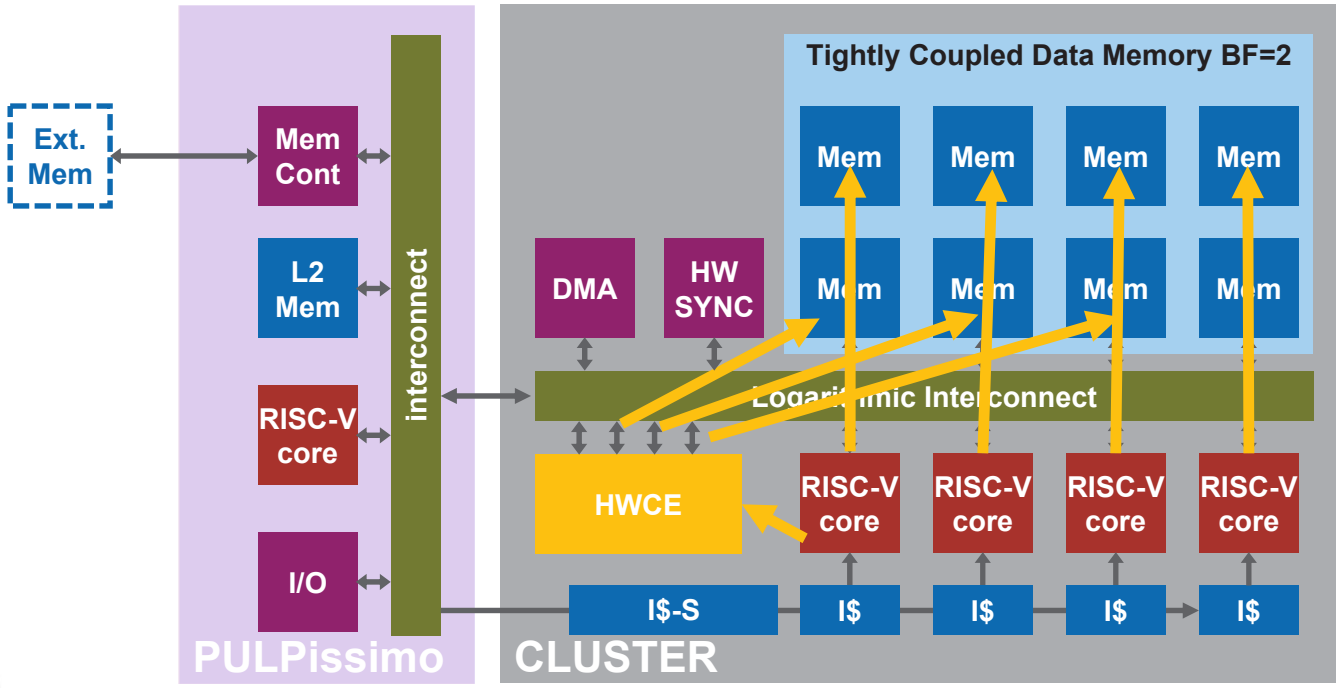
DORY
Deployment **O**riented to memo**RY**

PULP-NN
PULP **N**eural **N**etwork backend



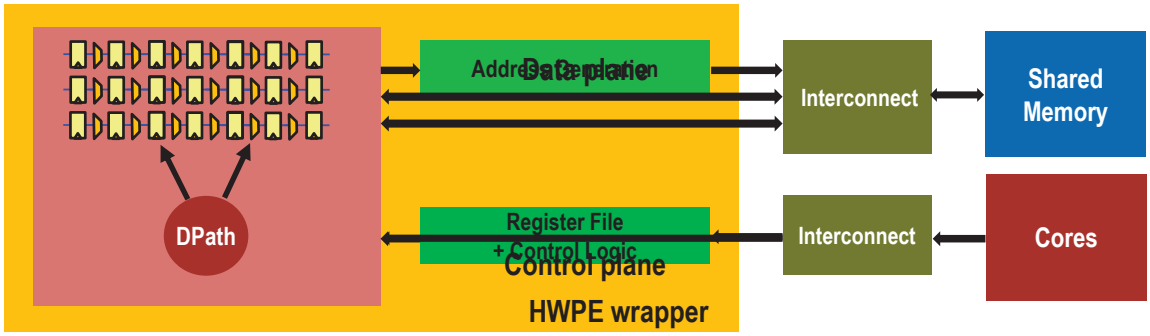
github.com/pulp-platform/nemo, /dory, /pulp-nn

What's next? Tightly-coupled HW Compute Engine



Acceleration with flexibility: zero-copy HW-SW cooperation

Hardware Processing Engines (HWPEs)



HWPE efficiency ($\frac{MAC}{A(mm^2),E(J),W(bps)}$) vs. optimized RISC-V core

1. Dedicated control (no I-fetch) with shadow registers (overlapped config-exec)
2. Specialized high-BW interco into L1 (on data-plane)
3. Specialized datapath: supporting configurable & aggressive quantization

Binary-Based Quantization (BBQ)

QNN layer :

$$\mathbf{y}(k_{out}) = \boldsymbol{quant}\left(\sum_{k_{in}} \left(\mathbf{W}(k_{out}, k_{in}) \otimes \mathbf{x}(k_{in})\right)\right)$$

INT32 accumulator

Q-bit output fmaps

M-bit weights

N-bit input fmaps

Many $M \times N$ bits products...

... but one $M \times N$ **1-bit** products!

$$\mathbf{y}(k_{out}) = \boldsymbol{quant}\left(\sum_{i=0..M}\sum_{j=0..N}\sum_{k_{in}} 2^i 2^j \left(\mathbf{W_{bin}}(k_{out}, k_{in}) \otimes \mathbf{x_{bin}}(k_{in})\right)\right)$$

power-of-2 scaling factors

Q-bit output fmaps

1-bit weights

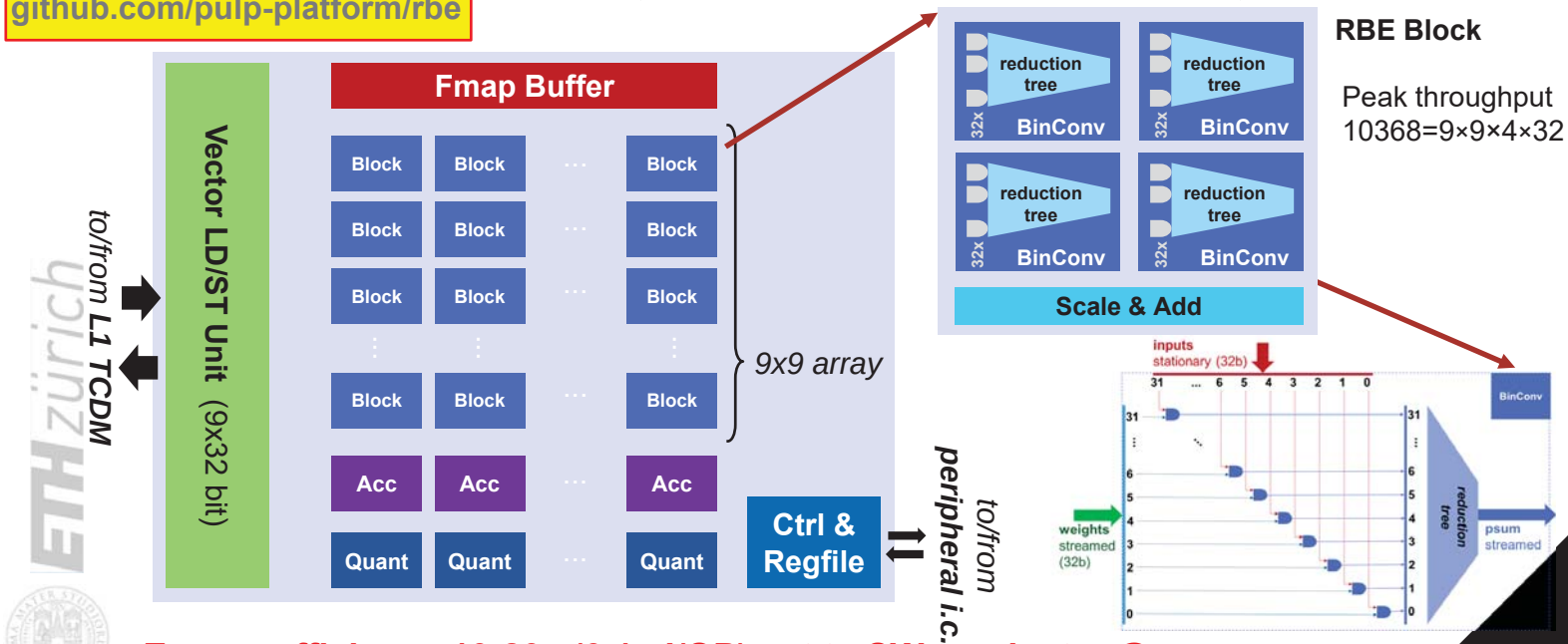
1-bit input fmaps

One quantized NN can be emulated by superposition of power-of-2 weighted $M \times N$ binary NN

Reconfigurable Binary Engine

$$\mathbf{y}(k_{out}) = \boldsymbol{quant}\left(\sum_{i=0..M}\sum_{j=0..N}\sum_{k_{in}} 2^i 2^j \left(\mathbf{W_{bin}}(k_{out}, k_{in}) \otimes \mathbf{x_{bin}}(k_{in})\right)\right)$$

github.com/pulp-platform/rbe

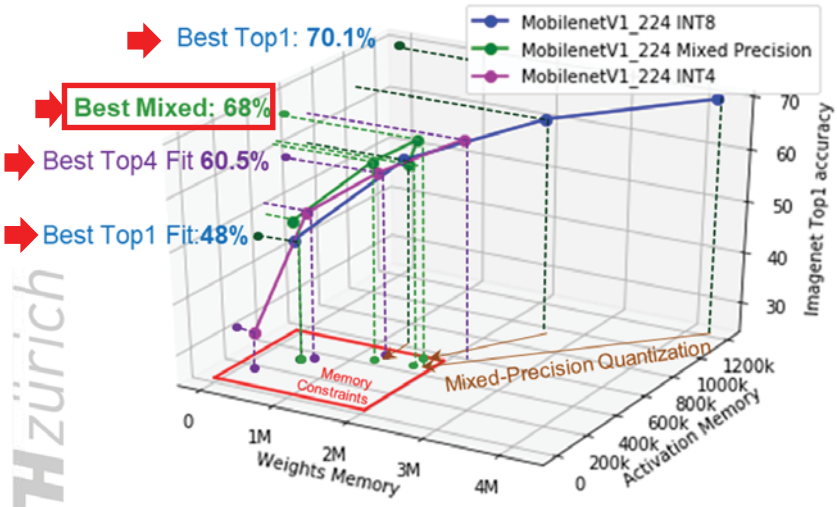


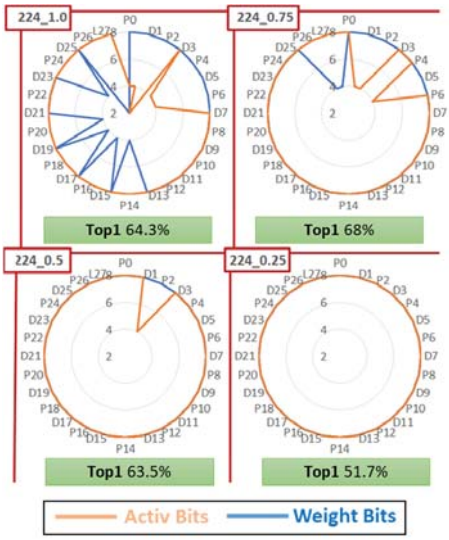
Energy efficiency 10-20x (0.1pJ/OP) wrt to SW on cluster @same accuracy

Mixed-Precision Quantized Networks – CMIX-NN

[Capotondi et al. TCAS II, 2020]

Apply **tensor-wise quantization** to **fit memory constraints** with **low accuracy drop**





Rule-based bit precision selection based on memory constraints

Avgbit 6.7 Act

Avgbit 5.9 Wgt

Only **-2%** wrt most accurate INT8 mobilenetV1 (224_1.0) which does not fit on-chip

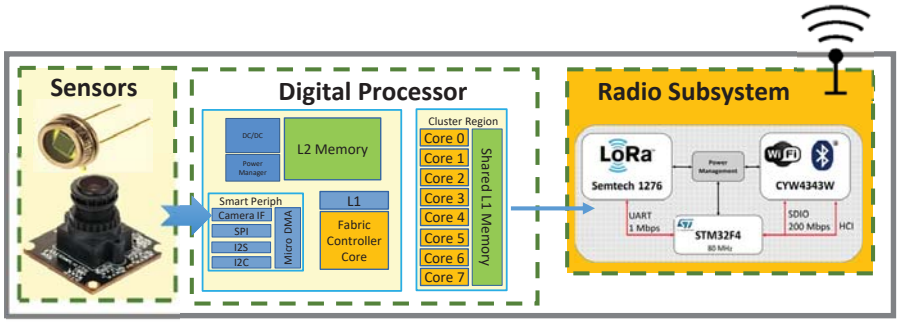
+8% wrt most accurate INT8 mobilenetV1 fitting on-chip (192_0.5)

+7.5% wrt most accurate INT4 mobilenetV1 (224_1.0) fitting on chip

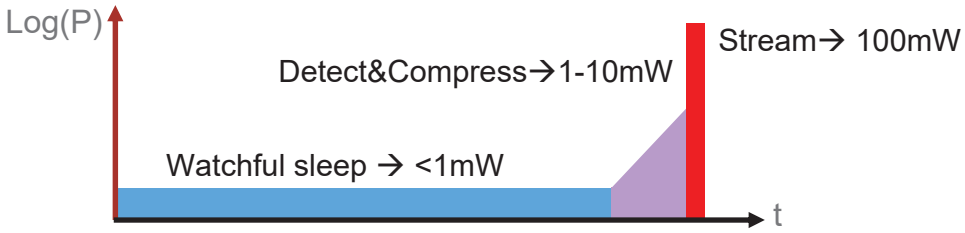
22

Not only Efficiency: Achieving sub-mW Average Power?

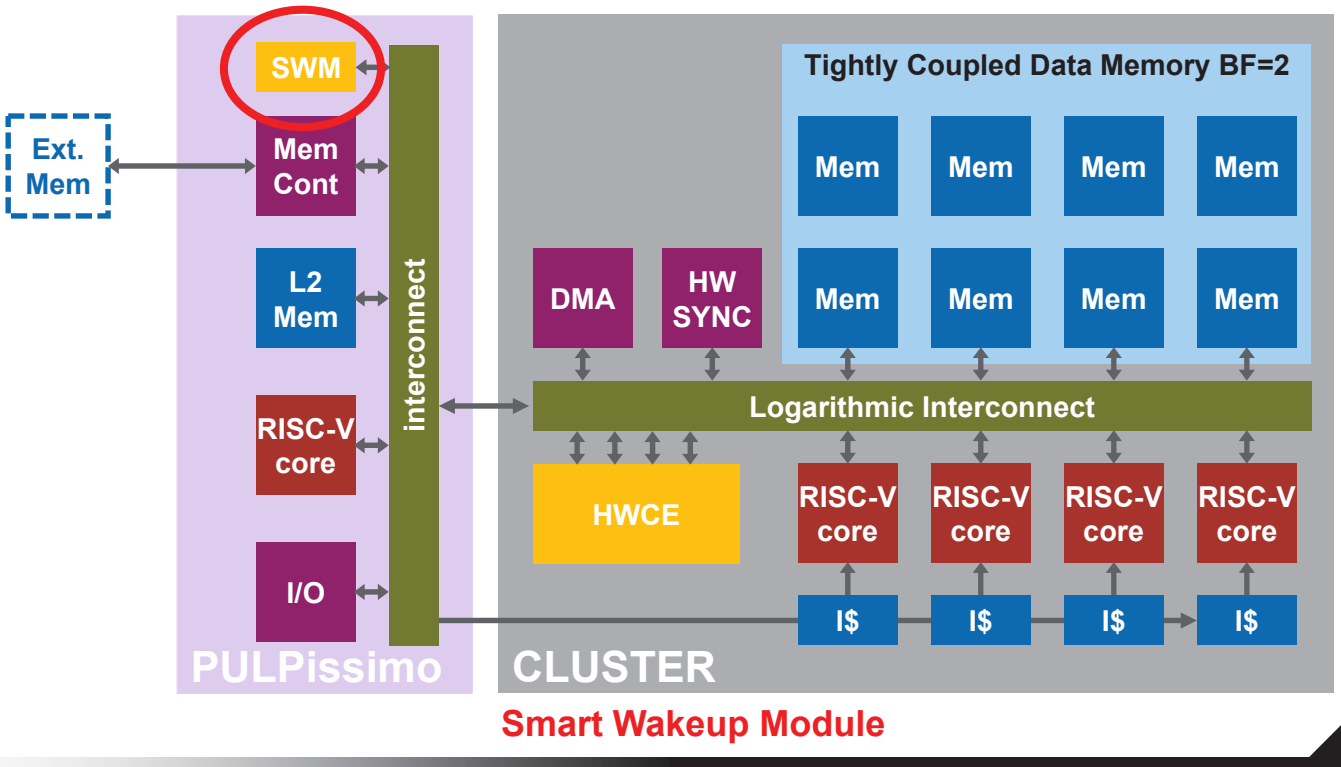
1mW average power with 10mW active power (10GOPS @ 1pJ/OP) → **sub mW sleep**



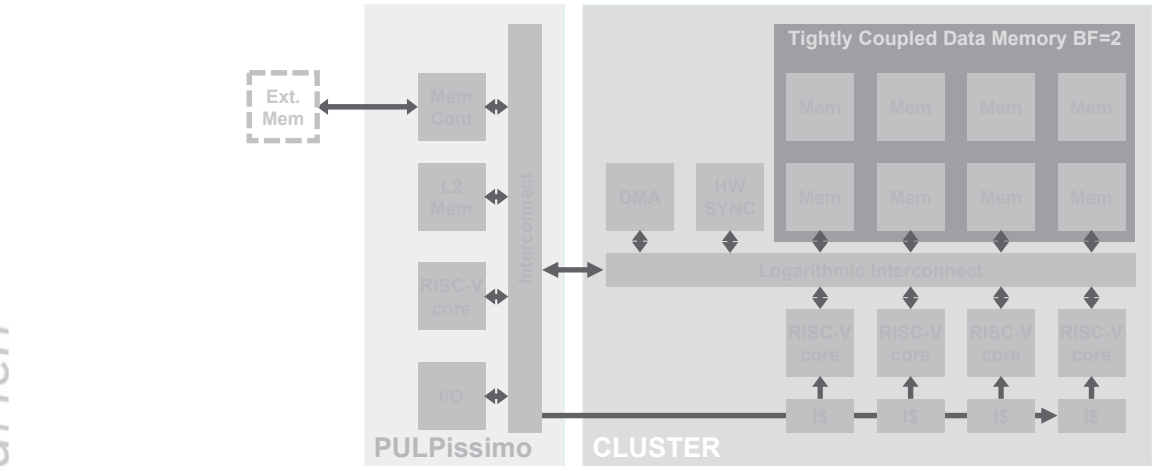
Duty cycling not acceptable when input events are asynchronous → **watchful Sleep**

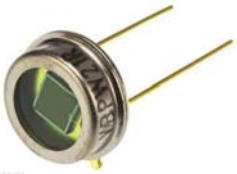


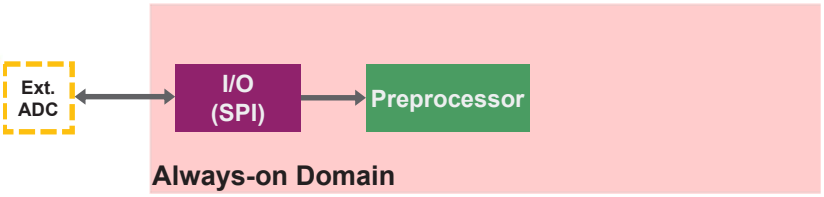
Need µW-range always-on Intelligence



HD-Based smart Wake-Up Module

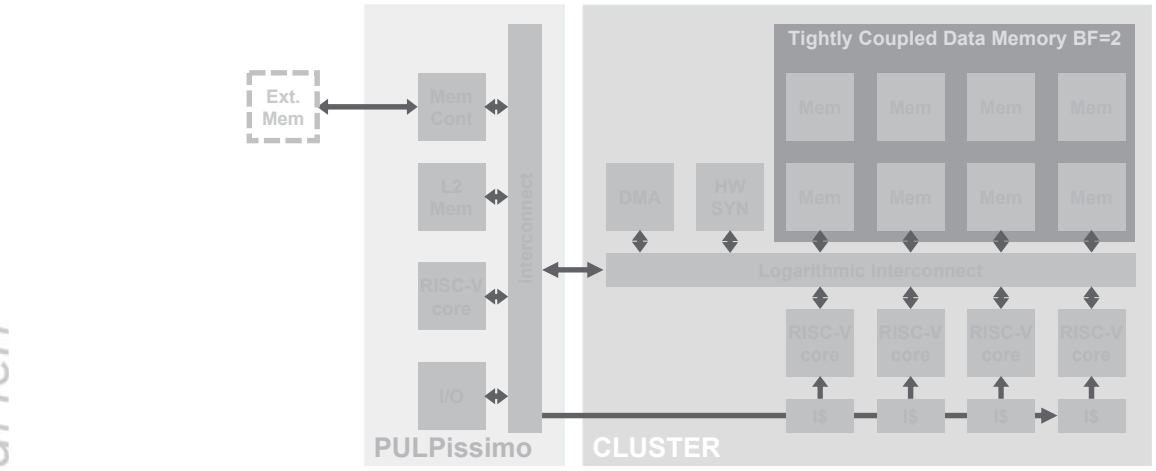


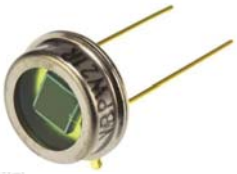


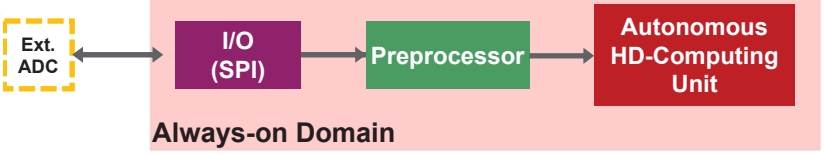


25

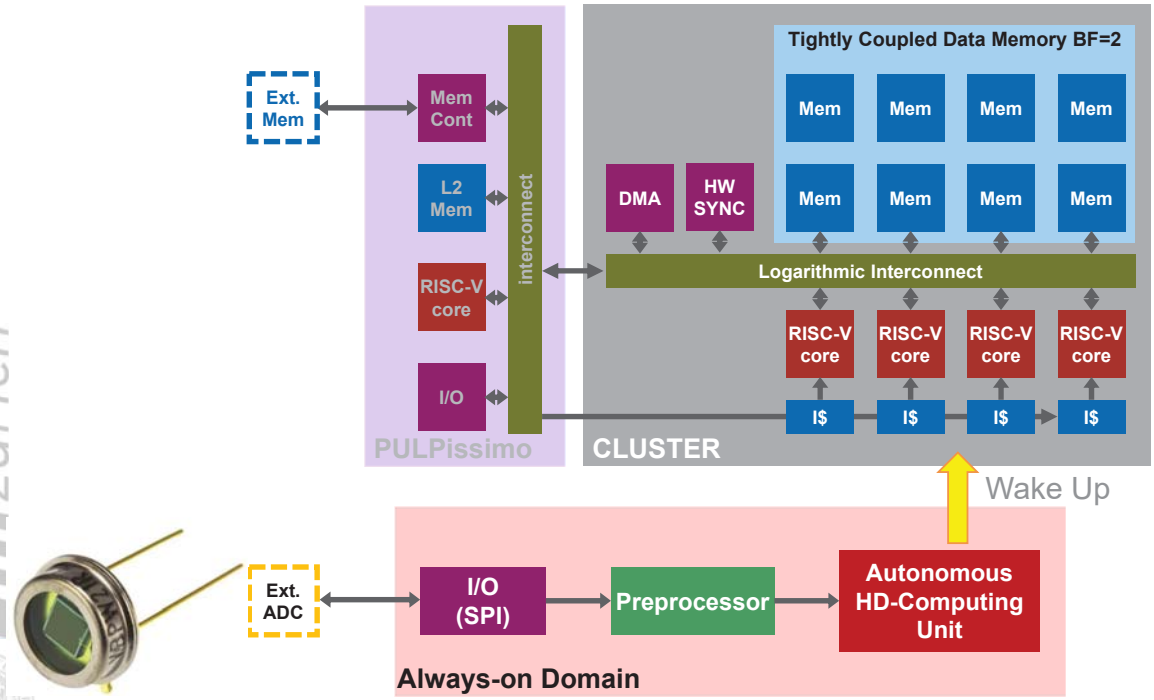
HD-Based smart Wake-Up Module



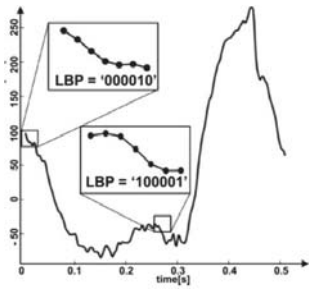




HD-Based smart Wake-Up Module



Not Only CNNs: Hyper-Dimensional Computing



Mapping → [0 1 0 1 ..... 1] (1st 2nd 3rd 4th ... 1000th)

Low Dimensional Input Data (e.g. 7-bit LBP)

[0 1 0 1 ..... 1]
[1 1 1 0 ..... 1]
[1 1 0 0 ..... 0]
[0 1 1 1 ..... 1]

HD-Encoding

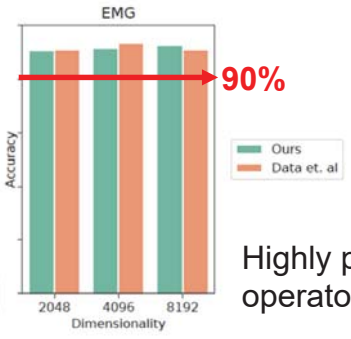
- Component-wise Majority
- XOR
- Permutation

Search Vector: [1 1 0 1 ..... 1]

Similarity Search (e.g. Hamming Distance)

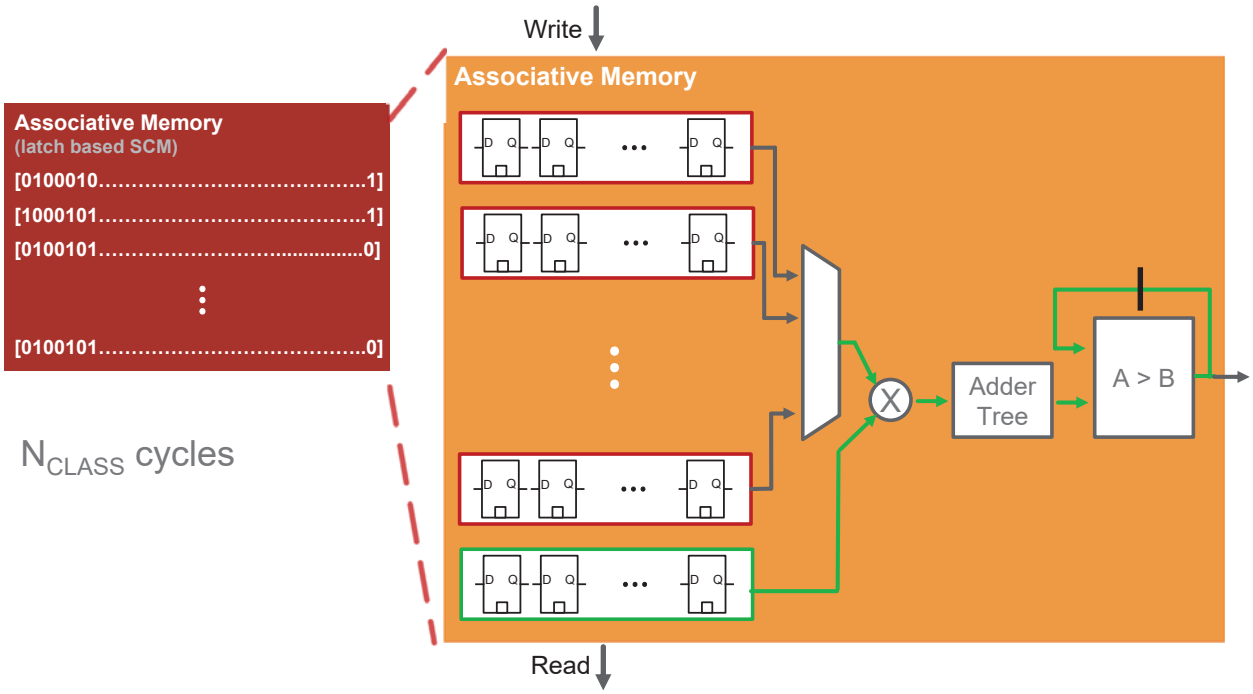
Prototype Vectors

Associative Memory
[0 1 0 1 ..... 1]
[1 1 1 0 ..... 1]
[1 1 0 0 ..... 0]
[0 1 1 1 ..... 1]
[1 1 1 1 ..... 1]
[0 1 0 1 ..... 1]
[0 1 0 1 ..... 1]



Highly parallel, fault-tolerant binary operators, assoc-min-distance search → Merge storage & computation i.e. **In-memory computing**

In-memory Hyperdimensional Computing

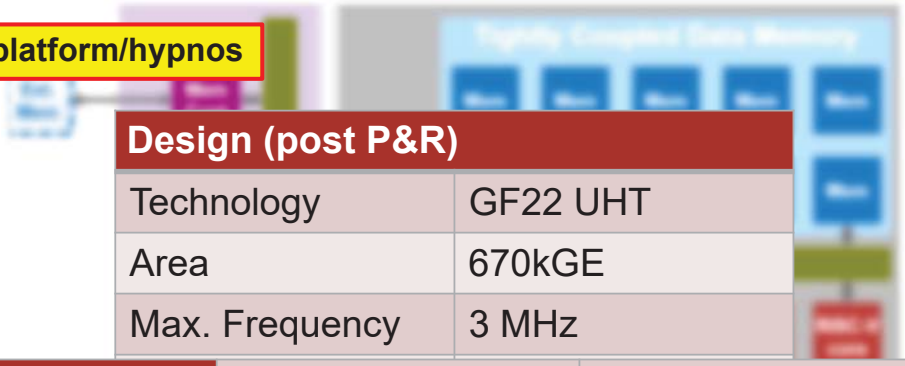


N_{CLASS} cycles

HD-Based smart Wake-Up Module - Hypnos

[Eggiman et al. TCAS22]

github.com/pulp-platform/hypnos



Design (post P&R)	
Technology	GF22 UHT
Area	670kGE
Max. Frequency	3 MHz

Implemented with lowest leakage cell library (UHVT)

f_{clk}	32kHz	200kHz
max. sampling rate	150 SPS/Channel	1kSPS/Channel
$P_{SWU, dynamic}$	0.99uW	6.21uW
$P_{SWU, leakage}$	0.7uW	0.7uW
$P_{SPI, dynamic}$	1.28uW	8.00uW
$P_{SWU, total}$ **Measured**	**2.97uW**	**14.9uW**

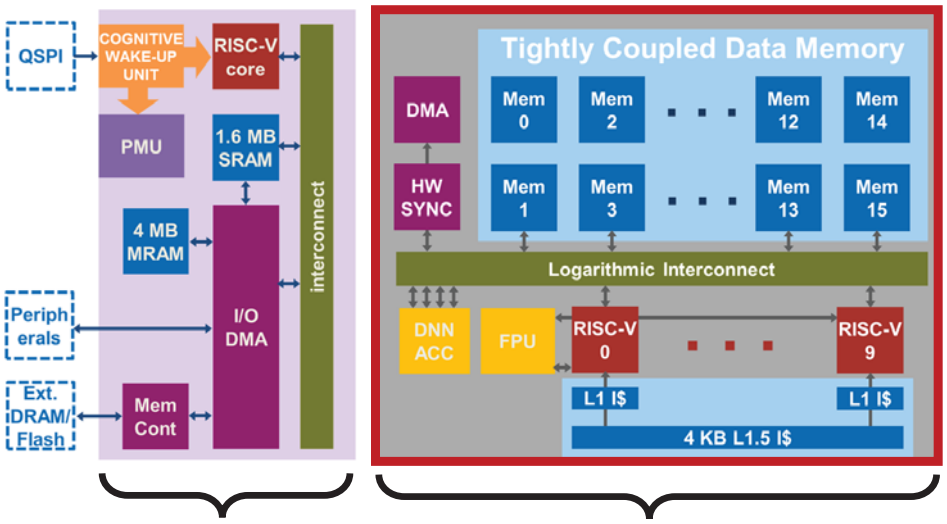
ETH zürich

30

All together in VEGA: Extreme Edge IoT Processor

[Rossi et al. ISSCC21]

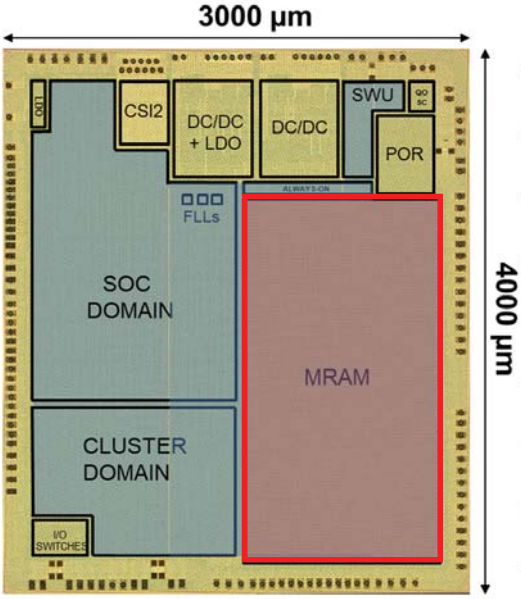
- RISC-V cluster (8cores +1) 614GOPS/W @ 7.6GOPS (8bit DNNs), 79GFLOPS/W @ 1GFLOP (32bit FP appl)
- Multi-precision HWCE(4b/8b/16b) 3×3×3 MACs with normalization / activation: 32.2GOPS and 1.3TOPS/W (8bit)
- 1.7 µW cognitive unit for autonomous wake-up from retentive sleep mode



In cooperation with GREENWAVES TECHNOLOGIES

All together in VEGA: Extreme Edge IoT Processor

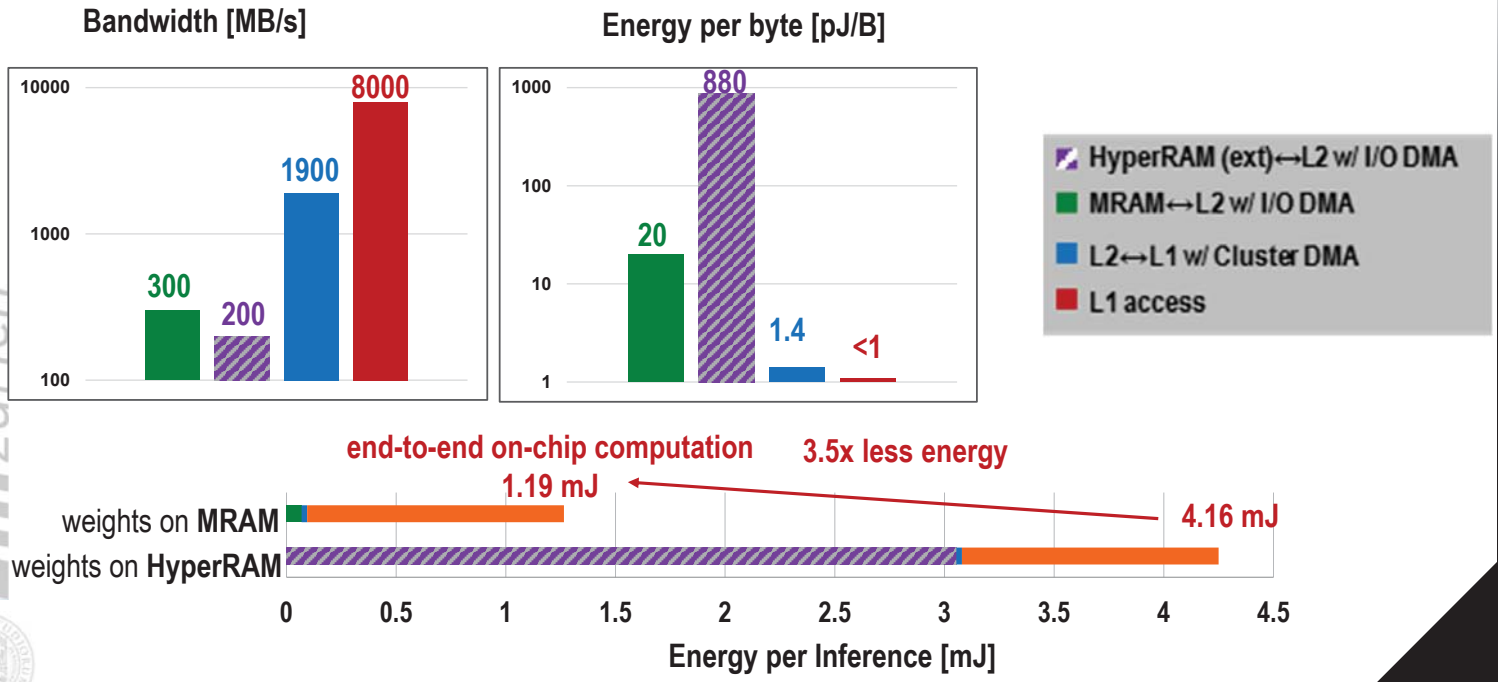
- RISC-V cluster (8cores +1) 614GOPS/W @ 7.6GOPS (8bit DNNs), 79GFLOPS/W @ 1GFLOP (32bit FP appl)
- Multi-precision HWCE(4b/8b/16b) 3×3×3 MACs with normalization / activation: 32.2GOPS and 1.3TOPS/W (8bit)
- 1.7 µW cognitive unit for autonomous wake-up from retentive sleep mode
- **Fully-on chip DNN inference with 4MB MRAM (high-density NVM with good scaling)**



Technology	22nm FDSOI
Chip Area	12mm^2
SRAM	1.7 MB
MRAM	4 MB
VDD range	0.5V - 0.8V
VBB range	0V - 1.1V
Fr. Range	32 kHz - 450 MHz
Pow. Range	1.7 µW - 49.4 mW

In cooperation with GREENWAVES TECHNOLOGIES ➡ **GAP9**

Full DNN Energy (MobileNetV2)



Multi-tasking autonomous nano-drone with GAP9

Autonomous exploration

Object Recognition

Obstacle Avoidance

Rescue reconnaissance

Safe indoor human-drone interation

ETH zürich

34

What's next? Heterogeneous Accelerators

The *Kraken*: TCNs and SNNs at The Extreme Edge

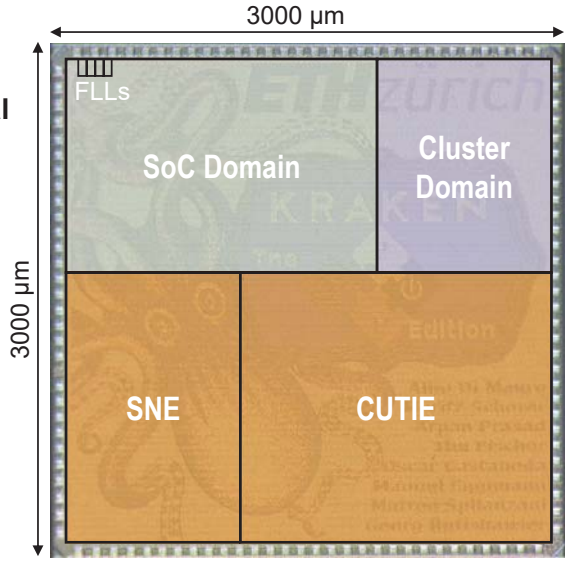
- RISC-V Cluster (8 Cores + 1)
- **CUTIE – dense ternary neural network accelerator**
- **SNE – energy-proportional spiking neural network accelerator**
- DVS Interface for hardware support of event-based vision
- PULPO – Floating point linear algebra accelerator

Technology	22 nm FDSOI
Chip Area	9 mm^2
SRAM SoC	1 MB
SRAM Cluster	128 KB
VDD range	0.55 V - **0.8 V**

What's next? Heterogeneous Accelerators

The *Kraken*: TCNs and SNNs at The Extreme Edge

- RISC-V Cluster (8 Cores + 1)
- **CUTIE – dense ternary neural network accelerator**
- **SNE – energy-proportional spiking neural network accelerator**
- DVS Interface for hardware support of event-based vision
- PULPO – Floating point linear algebra accelerator

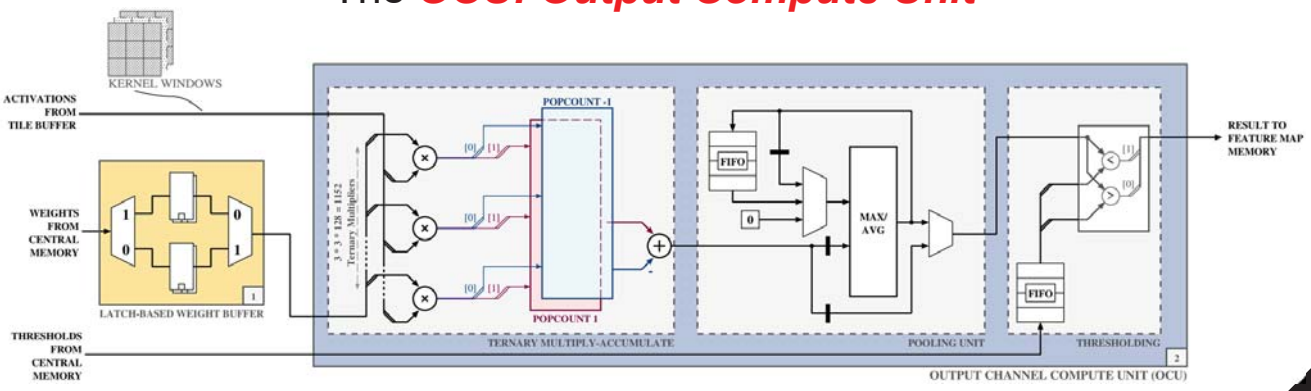


Technology	22 nm FDSOI
Chip Area	9 mm^2
SRAM SoC	1 MB
SRAM Cluster	128 KB
VDD range	0.55 V - **0.8 V**

CUTIE: Minimize Switching Activity & Data Movement

- KxK window on all input channels unrolled, cycle-by-cycle sliding
- All weights for an output channel are held stationary in local buffer (latch-based)
- Completely unrolled inner products vs. systolic MAC →one output activation per cycle!
- Zeros in weights and activations reduce switching activity
- One output per cycle → spatial smoothness of activations helps reducing switching

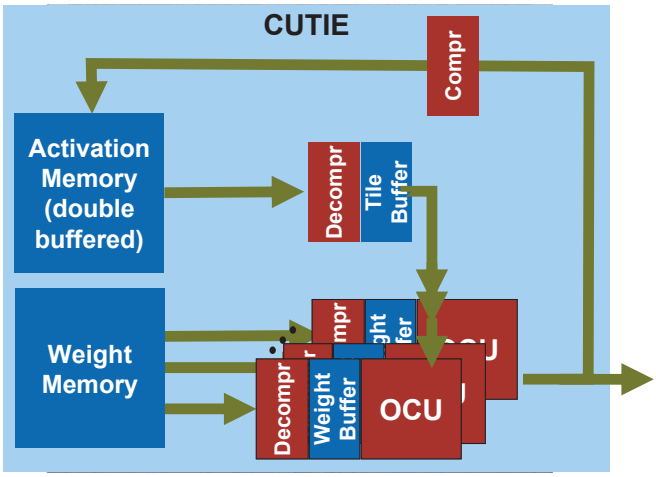
The ***OCU: Output Compute Unit***



Kraken`s CUTIE Implementation

- **Data in 1.6b/Tvalue with Comp/Decomp on the fly**
- **Highly parametrizable accelerator**
  - Channels, kernel shapes, pipeline depth, memory sizes, …
- **Configuration in Kraken**
  - 96 channels (OCUs)
  - 3x3 kernels
  - 64 x 64 pixels feature maps (158 KB)
  - 9 layers of weights (117 KB)
- **Lots of TMAC/cycle**
  - 96 OCUs, 96 Input channels, 3x3 kernels:

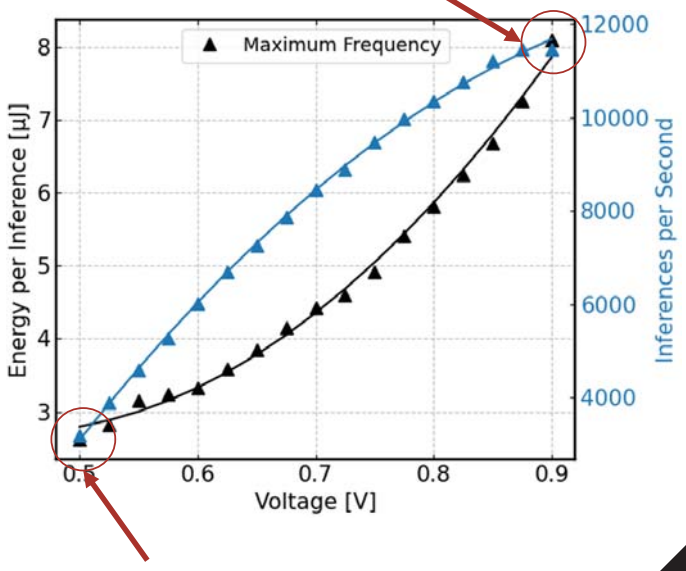
    96 * 96 * 3 * 3 = **82'944 TMAC/cycle**



Results – The numbers

- **Key implementation results**
  - Area: ~3 mm^2
  - Voltage: 0.5 – 0.9 V
- **Inference on CIFAR-10 - Ternary**
  - Accuracy: 86%
  - Energy per inference: 2.72µJ

8.1 µJ per Inference @ 0.85 V



2.72 µJ per Inference @ 0.5V

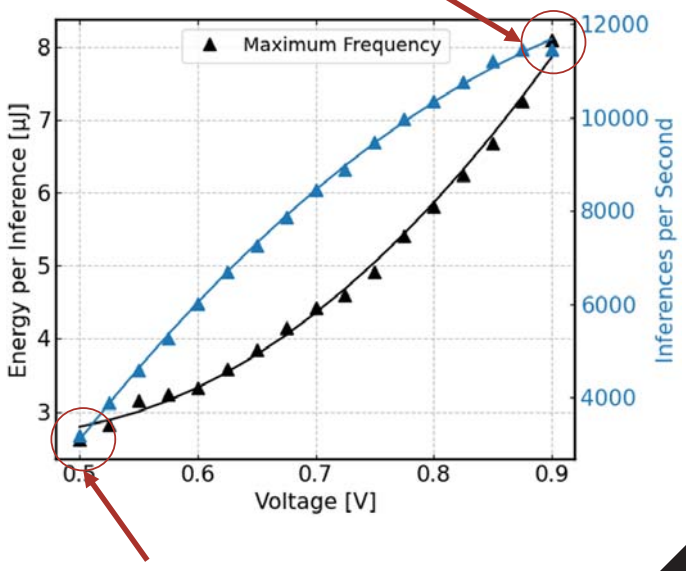
ETH zürich

39

Results – The numbers

- **Key implementation results**
  - Area: ~3 mm^2
  - Voltage: 0.5 – 0.9 V
- **Inference on CIFAR-10 - Ternary**
  - Accuracy: **86%**
  - Energy per inference: **2.72µJ**
- **Inference on CIFAR-10 - Binary**
  - Accuracy: 82%
  - Energy per inference: 4µJ!

8.1 µJ per Inference @ 0.85 V

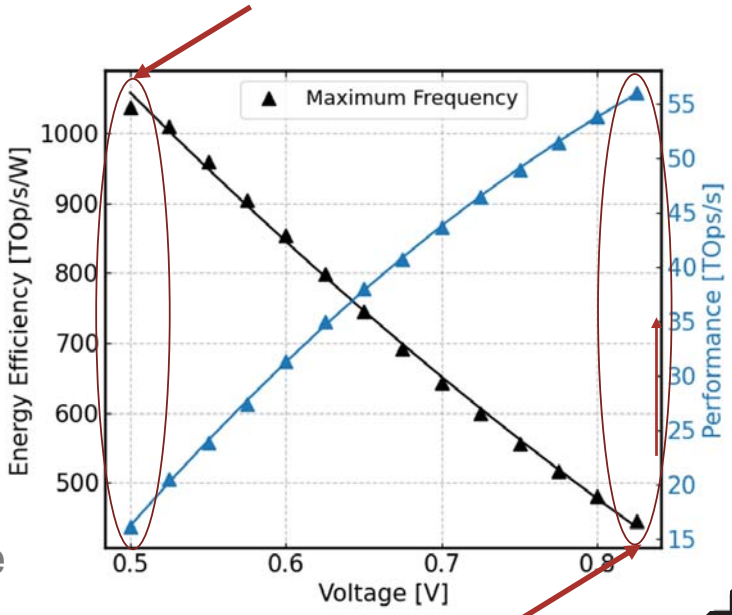


2.72 µJ per Inference @ 0.5V

Results – The numbers

- **Key implementation results**
  - Area: ~3 mm^2
  - Voltage: 0.5 – 0.9 V
- **Inference on CIFAR-10 - Ternary**
  - Accuracy: **86%**
  - Energy per inference: **2.72µJ**
- **Inference on CIFAR-10 - Binary**
  - Accuracy: 82%
  - Energy per inference: 4µJ!
- **Achievable Efficiency and Performance**
  - Peak Core Energy Efficiency: **1036 Top/s/W**
  - Peak Throughput: **55 TOp/s**

1036TOp/s/W, 15 TOp/s, 0.5 V



450 TOp/s/W, 55 TOp/s, 0.85 V

HW acceleration in perspective

Using 22FDX tech, NT@0.6V, High utilization, minimal IO & overhead

Energy-Efficient RV Core → **20pJ (8bit)**

ISA-based 10-20x →**1-5pJ (8bit)** → **XPULPV2 &V3**

Configurable DP 10-20x → **20-100fJ (4bit)** → **HWCE, RBE, NE**

Fully specialized DP 10-20x →**1-5fJ (ternary)** → **XNE, CUTIE***

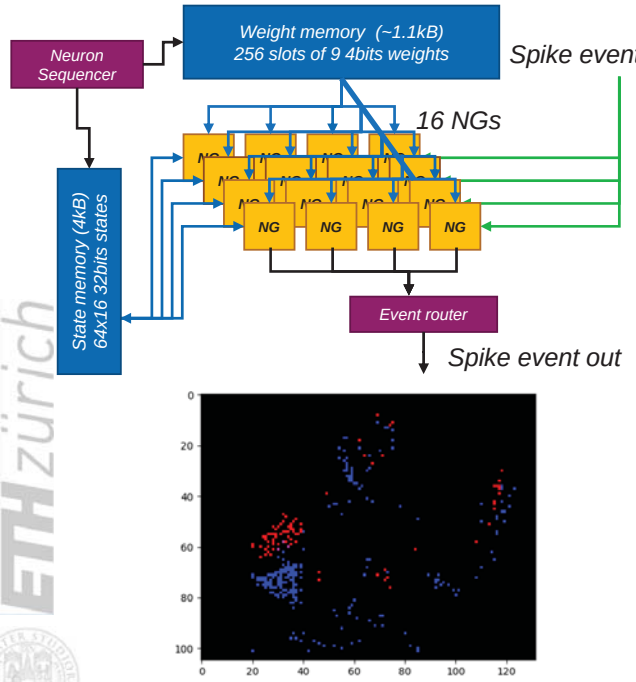
*sub 1fJ in 7nm

ETH zürich

42

What About Neuromorphic? Spiking NN Acceleration

SNE Engine (85 SOP/cycle)



- Engine: 16 Adaptive-LIF neuron data paths (NG). A NG executes one Synaptic Operation (SOP) per cycle
  - 1 SOP = 1 4b-ADD + 2 8b-MUL + 1 8b-ADD + 1 8b-CMP
  - Implements weight accumulation on the neuron membrane potential + leaky decay since the last membrane update + dynamic threshold adaptation + spike generation + membrane potential reset
- For fully connected layers one NG is time-shared for 64 virtual neurons
- Optimized buffering and neuron state update for 64x16 neurons in just 12 cycles for a 3x3 event receptive field
  - Equivalent number of 85 SOP/cycle per engine (682 SOP/cycle on 8 engines)

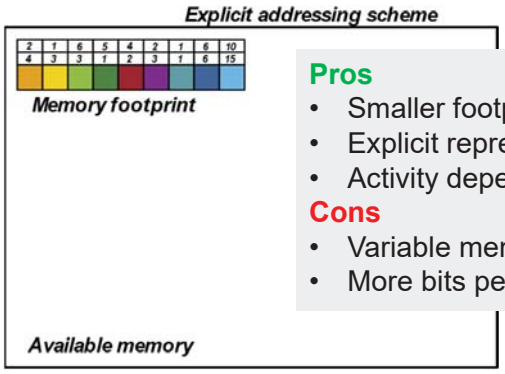
Useful? YES, when coupled with DVS (event-based) sensors

[Di Mauro et al. DATE22]

DVSI: Event Representation and Conversion

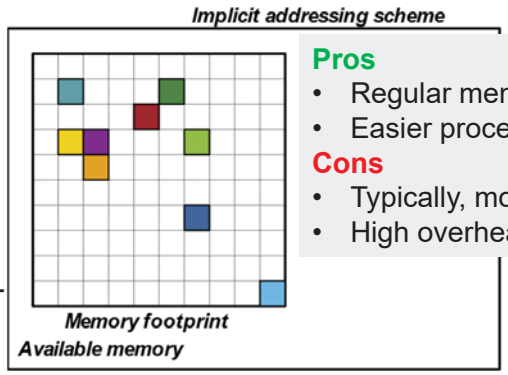
- The DVSI internal buffer uses an explicit coordinate list representation (COO)
- The sensor "reading" can be triggered at any rate (the more we wait, the more events we expect to read/buffer)

- The DVSI can output:
  - Linear streams of events COO-encoded (directly consumable by SNE)
  - Event-frames (suitable for more traditional frame-based processing, e.g. CUTIE)



Pros
- Smaller footprint
- Explicit representation
- Activity dependent

Cons
- Variable memory footprint
- More bits per evt

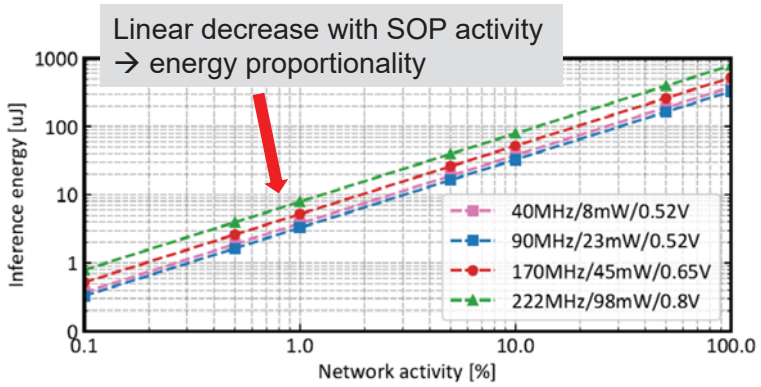


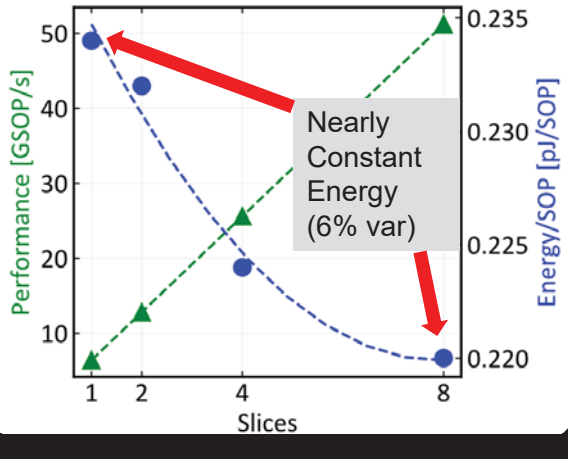
Pros
- Regular memory footprint
- Easier processing

Cons
- Typically, more bits per frame
- High overhead at low activity

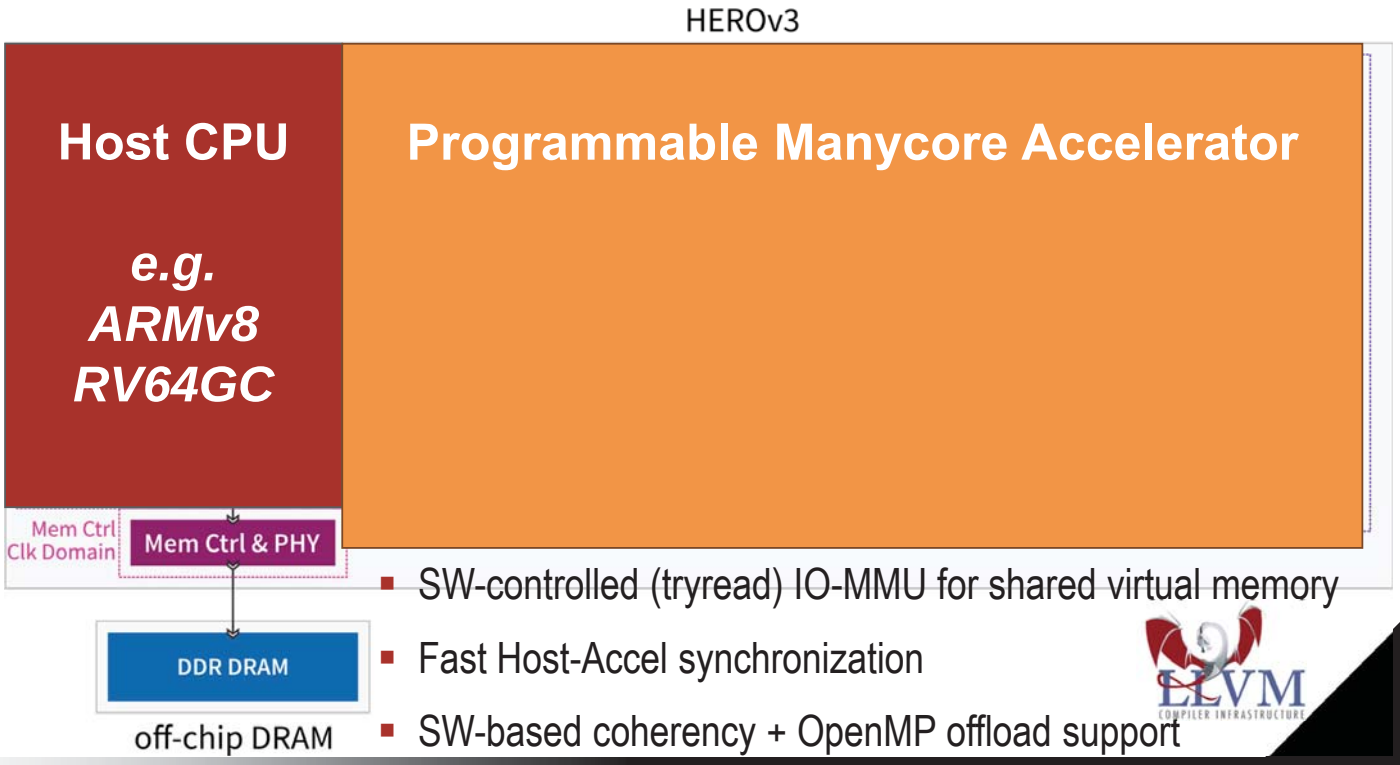
SNE Results

- Inference energy is indeed proportional to event activity → Energy-Proportional operation achieved
- Scalable architecture
  - From 1 to 8 slices linear performance increase
  - Energy/SOP grows by only 6%
- Leads the SoA by 1.7x in Energy/SOP
- Can work concurrently with CUTIE for SNN + TNN "fused" inference (never done so far)



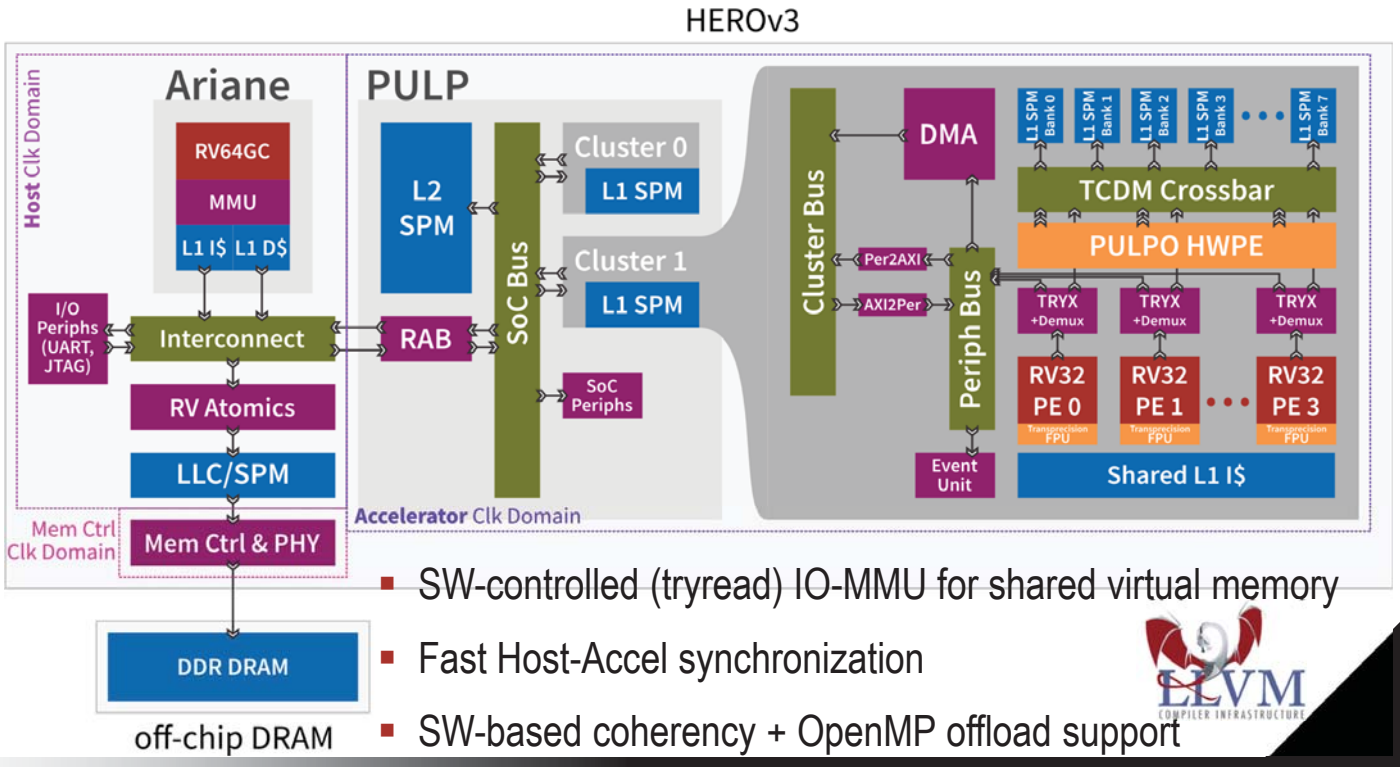


What's next: Scale up with HERO



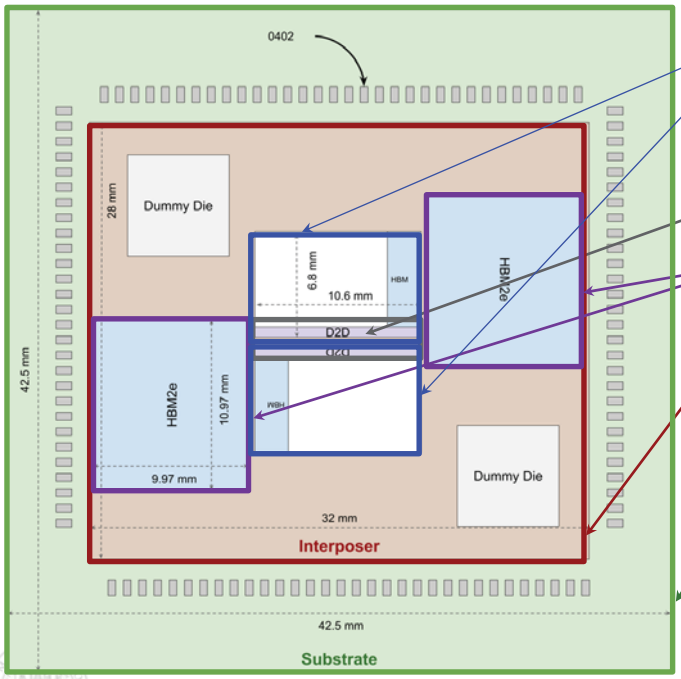
- SW-controlled (tryread) IO-MMU for shared virtual memory
- Fast Host-Accel synchronization
- SW-based coherency + OpenMP offload support

What's next: Scale up with HERO



- SW-controlled (tryread) IO-MMU for shared virtual memory
- Fast Host-Accel synchronization
- SW-based coherency + OpenMP offload support

Leveraging Chiplet Technology: Occamy



Dual-chiplet 12LP+
- Area: **~70mm^2**

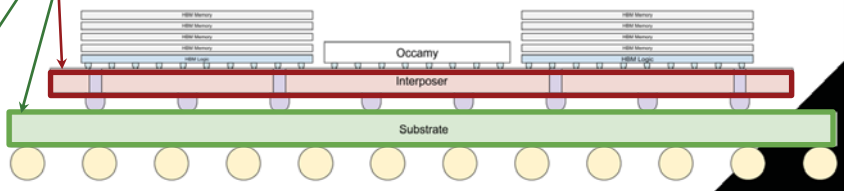
Chip2chip link

HBM2e memories

Interposer
- 65nm tech, passive

Substrate

GlobalFoundries
SYNOPSYS
Rambus



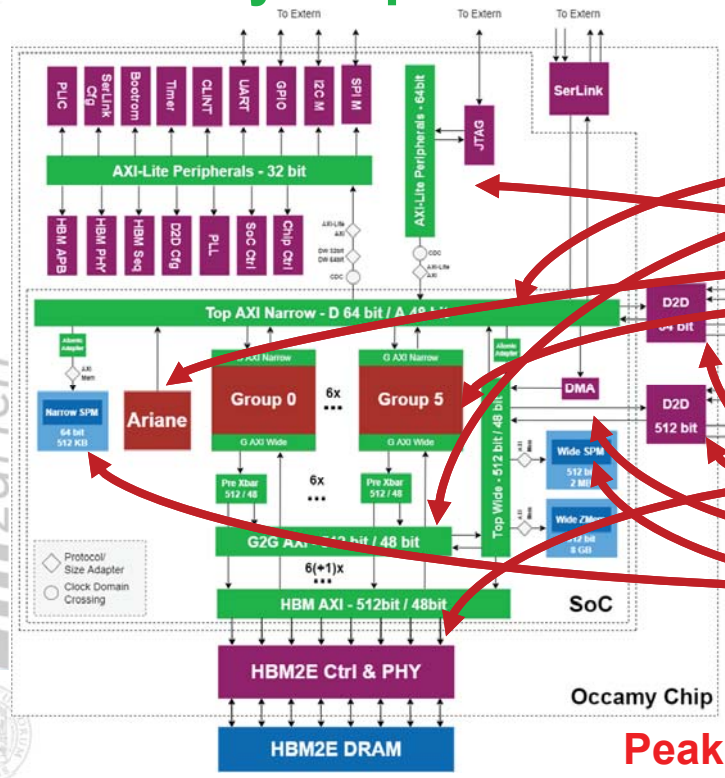
48

Occamy Chiplet

- GF12, target **1GHz** (typ)
- 2 AXI Interconnect Subsystems (multi-hierarchy)
  - 64-bit
  - 512-bit with "interleaved" mode
- Peripherals
- Linux-capable manager core CVA6
- 6 Quadrants: 216 cores/chiplet
  - 4 cluster / quadrant:
    - 8 compute +1 DMA core / cluster
    - 1 multi-format FPU / core (FP64,x2 32, x4 16/alt, x8 8/alt)
- 8-channel HBM2e (8GB) **512GB/s**
- D2D link (Wide, Narrow) **70+2GB/s**
- System-level DMA
- SPM (2MB wide, 512KB narrow)

Peak 384 GDPflop/s per chiplet!

Occamy Chiplet



- GF12, target 1GHz (typ)
- 2 AXI Interconnect Subsystems (multi-hierarchy)
  - 64-bit
  - 512-bit with "interleaved" mode
- Peripherals
- Linux-capable manager core CVA6
- 6 Quadrants: 216 cores/chiplet
  - 4 cluster / quadrant:
    - 8 compute +1 DMA core / cluster
    - 1 multi-format FPU / core (FP64,x2 32, x4 16/alt, x8 8/alt)
- 8-channel HBM2e (8GB) 512GB/s
- D2D link (Wide, Narrow) 70+2GB/s
- System-level DMA
- SPM (2MB wide, 512KB narrow)

Peak 384 GDPflop/s per chiplet!

ETH zürich

50

Closing thoughts – Open Platform for TinyML

PULP is an Open Platform

- For science … fundamental "research infrastructure"
  - Reduce "getting up to speed" overhead for partners
  - Enables fair and well controlled benchmarking
- For Business … it is truly disruptive
  - Reduces NRE + faster innovation path for startups (e.g. Greenwaves tech.), new business models (eg. OpenHWGroup), helps ollaboration with foundries (e.g. GF)

Heterogeneous & Flexible

- 1-3+ orders of magnitude improvement (wrt to efficient RV) by acceleration
  - ISA → Configurable → Fully customized + heterogeneous architectural combinations
- Focus on IO energy (memory, sensor) to achieve sub pJ/OP @ full platform
  - 3D-IC technology is a key enabler

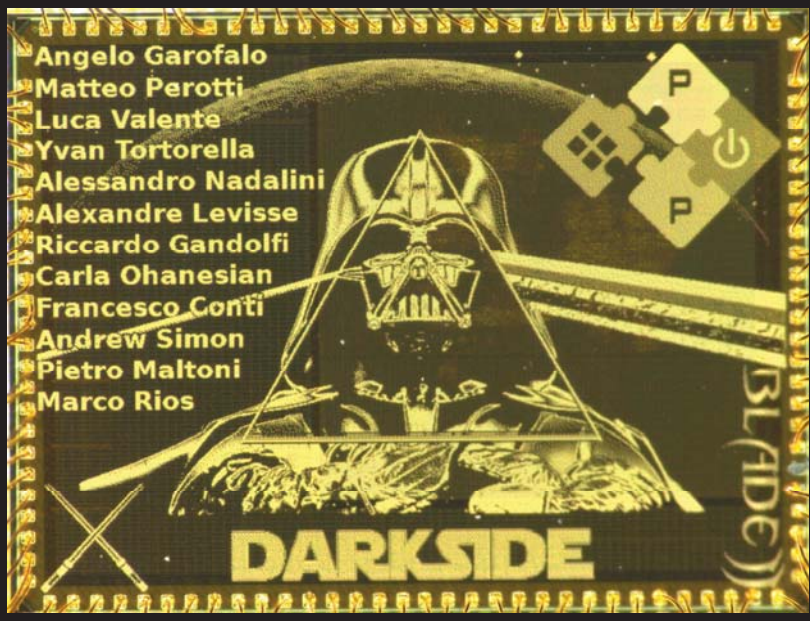
ETH zürich

51

PULP

Parallel Ultra Low Power

Luca Benini, Alessandro Capotondi, Alessandro Ottaviano, Alessio Burrello, Alfio Di Mauro, Andrea Borghesi, Andrea Cossettini, Andreas Kurth, Angelo Garofalo, Antonio Pullini, Arpan Prasad, Bjoern Forsberg, Corrado Bonfanti, Cristian Cioflan, Daniele Palossi, Davide Rossi, Fabio Montagna, Florian Glaser, Florian Zaruba, Francesco Conti, Georg Rutishauser, Germain Haugou, Gianna Paulin, Giuseppe Tagliavini, Hanna Müller, Luca Bertaccini, Luca Valente, Manuel Eggimann, Manuele Rusci, Marco Guermandi, Matheus Cavalcante, Matteo Perotti, Matteo Spallanzani, Michael Rogenmoser, Moritz Scherer, Moritz Schneider, Nazareno Bruschi, Nils Wistoff, Pasquale Davide Schiavone, Paul Scheffler, Philipp Mayer, Robert Balas, Samuel Riedel, Segio Mazzola, Sergei Vostrikov, Simone Benatti, Stefan Mach, Thomas Benz, Thorir Ingolfsson, Tim Fischer, Victor Javier Kartsch Morinigo, Vlad Niculescu, Xiaying Wang, Yichao Zhang, Frank K. Gürkaynak, all our past collaborators and many more that we forgot to mention



http://pulp-platform.org @pulp_platform

CPS&IoT'2022 Summer School
Budva, Montenegro, June 7-11, 2022

Green CPS and IoT for Green World

Lech Jóźwiak

L.Jozwiak@tue.nl

Outline

1. Introduction
2. Modern cyber-physical systems (CPS)
3. Importance of modern CPS and IoT
4. Challenges of advanced CPS development
5. Computing technology for advanced CPS
6. Environmental crisis and environmental footprint of CPS and IoT
7. Importance of advanced green CPS and IoT for environmental recovery
8. IoT for advanced green CPS
9. Quality-driven design of advanced green CPS
10. Conclusion

Introduction: Aims of this tutorial

- **The two main aims of this tutorial are the following:**
  - ***to make the participants aware of the necessity of green CPS and IoT***
  - ***to prepare the ground for the whole CPS&IoT'2021 Summer School***
- This means in particular:
  - to introduce several basic definitions related to CPS
  - to explain the necessity of green CPS and IoT
  - to sketch the CPS scene, what includes:
    - introduction to modern CPS and IoT, their importance, their ongoing revolution, and challenges of their development, and
    - explanation of the necessity of their holistic multi-objective quality-driven design
  - to introduce the methodology of quality-driven green system design

Introduction: Further reading for this tutorial

- L. Jóźwiak: Advanced Mobile and Wearable Systems, Microprocessors and Microsystems, Elsevier, Vol. 50, May 2017, pp. 202–221
- L. Jóźwiak: Quality-driven Design in the System-on-a-Chip Era: Why and how?, Journal of Systems Architecture, vol. 47, no. 3-4, Apr. 2001, pp. 201-224
- L. Jóźwiak: Life-inspired Systems and Their Quality-driven Design, Lecture Notes in Computer Science, Vol. 3894, 2006, Springer, pp. 1-16
- Jóźwiak, L.; Lindwer, M.; Corvino, R.; Meloni, P.; Micconi, L.; Madsen, J.; Diken, E.; Gangadharan, D.; Jordans, R.; Pomata, S.; Pop, P.; Tuveri, G.; Raffo, L. and Notarangelo, G.: ASAM: Automatic Architecture Synthesis and Application Mapping, Microprocessors and Microsystems journal, Vol.37, No 8, pp. 1002-1019, 2013
- Jóźwiak, L. and Jan, Y.: Design of Massively Parallel Hardware Multi-Processors for Highly-Demanding Embedded Applications. Microprocessors and Microsystems, Volume 37, Issue 8, November 2013, pp. 1155–1172.
- L. Jóźwiak and S.-A. Ong: Quality-driven Model-based Architecture Synthesis for Real-time Embedded SoCs, Journal of Systems Architecture, Elsevier Science, Amsterdam, The Netherlands, ISSN 1383-7621, Vol. 54, No 3-4, March-April 2008, pp. 349-368
- Many other papers of myself and my former Ph.D. students; many of them referenced in the above papers

Introduction: What is a system?

- A **system** is a ***complex whole composed of interrelated, interdependent and/or interacting items*** (parts or elements of a system) ***that are so intimately connected that they appear and operate as a single unit in relation to the external world*** (to other systems)
- **Three basic types of systems:**
  - ***unorganized system*** **-** a mechanical unsystematic conglomerate of objects
  - ***organized system*** **-** a systematic, relatively stable and law-governed composition of parts which properties cannot be reduced to the simple sum of the properties of its parts, but involve some new emerging properties resulting from complex composition of the parts' properties (e. g. a molecule, crystal, circuit, computer, machine), and
  - ***organic system*** **-** formed not as a composition of some ready-made parts, but being an ***integral whole*** with distinguishable parts that originate, develop and die together with the whole, and cannot preserve and demonstrate their complete quality without the whole (e. g. life organisms); the characteristic features of the organic systems are the self-development and self-reproduction
- In this presentation **organized systems** will be considered

Introduction: System organization and structure

- The **system organization** (composition) appropriately:
  - defines its parts
  - arranges the parts in relation to each other and to the whole, and
  - interconnects them to form the whole
- The term **system structure** designates the ***parts of a system arranged into a proper relation and appropriately interconnected*** according to a certain set of laws and/or rules in order to form a whole
- We will consider material systems
- **Since matter is active** and is in constant change, **the material systems are in constant change**, with only some relative and transient stability conditions
- Compositions of interrelated, interdependent or interacting single changes (transformations, actions) form **processes**
- **Process** is a relatively ***isolated composition of interrelated interdependent or interacting actions*** (transformations, changes)

6

Introduction: System = process © structure

- A given process can only perform (take place, occur) in particular relatively stabile conditions
- These conditions that make the process possible are created and guaranteed by the system **structure**
- The **system structure** is a relatively isolated, stable and slowly changing (in relation to the process) part of the universe in which a particular process (or a collection of co-operating processes) can take place
- A **system** is a ***unity of a process and structure*** in which this process takes place
- **System design** is an activity of ***defining an appropriate composition of the system process and structure***

7

Introduction: What are cyber-physical systems?

- **Cyber** comes from Greek adjective ***kyberneticos*** (***cybernetic***) that means skilled in steering or governing
- Already in ancient times people constructed various systems: the oldest known artificial automatically controlled system is probably a water clock invented by Ktesibios (285–222 BC) in Alexandria
- Form those times, the construction of machines (physical systems) and their controllers (cyber systems) continued and developed through the centuries
- Until the end of 19th century the controllers (cyber systems) were implemented as mechanical, hydraulic and pneumatic systems
- In the 20th century they started to be gradually replaced by the electric controllers, and later by the electronic controllers
- **Physical systems** are systems in which matter or energy acquisition, processing and transfer take place according to the lows of physics
- **Cyber systems** are ***(parts of) control systems***, i. e. information collecting, processing and communicating systems

8

Introduction : What are cyber-physical systems?

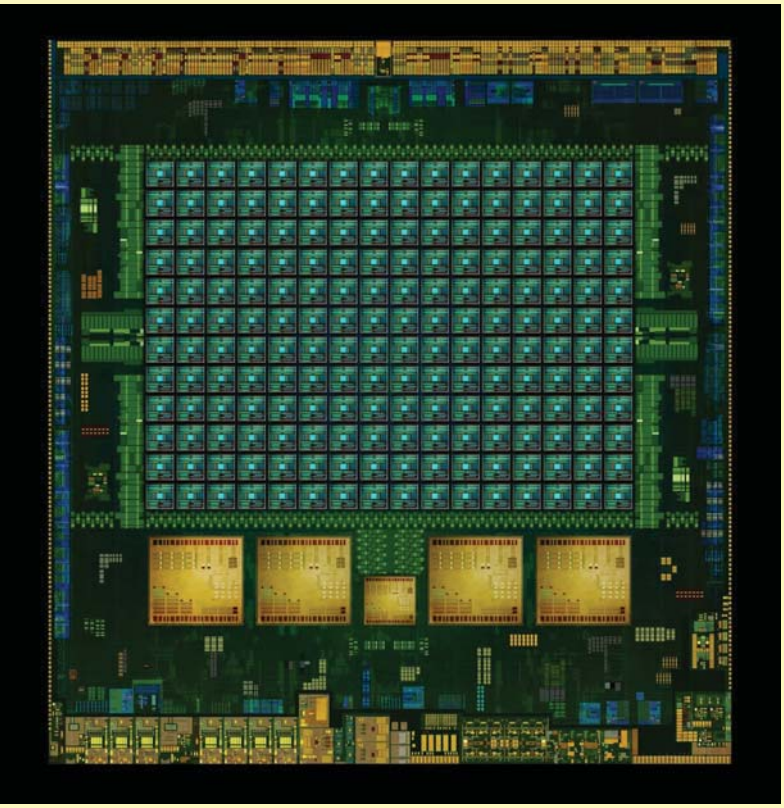
- **Cyber-physical system** (**CPS**) is a compound system engineered through integration of cyber and physical sub-systems or components and/or pre-existing component cyber-physical systems, so that it appears and operates as a single unit in relation to the external world (to other systems)

- Introduction of the transistor and integrated circuit technologies in the years 1950s and 1960s, correspondingly, enabled the ***ongoing microelectronics and information technology revolution*** that is till now progressing according to the Moore's low

- The recent revolutionary progress in computing platforms, communication, networking, sensors and actuators enables:
  - much more effective and efficient CPS for traditional applications, and
  - "smart", sophisticated and affordable CPS for numerous new applications, e.g. smart robots, homes, cars, wearable and implantable medical devices, etc.

9

Introduction: very complex MPSoCs



Source: ANANDTECH
(http://www.anandtech.com/show/7622/nvidia-tegra-k1)

- *Modern nano-dimension semiconductor technology enables implementation of a* ***very complex multiprocessor system on a single chip (MPSoC)***
- **This facilitates a rapid progress in:**
  - ***global networking***
  - ***(mobile) wire-less communication***
  - ***(mobile autonomous) embedded computing***

NVIDIA Tegra K1 massively parallel MPSoC for mobile applications

CPU: (4+1) Cortex-A15 cores

Kepler GPU: 192 CUDA GPU cores

10

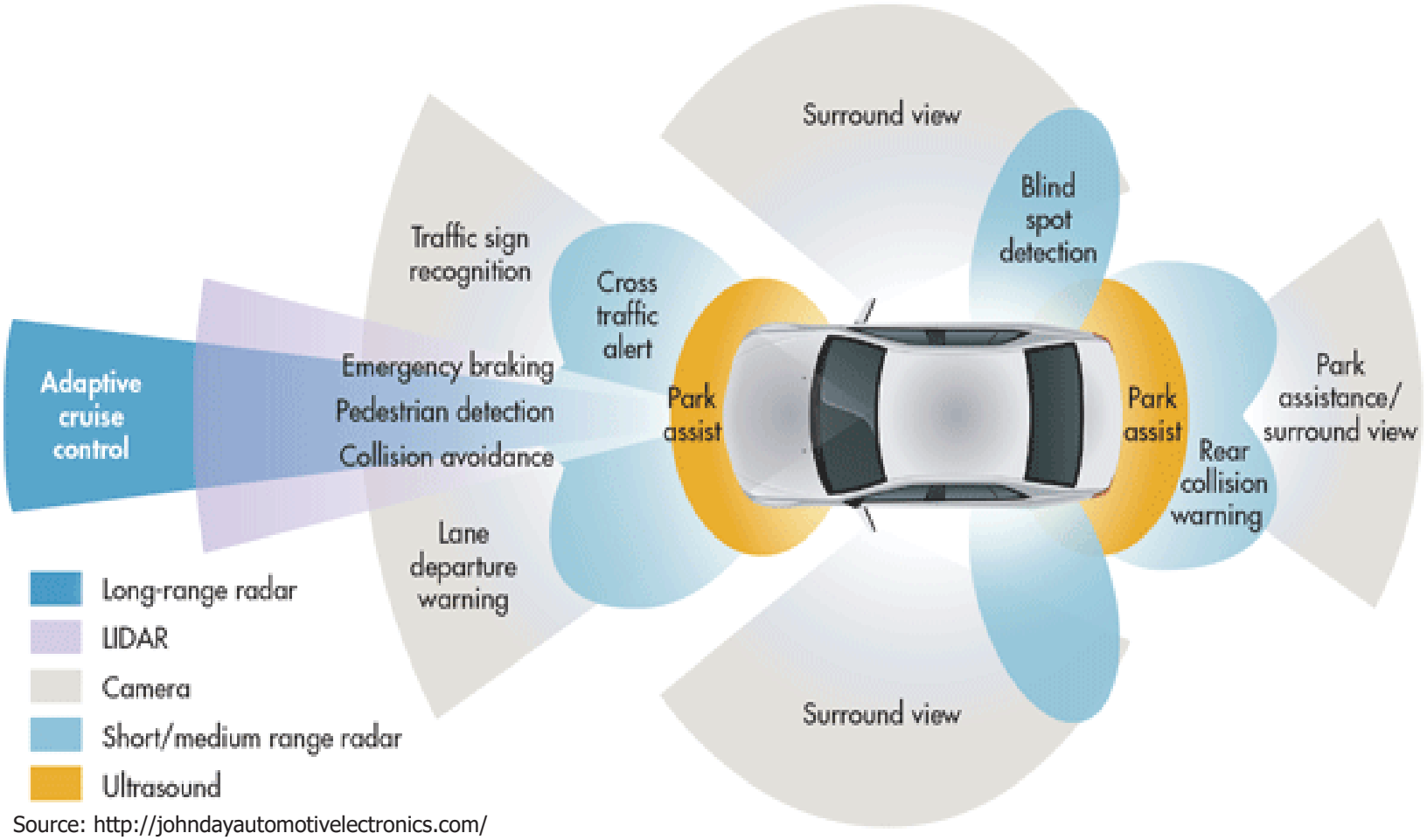
Introduction: cyber-physical technology revolution

- **The recent rapid developments in:**
  - system-on-a-chip technology
  - common global networking
  - wire-less communication
  - mobile and autonomous computing
  - miniaturized sensors and actuators
  - material technology

  created a **large discrepancy between what is possible and what is used nowadays**
- This discrepancy:
  - causes both a **very strong technology push** and **market pull** to create new or modified products and services, and
  - results in the ***cyber-physical technology revolution***
- Recently, a revolutionary transition has been started from the **internet of computers** to the **internet of smart (mobile) cyber-physical systems (CPS)**, called **Internet of Things (IoT)**

11

Examples of modern CPS: autonomously-driving cars



Source: http://johndayautomotivelectronics.com/

Examples of modern CPS: smart wearables

Examples of CPS: wearable virtual and augmented reality

Source: https://www.oculus.com

Source: http://www.technodo.com/

Examples of modern CPS: smart miniaturized implants and pill-size medical devices

modern 10 times smaller pace-makers

A new wave of the information technology revolution has arrived that creates much more coherent and fit to use CPS and connects them to form the IoT

15

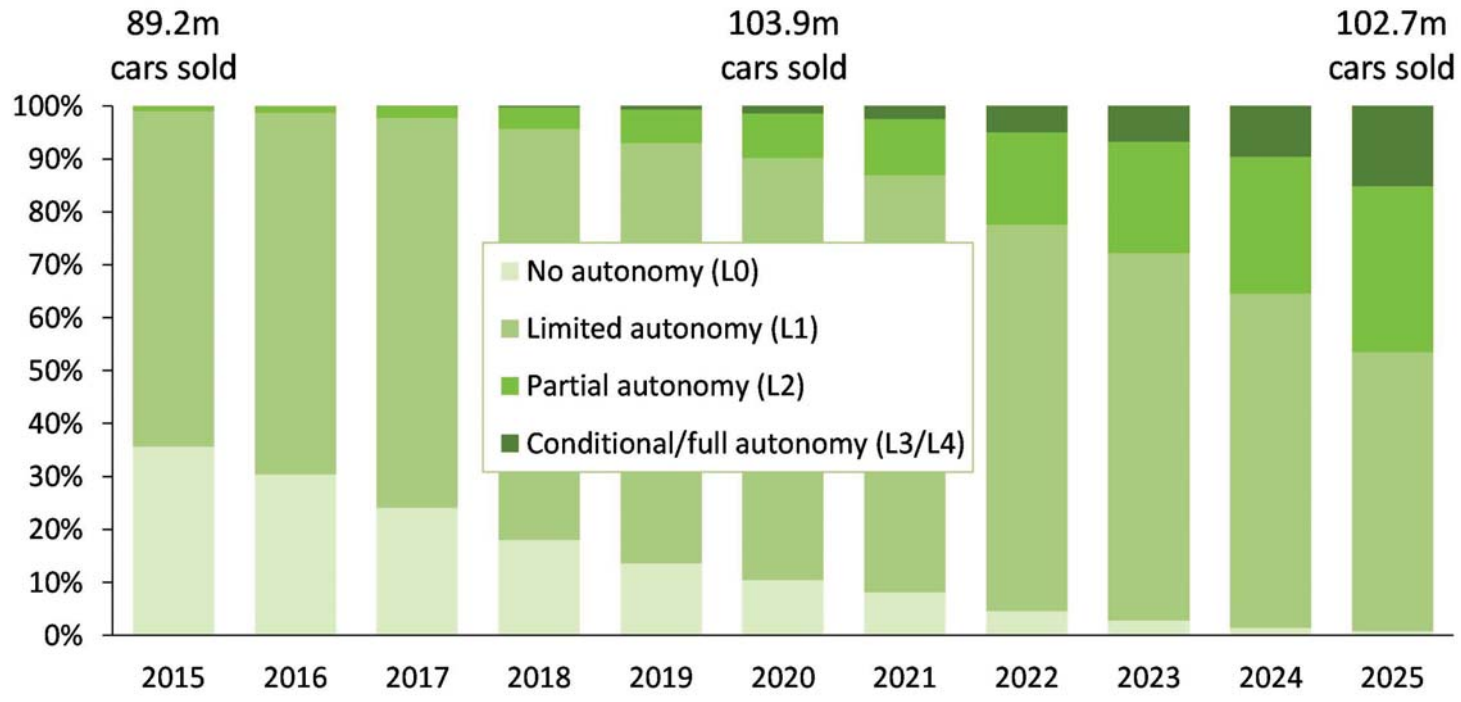
Importance of modern CPS

- **Application areas of mobile CPS** cover ***virtually all socially important application sectors***, including:
  - *consummer applications* , e.g. mobile computing, communication, localization, navigation, gaming, entertainment, fashion, etc.
  - *extension or replacement of human capabilities*, e.g. tele-operation, personal assistance, artificial limbs, implants, etc.
  - *social systems*, e.g. smart health-care and other numerous health-care applications, assisted leaving, law enforcement, public safety, military, etc.
  - *transportation and automotive*, e.g. traffic control, navigation, tracking, communication, mobile fares and personalized customer service, assisted/autonomous driving, etc.
  - *industrial, safety, security and military applications* , e.g. mobile real-time in-the-field surveillance, monitoring, inspection, repair, robotics, instruction, assistance, etc.
  - *commercial applications*, e.g. mobile inventory tracking and customer service, wearable augmented reality and other systems for touristic applications, and many others
- **The economic and societal importance of mobile CPS is very high and rapidly increases**

16

Rapid growth of the modern mobile CPS and IoT markets

Worldwide car sales forecast by level of autonomy



Source: Canalys estimates, Autonomous Vehicle Analysis, December 2016

canalys

17

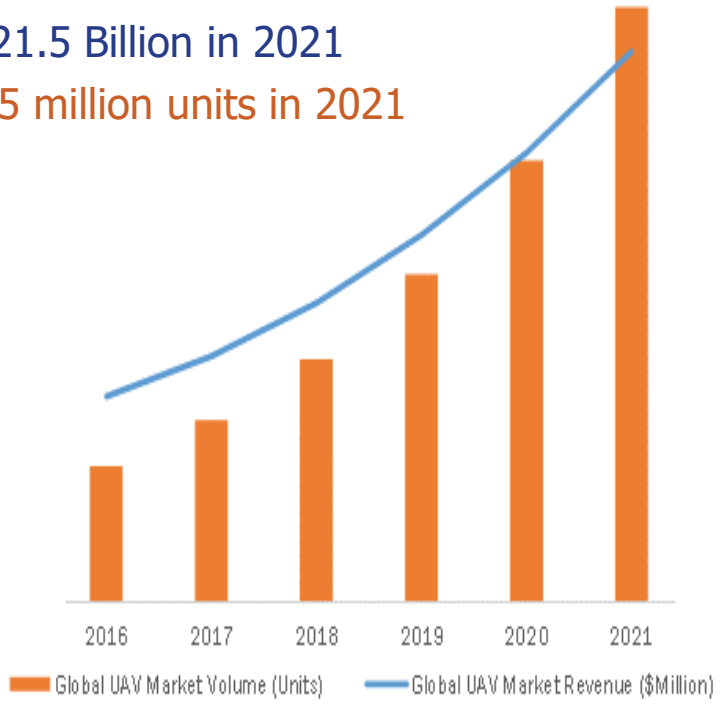
Rapid growth of the modern mobile CPS and IoT markets

Global unmanned aerial vehicle (UAV) market

$8 billion in 2016

$21.5 Billion in 2021

>5 million units in 2021



- **The fastest growing market** of all mobile sectors is this **of smart wearable devices**:
  - $14 billion and 123 million devices in 2016
  - $34 billion and 411 million devices in 2020

  (CCS Insight, February 2016)

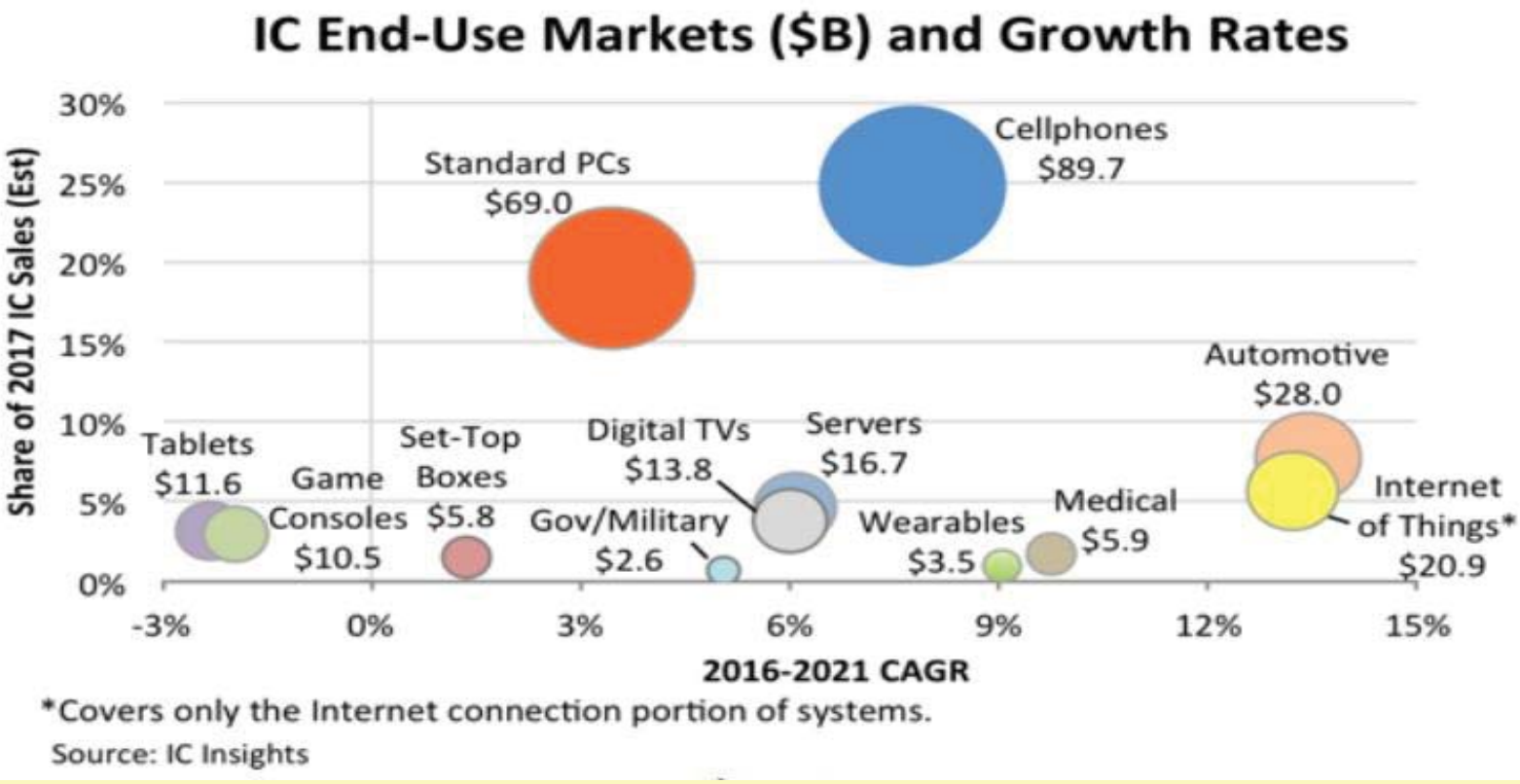
Source: BIS Research, January 2018

18

Rapid growth of the CPS and IoT markets

- The number of connected IoT devices was 11.3 billion in 2020
- IoT Analytics forecasts that there will be 27.1 billion connected IoT devices by 2025, and Cisco 14.6 billion machine-to-machine IoT connections by 2022.
- Allied Market Research finds that the global IoT industry generated sales of $740.5 billion in 2020, and is estimated to reach $4,421.6 billion by 2030
- This corresponds to the growth rate at a CAGR of 19.6% between 2021 and 2030
- In 2020 the strongest contributor to the global IoT market was the industrial manufacturing segment, accounting for more than 25% of the market, and it is expected to maintain its leadership during the forecast period
- However, the healthcare segment is expected to grow at the highest CAGR of 26.2% between 2021 and 2030
- The ESD Alliance reported that the electronic (embedded) system design (ESD) industry had the total market revenue of $13.2 billion in 2021, with the revenue grow rate of 15.8% comparing to 2020, with the fastest growing computer-aided engineering, printed circuit board and multi-chip module, semiconductor intellectual property, and services

Rapid growth of the **chip market** for CPS and IoT

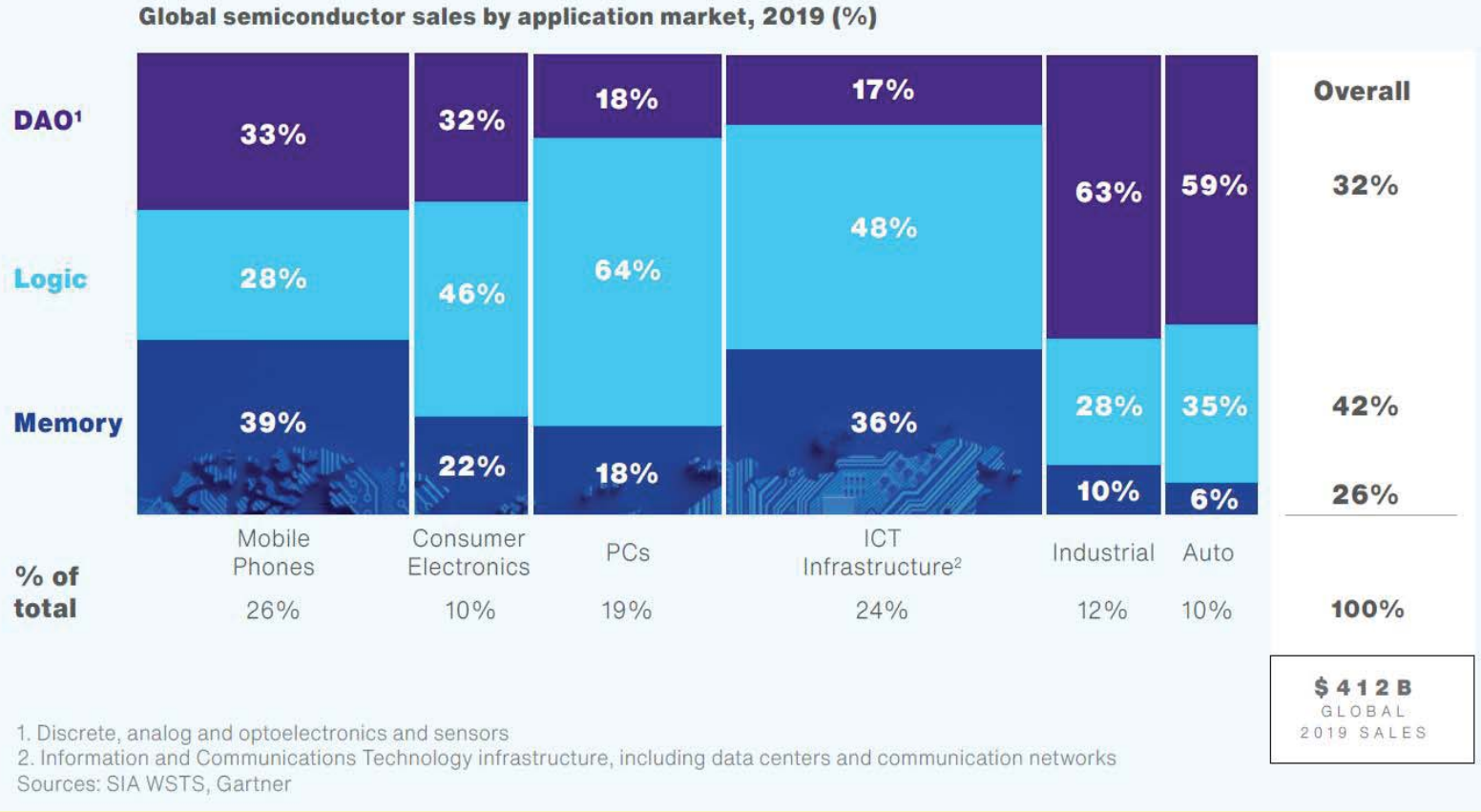


Source: IC Insights

- The fastest-growing chip markets are automotive, IoT, medical and wearables

20

Semiconductor market related to CPS and IoT in 2019



Source: SIA WSTS and Gartner

- PCs account for only 19%, while a large majority of the rest is related to CPS and IoT

21

Semiconductor market related to CPS/IoT in 2021/2022

- According to Semiconductor Industry Association (SIA), the global semiconductor industry sales in 2021 increased by 26.2% compared to the 2020 to the highest-ever annual value of $556 billion
- A record number of 1.15 trillion semiconductor units were shiped in 2021
- Sales in categories related to CPS and IoT were as follows:
  - micro-ICs (including microprocessors) increased 15.1% to $80.2 billion
  - logic increased by 30.8% to $154.8 billion
  - memories incresed 30.9% to $153.8 billion
  - analog semiconductors (commonly used in vehicles, consumer goods, and computers) increased 33.1% to $74 billion
  - automotive ICs increased 34.3% to a record high of $26.4 billion
- According to World Semiconductor Trade Statistics (WSTS), in 2022 the global semiconductor market is expected to increase by 10.4%, which corresponds to annual sales of US$ 613.5 billion
- The growth will be mainly driven by sensors 17.2%, logic with 17.1% and analog with 14.1%

22

Challenges: unusual complexity and ultra-high demands

- The huge and rapidly developing markets of sophisticated mobile CPS represent **great opportunities**
- These opportunities come with a price of:
  - **unusual system complexity** and **heterogeneity**, resulting from ***convergence and combination of various applications and technologies*** in one system or even on one chip, and
  - **stringent and difficult to satisfy requirements** of modern applications
- **Smart cars, drones and various wearable systems**:
  - involve **big instant data** from multiple complex sensors (e.g. camera, radar, lidar, ultrasonic, sensor network tissues, etc.) and from other systems, used for mobile vision, imaging, virtual or augmented reality, etc.
  - are required to provide **continuous autonomous service in a long time**
  - are **safety-critical**
- In consequence, they demand a **guaranteed (ultra-)high performance** and/or **(ultra-)low energy consumption**, while requiring a **high reliability, safety and security**

23

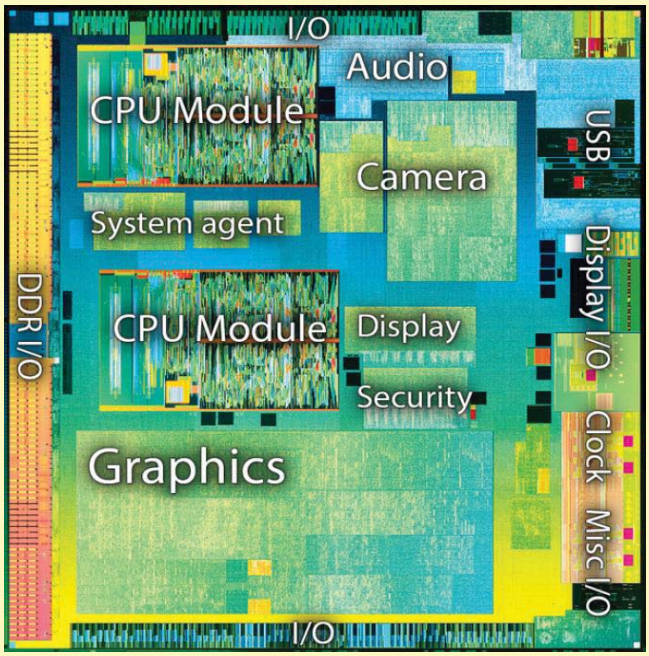
Challenges: application parallelism and heterogeneity

- The modern complex applications that require ultra-high performance and/or ultra-low energy consumption:
  - are from their very nature **heterogeneous**
  - include numerous different algorithms involving **various kinds of massive parallelism**: data parallelism, and task-level, instruction-level and operation-level functional parallelism
- To adequately serve these applications:
  - **heterogeneous computation platforms** have to be exploited
  - processing engines with **parallel multi-processor macro-architectures** and **parallel processor micro-architectures** have to be constructed
  - different parts of complex applications involving different kinds of parallelism have to be implemented with corresponding different application-part specific parallel hardware
  - multiple different or identical processors, each operating on a (partly) different data sub-set, have to work concurrently to realize the ultra-high throughput and ultra-low energy consumption

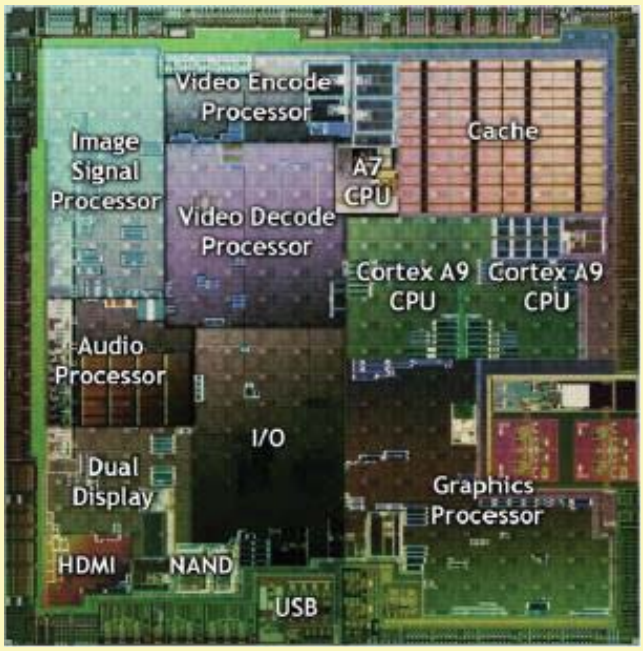
24

Challenges: application complexity, parallelism and heterogeneity

To implement the highly-demanding complex heterogeneous CPS applications ***complex heterogeneous MPSoCs*** *are needed*



Intel Atom Z3770*

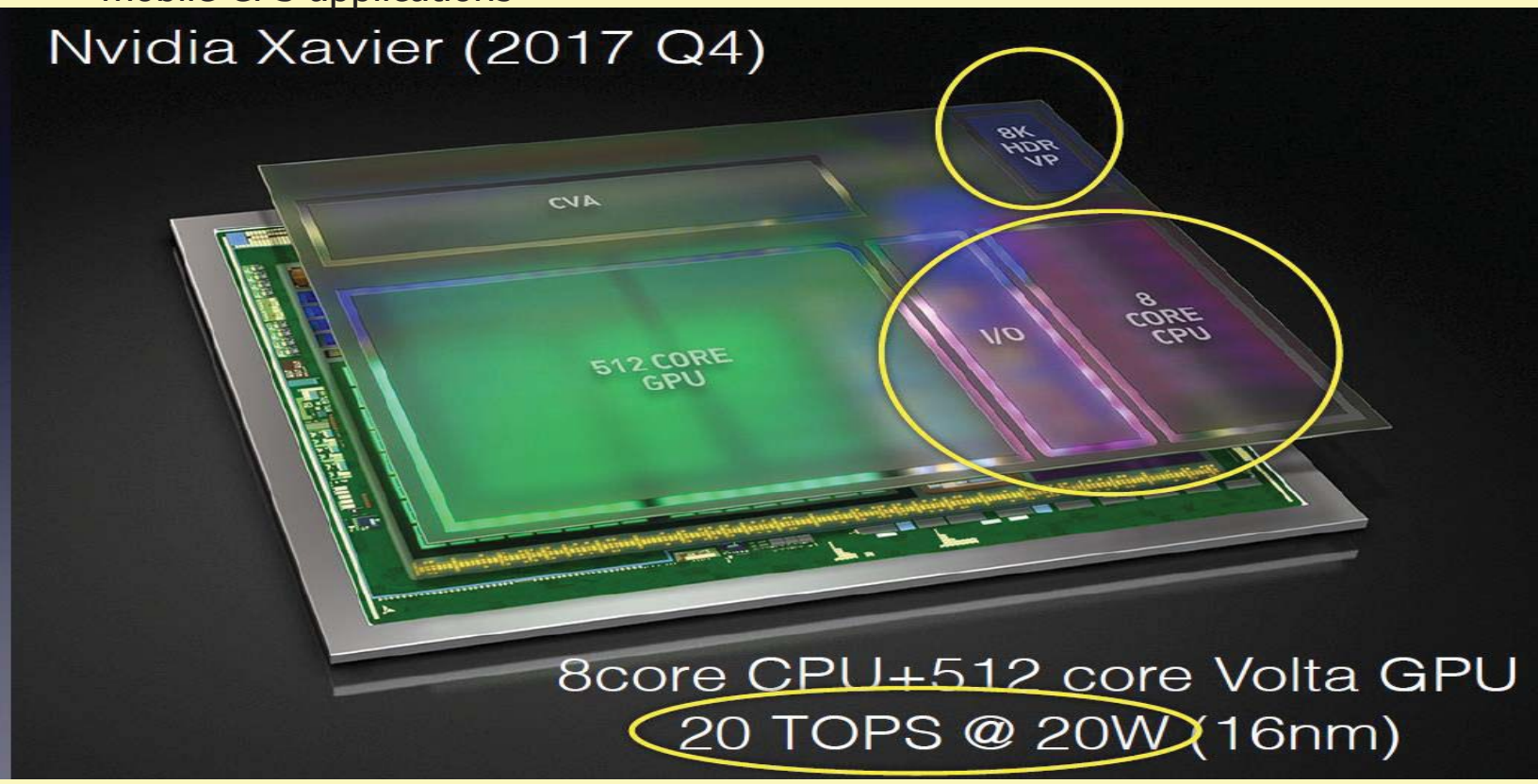


Nvidia Tegra 2+

*Source: http://tweakers.net/reviews/3162/2/intels-atom-bay-trail-de-eerstenieuwe-atom-in-vijf-jaar-zes-verschillende-bay-trails.html
+Source: http://www.anandtech.com/show/4144/lg-optimus-2x-nvidia-tegra-2-reviewthe-first-dual-core-smartphone/3

Challenges: application complexity, parallelism and heterogeneity

NVIDIA's advanced massively parallel heterogeneous MPSoC for ADAS and similar mobile CPS applications



Source: Albert Y.C. Chen, Viscovery

Challenges: criticality of applications

- Cyber-physical systems influence our life to a higher and higher degree
- Therefore, the society expectations regarding them grow rapidly
- Due to CPS common usage in various kinds of technical, social and biological applications, and their growing influence, **we and the life on the Earth more and more depend and rely on these systems**:
  - their ***quality*** is becoming ***more and more critical***
  - many ***applications considered previously as non-critical are becoming critical***
- Due to the rapidly growing share of the highly demanding embedded and CPS applications, ***higher demands are becoming much more common***
- Due to the multiple reasons just discussed, and specifically, due to the rapidly growing system and silicon complexity and diversity, it will be ***more and more difficult to guarantee the systems' quality***
- This is a **new difficult situation** that cannot be adequately addressed without an **adequate design methodology** and **electronic design automation**

27

Quality-driven Model-based Design

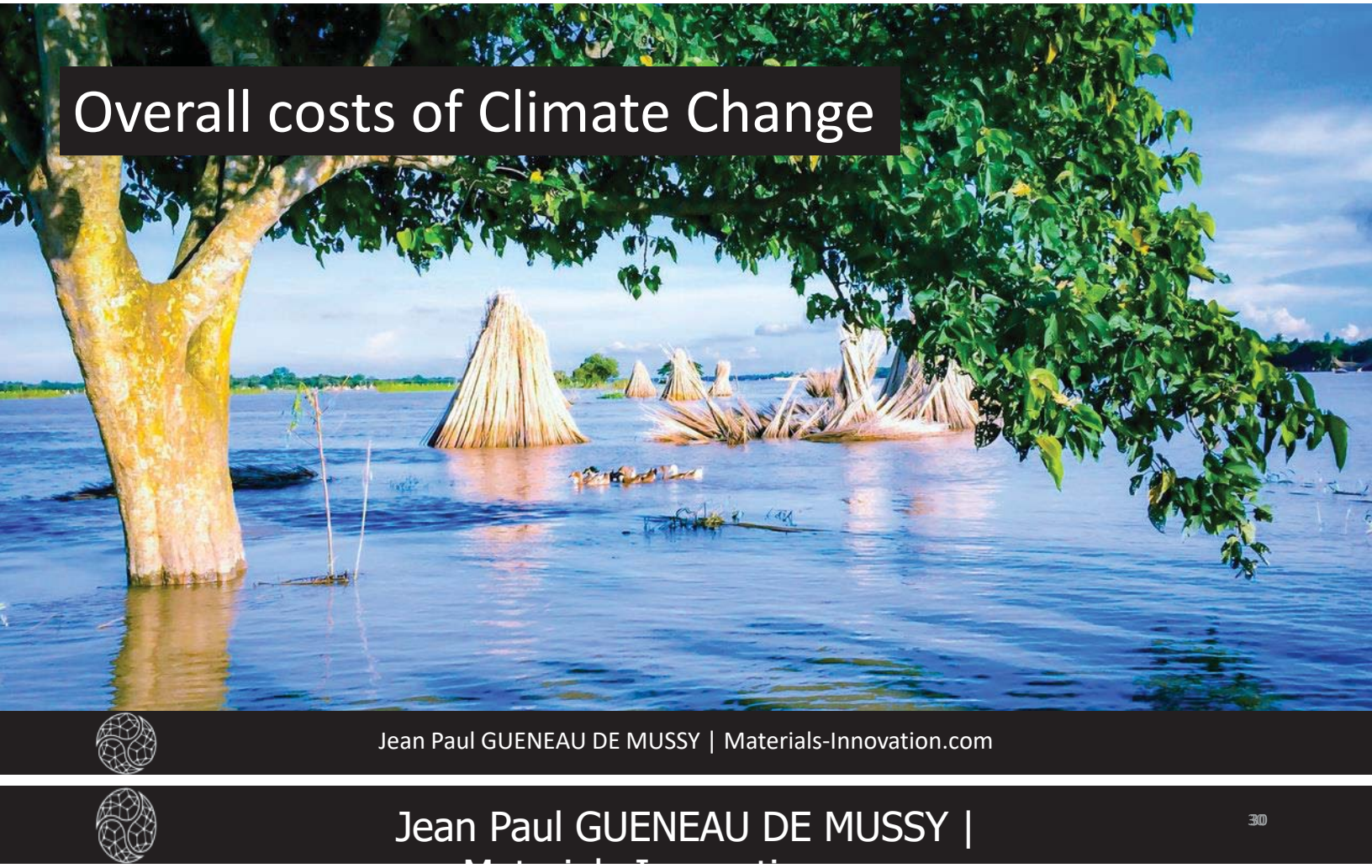
- When considering a **system and design methodology adaptation** to the situation in the field of modern CPS, we have first to ask: *what general system approach and design approach seem to be adequate to solve the listed problems and overcome the challenges*?
- **Predicting the current situation,** more than 20 years ago I proposed such **system paradigm** and **design paradigm**, i.e. the paradigms of:
  - **life-inspired systems** and **quality-driven design**, and
  - the **methodology of quality-driven model-based system design** based on them
- From that time my research team and our industrial and academic collaborators were researching the **application of this methodology** to the design and design automation of embedded processors, MPSoCs and CPS, and this **research confirmed the adequacy of the quality-driven design methodology**
- For "Outstanding Achievements and Contributions to Quality of Electronic Design" I was awarded the Honorary Fellow Award by the International Society for Quality Electronic Design (San Jose, CA, USA, 2008)

28

Quality-driven Design, CPS and IoT for making high-quality systems

- When using the quality-driven design methodology to develop the modern high-quality collaborating cyber-physical systems, in which the sophisticated cyber systems (controllers) are tightly integrated with the controlled by them physical, social and life systems, we have a great chance to much better control and optimize the social, physical and life systems than we did it till now
- ***With modern CPS and IoT technology we have a great chance to significantly improve most systems used by us or that we are part of***
- **We also have no chance to not do this**
- ***Our social, physical and life systems have to be significantly and immediately improved***
- **Why?**
- Please watch the following few slides that I got from my friend Jean Paul Gueneau de Mussy, Sustainability and Innovation Expert, CEO of Materials and Systems Innovation Company, https://materials-innovation.com/

29

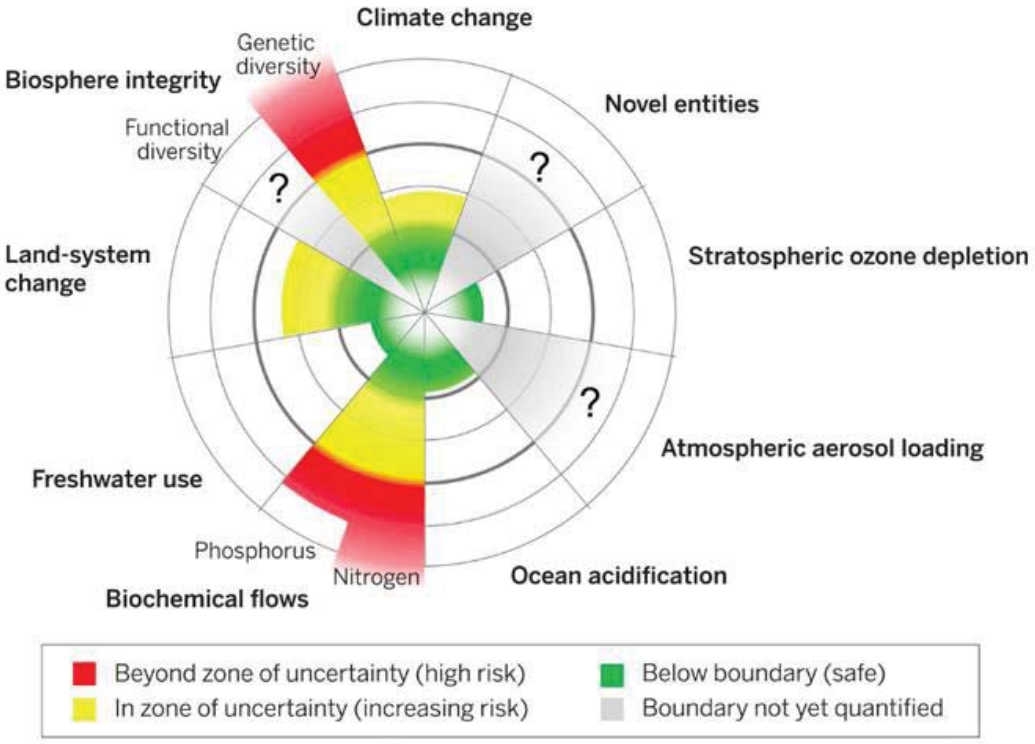

Overall costs of Climate Change
Jean Paul GUENEAU DE MUSSY | Materials-Innovation.com
Jean Paul GUENEAU DE MUSSY |
30

Jean Paul GUENEAU DE MUSSY | Materials-Innovation.com

Jean Paul GUENEAU DE MUSSY |

31

Planetary Boundaries



Johan Rockström et all, February 2017, Volume 46, Issue 1, pp 4–17

Huge destruction, chaos, no care for long-term consequences

- These were only a few examples of what was done wrong for a long time with our economic, social, technical and life systems on a global scale, and what resulted in a **huge destruction on a global scale**
- This huge destruction is a result of systemic drawbacks of the traditional economy and very many bad decisions made by numerous governments and companies for a short-term profit only, without accounting for long-term consequences
- Example: the wild chaotic globalization, without carefully designed interfaces and collaboration between very different economic/political systems in different parts of the World and between companies from the very different systems
- Globalization is unavoidable, but the actual costs of the wild globalization were not pay by those who profited, but by the poverty of others and destruction of the World
- The not well regulated and controlled inefficient collaboration chains and related material, product and waste flows of the wild globalization resulted in inefficient use of resources, environment destruction and pollution, climate change, bio-diversity loss, etc.

33

Huge destruction, chaos, no care for long-term consequences

- Covid-19 pandemics demonstrated the problems sharply
- Example: Due to globalization multiple supply chains became very complicated and very long, often crossing borders of several countries; due to Covid-19 pandemics, protectionism, etc. many chains were broken or function inefficiently
- For instance, current chip shortages for 5G, automotive, industrial machinery, electrical equipment, servers, etc. highlighted the supply competition among different countries and industries, and the necessity of making the critical supply chains less complicated, shorter, better controlled and more resilient
- The manufacturing of the global chip supply chains is mainly concentrated in East Asia, and manufacturing in the most advanced nodes below 10nm in Taiwan and South Korea.
- The decisions on the concentration of the critical manufacturing in one or two countries were almost only based on profit, without accounting for the fact that East Asia is a region of political conflicts and natural disasters
- The only-profit-driven wild globalization and chaotic resource exploitation results in a rapidly increasing fierce competition among different countries and industries for scarce resources, environment destruction and pollution

34

Broader context of the destruction

- Without understanding the broader context of the destruction we will not be able to effectively recover from it
- **The world is in constant war**: of **evil** against **good**.
- This war is "**eternal**" and has different phases of:
  - "**cold**" war, in the sense of moral, political, economic, etc., war

  and
  - "**hot**" war, in the sense of military conflict, revolution, and other types of enslavement and exploitation of people or destruction and looting of nature and all what humans created.
- Now this war between good and evil is a war between:
  - the world of civilization achievements being humanistic and ecological values, moral and social norms such as: human rights, democracy, self-governance, fair division of welfare, nature protection, etc.,
  - and
  - the backward old-fashioned world, negating humanistic and ecological values, negating moral and social norms such as: human rights, democracy, self-governance, fair division of welfare, nature protection, etc.

35

Broader context of the destruction

- Now this war between good and evil is a war between:
  - the world based on the state of law build on humanistic and ecological values, in which all are equal, and which protects everyone, a world where the government elected by the whole society in free and democratic elections acts for the social good within the law, and everyone has free access to information,

  and
  - the world of lawlessness of a totalitarian regime, negating humanistic and ecological values, denying and destroying moral and social norms, destroying or enslaving people, destructing and looting nature and all what humans created, and where society does not have free access to information and is manipulated by totalitarian propaganda.

Broader context of the destruction

- **Where is the front line between good and evil in this war**?

- Some say that this war between good and evil is between:
  - the world of the "West" build on a socially advanced civilization based on humanistic and ecological values, and social norms such as: human rights, democracy, self-governance, nature protection, etc.

  and
  - some other parts of the world where these rights and norms are not actually accepted and not followed by rulers and influential people.

- Is this the (whole) truth ???

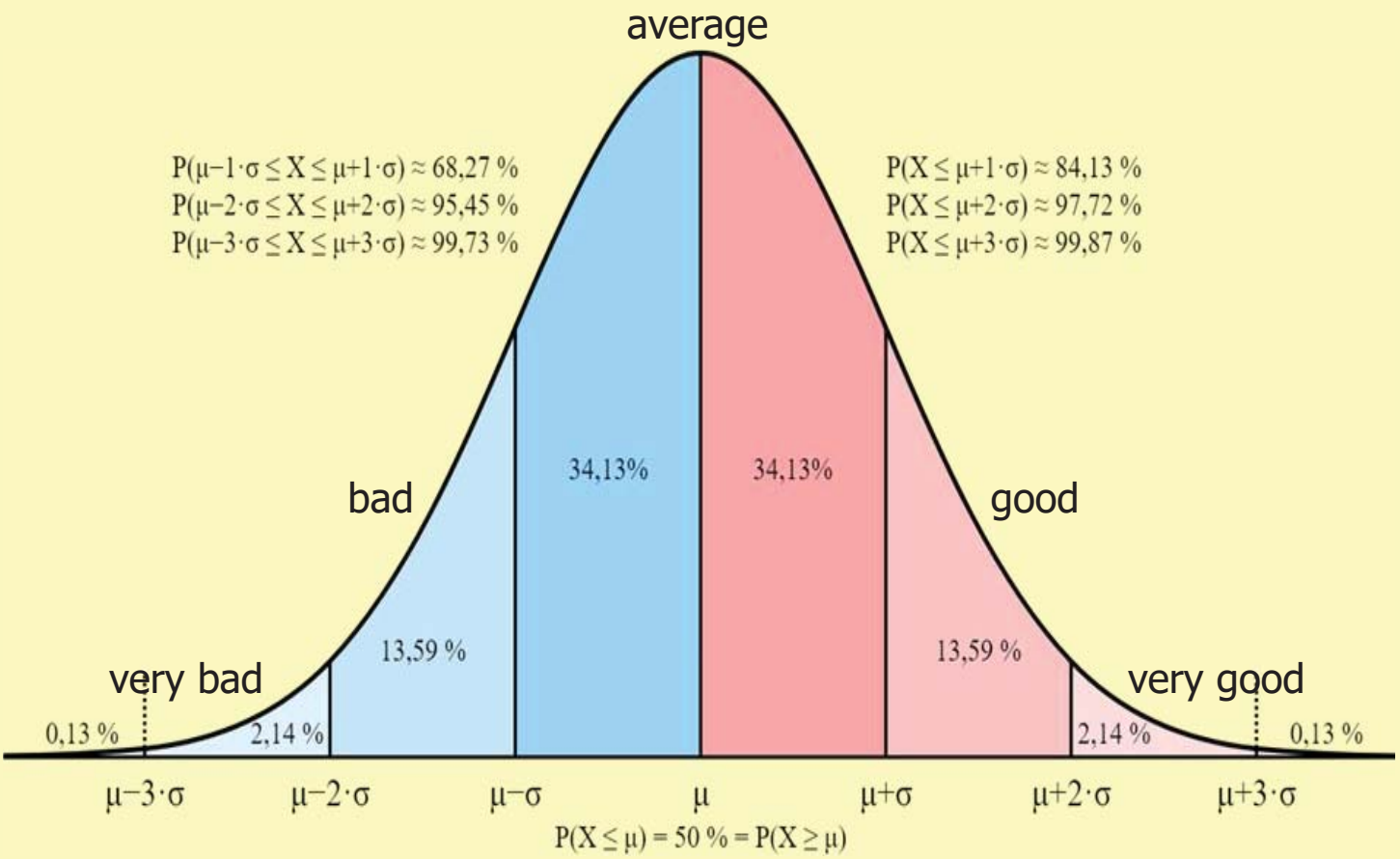
- **Definitely not !!!**

Broader context of the destruction

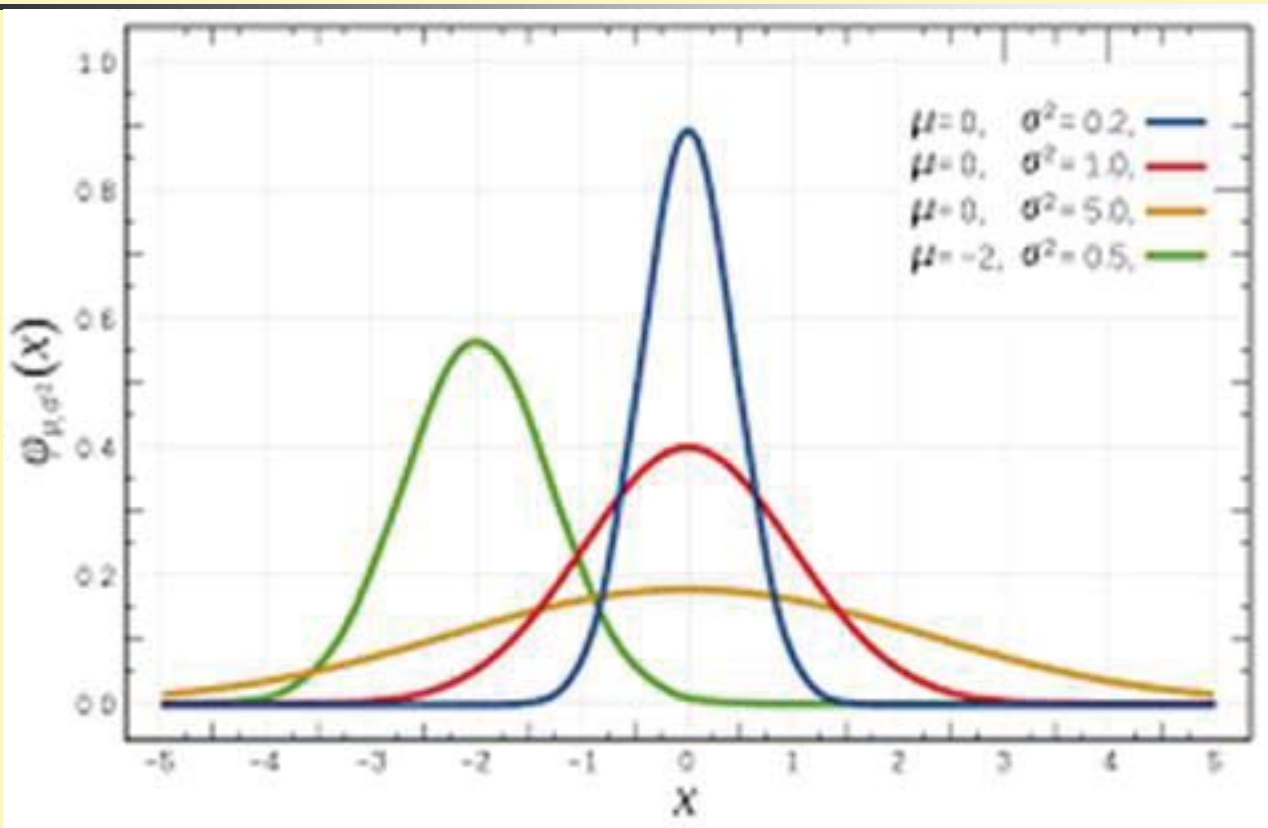
- In terms of a hot war, the front line between evil and good runs often between a totalitarian regime ruling a certain country and free nation of a neighbouring country
- The front line between evil and good is often between a totalitarian regime and a part of the society ruled by the totalitarian regime
- The front line between evil and good is often between a company owner not respecting people and environment, and the exploited company employees and destructed environment
- In general:
- **The war between good and evil is taking place all over the world, in every country and in every society**
- In each country and society, the distribution of characteristic features of people can be well modelled by a normal distribution
- In each country and society, there are "good" and "bad" people, but there are the most "average" people, less "good" and "bad", and only a small number of "very good" and "very bad" people.
- Likewise, with "smart" and "dumb" people

38

Broader context of the destruction



Broader context of the destruction



The parameters of the normal distribution can be different for each country and for each society

Source: Wikipedia

Broader context of the destruction

- Observe that "bad" and "stupid" people are in every country and in every society, but in different countries can be in different proportions
- In particular, the stronger totalitarian and longer-lasting totalitarian a country is, the more heavily manipulated the society of that country is and the more the mean value of the normal distribution shifts towards "evil" and "stupidity"
- Actual supporters of totalitarian regimes are usually people who are "bad" or bemused by totalitarian propaganda
- It is a common knowledge that one can influence people, their thinking and their characteristics
- Let us observe:
  - how important the role of free access to information and "real" education is, and
  - how disastrous is the lack of free access to information and propaganda instead of "real" education and information

41

Let us act on the side of good

- As people belonging to the best educated part of our societies, let us not only be well educated, but also "good" and "wise".
- **Let us be on the side of good** in this war between good and evil.
- Let's not wait for someone to win this war for us.
- Let us actively work against evil and do good in all the most effective and efficient ways available to us.
- Let us work for respecting the humanistic and ecological values, and for human rights, democracy, self-governance, fair division of welfare, nature protection, etc.
- Let us inform and educate people.
- As scientists and engineers: let us create "green" cyber-physical systems.
- **How to recover from the environmental disaster**?

42

EUROPE Recognizes the CLIMATE and POLUTION CRISIS and starts to take serious measures

EU President **Ursula von der Leyen** unveiled Europe's "**Green Deal**" plan to fight the crises on Dec. 11, 2019

It represents a stepwise incremental approach to solve the problems

43

How to recover from the disaster?

- The agreed in July 2020 Next Generation EU fund of €750 billion to recover from the crisis caused by the COVID-19 pandemics will be added to the regular EU budget for 2021–2027 to result in approximately €1824.3 billion
- As much as 30% of the total amount will be devoted to the climate and environment in compliance with the Paris Climate Agreement
- US also came back to the Paris Climate Agreement and devoted substantial funds to the climate and environment, and many other countries follow
- To recover from the disaster, ***a model of a well regulated and controlled effective and efficient system has to be applied to all kinds of systems, collaboration chains and related flows, implementing***:
  - **regenerative, circular and more local economy**

  and
  - **global ecology**
- In particular, ***this applies to collaboration chains and related material and information flows in CPS and IoT***
- ***What is circular regenerative economy?***

44

Traditional versus Circular Regenerative economy

- Traditional economy is characterised by assumption of unlimited growth; competition; intensive exploitation of and fighting for non-renewable scarce resources; and short-term profit maximalization, without taking care of the negative long-term economic, social and ecological consequences
- Traditional economy uses linear model: take scarce resources – make – use – dispose waste; it did not pay the actual costs of inefficient resource usage and of the pollution and destruction it made
- Circular regenerative economy is a systemic approach that aims to benefit all: business, society and environment, through:
  - quality-based growth, collaboration and partnership;
  - increasing use of renewable resources, resource sharing and gradually limiting the use of finite resources;
  - introducing biological cycles to regenerate living systems and technical cycles implementing product repair, reuse, sharing, remake, and recycling; and this way minimizing the use of scarce resources and regenerating the environment

45

Innovate applying circular economy and quality-driven design

- The principles of the circular regenerative economy are derived from the same source as the principles of my paradigms of life-inspired systems and quality-driven design
- They are derived from the observation of nature, and especially of structures and operations of living organisms, their populations and ecosystems that have demonstrated to effectively, efficiently and robustly work for many millions of years, and are a great source of inspiration
- Therefore, in relation to technical systems the principles of the circular regenerative economy repeat the main principles of the paradigms of life-inspired systems and quality-driven design proposed by me more than 20 years ago
- Implementation of the circular regenerative economy will require **many breakthrough innovations of processes and products**
- All those innovations will have to be designed and implemented
- ***When designing and implementing the innovative processes and products the methodologies of circular regenerative economy and quality-driven design should be used***

46

We have to recover from this disaster ASAP

- Innovations exploiting modern CPS and IoT technologies, circular regenerative economy and quality-driven design can significantly improve systems used by us or that we are part of
- Significantly improve does not mean to completely solve the environmental crises
- For this, the unnecessary and inefficient consumption has to be eliminated and all social systems have to be re-organized and made much more efficient
- The principles of circular regenerative economy and the quality-driven design methodology should be used to develop high-quality collaborating cyber-physical systems
- In these systems the sophisticated intelligent cyber systems (controllers) will be tightly integrated with the intelligently controlled and optimized physical, social and life systems
- This way, we have a great chance to much better control and optimize the social, physical and life systems than we did it till now
- This way, we can create green cyber-physical systems

47

Environmental footprint of cyber systems

- According to https://www.energuide.be, the average energy consumption and CO_2 footprint of a contemporary computer are the following:
  - desktop (basic peripherals included): 200 W/hour in work mode; used for 8h a day *consumes 600 kWh and emits 175 kg of CO_2 per year*;
  - laptop: 50 and 100 W/hour in work mode; used for 8h a day *consumes between 150 and 300 kWh and emits between 44 and 88 kg of CO_2 per year*;
  - in stand-by mode: the consumption/emission of both decrease to a third of the above.
- For microcontrollers (MCUs) and MPSoCs used in CPS, the story is much more complicated
- For them, the actual energy consumed depends on very many factors
- It is difficult to speak about an average energy consumption even for a given single MCU or MPSoC, because the energy consumption very much depends on the actual use and working conditions
- The power consumed by MCU or MPSoC grows with operating frequency, temperature, supply voltage and signal activity

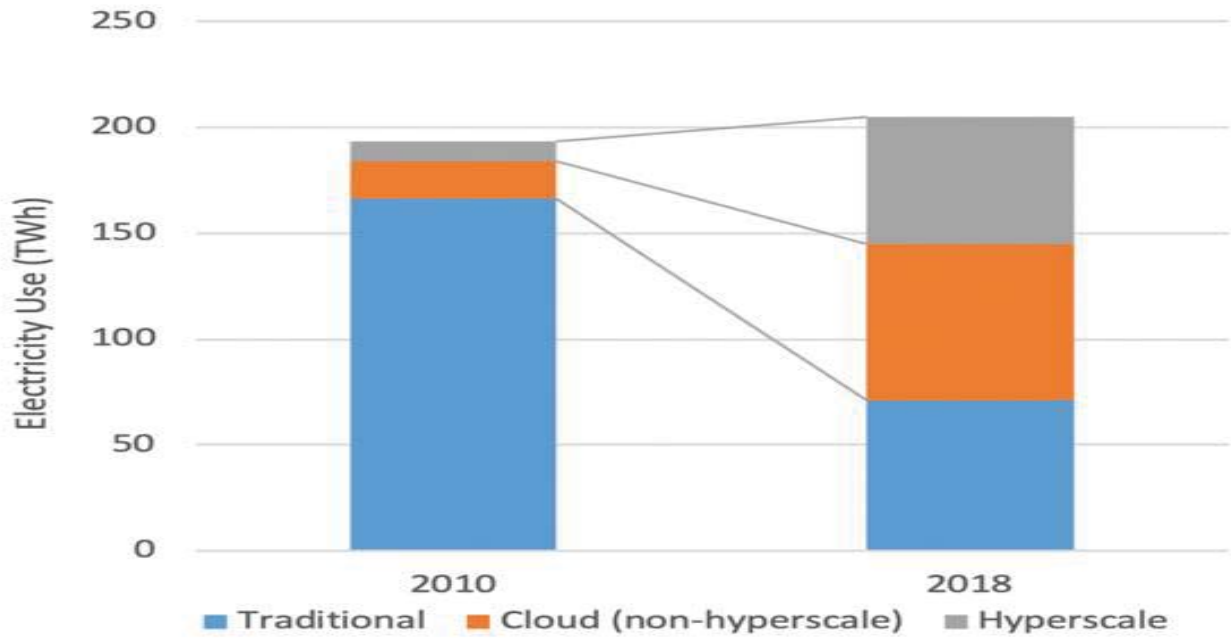
48

Environmental footprint of cyber systems

- Moreover, modern MCUs and MPSoCs often have several different active and energy saving modes (e. g. sleep, deep sleep, standby, etc.) and use the frequency and voltage scaling
- Finally, different MCUs and MPSoCs may have very different energy consumption characteristics, dependent on their architectures and implementation technologies, which in turn depend on the purposes/application fields which a given MCU or MPSoC is supposed to serve
- A simple ultra-low-power MCU for wearables can run in its active mode at much under 1W
- A complex MPSoC for automotive may use hundreds of Watts
- However, this is only a small part of the whole story
- The environmental footprint of cyber systems in CPS depends not only the embedded processors and their use, but on the usage of fog and cloud computers, and of the communication among all the computers as well

49

Environmental footprint of cyber systems



Source: https://energyinnovation.org/2020/03/17/

Figure 2. Estimated global data electricity use by data center type, 2010 and 2018. Source: Masanet et al. 2020.

- In 2018 global data centers consumed approximately 205TWh, what is more than the electric energy consumption of a medium country
- It represents 1% of global electric energy use and 0.3% of global CO_2 emission

50

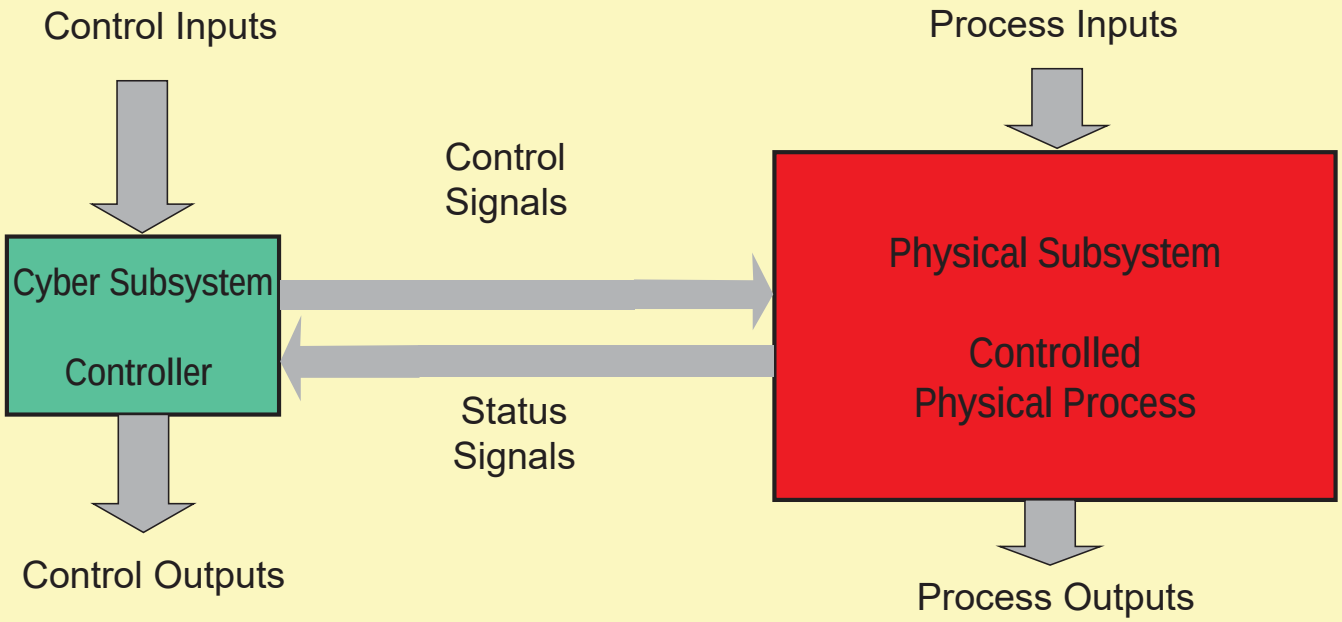
Environmental footprint of cyber systems

- Similarly, in 2019 global data transmission networks consumed around 250 TWh or somewhat more than 1% of global electric energy use, what corresponds to more than 0.3% of global CO_2 emission
- The demand for data center and network services is exponentially increasing.
- Between the 2019 and 2025, the number of IoT connections is expected to grow from 12 billion to 25 billion (https://www.gsma.com/mobileeconomy/wp-content/uploads/2020/03/GSMA_MobileEconomy2020_Global.pdf)
- To manage the environmental footprint of the CPS cyber systems, the exponential growth of CPS and IoT has to be compensated by efficient IoT organization and continuous energy efficiency improvements of embedded processors and MPSoCs, servers and storage devices, network processors and their software
- However, this is still only a small part of the whole story
- The environmental footprint of cyber systems depends not only on their use, but on their whole life cycle, including design, manufacturing, usage and disposal

51

Environmental footprint of cyber-physical systems

General Model of Cyber-Physical System



Usually:

Low Energy

High Energy

52

Environmental footprint of CPS

- The physical subsystem of CPS (implementing the controlled physical process) usually involves much larger material structures and flows, and several times more energy than the cyber subsystem (controller)
- The environmental and other effects are usually much larger from usage of the modern CPS and IoT technology to intelligently control and optimize the physical, social and life systems than from making green only the cyber systems
- We should make green the physical, social and life systems, as well as the cyber systems controlling them and the IoT connecting the collaborating CPS
- The environmental footprint of CPS and IoT depends on the whole CPS and IoT life cycle involving the CPS and IoT design, manufacturing, usage and disposal
- *Manufacturing* usually includes installation, testing and validation
- *Usage* often involves maintenance, repair and enhancement
- Let's start with IoT

53

Distribution of intelligence, computing resources, services and workloads in the IoT chierarchy

- To transform the big data from multiple sensors to the information being directly used for decisions, while satisfying the stringent requirements of the modern mobile systems, a careful distribution of information delivery and computation services among the different layers of IoT is needed
- For many reasons of primary importance, as:
  - real-time availability of local information
  - guaranteed real-time reaction
  - privacy, security, safety, reliability
  - minimization of energy used, communication traffic, costs, etc.

  a majority of computing and decision making related to advanced CPS should be performed locally in the IoT edge devices, in collaboration among various local IoT edge devices or just above the edge nodes, and not in the higher levels of fog or in cloud
- The higher levels of fog and cloud should only be asked for services if:
  - necessary information or computing resources are not available locally, and
  - reaction-time, security, safety, etc. allow for this

54

Distribution of intelligence, computing resources, services and workloads in the IoT chierarchy

- This requires implementation of advanced intelligent computations and sophisticated powerful embedded computing technology:
  - directly in the IoT edge devices related to the (complex) sensors and actuators, or
  - just above the edge nodes, where the information from different sensors can be combined and based on the combined information the control decisions can be taken and subsequently actuated
- Sophisticated and powerful edge computing has to be used requiring advanced intelligence, processing power and communication capabilities to be pushed towards the edge-nodes of IoT, where the data originate and information is used (i. e. to sensors, controllers and actuators)
- A very good example of the edge computing necessity is the **local** vehicle-to-vehicle and -infrastructure communication and collaboration necessary for autonomous driving
- In consequence, the **IoT for advanced CPS will be substantially different than Internet for other traditional targets**

55

Edge Computing, Intelligent Sensors, Edge AI and Edge ML

- This is the reason why Edge Computing, and specifically, intelligent sensors and actuators, as well as edge Artificial Intelligence (edge AI) and edge Machine Learning (edge ML) became very relevant and hot R&D topics recently
- Artificial intelligence (AI) is intelligence demonstrated by organized systems (e.g. machines), in contrast to "natural" intelligence demonstrated by organic systems (e.g. humans or animals)
- An intelligent system is a system that shows a goal-directed behavior
- AI system is a system that analyses the problem, and based on the analysis results, takes actions that maximize the chances of success to achieve the goal
- Machine learning (ML) is a learning implemented in machines through developing methods and algorithms that can "learn"", in the sense of being trained on some set of data, discovering the structure in data, or optimizing own performance for some set of problems through interacting with environment and processing feedback from the environment

Edge AI and Edge ML

- Usually, based on the training data machine learning methods/algorithms build/train a model which is then used to process additional data to make decisions or predictions
- Machine learning system is an organized system that implements one or more machine learning methods/algorithms
- Several types of learning approaches and models are known
- Depending on the nature of the input data and feedback used for learning the following three main machine learning approaches can be distinguished: supervised learning, unsupervised learning and reinforcement learning
- Supervised Learning:
  - the training data consists of sample inputs and the desired output for each sample input
  - the learning goal is to build/train a model that implements a general rule (function) that maps inputs to outputs

57

Edge AI and Edge ML

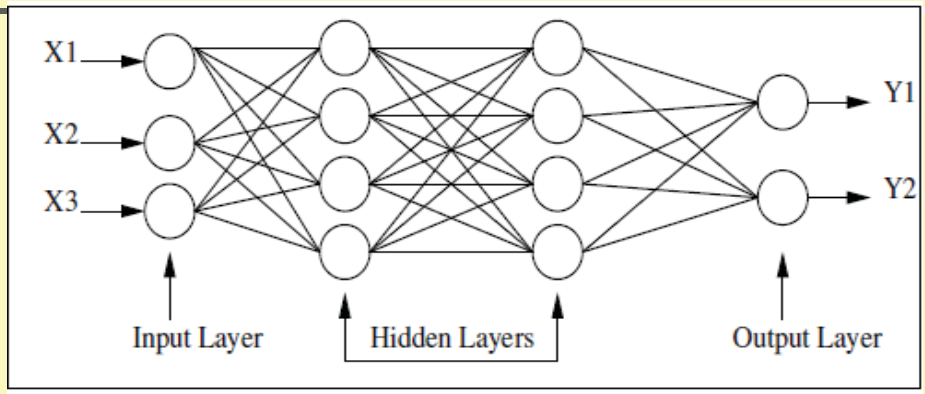
- Unsupervised Learning:
  - the training data consists of only inputs and no information on the desired outputs is given to the learning algorithm
  - the learning goal is to discover structure/patterns in the data (such as clustering, partitioning or other grouping) through finding and reacting to commonalities and differences in the data in order to react to new data based on the discovered structure and commonalities/differences
- Reinforcement Learning:
  - it does not need explicit training data (as e.g. some sample inputs and the desired outputs for the inputs);
  - it learns the optimal behaviour through interacting with its environment in which it has to achieve a certain goal (e.g. autonomously driving a car, planning an efficient and safe path for a robot, etc.) and observing how the environment responds to its actions;
  - it does it through rewarding desired actions and punishing non-desired once to discover the sequence of actions that maximize the cumulative reward that is usually expressed by a value function that defines the cumulative reward as a sum of the award of being in a certain state and the expected future award to be collected from that state further

Edge AI and Edge ML

- Although reinforcement learning methods/algorithms can be model-based or model-free (i.e. they can work without building an explicit model of the environment), a vast majority of ML methods/algorithms use various models, such as: artificial neural networks, support-vector machines, decision trees, belief networks, etc.
- In CPS and IoT, Machine Learning is used for a wide variety of important tasks, such as:
  - video and image processing
  - computer vision
  - speech processing and recognition
  - object (e.g. human, animal, car etc.) motion prediction
  - robot or vehicle path planning
  - and many more
- Machine Learning (ML) can be seen as a part of Artificial Intelligence (AI), although some researchers argue that they only have a large common part

59

Edge ML and Artificial Neural Networks (ANNs)



- ANN is a ML model involving nodes called neurons which are connected with edges called synapses that serve to transfer signals between neurons
- A neuron processes the received signals, when computing a non-linear function of the sum of its inputs, and then sends a signal to neurons to which its output is connected
- Neurons are organized into layers that may perform different transformations on their inputs
- Neurons and edges have weights that adjust during the learning process, and this way change the strength of the signals at a corresponding connections
- Neurons may also have a threshold, so that the signal is only sent from a given neuron if the aggregate signal of the neuron is higher than the threshold value

60

Edge ML and Deep Learning (DL)

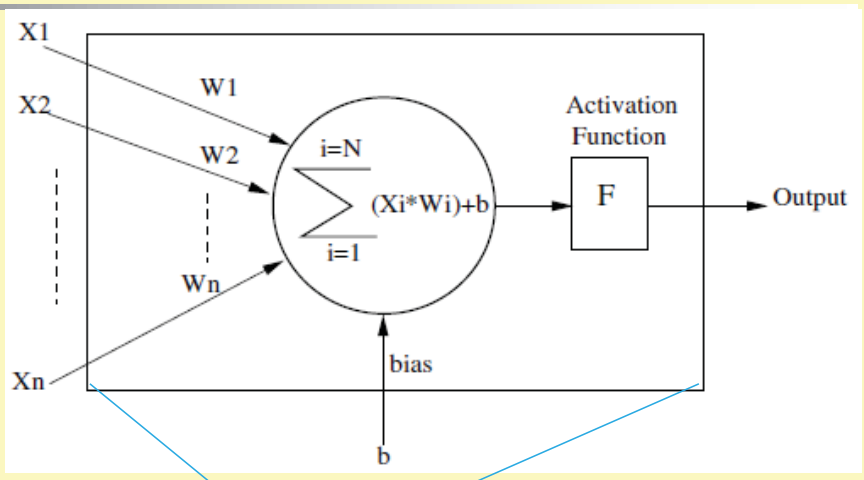
- The spectacular progress in the nano-dimension semiconductor technologies in the recent 10 years resulted in:
  - ability to implement much larger and more complex neural networks on a chip

  and
  - availability of powerful processing technology that makes it possible to efficiently use large data sets to train the networks
- In the recent 10 years the neural networks and related fields are undergoing a reincarnation in the form of the so called "deep learning"
- Deep learning refers to machine learning using complex ANNs with "multiple layers of nonlinear processing units and supervised or unsupervised learning of feature representations in each layer, with the layers forming a hierarchy from low-level to high-level features." (Wikipedia)
- Specifically, in 2015 several specific computing platforms for multi-layer Convolutional Neural Networks (CNNs) called Deep Neural Networks (DNN) have been developed, by Cognivue: G2-APEX IP - High Performance Image Cognition Processor Core, Qualcomm: Zeroth cognitive processor, Nvidia: Drive PX, Synopsys: DesignWare EV Processors, and by some other companies

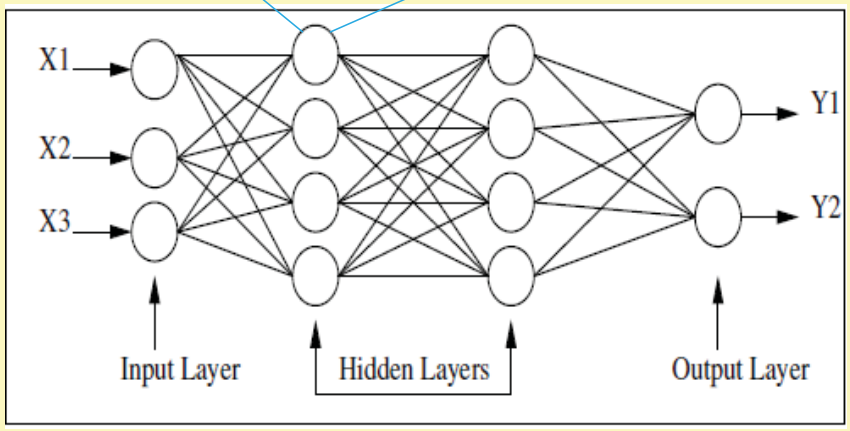
61

Convolutional Neural Networks (CNNs) and Deep Learning (DL)

- CNN is type of ANN in which neurons mainly compute convolutions on their input data



- CNN with several hidden layers is called a deep CNN (DNN)



62

New Edge Computing Platforms for ML and AI

- The interest in Machine (Deep) Learning and Artificial Intelligence is rapidly increasing (Research and Markets predicts that AI in IoT will reach a value of $14,8 billion by 2026)
- ML and AI technologies belong to main contributors to modern CPS and IoT, but they are also expected to substantially contribute to the solution of the environmental crises
- In the last two years many different new Edge computing platforms and accelerators for Deep Learning, other learning and other AI have been developed
- Synaptics developed Katana ultra-low power Edge AI SoC for a wide range of energy constrained IoT applications (e.g. sensors and edge devices in offices, factories, warehouses, robotics, farms, smart homes and cities, etc.)
- Katana has a heterogenous hexa-core architecture, with each core optimized for specific set of tasks, such as ANN image processing, audio/voice processing, control, etc. Its power efficiency mainly results from the combination of efficient specialized cores, sophisticated dynamic voltage and frequency scaling mechanism, and efficient (model) development and optimization tools.

63

New Edge Computing Platforms for ML and AI

- Silicon Labs implemented a low power AI/ML accelerator in the BG24 and MG24 Gecko wireless SoCs for IoT applications, such as such as medical, industrial, and smart home and town.
- It can speed up IoT AI/ML workloads up to four times with a resulting six-fold power savings compared to Cortex-M33, what makes BG24 and MG24 suitable for battery-operating IoT devices
- GreenWaves developed Gap9 ultra-low power neural network Edge processor suitable for battery-powered devices and optimized for advanced audio. The total power consumption for Gap9 can be as low as 1.8 mW
- quadric developed q16 processor implementing "Quadric architecture" being a general-purpose neural processing unit (GPNPU). Quadric's GPNPU architecture is claimed to be appropriate for a wide range of processing (including ML, general-purpose control, and signal processing) and delivering a high performance for both ML inference and conventional signal processing, as e.g. for computer vision
- Quadric announced that it starts marketing its edge AI processor architecture as a licensable intellectual property (IP).

64

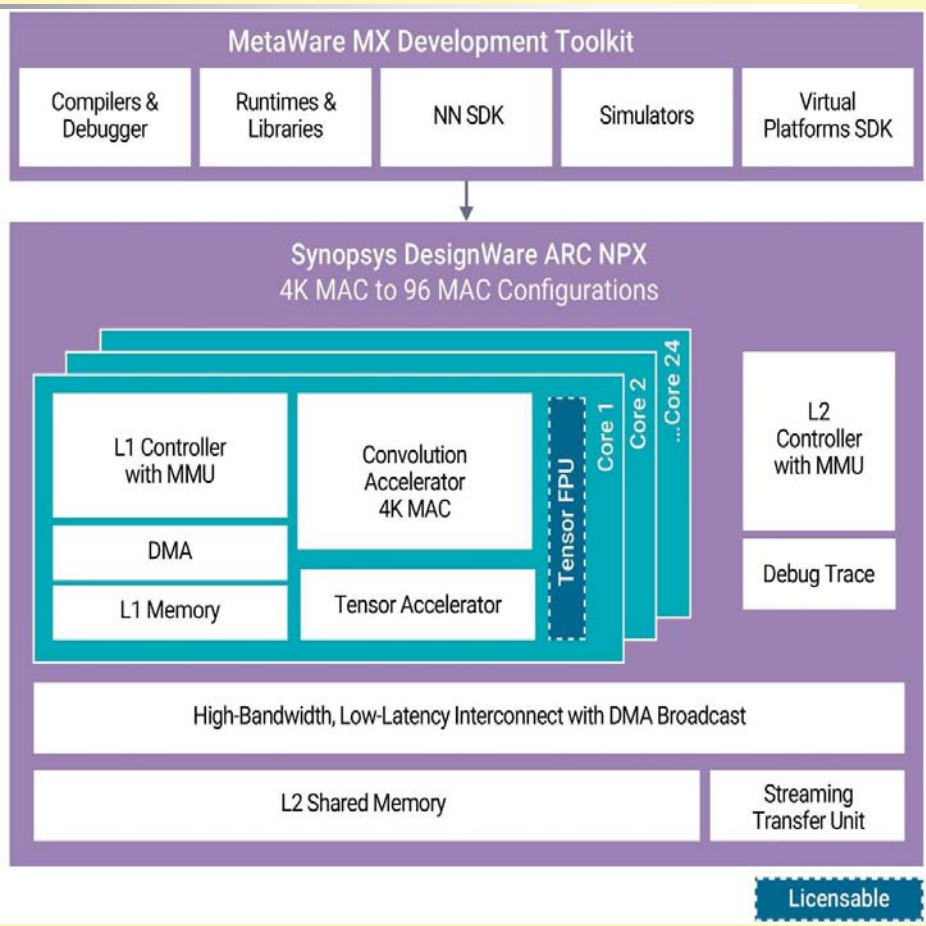
New Edge Computing Platforms for ML and AI

- Intuitive developed NU4000 high-performance heterogeneous multi-core SOC for robots, drones, VR/AR, etc. that involves a. o. a high-quality 3D depth engine, SLAM accelerators, strong Computer Vision engine and deep-learning CNN processor
- It is a powerful imaging, vision and AI computing platform with total performance exceeding 8 TOPS
- The CNN processor provides above 2 Terra OPS, and it enables processing of large deep CNNs (e.g. VGG16) at the rate of 40 frames (ROIs) per seconds at ~10 times less power than equivalent GPU, DSP or FPGA based processing
- For deep-learning with advanced CNN models the Synopsys DesignWare® ARC® EV processor IP is used in NU4000
- Synopsys DesignWare ARC NPX Neural Processor IP family provides a high-performance, power- and area-efficient IP solution for applications requiring AI supported computer vision (e.g. for object detection, scene segmentation, image quality improvement), and for broader range of applications (e.g. audio, natural language processing, etc.)

65

New Edge Computing Platforms for ML and AI

- The NPX6 NPU IP architecture is based on individual cores that can scale from 4K MACs to 96K MACs for a single AI engine performance of over 250 TOPS and over 440 TOPS with sparsity
- The IP includes hardware and software support for multi-NPU clusters of up to 8 NPUs achieving 3500 TOPS with sparsity
- An optional tensor floating point unit is available for applications benefiting from BF16 or FP16 inside the neural network.



Source: Synopsys

66

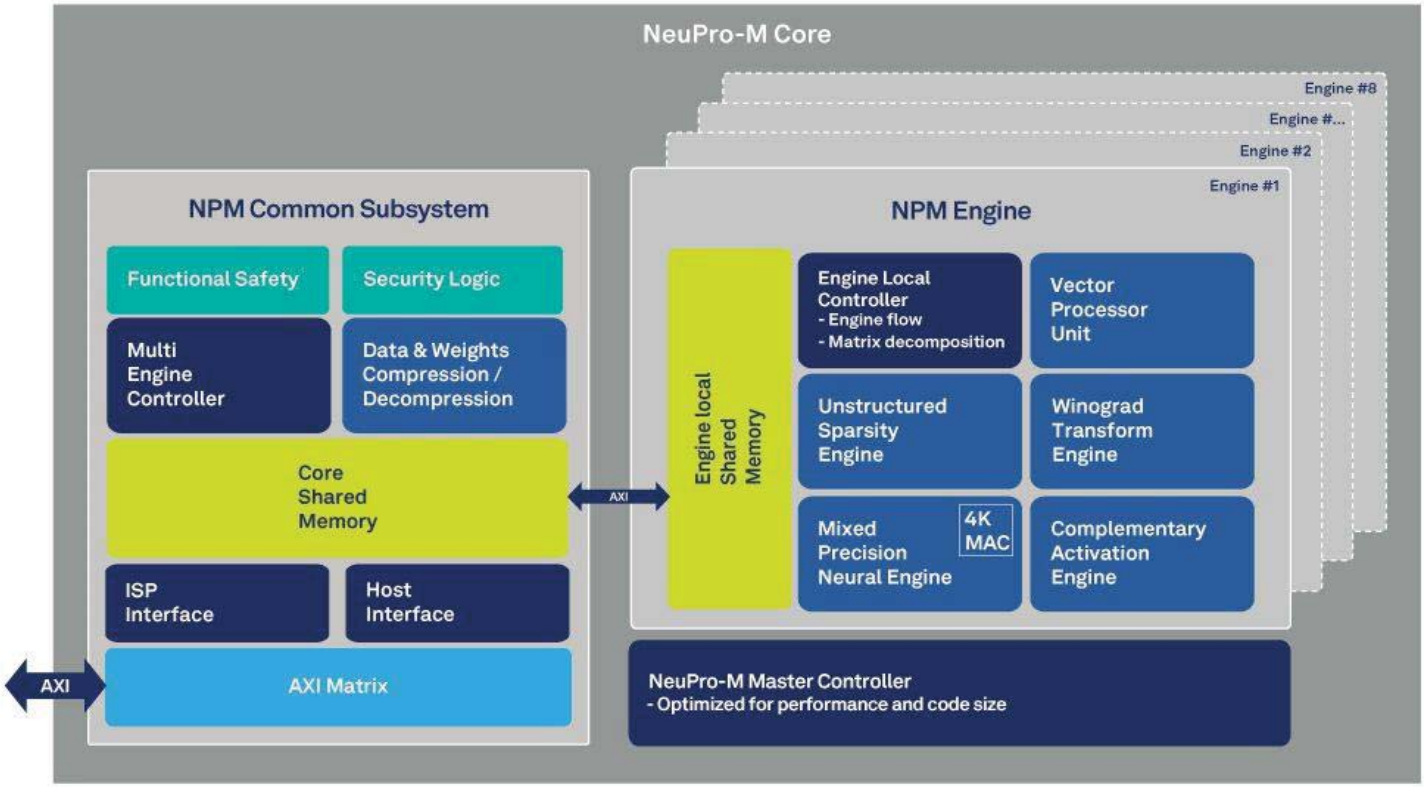
New Edge Computing Platforms for ML and AI

- CEVA's new NeuPro-M scalable heterogeneous high-performance AI processor architecture IP is focused on edge AI/ML applications where both the high-performance and power efficiency are of importance, such as automotive and other mobile systems, robotics, industrial IoT, etc.
- It achieves from 8 TOPS (with a single NPM engine) up to 160 TOPS (with 8 NPM engines) per core, and can reach above 1200 TOPS with multi-core instantiations, with a remarkable power efficiency of up to 24 TOPS/Watt
- It is accompanied by and CEVA Deep Neural Network (CDNN) development software including:
  - NeuPro-M system architecture planner for ANN deployment over NeuPro-M
  - ANN training optimizer

  and
  - CDNN compiler

67

New Edge Computing Platforms for ML and AI



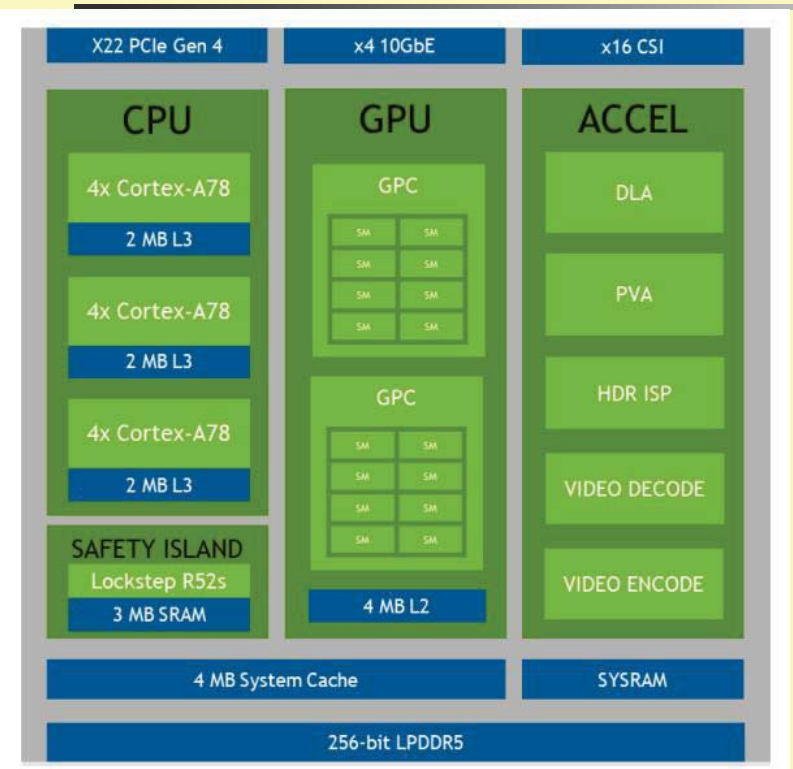
Source: CEVA

New Edge Computing Platforms for ML and AI

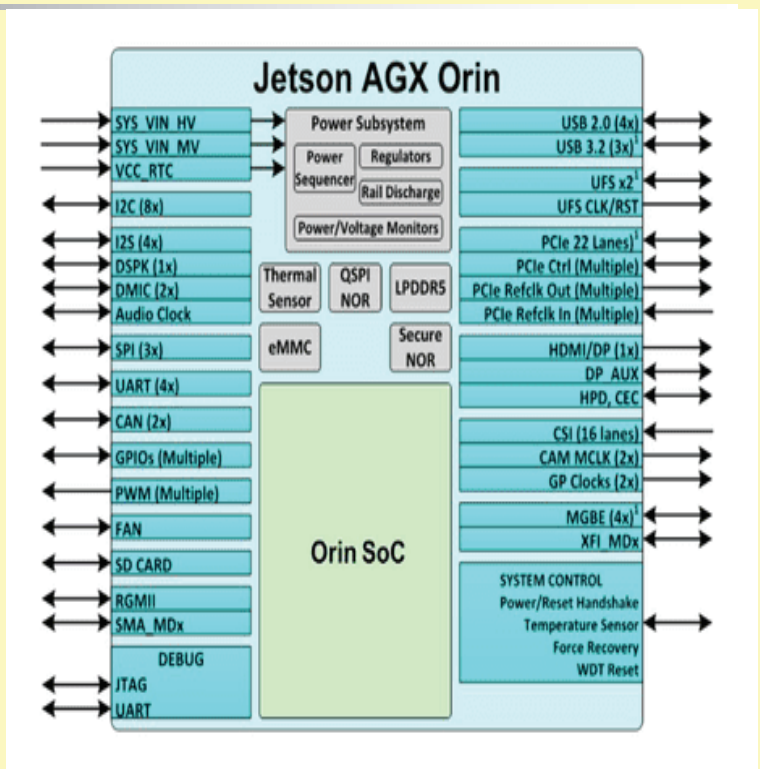
- NVIDIA introduces new Jetson AGX Orin System-on-Module (SoM) for powerful high-performance and energy-efficient AI/ML at the edge
- It is aimed at the most advanced applications requiring powerful embedded computing at the edge in such sectors as advanced medical devices, autonomous cars, autonomous delivery, logistics and factory robots, advanced UAVs, and other advanced autonomous systems, for highly demanding tasks of multi-sensor fusion, computer vision, motion prediction, path planning, natural language understanding, etc.
- Jetson AGX Orin delivers up to 200 TOPS AI performance, which is comparable to the performance of a GPU-based server, but has a size of only 100mm x 87mm and uses much less power (15 – 40 W)
- Jetson AGX Orin SoM is built around Orin SoC with Nvidia's GPU Ampere architecture with 1792 NVIDIA® CUDA® cores and 56 Tensor Cores in two Graphic Processing Clusters (GPCs), 8-core ARM Cortex-A78AE CPU, powerful HW deep learning accelerator (DLA) and vision accelerator (PVA), video encoder and video decoder
- High speed I/O and memory with 204 GB/s bandwidth make it possible to run multiple concurrent AI pipelines

69

New Edge Computing Platforms for ML and AI



Orin SoC Block Diagram



Jetson AGX Orin System-on-Module

Source: NVIDIA

New Edge Computing Platforms for ML and AI

- Mobileye introduces its EyeQ Ultra high-performance and low-power SoC aimed at autonomous vehicles and similar advanced applications
- EyeQ Ultra is fabricated in 5-nm process and delivers AI performance up to 176 TOPS at less than 100 W
- It has a very heterogeneous architecture involving several different types of cores tuned to different tasks involved in an L4 autonomous car, including:
  - 12 RISC-V CPU cores,
  - Arm GPU and VPU,
  - 4 types of Mobileye's proprietary accelerators involving 16 CNN accelerators, 8 CGRA-based cores, 16 VLIW/SIMD cores, and 24 barrel-threaded CPU cores,
  - video encoding/decoding cores, safety/security subsystem, two separate sensor subsystems: one camera-only, and the other one for radar and lidar, etc.
- Each of the two separate sensor subsystems can support a full operation, and this redundancy results in a more robust overall system.

New Edge Computing Platforms for ML and AI

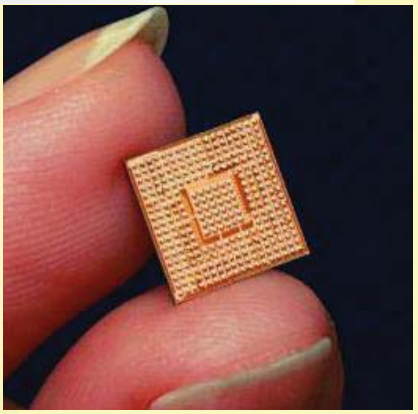
Source: Mobileye, an Intel Company

72

New Edge Computing Platforms for ML and AI

- GrAI Matter Labs (Eindhoven, Paris, San Jose) is introducing GrAI VIP Edge AI SoC for high-performance and energy-efficient AI/ML at the edge, aimed at near-sensor AI/ML based solutions in robotics, industrial automation, AR/VR, Smart Homes, Infotainment in automobiles, etc.
- GrAI VIP SoC is based on GrAICore™ neuron AI engine, and involves embedded ARM processors, and interfaces to be connected to a multitude of sensors (e.g. vision, sound, pressure, etc.) to enable Life-Ready AI
- GrAI VIP SoC is manufactured in 12nm TSMC process and has a 8mmx8mm compact package with memory included
- It can can execute complex AI applications based on advanced DNNs, such as ResNet-50, EfficientNet, SSD, Yolo, Unet, etc. with very low inference latencies (only few ms for ResNet-50) and low-power (< 0,5 W for ResNet-50)
- GrAIFlow™ SDK provides a complete toolkit to conceptualize, visualize, test, debug, and iterate user applications

New Edge Computing Platforms for ML and AI



GrAI VIP SoC

Source: GrAI Matter Labs

74

New Edge Computing Platforms for ML and AI: **Summary**

- Very many new Edge Computing Platforms for ML/AI have been developed recently, and the above discussed constitute only a subset of them
- There is no one computing platform that fits all applications
- Different AI/ML computing platforms with different performance, energy usage, size etc. are needed for different CPS and IoT applications having their different characteristics and different requirements
- For Edge Computing it is necessary to realize the required computations when guaranteeing the required performance at the minimum possible energy used
- Nevertheless, the platforms have much in common
- All the computing platforms for Edge AI/ML are heterogeneous and implement massive parallelism

75

What can be the future in Computing Platforms for ML and AI?

- Companies will introduce their next-generation more powerful platforms
- For example: Nvidia is introducing its Orin SoC, but it already announced its next-generation much more powerful Drive Atlan that is expected to process AI/ML loads with 1000 TOPS.
- European Processor Initiative (EPI) will perhaps propose some solutions, as EPI aims a.o. at Accelerator Processors and Edge Systems?
- More energy efficient implementation technologies will be used
- For example: The StorAIge project led by STMicroelectronics aims to develop and industrialize the FD-SOI 28nm process and a very efficient embedded Phase Change Memory (ePCM), what will enable high-performance and ultra-low power SoCs for Edge AI (the tearget is 10 TOPS/W)
- Other examples can be analog implementation and in-memory computing (e.g. DIANA chip from Dutch AI startup Axelera, Eindhoven), use of photonic IC technology, etc.
- Development and use of a more efficient "neuromorphic" computing that will much better mimic the biological brain processes than the currently used neural networks
- etc.

76

Main IoT Networking Technologies and Standards

- As earlier explained: the IoT for advanced CPS will be substantially different than Internet for other traditional targets
- Specifically, due to different application requirements in relation to connectivity (data rate, latency, etc.), deployment area, number of connected devices, energy consumption, safety, security, reliability, cost, etc. different networking technologies, standards and protocols will be used
- The following two kinds of IoT applications are distinguished in relation to two distinct areas of the requirement spectrum: Massive IoT and Critical IoT
- Massive IoT refers to applications that require a huge number (from thousands to milliards) of low-cost and low-energy devices often in remote locations, each generating a small number of (regularly) reported data, and that have relatively low throughput and latency requirements:
  - Aim: to efficiently transmit small amounts of data from the huge number of devices
  - Key requirements: sufficient network capacity, scalability, security and availability, wide and strong coverage, (ultra) low-power/energy, low cost
  - Example Applications: smart metering, smart building/city, smart grid, asset tracking, fleet management, wearables and part of e-health, process monitoring and optimization in indystry, environmental monitoring, climate monitoring and livestock tracking in agriculture, etc.

77

Main IoT Networking Technologies and Standards

- Critical IoT refers to time- and safety-critical applications that demand data delivery within a specified time and with required guarantees, and that usually involve fewer (up to thousands) complex costly devices, each generating/receiving large amount of data with high throughput and low latency requirements, and that have to withstand harsh/remote environments, as well as security threats and attacks:
  - Aim: to guarantee efficient transmission of large amount of data with high throughput and low latency in harsh environment and while facing security threats and attacks
  - Key Requirements: guaranteed high-bandwidth, low-latency, and very high security, safety, reliability, and availability, at low energy and acceptable cost
  - Example Applications: Autonomous Vehicles and V2X, UAVs, Robotics, Industry 4.0, telemedicine, VR/AR/MR applications, traffic and flight control and safety, critical part of smart city, etc.
- For massive IoT applications requiring:
  - low-power, wide area connectivity, security and availability, cellular network standards LTE-M and NB-IoT can be used
  - very low power from the device to send/receive data, very many connected devices/large area and lower cost, some LPWANs, as LoRa or Sigfox, can be used

78

Main IoT Networking Technologies and Standards

- For home appliances and similar consumer devices and applications WiFi, Bluetooth, Thread or Zigbee can be a satisfactory and low-cost solution, and the recently introduced Matter uses a combination of WiFi, Bluetooth Low Energy and Thread to enable devices and applications interoperability
- From the above it is clear that 5G is not always required and not always the best option for IoT
- However, 5G is indispensable for Critical IoT, as it provides Network Slicing, and much higher bandwidth, lower latency, lower power consumption, and higher safety, security and reliability than 4G
- Using Network Slicing the service provider can devote a part of the 5G radio spectrum to run a separate private wireless network for a company, or an NB-IoT massive service connecting thousands of sensors, or to enable higher bandwidth and lower latency for some highly demanding applications as autonomous vehicles or UAVs
- Allied Market Research reported that the global market of 5G infrastructure industry was $2.06 billion in 2020, and the market will grow to $83.62 billion by 2030, at a CAGR of 45.3 percent between 2021 and 2030

Quality-driven design approach

- To develop green collaborating CPS the principles of circular regenerative economy and the quality-driven design methodology should be used
- **System design is a *definition of the required quality***, i. e. a satisfactory answer to the following two questions:
  - **What new** (or modified) **quality is required**?
    and
  - **How can it be achieved**?
- Intuitively we feel that **quality** is here used in the sense of ***the totality of the (important) features the system has***
- So, **system design should define**:
  - **What is the required totality of the (important) system features?**
    and
  - **How to realize a system that has these all features**?
- In other words:
  - What process must be realized in a certain system and what structural and parametric features must have the system?
  - How can we build a system that will be able to realize this process and will have the required structural and parametric features?

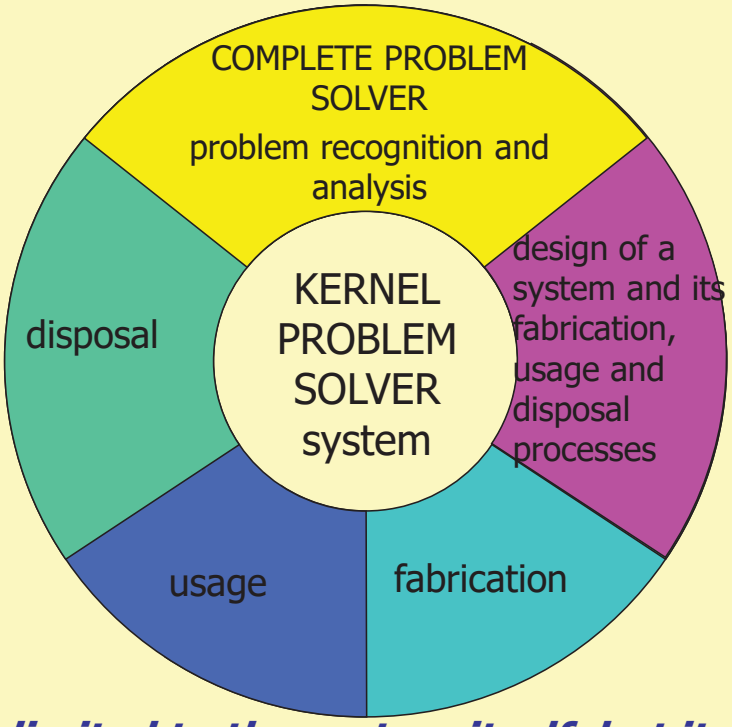
80

Quality

- Actually, **what is quality?**
- The most used and cited definitions of quality:
  - fitness for use (*Juran*)
  - conformance to requirements (*Crosby*)
  - quality is meeting the customers' expectations at a price they can afford (*Deming*)
  - the loss of quality is the loss a product causes to society after being shipped, other than any losses caused by its intrinsic functions (*Taguchi*)
  - the totality of features and characteristics of a product or service that bear on its ability to satisfy given needs (*American Society for Quality Control*)
  - the totality of features and characteristics of a product or service that bear on its ability to satisfy stated or implied needs (*ISO8402: Quality Vocabulary Part 1)*

Problems with the existing definitions of quality

they focus exclusively on a product being designed, while the original problem is solved by designing, fabrication, usage and disposing of the system



Quality cannot be limited to the system itself, but it must account for the complete problem solution, related to complete system life-cycle

82

Problems with the existing definitions of quality

- none of these definitions is precise enough to enable the systematic consideration, measurement and comparison of quality
- the assumption of perfectly known and inviolable customer's requirements is not acceptable, because the customer may specify the requirements poorly and such requirements may result in system which will create danger, damage environment or squander scarce resources
- **engineered systems** solve certain real-life problems, serve certain purposes – they are **purposive systems**
- quality of a purposive system can only be defined in relation to its purpose

83

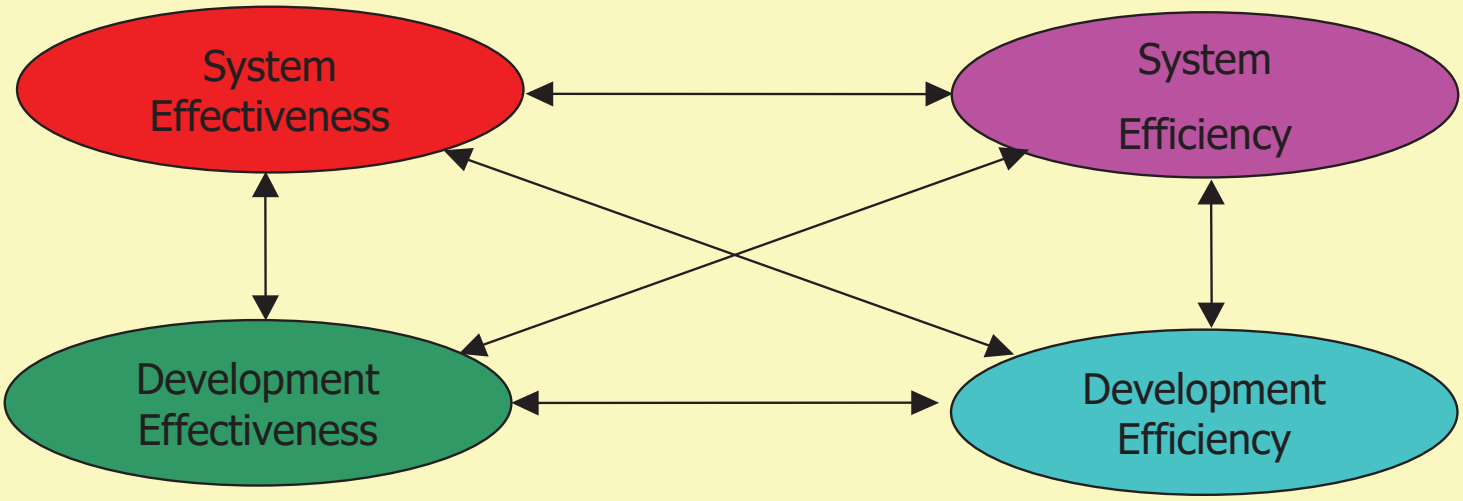
New quality definition proposed by me 20 years ago

__Quality__ of a purposive systemic solution is
its __total effectiveness and efficiency__
in solving of the real-life problem that defines the solution's purpose

- **Effectiveness** = the degree to which a solution attains its goals
- **Efficiency** = the degree to which a solution uses resources in order to realize its aims
- ***Effectiveness and efficiency of a systemic solution together decide its grade of excellence*** - **their aggregation expresses quality**
- Effectiveness and efficiency can be expressed in terms of measurable parameters, and in this way, quality can be modeled and measured
- In particular, the quality can be modeled in the form of *multi-objective decision models* involving measurable design parameters
- *The multi-objective decision models* and *design parameter estimators* enable application of the *multi-objective decision methods* for construction, improvement and selection of the most promising solutions

84

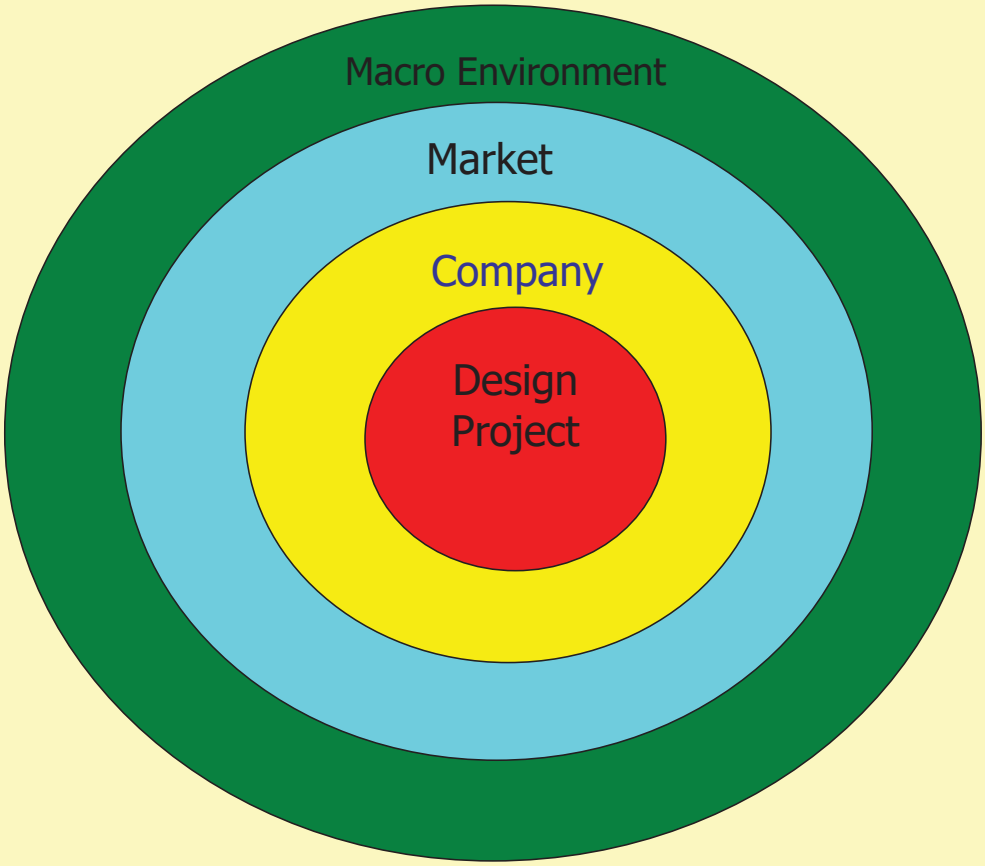
Quality-driven Design - Difficulties



Interactions and trade-offs between various parts and aspects of the total systemic solution

85

Quality-driven Design - Difficulties



Interactions of a design project with its context

86

Quality-driven Design - Difficulties

- Design does not concern the reality as it is, but as it will possibly be realized
- Quality recognition and formulation, i.e. recognition of the problem, as well as of the nature of its solution are ***subjective*** to a high degree
- The **contemporary system design problems** are ***complex***, ***multi-aspectual***, ***dynamic***, and ***ill-structured***:
  - there is no definitive formulation of the problem,
  - any problem formulation may be inconsistent,
  - formulations of the problem are solution dependent,
  - proposing and considering solutions is a means for understanding the problem, and
  - there is no definitive solution to the problem

87

Quality-driven Design - Difficulties

- The complex design problems are ill-defined
- It is very difficult to find precise relations between various aspects of the system effectiveness and between the different forms of energy and matter used to attain the system's aim, and even more difficult to express them as one uniform measure
- There are trade-offs as well between effectiveness and efficiency as among different their aspects
- The required quality or its perception can change in time

quality cannot be well defined,
but it can and should be modelled

88

Quality-driven Design - Design models

- *Well-structured models of the required/delivered quality* can serve to:
  - conceptualize, denote, analyse and communicate the customer's and designer's ideas
  - show that the requirements and designs are meaningful and correct
  - guide the design process
  - enable the explicit and well-organized design decision making
  - enable design automation
  - etc.

89

Quality-driven Design: Design problem-solving using models

- Since the system design problems are:
  - complex;
  - multi-aspect;
  - ill-defined,

  to solve them, ***all human concepts for dealing with complexity, diversity and ill-structure have to be applied***:
  - abstraction;
  - separation of concerns;
  - decomposition and composition;
  - generalization and specialization;
  - modelling;
  - simulation;
  - prototyping;
  - .....
- ***A design problem has to be converted into a system of simpler sub-problems***
- The solution to the original problem can then be achieved by solving the sub-problems and composing the sub-problem solutions into an aggregate solution

90

Quality-driven Design: Design problem-solving using models

- The problem decomposition and design modelling are to some degree subjective
- The design decision processes are also to some degree subjective, as they are influenced by the designers' value systems, feelings, believes, intuition etc.
- The design problem solving activity is performed under uncertainty, inaccuracy, imprecision and risk conditions, and in a dynamic environment

- ***System design has to be an evolutionary process*** in which analysis and modelling of problems; proposing their solutions; analysis, testing and validation of the proposals; learning and adapting are very important

91

Main concepts of the quality-driven design

- Designing ***top-quality systems is the aim*** of a design process
- ***Quality is modelled and measured*** (in particular, in the form of the multi-objective decision models) to enable invention and selection of the best alternatives and quality improvement
- ***Quality models are considered to be heuristics for setting and controlling the course of design***
- ***The design process is evolutionary*** and it basically **consists of**:
  - constructing the tentative quality models,
  - using them for constructing, improving and selecting of the tentative solutions,
  - analysing and estimating them directly and through analysis of the resulting solutions,
  - improving the models, and using them again to get improved solutions, etc.

92

Quality-driven Design: Limiting the design subjectivity

- **One of the main aims** of using the well-defined quality models in design is:

  Limiting the scope of subjective design decision making and ***enlarging the scope of reasoning-based decision making with clear and well-defined rational procedures*** which can be ***computerized***

- Too much subjectivity in design may result in solutions that either do not solve the actual real-life problem or do not do it in a satisfactory manner

- Limiting the design subjectivity in an appropriate manner, when enabling the creativity exploitation at the same time, *is necessary to arrive at the high-quality designs*

- The **main means for limiting the design subjectivity** is the ***design space exploration (DSE) with usage of the well-structured quality models***

93

Quality-driven Design: Limiting the design subjectivity

- **Exploration** of the abstract models of the required quality and more concrete solutions obtained with these models:
  - *gives much and more objective information* on the design problem, its possible and preferred solutions, and various models used in this process
  - *enhances exploitation of the designer's imagination, creativity, knowledge and experience*
- **Other important means for limiting the design subjectivity** and for **increasing quality** this way include:
  - appropriately organised team-work
  - benchmarking and comparison with both own previous designs and designs of competition
  - design analysis and validation
  - design reuse
  - government and branch **regulations and standards**

94

Quality-driven Design: Government regulations and standards

- ***Adequate government and branch regulations and standards are of primary importance for bringing into effect the green systems and green economy***
- Regulations and standards specify what is allowed or standard, and what is not
- They constitute general constraints for the industry and system designers that have to be satisfied by their designs, products and services
- Of course, particular systemic solutions satisfying these general constraints can still be very different, better or worse for the environment, but ***all systemic solutions have to satisfy the minimum required by the regulations and standards***
- Remember that the decisions made by companies and governments that caused the environmental destruction were mainly driven by short-term profit, without accounting for long-term consequences
- It would be naïve to expect that all companies and individuals will suddenly become environment-friendly without adequate regulations pressing them to do so

95

Quality-driven Design - Design requirements

- The general model of the required system's quality is represented by the ***system (design) requirements***
- System requirements can only be treated as *a non-perfect and tentative model of the required quality*
- Requirements and solutions obtained with their use are *subject to design and change*
- They should be confronted with the actual up-to-date needs many times during the design process, and replaced or modified, if necessary
- Design requirements model the design problem at a hand through *imposition of constraints and objectives in relation to the acceptable or preferred problem solutions*
- It is possible to distinguish **three sorts of requirements:**
  - ***functional***,
  - ***structural***, and
  - ***parametric***

96

Quality-driven Design - Design requirements

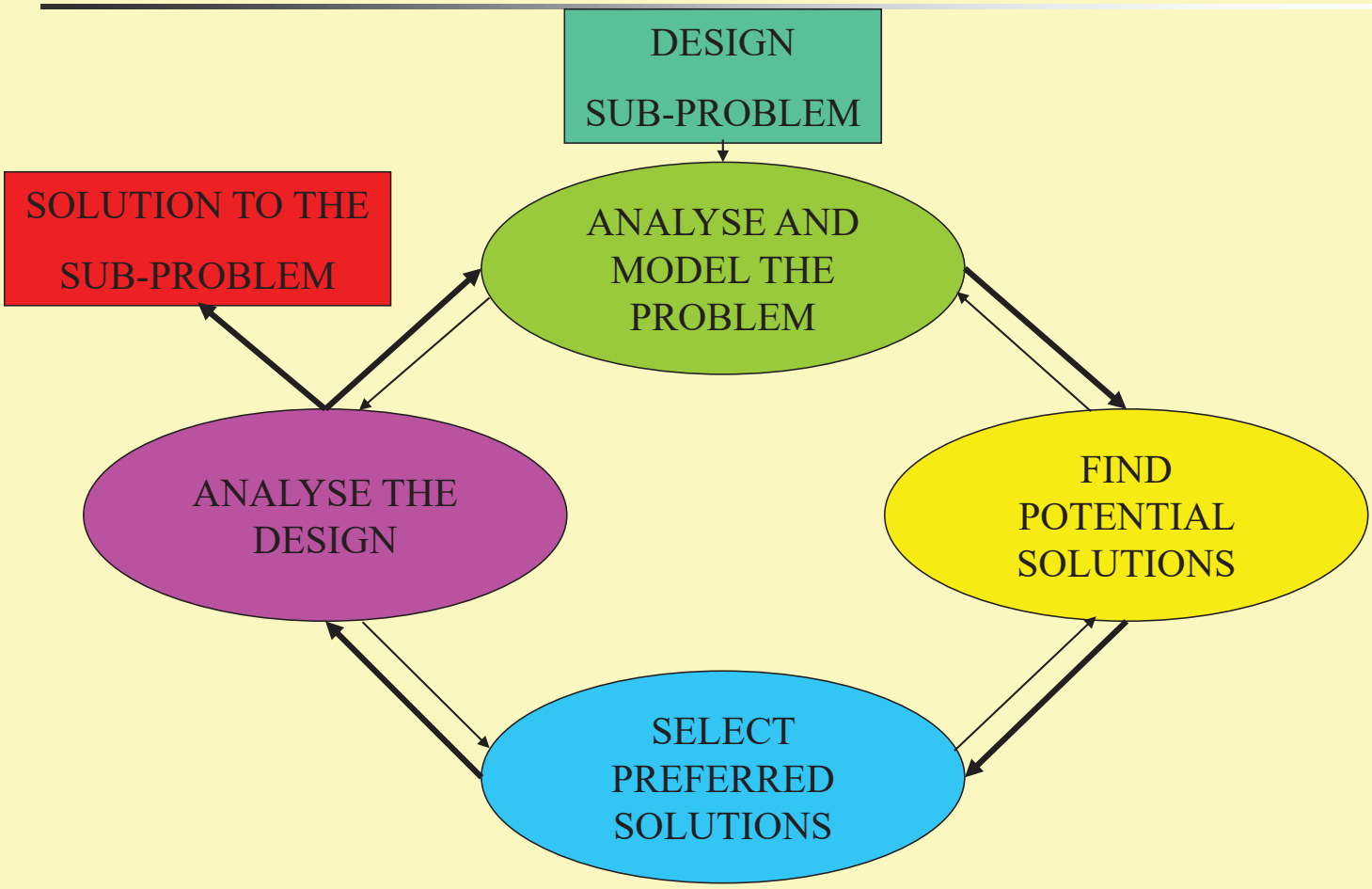
- All the three sorts of **requirements impose *limits on the structure of a required solution***, but they do it in different ways
- The ***structural requirements*** define the acceptable or preferred solution structures directly, by limiting them to a certain class or imposing a preference relation on them
- The ***parametric requirements*** define the structures indirectly, by requiring that the structure has such physical, economic or other properties (described by values of some parameters) as fulfil given constraints and satisfy stated objectives
- The ***functional requirements*** also define the structures indirectly, by requiring the structure to expose a certain externally observable behaviour that realizes the required behaviour

Quality-driven design space exploration (DSE)

- ***System design is an evolutionary quality engineering process*** in which the concepts of analysing and modelling problems, proposing their solutions, analysing and testing the proposals, learning and adapting are very important
- It **starts** with an ***abstract,*** and possibly *incomplete, imprecise,* and *contradictory*, ***initial quality model*** (initial requirements)
- It tries to **transform** the initial model into a ***concrete, precise, complete, coherent and directly implementable final quality model***
- Usually, the initial abstract model mostly involves some *behavioural and parametric characteristics* and to a lesser extend the structure definition
- The **final model** defines the ***system's structure explicitly***
- This structure supports the system's required behaviour and satisfies the parametric requirements

98

Generic model of the quality-driven design space exploration

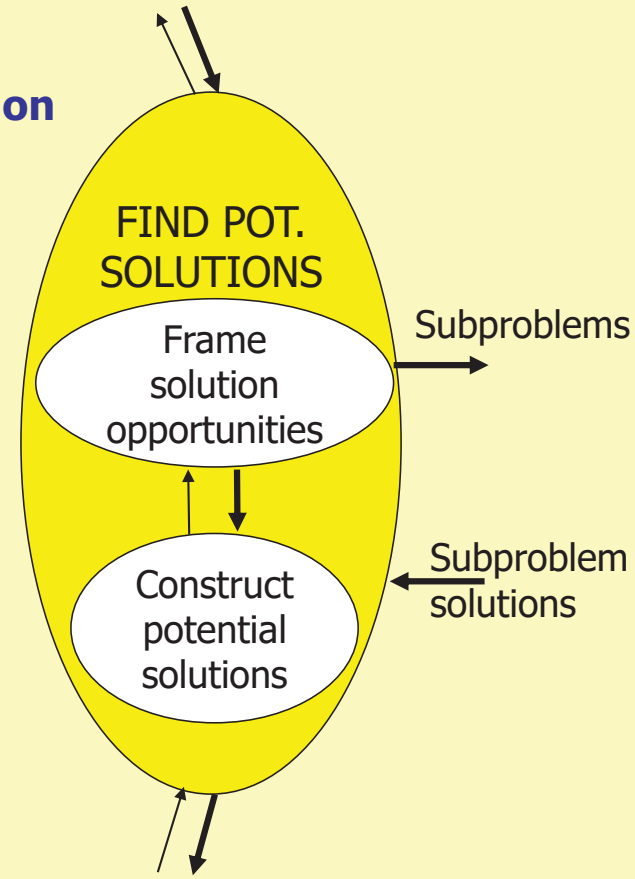


99

Generic model of the quality-driven design space exploration

- The **quality-driven design space exploration** basically consists of the alternating phases of:
  - ***exploration of the space of abstract models of the required quality***

    and
  - ***exploration of the space of the more concrete issue's solutions*** obtained with the chosen quality models



100

Quality-driven design space exploration

- In result of the design space exploration, the considered system is defined as an appropriate *decomposition into a network of sub-systems*
- Each sub-system solves a certain sub-problem
- All *sub-systems cooperating together solve the system design problem* by exposing the external *aggregate behaviour and characteristics* which *match the required behaviour and characteristics*
- The design process breaks down *a complex system* defined in *abstract and non-precise terms* into *a structure of cooperating sub-systems* defined in *more concrete and precise terms*, which are in turn further broken down to the ***simpler sub-systems that can be directly implemented with the elements and sub-systems at the designer's disposal***

101

Conclusion

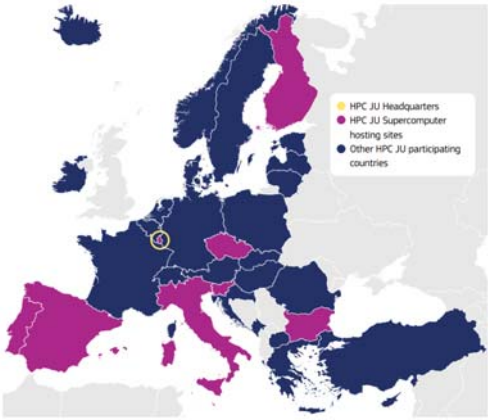
- Systemic drawbacks of the traditional economy and cumulation of bad decisions made by numerous governments and companies without accounting for long-term consequences resulted in the **huge global environmental disaster**
- To recover from the environmental disaster and further develop:
  - *a model of a well regulated and controlled effective and efficient system should be applied to all kinds of systems, collaboration chains and related flows*
  - *modern CPS and IoT technologies should be used to much better control and optimize the social, physical and life systems than till now*
  - *methodologies of circular regenerative economy and quality-driven design should be used to design the systems*
- Innovations exploiting modern CPS and IoT technologies, circular regenerative economy and quality-driven design can significantly improve systems used by us or that we are part of
- In this CPS&IoT Summer School you will have a unique occasion to be informed on and to discuss the most recent European R&D developments in CPS and IoT

102


European
Processor
Initiative
epi
European Processor Initiative
Framework Partnership Agreement In European Low-power Microprocessor Technologies

EU EXASCALE HPC STRATEGY

- March 2017, Rome: EC launched the ***EuroHPC declaration***
- November 2018, EuroHPC Joint Undertaking, a 1 billion Euro joint initiative between the EU and European countries to develop a World Class Supercomputing Ecosystem in Europe
- 13.7.2021.: EU Council established new EuroHPC JU
  - the 27 Member States, 6 other countries, 2 Private Members
  - €7 billion investment



EUROHPC JU AMBITIOUS MISSION

- **Supercomputers**
  - reaching the next frontier of high-performance computing: the acquisition of exascale supercomputers
- **Interconnectivity**
  - interconnection through terabit networks of this supercomputing infrastructure, as well as in allowing access from the cloud to a large number of public and private users from anywhere in Europe
- **Applications for life**
  - further development of novel scientific and industrial applications
- **Skills and engagement with business**
  - increased investment in skills, education and training in the use of HPC, co-investment with industry in the acquisition of dedicated systems and in the development of large-scale industrial applications, creation of HPC Centres of Excellence
- **Technology activities**
  - the development of high-end European technologies, for example in the European Processor Initiative (EPI)

DRIVERS OF THE EPI PROPOSAL

Societal challenges

- Climate change
- Cybersecurity
- Increasing energy needs
- Intensifying global competition
- Aging population
- Sovereignty (data, economical, embargo)

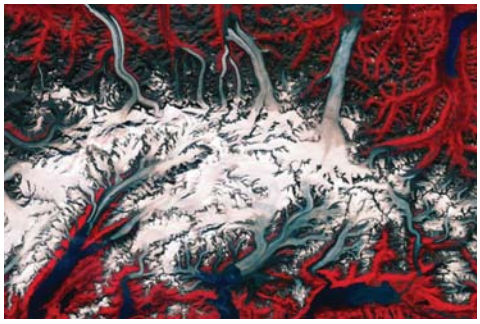


Image: https://www.compbiomed.eu/services/software-hub/

European Processor Initiative

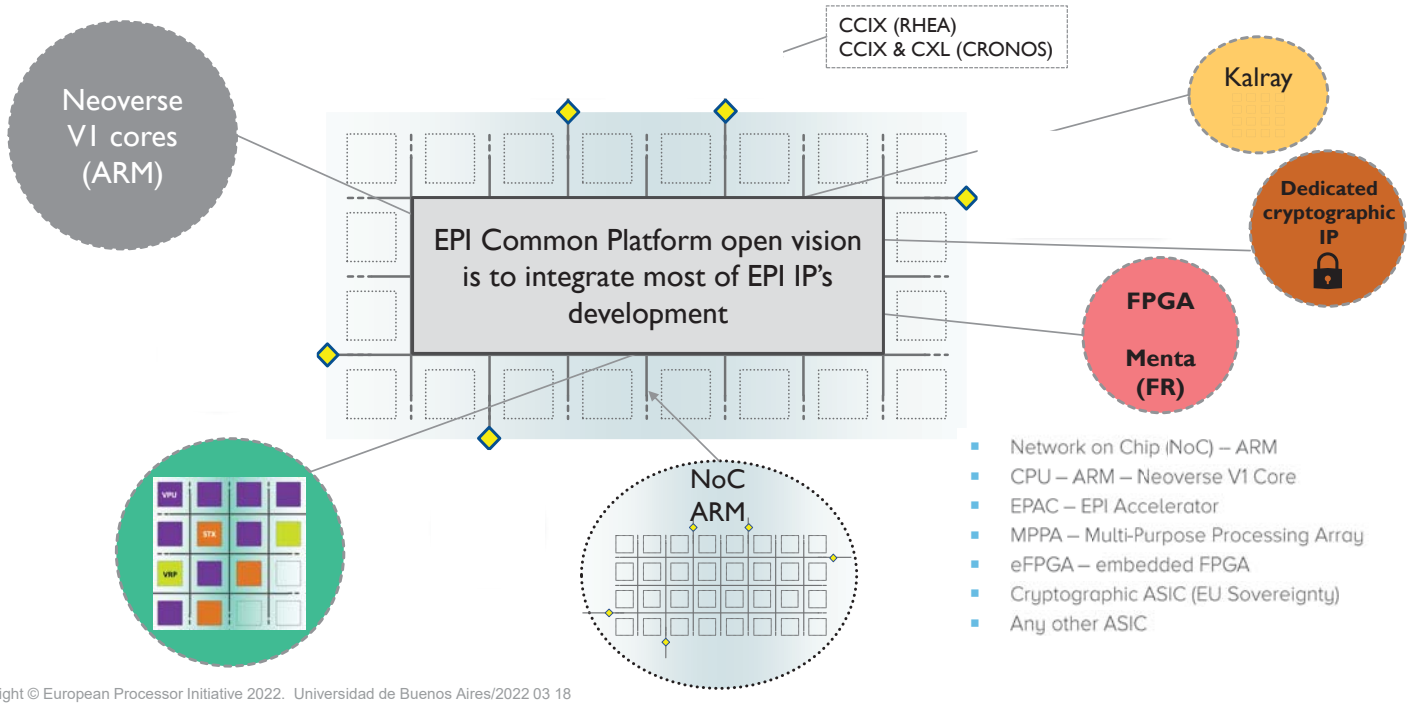
EPI PARTNERS

EPI OBJECTIVES

- Overall: **Develop a complete EU designed high-end microprocessor, addressing Supercomputing and edge-HPC segments**
- **Short-term objective**
  - supply the EU-designed microprocessor to empower the future Exascale machines
- **Long-term objective**
  - Europe needs a sovereign access to high-performance, low-power microprocessors, from IP to products
  - contribute to the emergence of Risc-V as an open alternative to proprietary chip standards
  - enable the emergence of an EU high-end processor industry (Arm & Risc-V based) that will have long term benefits

THE EPI TECHNOLOGY: COMMON PLATFORM

GPP AND COMMON ARCHITECTURE

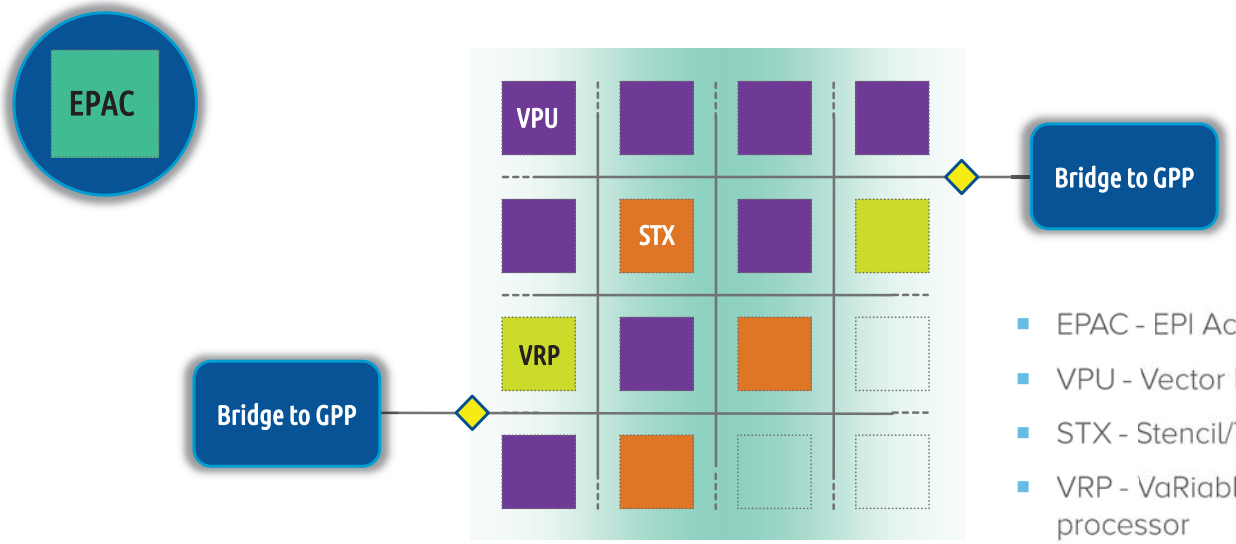


THE EPI TECHNOLOGY: ACCELERATORS

TOP10 (GREEN) OVER THE LAST 10 YEARS

	2009 – Nov.	2014 – Nov.	2020 – Nov.	2021 – Nov.
CPU only	9	5	2	0
CPU + ACC.	1	5	8	10

EPAC – RISC-V ACCELERATOR FOUNDATIONS

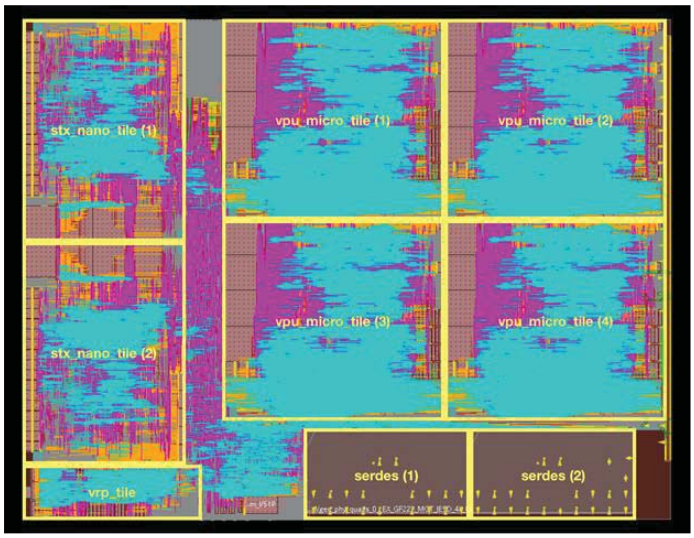


- EPAC - EPI Accelerator
- VPU - Vector Processing Unit
- STX - Stencil/Tensor accelerator
- VRP - VaRiable Precision co-processor

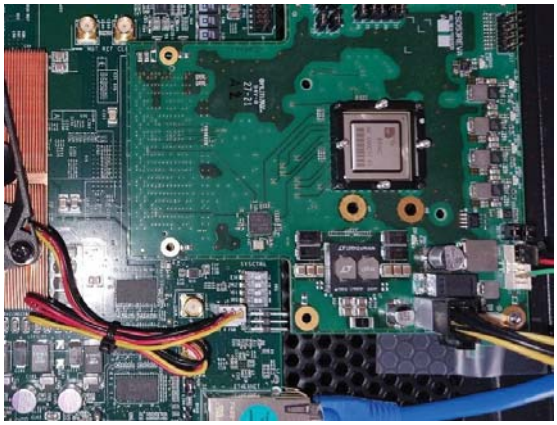
EPAC1.0

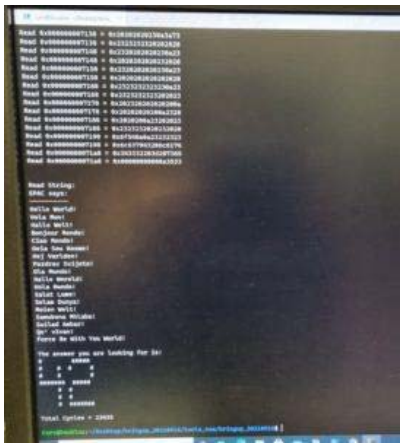
- EPAC test chip combines several accelerator technologies specialized for different application areas:
  - four vector processing micro-tiles (VPU) composed of an Avispado RISC-V core designed by SemiDynamics and a vector processing unit designed by Barcelona Supercomputing Center and the University of Zagreb
  - Home Node and L2 cache, designed respectively by Chalmers and FORTH
  - two additional accelerators:
    - the Stencil and Tensor accelerator (STX) designed by Fraunhofer IIS, ITWM and ETH Zürich
    - variable precision processor (VRP) by CEA LIST
  - All accelerators on the chip are connected with a very high-speed network on chip and SERDES technology from EXTOLL.

EPAC 1.0

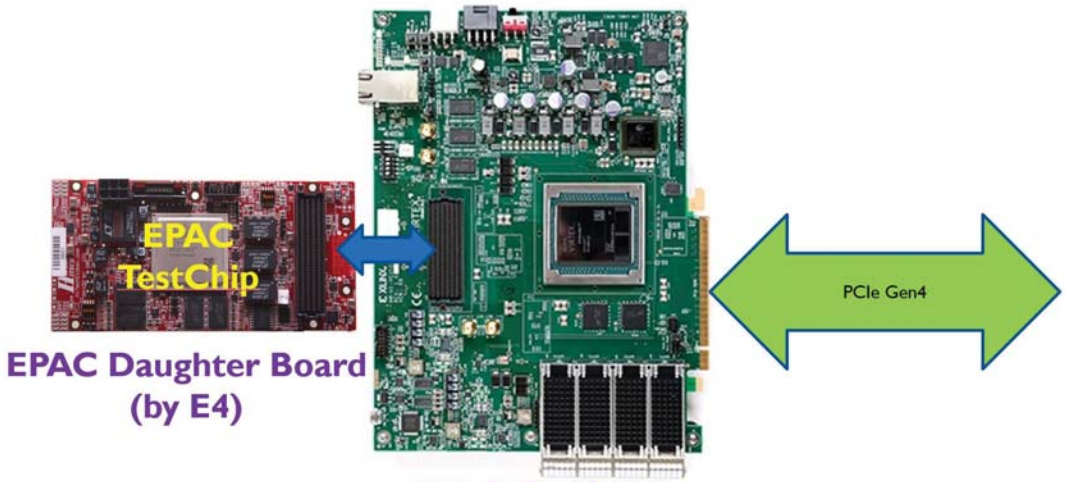


AND THE "HELLO WORLD" IN EU LANGUAGES





EPAC BOARD SUPPORT FOR EARLY TECHNOLOGY ADOPTERS



EPAC Daughter Board (by E4)

EPAC Carrier Board

Virtex UltraScale+ HBM VCU128 FPGA

19

EPI PHASE2

STRATEGIC GRANT AGREEMENT 2

EPI2 ROADMAP

EXPECTED OUTCOMES

- We expect, at the end of our second phase, to have
  - The first generation of our GPP validated & exposed to customers
  - The second generation of our GPP designed
  - Several flavours and versions of Risc-V accelerators developed and tested, for instance EPAC 1.5 & 2.0 test chips
- and, as indirect outcomes,
  - developed and validated systems that integrate that GPP into data centres
  - contributed to the emergence of Risc-V as an open alternative to proprietary chip standards
  - enabled the emergence of an EU high-end processor industry (Arm & Risc-V based) that will have long term benefits

EPI CONCLUSION

- Use of HPC and AI is cornerstone of successful address of societal and global challenges
- Future science, technologies and applications require processing of vast amount of data and there is a large need for efficient HPC
- HPC provides needed competitiveness for industry and society
- The expertise for developing high-end and complex processing units in Europe, after decades of dis-investment
- The European Processor Initiative aims to provide an EU HPC processor, accelerators and system/application design for exascale HPC systems in Europe and around the globe

www.european-processor-initiative.eu

@EuProcessor

European Processor Initiative

European Processor Initiative

UNIVERSITY OF ZAGREB

- Founded in 1669
- 29 Faculties, 3 Academies
- approx. 70,000 students
- 167 Undergraduate Programs
- 21 Integrated Programs
- 182 Graduate Programs
- 66 Doctoral Programs
- 146 Postgraduate Specialist Programs

- Studnets enrolled in the 1st year of study:11,500
- PhD degrees:400 / year

FACULTY OF ELECTRICAL ENGINEERING AND COMPUTING (FER)

- ~ 650 employees
- 12 departments
- 4000 students
- 450 PhD students
- Bachelor & Master Study Programs:
  - Electrical Engineering and Information Technology
  - Information and Communication Technology
  - Computing
- PhD Programs:
  - Electrical Engineering
  - Computing

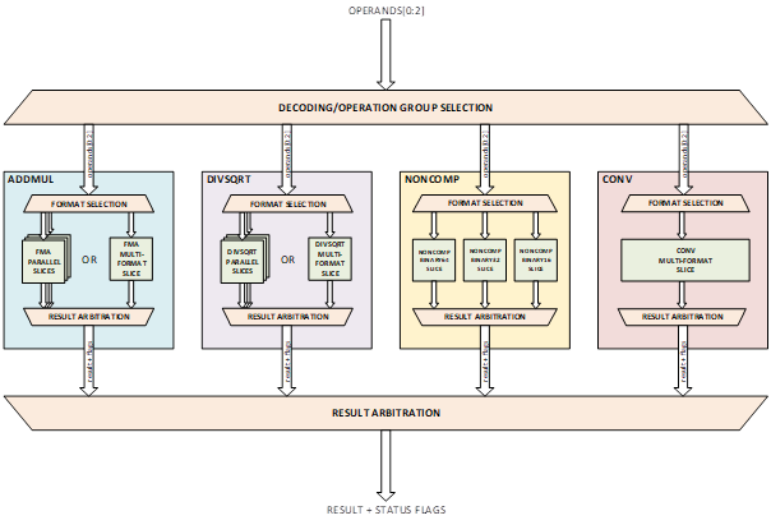
HPC ARCHITECTURES AND APPLICATIONS RESEARCH CENTER @ FER

- GPP/Accelerator architecture
- FAUST
  - Risc-V VE pipelined vector FPU
  - Implemented in EPAC VPU
- Risc-V based accelerators
  - SA
- Imaging apps/optimizations
  - Bolt65
  - Jaguar

FAUST – RISC-V PIPELINED VECTOR FPU

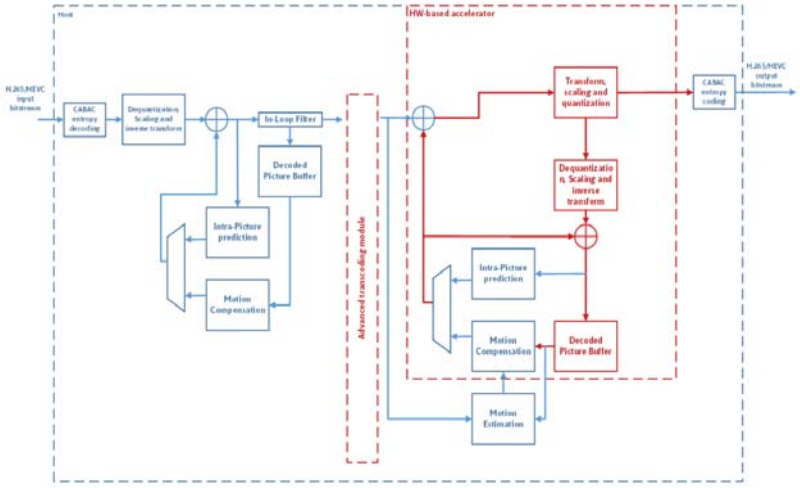
- **Compliance with IEEE 754-2019 Standard**
  - Only minor deviation
- **Supported all floating-point operations defined in RISC-V ISA**
  - RVV 1.0: all operations except reciprocal estimate operations
- **Floating-point formats**
  - binary16 (half precision format)
  - binary32 (single precision format)
  - binary64 (double precision format)
- **Rounding modes**
  - Round to nearest, ties to even
  - Round to nearest, ties to max magnitude
  - Round up
  - Round down
  - Round towards zero
- **Supported all IEEE 754 status flags**
  - Invalid operation
  - Divide by zero
  - Overflow
  - Underflow
  - Inexact
- **Supported subnormal numbers**
- **Support for vector unit integration**
  - Masking support
  - Handshake interface for data flow control to and from the floating-point unit

Parameterized design: configurable architecture and pipeline stages



BOLT65

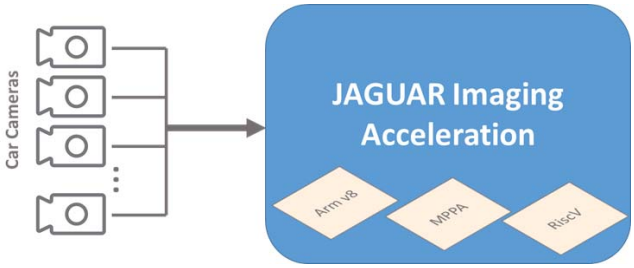
- Bolt65 is a HEVC/H.265 hardware/software suite
  - focus on Just-in-Time video processing
    - constrainted by processing time
  - clean room project created on FER - UNIZG
  - consists of encoder, decoder, transcoder
- Portability
  - written in C++
  - compiled and executed on ARM and x86
  - can be compiled for Linux and Windows
  - no external libraries used
- Optimizations
  - Partially optimized for AVX, SVE, NEON
  - Custom accelerator and GPU support



European Processor Initiative

JAGUAR

- Jaguar Imaging and AI Framework
  - 8/12-bit JPEG image codec
  - SW/HW accelerator kernels



REAL CAR DRIVING DETECTION RESULTS

Blind spot detection area

Blind spot alert

Detected moving object

THANK YOU FOR YOUR ATTENTION

COMP4DRONES

SS-CPS&IoT2022

Accelerator-Rich FPGA Architecture Exploration via a Programmable and Reconfigurable Overlay

Gianluca Bellochi[1], Daniel Madroñal[2], Alessandro Capotondi[1], Andrea Marongiu[1], Francesca Palumbo[2]

[1]Università degli Studi di Modena e Reggio Emilia
[2]Università degli Studi di Sassari

ECSEL Joint Undertaking
Electronic Components and Systems for European Leadership

This project has received funding from the ECSEL Joint Undertaking (JU) under grant agreement No 826610. The JU receives support from the European Union's Horizon 2020 research and innovation programme and Spain, Austria, Belgium, Czech Republic, France, Italy, Latvia, Netherlands.


COMP4DRONES will provide a **framework** of key enabling technologies for **safe and autonomous drones** that will leverage their **customization and modularity** for civilian services
COMP4DRONES
ECSEL JU GA No 826610
Website: ***comp4drones.eu***

AGENDA

1 Introduction

2 Methodology overview

3 MDC tool

4 OODK overlay

5 COMP4DRONES use case

6 Conclusions

AGENDA

1 Introduction

2 Methodology overview

3 MDC tool

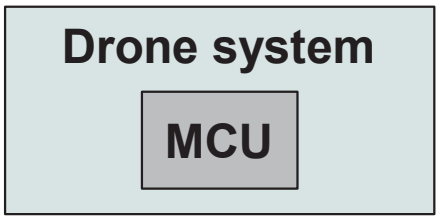
4 OODK overlay

5 COMP4DRONES use case

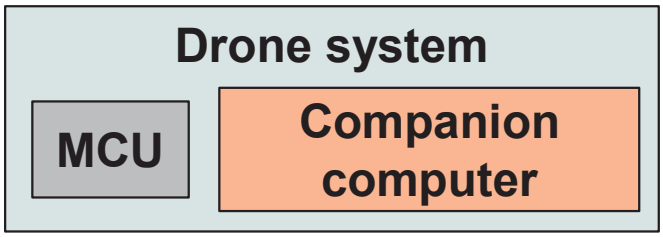
6 Conclusions

Introduction

Accelerator-rich paradigm



- The *"classic" set-up* comprises a **micro-controller unit (MCU)** that is used for control and actuation

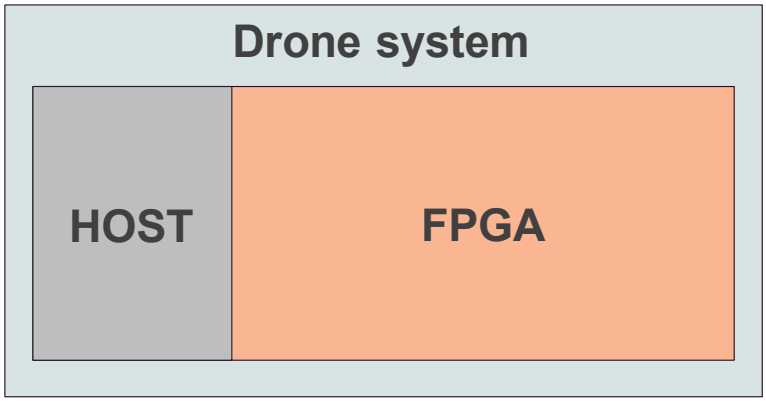


- *Current paradigm* envisions coupling a **MCU** with a **companion computer**
- **Heterogeneous solutions** (Nvidia Tegra TX2, Xilinx Zynq US+, ..) are increasingly used

5

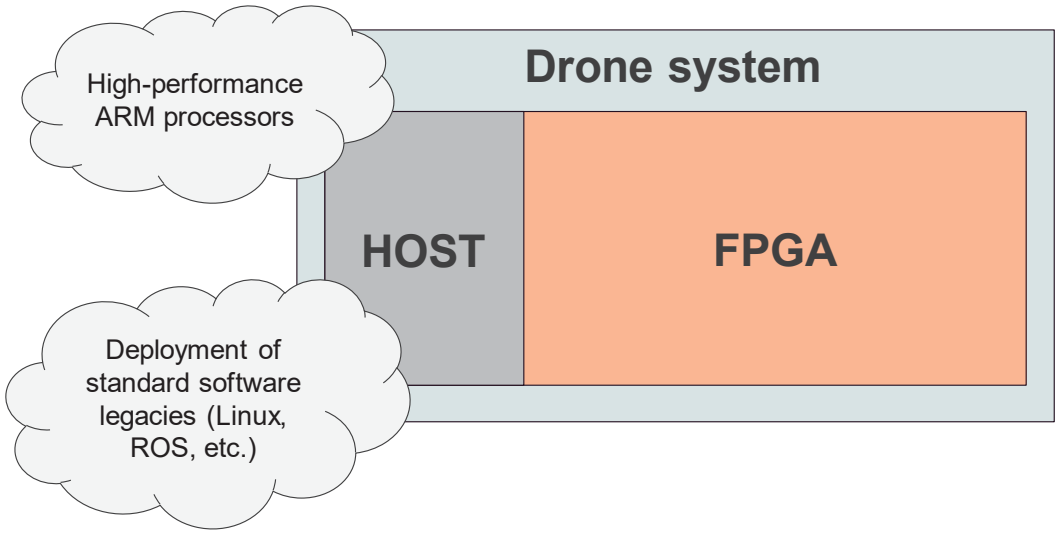
Introduction

Accelerator-rich paradigm



Introduction

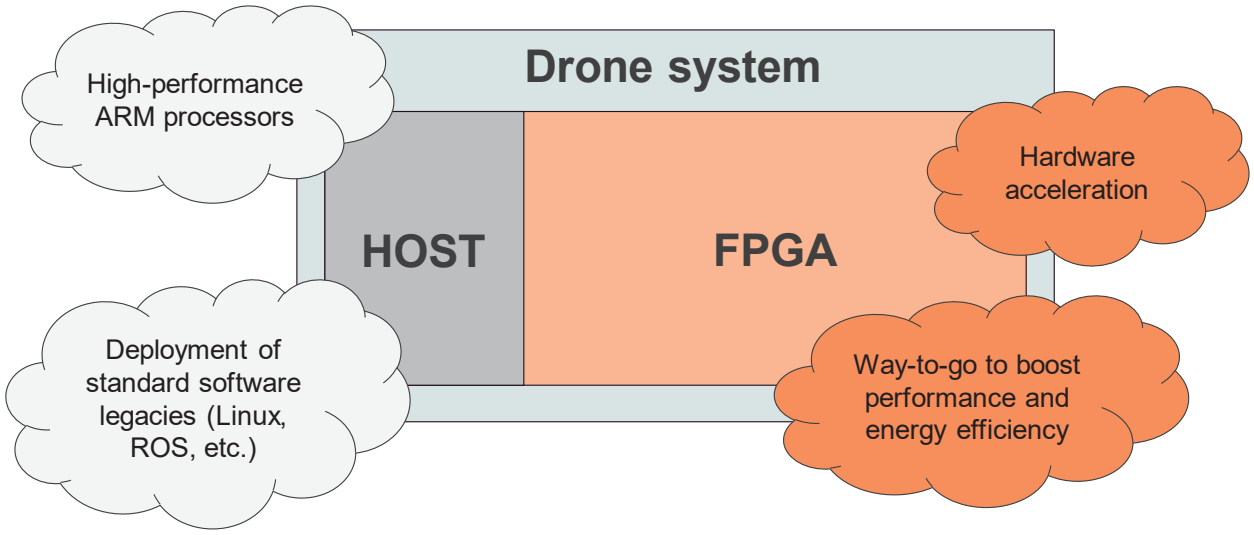
Accelerator-rich paradigm



7

Introduction

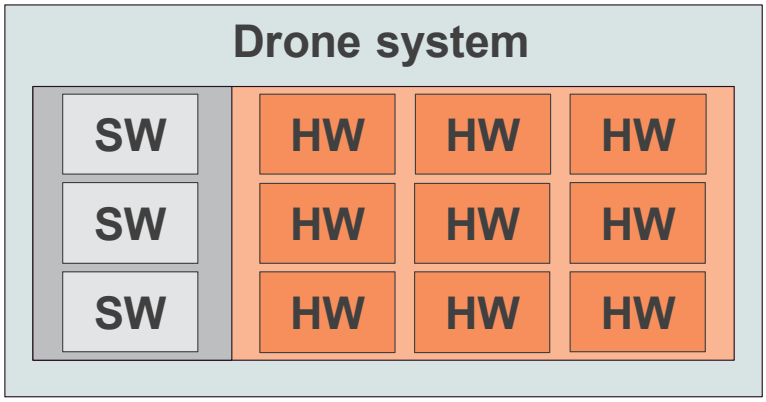
Accelerator-rich paradigm



8

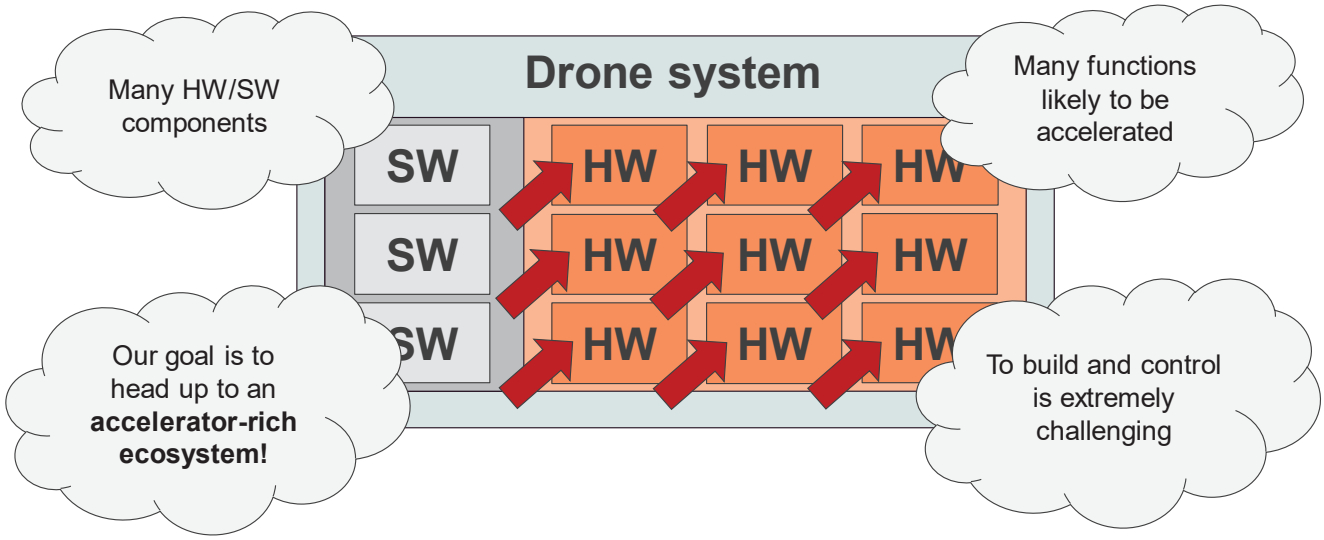
Introduction

Accelerator-rich paradigm



Introduction

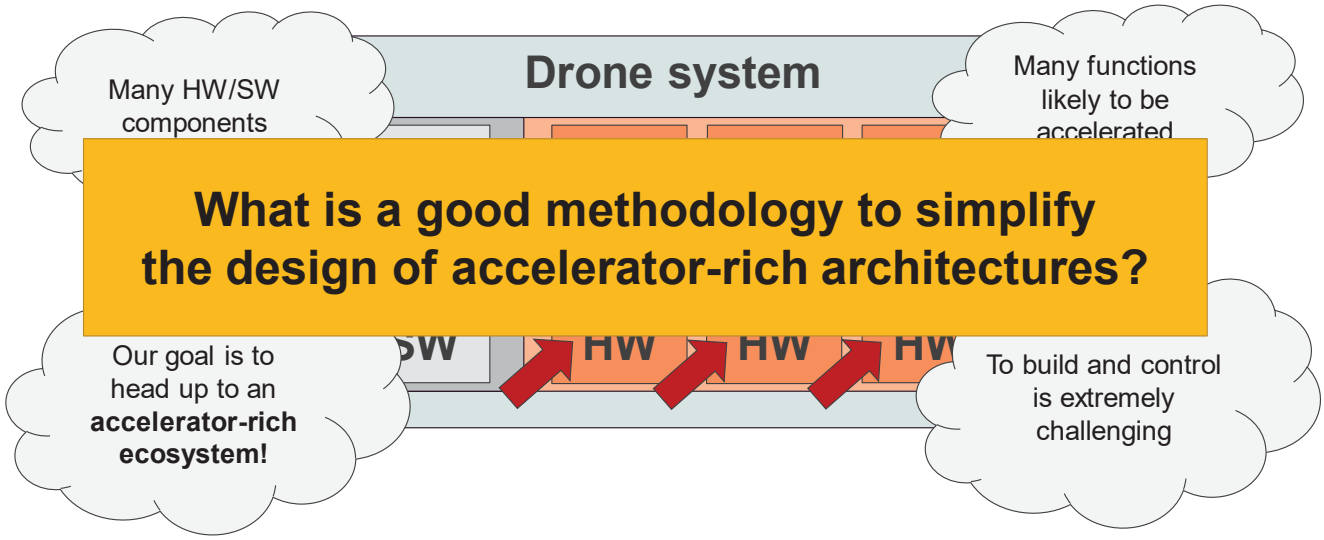
Accelerator-rich paradigm



10

Introduction

Accelerator-rich paradigm



11

Introduction

Motivation

What has to be simplified?

- *System-Level Design*
  - Build and evaluate accelerator-rich systems
    - Expensive
    - Time-consuming

Introduction

Motivation

COMP4DRONES

What has to be simplified?

- *System-Level Design*
  - Build and evaluate accelerator-rich systems
    - Expensive
    - Time-consuming

- *Design Space Exploration (DSE)*
  - Key effects only manifest at system-level
  - User knobs:
    - System optimization
    - Accelerator optimization

Introduction

Motivation

What has to be simplified?

- *System-Level Design*
  - Build and evaluate accelerator-rich systems
    - Expensive
    - Time-consuming
- *Design Space Exploration (DSE)*
  - Key effects only manifest at system-level
  - User knobs:
    - System optimization
    - Accelerator optimization
- *Accelerator Design*
  - Multi-functionality support
  - Multi working-point support

Introduction

Structure of the presentation

Introduction

Structure of the presentation

Step 1:

Overview of the proposed methodology (How to build a whole FPGA-based system starting from a dataflow specification)

Introduction

Structure of the presentation

Step 1:

Overview of the proposed methodology (How to build a whole FPGA-based system starting from a dataflow specification)

Step 2:

Accelerator definition and generation (MDC workflow)

Introduction

Structure of the presentation

Step 1:

Overview of the proposed methodology (How to build a whole FPGA-based system starting from a dataflow specification)

Step 2:

Accelerator definition and generation (MDC workflow)

Step 3:

Overlay connection and usage from SW (OODK workflow)

AGENDA

1 Introduction

2 Methodology overview

3 MDC tool

4 OODK overlay

5 COMP4DRONES use case

6 Conclusions

Methodology overview

High-level outline

1) Dataflow specification 2) Datapath merging and wrapper generation 3) Build the system

Methodology overview

High-level outline

1) Dataflow specification 2) Datapath merging and wrapper generation 3) Build the system

Prerequisites

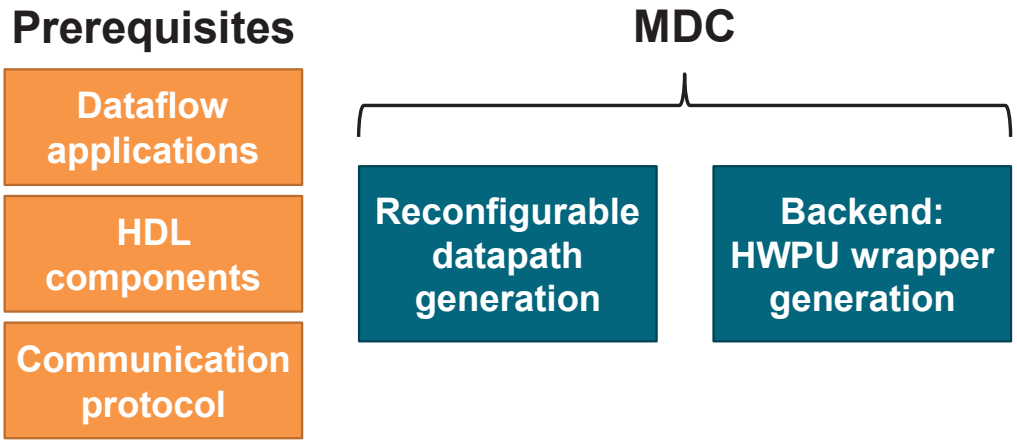
- **Dataflow applications**
- **HDL components**
- **Communication protocol**

21

Methodology overview

High-level outline

1) Dataflow specification 2) Datapath merging and wrapper generation 3) Build the system

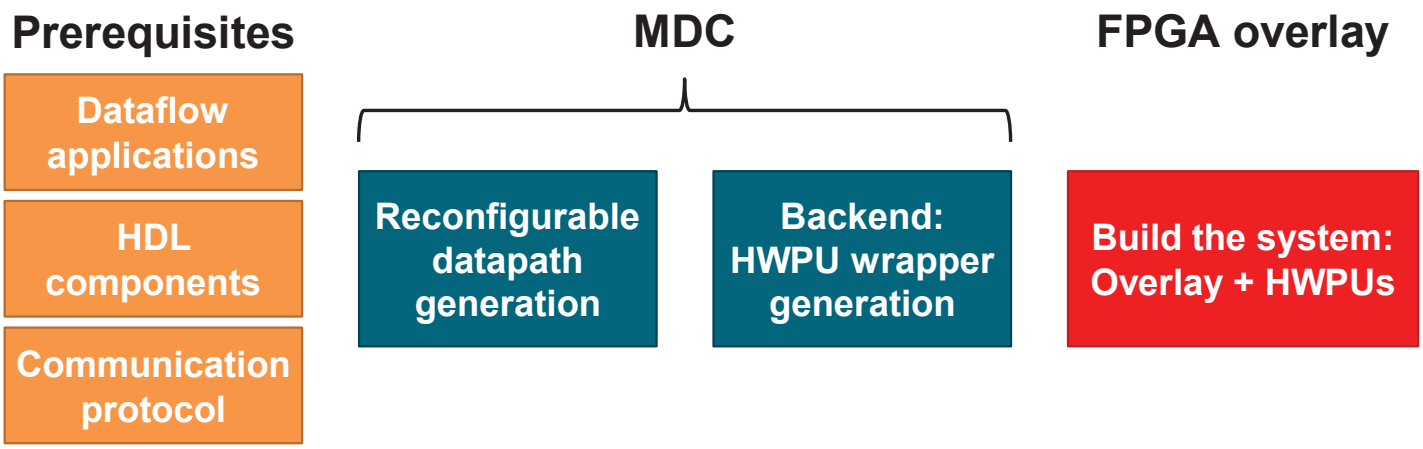


22

Methodology overview

High-level outline

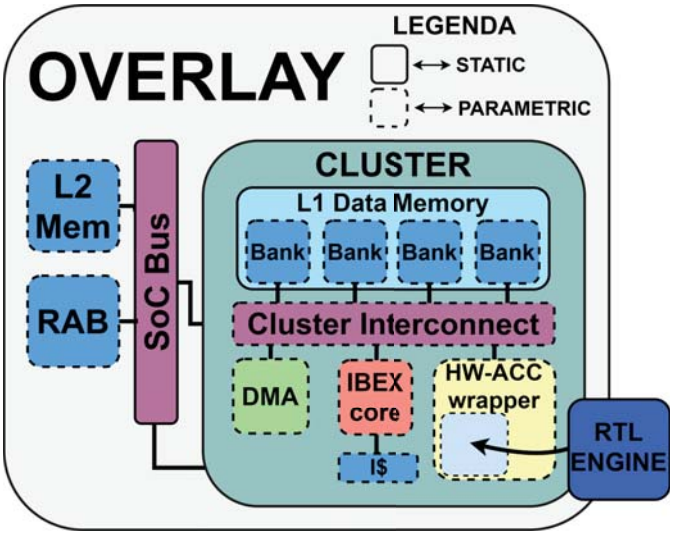
1) Dataflow specification 2) Datapath merging and wrapper generation 3) Build the system



23

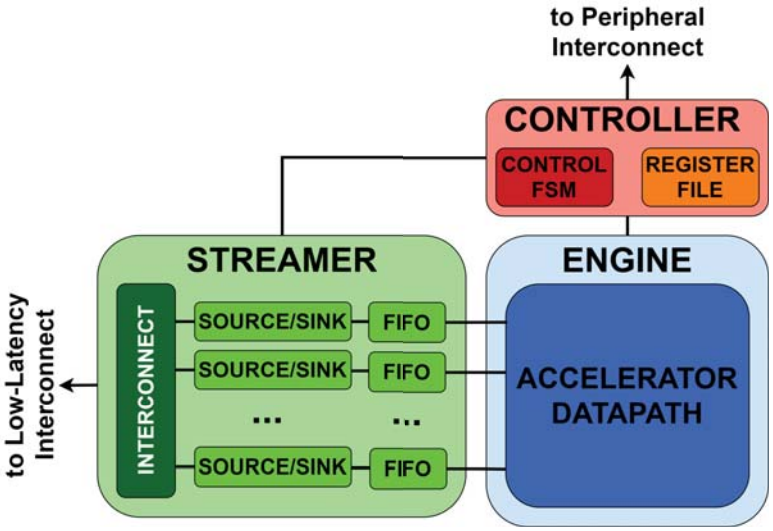
Methodology overview

FPGA overlay



Methodology overview

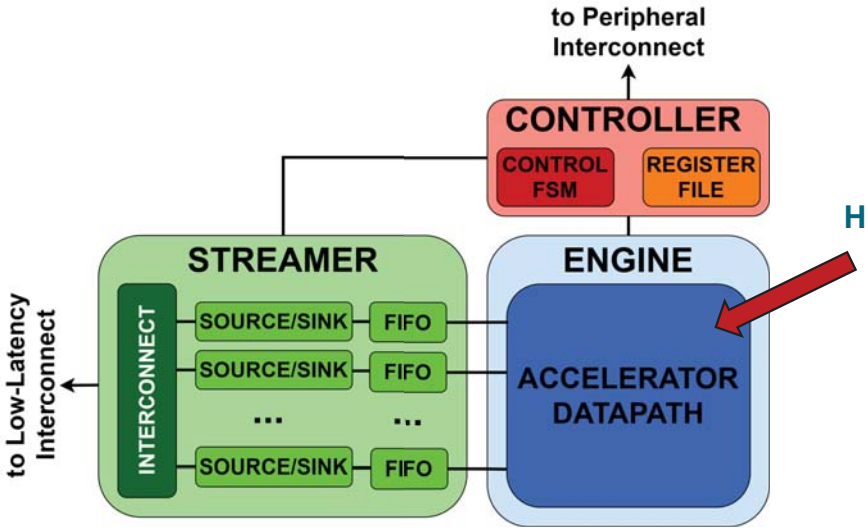
HWPU accelerator wrapper



25

Methodology overview

HWPU accelerator wrapper

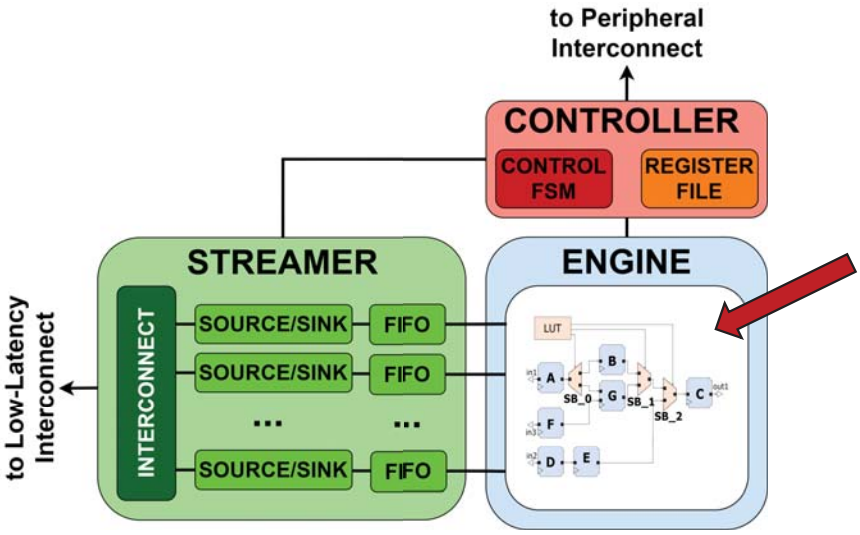


HW-mapped application
(MDC, HLS, etc.)

26

Methodology overview

HWPU accelerator wrapper



MDC-based reconfigurable application

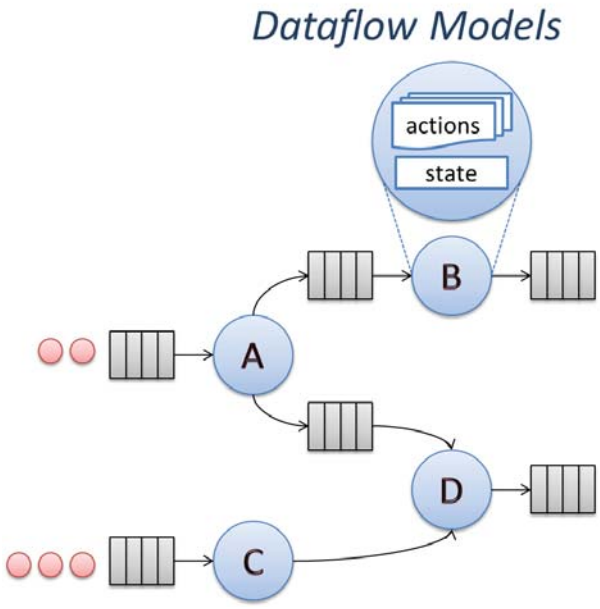
27

Methodology overview

App modeling

Methodology overview

App modeling

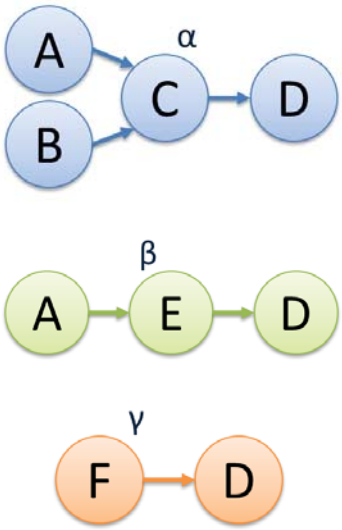


- Directed graph of actors (functional units)
- Actors exchange tokens (data packets) through dedicated channels

29

Methodology overview

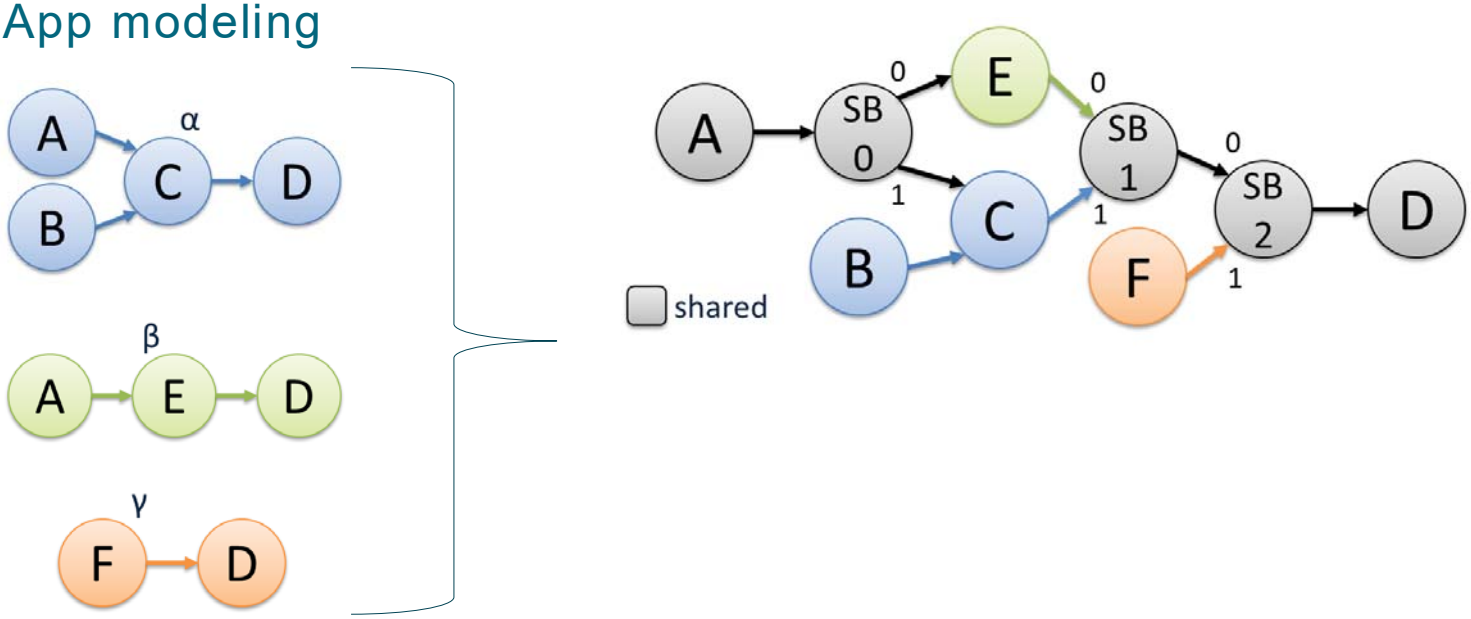
App modeling



30

Methodology overview

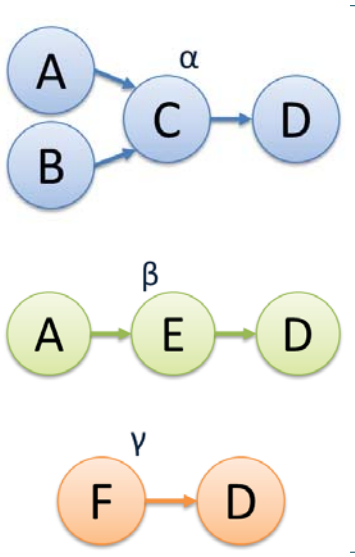
App modeling

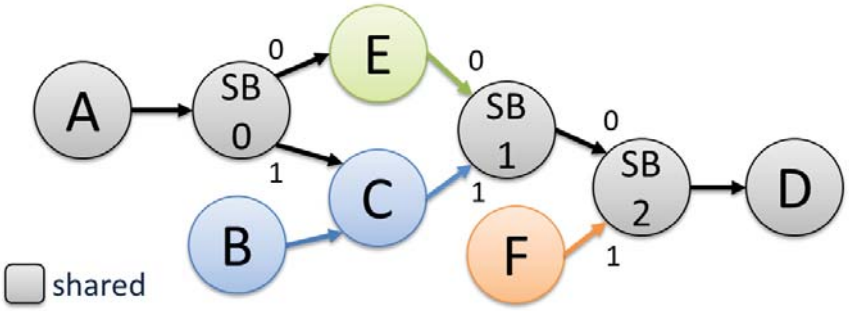


31

Methodology overview

App modeling



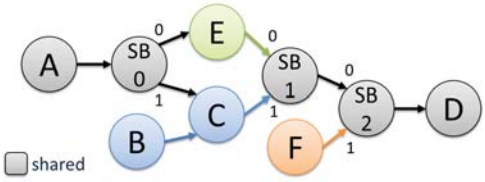


SB	0	1	2
α	1	1	0
β	0	0	0
γ	x	x	1

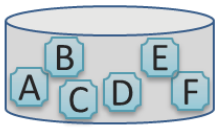
32

Methodology overview

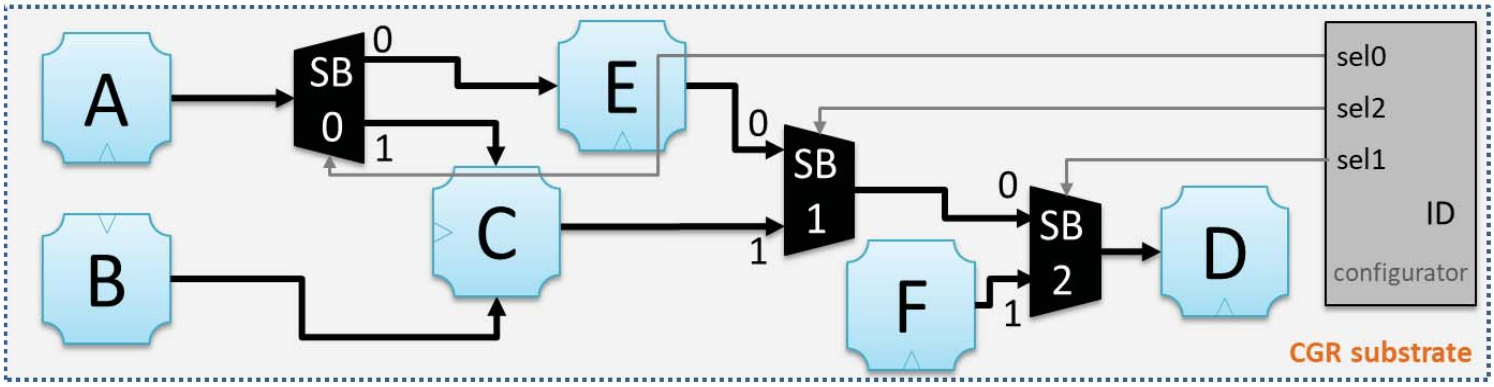
HW accelerator generation



HDL components library



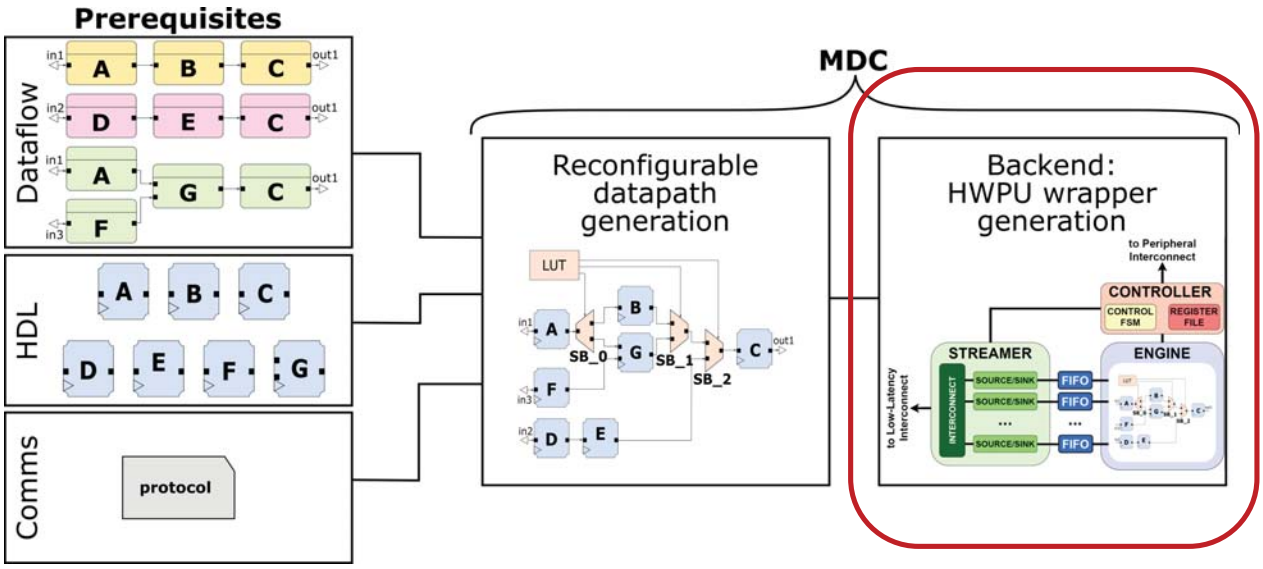
hardware communication protocol (XML)



33

Methodology overview

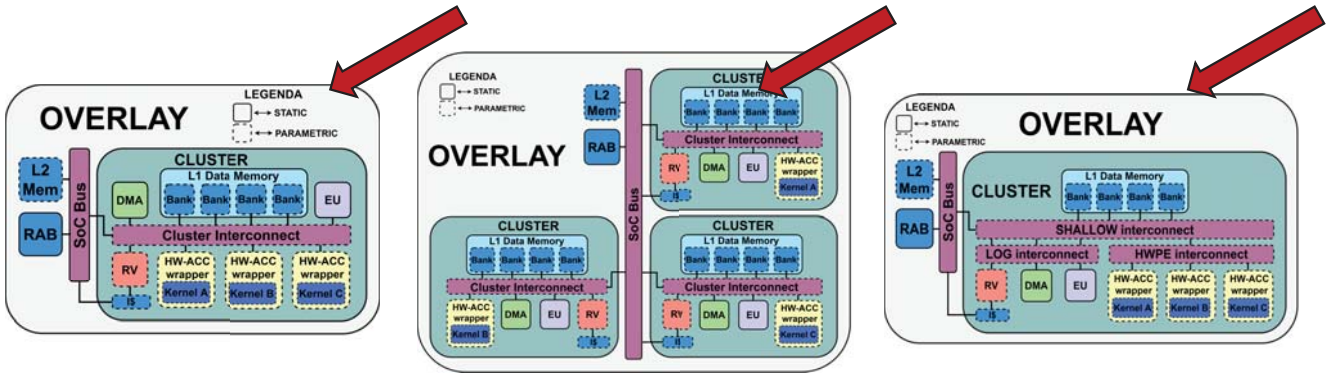
HW accelerator integration



Methodology overview

System generation

A subset of the generable accelerator-rich systems



Agile system-level design and exploration methodology

35

AGENDA

1 Introduction

2 Methodology overview

3 MDC tool

4 OODK overlay

5 COMP4DRONES use case

6 Conclusions

MDC

What application are we using in this tutorial?

MDC

What application are we using in this tutorial?

Edge detection using different kernels

INPUT IMAGE

MDC

What application are we using in this tutorial?

Edge detection using different kernels

INPUT IMAGE

SOBEL

MDC

What application are we using in this tutorial?

Edge detection using different kernels

INPUT IMAGE

SOBEL

ROBERTS

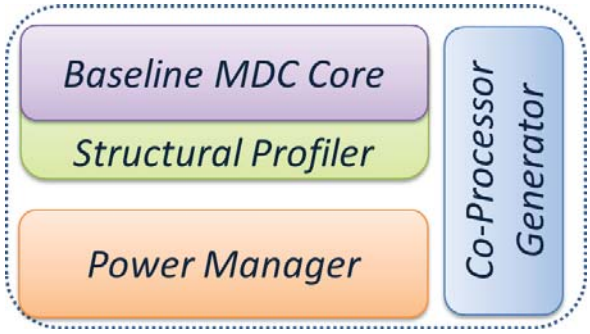
40

MDC

Multi-Dataflow Composer concepts

MDC

Multi-Dataflow Composer concepts

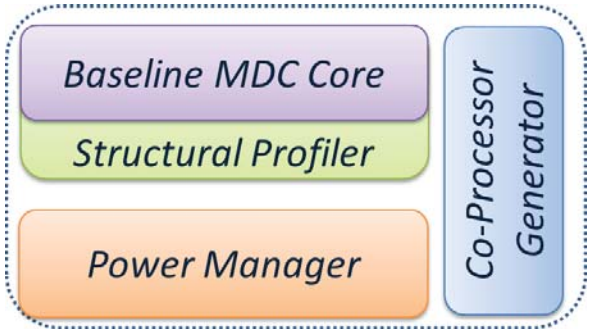


MDC *design suite:*
https://github.com/mdc-suite

42

MDC

Multi-Dataflow Composer concepts



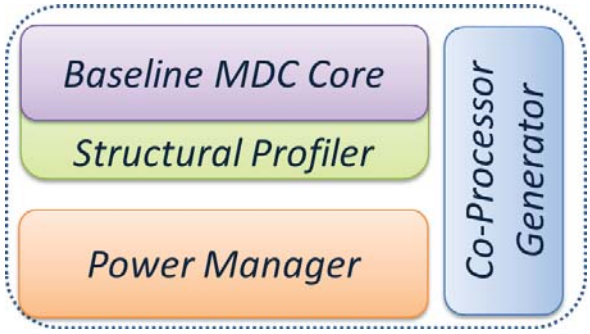
Baseline MDC Core: Datapath merging and CGR generation

MDC *design suite:*
https://github.com/mdc-suite

43

MDC

Multi-Dataflow Composer concepts



Baseline MDC Core: Datapath merging and CGR generation

Structural Profiler: DSE for optimal CGR composition

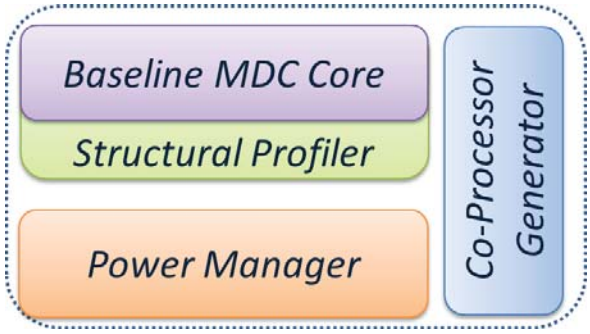
Power Manager: Clock and power gating by regions

MDC *design suite:*
https://github.com/mdc-suite

44

MDC

Multi-Dataflow Composer concepts



Baseline MDC Core: Datapath merging and CGR generation

Structural Profiler: DSE for optimal CGR composition

Power Manager: Clock and power gating by regions

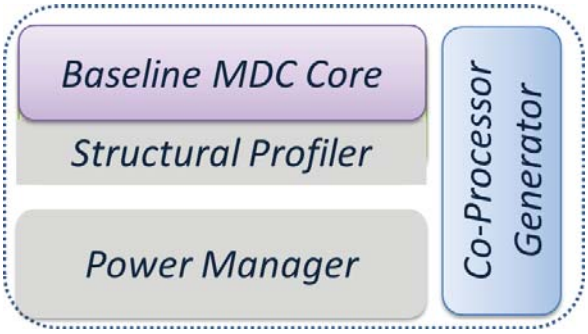
Co-Processor Generator: Wrapper to connect accelerator and processor

MDC *design suite:*
https://github.com/mdc-suite

45

MDC

Multi-Dataflow Composer concepts



Baseline MDC Core: Datapath merging and CGR generation

Structural Profiler: DSE for optimal CGR composition

Power Manager: Clock and power gating by regions

Co-Processor Generator: Wrapper to connect accelerator and processor

MDC *design suite:*
https://github.com/mdc-suite

Relevant for this tutorial

46

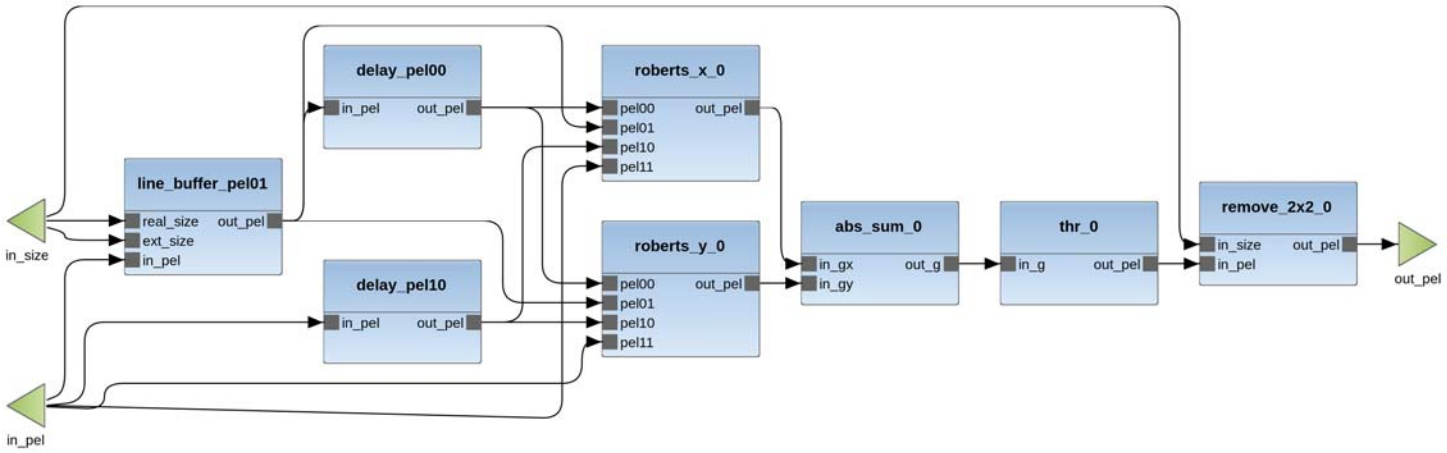
MDC
Modelling applications

47

MDC

Modelling applications: MDC interface

Network: Roberts

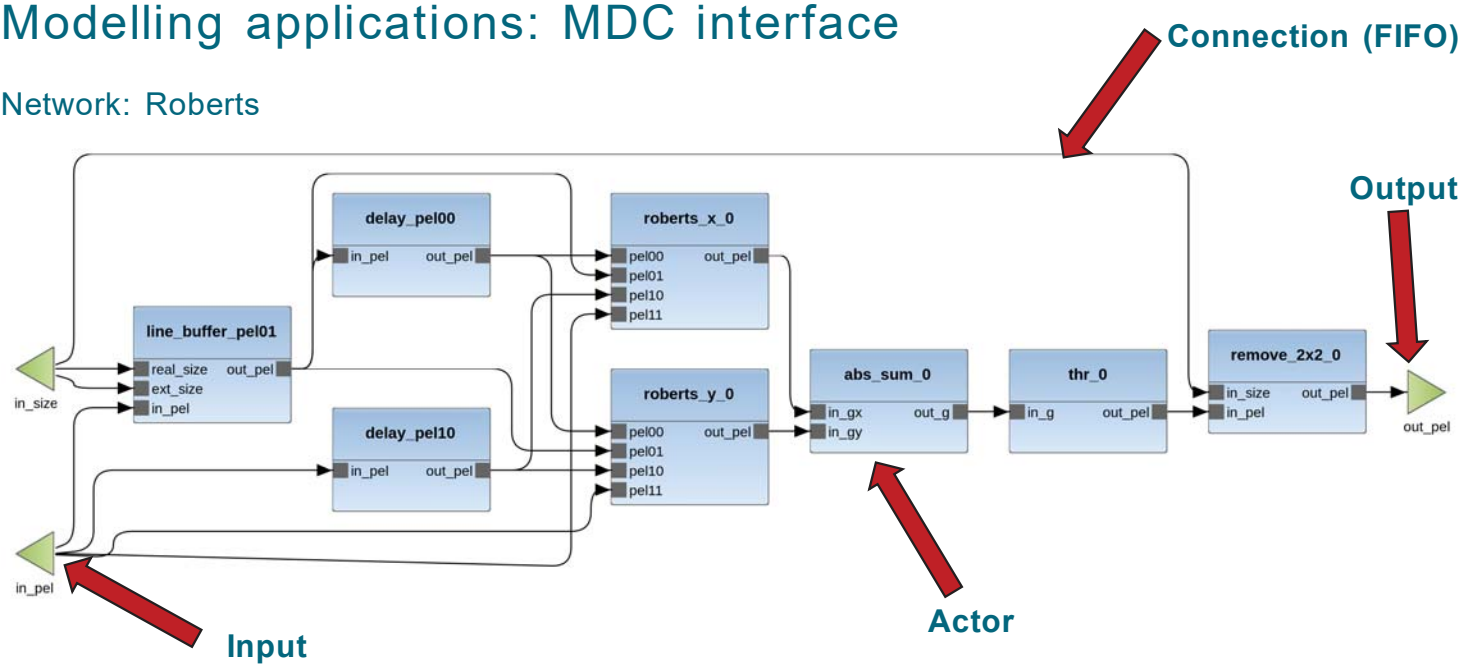


48

MDC

Modelling applications: MDC interface

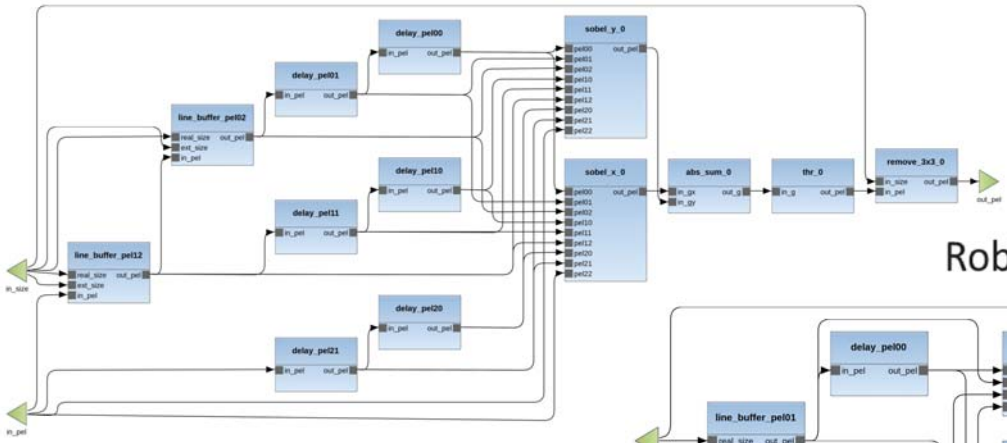
Network: Roberts



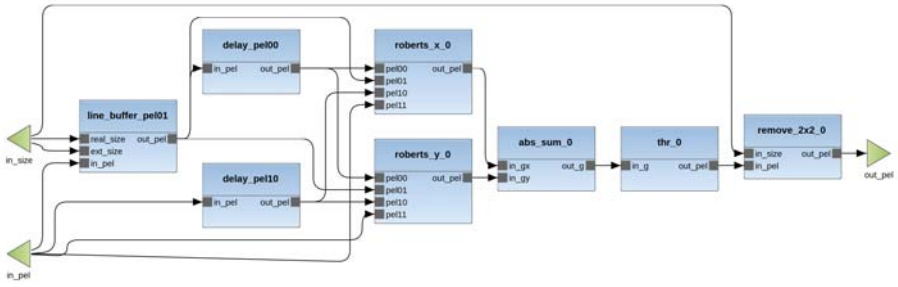
MDC

Modelling applications: MDC interface

Sobel dataflow



Roberts dataflow



MDC

Baseline MDC Core: app analysis

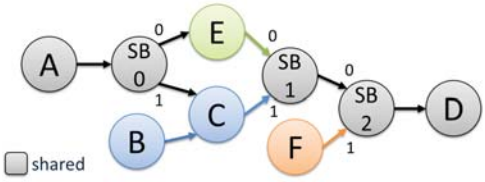
actor	Sobel	Roberts	NS	S
Forward2x2	0	1	1	0
Forward3x3	1	0	1	0
Delay	6	2	4	2
LineBuffer	2	1	1	1
LeftShifter	4	0	4	0
Subtractor	6	2	4	2
Adder3x1	2	0	2	0
Multiplier	2	2	0	2
Adder2x1	1	1	0	1
Sqrt	1	1	0	1
Align2x2	0	1	1	0
Align3x3	1	0	1	0
Total	26	11	19	9

NS = Non Shareable, **S** = Shareable

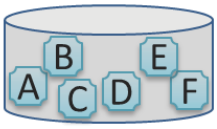
51

MDC

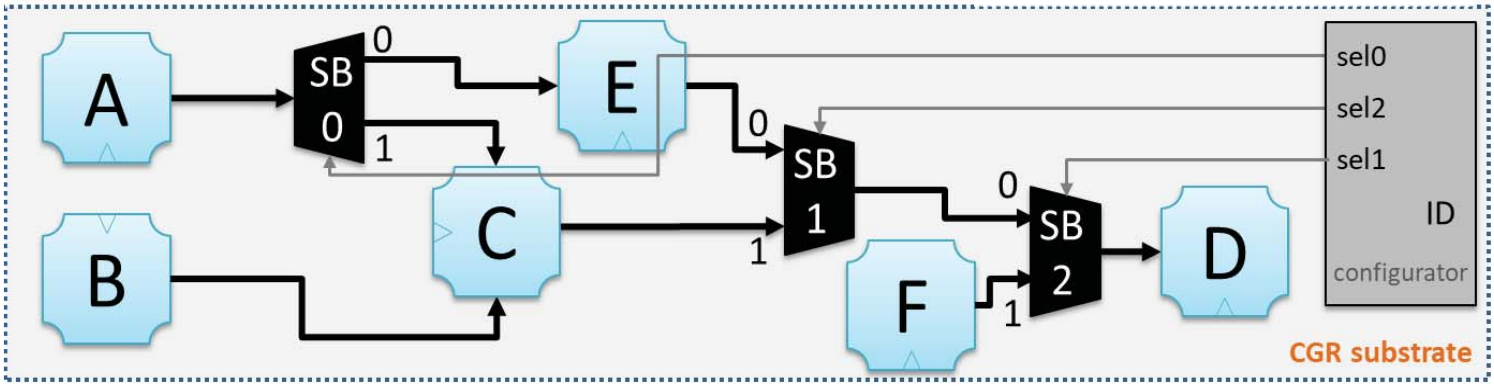
Baseline MDC Core: multi-dataflow



HDL components library



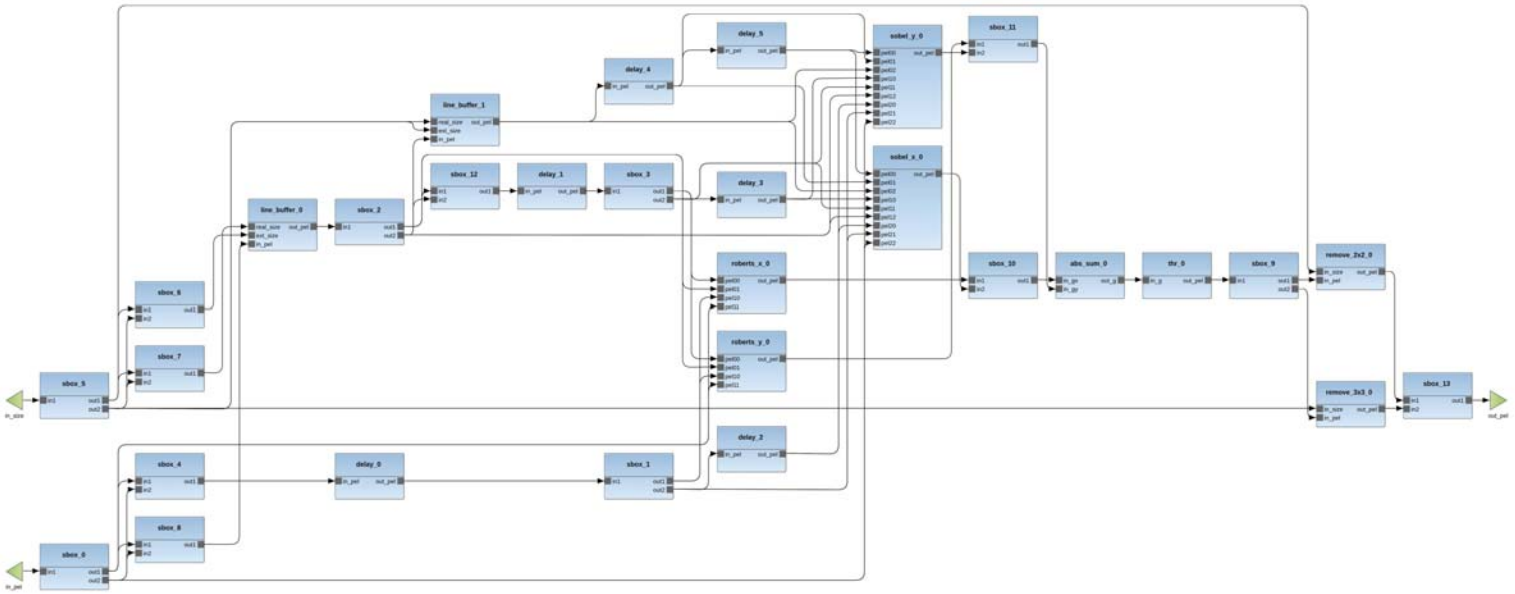
hardware communication protocol (XML)



52

MDC

Baseline MDC Core: multi-dataflow



53

MDC

Co-processor generation

Used to generated the wrapper compatible with the HWPU

- We need to define the communication protocol
  - Among actors
  - With the wrapper
- Different accelerators may have different communication protocols
  - They can coexist in the same system, since each HWPU is generated independently

MDC

Co-processor generation: Communication protocol

```
<protocol>
  <sys_signals>
    <signal id="0" net_port="clock" size="1" kind="input" is_clock=""/>
    <signal id="1" net_port="reset" size="1" kind="input" is_resetn=""/>
  </sys_signals>
  <actor>
    <sys_signals>
      <signal id="0" port="ap_clk" size="1" net_port="clock"/>
      <signal id="1" port="ap_rst_n" size="1" net_port="reset"/>
    </sys_signals>
    <comm_signals>
      <signal id="0" port="TDATA" channel="data" size="variable" kind="input" dir="direct"/>
      <signal id="1" port="TDATA" channel="data" size="variable" kind="output" dir="direct"/>
      <signal id="2" port="TREADY" channel="rd" size="1" kind="input" dir="direct"/>
      <signal id="3" port="TVALID" channel="wr" size="1" kind="output" dir="direct"/>
      <signal id="4" port="TVALID" channel="valid" size="1" kind="output" dir="reverse"/>
      <signal id="5" port="TREADY" channel="full" size="1" kind="input" dir="reverse"/>
    </comm_signals>
  </actor>
  <predecessor>
    <name>fifo_small</name>
    <sys_signals>
      <signal id="0" port="clk" size="1" net_port="clock"/>
      <signal id="1" port="rst" size="1" net_port="reset"/>
    </sys_signals>
    <comm_parameters>
      <parameter id="0" name="depth" value="bufferSize"/>
      <parameter id="1" name="size" value="variable"/>
    </comm_parameters>
    <comm_signals>
      <signal id="0" port="datain" channel="data" size="variable" kind="input" dir="direct"/>
      <signal id="1" port="dataout" channel="data" size="variable" kind="output" dir="direct"/>
      <signal id="2" port="enr" channel="rd" size="1" kind="input" dir="reverse"/>
      <signal id="3" port="enw" channel="wr" size="1" kind="input" dir="direct"/>
      <signal id="4" port="valid" channel="valid" size="1" kind="output" dir="direct"/>
      <signal id="5" port="full" channel="full" size="1" kind="output" dir="reverse"/>
    </comm_signals>
  </predecessor>
  <wrapper>
    <comm_signals>
      <signal id="0" channel="data" size="variable" mapping="data"/>
      <signal id="1" channel="rd" size="1" mapping="pop"/>
      <signal id="2" channel="wr" size="1" mapping="push"/>
      <signal id="3" channel="valid" size="1" mapping="empty"/>
      <signal id="4" channel="full" size="1" mapping="full"/>
    </comm_signals>
  </wrapper>
</protocol>
```

MDC

Co-processor generation: Communication protocol

1. System signals

1

```
▼<protocol>
 ▼<sys_signals>
    <signal id="0" net_port="clock" size="1" kind="input" is_clock=""/>
    <signal id="1" net_port="reset" size="1" kind="input" is_resetn=""/>
  </sys_signals>
 ▼<actor>
  ▼<sys_signals>
     <signal id="0" port="ap_clk" size="1" net_port="clock"/>
     <signal id="1" port="ap_rst_n" size="1" net_port="reset"/>
   </sys_signals>
  ▼<comm_signals>
     <signal id="0" port="TDATA" channel="data" size="variable" kind="input" dir="direct"/>
     <signal id="1" port="TDATA" channel="data" size="variable" kind="output" dir="direct"/>
     <signal id="2" port="TREADY" channel="rd" size="1" kind="input" dir="direct"/>
     <signal id="3" port="TVALID" channel="wr" size="1" kind="output" dir="direct"/>
     <signal id="4" port="TVALID" channel="valid" size="1" kind="output" dir="reverse"/>
     <signal id="5" port="TREADY" channel="full" size="1" kind="input" dir="reverse"/>
   </comm_signals>
 </actor>
▼<predecessor>
   <name>fifo_small</name>
  ▼<sys_signals>
     <signal id="0" port="clk" size="1" net_port="clock"/>
     <signal id="1" port="rst" size="1" net_port="reset"/>
   </sys_signals>
  ▼<comm_parameters>
     <parameter id="0" name="depth" value="bufferSize"/>
     <parameter id="1" name="size" value="variable"/>
   </comm_parameters>
  ▼<comm_signals>
     <signal id="0" port="datain" channel="data" size="variable" kind="input" dir="direct"/>
     <signal id="1" port="dataout" channel="data" size="variable" kind="output" dir="direct"/>
     <signal id="2" port="enr" channel="rd" size="1" kind="input" dir="reverse"/>
     <signal id="3" port="enw" channel="wr" size="1" kind="input" dir="direct"/>
     <signal id="4" port="valid" channel="valid" size="1" kind="output" dir="direct"/>
     <signal id="5" port="full" channel="full" size="1" kind="output" dir="reverse"/>
   </comm_signals>
 </predecessor>
▼<wrapper>
  ▼<comm_signals>
     <signal id="0" channel="data" size="variable" mapping="data"/>
     <signal id="1" channel="rd" size="1" mapping="pop"/>
     <signal id="2" channel="wr" size="1" mapping="push"/>
     <signal id="3" channel="valid" size="1" mapping="empty"/>
     <signal id="4" channel="full" size="1" mapping="full"/>
   </comm_signals>
 </wrapper>
</protocol>
```

MDC

Co-processor generation: Communication protocol

1. System signals
2. Actor signals

1

2

```
<protocol>
 <sys_signals>
   <signal id="0" net_port="clock" size="1" kind="input" is_clock=""/>
   <signal id="1" net_port="reset" size="1" kind="input" is_resetn=""/>
 </sys_signals>
 <actor>
  <sys_signals>
     <signal id="0" port="ap_clk" size="1" net_port="clock"/>
     <signal id="1" port="ap_rst_n" size="1" net_port="reset"/>
   </sys_signals>
  <comm_signals>
     <signal id="0" port="TDATA" channel="data" size="variable" kind="input" dir="direct"/>
     <signal id="1" port="TDATA" channel="data" size="variable" kind="output" dir="direct"/>
     <signal id="2" port="TREADY" channel="rd" size="1" kind="input" dir="direct"/>
     <signal id="3" port="TVALID" channel="wr" size="1" kind="output" dir="direct"/>
     <signal id="4" port="TVALID" channel="valid" size="1" kind="output" dir="reverse"/>
     <signal id="5" port="TREADY" channel="full" size="1" kind="input" dir="reverse"/>
   </comm_signals>
 </actor>
 <predecessor>
   <name>fifo_small</name>
  <sys_signals>
     <signal id="0" port="clk" size="1" net_port="clock"/>
     <signal id="1" port="rst" size="1" net_port="reset"/>
   </sys_signals>
  <comm_parameters>
     <parameter id="0" name="depth" value="bufferSize"/>
     <parameter id="1" name="size" value="variable"/>
   </comm_parameters>
  <comm_signals>
     <signal id="0" port="datain" channel="data" size="variable" kind="input" dir="direct"/>
     <signal id="1" port="dataout" channel="data" size="variable" kind="output" dir="direct"/>
     <signal id="2" port="enr" channel="rd" size="1" kind="input" dir="reverse"/>
     <signal id="3" port="enw" channel="wr" size="1" kind="input" dir="direct"/>
     <signal id="4" port="valid" channel="valid" size="1" kind="output" dir="direct"/>
     <signal id="5" port="full" channel="full" size="1" kind="output" dir="reverse"/>
   </comm_signals>
 </predecessor>
 <wrapper>
  <comm_signals>
     <signal id="0" channel="data" size="variable" mapping="data"/>
     <signal id="1" channel="rd" size="1" mapping="pop"/>
     <signal id="2" channel="wr" size="1" mapping="push"/>
     <signal id="3" channel="valid" size="1" mapping="empty"/>
     <signal id="4" channel="full" size="1" mapping="full"/>
   </comm_signals>
 </wrapper>
</protocol>
```

MDC

Co-processor generation: Communication protocol

1. System signals
2. Actor signals
3. FIFO parameters

MDC

Co-processor generation: Communication protocol

1. System signals
2. Actor signals
3. FIFO parameters
4. FIFO signals

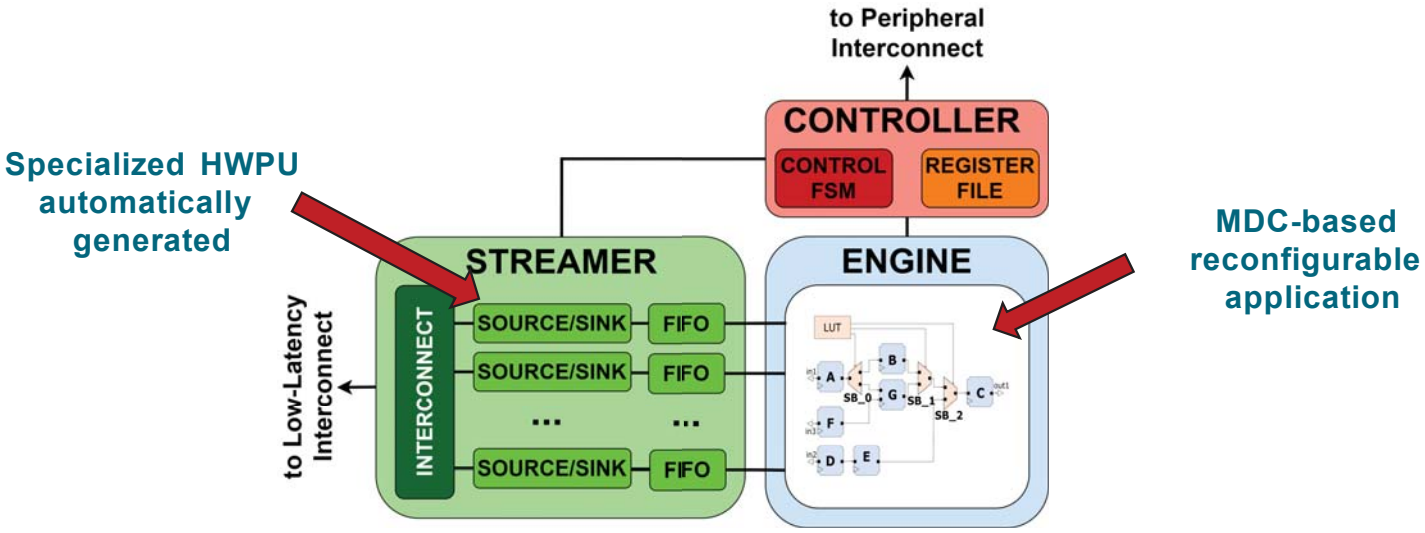
MDC

Co-processor generation: Communication protocol

1. System signals
2. Actor signals
3. FIFO parameters
4. FIFO signals
5. Wrapper signals

MDC

Co-processor generation: HWPU generated by MDC



61

MDC + OODK

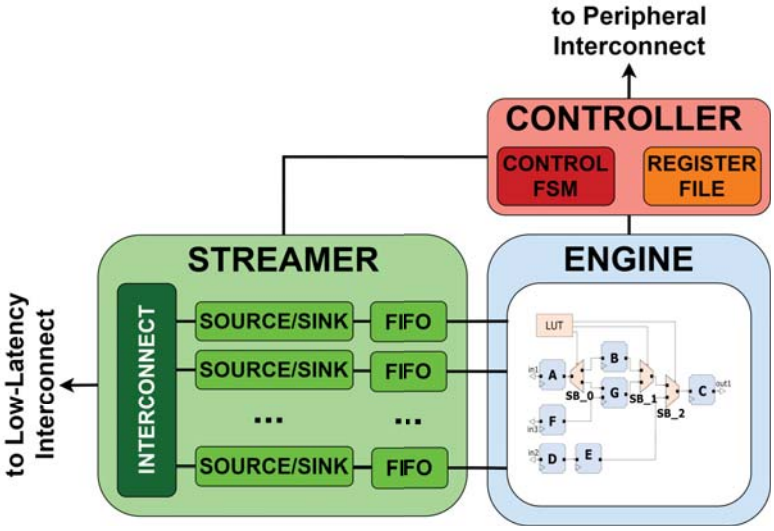
HWPU accelerator wrapper

Streamer

- Specialized DMA controller that transforms streams into memory accesses

Controller

- Register file to host runtime parameters
- Control FSM for coarse-grained control/(re-)configuration



Hardware Processing Unit

62

MDC + OODK

HWPU accelerator wrapper

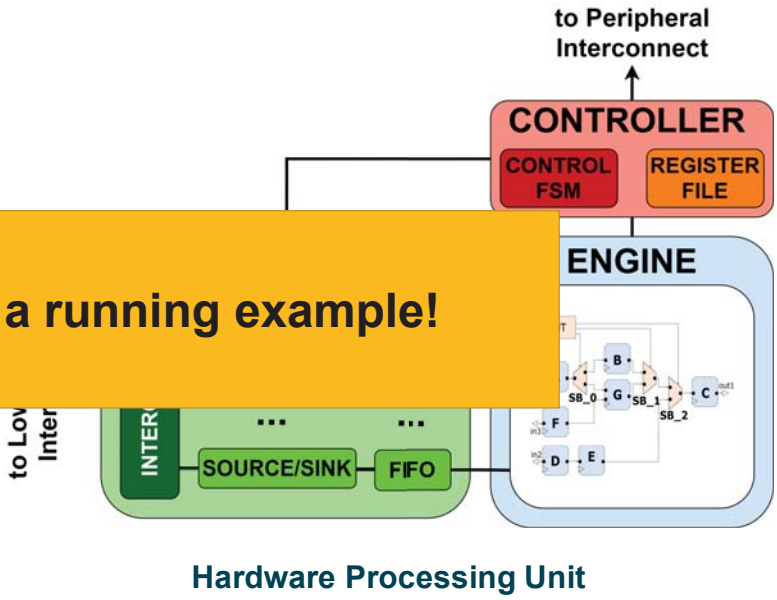
Streamer

- Specialized DMA controller that transforms streams into memory accesses

Controller

- Registe
- Control
  control/

It's now time to see a running example!



Hardware Processing Unit

63

AGENDA

1 Introduction

2 Methodology overview

3 MDC tool

4 OODK overlay

5 COMP4DRONES use case

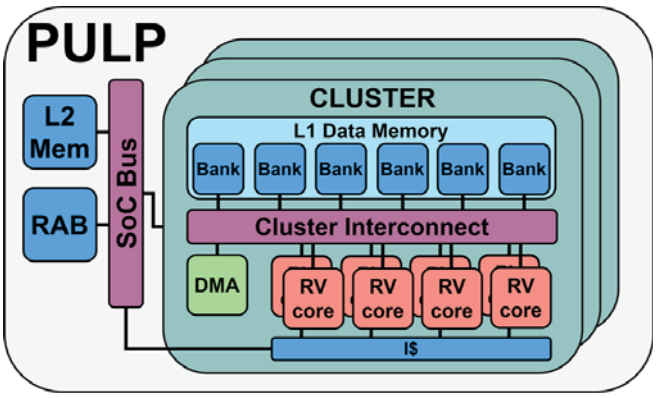
6 Conclusions

OODK

Starting point

PULP architecture

- PULP stands for «Parallel Ultra Low Power»
- Open and Scalable HW/SW research and development platform
- Cluster-based architecture
- RISC-V ISA compliant



Website: ***pulp-platform.org***

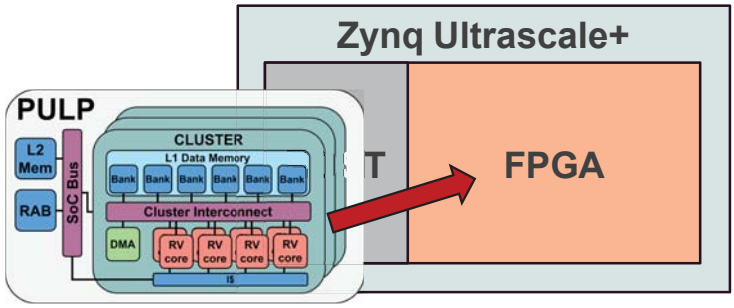
65

OODK
Starting point

COMP4DRONES

HERO

- FPGA emulation of heterogeneous and massively parallel PULP systems
- Instantiable with COTS FPGA-based heterogeneous SoCs



Website: ***pulp-platform.org***

RISC-V

Kurth, A., Capotondi, A., Vogel, P., Benini, L., & Marongiu, A. (2018)
HERO: An open-source research platform for HW/SW exploration of heterogeneous manycore systems.

66

OODK

FPGA overlay

COMP4DRONES

What is it?

- Hardware abstraction layer
- Overlays the original FPGA fabric → Hides hardware details

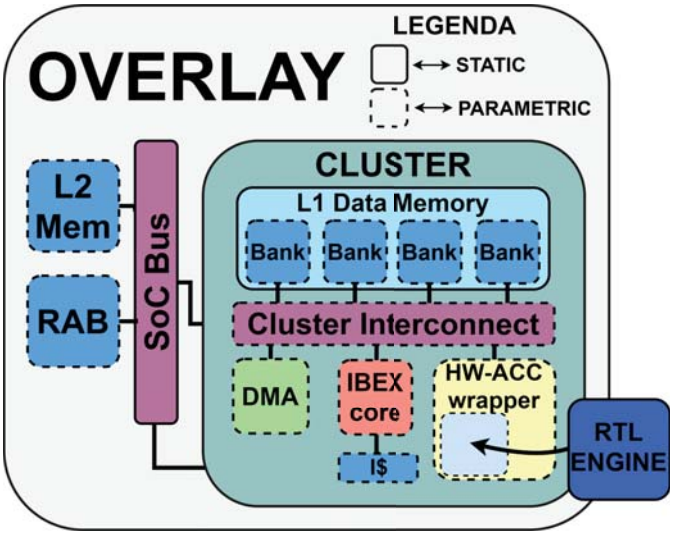
Features:

- Parametrized HW → Flexible design of custom architectures
- Abstracted design flow → Improved design productivity
- Programmable via standard APIs for heterogeneous compute platforms (e.g. OpenMP)

Bellocchi, G., Capotondi, A., Conti, F., & Marongiu, A. (2021)
A RISC-V-based FPGA Overlay to Simplify Embedded Accelerator Deployment

67

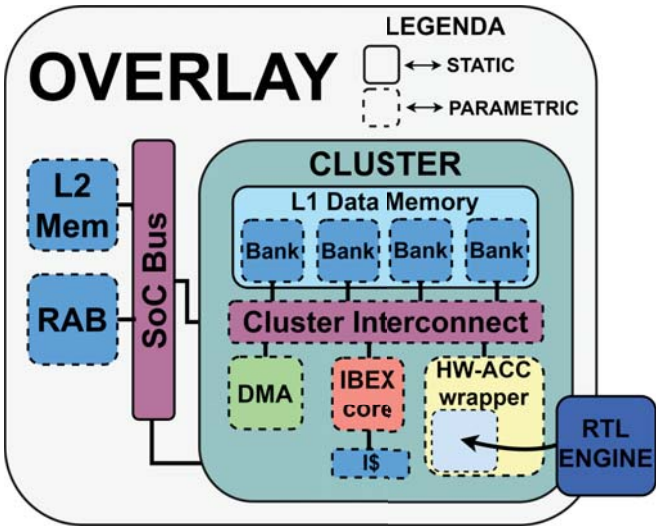
OODK
Architecture



OODK
Architecture

System-on-Chip (SoC) domain

- Cluster
  - Multi and single-cluster architectures
  - Agile integration of different accelerators
- L2 memory
  - Data and instruction memory
- Remapping address block (RAB)
  - An IO-MMU for translation of virtual addresses
- SoC bus
  - Highly-scalable interconnect

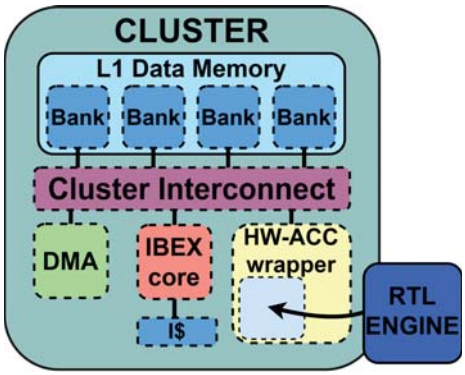


69

OODK
Architecture

Cluster domain

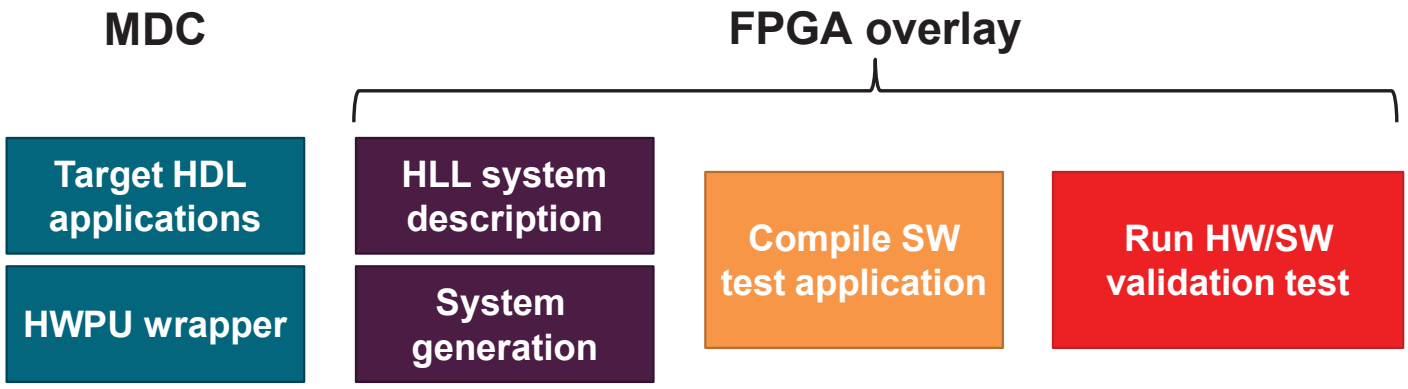
- HW accelerators
  - MDC-based HWPU
- RISC-V core
  - Tightly-coupled SW control - Accelerator routines, data management policies, etc.
  - L1 Instruction cache
- DMA
  - Specialized core for efficient L2 ↔ L1 data transfers
  - Support for 2D and 1D data transfers
- L1 data memory
  - Multi-banked scratchpad data memory (not a cache!)
- Cluster interconnect
  - Highly-scalable logarithmic interconnect + Peripheral bus



70

OODK

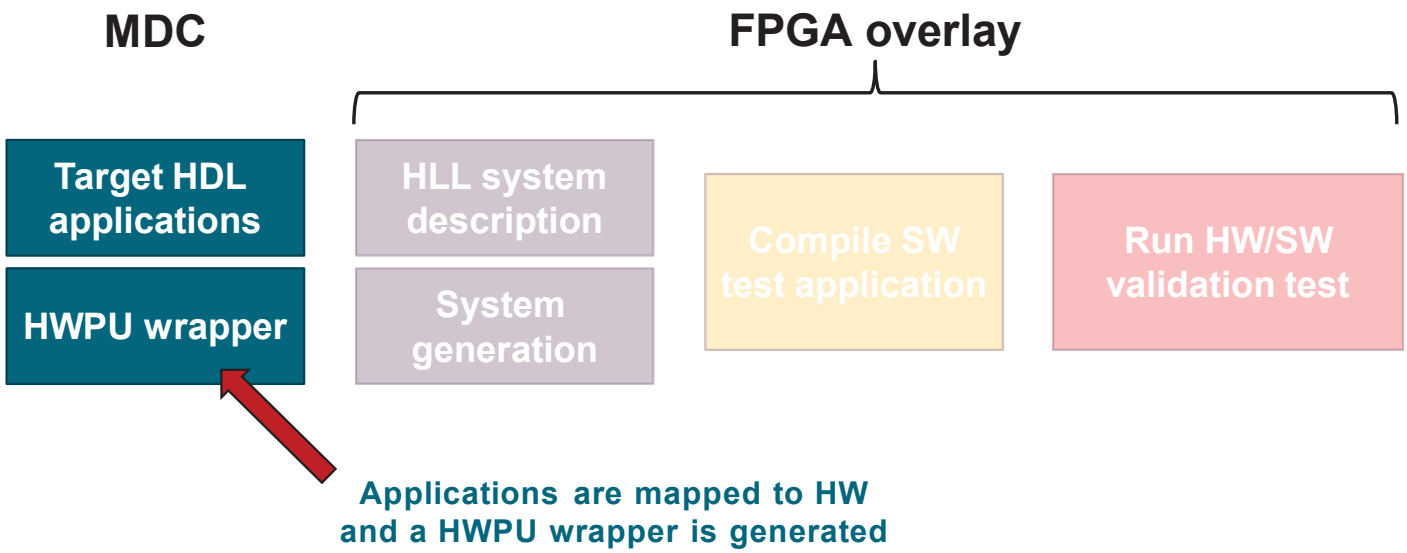
HW/SW co-design and verification



71

OODK

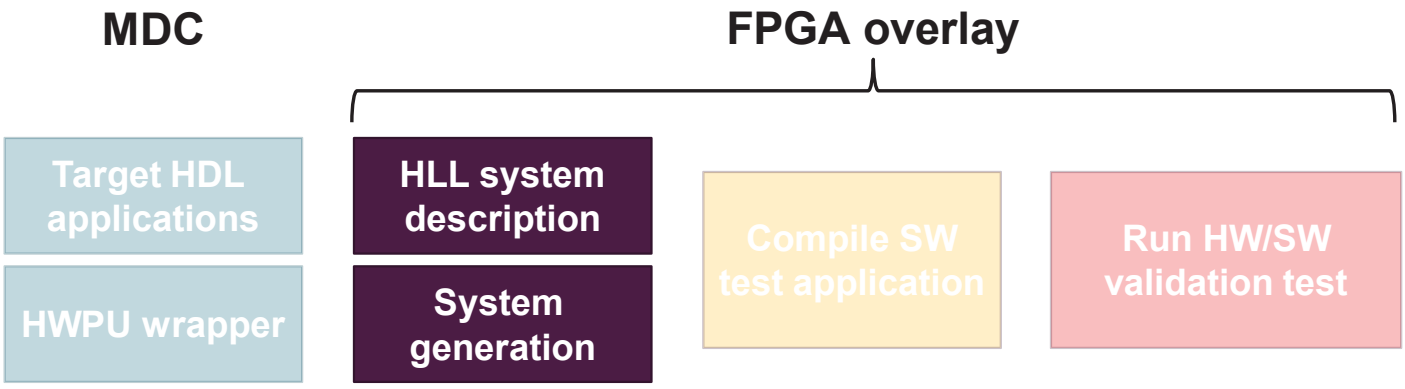
HW accelerator generation and integration



72

OODK

System generation



73

OODK

System generation

To choose a proper way of interconnecting accelerators is a primary requirement

- Which type of interconnect topology better fits our needs?
- What about the clustering level?
- How do accelerators mutually work?
  - Accelerators can either work in parallel or sequentially

Generation principles

- User knobs:
  - ***System optimization***
    - ✓ Memory hierarchy, control cores, DMA, etc.
    - ✓ Accelerator interconnections (generic vs. application-specific interconnects)
    - ✓ Accelerator scheduling (concurrent, serial or mixed scheduling)

74

OODK

System generation

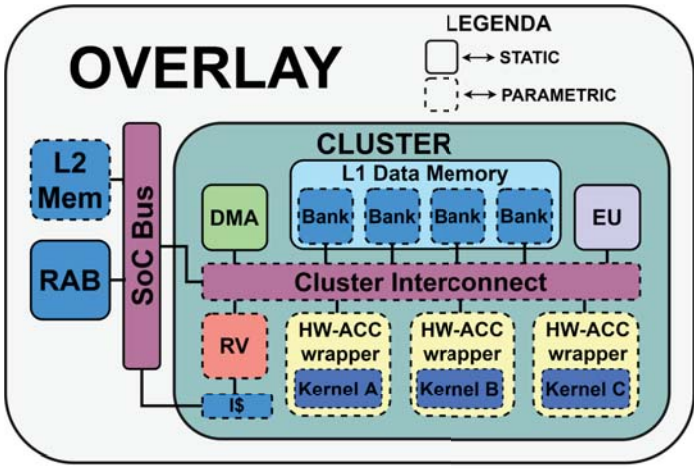
1. System information
2. Cluster information
3. HW accelerators interconnection
   - ❖ Logarithmic interconnect
   - ❖ Heterogeneous interconnect

```
class oodk_specs:

    def system(self):                                   1
        self.oodk_config                = 'ex_1_sys_gen'
        return self

    def cluster_0(self):                                2
        self.cl_offset                  = 0
        self.core                       = [ 'ibex', 1 ]
        self.tcdm                       = [ 32 , 128]
        self.lic                        = [ [ 'kernel_A' , 'hwpu'],   3
                                            [ 'kernel_B' , 'hwpu'],
                                            [ 'kernel_C' , 'hwpu']]
        self.hci                        = [ ]
        return self
```

OODK

Example #1 – Connection to Cluster Interconnect

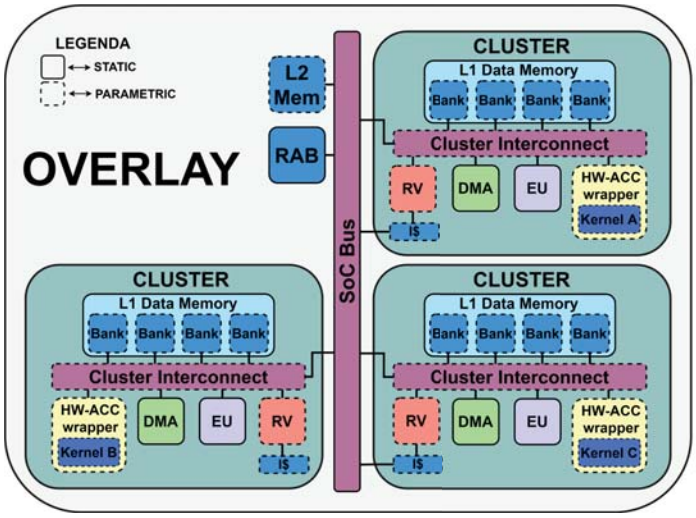
```
class oodk_specs:

    def system(self):
        self.oodk_config                = 'ex_1_sys_gen'
        return self

    def cluster_0(self):
        self.cl_offset                  = 0
        self.core                       = [ 'ibex', 1 ]
        self.tcdm                       = [ 32 , 128]
        self.lic                        = [ [ 'kernel_A' , 'hwpu'],
                                            [ 'kernel_B' , 'hwpu'],
                                            [ 'kernel_C' , 'hwpu']]
        self.hci                        = [ ]
        return self
```



OODK

Example #2 – Multi-Cluster Interconnection

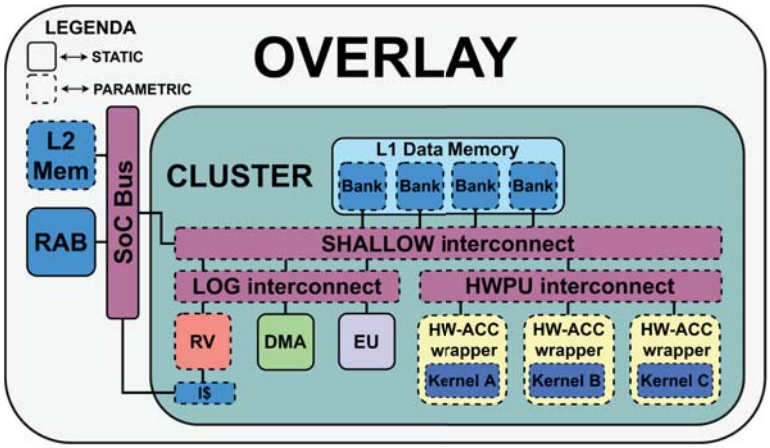
```
class oodk_specs:

    def system(self):
        self.oodk_config                = 'ex_2_sys_gen'
        return self

    def cluster_0(self):
        self.cl_offset                  = 0
        self.core                       = [ 'ibex', 1 ]
        self.tcdm                       = [ 32 , 128]
        self.lic                        = [ [ 'kernel_A' , 'hwpu']]
        self.hci                        = [ ]
        return self

    def cluster_1(self):
        self.cl_offset                  = 0
        self.core                       = [ 'ibex', 1 ]
        self.tcdm                       = [ 32 , 128]
        self.lic                        = [ [ 'kernel_B' , 'hwpu']]
        self.hci                        = [ ]
        return self

    def cluster_2(self):
        self.cl_offset                  = 0
        self.core                       = [ 'ibex', 1 ]
        self.tcdm                       = [ 32 , 128]
        self.lic                        = [ [ 'kernel_C' , 'hwpu']]
        self.hci                        = [ ]
        return self
```



OODK

Example #3 – Heterogeneous Interconnection

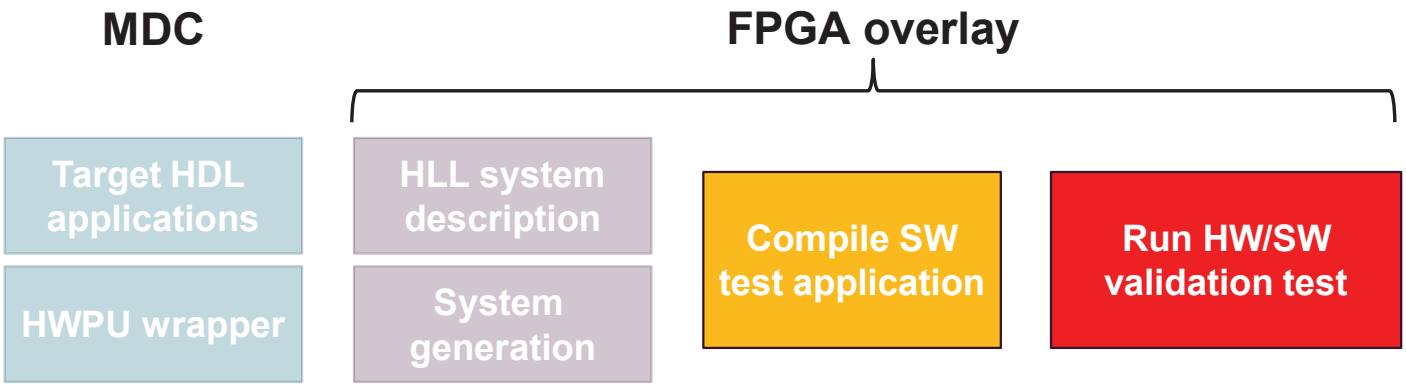
```
class oodk_specs:

    def system(self):
        self.oodk_config                = 'ex_3_sys_gen'
        return self

    def cluster_0(self):
        self.cl_offset                  = 0
        self.core                       = [ 'ibex', 1 ]
        self.tcdm                       = [ 32 , 128]
        self.lic                        = [ ]
        self.hci                        = [ [ 'kernel_A' , 'hwpu'],
                                            [ 'kernel_B' , 'hwpu'],
                                            [ 'kernel_C' , 'hwpu']]
        return self
```



78

OODK

System generation



79

OODK

System generation

Test application

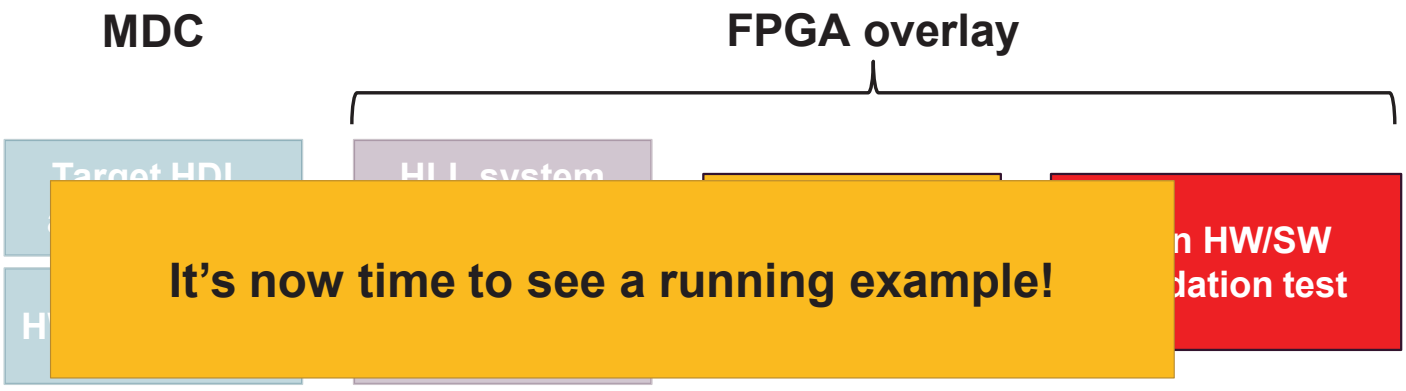
- Baremetal software test
- Compiled for the OODK system
- A template version is generated together with the system itself

HW/SW validation test

- RTL simulation
  - Before to head up to the FPGA set-up, the generated designs are tested in QuestaSim testbench
  - The real behavior of the baremetal application is tested
    - The RISC-V core executes the test application
    - The accelerators functionality is validated with synthetic stimuli

OODK

System generation



81

AGENDA

1 Introduction

2 Methodology overview

3 MDC tool

4 OODK overlay

5 COMP4DRONES use case

6 Conclusions

Current application: C4D

COMP4DRONES

Current application: C4D

Development and assessment of Smart and Precision Agriculture Technologies to enable:

1. **Improve non-real time actions**, i.e. forecast on production volume and optimized water management.
2. **Real-time field monitoring and inspection**, i.e. automatic disease detection and cross-correlation of plants indexes;
3. **Prompt on-field intervention**, i.e. customized spot spraying;

84

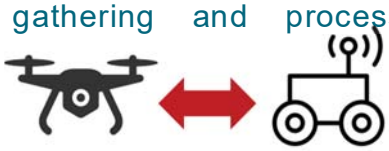
Current application: C4D

Development and assessment of Smart and Precision Agriculture Technologies to enable:

1. **Improve non-real time actions**, i.e. forecast on production volume and optimized water management.
2. **Real-time field monitoring and inspection**, i.e. automatic disease detection and cross-correlation of plants indexes;
3. **Prompt on-field intervention**, i.e. customized spot spraying;

TECHNICAL SET-UP

Tandem of cooperative autonomous vehicles composed of a field rover, responsible of gathering and processing field data, and a spraying drone



85

Current application: C4D motivation

Current application: C4D motivation

USER NEEDS

1. **Use as little pesticides**: Proper assessment of health status & on spot interventions
2. **Waste as little water as possible**: Precise growth assessment

87

Current application: C4D motivation

USER NEEDS

1. **Use as little pesticides**: Proper assessment of health status & on spot interventions
2. **Waste as little water as possible**: Precise growth assessment

88

Current application: C4D motivation

USER NEEDS

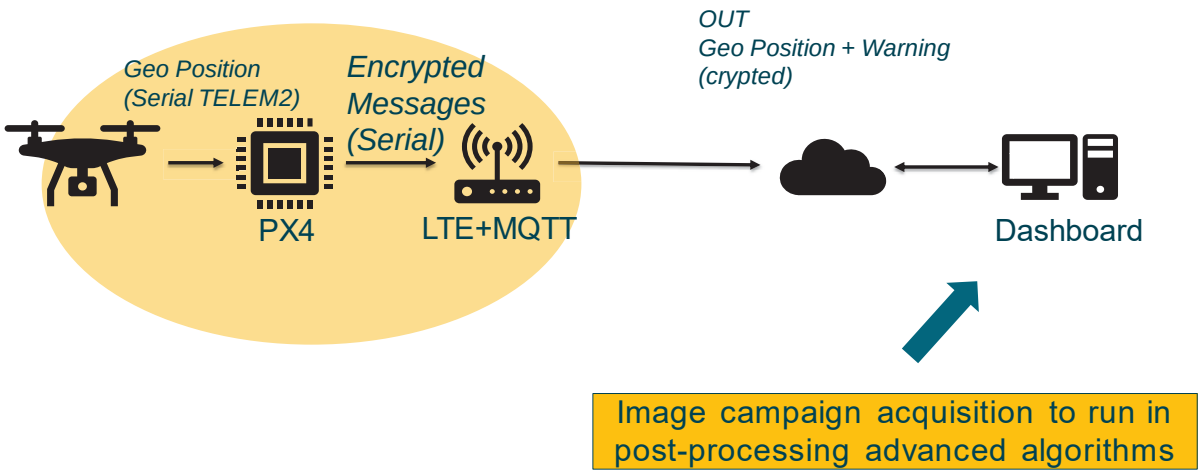
1. **Use as little pesticides**: Proper assessment of health status & on spot interventions
2. **Waste as little water as possible**: Precise growth assessment

EXPECTED BENEFITS

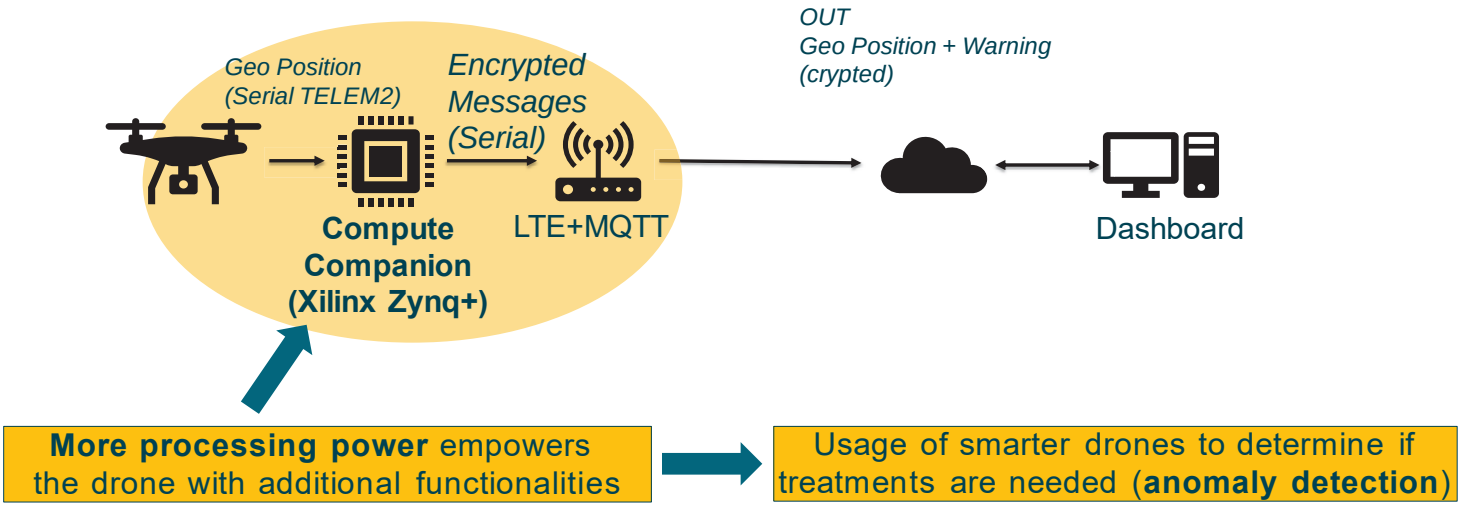
1. Reduced impact on the environment
2. Reduced human effort
3. Improved usability of advanced technologies by non-expert operators

89

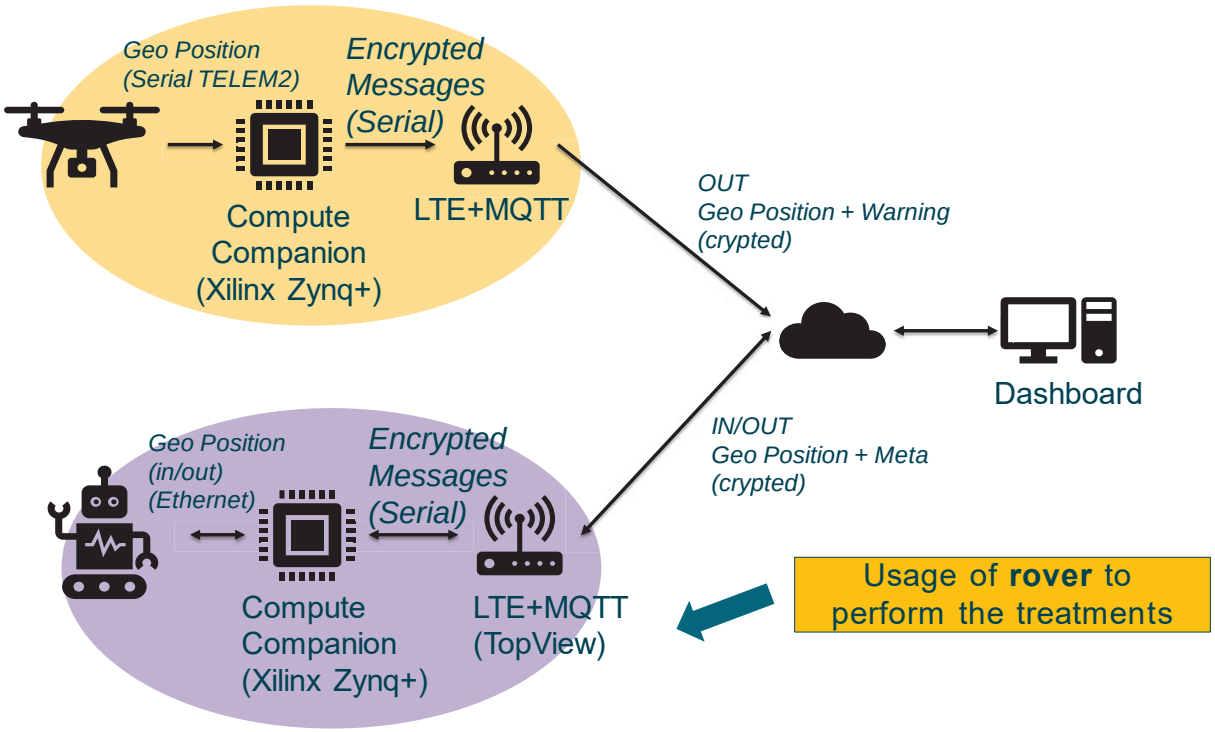
Current application: Baseline



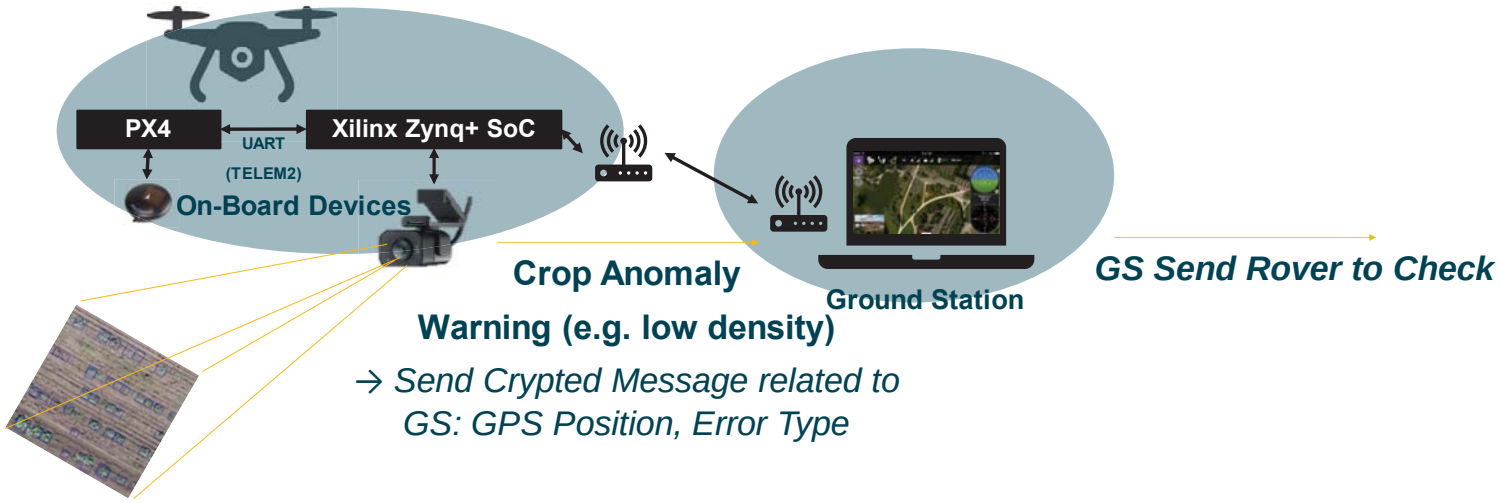
Current application: Scenario 2



Current application: Scenario 3

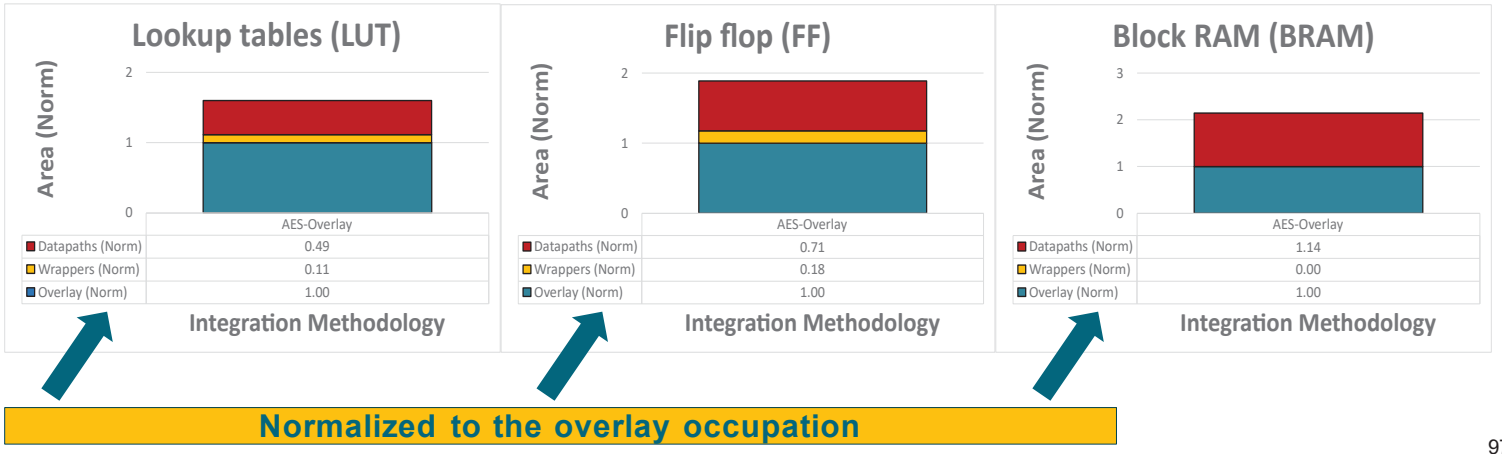


Current application: Scenario 3



C4D methodology experimental results

- ✓ **Overall x2 speedup** when comparing SW vs HW implementation of the AES algorithm
- ✓ OODK+AES has been implemented targeting a ZU9EG SoC with a resource cost of:
  - ➢ **~43.7% LUTs**
  - ➢ **~11.7% FFs**
  - ➢ **~13.2% BRAMs**



97

AGENDA

1 Introduction
2 Methodology overview
3 MDC tool
4 OODK overlay
5 COMP4DRONES use case
6 Conclusions

Conclusions

- ✓ Simplified design of HW accelerators through MDC
- ✓ Multi-functionality, multi working-point and reconfiguration support for CGRAs
- ✓ Support for accelerators generated with different tools (e.g., CAPH, HLS)
- ✓ Agile methodology for the design and exploration of accelerator-rich systems
- ✓ Simplified validation and deployment of the generated HW/SW system
- ✓ Practical use case: COMP4DRONES

COMP4DRONES

SS-CPS&IoT2022

Accelerator-Rich FPGA Architecture Exploration via a Programmable and Reconfigurable Overlay

Gianluca Bellochi[1], Daniel Madroñal[2], Alessandro Capotondi[1], Andrea Marongiu[1], Francesca Palumbo[2]

[1]{gianluca.bellochi, alessandro.capotondi, andrea.marongiu}@unimore.it
[2]{dmadronalquin, fpalumbo}@uniss.it

ECSEL Joint Undertaking
Electronic Components and Systems for European Leadership

This project has received funding from the ECSEL Joint Undertaking (JU) under grant agreement No 826610. The JU receives support from the European Union's Horizon 2020 research and innovation programme and Spain, Austria, Belgium, Czech Republic, France, Italy, Latvia, Netherlands.


COMP4DRONES

From Embedded Systems To Swarms: Opportunities And Challenges

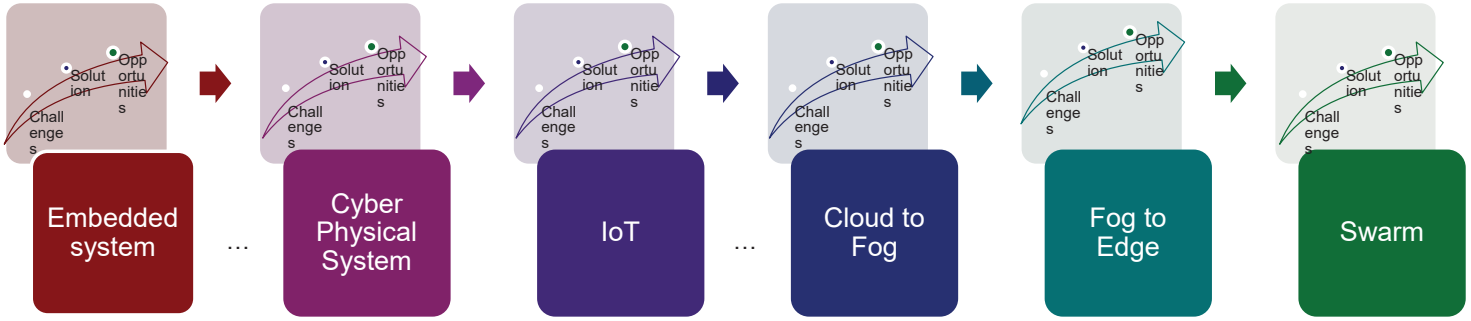
Morayo Adedjouma, Reda Nouacer
CEA LIST - DILS

CPS&IoT'2022 Summer School on Cyber-Physical Systems and Internet-of-Things
Budva, Montenegro, June 7-11, 2021

AGENDA

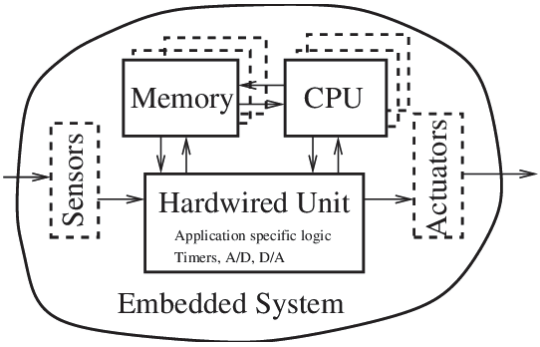
- From Embedded systems to Cyber-Physical Systems (CPS)
- From Cyber-Physical Systems to Internet of Things (IoT)
- Example from CPS4EU project
- Through IoT Architecture
  - From Cloud to Fog to Edge continuum
- Swarm computing

A CYCLIC, ITERATIVE EVOLUTION PROCESS



AN EMBEDDED SYSTEM IS…

- **Custom-built special computer used for a specific purpose**
- Combination of
  - computer **processor**,
  - computer **memory**,
  - and **input/output** peripheral **devices**
- Use **real-time operating system (RTOS)** to communicate with the hardware
- ***"Embedded"*** as part of a complete device often including electrical or electronic hardware and mechanical parts



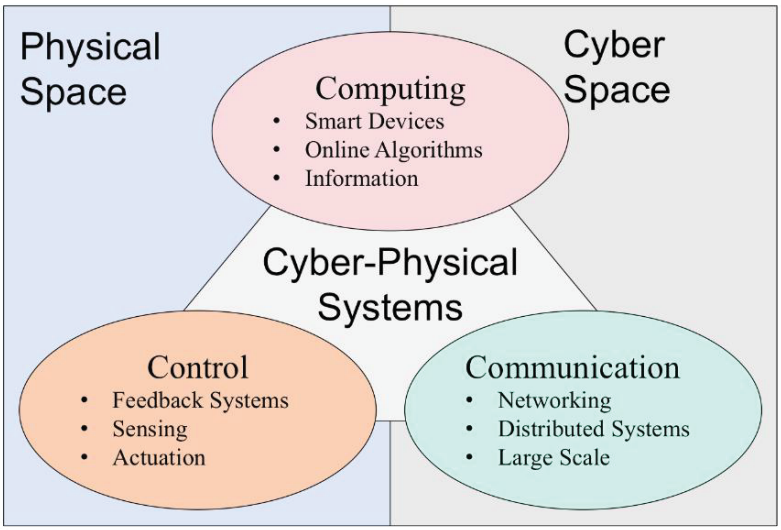
AN EMBEDDED SYSTEM …

- Range in size from portable personal devices to bigger machines
- Mainly constitute subsystems of other machines like automobile, aircraft, household appliances… the cyber physical system (CPS)

A CYBER PHYSICAL SYSTEM IS…

- **Cyber-physical** systems (CPS) combine, and build on, different elements including embedded systems, cybernetics, distributed control, sensor networks, control theory and systems engineering artefacts

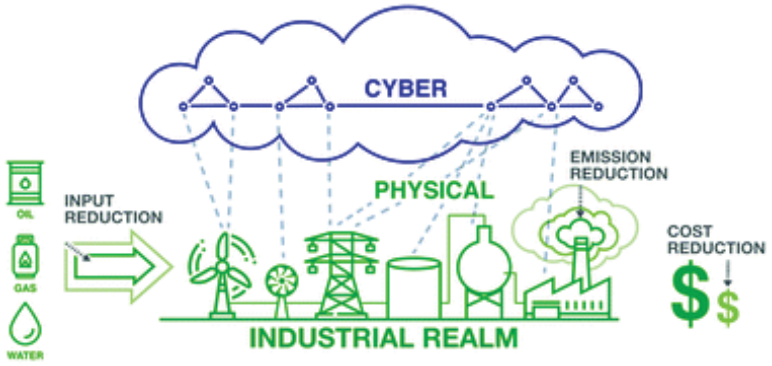
- Requires three fundamental attributes:
  - **Communication**
  - **Control**
  - **Computing**



A CYBER PHYSICAL SYSTEM …

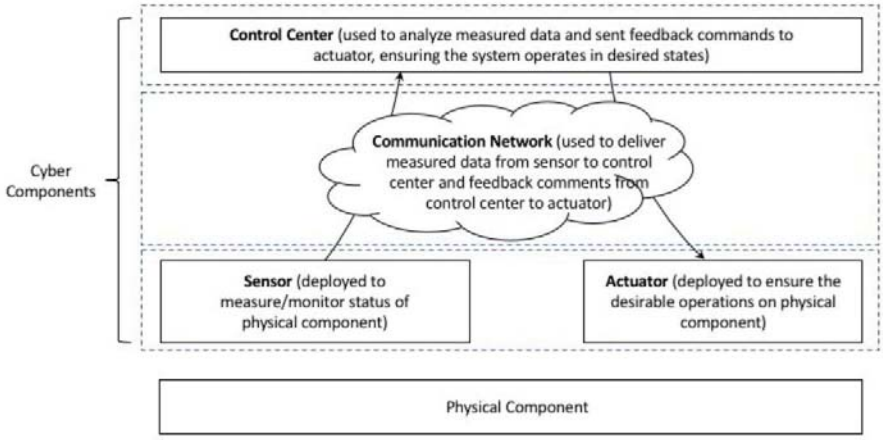
- **CPS is closely coupled to its physical environment**
  - Embedded computers and networks monitor and control the physical processes, with feedback loops where physical processes affect computations and vice versa

- **Two categories of CPS**
  - Autonomous (AI-based) systems, capable of making decisions and operating independently.
  - closed-loop human machine systems, able to learn from the environment including the human, to make decisions in real-time, but the human remains an integral part of the system's decision-making process.



ARCHITECTURE AND FUNCTIONS OF CPS

- **3 layers vertical architecture**
  - The sensor/actuator layer is used to collect real-time data and execute commands
  - The communication layer is used to deliver data to the application (control) layer and commands to the sensor/actuator layer
  - The application (control) layer is used to analyze data and make decisions.



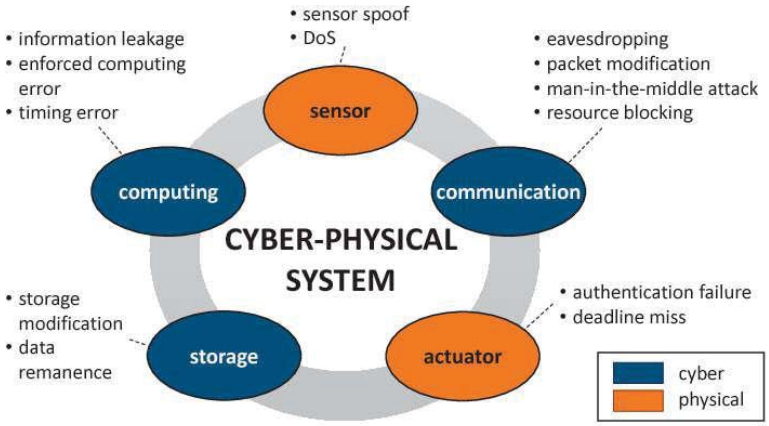
From Embedded-Systems towards swarms: opportunities and challenges | 8

CPS OPPORTUNITIES AND CHALLENGES

- **CPS requires three main properties: *safety*, *security*, and *sustainability***

- **Opportunities:**
  - Fortify and increase the efficacy of traditionally physical system based on computing
  - Achieve significant benefits in terms of cost, performance and overall life-cycle sustainability
  - Focus on effective reliable, accurate, real-time and secure data transmission and control within the CPS

- ***Challenges:***
  - Intermittent power supply and unknown load characteristics due to unpredictability of physical environment
  - Vulnerability to cyber attacks on the control elements, network or physical systems
  - Privacy issues

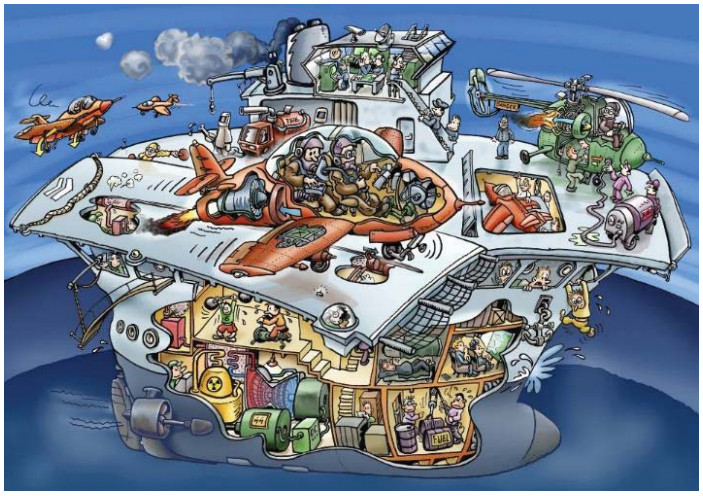


From Embedded-Systems towards swarms: opportunities and challenges

FROM CPS TO SYSTEMS OF SYSTEMS AND IOT

- A **CPS** can be an **element of an super – CPS …** with an increasing complexity, cumulative ambiguity, etc.
  - An aircraft carrier is an airport, a boat, a town, …
  - A carrier battle group is an aircraft carrier, airplanes, helicopters, a supply ship, submarines, an anti-submarine frigate an air defense frigate

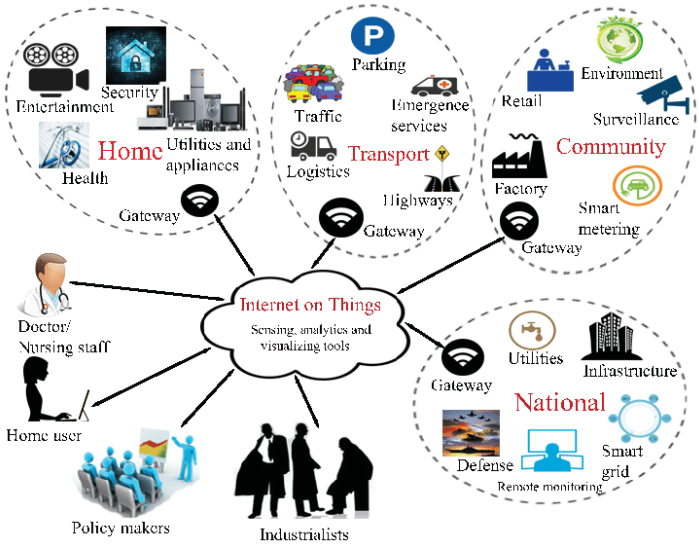
→ we talk about **Internet of Things (IoT) system**



Aircraft Carrier [10]

INTERNET OF THINGS (IOT) IS …

- Platform where physical objects (or groups of such objects) are connected to the internet or other heterogeneous networks, so they can interact, collect and exchange data with each other, learn from each other's experiences
- **IoT** is viewed as an internet of many CPS
- Focus on effective resource sharing and management, interface among different networks, massive-scale data and big data collection and storage, data mining, data aggregation and information extraction, high quality of network quality of service



IOT TRENDS

- Sensing and computing
  - Access to low-cost, low-power sensor technology
  - Opportunities for more direct integration of the physical world into computer-based systems

- Machine learning and analytics
  - Advances in machine learning and analytics, and access to vast amounts of diverse data, businesses get insights faster and easier.

- Connectivity
  - Explosive growth of devices connected and controlled by the Internet
  - Proliferation of network protocols for the Internet ease connection of sensors to the other "things" for more efficient data transfers.

CPS4EU PROJECT OVERVIEW

What is at stake with the CPS4EU Project ?

- Transition from linear eco-systems to networked eco-systems
- Digitization, AI, CPS/IOT, Edge computing, Connectivity, 5G, Software updates Over The Air
- Safety, Security (Cyber-), Privacy, Ethics, Export rules
- Low power consumption, SWAP*
- Seamless development process (Digital twins, Model-based engineering, Security by design)
- Management of project complexity

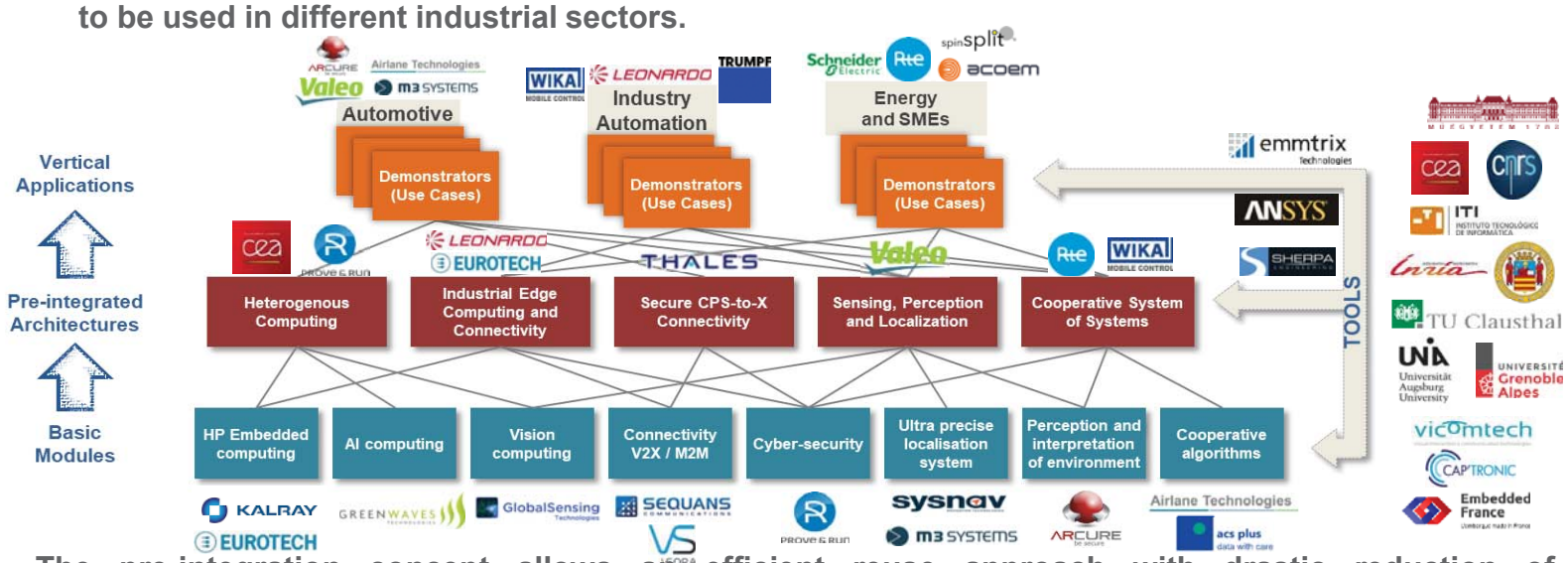
CPS4EU is about:

- Innovations in Cyber Physical Systems
- Components and Tools for 5 Pre-Integrated Architectures
- 16 Practical Use Cases
- New way of working and doing business with partners

(*) SWAP: Size Weight And Power

STRONG AND BALANCED CONSORTIUM

- 36 Partners from 5 European Countries, equally distributed among Large Enterprises, SMEs and Academics
- From basic modules to vertical applications: Pre-integrated architecture allow basic CPS modules to be used in different industrial sectors.

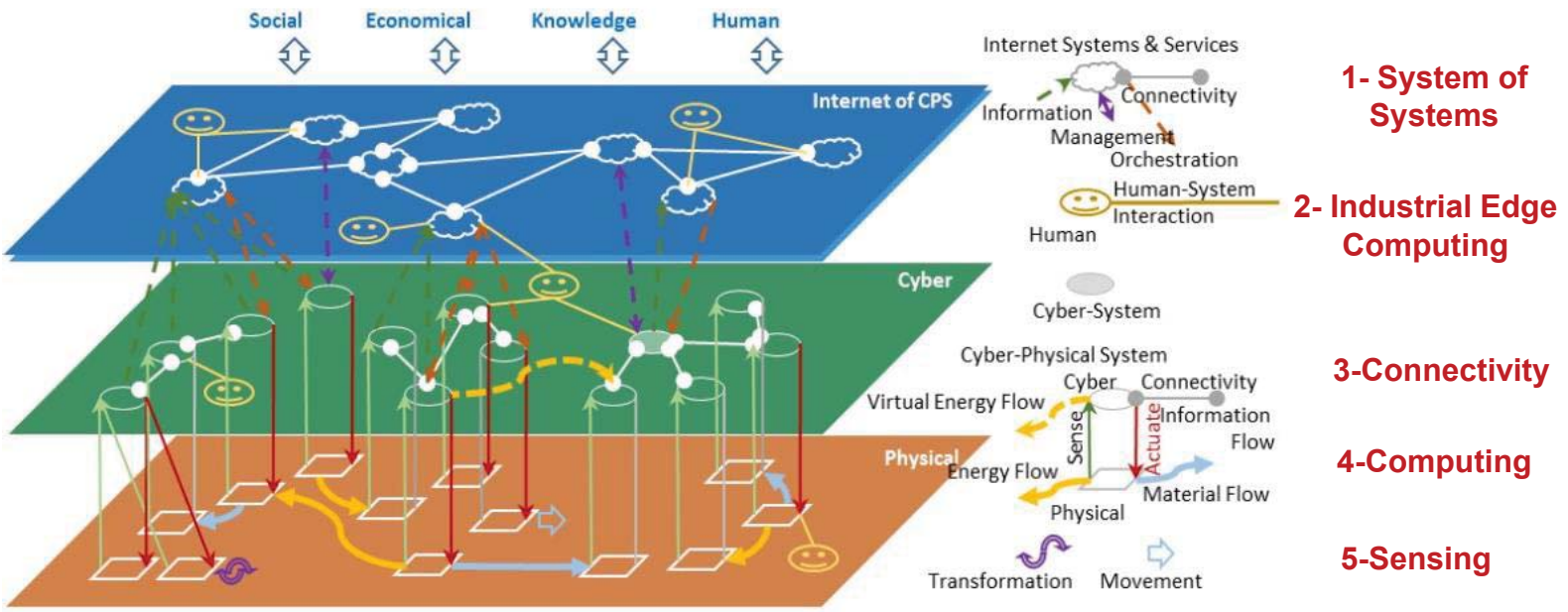


- The pre-integration concept allows an efficient reuse approach with drastic reduction of implementation effort for both CPS module providers and users.

CPS4EU – NIST FRAMEWORK FOR SYSTEM

CPS4EU – European project

Propose 5 patterns as Pre-Integrated Architecture (PIARCH) for CPS



1- System of Systems

2- Industrial Edge Computing

3-Connectivity

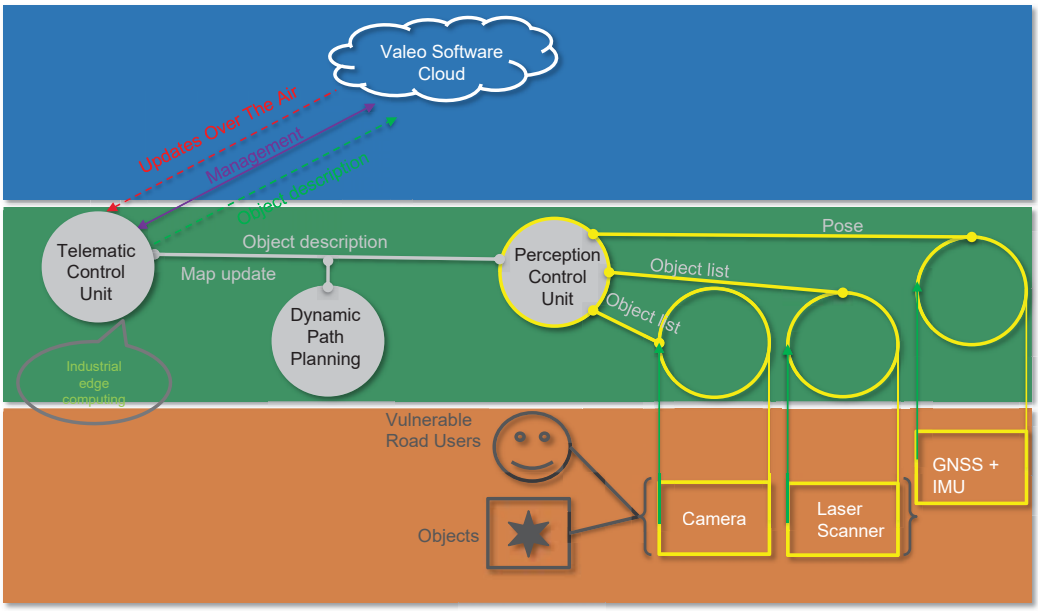
4-Computing

5-Sensing

From Embedded-Systems towards swarms: opportunities and challenges | 15

CPS4EU – USE CASE DESCRIPTION WITH PRE-INTEGRATED ARCHITECTURES

Automotive Use Case : Smart sensors for AD Level 4 - Valeo



Note: the example is related to the usage and update of Digital Maps by the perception control unit

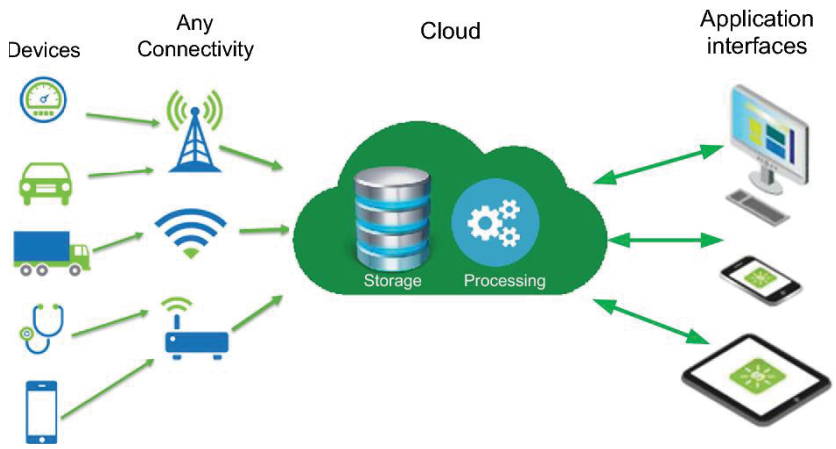
AD L4: the Automated Driving System (ADS) is fully responsible of the Dynamic Driving Task (DDT) within the Operating design Domain

GNSS: Global Navigation Satellite System

IMU: Inertial Measurement Unit

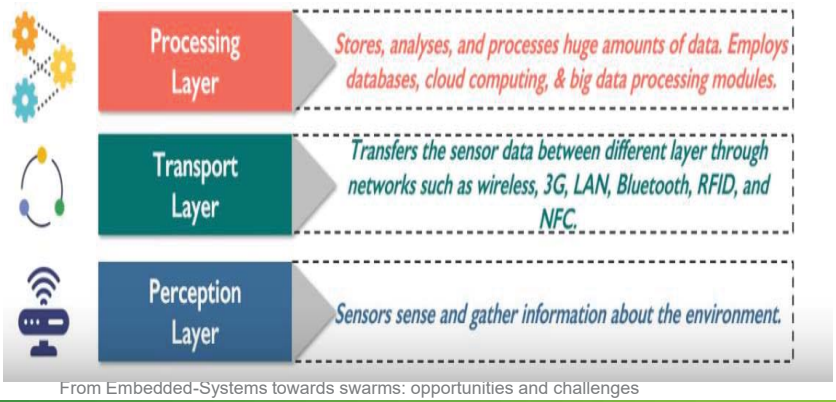
IOT OPPORTUNITIES

- **Big data and Cloud platforms**
  - Vast amounts of data produced by the IoT required adequate infrastructure for storage.
  - Businesses and consumers may access to the infrastructure they need to scale, without the hassle of managing it.



BASIC IOT ARCHITECTURE: 3-LAYER ARCHITECTURE

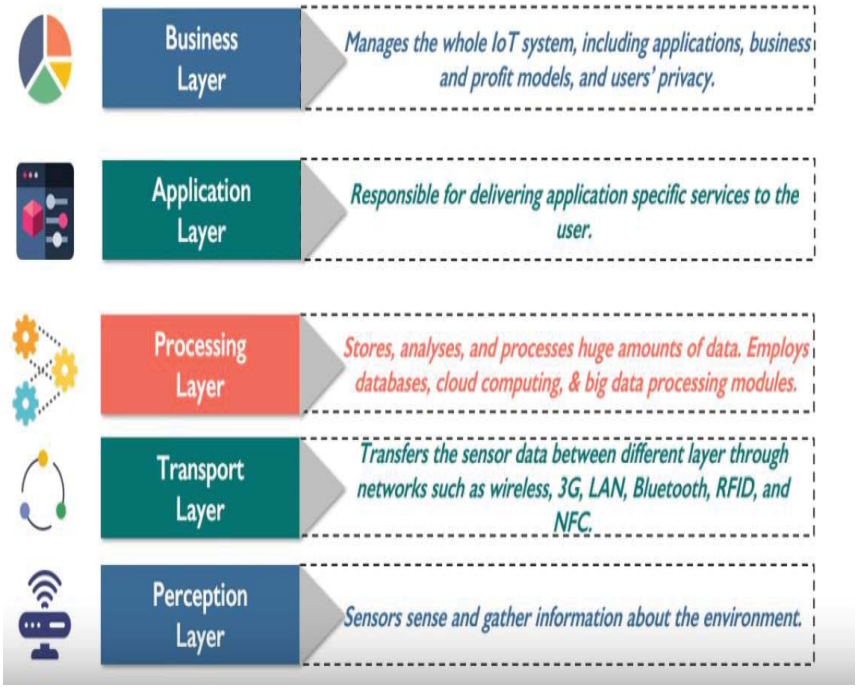
- **Cloud-centric IoT architecture**, where almost all of the IoT data processing is done on the cloud or a remote server.
- Comprises
  - **Perception layer** – Sensors, actuators and edge devices that interact with the environment
  - **Transport Layer** – Discovers, connects and translates devices over a network and in coordination with the application layer
  - **Processing Layer** – Data processing and storage with specialized services and functionality for users



From Embedded-Systems towards swarms: opportunities and challenges

5-LAYER IOT ARCHITECTURE (1/2)

- Builds upon the three layer approach with the addition of :
  - **A Business Layer** - Manages the entire IoT system, its functionality, applications, and business models.
  - **A Application Layer** - Provides application specific services to users



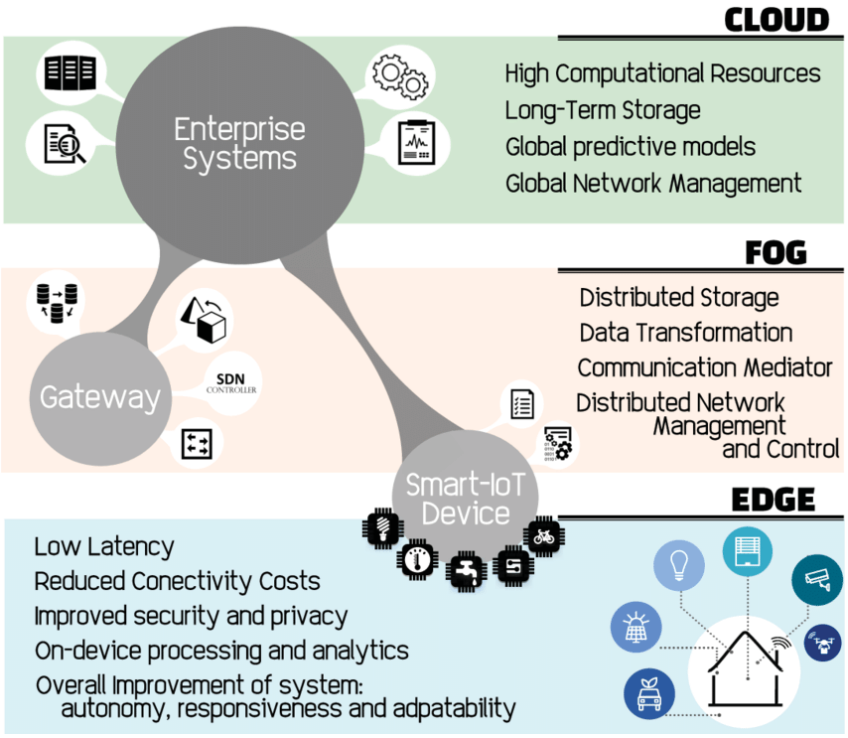
From Embedded-Systems towards swarms: opportunities and challenges | 19

5-LAYER IOT ARCHITECTURE (2/2)

- **Benefits**
  - Provide consistent value to the business and end users.
  - Filter down massive data sets and thus conserve resources.

- **Drawbacks:**
  - Push network bandwidth requirements to the limit
  - Security threats and privacy issues
  - Latencies
  - Analysis & (critical /sensitive) application deployment limitations regarding
    - Dependability concerns
    - Embarqualibility

FROM CLOUD TO FOG/EDGE CONTINUUM



* Alam, Muhammad & Rufino, João & Ferreira, Joaquim & Ahmed, Syed & Shah, Nadir & Chen, Yuanfang. (2018). Orchestration of Microservices for IoT Using Docker and Edge Computing. IEEE Communications Magazine. 56. 118-123.

FROM CLOUD TO FOG/EDGE CONTINUUM

Key enablers

- 5G – Speed and low latency
- Need for Near Real-Time Response
- Container technology like Docker and Kubernetes
- Service and Data mesh
- Software-Defined Networking (SDN)
- Digital Twin

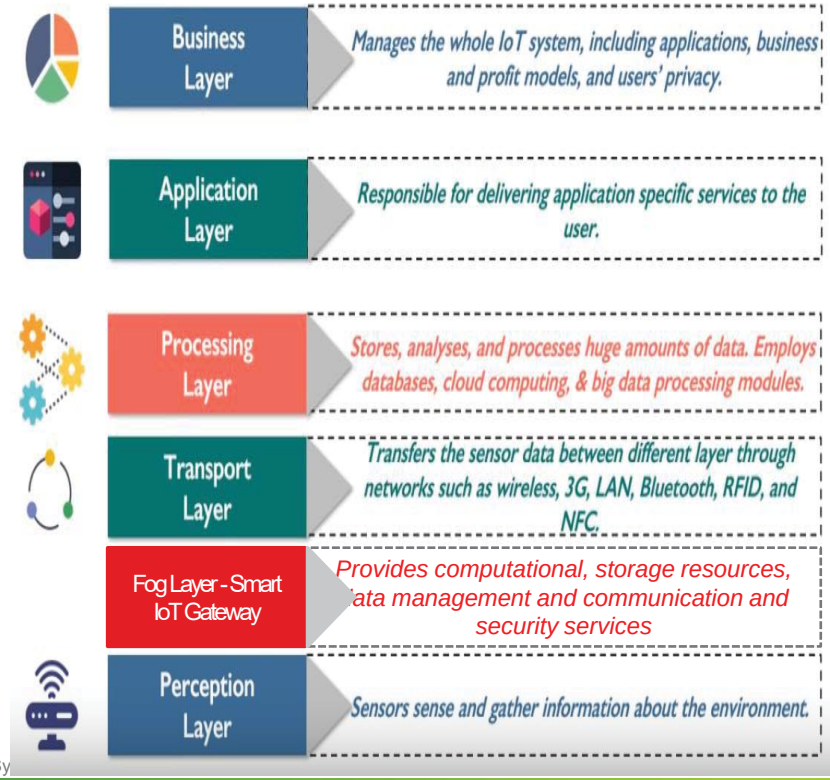
Key enablers

- Maturity and adoption of Industrial IoT
- Industrialization of IoT Sensors
- Multi-Access Edge Computing (MEC)
- Extended Reality (XR)
- Privacy-oriented technology
- Robotics
- Heterogeneous hardware
- …

FOG COMPUTING IOT ARCHITECTURE (1/2)

- Moves certain IoT services, like monitoring and pre-processing to a Fog-layer, closer to the edge to enable faster local decision making and automation.
- May be physical or virtual
- Comprises as sub-layers:

Security Layer	*Encrypt / decrypt data.*
Storage Layer	*Store files or data with localized relevance.*
Pre-Processing	*Filter, processes, analyze and reduce edge data or process commands or subscriptions from the cloud.*
Monitoring	*Monitor power, resources, responses, and services, access.*



From Embedded-Sy

FOG COMPUTING IOT ARCHITECTURE (2/2)

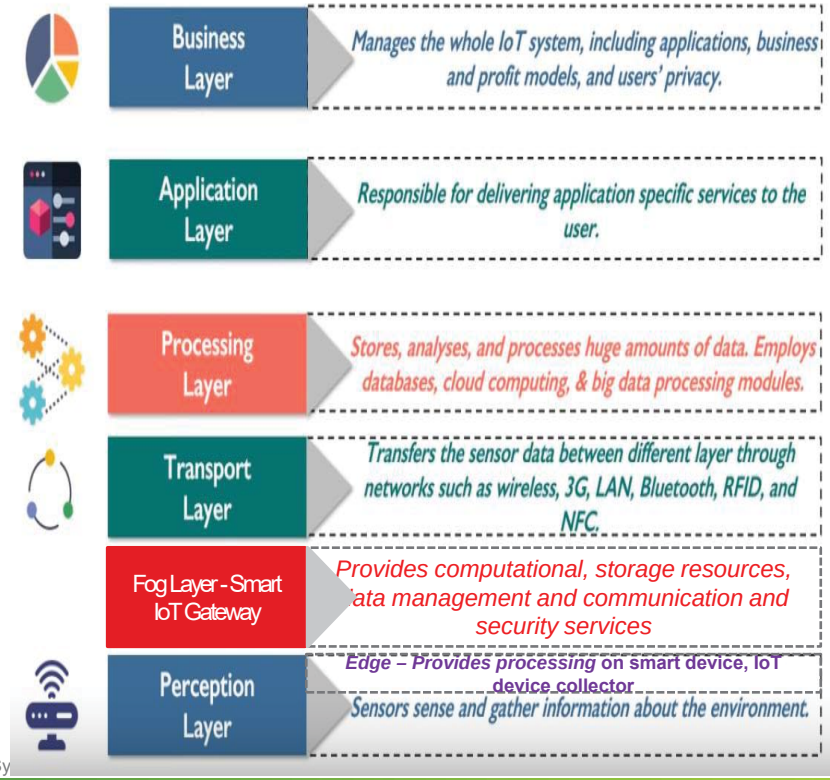
- **Benefits**
  - Address requirements surrounding real-time performance, security and efficiency
  - May be federated to provide horizontal expansion of the functionality over disperse geolocations
  - Reduce bandwidth requirements between gateway and cloud while reducing the resource consumption on the cloud
  - May be configured to communicate directly with other fog nodes to create a mesh that can bypass cloud completely

- **Drawbacks**
  - Increases the complexity of the IoT architecture
  - Introduce more potential points of failure with additional steps and data conversions

EDGE COMPUTING IOT ARCHITECTURE (1/3)

- Distributed computing paradigm closely related to fog computing
- keep certain processing capabilities and functionality closer to **perception layer nodes**
- Can extent to fog, transport and processing layers if powerful
- Operates on "instant data", i.e. real-time data generated by sensors or users.

  While cloud computing operates on big data

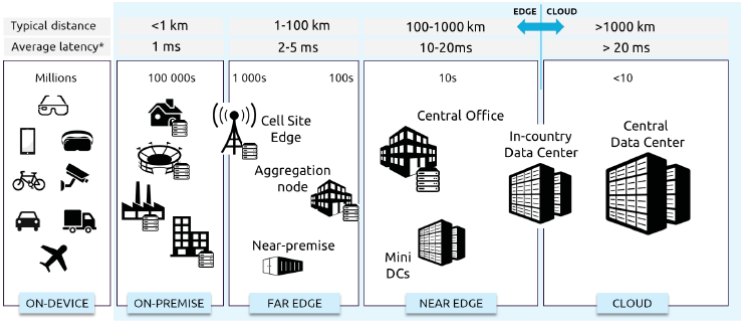


From Embedded-Sy

EDGE COMPUTING IOT ARCHITECTURE (2/3)

Different typologies:

- From microcontrollers to constrained lightweight edge devices to heavy edge nodes with more resources to (micro) datacenter
- From few to many distributed edge nodes

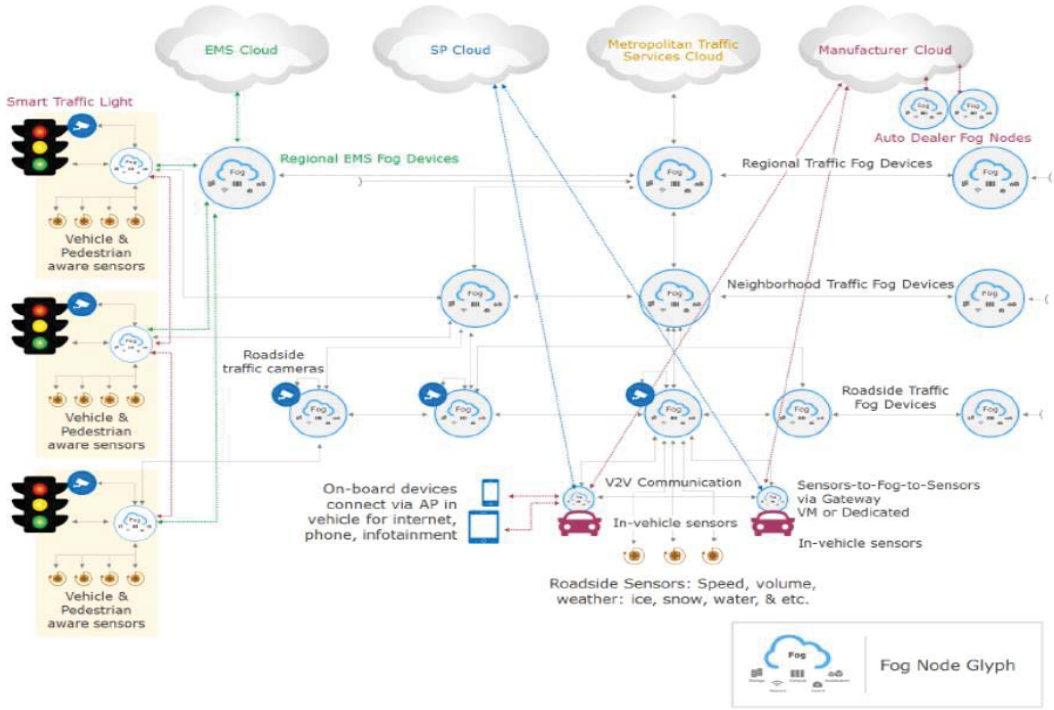


Edge as driver of cloud computing

- Extension of AI and IoT
- Value creation by Multi-Partner / Multi-Company Solutions
- Revolutionizes technologies like 5G, robotics, XR, and other connected devices

EXAMPLE : FROM CLOUD TO FOG TO EDGE

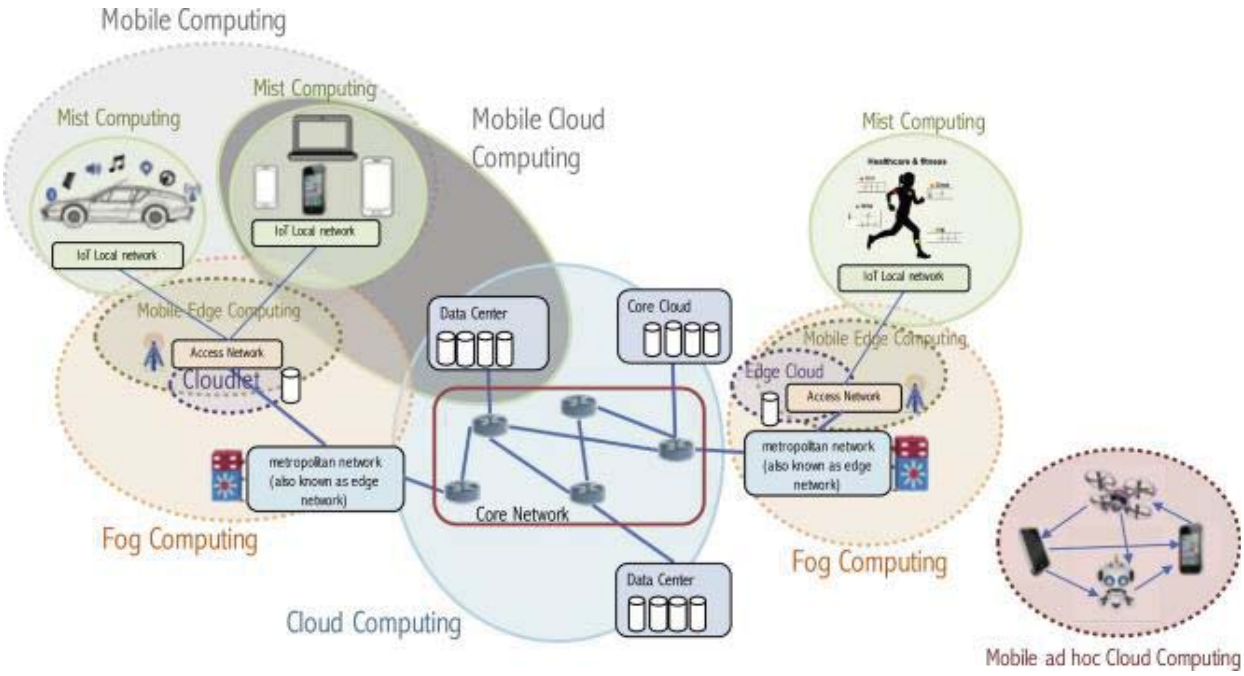
Smart cars and traffic control



* Fog computing, Couse of Università degli Studi di Roma "Tor Vergata"Dipartimento di Ingegneria Civile e Ingegneria Informatica, Valeria Cardellini, 2016/17

EXAMPLE : FROM CLOUD TO FOG TO EDGE

Global metropolitan network



* Ashkan Yousefpour, Caleb Fung, Tam Nguyen, Krishna Kadiyala, Fatemeh Jalali, Amirreza Niakanlahiji, Jian Kong, Jason P. Jue, All one needs to know about fog computing and related edge computing paradigms: A complete survey, Journal of Systems Architecture, Volume 98,2019,Pages 289-330

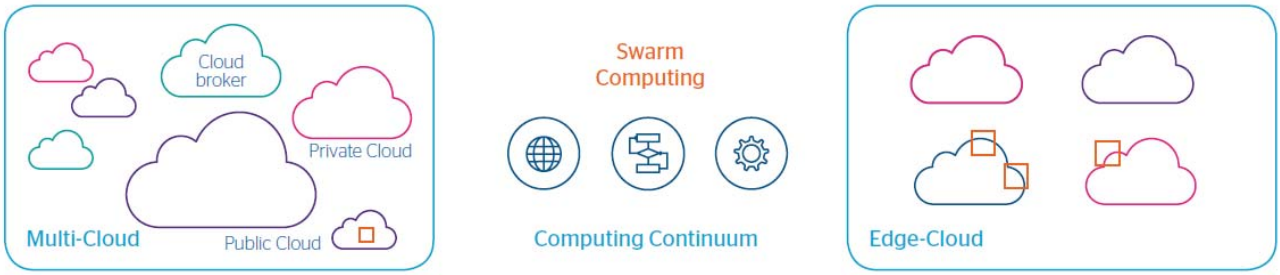
EDGE COMPUTING IOT ARCHITECTURE (3/3)

- **Benefits**
  - Greater capability to reduce localized latencies
  - Allows computing decentralization of computing
  - Greater data privacy
  - Allows for mesh networking off of the cloud

- **Issues**
  - Higher complexity of deployment, management, and orchestration
  - Additional design constraints for data integration resulting from Operational and Information Technologies incompatibility
  - Integration of legacy systems
  - Development, Integration and Operational Costs

TOWARDS SWARM COMPUTING CONCEPT

Swarm computing is a combination of complex multi-cloud architectures with edge computing, as a temporary, limited organization, automatically assembled on-demand to address specific needs.



- Multi-cloud is as a combination of public, private or hybrid cloud solutions working together as a whole to deliver digital capabilities
- Edge computing is as a large network of interconnected devices exchanging and interacting among themselves to fulfil a collective goal (swarm coordination)
  - can be combined and interconnected among them and to diverse clouds

SWARM COMPUTING PRINCIPLES

- Use highly distributed, self-organizing systems of agents that work collaboratively towards a defined outcome.
- Each agent within the system must be able to sense something about its physical or digital environment and only interacts with its local environment.
- Data gathered from a swarm agent may be processed locally or transferred to another specialized central processing application automatically
- The aggregate behavior of the agents leads to the emergence of 'intelligent' global behavior.

SWARM COMPUTING CONCEPT (1/2)

Benefits

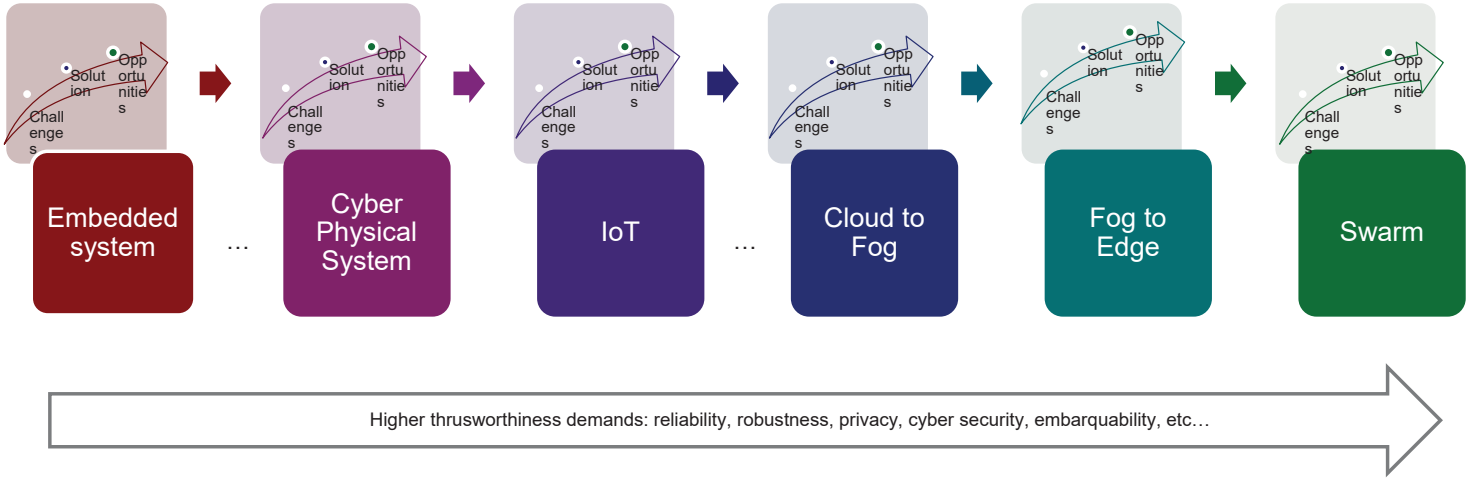
- Creates a dynamic eco-systems of cyber-physical devices and clouds, each adding to the collective capability and insight
- Allows distributed operations and interactions to adapt according to context across simple devices
- Improves efficiency, flexibility and reliability of service provision through:
  - Enabling ad-hoc collaborations, which help built service networks
  - Optimizing delivery schemes and communication patterns, which allow information and services to be shared and exchanged
  - Creating reliability and dependability from volatile resources, which help manage uncertainty
- Complements other forms of artificial intelligence

SWARM COMPUTING CONCEPT (2/2)

Issues

- Complexity of agent-based programming
- Complexity of swarm integration with other centralized control mechanisms
- Security concern due to the distributed control of individual agents within a swarm
- Lack of standardized communication protocols
- Possibility of non-deterministic behaviors, including unexpected or out of control "emergent" behaviors
- Possibility of adverse influence by rogue components on swarm behaviors

A CYCLIC, ITERATIVE EVOLUTION PROCESS



REFERENCES

European industrial technology roadmap for the next generation cloud-edge offering, European Champions Alliance, May 2021

Swarm computing, realization of Computing Continuum, Atos Vision for Future Cloud, Enric Pages, Atos, Research & Innovation, Next Generation Cloud Lab, 2021

Swarm Computing Concepts, technologies and architecture, Whitepaper, Atos scientific community, 2018, https://atos.net/wp-content/uploads/2018/12/atos-swarm-computing-white-paper.pdf

Edge Computing: The Future of Cloud, Gaurav Aggarwal, July 2021, https://gauravagg2016.medium.com/edge-computing-the-future-of-cloud-8fdc321bb601

Driving Industry 4.0 at Distributed Edges with Cloud Orchestration, Industrial Internet consortium, SAP SE, 2021

Architectures in the IoT civilization, Adam Calihman , January 2019, https://www.netburner.com/learn/architectural-frameworks-in-the-iot-civilization/

Cyber-Physical Systems (CPS) vs. IoT, Kwanwoo Lee, 2018, https://kwanwooleecom.wordpress.com/2018/03/03/cyber-physical-systems-cps-vs-iot/

Onik, Md Mehedi Hassan & KIM, Chul-Soo & Yang, Jinhong. (2019). Personal Data Privacy Challenges of the Fourth Industrial Revolution. 635-638. 10.23919/ICACT.2019.8701932.

Inderwildi, Oliver & Zhang, Chuan & Wang, Xiaonan & Kraft, Markus. (2020). The impact of intelligent cyber-physical systems on the decarbonization of energy. Energy & Environmental Science. 13. 10.1039/C9EE01919G.

Applications of Embedded Systems, theenggprojects, 2018, https://www.rs-online.com/designspark/applications-of-embedded-systems-1,

What is an embedded system?, EmmaAshely , June 2011, https://www.rs-online.com/designspark/what-is-an-embedded-system

Alam, Muhammad & Rufino, João & Ferreira, Joaquim & Ahmed, Syed & Shah, Nadir & Chen, Yuanfang. (2018). Orchestration of Microservices for IoT Using Docker and Edge Computing. IEEE Communications Magazine. 56. 118-123.

From Embedded Systems To Swarms: Opportunities And Challenges

Morayo Adedjouma, Reda Nouacer
CEA LIST - DILS

CPS&IoT'2022 Summer School on Cyber-Physical Systems and Internet-of-Things
Budva, Montenegro, June 7-11, 2021

Letizia Jaccheri

Software for a better society

Menti.com 97309805

https://www.ntnu.edu/employees/letizia.jaccheri

- Master in Computer Science Universita' di Pisa 1988 – supervisor V. Ambriola
- PhD Software Engineering Politecnico di Torino 1995 – supervisor S. Gai
- In Norway as exchange student in 1989 – supervisor R. Conradi
- Programmer for two years in late 80'
- Professor at NTNU since 2002
- Department head from 2013 to 2017
- Adjunct Professor at UiT since 2019
- Independent Director of Reply SPA (with 6000 employees) in 2015-2018
- ACM Distinguished Speaker (one out of 200 since 2018) https://speakers.acm.org/speakers/
- Årets døråpner 2019 Trondheim
- Two gender equality prizes in 2021
  - ODA network
  - NTNU gender

2018 Keynote Pakistan

Letizia Jaccheri

CERN 1985

2003 RIK Kunst og Teknologi

From PhD to Professor

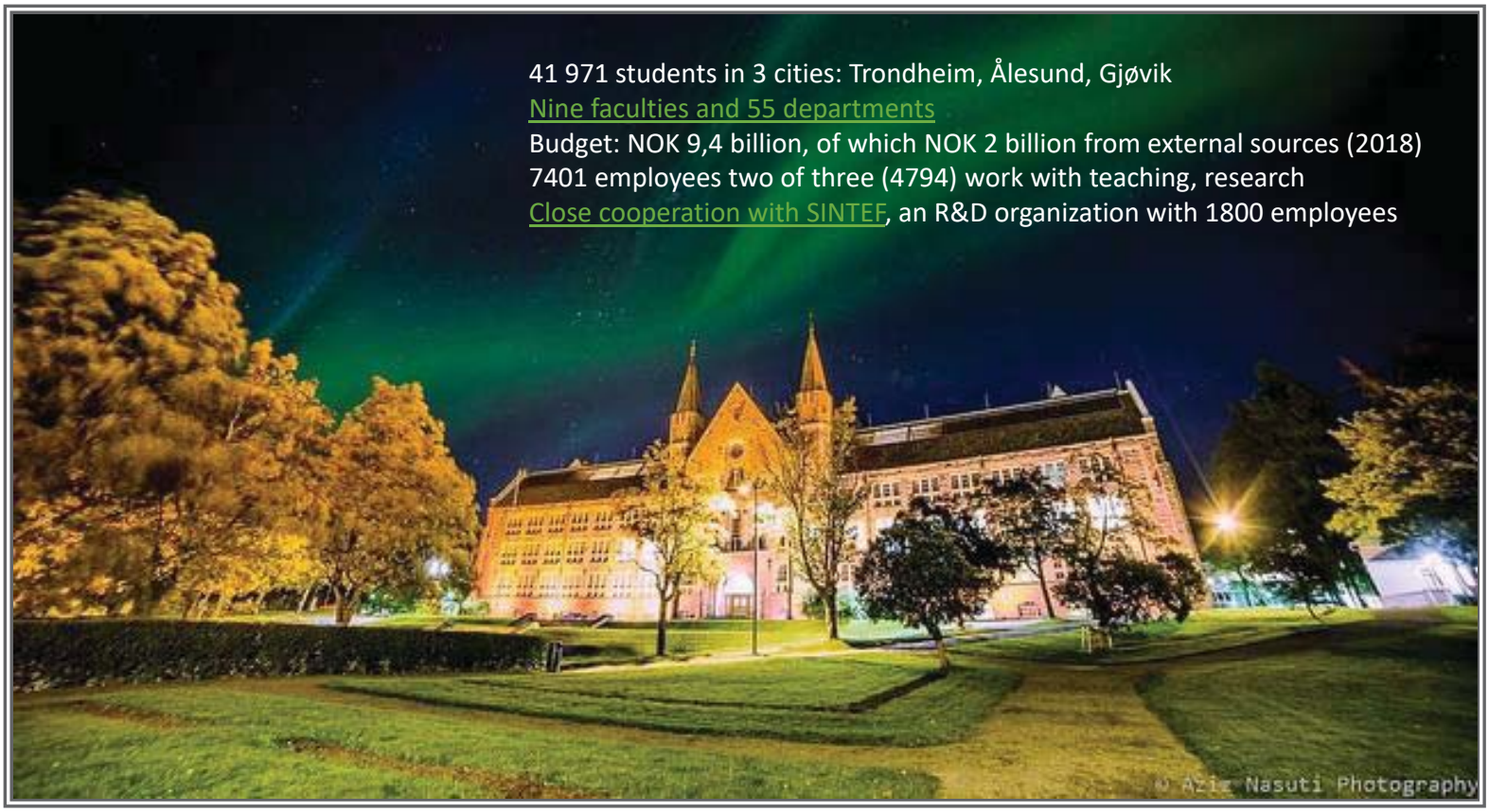
- Conferences & network building
- Stays abroad
- Teaching and dissemination (prize in 2006 for Bok Kjærlighet og Computer see letiziajaccheri.org)
- Research & supervision
- Activist for my values: gender, art, UN goals

Department of Computer Science

- Several study programs:
- 500 students each year
- 350 employees including PhDs

2013 played in meeting room
2016 out of Univ
All countries in the worls
2019 Stock Exchange > 5 Billions NOK

Search engine 1997
2008 Microsoft 6,6 Billions NOK for Fast


41 971 students in 3 cities: Trondheim, Ålesund, Gjøvik
Nine faculties and 55 departments
Budget: NOK 9,4 billion, of which NOK 2 billion from external sources (2018)
7401 employees two of three (4794) work with teaching, research
Close cooperation with SINTEF, an R&D organization with 1800 employees
© Aziz Nasuti Photography

Menti – three questions about the audience
Menti.com 97309805

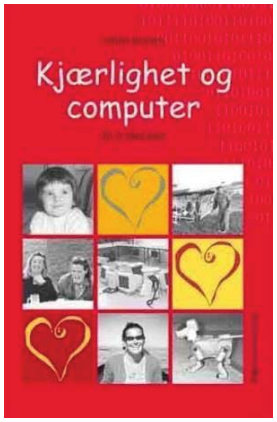
- Pisa, Italy, 60's
- CEP, first Italian Computer
- Social good and culture

17 PARTNERSHIPS FOR THE GOALS

7

1989 Pisa Keith Haring

8



letiziajaccheri.org

Try to understand the relation between life and software

Supervision (facts)

Phd now	5, 2.f, 3.m
Phd total	20, 10.f, 10.m
Phd opponent	20, 10.f, 10.m
Postdoc	7, 1.f, 6.m
Total	52, 23.f, 29.m

https://sbs.idi.ntnu.no/

Projects

E-ladda	Early Language Development	MSCA ITN	Mila Vulcanova
Craft	Creating Actionable Futures	HORIZON-MISS-2021-CIT-01-02	Annemie Wychmans
BALANSE IDUN	From PhD to Professor	NFR 2019 - 2022	Letizia Jaccheri
COST Action CA19122 Eugain	European Network for Gender Balance in Informatics	COST Action 2020 - 2024	Action Chair Letizia Jaccheri


Yiru Wang

SUSTAINABLE CITIES AFTER COVID-19?
INTERNATIONAL HACKATHON

Would you like to experience innovation and entrepreneurship in a global team? Join us in our international 2-day Hackathon with students from Norway and China, and make a real impact in creating safe, resilient, sustainable and gender-equal cities!

Challenge: The Covid-19 pandemic has touched all of humanity. In a post-Covid world, how can we use artificial intelligence (AI) to dramatically boost progress towards achieving the SDGs?

WHO CAN JOIN?

We're looking for all students with passion and motivation! Useful additional skills include sustainability, management, architecture, urban development, programming and more. You can join as an individual or a team.

HOW TO JOIN

Registration for the hackathon as a project leader or participant:
ipit.network/hackathon-intelligent-cities-after-covid-19/

Partners:

For NTNU students, we will provide

TIMELINE

- 8th March — Q&A session for all potential candidates
- 10th March — Deadline for submission of project ideas by group leaders
- 12th March — Presentation of project ideas
- 15th March — Deadline for registration of participants
- 16th March — Independent division of students into mixed and diverse groups
- 20-21st March — Hackathon weekend!

Prizes: Vouchers to spend on training or skills development!
Voucher value:
1st Prize: 15.000 CN¥ (ca. 20.000 NOK)
2nd Prize: 7.500 CN¥ (ca. 10.000 NOK)
3rd Prize: 3.700 CN¥ (ca. 5.000 NOK)

Sbs.idi.ntnu.no
Cybersecurity for children
Blockchain and UN goals
Software and gender

Jaccheri, M. L., & Conradi, R. (1993). Techniques for process model evolution in EPOS. IEEE Transactions on Software Engineering, 19(12), 1145-1156.

Carver, Jeffrey C; Jaccheri, Letizia; Morasca, Sandro; Shull, Forrest; A checklist for integrating student empirical studies with research and teaching goals, Empirical Software Engineering 15.1 (2010): 35-59.

Berg, V., Birkeland, J., Nguyen-Duc, A., Pappas, I., & Jaccheri, L. (2018). Software Startup Engineering: A Systematic Mapping Study Journal of Systems and Software 144 (2018): 255-274.

Menti – three questions about software

Menti.com 97309805


art

Art and Technology have been in contact for centuries

Between 1800 and 1900:
camera, film, and telephone
Romanticism (Goethe and Beethoven)
After 1900 *Modernism*
Filippo Tommaso Marinetti *Futurist*
Duchamp

Video The Lumiere Brothers' - First films (1895) on youtube

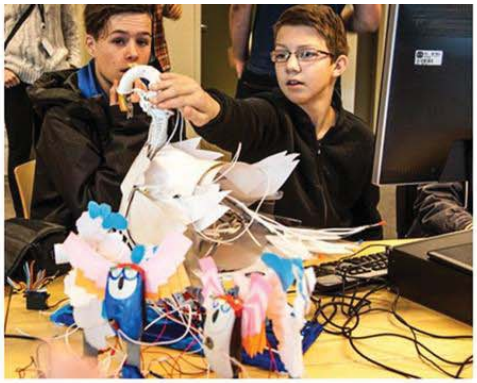

1965 First Computer Art Exhibition
Trifonova, Anna; Jaccheri, Letizia; Bergaust, Kristin;
Software engineering issues in interactive installation art,
2008International Journal of Arts and Technology 1.1 (2008): 43-65.

How can we improve the development process of software dependent artworks and projects, in terms of software development, maintenance, upgrade and usability of the artwork?

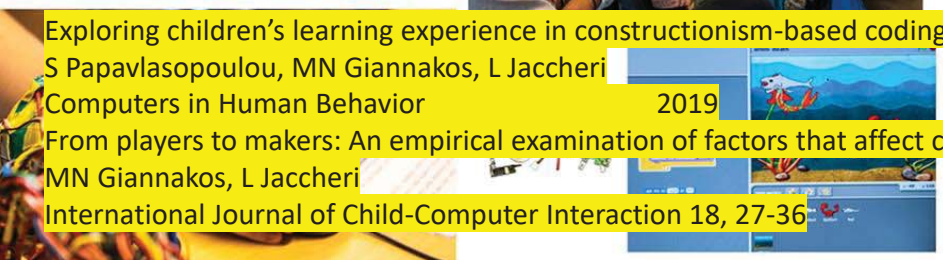
Ahmed, Salah Uddin; Jaccheri, Letizia; M'kadmi, Samir; Sonic onyx: Case study of an interactive artwork, 2009International Conference on Arts and Technology. Springer, Berlin, Heidelberg, 2009.

From players to makers: which factors do affect creative game development?

Papavlasopoulou, S., Giannakos, M. N., & Jaccheri, L. (2017). Empirical studies on the Maker Movement, a promising approach to learning: A literature review. *Entertainment Computing*, *18*, 57-78.

KODELØYPA
2014 - ONGOING



Art & Recycling
Coding & Interaction

Exploring children's learning experience in constructionism-based coding activities through design-based research
S Papavlasopoulou, MN Giannakos, L Jaccheri
Computers in Human Behavior 2019

From players to makers: An empirical examination of factors that affect creative game development
MN Giannakos, L Jaccheri
International Journal of Child-Computer Interaction 18, 27-36

Girl project ADA

2004	2019
7%	36%

Average %-share 5-year integrated Master
- Computer Sciences
- Communication Technologies
- Cybernetics and Robotics
- Electronics System design and innovation

- Calling the applicants
- Welcome day
- 8th March - Women´s Day
- Networking lunches
- Programming courses
- Mountain hiking
- CodeHubs
- PhD-party

- Invite girls from high school from all over the country
- 3 days
- Presentations and workshops
- Personal meeting with rolemodels
- Meeting students
- Break down stereotypes
- Hands-on experiences

www.ntnu.edu/girls


Solution?
The Girl Project Ads!

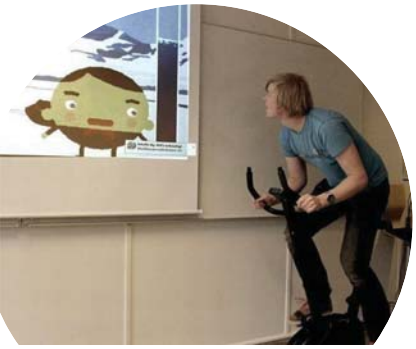
Menti – three questions about art

Menti.com 97309805

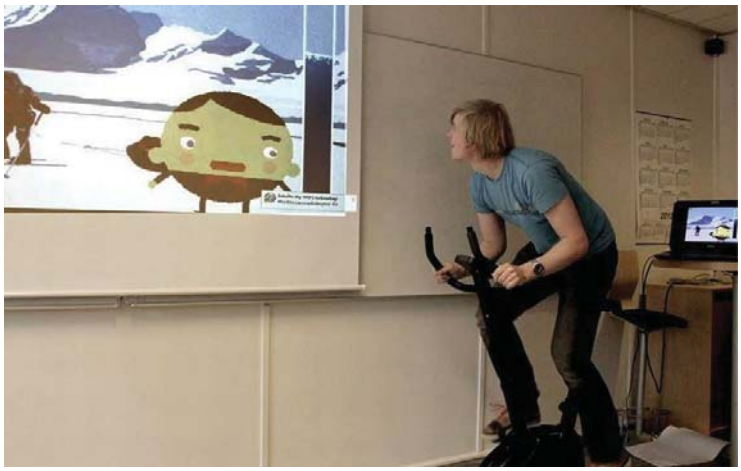
Sustainability goals

Liv Arnesen

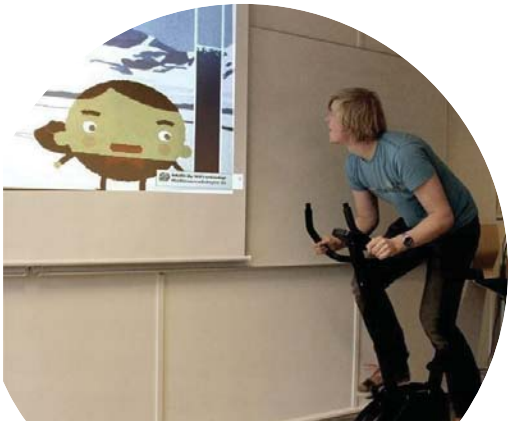
How can we design and evaluate software that becomes a medium to engage and inform the user?



From art to sustainability



Hagen, K., Chorianopoulos, K., Wang, A. I., Jaccheri, L., & Weie, S. (2016, May). Gameplay as exercise. In Proceedings of the 2016 CHI conference extended Abstracts on human factors in computing systems (pp. 1872-1878).



SOCRATIC

The main goal of SOCRATIC project is to provide citizens and organizations a collaborative space where they can identify innovative solutions to achieve the Sustainable Development Goals set by the United Nations.

UN Goal 5:

Achieve **gender equality** and empower all **women** and girls. Ending all discrimination against **women** and girls is not only a basic human right, it's crucial for sustainable future; it's proven that empowering **women** and girls helps economic growth and development.

Software engineering and gender

- People decide requirements
- People develop solutions for people
- People Interact with systems

Software engineering and gender

- People decide requirements
- People develop solutions for people
- People Interact with systems

Question: who decides the requirements and for which people?

Designing Software to Prevent Child Marriage Globally, J Brevik, L Jaccheri, JCT Vidal, Proceedings of the 18th ACM International

Software engineering and AI

People
- Decide requirements
- Develop solutions for people
- Interact with systems

Computer system
- learns

腾讯
Tencent

big data

Question: who decides how the system will learn?

Technology with gender biases

ACTIVITY TRACKERS THAT FAIL TO MEASURE STEPS IN THE, PREDOMINANTLY FEMALE, ACTIVITY OF PUSHING A STROLLER.

TRANSPORT NETWORKS THAT IGNORE THE SO-CALLED "MOBILITY OF CARE"

AI RECRUITING TECHNOLOGY DEVELOPED TRAINED PREDOMINANTLY ON MEN'S RÉSUMÉS

EU REPORT OF THE EXPERT GROUP "INNOVATION THROUGH GENDER"

Do these considerations apply to Software Development companies in your country?

Engineering and Technology: on average, in the whole of Europe, women take less than 15% of the full professor positions

Figures show that in 2016, an overwhelming majority (83.3%) of ICT specialists employed in the EU were men.

Skills and talent gap: 53% of European employers say they face difficulties in finding the right people with the right qualifications.

Software is

- Intelligence
- Interconnection
- Efficiency
- Experience

Menti – three questions about sustainability

Menti.com 97309805


European Network
for Gender Balance
in Informatics
EUGAIN
eugain.eu


1 NO POVERTY
G 1 2 3 4 5
G 1 2 3 4 5
2 ZERO HUNGER
G 1 2 3 4 5
G 1 2 3 4 5
3 GOOD HEALTH AND WELL-BEING
G 1 2 3 4 5 6 7 8 9
G 1 2 3 4 5 6 7 8 9
4 QUALITY EDUCATION
G 1 2 3 4 5 6 7
G 1 2 3 4 5 6 7
5 GENDER EQUALITY
G 1 2 3 4 5
G 1 2 3 4 5
6 CLEAN WATER AND SANITATION
G 1 2 3 4 5 6
G 1 2 3 4 5 6
7 AFFORDABLE AND CLEAN ENERGY
G 1 2 3
G 1 2 3
8 DECENT WORK AND ECONOMIC GROWTH
G 1 2 3 4 5 6 7 8 9 10
G 1 2 3 4 5 6 7 8 9 10
9 INDUSTRY, INNOVATION AND INFRASTRUCTURE
G 1 2 3 4 5
G 1 2 3 4 5
10 REDUCED INEQUALITIES
G 1 2 3 4 5 6 7
G 1 2 3 4 5 6 7
11 SUSTAINABLE CITIES AND COMMUNITIES
G 1 2 3 4 5 6 7
G 1 2 3 4 5 6 7
12 RESPONSIBLE CONSUMPTION AND PRODUCTION
G 1 2 3 4 5 6 7 8
G 1 2 3 4 5 6 7 8
13 CLIMATE ACTION
G 1 2 3
G 1 2 3
14 LIFE BELOW WATER
G 1 2 3 4 5 6 7
G 1 2 3 4 5 6 7
15 LIFE ON LAND
G 1 2 3 4 5 6 7 8 9
G 1 2 3 4 5 6 7 8 9
16 PEACE, JUSTICE AND STRONG INSTITUTIONS
G 1 2 3 4 5 6 7 8 9 10
G 1 2 3 4 5 6 7 8 9 10
17 PARTNERSHIPS FOR THE GOALS
G 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19
G 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19
Positive impact
No positive impact
Negative impact
No negative impact

Orges Cico, Letizia Jaccheri

NTNU
Norwegian University of Science and Technology

Anh Nguyen Duc

University of South-Eastern Norway

Software Sustainability in Customer-Driven Courses

Software is becoming ever more ubiquitous and sustainability is an emergent topic

Ensuring dissemination of sustainable software developments

We review of project descriptions from a project-based course

Customers' project descriptions

- Social sustainability moderately addressed
- Technical sustainability addressed by a little more than half of total projects
- Environmental and economic sustainability are not addressed

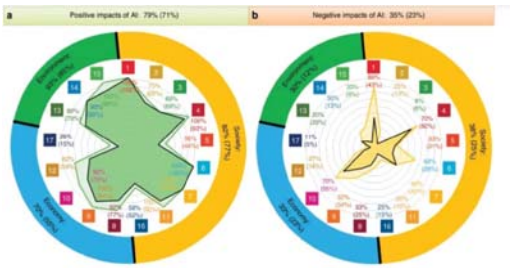
Cico, O., Jaccheri, L. and Duc, A.N., 2021, June. Software Sustainability in Customer-Driven Courses. In *2021 IEEE/ACM International Workshop on Body of Knowledge for Software Sustainability (BoKSS)* (pp. 15-22). IEEE.

Research work about sustainability

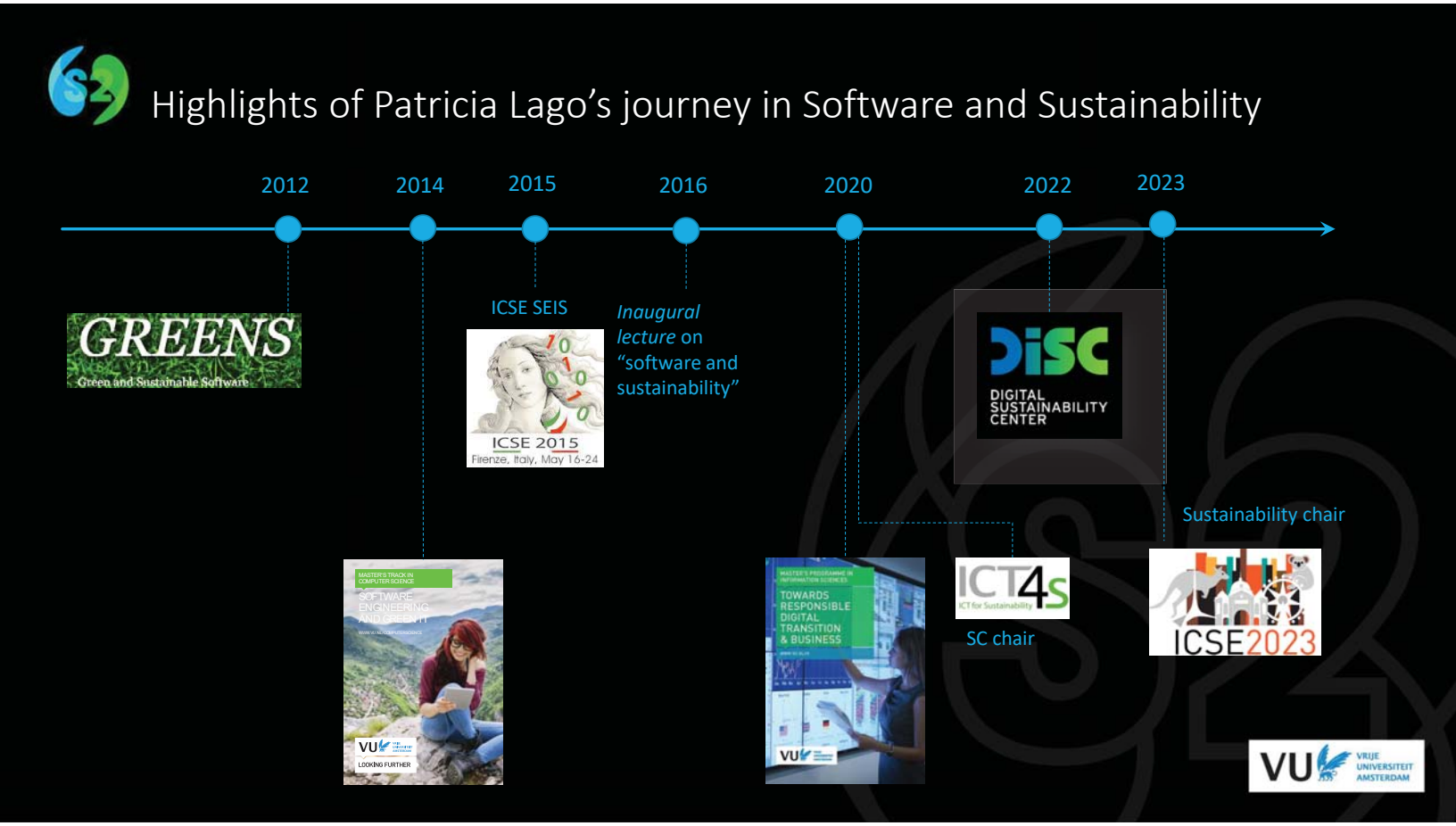
- Patricia Lago

GREENS
Green and Sustainable Software

- Ricardo Vinuesa



- Birgit Penzenstadler


Highlights of Patricia Lago's journey in Software and Sustainability
2012
2014
2015
2016
2020
2022
2023
GREENS
Green and Sustainable Software
ICSE SEIS
ICSE 2015
Firenze, Italy, May 16-24
Inaugural lecture on "software and sustainability"
DISC
DIGITAL SUSTAINABILITY CENTER
Sustainability chair
SOFTWARE ENGINEERING AND GREEN IT
VU
LOOKING FURTHER
TOWARDS RESPONSIBLE DIGITAL TRANSITION & BUSINESS
ICT4S
ICT for Sustainability
SC chair
ICSE2023
VU
VRIJE UNIVERSITEIT AMSTERDAM

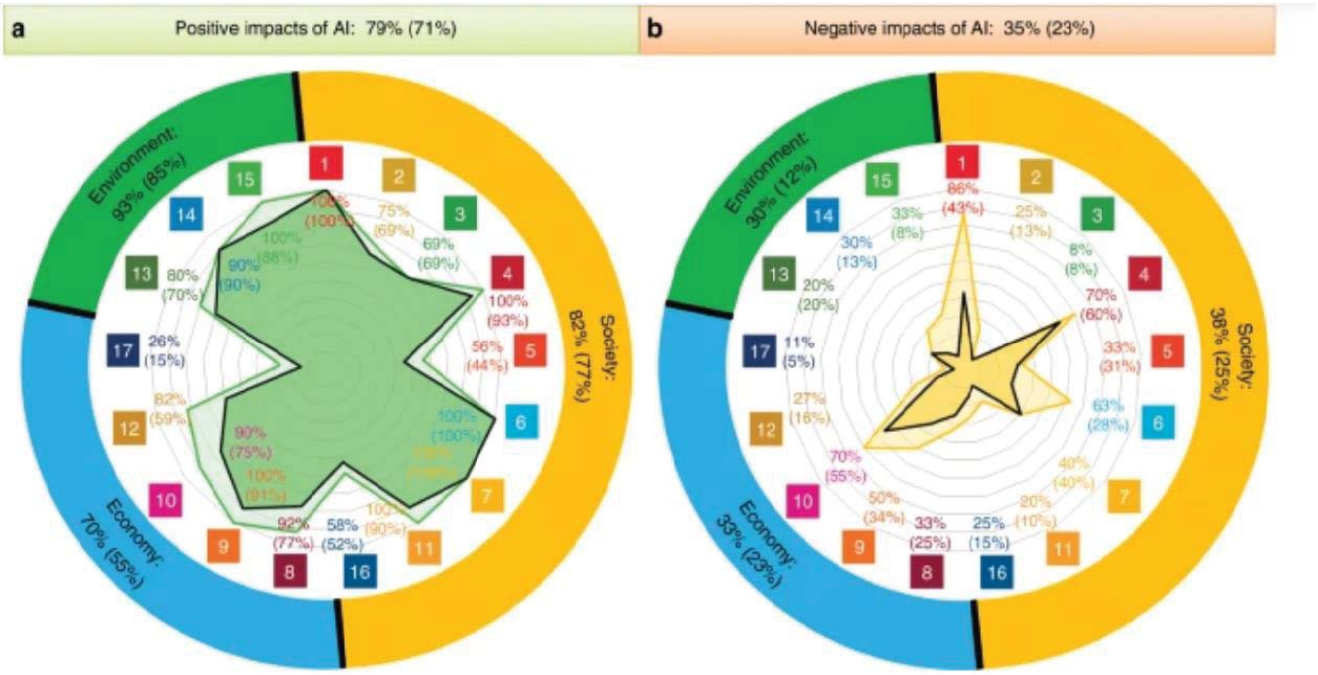

"LIKE PERFORMANCE, RELIABILITY, SECURITY,
SUSTAINABILITY DOES NOT JUST HAPPEN
UNLESS WE PLAN FOR IT."
Patricia Lago @ 2016 Inaugural speech
VU VRIJE UNIVERSITEIT AMSTERDAM

Vinuesa, R. et al (1)

Research interdisciplinary

- Ricardo Vinuesa, Hossein Azizpour, Iolanda Leite, Madeline Balaam, Virginia Dignum, Sami Domisch, Anna Felländer, Simone Daniela Langhans, Max Tegmark & Francesco Fuso Nerini
- All sustainable goals
- AI
- 10 authors from
  - Sweden, New Zeland, Germany, Usa, Spain
  - Mechanics, robotics, interaction design, AI, Fishery and ecology, zoology, energy

Vinuesa, R. et al (2)



Vinuesa, R. et al (3)

- Supplementary Data
- SDGs: Society, Economy, Environment
- Targets

There is another important shortcoming of AI in the context of SDG 5 on gender equality: there is insufficient research assessing the potential impact of technologies such as smart algorithms, image recognition, or reinforced learning on discrimination against women and minorities. For instance, machine-learning algorithms uncritically trained on regular news articles will inadvertently learn and reproduce the societal biases against women and girls, which are embedded in current languages. Word embeddings, a popular technique in natural language processing, have been found to exacerbate existing gender stereotypes[2].

Bolukbasi, T., Chang, K.-W., Zou, J., Saligrama, V. & Kalai, A. Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. Adv. Neural Inf. Process. Syst. 29, 4349–4357 (2016).

Karlskrona Alliance (est. 2014)

1) Manifesto

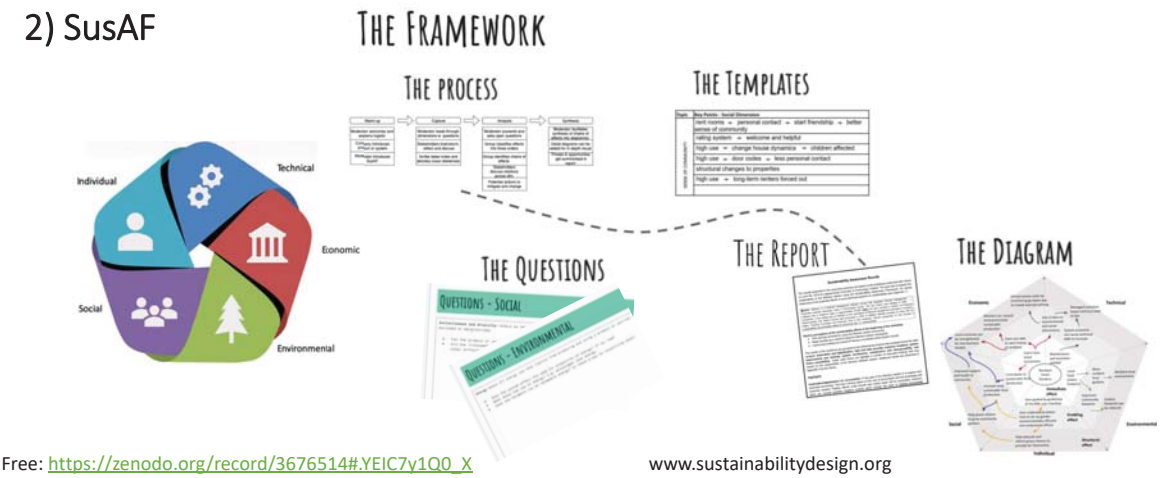
KARLSKRONA MANIFESTO
FOR SUSTAINABILITY DESIGN

As designers of software technology, we are responsible for the long-term consequences of our designs.

As software practitioners and researchers, we are part of the group of people who design the software systems that run our world. Our work has made us increasingly aware of the impact of these systems and the responsibility that comes with our role, at a time when information and communication technologies are shaping the future. We struggle to reconcile our concern for planet Earth and society with the work that we do. Through this work we have come to understand that we need to redefine the narrative on sustainability and the role it plays in our profession.

Read the Manifesto | Become a signatory

2) SusAF



Free: https://zenodo.org/record/3676514#.YEIC7y1Q0_X

www.sustainabilitydesign.org

Duboc, L., Penzenstadler, B., Porras, J., Kocak, S. A., Betz, S., Chitchyan, R., ... & Venters, C. C. (2020). Requirements engineering for sustainability: an awareness framework for designing software systems for a better tomorrow. Requirements Engineering, 1-24.

Advice

- To work with UN goal 5
  - Know your numbers and Set your goals
  - Keep and empower the female you have
  - LGBT+ (LGBT stands for lesbian, gay, bisexual and transgender/transsexual people)
  - Look for funds, connections, projects
  - Celebrate
  - Document
  - Everything is research
- To be an happy professor
  - Say yes and say no
  - Keep and empower the students you have
  - Connect education and research
  - Look for funds, connections, projects
  - Celebrate
  - Document
  - Everything is research
  - Be happy for the success of other people, expecially your (ex)students

The Distinguished Speakers Program is made possible by

Association for Computing Machinery

Advancing Computing as a Science & Profession

For additional information, please visit http://dsp.acm.org/

Distinguished Speaker Association for Computing Machinery (ACM).
https://speakers.acm.org/speakers/
ACM 100,000 members
ACM Speakers: IBM, Microsoft, Stanford University, Carnegie Mellon, University of British Columbia, Tsinghua University
Two ACM Distinguished Speakers in Norway
Letizia with her lecture From Software through Art to Social Entrepreneurship.
There are a total of 200 distinguished speakers world wide.

References

- Jaccheri, M. L., & Conradi, R. (1993). Techniques for process model evolution in EPOS. IEEE Transactions on Software Engineering, 19(12), 1145-1156. SW
- Carver JC, Jaccheri L, Morasca S, Shull F. A checklist for integrating student empirical studies with research and teaching goals. Empirical Software Engineering. 2010 Feb;15(1):35-59. SW
- Berg, V., Birkeland, J., Nguyen-Duc, A., Pappas, I. O., & Jaccheri, L. (2018). Software startup engineering: A systematic mapping study. Journal of Systems and Software, 144, 255-274. SW
- Berg, V., Birkeland, J., Nguyen-Duc, A., Pappas, I. O., & Jaccheri, L. (2020). Achieving agility and quality in product development-an empirical study of hardware startups. Journal of Systems and Software, 167, 110599. SW
- Trifonova, A., Jaccheri, L., & Bergaust, K. (2008). Software engineering issues in interactive installation art. International Journal of Arts and Technology, 1(1), 43-65. ART
- Hagen, K., Chorianopoulos, K., Wang, A. I., Jaccheri, L., & Weie, S. (2016, May). Gameplay as exercise. In Proceedings of the 2016 CHI conference extended Abstracts on human factors in computing systems (pp. 1872-1878). UN G 3
- Papavlasopoulou, S., Giannakos, M. N., & Jaccheri, L. (2017). Empirical studies on the Maker Movement, a promising approach to learning: A literature review. Entertainment Computing, 18, 57-78. UN G 4 & 5
- Jaccheri, L., Pereira, C., & Fast, S. (2020). Gender Issues in Computer Science: Lessons Learnt and Reflections for the Future. In 2020 22nd International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC) (pp. 9-16). IEEE. UN G 5
- Cico, O., Jaccheri, L., Nguyen-Duc, A., & Zhang, H. (2021). Exploring the intersection between software industry and Software Engineering education-A systematic mapping of Software Engineering Trends. Journal of Systems and Software, 172, 110736. Software
- Pappas I, Jaccheri ML, Mikalef P, Giannakos M. Social innovation and social entrepreneurship through big data: developing a reseach agenda. InThe 11th Mediterranean Conference on Information Systems (MCIS) 2017. Association for Information Systems. UN Gs
- Brevik J, Jaccheri L, Vidal JC. Designing software to prevent child marriage globally. In Proceedings of the 18th ACM International Conference on Interaction Design and Children 2019 Jun 12 (pp. 452-457). UN G 5
- Lund EH, Jaccheri L, Li J, Cico O, Bai X. Blockchain and sustainability: A systematic mapping study. In2019 IEEE/ACM 2nd International Workshop on Emerging Trends in Software Engineering for Blockchain (WETSEB) 2019 May 27 (pp. 16-23). IEEE. UN Gs
- Cico, O., Jaccheri, L. and Duc, A.N., 2021, June. Software Sustainability in Customer-Driven Courses. In *2021 IEEE/ACM International Workshop on Body of Knowledge for Software Sustainability (BoKSS)* (pp. 15-22). IEEE. UN Gs


MECO 2022

11th Mediterranean Conference on Embedded Computing

7-10 June 2022, Budva, Montenegro

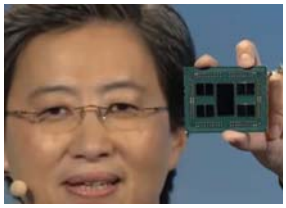
Extending Performance and Reliability via Modular FPGA Clusters

Roberto Giorgi
University of Siena, Italy
http://www.dii.unisi.it/~giorgi

MOTIVATION

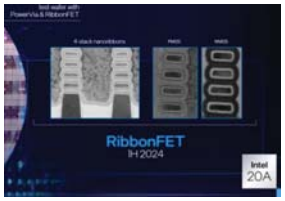
- Several embedded applications and systems are deployed in scenarios where the system may be subject to several sources of faults
  - E.g., mission-critical applications, space, autonomous cars, …..
- Systems use more and more smaller microelectronics (e.g., 2 nm)
- PLATFORMS
  - Multicore and tightly interconnected Multicores

What the Electronics provides us

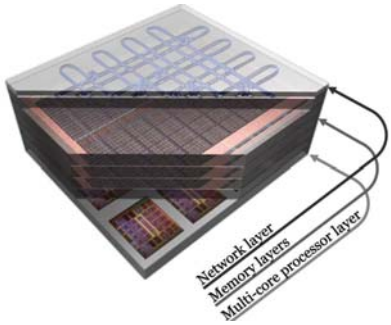


- Integration: 3D stacking → Chiplets + interposer (2.5D), DDR4/DDR5 are 3D stacked

- Technology node: 5(-)nm → 5nm (2nm projected for 2024)



- Transistor: FinFET → FinFET (GAA-FET/RibbonFET projected)



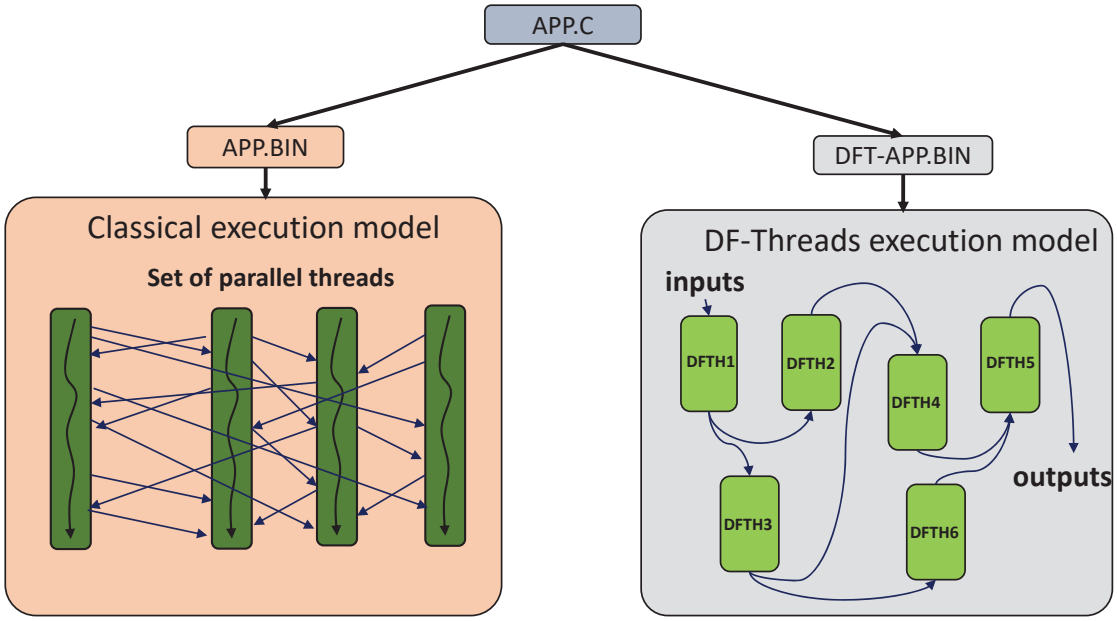
G. Hendry, K. Bergman, "Hybrid On-chip Data Networks", HotChips-22, Stanford, CA – Aug. 2010

→As systems get smaller, faults may become much more frequent

DF-THREADS

(Data Flow Threads)

How DF-Threads work (simplified overview)

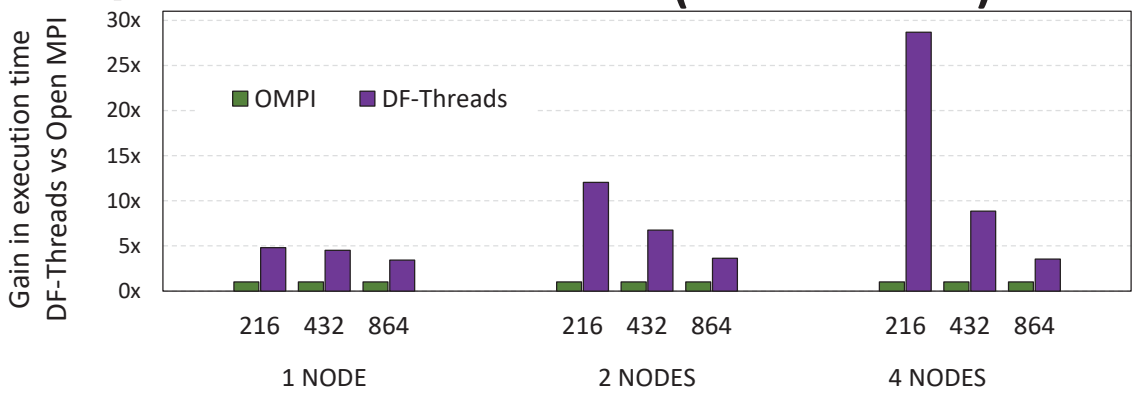


- Every instruction can write in any memory location
- High synchronization activity
- Coherency needed on multicores or DSMs
- 1 single instructions failing →the whole system to fails

- Regularization of data exchange
- Less synchronization activity
- No coherency needed
- Idempotency property (more resilience, lighter checkpointing)

Performance Motivation

→ DATAFLOW-THREADS (DF-THREADS)



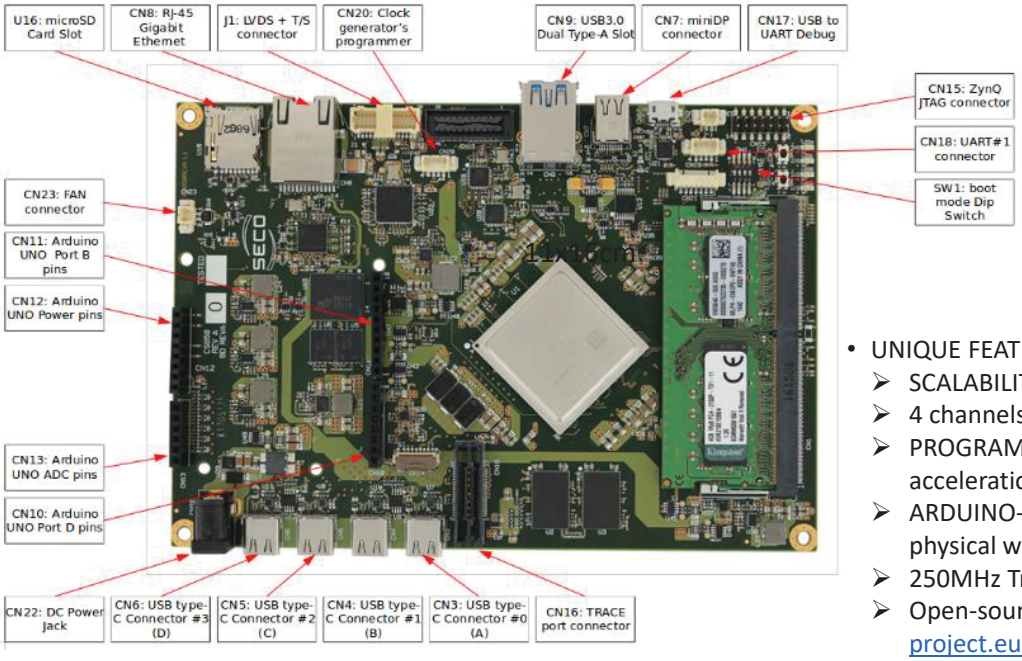
Gain in execution time of DF-Threads compared to Open MPI. The benchmark is Blocked Matrix Multiplication for different sizes of the square matrices (216, 432, 864 – the block size is 8 elements) and 1, 2, 4 nodes. FULL-SYSTEM simulation.

- Standard stacks (MPI) have much overhead
- Checkpointing and lockstep-execution may introduce lot of overhead
- Proposal: using a different EXECUTION model by grouping data processing at the THREAD level

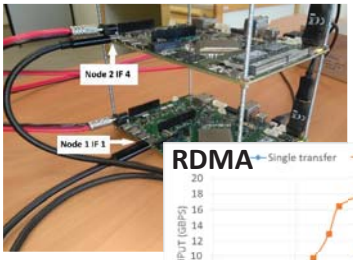
R. Giorgi, "Scalable Embedded Computing through Reconfigurable Hardware: comparing DF-Threads, Cilk, OpenMPI and Jump", ***ELSEVIER Microprocessors and Microsystems***, vol. 63, Aug. 2018, pp. 66-74.

THE AXIOM BOARD

2015-2018 -- 4Meuro funding
Coordinator: Roberto Giorgi
University of Siena (Coordinator Partner)



Bidirectional links



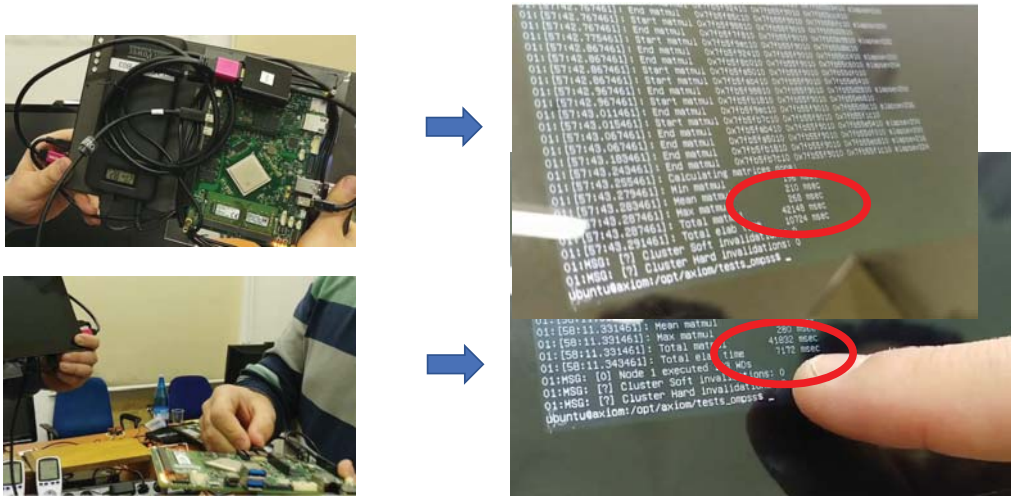
- UNIQUE FEATURES
  - SCALABILITY via USB-C cable: building clusters up to 255 boards
  - 4 channels on USB-C cable @ 18Gbps (custom protocol)
  - PROGRAMMABILITY via OpenMP: TRANSPARENT FPGA acceleration + CLUSTER distribution
  - ARDUINO-UNO socket on board for easy interfacing with the physical world
  - 250MHz Trace port – Lauterbach compatible
  - Open-source software stack + BSP: https://git.axiom-project.eu/ (10^6+ Lines of C Code!)

http://www.axiom-project.eu/AXIOM_BOARD_GUIDE.pdf

✓ DELIVERED March 2018

On-the-fly: node addition/removal

- 1 board → ~11s, 2 board(USB-C connected)→ ~7s (no optimizations)
- SAME board type, SAME software, NO programming efforts
- Experimenting embedded and cyber-physical scenarios

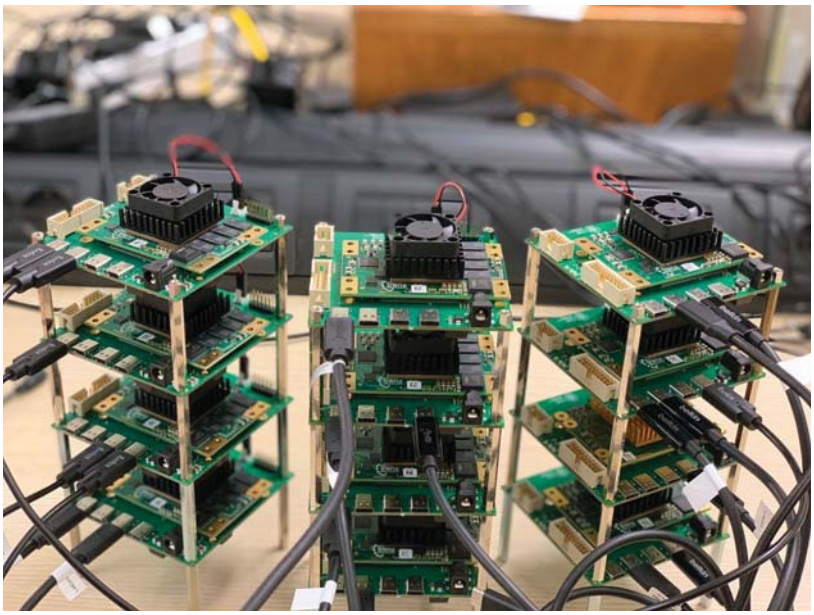


NOTE: this is using a SHARED-MEMORY programming model based on OpenMP (not message-passing!) ACROSS BOARDS

R. Giorgi, "Scalable Embedded Computing through Reconfigurable Hardware: comparing DF-Threads, Cilk, OpenMPI and Jump", ***ELSEVIER Microprocessors and Microsystems***, vol. 63, Aug. 2018, pp. 66-74.

NEW HARDWARE

12-node GLUON-B cluster (2020-22)

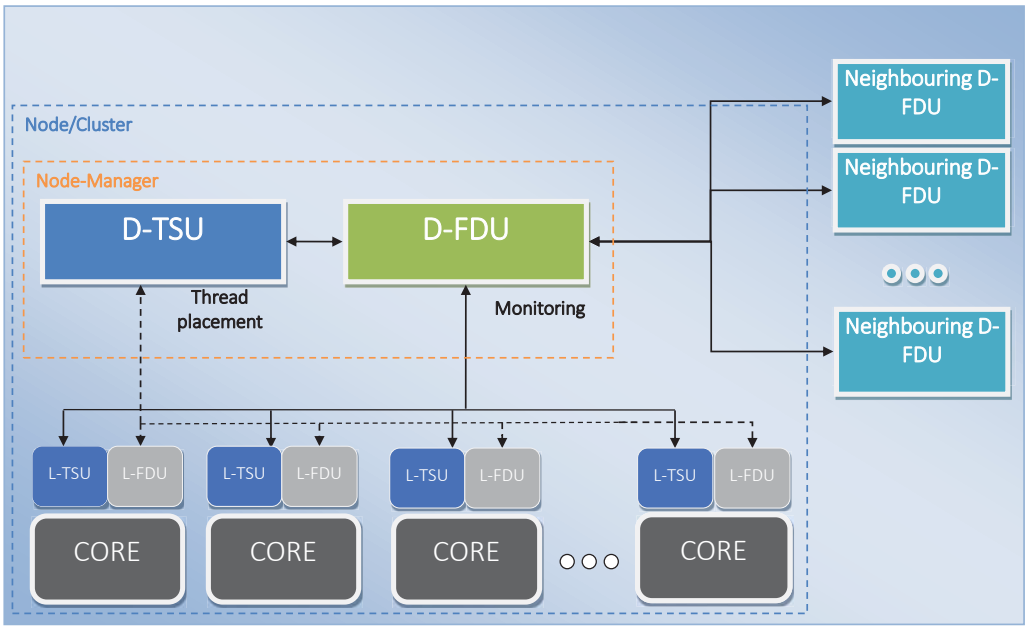


Simplified interconnects: arbitrary topologies are possible up to 255 boards
An interconnect self-discovery algorithm creates local routing tables (by EVIDENCE)

Architecture of a Dataflow Fault Detection Unit (FDU)

Fault Detection Unit (FDU) – System Architecture

TSU → supports the Dataflow Execution (what seen so far)
L-FDU+L-TSU → core level, D-TSU+D-TSU → node level

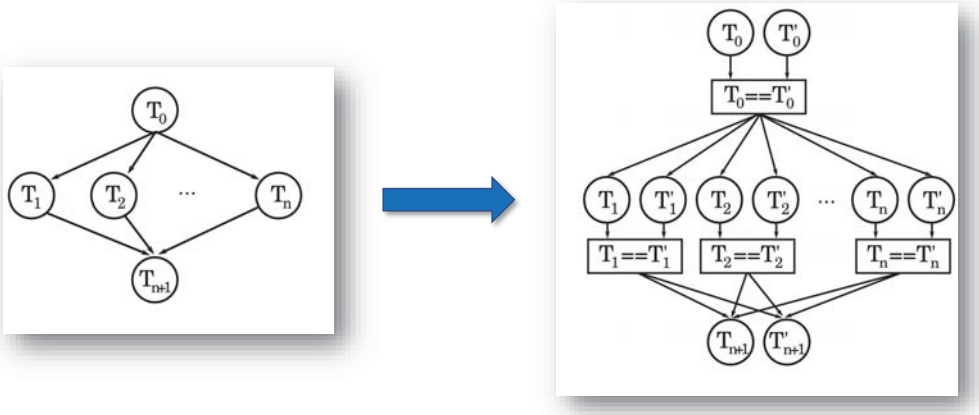


FDU Overall Objectives

- Quantitative Objectives
  - QO1: lowering the number of faults by 90% compared to the same overall multi-/many-core processor without reliability techniques
  - QO2: running parallel dataflow programs on a chip in a graceful degradation style even when more than 50% of the cores are permanently faulty as long as the cores are still interconnected
- Focus: fault detection of permanent, intermittent and transient faults

Core-Level Fault Detection by Dataflow Double Execution (DDX)

double execution approach of coarse-grained dataflow threads is a loosely-coupled thread-level redundant execution scheme, which exploits the dataflow execution model and provides support for parallel dataflow applications in order to detect permanent, intermittent, and transient faults



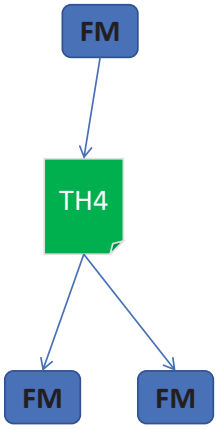
Exploits the **DF-Threads Execution Model** for

- **Input Replication**
- **Redundant Thread Synchronization**
- **Output Comparison**

DDX only needs to duplicate the ***continuation*** (cf. next slide) of a thread: after the continuation has been copied, both threads can be independently scheduled to different cores for execution, while sharing the same thread frame as input data → this means we can additionally exploit data locality by sharing

[Weis14-ijpp-Architectural Support for Fault Tolerance in a Teradevice Dataflow System]

DF-Threads



- DF-thread
  - A function that expects no parameters and returns no parameters.
    - The body of this function can refer to any memory location for which it has got the pointer through some API function calls (e.g., df_schedule, df_ldframe, ...); a DF-Thread is identified by an object of type dft_t (DF-Thread identifier). In other words:

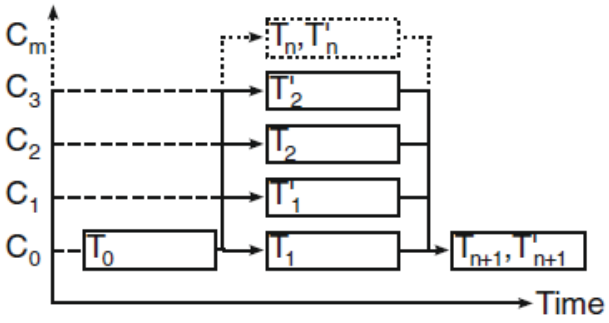
```
typedef void (*dft_t)(void)
```

- INPUT_FRAME, OUTPUT_FRAME (a.k.a. CONTINUATIONS)
  - INPUT_FRAME: a buffer which is allocated in the local memory and contains the input values for the current DF-Thread.
  - OUTPUT_FRAME: a buffer which is allocated in the local memory and contains values to be used by other DF-Threads (consumer DF-Threads)

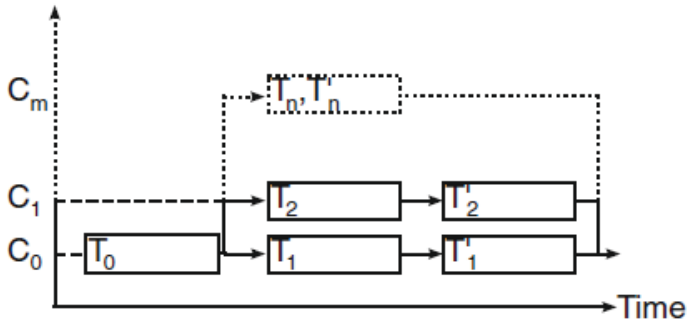
- SYNCHRONIZATION_COUNT
  - A number which is initially set to the number of input values (or events) needed by a DF-Thread. The SYNCHRONIZATION_COUNT must be decremented each time the expected data is written in an OUTPUT_FRAME

Dynamic spatial and temporal redundancy of DDX

REPLICATION IN SPACE

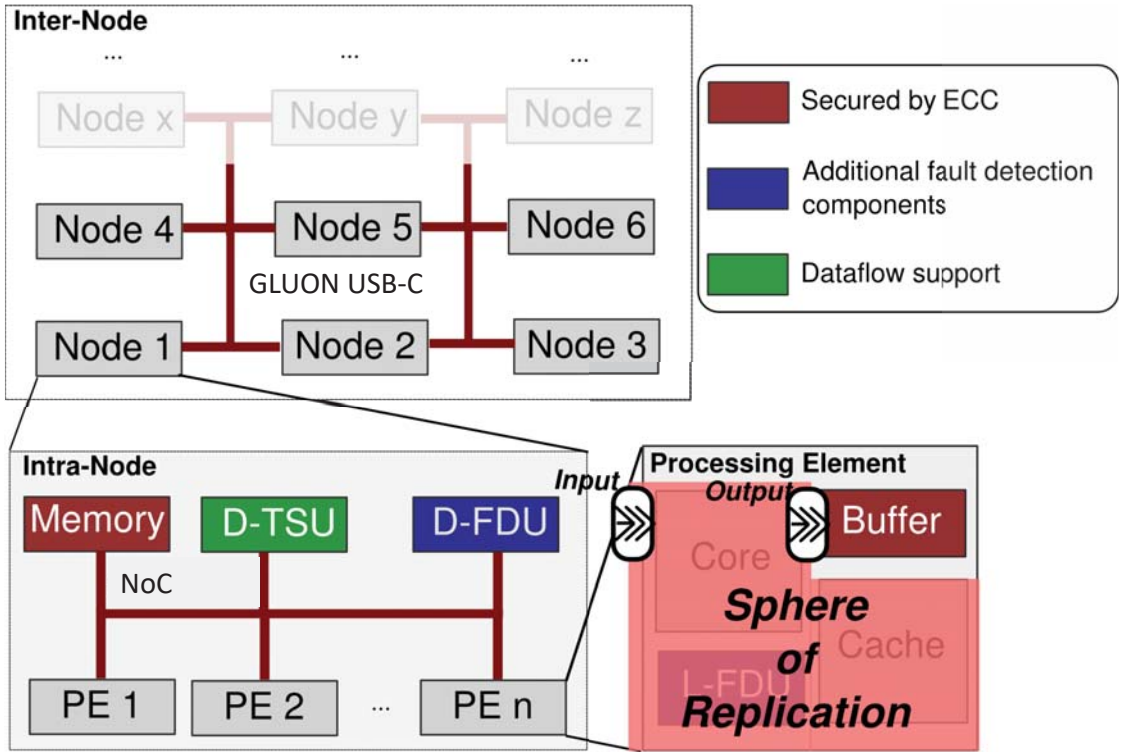


REPLICATION IN TIME



- If it is not possible to schedule all waiting threads to idle cores, since the thread-level parallelism of the original program is using all cores, the scheduler will execute the threads in a temporal redundant way instead of a spatial redundant way

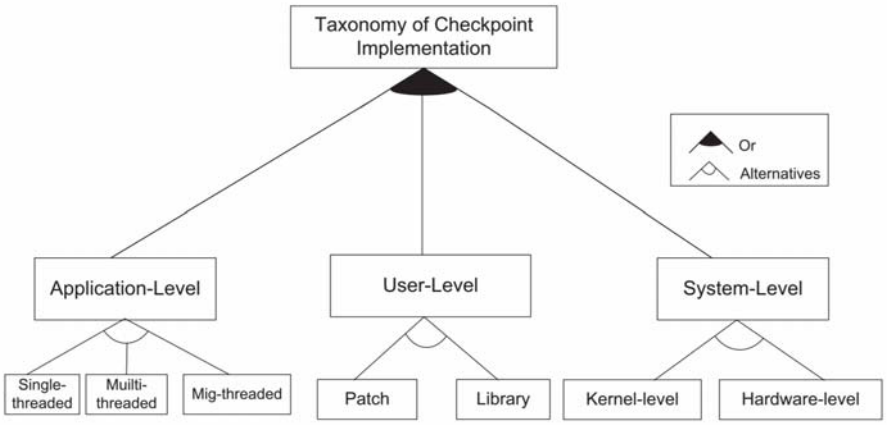
Sphere of Replication for DDX



RELATED WORK

Fault tolarence and Checkpointing techniques

- CRAK [Zhong01]
- Epckpt [Pinheiro01]
- Condor [CondorTeam01]
- Libckpt [Plank95]
- Cockeck [Stallner96]
- DMTCP [Ansel09]
- BLCR [Duell02]
- Ckpt [Zandy02]
- Dynamite [Overenider96]
- CHPOX [Sudakkov07]
- Porch [Ramkumar97]
- Esky [Gibson12]
- LAM-MPI [Sankaram04]
- CryoPID [Blackham04]
- Libtchpt [William01]
- Score [Takahashi00]
- FT-MPI [Fagg00]
- DejaVu [Ruscio07]
- C3 [Schulz04]
- MPICH-V [Bosilica02]



[Egwutuoha13-JS-A survey of fault tolerance mechanisms and checkpoint/restart implementations for high performance computing systems]

XKAAPI

- **Dataflow and Work-Stealing in KAAPI**
- The Kernel for Adaptive, Asynchronous Parallel Interface (KAAPI) is "a C++ library that allows to program and execute multithreaded computations with dataflow synchronization between threads; the library is able to schedule programs at fine or medium granularity in a distributed environment."
- In the KAAPI execution model "a multi-processor system is viewed as a collection of so-called *K*-processors, which can be thought of as kernel threads. A process may consist of several *K*-processors. A *K*-processor in turn executes so-called *K*-threads, which can be thought of as application-level user threads. On a *K*-processor only one *K*-thread is active at a given time. The thread of control is a sequence of non-interruptible tasks. A *K*-processor becomes idle if there are no ready-tasks, i.e. either all tasks have finished execution or they are waiting for data as the result of synchronization. Under the work-stealing strategy, an idle *K*-processor tries to steal a task of a *K*-thread from a randomly selected *K*-processor called *victim*."

- Similar to DF-Threads but done at kernel level; in DF-Threads we have two flavors: i) a user-space implementation and ii) a hardware supported implementation via a coprocessing unit combination (TSU+FDU)

[Jafar05-europar - Jafar, S., Gautier, T., Krings, A., louis Roch, J.: A checkpoint/recovery model for heterogeneous dataflow computations using work-stealing. In: Cunha, J.C., Medeiros P.D.(eds.)Euro-Par -LNCS, vol. 3648, pp. 675–684. Springer, 2005]

XKAAPI checkpointing

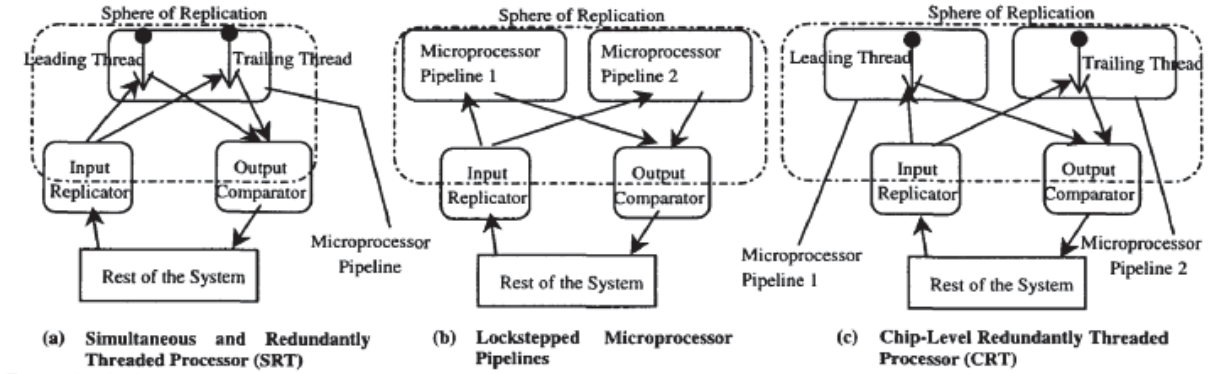
- "Definition of a checkpoint
- A copy of the dataflow graph G represents a consistent global checkpoint of the application. In this research, checkpoints are with respect to a process, and consist of a copy of its local Gi must representing the stack. The checkpointing protocol must ensure that checkpoints are created in such a fashion that G is always a consistent global application state, even if only a single process is rolled back.

- The checkpoint of Gi itself consists of the entries of the process stack, i.e., its tasks and their associated inputs, and not of the task execution state on the processor itself. Understanding this difference between the two concepts is crucial. Checkpointing the tasks and their inputs simply requires to store the tasks and their input data as a dataflow graph. On the other hand, checkpointing the execution of a task usually consists of storing the execution state of the processor as defined by the processor context, i.e., the processor registers such as program counters and stack pointers as well as data. In the first case, it is possible to move a task and its inputs, assuming that both are represented in a platform-independent fashion. In the latter case the fact that the process context is platform-dependent requires a homogeneous system in order to perform a restore operation or a virtualization of this state"
- How this extends to a generic **shared-memory computation**?

[Jafar05-iceit-Theft_induced_checkpointing_for_reconfigurable_dataflow_applications]

RMT – Redundant Multi-Threading

Dual SMT pipeline (via RMT) | Dual-core RMT lockstepping | Dual-core RMT (called CRT)



(a) Simultaneous and Redundantly Threaded Processor (SRT) (b) Lockstepped Microprocessor Pipelines (c) Chip-Level Redundantly Threaded Processor (CRT)

Figure 1. Fault Detection Using SRT, Lockstepped Microprocessors, and CRT. Specifically, in our implementations, the microprocessor pipelines, input replicators, and output comparators are on the same chip. The "rest of the system" is split into on-chip components (L2 cache, memory controllers, on-chip router) and off-chip components (memory, disks, other I/O devices).

From [Mukherjee02-isca] Different RMT implementations

- RMT performs duplication of threads, feeding them with identical inputs, but the outputs are still compared at instruction level like in lockstep execution
- RMT detects both transient and permanent faults
- RMT operates on sequential programs

[Mukherjee02-isca-Detailed design and evaluation of redundant multithreading alternatives]

TLR – Thread-Level Redundancy

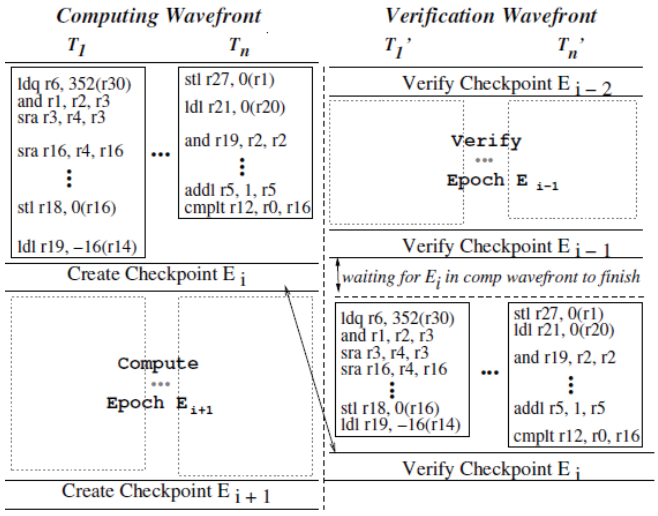


Figure 2. The operation of the TLR architecture: the computing and verification wavefronts and the epochs. Due to branch predictions and prefetching provided by the computing wavefront, threads in the verification wavefront have different timing and are generally faster.

- TLR compares the state of two wavefronts at epoch boundaries (typically thousands of instructions)
- The "state" of the thread could be very complex and involve many hardware resources (registers, caches, memory, …)
  - Practical solution: use store buffers
- Non-determinism has to be suppressed to compare the wavefronts

[Rashid08-hpca-Supporting_highly_decoupled_thread_level_redundancy_for_parallel_programs]

LBRA – Log-based Redundant Architecture

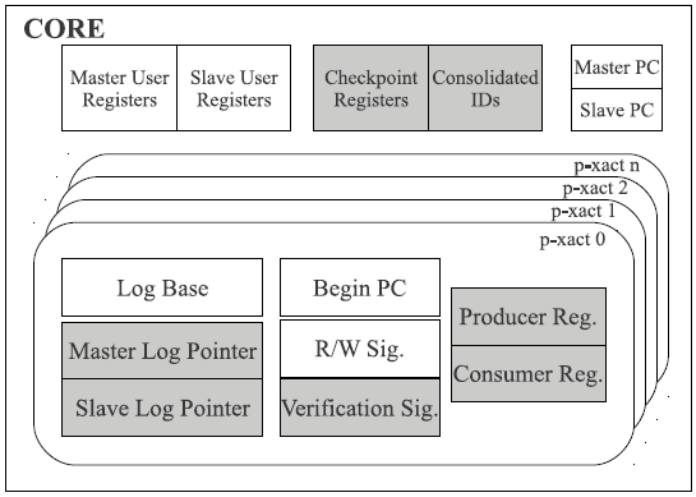
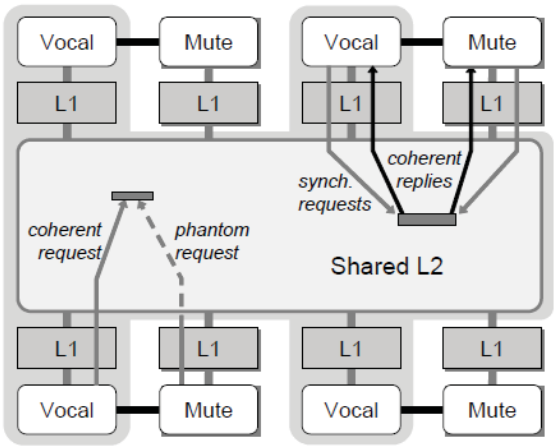


Figure 1. LBRA Hardware Overview. Shadowed boxes represent the added structures.

- Similar to RMT: two hardware threads are executed in a redundant way
- Redundant threads are assimilated as Virtual Transactions
- Using Hardware Transactional Memory (HTM) buffers to detect if two virtual transaction are experiencing faults by comparing verification signatures
- Producer/Consumer registers track the shared areas of a thread to avoid spreading of faults
- 5% slowdown in average on some PARSEC and SPLASH-2 benchmarks

[Sanchez10-hipc-A log-based redundant architecture for reliable parallel computation]

REUNION



- does not require complex hardware to duplicate data; key idea: data is naturally fetched on redundant cores
- if loaded values are different: an error has occurred and detected (similar to a data-race)
- it requires lock-stepping in the case of data-races between redundant stores, increasing communication among cores and complicating redundant thread management

[Smolens06-micro-Reunion_ Complexity_Effective Multicore Redundancy]

DCC - Dynamically Coupled Cores

- DCC is an architectural technique that allows arbitrary CMP cores to verify each other's execution while requiring no static core binding at design time or dedicated communication hardware

- Sanchez et al., later showed hat DCC requires fast result comparison, which makes its use in tiled architectures, which communicate by a network-on-a-chip, inefficient and may induces high overhead for redundant execution

[LaFreida07-dns-Utilizing_Dynamically_Coupled_Cores_to_Form_a_Resilient_Chip_Multiprocessor]

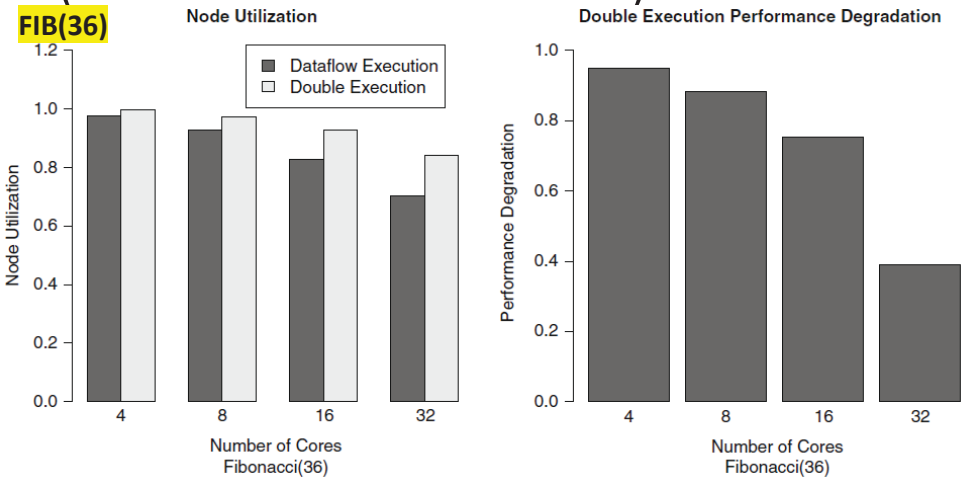
Advantages of DDX over conventional lock-step architectures

- Result comparison can be restricted to data that is consumed by subsequent threads
- Result propagation is only required when a thread has finished execution, which inherently supports deferred memory updates
- Redundant threads are synchronized at thread level
  - This enables the exploitation of the scalability of the dataflow model for redundant thread pairs
  - In particular, the D-TSU scheduler can take advantage of underutilized cores

DDX EVALUATION

DDX node utilization and net performance degradation (in case of zero faults)

FIB(36)

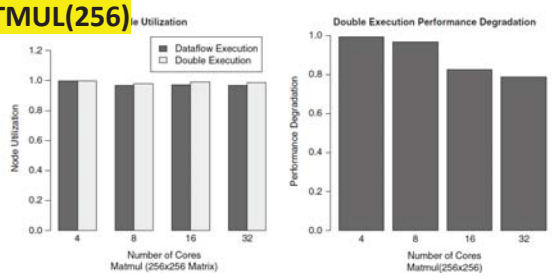


$$Performance\ Degradation = \frac{(TDX - TDF)}{TDF}$$
(i.e., overhead)

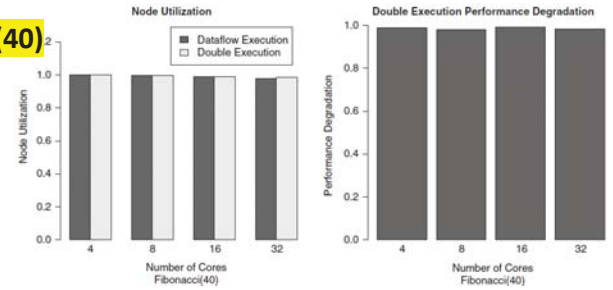
TDX=TIME for DATAFLOW DOUBLE EXECUTION
TDF=TIME for regular DATAFLOW EXECUTION

- In classical lockstepping, we have always a degradation greater than 1
- In DDX the degradation can be less than 1 (as we can use idle time of available cores) and it improves (it's decreasing) if the system is not fully loaded at any time

MATMUL(256)

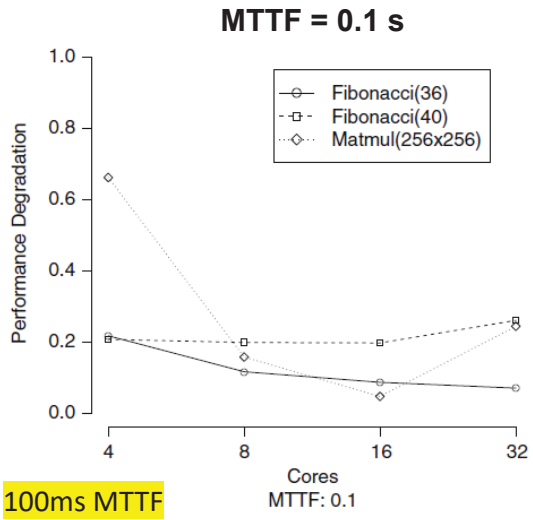


FIB(40)

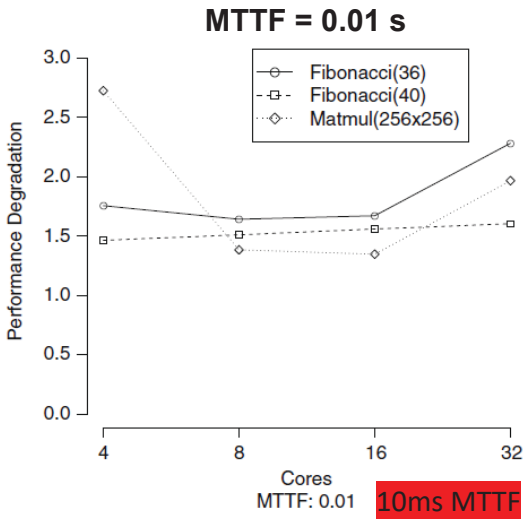


[Weis15-ijpp]

Performance degradation of pure thread restart recovery (when injecting faults, but no DDX!)



100ms MTTF



10ms MTTF

- Although an MTTF of 0.1 is already an artificially high failure rate, the DDX thread restart recovery can be efficiently used, even with increasing failure rates

$$Performance\ Degradation = \frac{(TDX - TDF)}{TDF}$$

TDX=TIME for DOUBLE EXECUTION
TDF=TIME for regular DATAFLOW EXECUTION

[Weis15-ijpp]

DF-Thread benchmark characteristics

- Further analyzing the two scenarios with more benchmarks
  - High utilization – large number of DF-threads

Benchmark	*HighUtil* Input Set
Fibonacci	n: 36, cut-off: 20
Matmul	Blocks: 12×12, Block Size: 16×16
Sparse LU	Matrix Size: 512×512, Block Size: 16×16
Cholesky	Matrix Size: 512×512, Block Size: 16×16

	HighUtil			
	Fibonacci	Matmul	Sparse	Cholesky
# Thrds	5,168	1,728	4,005	7,536
# `tread`	56,841	1,369,153	90,499	200,505
# `twrite`	5,167	1,369,154	7,915	17,920

  - Low utilization – smaller number of DF-threads

Benchmark	*LowUtil* Input Set
Fibonacci	n: 35, cut-off: 28
Matmul	Blocks: 6×6, Block Size: 16×16
Sparse LU	Matrix Size: 256×256, Block Size: 16×16
Cholesky	Matrix Size: 256×256, Block Size: 16×16

	LowUtil			
	Fibonacci	Matmul	Sparse	Cholesky
# Thrds	68	217	701	1,208
# `tread`	741	175,753	12,772	27,645
# `twrite`	67	175,754	1,083	2,432

[Weis15-phd_thesis-Fault_Tolerant Coarse_Grained Data_Flow Execution]

Graceful degradation of DDX w/permanent faulty cores

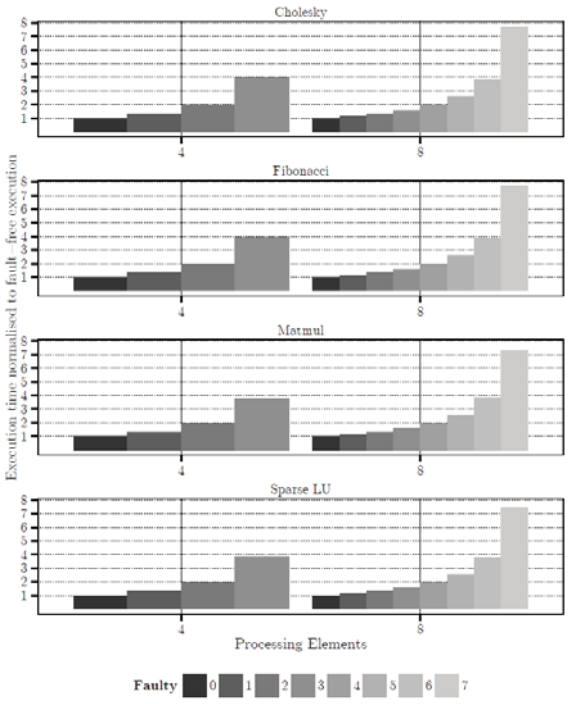


Figure 7.8: Graceful degradation of double execution, when 0–7 PEs are permanen faulty in the 4 and 8 PE systems (*HighUtil*).

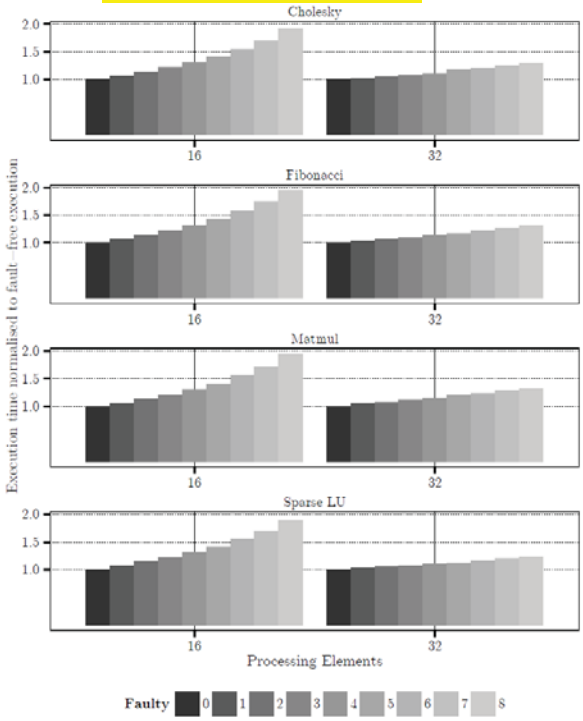
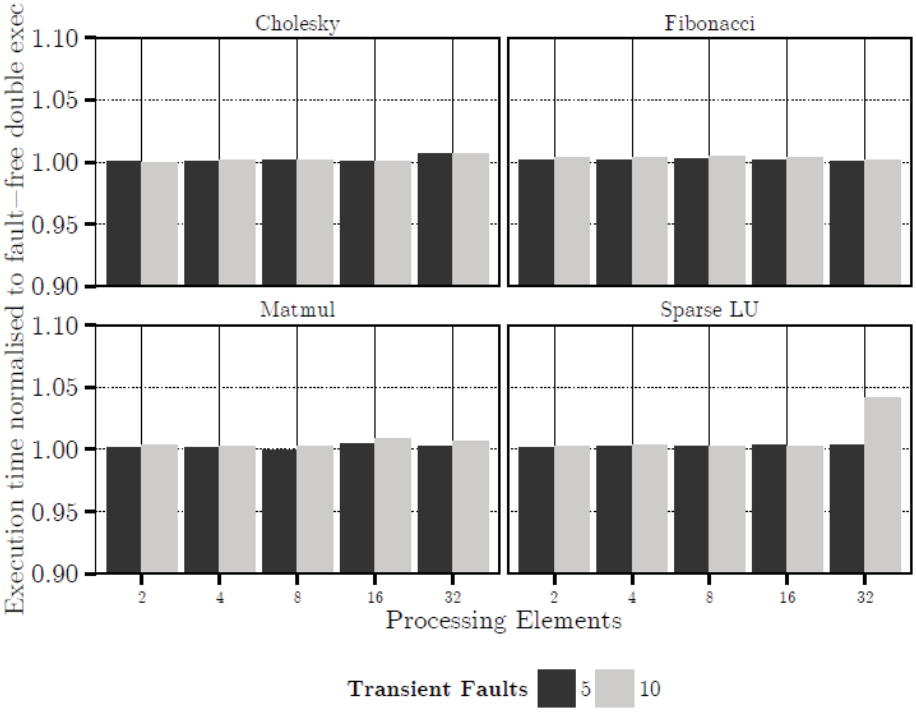


Figure 7.9: Graceful degradation of double execution, when 0–8 PEs are permanent faulty in the 16 and 32 PE systems (*HighUtil*).

[Weis15-phd_thesis-Fault-Tolerant Coarse-Grained Data-Flow Execution]

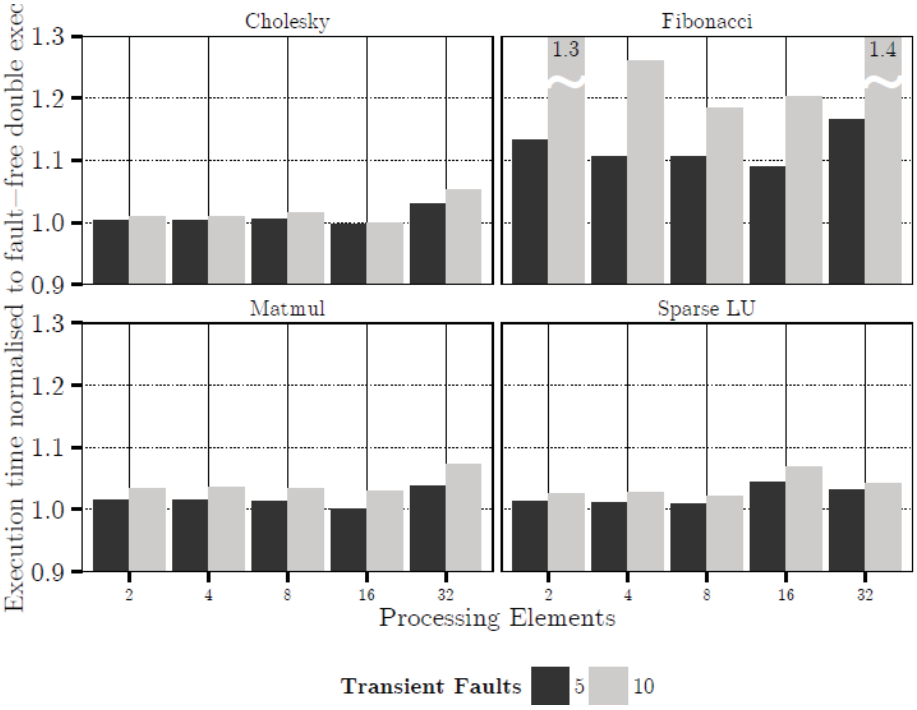
Transient faults (high utilization)



- Execution time normalized to fault-free DDX
  - We inject faults at a fixed time interval of 10 us
  - After a fixed number of faults, the system stops the injection and proceeds without faults
  - We investigate the impact of 5 and 10 faults on the simulated system configurations
- Thread restart recovery has a negligible effect

[Weis15-phd_thesis-Fault_Tolerant Coarse_Grained Data_Flow Execution]

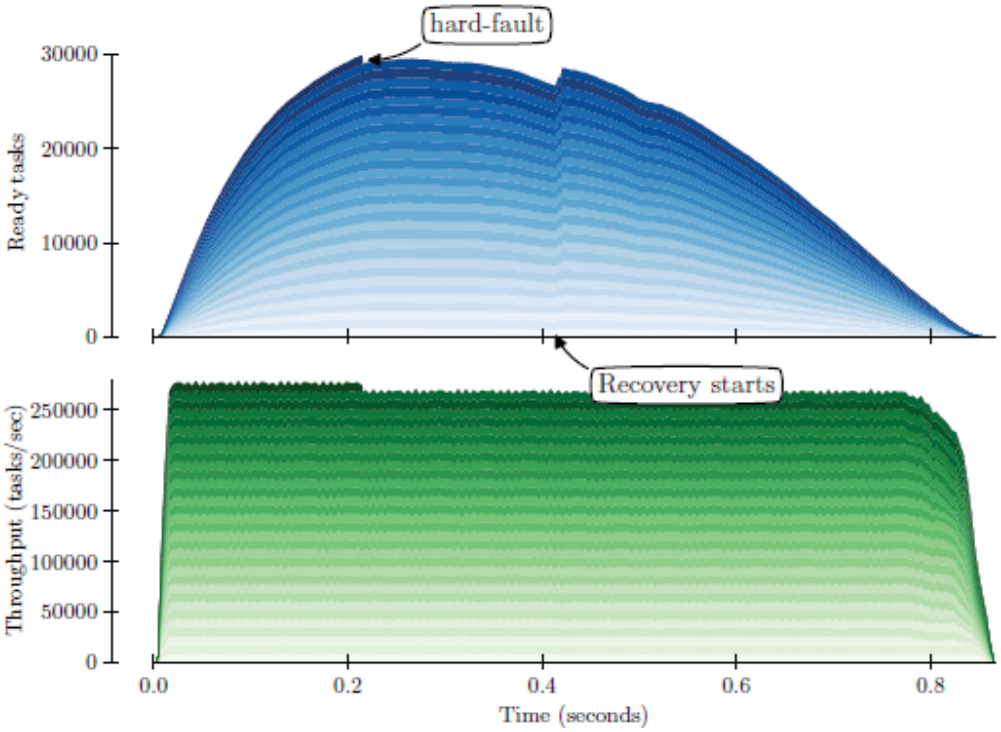
Transient faults (low utilization)



- Execution time normalized to fault-free DDX
  - We inject faults at a fixed time interval of 10 us
  - After a given number of faults, the system stops the injection and proceeds without faults
  - We investigate the impact of 5 and 10 faults on the simulated system configurations
- The Fibonacci benchmark, which executes only 68 threads, suffers from a significant overhead, in particular when 10 threads are restarted

[Weis15-phd_thesis-Fault_Tolerant Coarse_Grained Data_Flow Execution]

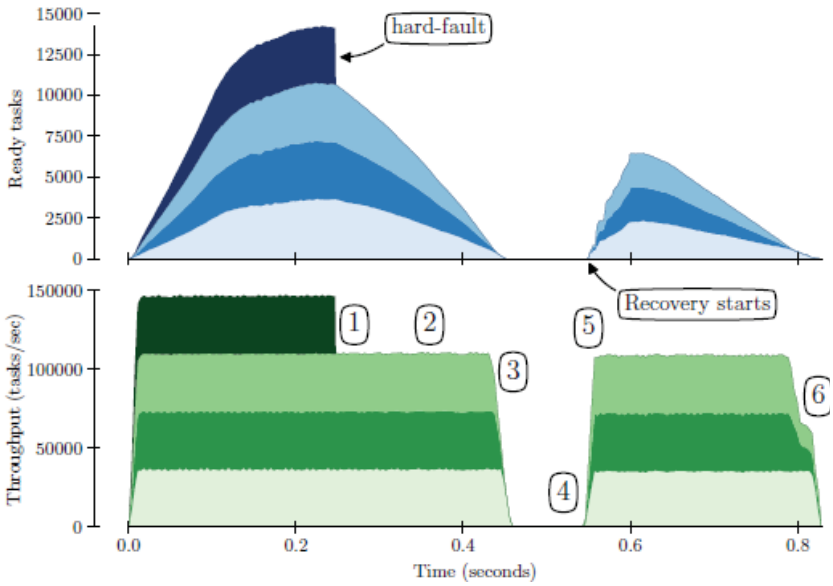
On-the-fly recovery from a hard fault on 32 nodes



- A single application was utilizing all the 32 nodes during the fault
- It remains unaware as the system detects the fault and automatically recovers and redistributes the abandoned task to the remaining cores

[Fuchs18-ms_thesis-Fault-Tolerant Operating System for Many-core Processors]

Hard-fault recovery process

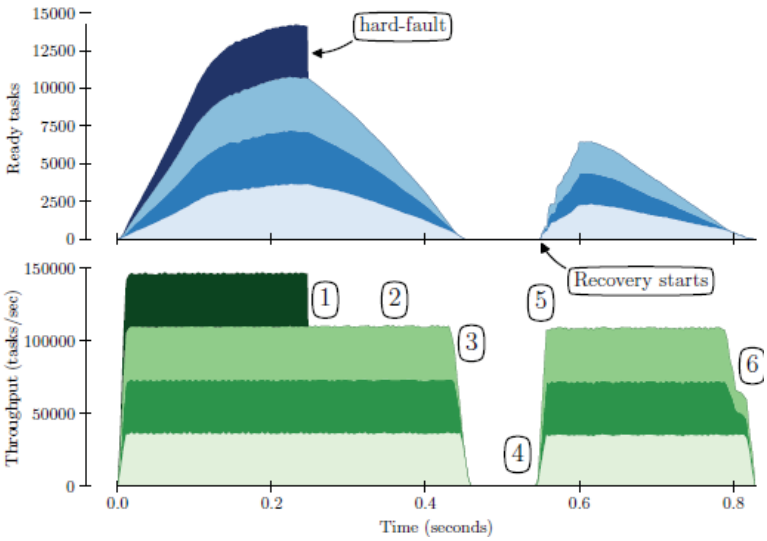


1. A hard-fault is injected into one of the nodes
2. The remaining nodes still have plenty of work to do that does not depend on results from the abandoned tasks of the failed node
3. Throughput begins to fall until it reaches zero when there are no ready tasks left to execute
4. The FDU detects missing heartbeats from the faulty node, so the work from the failed node is rebalanced
5. Throughput reaches 75% of the previous value as only 3 out of 4 nodes are now active
6. All work is complete

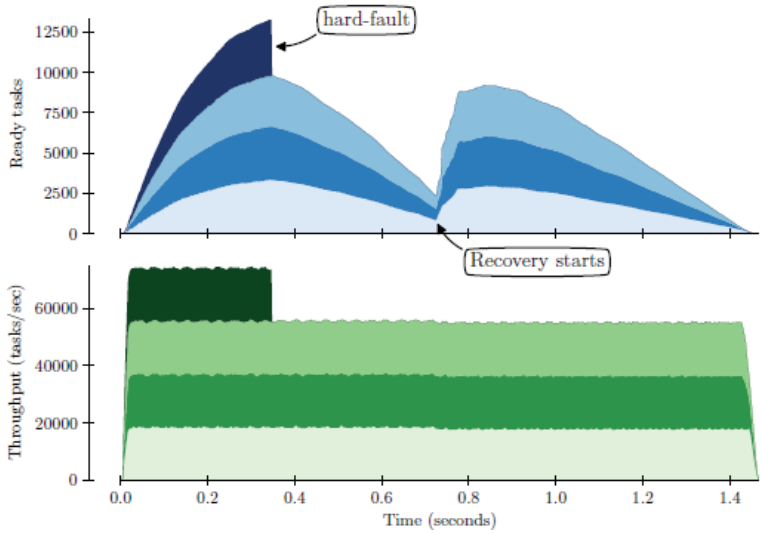
- A single application is running on 128 cores (4 nodes with 32 cores each)
- There are ~86k threads to execute with an 807 us average duration per thread
- Recovery time is about 400 us

[Fuchs18-ms_thesis-Fault-Tolerant Operating System for Many-core Processors]

Recovery with longer thread duration



870 us average thread duration



1600 us average thread duration

[Fuchs18-ms_thesis-Fault-Tolerant Operating System for Many-core Processors]

Conclusions

- Dataflow Threads (DF-Threads) have the potential to provide several benefits
  - Reduced (or eliminated) contention, no need for cache coherency
  - Improved performance due to reduced synchronization hardware and software
  - Graceful degradation of performance in case of faults, the system continues to work!

THANKS FOR YOUR KIND ATTENTION

Roberto Giorgi

University of Siena, Italy

Extending Performance and Reliability via Thread-Level Dataflow Management

Edge computing: the BRAINE Solution

Big data pRocessing and Artificial Intelligence at the Network Edge

Filippo Cugini, CNIT, IT
Vojtěch Janů, CTU, CZ
Martin Ron, Factorio Solutions, CZ

BRAINE: Big data pRocessing and Artificial Intelligence at the Network Edge
H2020 ECSEL JU Grant No. 876967
https://www.braine-project.eu/

Outline

- The BRAINE project
  - Objectives
  - Edge Micro Data Center
  - Use case overview
- BRAINE Network programmability at the edge
  - Openflow
  - P4
  - P4 at the edge: applications
- BRAINE Use case on Factory 4.0
  - Motif discovery
  - Multi-Agent production planning

BRAINE Project Overview

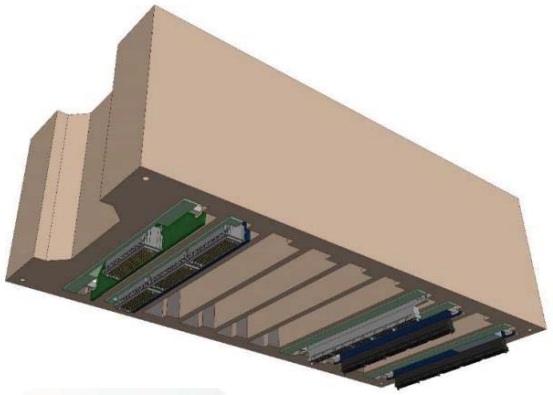
- H2020 ECSEL RIA
- Start Date: May 1st, 2020
- Duration: 36 months
- 27 Partners from 14 Countries
- Budget: 16.3 M

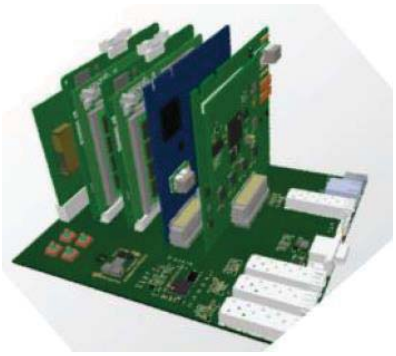
BRAINE Goal

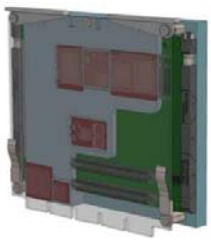
- The BRAINE project's overall aim is to **boost** the development of **Artificial Intelligence (AI)** at the **Edge**
- BRAINE's overall aim will be reached by targeting specific fine-grained goals:
  - Developing an **energy efficient Edge Micro Data Center (EMDC)** that offers Big Data processing and AI capabilities at the Edge.
  - Devising an Edge Computing infrastructure that offers control, computing, **acceleration**, storage, and 5G networking at the Edge for low latency applications.
  - Developing a **secure**, distributed and partly-autonomous system that takes data privacy and sovereignty into account on each and every decision regarding workload placement, data transfer, and computation.
  - Testing and demonstrating the effectiveness and generality of the BRAINE approach by evaluating **multiple real-world use cases** and scenarios that exhibit the required scalability, security, efficiency, agility, and flexibility concerns.

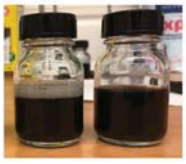
BRAINE Edge Micro Data Center (EMDC) Hardware

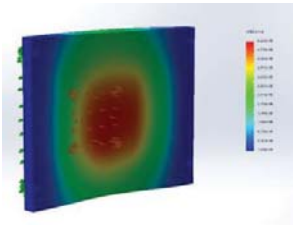
- BRAINE will develop a modular Edge Micro Data Center (EMDC) including:
  - Heterogeneous and modular platform encompassing **HW Acceleration** (CPU, GPU, FPGA)
    - ✓ Board design
    - ✓ Cluster design
  - Innovative **cooling system** (specific design for no energy cooling, graphene nano-fluids, etc.)
  - Non-volatile **memory**
  - Embedded **programmable networking** capibilities (6.4Tb/s ASIC, smart NICs)
  - Integration with **5G** and optical metro connectivity for efficient movement of data
  - Embedded **security** developed for the AI hardware (data integrity, confidentialy, cryptographic functionalities, etc)







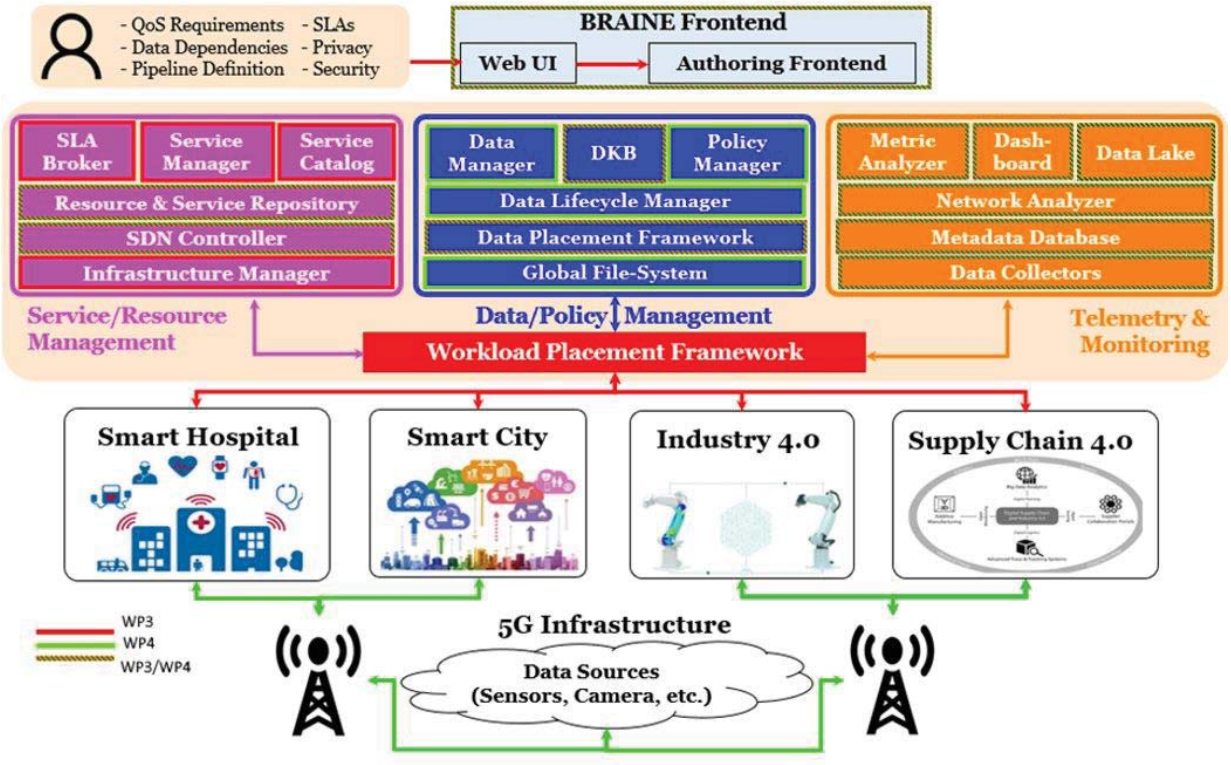




BRAINE Edge Micro Data Center (EMDC)

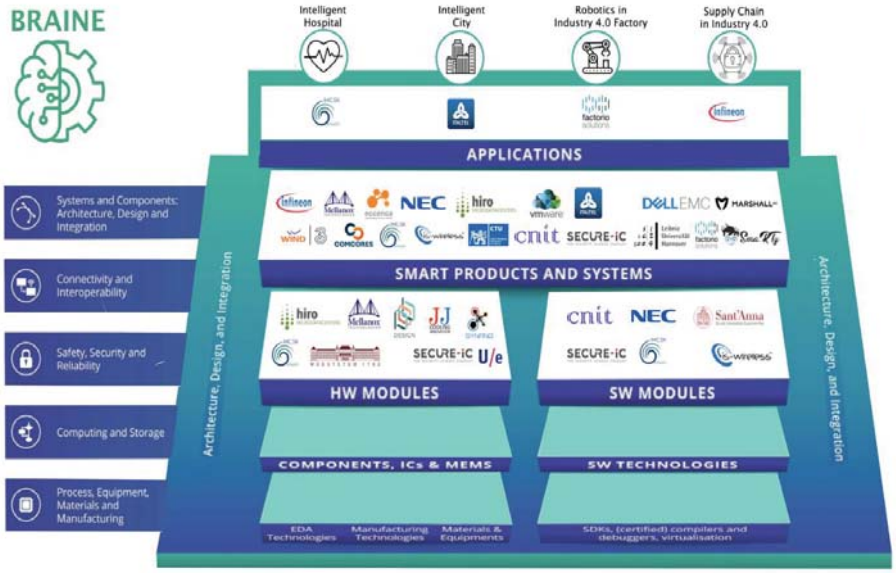
BRAINE EMDC SW

- BRAINE will develop an efficient data management and control system supporting AI



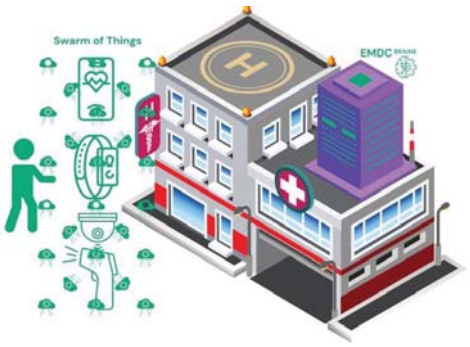
BRAINE Use Cases

- BRAINE will demonstrate edge computing enabling AI through four use cases (UC) :
  - Healthcare Assisted Living
  - Robotics and Factory 4.0
  - Industry 4: SemiConductor Supply Chain.
  - Smart cities/Campus: Multi-tenant real-time AI video analytics



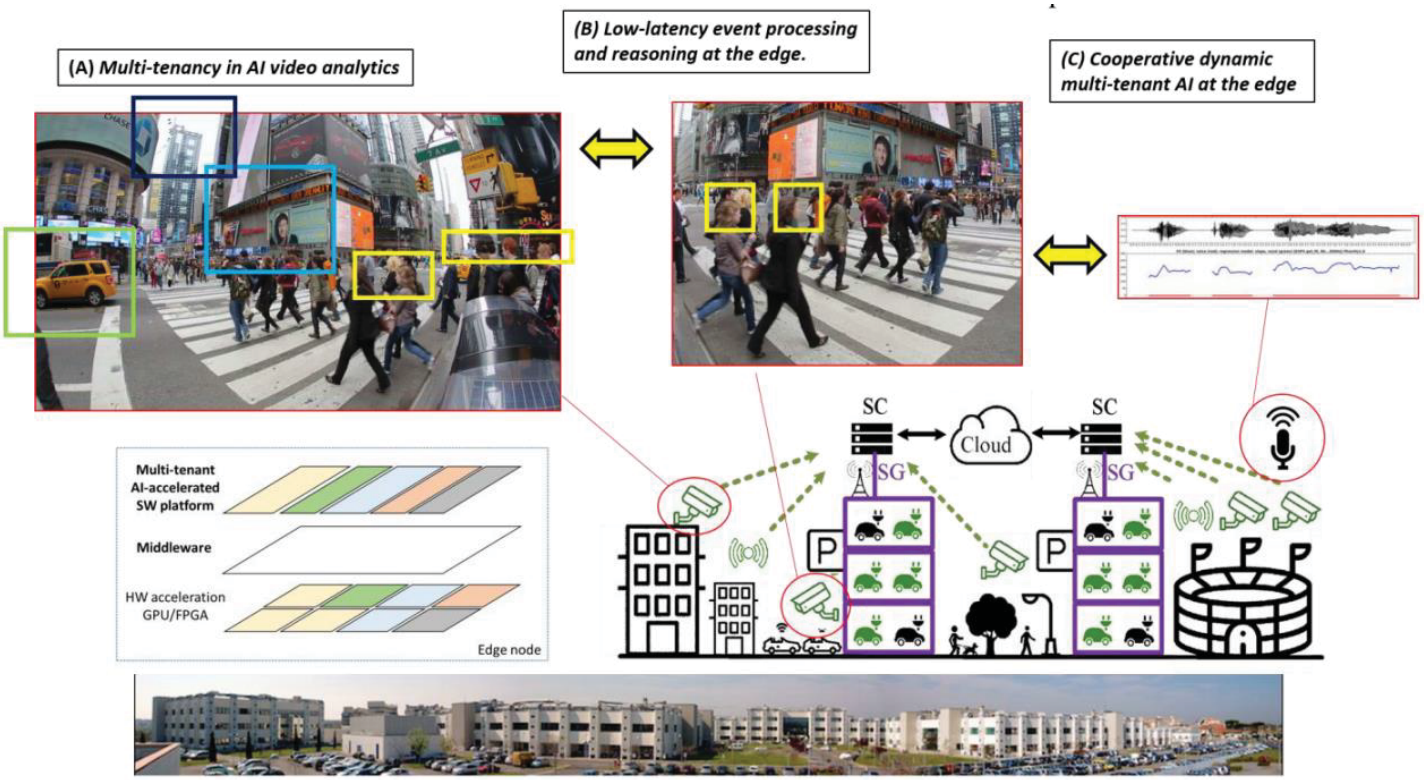
BRAINE Use Case: Healthcare Assisted Living

- Intelligent Hospital & Remote Patient Monitoring
- Specific objective: AI Digital Twins for patients
  - Cyber-bio-physical model of the system
  - Heterogeneous and innovative sensors (wearable bracelets, home sensors, beacons, medical diagnostic equipment, mobiles, etc)
  - Automated parameter monitoring (pulse, ECG, HR, PPG, SpO2, BP, Galvanic skin, respiration, body temperature, emotional status), including environmental data (room temperature, humidity, etc.)
  - Real-time AI-driven analysis diagnose abnormalities in order to predict and identify emergencies.



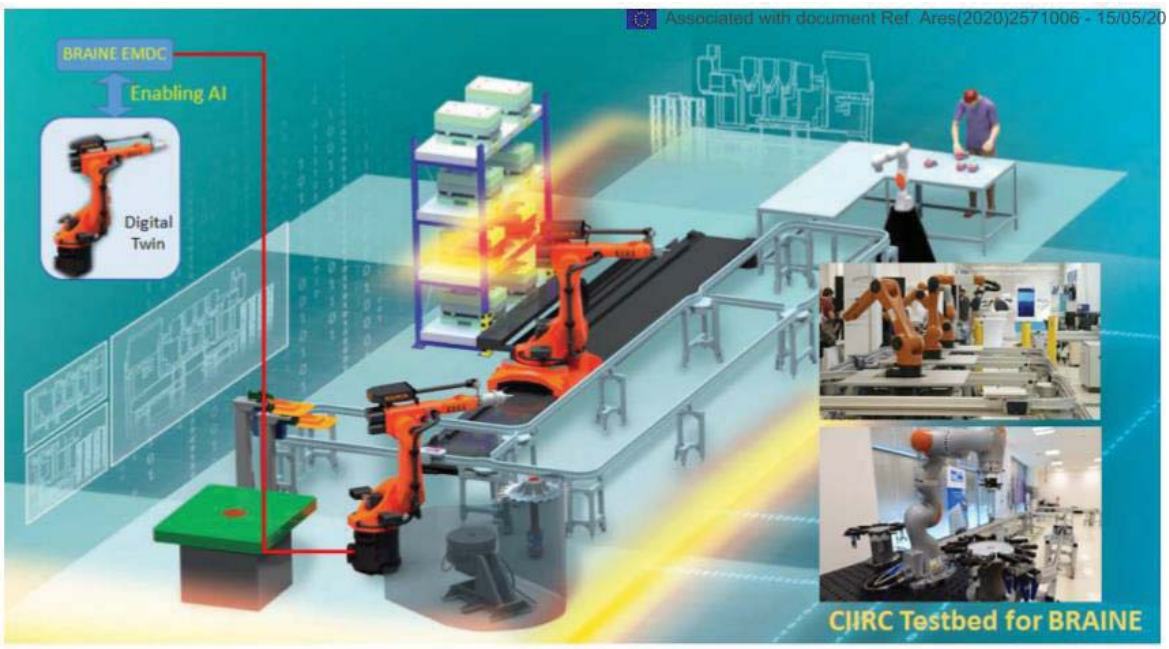
BRAINE Use Case: Smart cities

- Final demo in 2023



BRAINE Use Case: Robotics and Factory 4.0

- Digital twin of a robotic actuator network, at the edge



BRAINE Use Case: Industry 4.0 in SemiConductor Supply Chain

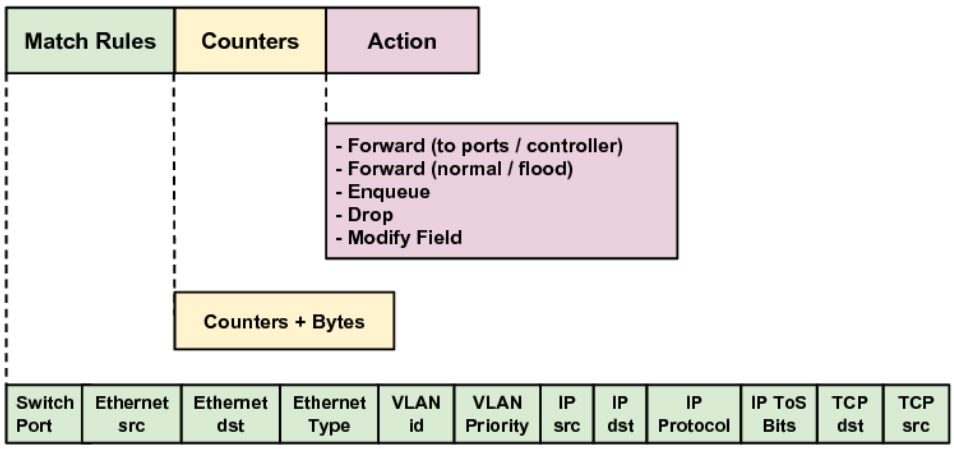
- Move current cloud-based semiconductor supply chains and manufacturing to the edge for time-saving data generation and real-time processing

Outline

- The BRAINE project
  - Objectives
  - Edge Micro Data Center
  - Use case overview
- BRAINE Network programmability at the edge
  - Openflow
  - P4
  - P4 at the edge: applications
- BRAINE Use case on Factory 4.0
  - Motif discovery
  - Multi-Agent production planning

Network programmability: openflow

- Match-action flow rules over the entire packet header stack allows dynamic behavior and network functions on the same device



Network programmability: openflow

BRAINE

L2 Switch

Switch Port	MAC src	MAC dst	Eth type	VLAN ID	IP Src	IP Dst	IP Prot	TCP sport	TCP dport	Action
*	*	00:1f:..	*	*	*	*	*	*	*	port6

Network programmability: openflow

VLAN Switching

Switch Port	MAC src	MAC dst	Eth type	VLAN ID	IP Src	IP Dst	IP Prot	TCP sport	TCP dport	Action
*	*	00:1f..	*	vlan3	*	*	*	*	*	port6, port7, port9

Network programmability: openflow

	Switch Port	MAC src	MAC dst	Eth type	VLAN ID	IP Src	IP Dst	IP Prot	TCP sport	TCP dport	Action
Routing	*	*	*	*	*	*	5.6.7.8	*	*	*	port6

Network programmability: openflow

TCP Flow Switch

Switch Port	MAC src	MAC dst	Eth type	VLAN ID	IP Src	IP Dst	IP Prot	TCP sport	TCP dport	Action
port3	00:20..	00:1f..	0800	vlan1	1.2.3.4	5.6.7.8	4	17264	80	port6

Network programmability: openflow

Firewall

Switch Port	MAC src	MAC dst	Eth type	VLAN ID	IP Src	IP Dst	IP Prot	TCP sport	TCP dport	Forward
*	*	*	*	*	*	*	*	*	22	drop

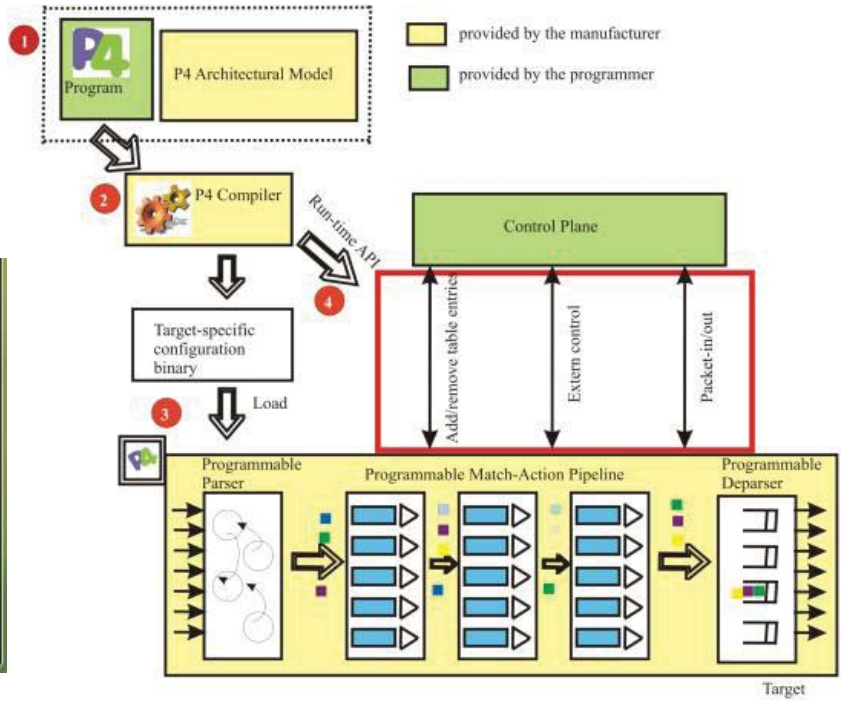
Openflow: limitations

- Match-action instructions (flow rules):
  - stateless and limited to standard protocol headers
- Switch data plane:
  - proprietary vendor-locked solutions, STATIC
- Switch pipelines
  - fixed
- Tradeoffs to handle performance
- Functions: Software-based (CPU) vs hardware-based processing (ASIC)
- Context-based forwarding/processing not possible online
  - demanded to Controller -> big scalability issues at high rates

Programmable networking: P4

- The P4 technology has been conceived to program pipelines and functions of a switch
- High-level, aims to be platform agnostic, highly re-configurable

```
table routing {
  key = { ipv4.dstAddr : lpm; }
  actions = { drop; route; }
  size : 2048;
}
control ingress() {
  apply {
    routing.apply();
  }
}
```



Programmable networking: P4 language

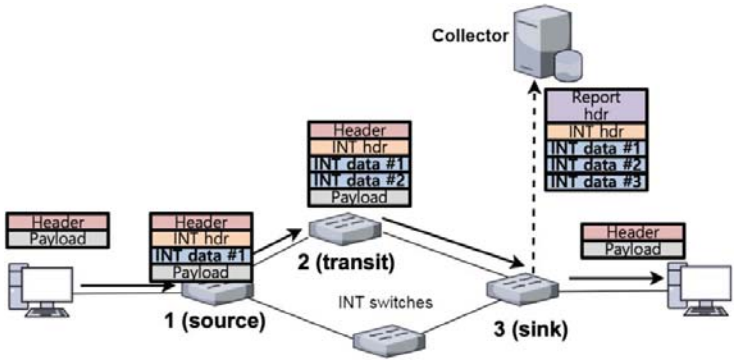
- Custom pipelines with conditional execution
  - allowed complex conditional control and dedicated per-packet treatment, packet cloning and recirculation features
- Custom flow tables
  - customize flow entries
- Custom actions
  - improved operation flexibility
- Stateful objects management (meters, registers, counters)
  - enable Finite state machines (FSM), context/history-based decisions, computational algorithms
- Programmable packet metadata
  - enriched packet extra-information processing
- Programmable extra headers
  - new custom protocols/stacks

P4-enabled functionalities

A. Monitoring and Telemetry
B. Latency-aware scheduling and forwarding
C. 5G function acceleration
D. Cyber-security

A. Monitoring and Telemetry

- The most successful P4 application provides advanced monitoring and telemetry capabilities.
- For example, **In-band Telemetry (INT)** enables the introduction of custom packet headers including metadata such as timestamps and the time spent in the traversed queues.
- This enables accurate monitoring across the whole network, potentially leading to improved traffic engineering solutions.
- INT can be enforced at the IoT/Terminal



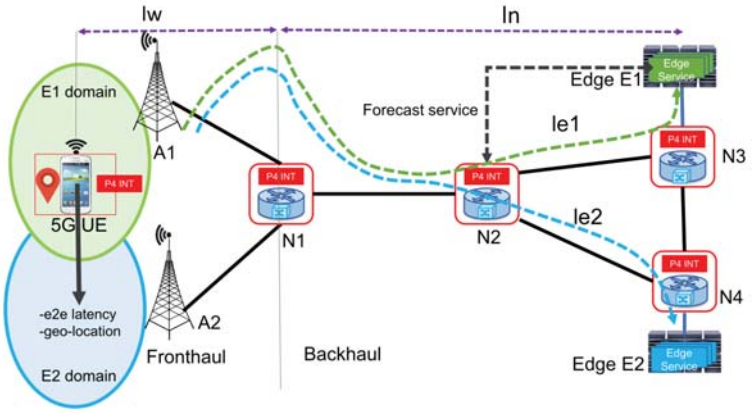
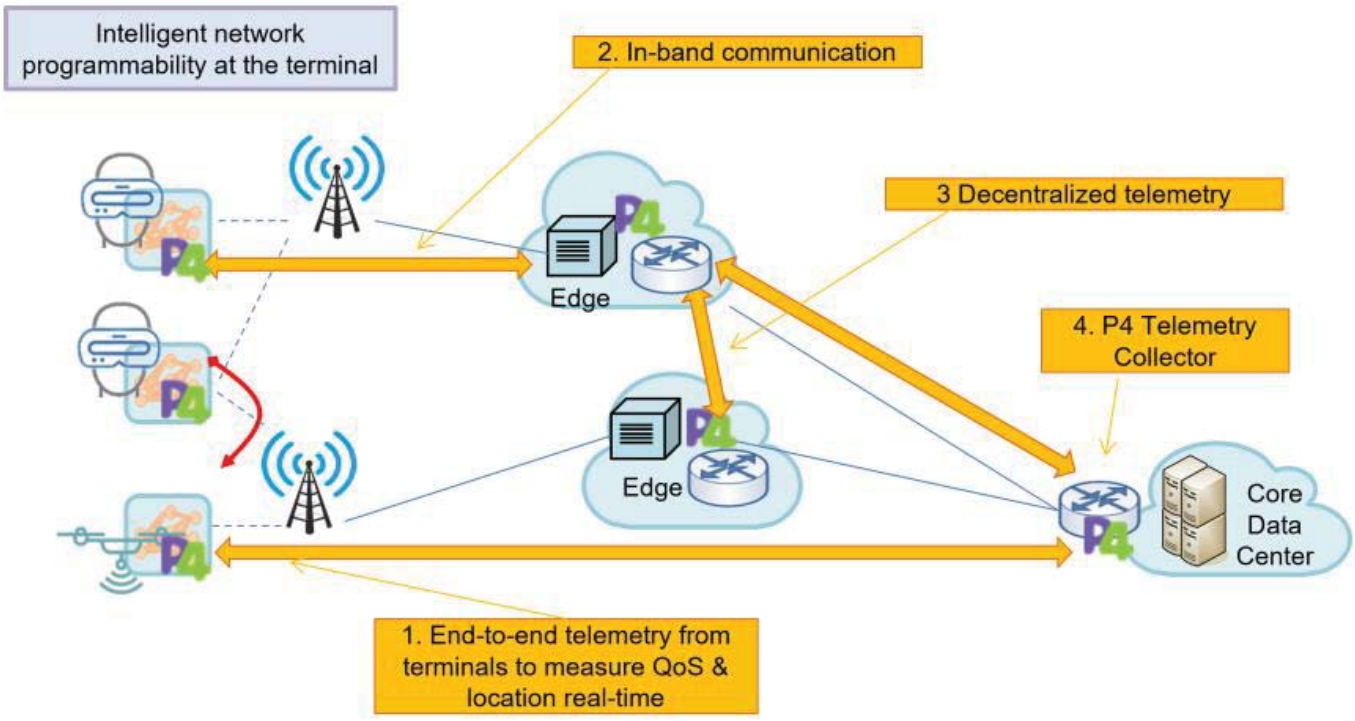


Image source: https://onlinelibrary.wiley.com/doi/full/10.1002/nem.2080

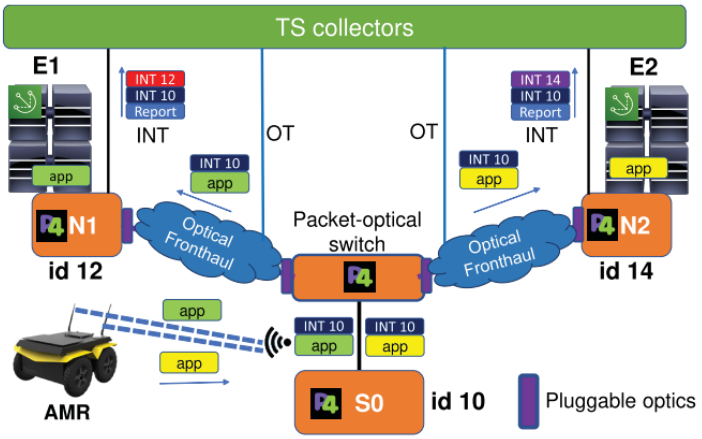
A. Monitoring and Telemetry

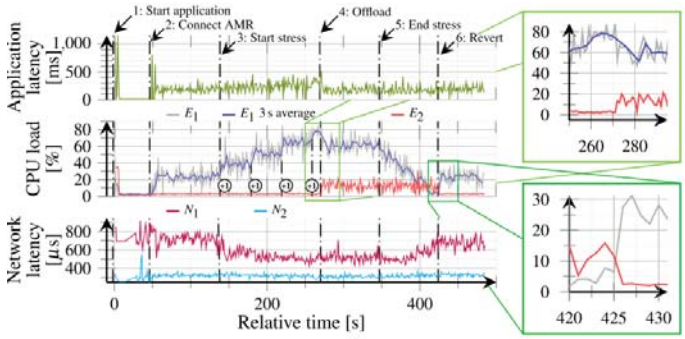
- Telemetry enhanced as in-band communication channel among network nodes



B. Latency-aware scheduling and forwarding

- In-band telemetry applied to serverless edge services
- After detection of an alarm condition, serverless function replacement has been completed in 9 ms

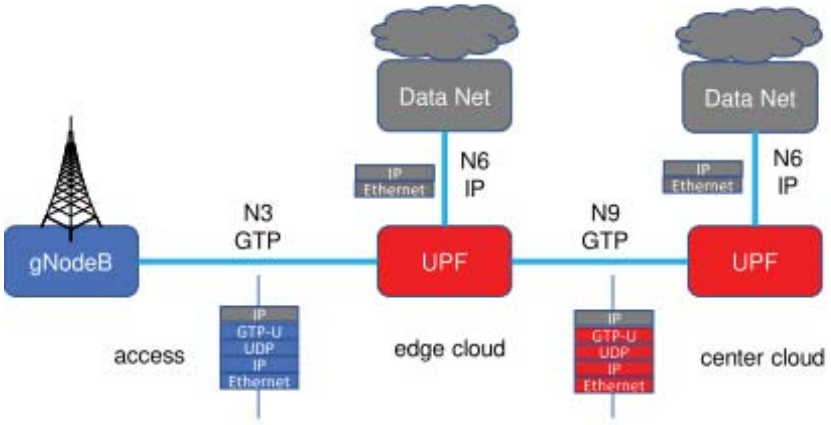




I. Pelle, et al, "Fast Edge-to-Edge Serverless Migration in 5G Programmable Packet-Optical Networks", OFC 2021

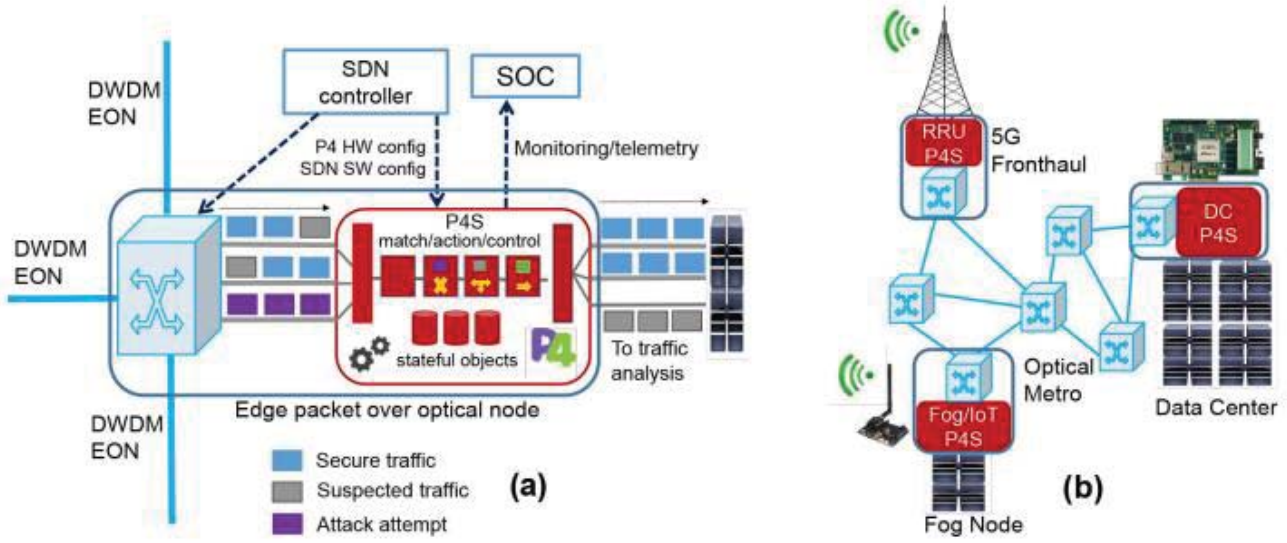
C. 5G function acceleration

- P4 has the capability to offload specific 5G functions, such as the User Plane Function (UPF), directly performing protocol encapsulation/decapsulation function and traffic steering



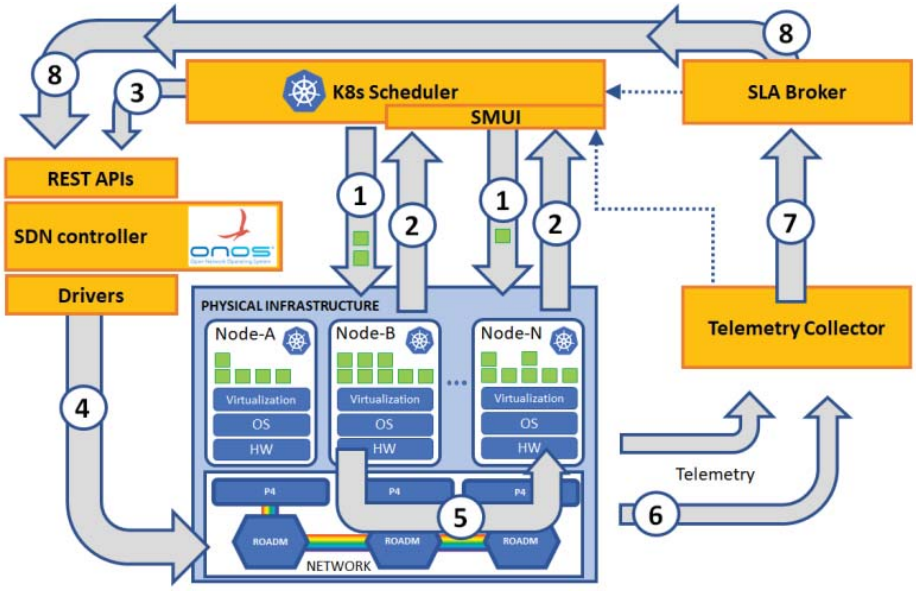
D. Cyber-security

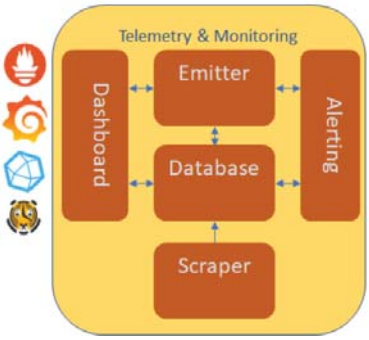
- The stateful capabilities of P4 nodes enables the implementation of in-network firewalling solutions operating at wire speed and potentially deployable in all metro nodes.
- This would constitute a distributed security barrier across the entire network.



BRAINE Closed-loop solution

- Closed-loop K8Sscheduler-controller-infrastructure-telemetry-SLAbroker





BRAINE Dissemination

- www.braine-project.eu
- https://www.instagram.com/braine_project/
- https://twitter.com/BraineProject
- https://www.linkedin.com/company/braineproject

Outline

- The BRAINE project
  - Objectives
  - Edge Micro Data Center
  - Use case overview
- BRAINE Network programmability at the edge
  - Openflow
  - P4
  - P4 at the edge: applications
- BRAINE Use case on Factory 4.0
  - Motif discovery
  - Multi-Agent production planning

Acknowledgment

This project has received funding from the ECSEL Joint Undertaking (JU) under grant agreement No 876967. The JU receives support from the European Union's Horizon 2020 research and innovation programme

https://www.braine-project.eu/

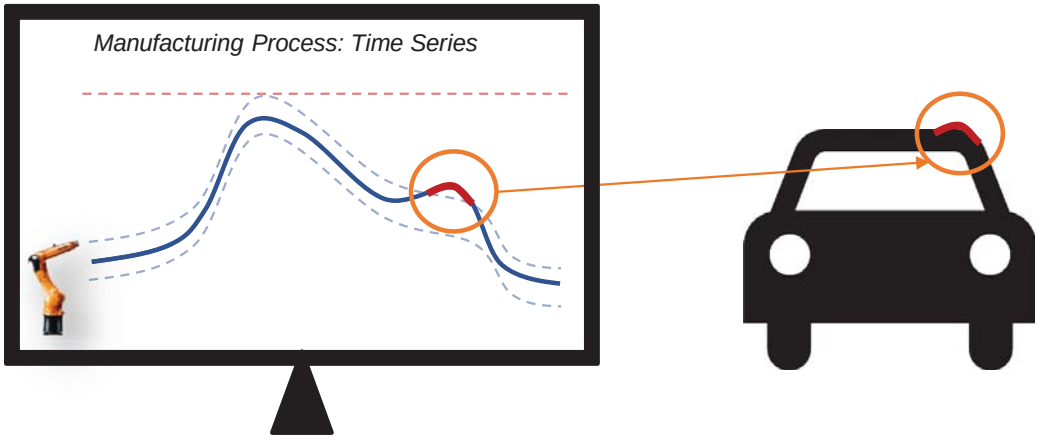
Use Case: Motif Discovery (MOD)

Martin Ron, FS, martin.ron@factorio.cz

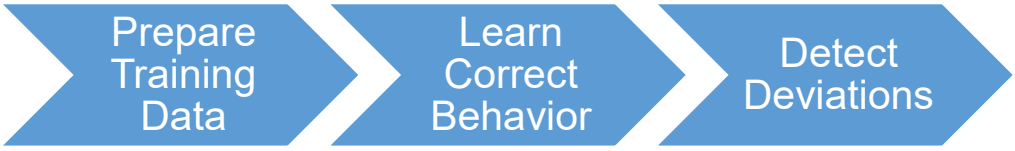
BRAINE: Big data pRocessing and Artificial Intelligence at the Network Edge
H2020 ECSEL JU Grant No. 876967
https://www.braine-project.eu/

Motif Discovery in Detail

- Motivation:
  - Connect manufacturing process shape with products.
  - Deviations in process shape may be associated with product quality loss.

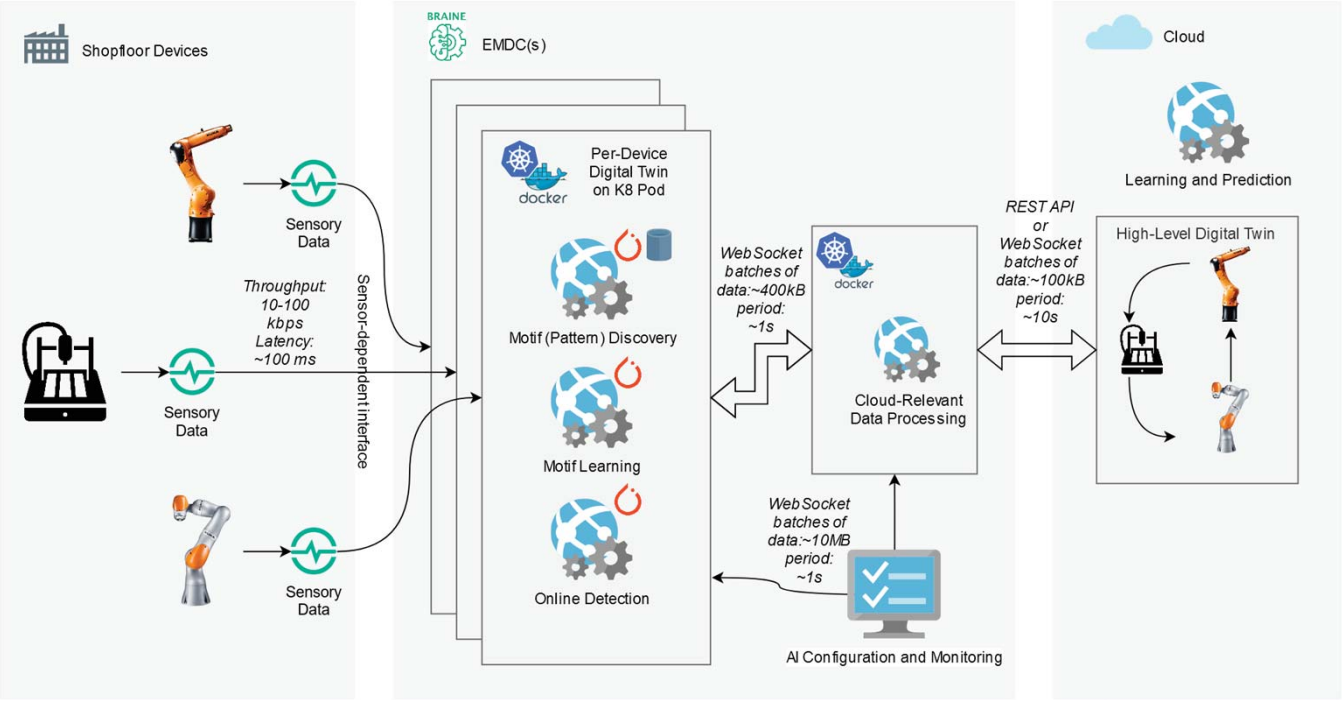


- To achieve this, we need to:



Platform: Motif Discovery

- Each device has its own Digital Twin (DT) model running in separate K8 pod.
- Sensor-dependent interface can be e.g., OPC UA, TCP/IP or UDP stream or through some database exchange.

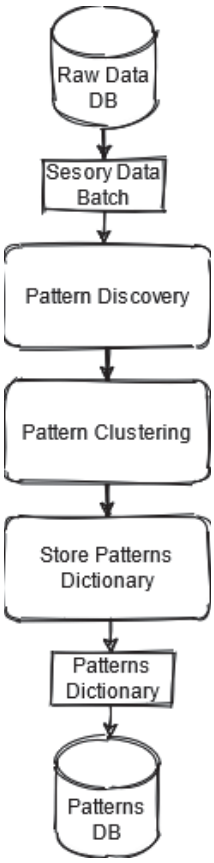


MOD Workflow | Connecting Device and Database

- From MOD tool perspective, the device is a collection of semanticly related sensory data-streams.
  - It can be, e.g., RPMs of a motor in a room and the room's temperature if they correlate. This would be a device made of two data-streams.
- Every following step requires the device to be configured.
- User specifies:
  - Data-streams of all sensors of device – connects to their online sources.
  - Database for storing the data on the Edge – raw sensory data are stored there.
  - Database for storing segmented time series – segments of sensory data.
  - Database for storing detection models – models coming from ML task on top of the sensory data.

MOD Workflow | Discovery Phase

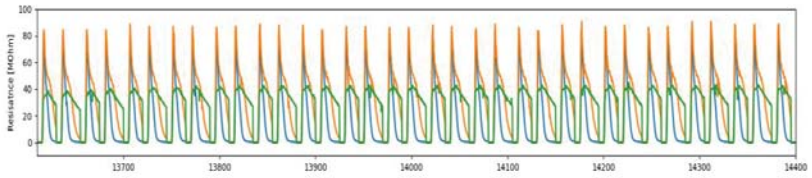
- Overview:
  - Discovery phase searches batches of continuous data for repeated patterns and segments them into structured datasets.
- Pre-requisite:
  - Device data were collected for some time and are available in the database.
- User steps:
  - User invokes Discovery on a single device on selected historic horizont.
  - Discovery is any-time algorithm, the more time it gets, the better results it yields.
  - User asks for results, MOD presents him dictionary of proposed patterns, user can dismiss the proposal (continue searching) or accept it.
  - Accepted dictionary is persistently stored.
    - ✓ It serves as a parameter for following learning phase.



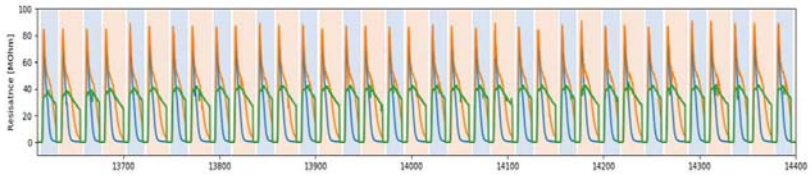
MOD Workflow | Discovery Phase

- Motif (pattern) discovery and then segmenting batched continuous sensory data-stream.

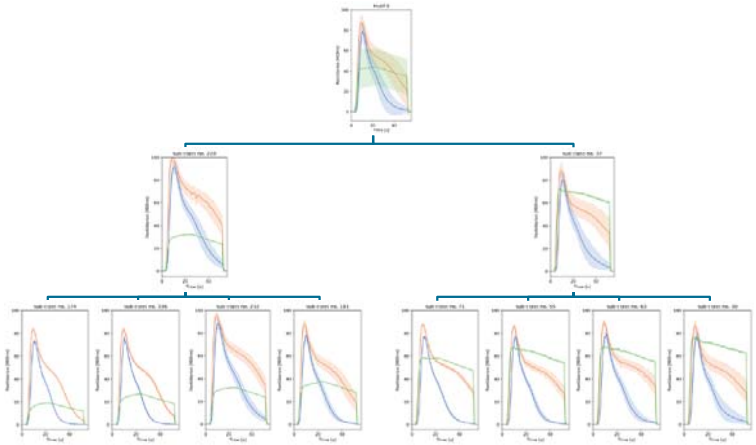
Raw Datastream:

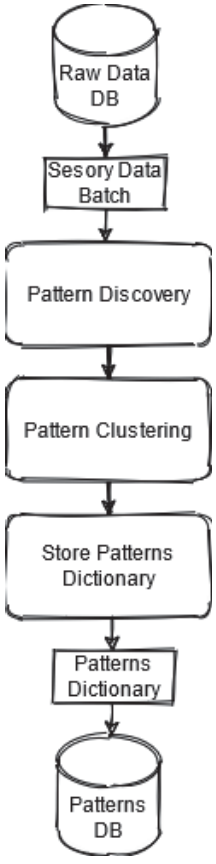


Segmented Motifs:



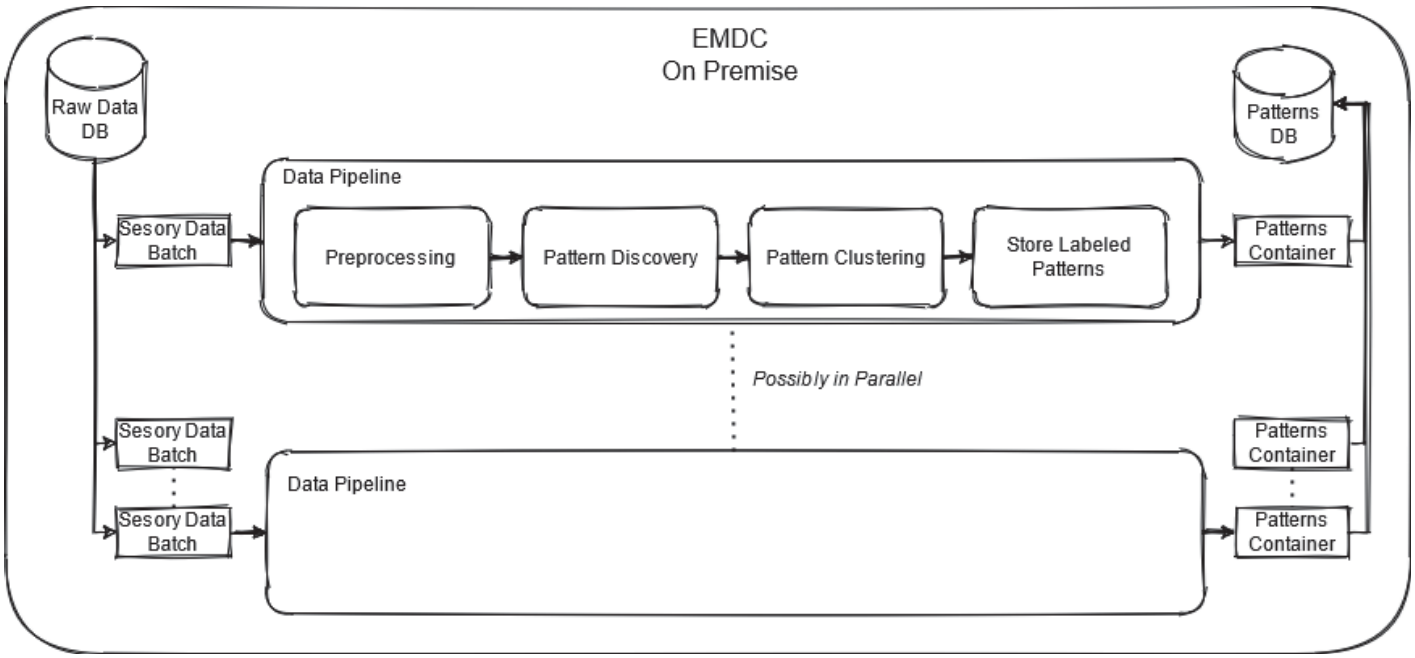
Clusttered Motifs (Dictionary):





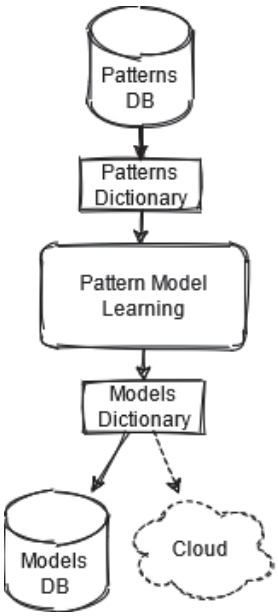
MOD Workflow | Discovery Phase | Pipeline Abstraction

- Parallelizable pipeline, at least one pipeline for multiple batches.
- #Batches:#Pipelines = M:N, but usually M:1.



MOD Workflow | Learning Phase

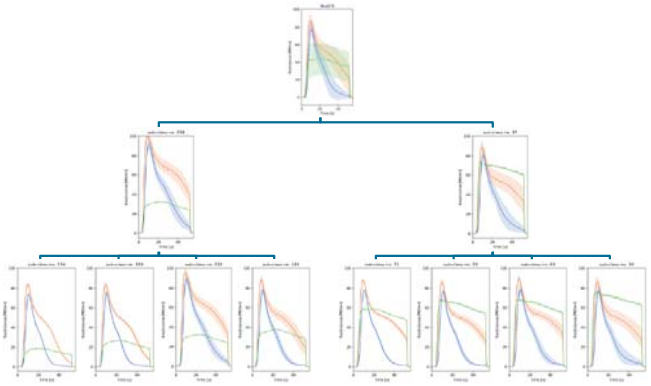
- Overview:
  - Learns what particular patterns look like and their deviations from structured dataset
- Pre-requisite:
  - Dictionary of segmented patterns.
- User steps:
  - User invokes Learning on a single device on selected dataset (dictionary of patterns).
  - Learning runs for certain time (depends on amount of data).
  - Predictive detection model is persistently stored.
    - ✓ It serves as a parameter for following Online detection phase.

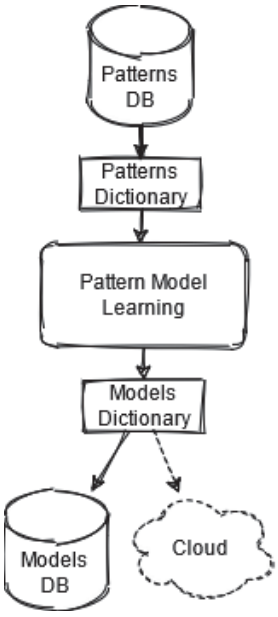


MOD Workflow | Learning Phase

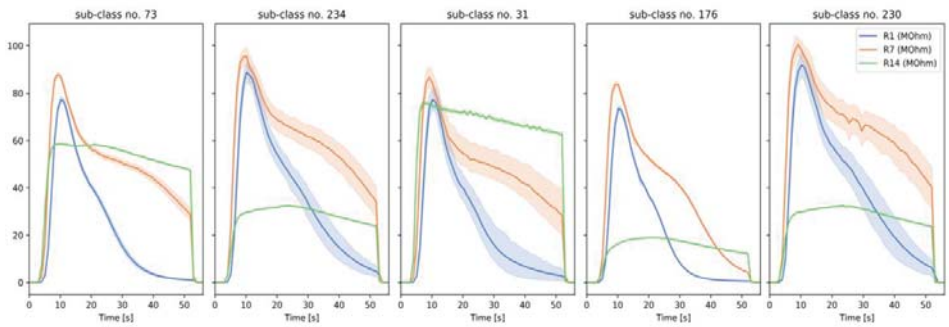
- Learn model for each of discovered motifs.

Clusttered Motifs (Dictionary):



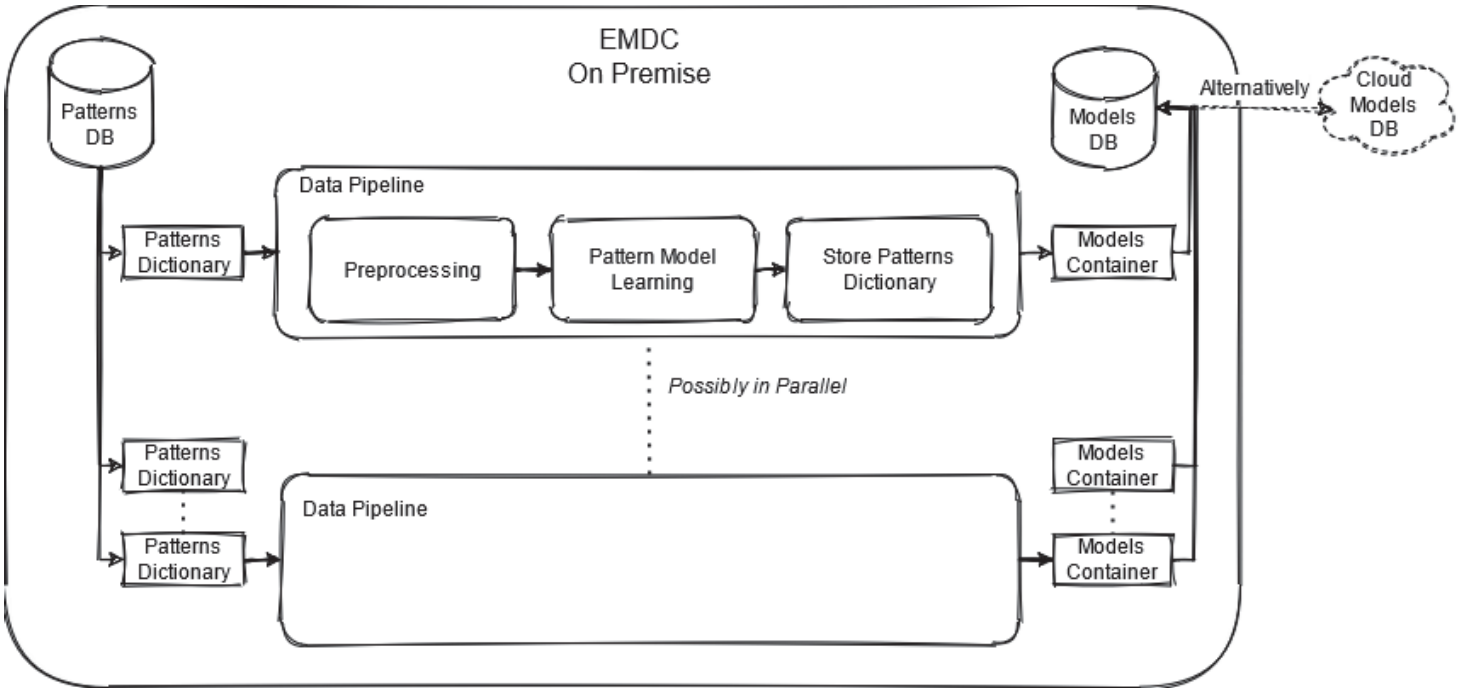


Learned Models:



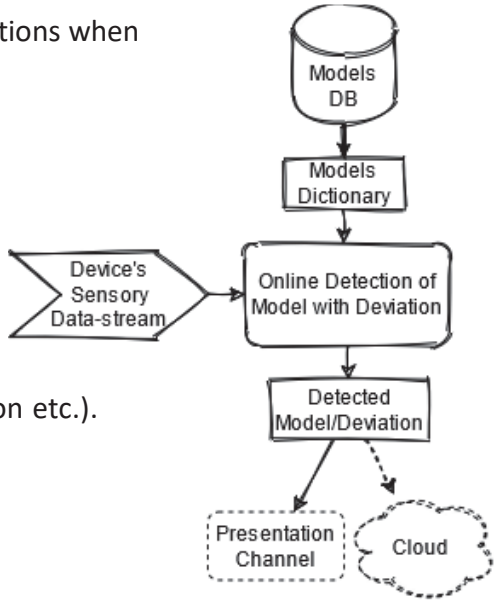
MOD Workflow | Learning Phase| Pipeline Abstraction

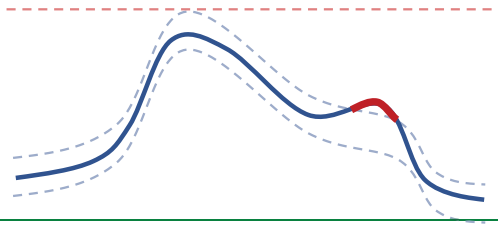
- Parallelizable pipeline, at least one for multiple models trained in series.
- #PatternCollections:#Pipelines = M:N, but usually M:1.



MOD Workflow | Online Detection Phase

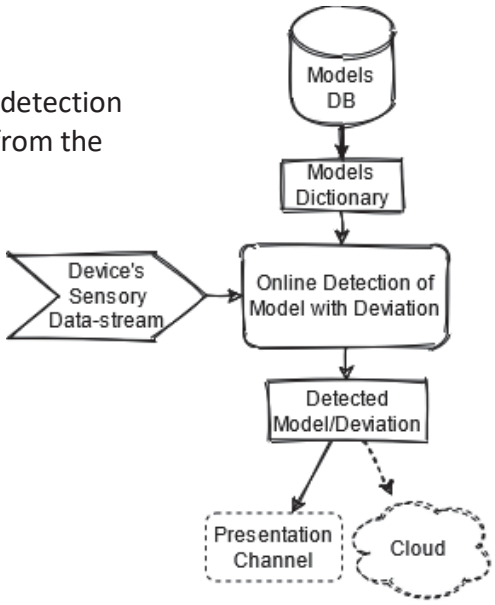
- Overview:
  - Monitoring watchdog system supervises the device and takes actions when deviation occurs.
- Pre-requisite:
  - Predictive model of discovered patterns is available.
  - Device is feeding the data-streams with current sensory data.
- User steps:
  - User configures presentation channel (thin client, mail notification etc.).
  - User defines what happens with deviations.
  - User starts online detection.
  - Detection runs continuously 24/7.
    - ✓ When deviation occurs, defined actions happen (notification, labeled segment is stored into DB etc.)
  - User can stop online detection when needed.





MOD Workflow | Online Detection Phase – Digital Twin

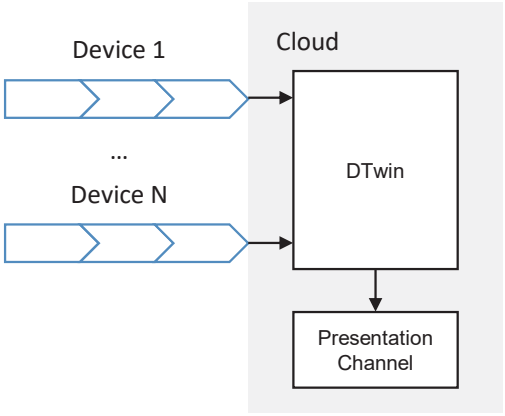
- Overview:
  - Similar to Online Detection Phase.
  - The Digital Twin resides in Cloud and consumes calls from online detection whenever a known pattern is detected nad thus segmented out from the stream.

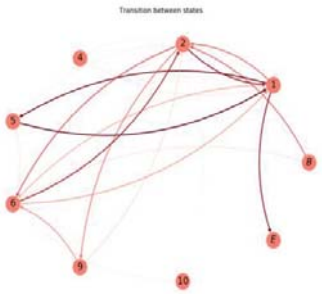


- Pre-requisite:
  - Predictive model of discovered patterns is available.
  - Device is feeding the data-streams with current sensory data.
- User steps:
  - User configures connection to the Cloud application.
  - User defines what happens with deviations.
  - User starts online detection.
  - Detection runs continuously 24/7.
    - ✓ When deviation occurs, defined actions happen (notification, labeled segment is stored into DB etc.)
  - User can stop online detection when needed. This notifies Cloud about planned disconnect.

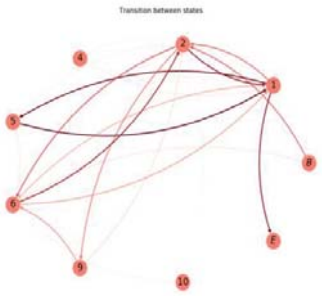
MOD Workflow | Online Detection Phase – Digital Twin

- Finite-state Automaton or Hidden Markov model.
- Stochastic timed transitions between discrete states.





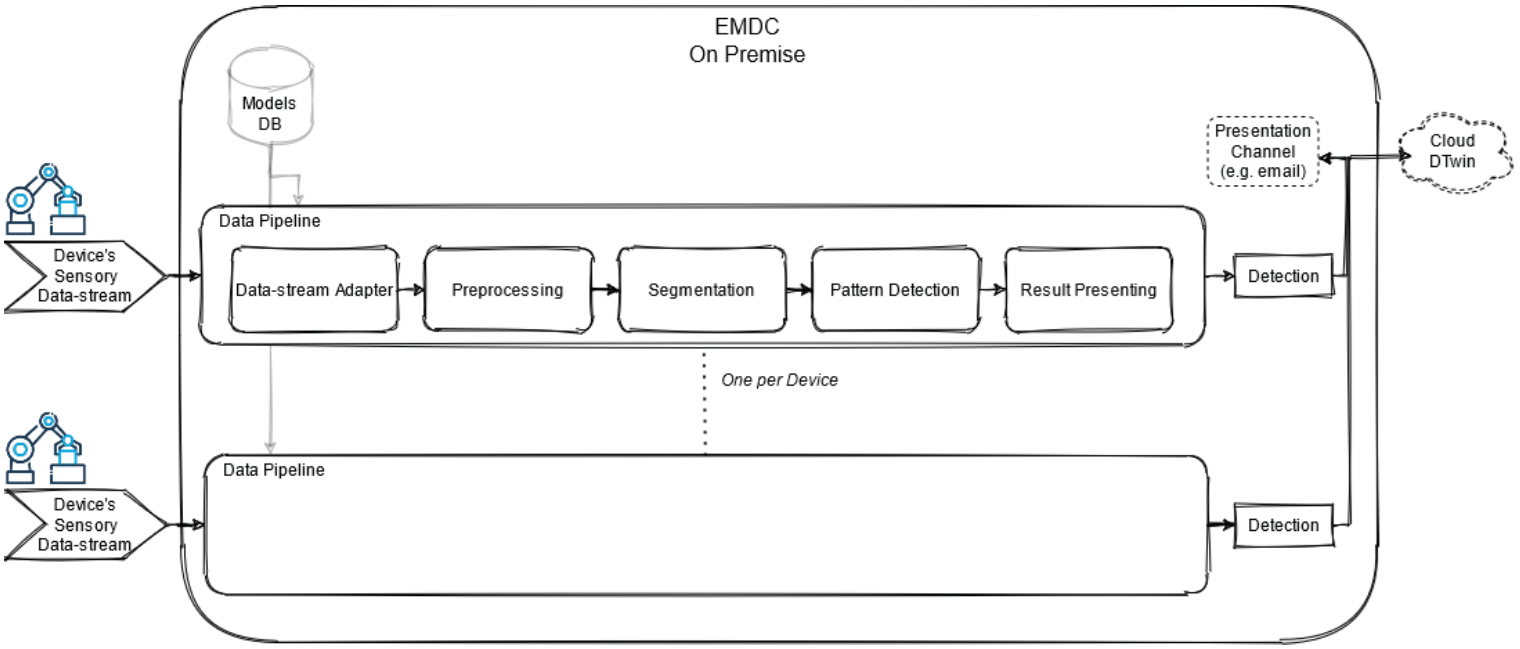
Device 1



Device N

MOD Workflow | Detection Phase|Pipeline Abstraction

- One pipeline per one IoT device.
  - Models DB used for configuring the pipeline.
- #Devices:#Pipelines = 1:1.



Questions

Acknowledgment

BRAINE

This project has received funding from the ECSEL Joint Undertaking (JU) under grant agreement No 876967. The JU receives support from the European Union's Horizon 2020 research and innovation programme

https://www.braine-project.eu/

Multiagent production planning

Vojtěch Janů, CTU, vojtech.janu@cvut.cz

BRAINE: Big data pRocessing and Artificial Intelligence at the Network Edge
H2020 ECSEL JU Grant No. 876967
https://www.braine-project.eu/

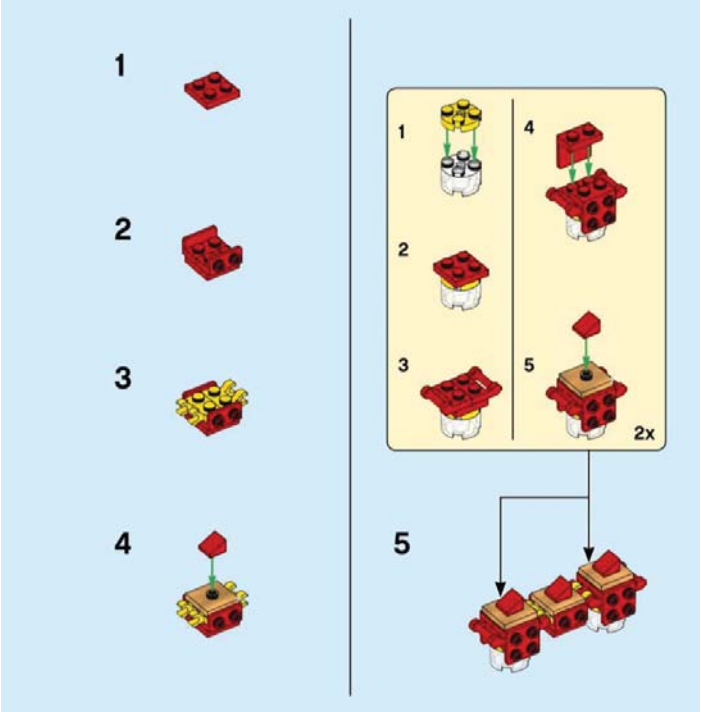
Goals

- Manufacturing of customized products
- Small batches
- Different products on one production line
- Maximum flexibility
- Continuous production (Autonomous error handling)

Domain

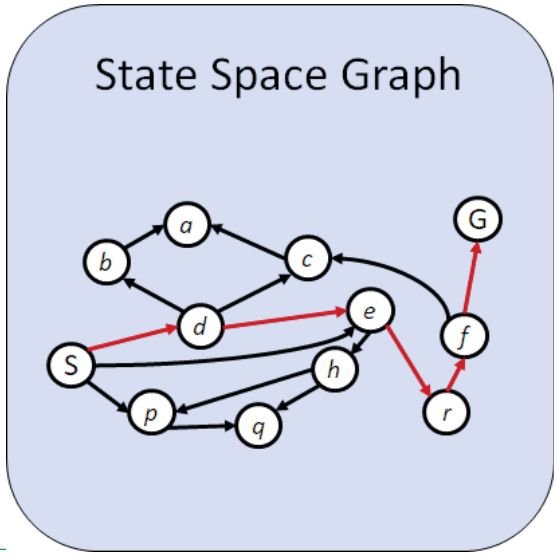
- In production one step of planning consists of
  - Resource transportation
  - Product transportation
  - Desired operation

- steps branching and chaining



Problem

- Common solution to the production planning -> use planner
- Planner advantages:
  - Can find optimal solution
  - Multiple existing algorithms and solvers
- Planner disadvantages:
  - Long planning time (grows exponentially with number of states e.g. machines)
  - new product order -> new plan
  - Change to the process (runtime error) -> new plan



solution

- Agent based reactive planning
- Reactive planning advantages
  - Planning on the go -> natural reactivity
  - Distributed computation -> Low planning time
- Reactive planning disadvantages
  - Only suboptimal
  - No existing up-to-date implementation

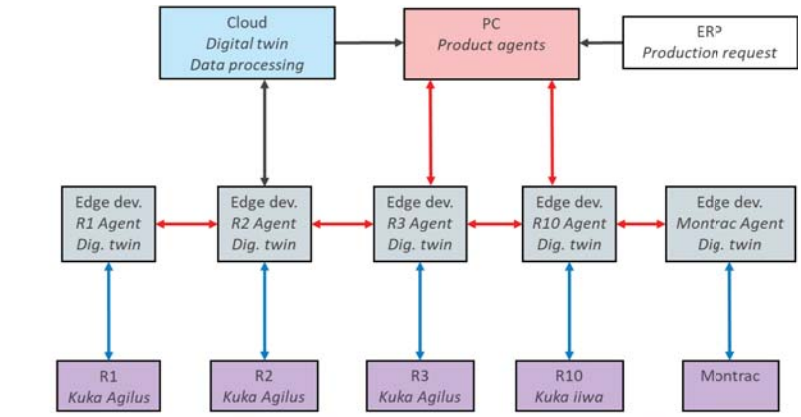
What we do

- Development of Agent based multiagent architecture
  - Build on top of production ready IT technologies (Kubernetes, Docker/Containerd, Spring, Zookeeper, RabbitMQ…)
  - Inspired by enterprise microservice architecture
  - Inspired by FIPA agent architecture
  - Incorporate Industry 4.0 technologies (OPC UA, MQTT,…)

- Devise methods for flexible manufacturing and production planning
  - Focus on SME

Agent based reactive planning

- Each agent provides some capabilities to others (acts in FIPA)
  - Capabilities can be physical eg. Mount
- Agents can query and respond to other agents
- Some agents (product agents) initiate the coversation

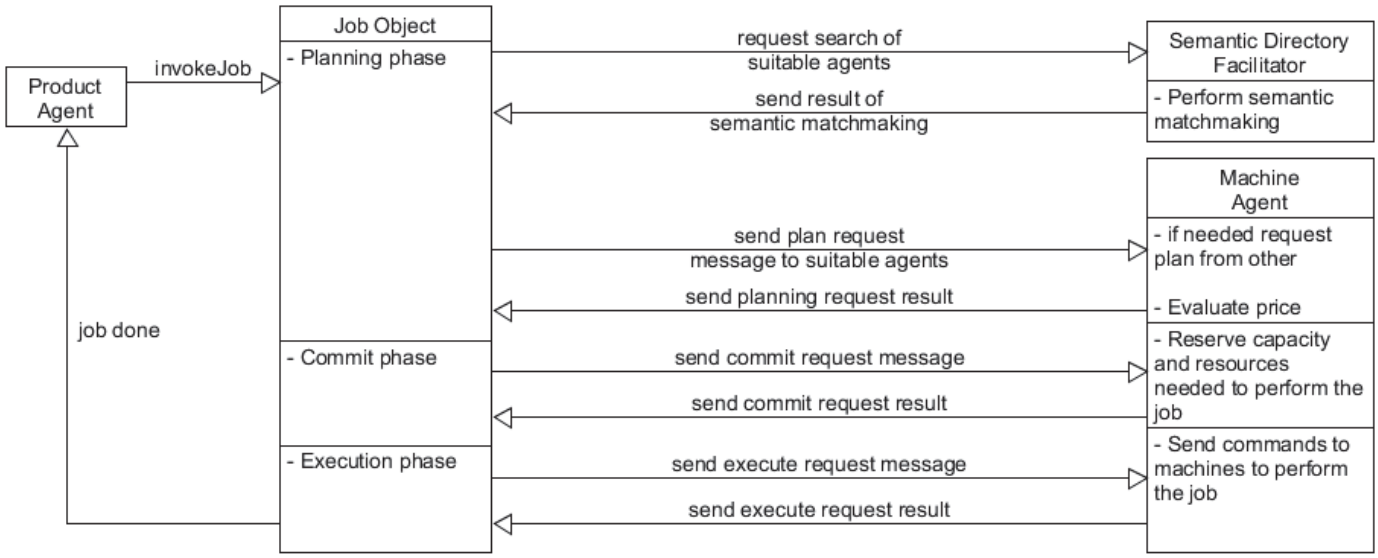


Product description

- Product description in steps:
  1. Place board
  2. Place first dragon
  3. Place second dragon
  4. Place third dragon

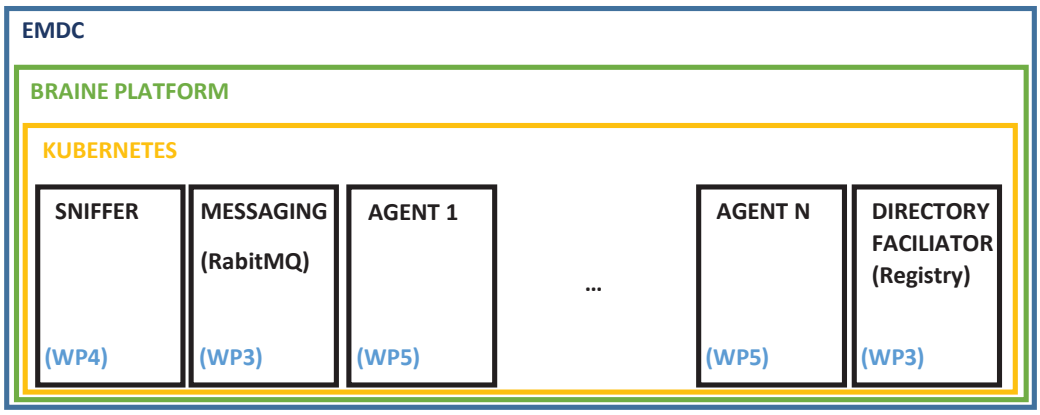
Negotiation

- Plan – Commit - Execute



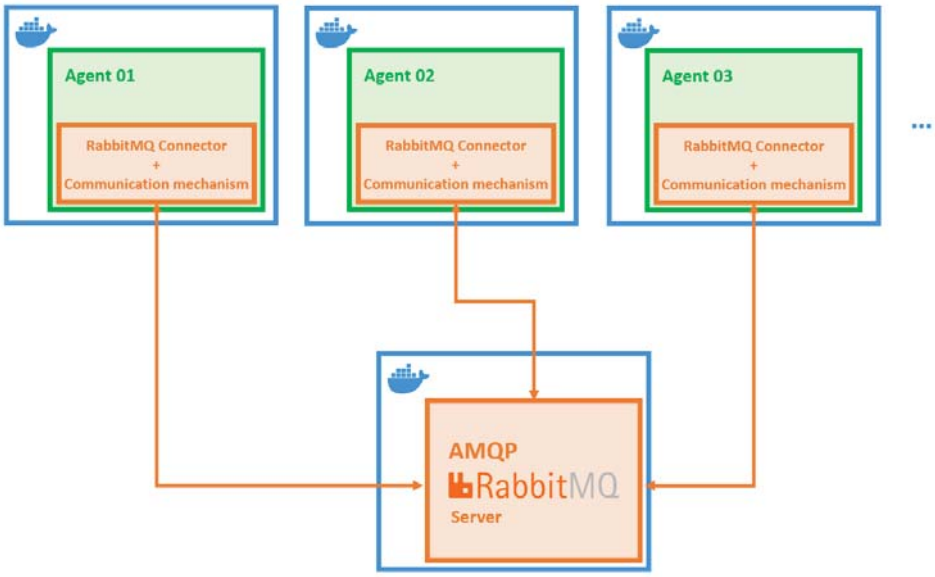
Multi-agent platform in Braine

- Multiagent platform can run across multiple EMDC (Edge Micro Data Center)
- Utilizes functionality of Braine such as Workload placement, SLAs, etc.



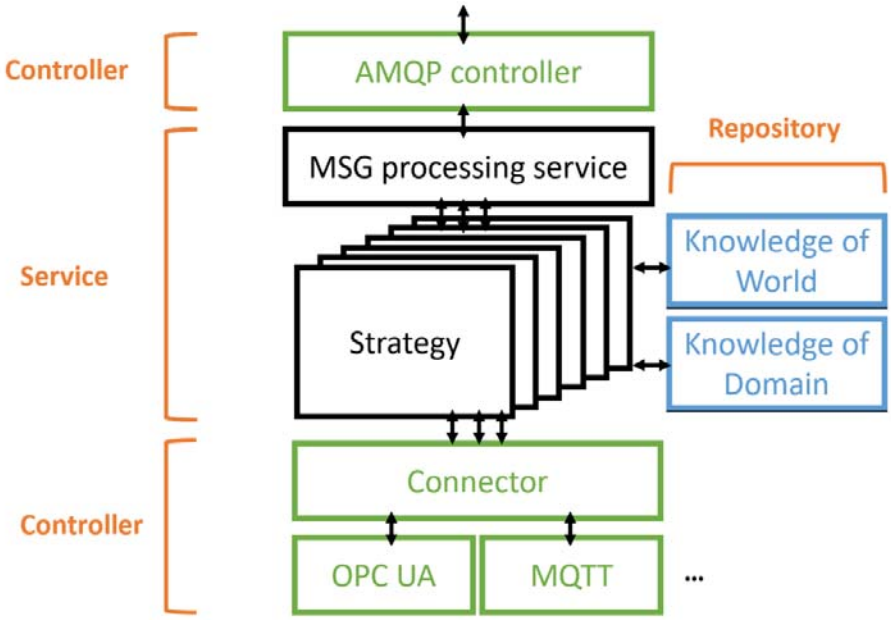
Communication

- Uses RabbitMQ as a broker
- Uses AMQP as a communication protocol



Agent Architecture

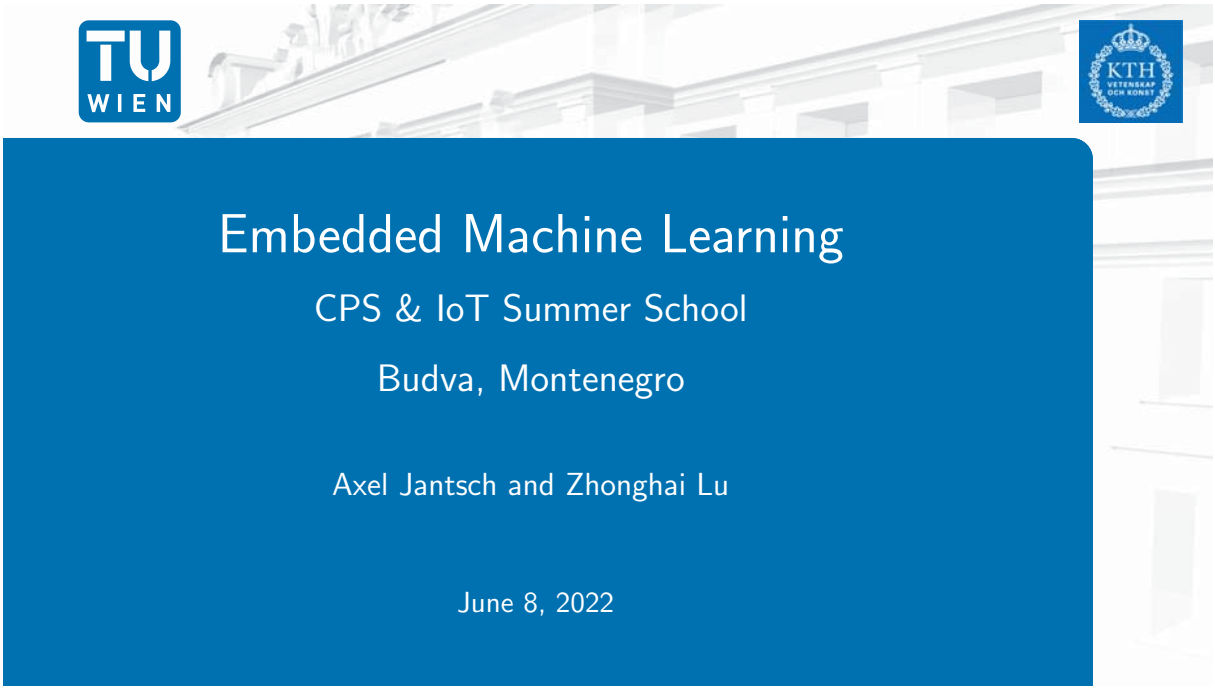
- Agent architecture allows asynchronous message handling



Acknowledgment

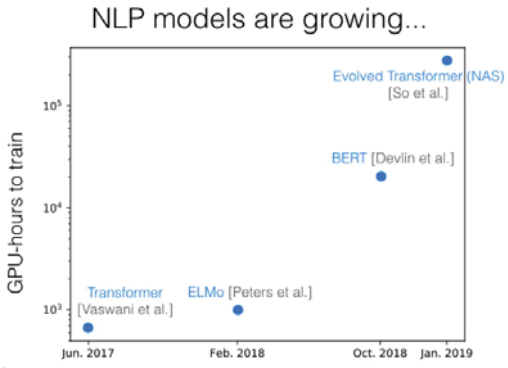
This project has received funding from the ECSEL Joint Undertaking (JU) under grant agreement No 876967. The JU receives support from the European Union's Horizon 2020 research and innovation programme

https://www.braine-project.eu/


TU
WIEN
KTH
VETENSKAP
OCH KONST
Embedded Machine Learning
CPS & IoT Summer School
Budva, Montenegro
Axel Jantsch and Zhonghai Lu
June 8, 2022

Machine Learning is Resource Intensive

- NAS based training is beyond the reach of most organizations

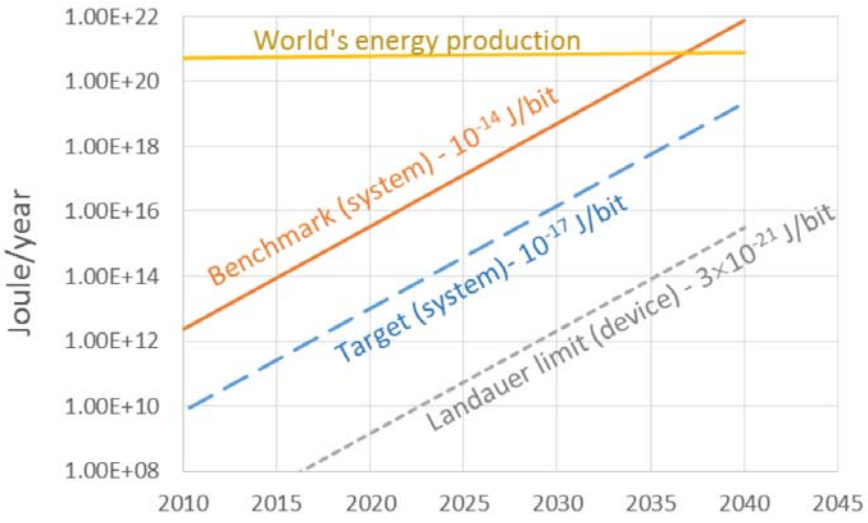


Full architecture search for a big transformer model requires

- 979M training steps and
- 32,623hours of TPU or 274,120 hours on 8 P100 GPUs,
- carbon footprint equivalent to the **lifetime of 5 US cars**.

Emma Strubell, Ananya Ganesh, and Andrew McCallum. "Energy and Policy Considerations for Deep Learning in NLP". In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Florence, Italy: Association for Computational Linguistics, July 2019, pages 3645–3650

ML is Resource Usage is Unsustainable

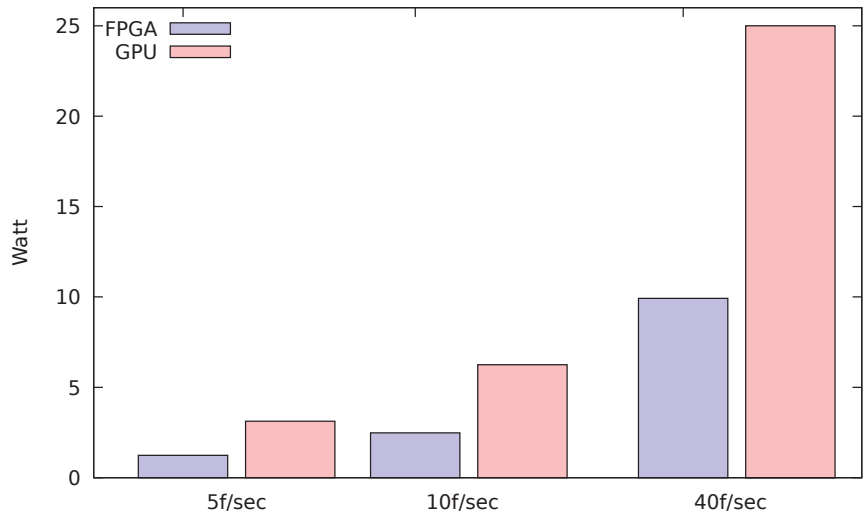


www.ict.tuwien.ac.at

SIA - SRC. *Rebooting the IT Revolution: A Call to Action.* Technical report. Semiconductor Industry Association and Semiconductor Research Corporation, Sept. 2015

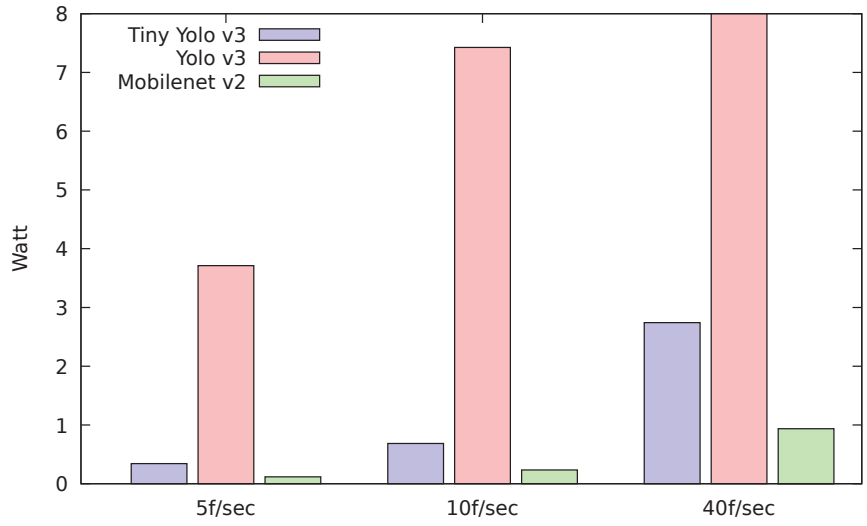
3

Power Consumption in Inference



VGG16 applied to the ImageNet data set based on published papers.

Power Consumption in Inference



Object detection on the NCS2 platform; own measurements.

What is Special About "Embedded"?

Resource limitations

	Embedded	Server farm
Computation [flop]	$30 - 1800 \cdot 10^{12}$	$86 \cdot 10^{18}$
Memory [bit]	10^{10}	10^{15}
Power [W]	5-100	$10^3 - 10^6$
Energy [Wh]	48-1000	$200 \cdot 10^6$

Computation Embedded refers to an Nvidia Jetson Nano running 1 min and 1 hour, respectively.
Computation server refers to the computation needed for the 40 day experiment with AlphaGo Zero
Energy embedded refers to a mobile phone and to a car battery, respectively.
Energy server refers to the 40 day experiment for AlphaGo Zero.

Embedded Machine Learning

Part I
Exploring Parallelization for Multi Layer Perceptrons

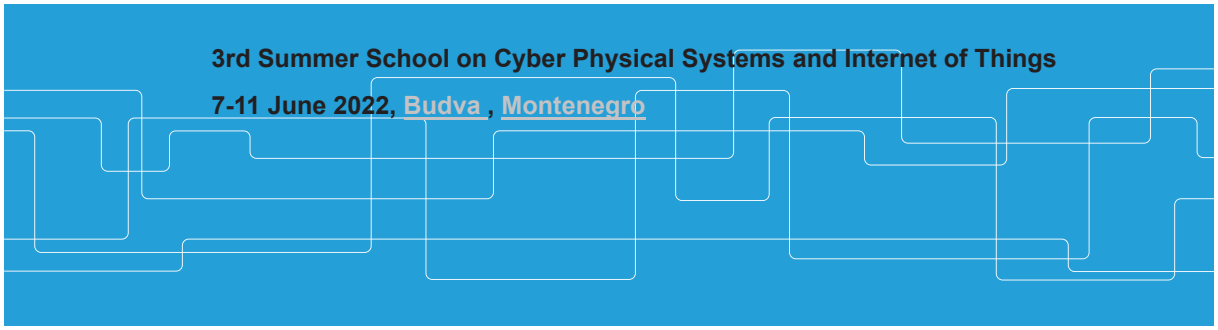
Zhonghai Lu, KTH

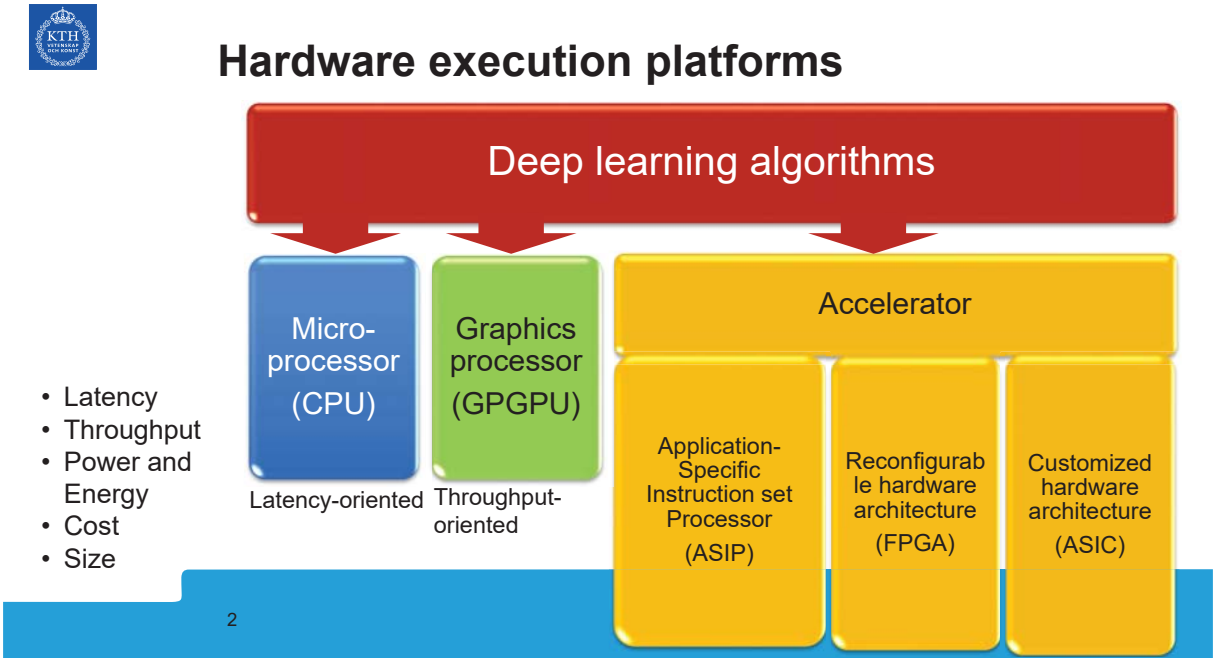
Part II
Exploring the Design Space

Axel Jantsch, TU Wien

Exploring Parallelization for MLP Hardware Acceleration

Prof. Zhonghai Lu, zhonghai@kth.se
KTH Royal Institute of Technology




Hardware execution platforms
Deep learning algorithms
Micro-processor (CPU)
Graphics processor (GPGPU)
Accelerator
Application-Specific Instruction set Processor (ASIP)
Reconfigurable hardware architecture (FPGA)
Customized hardware architecture (ASIC)
Latency-oriented
Throughput-oriented
• Latency
• Throughput
• Power and Energy
• Cost
• Size
2

Multi-Layer Perceptron (MLP) is important

- MLPs are universal function approximators as shown by Cybenko's theorem*.
  - They can be used to create mathematical models by regression analysis.
- MLPs represent 61% of inference workloads in Google's data centers.
  - One MLP is **BrainRank**, the algorithm for ranking search results.
- A major AI model for Facebook's products and services including ads, news feed, search, sigma etc.

- Cybenko, G. "Approximation by superpositions of a sigmoidal function". Mathematics of Control, Signals, and Systems, 2(4), 303–314, 1989.
- N. P. Jouppi, C. Young, N. Patil, D. Patterson, G. Agrawal et al.,"In-datacenter performance analysis of a tensor processing unit," ISCA 2017, pp. 1–12.
- K. Hazelwood et al., "Applied Machine Learning at Facebook: A Datacenter Infrastructure Perspective," HPCA 2018, pp. 620-629,

Inference workloads in Google's Data Centers

- MLPs represent 61% of total inference workloads while CNNs and LSTMs take 19% and 5%, respectively.

Name	*LOC*	*Layers*					*Nonlinear function*	*Weights*	*TPU Ops / Weight Byte*	*TPU Batch Size*	*% of Deployed TPUs in July 2016*
		FC	*Conv*	*Vector*	*Pool*	*Total*					
MLP0	100	5				5	ReLU	20M	200	200	61%
MLP1	1000	4				4	ReLU	5M	168	168	
LSTM0	1000	24		34		58	sigmoid, tanh	52M	64	64	29%
LSTM1	1500	37		19		56	sigmoid, tanh	34M	96	96	
CNN0	1000		16			16	ReLU	8M	2888	8	5%
CNN1	1000	4	72		13	89	ReLU	100M	1750	32	

Table 1. Six NN applications (two per NN type) that represent 95% of the TPU's workload. The columns are the NN name; the number of lines of code; the types and number of layers in the NN (FC is fully connected, Conv is convolution, Vector is self-explanatory, Pool is pooling, which does nonlinear downsizing on the TPU; and TPU application popularity in July 2016. One DNN is RankBrain [Cla15]; one LSTM is a subset of GNM Translate [Wu16]; one CNN is Inception; and the other CNN is DeepMind AlphaGo [Sil16][Jou15].

- One MLP is the heart of **BrainRank**, the algorithm for ranking search results.

N. P. Jouppi, C. Young, N. Patil, D. Patterson, G. Agrawal et al.,"In-datacenter performance analysis of a tensor processing unit," ISCA 2017, pp. 1–12.

Facebook services leveraging ML

- **Ads**: determines which ads to display to a given user.
- **News Feed**: ranking algorithms help people see the stories that matter most to them first, every time they visit Facebook.
- **Search**: launches a series of distinct and specialized sub-searches to the various verticals, e.g., videos, photos, people, events, etc.
- **Sigma** is the general classification and anomaly detection framework for internal applications.

Models	Services
Support Vector Machines (SVM)	Facer (User Matching)
Gradient Boosted Decision Trees (GBDT)	Sigma
Multi-Layer Perceptron (MLP)	Ads, News Feed, Search, Sigma
Convolutional Neural Networks (CNN)	Lumos, Facer (Feature Extraction)
Recurrent Neural Networks (RNN)	Text Understanding, Translation, Speech Recognition

TABLE I
MACHINE LEARNING ALGORITHMS LEVERAGED BY PRODUCT/SERVICE.

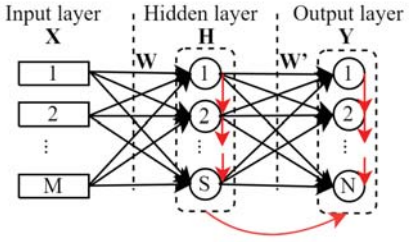
K. Hazelwood et al., "Applied Machine Learning at Facebook: A Datacenter Infrastructure Perspective," 2018 IEEE International Symposium on High Performance Computer Architecture (HPCA), 2018, pp. 620-629,

Outline

- MLP hardware designs
  - *Intra-layer parallelization*: layer-after-layer processing, but computing multiple neurons of the same layer in parallel
  - *Inter-layer parallelization*: break layer-after-layer processing, enabling processing multiple neurons of two layers in parallel
- Evaluation
- Conclusion

MLP(M, N, S)

- Different representations of MLP: graphical, math, software

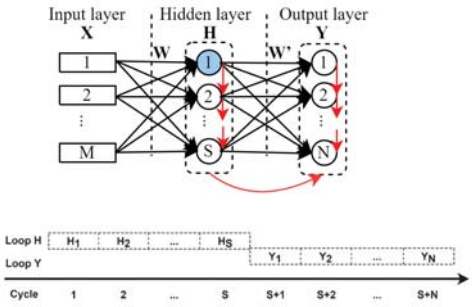


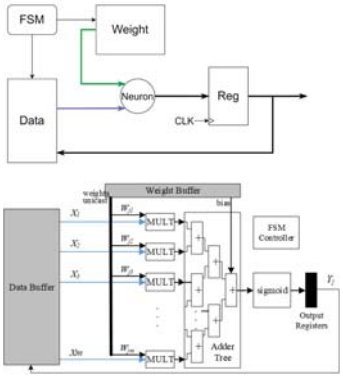
$$\vec{H} = f(W_{S\times M} \times \vec{X} + \vec{b})$$
$$\vec{Y} = f(W'_{N\times S} \times \vec{H} + \vec{b'})$$

```
hidden layer   for (int j = 0; j < S; j++)          //loop S1
                   for (int i = 0; i<M; i++)        //loop M
                       H[j] += W[j,i]*X[i];
output layer   for (int k = 0;k < N; k++)          //loop N
                   for(int j = 0; j < S; j++)       //loop S2
                       Y[k] += W'[k,j]*H[j];
```

Sequential layer-after-layer processing

Design with a single hardware neuron



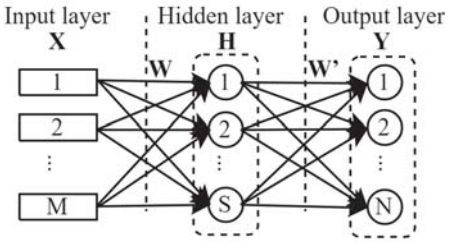


- **Fully serialized execution:** Temporally reuse one hardware neuron.
- Most area-efficient

- Latency: S+N cycles
- Area: one hardware neuron

Parallelism in MLP

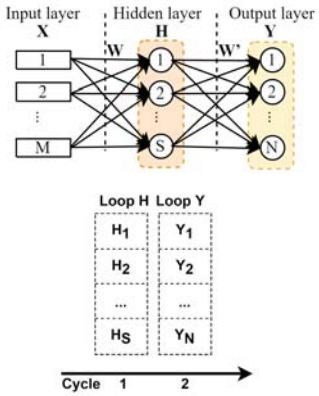
- Each layer is parallel computation
- Layer to layer execution is dependent

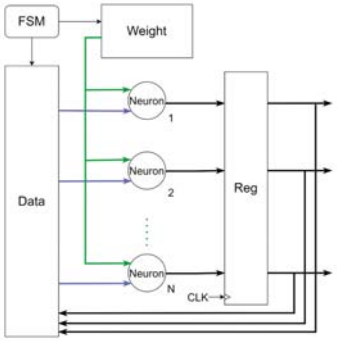


- How to exploit parallelism?
- What kinds of parallelization possible?
  - Intra-layer parallelization
  - Inter-layer parallelization

9

Design with one-layer hardware neurons



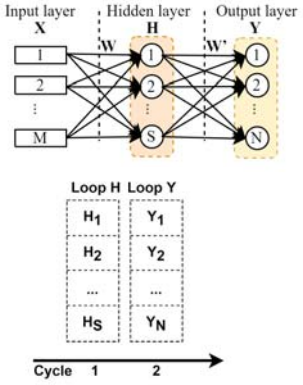


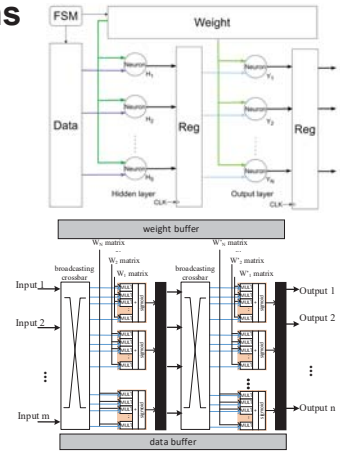
- Intra-layer parallelization: each layer is fully parallel
- No pipeline

- Latency: 2 cycles
- Area: S or N hardware neurons
- Throughput: 0.5

10

Design with all hardware neurons



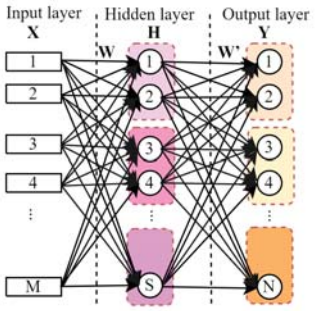


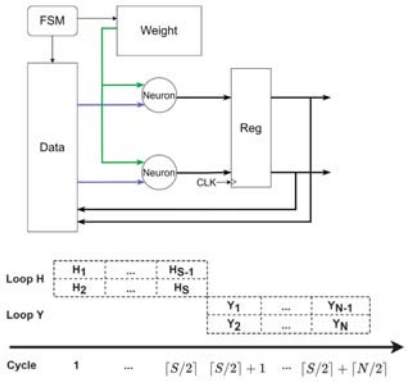
- Intra-layer parallelization: each layer is fully parallel
- Layer pipeline (2 pipeline stages)

- Latency: 2 cycles
- Area: S+N hardware neurons
- Throughput: 1

11

Intra-layer parallelism in different granularity





- Partition computation of one layer into multiple computational blocks
- Intra-layer optimization: Area = 2 hardware neurons, Latency $\left\lceil \frac{S}{2} \right\rceil + \left\lceil \frac{N}{2} \right\rceil$ cycles

Pareto front

- Now we have a classic trade-off space to play with intra-layer parallelization

  1 neuron, $S + N$ cycles

  2 neurons, $\left\lceil \frac{S}{2} \right\rceil + \left\lceil \frac{N}{2} \right\rceil$ cycles

  4 neurons, $\left\lceil \frac{S}{4} \right\rceil + \left\lceil \frac{N}{4} \right\rceil$ cycles

  …

  However, all these are layer-after-layer execution.

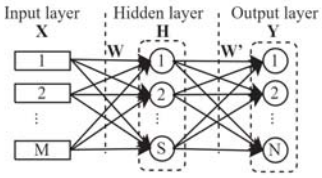
- Can we break the layer-after-layer execution, parallelize the two-layer computation?
- If and what do we gain by doing so?

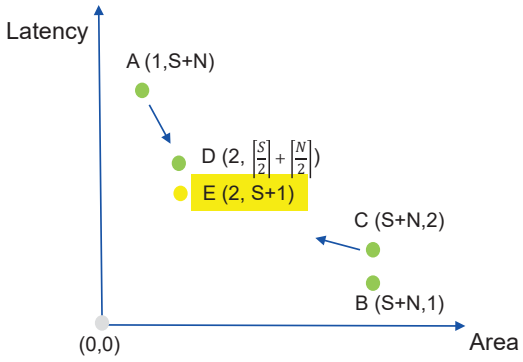
13

The Pareto front for the MLP designs

- A: area = 1 neuron, latency = S+N cycles
- B: area = S+N neurons, latency = 1 cycle (no pipeline)
- C: area = S+N neurons, latency = 2 cycles (pipelined)
- D: area = 2 neurons, latency $\left\lceil \frac{S}{2} \right\rceil + \left\lceil \frac{N}{2} \right\rceil$ cycles

MLP(M, S, N)

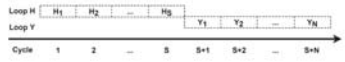




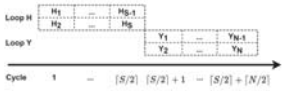
Analysis of layer-after-layer execution

1) The layer-after-layer processing results in sequential execution among layers, thus prevents the computation overlapping between layers and the further performance improvement beyond intra-layer optimization.

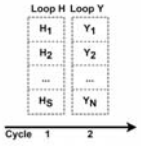
2) The layer-after-layer processing cannot immediately reuse the computed neuron, resulting in unnecessary buffer writes and reads for early computed neurons in the previous layer.



A. One-neuron design



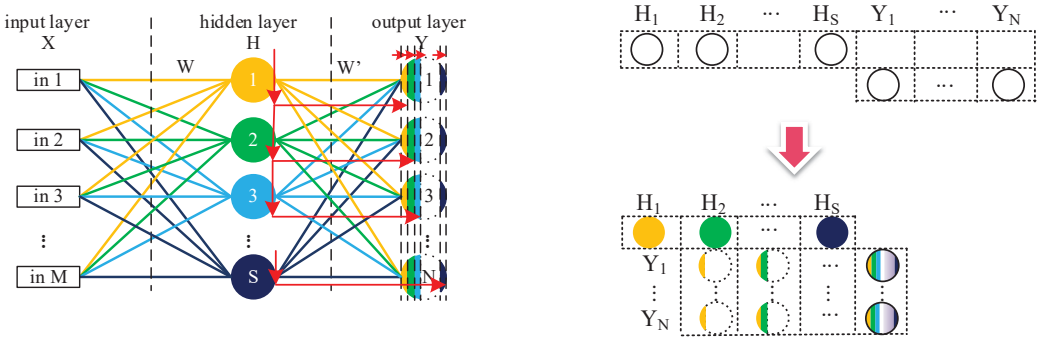
B. Two-neuron design
(Intra-layer parallelization)



D. All-neuron design

15

Inter-layer optimization: Parallelize the two-layer computation



Can we parallelize the processing of two layers at the same time?

- Yes, after the minimal dependence is resoled.

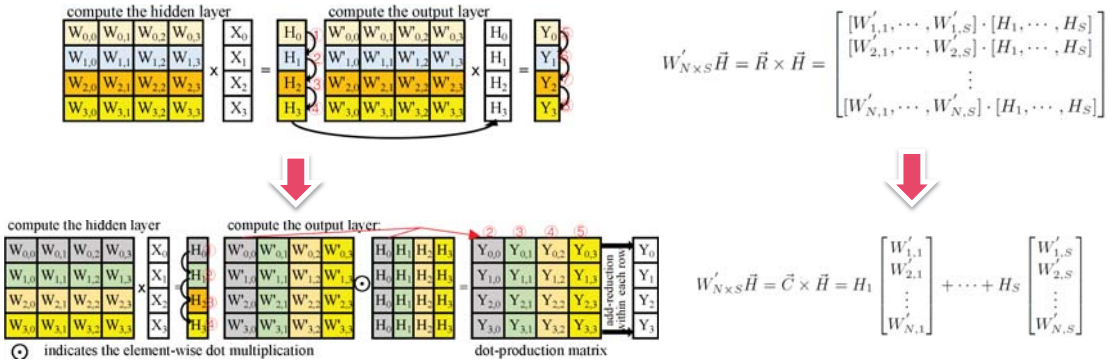
16

Benefits of the inter-layer parallelism

The benefits of the inter-layer optimization and overlapped layer processing:

- Immediate inter-layer data reuse
- Partial sum reuse in the output layer
- Extensible for more than one hidden layer

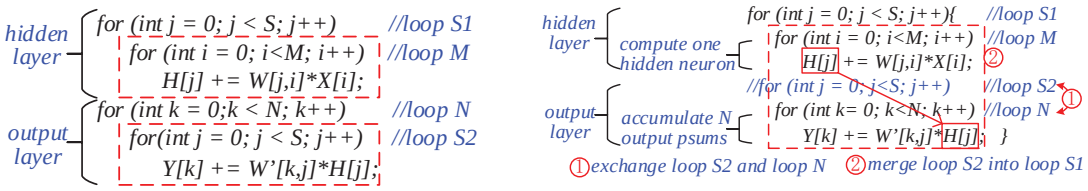
Inter-layer optimization in math



How is the output layer computed in mathematical representation?

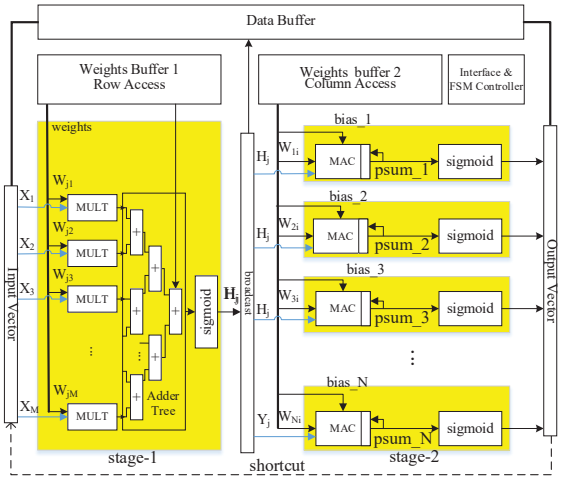
18

Inter-layer optimization in software



1. No data dependency between iterations of *loop_S2, i.e.* different *loop_Ns*.→ exchnage loop_S2 and loop_N.
2. New *loop_S2* can be merged into *loop_S1*. H[j] is used immediately after its calculation done. → the two layers' execution is combined into one iteration of the same loop. That's what we call **inter-layer optimization**.

Two-neuron architecture



Reference design is a two-neuron architecture, approximately two equivalent hardware neurons.

- 2 stages:
  - Loop_M → stage 1
  - Loop_N → stage 2
- Stage 1:
  - Compute one complete neuron each time
- Stage 2:
  - Accumulate N partial sum of N different neurons.

Comparisons

For a typical MLP(M, S, N)

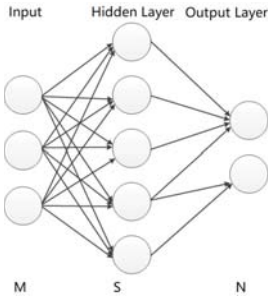
TABLE I
COMPARISON OF THE THREE MLP ARCHITECTURES

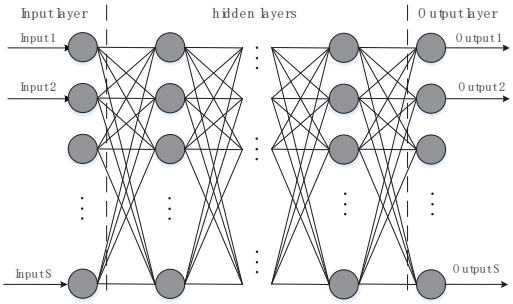
Item	Single-Neuron	Two-Neuron	All-Neuron
Area Cost (Unit)	MULT:M ADD:M SIG:1;REG:1	MULT:M, MAC:N ADD:M SIG:N+1;REG:N+1	MULT:S×(M+N) ADD:S×(M+N) SIG:S+N;REG:S+N
Performance	(S+N) cycles	(S+1) cycles	2 cycles
Weights BW (data/cycle)	M+1	M+1 for Weight buffer1 N for Weight buffer 2	S×M+S (hidden layer) N×S+N (output layer)

[a]Other area costs such as controllers are not considered here.

From MLP to DNN

Can the inter-layer optimization be applied to many hidden layers?

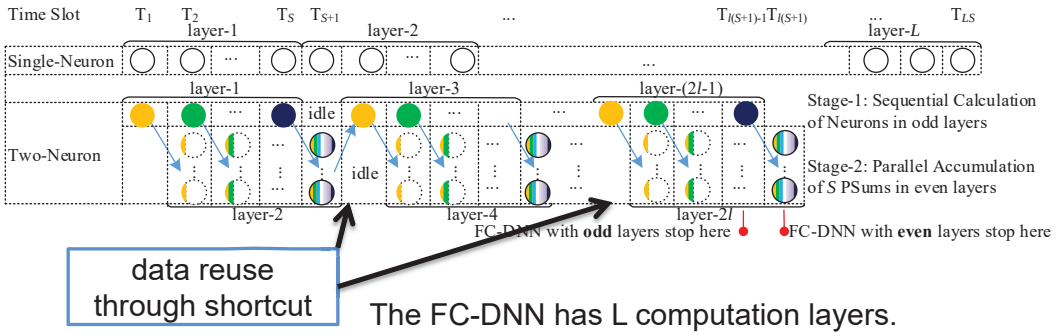




Assume that $M = S_1 = S_2 = \ldots = N = S$, where S_i is the number of neurons in the i^{th} layer.

22

Execution pipeline of a FC-DNN on two-neuron



The FC-DNN has L computation layers.

- if L is odd : (L+1)(S+1)/2-1 slots
- if L is even: L(S+1)/2 slots

23

Comparisons for FC-DNN

COMPARISON OF THE THREE ARCHITECTURES FOR FC-DNN.

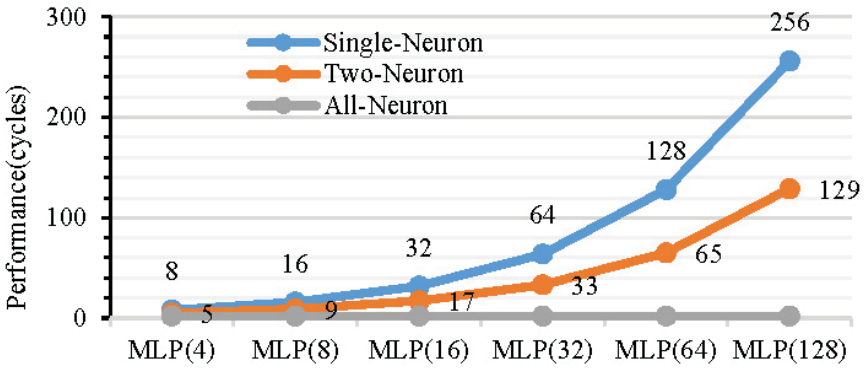
Item	Single-Neuron	Two-Neuron	All-Neuron
Performance	LS	$\lceil L/2 \rceil \times (S+1)$*	L
Speedup $[L \uparrow S \uparrow]$	1	2	S
Relative Area	1	2	LS
RPAP Ratio	1	1	L

L computation layers, S neurons per layer.

RPAP: Relative Performance-Area Product.

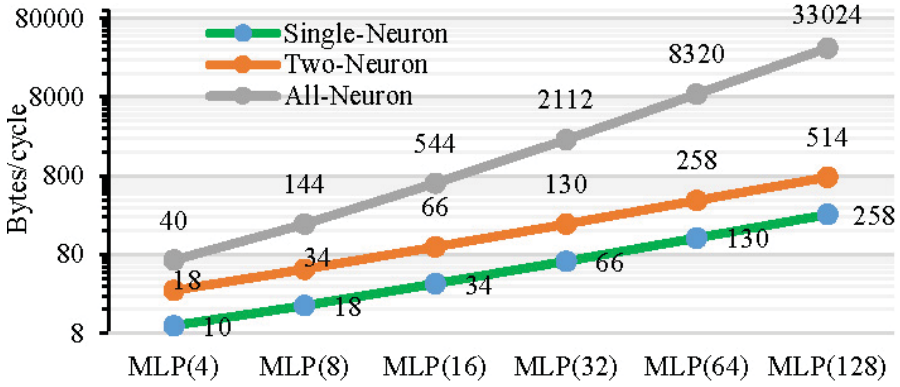
* When L is odd, there is "-1" in the performance formula.

Evaluation of performance scalability



Direct results from the analysis table (L=2).

Evaluation of weight bandwidth



Direct results from the analysis table (L=2).

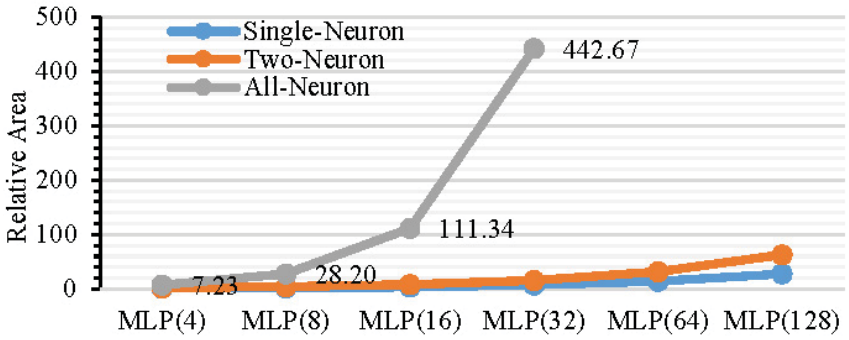
MLP(N) to represent MLP(N,N,N)

26

Hardware synthesis and results

- Verilog® RTL, Design Compiler®, No P&R.
- @200 MHz for all the configurations
- 16-bit fixed-point integer
- let M = S = N, MLP(N) to represent MLP(N;N;N)
- MLP(4)/ MLP(8)/ MLP(16)/ MLP(32)/ MLP(64)/ MLP(128) are evaluated.
  - For All-Neuron architecture, MLP(64) and MLP(128) are omitted because of too much area and too many connection wires. Synthesis cannot finish properly.

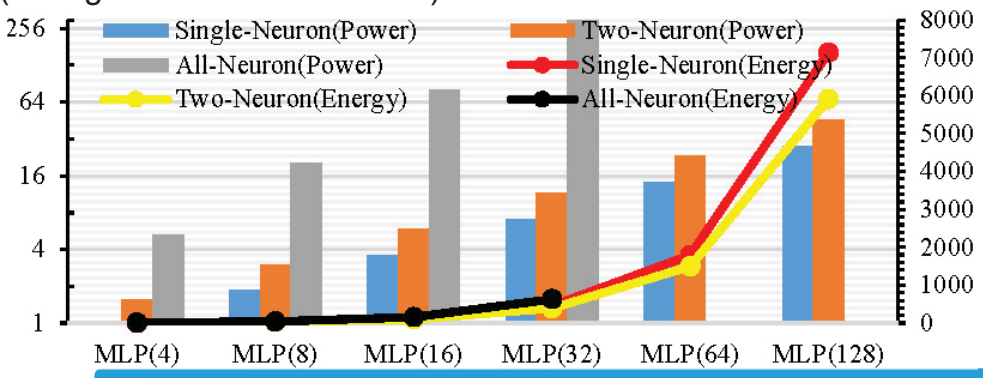
Relative Area Cost over Single-Neuron



Numbers on the grey curve are for All Neuron.

Relative power and energy

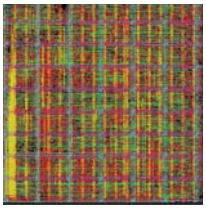
Relative power (the left Y-axis and the bars) and energy consumption (the right Y-axis and the lines).



29

Comparisons with a related work

Comparison with RNA[11], a PE-based Reconfigurable Neural Architecture, @65nm, 500MHz



Physic view of Two-Neuron

COMPARISON WITH A STATE-OF-THE-ART MLP ACCELERATOR

Item	Two-Neuron MLP(8, 8, 8)	RNA [11]
Multiplier	8	-
MACs	8	16
Adders	8	16
Sigmoid	9	16
Frequency	500MHz@65nm	500MHz@65nm
Computation Area(P&R)	96,720 um^2	183,184 um^2
Execution time of MLP(8, 8, 8)	9	16

RNA: Reconfigurability for several aspects, Scalability for CNN
Area: Only computational parts concerned.

• F. Tu, S. Yin, P. Ouyang, L. Liu, and S. Wei, "Reconfigurable architecture for neural approximation in multimedia computing," IEEE Transactions on Circuits and Systems for Video Technology, vol. 29, no. 3, pp. 892–906, 2019.

Conclusion

- MLPs can be parallelized for intra-layer and inter-layer computation.
  - Intra-layer parallelization follows sequential layer-after-layer processing while exploiting parallelism *per layer*.
  - Inter-layer parallelization breaks the sequential layer-after-layer processing while exploiting parallelism *cross layer*.
- Inter-layer optimization allows overlapped layer processing, beneficial for low-latency small-area implementations.
  - The two-neuron reference design is evaluated together with one-neuron, all-neuron designs.
  - It opens new opportunity to improve the performance and energy efficiency of MLP hardware, besides the conventional approach.

Thanks for your attention!

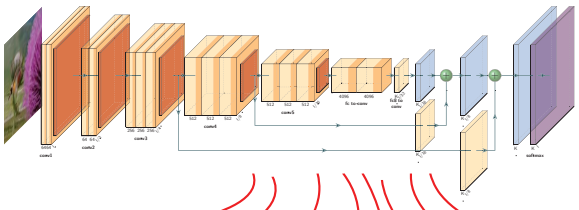
Shenggang Chen and Zhonghai Lu. "Hardware Acceleration of Multi-layer Perceptron Based on Inter-layer Optimization". IEEE 37th International Conference on Computer Design (ICCD), November 2019.

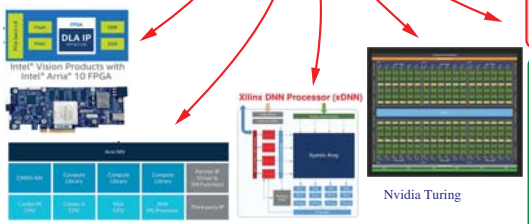
Part II

Exploring The Design Space

Axel Jantsch, TU Wien

Design Space





DNN Choices

Convolutional layers
Filter kernels
Number of filters
Pooling layers
Filter shape
Stride
Fully connected layer
Number of layers
Regularization
etc.

Mapping Choices

Neuron pruning
Data type selection
Approximation
Retraining
Connection pruning
Weight sparsifying
Regularization
etc.

Platform Choices

Platform Selection
Reconfiguration
Batch processing
Deep pipelining
Resource reuse
Hierarchical control
Processing unit selection
Memory allocation
Memory reuse
etc.

www.ict.tuwien.ac.at

TU WIEN

2

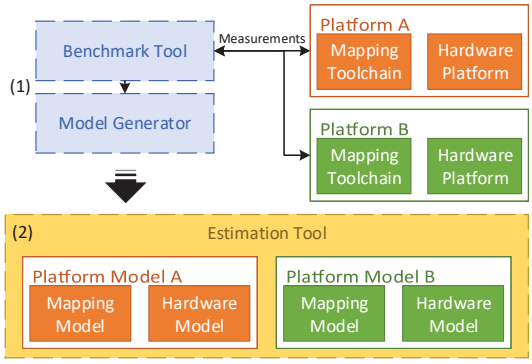
Outline

1. Estimation
2. Power Profiling
3. Traffic Light Controller Case Study

ESTIMATION

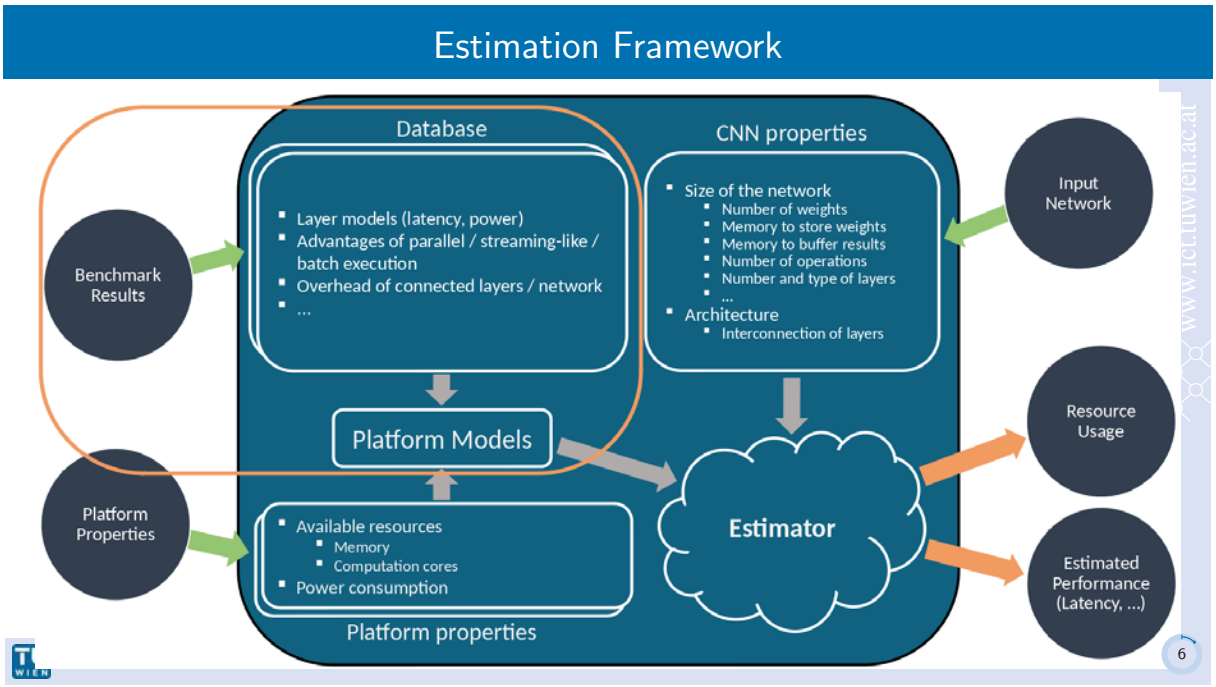
Estimation

- Two leading performance estimation tools: ANNETTE and Blackthorn
- For NCS2, Xilinx FPGA, and Jetson
- Combine analytic, statistical model and partial measurements



M. Wess, M. Ivanov, C. Unger, A. Nookala, A. Wendt, and A. Jantsch. "ANNETTE: Accurate Neural Network Execution Time Estimation With Stacked Models". In: *IEEE Access* 9 (2021), pages 3545–3556

Martin Lechner and Axel Jantsch. "Blackthorn: Latency Estimation Framework for CNNs on Embedded Nvidia Platforms". In: *IEEE Access* (2021)


Estimation Framework
Database
Layer models (latency, power)
Advantages of parallel / streaming-like / batch execution
Overhead of connected layers / network
...
CNN properties
Size of the network
Number of weights
Memory to store weights
Memory to buffer results
Number of operations
Number and type of layers
...
Architecture
Interconnection of layers
Benchmark Results
Input Network
Platform Models
Platform Properties
Available resources
Memory
Computation cores
Power consumption
Platform properties
Estimator
Resource Usage
Estimated Performance (Latency, ...)
www.ict.tuwien.ac.at
6

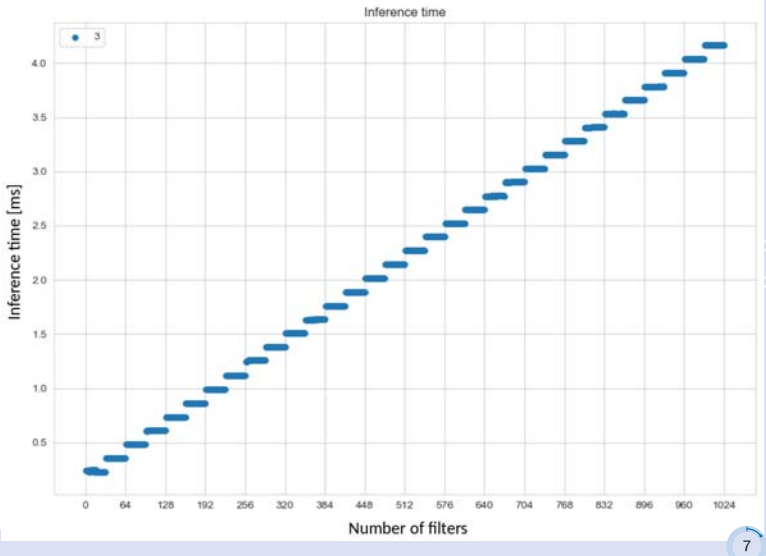
Inference Run Time Estimation

Assumption:

- Inference time as a function of problem size is a combination of step and linear functions due to limited parallel resources.
-

Example:

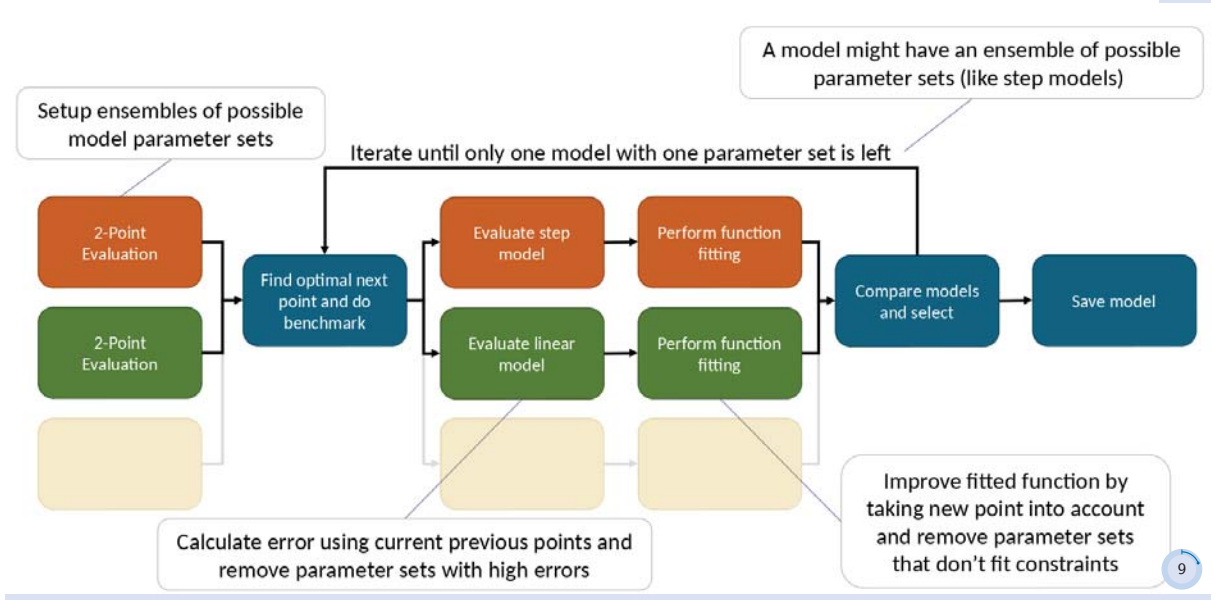
- Single convolutional layer sweep
- 32x32x64 with k filter and kernel size 3



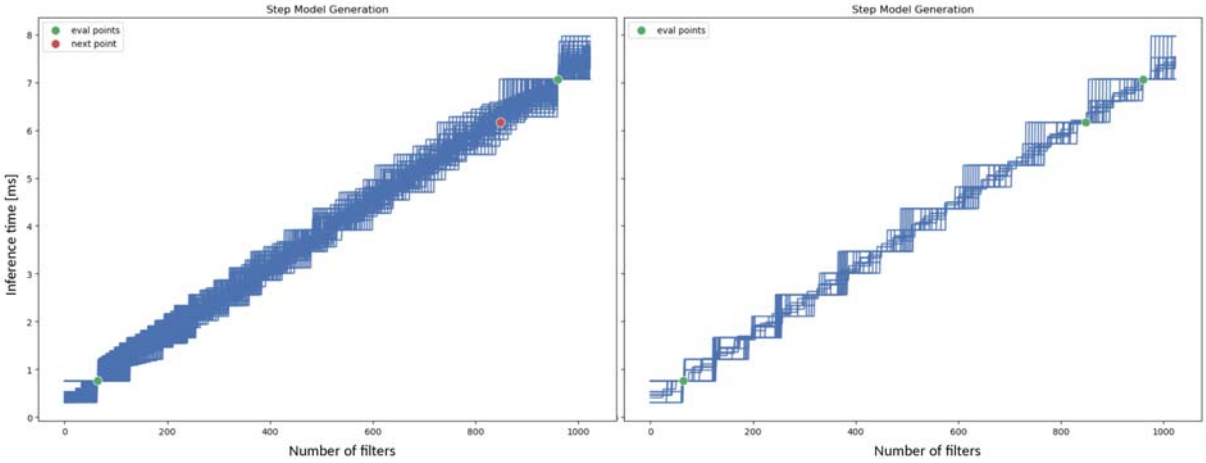
Inference Run Time Estimation

- Assumption:
  The inference time can be approximated by a combination of linear and step functions for each dimension, such as filter, channels, etc.
- Determining the function based on selected measurements
- Goals: automatic computation of estimation functions for latency, power consumption and various platforms.

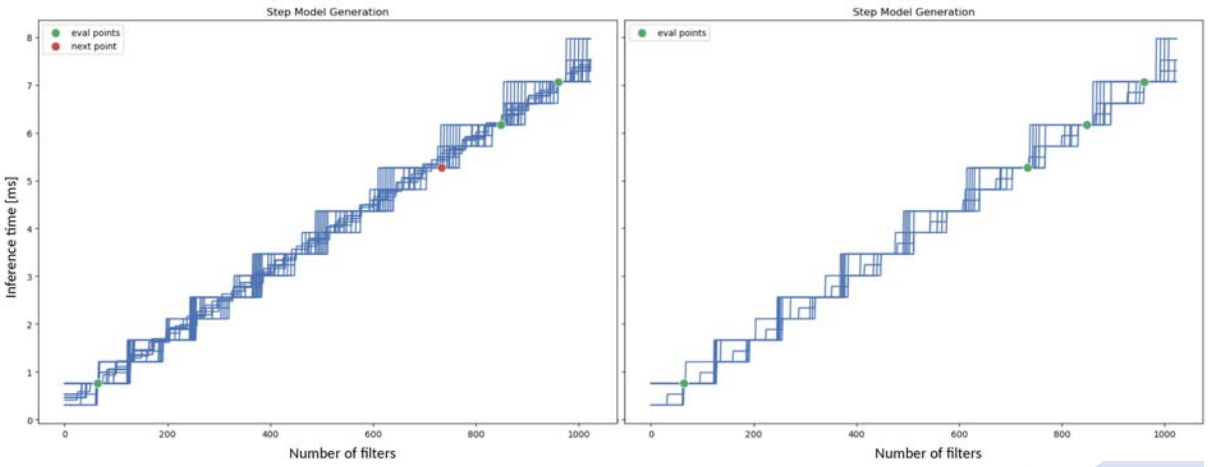
Automatic Estimation Function Generation



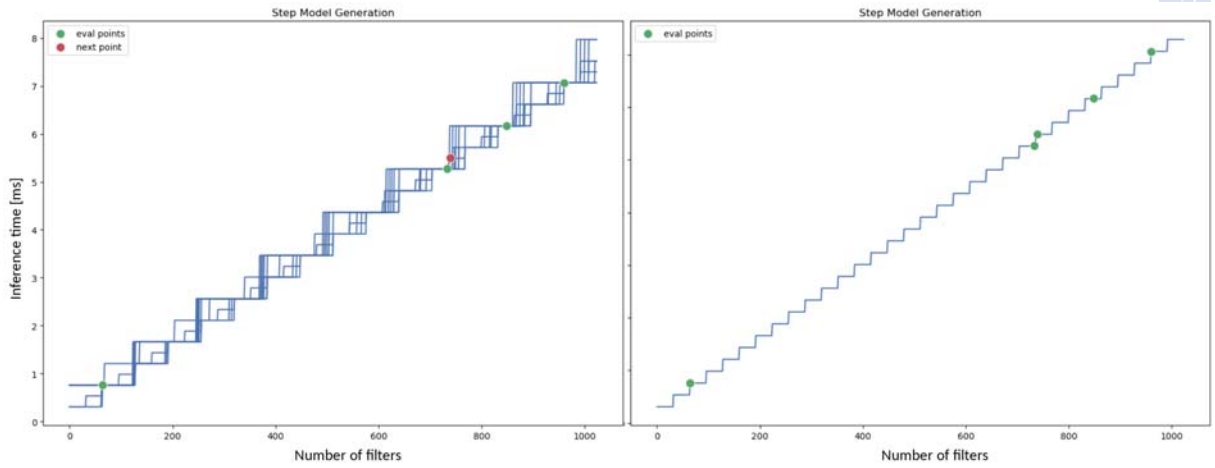
Iterative Refinement



Iterative Refinement

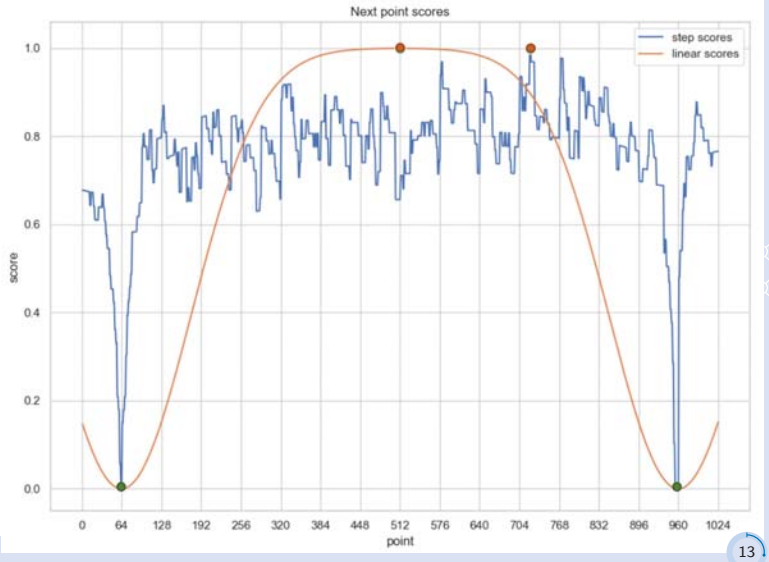


Iterative Refinement



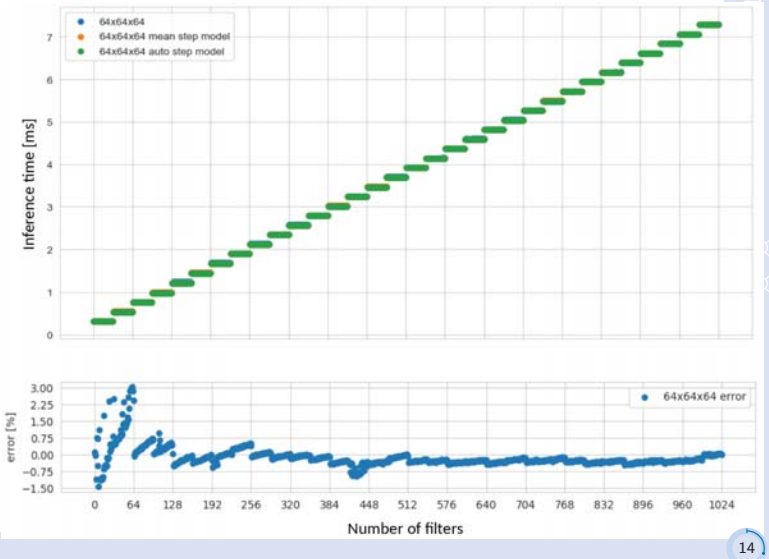
Next Point Selection

- Linear function criteria:
  - Point furthest away from previous points
- Step function criteria
  - Point with most unique discrete levels
  - Point with largest range of values
  - Point farthest away from previous points
- Next point selection: Point with highest score



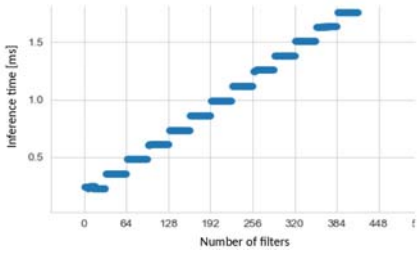
Method Evaluation

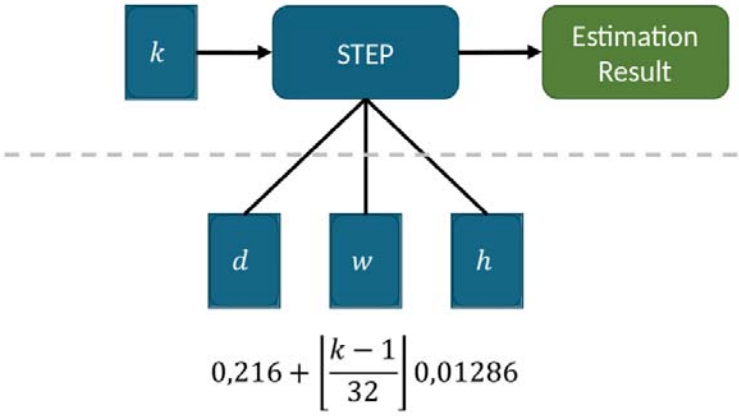
- Results after 3 iterations (5 meausurment points)
- Execution times:
  - Full sweep: 3-4 h
  - Proposed approach: 2-5 minutes



2D Example

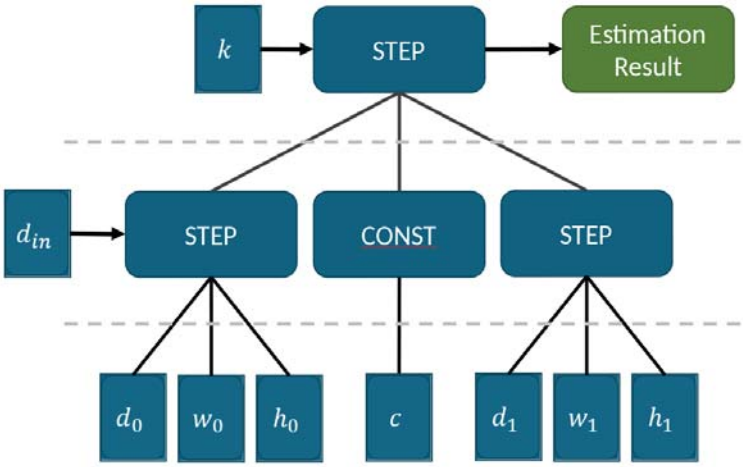
- Phase 1: Estimate function in single dimension: number of filters
- Result: step function





2D Example

- Phase 2: Test how d, w and h behave in the next dimension
- Next dimension: input channels d_{in}
- Result:
  - Step function: $d_{0=0.1418}$, $w_0 = 8$, $h_0 = 0.0106$
  - Constant: $c = 32$
  - Step function: $d_1 = 0.044$, $w_1 = 8$, $h_1 = 0.0121$



$$0.1418 + \left\lfloor \frac{d_{in}-1}{8} \right\rfloor 0.0106 + \left\lfloor \frac{k-1}{32} \right\rfloor \left(0.044 + \left\lfloor \frac{d_{in}-1}{8} \right\rfloor 0.0121\right)$$

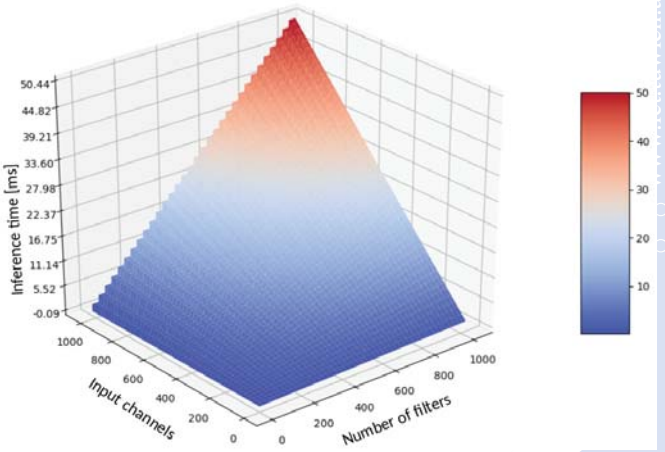
16

2D Example

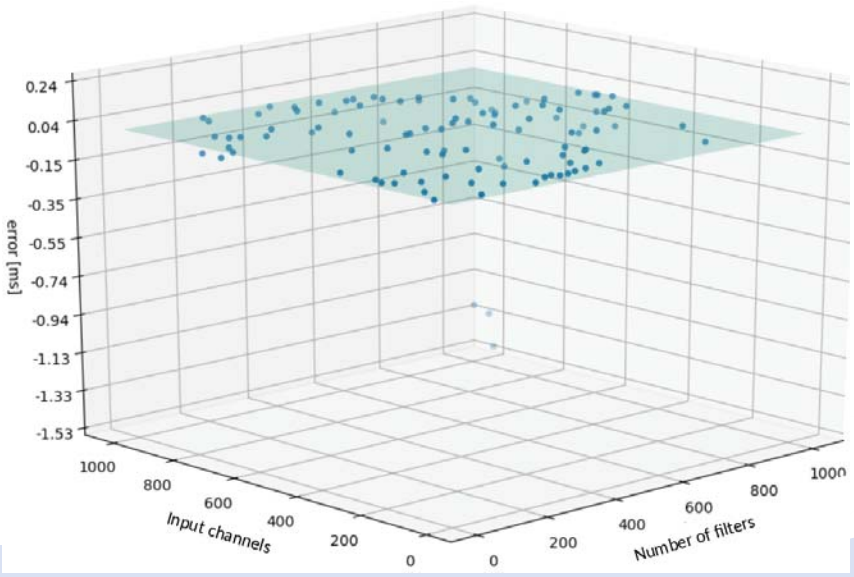
Generated model:

$$f(d_{\text{in}}, k) = 0.1418 + \lfloor \frac{d_{\text{in}} - 1}{8} \rfloor 0.0106 + \lfloor \frac{k-1}{32} \rfloor \left(0.044 + \lfloor \frac{d_{\text{in}} - 1}{8} \rfloor 0.0121\right)$$

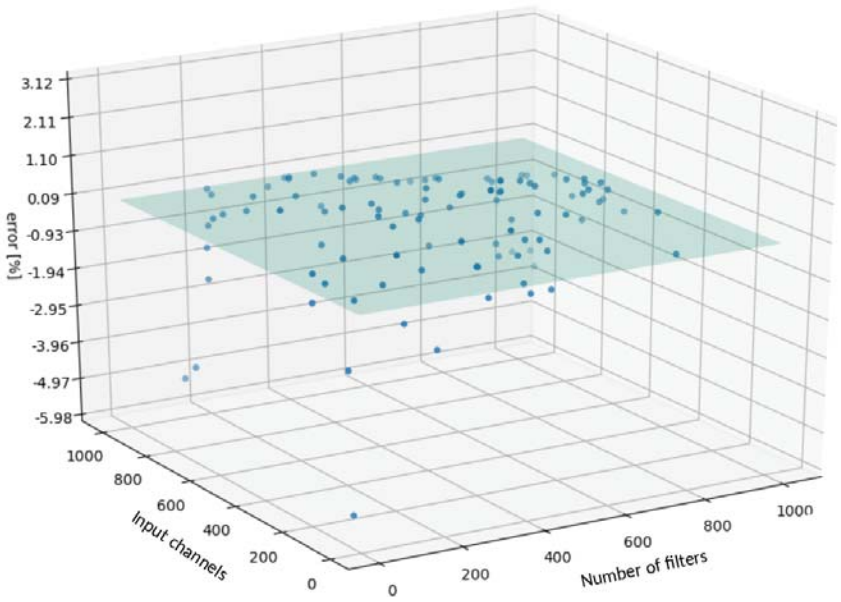
- Meausrement points: 112
- Execution time: 32 minutes



2D Example - Error



2D Example - Error

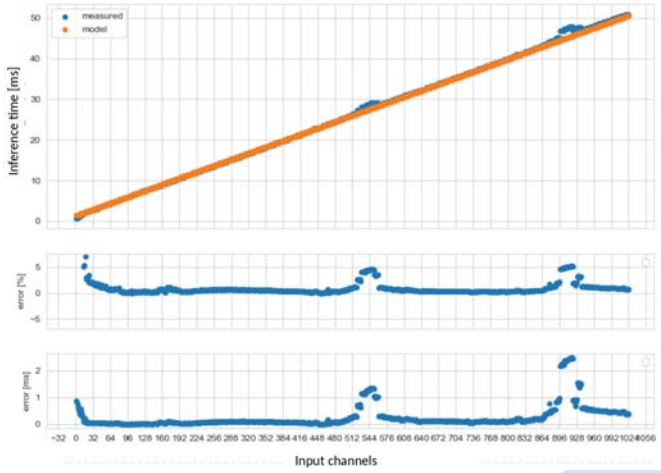


2D Example - Error

Slice through 2D plane at $k = 1024$

$$
\begin{aligned}
&f(d_{\text{in}}, k) \\
&\quad = 0.1418 + \lfloor \frac{d_{\text{in}} - 1}{8} \rfloor 0.0106 \\
&\quad + \lfloor \frac{k-1}{32} \rfloor \left(0.044 + \lfloor \frac{d_{\text{in}} - 1}{8} \rfloor 0.0121\right)
\end{aligned}
$$

$$
\begin{aligned}
&f(d_{\text{in}}, 1024) \\
&\quad = 1.5058 + \lfloor \frac{d_{\text{in}} - 1}{8} \rfloor 0.3857
\end{aligned}
$$

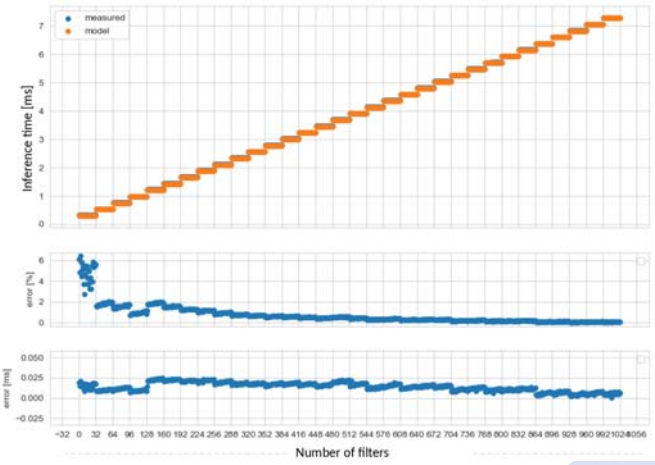


2D Example - Error

Slice through 2D plane at $d_{\text{in}} = 128$

$$
\begin{aligned}
&f(d_{\text{in}}, k) \\
&\quad = 0.1418 + \lfloor \frac{d_{\text{in}} - 1}{8} \rfloor 0.0106 \\
&\quad + \lfloor \frac{k-1}{32} \rfloor \left(0.044 + \lfloor \frac{d_{\text{in}} - 1}{8} \rfloor 0.0121\right)
\end{aligned}
$$

$$
\begin{aligned}
&f(128, k) \\
&\quad = 0.3008 + \lfloor \frac{k-1}{32} \rfloor 0.2255
\end{aligned}
$$



Latency Estimation Summary

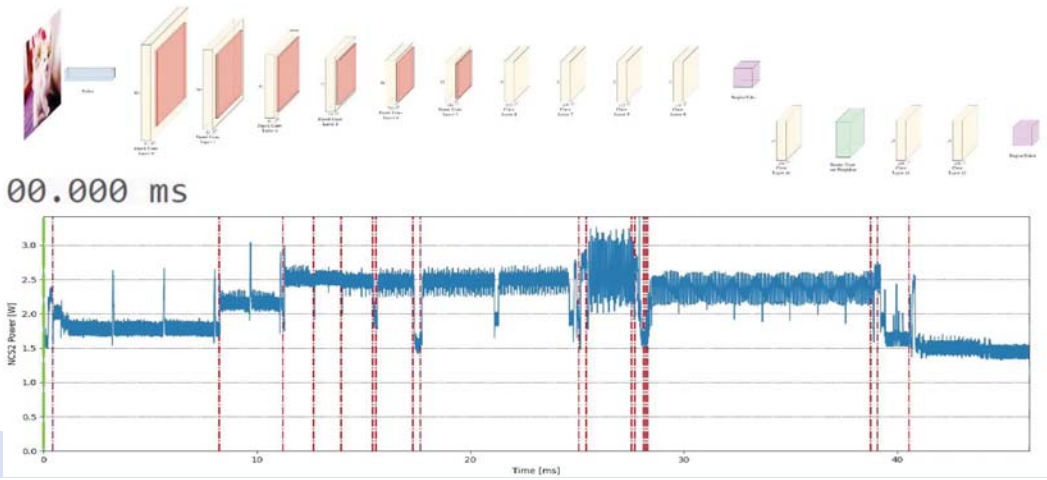
- Exploiting the discrete nature of HW resources
- Fast estimation function for latency based on linear an step functions
- Automatic derivation of estimation for a new platform
- Results for several platforms are robust

Network	Estimation Error [%]			
	NCS2	ZCU102	Jetson Nano	Jetson TX2
YoloV3	4.1	3.2	-	-
MobileNetV2	4.3	4.2	3.6	4.2
ResNet50	8.2	1.2	2.4	2.8
FPN Net	9.3	7.5	-	-
AlexNet	5.2	4.8	5.5	6.6
VGG16	11.3	6.2	0.5	1.4

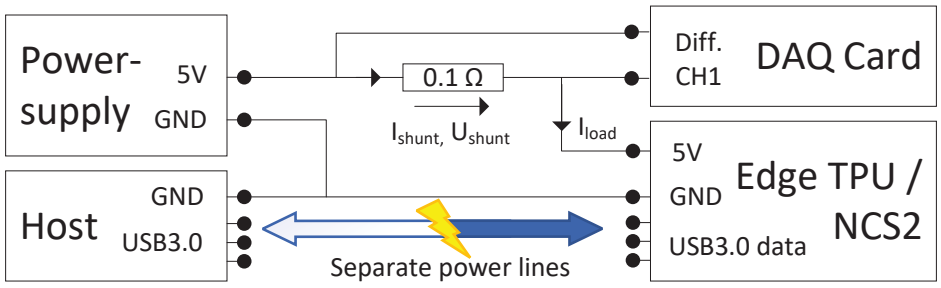
POWER PROFILING

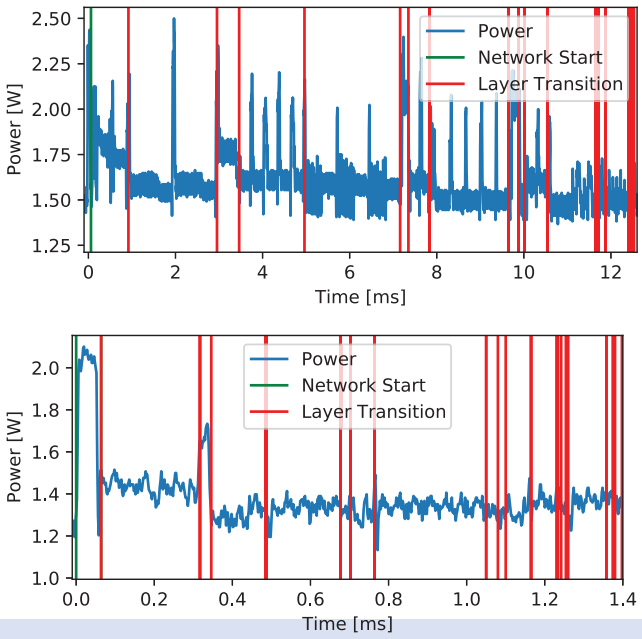
Power and Performance Profiling

Yolov3-tiny power profile on NCS2

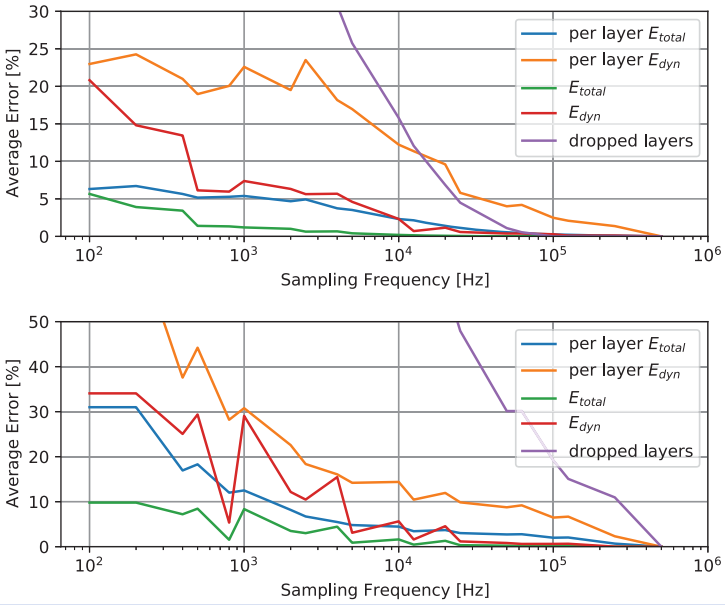


Experimental Setup



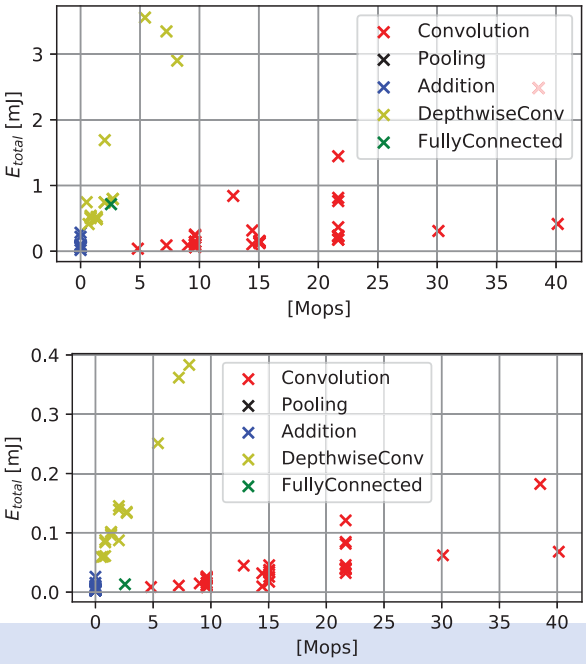


MobileNetV2 on NCS2 and Coral Edge TPU



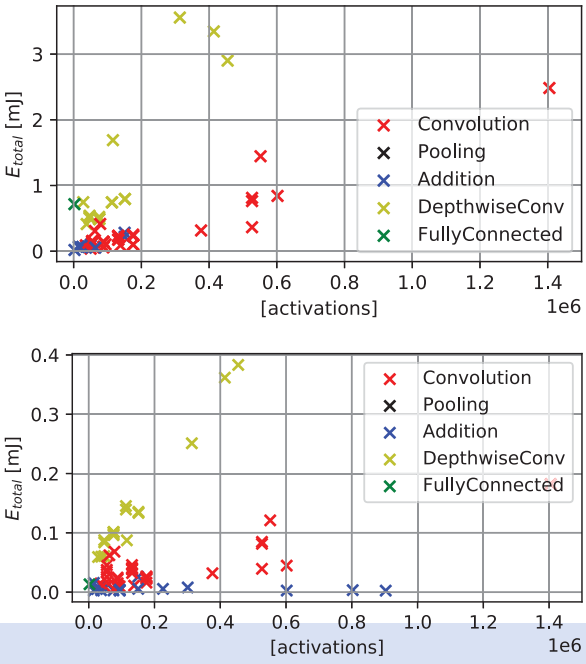
MobileNetV2 on NCS2 and Coral Edge TPU

The error in % with respect to 500 kHz sampling frequency.



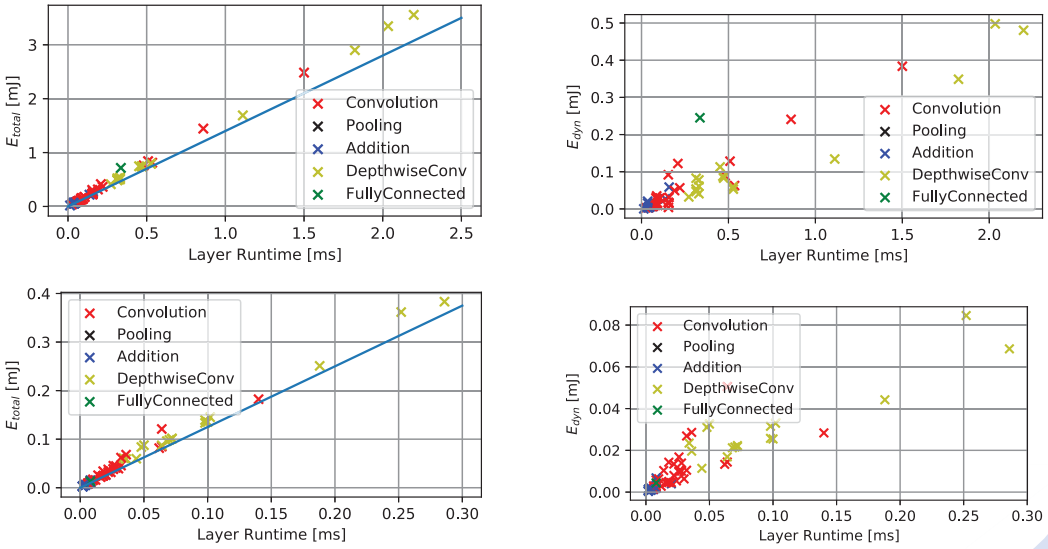
MobileNetV2 on NCS2 and Coral Edge TPU

Energy versus number of operations.



MobileNetV2 on NCS2 and Coral Edge TPU

Energy versus number of activations.



MobileNetV2 on NCS2 and Coral Edge TPU; Energy versus latency.

Profiling Results

HW	Network	*nireq*	F_{thr}(fps)	T_{lat}(ms)	P (mW)	E_{total}(mJ)	E_{base}(mJ)	E_{dyn}(mJ)	E/Gop(mJ)	$E/Mpar$(mJ)
NCS2	Tiny YOLOv3	1	21.2	41	2165	101.93	65.91	36.02	18.32	11.52
		2	35.3	52	2670	75.55	39.61	35.94	13.58	8.54
	5.6 Gop	3	43.1	46	2995	69.42	32.45	36.97	12.47	7.85
	8.8 Mpar	4	43.1	44	2954	68.54	32.48	36.06	12.32	7.75
	YOLOv3	1	2.6	363	2505	960.92	537.04	423.88	14.69	15.61
		2	4.4	400	3413	769.61	315.69	453.92	11.76	12.50
	65.8 Gop	3	4.7	425	3615	764.89	296.22	468.67	11.69	12.42
	61.6 Mpar	4	4.9	390	3604	742.50	288.43	454.07	11.35	12.06
	MobileNetV2	1	49.3	21	1806	36.60	28.37	8.23	60.84	10.55
		2	87.2	23	2118	24.29	16.06	8.23	40.38	7.00
	0.6 Gop	3	90.4	31	2164	23.95	15.49	8.46	39.81	6.90
	3.4 Mpar	4	92.4	53	2162	23.39	15.15	8.24	38.88	6.74

HW	Network	*Freq*	F_{thr}(fps)	T_{lat}(ms)	P (mW)	E_{total}(mJ)	E_{base}(mJ)	E_{dyn}(mJ)	E/Gop(mJ)	$E/Mpar$(mJ)
Edge TPU	Tiny YOLOv3	std	46.3	22.3	1407	30.40	22.28	8.12	5.46	3.44
		max	51.0	19.6	1528	29.95	20.21	9.73	5.38	3.39
	YOLOv3	std	6.3	158.3	1519	240.50	163.27	77.23	3.68	3.91
		max	7.0	142.0	1657	235.36	147.29	88.06	3.60	3.82
	MobileNetV2	std	331.3	3.0	1422	4.29	3.11	1.18	7.13	1.24
		max	512.3	1.9	1658	3.23	2.02	1.21	5.37	0.93

Power and Performance Profiling

- NCS2, Edge TPU and Nvidia platforms
- Detailed, per layer latency and power profiling
- Number of operations is a poor predictor for latency and energy
- Latency and energy usage correlate fairly well
- Hardware setting have significant influence
- 100 kHz sampling frequency is required for 5 % accuracy

TRAFFIC LIGHT CONTROLLER CASE STUDY

Traffic Light Controller

Data set:

- training: 19087 images
- positive examples 47%
- validation: 13184
- positive examples 26%
- Resolution: 1280x720
- Issue: Validation 4h/network
  → validation set: 1319

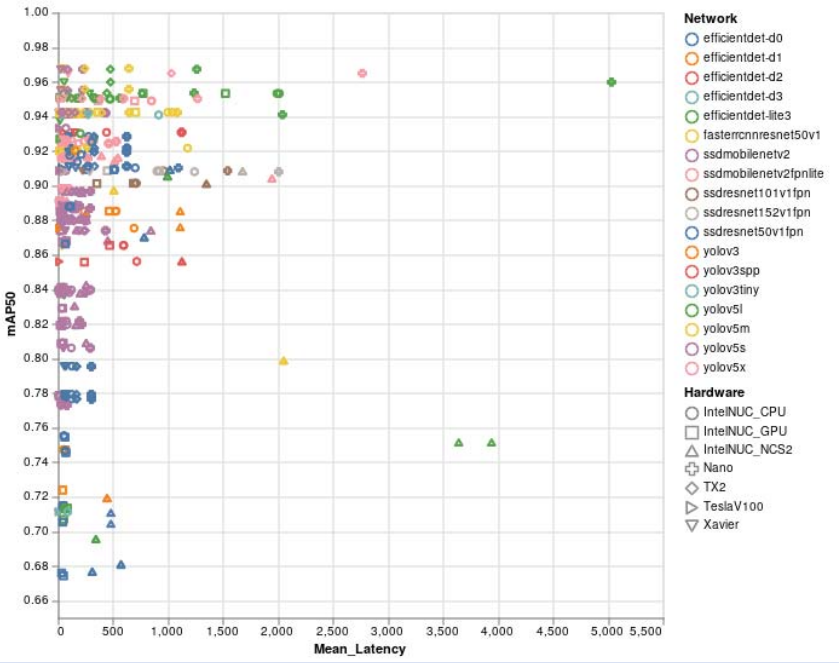
Platforms under Study

Name	Performance [T op/s]	Memory [GB]	Power [W]	Cost [€]
NVIDIA Xavier AGX	32	16	10–30	800
NVIDIA Jetson TX2	1.3	4	7.5–15	260
NVIDIA Jetson Nano	0.5	4	5–10	120
Intel NCS2	1	0.5	5	80
Intel NUC CPU (i7-8650U)	22.4	32	15	600
Intel NUC GPU (Intel UHD 620)	0.8	32	15	600
Tesla V100	130	32	250	>1000

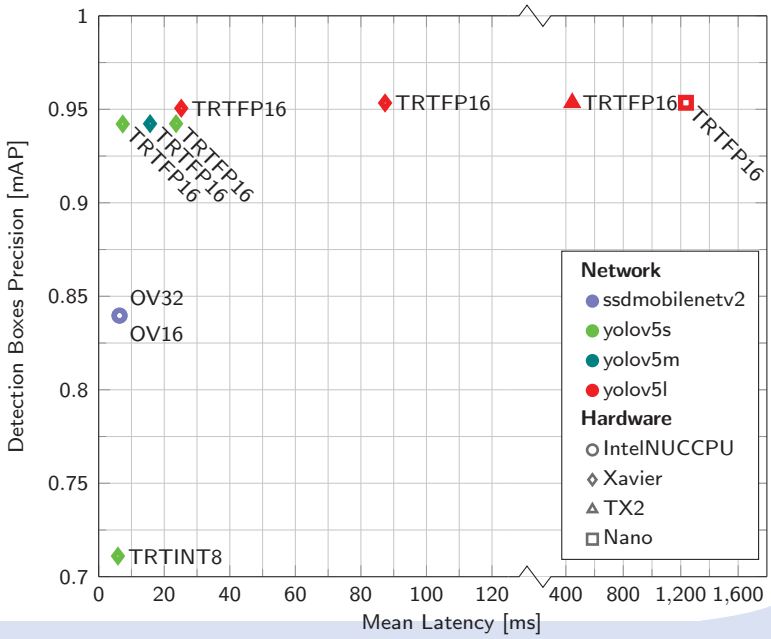
Networks under Study

Name	Framework used	No of parameters (10^6)
ssdmobilenetv2fpnlite	Tensorflow	2.8
efficientdet-d0	Tensorflow	3.9
ssdmobilenetv2	Tensorflow	4.5
yolov5s	Pytorch	7.0
yolov3tiny	Pytorch	8.6
yolov5m	Pytorch	21.0
yolov5l	Pytorch	46.6
ssdresnet50v1fpn	Tensorflow	50.7
yolov3	Pytorch	61.4
yolov3spp	Pytorch	62.5
ssdresnet101v1fpn	Tensorflow	69.7
ssdresnet152v1fpn	Tensorflow	85.3
yolov5x	Pytorch	87.1

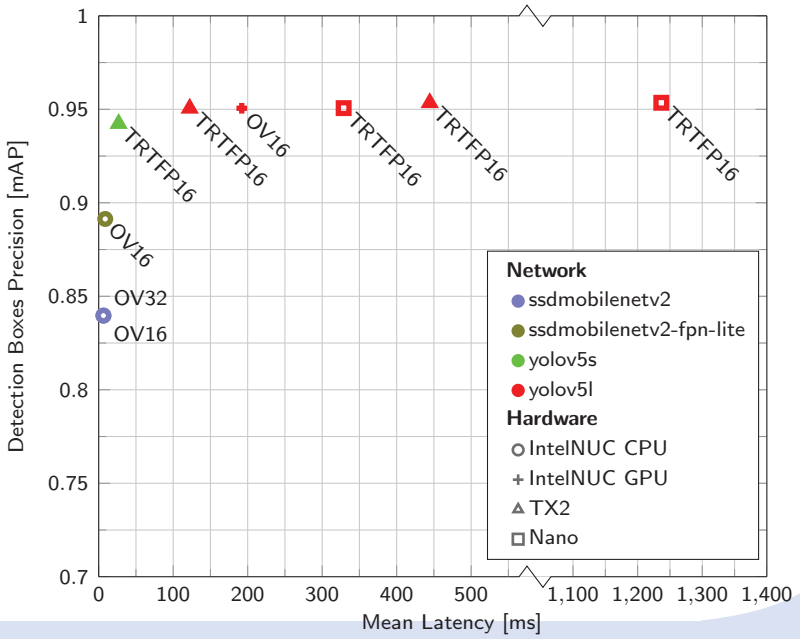
All solutions



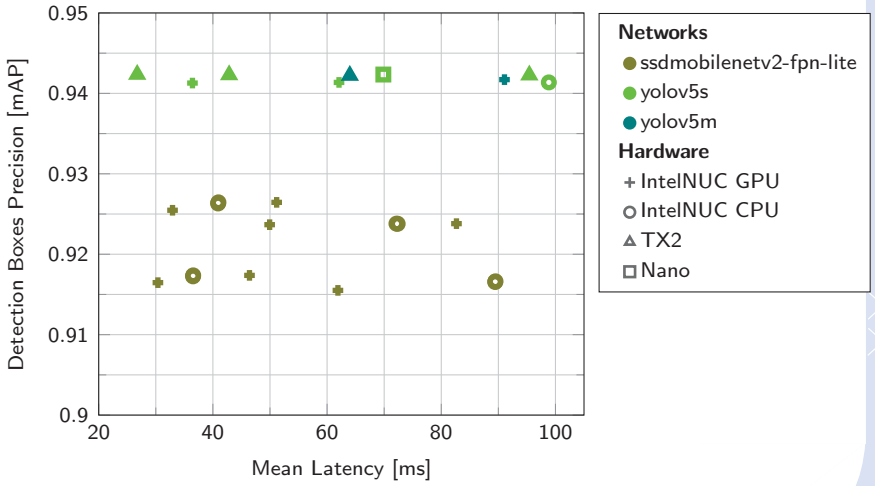
Pareto-optimal solutions



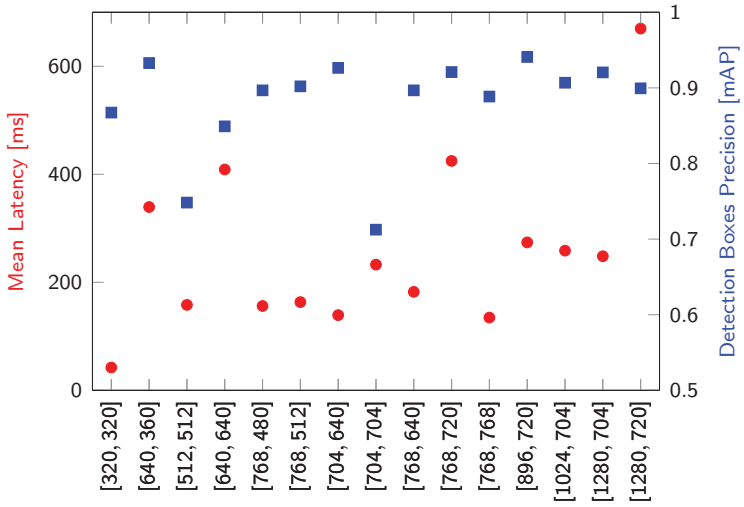
Pareto-optimal solutions under cost constraints



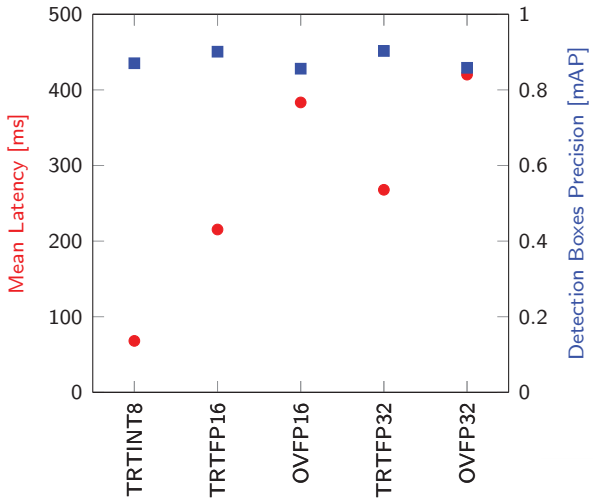
Solutions under cost constraints, latency ≤ 100 ms and mAP50 ≥ 0.9.



Impact of image resolution



Impact of image Quantization



Traffic Light Case Study - Conclusions

- Yolo v5s is the most suitable network; Yolo v5m and MobileNetV2 are also reasonable;
- Four platforms are reasonable choices: IntelNUC GPU, IntelNUC CPU, Jetson Nano, TX2.
- For very low latency under 50 ms TX2 and IntelNUC GPU are preferable platforms with an image resolution of 640×360 pixels.
  If high accuracy is prioritized, TX2 is the winner in this group, delivering 0.943 mAP with 640×360 image resolution and TRTFP16 quantization.

Results, publications, demos, code on

eml.ict.tuwien.ac.at

CD-Lab for Embedded Machine Learning

Duration	7 years, Oct 2019 - Sept 2026
Partner	TU Wien, TU Graz, AVL, Mission Embedded, Siemens
3 WPs	WP1 Embedded Platforms (TUW, Mission Embedded) WP2 DNN Architecture and Optimization (TUW, Siemens) WP3 Continuous Learning (TUG, AVL)
Budget	2.8 M€, 400 k€/year
People	Funded: 2 Postdocs, 5 PhD Students, 3 MSc Students Total: 2 Postdocs, 5 PhD Students, 14 MSc+BSc Students

www.ict.tuwien.ac.at

45

¿ Questions ?

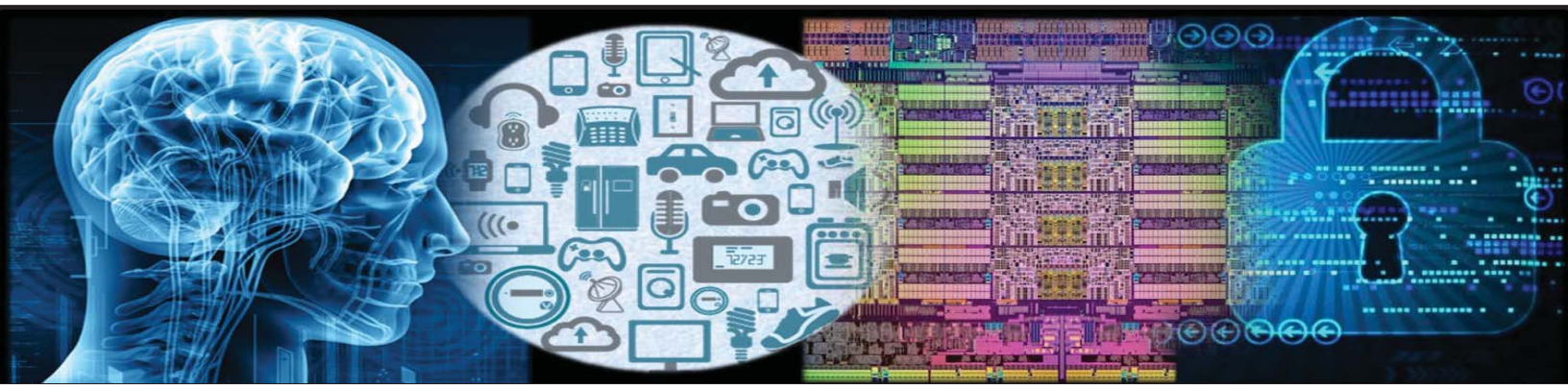
References I

- M. Wess, M. Ivanov, C. Unger, A. Nookala, A. Wendt, and A. Jantsch. "ANNETTE: Accurate Neural Network Execution Time Estimation With Stacked Models". In: *IEEE Access* 9 (2021), pages 3545–3556.
- Martin Lechner and Axel Jantsch. "Blackthorn: Latency Estimation Framework for CNNs on Embedded Nvidia Platforms". In: *IEEE Access* (2021).
- SIA - SRC. *Rebooting the IT Revolution: A Call to Action*. Technical report. Semiconductor Industry Association and Semiconductor Research Corporation, Sept. 2015.
- Emma Strubell, Ananya Ganesh, and Andrew McCallum. "Energy and Policy Considerations for Deep Learning in NLP". In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Florence, Italy: Association for Computational Linguistics, July 2019, pages 3645–3650.

Embedded Machine Learning for the Edge:

From Algorithms to Architectures

M. Shafique (Director, eBrain Lab)
New York University (NYU) Abu Dhabi, UAE
NYU Tandon School of Engineering, USA



Who Ruled the World!

Age of Power

Man-Power (#), Skills, Strength, Courage, etc.

Age of Resources and Industry

Fuel, Industrial Tech., Economic Politics, etc.

Age of Data and AI

Data is the New Fuel

Innovation in Technology is the New Politics

Nation-wide Race for Dominance in AI

3

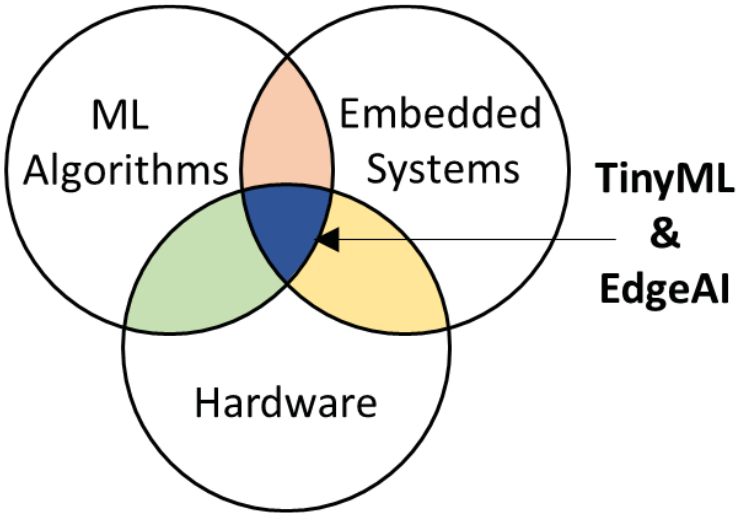
Outline

- What is TinyML / EdgeAI?
- Applications
- Cross-Layer Design Flow
- Future Research Directions

TinyML and EdgeAI: Unique Features?

Performing on-device data analytics at extremely low power

- Fastest-growing field of machine learning
- Combination of embedded systems, algorithms and hardware
- On-device machine learning
- Always-on use-cases
- Battery-operated devices
- Scalable to trillions of sensors



5

TinyML and EdgeAI

- Fundamentally different from machine learning in the cloud

	Cloud AI	→ Edge/Mobile AI	→ TinyML
Hardware	NVIDIA DGX A100	Samsung S20, NVIDIA Jetson	STM32F769 Microcontroller
Memory	1 TB System Memory + 320 GB GPU Memory	2 - 12 GB	~512 KB
Storage	>15 TB	16 - 512 GB	~2 MB
Applications	Model Training, Big Data Analytics	Data Processing, Continual Learning	In-/near-sensor processing
		Tight Constraints	**Extreme Constraints**

Smart Cyber Physical Systems & Internet-of-Things

Smart Automobiles
http://www.it5g.com/latest-software-

AI / ML is inevitable, we have to efficiently infer knowledge from the big data, and *derive predictions*

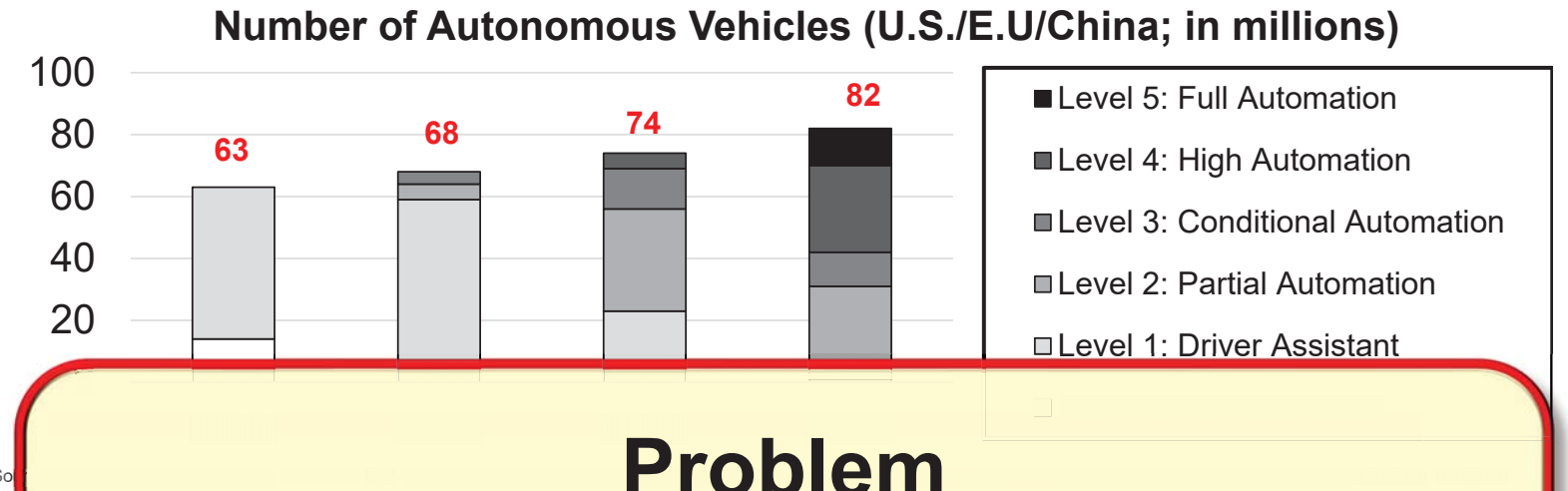
Industry 4.0:
Smart Industrial Automation
https://vimeo.com/145877805

Smart Houses
https://www.linkedin.com/pulse/smart-homes-private-secure-future-intelligent-home-tripti-jha

Smart Grids
http://solutions.3m.com/wps/portal/3M/en_EU/SmartGrid/EU-Smart-Grid/

7

Autonomous Cars: The Big Data Processing Challenge!



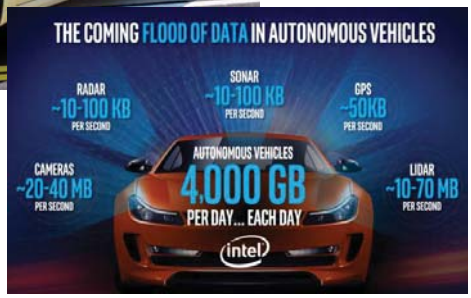
So

Problem

AI on Big Data@Edge => Complexity²

- ❑ Radar: ~10-100KB/sec
- ❑ Sonar: ~10-100KB/sec
- ❑ Camera: ~20-40MB/sec
- ❑ GPS: ~50KB/sec

4000 GB per day



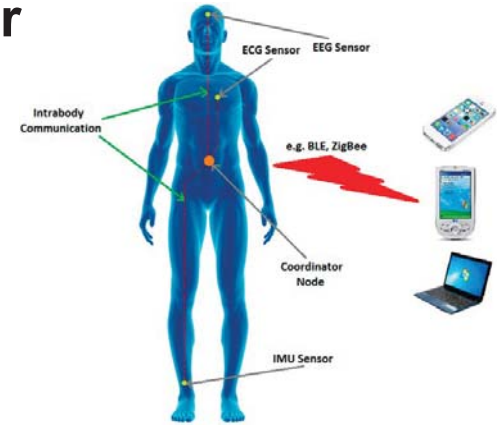
Sources:
https://www.networkworld.com/article/3147892/one-autonomous-car-will-use-4000-gb-of-dataday.html

8

Smart CPS & IoT => The Robustness Challenge!

… should consider

- ❑ **Robustness**
  - ❑ Reliability
  - ❑ Security
- ❑ **Performance**
  - ❑ Throughput
  - ❑ Latency
- ❑ **Others**
  - ❑ Adaptability
  - ❑ Safety
  - ❑ Privacy
  - ❑ Interoperability



Smart Healthcare
(Energy and time constraints)

Norwegian C-130 crash (2012)
https://en.wikipedia.org/wiki/2012_Norwegian_C-130_crash

Failure of F-22 Raptor (2007)
http://www.dailytech.com/Lockheeds+F22+Raptor+Gets+Zapped+by+International+Date+Line/article6225.htm

Satellite imagery of the Northeastern United-States taken before and during the blackout

Toronto, on the evening of August 14, 2003

Northeast blackout of 2003
https://en.wikipedia.org/wiki/Northeast_blackout_of_2003

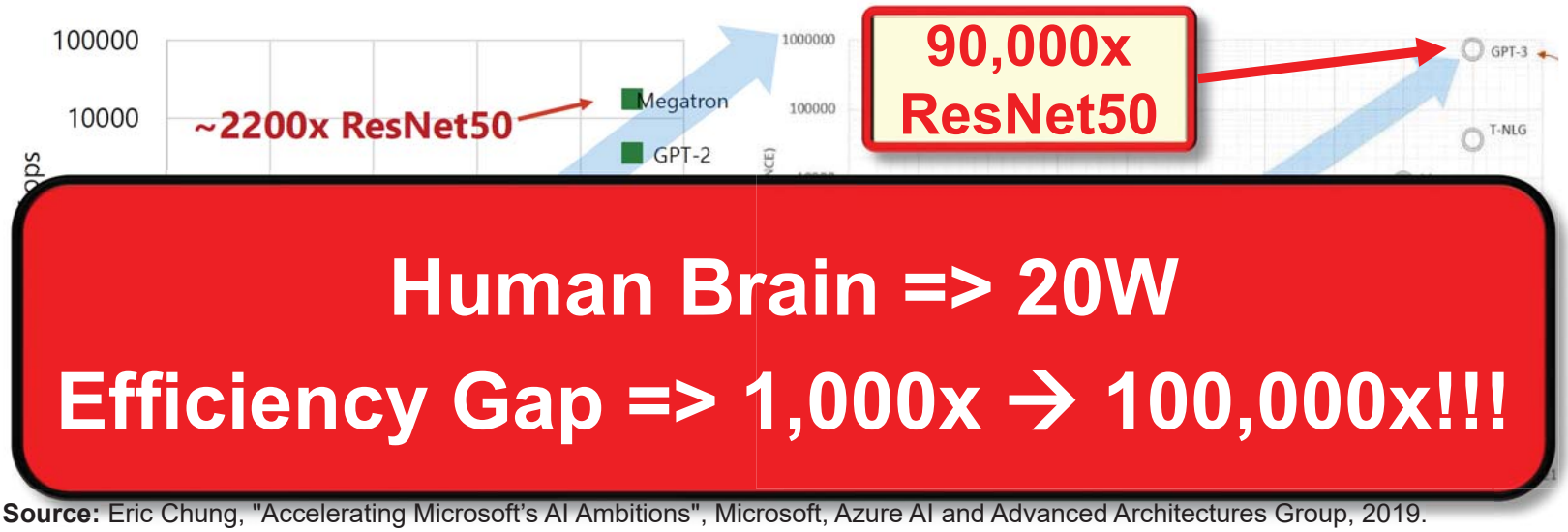
Hacking Jeep Cherokee 4x4 (2015)

Sent the instructions through Entertainment systems
- Change the in-car temperature
- Control the steering
- Control the braking system

https://www.ophtek.com/4-real-life-examples-iot-hacked/
https://www.wired.com/2015/07/hackers-remotely-kill-jeep-highway/

9

Complexity: Exponential Growth in Model Sizes!



Source: Eric Chung, "Accelerating Microsoft's AI Ambitions", Microsoft, Azure AI and Advanced Architectures Group, 2019.
Source: https://www.microsoft.com/en-us/research/blog/a-microsoft-custom-data-type-for-efficient-inference/.

Challenging Question

How to process huge amount of data in robust & energy-efficient way, while considering tinyML / EdgeAI constraints?

10

Robustness for Machine Learning: News Feed

Beware: Galaxy S10's Facial Recognition Easily Fooled with a Photo

Jesus Diaz · Freelance Writer
Updated Mar 11, 2019

Self-driving Uber kills Arizona woman in first fatal crash involving pedestrian

Tempe police said car was in autonomous mode at the time of the crash and that the vehicle hit a woman who later died at a hospital

Hackers trick a Tesla into veering into the wrong lane

https://www.youtube.com/watch?v=a7L51u23YoM

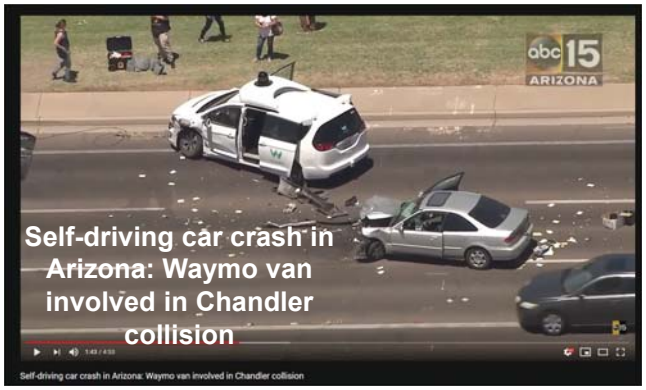
Tesla Model 3: Autopilot engaged during fatal crash

BBC

17 May 2019

Tesla driver dies in first fatal crash while using autopilot mode

The autopilot sensors on the Model S failed to distinguish a white tractor-trailer crossing the highway against a bright sky



https://www.technologyreview.com/f/613254/hackers-trick-teslas-autopilot-into-veering-towards-oncoming-traffic/

13

Adversarial Attacks on Tesla Autopilot by Tencent Keen Security Lab

Digital Adversarial Examples

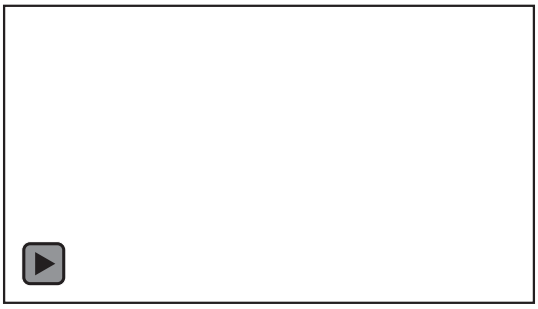
- Insert the noise into the DNN input

Rainy Score: 0.0113

Adversarial Noise

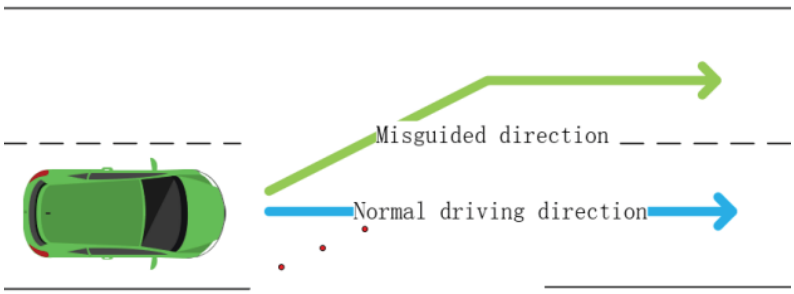
Rainy score: 0.8204

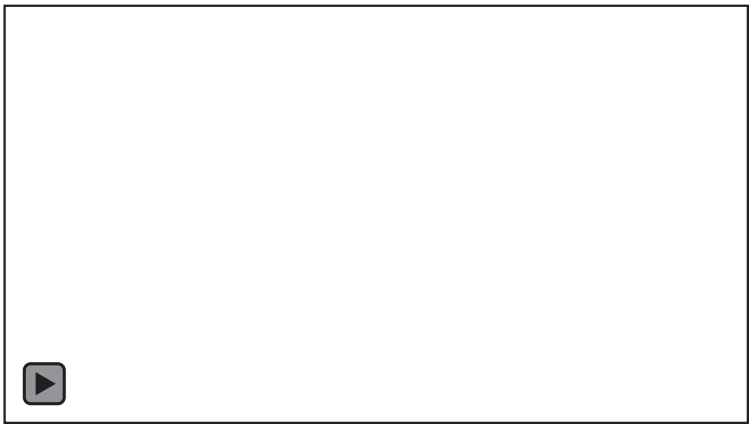
Black-Box Attack



Physical World Adversarial Examples

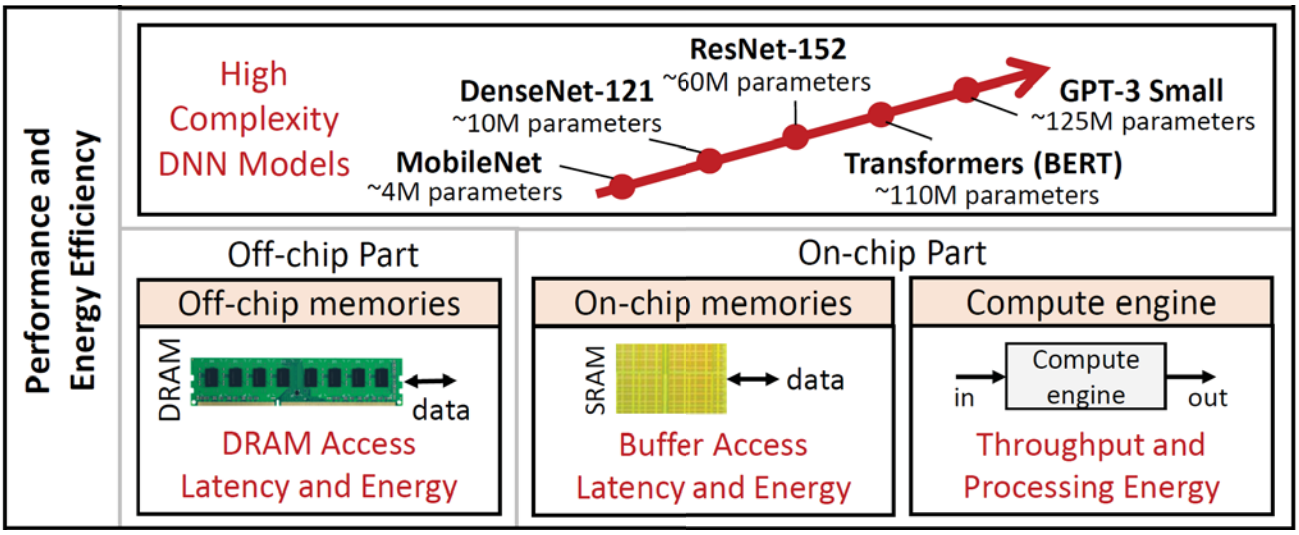
- Place the small stickers on the ground





Tencent Keen Security Lab, "Experimental Security Research of Tesla Autopilot" Technical Report 2019-03

Overview of Challenges for EdgeAI & tinyML



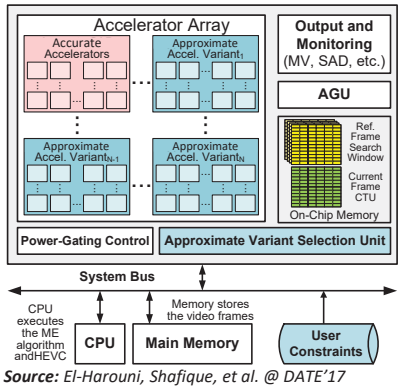
15

Cross-Layer Design Flow

- Frameworks enable seamless integration of algorithms and optimizations at all layers, developed by the community.
  - Design and optimize ML models for ultra-low power devices

- Hardware accelerators
  - Specialized hardware for accelerating vector/matrix multiplication
- DNN Optimization
  - Neural Architecture Search (NAS), Pruning and Quantization

Embedded AI @ eBrain Lab:
A Multi-Dimensional Research Challenge

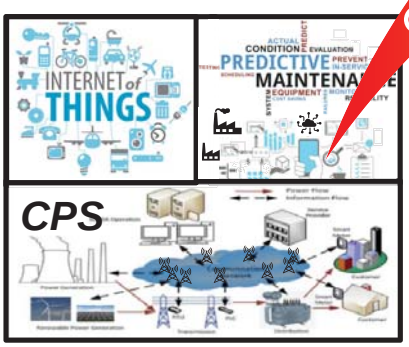


Source: *El-Harouni, Shafique, et al. @ DATE'17*

Source: *NVIDIA Jetson Nano*

Source: *Intel, Loihi Chip*

Source: *ARDUINO*



Software (GPUs)

Accelerators + Approximate Computing

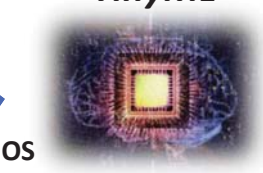
Deep Learning Architectures

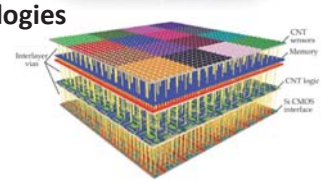
Hardware-Aware Neural Architecture Search (NAS) + Optimization

Neuromorphic Architectures

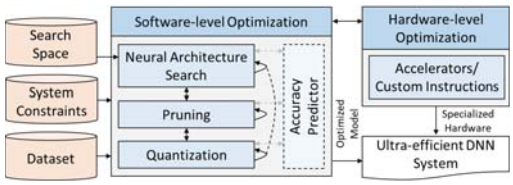
Post-CMOS Technologies

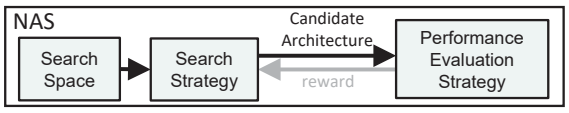
TinyML

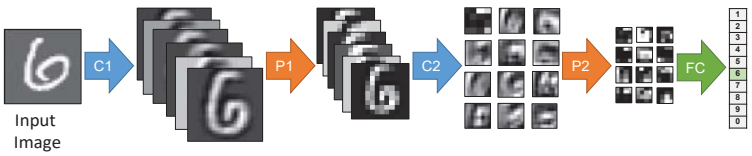


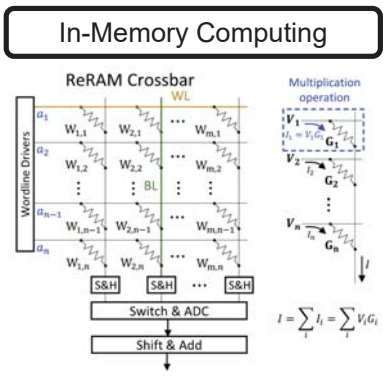


Source: *M. M. Shulaker et al., Nature 547, 74 (2017) and R. Mark Wilson; Physics Today 70, 14-16 (2017)*



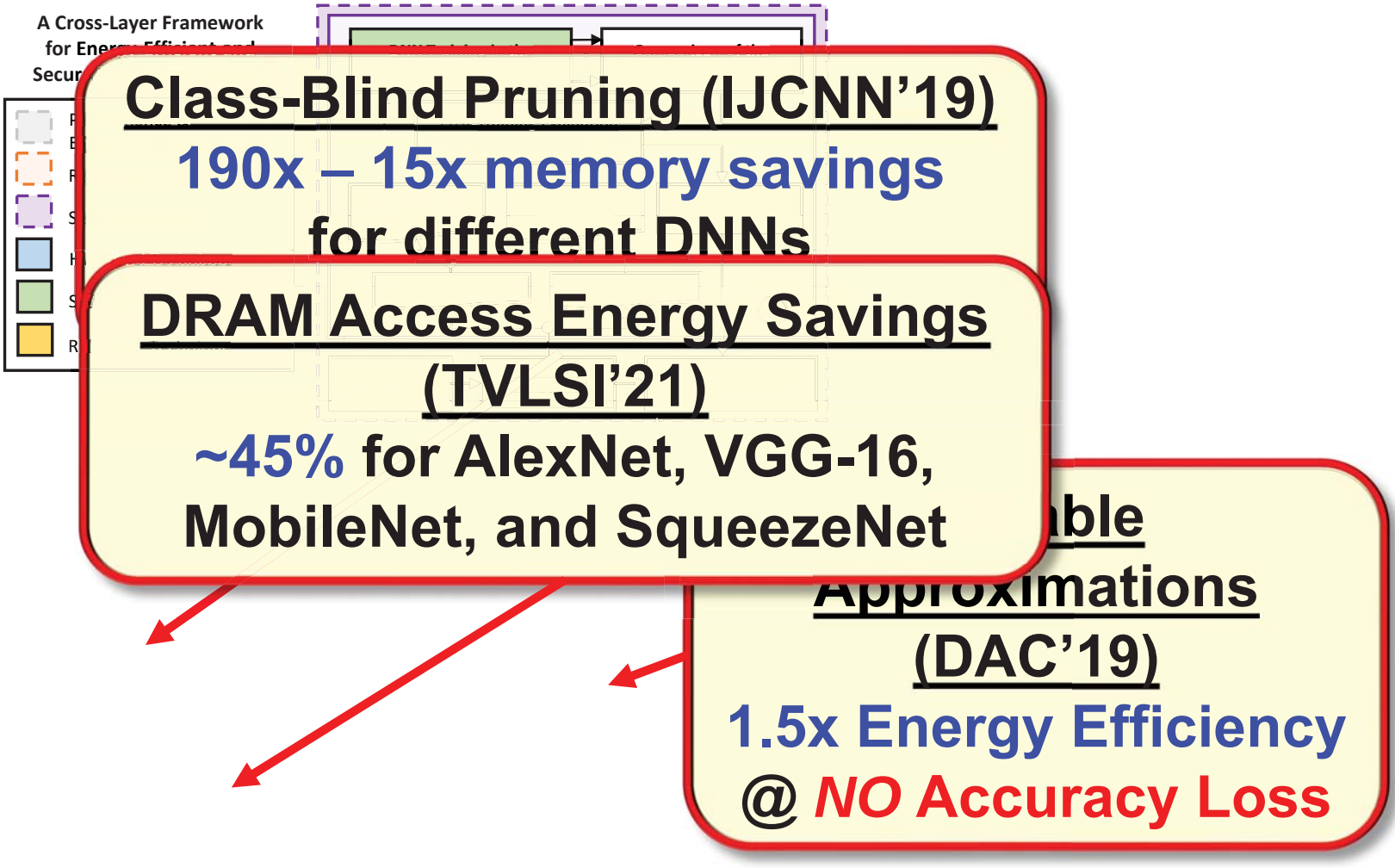






Source: Hanif, Manglik, Shafique @ IEEE Access'20

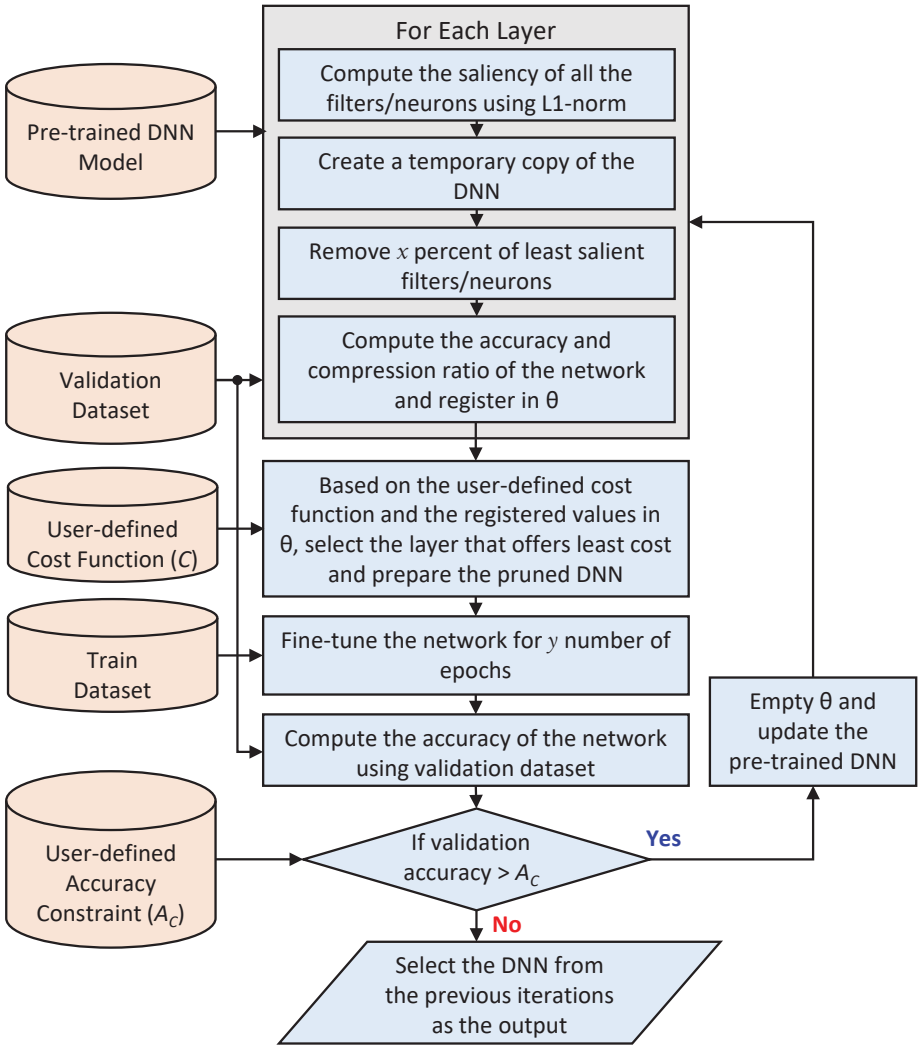
Our Cross-Layer TinyML and Edge AI Framework: An Overview



18

Structured Pruning Methodology

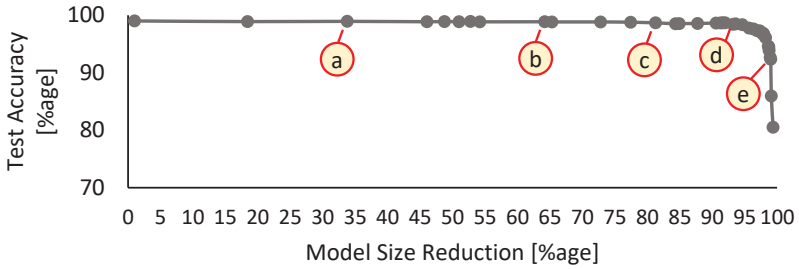
- Step 1: **Compute the sensitivity** of the layers of the given DNN to pruning using a user-defined cost function
- Step 2: **Remove x percent filters/neurons** from the least sensitive layer
- Step 3: **Fine-tune the network** for y number of epochs
- Step 4: **Compare the accuracy** with the defined accuracy constraint
- Step 5: **Continue pruning** if the accuracy is greater than the defined constraint



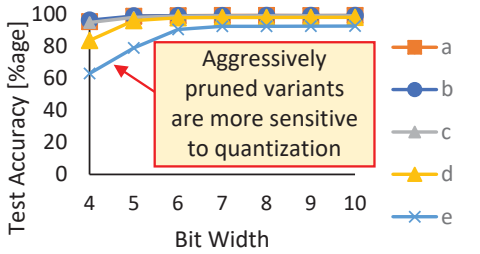
22

Results using LeNet-5 trained with MNIST Dataset

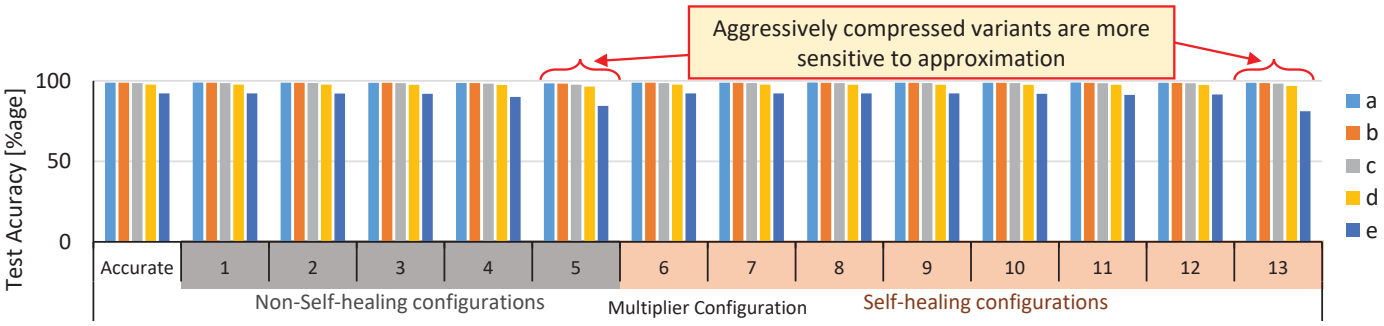
1. Structured Pruning



2. Quantization

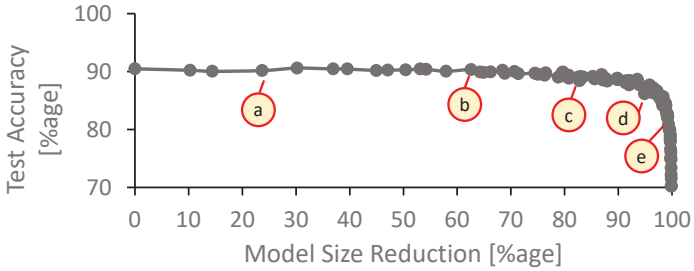


3. Hardware Approximation

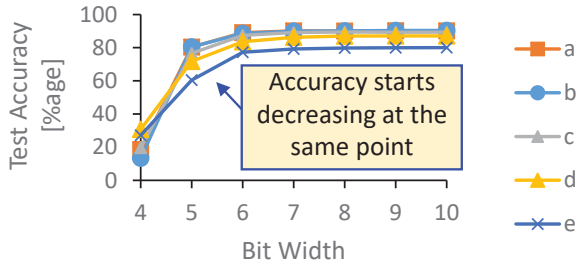


Results using VGG11 trained with Cifar10 Dataset

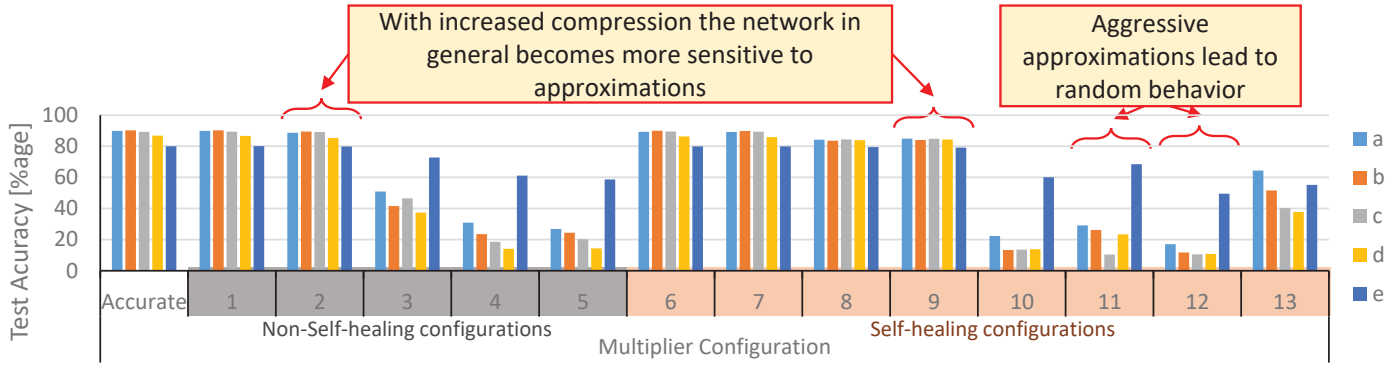
1. Structured Pruning



2. Quantization

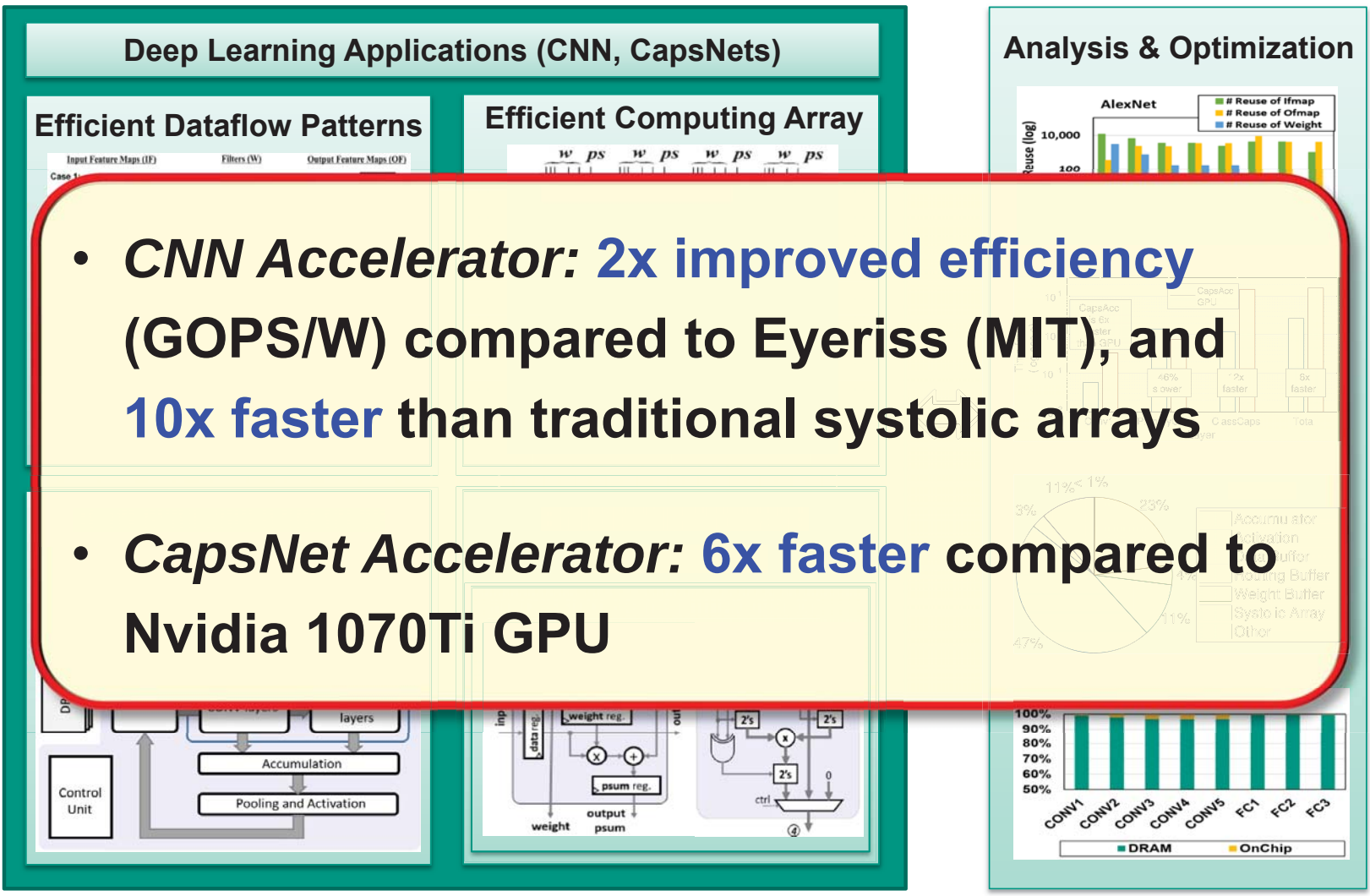


3. Hardware Approximation



25

Energy-Efficient Deep Learning Architectures



- *CNN Accelerator:* **2x improved efficiency (GOPS/W) compared to Eyeriss (MIT), and 10x faster than traditional systolic arrays**
- *CapsNet Accelerator:* **6x faster compared to Nvidia 1070Ti GPU**

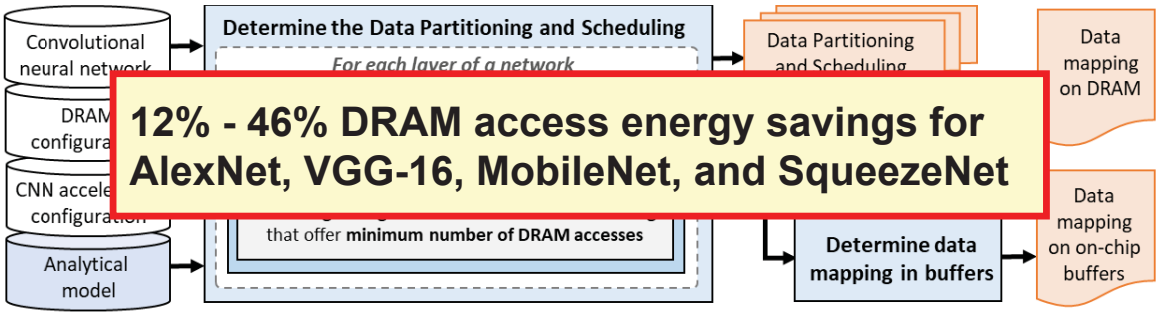
Hanif, Shafique et al. DAC'19, Hanif, et al: MPNA@arXiv'18, Marchisio, Shafique et al., DATE 2019

29

Memory Optimizations

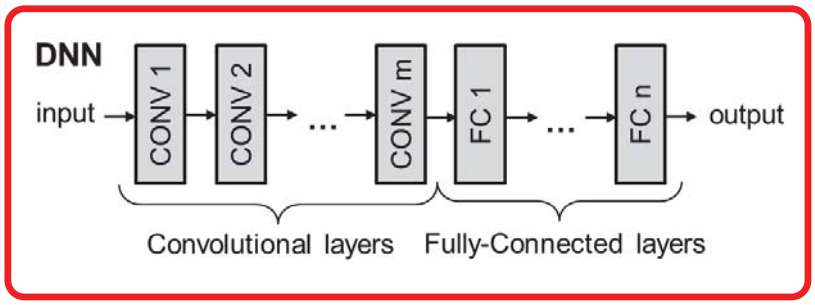
Energy-Efficient Memory Accesses for DNN Accelerators (IEEE TVLSI'21)

1

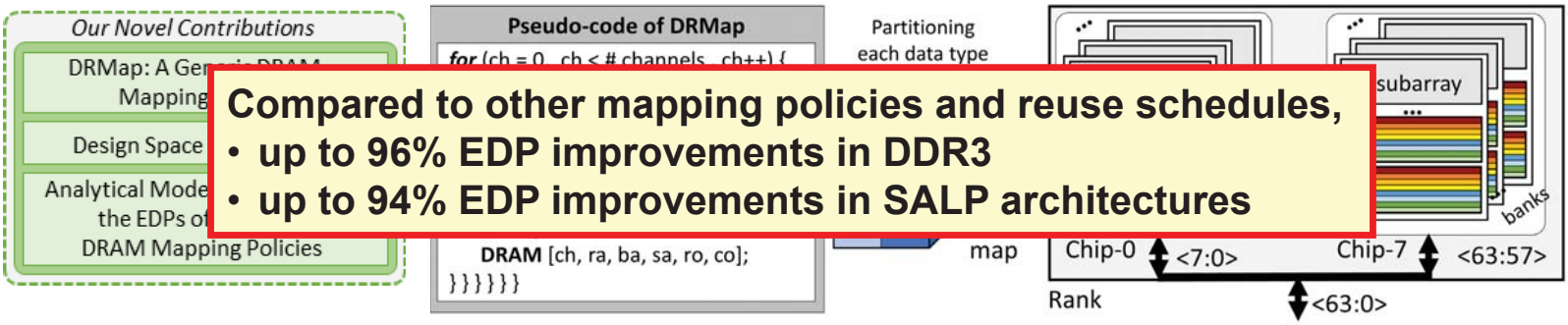


12% - 46% DRAM access energy savings for AlexNet, VGG-16, MobileNet, and SqueezeNet

2



Generic DRAM Mapping for Energy-Efficient DNNs (DAC'20)



Compared to other mapping policies and reuse schedules,
- **up to 96% EDP improvements in DDR3**
- **up to 94% EDP improvements in SALP architectures**

31

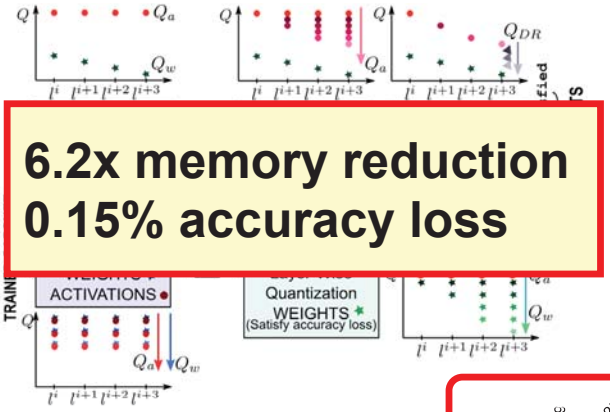
STP vs. Resizing

STP **Resizing**

[Theo and Shafique, et al. @ISVLSI'19]

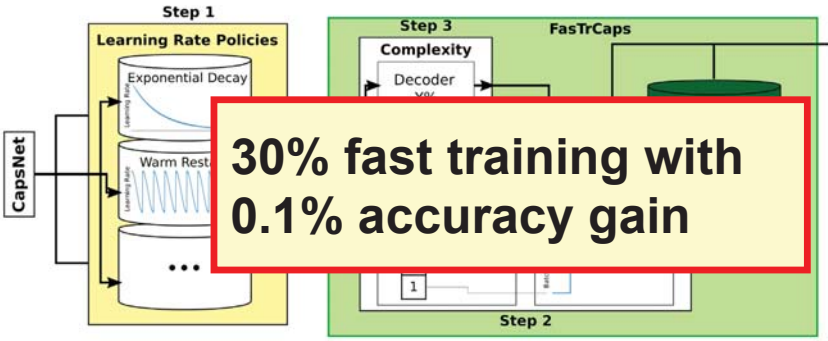
33

Capsule Networks Research



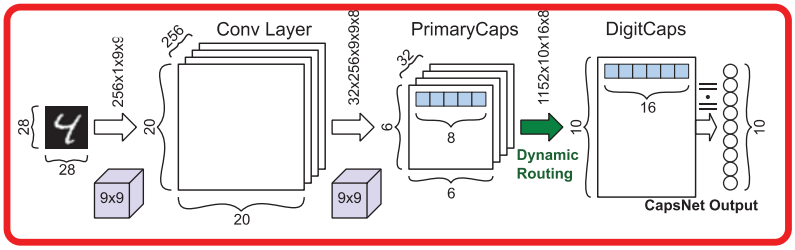
6.2x memory reduction
0.15% accuracy loss

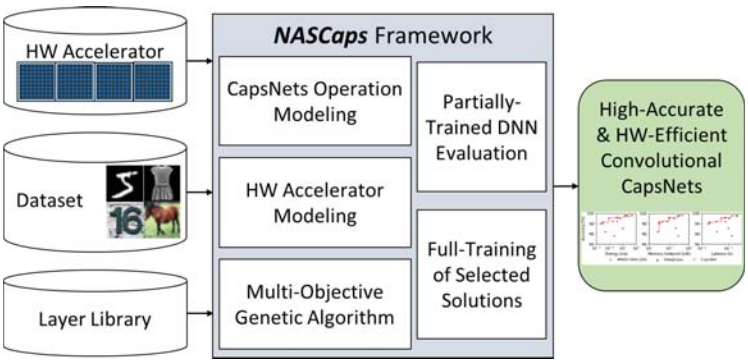
QCaps: Quantization Framework (DAC'20) 1



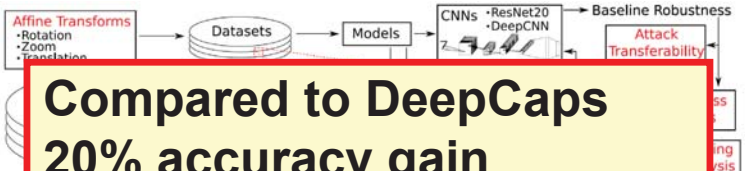
30% fast training with
0.1% accuracy gain

FasTrCaps: Fast Training (IJCNN'20) 2





NASCaps: NAS Framework for CapsNet (ICCAD'20) 3

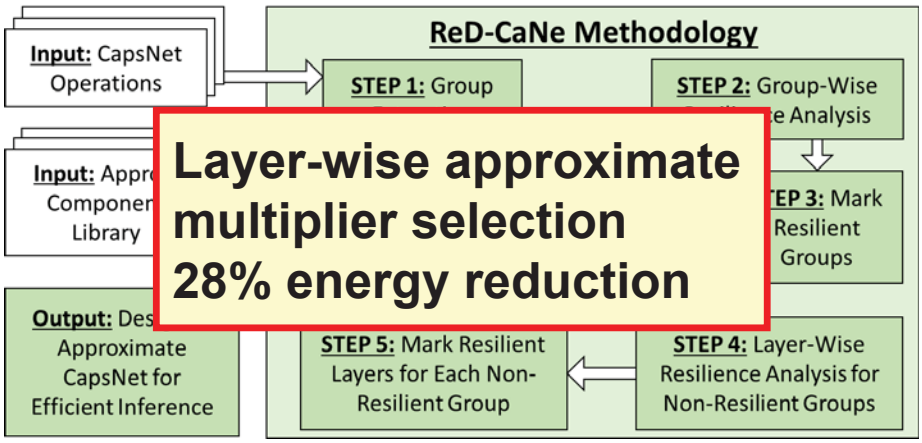


Compared to DeepCaps
20% accuracy gain
52% energy saving
30% reduced memory
64% lower latency

RobCaps: Security & Robustness (under Review) 4

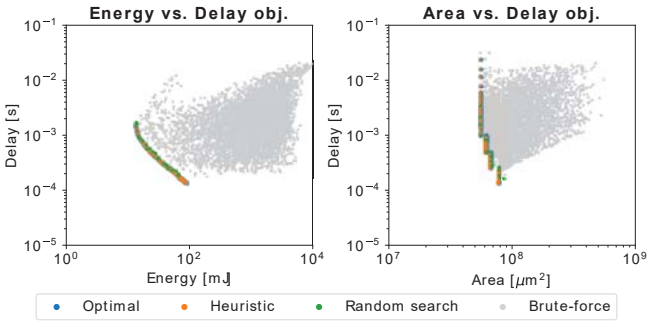
34

Capsule Networks Research

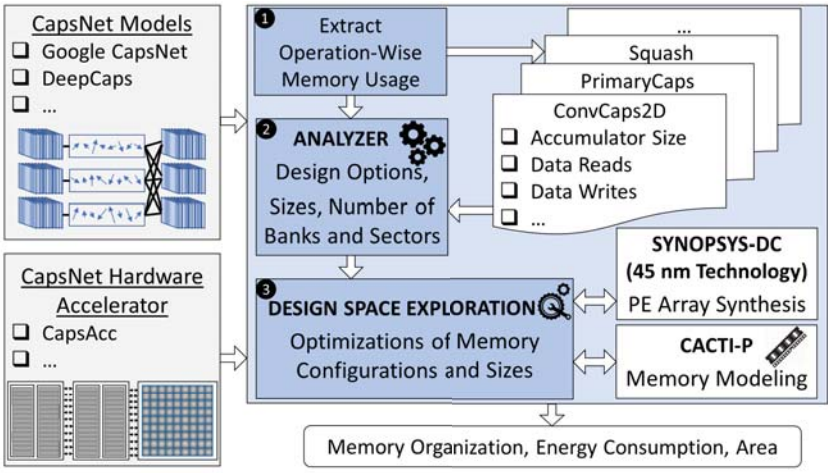


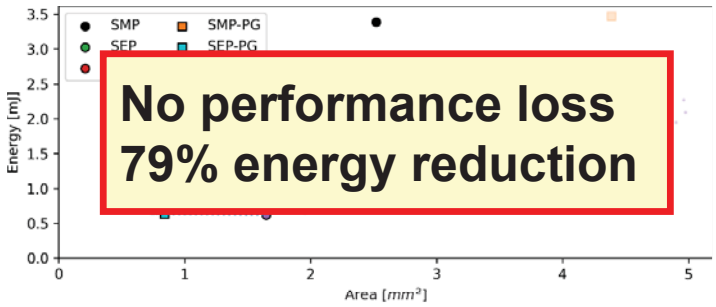
Layer-wise approximate multiplier selection
28% energy reduction

5 **Approximate CapsNet Design (DATE'20)**



6 **DSE of the PE Array for CapsNet Accelerators (IEEE TVLSI'21)**



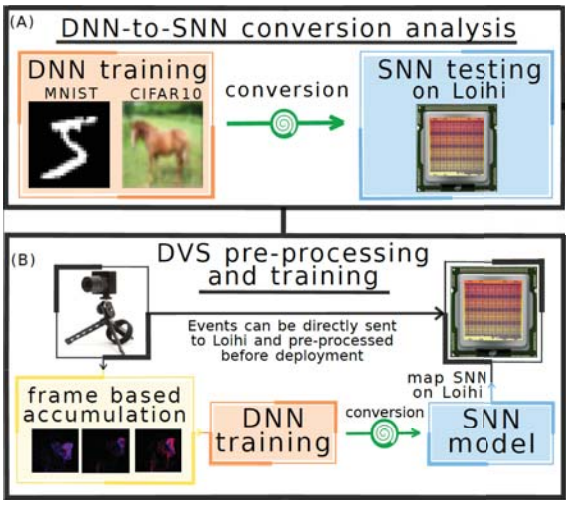


No performance loss
79% energy reduction

DESCNet: Scratchpad Memory Design for CapsNet Hardware (IEEE TCAD'20) 7

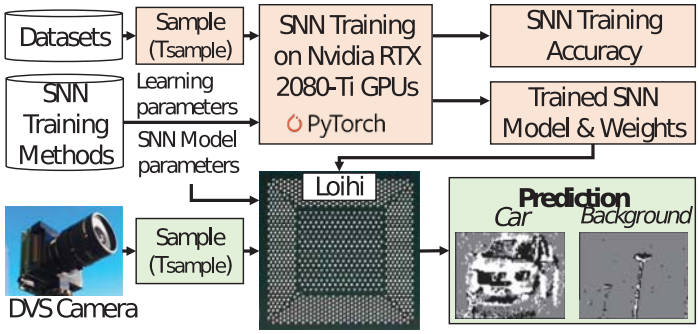
35

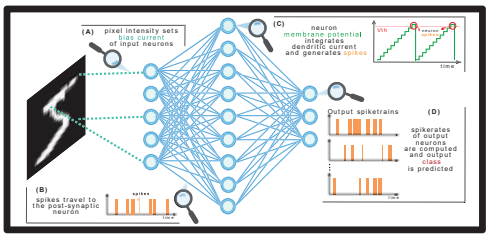
Neuromorphic Computing using Intel's Loihi



SNN Mapping over Intel's Loihi Processor (IJCNN'20)

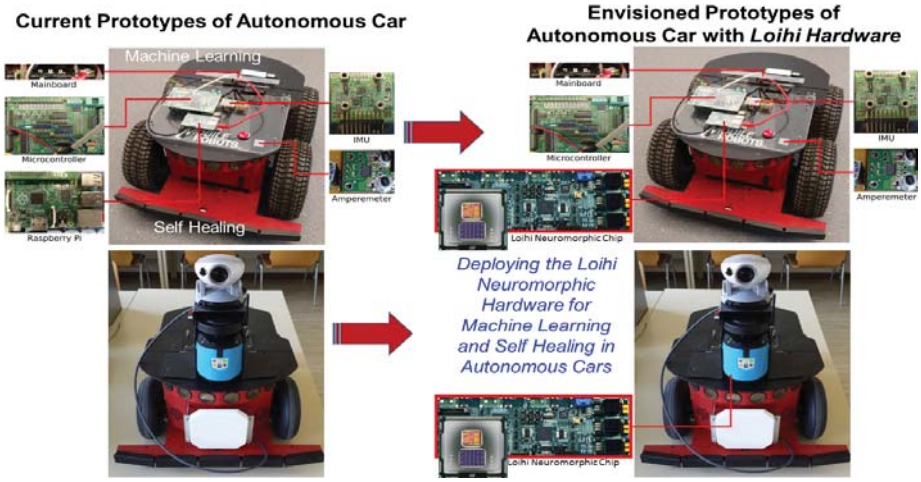
1





2 DVS-Based Car vs. Background Classification on Intel's Loihi (IJCNN'21)

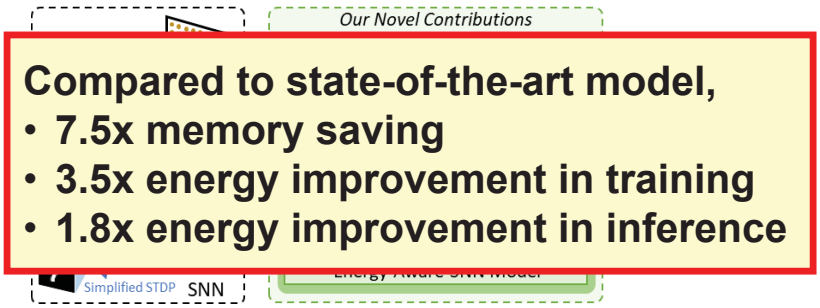
Autonomous Driving



Smart Farming

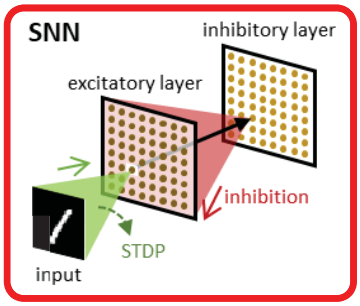
36

Spiking Neural Networks Research



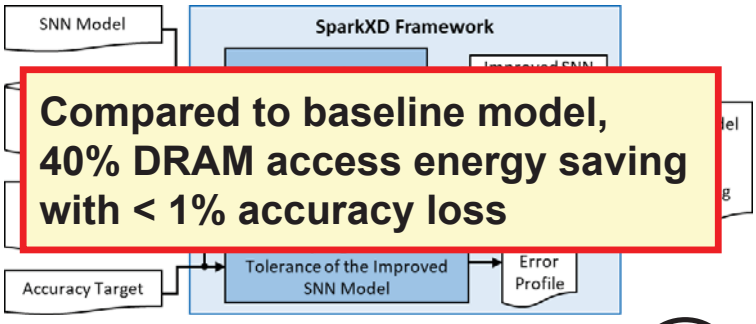
Compared to state-of-the-art model,
- **7.5x memory saving**
- **3.5x energy improvement in training**
- **1.8x energy improvement in inference**

Energy-Aware Optimizations and Learning Methods (IEEE TCAD'20) 1



SNN with Unsupervised Continual Learning (DAC'21) 3

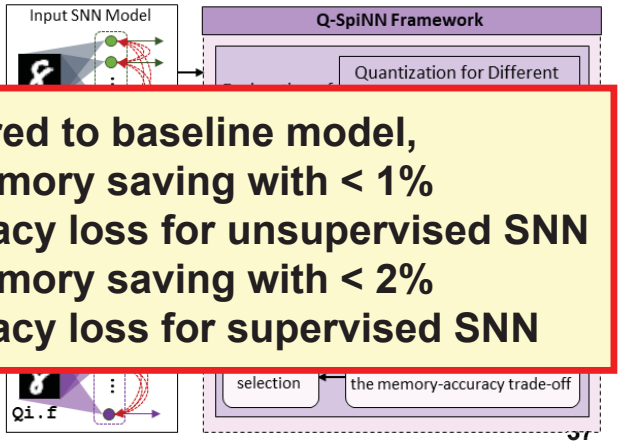
Compared to state-of-the-art model,
- **51% energy saving in training**
- **37% energy saving in inference**
- **21% accuracy gain for the most recently learned task**
- **8% accuracy gain for the previously learned tasks**



Compared to baseline model, 40% DRAM access energy saving with < 1% accuracy loss

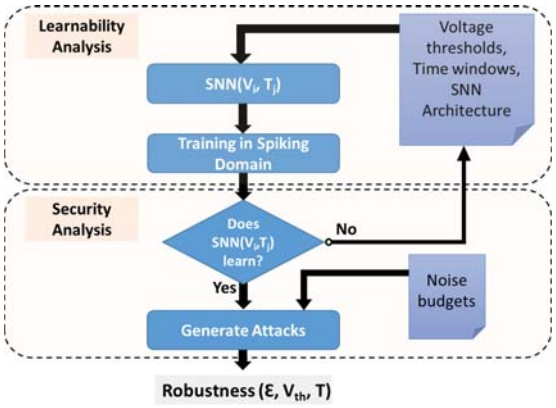
Resilient and Energy-Efficient SNN Inference (DAC'21) 2

Quantization for SNNs (IJCNN'21) 4



Compared to baseline model,
- **4x memory saving with < 1% accuracy loss for unsupervised SNN**
- **2x memory saving with < 2% accuracy loss for supervised SNN**

Security for SNNs & Neuromorphic Computing

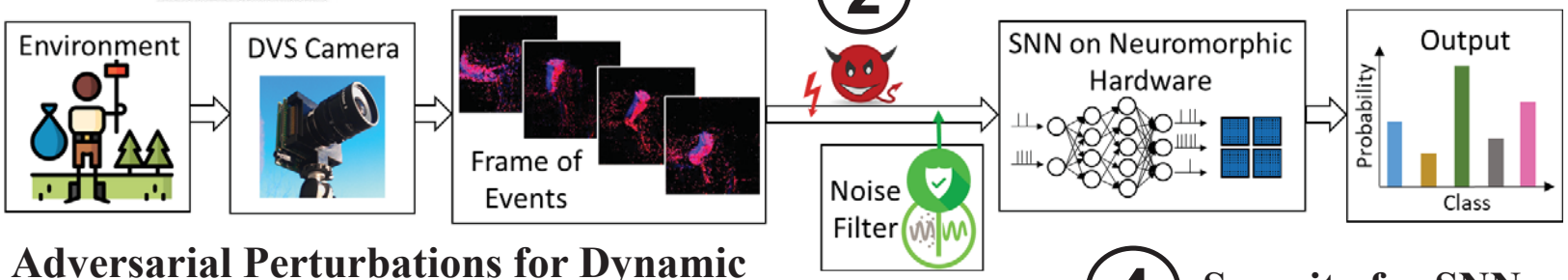


Robust SNN Design against Adversarial Attacks (DATE'21)

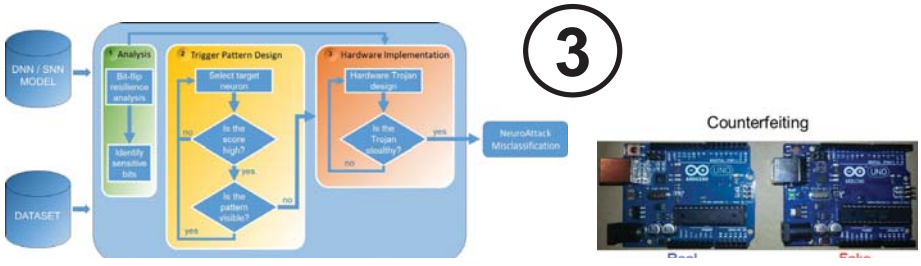
1

Same clean accuracy than CNN

75% higher accuracy for large perturbations



Adversarial Perturbations for Dynamic Vision Sensors (IROS'21, IJCNN'21)

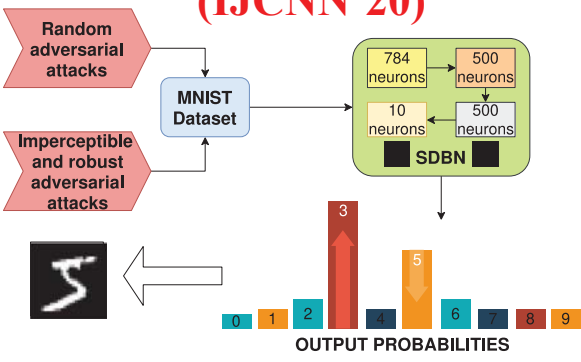


3

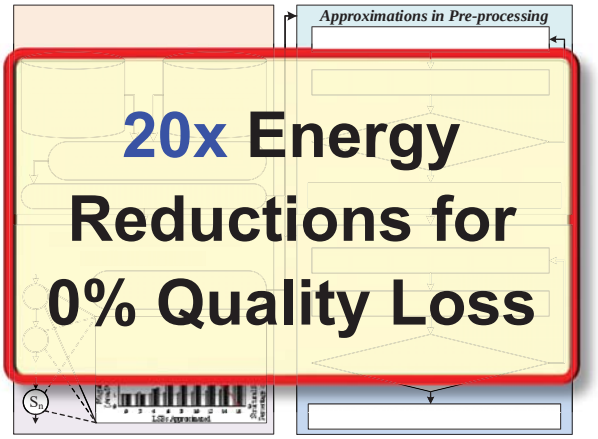
Fault-Injection Attacks (IJCNN'20)

4

Security for SNNs (IJCNN'20)



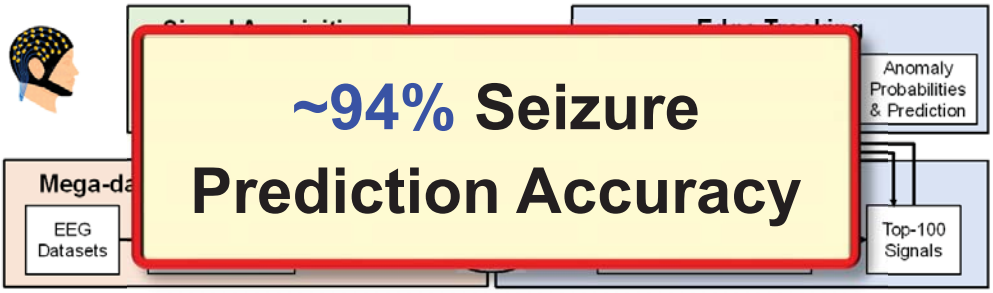
Energy-Efficient IoT-Healthcare and AI



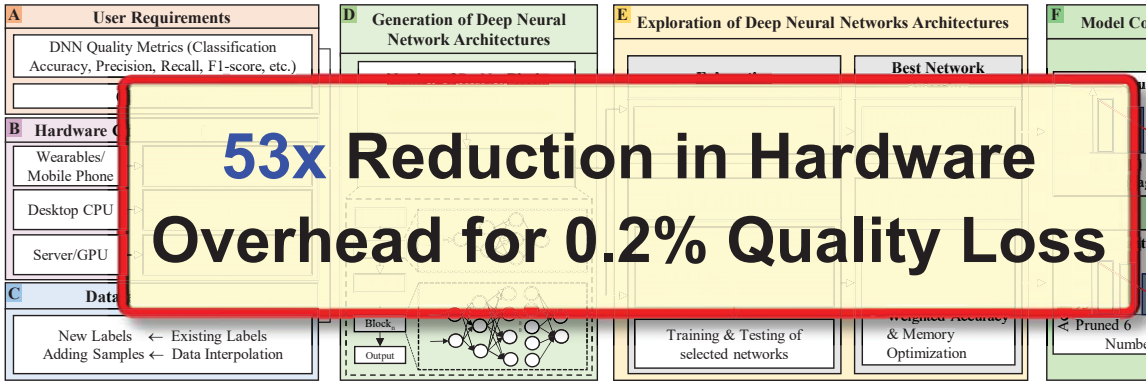
20x Energy Reductions for 0% Quality Loss

Methodology for Approx. Bio-signal Processing: **DAC'19** ①

Cloud-Edge Framework for EEG Monitoring and Real-time Anomaly Prediction: **DAC'20** ②



~94% Seizure Prediction Accuracy



53x Reduction in Hardware Overhead for 0.2% Quality Loss

NAS for HW-Constrained Healthcare DNNs: **(IEEE IoT'21)** ③

39

EdgeAI for Healthcare: Moore4Medical EU Project

Src: Google Images

Next Generation Ultrasound

- ❑ Data Acquisition
- ❑ 3D Reconstruction
- ❑ Edge Processing

- ❑ AI algorithms for detecting fetus' anatomical features
- ❑ Hardware accelerator for high throughput feature extraction
- ❑ Closed-loop system for real-time user feedback

- ❑ Investigating DL architectures and statistical ML techniques for classification, segmentation, and anatomical feature extraction
- ❑ Evaluating requirements of proposed algorithms to develop energy-efficient hardware accelerators for edge processing
- ❑ Develop **FPGA prototype** to demonstrate the efficacy of the accelerator and deployability of the HW-SW system

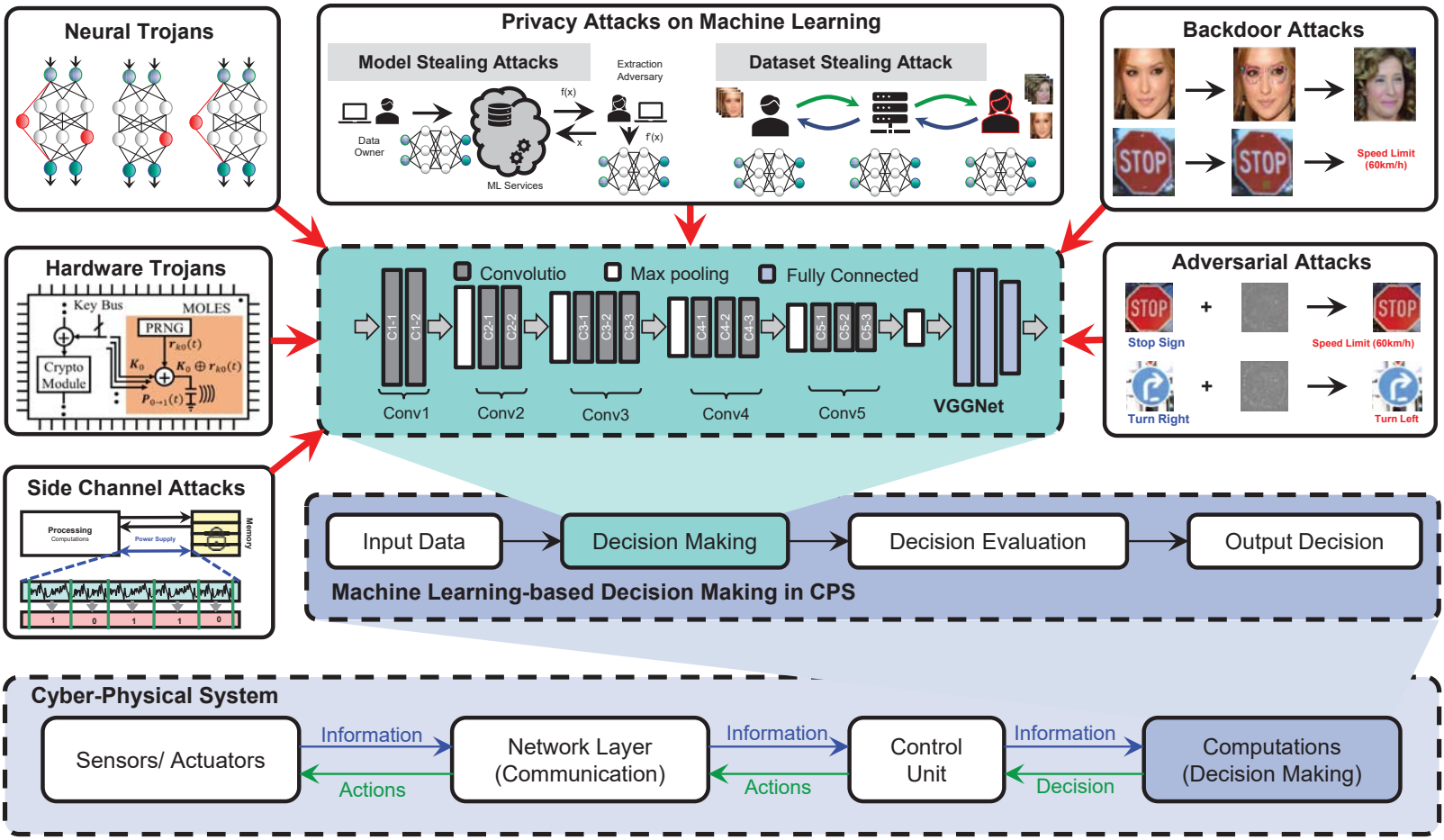
PHILIPS

TU WIEN TECHNISCHE UNIVERSITÄT WIEN

Moore4Medical

40

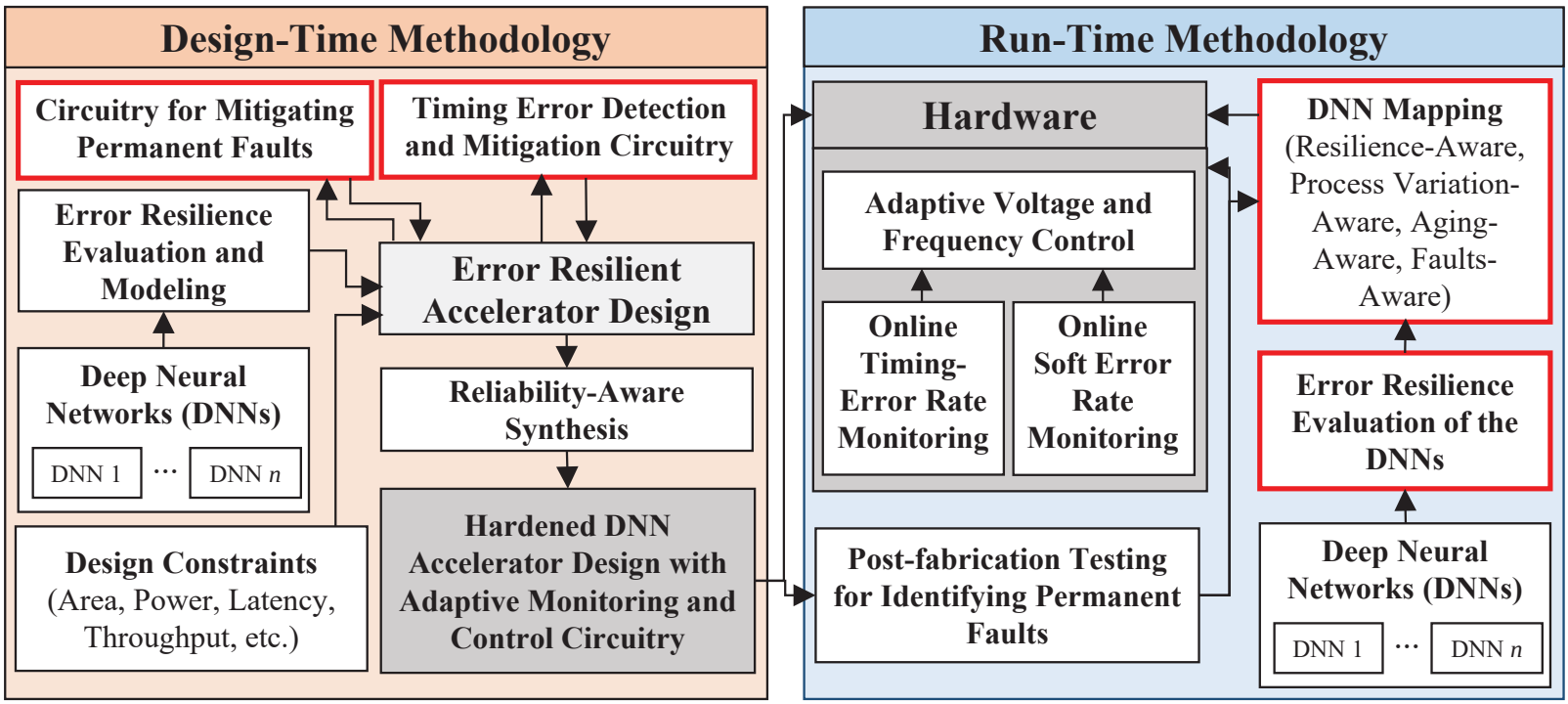
ML Security Research @ eBrain Lab



- M. A. Hanif, F. Khalid, R. V. W. Putra, S. Rehman, M. Shafique, "Robust Machine Learning Systems: Reliability and Security for Deep Neural Networks", in IOLTS-2018, Platja d'Aro, Spain, pp. 257 - 260.
- F. Kriebel, S. Rehman, M. A. Hanif, F. Khalid, M. Shafique, "Robustness for Smart Cyber-Physical Systems and Internet-of-Things: From Adaptive Robustness Methods to Reliability and Security for Machine Learning", ISVLSI-2018, Hong Kong, China, pp. 581-586.

46

ML Dependability Research @ eBrain Lab



47

Future Research Directions

- New computing paradigms such as near-/in-memory computing and approximate computing
- It is not all about deep learning. Conventional machine learning and spiking models may also be better in several scenarios.
- Optimization frameworks for all types of systems, as the selection is limited in some scenarios due to other constraints, e.g., cost.
- Novel techniques for training and optimizing machine learning models
- Interpretability, Explainability, Fairness, Robustness of models
- Formal analysis and verification for safety critical systems

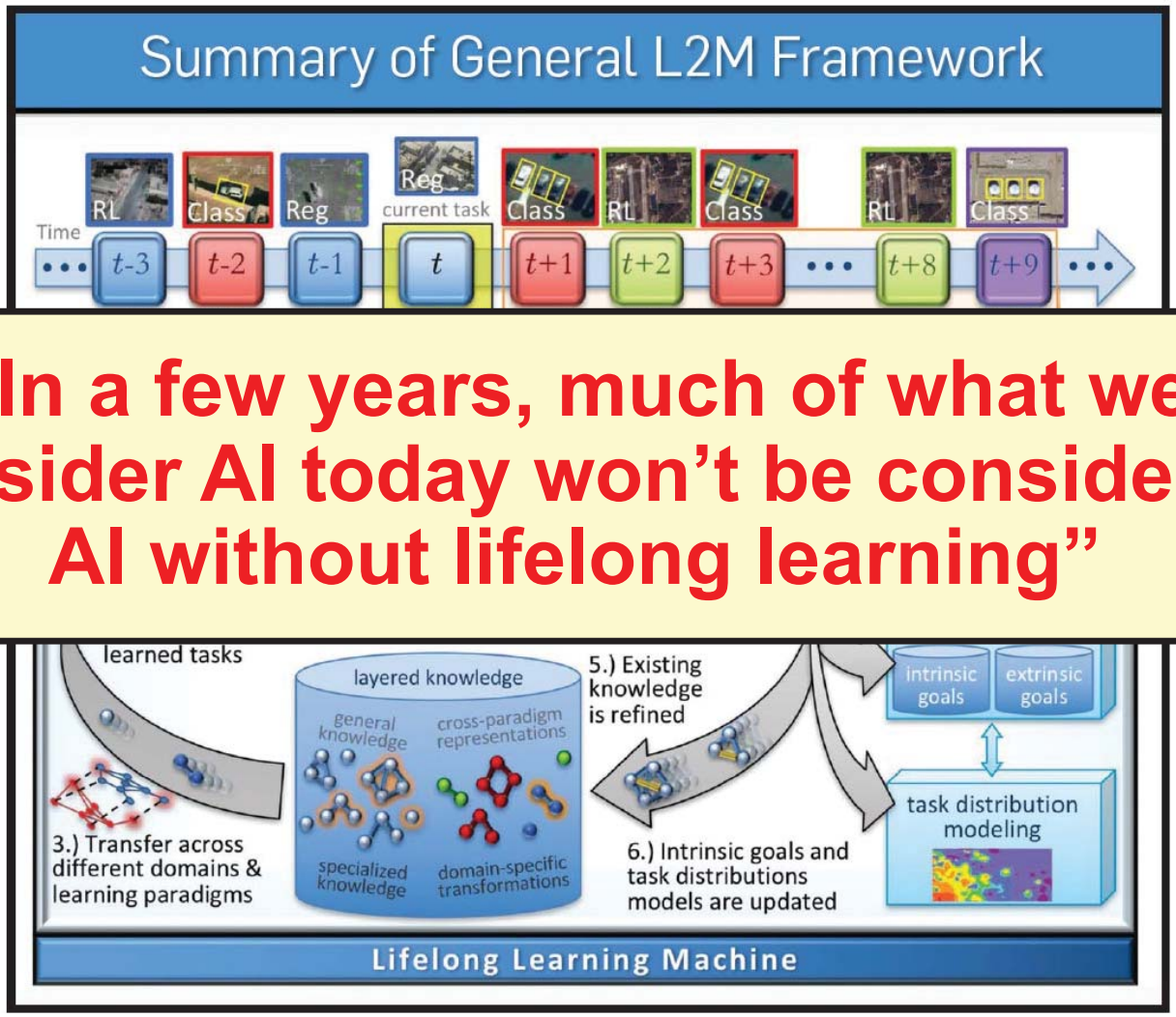
Summary

- **Artificial Intelligence** has proliferated almost everywhere, that's for a good reason! => *the big data challenge!*
  - Cloud, Fog, Edge, …, In-Sensor / In-Situ
- **Required:** High-Throughput, Energy-Efficient, & Robust Designs
- **Our System-Level Framework**
  - Optimizations across the Software & Hardware stacks
  - Specialized hardware accelerators, dataflows, memory, self-healing approximations, hardware-aware NAS, …
  - Selective Tile Processing for energy-efficient object detection
  - **Robustness**
    - Analyzing security attacks and hardware-level faults.
    - New attacks and defense mechanisms for Deep Learning systems

A system level approach requires bridging the gap between the AI/ML community & System designers (HW + SW)

50

Lifelong Learning in Artificial Neural Networks



"In a few years, much of what we consider AI today won't be considered AI without lifelong learning"

Data and image source: "Lifelong Learning in Artificial Neural Networks" in Communications of the ACM

51

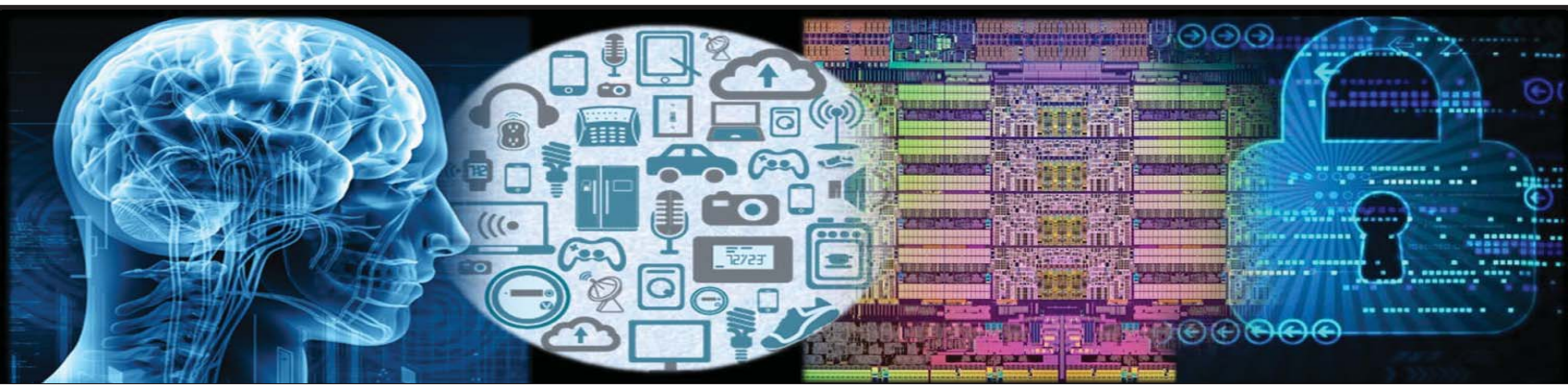
Thank You!

Questions?

M. Shafique

Director, eBrain Lab

muhammad.shafique@nyu.edu



Microelectronics Engineering Group
University of Cantabria

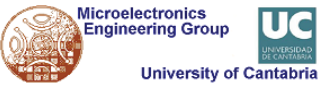
COMP4DRONES

GOBIERNO DE ESPAÑA MINISTERIO DE CIENCIA E INNOVACIÓN

Modeling, Design and Implementation of drone-based Services

Eugenio Villar
University of Cantabria

SS-CPSIoT 7-11 June 2022 Budva , Montenegro



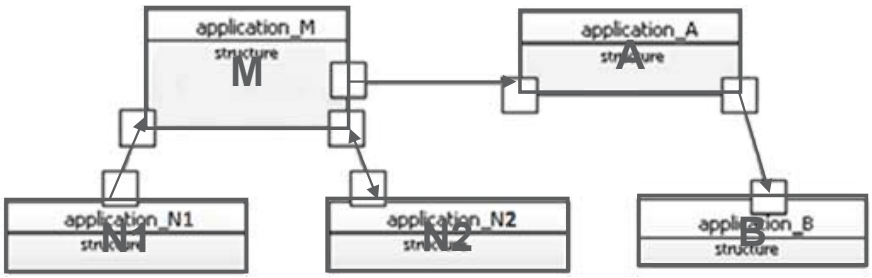
Agenda

- Introduction
- Model-Driven Design of CPSoS
  - Drone-based Services
- Design Verification and Performance Analysis
- Experimental Results
- Conclusions
- Demos

- Slides can be found at:
  - https://www.slideshare.net/EugenioVillar/

Introduction

- Model-Driven Design (MDD)
  - High-abstraction level
  - Mature SW engineering methodology
- UML
  - Standard, any (user-defined) MoC, any language
  - Natural way to capture system architecture
  - Semantic lacks
  - Domain-specific profiles
  - MetaMorph
    - OpenSource, any (user-defined) MoC, language agnostic

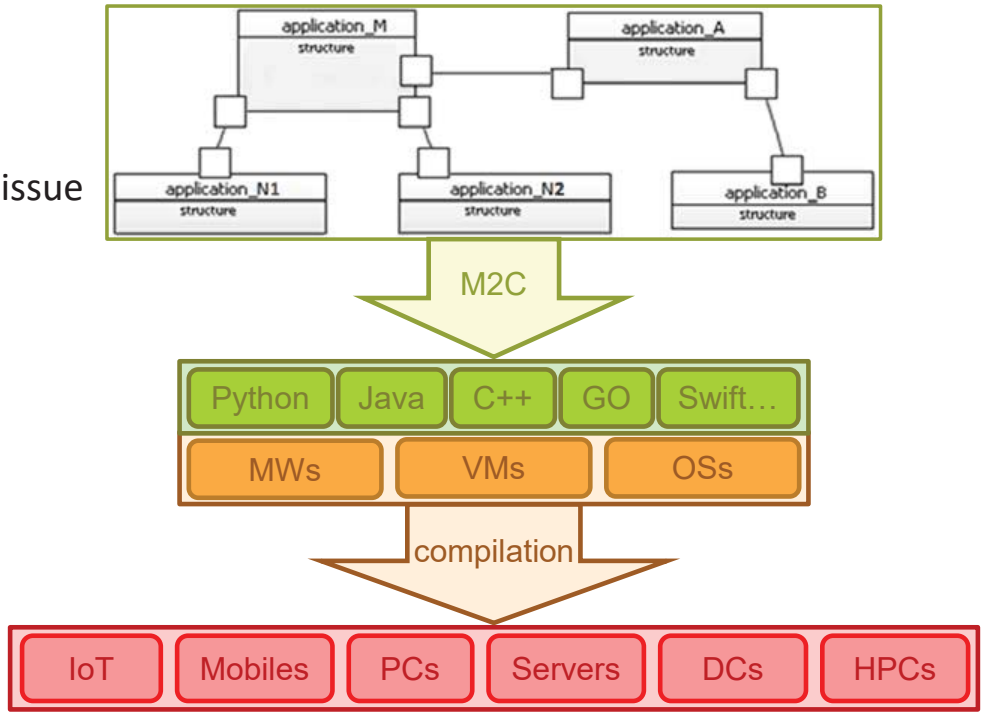


Microelectronics Engineering Group
University of Cantabria

GOBIERNO DE ESPAÑA MINISTERIO DE CIENCIA E INNOVACIÓN
COMP4DRONES

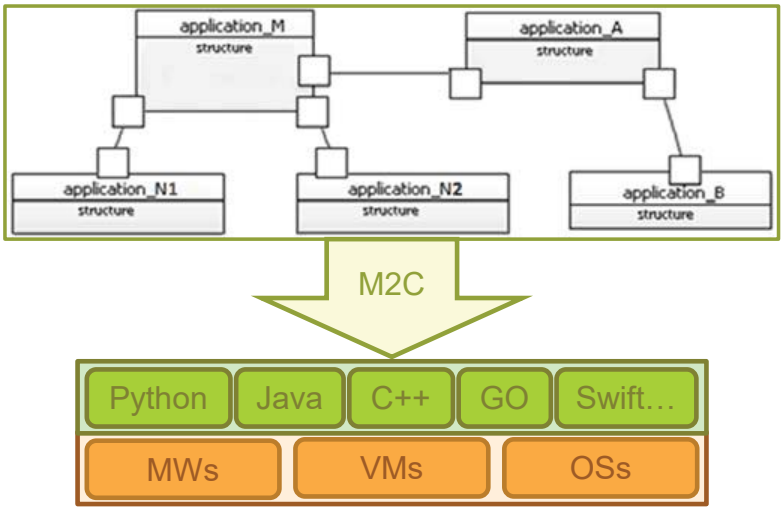
Introduction

- Model-Driven Design (MDD)
  - Abstraction of the platform
  - SW development on APIs
    - MWs, VMs, OSs….
  - Simulation is not always an issue
    - SiL & HiL Verification
  - Performance is not key



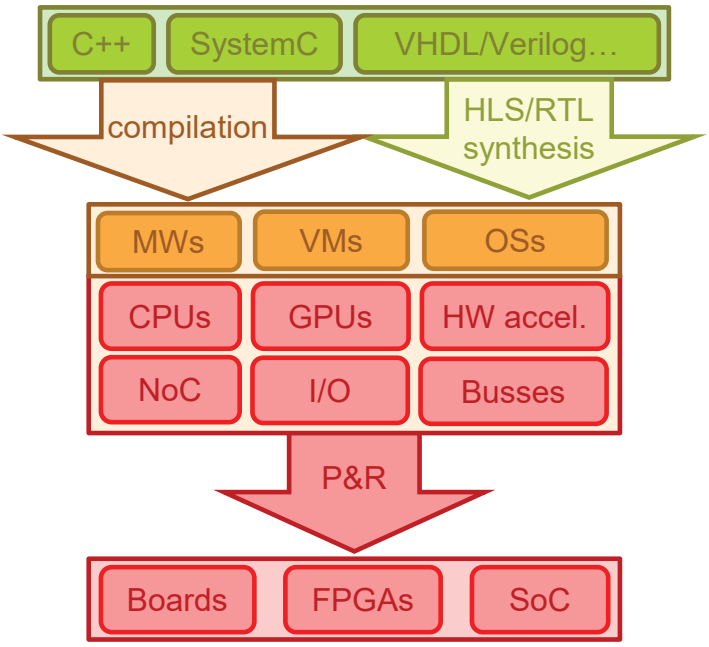
Introduction

- State-of-the-Art in Simulation-Based MDD
  - Matlab-Simulink
    - Proprietary
    - only one MoC, M language
    - Application to UAVs
      - Autopilot + Physics
    - ROS toolbox
  - AMeSIM/ANSyS
    - Proprietary, only one MoC
  - CoFluent
    - Proprietary
    - a few MoCs, C/C++ language
  - Ptolemy II
    - Academic, any MoC, C/C++ inside a Java block
  - HEPSYCODE
    - Academic, several MoCs, SystemC
  - ...



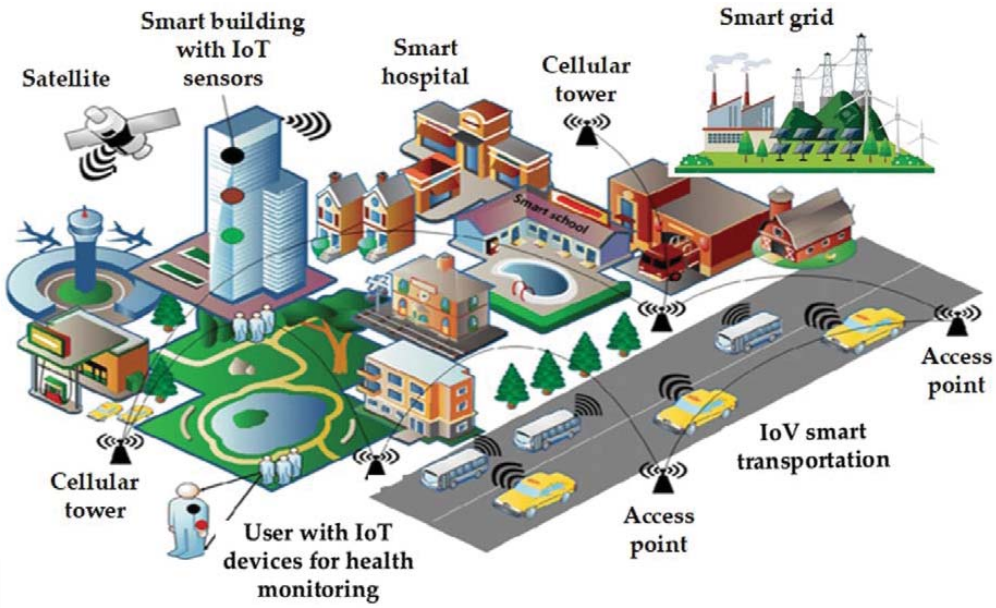
Introduction

- Embedded System Design (ESD) & Electronic System-Level Design (ESL)
  - Model of the platform
    - At different abstraction levels
  - Platform-based HW/SW co-design
  - Simulation is key
    - At different abstraction levels
  - Performance is key
    - Cyber-Physical interaction
    - Real-Time Systems



Introduction

- Services provided on computing platforms of many kind
- Programming the Internet of Everything
  - In close interaction with the physical world => CPSoS-IoT
  - Full abstraction of the computing platform is no longer possible

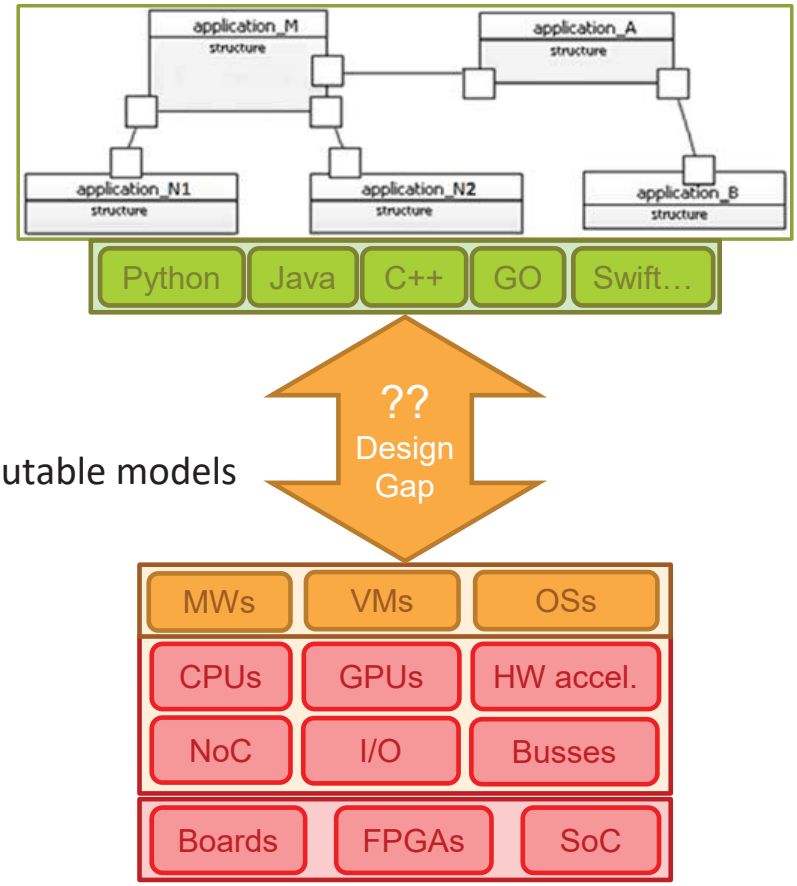


Microelectronics Engineering Group
University of Cantabria

GOBIERNO DE ESPAÑA MINISTERIO DE CIENCIA E INNOVACIÓN
COMP4DRONES

Introduction

- Gap between MDD & ESD/ESL
- S3D: Single-Source System Design
  - Linking MDD with ESD/ESL
  - Simulating the PSM
    - At different abstraction levels
  - Estimating performance
    - Depending on the HW Platform
  - Automatic generation of the executable models
  - Design-Space Exploration

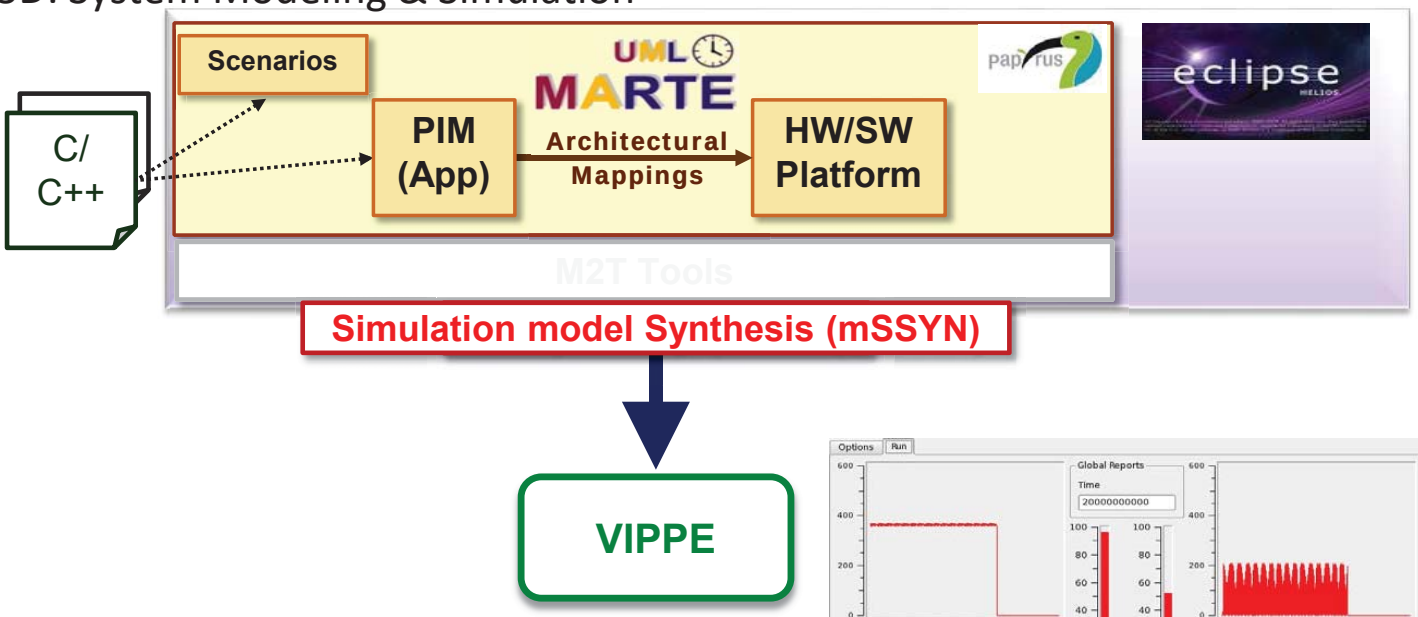


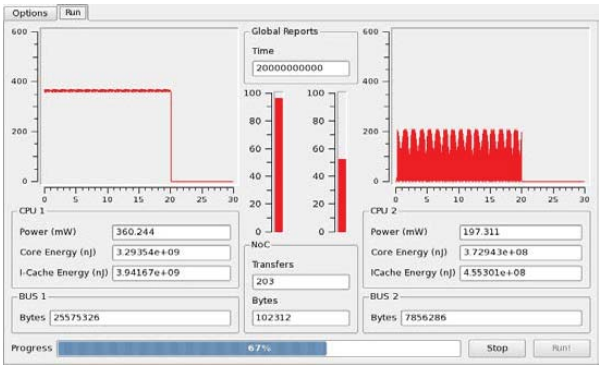
Microelectronics Engineering Group
University of Cantabria

GOBIERNO DE ESPAÑA MINISTERIO DE CIENCIA E INNOVACIÓN
COMP4DRONES

Introduction

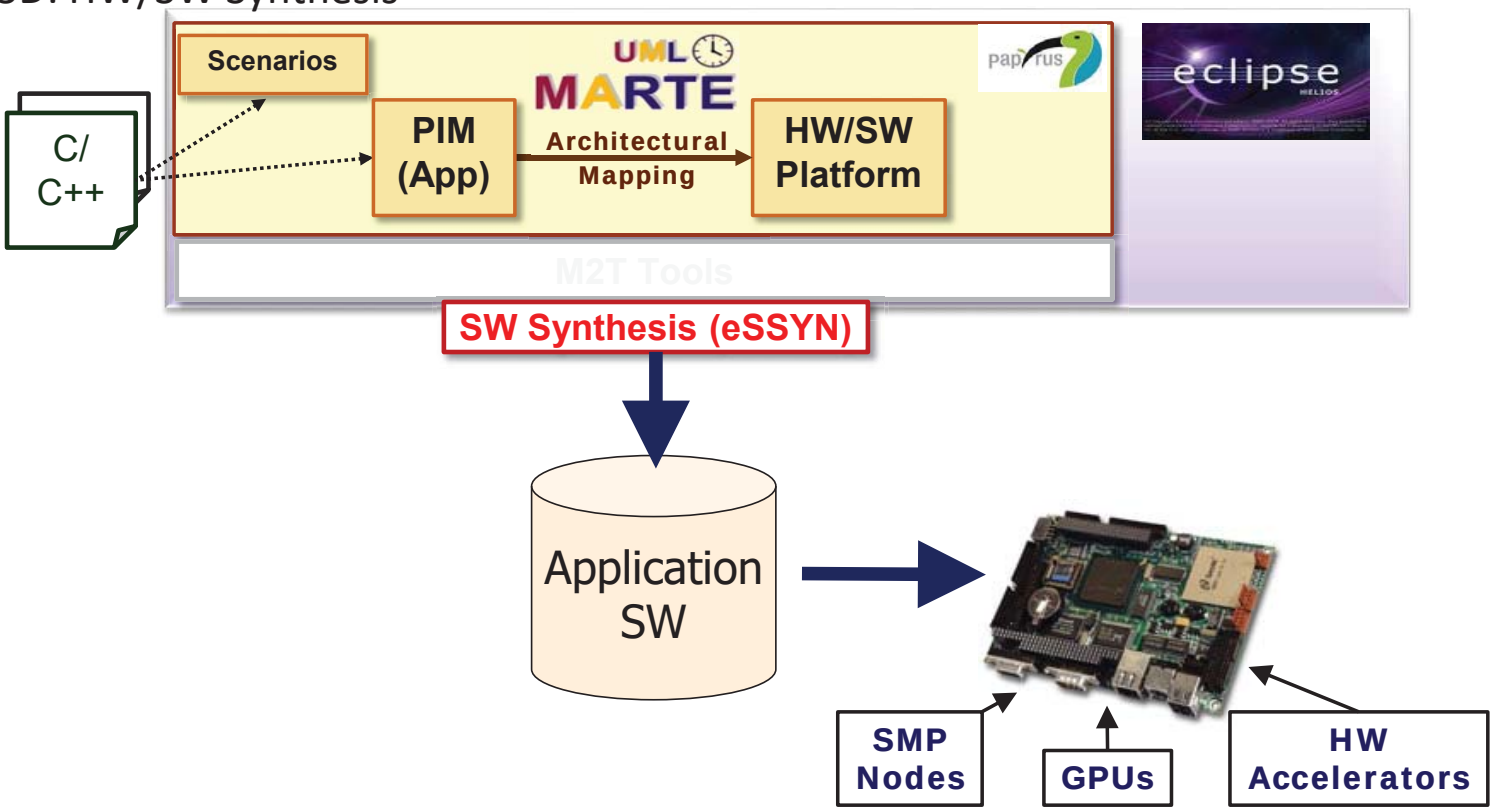
- S3D: System Modeling & Simulation





Introduction

- S3D: HW/SW Synthesis

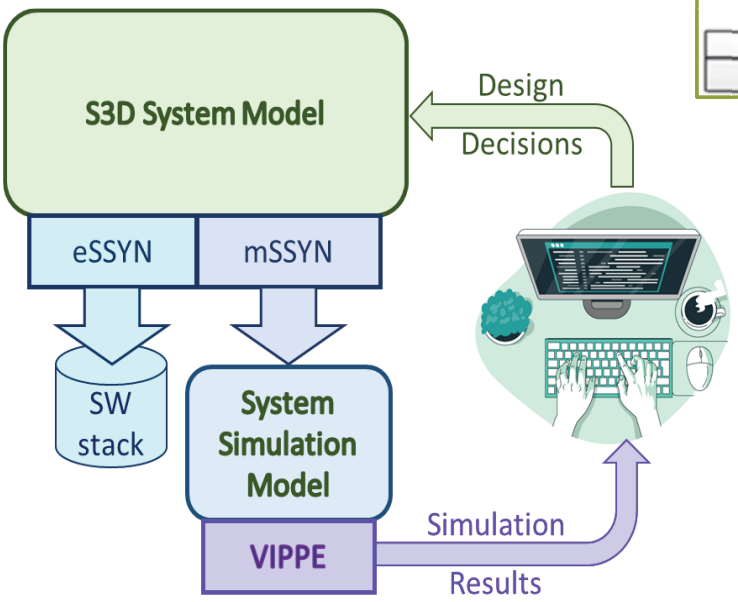


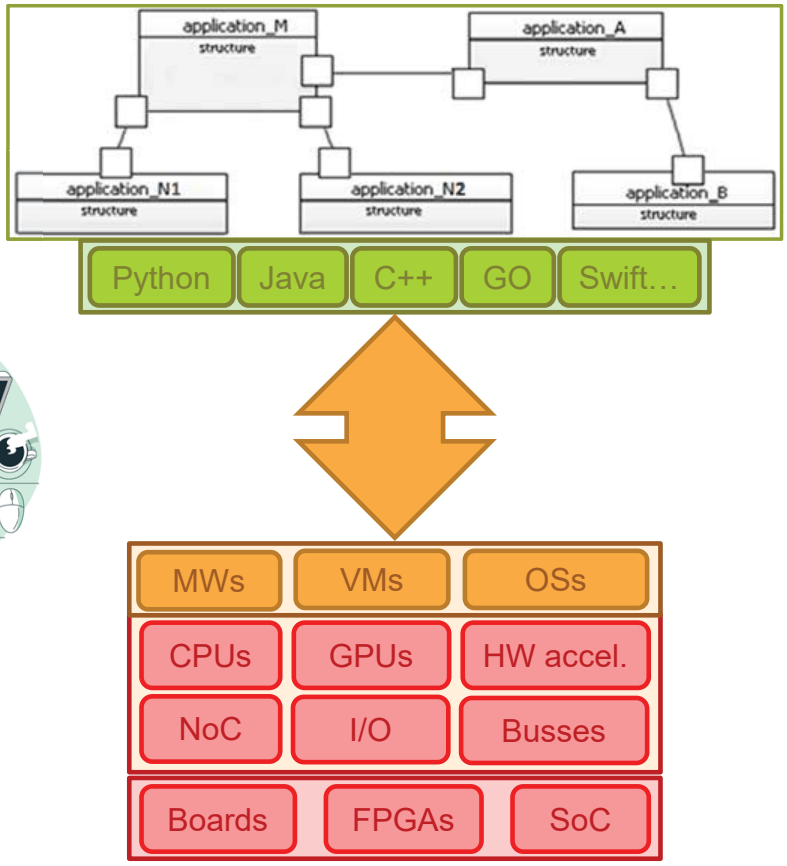
Microelectronics Engineering Group
University of Cantabria

COMP4DRONES

Introduction

- S3D: Design-Space Exploration
  - Linking MDD with ESL



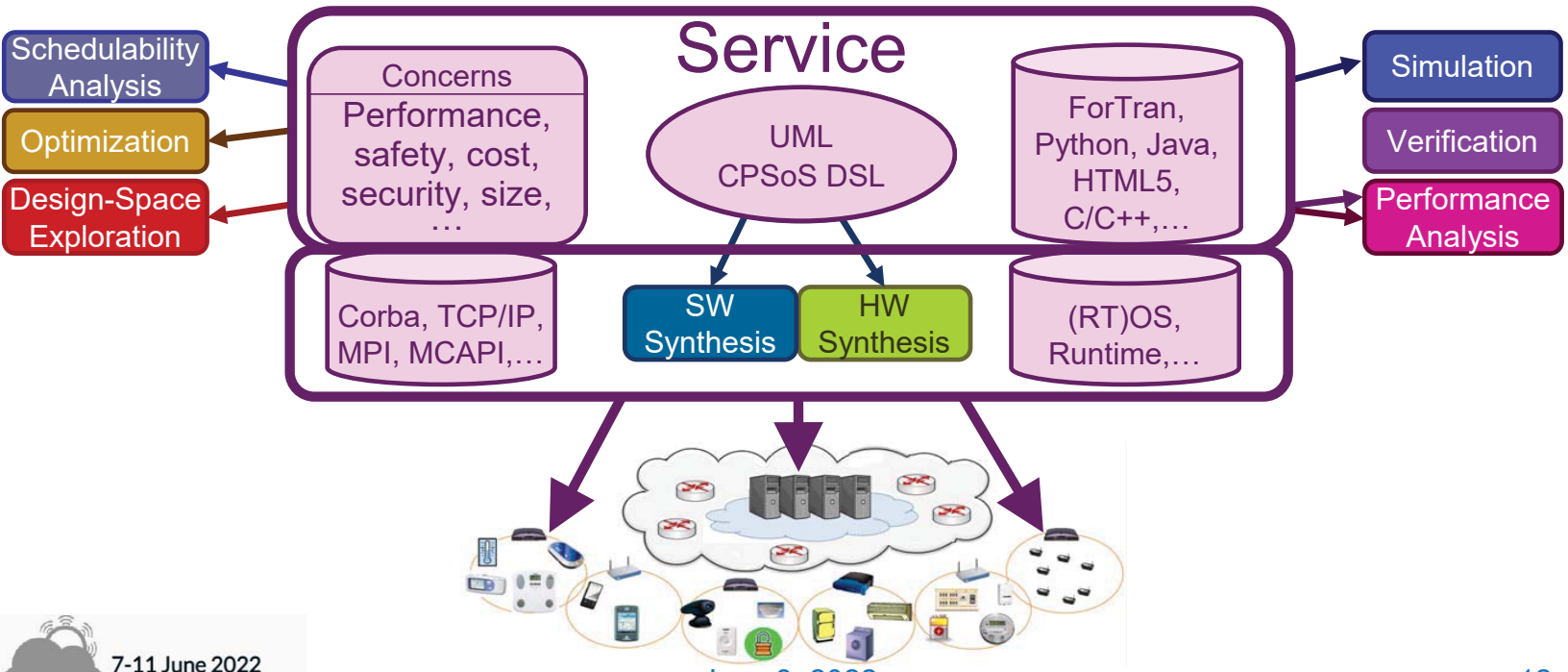


Microelectronics Engineering Group
University of Cantabria

GOBIERNO DE ESPAÑA MINISTERIO DE CIENCIA E INNOVACIÓN
COMP4DRONES

Introduction

- S3D: Programming the Internet of Everything
- Services provided on heterogeneous computing platforms of many kind
- Impact on functional and non-functional properties of the execution platform

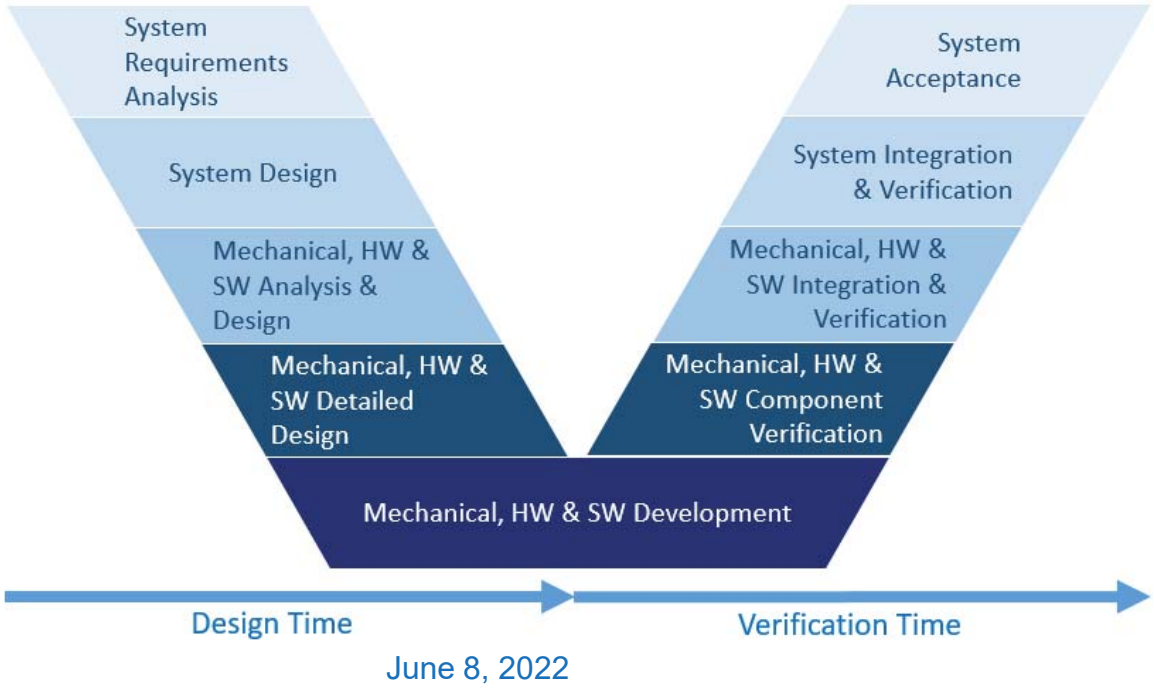


Microelectronics Engineering Group
University of Cantabria

GOBIERNO DE ESPAÑA MINISTERIO DE CIENCIA E INNOVACIÓN
COMP4DRONES

Model-Driven Design of Cyber-Physical SoS

- Traditional V-Cycle for Mechatronic Systems
  - Design & Analyze
  - Develop (Platform-Dependent)
  - Verify

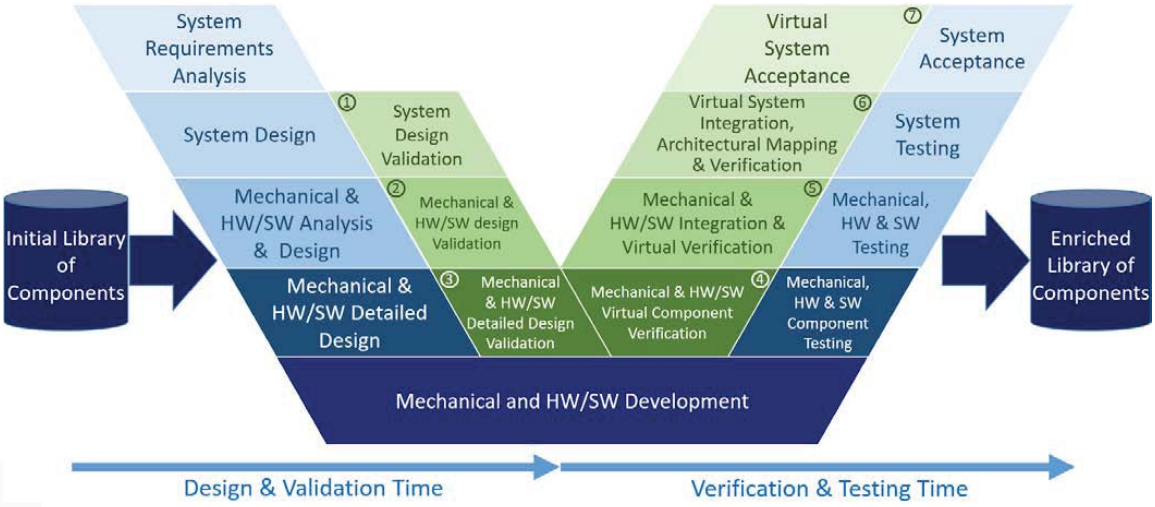


Microelectronics Engineering Group
University of Cantabria

GOBIERNO DE ESPAÑA MINISTERIO DE CIENCIA E INNOVACIÓN
COMP4DRONES

Model-Driven Design of Cyber-Physical SoS

- New Validation/Verification V-Cycle
  - Reusability
  - Simulation-based Analysis & Design
  - Design-Space Exploration
    - Performance Analysis
  - Architectural mapping

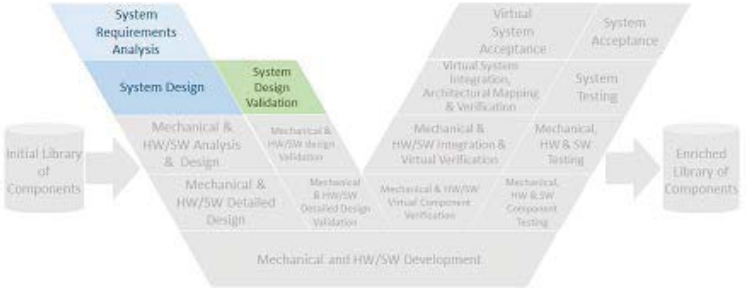


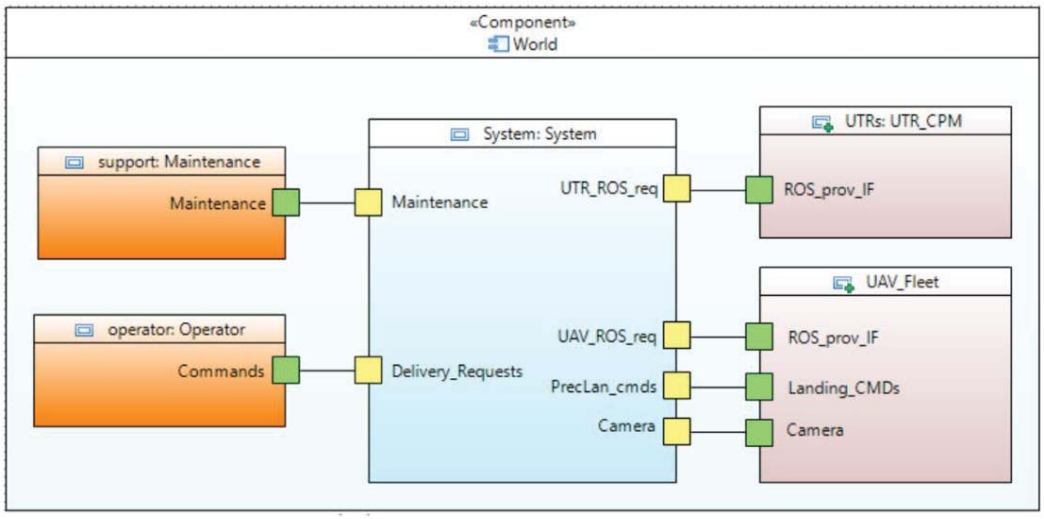
Microelectronics Engineering Group
University of Cantabria

GOBIERNO DE ESPAÑA MINISTERIO DE CIENCIA E INNOVACIÓN
COMP4DRONES

Model-Driven Design of Cyber-Physical SoS

- System Design & Validation
  - System Interface
  - Domain & System Requirements
    - Functional Specification
    - Input/Output Rates and Delays
  - Test-Bench Development
  - Minimal Functionality
    - System Validation



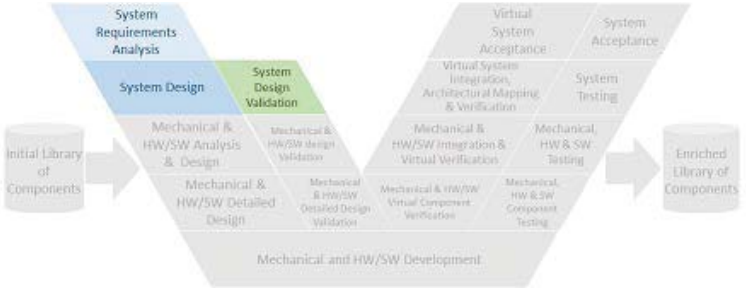


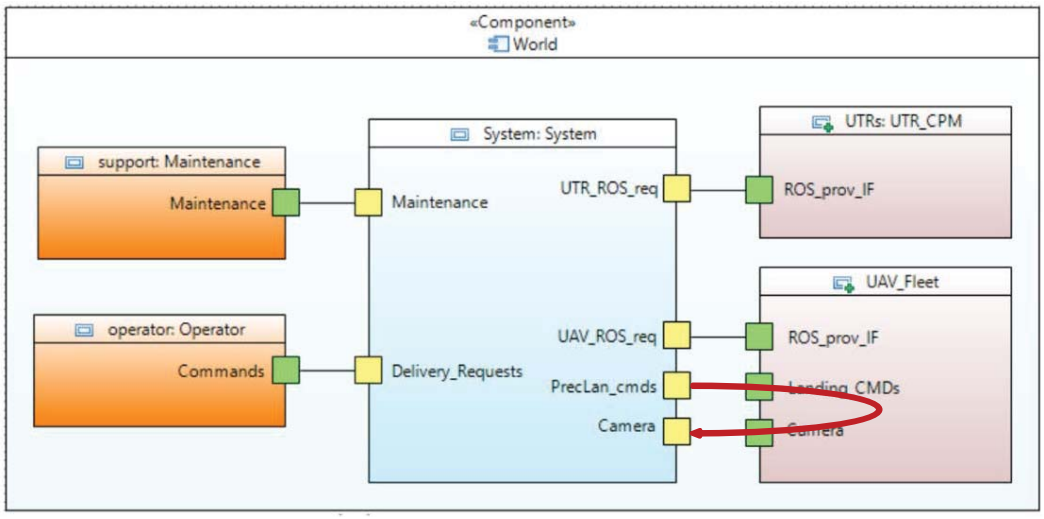
Microelectronics Engineering Group
University of Cantabria

GOBIERNO DE ESPAÑA MINISTERIO DE CIENCIA E INNOVACIÓN
COMP4DRONES

Model-Driven Design of Cyber-Physical SoS

- CPS: Digital Behavior in a Physical World
  - System Model (Specification)
    - The implementation is as good as similar to the model
  - Environment Model
    - The model is as good as similar to reality
  - Close-loop behavior can be extremely difficult to model



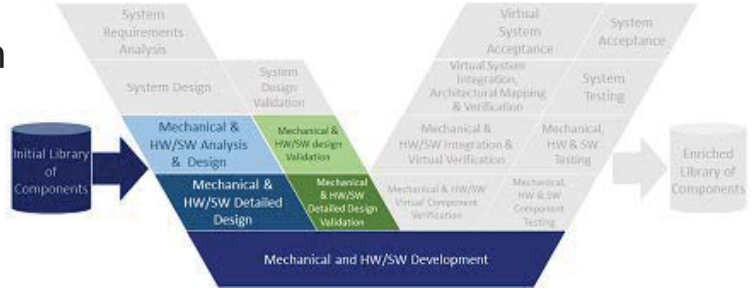


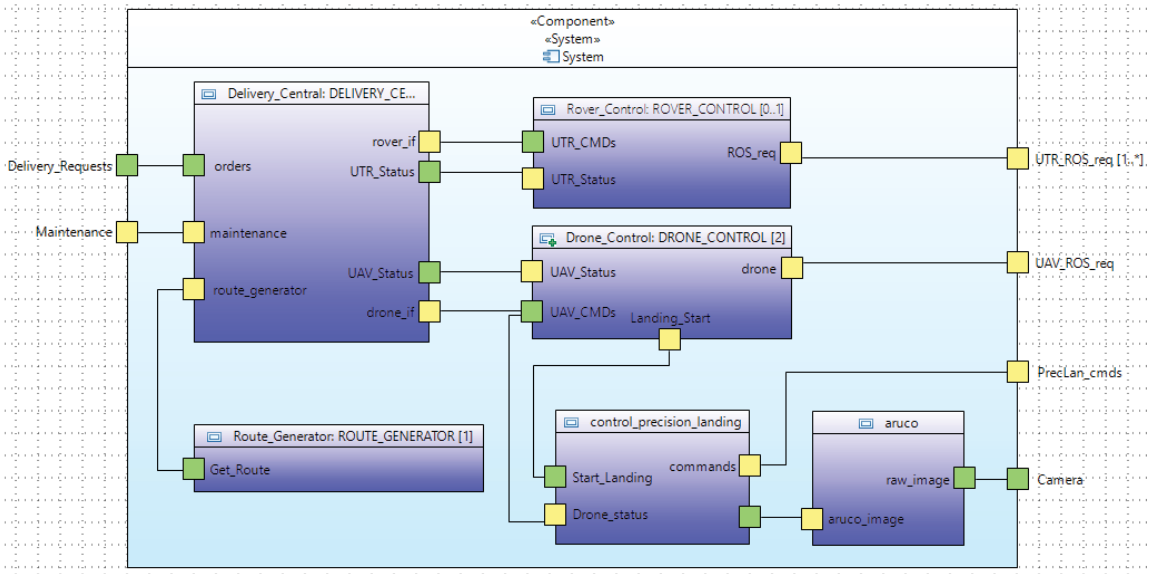
Microelectronics Engineering Group
University of Cantabria

GOBIERNO DE ESPAÑA MINISTERIO DE CIENCIA E INNOVACIÓN
COMP4DRONES

Model-Driven Design of Cyber-Physical SoS

- Mechanical & HW/SW Design & Validation
  - Sub-System & Component Interface
  - Sub-System & Component Requirements
    - Input/Output Rates and Delays
  - Minimal Sub-System & Component Functionality => System Validation
  - Reusability



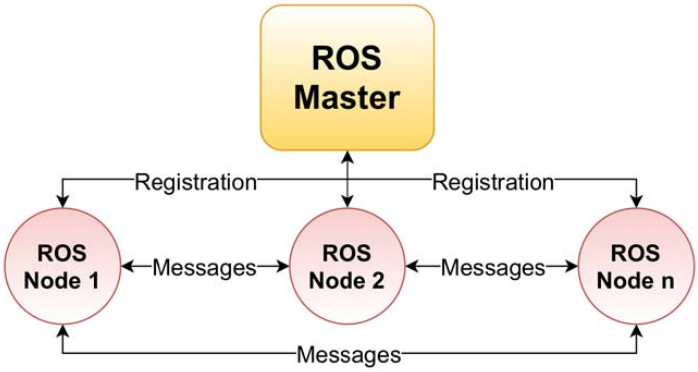


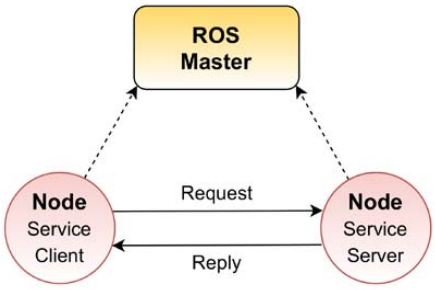
SS-CPSIoT 7-11 June 2022 Budva, Montenegro

June 8, 2022

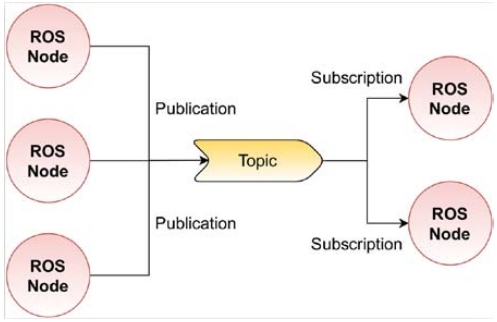
Model-Driven Design of Cyber-Physical SoS

- ROS
  - De-facto standard for robot applications
- ROS Infrastructure
  - ROS nodes
    - Processes that perform a certain computation
  - ROS Master (Core)
    - Nodes registration
    - Communication manager





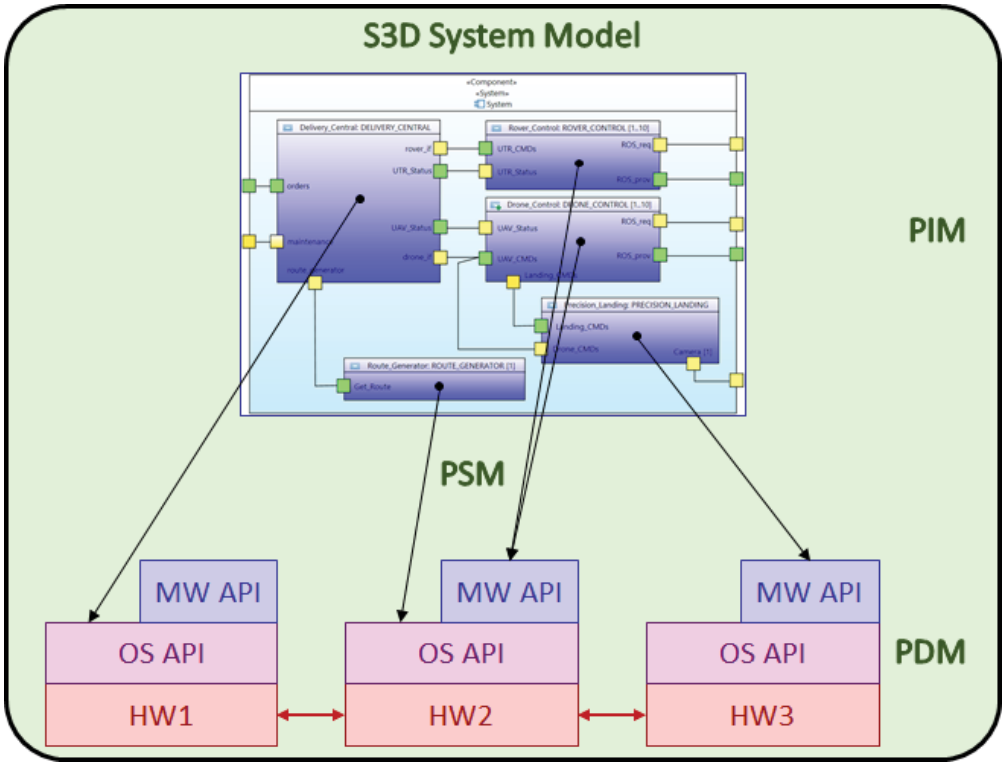
Client-Server (RPC)



Publish-Suscriber

Model-Driven Design of Cyber-Physical SoS

- Platform-Based Design
  - POSIX API
  - ROS API

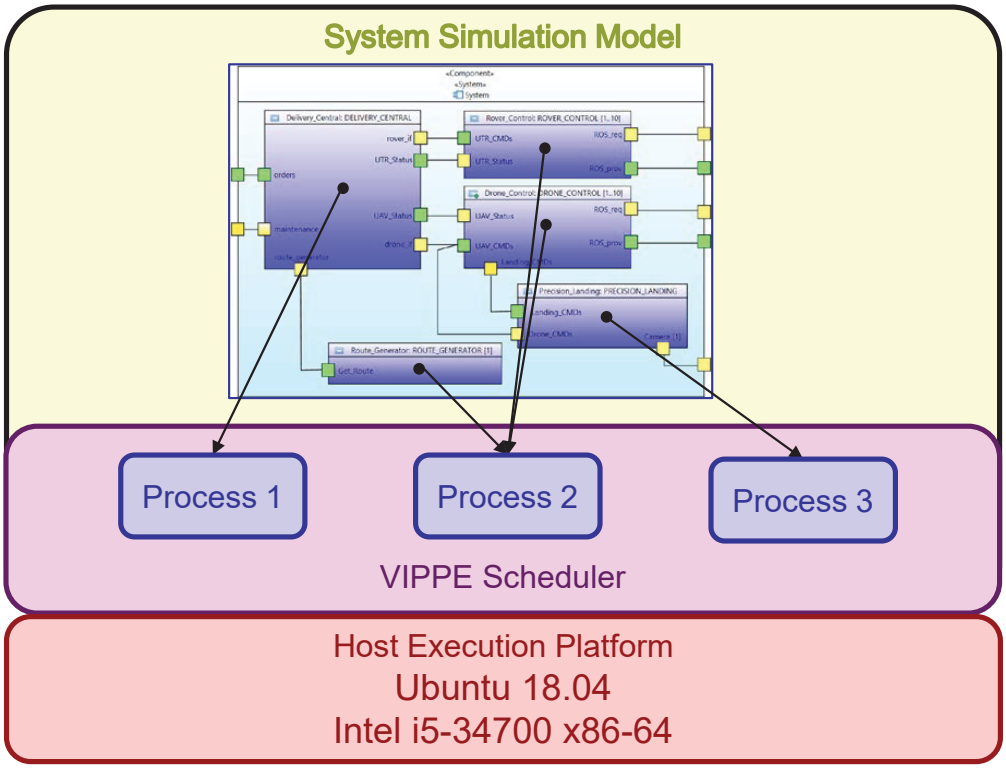


Microelectronics Engineering Group
University of Cantabria

GOBIERNO DE ESPAÑA MINISTERIO DE CIENCIA E INNOVACIÓN
COMP4DRONES

Model-Driven Design of Cyber-Physical SoS

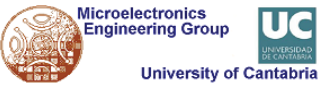
- Platform-Based Simulation
  - POSIX API
  - ROS API



Model-Driven Design of Cyber-Physical SoS

- Drone-Independent ROS Interface
  - Extensible to robots with similar functionality

- **void goTakeoff(int alt)**: When called, the drone rises to the indicated height, switches the drone into guided mode, and will return when it is in position.
- **void RTL()**: The drone returns to the starting point and ends when the drone is on the ground.
- **void goPoint(double lat, double lon, double alt)**: It will set the indicated point as destination and will end when it reaches the indicated position. This function uses the guided mode by default and if it was not in that mode it changes it when it is called.
- **void setManual(double x, double y, double z, double r)**: This function will indicate the position of the joystick as if it were a manual pilot. The arguments passed are X, Y, R to indicate pitch, roll, and yaw respectively and the valid values for these arguments are from -1000 to 1000. Argument Z is the accelerator with valid values between 0 and 1000. This function will terminate immediately so it needs to be called periodically. This function is usable in Stabilized mode and AltHold mode, if neither of these were previously selected, it will automatically switch to AltHold mode.
- **void modeStabilize()**: It will switch to a stabilized mode for joystick control. This method requires a quick feedback loop and is not recommended for general use, but is available for specific tasks. Switching to this mode in mid-flight, without knowing the acceleration required to maintain altitude, can cause the drone to go down quickly.
- **void modeAltHold()**: Same as modeStabilize(), but as long as the throttle is held at 50% (500) the altitude will be constant.
- **void land()**: The drone lands where it is and returns when the drone is perched on the ground.
- **double getAlt()**: This function returns the altitude of the drone when asked.
- **void getPosition(double *latitude, double *longitude, double *altitude)**: This function returns the position in the indicated variables.
- **bool isFlying()**: This function returns true if the drone is flying.
- **void shutdown()**: This function is useful for simulation and will close the flight controller.



Model-Driven Design of Cyber-Physical SoS

- Drone-Independent ROS Interface
  - Reusable ROS code
  - Code length reduction

Reduction in program length			
Drone	Drone-specific code (LoCs)	CoDIn (LoCs)	Reduction (%)
Ardupilot	343	233	32.07
Px4	357		34.73

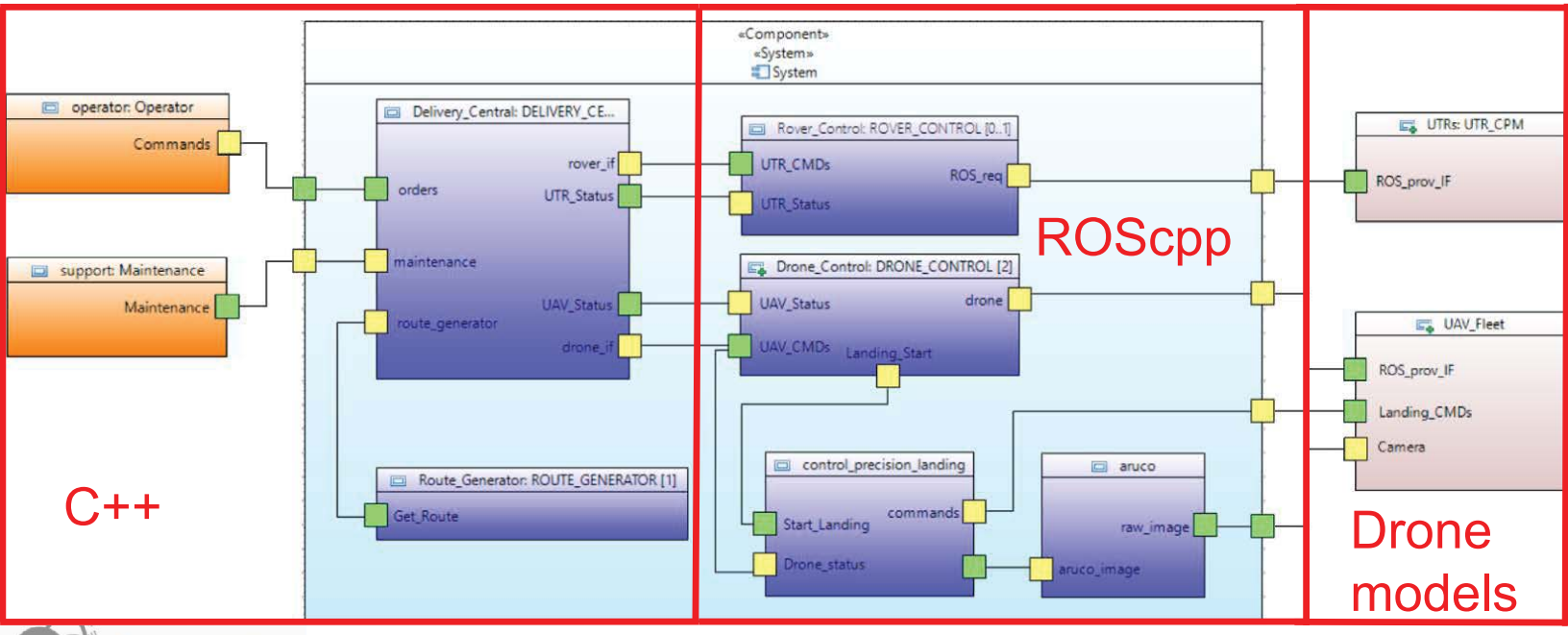
  - Decrease in simulation speed

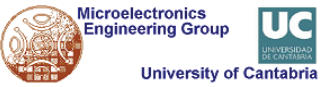
Reduction in simulation speed			
Drone	Drone-specific code (s/s)	CoDIn (s/s)	Reduction (%)
Ardupilot	1.9	1,85	2,6

Microelectronics Engineering Group
University of Cantabria

Design Verification & Performance Analysis

- S3D components in a drone-based service
  - C++ components
  - ROScpp components
  - Drone and robot models





Design Verification & Performance Analysis

- Multi-Level Simulation & Performance Analysis
  - C++ and ROScpp components

Abstraction Levels for C++ & ROS cpp components			
Level	**Code**	**Timing/Energy**	**ROS infrastructure**
MN	Minimal	No	Yes/No
MC	Minimal	Constant	
FC	Full code	Constant	
FD	Full code	Data-dependent	

  - Simulation of ROScpp components without the (slow) ROS infrastructure
    - Functional ROS
    - Direct links among publishers and subscribers

Microelectronics Engineering Group
University of Cantabria

Design Verification & Performance Analysis

- Native Simulation: Flaxibility + accuracy

Global variable
int Sim_Time = 0;

```
…
Overflow = 0;
s = 1L;
for (i = 0; i < L_subfr; i++) {
    Carry = 0;
    s = L_macNs(s, xn[i], y1[i]);
    if (Overflow != 0) {
        break; }}
if (Overflow == 0)  {
    exp_xy = norm_l(s);
    if (exp_xy<=0)
        xy = round(L_shr (s, -exp_xy));
    else
        xy = round(L_shl (s, exp_xy));}
mutex_lock(mutex_name);
…
```

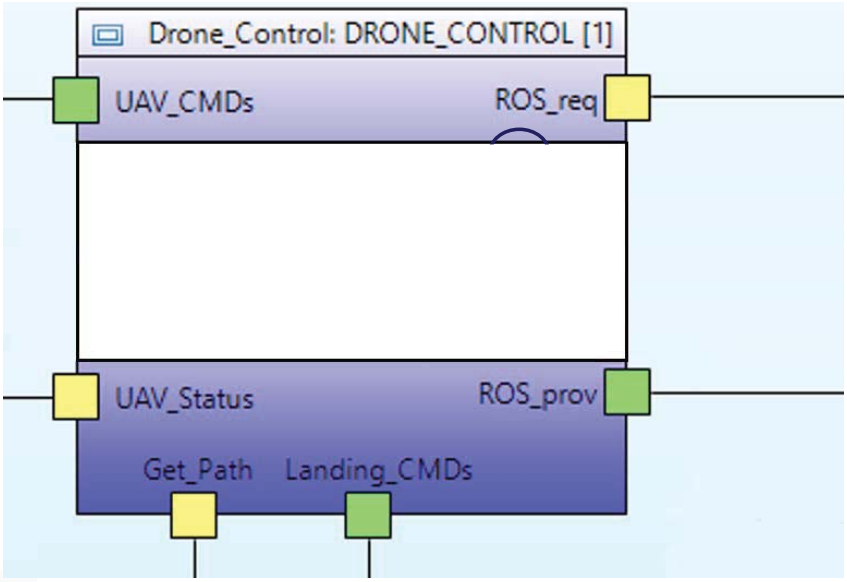
Sim_Time += T_B();
Sim_Time += T_B();
Sim_Time += T_B();
Sim_Time += T_B();
Sim_Time += T_B();
Sim_Time += T_B();
wait included
Sim_Time += T_{SYS}();

T_B() is a function of
- # of binary instructions
- type of instructions
- # of cache misses
- frequency
- …
- even
- data dependencies

T_{SYS}() is a function of
- preemptions
- conflicts in the bus…

Design Verification & Performance Analysis

- Performance Analysis of ROScpp components
  - Native simulation of C++ code
  - Constant time/energy for ROS method calls
    - Dependent on the CPU
    - Dependent on the number of nodes and subscribers
    - Part to be assigned to the component

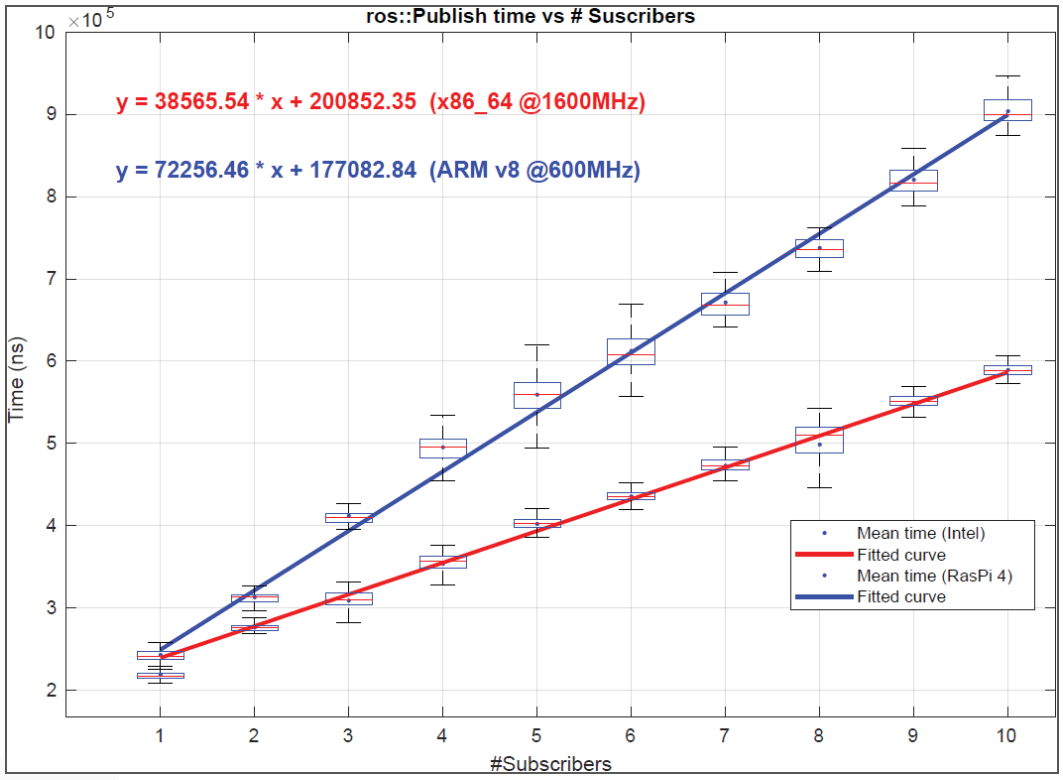


Microelectronics Engineering Group
University of Cantabria

GOBIERNO DE ESPAÑA MINISTERIO DE CIENCIA E INNOVACIÓN
COMP4DRONES

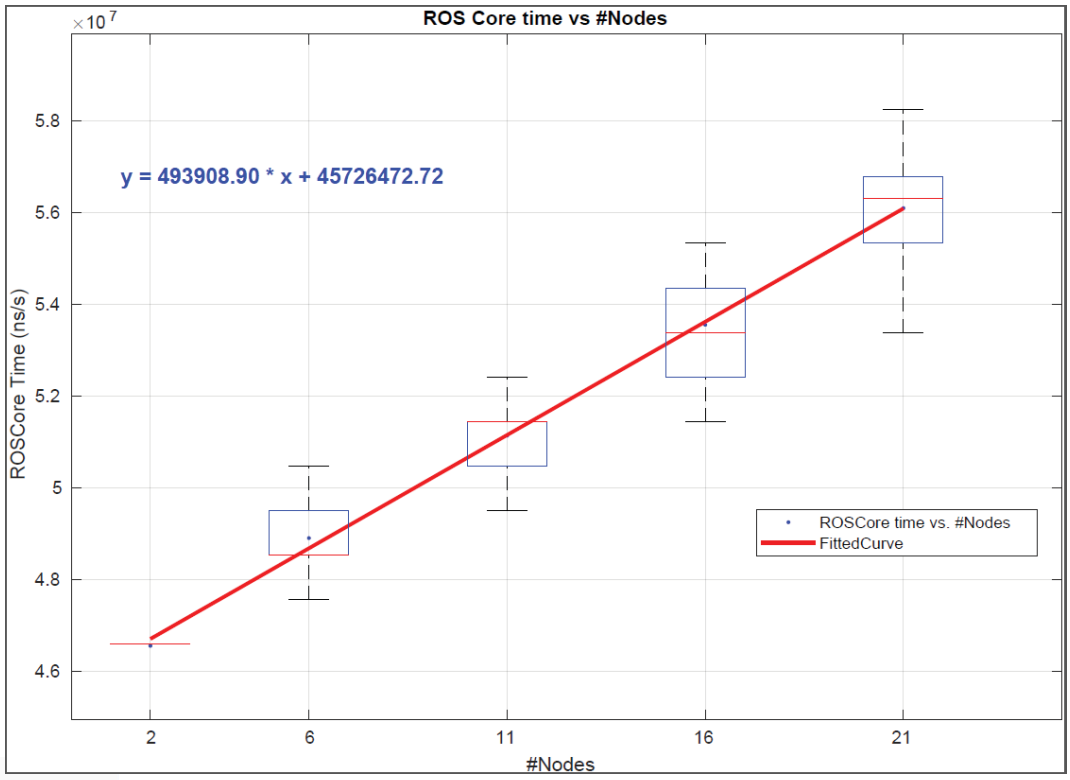
Design Verification & Performance Analysis

- Performance Analysis of ROScpp components
  - Time/energy for ROS method calls at the component



Design Verification & Performance Analysis

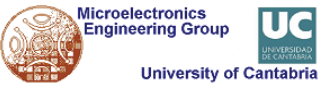
- Performance Analysis of ROScpp components
  - Time/energy for ROS method calls at ROScore



Design Verification & Performance Analysis

- Performance Analysis
  - Estimation Error

	Frequency (MHz)	Estimated Time (ms)	Measured Time (ms)	Estimation Error (%)
Intel	1600	485.87	582.58	16.60
	3000	318.62	461.46	30.95
ARM	600	521.46	904.95	42.38
	1500	233.96	398.26	41.25



Design Verification & Performance Analysis

- Performance Analysis
  - Error reduction

	Frequency (MHz)	Total Application time (ms)	%ROS code	Impact of ROS estimation error (%)	Impact of no estimation	Improvement
Intel	1600	22,875	2.55	0.42	2.55	83.5%
	3000	22,018	2.10	0.65	2.10	69.0%
ARM	600	213,446	0.42	0.18	0.42	57.1%
	1500	87,688	0.45	0.19	0.45	57.8%

  - Improvement increases as the ROS percentage increases

Design Verification & Performance Analysis

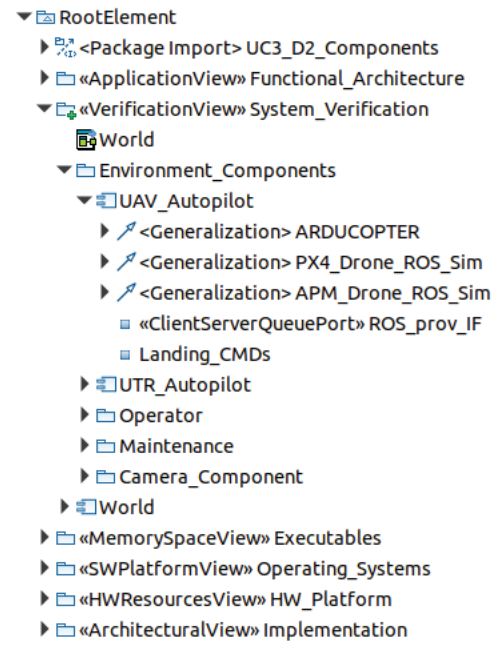
- Multi-Level Simulation & Performance Analysis
  - Drone models

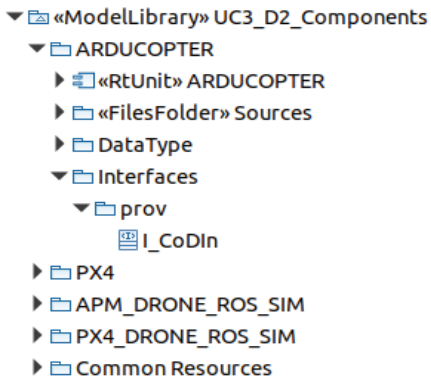
Abstraction Levels for drone models			
Level	**Drone model**	**Physical model**	**ROS infrastructure**
FN	Functional	No	No
FY	Functional	No	Yes
AY	Autopilot	Yes	Yes
AM	Autopilot	Electro-Mechanical	Yes

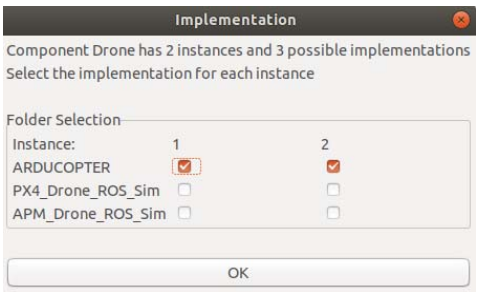
Microelectronics Engineering Group
University of Cantabria

Design Verification & Performance Analysis

- Drone and robot models
  - Components in the S3D 'ModelLibrary'
  - Instantiation in the 'VerificationView'







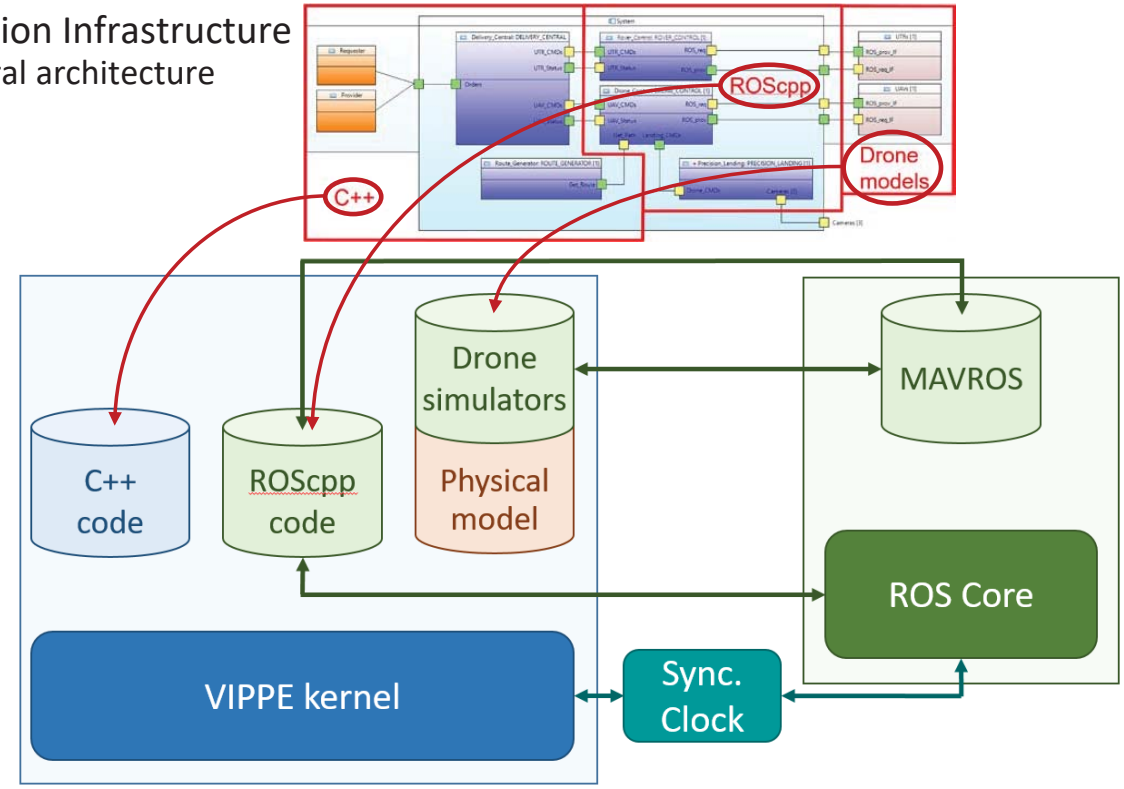
  - Selection for synthesis of the executable model

Microelectronics Engineering Group
University of Cantabria

GOBIERNO DE ESPAÑA MINISTERIO DE CIENCIA E INNOVACIÓN
COMP4DRONES

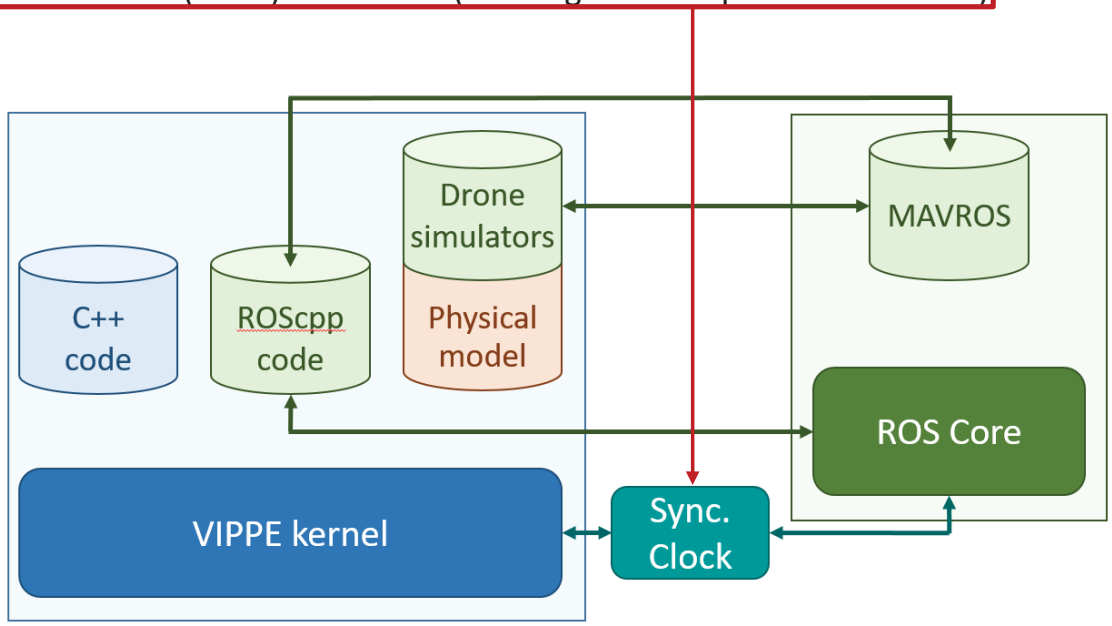
Design Verification & Performance Analysis

- Multi-Level Simulation & Performance Analysis
  - Simulation Infrastructure
    - General architecture



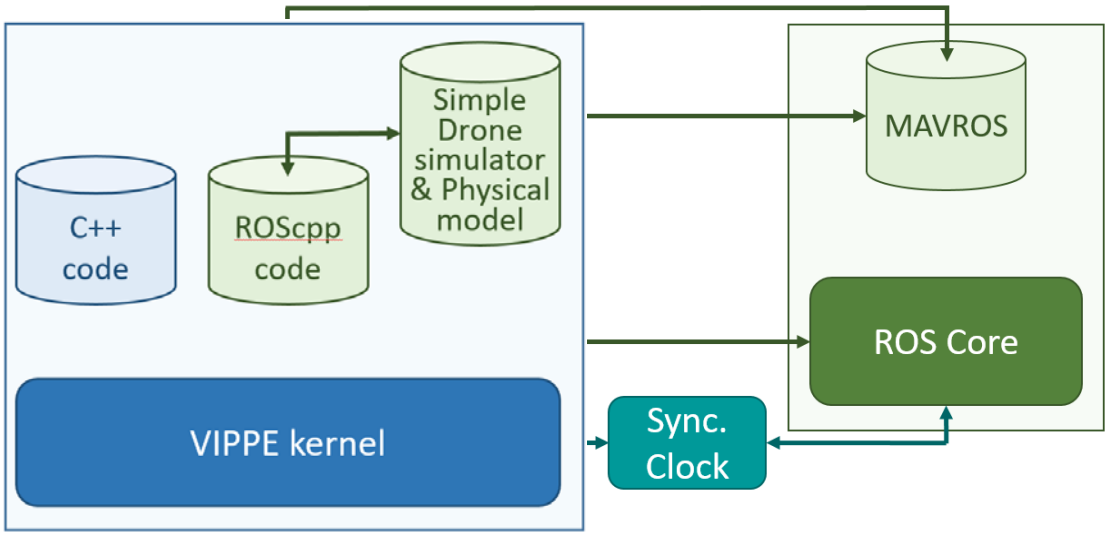
Design Verification & Performance Analysis

- Multi-Level Simulation & Performance Analysis
  - Simulation Infrastructure
    - General architecture
    - Real-Time (RT) simulation- simulation time = simulated time (SnT = SdT)
    - As Fast As Possible (AFAP) simulation (SnT as greater as possible than SdT)



Design Verification & Performance Analysis

- Multi-Level Simulation & Performance Analysis
  - Simulation Infrastructure
    - Functional drone modeling
    - Without ROS (FN)
    - Any C++ and ROScpp models (MN + MC + FC + FD)

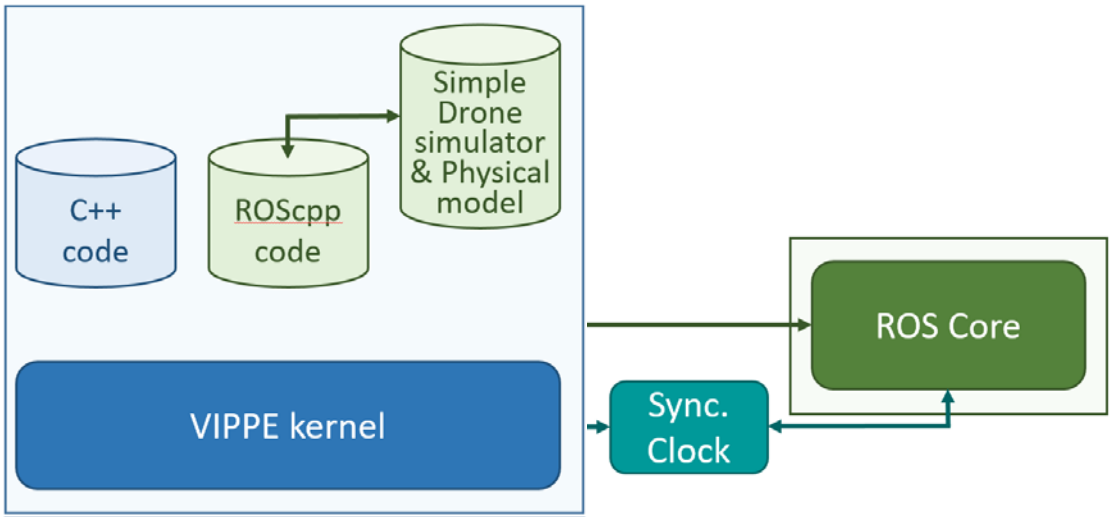


Microelectronics Engineering Group
University of Cantabria

GOBIERNO DE ESPAÑA MINISTERIO DE CIENCIA E INNOVACIÓN
COMP4DRONES

Design Verification & Performance Analysis

- Multi-Level Simulation & Performance Analysis
  - Simulation Infrastructure
    - Functional drone modeling
    - With ROS (FY)
    - Any C++ and ROScpp models (MN + MC + FC + FD)

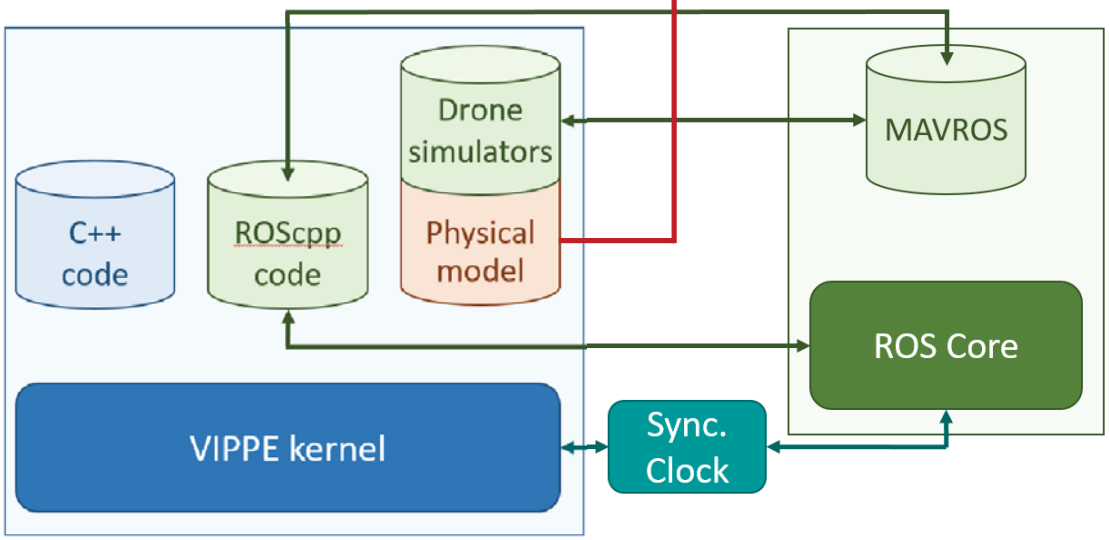


Microelectronics Engineering Group
University of Cantabria

GOBIERNO DE ESPAÑA MINISTERIO DE CIENCIA E INNOVACIÓN
COMP4DRONES

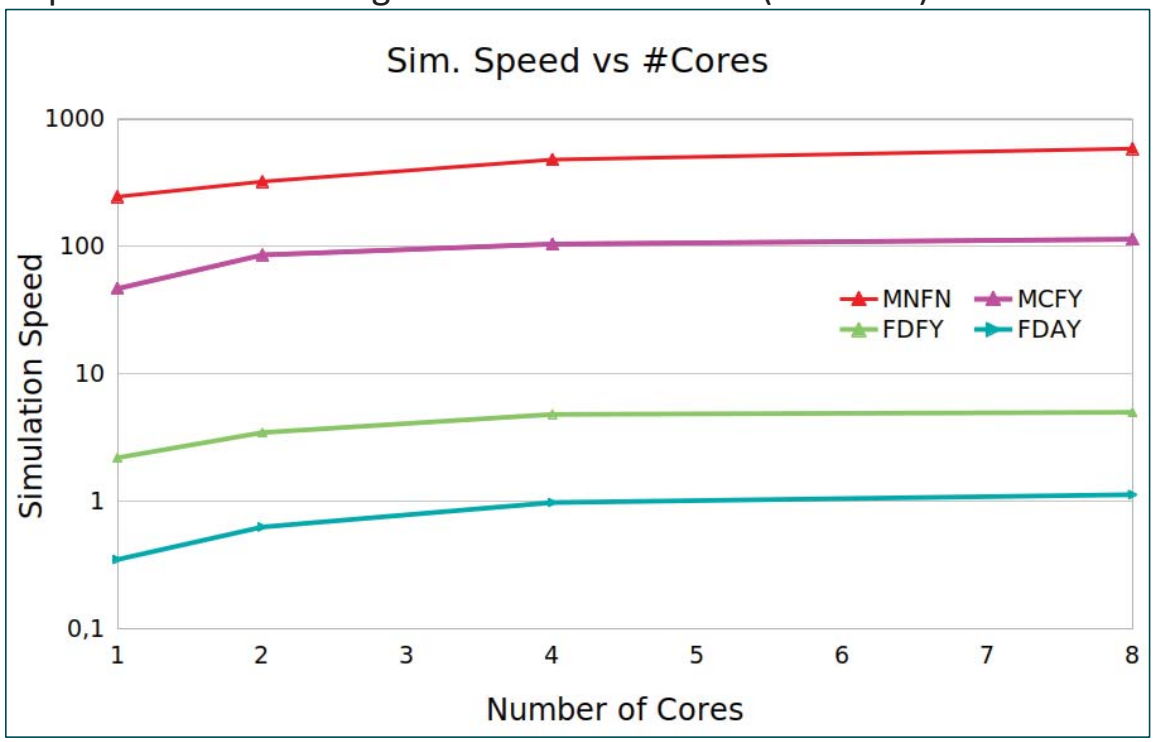
Design Verification & Performance Analysis

- Multi-Level Simulation & Performance Analysis
  - Simulation Infrastructure
    - Realistic drone modeling (Autopilot + Physics)
    - With ROS (AY + AM)
    - Any C++ and ROScpp models (MN + MC + FC + FD)
    - With or without 3D Graphics



Simulation Results

- Multi-Level Simulation
  - Impact of an increasing number of host CPUs (8 drones)

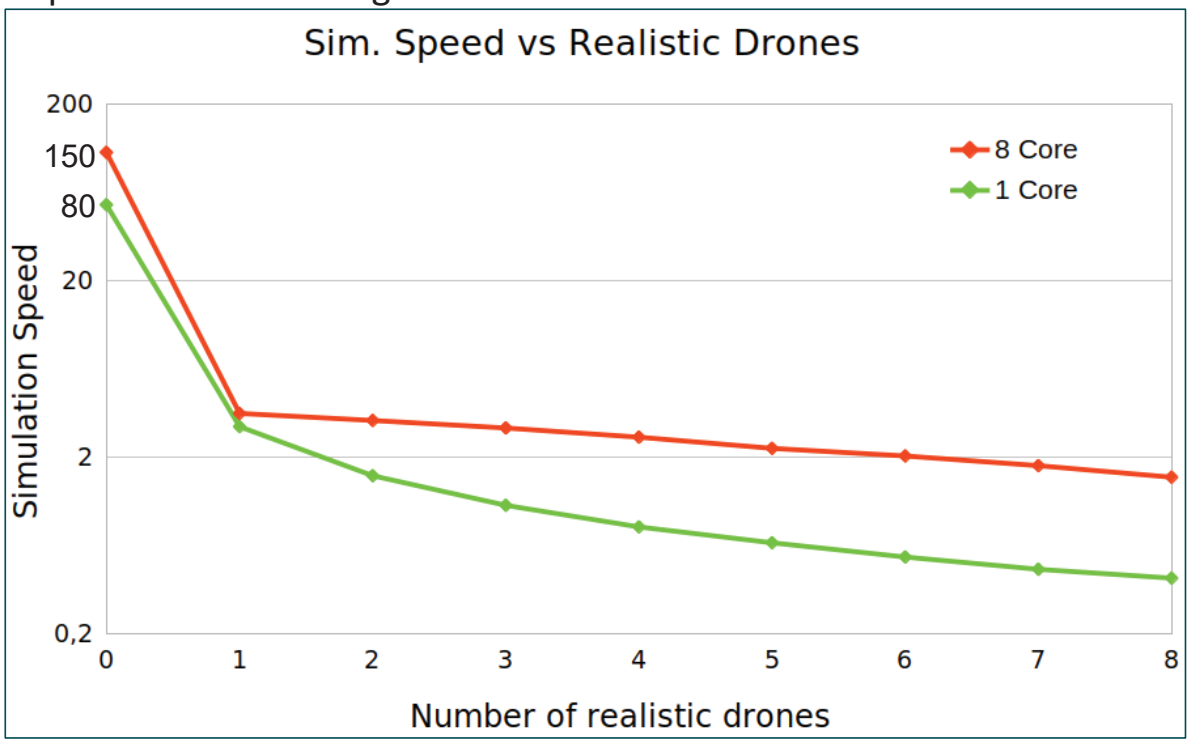


Microelectronics Engineering Group
University of Cantabria

GOBIERNO DE ESPAÑA MINISTERIO DE CIENCIA E INNOVACIÓN
COMP4DRONES

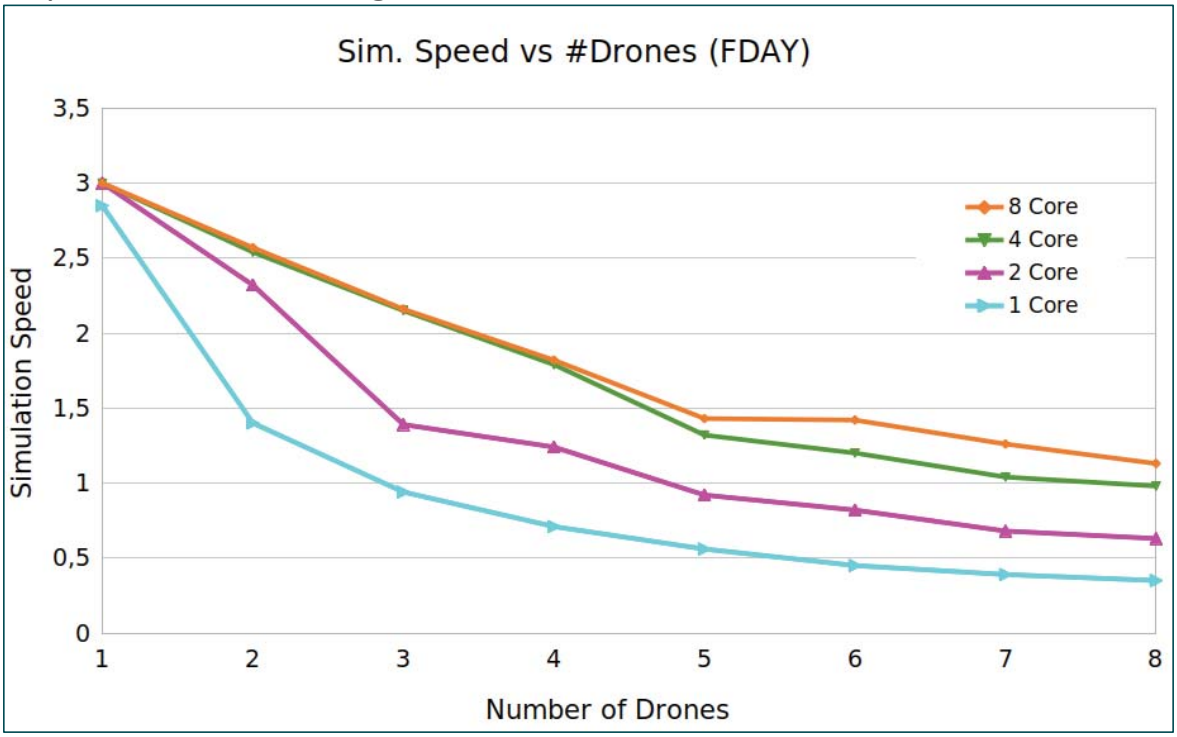
Simulation Results

- Multi-Level Simulation
  - Impact of an increasing number or realistic vs functional drones



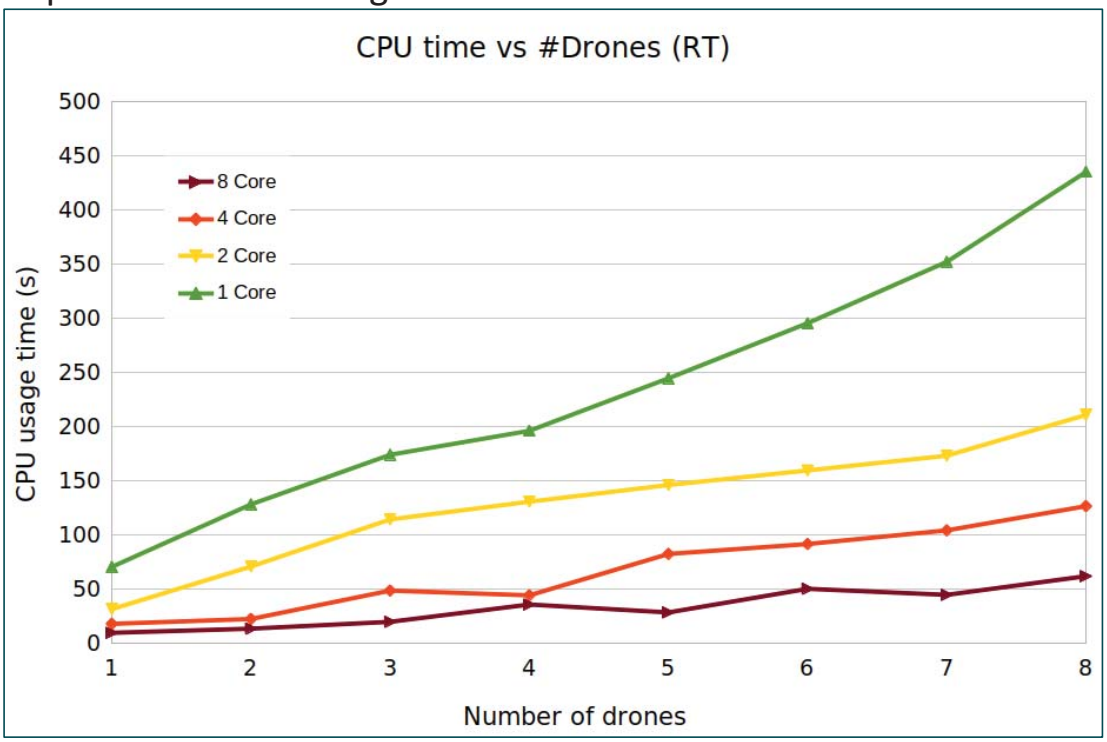
Simulation Results

- Multi-Level Simulation & Performance Analysis
  - Impact of an increasing number of realistic drones



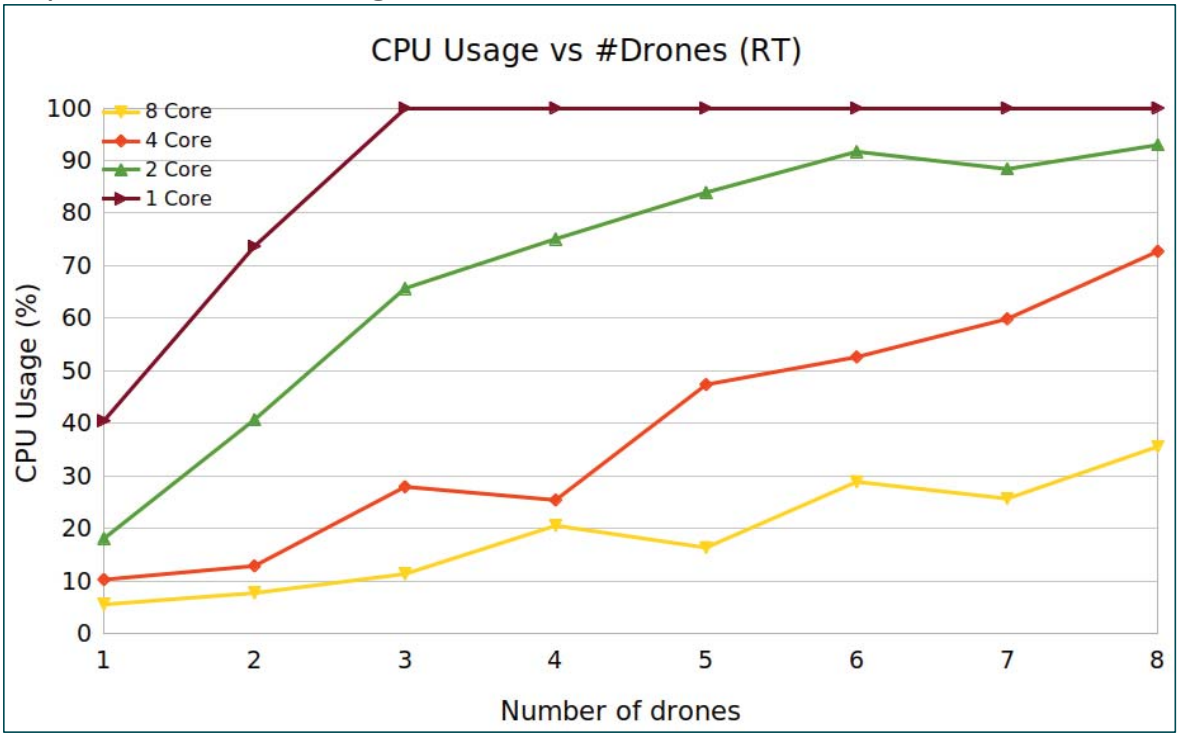
Simulation Results

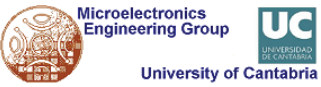
- Real-Time Simulation in seconds
  - Impact of an increasing number of realistic drones



Simulation Results

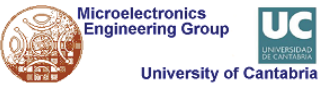
- Real-Time Simulation in % of CPU usage
  - Impact of an increasing number of realistic drones





Conclusions

- Services based on CPS demand new design methods and tools
  - e.g. Drone-based Services
  - Close interaction between the physical world and the digital electronics
- Model-Driven System Design is a powerful candidate
  - HiL & SiL are not enough
  - Model in the loop (MiL) is required
  - Extension of the classical V-Cycle
- Multi-Level Simulation is key in designing drone-based services
  - As Fast As Possible vs Real-Time
- Drones are just pieces inside a complex, distributed functionality
- S3D is a valid approach towards MDD of drone-based services



Acknowledgement

- Last Research Results from many people

- Hector Posadas
- Javier Merino
- Raul Gonzalez
- Jose Maria Gandara
- … and the rest of the Microelectronics Engineering Group

Any comment/question?

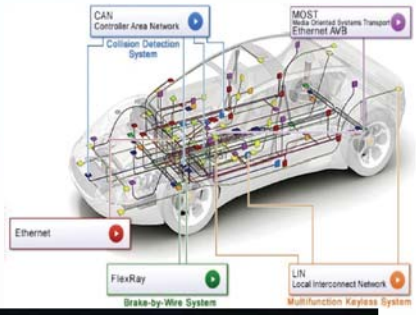
Synthesis of Run-time Monitors for Safe and Secure Industrial Systems

Dimitrios Serpanos, CTI & University of Patras
Stavros Koubias, University of Patras & ISI/ATHENA

Introduction

- Industrial Control Systems (ICS) are Cyber-Physical Systems (CPS)
- CPS combine computation and physics (interdisciplinary area: algorithms, logic, control, …)
- Operational Technology (OT): hardware and software that monitors, controls and manages systems and processes in an industrial setting
- Process-dependent (plants)
- New threat models (false data injection)
- Solution challenges (process-dependence, real-time, continuous operation)

Industrial Control Systems



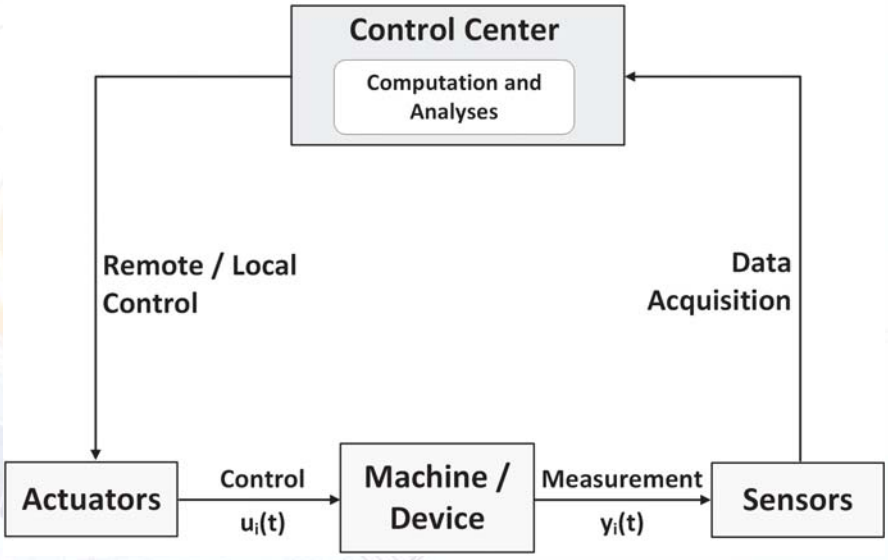
Incidents

- Aurora (2007)
- Stuxnet (2010)
- Mirai (2016)
- Petya/NotPetya (2016/2017)
- Industroyer2 (2022)

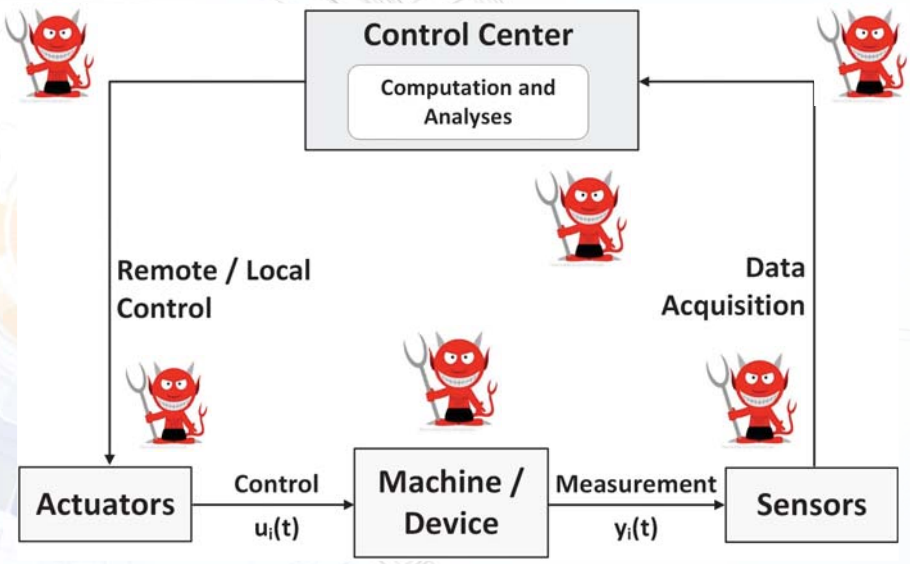
IT vs. OT

	Information Technology	Operational Technology
Purpose	Process transactions, provide information	Control or monitor physical processes and equipment
Architecture	Enterprise wide infrastructure and applications (generic)	Event driven, real time, embedded hardware and software (custom)
Interfaces	GUI, web browser, terminal and keyboard	Electromechanical, sensors, actuators, coded displays, hand-held devices
Ownership	CIO and IT	Engineers, technicians, operators and managers
Connectivity	Corporate network, IP based	Control networks, hardwired twisted pair and IP based
Role	Supports people	Controls machines

ICS Control Loop

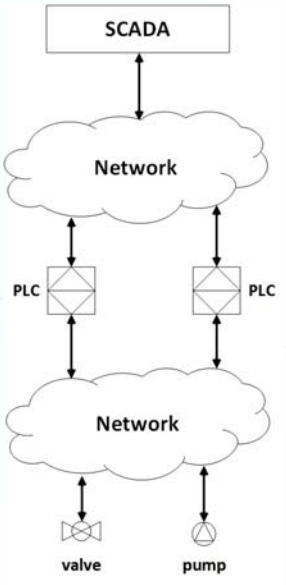


ICS Control Loop Attacks



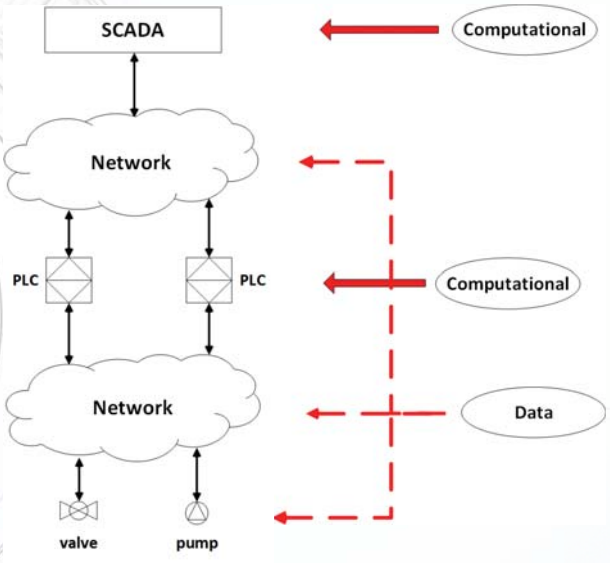
ICS Computational Structure & Requirements

- Hierarchical structure
- Heterogeneous technologies
- Autonomy
- Continuous operation/fail-safe
- Dependability
- Dependence on large number of input devices
- Large installation base (legacy systems)
- Increasing connectivity



Attacks on ICS

- Resilience
- Continuous operation under attack
- Attack mitigation
- Fast recovery after attack
- System evolution without disruption



Safety and Security

Safety

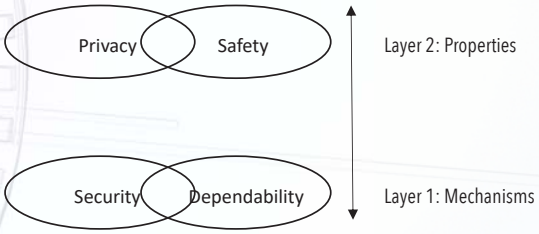
- Safety properties
  - Maintain well-defined state that corresponds to safe operation
- Safety typically expressed as requirements on control loop
- Security is related to safety:
  - Data integrity

Security

- Confidentiality
- Integrity
- Authentication
- Access control
- Non-repudiation
- Dependability
- Safety
- Privacy

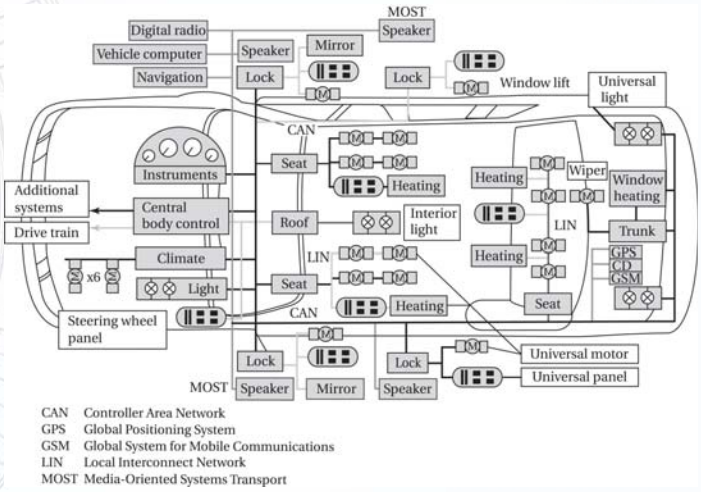
Security property layers

- Security and dependability are mechanisms
- Privacy and safety are system properties
  - Requirements for processes, applications, services
- Privacy and safety depend on security
- Threats:
  - Computational
  - Data



High complexity systems

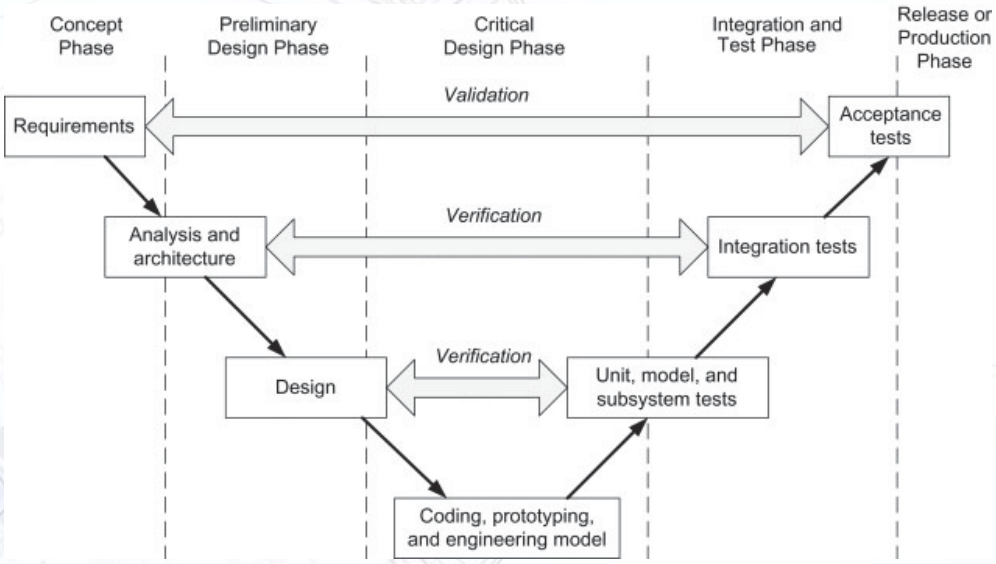
- Ford F150 ships with 150 million lines of code
- Boeing 787 ships with 7 million lines of code



Types of attacks and failures

- Computational attacks
  - Viruses, trojans, worms, …
- Communication attacks
  - Message deletion/alteration, disruption, (D)DoS
- False data injection
  - New type of attack (CPS)

V model


Concept Phase
Preliminary Design Phase
Critical Design Phase
Integration and Test Phase
Release or Production Phase
Validation
Requirements
Acceptance tests
Verification
Analysis and architecture
Integration tests
Verification
Design
Unit, model, and subsystem tests
Coding, prototyping, and engineering model

Field Operation - Monitoring

- Safety/security monitors
  - Computational attacks
  - Communications/network attacks
  - False data injection attacks
  - Failures
- Monitor security
- Runtime operation

Approach - Strategy

- **Build it right and continuously monitor**
  - *US Federal Government Strategy*
- Our approach
  - Programmable (executable) specification with security properties
    - Secure by design
  - Middleware monitoring process (app) execution
    - ARMET compares app and specification execution
  - Specification includes defense against identified process vulnerabilities
    - Vulnerability analysis against false data injection attacks

- Define executable process specification
- Augment with all necessary invariants
- Refine to a single behavioral spec (program)

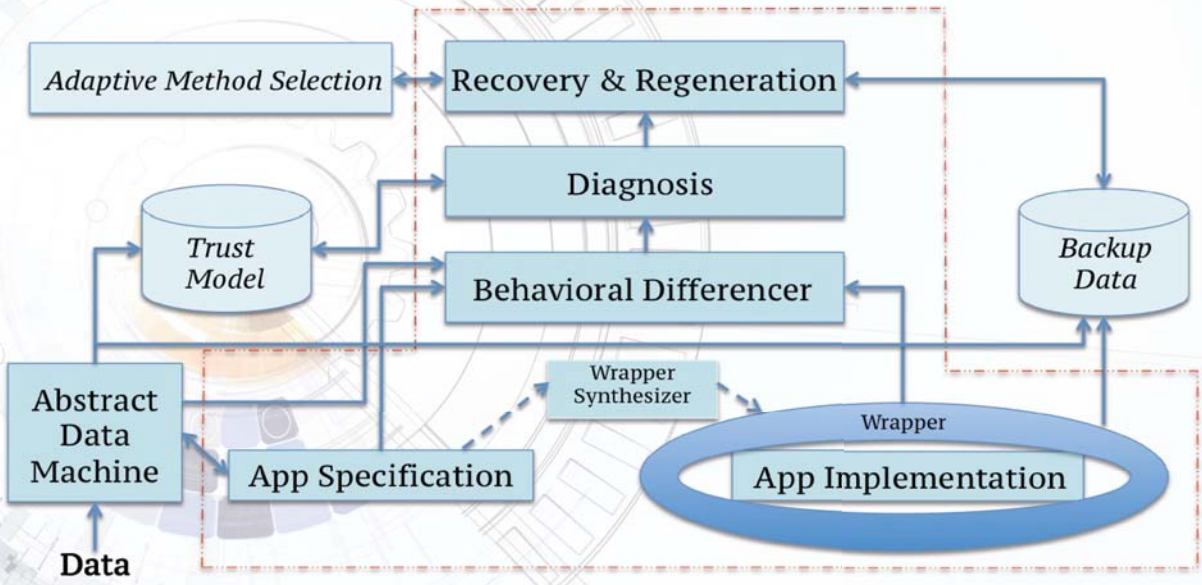
Build it right

- Include implementation and specification to middleware (ARMET)
- Compare predictions (spec) and observations (implementation)

Continuously monitor

- Identify inconsistencies – diagnose - recover

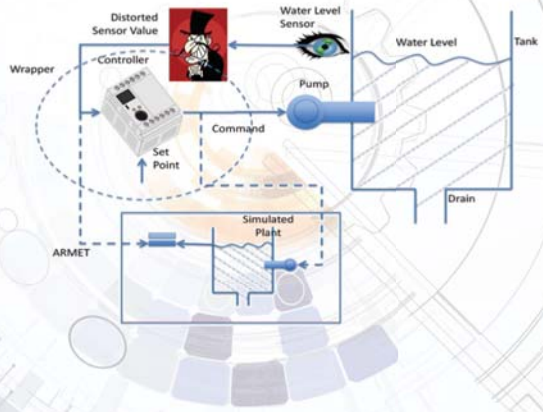
ARMET

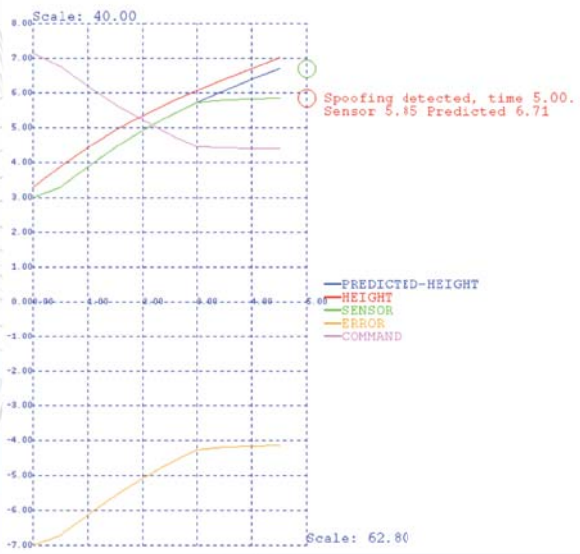

Adaptive Method Selection
Recovery & Regeneration
Diagnosis
Trust Model
Behavioral Differencer
Backup Data
Abstract Data Machine
Wrapper Synthesizer
Wrapper
App Specification
App Implementation
Data

ARMET - Middleware

- Self-aware system
  - Self-awareness through dependency-directed reasoning
- System is allowed to only behave legally
  - Continuous monitoring of prediction/observation consistency
  - IF inconsistency, THEN diagnosis
  - Recovery (safe state from alternate, reliable source)
- Detection of unknown attacks
  - Inconsistency between predictions and observations
- System adaptability to evolutionary constraints
  - ICS-CERT standards, security and privacy policies, etc.
  - Specify policies as legal behavior & monitor behavioral consistency

Example: Water tank

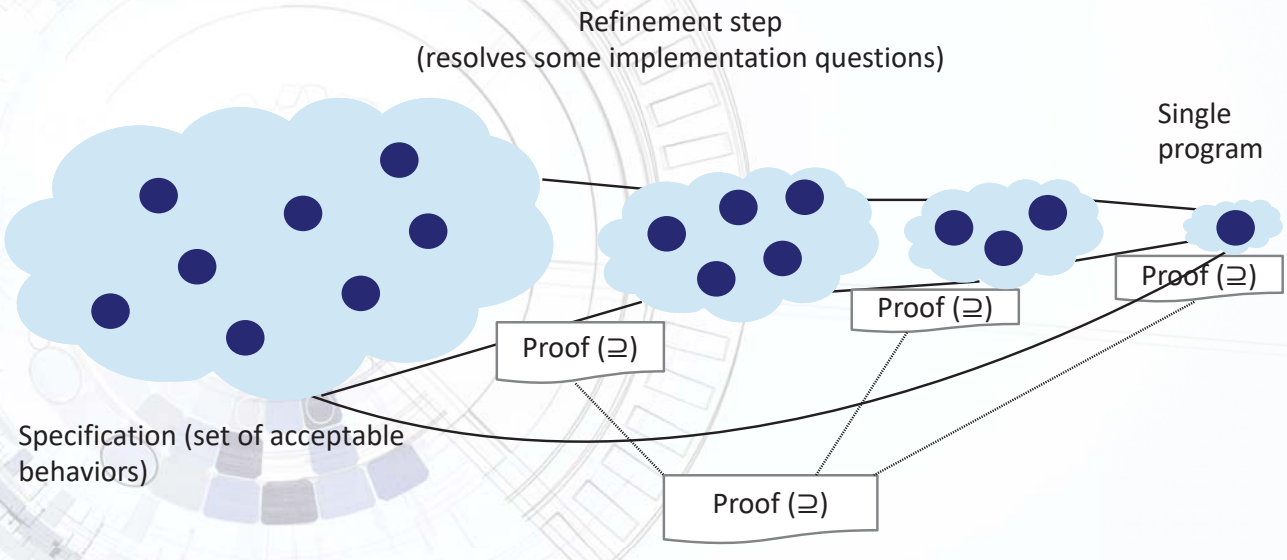




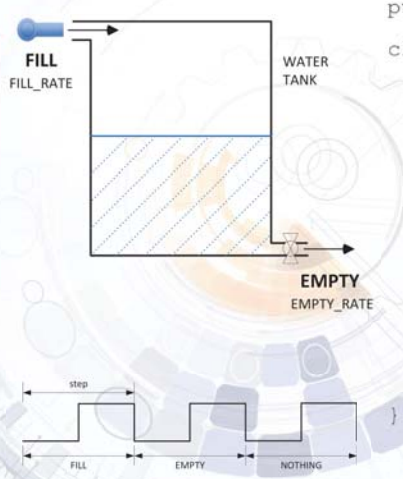
Synthesis of Monitor – Computational Attack

- Modular structure (3 parts)
- Application specification
- Application code
- Comparison of synchronized execution (observations vs. predictions)

Program derivation

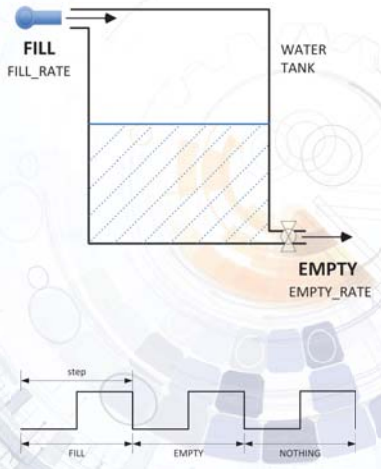


Example: Water tank (Spec)




```
public enum Action { NOTHING, FILL, EMPTY }

class WaterTankSpec {
    private int water_level = 0;

    public void newSensorReading(int reading) {
        if (abs(reading - water_level) > SENSOR_ACCURACY)
            water_level = {n | True};
    }

    public Action timestep(int target_level) {
        Action act = {a | (a = FILL → water_level + FILL_RATE ≤ TANK_MAX)
                        ∧ (a = EMPTY → water_level - EMPTY_RATE ≥ 0)};
        if (act == FILL)
            water_level += FILL_RATE;
        else if (act == EMPTY)
            water_level -= EMPTY_RATE;
        return act;
    }
}
```

Example: Water tank (Code)




```
public enum Action { NOTHING, FILL, EMPTY }

class WaterTank {
   private int water_estimate = 0;

   public void newSensorReading(int reading) {
      water_estimate = reading;
   }

   public Action timestep(int target_level) {
      if (water_estimate < target_level
      && water_estimate + SENSOR_ACCURACY + FILL_RATE < TANK_MAX) {
         water_estimate += FILL_RATE; return FILL;
      } else if (water_estimate > target_level
         && water_estimate - SENSOR_ACCURACY - EMPTY_RATE >= 0) {
         water_estimate -= EMPTY_RATE; return EMPTY;
      } else
         return NOTHING;
   }
}
```

False Data Injection

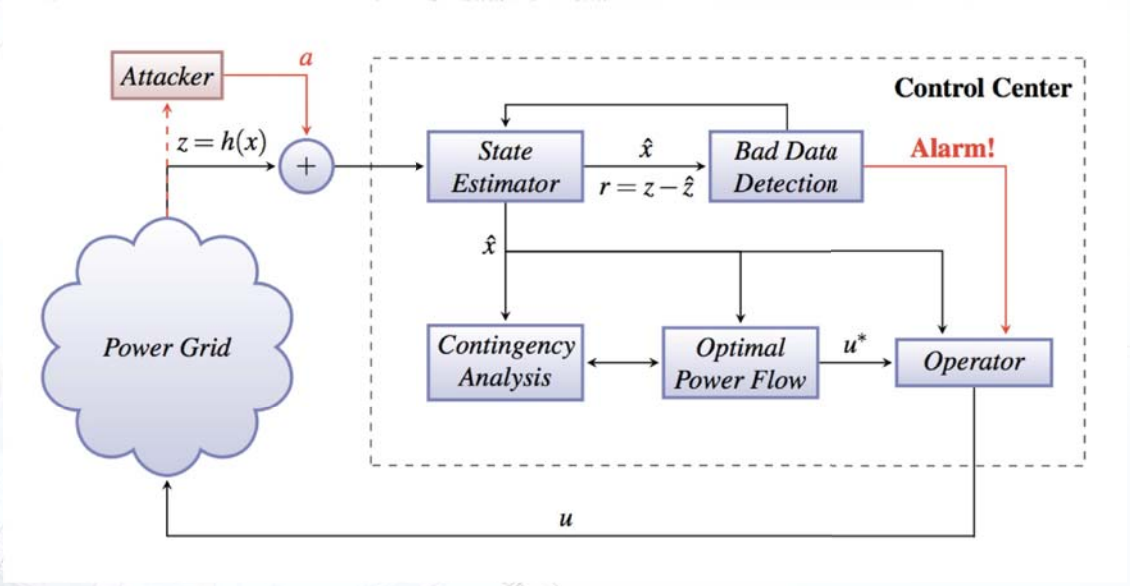
- False data injection attacks and failures
  - Feed fake/faulty measurement data to the system
  - Avoid being detected as bad data
  - Mislead controllers
  - Attacks can be local (each control unit) or global (the whole control network)

- Defense: methods for data estimation formalizing
  - Plants, sensors, actuators, channels, control software
  - Attacks, defenses, detection

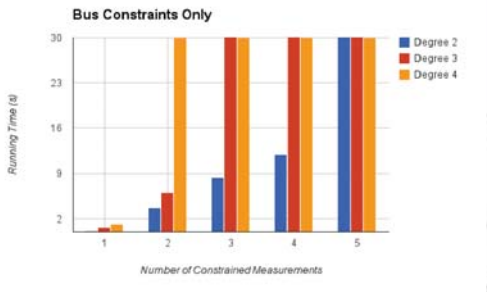
Threat Analysis for False Data Injection

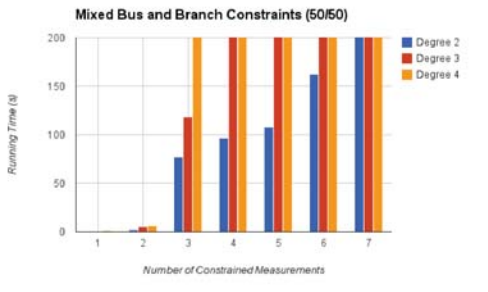
- Assumption
  - Process P(x)
  - There is a monitor mon(x,y) [x= process variables, y= measurements]
- Write satisfiability expression for process
  - FDI(y)= There_exists x : pass_monitor(x,y) AND NOT correct_reading(x,y)
  - Solve for satisfiability of FDI(y)
    - IF FDI(y) is satisfiable with injected values, THEN there exists attack
- Available tool today: dReal

Example: Power grid state estimation



Benchmarks


Bus Constraints Only
Degree 2
Degree 3
Degree 4
Running Time (s)
30
23
16
9
2
1
2
3
4
5
Number of Constrained Measurements


Mixed Bus and Branch Constraints (50/50)
Degree 2
Degree 3
Degree 4
Running Time (s)
200
150
100
50
0
1
2
3
4
5
6
7
Number of Constrained Measurements

Syntesis of Monitor for False Data Injection

- Executable model
- Constraints on measurements
- Runtime verification for consistency
- Part of ARMET

Conclusions – Future Directions

- ICS safety and security are extremely challenging
  - New threats (FDI)
  - Process-dependence
- Need for formal models of everything…
  - Devices, systems, hardware/software, processes
- Vulnerability analysis for specs
- Monitor for all theats and failures
- Behavioral analysis addresses needs and process-dependence
- Significant challenges to defenses due to
  - Methods are process-dependent
  - Formal models for processes are challenging (e.g. consider reverse osmosis plant)
  - Automation of monitor synthesis

Embedded Systems Modeling, Analysis and Automatic Code Generation with AADL and RAMSES (Hands-on Tutorial)

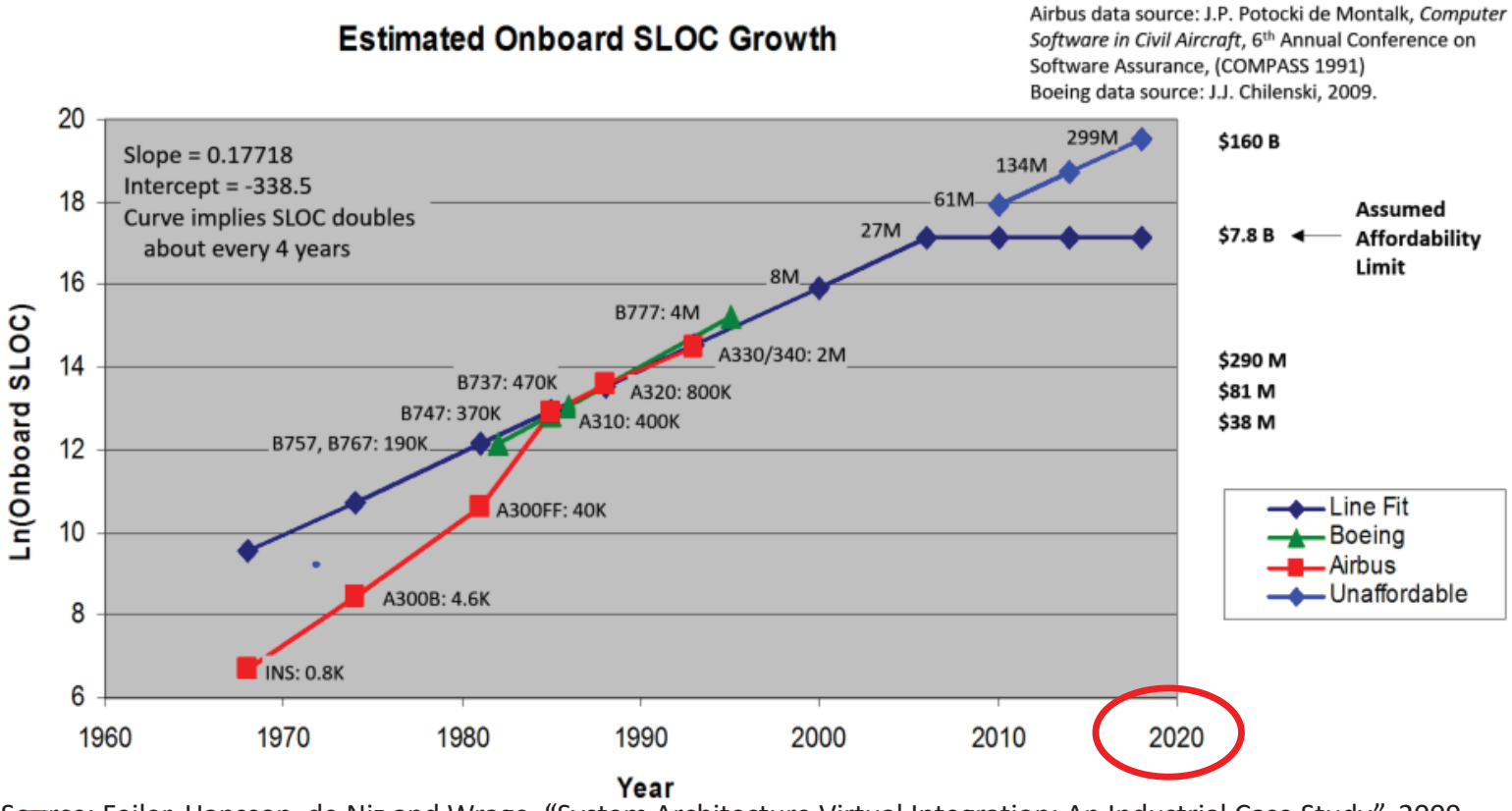
CPS&IoT'2022 Summer School

Dominique Blouin, Anish Bhobe and Etienne Borde
LTCI Lab, Telecom Paris
Institut Polytechnique de Paris, France
dominique.blouin@telecom-paris.fr
anish.bhobe@ip-paris.fr
etienne.borde@telecom-paris.fr

Outline

- Model-Based Engineering
- Introduction to AADL
- Timing Analyses with AADL
- Automatic Code Generation with RAMSES
- Introduction of the Hands_on Exercise

Increasing Systems Complexity and Unaffordable Development Costs



Source: Feiler, Hansson, de Niz and Wrage. "System Architecture Virtual Integration: An Industrial Case Study", 2009.

Non-Linear Development Effort Increase: F35 versus F16 Example

- **F35 SLOC / F16 SLOC ~ 175**
- **F35 Effort / F16 Effort ~ 300**
  - Source: SAVI Project (https://savi.avsi.aero/)

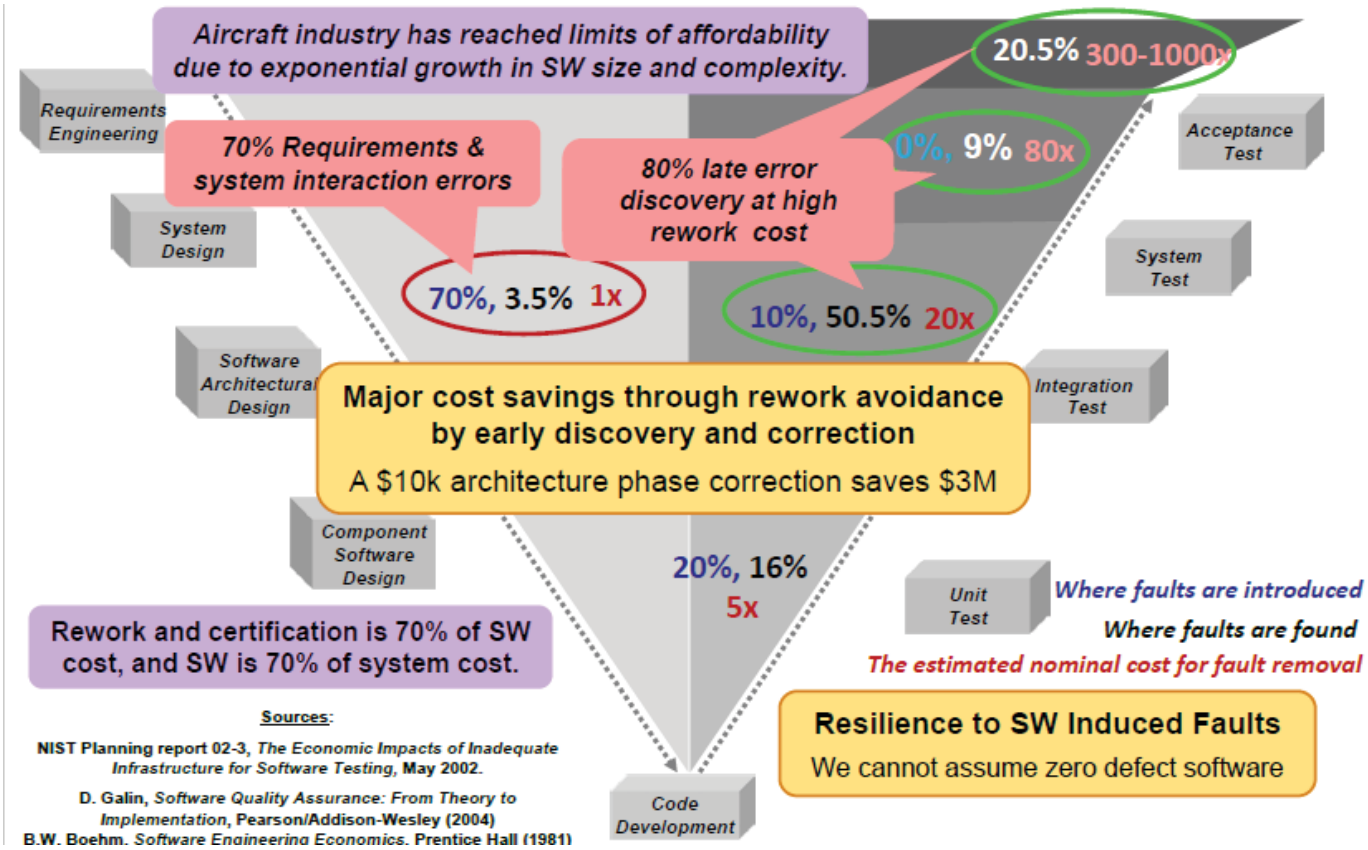
A400M:

Over 10 years delayed (2013)

6.2 billion euros over budget (30% overrun)

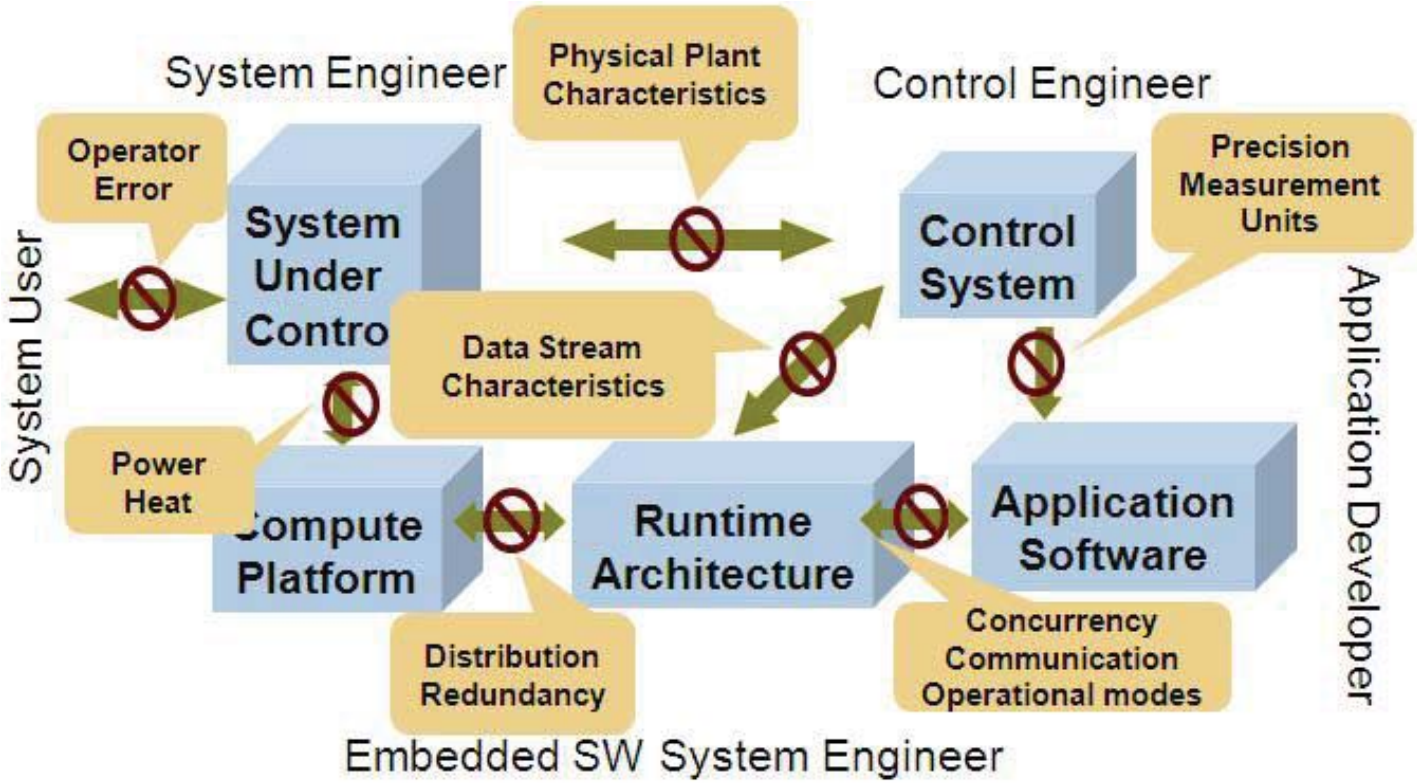
Source: https://www.rt.com/business/airbus-a400m-france-delays-561/

Costs Origins: Errors Introduced Early



Source: P. Feiler and J. Delange, "Design and Analysis of Cyber-Physical Systems: AADL and Avionics Systems", 2013

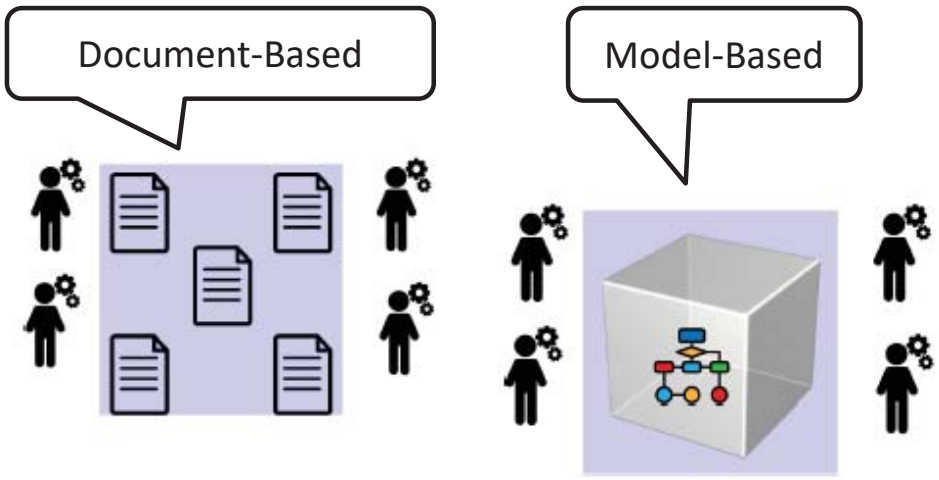
Mismatched Assumptions in Collaborative Engineering



Source: P. Feiler, "Model-based validation of safety-critical embedded systems", 2010

New Paradigm: Model-Based Engineering (MBE)

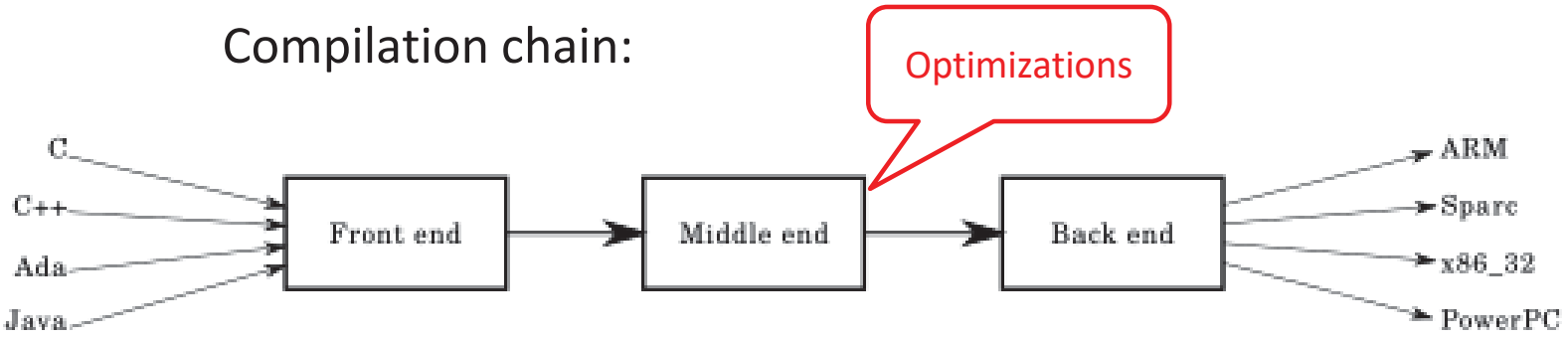
Paradigm shift: From natural language documents to models



Provide common vocabulary
Enforce more precision
Allow building tools to process specifications (models)
Allow detecting errors / inconsistencies *early*
Quite effective for avionics development (> 25 % costs reduction)

Analogy with Code Compilation

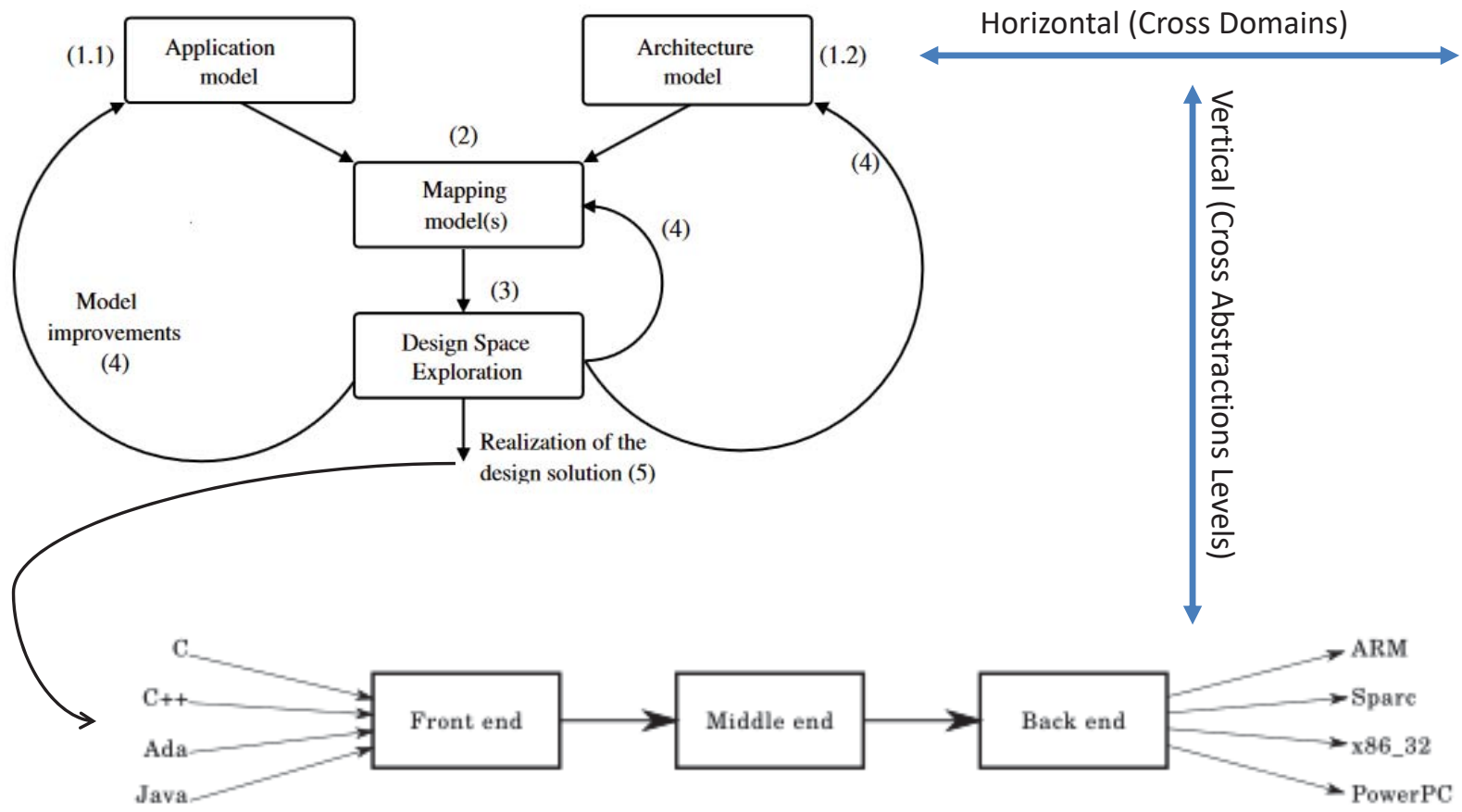
Compilation chain:



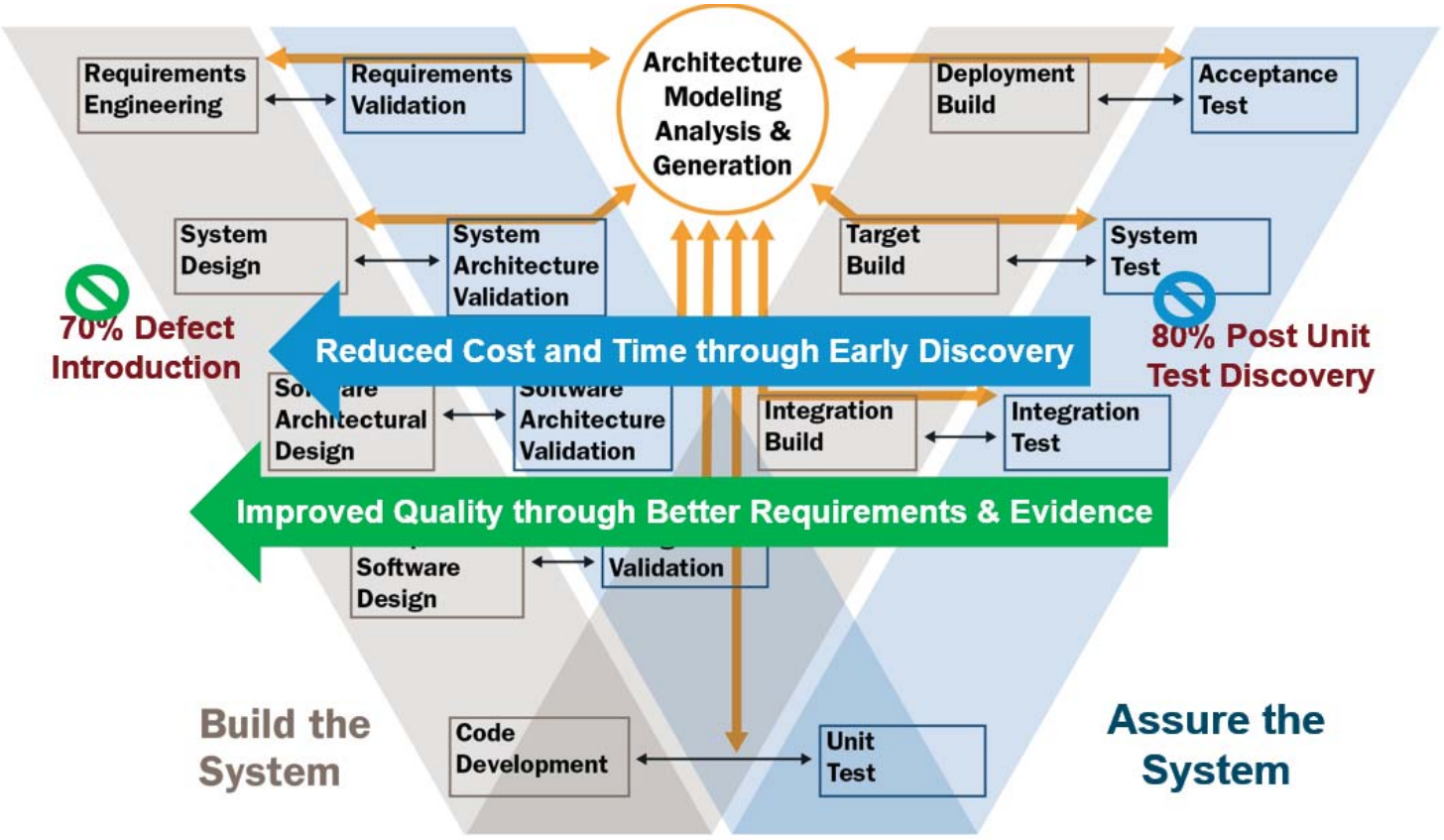
Key points:

- Increase level of abstraction
- Execution platform independent

Separation of Concerns



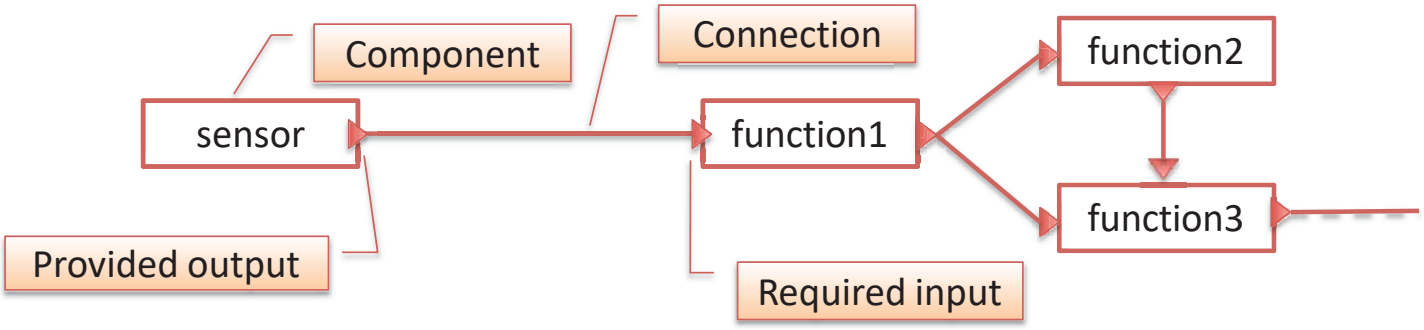
V-Cycle Model with Model-Based Virtual Integration Activities



Source: J. McGregor, P. Gluch and P. Feiler, "Analysis and Design of Safety-critical, Cyber-physical Systems", 2017.

Components-based Architecture Models

- Architecture models represent the **organisation of a computer system** as a **set of components** and their **interactions**
- Main artefacts: boxes and arrows
  - ⌘ Components : main elements of the design
  - ⌘ Interfaces : what components offer and what they need
  - ⌘ Connections : resolve components needs



- Then drawing becomes programming… or at least designing… Nothing new conceptually.
- What about the semantic?

Architecture Description Languages (ADL)

- Formal ADLs, i.e. base on a mathematically sound definition of their semantic
  - Meant to formally verify/prove expected properties of a computer system
  - Wright, Data flow graphs, state machines, …
- Domain specific ADLs
  - Meant to describes the design and implementation of computer systems constituents
  - UML 2, AUTOSAR, ***AADL***
- Abstract ADLs
  - Meant to describe the organization of a computer system without providing a precise semantics
  - ArchJava, Fractal

Note: some ADLs are standardised (e.g. UML, AADL), which provides a common understanding of the notation to the cost of slow evolutions through a committee.

Outline

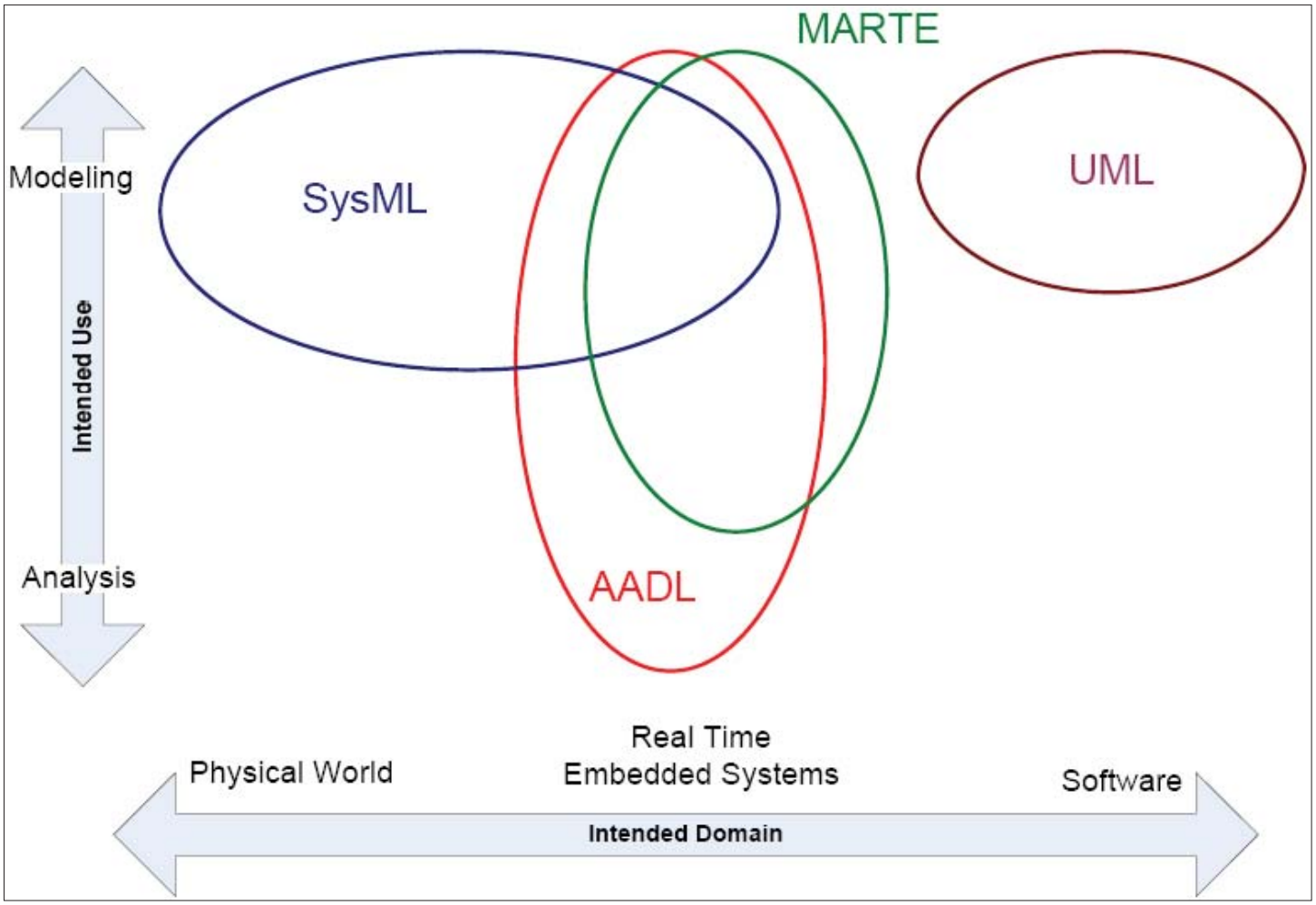
- Model-Based Engineering
- Introduction to AADL
- Timing Analyses with AADL
- Automatic Code Generation with RAMSES
- Introduction of the Hands-on Exercise

AADL (Architecture Analysis and Design Language)

- An ADL for real--time embedded systems:
  - Uses the principles of a concrete DSL (components, interfaces and connections)
  - Define properties for real--time and embedded systems analysis
    - Scheduling policy, compute execution time, latency…
    - Software components to hardware components allocation

- Objective: assist the design of such systems
  - Standardized semantics (formulated with natural language)
  - Textual and graphical syntax
  - Strongly typed (components category, composition rules, …)
  - Extendable (property definition language and annexes)

Comparison with other Architecture Description Languages (ADL)



Source: Steven P. Miller, AADL standards winter meeting, 2011

General Characteristics

- Components are the main modeling entities
- The standard defines categories of components (keywords of the language); examples of categories a component can belong to:
  - System
  - Process, thread, data…
  - Processor, memory, bus…
- Components definition is divided into types, implementations, and subcomponents
  - Type: how the component is viewed from outside (e.g. interaction interfaces)
  - Implementation: internal structure of the component (e.g. subcomponents)
  - Subcomponents: instances of components, starting from a root system implementation
- Characteristics of components are structured into sections (e.g. **features, properties, subcomponents**) identified by keywords of the language
- Details:
  - Components can be declared in any order
  - The language is case insensitive

Language Constructs: Running Example

- Initial plan: a really complex system

- Actual case-study
  - Smaller but still representative

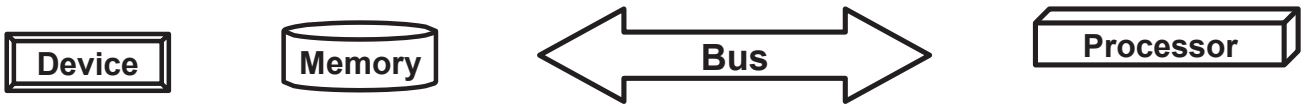
System Level Viewpoint Categories

- Two categories: system and abstract

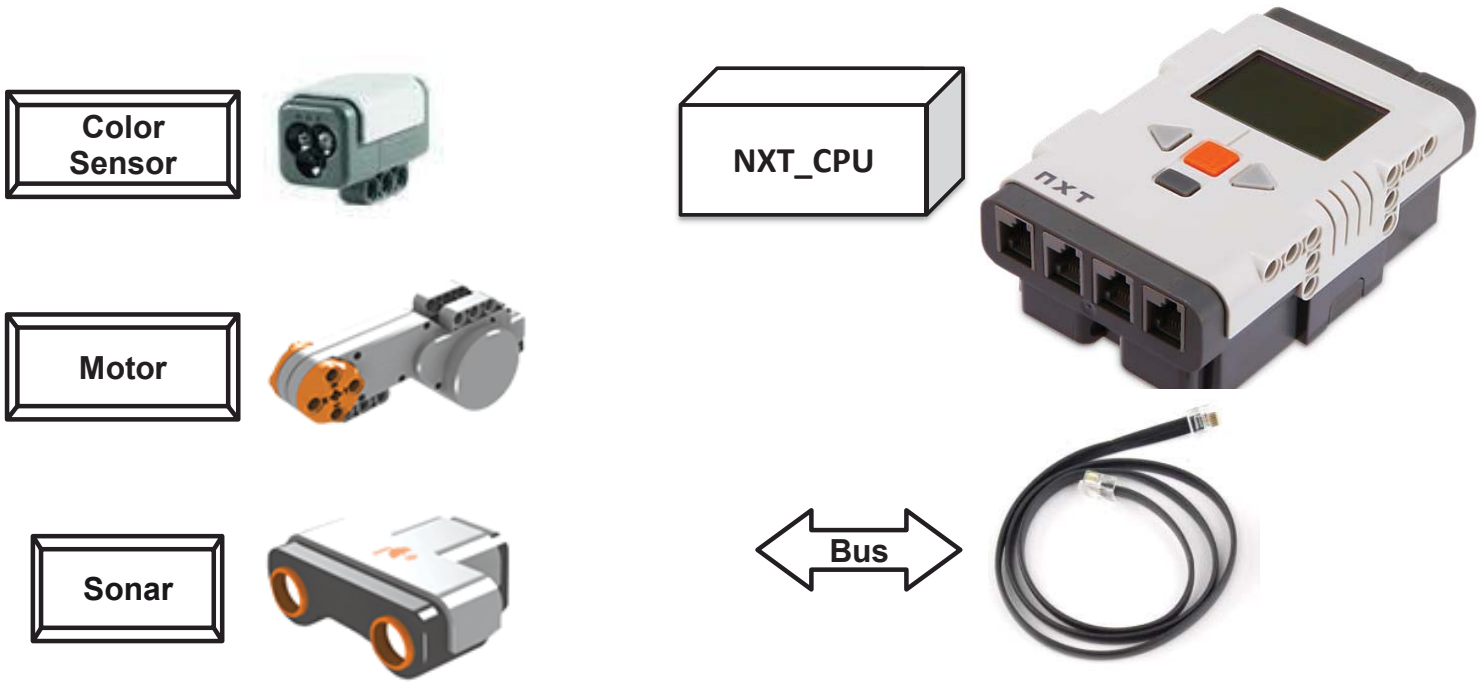
- Different possible objectives
  - Represent, from a very abstract view point, the main constituent of the system, their interfaces and connections.
  - System:
    - aggregate, by composition, subcomponents describing the execution platform and subcomponents describing the software architecture.
    - Define the main operational modes of the system
  - Abstract:
    - Define structure and interaction without knowing yet the nature of the component

Execution Platform Viewpoint Categories

- Instances of components categories dedicated to the execution platform description
  - ⌘ processor : hardware computation unit + tasks scheduling capabilities
  - ⌘ memory : storage component (may be RAM, hard disk drive, cache, etc.)
  - ⌘ bus : physical communication link (network cable, etc.)
  - ⌘ device : interface with the physical environment of the system (sensors/ actuators)



Example Execution Platform Components



Software Architecture Viewpoint Categories

- AADL component categories for software
  - Data: information that can be exchanged among software components
  - Subprogram: sequentially executable software, like functions in C programming language
  - Thread: task executing a sequence of functions
  - Process: memory address space allocated for the execution of its thread subcomponents

- These categories focus on operating system and programming elements

Example of Software Components

```
data Light
end Light;
```

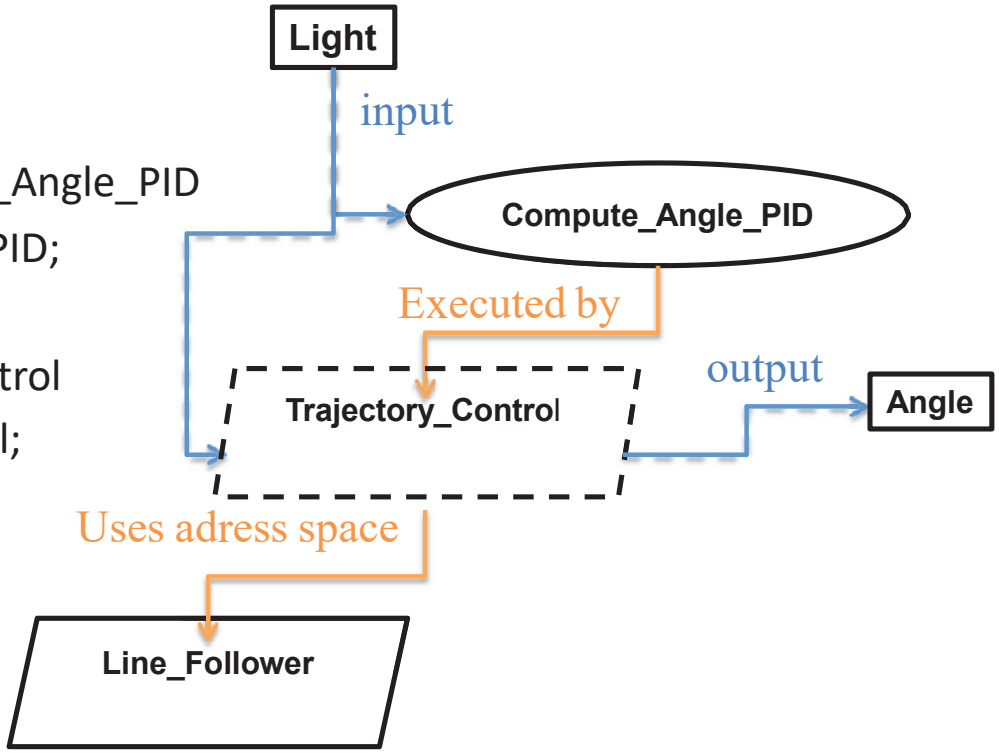

```
subprogram Compute_Angle_PID
end Compute_Angle_PID;
```


```
thread Trajectory_Control
end Trajectory_Control;
```


```
process Line_Follower
end Line_Follower
```

Example of software components

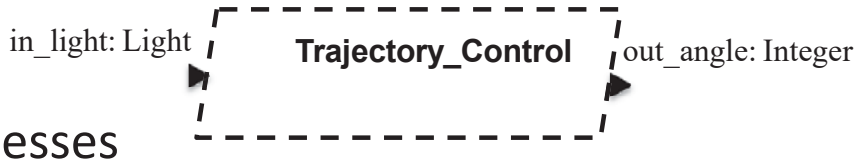
```
data Light
end Light;

subprogram Compute_Angle_PID
end Compute_Angle_PID;

thread Trajectory_Control
end Trajectory_Control;

process Line_Follower
end Line_Follower
```



How to represent these interactions and allocations in AADL?

First, Define Component Interfaces (Component Types in AADL)

- Parameters
  - **in**, **out**, or **inout**
  - Usable for subprograms
- Requires or provides data access
  - Usable for subprograms and threads
- Ports
  - **in**, **out**, or **in out**
  - **Data**, **Event** or **Event Data**
  - Usable for threads and processes



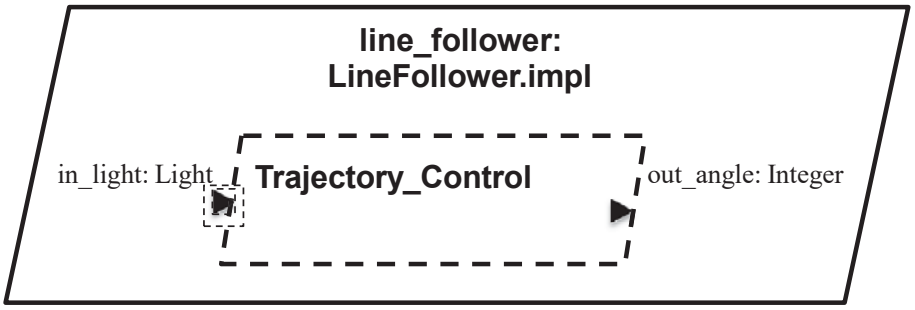
Semantical Differences among Features of Software Components

- Data port versus Event Data Port
  - **data port :** single value shared among components (no queueing).
  - **event** or **event data port** : multiple values queued.

- Data Port versus Data access
  - Data access allows access to the data at anytime during the execution of a task / subprogram
  - Data port defines the following semantics:
    - Data becomes available on an input port when the thread starts its execution. Data not used in the previous execution of the thread is lost. Data is not updated during the execution of the task.
    - Data produced on an output port are sent to the recipient port at the end of the producer task.

Second, need to compose

- Thread subcomponents in process implementation:




```
process implementation Proc.simple
   subcomponents
       C_Th   : thread ContrThread.Impl;
end Proc.simple;
```

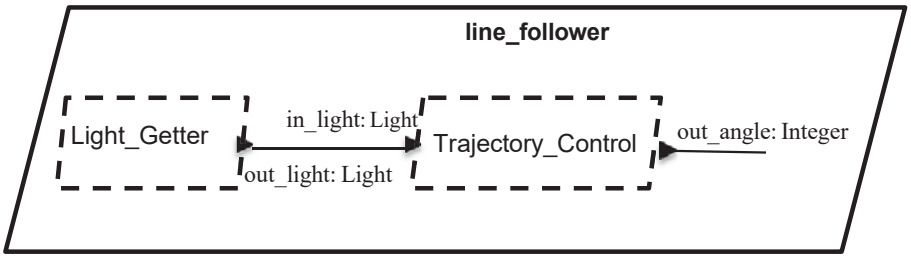
- Subprogram calls in call sequences

```
thread implementation ContrThread.Impl
calls
   call1:
   {
     cp   : subprogram  computePID;
     cs   : subprogram computeSpeed;
   };
end ContrThread.Impl;
```

Sequence of subprogram calls

Third, need to connect

- Components features are connected hierarchically
  - ⌘ Thread subcomponents in the process




```
process implementation Proc.simple
  subcomponents
    C_Th   : thread ContrThread.Impl;
  connections
    c2: port Bg_Th.out_light -> C_Th.in_light;
end Proc.simple;
```

  - ⌘ Subprogram calls in the thread but also the thread features with the subprogram calls

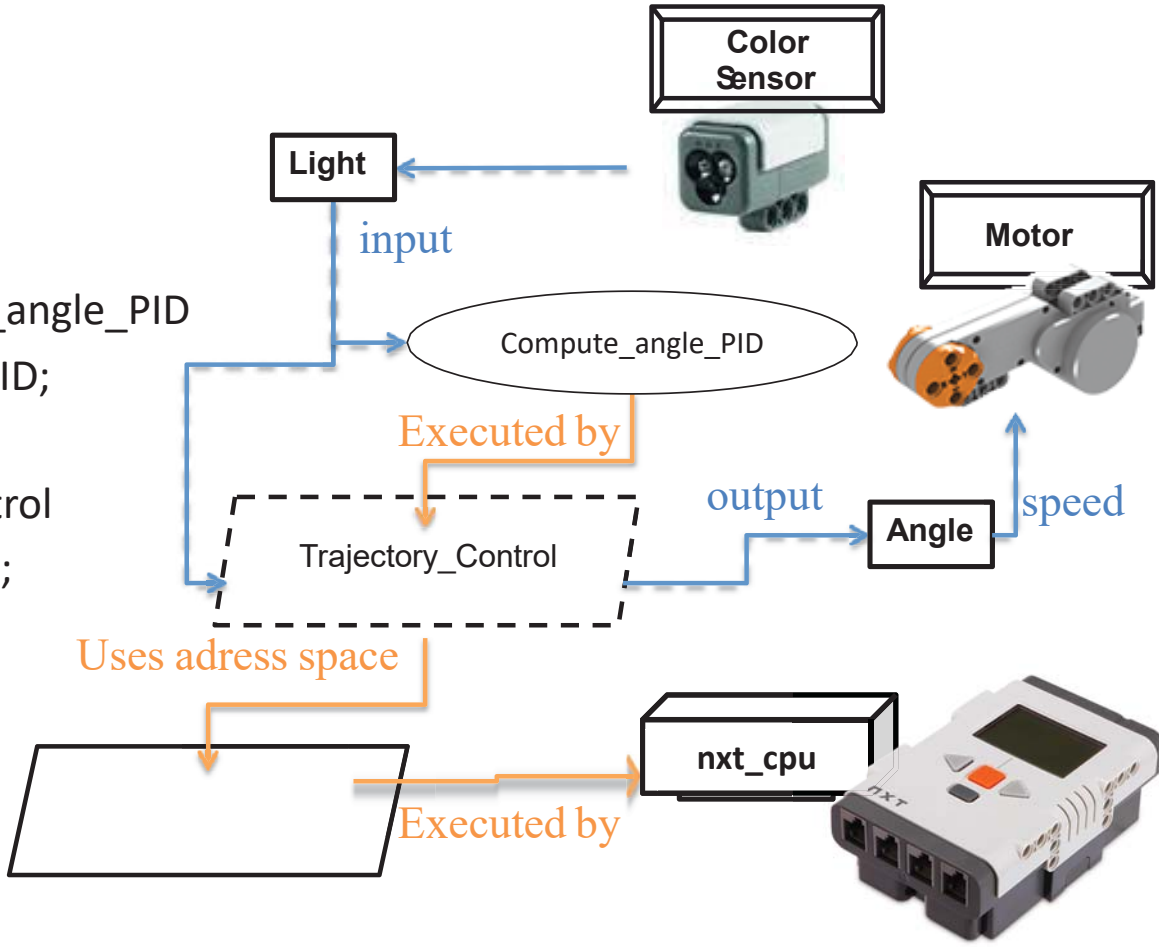
```
thread implementation ContrThread.Impl
calls
  call1:
  {
    cp   : subprogram computePID;
    cs   : subprogram computeSpeed;
  };
connections
  cc0 : parameter cp.currentLight -> in_light;
  cc1 : parameter cp.angle -> cs.angle;
end ContrThread.Impl;
```

Value received on the input port is passed as a parameter to the subprogram

Value produced by computePID is passed to compute Speed

Combining Execution Platform and Software Architecture Viewpoints

```
data Light
end Light;

subprogram compute_angle_PID
end compute_angle_PID;

thread trajectory_control
end trajectory_control;

process line_follower
end line_follower
```



A few words about properties

- Properties in AADL is a way to decorate your model
  - Properties can be associated to almost any element of your model.
  - The standard defines a property language: AADL users can define their own **property sets** (means to extend the possibilities of the language).
  - The AADL standard predefines a set of properties for most common used properties in ADLs and real-time embedded systems.
- Example of property definition

```
Actual_Processor_Binding: inherit list of reference (processor,
                                                     virtual processor,
                                                     system, device)
          applies to (thread, thread group, process,
                      system, virtual processor, device);
```

- Example of property association

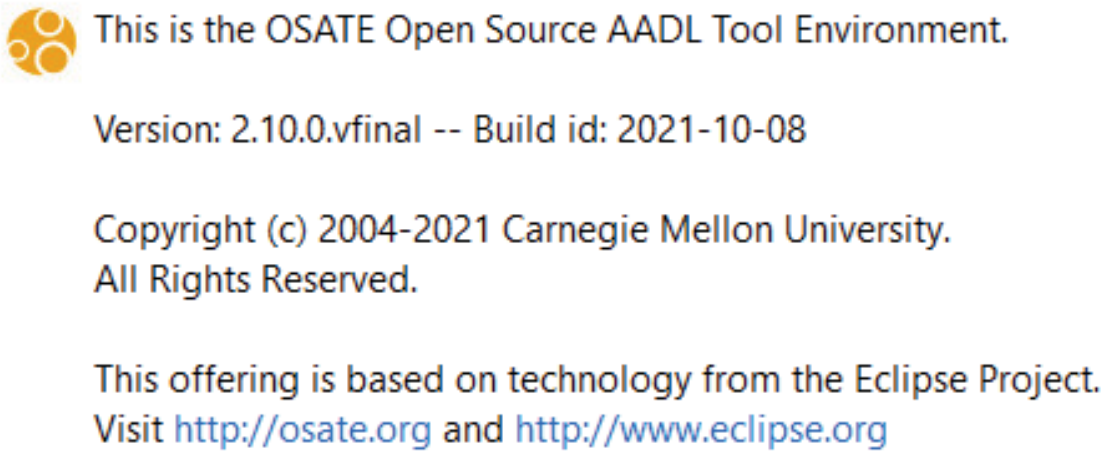
```
system implementation synchronous.others
  subcomponents
    my_platform : processor CPU;
    my_process : process my_process.impl;
  properties
    Actual_Processor_Binding => ( reference(my_platform) ) applies to my_process;
end synchronous.others;
```

Things we could not present

- AADL has much more than this
  - Lots of standardized properties
  - Modes, with mode--specific architectures (reconfiguration)
  - Flows, to analyse the worst--case latency and jitter of data
  - System/Hardware/Network configuration representation
  - Behavior as state machines (BA)
  - Errors and faults propagation (EMV2)

Open-Source AADL Tool Environment (OSATE)

- Developed at SEI (Carnegie Mellon University)
- Synchronized textual and graphical editors
- Eclipse-based (EMF, Ecore, Xtext, Graphiti, etc.)
- Actively maintained (more info at https://osate.org/)



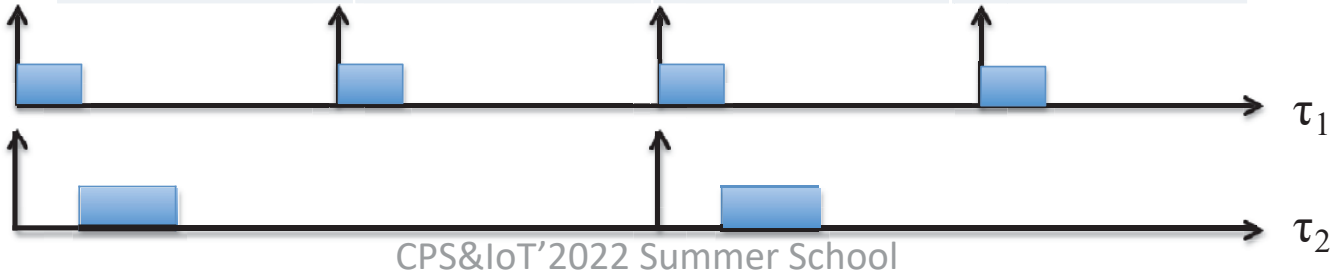
Outline

- Model-Based Engineering
- Introduction to AADL
- Timing Analyses with AADL
- Automatic Code Generation with RAMSES
- Introduction of the Hands-on Exercise

Principles of Response Time Analysis (RTA)

- Problem: the computation unit is shared among different tasks… Is that safe from a timing performance viewpoint?
  - Expressed as: tasks never miss their deadlines
- Assume a set of periodic tasks with a fixed priority scheduling configured using the Rate Monotonic Scheduling principles
  - RMS: the higher the frequency, the higher the priority
- RTA is meant to compute the worst–case response time of a task, based on:
  - The tasks period
  - The tasks worst case execution time (WCET) for the code of the task
  - The tasks deadline

Task	Period	WCET	Deadline
τ_1	10 ms	2 ms	6 ms
τ_2	20 ms	3 ms	10 ms



AADL Properties for RTA

- In terms of structure, we need
  - Threads, representing tasks
  - The processor on which threads are deployed
  - The property value assigning threads to a processor

- Scheduling properties are
  - *Dispatch_Protocol*, usually set with value ***periodic***
  - *Compute_Execution_Time* represents the execution time interval, where the upper bound is the WCET
  - The *Priority*, meaning is obvious.
  - *Deadline*, meaning is obvious.
  - *Scheduling_Protocol*, associated to the processor, defines the scheduling policy applied by the operating system running on the processor

Example (1/2)

```
PACKAGE synchronous_Pkg
PUBLIC
WITH Base_Types;

SYSTEM synchronous
END synchronous;

SYSTEM IMPLEMENTATION synchronous.others
  SUBCOMPONENTS
    my_platform : PROCESSOR CPU;
    my_process : PROCESS my_process.impl;
PROPERTIES
Actual_Processor_Binding =>
          ( reference(my_platform) )
            applies to my_process;
END synchronous.others;
```

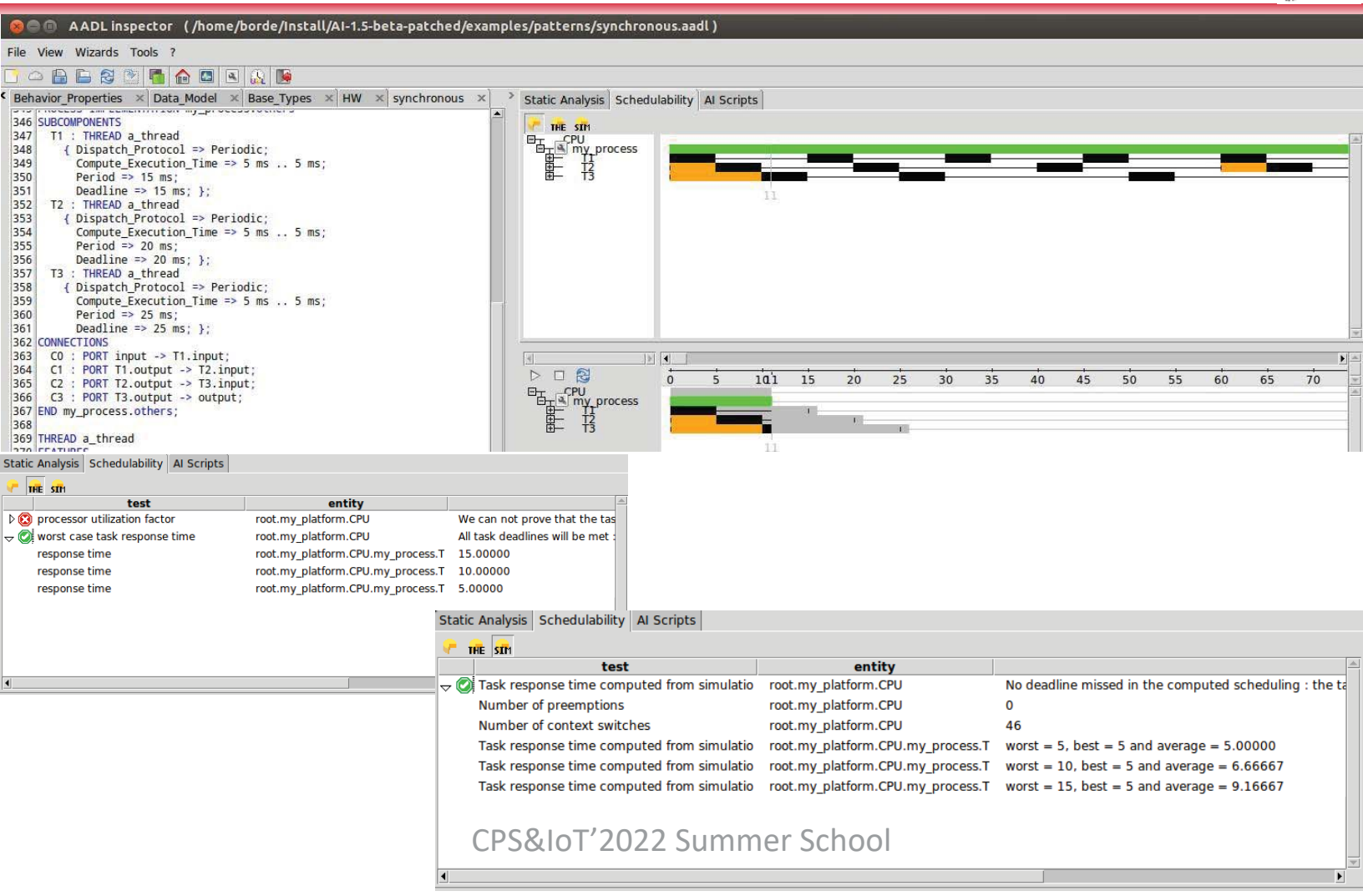
```
PROCESSOR CPU
PROPERTIES
    Scheduling_Protocol => (RMS);
END CPU;


PROCESS my_process
END my_process;
```

Example (2/2)

```
PROCESS IMPLEMENTATION my_process.impl
SUBCOMPONENTS
T1 : THREAD a_thread
   { Dispatch_Protocol => Periodic;
     Compute_Execution_Time=>5 ms..5 ms;
     Period => 15 ms;
     Deadline => 15 ms; };
T2 : THREAD a_thread
   { Dispatch_Protocol => Periodic;
     Compute_Execution_Time=>5 ms..5 ms;
     Period => 20 ms;
     Deadline => 20 ms; };
T3 : THREAD a_thread
   { Dispatch_Protocol => Periodic;
     Compute_Execution_Time=>5 ms..5 ms;
     Period => 25 ms;
     Deadline => 25 ms; };
END my_process.impl;
```

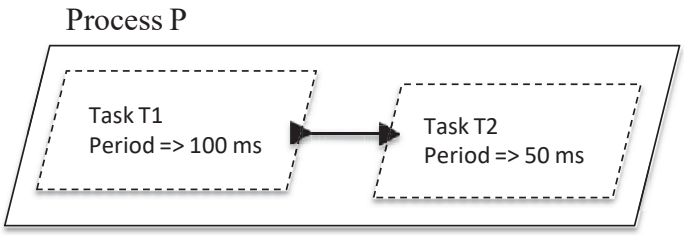
```
THREAD a_thread
END a_thread;


END synchronous_Pkg;
```

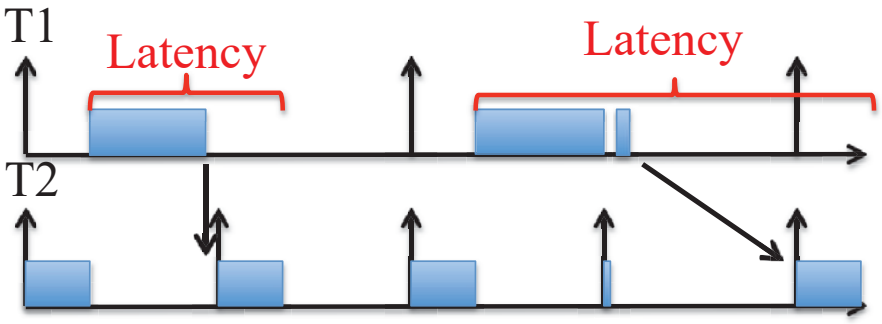
AADL Inspector



Communications through Data Ports

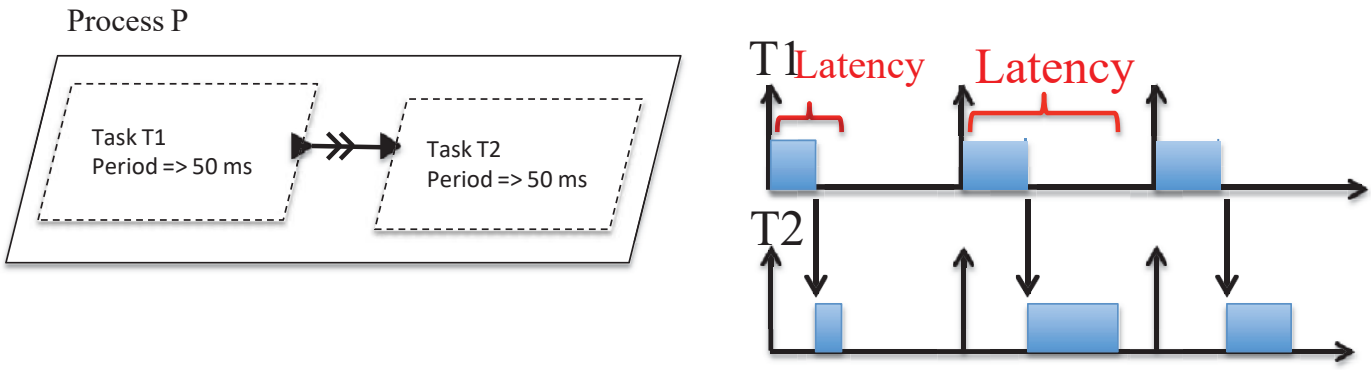
- Configured with the Timing property; possible values:
  - ⌘ Sampled (default): similar to a shared variable, except for the read/execute/write semantics
    - o Advantage: simplicity
    - o Disadvantage: undeterministic





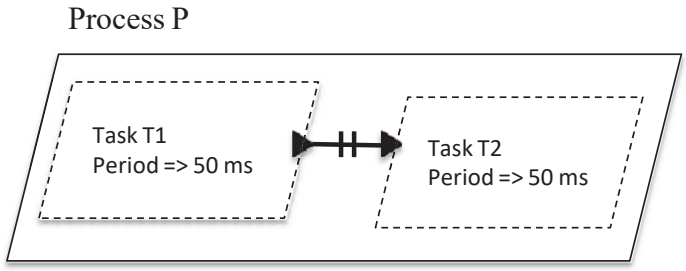
Communications through Data Ports

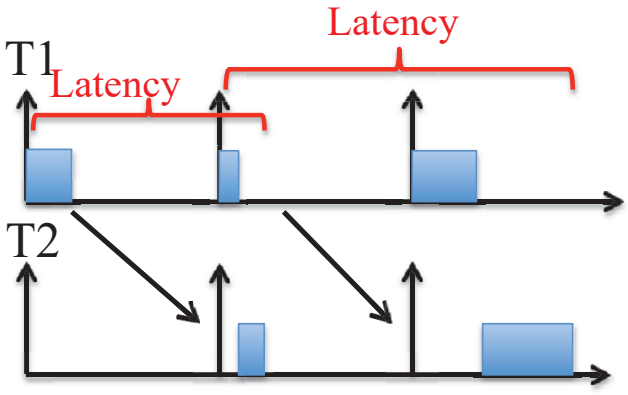
- Configured with the Timing property; possible values:
  - ⌘ Immediate: the recipient is not supposed to start until the output port of the connected thread has been updated
    - o Advantages: deterministic, reduces latency
    - o Disadvantages: put constraints on the scheduler and the model (no cycle, consistent periods)



Communications through Data Ports

- Configured with the Timing property; possible values:
  - ⌘ Delayed: the output port is updated at the deadline of its thread
    - o Advantages: deterministic, reduces jitter
    - o Disadvantages: increases latency

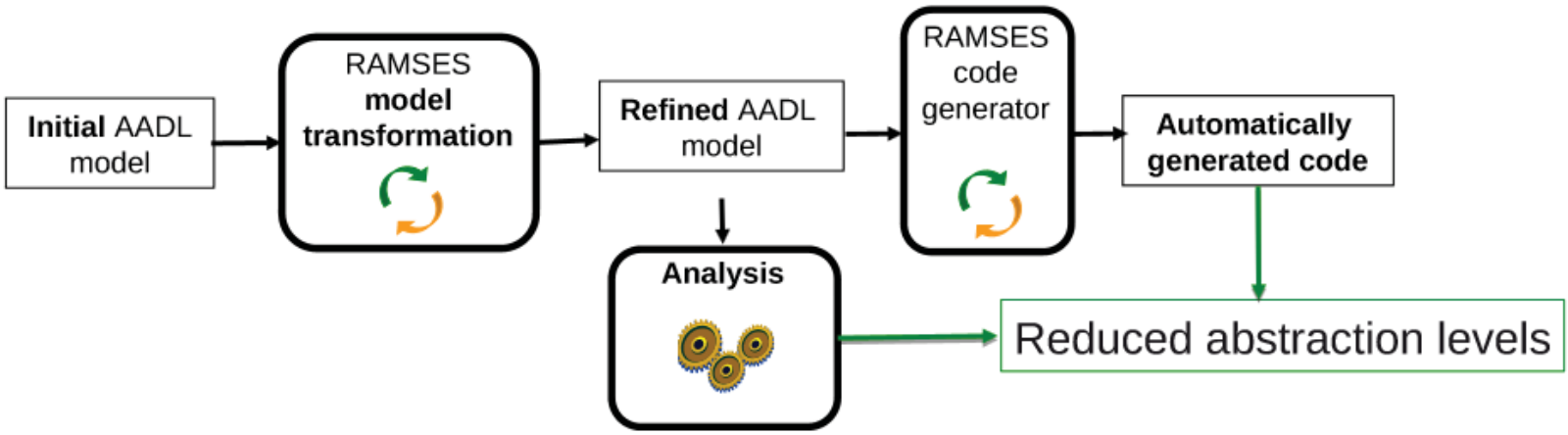




Outline

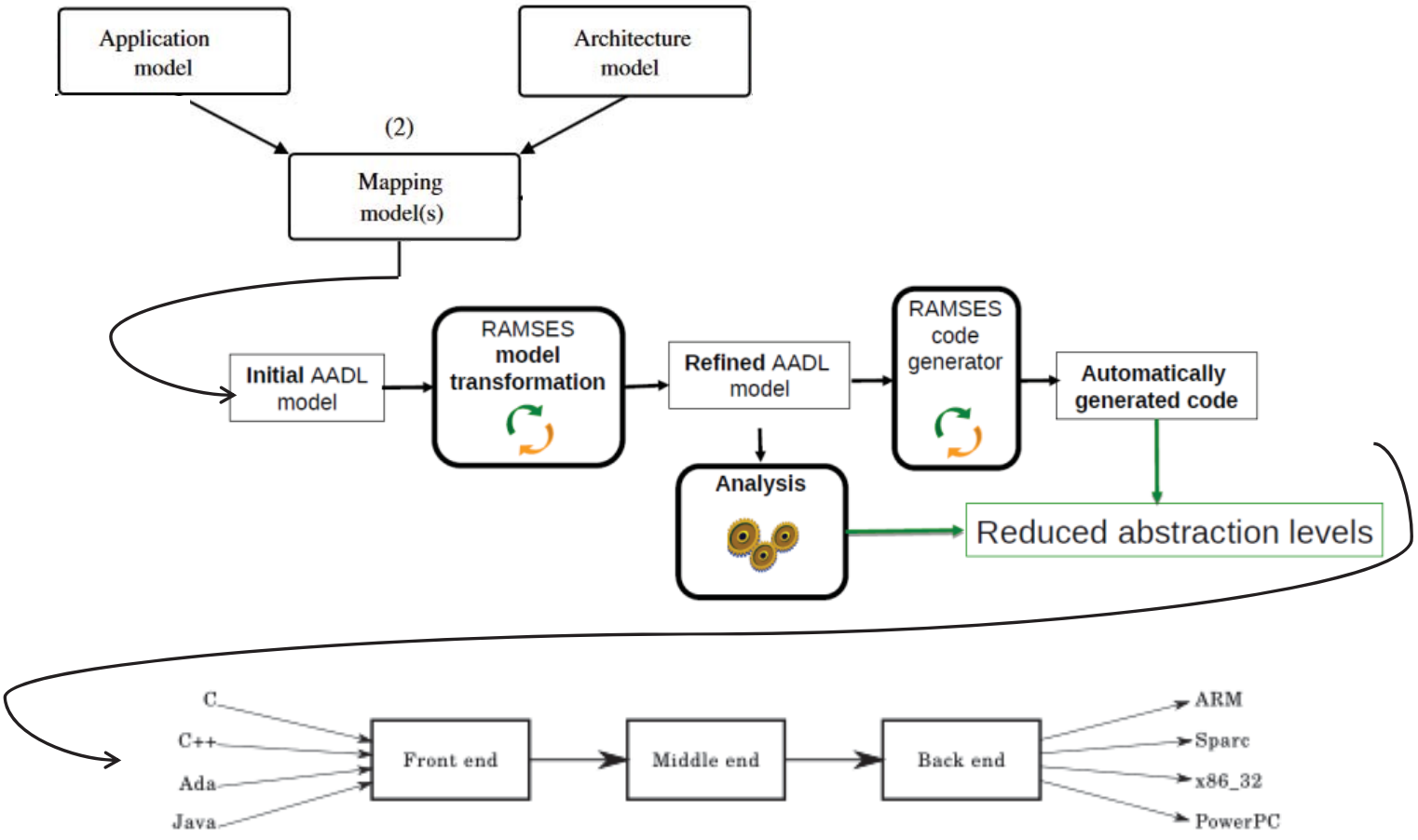
- ❑ Model-Based Engineering
- ❑ Introduction to AADL
- ❑ Timing Analyses with AADL
- ❑ Automatic Code Generation with RAMSES
- ❑ Introduction of the Hands-on Exercise

RAMSES (Refinement of AADL Models for Synthesis of Embedded Systems)

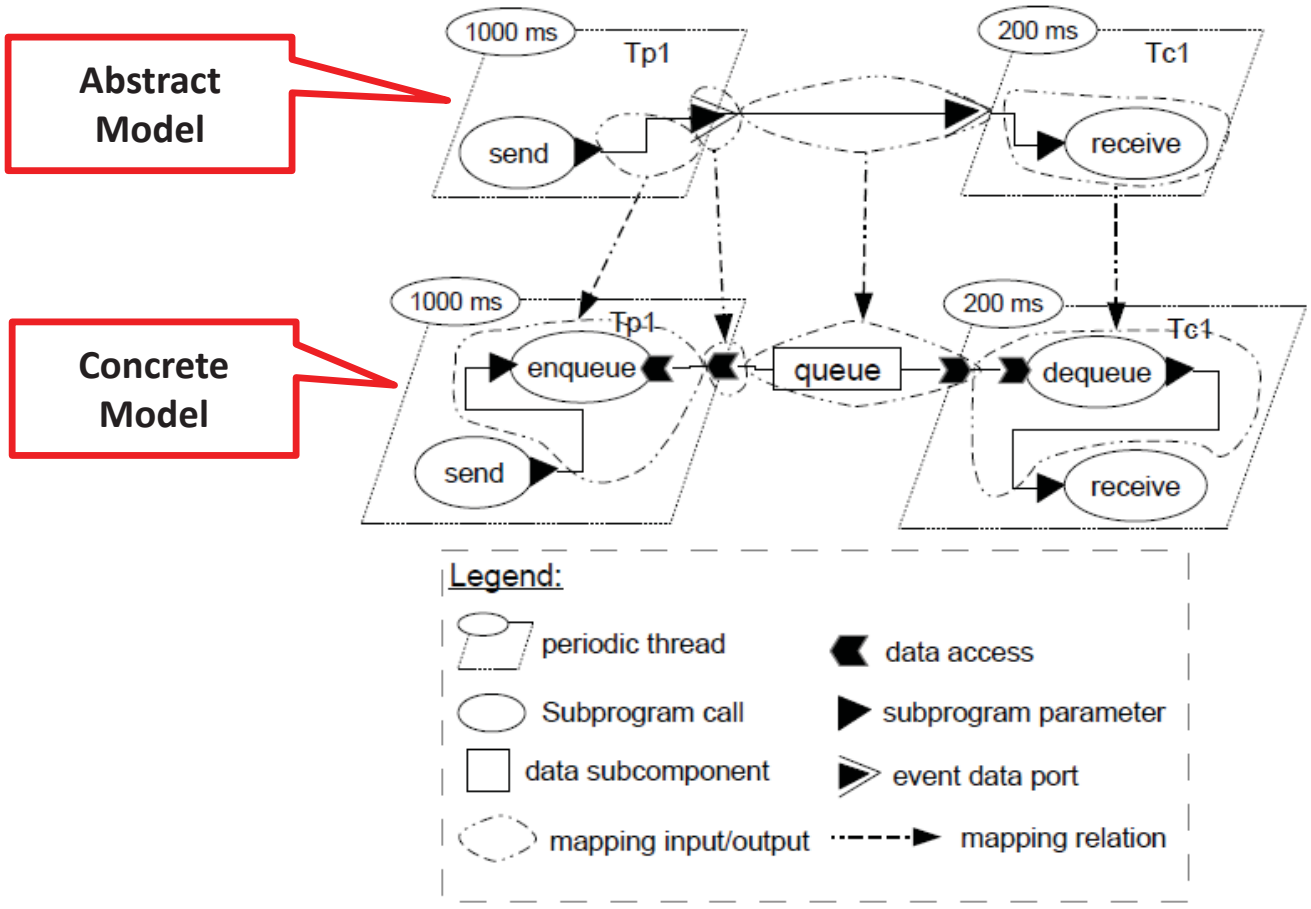


- More info at https://mem4csd.telecom-paristech.fr/

Model Refinement and Code Generation



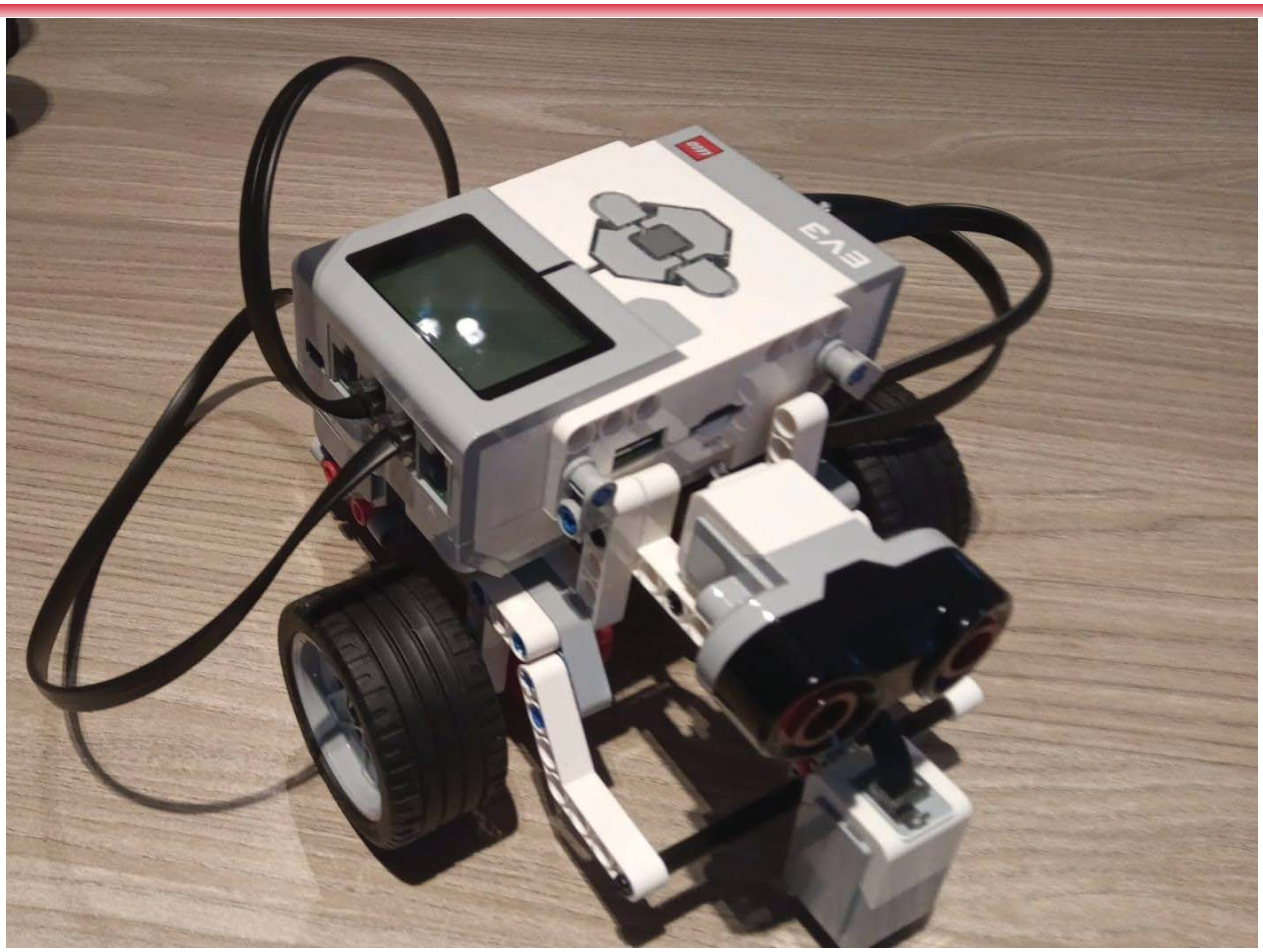
Example of RAMSES Refinement Rule



Outline

- ❑ Model-Based Engineering
- ❑ Introduction to AADL
- ❑ Timing Analyses with AADL
- ❑ Automatic Code Generation with RAMSES
- ❑ Introduction of the Hands-on Exercise

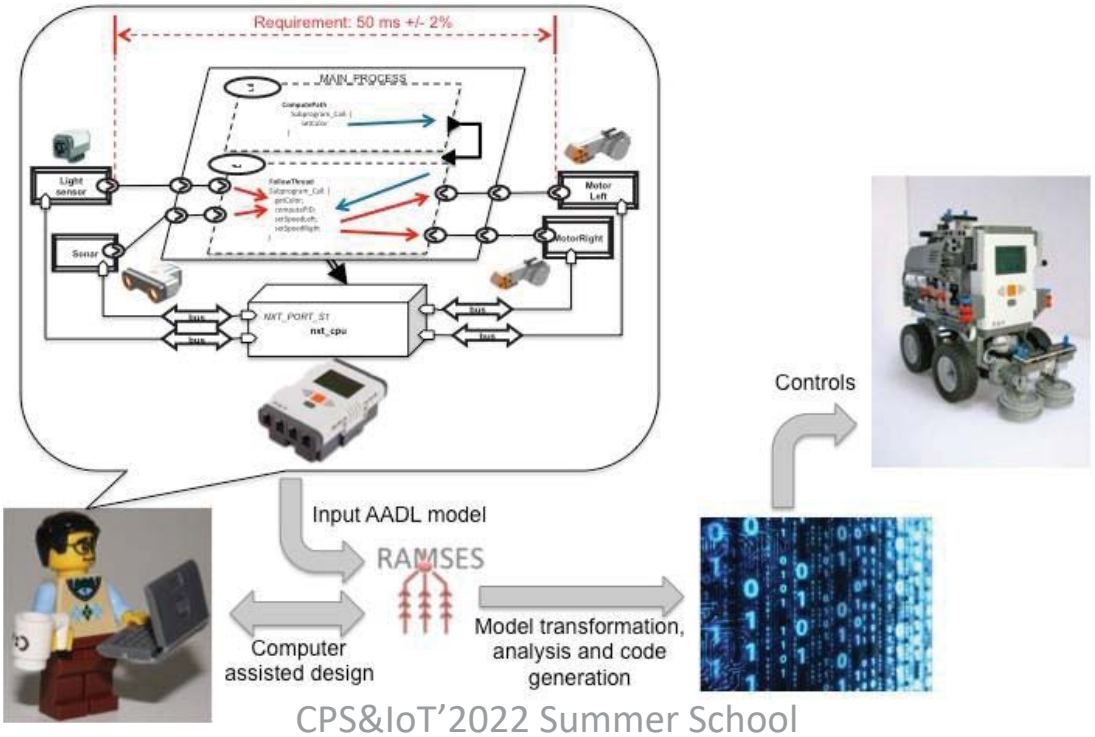
Case Study: EV3 Dev Minstorm Lego Robot



CPS&IoT'2022 Summer School

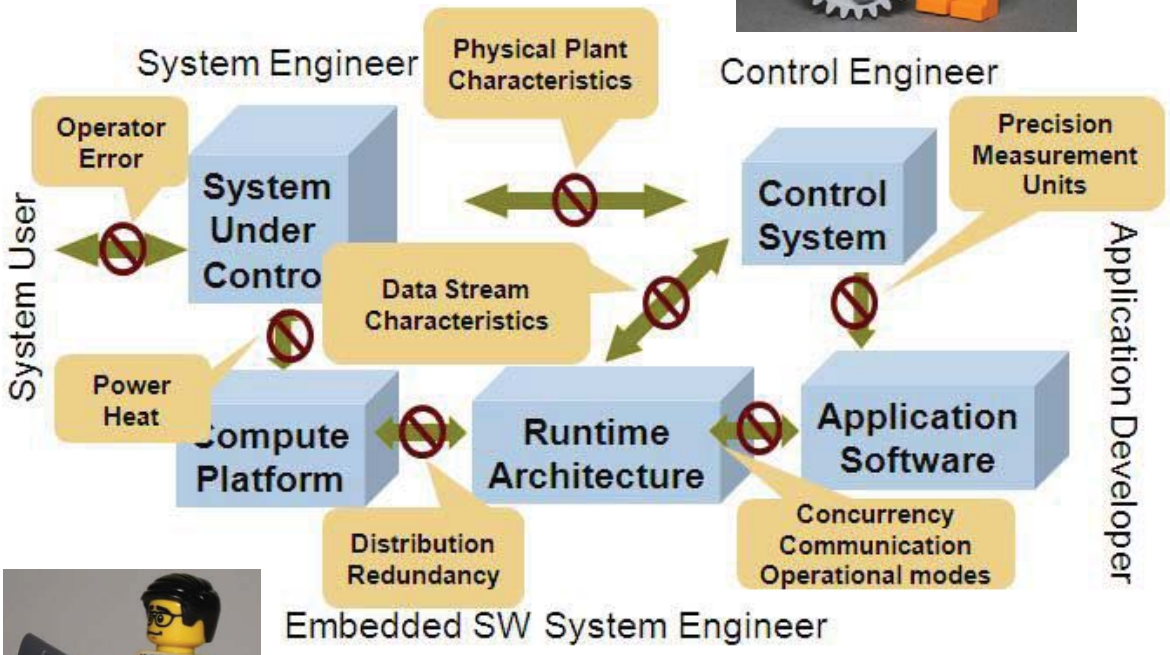
Objectives of the Hands-on Exercise

Understand how to model resource sharing in order to
1 – ensure it does not jeopardize the program execution because of performance issues
2 – define timing properties of communicating tasks
3 – integrate new functions such as an obstacle detection task

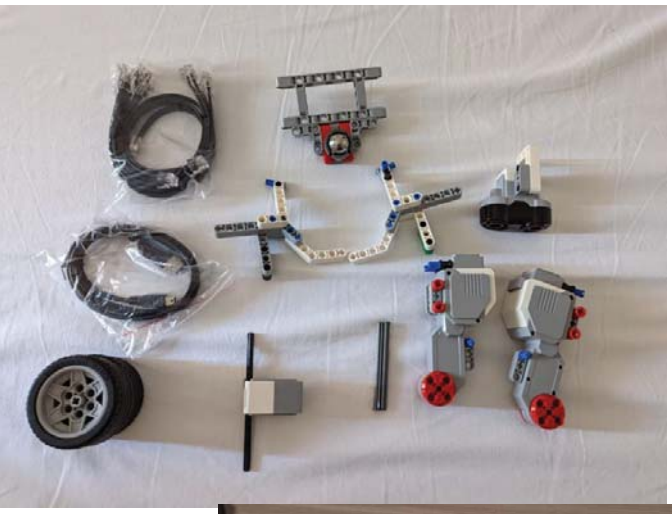


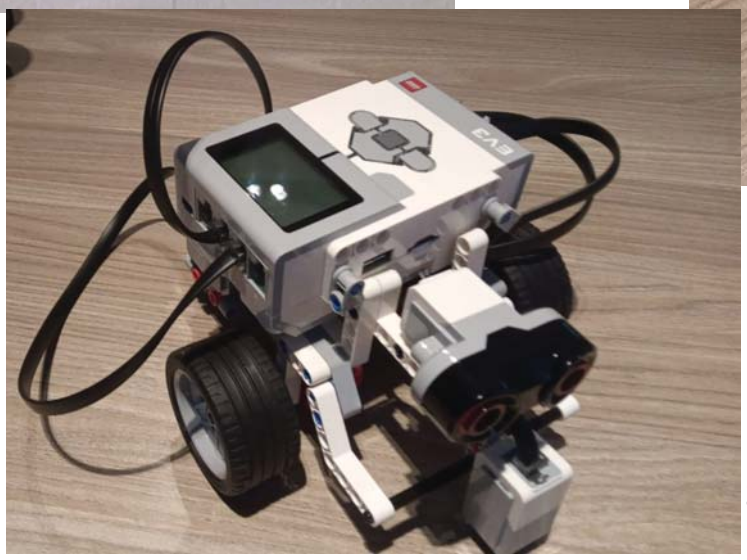
Engineering Teams: Two Domains

Mechanical Engineer



Mechanical Engineering Task: Assemble the Robot





School

Embedded Software Engineering Task: Develop the Embedded Software

- Follow instructions at
- https://mem4csd.telecom-paristech.fr/blog/index.php/training-schools/cps-iot-summer-school-2022/

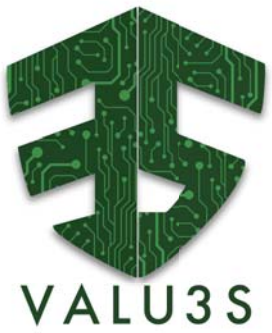
Making Teams

- 11 robots ➔ How many students per team?
- Requires Linux
- Provide a Virtual machine and executable to install VMWare
  - Password ramses

Thanks for your attention

Questions ???

CPS&IoT'2022 Summer School



How to design and tailer a perfect fitting verification and validation process for your CPS&IoT project

Thomas Bauer, Rupert Schlick, David Fürcho, Joseba Agirre, Bob Hruska, David Pereira, Jose Proença, Robert Sicher, Ales Smrcka, Ugur Yayan, Bernd Bredehorst, Christoph Schmittner, Behrooz Sangchoolie

thomas.bauer@iese.fraunhofer.de

09-06-2022 | 3rd Summer School on Cyber Physical Systems and Internet of Things 2022

Public

ECSEL Joint Undertaking

This project has received funding from the ECSEL Joint Undertaking (JU) under grant agreement No 876852. The JU receives support from the European Union's Horizon 2020 research and innovation programme and Austria, Czech Republic, Germany, Ireland, Italy, Portugal, Spain, Sweden, Turkey.

Agenda

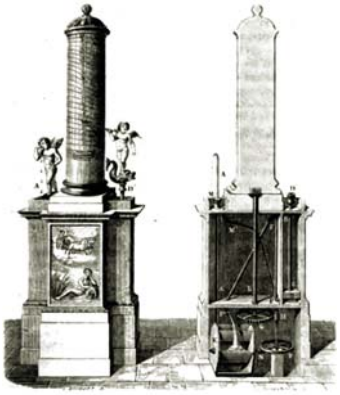
- Motivation
- Project Overview and Objectives
- Project Structure and Assets
- Use case and demonstrators
- V&V Framework
- V&V Methods
- V&V Tools+Workflows
- Standardization
- Closing
- Questions

Motivation

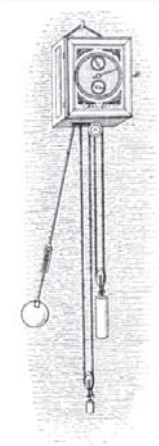
Motivation

- **Automated systems** have been around for so long

Ancient Persian clock
(16th century BC)



Ctesibius' water clock
(2nd century BC)



Christiaan Huygens'
pendulum clock
(1656)

Synchronous electric clocks
(1930s)

Digital clocks

Motivation

- **Automation** is heavily used in **safety-critical** systems









Motivation

- **Functionality** has been in the centre of attention



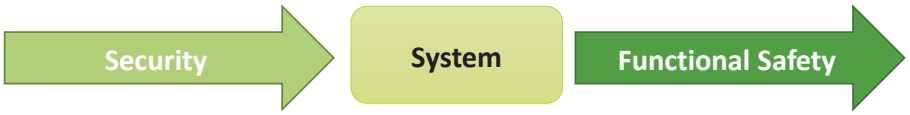






Motivation

- **Functionality** has been in the centre of attention
  - With rising complexity, unknown emerging properties of the system may come to the surface making it necessary to conduct thorough **verification** and **validation** of these systems.
  - To be introduced to the market, automated systems need to also be **Safe and Secure**



  - The high complexity of automated systems incurs an overhead on the verification and validation making it **time-consuming** and **costly**.

Project Overview and Objectives

High-level Project Objective

Design, **implement** and **evaluate** state-of-the-art **methods** and **tools** that reduce the **time** and **cost** needed to verify and validate automated systems with respect to **Safety** and **Security** requirements.


Automotive (3 UC)


Railway (2 UC)


Aerospace (1 UC)


Agriculture (1 UC)


Health (2 UC)


Industrial robotics (4 UC)

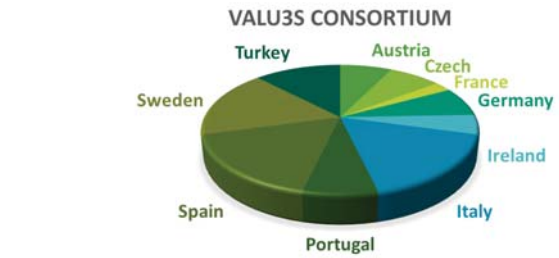
Project Overview

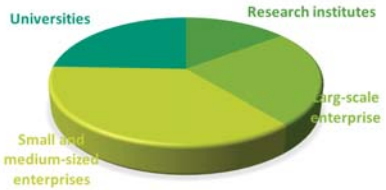
- VALU3S is funded by **ECSEL JU** under Horizon 2020 Work Programme
- Start date: **01/05/2020** Ending date: **30/04/2023** Duration: **36 months**
- The consortium consists of **41 partners** from **10 countries**
- The total VALU3S project cost is **25 857 454 €**

Countries

Industry

Academia



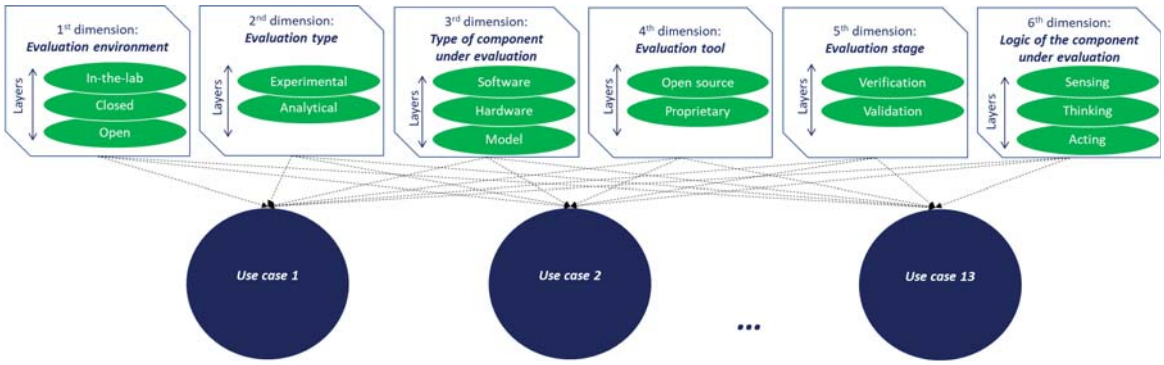


High-level Objective

Design, **implement** and **evaluate** state-of-the-art **methods** and **tools** that reduce the **time** and **cost** needed to verify and validate automated systems with respect to **Safety** and **Security** requirements.

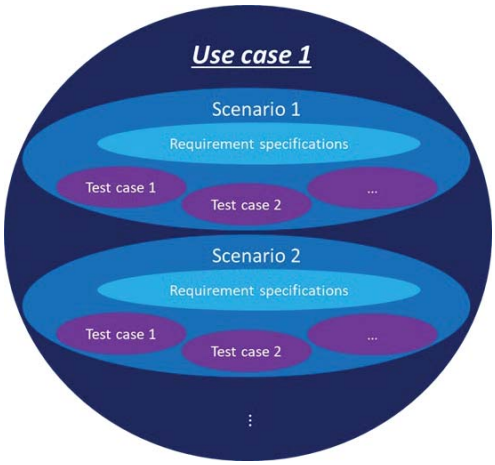
High-level Objective

Design, **implement** and **evaluate** state-of-the-art **methods** and **tools** that reduce the **time** and **cost** needed to verify and validate automated systems with respect to **Safety** and **Security** requirements.



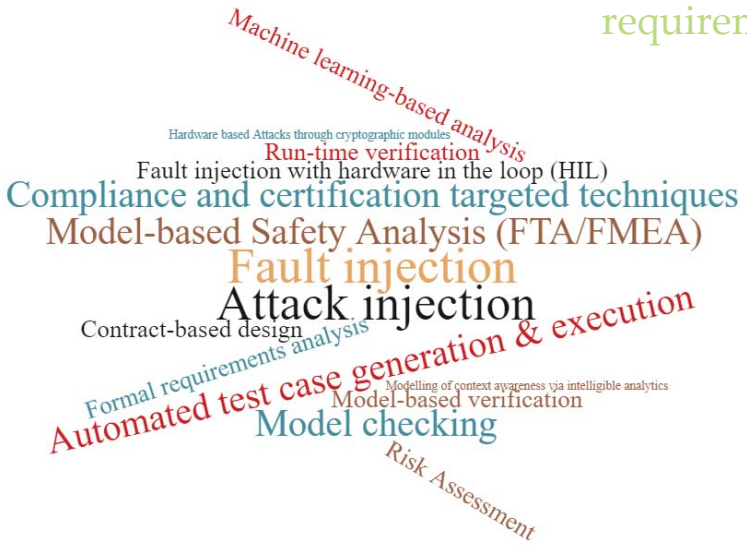
High-level Objective

Design, **implement** and **evaluate** state-of-the-art **methods** and **tools** that reduce the **time** and **cost** needed to verify and validate automated systems with respect to **Safety** and **Security** requirements.



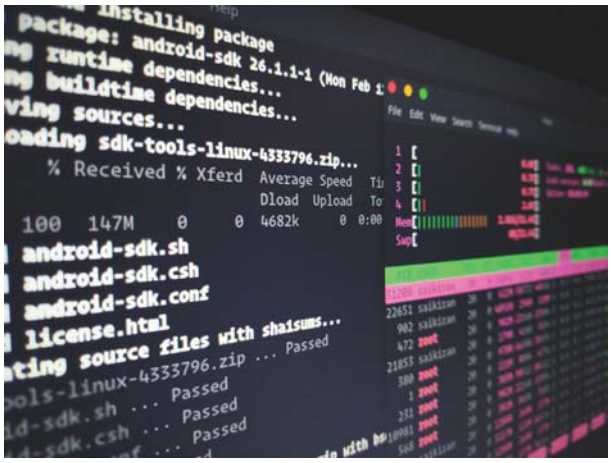
High-level Objective

Design, **implement** and **evaluate** state-of-the-art **methods** and **tools** that reduce the **time** and **cost** needed to verify and validate automated systems with respect to **Safety** and **Security** requirements.



High-level Objective

Design, **implement** and **evaluate** state-of-the-art **methods** and **tools** that reduce the **time** and **cost** needed to verify and validate automated systems with respect to **Safety** and **Security** requirements.



High-level Objective

Design, implement and **evaluate** state-of-the-art **methods** and **tools** that reduce the **time** and **cost** needed to verify and validate automated systems with respect to **Safety** and **Security** requirements.


Automotive (3 UC)


Railway (2 UC)


Aerospace (1 UC)


Agriculture (1 UC)


Health (2 UC)


Industrial robotics (4 UC)

Target Domains

- 13 use cases (UC) from 6 domains will be evaluated.

Automotive (3 UC)

Railway (2 UC)

Aerospace (1 UC)

Agriculture (1 UC)

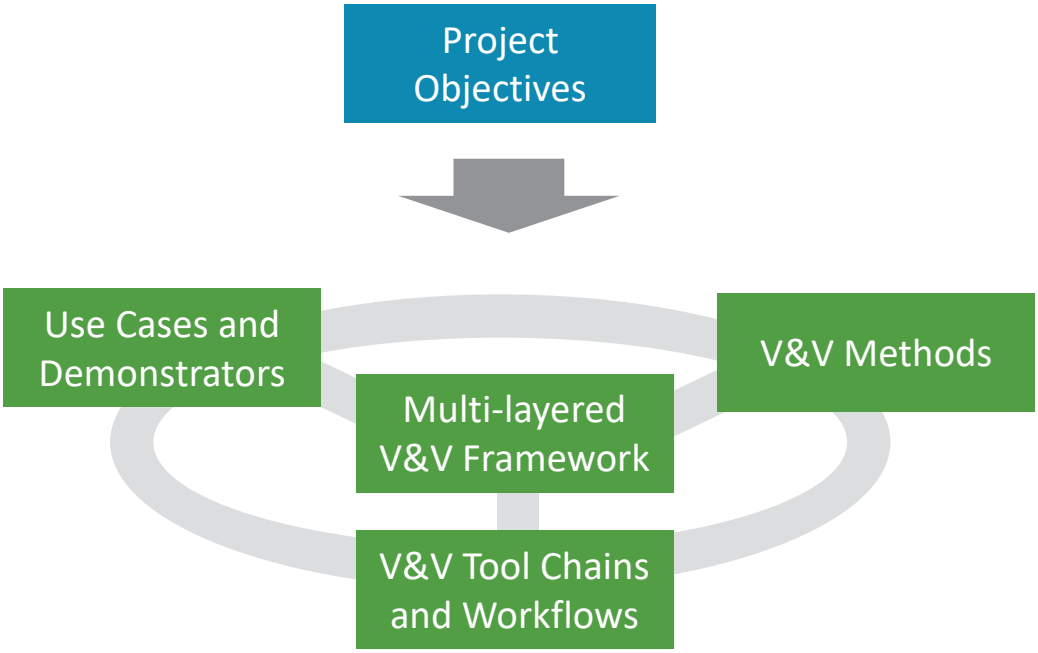
Health (2 UC)

Industrial robotics (4 UC)

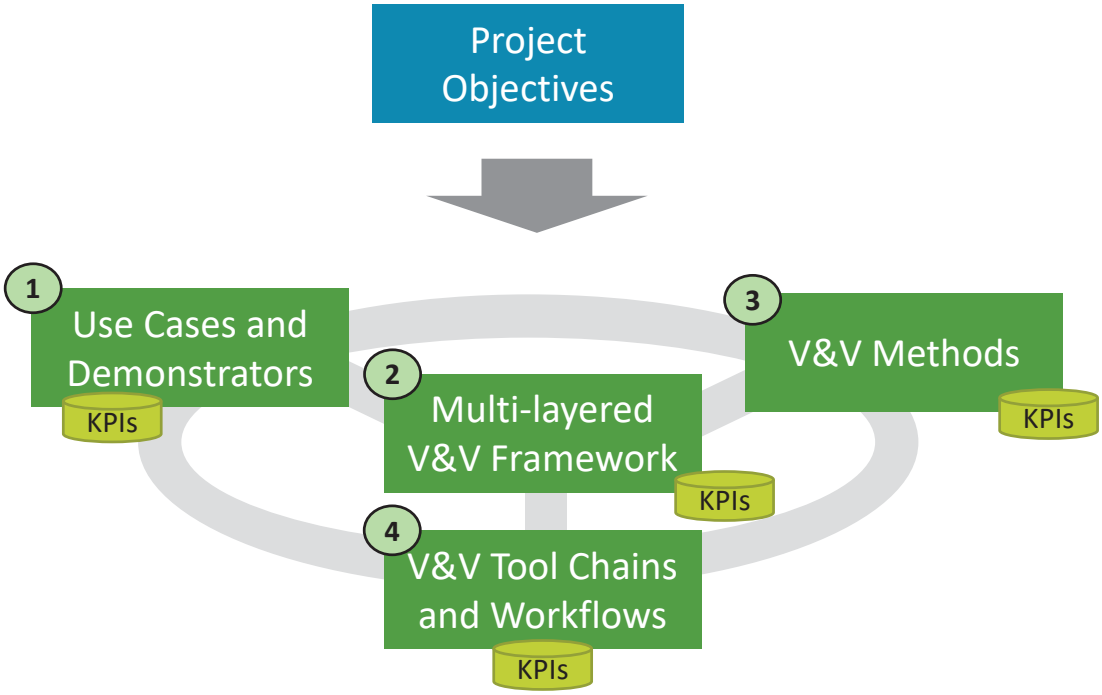
Definitions of Major Project Assets in VALU3S

- Use Case (UC)
  - Description of how users will implement and work with the developed solution in their industry specific productive environment
  - Outline of the solution from a user's point of view (incl. goals and solution steps)
- Key Performance Indicator (KPI)
  - Metrics measure and/or determine progress or the degree of fulfillment with respect to important objectives or critical success factors within an organization or project
- V&V Method
  - A particular procedure for V&V, especially a systematic or established one
- V&V Tool
  - computer program or technical asset that implements a V&V method or parts of it and often supports the automated the execution of a V&V method or parts of it.
- V&V Workflow
  - orchestrated and repeatable pattern of V&V activities that provide services or process information and consists of sequence of operations
- Demonstrator
  - Use case, utilising the results achieved by the project

Project Assets



Project Assets and KPIs



Project Asset 1
Use Cases and Demonstrators

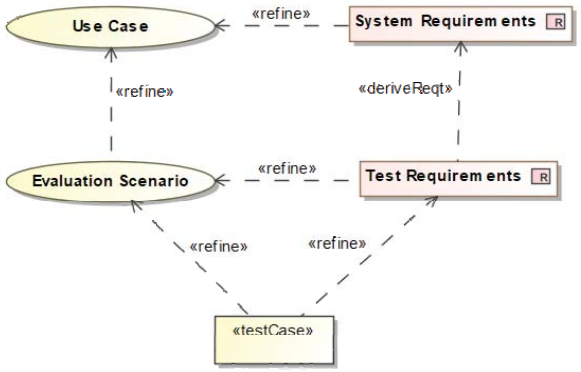
Project Asset 1: Use Cases and Demonstrators

- Use Case (UC)
  - Description of how users will implement and work with the developed solution in their industry specific productive environment
  - Outline of the solution from a user's point of view (incl. goals and solution steps)
- Demonstrator
  - Use case, utilising the results achieved by the project

Use case-driven approach

The main objectives of this approach

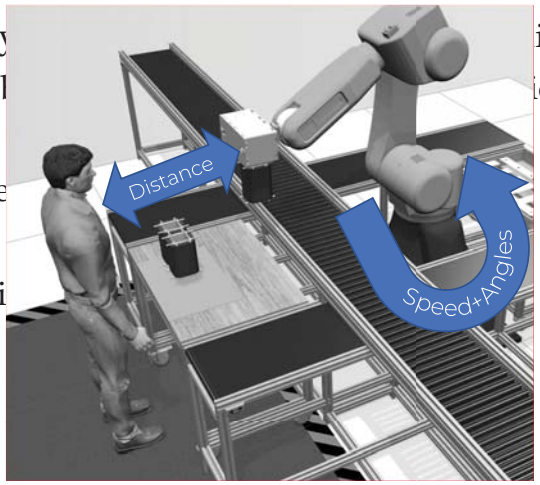
- (i) gaining insight into the evaluation
  VALU3S use cases;
- (ii) describing in detail the evaluation
  and the derivation of respective test r
- (iii) taking the repository of scenarios
  domains/the use cases.



Use Case Assets

Sample Use Case:
Human-Robot-Interaction in Semi-Automatic Assembly Processes

- **Use case definition**
  - Automated validation of design concepts and sy ine
  - Modeling, virtualization and simulation of cyl on line parts, worker, comm. networks, sensors)
  - Evaluation object: **virtual models** of a distribute
  - Quality characteristics: **fault tolerance**
  - Design and implementation of a virtual validati



Partners PUMACY TECHNOLOGIES | Fraunhofer IESE | Fraunhofer IIS | UCLM

Use Case Assets
Human-Robot-Interaction in Semi-Automatic Assembly Processes

Evaluation Scenarios

VALU3S_WP1_Industrial_13 - Corruption of input/output signal at robot gripper
VALU3S_WP1_Industrial_14 - Data manipulation in human-robot-interaction

Use Case Requirements

Test Cases

Evaluation Criteria

Eval_SCP_12 – Software fault tolerance robustness
Eval_SCP_13 – Simulation level system robustness

Eval_VV2: Test Coverage (regarding faults and fault combinations)

Sample Evaluation Criteria for Demonstration

Identifier	Eval_SCP_12
Description	Model-based software testing including fault injection to ensure fault-tolerant use case activitiy. Compliance through measurement and verification results.
Measured quantities	Number of detected faults (e.g. behavior, communication network, timing, algorithmic accuracy)

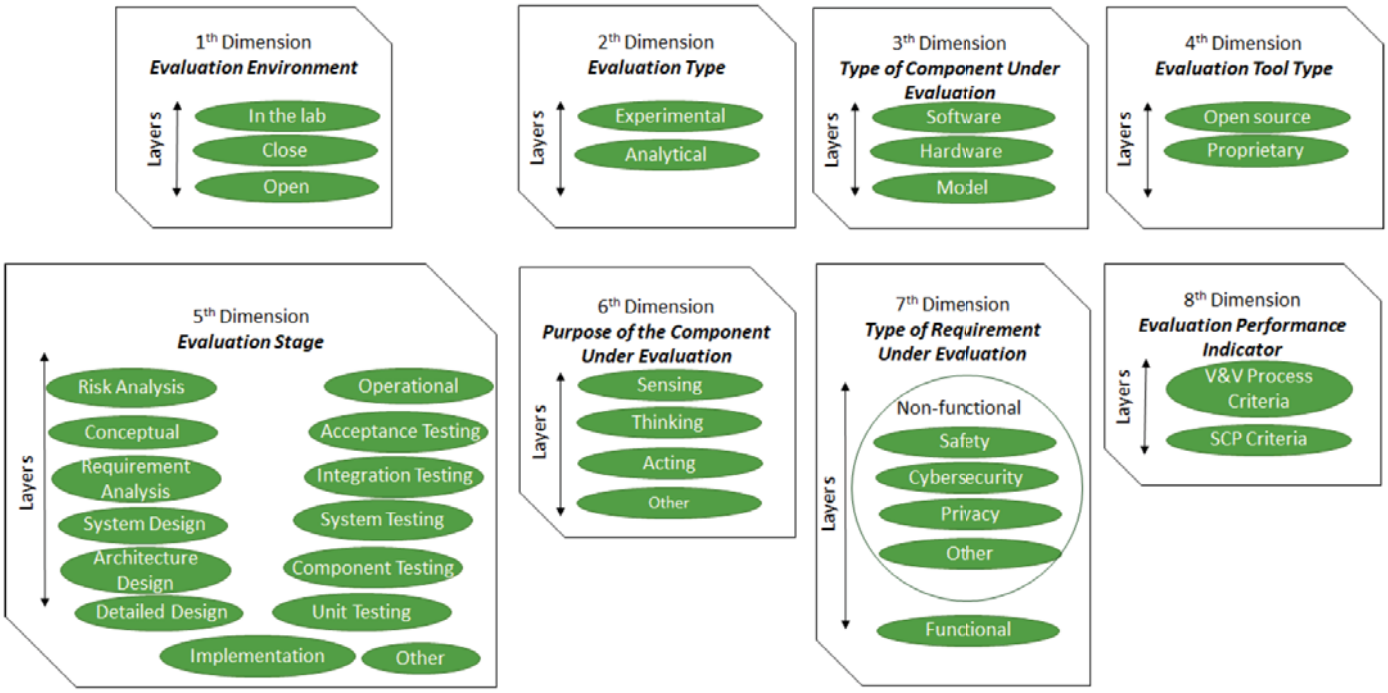
Identifier	Eval_VV_2
Description	Model-based software testing including fault injection to ensure fault-tolerant use case activitiy. Compliance through measurement and verification results.
Measured quantities	Number of test items covered detected faults (e.g. behavior, communication network, timing, algorithmic accuracy)

Project Asset 2
Multi-layered V&V Framework

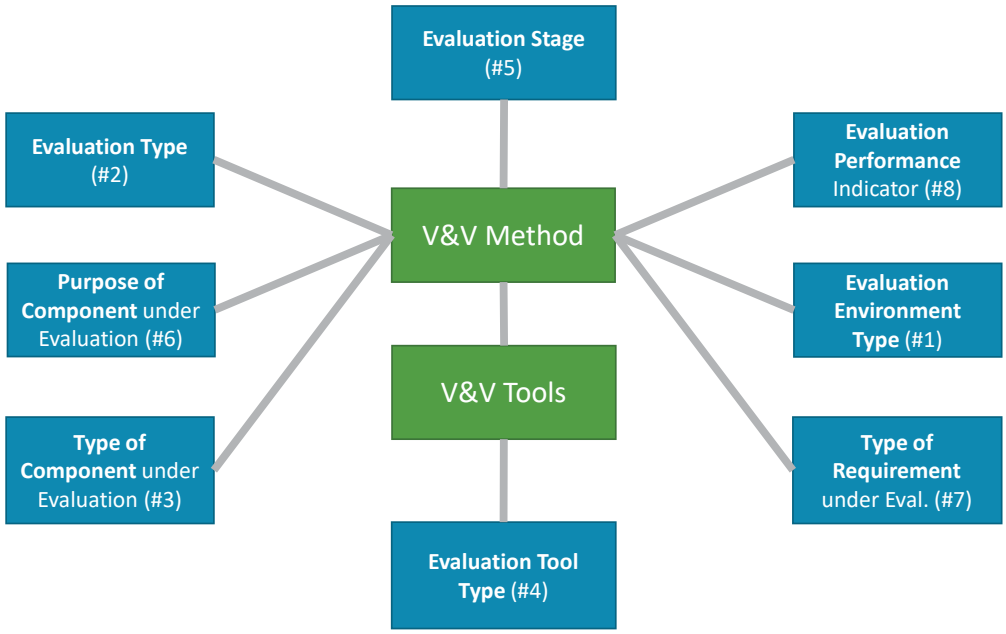
Project Asset 2: Multi-layered V&V Framework

- Goal: classification and storage of the different V&V elements that are created during the project
- storage of the V&V information in a uniform and homogeneous way
- defines what data related with each V&V activity must be collected and defines the data format.
- methodological framework, enabling the decomposition of elements and components required to conduct system V&V

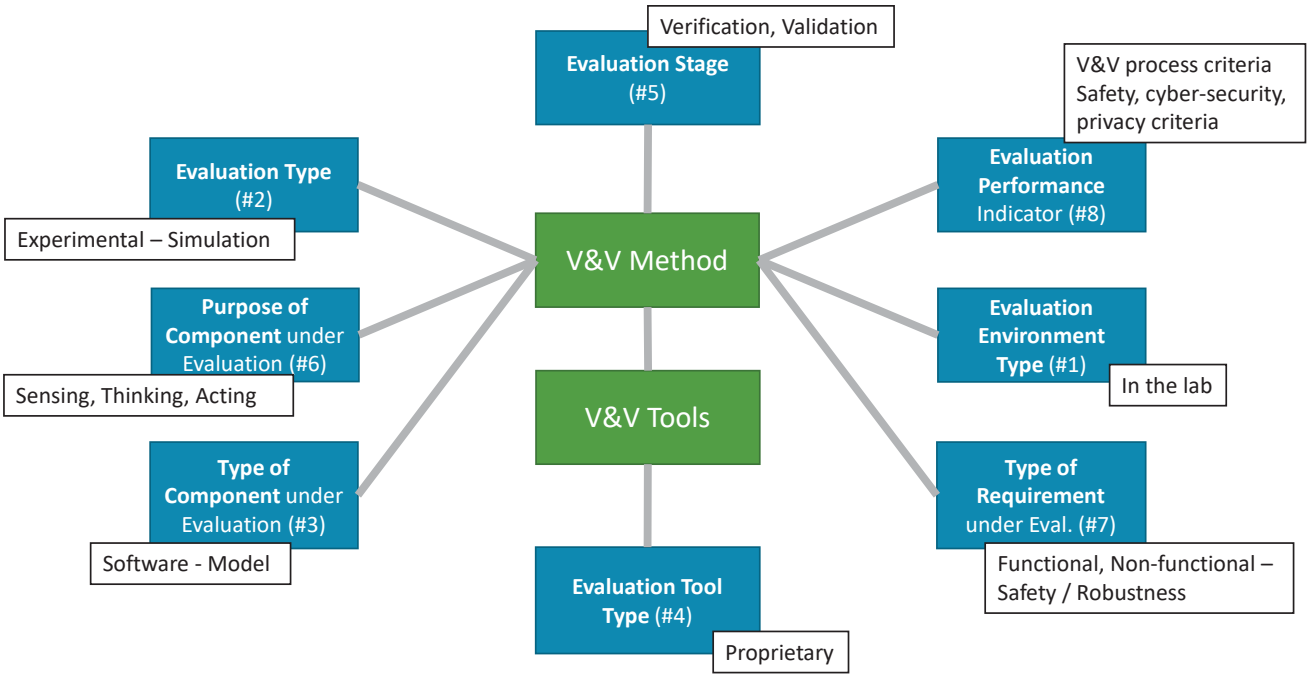
Multi-layered V&V Framework



Multi-layered V&V Framework



Multi-layered V&V Framework Example: FERAL in Context of UC4



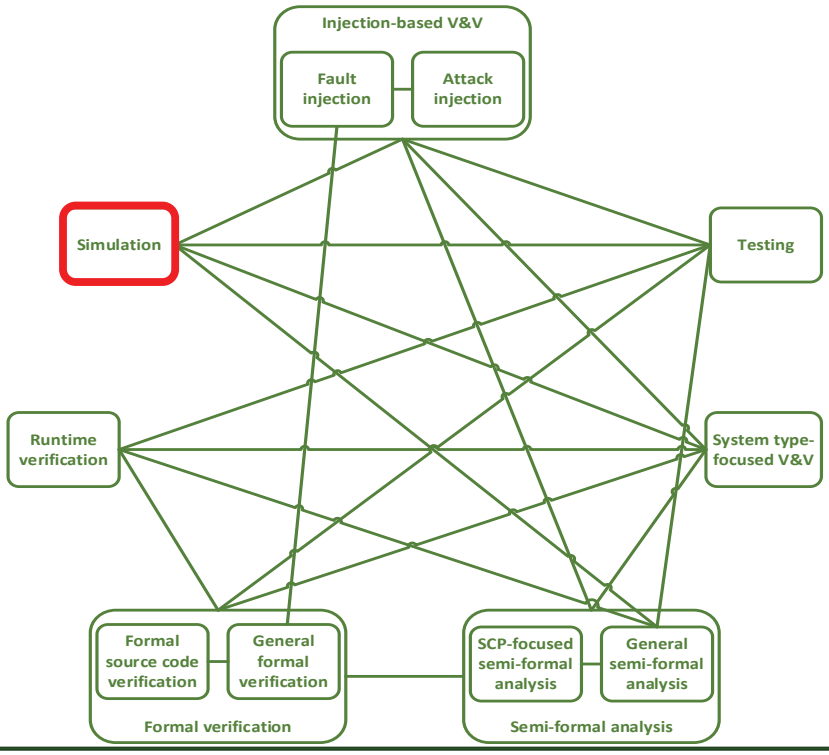
Project Asset 3
V&V Methods

- Definition
- Classification
- Gap Analysis and Improvement
- Combination

V&V-Methods

- Def. V&V method: A particular procedure for V&V, especially a systematic or established one

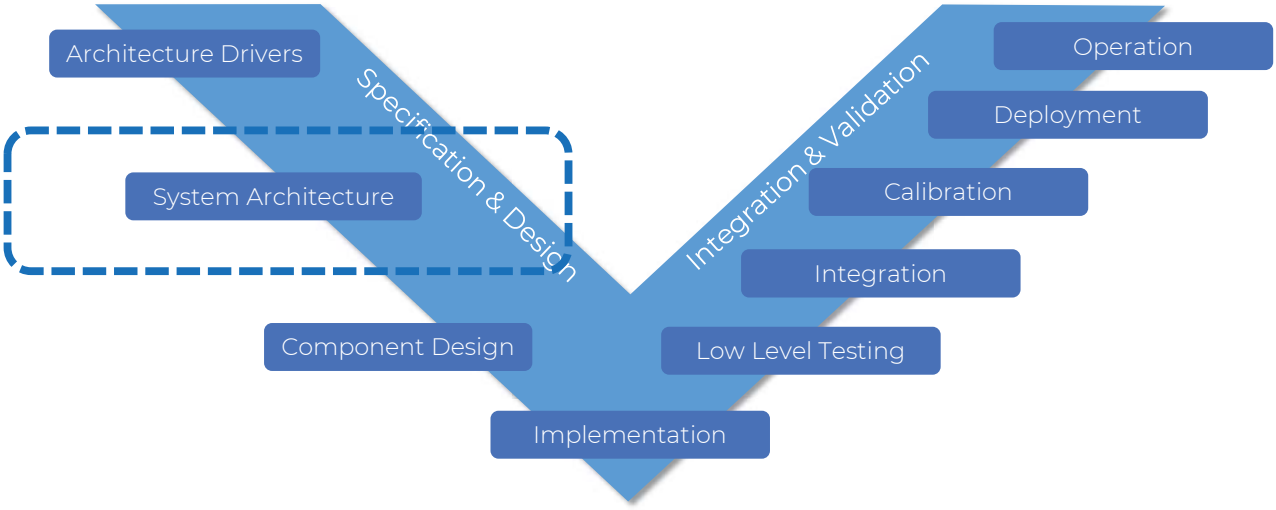
V&V Method: Classification



Class: Simulation-based V&V Methods

- Goals:
  - Enable early verification and validation at system design time
  - Replace system integration tests with expensive hardware test equipment by appropriate V&V activities at design time
- Main asset: Models
  - use of models that behave or operate like a given system to predict how the system would respond to defined inputs
- Challenges
  - Coverage of specific quality properties, which involve the development of dedicated simulation components
    - e.g. appropriate models of humans and their interactions with autonomous robots)
  - support to parallelize and accelerate execution and evaluation activities
    - cost-efficient development of simulation components and scenarios is a pre-requisite to fully exploit the advantages of early V&V at system design time.

Simulation-based V&V Methods in Development Processes



Example Method: Virtual Architecture Development and Simulated Evaluation of Software Concepts [VAD]

- Purpose: **Efficient** and reliable prototyping of complex systems involving cross-domain aspects by integrating heterogeneous components within holistic testing scenarios subject to goal-specific model fidelity and by systematically evaluating properties of interest in self-contained virtual runtime environments.
- Description
- **Relationship with other methods**: Model-based testing, Model-based robustness testing, Simulation-based verification.

Example Method: Virtual Architecture Development and Simulated Evaluation of Software Concepts [VAD]

- **VALU3S Goals**: Reliability; Fault Tolerance; Robustness [Safety]
- **Use case**: Human-Robot-Interaction in Semi-Automatic Assembly Processes [UC4 Pumacy] (Domain: Industrial automation)
- **Use case scenarios**
  - Industrial_13 - Corruption of input/output signal at robot gripper.
  - Industrial_14 - Data manipulation in human-robot-interaction

Example Method: Virtual Architecture Development and Simulated Evaluation of Software Concepts [VAD]

- **Strengths**
  - Enables early verification of the appropriateness of design decisions by providing executable simulation scenarios
  - Provides technical solutions for coupling heterogeneous system parts (i.e. different implementation formats and maturity levels) and communication protocols
  - Reuses and connects existing simulation tools
- **Limitations**
  - Initial abstraction level-dependent efforts for creating simulation scenarios and simulation components
  - Trade-off between accuracy and effort for finding an appropriate simulation model quality

Project Asset 4
V&V Tools and Workflows

Project Asset 4: V&V Tools and Workflows Definitions

- V&V Tool
  - computer program or technical asset that implements a V&V method or parts of it and often supports the automated the execution of a V&V method or parts of it.
- V&V Workflow
  - orchestrated and repeatable pattern of V&V activities that provide services or process information and consists of sequence of operations

Project Asset: V&V Tools
Tool Description: FERAL 1/2

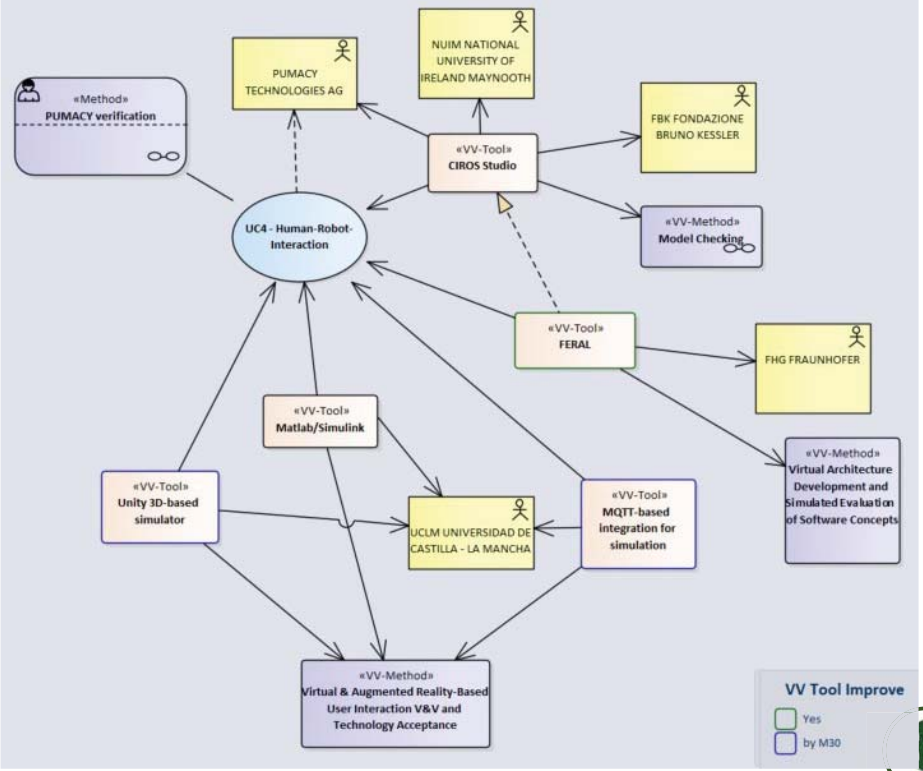
- **Short description:** FERAL is a **simulation framework** for
  - creating **virtual prototypes**
  - by **coupling simulation models and simulators**, existing code, and virtual hardware platforms and valuating
  - **across different abstraction levels** and in an early stage of the development process.
- **Partner**: Fraunhofer IESE
- **VALU3S Goals**: Reliability; Fault Tolerance; Robustness [Safety]
- **Use case**: Human-Robot-Interaction in Semi-Automatic Assembly Processes [UC4 Pumacy] (Domain: Industrial automation)
- **V&V Method**: Virtual Architecture Development and Simulated Evaluation of Software Concepts

Project Asset: V&V Tools
Tool Description: FERAL 2/2

- **Unique selling points**:
  - Enables early verification of design decisions by providing executable simulation scenarios
  - Provides technical solutions for coupling heterogenous system parts and communication protocols.
  - Reuses and connects existing simulation tools.
- **New features**:
  - Extending the co-simulation interface for Python (to enable the coupling of additional simulation tools (CIROS Simulation Framework)
  - Fault injection component for communication protocol and component behavior
- Link: https://www.iese.fraunhofer.de/en/services/digital-twin/feral.html

Use Case Tool Map

- Sample use case: Human-Robot-Interaction in Semi-Automatic Assembly Processes



Project Asset: V&V-Workflow

- Definition of V&V workflow
  Orchestrated and repeatable **pattern of V&V activities** that provide services or **process information** and consists of **sequence of operations**

Verification and Validation Modelling Language (VVML)

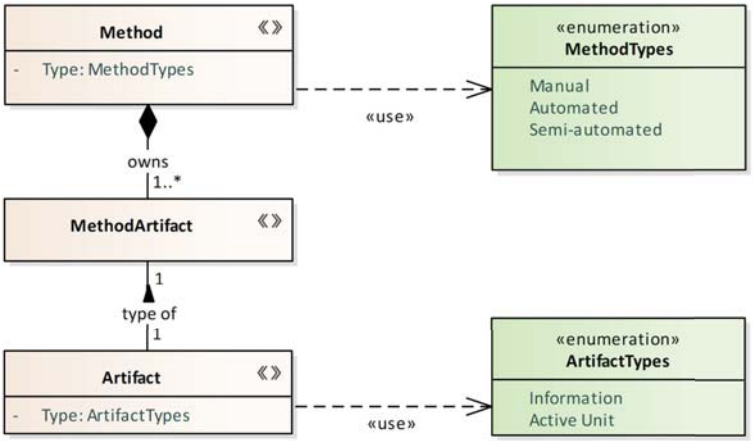
- **Stakeholder requirements**
  - simple and clear notation, i.e., providing few element types and few diagrams
  - based on behavior modelling approaches in software engineering
  - implementable in state of the practice modelling frameworks
  - exchange of artifacts between V&V methods
  - decomposition of V&V methods as implementation of sequences of lower level activities
  - composition of methods to higher level methods
  - preparation for automated and tool-supported analysis of V&V workflows

- **Domain-specific language** for modelling V&V workflows VVML
- **Tool:** Modelling Framework Enterprise Architect (EA) + Profile
- **Levels of modelling**
  - V&V Method Specification (→ base elements)
  - V&V Workflow Definition (→ sequence of activities and flow of artifacts)

V&V Workflows
Main Outputs

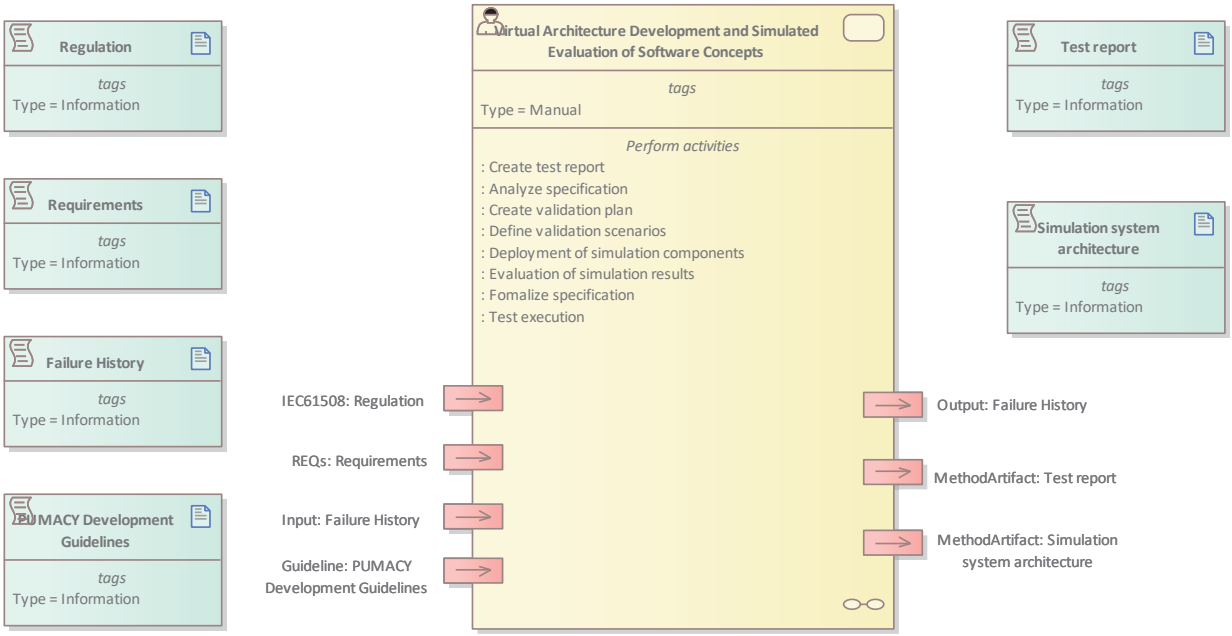
- Modelling Language VVML
- Tool-Support based on standard modelling framework Enterprise Architect (EA)
- Guidelines and handbook for V&V workflow modellers

VVML Elements



Tool Framework: Enterprise Architect

Tool-supported Design of VVML Elements



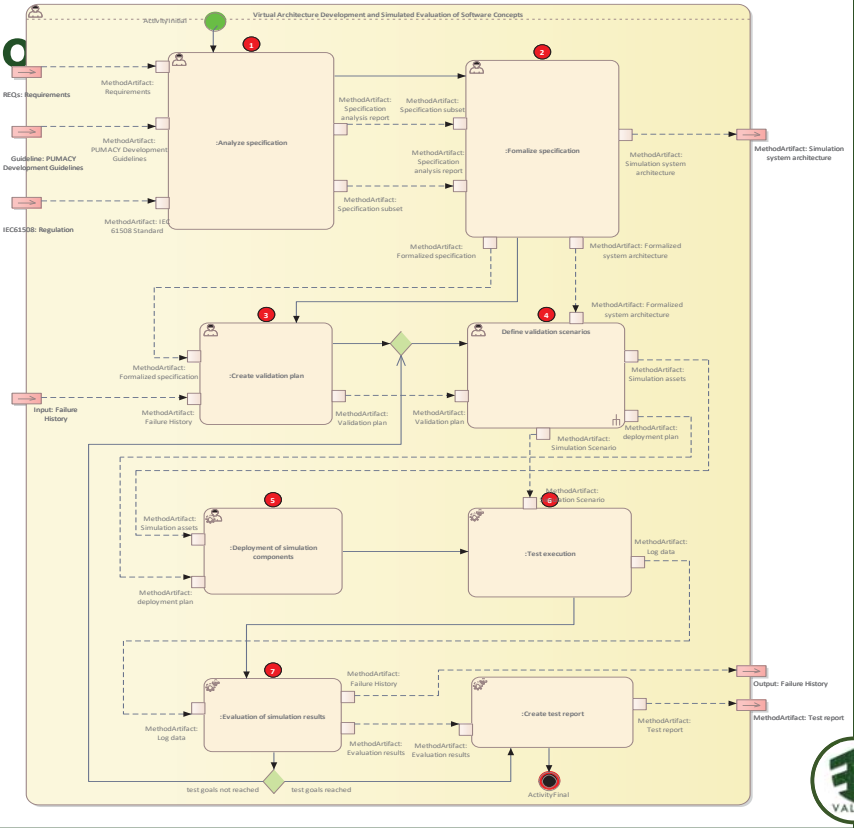
Tool Framework: Enterprise Architect

Design Elements of V&V Workflows

Definition of base assets for defining V&V workflows and sub-activities

Element	Decription
Start Workflow	Node that initiates the beginning of a workflow
Stop Workflow	Node that indicates the end of a workflow
Activity	Atomic action that is not further decomposed into steps
Call Behavior	Invocation of another method, which is further decomposed in another method workflow diagram
Activity Artifact	Activity interface for its input and output artifacts
Gateway	Branching of sequence flow based on condition
Fork / Join	Enables parallel sub-paths of sequence and artifact flows
Sequence Flow	Sequential connection of VVML activities
Artifact Flow	Exchange of artifacts between activities or from/to method interfaces

Tool-supported Design o



Tool Framework: Enterprise Architect

Project Asset: Standardization

Project Asset: Standardization

- Incorporating recommendations from industrial standards of project-related domains
  - Automotive
  - Railway
  - Aerospace
  - Industrial automation / robotics
  - Health / medical
  - Agriculture
- Exchange and involvement of partners in standardization group regarding VALU3S results

Standardization Landscape

		Domain					
Standard	**Comment**	**Agriculture**	**Aerospace**	**Automotive**	**Railway**	**Industrial Automation and robotics**	**Health**
IEC 61508	Domain independent basic safety standard. Security is partially considered (during risk analysis) and a maintenance phase with a discussion about the role of security is ongoing.	X	X	X	X	X	X
IEC TR 63069	Framework for the interaction from safety to security on a domain independent level.	X		X	X	X	X
IEC 62443-1	General describes overarching concepts, terms and metrics for secure IACS systems. Ongoing rework of some subparts.				X	X	
IEC 62443-2	Policies & Procedures present the management framework for implementation, patching and operation. Ongoing rework of some subparts.				X	X	
IEC 62443-3	The system level is aimed at Asset Operator and System Integrator and describes necessary activities and processes during the system engineering. Ongoing rework of some subparts.				X	X	
IEC 62443-4	The component level is for Product supplier and describes how to develop secure components for the integration in IACS. Ongoing rework of some subparts.				X	X	
ISO 26262	ISO 26262 Edition 2 was published in 2018 and focuses on functional safety for automotive systems. It could be applied to vehicles in the farming domain and the interaction with security (e.g. combining V&V) is included.	X		X			

Project Asset: Standardisation Recommendation from Standards

- ISO 26262 – 6 [SW-Part]
- Fault injection in software integration test:
  - Focus: test the appriateness of hardware-software interfaces related to safety mechanisms.
  - Injection targets
    - SW Architectural Design and HW-SW Interfaces
    - Safety-related functionality of integrated SW subsystem
    - Robustness and fault tolerance
    - Communication network and resources
  - Verify freedom from interference.

Publications

- General Papers
  - J.A. Agirre et al., The VALU3S ECSEL project: Verification and validation of automated systems safety and security, Microprocessors and Microsystems. Vol. 87, 2021
  - T. Bauer et al., Cross-domain Modelling of Verification and Validation Workflows in the Large Scale European Research Project VALU3S, SAMOS 2021
  - J.L. de la Vara et al., A Proposal for the Classification of Methods for Verification and Validation of Safety, Cybersecurity, and Privacy of Automated Systems. QUATIC 2021: 325-340
- List of Publications
  - https://valu3s.eu/publications/

Summary and Outlook

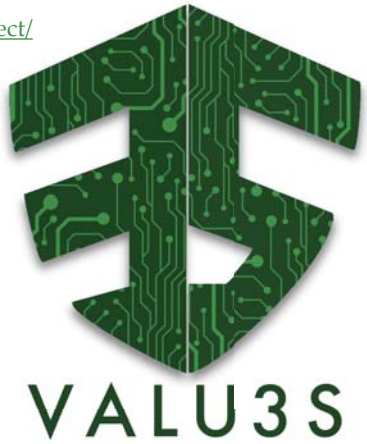
- Current status
  - Use-case driven approach has been defined and implemented through all WPs
  - V&V framework, methods, and workflow modelling approach modelling are defined and in use
  - V&V tool are being developed and integrated
  - Initial evaluation of V&V methods, tools, and framework has been conducted
- Next steps
  - Continuous monitoring regarding project objectives and use case goals
  - Adaptation and improvements of assets (esp. tooling) based on initial results
  - Creating and sharing reusable solutions (methods, workflows, tools) for the V&V community

Follow us on:

https://www.linkedin.com/company/valu3s-project/

https://twitter.com/valu3s_project

https://www.youtube.com/channel/UCBvhaW8hkWgopiJWbFBrIFQ



Verification and Validation of Automated Systems' Safety and Security

www.valu3s.eu

Further Questions?

Thomas Bauer
Fraunhofer IESE
thomas.bauer@iese.fraunhofer.de

This project has received funding from the ECSEL Joint Undertaking (JU) under grant agreement No 876852. The JU receives support from the European Union's Horizon 2020 research and innovation programme and Austria, Czech Republic, Germany, Ireland, Italy, Portugal, Spain, Sweden, Turkey.

Intelligent Secure Trustable Things

How to develop trustworthy smart systems?

Framework to facilitate Trustworthiness of Smart Systems for End Users

Nikolai Ebinger Peter Moertl

InSecTT has received funding from the ECSEL Joint Undertaking (JU) under grant agreement No 876038. The JU receives support from the European Union's Horizon 2020 research and innovation programme and Austria, Sweden, Spain, Italy, France, Portugal, Ireland, Finland, Slovenia, Poland, Netherlands, Turkey

Smart systems can cause problems and have unintended consequences

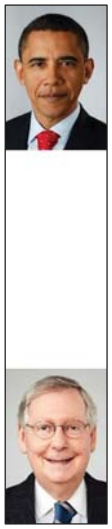
"Employment office algorithm: Researchers warn of discrimination and complain about the lack of transparency"
– derstandard.at (25.02.2020)

https://www.ams.at/arbeitsmarktdaten-und-medien/medien/presse-fotos/-service-fuer-arbeitsuchende

Smart systems can cause problems and have unintended consequences

"Twitter apoalogises for 'racist' image-cropping algorithm"
– theguardian.com (21.09.2020)

Entire picture:

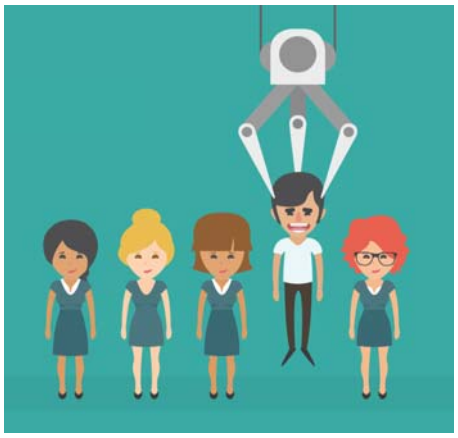


Twitter preview (2020):

https://pbs.twimg.com/media/EiT7aXCU0AANjKi?format=jpg&name=4096x4096
https://web.archive.org/web/20200919220517/https://twitter.com/bascule/status/1307440596668182528

Smart systems can cause problems and have unintended consequences

"Amazon scrapped 'sexist AI' tool"
– bbc.com (10.10.2018)



https://www.data-traction.at/wp-content/uploads/2019/05/ams_algo_v1.png

Smart systems can cause problems and have unintended consequences

"Employment office algorithm: Researchers warn of discrimination and complain about the lack of transparency"
– derstandard.at (25.02.2020)

"Twitter apoalogises for 'racist' image-cropping algorithm"
– theguardian.com (21.09.2020)

"Amazon scrapped 'sexist AI' tool"
– bbc.com (10.10.2018)

Need for developing smart systems in a human-centered, ethical and trustworthy way

Objectives

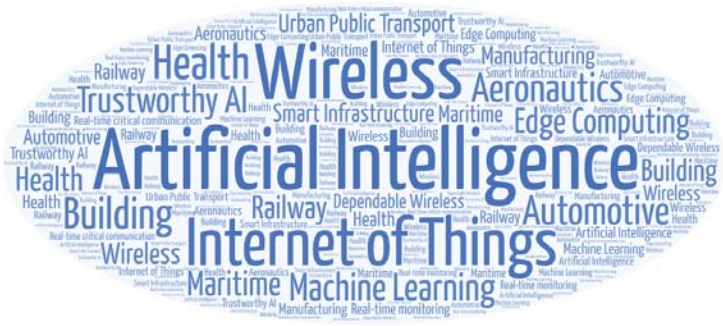
- Examples on problems caused by smart systems that show the need for developing smart systems in a human-centered way
- Overview on guidelines for ethical and trustworthy AI
  - EU Ethics Guidelines (2019)
  - ISO standard 24028 at stage 60:60 (2020) (trustworthiness in AI)
  - EU guidelines and proposed regulations for trustworthy AI (2021)
- Contextualization of technical systems
- Examples on considering user requirements in development of smart systems
- How to develop smart systems in a human-centered way (InSecTT framework)

https://www.data-traction.at/wp-content/uploads/2019/05/ams_algo_v1.png

InSecTT – Intelligent Secure Trustable Things

- InSecTT – Intelligent Secure Trustable Things
- The project aims at creating trust in AI-based intelligent systems and solutions
- Trust and trustworthiness are investigated from the human´s perspective
- Find details on **insectt.eu**

54 partners from 12 countries



Focus of InSecTT

What does trustworthy and ethical mean?

Trust

"... the attitude that an agent [smart system] will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability" Lee & See (2004)

Ethical

"Ethical behavior is based on written and unwritten codes of principles and values held in society."

"Ethics reflect beliefs about what is right, what is wrong, what is just, what is unjust, what is good, and what is bad in terms of human behavior." LumenCandela

➡ Trustworthiness and ethical principles for AI emerged

LumenCandela: https://courses.lumenlearning.com/boundless-management/chapter/ethics-an-overview/

EU Ethics Guidelines
Requirements for ethical AI

ISO Approach on trustworthy AI: Standard 24028

1 - Scope and definitions
2 - Normative references
3 - Terms and definitions
4 - Overview
5 - Existing frameworks applicable to trustworthiness
6 - Stakeholders
7 - Recognitions of high-level concerns
8 - Vulnerabilities, treats and challenges
9 - Mitigation measures
10 - Conclusion


TECHNICAL REPORT
ISO/IEC TR 24028
First edition 2020-05
Information technology — Artificial intelligence — Overview of trustworthiness in artificial intelligence
Reference number ISO/IEC TR 24028:2020(E)
© ISO/IEC 2020

ISO Standard 24028

Definition of Stakeholders

"Different stakeholders can hold differing views of the relative importance of different proposed characteristics for a trustworthy AI"

- Standardization of terms and a conceptual framework for trustworthy AI allows for understanding and communication between different stakeholders
- Proposed stakeholder types are defied based on roles in the AI value chain:
  - Data source, AI system developer, AI producer, AI user, AI tools and middleware developer, test and evaluation agency
- Stakeholder can differ in their views on trustworthy AI based on background and values
  - Principles proposed by the European Commission's High Level Expert Group working paper on Trustworthy AI build on the European Charter of Fundamental Rights
  - Different worldviews, such as Western Ethics, Buddhism, Ubuntu, Shinto, could bring the need to be considered in communicating trustworthy AI characteristics at a global level

Need to define and consider relevant stakeholder in order to apply recommendations in an adequate way

ISO Standard 24028

Explainability

"Although the explainability alone is not sufficient to guarantee the transparency of an AI system, it is an important component of a transparent AI system"

Explanations…

- … can refer to the AI itself and the results by the system
- … should depend on the recipients and their current understanding
- … can be presented before the use (ex-ante) or after the use (ex-post)

Ex-ante: to establish trust that the system is well designed and serves its purpose

- Establishes trust with the users and motivate the use of the AI system in the first place

Ex-post: needed to explain specific algorithmic results and the circumstances they were made in

- Ensures transparency

➡ Ex-ante and ex-post explanations should be consistent

ISO Standard 24028

Approaches to Explainability

- Explanations can be generated in different stages of AI development:
  - Pre-modelling: Understanding the data before building the model
  - Modelling: AI models can explain their decisions or are inherently interpretable
  - Post-modelling: explanations about decisions of non-interpretable AI models

  ➡ Explanations can be locally by a specific example of input/output or globally by explaining the general concept

- Different types of explanations can be used:
  - Causal: "How something functions"
  - Epistemic: "How we know it functions"
  - Justificatory: "On what ground it functions" or justificatory

  ➡ Formulating explanations brings the need for a trade-off between accuracy and understandability

ISO Standard 24028

Controllability

- Controllability can be addressed by implementing reliable mechanisms for the operator to take over control from the AI system
  - Need to clarify "who is offered what control over whose AI systems where multiple stakeholders are involved (e.g., the service provider or product vendors, the provider of the constituent AI, the user or an actor with regulatory authority)"
- Need to integrate human-in-the-loop control points in the AI system lifecycle to ensure reliable decision-making:
  - Decision makers with autonomy in the final decision-making process when taking into account the outcomes of AI systems
  - Domain experts given the opportunity to provide feedback to re-assess the AI system, to explain why it works in a certain way but also to improve the operation of the system

Proposed regulations for AI (2021)

EUROPEAN
COMMISSION

Brussels, 21.4.2021
COM(2021) 206 final

2021/0106 (COD)

Proposal for a

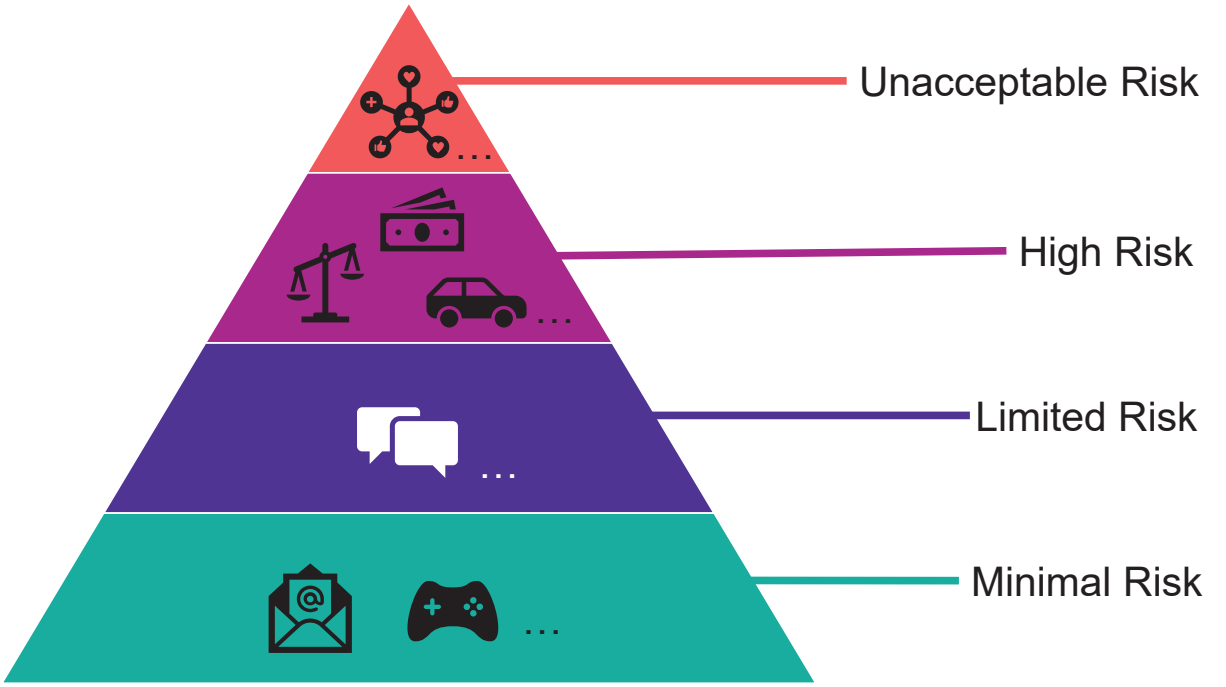
REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL

LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS

{SEC(2021) 167 final} - {SWD(2021) 84 final} - {SWD(2021) 85 final}

Proposed AI Regulations (European Commission)

Differentiation between **Four Risk Levels**



Ban

Strict obligations

Specific transparency obligations

Free use

Proposed AI Regulations (European Commission)

Today´s Focus

Unacceptable Risk	Ban
High Risk	Strict obligations
Limited Risk	Specific transparency obligations
Minimal Risk	Free use

- Risk management system
- Record-keeping
- Transparency
- Human oversight
- Accuracy, robustness and cybersecurity obligations

Proposed AI Regulations (European Commission)

Article 9 – Risk Management System

- Risk management system needs to be included throughout the entire lifecycle
  - Risk analysis of "known and foreseeable risks" shall be performed
  - These risks should be estimated considering use according to the intended purpose but also in case of "reasonably foreseeable misuse"
  - Risk evaluation based on post-market monitoring and adaption of risk management measures shall be applied if needed
- Identified residual risk shall be deemed acceptable and communicated to the user
  - Risk management measures shall reduce residual risk as far as possible
  - Where appropriate, mitigation and control mechanisms need to be used if the risk cannot be eliminated
- For the purpose appropriate risk management measures shall be identified by testing throughout the development

Article 13 – Transparency and provision of information to users

Proposed AI Regulations (European Commission)

"High-risk AI systems shall be designed and developed in such a way to ensure that their operation is sufficiently ***transparent to enable users to interpret the system's output and use it appropriately****…"*

- High-risk AI shall be accompanied by for the user **understandable instructions**
- Various required aspects about which the user must be informed are defined
  - Identity and the contact details of the provider
  - Intended purpose of the high-risk AI
  - Level of accuracy, robustness and cybersecurity
  - Known and foreseeable circumstances that…
    - …could impact expected level of accuracy, robustness and cybersecurity
    - …may lead to risks to the health and safety or fundamental rights
  - "when appropriate, specifications for the input data, or any other relevant information in terms of the training, validation and testing data sets used"
  - The expected lifetime and necessary maintenance/care measures to ensure the proper functioning of that AI (e.g., software updates)

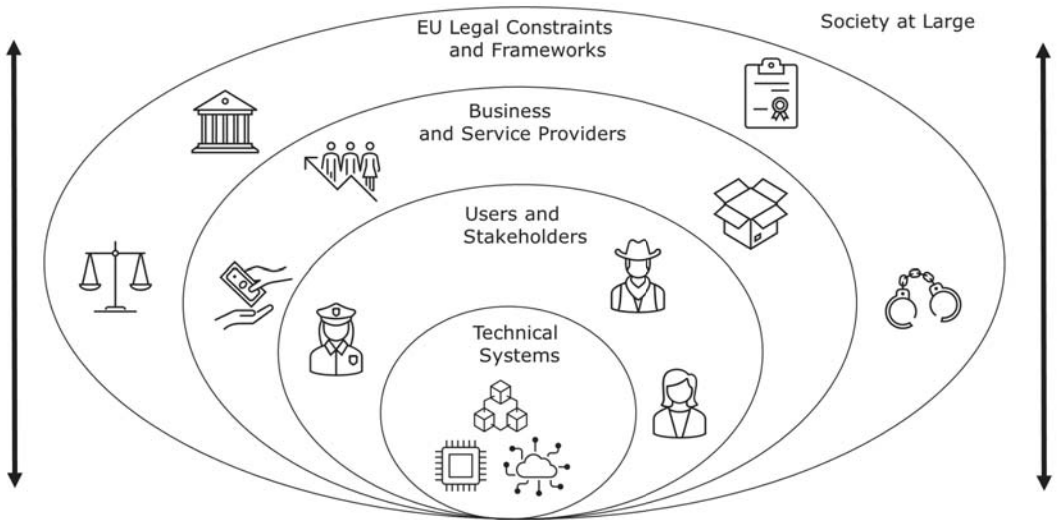
Proposed AI Regulations (European Commission)

Article 14 - Human Oversight

"High-risk AI systems shall be designed and developed in such a way, including with appropriate human-machine interface tools, ***that they can be effectively overseen by natural persons during the period in which the AI system is in use****"*

- Measures to achieve **human oversight on a cognitive level**
  - Ensure that the user has a full understanding of the AI´s capabilities/limitations
  - Enable the user to monitor, detect and address anomalies/dysfunctions/unexpected performance
  - Design with awareness of automation bias
    - Address the user´s tendency to initially overtrust technical systems
  - Enable the user to correctly interpret the system´s output
- Measures to achieve **human oversight on a behavioral level**
  - Provide the possibility to intervene/interrupt on AI operation through „stop" button or similar
  - Allow stakeholder to disregard/override/reverse output of AI

Contextualization of technical systems



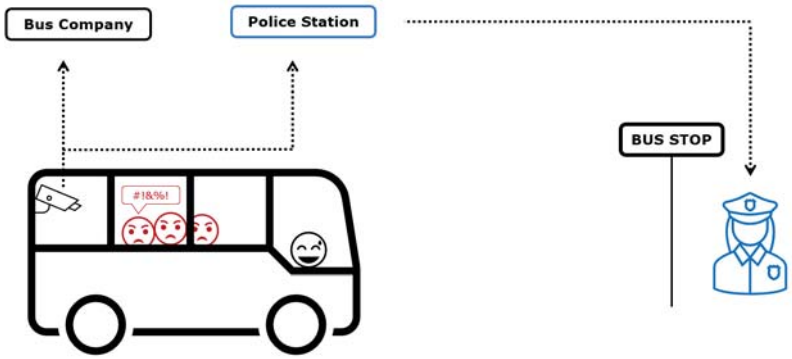
Intelligent Secure Trustable Things

Part 2 - Solutions

Developing smart Systems in a Human-Centered Way

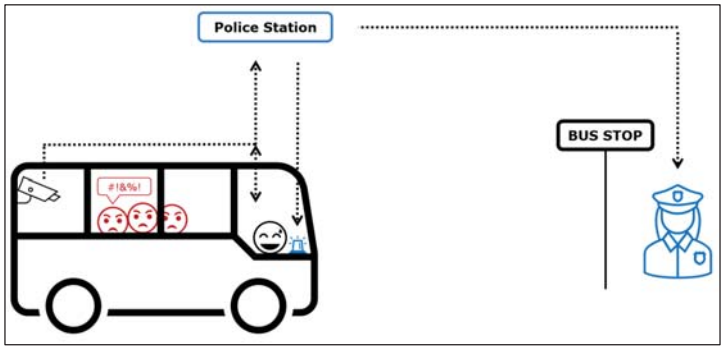
InSecTT has received funding from the ECSEL Joint Undertaking (JU) under grant agreement No 876038. The JU receives support from the European Union's Horizon 2020 research and innovation programme and Austria, Sweden, Spain, Italy, France, Portugal, Ireland, Finland, Slovenia, Poland, Netherlands, Turkey

Example of applying ECCOLA



Example of applying ECCOLA

In a workshop we discussed trustworthiness risks for a use case using „ECCOLA“:



Identified trustworthiness risks:

- Insufficient transparency for different stakeholder
- Privacy of passengers and other road users
- Sharing of some sensitive data seems required
- Observed possible tension between different trustworthiness risks

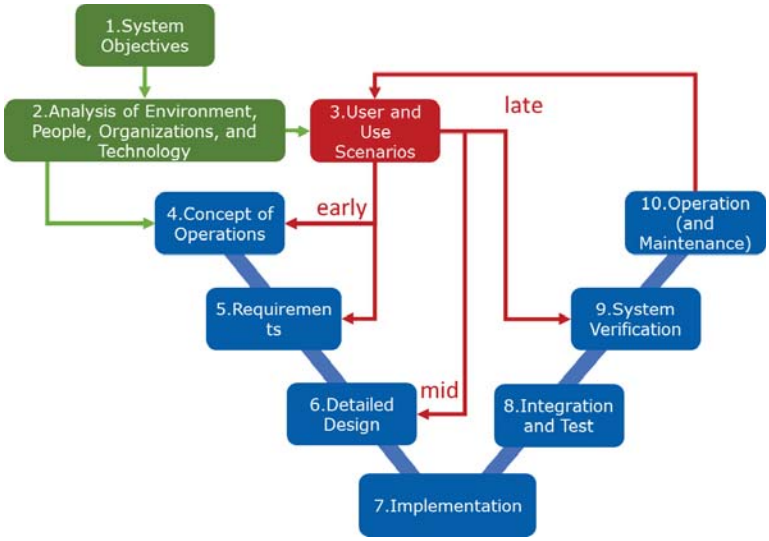
Framework for developing smart Systems in a Human-Centered Way

- So, how to develop trustworthy ethical AI systems?
- The single overarching theme
  - is to formulate trustworthiness
  - and ethics requirements
- early on into the development process and
- then manage them during the duration of the product life-cycle.
- Not dissimilar to a achieving a safe system
  - Much larger than just technical
    - Need to bring requirements that are not visible in the design of the components.
    - The safety challenges of an airplane are being observed in the real world and then shape the safety requirements on the components and organizational processes.

HSI Framework for Building Trustworthy AI Systems

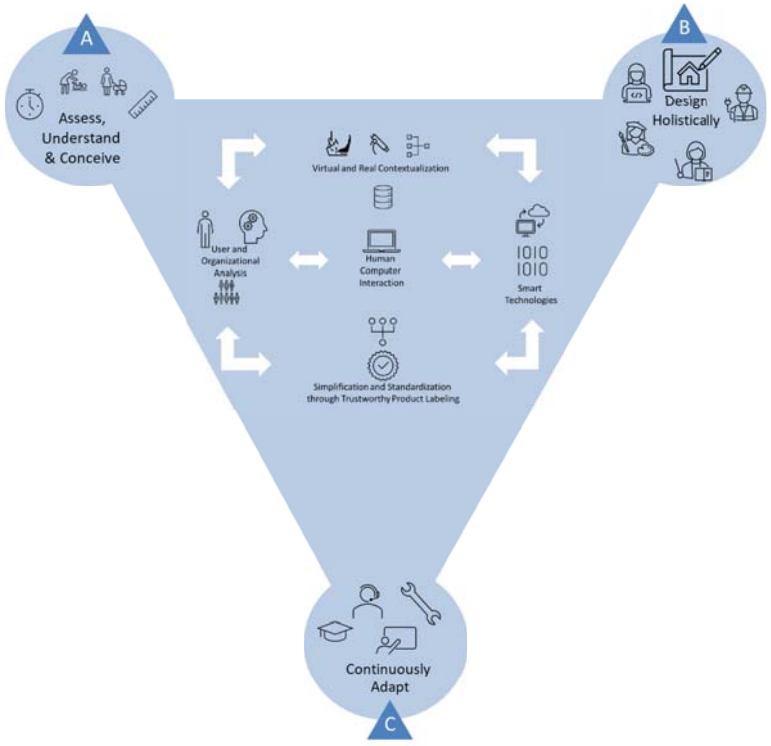
1. Of critical importance is the start of the system development with a system objective and mission
   - to derive the necessary contextualization of how and where the system will be used.
   - It is not possible to derive ethical or trustworthiness requirements for a computer chip, if its intended application, use, and use context is unknown.
2. Steps 2 through 5 collect the necessary information about the involved technologies and human stakeholders and formulate
   - Scenario based methods
   - Prime vehicle for requirements that are derived from the concept of operations.
   - Specifically, acceptable explainability methods are derived from knowledge about the user.

- Use scenarios inform design, verification, and are updated through actual operations…
  - Red thread through holistic system behavior



HSI Framework for Building Trustworthy AI Systems

- To successfully bring trustworthiness and ethical requirements into the R&D processes early on
  - organizational structures and responsibilities need to be defined upfront.
  - Three interconnecting cornerstones are critical:
    - (A) a research organization
      - Capabilitiy to extract and represent contextual requirements
    - (B) a holistic development organization
      - strong multi-disciplinary solutions
    - (C) a life-cycle long learning and maintenance operations
      - To address continuously changing aspects of the system.
  - Cornerstones are linked via
    - Virtualization tools,
    - Iterative design processes and refinement, and
    - Standardization schemes to aid collaboration and team work



HSI Cornerstones

A. "Assess, Understand, Conceive"

- provide information about the intended use situation for the development of smart systems
- To achieve sustainable acceptance and use.
- Technical feasibility and cost-effectiveness are thereby concept-forming factors equal to the usage situation information, this is a novelty here.

B. "Design Holistically"

- translates the vision from cornerstone 1 into a holistic design of the system.
- Holistic: orchestrated teams of multidisciplinary specialists work together
- This serves as a point of convergence across the disciplines and teams.

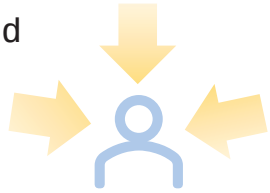
C. "Continuously Adapt"

- continuous adaptation and updating of products
  - education of users during the life cycles of smart technologies
  - Effective adaptations require detailed information about user and usage conditions.

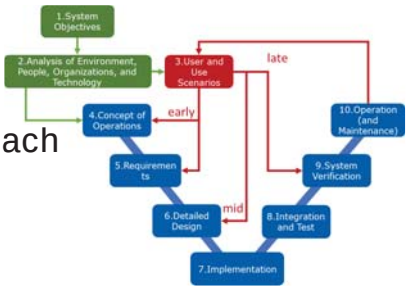
Conclusions

- High-level requirements for trustworthy and ethical AI are defined by several institutions and groups
- Approaches were already undertaken in the past to make requirements applicable

There is a need for holistic human centered development approaches



- The InSecTT framework guides through a process for developing trustworthy smart systems in a user centered approach



Intelligent Secure Trustable Things

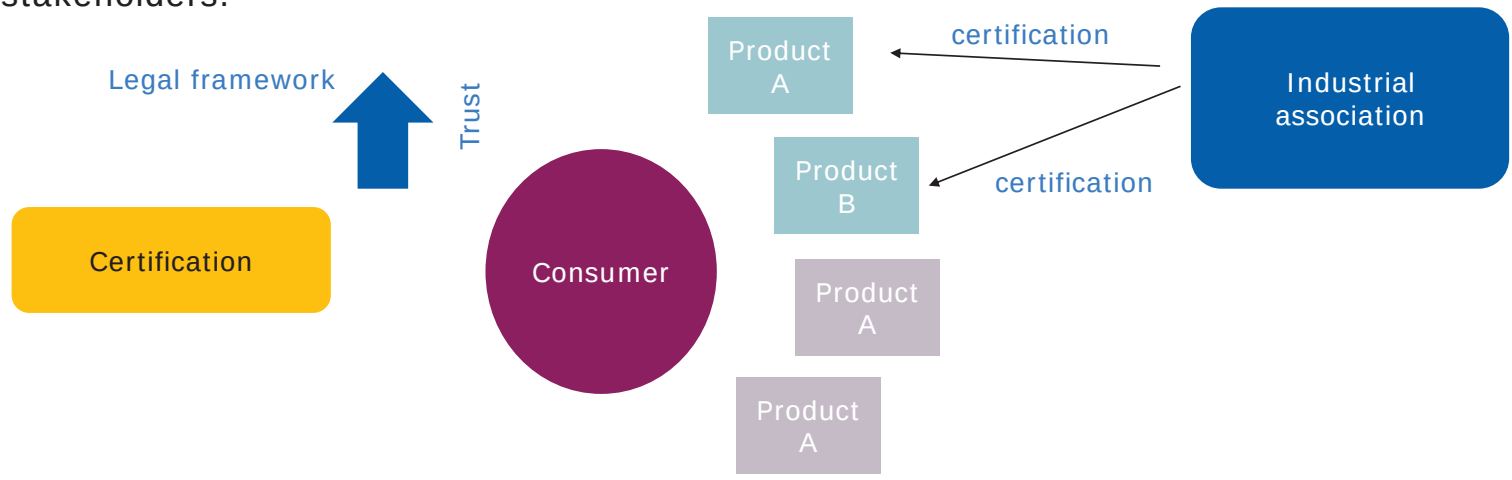
Reference architecture for trusted AIoT systems: certification, standardization and regulation

Ramiro Robles

InSecTT has received funding from the ECSEL Joint Undertaking (JU) under grant agreement No 876038. The JU receives support from the European Union's Horizon 2020 research and innovation programme and Austria, Sweden, Spain, Italy, France, Portugal, Ireland, Finland, Slovenia, Poland, Netherlands, Turkey

Certification and consumer trust

- Certification is the process by which a government or industrial association or regulatory body qualifies a product, service, goods or processes against predefined standards, norms or recommendations.
- This certification process aims to create trust in the end consumer or in involved stakeholders.

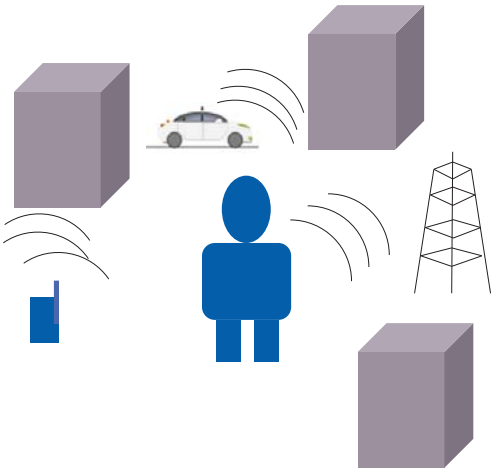


Certification and consumer trust

- The end consumers trust more in a product if it is certified by an industrial association or a regulatory body.
- The certification process can be complex, and it varies according to the type of product, country or region of world, and the consumer idiosyncrasy.
- Globalization has created the need to harmonize certification, standardization and regulation of multiple products, services or processes across countries.
- Labels, seals, marks, are the visible result of a product certification

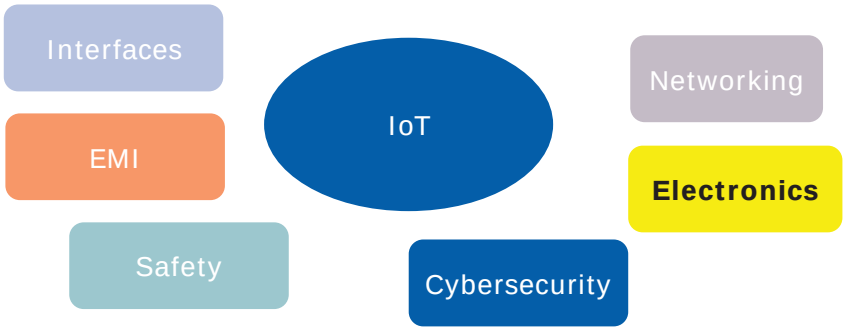
The Internet of Things

- A new paradigm for using small embedded processors, sensors, actuators in every-day objects, machines or wearable devices communicating with each other or with Edge/cloud infrastructure
- Objective: Bridge the gap between the physical world of objects and the virtual world of computers
- Huge implications on automation, control, Sensing of environmental parameters, machine to machine communication, automated industrial processes, etc
- Regarded as the new industrial revolution.
- Real impact of Internet and Wireless technology will be with the IoT or Internet of Everything (IoE)



The need for certification on the Internet of Things

- IoT promises a new generation of devices with embedded processors with networking capabilities.
- New services, products and issue will be created by the upcoming technological advances.
- Great challenge for certification entities due to the wave of millions of products, services and implications to human lives that the services are expected to bring.
- More complex certification issues. Products,

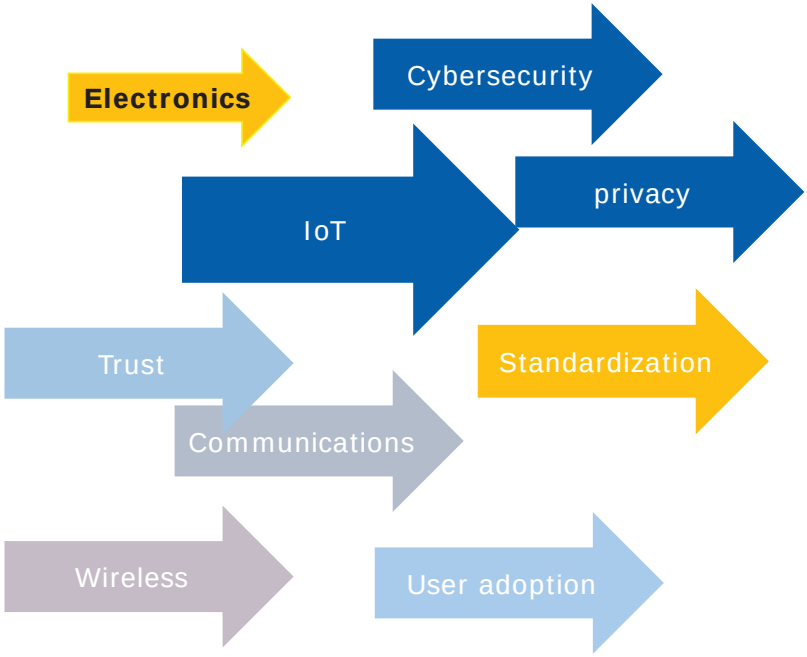


How did we get here?

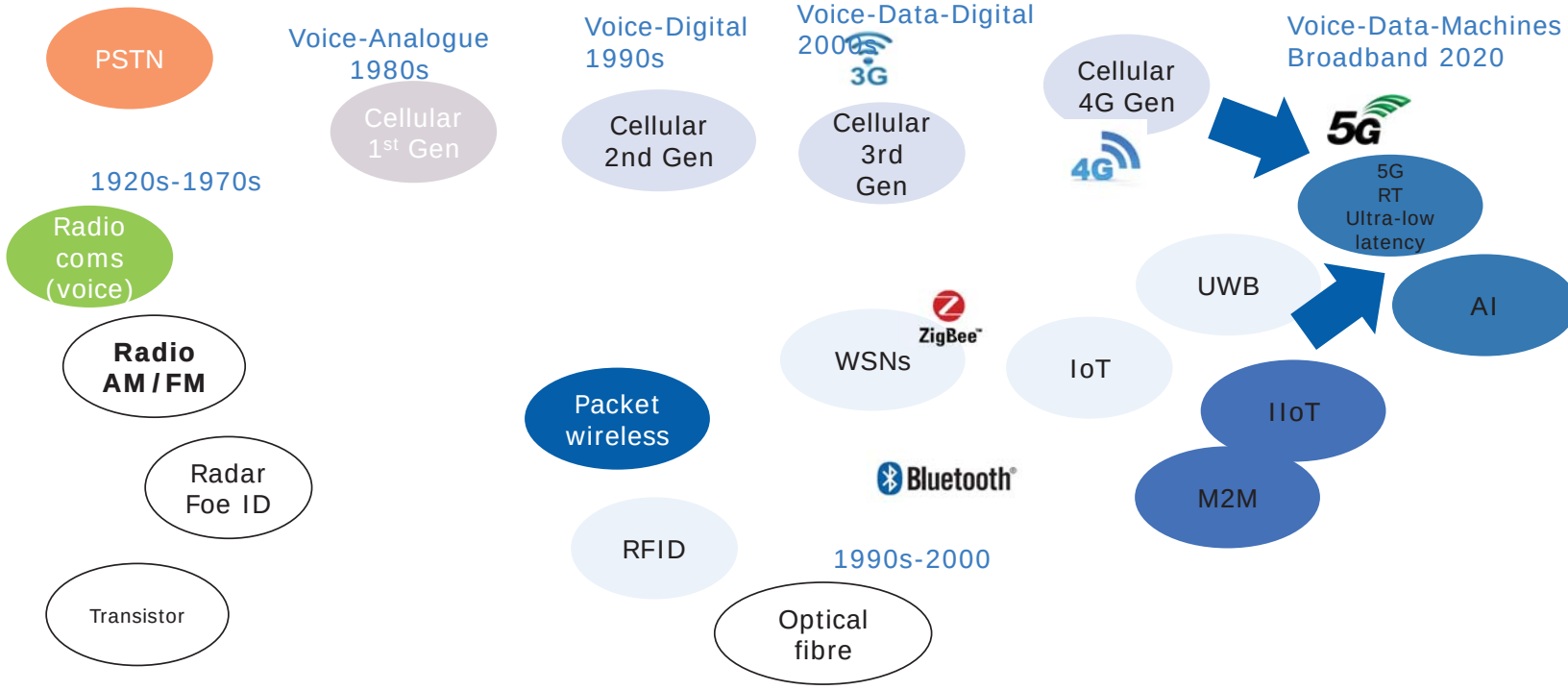
- IoT is an evolution of **multiple technologies**, market acceptance, and socio-economical and political landscape.
- **Electronic components** become more efficient and reduced in size. They are now cheap enough to be embedded in small devices hosting sensors and actuators that can detect/sense/modify environmental variables or control machines.
- **Communication technologies** are also evolving to deal with constrained resource devices, low encryption capabilities, and in future scenarios real-time machines and applications.
- **User adoption** has been evolving. The adoption of mobile communications provides a predecessor that can indicate openness of the market.
- At the political and social levels great efforts are being made to push for **standardization and certification of the potential new devices**. This obviously has strategic global implications and great national security risks that different countries are taking seriously.

How did we get here?

- Privacy by design
  - GDPR
  - Europrise seal
- IoT trustworthiness label
- Cybersecurity budget in Europe increase substantially
  - Creation and strengthening of the ENISA

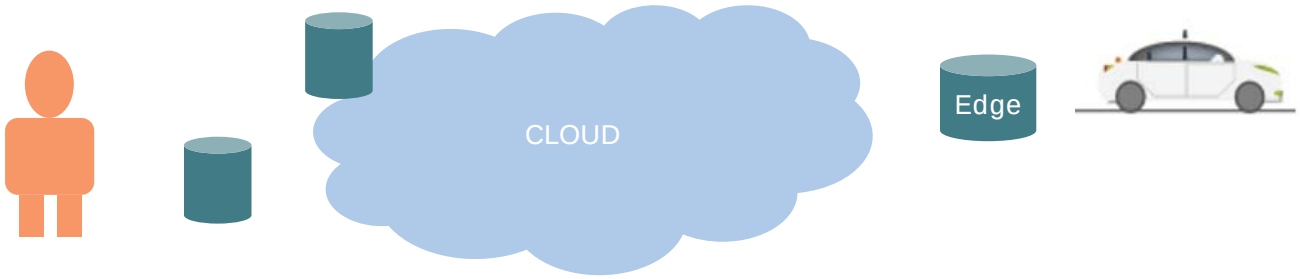


Technology evolution



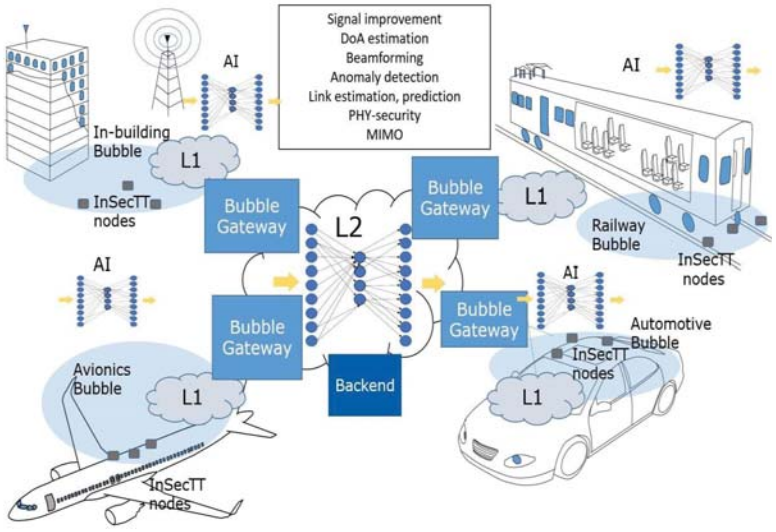
Cloud /Edge processing

- Recent advances in Internet transport technology, storage and capacity have pushed for the use of centralized application and data processing.
- Large number of sensor readings and data can be processed in the same location (cloud)
- More accuracy is achieved by using large amounts of data for optimization (Big Data)
- Recent years propose the use of Edge processing to reduce latency and security



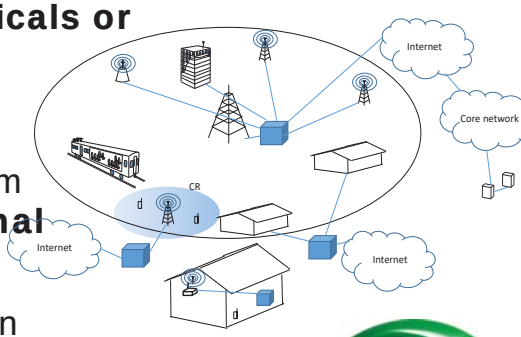
Artificial Intelligence and IoT

- **Artificial intelligence** is experiencing a boom in the last few years. Ability to learn from existing data and predict afterwards.
- It matches perfectly the **IoT as a source of data to be processed in the cloud or Edge**
- Machine learning can be used to refine, detect, predict events based on the collected information from objects and embedded processors.
- The fusion of AI and IoT, also called **AIoT, is one of the main technologies that will boom in the coming years. It promises to bring IoT to a new level**. For example: critical industrial applications



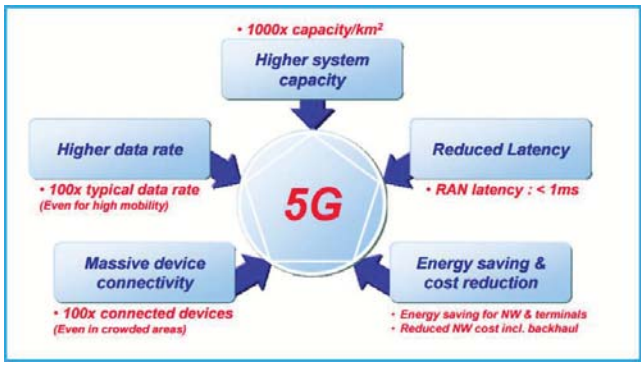
What is 5G?

- 5G is the **next generation of wireless cellular infrastructure**.
  - Convergence of cellular, wireless sensor, IoT, M2M and CPS systems
  - New standards for mobile comms are named as generations
- A generation is a set of complex **industrial technology standards** that lead to **business**, **marketing**, and multiple sets of **verticals or applications**
- Generations may differ from region to region
- They are endorsed by the technological leaders of the telecom industry, and they have **strategic regional and international importance**
- Generations are needed because there is a natural advance in technology that allows new features
  - **Features can be enabled by market need** or
  - **Features can enable market needs**



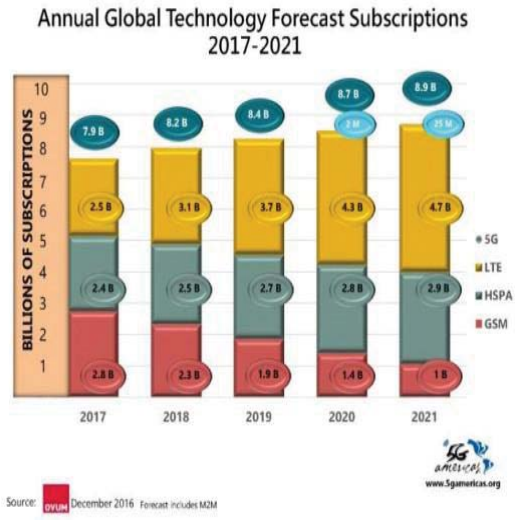
Objectives

- 5G has four main **objectives beyond 3G/4G systems**:
  - **higher capacity,**
  - **ultra-low latency for machine-type traffic support, and**
  - **dense object connectivity demand (IoT or Internet of things)**



Motivations

- 50 billion devices will be connected to the cloud in 5-10 years time
- **Industrial automation** (M2M) and cyber-physical systems are expected to proliferate thanks to wireless pervasive connectivity (e.g., automated driving and structure health monitoring)
- Current cellular technologies cannot cope with the **scalability** of large numbers of "things" connected wirelessly with **ultra-low latency**
- 3G/4G solutions were designed for **human users**, not for machines.
- **WSNs cannot achieve cellular coverage** service for industrial IoT



Is 5G bad for health?

- 5G will use a mmwave spectrum
  - Shorter wavelengths can have higher impact on smaller objects and human organs. However, the mmwave signals decay really fast in space.
  - **Cell sizes** are much smaller than mm waves. Cell mutation is not likely to occur.
  - 5G also uses **MIMO technology** that can potentially reduce Tx power, can also avoid body absorption.
  - 5G technology is also more **efficient in terms of power**
  - Mm wave spectrum is expected to be used in highly dense urban scenarios, not for home or indoor environments.
  - KEEP and EYE on the product: **exposure x power x frequency**
- **CONCLUSION**:
  - **More studies need to be conducted, but the issues raised by some groups seem to be not likely or with low probability**

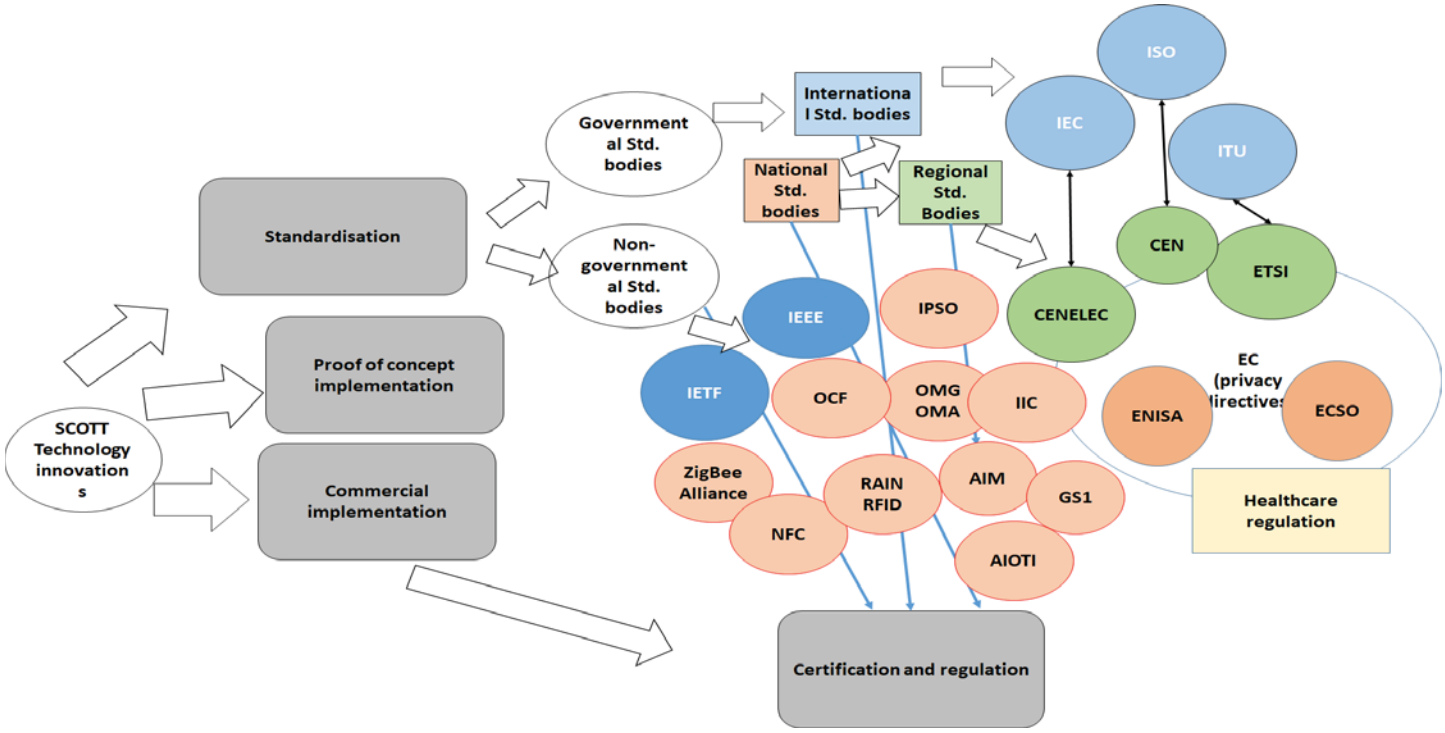
5G and IoT

- The advent of 5G means
  - objects can have a direct access to the cloud/Edge
  - Low cost per device/area
  - ultra low latency,
  - wider coverage than with other technologies
  - Higher capacity,
  - higher scalability
  - Experience with mobility, handover, roaming, etc.

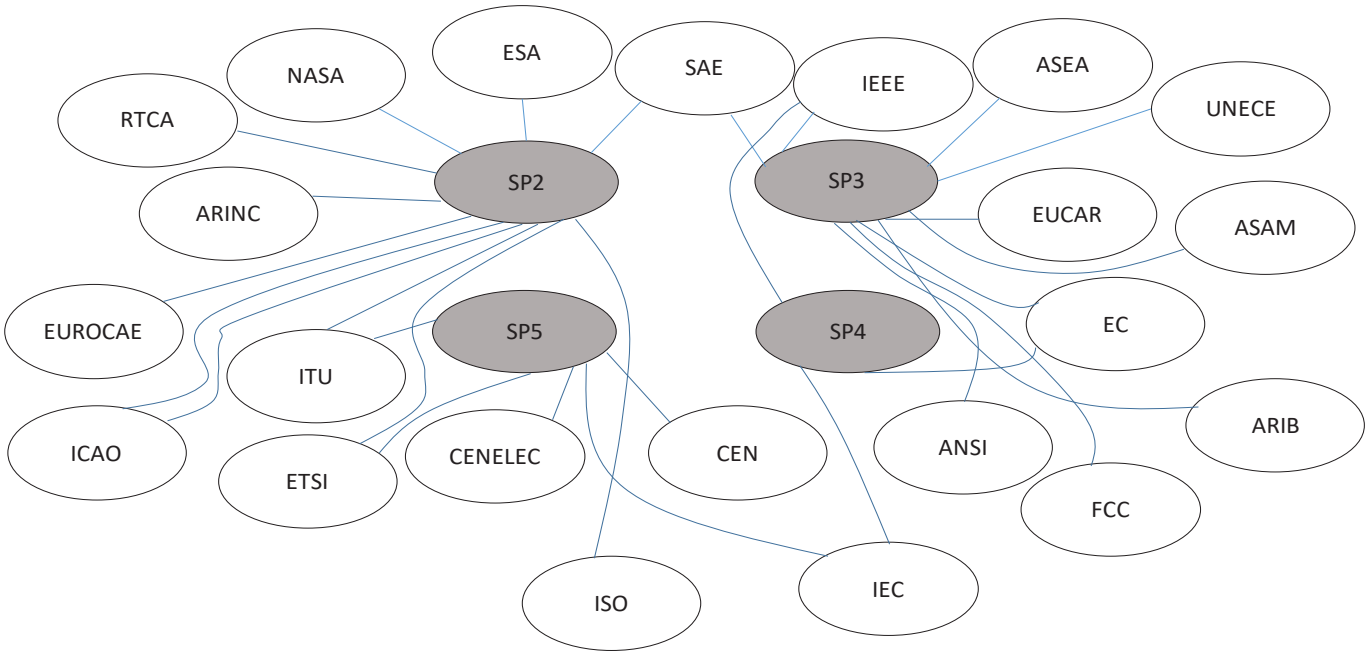
Standardization

- The process of making something conform to a standard
- Industrial associations, regulation bodies, government entities agree on **standards or norms** for different products, services, goods, etc.
- Certification is based usually on standards. Not usually conducted by the same entity. **Standards are released so that product conform to these guidelines**
- Standards are the basis of modern economy
  - Rules for interoperability, legitimate and fair competition between products, goods, services, etc
- Standards are the basis of the **boom of telecommunication systems**
  - Interface definition, compatibility between components, quality of service ensured, etc.

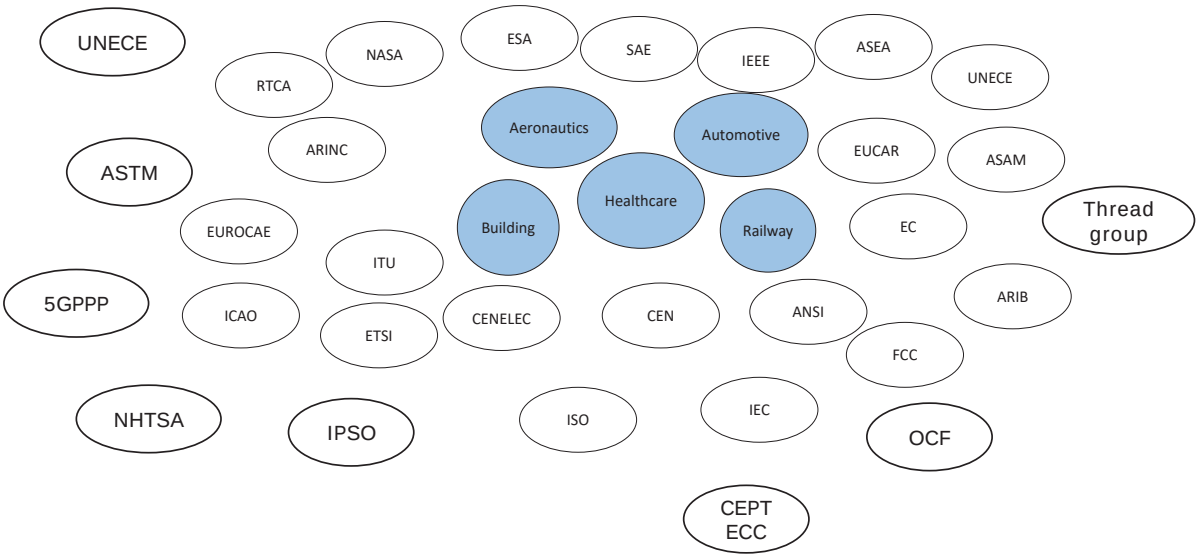
Standardisation and regulation bodies



Per industrial domain

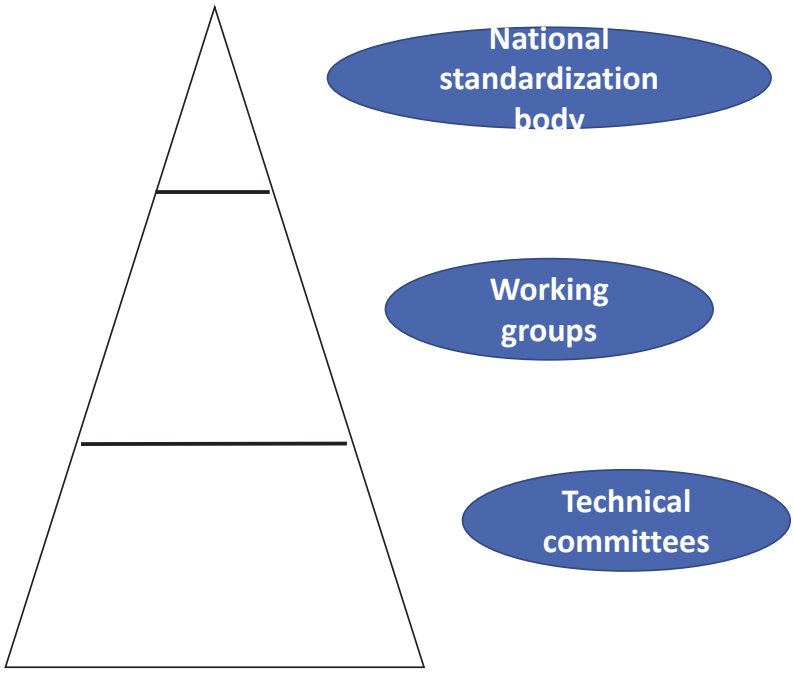


Regulation framework

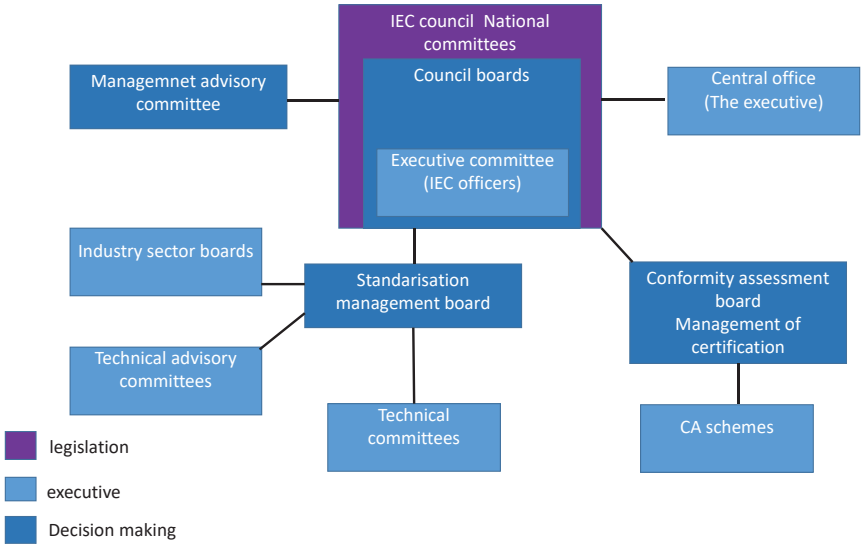


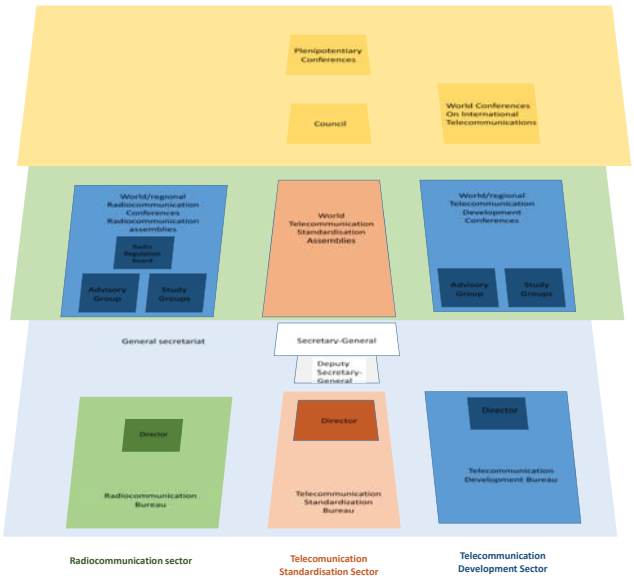
Typical structure for standardisation national bodies

- Each country has a pyramidal standardization and regulation framework
- There is a need to approve, create or correlate international standards to be adopted at te national levels.
- Usually the work is split into technical committees and working groups.

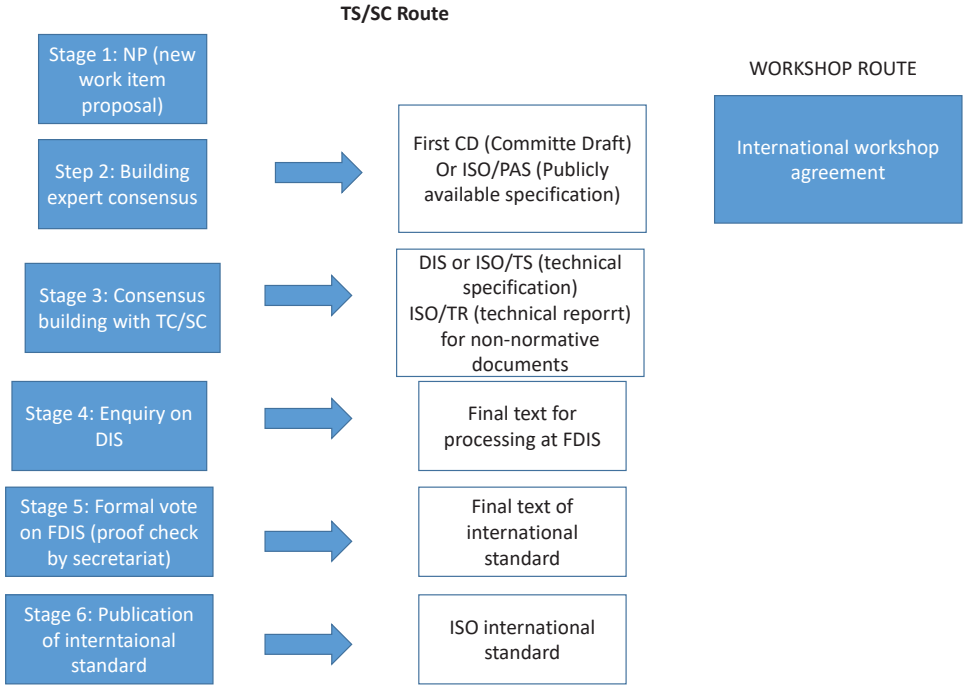


Examples of organization of Standardization and regulation bodies



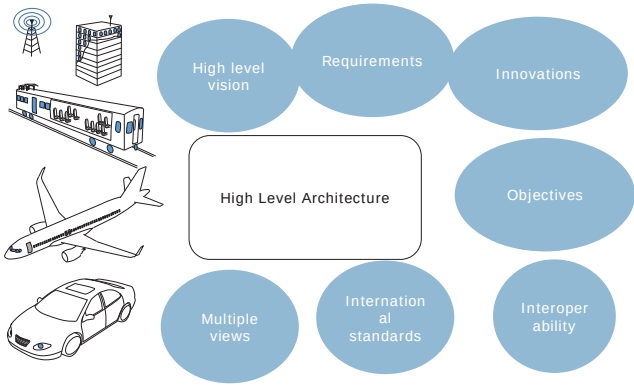


Example of standardization process (ISO)



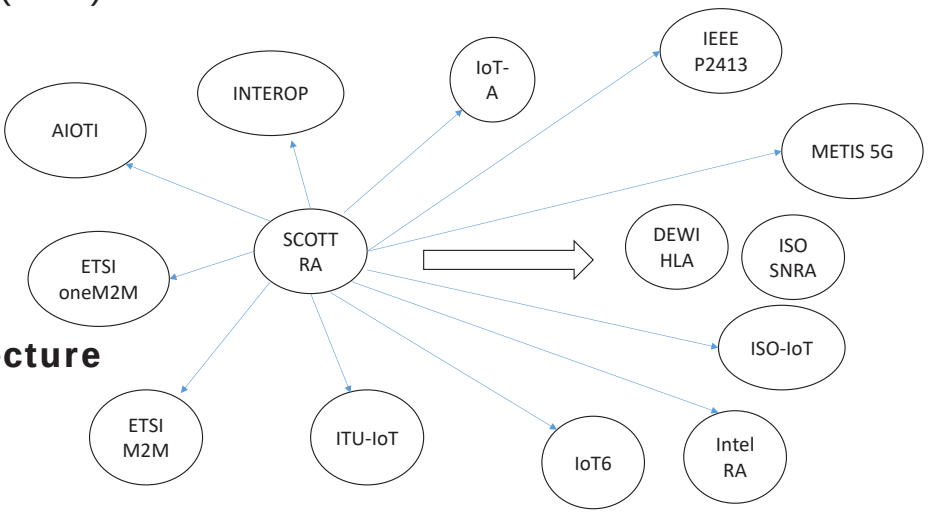
What is a reference architecture?

- A generic architecture useful for design particular instances and use cases aligned with international standards and guidelines
- A high-level overview for infrastructure design and application overview
- Not a simple conexion of boxes anymore
- Modern reference architectures are a complex collection of views or perspectives, each one providing particular insights that highlight different feature for different stakeholders
- Reference architecture ae basis for alignment, design, validation, verification standardization and certification of modern IoT systems.



Modern IoT architectures

- IoT Architecture reference model (ARM)
- IEEE IoT architecture
- ISO reference architecture
- ITU reference architecture
- ETSI M2M architecture
- **AIOTI architecture**
- **DEWI/SCOTT/InSecTT architecture**

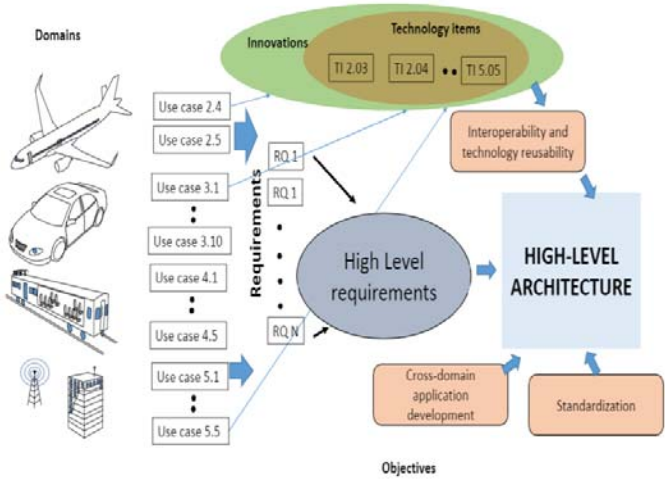


DEWI/SCOTT/InSecTT projects

- DEWI (Dependable Embedded Wireless Infrastructure)
  - Dependability of sensor networks
  - Interoperability
  - 15 use cases, 10 technology items, 57 partners
- SCOTT (Secure Connected Trustable Things)
  - Security, privacy, trust
  - Internet of Things
  - 15 use cases, 10 technology items, 57 partners
- InSecTT (Intelligent Secure Connected Trustable Things)
  - Artificial Intelligence , Edge Computing
  - Internet of Things
  - 15 use cases, 10 technology items, 57 partners

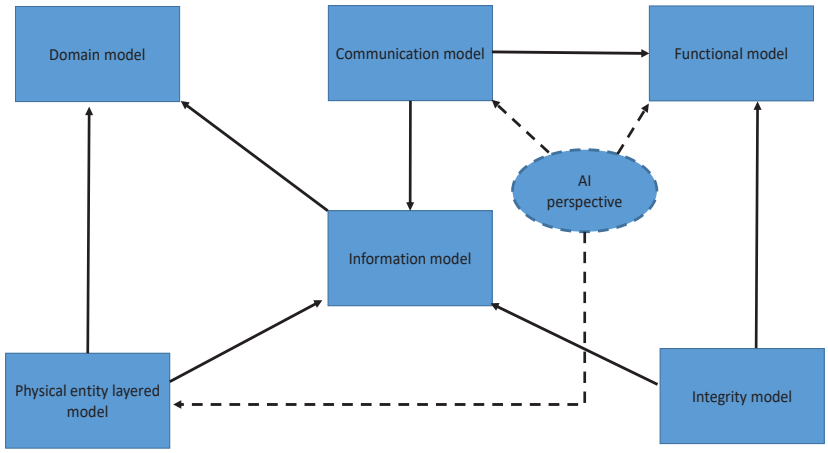
DEWI/SCOTT/InSecTT Reference architecture

- **Definition**: "A set of guidelines for infrastructure organization of IoT use cases supporting the objectives of the projects"
- The framework for a **high-level analysis** of all building blocks of use cases in different industrial domains
- **Interface** and **vulnerability** analysis per layer and entity.
- Framework for **reusability** and **cross-domain** interpretation
- High level perspective of use case requirements, road-map, and forecast analysis
- Compilation of **expertise** accumulated across different **use cases** in different **industrial domains**.
- Framework for **standardization** needs in detail (forensic analysis)



INSECTT Reference architecture

- The INSECTT Reference architecture consists of multiple views or perspectives of a generic IoT system
- The multiple views approach is useful for modern IoT use cases with multiple stakeholders
- The INSECTT RA consists of
  - Entity model
  - Functionality Model
  - Information Model
  - Domain Model
  - Communication model
  - Ontology model

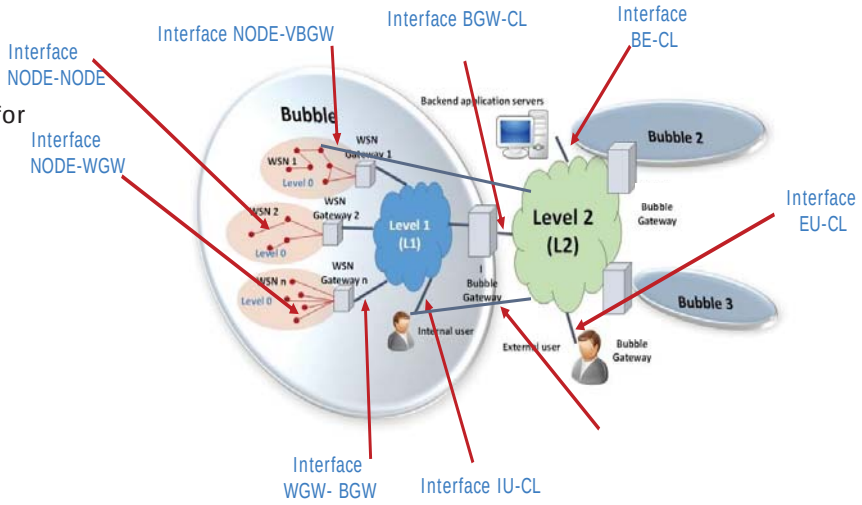


Entity model

- **Bubble** is a high-level WSN with a unique gateway (Bubble Gateway), controlling wireless and wireline infrastructure.
- Attributes:
  - Interoperability (single protocol or semantics model for interoperability)
  - Integration of new and legacy critical industrial

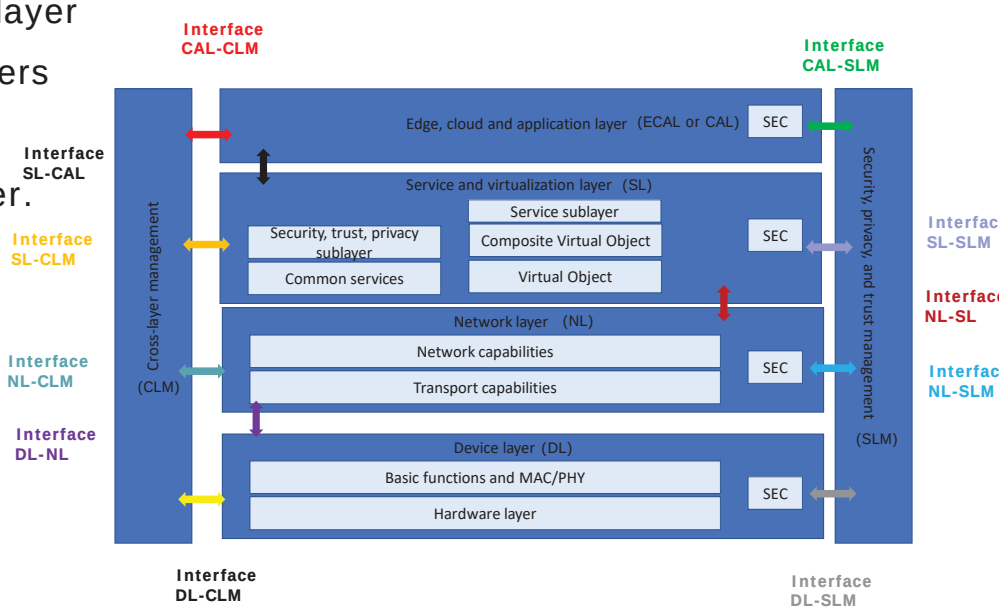
sensors to a modern IoT infrastructure

- Three-level organization
  - L0 Wireless
    - Nodes and WSN Gateway
  - L1 wireline- existing critical infrastructure
    - For example: aeronautical internal bus, CAN bus
  - L2 interoperability
    - Cloud, Edge servers. The Bubble Gateway can also act as fog or Edge server.



Functional model

- Hybrid model combining ISO, ITU, AIOTI, SNRA, and AIOTI
- Security functionalities on each layer
- Software interfaces between layers
- HW interfaces between entities
- Trustworthiness metrics per layer.



The Bubble and high-level architecture evolution

Based on ISO/SNRA
Interoperability ETSI M2M, IoT-ARM
L0/L1/L2 layering for Wireless/wireline
Model

Dependability inside the bubble
Integration Wireless/wireline industrial WSN and IoT
Cross-domain reusability
Interoperability
Integrated sensors into IoT

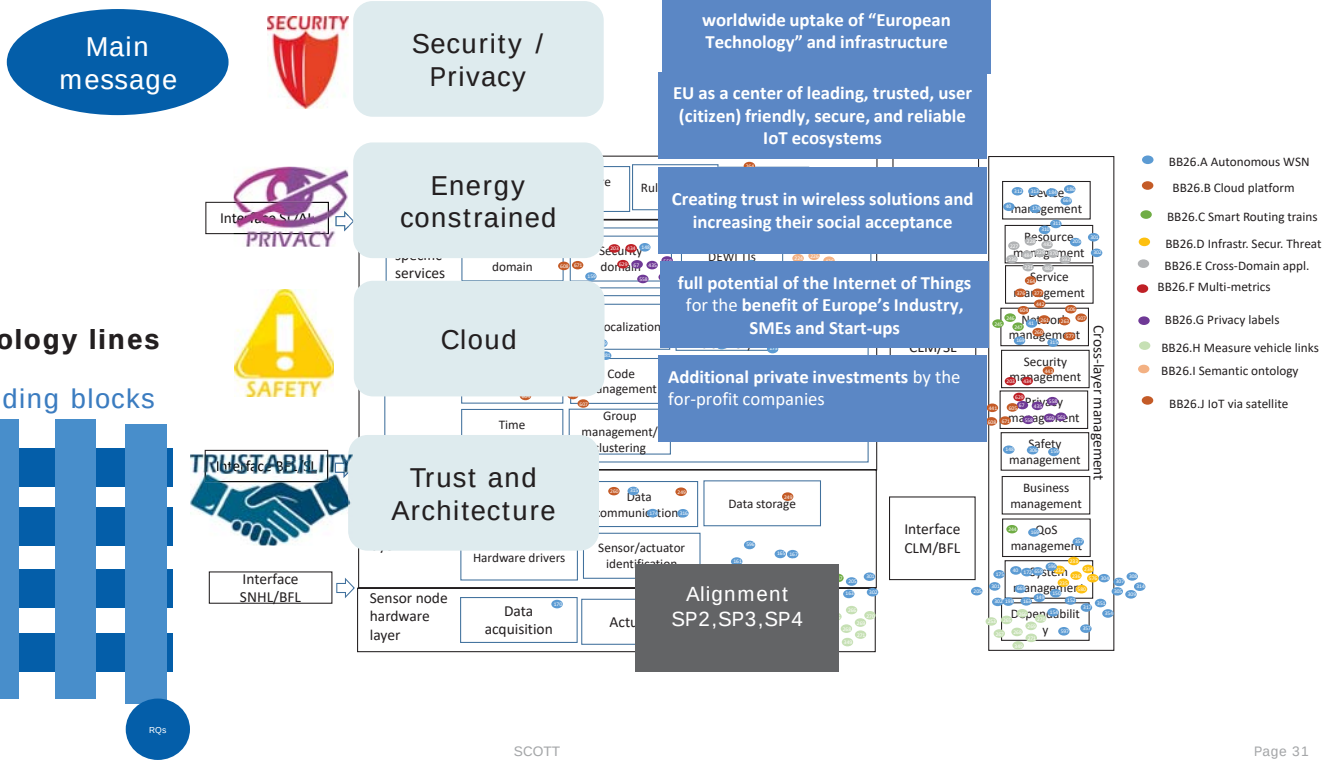
Full IoT architecture (around the bubble)
Hybrid ISO SNRA ITU, ISO, AIOTI, IEEE IoT architectures
L0/L1/L2 layering for Wireless/wireline
Security sublayers and processes

Dependability inside the bubble
Integration Wireless/wireline industrial WSN and IoT
Cross-domain reusability
Interoperability
Integrated sensors into IoT
Trustworthiness and security metrics
Bubble gateway as Edge processor
Inter-bubble communications based on trust indicator
Blockchain compatibility

Full IoT architecture (around the bubble)
Hybrid ISO SNRA ITU, ISO, AIOTI, IEEE IoT architectures
L0/L1/L2 layering for Wireless/wireline
Security sublayers and processes
Specific AI models and impact analysis

Virtualized Bubble
Multiple connections inside the Bubble
Long and short-range communications
Direct cloud connections inside the bubble an for internal users

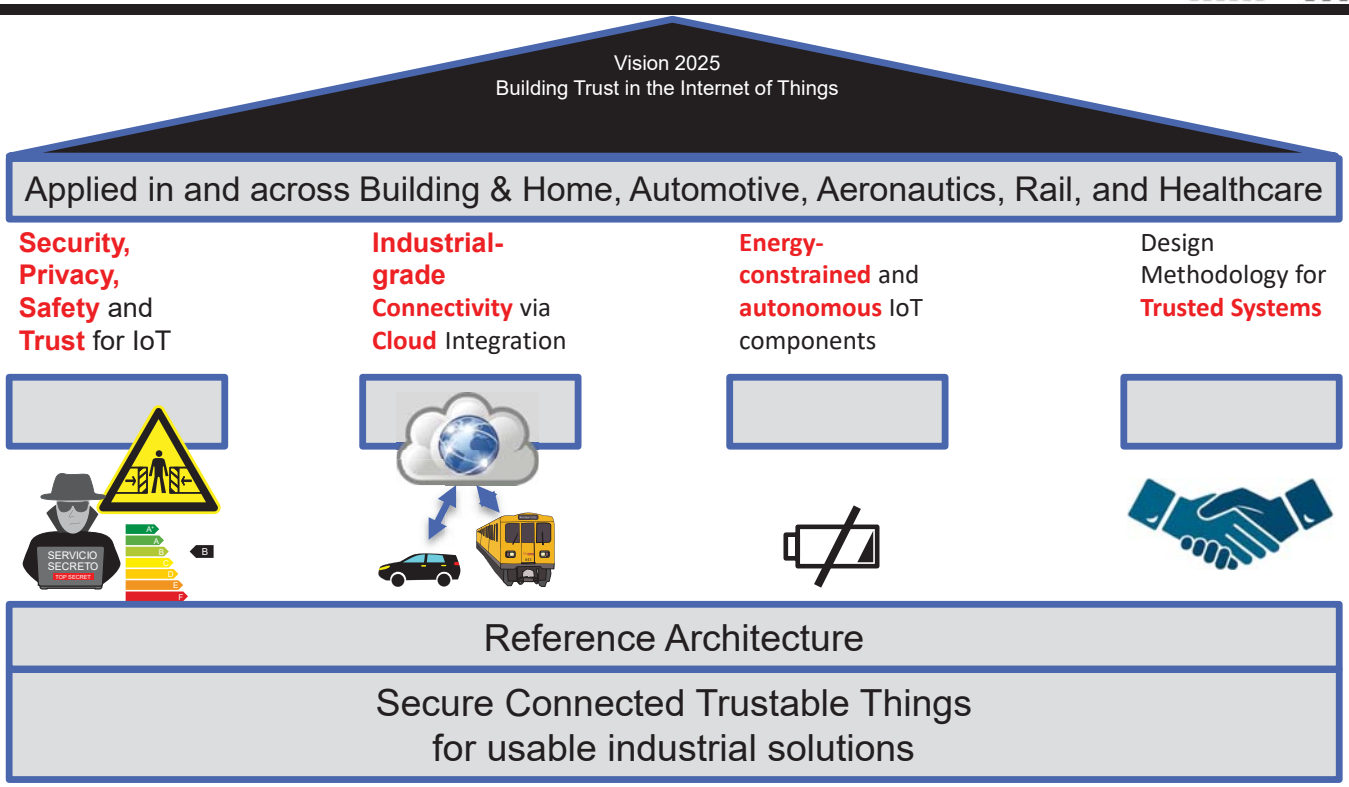
Impact and main message



Key Messages

- Boosting **Security, Privacy, Safety and Trust** for IoT
- Ensuring **Industry-compliant Connectivity** via **Cloud** Integration
- Developing Innovative **Energy-constrained** and **Autonomous** IoT Components
- Providing a **Reference Architecture** for Secure Connected Trustable Things demonstrated across 5 Domains
- Design a scientifically sound yet practical **Methodology** for developing **Trusted Systems**

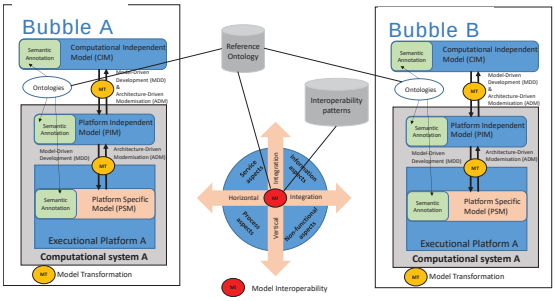
SCOTT Vision

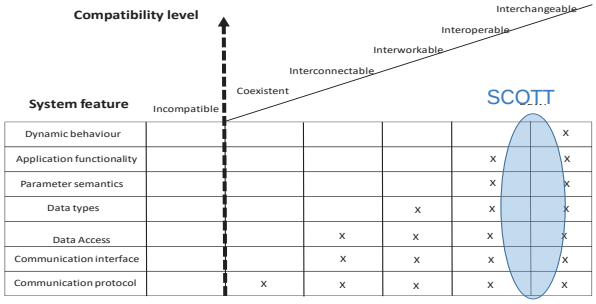


SCOTT Bubble

- SCOTT envisions an Internet of Bubbles, each bubble encapsulating legacy or state of the art industrial technologies (wireless and wireline) for sensor and actuators.
  - The encapsulation enforces dependability and security inside each domain network.
- Interoperability in SCOTT: Re-use of building-blocks, Co-operability in wireless context, and Co-existence with other wireless standards. (Mapping to the IEC concept of interoperability)
- Model driven design helps us introduce interoperability together with non-functional requirements such as security, privacy and trustiness in the Bubble
- Ontologies at the three different modeling levels of the INTEROP architecture match the Bubble infrastructure and the SCOTT framework for security, privacy and trustiness

INTEROP perspective of Bubbles





IEC interoperability mapping

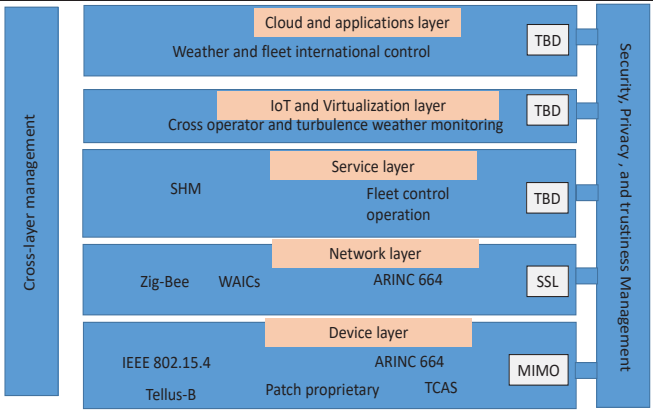
What do you get by following the Bubble specs?

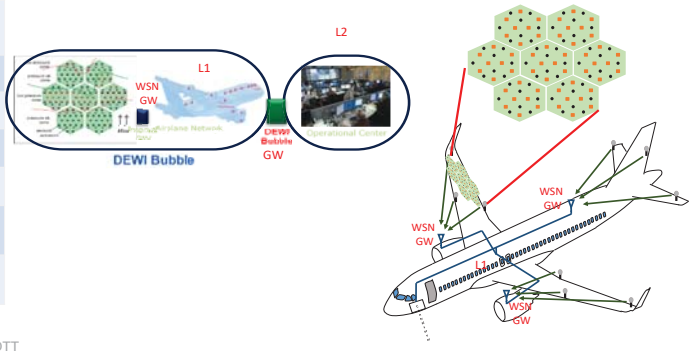
- **Guidelines** to achieve **dependability, security, safety, privacy and trustworthiness** inside the bubble
- Specific measures for interaction between **Wireless and wireline infrastructure with real time constraints**
- **Cross-domain interoperability**
- **International IoT standards compatibility**
- **Trustworthiness multi-metrics evaluation** (extension of ARMOUR and SCOTT metrics)
- **Privacy and trustworthiness by design** approach
- Collected experience of real **industrial use cases**
- Per layer, per interface and per entity **trustworthiness metrics analysis**
  - **Extension or ARMOUR and ETSI metrics (CWSS, CVSS)**
- Integrated **trust methodology** to include end user and stakeholder perspective

Alignment of use cases

- Example of alignment of an industrial use case (aeronautics)
- Dense sensor flow control
- **Entity, functionality, entity vs. functionality**

		L0 device to/from L0 GW		L0 Device to/from L0 Device		L0 GW to/from L1 GW		Bubble L1 to/from Internal User		Bubble GW to/from Service provider/Cloud	
Cloud and applications layer		-	N/A	-	-	N/A	-	Operator fleet control	-	Operator fleet control	Encryption
IoT Virtualization layer		-	N/A	-	-	WAIC server	-	Avionics layer	-	Avionics layer	PHY-access
Service layer	Security, trustability and privacy	SSL	SSL	-	-	SSL	-	SSL	-	SSL	-
	Common services	Flow control	HTTPS	-	-	WAICs	-	WAICs	-	HTTPS	-
Network layer	Transport	UDP	SSL	-	-	VL	-	VL	-	HTTP	-
	Network	ZigBee	Encryption	-	-	ARINC 664	-	ARINC 664	-	IP	-
Device layer	Basic functions and MAC/PHY layers	IEEE 804.15.4	MIMO-based	-	-	ARINC 664	-	ARINC 664	-	Ethernet	TBD
	Hardware layer	Pressure sensors, micropumps, TTP	Compression	-	-	ARINC 664	-	ARINC 664	-	Ethernet	-

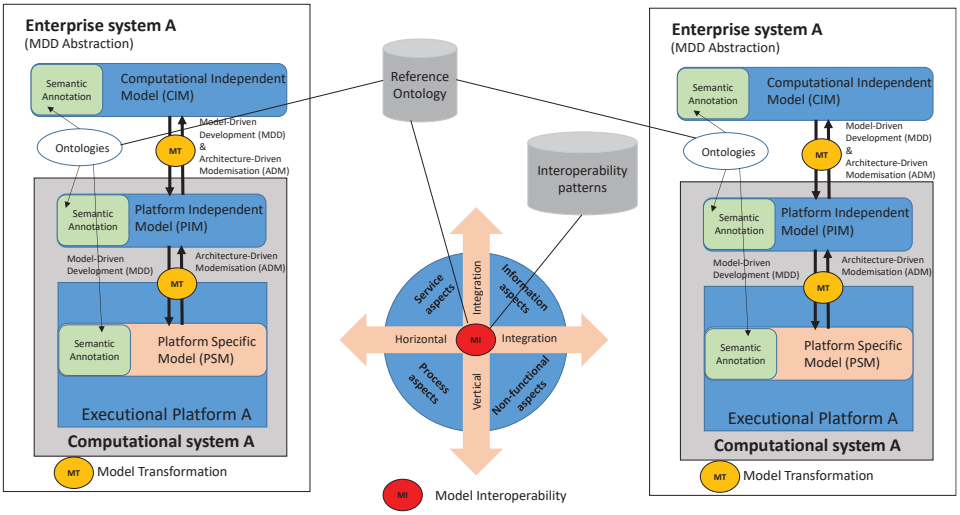




Reference model for conceptual integration (model driven)

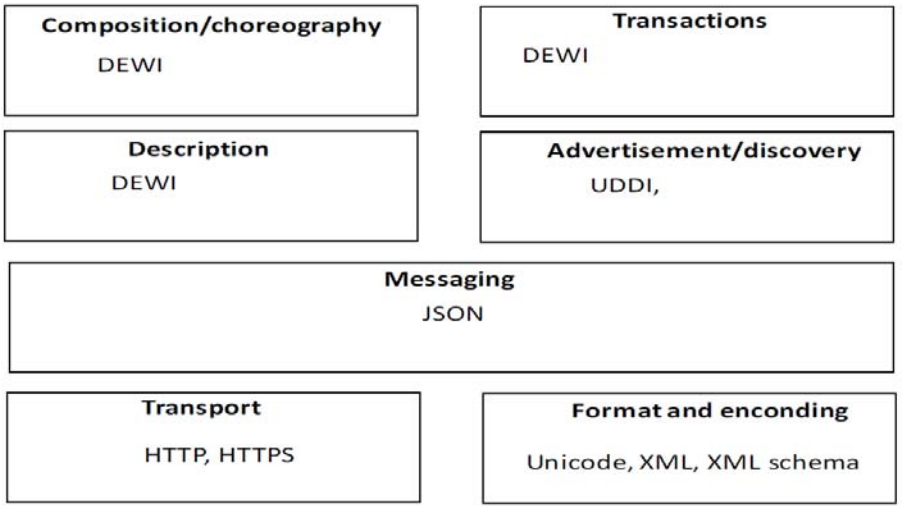
- **Model driven architectures for interoperability** allow us to introduce security solutions in the interaction between entities or layers inside the same entity
- The **INTEROP project** proposed the model driven interoperability view with a three – layer ontology approach to catch the different stages of model development.

- **CIM (Computational Independent Model)**
- **PIM (Platform Independent Model)**
- **PSM (Platform Specific Model)**



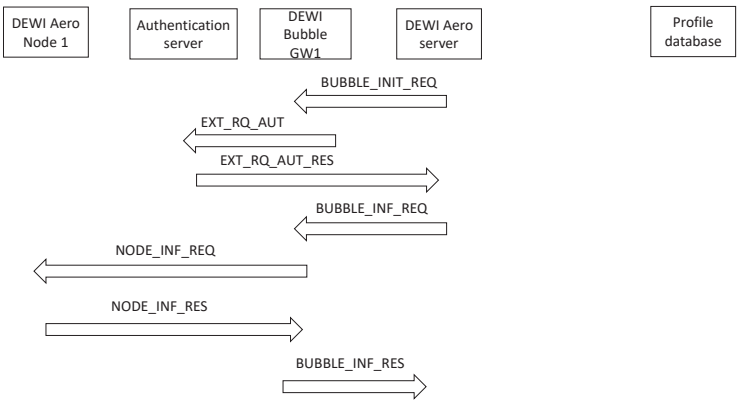
Information model

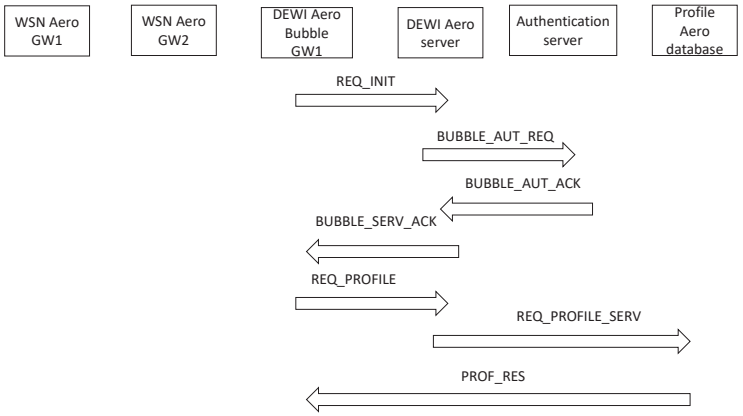
- **The information model describes how the information between entities or layers is exchanged, described, organized, transported, etc.**
- **Format / encoding**
- **Transport.** How to send the information to the destinations
- **Messaging.** How information will be understood by the receiver
- **Description.** How resources are described
- **Advertisement.** How resources are advertised
- **Composition. Business layer**
- **Transactions**



Communication model

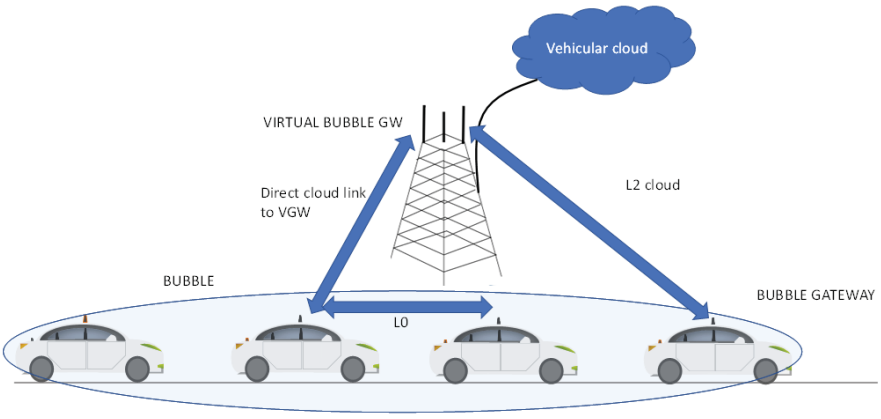
- The description of communication protocols between the entities and functionalities of the architecture

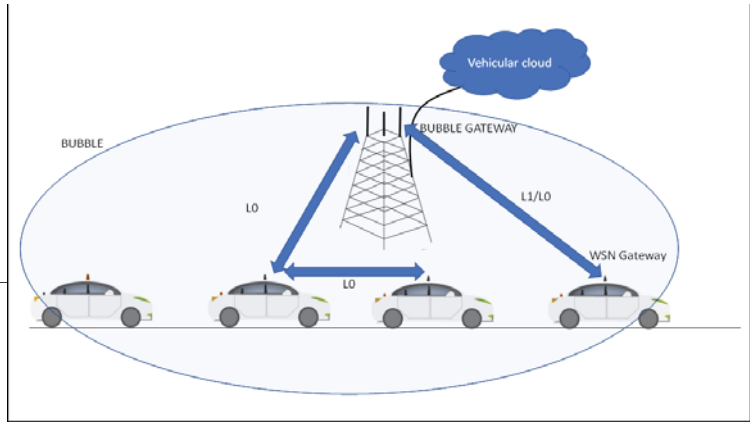




Example Vehicle platoon

- In a vehicle platoon, different elements can be assigned to different responsibilities according to the Bubble model:
  - Option 1: BS as Edge BGW
  - Option 2: Leader as EBGW



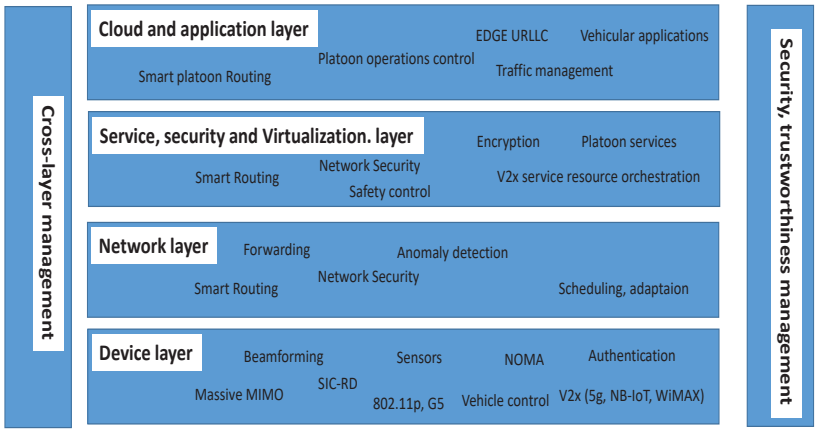


- Two types of links are recognized: V2V and V2I

Example vehicle platoon

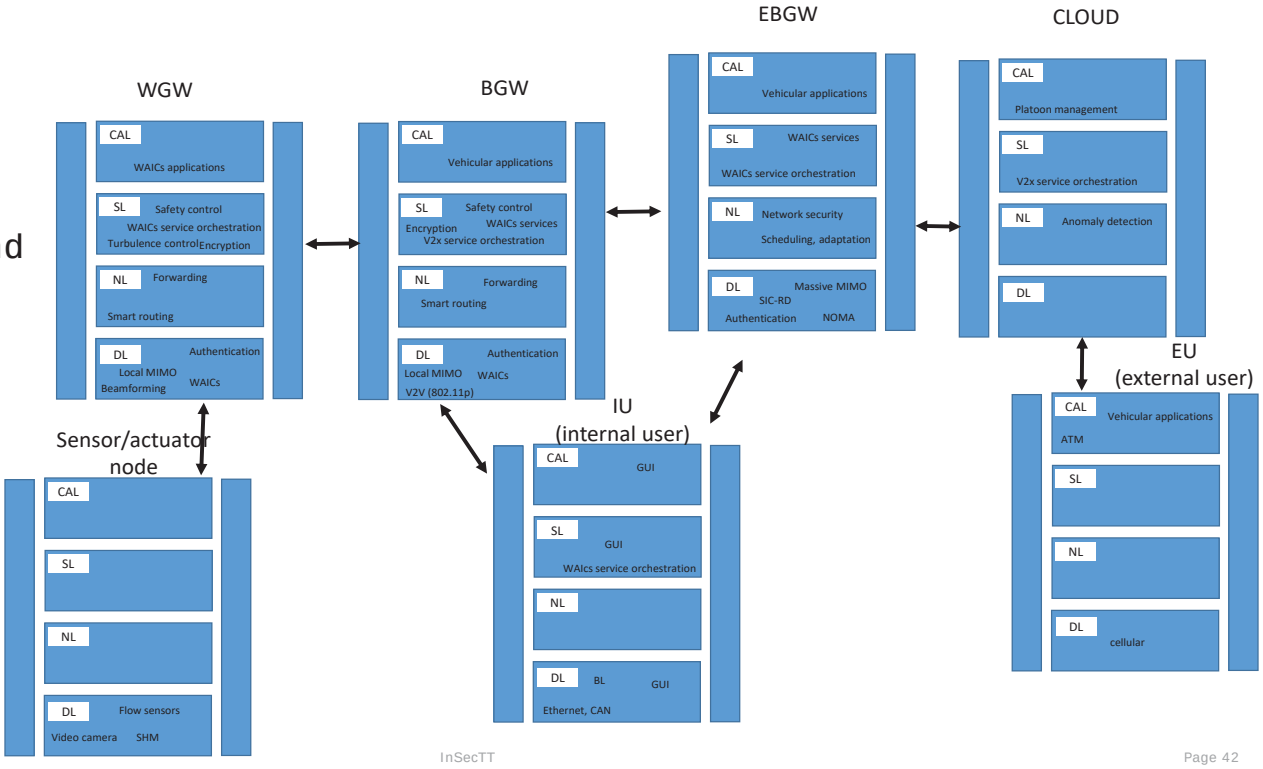
- The different functionalities can be arranged according to the functionality layered model.
- Interfaces between different functionalities or grousps of functionalities can be analyzed in a preliminary fashion

- The different functionalities can be arranged according to the functionality layered model.
- Interfaces between different functionalities or grousps of functionalities can be analyzed in a preliminary fashion



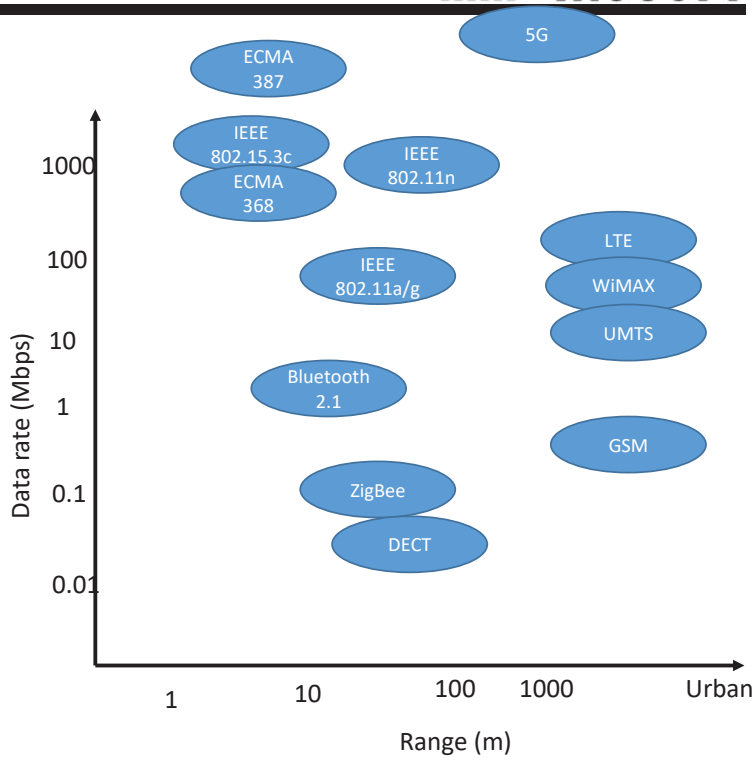
Example vehicle platoon

- Hybrid view
- Entity vs functionality.
- Detailed map of functionalities and interfaces



Level 0 interface technologies

- Multiple L0 technologies exist in the market with different properties.
- Data rate versus coverage distance is one of the main evaluation criteria
- Other criteria include scalability, latency, power consumption, etc.
- Complex decision making might be necessary to select the bets interface
- These L0 technologies are usually associated to other functionalities and standards of the architecture
- Tendency to cover higher speeds over wider ranges



PHY-layer requirements per industrial domain

- Each domain has PHY-layer properties that must be matched to the L0 technology to be used.
- The environment determines several of the parameters to select the most appropriate transmission technology.

- Industrial domains present different physical features that determine propagation issues such as multipath, path-loss, time variations

Feature/domain	Aeronautical		Automotive		Rail		Building	
	Inside	Outside	Inside	Outside	Inside	Outside	Inside	Outside
Multipath (frequency selective)	Dense	Not dense	Dense	Dense/regular	Dense	Not dense	Dense	Dense/ regular
Fast fading	No	Maybe	No	Maybe	No	Maybe	No	No
Path loss Exponent	Low	High	Low	High	Low	High	High	Low
Interference sensitivity	High	High	High	High	High	High	Medium	High
Interference to other systems	High	Low	High	Medium	High	Low	High	High
MIMO feasibility	Yes	Yes	Maybe	Yes	Yes	Maybe	Yes	Yes

MAC – layer requirements

- Each domain has MAC-layer properties that must be matched to the L0 technology to be used.
- The environment determines several of the parameters to select the most appropriate transmission technology.

Feature/domain	Aeronautical		Automotive		Rail		Building	
	Inside	Outside	Inside	Outside	Inside	Outside	Inside	Outside
Expected Data rate (depends on application) medium term	Expected medium-high	Expected Low-medium	Expected medium-high	Expected medium-high	Medium	Medium	High	Medium
Contention based/centralized	Both	Contention	Centralized	Contention	Centralized	Contention adhoc	Both	Both
Density of nodes	Medium to high	Low to medium (high for SHM)	High	Medium	Medium	Medium	High	Medium
Resources needed	High	High	Medium	High	High	Medium	High	Low

PHY-layer features

- Each technology can be selected to match the particular needs of each use case

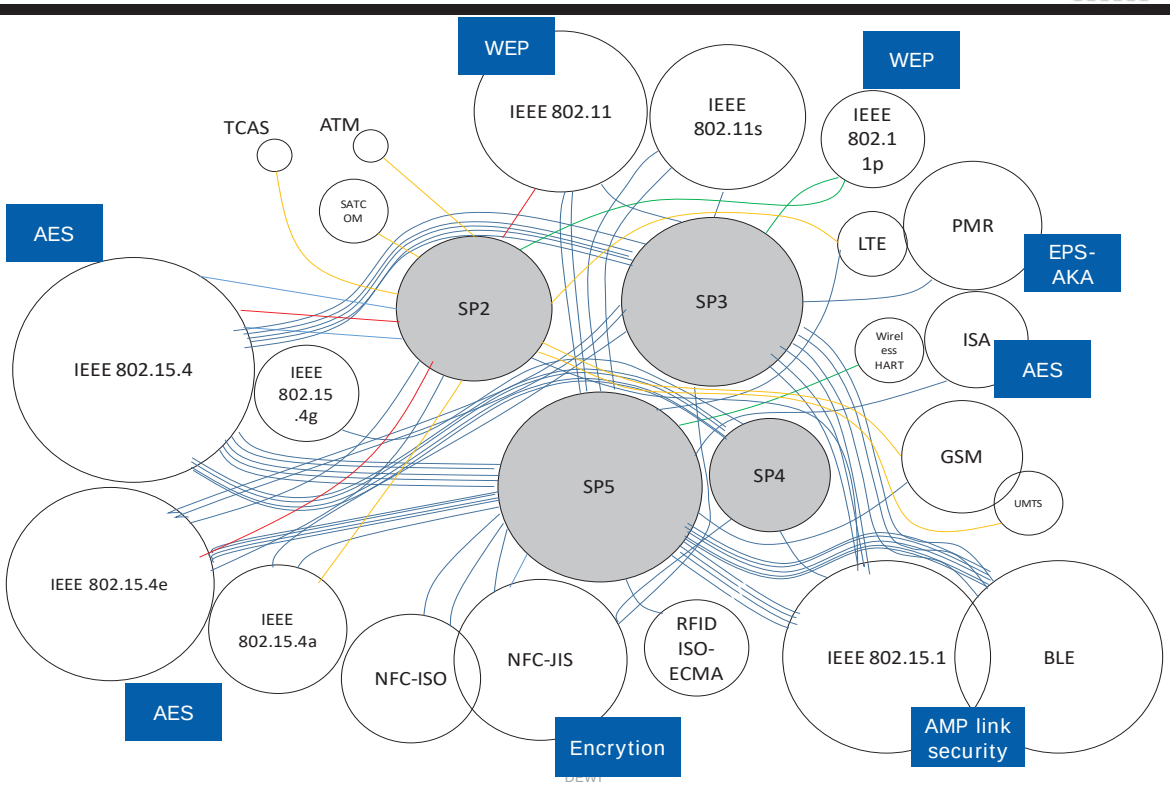
Feature/technology	802.15.4			802.11		ECMA standards		802.15.3	Bluetooth	
	ZigBee	ISA	15.4c	11.a,b,g	11n,	368	387	15.3	15.1	BLE
Multipath immunity	Regular	Regular	High	Regular	Regular	High	High	High	Regular	Regular
Fast fading immunity	Low	Regular	Very High	Low	Low	Very high	High	Very high		
Modulation /frequency(GHz)	DSCDMA (2.4)	DSCDMA (2.4)	UWB (3-6)	OFDM (2.4)	OFDM (5)	UWB	OFDM (60)	UWB (60)	FHSS (2.4)	FHSS (2.4)
Range	10-100m	10-100	10-100	10-100m	10-100m	10 m	5-10 m	1-10m	10-50m	10-20m
Jamming Interference sensitivity	Low	Low	Very low	High	High	Low	Low	Low	High	High
Interference to other systems	Low	Low	Very low	High	High	Low	Low	High	High	High
Availability	Very good	Good	Likely	Very good	Very good	Good	?	Likely	Very good	Very good
Other Signal characteristics	NLOS	NLOS	NLOS	NLOS	Indoor Low resources	Indoor high resource	LOS	LOS	NLOS	LOS
Cost	Medium	High	Very high	Low	High	High	High	High	Low	Low
Traffic density (Mbps/S/m2)	?	?	?	?	1.4	2.4	?	16	?	?

MAC-layer features (COTS technologies)

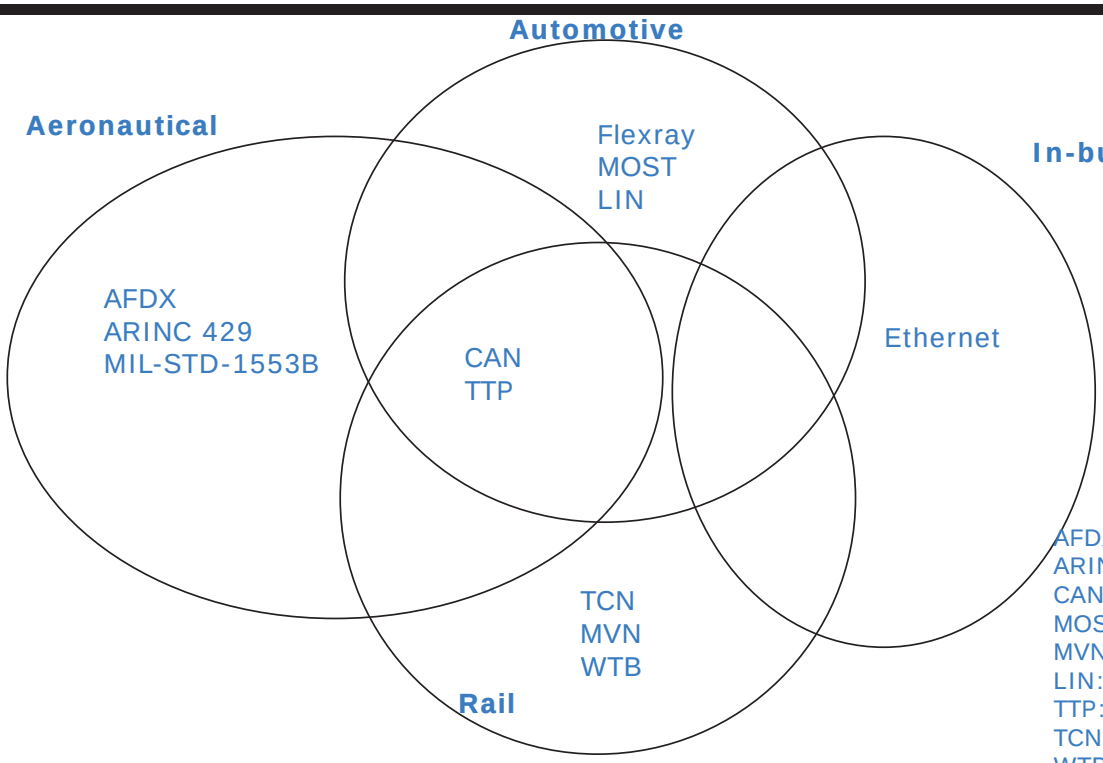
- Each technology can be selected to match the particular needs of each use case

Feature/ technology	802.15.4			802.11		ECMA standards		802.15.3	Bluetooth	
	ZigBee	ISA	15.4c	11.a,b,g	11n	368	387	15.3	15.1	BLE
Access	CSMA	CSMA	CSMA	CSMA	CSMA	FD/TDMA	FD/TDMA	CSMA	Master /slave	Master /slave
Discovery	Beacon	Beacon	Beacon	Beacon	Beacon	?	?	Beacon	On demand	On demand
Scalability	10-10000	10-10000	10-10000	10-100	10-100	?	?	?	10-10000	10-10000
Frame length	15ms	15ms	15ms	variable	10-100m	10 m	?	?	??	?
Multi-hop	Yes	Yes	Yes	Weak	Weak	?	?	Yes	No	No
Self configurability	Yes	Yes	No	Weak	Weak	?	?	?	Yes	Yes

Wireless standards mapping

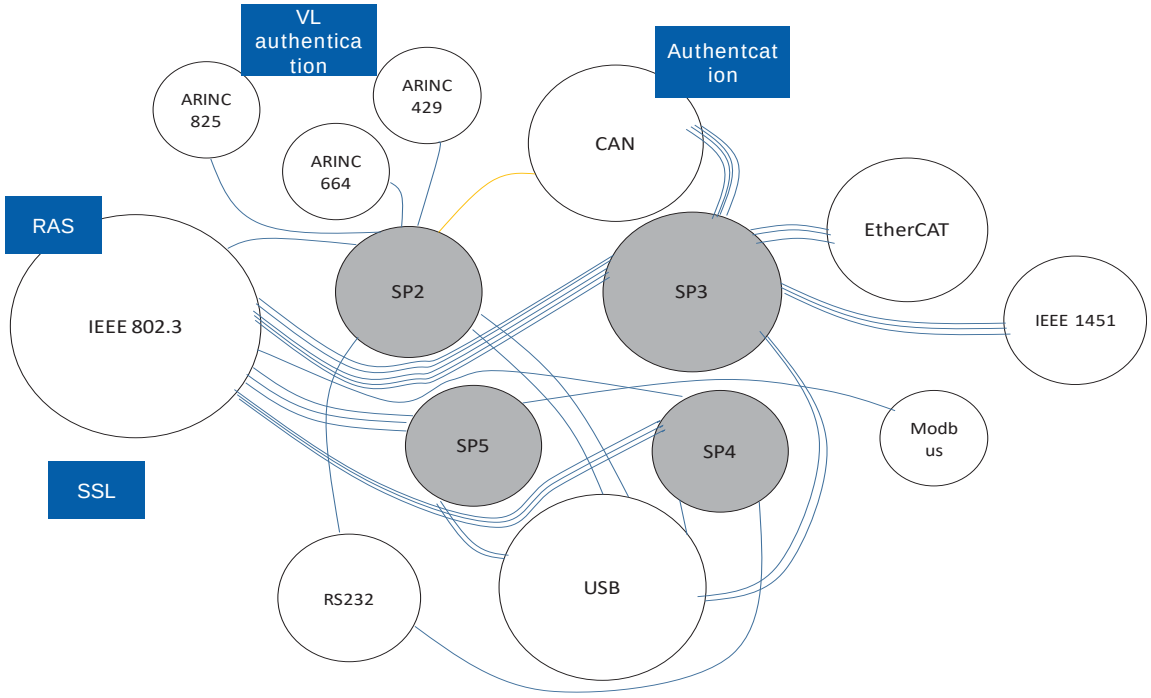


Existing infrastructure (buses)

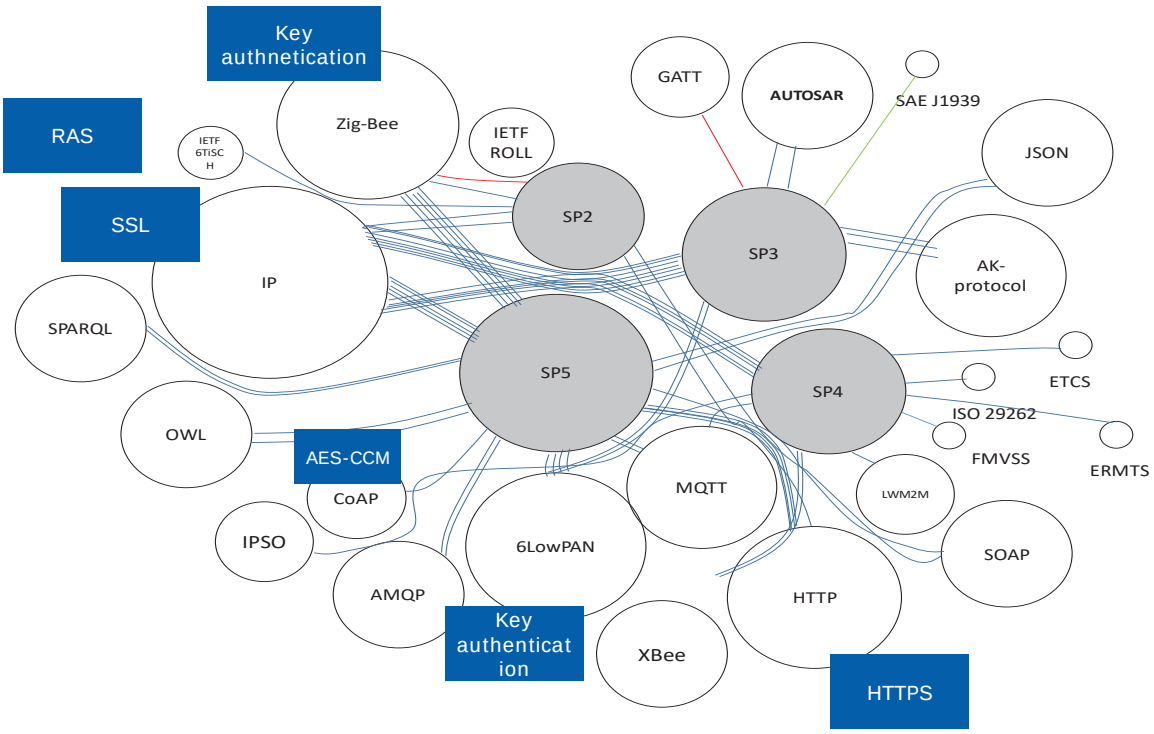


AFDX: Avionics Full-Duplex Switched Ethernet
ARINC: Aeronautical Radio, Incorporated
CAN: Controller Area Network
MOST: Media Oriented Systems Transport
MVN: Multifunction Vehicle Network
LIN: Local Interconnect Network
TTP: Time Triggered Protocol
TCN: Train Communication Network
WTB: Wired Train Bus

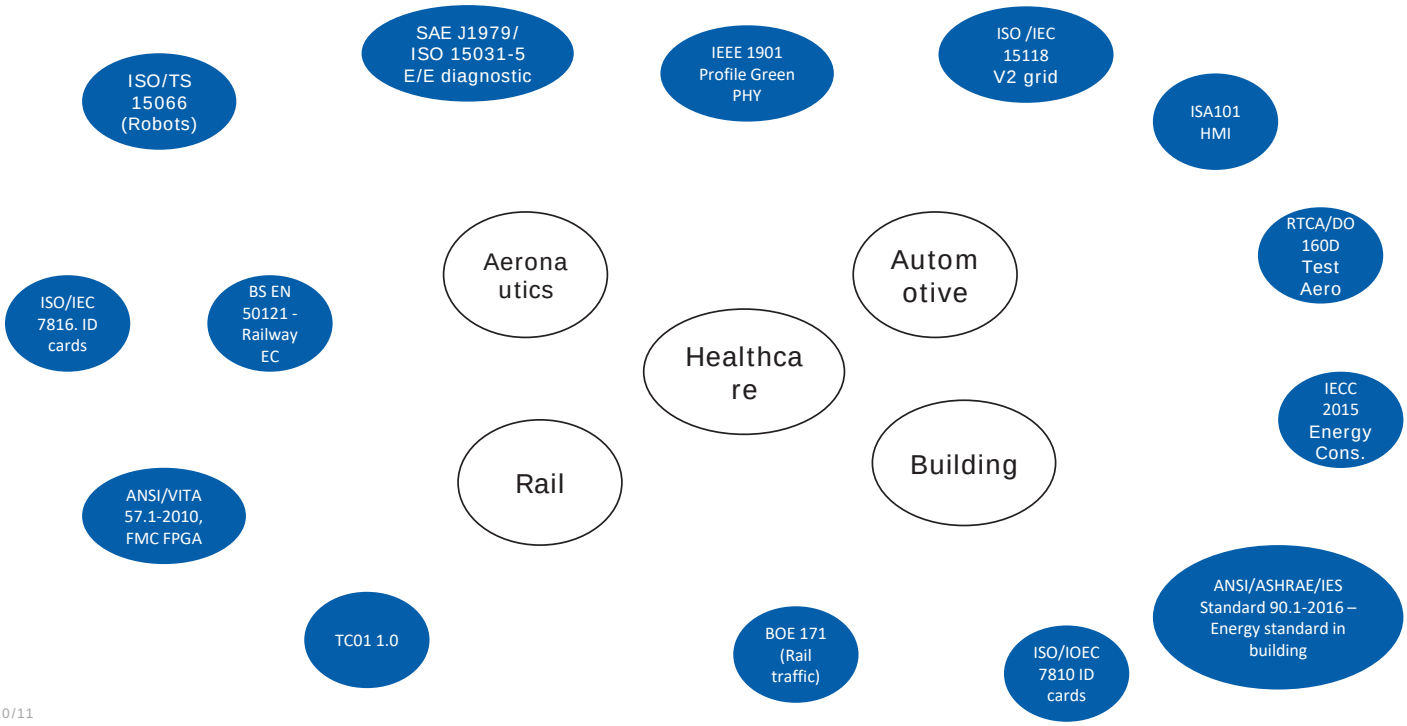
Wireline standards



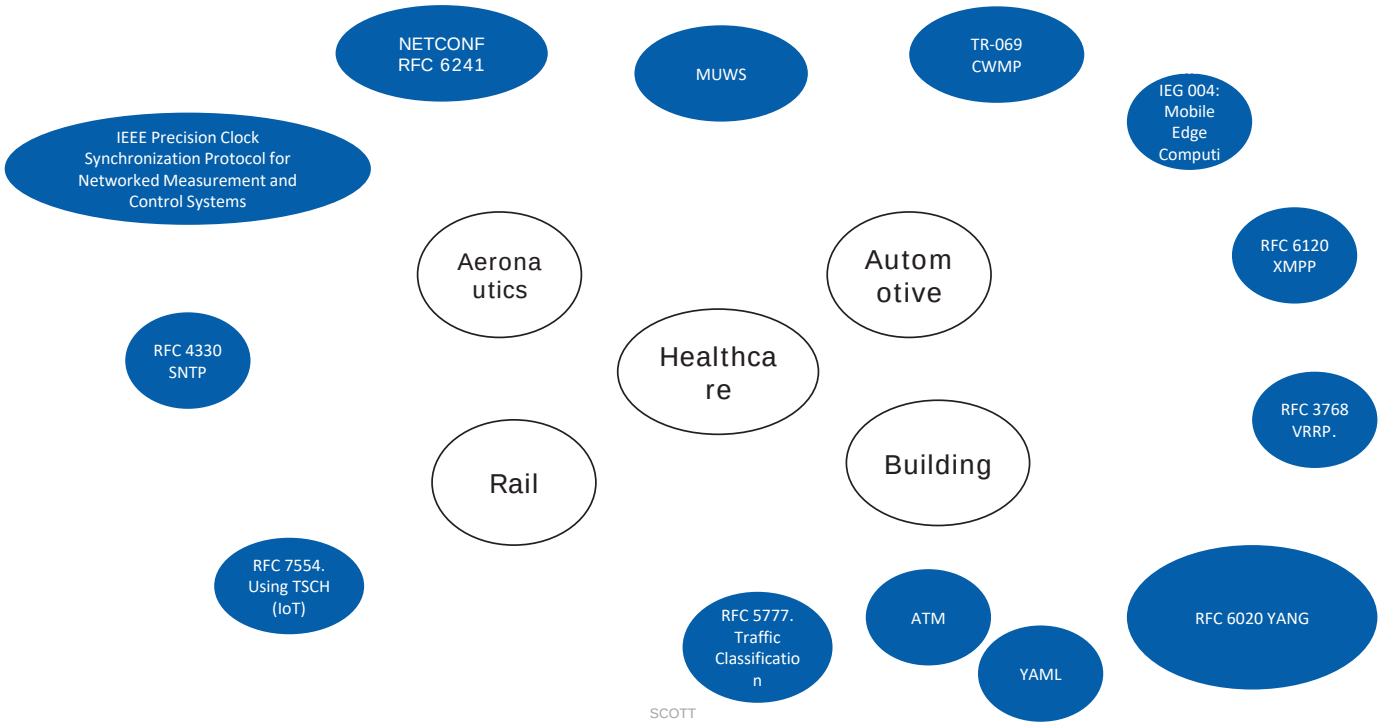
Higher layer standards



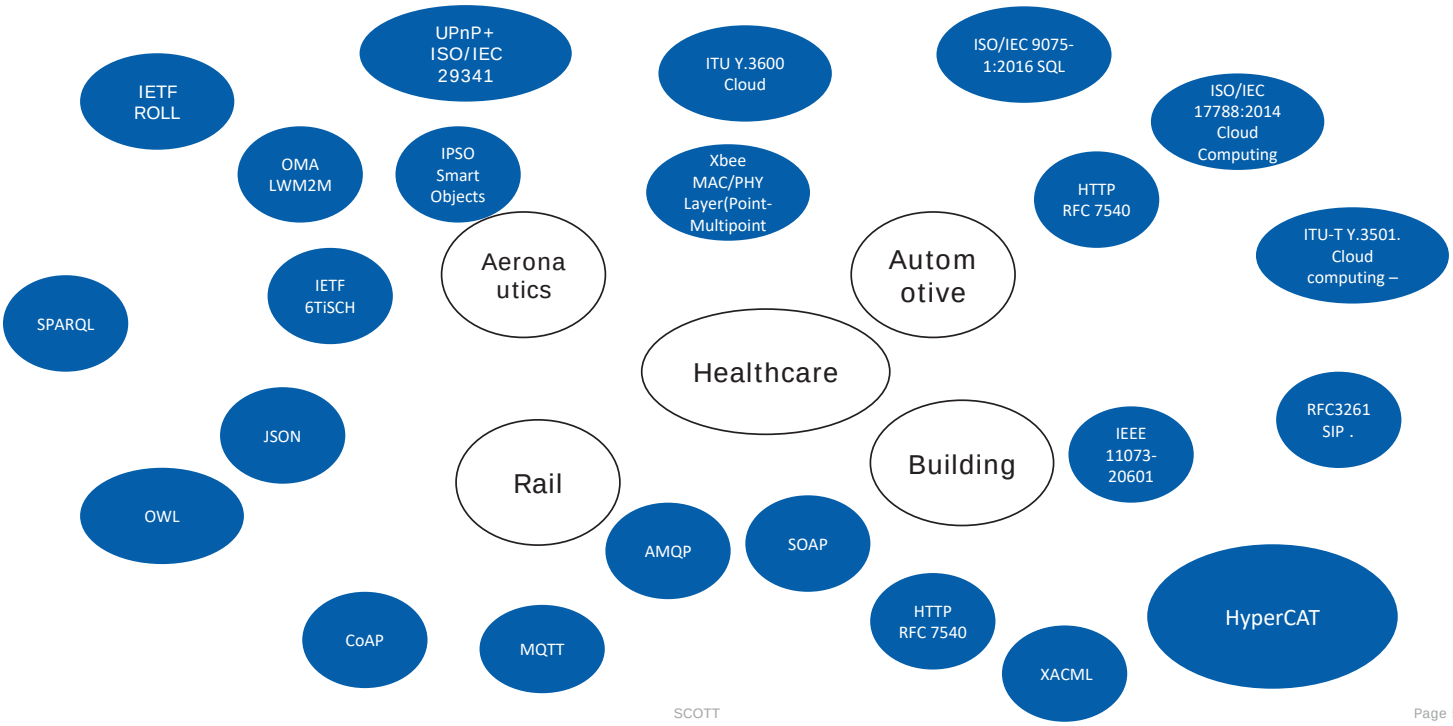
Device layer standards



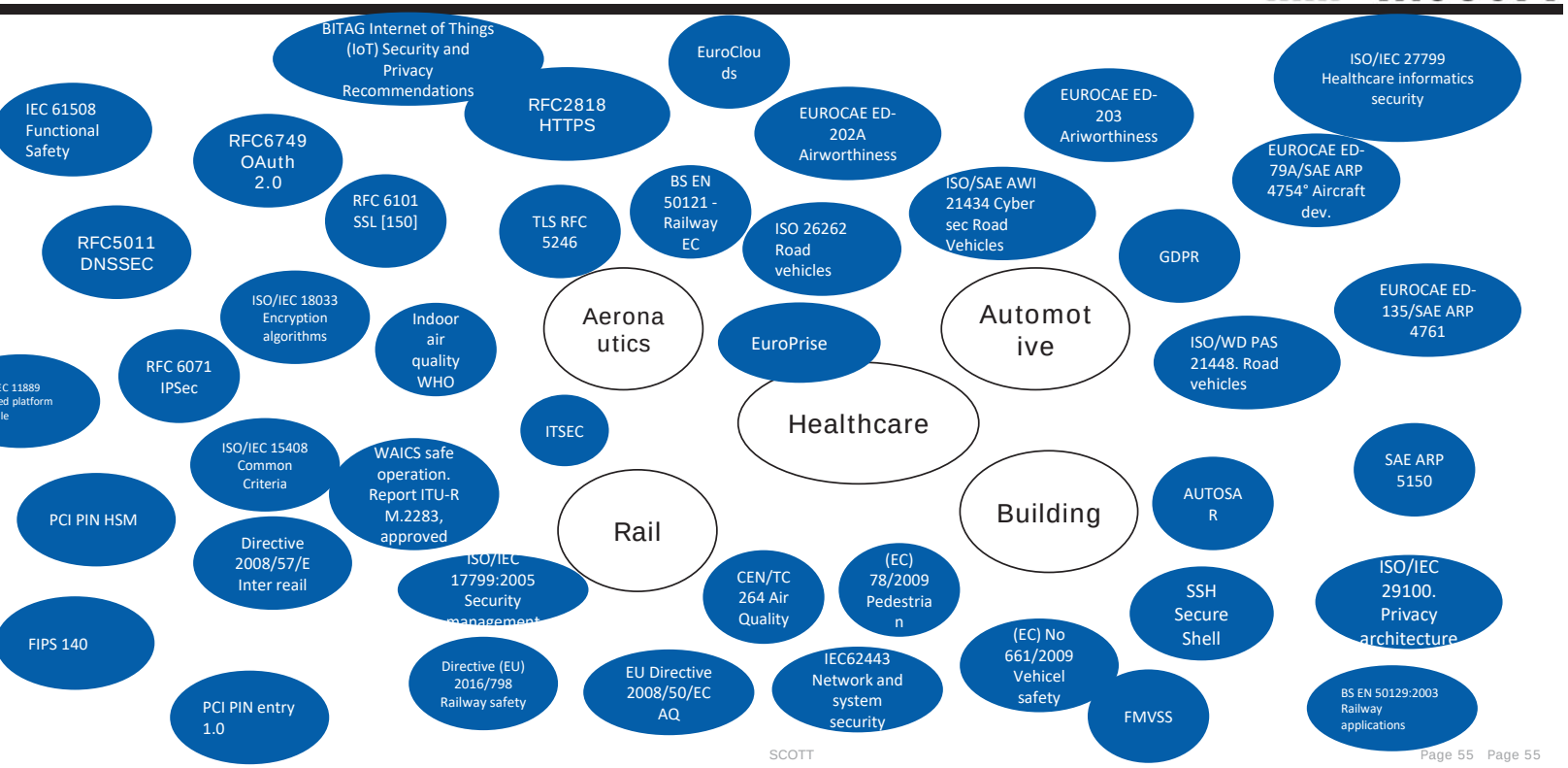
Network and service layer standards



IoT and cloud standards

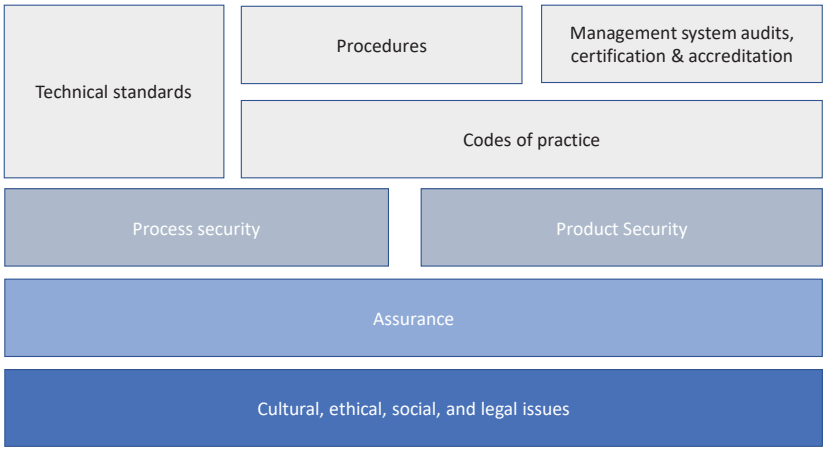


Security and safety standards

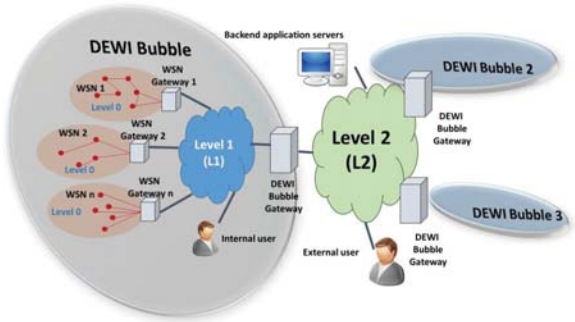


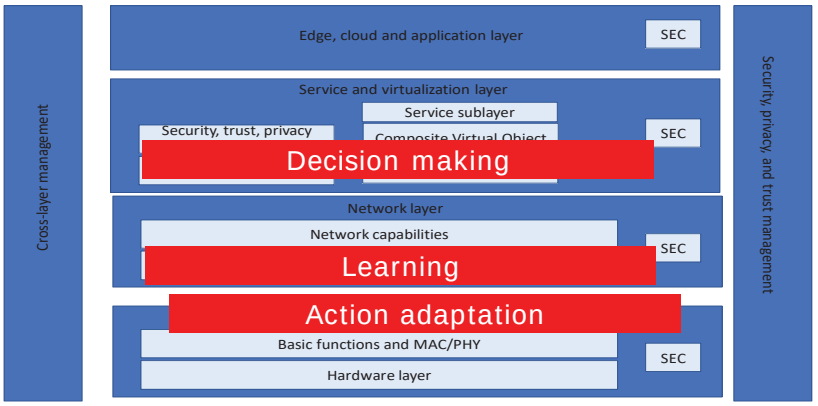
Security standards

- Security standards can be grouped as shown in the figure
- Each standard can be abstracted into different types of metrics to create security classes



Artificial intelligence in the reference architecture

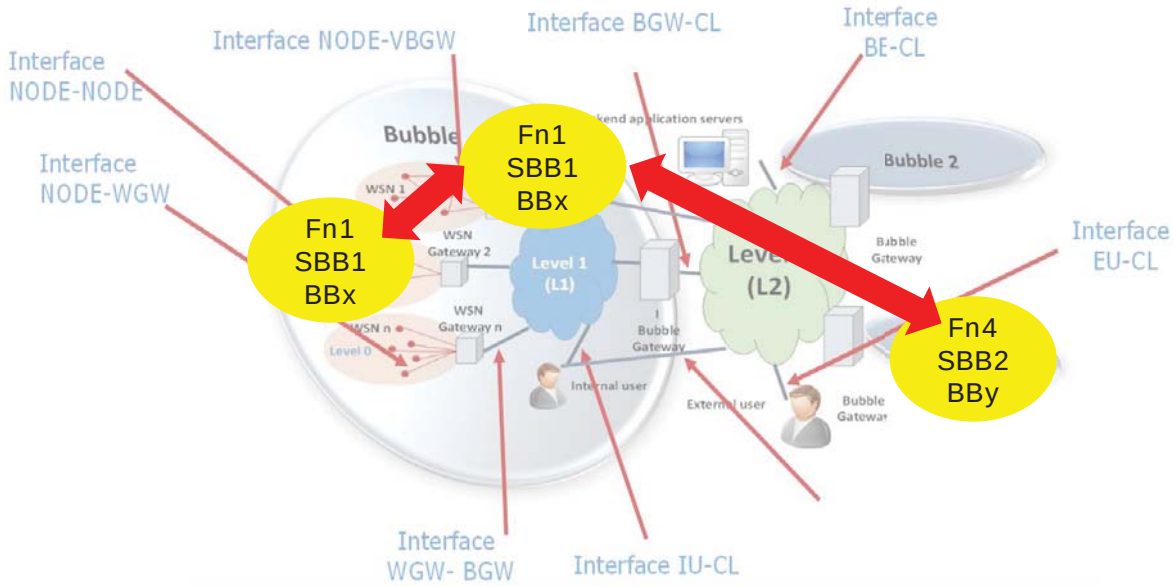




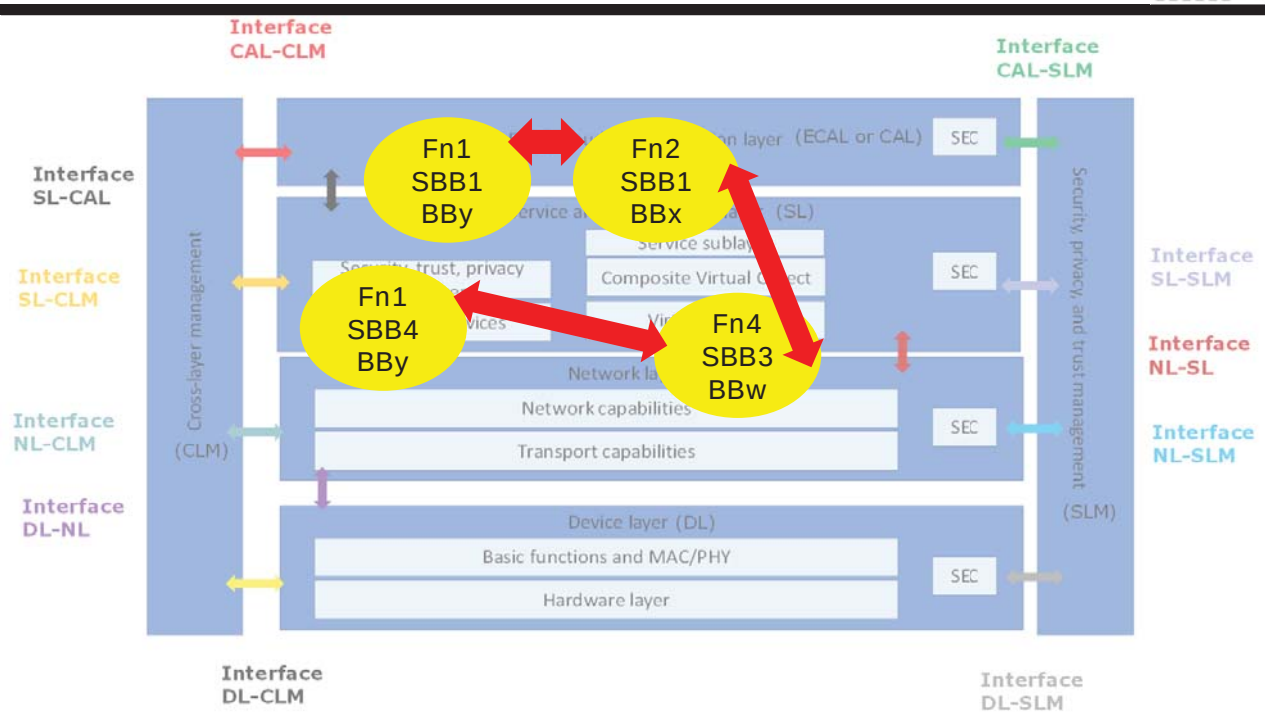
	Data source (e.g training data)	Preprocessing	Processing	Action	Storage
AI-Channel estimation	Device layer (PHY)	Device layer (PHY)	EDGE-CAL	Device layer (PHY)	EDGE/CAL
AI-conflict resolution	Device layer (PHY/MAC)	Device layer (PHY/MAC)	Device layer EDGE GW	Device layer (PHY/MAC)	EDGE/CAL

Sub-building block for AI development

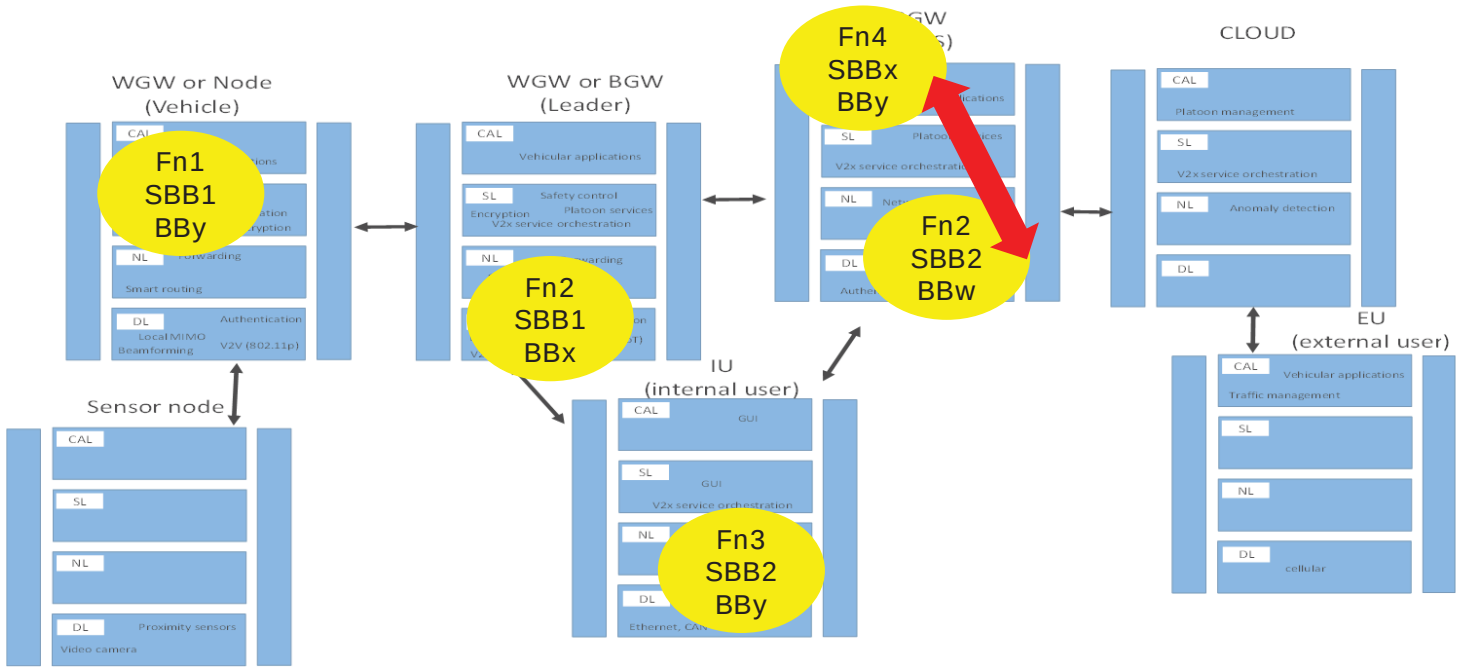
Entity model view (Model for mapping of sub-building blocks)



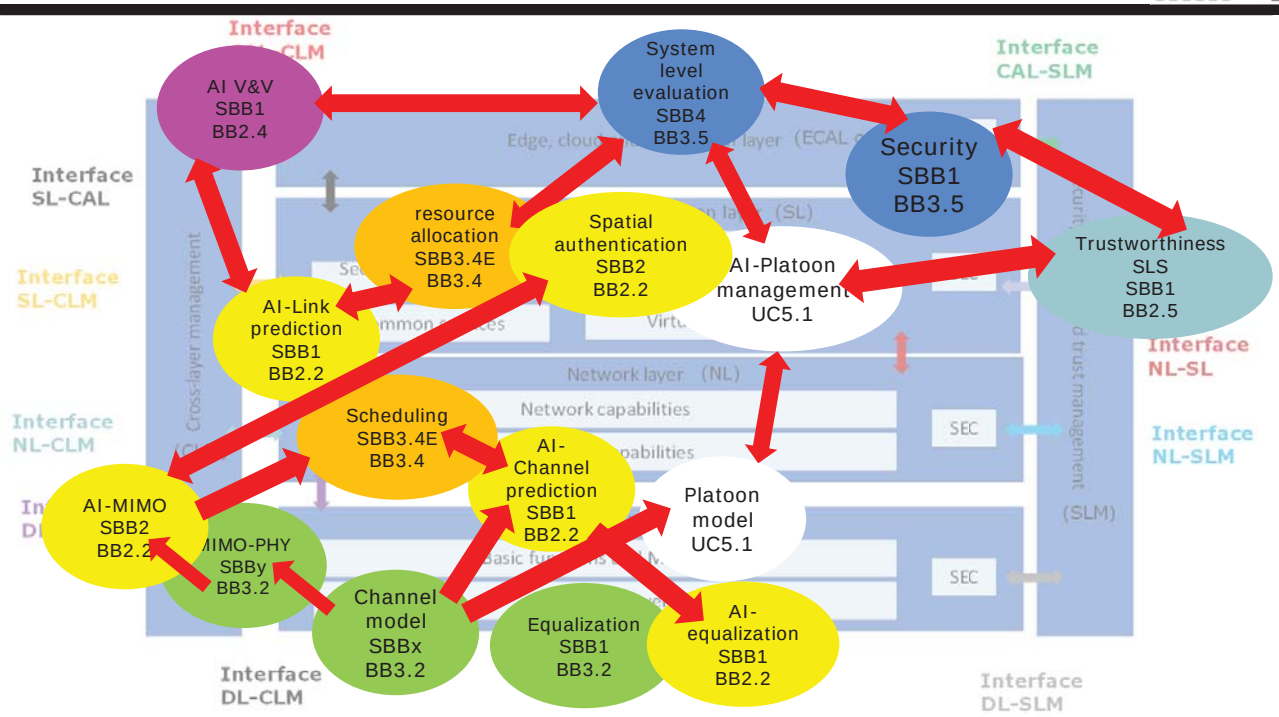
BB and subBB model for mapping on the functional view



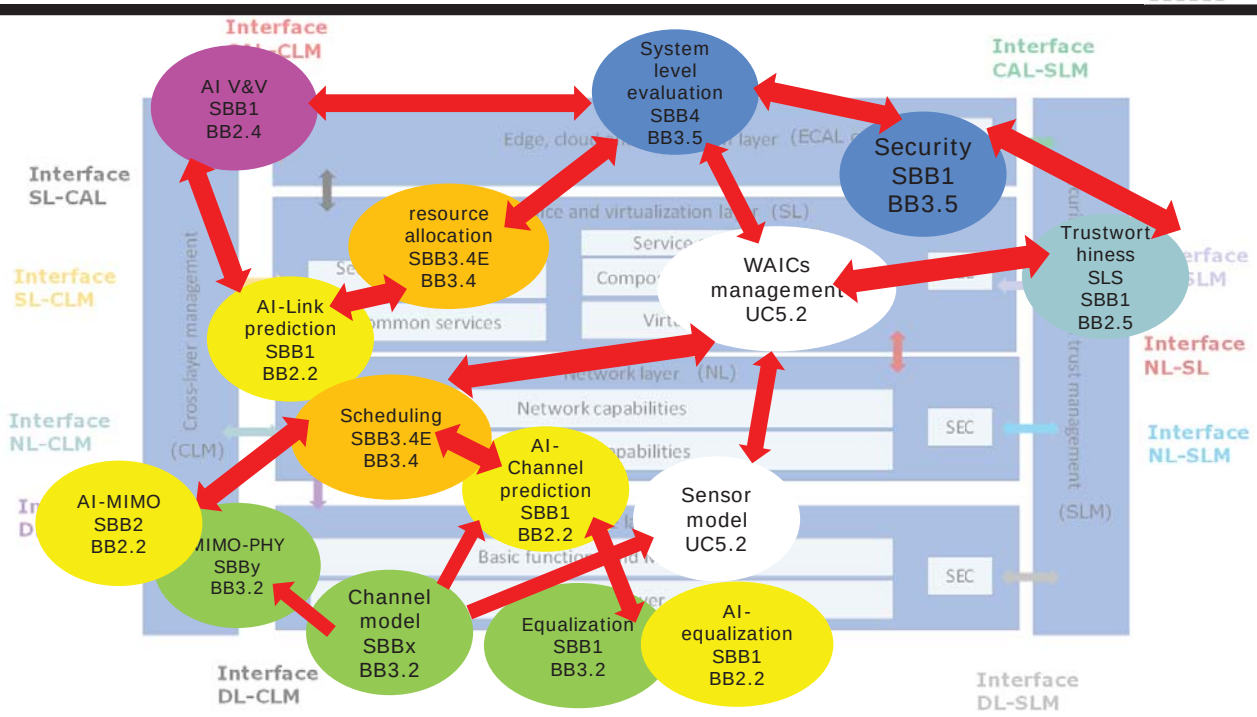
Entity vs functional view (Model for mapping of sub-building blocks)



BB mapping UC 5.1 Wireless platooning

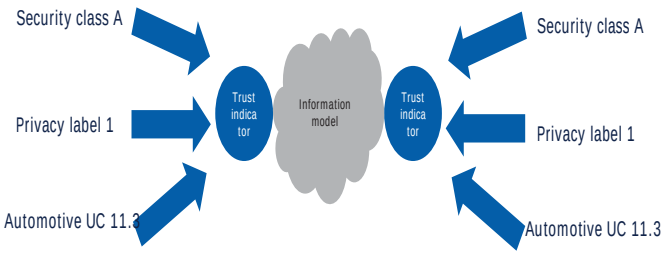


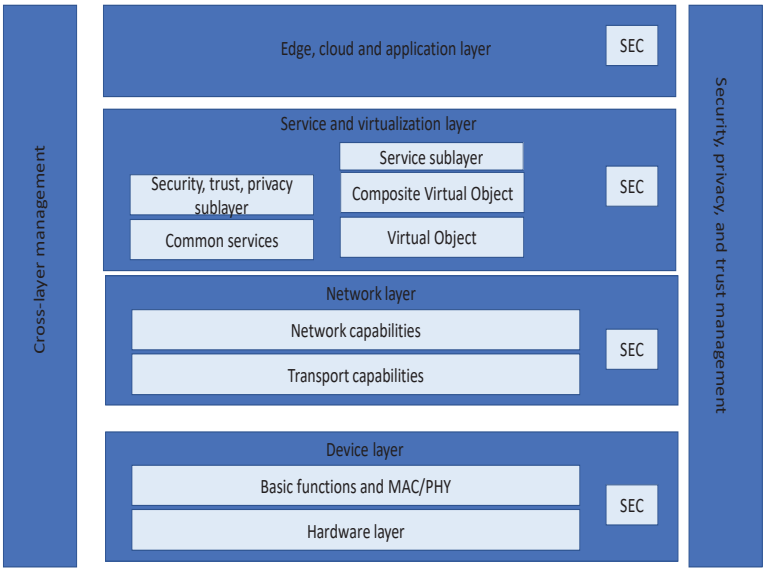
BB mapping UC 5.2 Wireless avionics



Security classes and trust indicators

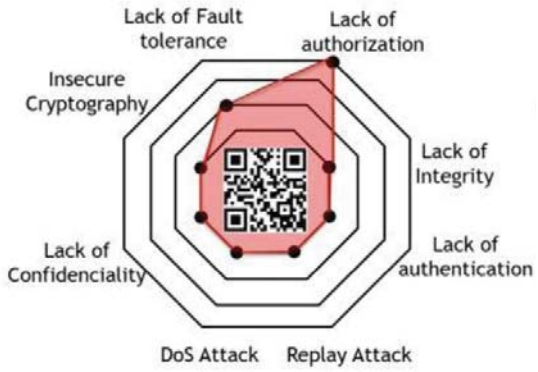
- Virtualization layer included in final version of the RA
- Level 2 Trust protocol adaptation as an extension of security classes, multi-metrics and privacy labels
  - Trust indicator for communication between bubbles





Security classes and trust indicators

- ARMOUR multi-metrics framework
- SCOTT security classes
- Trust metrics as extension of security analysis across layers

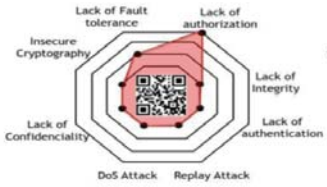


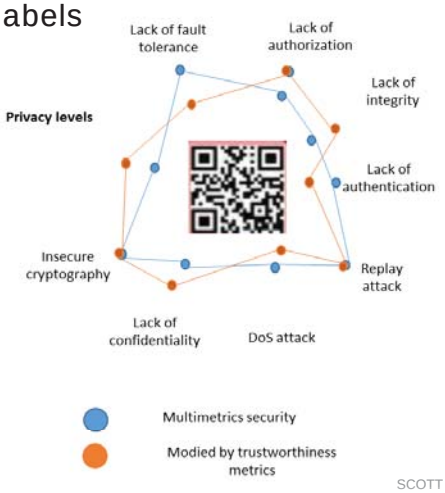
No.	Functionality	RQs and BB	Security Tuple	Trust model metric	Layer	Entity	Functionality mapping
001	PHY-layer encoding	RQ255,	[HHHHHHL]	CWSM	DL	L0	405,290
002	PHY-layer encryption	RQ34, 67,249,	[MHHHHHL]	Neural Network (NN)	DL	L0	32
003	Authentication	RQ33, 564	[HHMHHHL]	CWSM	SL	L0/L1/L2	356
004	Interference rejection (MIMO)	RQ334,654.	[HHMMHHL]	NN	DL	L0	45

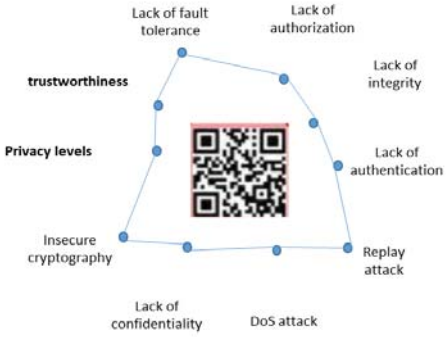
- Multi criteria decision making for extension to trust
- SCOTT and ISO trustworthiness framework

Privacy, multi-metrics, and trustworthiness labels

- Extension of existing labels
  - Risk assessment methodology
  - Models and IoT risk database
  - ETSI, ARMOUR, STRIDE methodologies
  - GDPR for privacy labels

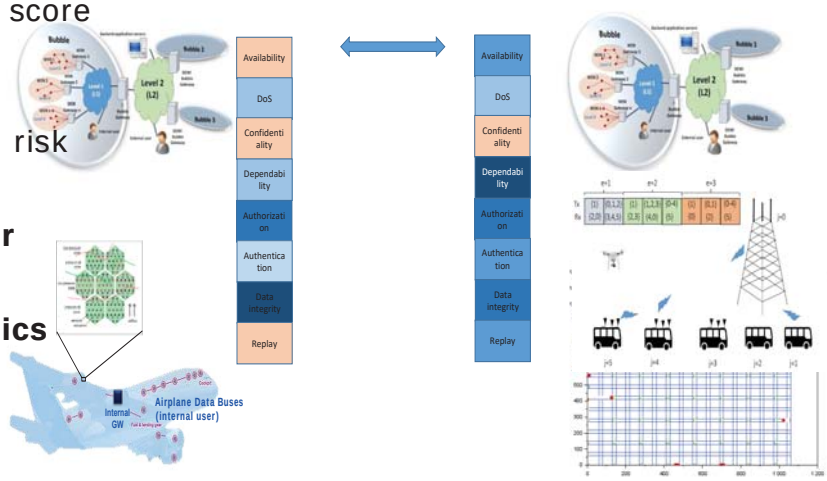






Certification-related aspects related to the RA

- Trustworthiness **labels and online certificates**
  - Extended **ARMOUR**, ISO and ETSI metrics
- Common weakness and common vulnerability score systems (**CWSS and CVSS**) applied to the reference architecture
- **Per layer and per entity** security threat and risk analysis
- **Trustworthiness vector** and **trust indicator** demonstrated in use cases
- Bubble to Bubble communications **trust metrics** adaptation
- **AI-based security metrics calculation**
- **User-friendly** security classes calculation
- **Modified CWSS and CVSS s**tudied in different use cases



Trustworthiness certification

Questions?

Ramiro Samano Robles
CISTER/ISEP, Portugal
rasro@isep.ipp.pt

AGENDA

- CYBERENG Project
- OVERVIEW CYBERSECURITY MANAGEMENT
- OVERALL: CYBERSECURITY MANAGEMENT
- REGULATION
- STANDARDS
- APPLICATION

What is CYBERENG?

2-year (2020-2022) Erasmus+ funded project

CYBERENG stand for *"ECQA Certified Cybersecurity Engineer and Manager"*

Project aims to define **needs** and **training** for cybersecurity engineer and manager.

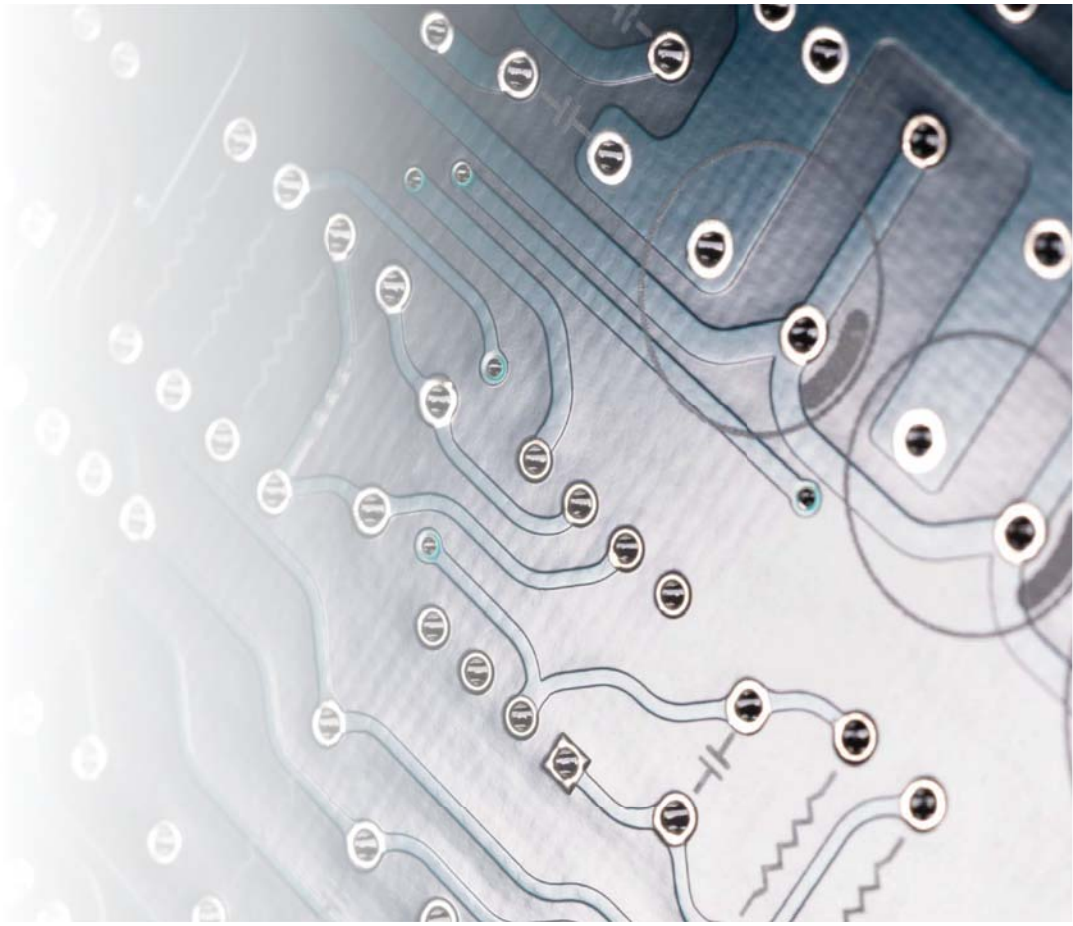
Main outcome is to develop **training courses and material** for the industry and universities.

CYBERENG

With the support of the Erasmus+ Programme of the European Union

Why Cyber Security?

- Project will tackle the demanded skills and competences for cyber security in **automotive domain** since the sector is becoming more and more dependent on the software.
- Cyber security and digitalization are among the main driving factors of the **automotive industry** as well as the transition to the electric vehicles which will require even more software-oriented approach.

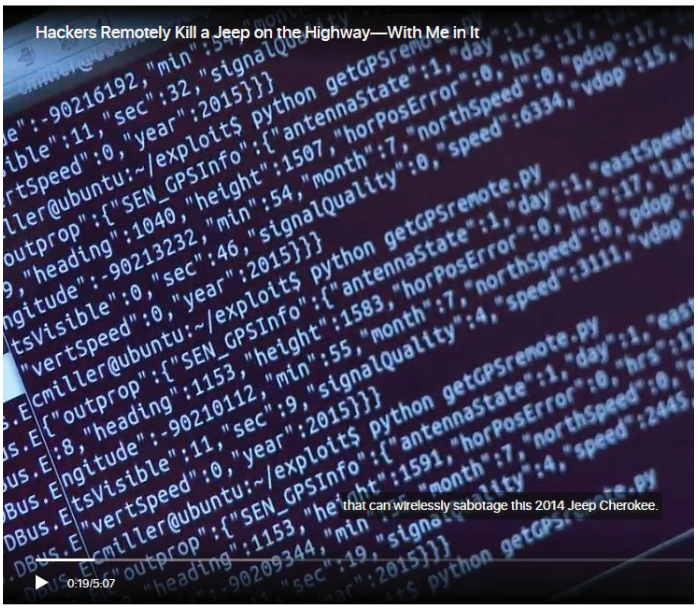


Starting Scenario

- Modern cars are controlled >90% by computers and software. Cars are connected by Wlan, Car2X, connectors and buses (e.g. OBD), etc. to the infrastructure. These interfaces can be exploited for attacks.

https://de.qaz.wiki/wiki/Jeep_Cherokee_(KL)

Software-Hack Jeep Cherokee (KL) leads to 1,5 million vehicles recall.



ackers Remotely Kill a Jeep on the H
/ith Me in It

Vulnerabilities in Modern Vehicles

- Cars have a fixed IP address for the connectivity gateway.
  - Car manufacturers (Tesla, Toyota, Honda, BMW etc) become a Local Internet Registry (LIR). This means that car makers get blocks of Internet IP address space.
  - Gateways are like small Internet servers (Linux) which can be attacked if there is no proper firewall, encryption mechanism.
  - Car computers (e.g. steering) react on commands on the bus. And commands can be faked (in cyber language spoofed).

Examples from 2020 in Automotive

January

- *4,118 vehicles were stolen in India with electronic devices that enabled the thieves to bypass the engine control module,* ***unlock the vehicle, start the engine****, and access the vehicles' computer*
- *A Mobileye 630 PRO and Tesla Model X hack fooled the ADAS and autopilot systems to* ***trigger the brakes and steer into oncoming traffic***

February

- *19 vulnerabilities were found in a Mercedes-Benz E-Class car, allowing hackers to* ***control the vehicle remotely****, including opening its doors and starting the engine*

April

- *Hackers* ***took full control of an OEM's corporate network*** *by reverse engineering a vehicle's TCU and using the telematics connection to infiltrate the network*

...

August

- *A hacker was able to gain* ***control over Tesla's entire connected vehicle fleet*** *by exploiting a vulnerability in the OEM's server-side mechanism*
- More than **300 vulnerabilities were found in over 40 ECUs developed by 10 Tier-1** companies and OEMs

Quelle: Upstream Security Global Automotive Cybersecurity Report 2021

Regulation for Car Makers

- Car homologation will require from 2021 onwards a declaration of Compliance for Cybersecurity.
- The regulation requires several methods to be implemented in automotive engineering projects dealt with in the three job roles cybersecurity manager, cybersecurity engineer, cybersecurity tester.


United Nations ECE/TRANS/WP.29/2020/79 REVISED
Economic and Social Council
Distr.: General
23 June 2020
Original: English

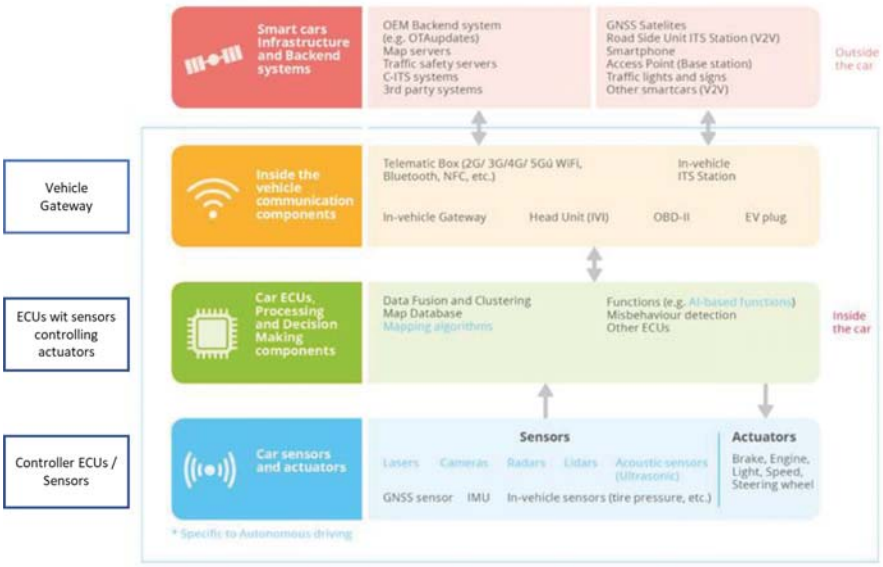
Economic Commission for Europe
Inland Transport Committee
World Forum for Harmonization of Vehicle Regulations

Proposal for a new UN Regulation on uniform provisions concerning the approval of vehicles with regards to cyber security and cyber security management system

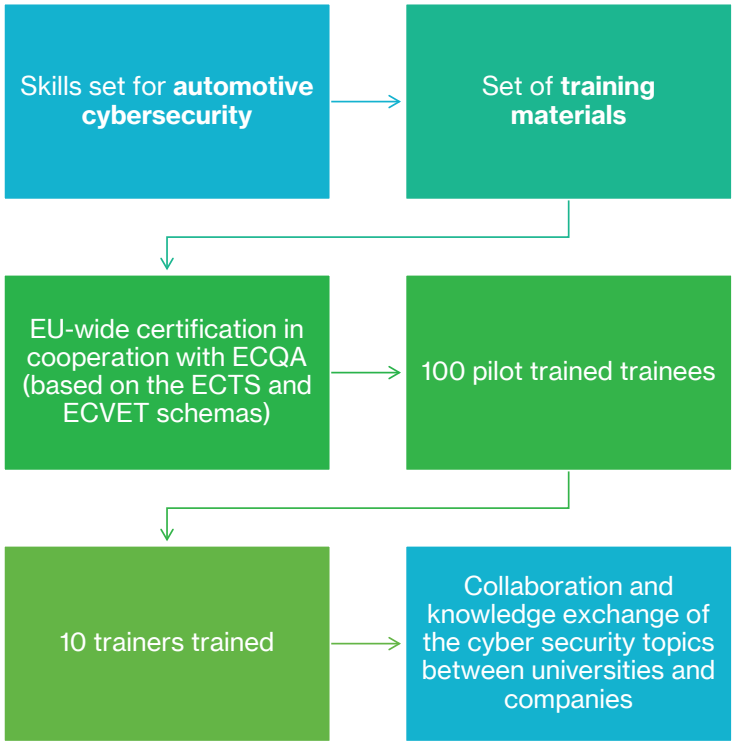
enisa
EUROPEAN UNION AGENCY FOR CYBERSECURITY

ENISA: High-Level Smart Cars Reference Model

- Asumes Ethernet and domain servers in a car.
- Reality:
  - Most cars still work with CAN FD, Flexray, and one gateway and no IP addresses per ECU.
  - Most cars still have intelligence in the actuator
- **and still cybersecurity to be achieved**



Cybereng Goals



Partner Organizations

VSB TECHNICAL UNIVERSITY OF OSTRAVA

ISCN
International Software Consulting Network

REALSECURITY

TU Graz

Elektrobit

AIT AUSTRIAN INSTITUTE OF TECHNOLOGY

CYBERENG

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

With the support of the Erasmus+ Programme of the European Union

Project Outcomes – Intellectual Outputs

IO1: Study about the requirements for an ECQA Certified Cybersecurity Engineer and Manager – Online, Publication

IO2: Skills set for an ECQA Certified Cybersecurity Engineer and Manager – Automotive Sector based on ECQA skills definition standards – Online

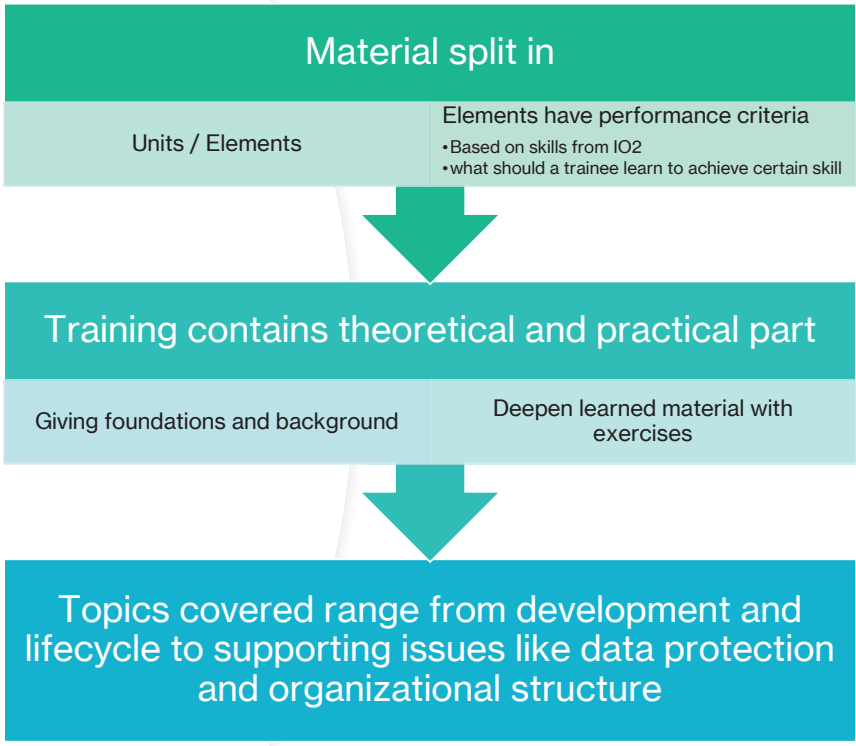
IO3: Training Material for the ECQA Certified Cybersecurity Engineer and Manager – Automotive Sector skills set

IO4: Online Training Campus

IO5: Certification Framework and Exams (based on ECQA guides)

CYBERENG

IO3: Training Material for the ECQA Certified Cybersecurity Engineer and Manager – Automotive Sector skills set



The following is a (very) shortened intro into:
- cybersecurity management
- cybersecurity engineering
full set of training will be a multi-day training for automovie cyberseucirty engineers or manager

OVERVIEW CYBERSECURITY MANAGEMENT

The organization institutes, governance, and a cybersecurity culture shall boost cybersecurity engineering, including:	The organization shall demonstrate and maintain organization-specific rules and procedures to:	The organization is responsible for implementing management systems for cybersecurity, particularly a quality management system, and managing the technologies used in cyber-engineering.
• cybersecurity awareness management, • competence management, and • continuous improvement.	• enable the implementation of the requirements according to ISO 21434; • support the implementation of the corresponding activities.	

CYBERENG

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

With the support of the Erasmus+ Programme of the European Union

OVERVIEW CYBERSECURITY MANAGEMENT

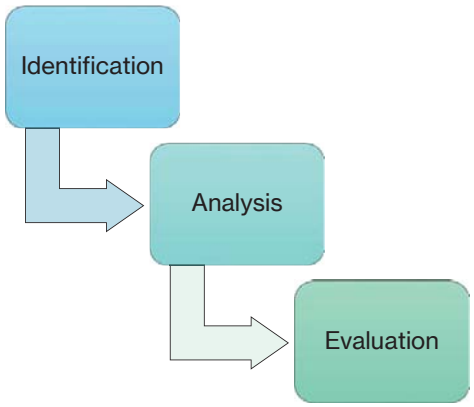
- The organization shall create and maintain a quality management system following **international standards**, or **equivalent**, to support cybersecurity engineering, including:
  - change management
    - Manage changes in items and their components so that the relevant cybersecurity objectives and requirements continue to be satisfied.
  - Documentation management
    - A work product can be merged or mapped to additional documentation repositories.
  - Configuration management
  - Requirements management
    - IATF 16949 in conjunction with:
      - ISO 9001; ISO 10007, Automotive SPICE, the ISO/IEC 33000 series of standards, ISO/IEC/IEEE 15288, and ISO/IEC 12207

The organization is responsible for implementing management systems for cybersecurity, particularly a quality management system, and managing the technologies used in cyber-engineering.

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.
With the support of the Erasmus+ Programme of the European Union

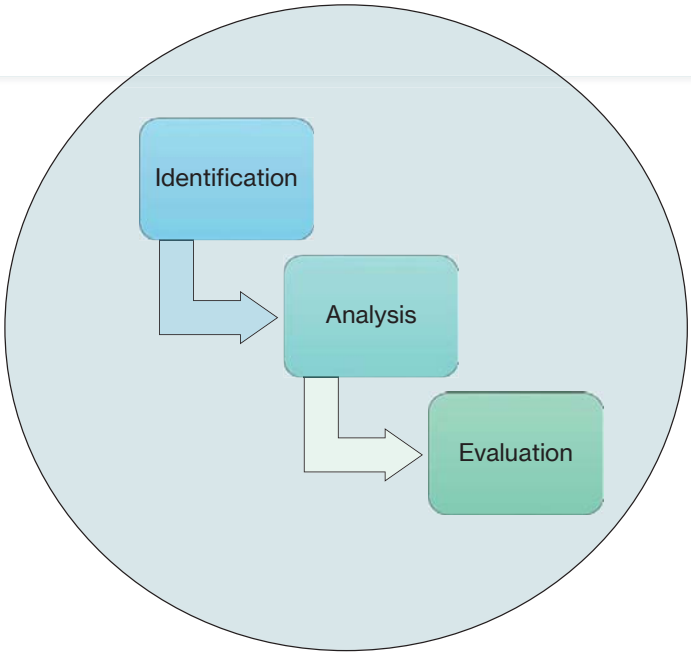
OVERVIEW CYBERSECURITY MANAGEMENT

The overall cybersecurity risk management of an organization is implemented in accordance with both this clause and ISO 31000 and applies throughout all phases.



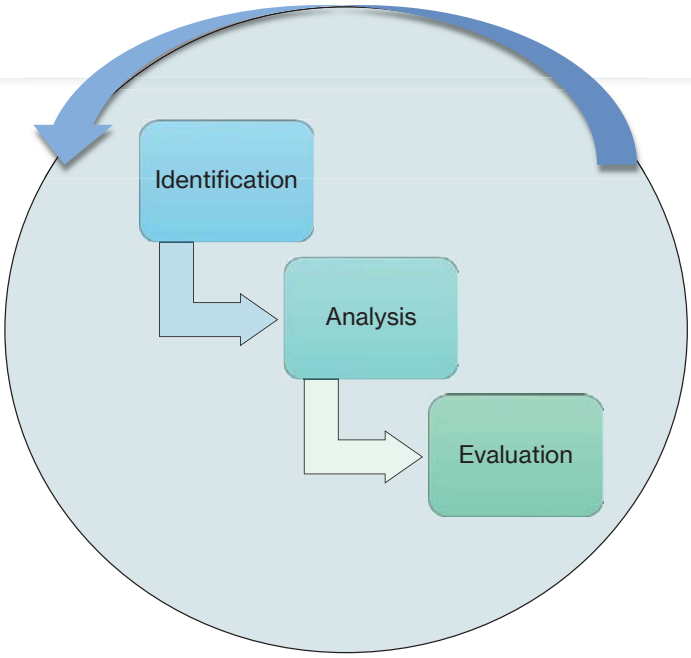
OVERVIEW CYBERSECURITY MANAGEMENT

The overall cybersecurity risk management of an organization is implemented in accordance with both this clause and ISO 31000 and applies throughout all phases.



OVERVIEW CYBERSECURITY MANAGEMENT

The overall cybersecurity risk management of an organization is implemented in accordance with both this clause and ISO 31000 and applies throughout all phases.



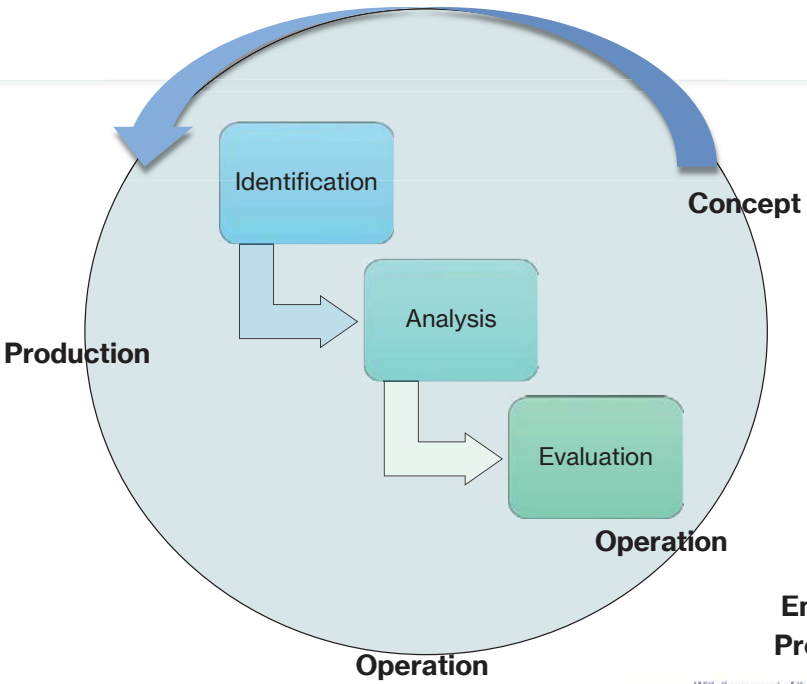
CYBERENG

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

With the support of the Erasmus+ Programme of the European Union

OVERVIEW CYBERSECURITY MANAGEMENT

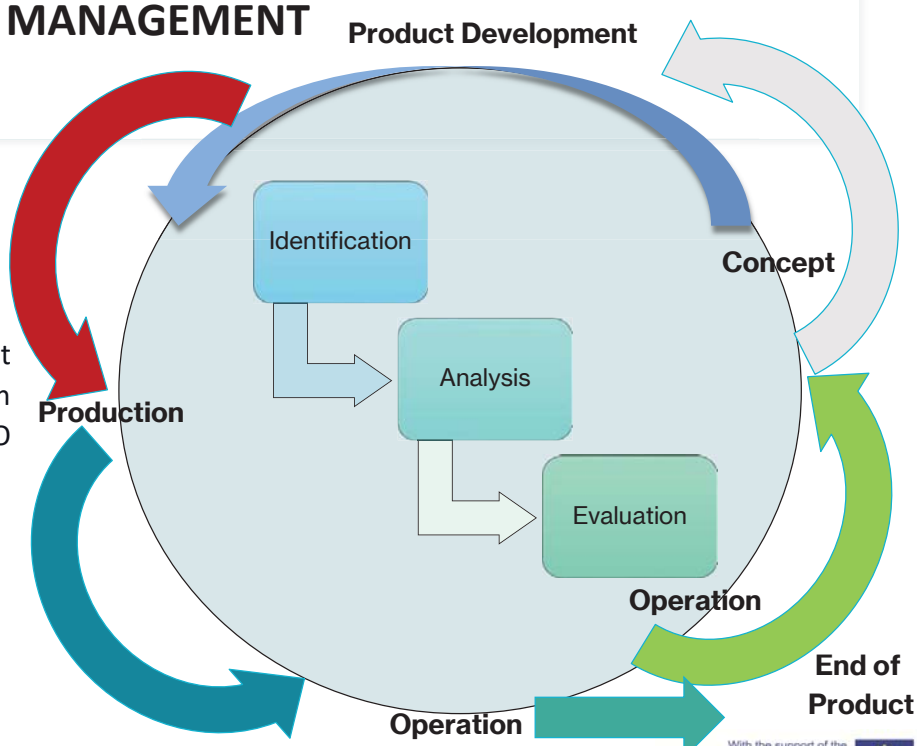
The overall cybersecurity risk management of an organization is implemented in accordance with both this clause and ISO 31000 and applies throughout all phases.



CYBERENG
The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.
With the support of the Erasmus+ Programme of the European Union

OVERVIEW CYBERSECURITY MANAGEMENT

The overall cybersecurity risk management of an organization is implemented in accordance with both this clause and ISO 31000 and applies throughout all phases.

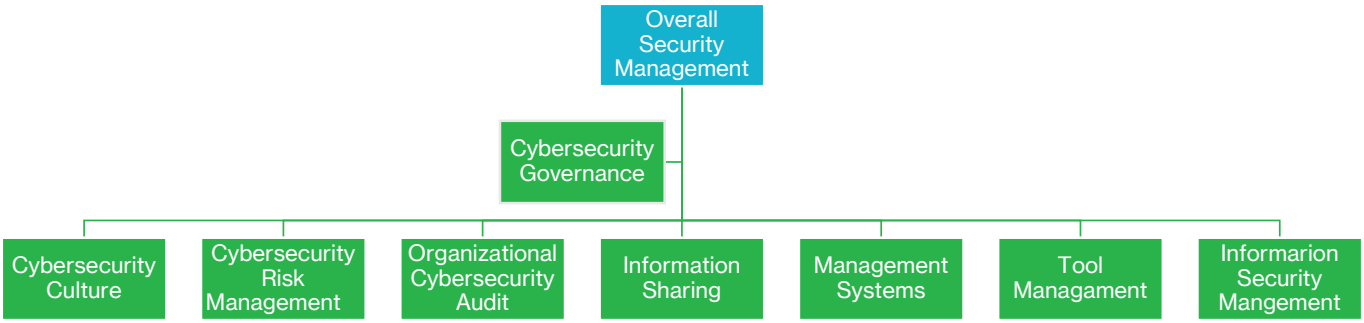


CYBERENG

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

With the support of the Erasmus+ Programme of the European Union

OVERALL: CYBERSECURITY MANAGEMENT



CYBERSECURITY GOVERNANCE

The organization shall specify a cybersecurity policy that contains

acknowledgement of road vehicle cybersecurity risks; and

the executive management's commitment to manage the relevant risks.

The organization shall establish and maintain organization-specific rules and processes to:

enable the implementation of the requirements according to ISO 21434; and

Support the implementation of the related activities.

The organization shall assign and communicate the responsibilities to achieve and maintain cybersecurity; and assign the corresponding authority.

The organization shall provide the resources to address cybersecurity

The organization shall identify disciplines related to, or interacting with, cybersecurity and establish and maintain communication channels between those disciplines in order to:

determine if and how cybersecurity will be integrated into existing processes; and

coordinate the exchange of relevant information.

The organization shall define the risk values (1 to 5) in a risk matrix

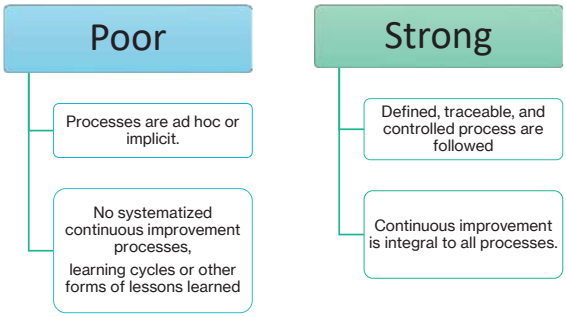
CYBERENG

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

With the support of the Erasmus+ Programme of the European Union

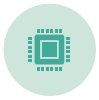
CYBERSECURITY CULTURE

- The organization shall foster and maintain a cybersecurity culture
  - According to ISO 21434, the following present some examples according to the poor and strong cybersecurity culture



- The organization shall ensure the persons within the organization that are involved in cybersecurity have the awareness to fulfill their responsibilities.

- The organization shall institute and maintain a continuous improvement process, such as:



Learning from previous cybersecurity experiences, including experiences gathered by field monitoring and observation of internal and external information;

Learning from information obtained regarding products of a similar application in the field;

Deriving improvements to be applied during subsequent cybersecurity activities;

Communicating lessons learned to the appropriate persons; and

Checking the adequacy of its rules and processes

CYBERENG
The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.
With the support of the Erasmus+ Programme of the European Union

CYBERSECURITY CULTURE

Cybersecurity risk management shall be following ISO 31000.

The organization may align its cybersecurity risk management and its corporate risk management.

EXAMPLE: During the development phase, an assumption is made that a specific cryptographic algorithm is secure, but in the field, it is discovered via cybersecurity monitoring that the algorithm is no longer secure. Vulnerability management and incident response are used to manage this issue.

CYBERENG

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

With the support of the Erasmus+ Programme of the European Union

ORGANIZATIONAL CYBERSECURITY AUDIT

A cybersecurity audit shall be performed to independently judge whether the organizational processes achieve the objectives according to ISO 21434.

According to a quality management system standard, such a cybersecurity audit can be included in or combined with an audit.

Independence can be based on, for example, IATF 16949 in conjunction with ISO 9001 or ISO 26262.

The person that performs the audit can be internal or external to the organization.

More details in Cybersecurity auditing will be addressed in U2.E2.

CYBERENG

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

With the support of the Erasmus+ Programme of the European Union

INFORMATION SHARING

The organization shall define the circumstances under which sharing is required, permitted, and prohibited within and outside the organization.

- A list of the types of cybersecurity information that can be shared
- An approval process for sharing;
- Requirements for redacting and sanitizing information;
- Rules for source attribution; and
- Consultation and types of communications permitted.

MANAGEMENT SYSTEMS

Change Management;

- The scope of change management in cybersecurity is to manage changes in items and their components so that the applicable cybersecurity goals and requirements continue to be fulfilled.

Documentation Management

- A work product can be combined or mapped to different documentation repositories.

Configuration Management; and

Requirements Management

TOOL MANAGEMENT

Tools that can impact the cybersecurity of an item, system, or component shall be managed.

EXAMPLE 1: Tools used for concept or product development, such as model-based development, static checkers, verification tools.	EXAMPLE 2: Tools used during production, such as a flash writer end of line tester.	EXAMPLE 3: Tools used for maintenance, such as an on-board diagnostic tool or reprogramming tool.

An appropriate environment to support remediation actions for the cybersecurity incident response

EXAMPLE 4: A testing environment for reproducing the vulnerability.	EXAMPLE 5: Forensic methods

CYBERENG

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

With the support of the Erasmus+ Programme of the European Union

INFORMATION SECURITY MANAGEMENT

The relevant information security properties of the work products required by the cybersecurity plan should be managed by an information security management system.

EXAMPLE: Work products can be stored on a file server that protects them from unauthorized alteration or deletion.

CYBERSECURITY REGULATION

Image credit: tag-cyber (https://www.tag-cyber.com/media/charlie-ciso)

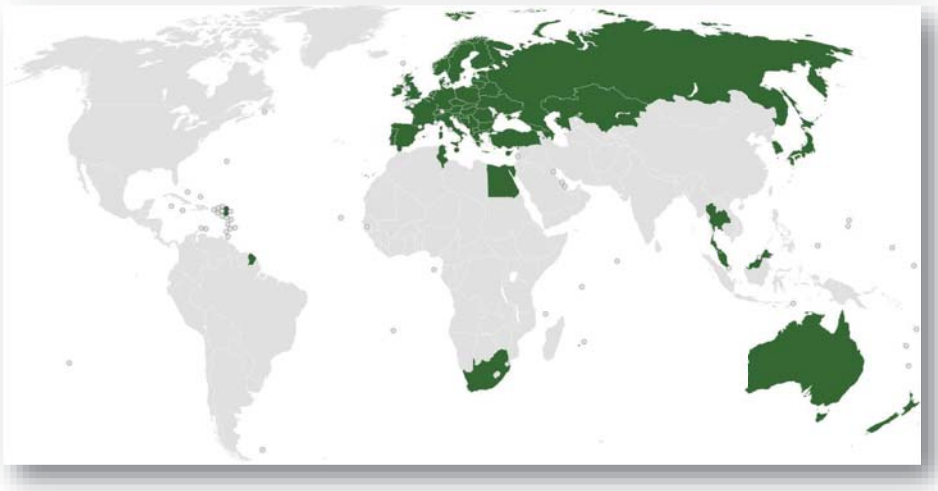
CYBERENG

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

With the support of the Erasmus+ Programme of the European Union

UNECE WORLD FORUM FOR HARMONIZATION OF VEHICLE REGULATIONS

- UNECE WP29 defines **requirements** for **type approval**
- Members are:
  - Type approval authorities
  - Certification bodies
  - OEM and Tier 1
- Delivered two draft regulations on:
  - **Cyber security**
  - Software updates



UNECE WP 29 DRAFT REGULATION ON CYBER SECURITY

- **Vehicle manufacturer**, **suppliers** and **service providers** need a Cyber Security Management System (CSMS)

- CSMS covers **distributed development, production,** and **post-production**
  - **Management** of cyber security in the **organization**
  - **Management** of risks to the **vehicle**
  - **Verification** of risk management
  - **Management** of **new** cyber **threats** and **vulnerabilities**

Cyber Security Management System → Post-Production Phase → Vehicle Type Approval

views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.
e support of the Erasmus+ Programme of the European Union

UNECE WP 29 DRAFT REGULATION ON CYBER SECURITY

- **Compliance** with the regulation is **maintained** through the **vehicle lifecycle**
  - **Monitoring** of changes in the **threat landscape** and vulnerabilities.
  - **Implemented** security measures need to be **monitored** for **effectiveness**.
  - **Changing** circumstances should **not impact safety** and **availability**.

Image credit: tag-cyber (https://www.tag-cyber.com/media/charlie-ciso)

Cyber Security Management System → Post-Production Phase → Vehicle Type Approval

views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.
...e support of the Erasmus+ Programme of the European Union

UNECE WP 29 DRAFT REGULATION ON CYBER SECURITY

- **Vehicle type approval requires certified CSMS** for vehicle manufacturer, suppliers and service providers
  - CMSC certificate is **valid for three years**

- **Verified evidence** for **cyber security** of the vehicle type from the **full supply chain**
  - How known **vulnerabilities** and **threats** are **considered** in the **risk assessment**
  - **Risk assessment** considers the **whole vehicle and interactions**
  - Elements are designed in a way and protected by security measures so that the **risk is reduced to an acceptable level**
  - **Tracing** from **identified risk to implemented mitigation to testing**
  - **Dedicated** and **protected environment** for storage or execution of **aftermarket software, services, applications**, or **data**

CYBERENG
views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.
e support of the Erasmus+ Programme of the European Union

TIMELINE - AUTOMOTIVE

7.3. Requirements for vehicle types

7.3.1. The manufacturer shall have a valid Certificate of Compliance for the Cyber Security Management System relevant to the vehicle type being approved.

However, for type approvals prior to 1 July 2024, if the vehicle manufacturer can demonstrate that the vehicle type could not be developed in compliance with the CSMS, then the vehicle manufacturer shall demonstrate that cyber security was adequately considered during the development phase of the vehicle type concerned.

Image credit: UNECE (https://www.unece.org/fileadmin/DAM/trans/doc/2020/wp29grva/GRVA-06-19r1e.pdf)

CYBERSECURITY STANDARDS

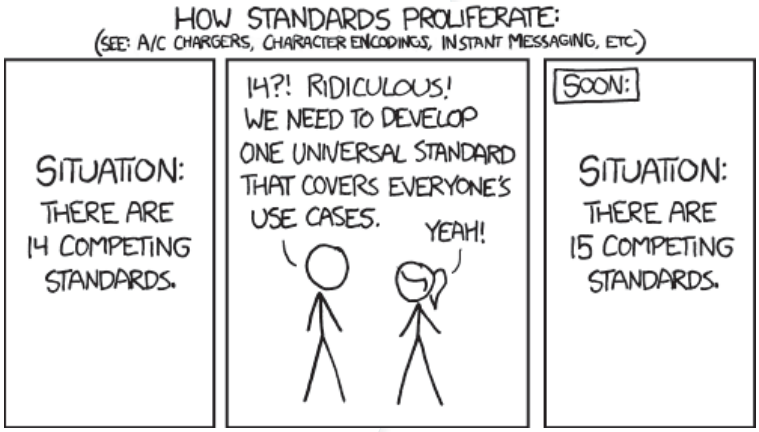


Image credit: XCKD (https://xkcd.com/927/)

CYBERENG

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

With the support of the Erasmus+ Programme of the European Union

AUTOMOTIVE

Cybersecurity Standards

ISO/SAE DIS 21434 ROAD VEHICLES — CYBERSECURITY ENGINEERING

- Requirements for cybersecurity
- Focus on risk management
- Considering engineering, production, operation, maintenance, and decommissioning
- For series production road vehicle electrical and electronic (E/E) systems, their components and interfaces
- Don't prescribe specific technology or solutions related to cybersecurity

CYBERENG

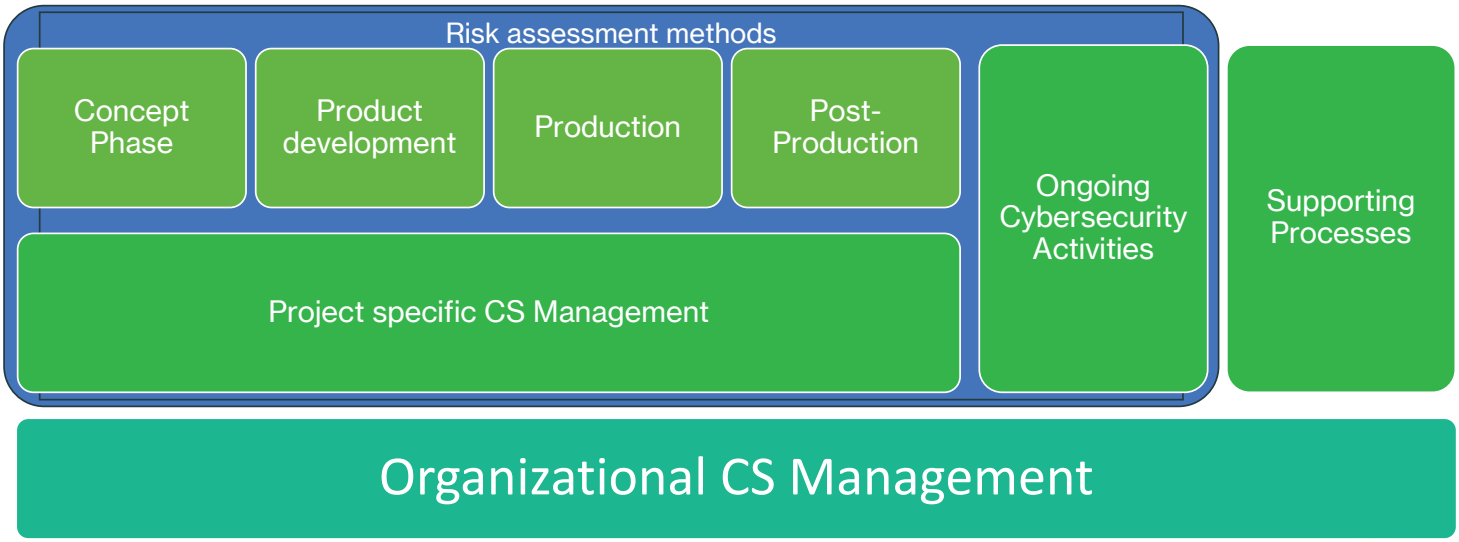
With the support of the Erasmus+ Programme of the European Union

ISO/SAE CD 21434 ROAD VEHICLES — CYBERSECURITY ENGINEERING

CYBERENG
The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.
With the support of the Erasmus+ Programme of the European Union

CONTENT

ISO/SAE CD 21434 Road Vehicles — Cybersecurity engineering



CONTENT

ISO/SAE CD 21434 Road Vehicles — Cybersecurity engineering

4. **General considerations**
   - a) Overview
   - b) How to read this standard
   - c) Safety & Security
5. Overall Cybersecurity Management
6. Project dependent cybersecurity management
7. Risk assessment methods
8. Concept phase
9. Product development
10. Production
11. Operations
12. Maintenance
13. Decommissioning
14. Supporting processes

Annexes

CYBERENG

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

With the support of the Erasmus+ Programme of the European Union

CONTENT

ISO/SAE CD 21434 Road Vehicles — Cybersecurity engineering

4. **General considerations**
5. **Overall Cybersecurity Management** →
6. Project dependent cybersecurity management
7. Risk assessment methods
8. Concept phase
9. Product development
10. Production
11. Operations
12. Maintenance
13. Decommissioning
14. Supporting processes

Annexes

a) cybersecurity (risk) management within the organization
b) cybersecurity governance and cybersecurity culture in the organization
c) information sharing and vulnerability disclosure

CYBERENG

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

With the support of the Erasmus+ Programme of the European Union

CONTENT

ISO/SAE CD 21434 Road Vehicles — Cybersecurity engineering

4. **General considerations**
5. **Overall Cybersecurity Management**
6. **Project dependent cybersecurity management**
7. Risk assessment methods
8. Concept phase
9. Product development
10. Production
11. Operations
12. Maintenance
13. Decommissioning
14. Supporting processes

Annexes

a) responsibilities regarding cybersecurity
b) plan of cybersecurity activities
c) Management of identified cybersecurity vulnerabilities
d) cybersecurity case
e) cybersecurity assessment

CYBERENG

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

With the support of the Erasmus+ Programme of the European Union

CONTENT

ISO/SAE CD 21434 Road Vehicles — Cybersecurity engineering

4. **General considerations**
5. **Overall Cybersecurity Management**
6. **Project dependent cybersecurity management**
7. **Risk assessment methods**
8. Concept phase
9. Product development
10. Production
11. Operations
12. Maintenance
13. Decommissioning
14. Supporting processes

Annexes

a) Identify assets, cybersecurity properties and damage
b) identify threat scenarios
c) rate the impact of damage scenarios
d) identify and analyze cybersecurity vulnerabilities
e) identify attack paths
f) Assessment of attack likelihood
g) determine risk level
h) select risk treatment

CYBERENG
The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.
With the support of the Erasmus+ Programme of the European Union

CONTENT

ISO/SAE CD 21434 Road Vehicles — Cybersecurity engineering

4. **General considerations**
5. **Overall Cybersecurity Management**
6. **Project dependent cybersecurity management**
7. **Risk assessment methods**
8. **Concept phase**
9. Product development
10. Production
11. Operations
12. Maintenance
13. Decommissioning
14. Supporting processes

Annexes

a) Cybersecurity relevance
b) Item description
c) Threat scenarios for the item
d) Risk level
e) Risk treatment
f) Cybersecurity goals
g) Cybersecurity requirements
h) Rationale for cybersecurity requirements
i) Allocate cybersecurity requirements
j) Verify cybersecurity concept

CYBERENG
The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.
With the support of the Erasmus+ Programme of the European Union

CONTENT

ISO/SAE CD 21434 Road Vehicles — Cybersecurity engineering

4. **General considerations**
5. **Overall Cybersecurity Management**
6. **Project dependent cybersecurity management**
7. **Risk assessment methods**
8. **Concept phase**
9. **Product development**
10. Production
11. Operations
12. Maintenance
13. Decommissioning
14. Supporting processes

Annexes

a) Cybersecurity architectural design
b) Compliance of architectural design with cybersecurity requirements and concept
c) Appropriate cybersecurity controls
d) Vulnerabilities in the design
e) Evidence for cybersecurity
f) Cybersecurity goals and cybersecurity claims
g) Item satisfies the cybersecurity goals
h) Residual risk
i) Release for post-development

CONTENT

ISO/SAE CD 21434 Road Vehicles — Cybersecurity engineering

4. **General considerations**
5. **Overall Cybersecurity Management**
6. **Project dependent cybersecurity management**
7. **Risk assessment methods**
8. **Concept phase**
9. **Product development**
10. **Production**
11. Operations
12. Maintenance
13. Decommissioning
14. Supporting processes

Annexes

a) Cybersecurity requirements from concept and product development phases implemented in the product
b) Cybersecurity processes to prevent the introduction of cybersecurity vulnerabilities in production phase.

CYBERENG

With the support of the Erasmus+ Programme of the European Union

CONTENT

ISO/SAE CD 21434 Road Vehicles — Cybersecurity engineering

4. **General considerations**
5. **Overall Cybersecurity Management**
6. **Project dependent cybersecurity management**
7. **Risk assessment methods**
8. **Concept phase**
9. **Product development**
10. **Production**
11. **Operations**
12. Maintenance
13. Decommissioning
14. Supporting processes

Annexes

a) Collection of cybersecurity information
b) Triage of cybersecurity information
c) Processes for assessing cybersecurity events
d) Processes for cybersecurity events that are elevated to a cybersecurity incident.

CYBERENG

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

With the support of the Erasmus+ Programme of the European Union

CONTENT

ISO/SAE CD 21434 Road Vehicles — Cybersecurity engineering

4. **General considerations**
5. **Overall Cybersecurity Management**
6. **Project dependent cybersecurity management**
7. **Risk assessment methods**
8. **Concept phase**
9. **Product development**
10. **Production**
11. **Operations**
12. **Maintenance**
13. Decommissioning
14. Supporting processes

Annexes

a) Cybersecurity requirements and responsibilities relating to service and repair
b) Provision of information on service and repair to preserve the cybersecurity of the product
c) Cybersecurity during and after updates

CYBERENG

With the support of the Erasmus+ Programme of the European Union

CONTENT

ISO/SAE CD 21434 Road Vehicles — Cybersecurity engineering

4. **General considerations**
5. **Overall Cybersecurity Management**
6. **Project dependent cybersecurity management**
7. **Risk assessment methods**
8. **Concept phase**
9. **Product development**
10. **Production**
11. **Operations**
12. **Maintenance**
13. **Decommissioning**
14. Supporting processes

Annexes

a) Ensure that products are decommissioned in a secure manner
b) Methods to communicate end of support

CYBERENG

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

With the support of the Erasmus+ Programme of the European Union

CONTENT

ISO/SAE CD 21434 Road Vehicles — Cybersecurity engineering

4. **General considerations**
5. **Overall Cybersecurity Management**
6. **Project dependent cybersecurity management**
7. **Risk assessment methods**
8. **Concept phase**
9. **Product development**
10. **Production**
11. **Operations**
12. **Maintenance**
13. **Decommissioning**
14. **Supporting processes**

Annexes

a) Quality and information security management to support cybersecurity
b) Interactions, dependencies and responsibilities between customers and suppliers for cybersecurity
c) Tools used during the lifecycle do not adversely affect cybersecurity.

CYBERENG

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

With the support of the Erasmus+ Programme of the European Union

CONTENT

ISO/SAE CD 21434 Road Vehicles — Cybersecurity engineering

4. General considerations
5. Overall Cybersecurity Management
6. Project dependent cybersecurity management
7. Risk assessment methods
8. Concept phase
9. Product development
10. Production
11. Operations
12. Maintenance
13. Decommissioning
14. Supporting processes

Annexes

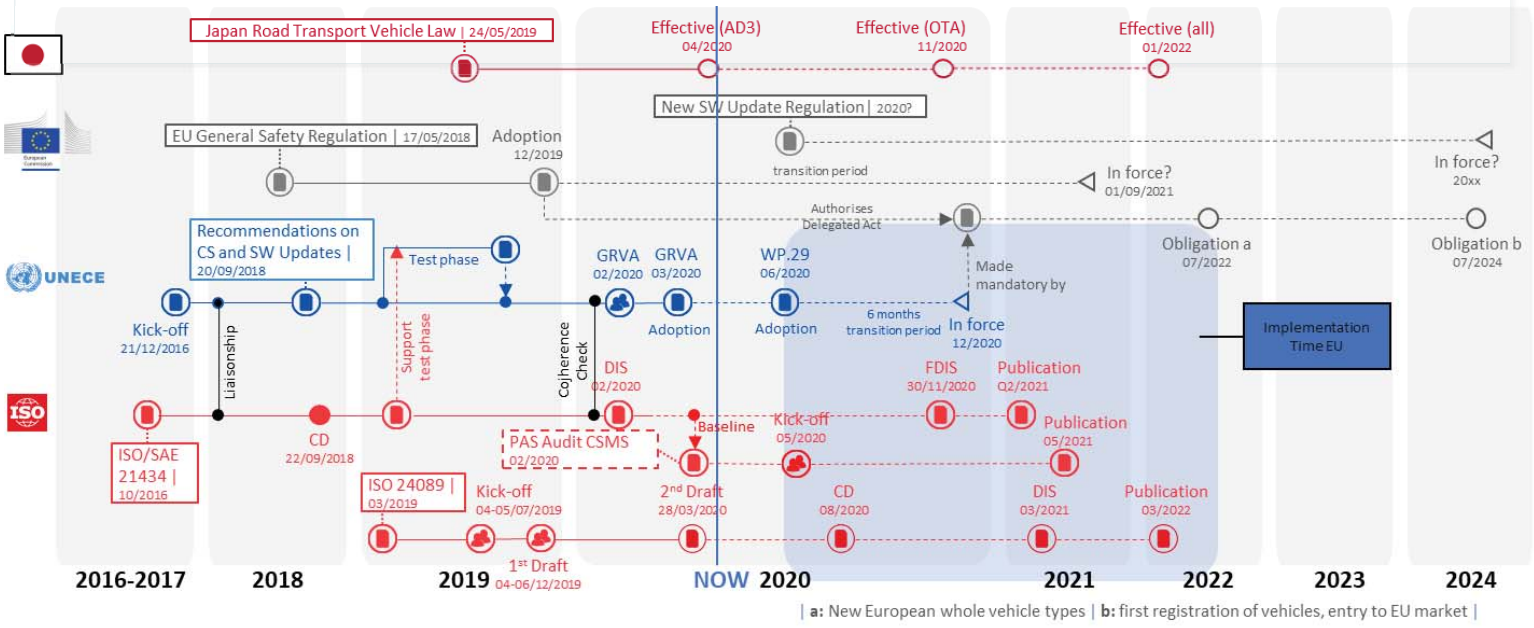
a) Summary of activities
b) Examples of cybersecurity culture
c) Examples for tailoring and distributed activities
d) Interface agreement example
e) Example for cybersecurity relevance
f) Verification & Validation methods
g) Cybersecurity assurance level
h) Impact ratings
i) Attack Feasibility rating

ONGOING DEVELOPMENTS

Automotive

- ISO/AWI 24089 Road vehicles — Software update engineering
  - Upcoming standard for automotive software updates
- ISO/WD PAS 5112 Road vehicles — Guidelines for auditing cybersecurity engineering
  - New development, describing how to audit a cybersecurity process

Timeline



CYBERENG
The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.
With the support of the Erasmus+ Programme of the European Union

ISO/SAE 21434

Risk Management Scenario with ThreatGet

CYBERENG

With the support of the Erasmus+ Programme of the European Union

SCENARIO

- Abbreviated
- Contains the step till security goals

SCENARIO

- Electronic Steering Column Lock with "keyless go"
  - The system should deactivate and activate the car lock based on proximity of key and operation of door handle
  - The system should allow start of the engine when the key is inside the car
  - The system is integrated with other systems
  - Key stores the latest 50 car accesses

ITEM DEFINITION

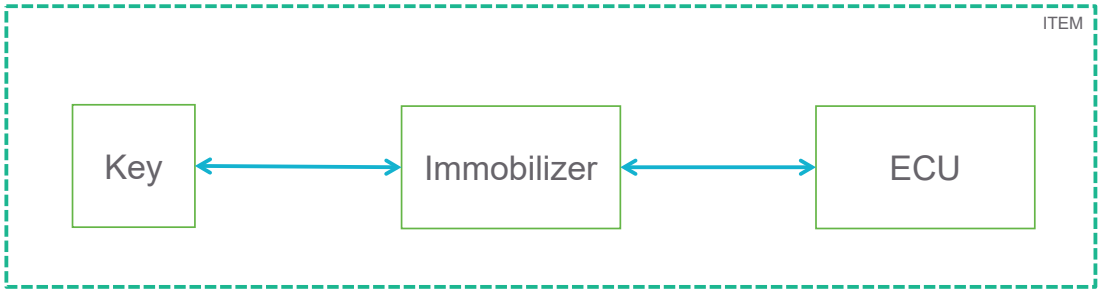
- Includes
  - Item boundary
  - Function
  - Preliminary Architecture

CYBERENG

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

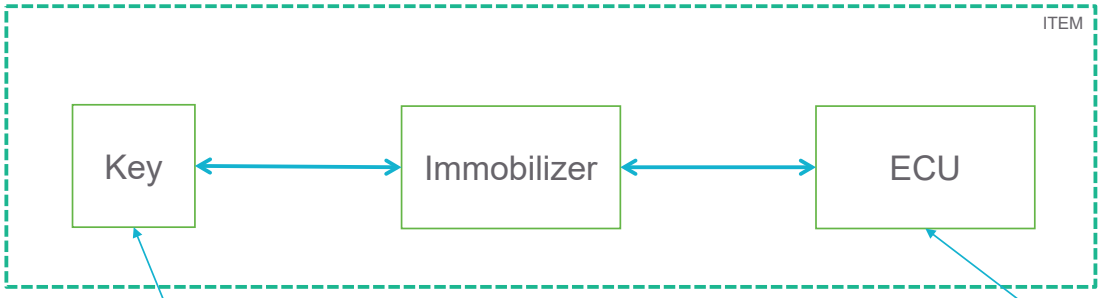
With the support of the Erasmus+ Programme of the European Union

ITEM DEFINITION



CYBERENG
The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.
With the support of the Erasmus+ Programme of the European Union

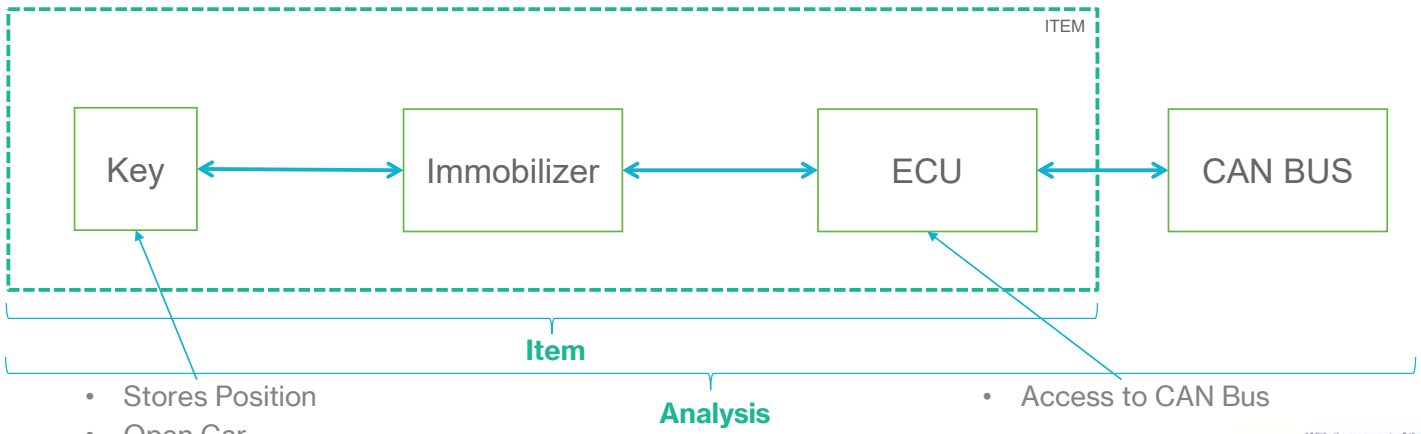
ITEM DEFINITION



- Stores Position
- Open Car

- Access to CAN Bus

ITEM DEFINITION

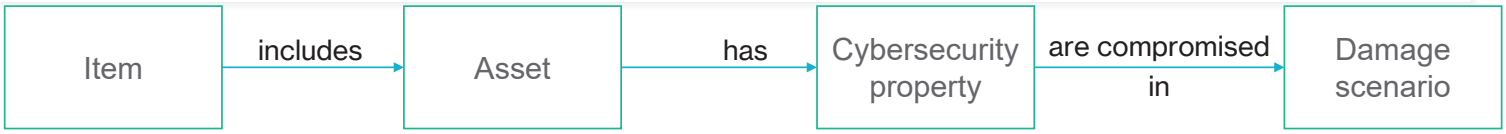


ITEM DEFINITION

Image credit: tag-cyber (https://www.tag-cyber.com/media/charlie-ciso)

CYBERENG
The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.
With the support of the Erasmus+ Programme of the European Union

WHAT DO WE PROTECT



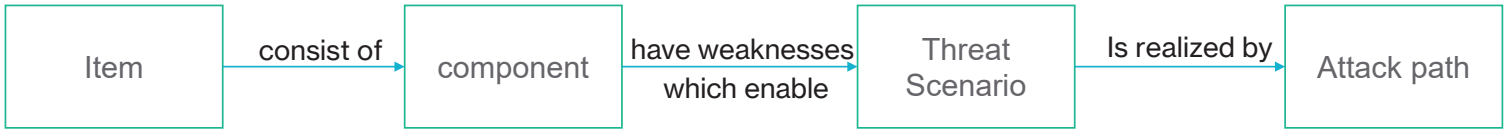
Utilization of SAHARA: Apply STRIDE to data and functions in order to identify assets, their cybersecurity properties and impact of damage scenarios

- **Stores Position**
  - (I) Disclosure of position information would impact driver => Confidentiality
- **Open Car**
  - (S) Spoofed signal could allow attacker to open car => Integrity
- **Access to CAN Bus**
  - (T) Tampering of data transmitted over CAN could impact vehicle operation => Integrity

For more information about SAHARA: https://www.researchgate.net/profile/Georg_Macher/publication/291349648_SAHARA_A_security-aware_hazard_and_risk_analysis_method/links/5deb8535299bf10bc346a9f8/SAHARA-A-security-aware-hazard-and-risk-analysis-method.pdf
For more information about STRIDE: https://en.wikipedia.org/wiki/STRIDE_(security)

CYBERENG
The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.
With the support of the Erasmus+ Programme of the European Union

WHAT COULD ATTACK US



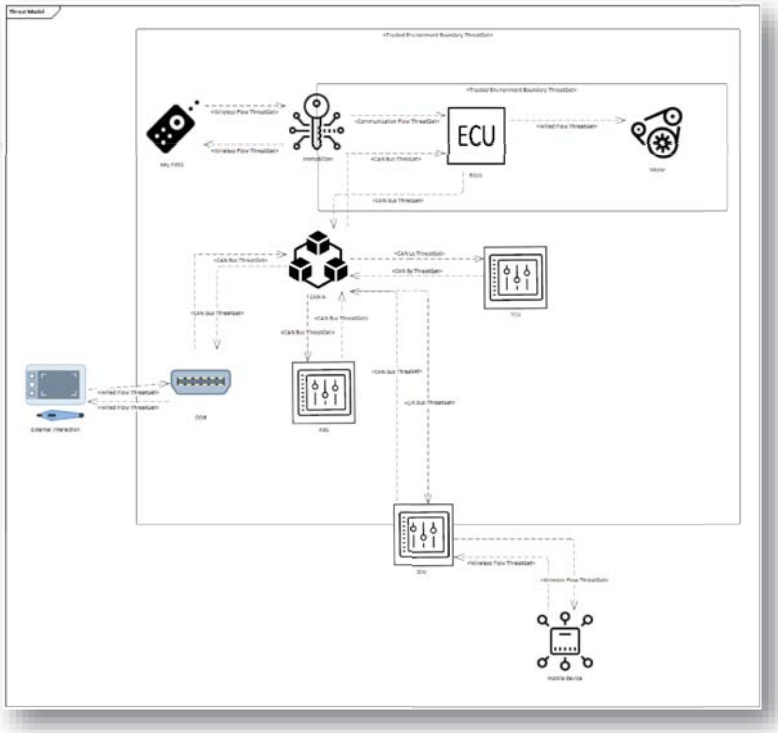
Utilization of Threat Modeling

- System Model with known security properties
- Threat Model
  - Combination => all potential Threats to the System

Automotive Threat Modeling:
https://www.researchgate.net/profile/Christoph_Schmittner/publication/312189316_Threat_Modeling_for_Automotive_Security_Analysis/links/58dbb49f92851c611d024a66/Threat-Modeling-for-Automotive-Security-Analysis.pdf
ThreatGet: Threat modeling based approach for automated and connected vehicle systems (https://ieeexplore.ieee.org/abstract/document/9094555)

WHAT COULD ATTACK US

- System Model for Threat Analysis requires more details
  - Assets are added as properties
    - Adding Privacy Asset to Key Fob



Automotive Threat Modeling:
https://www.researchgate.net/profile/Christoph_Schmittner/publication/312189316_Threat_Modeling_for_Automotive_Security_Analysis/links/58dbb49f92851c611d024a66/Threat-Modeling-for-Automotive-Security-Analysis.pdf
ThreatGet: Threat modeling based approach for automated and connected vehicle systems (https://ieeexplore.ieee.org/abstract/document/9094555)
ThreatGet: https://www.threatget.com/

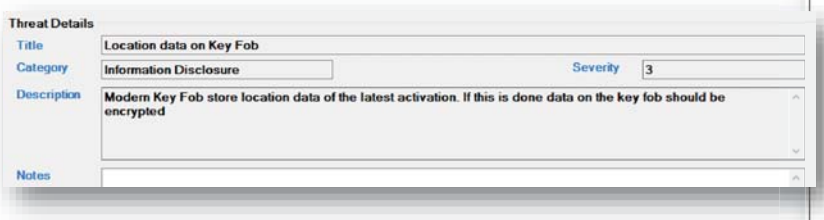
CYBERENG

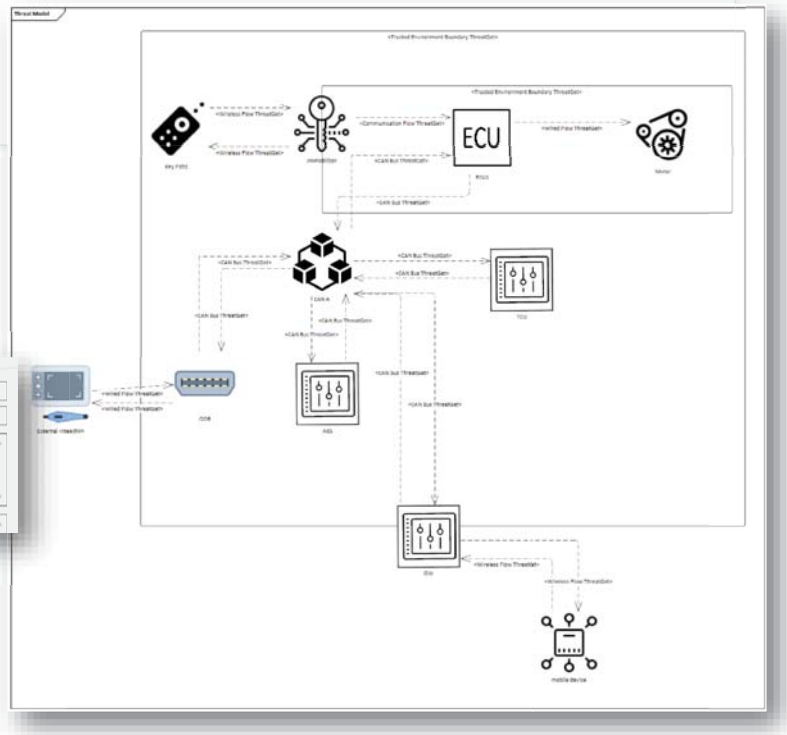
The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

With the support of the Erasmus+ Programme of the European Union

WHAT COULD ATTACK US

- System Model for Threat Analysis requires more details
  - Assets are added as properties
    - Adding Privacy Asset to Key Fob





Automotive Threat Modeling:
https://www.researchgate.net/profile/Christoph_Schmittner/publication/312189316_Threat_Modeling_for_Automotive_Security_Analysis/links/58dbb49f92851c611d024a66/Threat-Modeling-for-Automotive-Security-Analysis.pdf
ThreatGet: Threat modeling based approach for automated and connected vehicle systems (https://ieeexplore.ieee.org/abstract/document/9094555)
ThreatGet: https://www.threatget.com/

CYBERENG

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

With the support of the Erasmus+ Programme of the European Union

NEXT STEPS

Summary for concept phase

- Rank **Safety, Financial, Operational, Privacy Impact** to Road User in four levels
- Rank **Feasibility** of Attack scenarios in four levels
- Combine Impact and Feasibility to rank risk in five levels
- Decide on Security goals
  - Can be stated in Cybersecurity properties and Asset (and documents as Notes in ThreatGet)
    - **Protect confidentiality of location data on key fob against attacks trough configuration interface**
- Break down cybersecurity goals to (technical) requirements and assign
  - Think about longevity

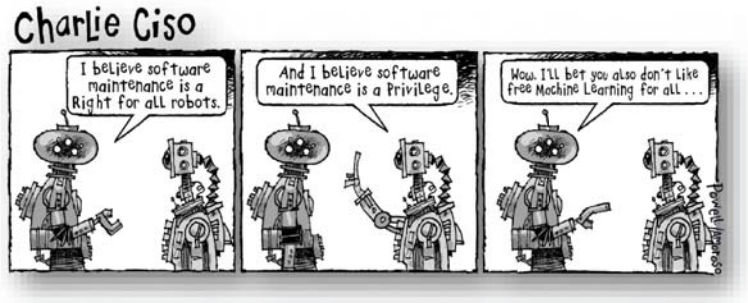


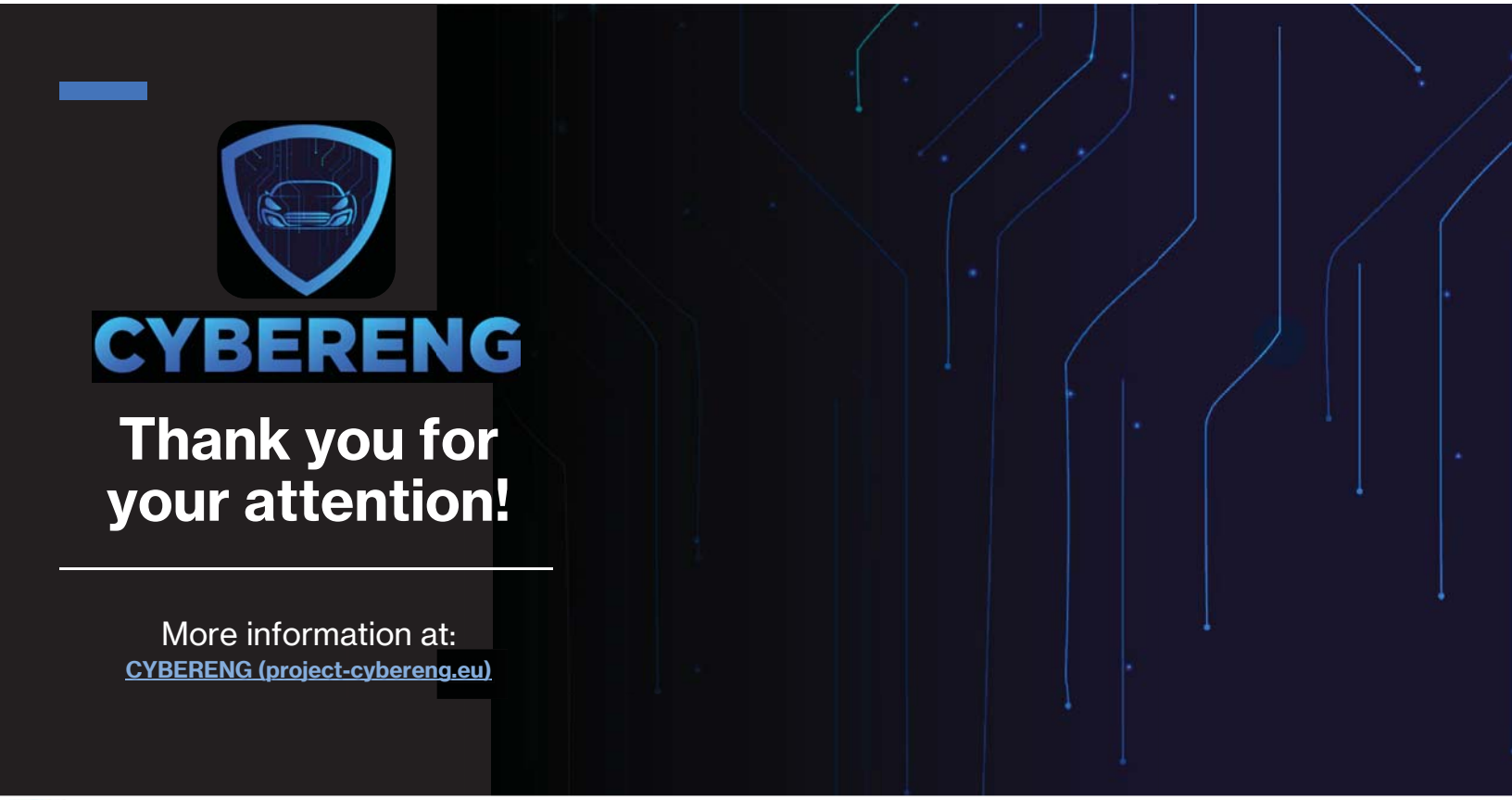
Image credit: tag-cyber (https://www.tag-cyber.com/media/charlie-ciso)

SUMMARY

SUMMARY

- Security is still a novel topic for many domains
- Standards are existing, but practical experience, methods and processes are missing
- Topic is important due to upcoming regulations

- **=> Cybereng will provide cybersecurity expertise for the Automotive domain**



CYBERENG
The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.
With the support of the Erasmus+ Programme of the European Union

Secure and Reliable Smart Cyber Physical Systems

Dr. Samir Ouchani
Lineact CESI, Aix-en-Provence
The 3rd Summer School on Cyber Physical Systems and Internet of Things (SS-CPS&IoT 2022)

Budva, Montenegro, June 7-10, 2022

Outline

Research Statement

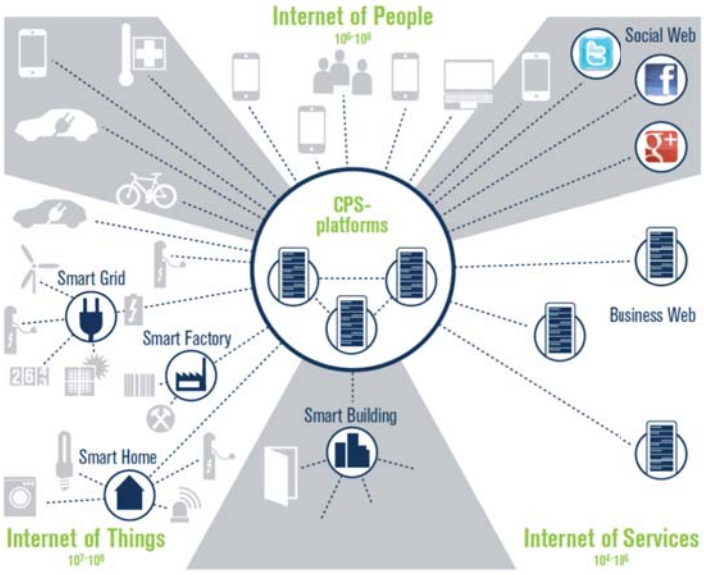
Secure and Reliable SCPS Framework

- SCPS Formalism
- Smart Attacks
- Security Reinforecement
- Security Assessment
- Applications

Research Directions

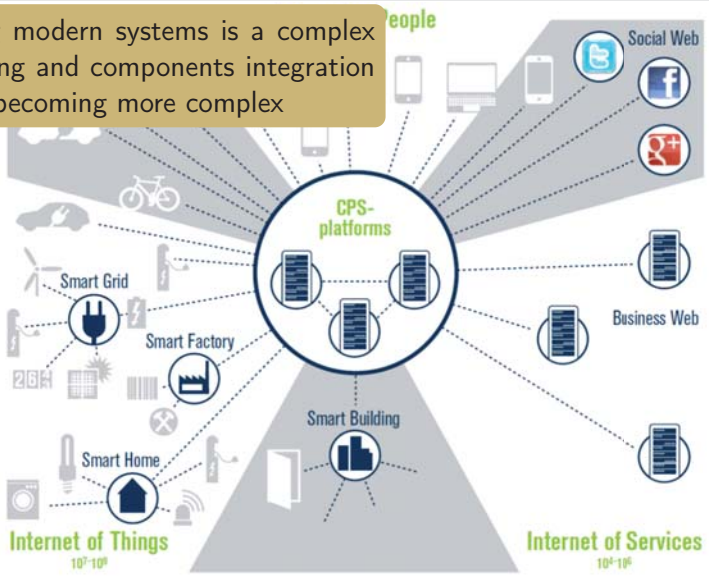
- Collaborations
- Research Interests
- Research Projects

Research Statement



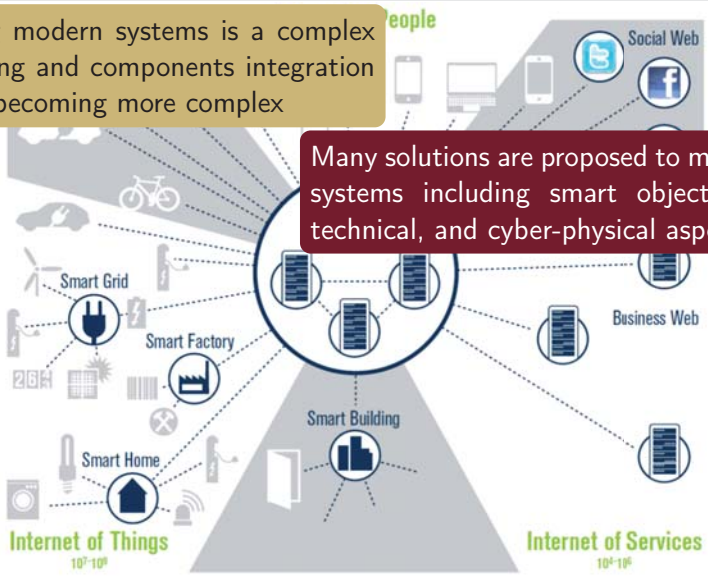
Research Statement

Designing modern systems is a complex undertaking and components integration are even becoming more complex



Research Statement

Designing modern systems is a complex undertaking and components integration are even becoming more complex

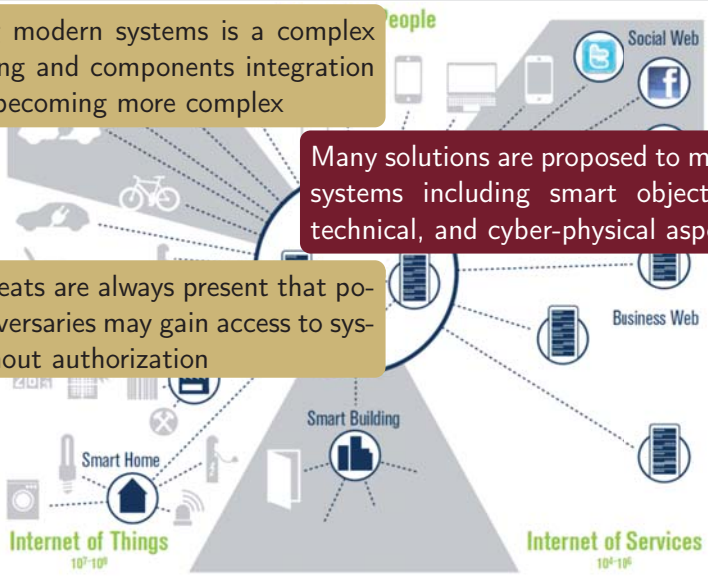


Research Statement

Designing modern systems is a complex undertaking and components integration are even becoming more complex

Many solutions are proposed to model large systems including smart objects, social-technical, and cyber-physical aspects

Major threats are always present that potential adversaries may gain access to systems without authorization



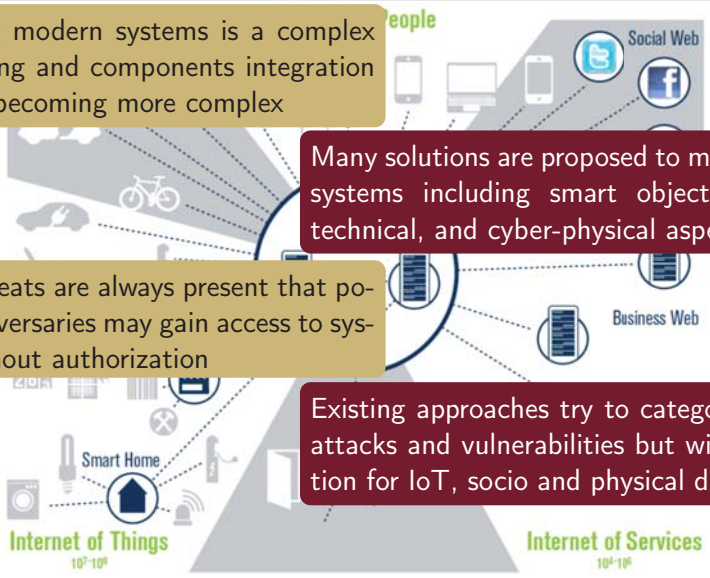
Research Statement

Designing modern systems is a complex undertaking and components integration are even becoming more complex

Many solutions are proposed to model large systems including smart objects, social-technical, and cyber-physical aspects

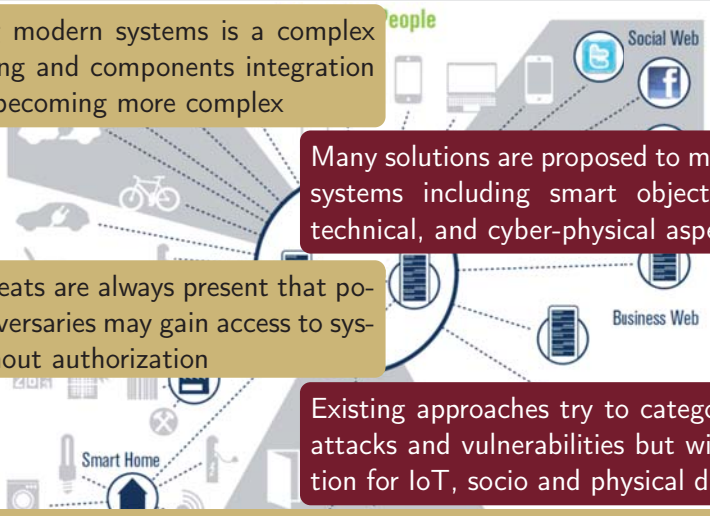
Major threats are always present that potential adversaries may gain access to systems without authorization

Existing approaches try to categorize such attacks and vulnerabilities but with limitation for IoT, socio and physical dimensions



Research Statement

Designing modern systems is a complex undertaking and components integration are even becoming more complex



Many solutions are proposed to model large systems including smart objects, social-technical, and cyber-physical aspects

Major threats are always present that potential adversaries may gain access to systems without authorization

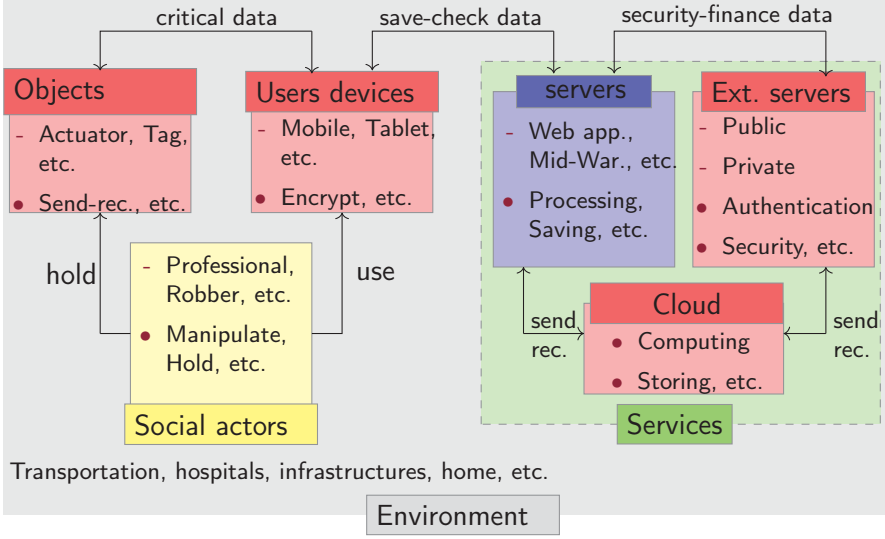
Existing approaches try to categorize such attacks and vulnerabilities but with limitation for IoT, socio and physical dimensions

How modeling, ensuring the functional correctness, constraining security policies, and gauging security in Smart/Industrial CPS ?

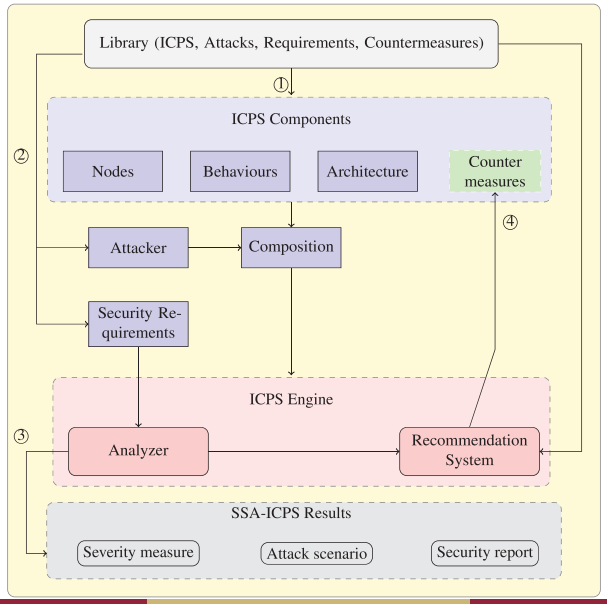
A fully Automatic Framework for Functional and Security Analysis of Smart CPS **is vital**!!!

Secure and Reliable SCPS Framework | Architecture

Architecture



Methodology



A modeling formalism to cover the system environment and its entities

Formal Model

A system S is the tuple $\langle Obj, Srv, Act, Env, Prot \rangle$:

- The connected objects (Obj),
- The environment (Env),
- The client-server applications and services (Srv),
- The social actors (Act), and
- The communication protocols ($Prot$) that ensure the interaction and the communication between the different types of IoT entities

Objects

An object can be physical as digital with specific abilities : container, lockable, movable or/and destroyable.

Object

An object Obj is a tuple $\langle O, attr_O, Actuator_O, \Sigma_O, Beh_O \rangle$, where :

- O is a finite set of tags identifying the objects,
- $attr_O : O \to 2^{\mathbb{T}}$ returns the attributes of an object,
- $Actuator_O : O \to L \times 2^O \times O \times \mathbb{B}$ returns the status of an object $\langle loc_O, cont_O, key_O, locked_O \rangle$,
- Σ_O is a finite set of atomic actions, where : $\Sigma_O = \{\texttt{Start}_O, \texttt{Terminate}_O, \texttt{Send}_O(o,o'), \texttt{Receive}_O(o,o'), \texttt{Update}_O(o,o'), \texttt{Loc}$ $o,o' \in O$ and $l,l' \in L\}$
- $Beh_O : O \to \mathscr{L}_O$ returns the behaviour of an object given by : $B_O ::= \texttt{Start}_O \cdot B_O \cdot \texttt{Terminate}_O \mid \alpha_O \cdot B \mid \alpha_O +_{g_o} \alpha'_O \mid \alpha_O$.

Services

A service Srv ensures a client-server based architecture : client applications, computation servers and web services.

Service

Srv is $\langle V, O_V, srv_V, \Sigma_V, Beh_V \rangle$, where :

- V is a finite set of computing and storage services v, v', etc.
- O_V is a finite set of physical objects hosting services from V.
- $srv_V : O_V \to 2^V$ assigns for a given object a set of services.
- Σ_V is a finite set of actions supported by a service V, where : $\Sigma_V = \{\texttt{Start}_V, \texttt{Terminate}_V, \texttt{Send}_V(o,o'), \texttt{Receive}_V(o,o'), \texttt{Update}_V(o,o'), \texttt{Lock}_V(o,o'), \texttt{Unlock}_V(o,o') : o,o' \in O\}$,
- $Beh_V : O_V \to \mathscr{L}_V$ returns the behaviour of an object hosting a service : $B_V ::= \texttt{Start}_V \cdot B_V \mid \alpha_V +_{g_V} \alpha'_V \mid \alpha_V$.

Actors

Actors can be human being or smart robot agents.

Actor

Act is $\langle A, categ_A, \Sigma_A, Bev_A \rangle$ where :

- A is a finite set of actors.
- $categ_A : A \to \mathbb{C}$ returns the category of an actor.
- $Actuator_A : A \to L \times 2^O$ returns the location ($loc_A \in L$) and the possessed objects ($poss_A \subseteq 2^O$) by an actor.
- The finite set of the actors actions Σ_A encloses all actions that can be executed by an agent.
  $\Sigma_A = \{\texttt{Start}_A, \texttt{Moving}_A(l,l'), \texttt{Lock}_A(o,o'), \texttt{Unlock}_A(o,o'), \texttt{Send}_A(o,x),$
  $\texttt{Receive}_A(o,x), \texttt{Update}_A(o,o'), \texttt{Terminate}_A : l,l' \in L$ and $o,o' \in$
  O and $a \in A$ and $x \in L \cup O \cup A\}$
- $Bev_A : A \to \mathscr{L}_A$ expresses the behaviour of an actor by
  $B ::= Stop \mid \alpha_A.B \mid B + B \mid B +_g B \mid B +_p B.$

Environment

Env can be any human body or other natural species, or even a physical space that hosts objects.

Environment

Env is a tuple $\langle E, L, O_E, Actuator_E \rangle$, where :

- E is a finite set of environments denoted by e, e', etc.
- L is a finite set of locations (l, l', etc.).
- O_E is a finite set of physical objects of type container.
- $Actuator_E : O_E \times O_E \rightarrow 2^O$ returns the set of objects linking containers by physical objects (e.g. doors connecting two rooms).

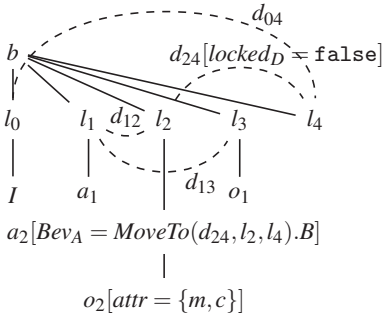
Interaction Protocol

$Prot$ orchestrates the communication between entities.

- $Prot$ is a tuple $\langle Prot_{h,o}, Prot_{o,o}, Prot_{o,s} \rangle$ where $Prot_{h,o}$ ensures the communications between social actors and the objects, $Prot_{o,o}$ between objects, $Prot_{o,s}$ between objects and services on servers
- A state $S = \langle S_O, S_V, S_A, S_E \rangle$ is an instance of $\langle Obj, Srv, Act, Env \rangle$ composed from states of objects, services, actors, and the environment
- The transitions between states are denoted by $S \overset{\ell,c,p}{\hookrightarrow} S'$, ℓ names the action to be executed with a cost c and a probability p

Interaction Protocol

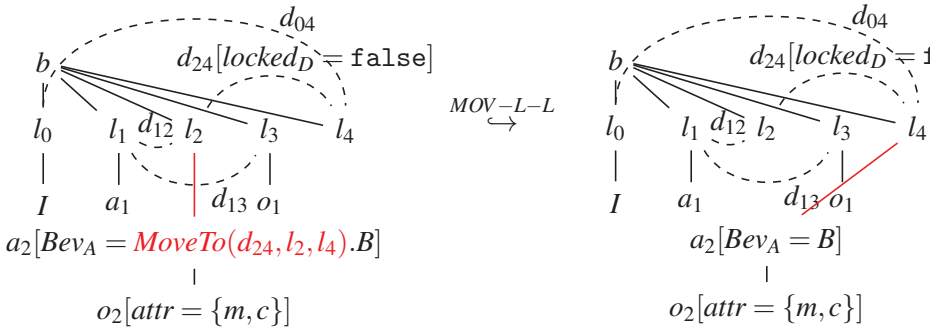
Example of a state



- A state is represented as a labelled multi-graph
- One initial vertex represents the name of the system
- Nodes are location, actors and objects
- Edges show the relation between the entities

Interaction Protocol

- This execution rule shows moving a_2 from l_2 to l_4 (MOV-L-L) : $\overset{moving_{a_2}(d_{24},l_2,l_4),c,p}{\hookrightarrow}$



Interaction Protocol

SYN-O-O two physical objects o and o' exchange a digital object o''

$$\frac{\begin{array}{c} Beh_O(o) = \texttt{Send}_O(o', [\![o'']\!]).Beh'_O(o) \wedge o'' \in cont_O(o) \wedge [\![o'']\!] \neq \varepsilon_o \\ Beh_O(o') = \texttt{Receive}_O(o''', [\![o'']\!]).Beh'_O(o') \wedge o''' \in cont_O(o) \wedge p \notin attr_O(o'') \end{array}}{\begin{array}{c} \langle\langle o, -, <-, \{o'', [\![o'']\!]\} >, -\rangle, \langle o', -, <-, \{o''', [\![o''']\!]\} >, -\rangle\rangle \overset{\texttt{Send}_O(o,o',[\![o'']\!]),c,p}{\hookrightarrow} \\ \langle\langle o, Beh'_O(o), <-, \{o'', [\![o'']\!]\} >, -\rangle, \langle o', Beh'_O(o'), <-, \{o''', [\![o'']\!]\} >, -\rangle\rangle \end{array}}$$

Interaction Protocol

REC-A-O an actor a takes an object o' from an object o.

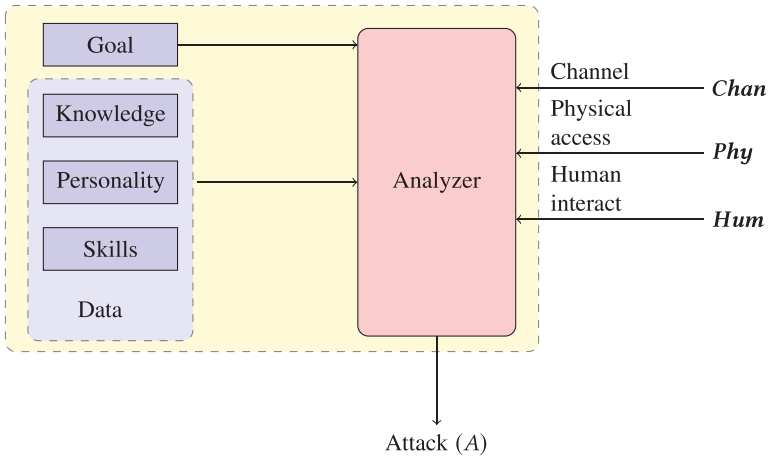
$$\frac{\begin{array}{c} Bev_A(a) = \texttt{Receive}_A(o,o').Bev'_A(a) \wedge loc_A(a) = loc_O(o) \\ \neg locked_O(o) \wedge o' \in cont_O(o) \wedge p \in attr_O(o') \end{array}}{\begin{array}{c} \langle\langle a,-,<-,->,-\rangle,\langle o,-,<-,\{o'\}>,-\rangle\rangle \overset{\texttt{Receive}_A(a,o,o'),c,p}{\hookrightarrow} \\ \langle\langle a,Bev'_A(a),<-,\{o'\}>,-\rangle,\langle o,Beh'_O(o),<-,->,-\rangle\rangle \end{array}}$$

Interaction Protocol

LOC-O-O encrypts an object o' by an object o using o''.

$$\frac{Beh_O(o) = \texttt{Lock}_O(o',o'').Beh'_O(o) \wedge \{o',o''\} \subset cont_O(o) \wedge [\![o',o'']\!] \neq \varepsilon_o}{\langle\langle o,-,<-,\{o',o''\}>,-\rangle,\langle o',-,<-,->,\neg locked_O(o')\rangle\rangle \overset{lock_O(o,o',o''),c,p}{\hookrightarrow} \langle\langle o,Beh'_O(o),<-,\{o',o''\}>,-\rangle,\langle o',-,<-,->,locked_O(o')\rangle\rangle}$$

ICPS Attacker



ICPS Attacker

Data ω of the attacker is a tuple $\langle K, P, S \rangle$, where K represents its knowledge, P is the personality of the attacker, and S is the set of skills.
Knowledge $\mathscr{K} = \langle \mathscr{M}, \mathscr{S}ec, \mathscr{A}lg \rangle$ contains the system model $(\mathscr{M})$, a secure information $(\mathscr{S}ec)$, and algorithms of control or security techniques $(\mathscr{A}lg)$
The **personalty** is a vector $\mathscr{P}er$, where each element is an emotion e in the state i. $\mathscr{P}er = \begin{vmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{vmatrix}$ where $\forall i \in [1,n]$, $e_i = \begin{vmatrix} 1 & positive \\ 0 & absence \\ -1 & negative \end{vmatrix}$
An **attacker goal** $\mathscr{G} = \{g_1, g_2, \ldots, g_n\}$ is a **Tree** representing a pre-ordered set of actions or small attacks where each action has a goal

ICPS Attacker

The **Analyzer** selects actions to earn more rewards by taking $\mathscr{M}_{mdp} = (\mathscr{S}, \mathscr{A}, s_0, \mathscr{R}, \mathscr{P}, \gamma)$ as input and producing the set of actions π, where :

- $\mathscr{S}$ is a set of finite states s_2, s_2, etc, $\mathscr{A}$ is a set of finite actions a_1, a_2, etc, and s_0 is the initial state.
- $\mathscr{R}(s)$ is a reward function that returns the utility for each state s by using the Weight Sum Method $\mathscr{R}(s) = \sum_{j=1}^{m} w_j c_{ij}$, $i = 1, 2, ..., n$ where c is a set of criterion to select an action and w denotes the relative weight of importance of c.
- $\mathscr{P}(s, a, s')$ depends on the value of $\mathscr{R}(s')$ and the sum of all the $\mathscr{R}(s')$ for the successors of the state s. $\mathscr{P}(s, a, s') = \frac{\mathscr{R}(s')}{\sum_{\forall s' \in succ(s)} \mathscr{R}(s')}$
- γ is a discount factor $0 < \gamma < 1$.

A formal language to express security policies and security requirements

Security Policies and Security Requirements

- A *security policy* is a statement on what may never happen in the system execution
- A *security requirement* is a desirable security property that we would like to be valid despite specific threats
- We express policies and requirements using the language of *security statements*

Definition

A *security statement* is any expression in the language $L(\varphi)$, so defined :

$$\begin{array}{lll} \varphi & ::= & \texttt{true} \mid \varphi_{SP} \mid \varphi \wedge \varphi' \mid \neg\varphi \mid \bigcirc \varphi \mid \varphi \ \mathrm{U} \ \varphi' \\ \varphi_{SP} & ::= & \varphi_{SP} \wedge \varphi'_{SP} \mid \neg\varphi_{SP} \mid d \in conn(l,l') \mid o \in key_D(d) \mid \\ & & (x,a) \in Hist_D(d) \mid (x,a) \in Hist_O(o) \mid z \in |type_O(o) \mid \\ & & y \in attr(o) \mid loc_O(o) = l \mid o \in key_O(o') \mid \\ & & o \in cont_O(o') \mid loc_A(a) = l \mid o \in poss_A(a) \end{array}$$

Security Policies and Security Requirements

- The formal semantics is defined in term of the function $[\![\cdot]\!]_S$ returning the truth value of a security statement φ_{SP} in a state S.
- $[\![d \in conn(l,l')]\!]_S$ iff $(l,d,l') \in C$; door d connects location l and l'
- $[\![y \in attr(o)]\!]_S$ iff $y \in attr(o)$; y is an attribute of the object o
- $[\![loc_O(o) = l]\!]_S$ iff $(l,o) \in (E)^+$; object o is in location l
- $[\![o \in poss_A(a)]\!]_S$ iff $(a,o) \in (E)^+$; object o is possessed by agent a

Security Policies and Security Requirements

- A security statement φ is valid in $\mathscr{S}$, written $\mathscr{S} \models_I \varphi$, when $\texttt{Traces}(\mathscr{S}) \subseteq \texttt{Words}(\varphi)$, where $\texttt{Traces}(\mathscr{S})$ is the set of all *prefix closed traces* of $\mathscr{S}$
- For the set of ω-words $\texttt{Words}(\varphi) = \{\rho \in (2^{\varphi_{SP}})^{\omega} : \rho \models_I \varphi\}$, $\models_I \subseteq (2^{\varphi_{SP}})^{\omega} \times L(\varphi)$ is the smallest relation satisfying
  - $\rho \models_I \texttt{true}$
  - $\rho \models_I \varphi_{SP}$ iff $[\![\varphi_{SP}]\!]_{\rho[0]}$
  - $\rho \models_I \neg\varphi$ iff $\rho \not\models_I \varphi$
  - $\rho \models_I \varphi_1 \wedge \varphi_2$ iff $\rho \models_I \varphi_1$ and $\rho \models_I \varphi_2$
  - $\rho \models_I \bigcirc\varphi$ iff $\rho[1\ldots] \models_I \varphi$
  - $\rho \models_I \varphi_1 \mathrm{U}\ \varphi_2$ iff $\exists j \geq 0 : \rho[j\cdots] \models_I \varphi_2$and $\rho[i\cdots] \models_I \varphi_1,\ \forall 0 \leq i < j$

Security Policies and Security Requirements

- A security policy is a safety property or a negation of a liveness property

Definition

A *policy* is a security statement of the form $\Box\neg\varphi_{SP}$ or $\neg\Box(\varphi_{SP} \rightarrow \Diamond\varphi_{SP})$.

- We do not impose any restrictions on the expression of security requirements

Definition

A *requirement* is a security statement.

Automatically Constraining Security Policies

Policy Constrained Semantics

- By definition an intruder is freed from playing by the rules

Definition (Honest Trace)

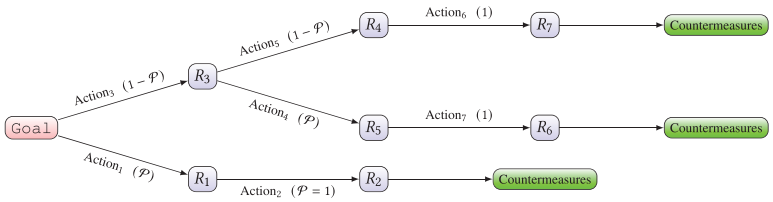
An *honest trace* is a trace whose underlying sequence of states, $S_0 \cdot \ldots \cdot S_i \cdot S_{i+1} \cdot \ldots$ is such that $(S_i, S_{i+1}) \in \hookrightarrow$, for all $i \geq 0$ and where the label of $\hookrightarrow$ is not the intruder's ID.

- For the set of traces $\texttt{Traces}_H(\mathscr{S})$ in $\mathscr{S}$ of honest agents (H), we consider the set of traces satisfying a given security statement

Definition (Trace satisfying φ)

A trace satisfying φ is a trace in $\texttt{traces}(\mathscr{S}, \varphi) = \texttt{Traces}(\mathscr{S}) \cap \texttt{Words}(\varphi)$. An honest trace satisfying φ is a trace in $\texttt{traces}_H(\mathscr{S}, \varphi) = \texttt{Traces}_H(\mathscr{S}) \cap \texttt{Words}(\varphi)$.

Policy Constrained Semantics



Policy Constrained Semantics

- In an honest trace, the affectedness distinguishes the requirements whose validity can be changed if the policy is enforced from those whose validity is unchanged by it.

Definition (Requirements/Policies Affectedness)

Let φ be a requirement, π a policy, and $\mathscr{S}$ models the executions. We say that φ is affected by π in $\mathscr{S}$, and we write it $\varphi \leftharpoonup \varphi'$, when $\texttt{traces}_H(\mathscr{S}, \varphi) \subseteq \texttt{traces}_H(\mathscr{S}, \neg\pi) \neq \emptyset$.

- In $\mathscr{S}_{|\pi}$ where $\mathscr{S}$ is enforced by π, no requirement must change its validity

Definition

The system $\mathscr{S}$ constrained by π is a new $\mathscr{S}' = \langle S', S_0, \hookrightarrow' \rangle$ satisfying.

- If $\mathscr{S} \not\models_H \pi$ then $\mathscr{S}' \models_H \pi$;
- For all p such that $p \not\leftharpoonup \pi$, if $\mathscr{S} \models_H p$ then $\mathscr{S}' \models_H p$.

Policy Constrained Semantics

- Procedure Reduce produces $\mathscr{S}_{|\pi}$ from $\mathscr{S} = \langle S, S_0, \hookrightarrow \rangle$ and π such that $p \not\vdash \pi$. Reduce $(\mathscr{S}, \pi) \rightarrow \mathscr{S}_{|\pi}$ distinguishes two cases

- Case of $\pi = \Box\neg\varphi_{SP}$:
  - **Forall** $S_i \in S : \exists \rho \in \mathtt{Traces}_H(\mathscr{S})$, $\rho = S_0 \cdots S_i \cdots$ **and** $\rho[i \cdots] \models \varphi_{SP}$ **Do** $S' := S \backslash \{S\}$;
  - **Forall** $(S', S_i) \in \hookrightarrow$ **Do** $\hookrightarrow' := (\hookrightarrow \backslash \{(S', S)\}) \cup \{(S', S')\}$;
  - **Forall** $(S_i, S') \in \hookrightarrow$ **Do** $\hookrightarrow' := (\hookrightarrow \backslash \{(S', S_i)\})$.

- Case of $\pi = \neg\Box\varphi_{SP} \rightarrow \Diamond\varphi_{SP}$:
  - **Forall** $S_i, S_j \in S : \exists \rho \in \mathtt{Traces}_H(\mathscr{S})$ $\rho = S_0 \cdots S_i \cdots S_j \cdots$ **and** $\rho[i \cdots] \models \pi\}$ **Do** $\hookrightarrow' := (\hookrightarrow \backslash \{(S_{j-1}, S_j)\}) \cup \{(S_{j-1}, S_{j-1})\}$.

Policy Constrained Semantics

- $\models_H$ stresses that the policy is enforced on the system's execution without the interference of the intruder
- The constrained system will be secure only when $\texttt{Traces}(\mathscr{S}_{|\pi}) \subseteq \texttt{Words}(p)$

Proposition (Soundness)

$\mathscr{S}' = Reduce(\mathscr{S}, \pi)$ is a system constrained by π.

- Each **For all**'s has at most $O(|S|^2)$ iterations

Proposition

$Reduce(\mathscr{S}, \pi)$ can be implemented with worst-case time complexity $O(|S|^2 \cdot check(\pi))$. Here, $check(\pi)$ is the complexity of checking π.

Automatically Evaluating Security

Abstraction

- Ψ considers a PCTL expression ϕ to be verified on S
- Σ_ϕ is the set of the atomic propositions of ϕ s.t. $\Sigma_\phi \subseteq \mathscr{N}$

Definition

For a given System $S \uparrow_{\mathtt{a}} S'$ and a PCTL expression ϕ such that $\Sigma_\phi \subseteq \mathscr{N}$, we have

- $\forall \mathtt{a_x} \notin \Sigma_\phi \wedge \mathtt{a_x} \in \mathscr{N} \cup \mathscr{N}' : \Psi(\mathtt{a_x} \rightarrowtail \mathtt{N}) = \mathtt{N}$.
- $\Sigma_\phi \cap \mathscr{N}_{S'} = \emptyset : \Psi(S \uparrow_{\mathtt{a}} S') = S$.

Reduction

- After abstraction, the size of S will be reduced
- Υ develops a set of reduction rules to compact more the resulted S

Definition

For a system S, we define a set of reduction rules that are applicable on the artifacts $\|$, $|$, $\blacklozenge$, and $\Diamond$ as follows.

- $\Upsilon(\|(a_1, \|(a_2, a_3))) = \|(a_1, a_2, a_3)$,
- $\Upsilon(|(a_1, |(a_2, a_3))) = |(a_1, a_2, a_3)$,
- $\Upsilon(\blacklozenge(a_1, \blacklozenge(a_2, a_3))) = \blacklozenge(a_1, a_2, a_3)$,
- $\Upsilon(\Diamond_p(a_1, \Diamond_{p'}(a_2, a_3))) = \Diamond_{p.p', p.(1-p'), (1-p).(1-p')}(a_1, a_2, a_3)$,
- $\Upsilon(\Diamond_g(a_1, \Diamond_{g'}(a_2, a_3))) = \Diamond_{g \wedge g', \neg g \wedge g', \neg g \wedge \neg g'}(a_1, a_2, a_3)$.

Property decomposition

- The decomposition operator "$\natural$" decomposes the PCTL property ϕ into local ones $\phi_{i:0\leq i\leq n}$ over S_i with respect to the call behavior actions $a_{i:0\leq i\leq n}$ (interfaces)
- The operator "$\natural$" is based on substituting the propositions of S_i to the propositions related to its interface a_{i-1}
- We denote by $\phi[y/z]$ substituting the atomic proposition "z" in the PCTL property ϕ by the atomic proposition "y"

Definition (PCTL Property Decomposition)

Let ϕ be a PCTL property to be verified on $S_1 \uparrow_a S_2$. The decomposition of ϕ into ϕ_1 and ϕ_2 is denoted by $\phi \equiv \phi_1 \natural_a \phi_2$ where AP_{S_i} are the atomic propositions of S_i, then :

1. $\phi_1 = \phi([l_a/AP_{S_2}])$, where l_a is the atomic proposition related to the action a in S_1.
2. $\phi_2 = \phi([\top/AP_{S_1}])$.

Property decomposition

- The first rule is based on the fact that the only transition to reach a state in S_2 from S_1 is the transition of the action l_a (BH-1)
- The second rule ignores the existence of S_1 while it kept unchanged till the execution of BH-2

Property

The decomposition operator $\natural$ is associative for $S_1 \uparrow_{a_1} S_2 \uparrow_{a_2} S_3$, i.e. :

$$\phi_1 \natural_{a_1} (\phi_2 \natural_{a_2} \phi_3) \equiv (\phi_1 \natural_{a_1} \phi_2) \natural_{a_2} \phi_3 \equiv \phi_1 \natural_{a_1} \phi_2 \natural_{a_2} \phi_3 .$$

Composition Verification

- For the verification of ϕ on $S_1 \uparrow_{a_1} S_2$, we deduce the satisfiability of ϕ from the satisfiability of local properties ϕ_1 and ϕ_2 obtained by the operator $\natural$.

Theorem (Compositional Verification)

The decomposition of the PCTL property ϕ by the decomposition operator $\natural$ for $S_1 \uparrow_{a_1} S_2$ is sound, i.e. :

$$\frac{S_1 \models \phi_1 \qquad S_2 \models \phi_2 \qquad \phi = \phi_1 \natural_{a_1} \phi_2}{S_1 \uparrow_{a_1} S_2 \models \phi}$$

Composition Verification

- We generalize the satisfiability of ϕ on S with n call behaviors

Proposition (CV-Generalization)

Let ϕ be a PCTL property to be verified on S, such that :
$S = S_0 \uparrow_{a_0} \cdots \uparrow_{a_{n-1}} S_n$ and $\phi = \phi_0 \natural_{a_0} \cdots \natural_{a_{n-1}} \phi_n$, then :

$$\frac{S_0 \models \phi_0 \cdots S_n \models \phi_n \qquad \phi = \phi_0 \natural_{a_0} \cdots \natural_{a_{n-1}} \phi_n}{S_0 \uparrow_{a_0} \cdots \uparrow_{a_{n-1}} S_n \models \phi}$$

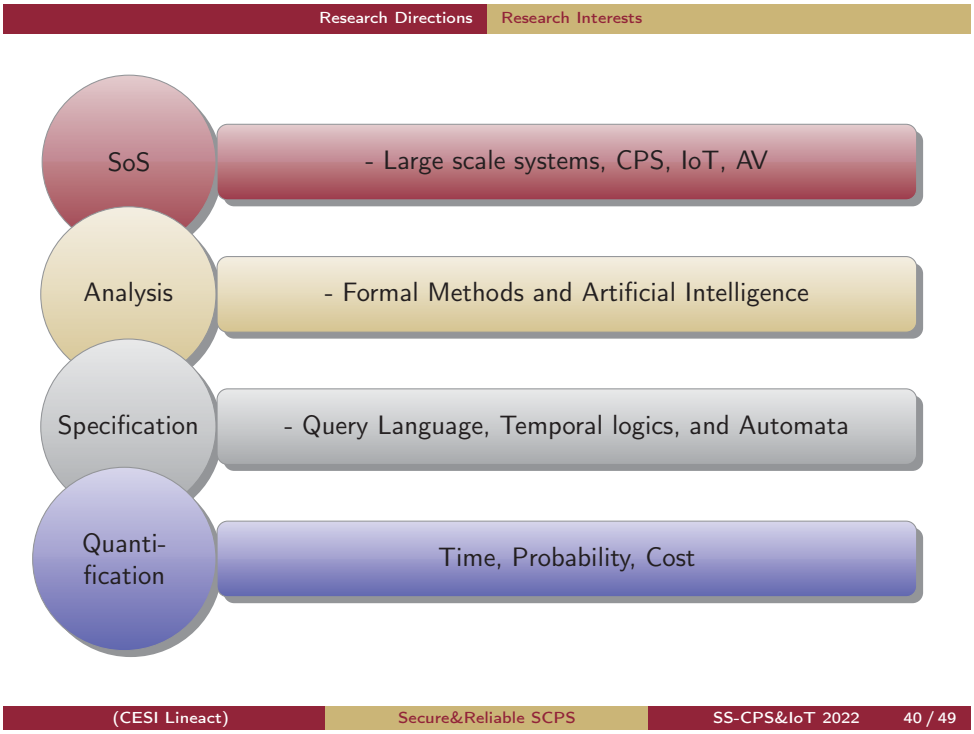
Transformation to PRISM

- $\mathscr{T}_P$ assigns for each entity an equivalent PRISM code fragment
- $o_{o_2} \stackrel{def}{=}$ the object o possess o_2, l_a and l_o present the locations of a and o, and p_{o_3} precises the physicality attribute of o_3.

$$\mathscr{T}_P(\alpha) = \begin{cases} [Syn_{o_2}]o_{o_2} \wedge o_{1_{o_3}} \wedge \neg p_{o_2} \wedge \neg p_{o_3} \rightarrow (o'_2 = o_2); \\ [Syn_{o_2}]o_{o_2} \wedge o_{1_{o_3}} \wedge \neg p_{o_2} \wedge \neg p_{o_3} \rightarrow (o'_3 = o_2); \\ \texttt{iff}:\texttt{Send}_O(o_1,o_2) \in \Sigma_O^{o_1}, \texttt{Receive}_O(o_3,o_2) \in \Sigma_O^{o_2}. \\ [Tak_{o_1}]l_a = l_o \wedge o_{o_2} \wedge \neg lock_o \wedge p_{o_2} \rightarrow (a'_{o_2} = \top); \\ [Tak_{o_1}]l_a = l_o \wedge o_{o_2} \wedge \neg lock_o \wedge p_{o_2} \rightarrow (o'_{o_2} = \bot); \\ \texttt{iff}:\texttt{Receive}_A(o,o_2) \in \Sigma_A^a. \\ [loc_{o_1}]o_{o_1} \wedge o_{o_2} \wedge \neg k_{o_1} \wedge p_{o_1} = p_{o_2} \rightarrow (k'_{o_1} = \top); \\ [loc_{o_1}]o_{o_1} \wedge o_{o_2} \wedge \neg k_{o_1} \wedge p_{o_1} = p_{o_2} \rightarrow (o'_{o_1} = \top); \\ \texttt{iff}:\texttt{Lock}_O(o_1,o_2) \in \Sigma_O^o. \end{cases}$$

Applications

- Utah Water-supply
- Maroochy Shire Sewage Sill
- Model-based systems : SysML, and UML
- Communication protocols
- Smart emergency rooms
- Smart cities
- Smart factories


Research Directions
Research Interests
SoS
- Large scale systems, CPS, IoT, AV
Analysis
- Formal Methods and Artificial Intelligence
Specification
- Query Language, Temporal logics, and Automata
Quanti-
fication
Time, Probability, Cost
(CESI Lineact)
Secure&Reliable SCPS
SS-CPS&IoT 2022
40 / 49

Preserving Privacy in IoT-based Systems

Preserving Privacy in IoT-based Systems

Objectives

A mechanism that preserves data of IoT while ensuring the systems requirements and respecting the system's services partitions

Plan

- **State of the art :** modeling IoT systems and communication protocols, analyzing massive data in IoT, and describing and enforcing data privacy
- **Formalization** : Proposing a formalism that models precisely IoT and expressing very well data privacy
- **Preservation** : Proposing a framework the preserves data privacy in IoT that allows the description formalism developed previously
- **Application** : Applied the contributions on case studies : medical systems (ICOST 2020) and providing a prototype

Security by Construction for Smart Cities

Security by Construction for Smart Cities

Objectives

Construct a robust and secure BC for Smart Cities.

Plan

- **Optimization** : Find the optimal configuration of the smart objects distribution
- **Communication** : Develop a flexible architecture and communication protocol for a smart city
- **Security** : Adapt a distributed security mechanism for the proposed architecture (blockchains, IOTA, etc.)
- **Application** : Applied the contributions on case studies and providing a prototype

Security by Construction for Smart Cities

Security by Construction for Smart Cities

Objectives

Construct a robust and secure BC for Smart Cities.

Plan

- **Optimization** : Find the optimal configuration of the smart objects distribution
- **Communication** : Develop a flexible architecture and communication protocol for a smart city
- **Security** : Adapt a distributed security mechanism for the proposed architecture (blockchains, IOTA, etc.)
- **Application** : Applied the contributions on case studies and providing a prototype

Security by Construction for Smart Cities

Security by Construction for Smart Cities

Objectives

Continuation : Data with smart decision supports

References (PhD W. M. Dahmane, J+SJ+4C/3Y)

- Towards a reliable smart city through formal verification and network analysis. Comput. Commun. 180 : 171-187 (2021)
- Guaranteeing Information Integrity Through Blockchains for Smart Cities. MEDI 2021 : 199-212
- A BIM-based framework for an Optimal WSN Deployment in Smart Building. NOF 2020 : 110-114
- Security Implementation and Verification in Smart Buildings. CITSC 2019 : 51-56
- Security Implementation and Verification in Smart Buildings. CITSC 2019 : 51-56
- A Smart Living Framework : Towards Analyzing Security in Smart Rooms. MEDI 2019 : 206-215

Low Cost Security for IoT Systems

Low Cost Security for IoT Systems

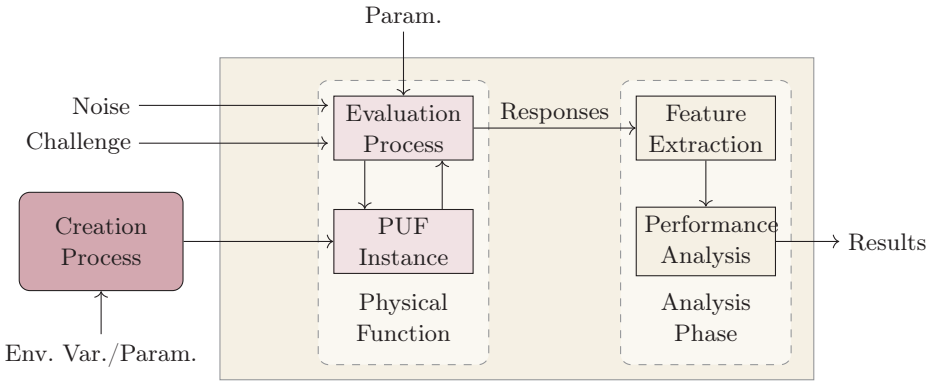
Objectives

Design and implement a robust and secure low cost protocol based on IoT chips randmoness.

Low Cost Security for IoT Systems

Objectives

Design and implement a robust and secure low cost protocol based on IoT chips randmoness.



Low Cost Security for IoT Systems

Low Cost Security for IoT Systems

Objectives

Design and implement one-way function based on IoT chips.

Methodology

- **Security test** : evaluating metrics, i.e. uniqueness, un-clonability, randomness, stability, evaluability, unpredictability, and one-wayness.
- **Key generation** : extracting keys from responses of the PUF by developing algorithms based on security sketches and fuzzy extractors.
- **Security gauging** : modeling and measuring the successfulness ratio of attacks : Brute-force, birth-day, replay, etc.
- **Application** : design an authentication and identification protocol based on the proposed PUF.

Low Cost Security for IoT Systems

Low Cost Security for IoT Systems

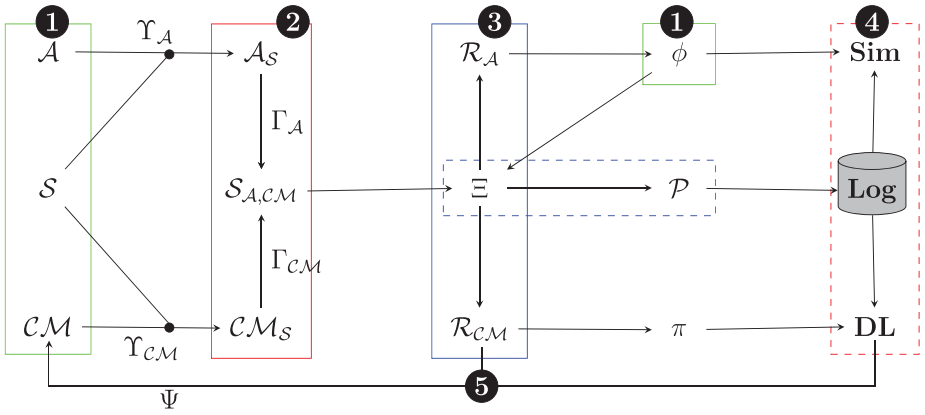
Objectives

Continuation : Integrating the protocol with the BC for Smart Transportation

References (F. Zerrouki, 5SJ+5C)

- Towards a Foundation of a Mutual Authentication Protocol for a Robust and Resilient PUF-Based Communication Network. FNC/MobiSPC 2021 : 215-222
- A Low-Cost Authentication Protocol Using Arbiter-PUF. MEDI 2021 : 101-116
- A Generation and Recovery Framework for Silicon PUFs Based Cryptographic Key. MEDI Workshops 2021 : 121-137
- Quantifying Security and Performance of Physical Unclonable Functions. IoTSMS 2020 : 1-4
- Towards an automatic evaluation of the performance of physical unclonable function. AIRES 2020 : 775-781

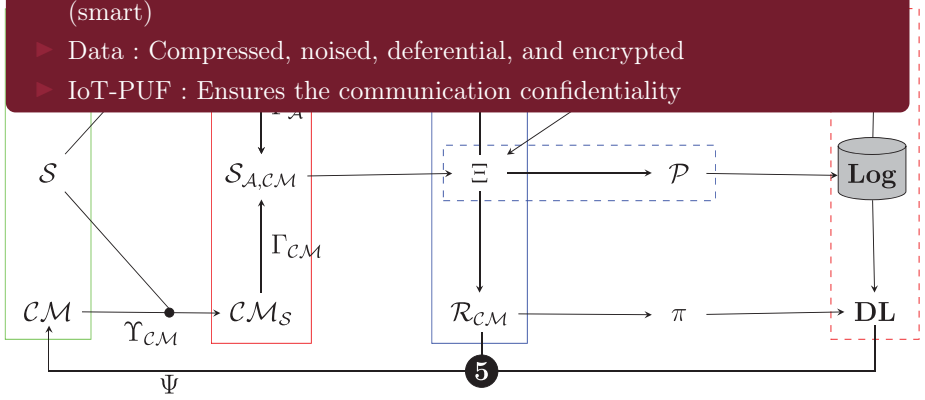
Security Assurance for Smart CPS



Research Directions | FA-SCPS

Security Assurance for Smart CPS

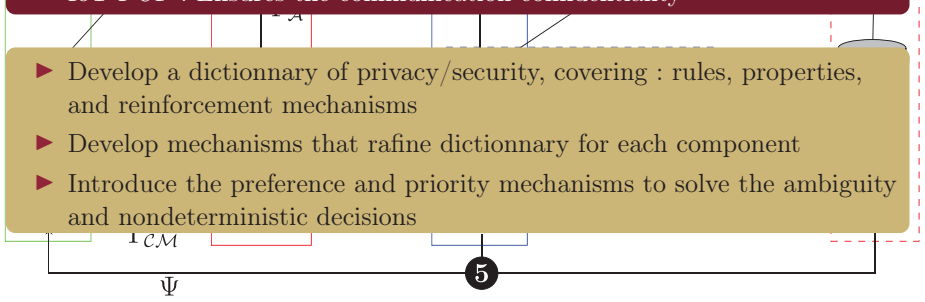
- System : Distributed, decentralized, blockchain, autonomuous nodes (smart)
- Data : Compressed, noised, deferential, and encrypted
- IoT-PUF : Ensures the communication confidentiality



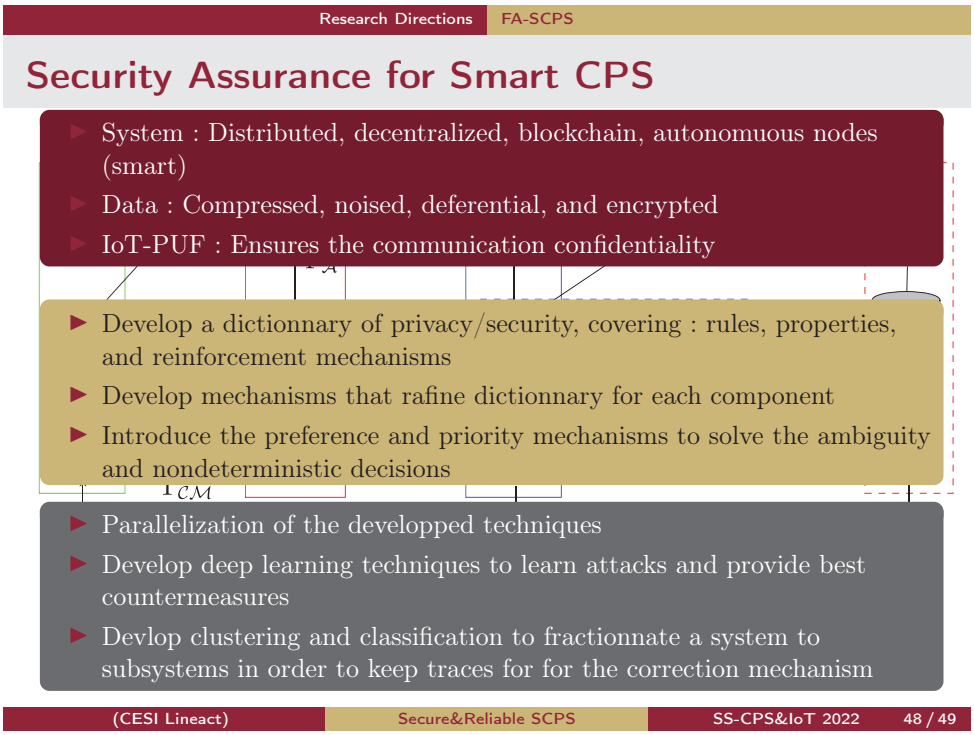
Research Directions | FA-SCPS

Security Assurance for Smart CPS

- System : Distributed, decentralized, blockchain, autonomuous nodes (smart)
- Data : Compressed, noised, deferential, and encrypted
- IoT-PUF : Ensures the communication confidentiality



- Develop a dictionnary of privacy/security, covering : rules, properties, and reinforcement mechanisms
- Develop mechanisms that rafine dictionnary for each component
- Introduce the preference and priority mechanisms to solve the ambiguity and nondeterministic decisions


Research Directions
FA-SCPS
Security Assurance for Smart CPS
System : Distributed, decentralized, blockchain, autonomuous nodes (smart)
Data : Compressed, noised, deferential, and encrypted
IoT-PUF : Ensures the communication confidentiality
Develop a dictionnary of privacy/security, covering : rules, properties, and reinforcement mechanisms
Develop mechanisms that rafine dictionnary for each component
Introduce the preference and priority mechanisms to solve the ambiguity and nondeterministic decisions
Parallelization of the developped techniques
Develop deep learning techniques to learn attacks and provide best countermeasures
Devlop clustering and classification to fractionnate a system to subsystems in order to keep traces for for the correction mechanism
(CESI Lineact)
Secure&Reliable SCPS
SS-CPS&IoT 2022
48 / 49

Thank you, Questions !

BRAIN-IoT
model-Based fRamework for dependable sensing and Actuation in INtelligent decentralized IoT systems

CPS&IoT'2022 Summer School

MACHINE LEARNING MEETS FORMAL METHODS

GENERATION AND VERIFICATION OF LEARNED STOCHASTIC AUTOMATA

Baouya Abdelhakim
VERIMAG (UGA)

Budva, Montenegro, 7-11 June, 2022

OUTLINE

- ML meets formal methods : Why and How
- BIP framework
- From data to automata : a Hybrid Approach
- SMC in a nutshell
- Conclusion

ML MEETS FORMAL METHODS : WHY AND HOW

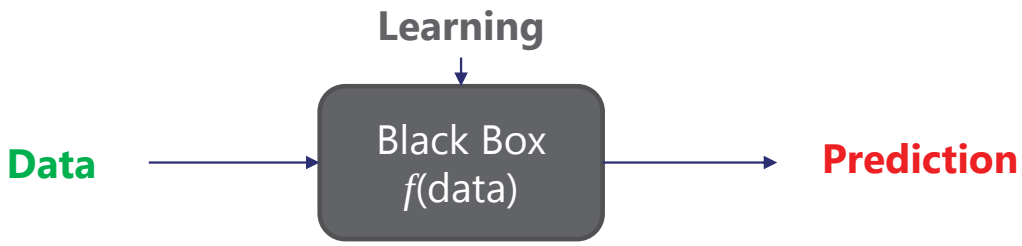
MACHINE LEARNING MEETS FORMAL METHODS BRAIN-IoT Generation and verification of learned stochastic automata

CONTEXT

Machine Learning techniques (ML): Ant Colonies, Neural Networks, Genetic Algorithms…..

ML techniques offer the possibility to make a prediction starting from historical data.

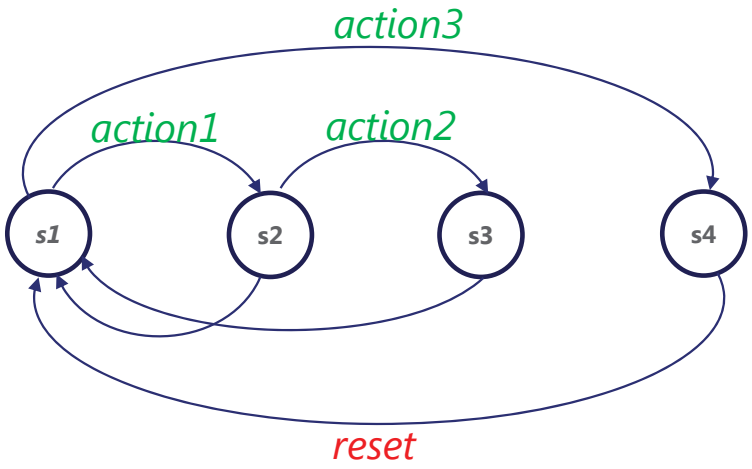
ML techniques allow **formally** reasoning if the learned model is amenable to verification.



BUILDING AUTOMATA-BASED MODEL

2002

Peled Doron et al,
Black box checking

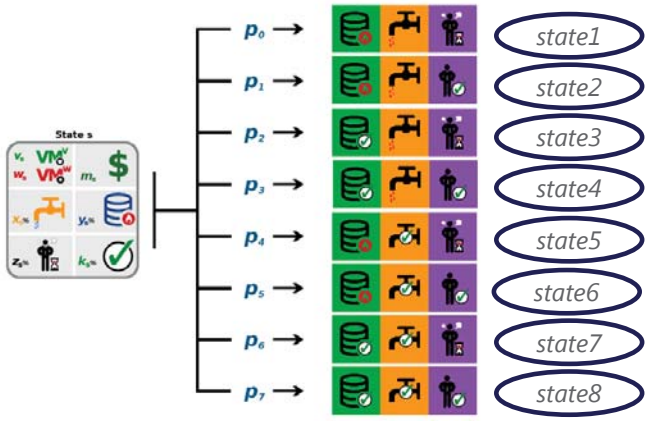


- Number of states N is ***learned***
- ***Simulate*** to find out the action types and states number

BUILDING AUTOMATA-BASED MODEL

2016

Naskos Athanasios et al,
Online Analysis of security risks in elastic cloud applications

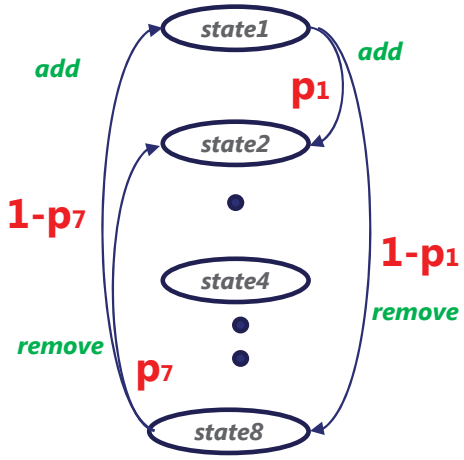


	data leakage		data loss
	no data leakage		no data loss
	performance violation	VM	VM type (t)
	no performance violation	$	Dep. cost

BUILDING AUTOMATA-BASED MODEL

2016

Naskos Athanasios et al,
Online Analysis of security risks in elastic cloud applications



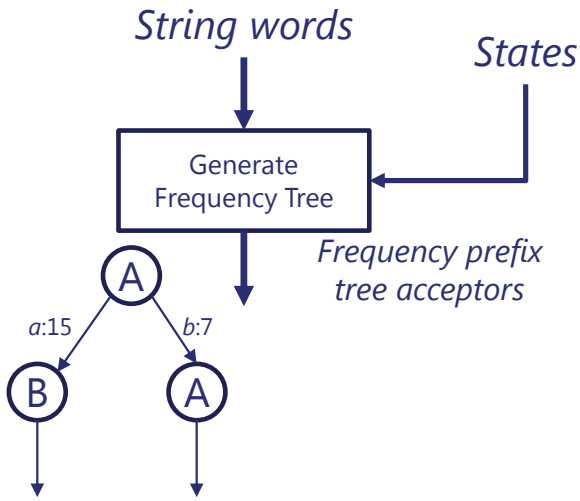
- Number of states N is ***semantically set as input***
- ***Probabilities*** *are computed and* ***Actions*** are recorded from past experiences

BUILDING AUTOMATA-BASED MODEL

2016

Mao H., Larsen K.

Learning deterministic probabilistic automata from a model checking perspective

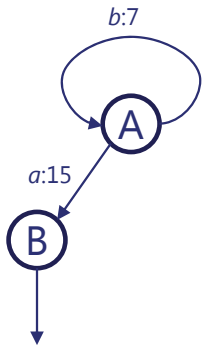


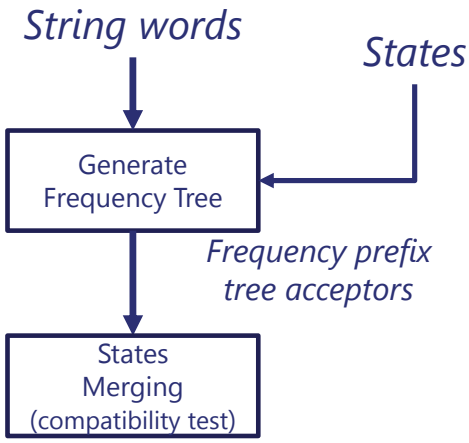
BUILDING AUTOMATA-BASED MODEL

2016

Mao H., Larsen K.

Learning deterministic probabilistic automata from a model checking perspective



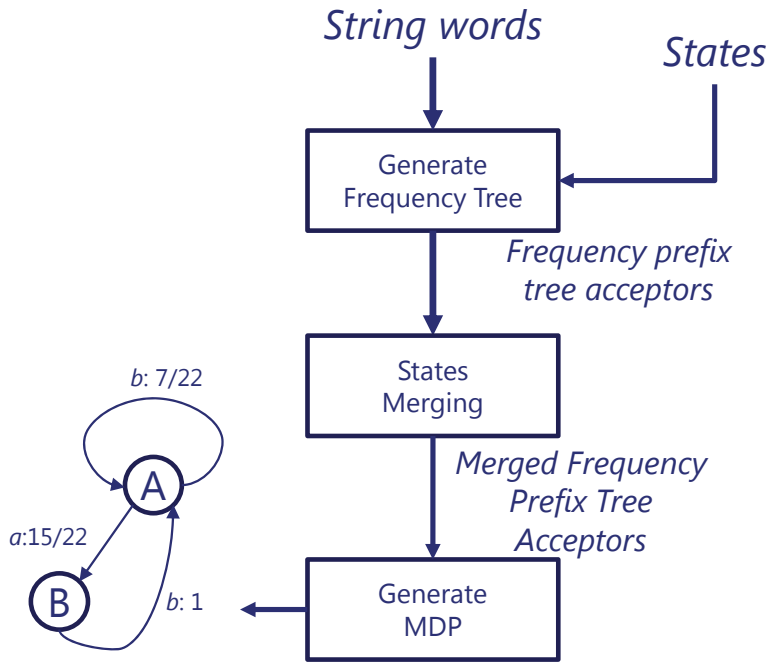


BUILDING AUTOMATA-BASED MODEL

2016

Mao H., Larsen K.

Learning deterministic probabilistic automata from a model checking perspective

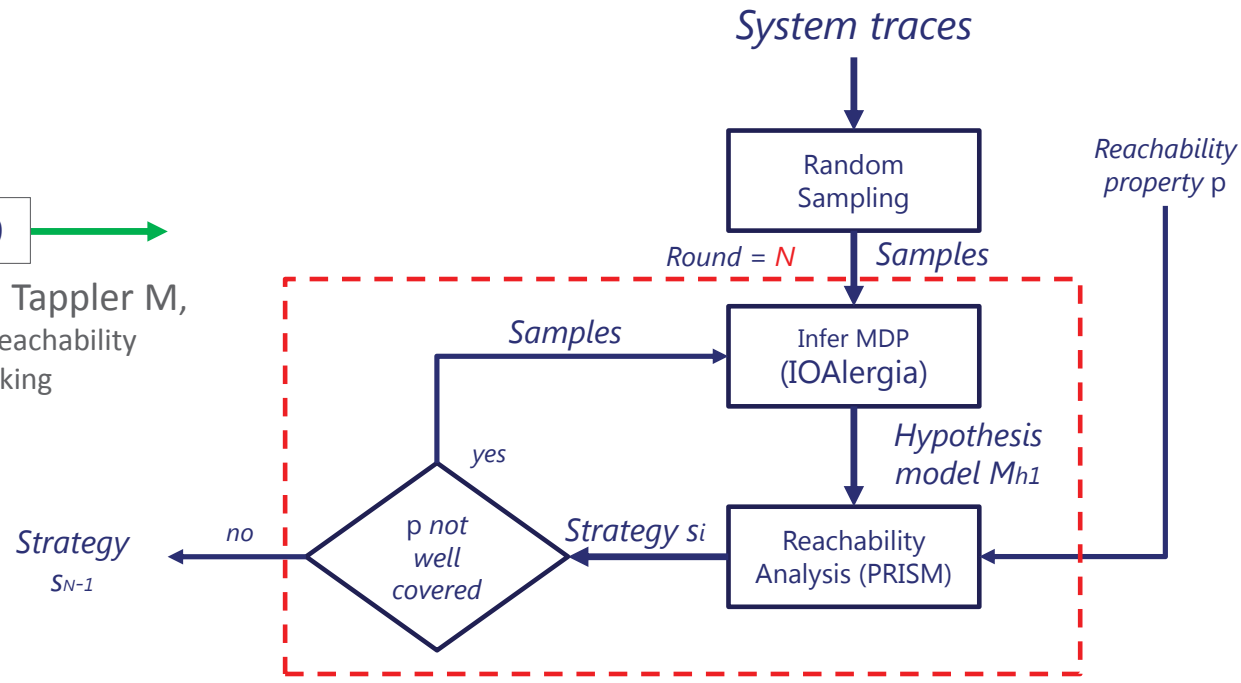


- Number of MDP states is ***the identified String words.***

BUILDING AUTOMATA-BASED MODEL

2019

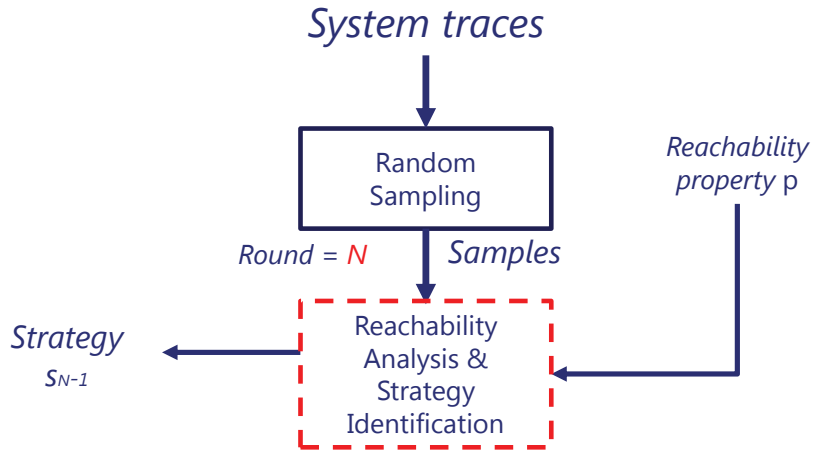
Aichernig B, Tappler M,
Black-box reachability checking



BUILDING AUTOMATA-BASED MODEL

2019

Aichernig B, Tappler M,
Black-box reachability checking



- ***Iterative approach until the relevant strategy is detected.***
- ***MC** is applied to* estimate the probability of reaching p.

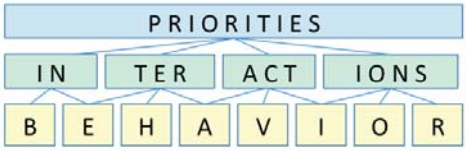
BIP FRAMEWORK

MACHINE LEARNING MEETS FORMAL METHODS BRAIN-IoT Generation and verification of learned stochastic automata

BIP FRAMEWORK = LANGUAGE + CODE GENERATION

Components = Layered composition of :

- **B**ehavior: atomic function unit (automata+ actions code).
- **I**nteractions: corporation between actions and behavior.
- **P**riorities : conflict resolution between interactions.

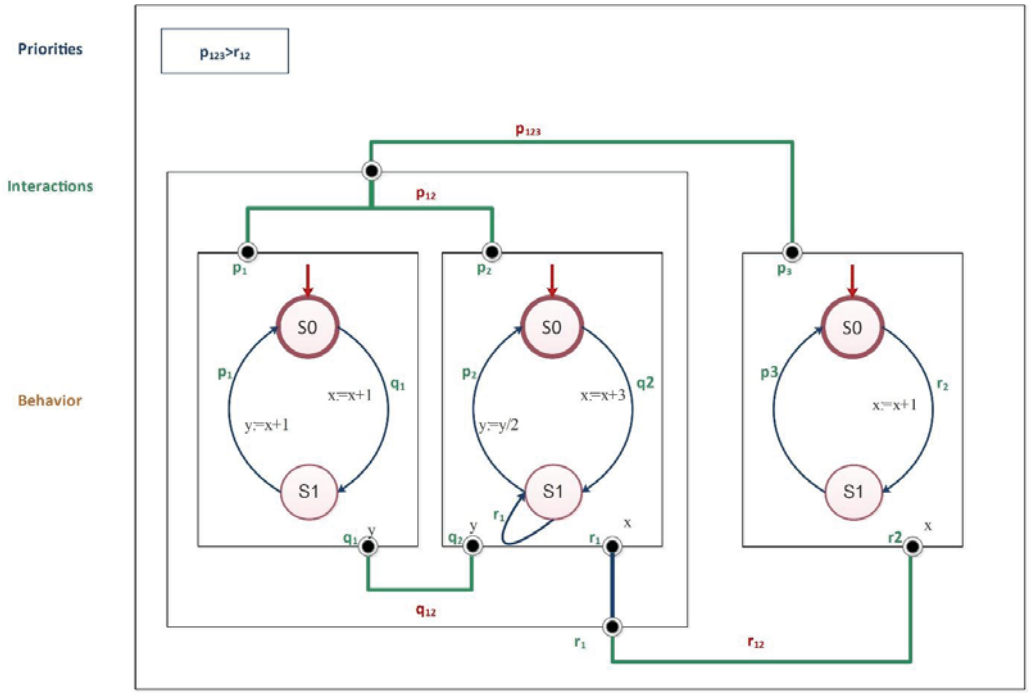


Model-based and component-based design:

- Rigorous operational semantics.
- Providing automated support for validation and performance analysis.
- Providing automated support for implementation on given platforms.

Stochastic BIP: Extends the BIP models with probabilistic variables.

BIP MODELS IN A NUTSHELL



FROM LTS TO ATOMIC COMPONENT

Definition 1: LTS *is a tuple* ⟨**Q**, **Act**, **→**, q_0⟩*:*

- **Q** *is a set of states,*
- **Act** *is a set of action names labelling the transitions,*
- →⊂ Q × Act × Q *is a set of labelled transitions,*
- q_0 ⊆ Q *is the initial state.*

Definition 2: *An atomic component* B=⟨**S**, **P**, **T**, s_0, $\mathcal{V}$⟩ *is an LTS extended with variables:*

- $\mathcal{V}$ = $\{v_0, \ldots, v_n\}$ *is a set of local variables,*
- ⟨**S**, **P**, **T**, s_0⟩ *is a labelled transition system,* S *is a set of states,* P *a set of communication ports,* T *is a set of transitions of the form (*s, p, g, f, s′*) where* s, s′ ∈ S, p ∈ P, g ∈ Eval($\mathcal{V}$) *is a guard, and* f ∈ Func($\mathcal{V}$) *is an update function on a subset of* $\mathcal{V}$*, and*
- s_0 ∈ S *is the initial state.*

FROM LTS TO ATOMIC COMPONENT

Let $\mathbb{D}$ be a finite universal domain. Given a set of variables $\mathcal{V}$, we define valuations for variables as a set of functions X : $\mathcal{V} \rightarrow \mathbb{D}$ that associate each variable in $\mathcal{V}$ with a value in $\mathbb{D}$,

Definition 3: [semantics]*The atomic component* **B** = ⟨**S**, **P**, **T**, s_0, $\mathcal{V}$⟩ *is an LTS* ⟨**Q**, **Act**, →, q_0⟩, *such that:*

- **Q** ⊆ S × $\mathbb{D}$,
- **Act** ⊆ P *is a set of transitions labels,*
- → *is a set including transitions of the form* ((s;X), p, (s';X')) *written such that,*
  var(p) ∈ X, g(X) *evaluates to true and* X' *is the new valuation,*
- q_0 = (s_0; X_0) ⊆ Q.

BIP OPERATIONAL SEMANTICS

ENB-UPD 1:

$$\frac{p\in P,\ var(p)\in X,\ Dom(X)=Dom(X'),X\neq X',\ [\![g(X)]\!]=\top}{(s;X)\xrightarrow{p}(s';X')}$$

ENB-UPD 2:

$$\frac{p_1,p_2\in P,var(p_1)\in X_1,var(p_2)\in X_2,Prior(p_1)\geq Prior(p_2),\ [\![g(X_1)]\!]=\top,[\![g(X_2)]\!]=\top}{(s_1:X_1)\xrightarrow{p_1}(s_1':X_1')}$$

SYNC:

$$\frac{p\in P,\ var(p)\in X,\ Dom(X)=Dom(X'),\ B_1=(s_1:X_1)\xrightarrow{p}(s_1':X_1'),B_2=(s_2:X_2)\xrightarrow{p}(s_2':X_2'),[\![g(X_1)]\!]=\top,[\![g(X_2)]\!]=\top}{((s_1,s_2):X_1\cup X_2)\xrightarrow{p}((s_1',s_2'):X_1'\cup X_2')}$$

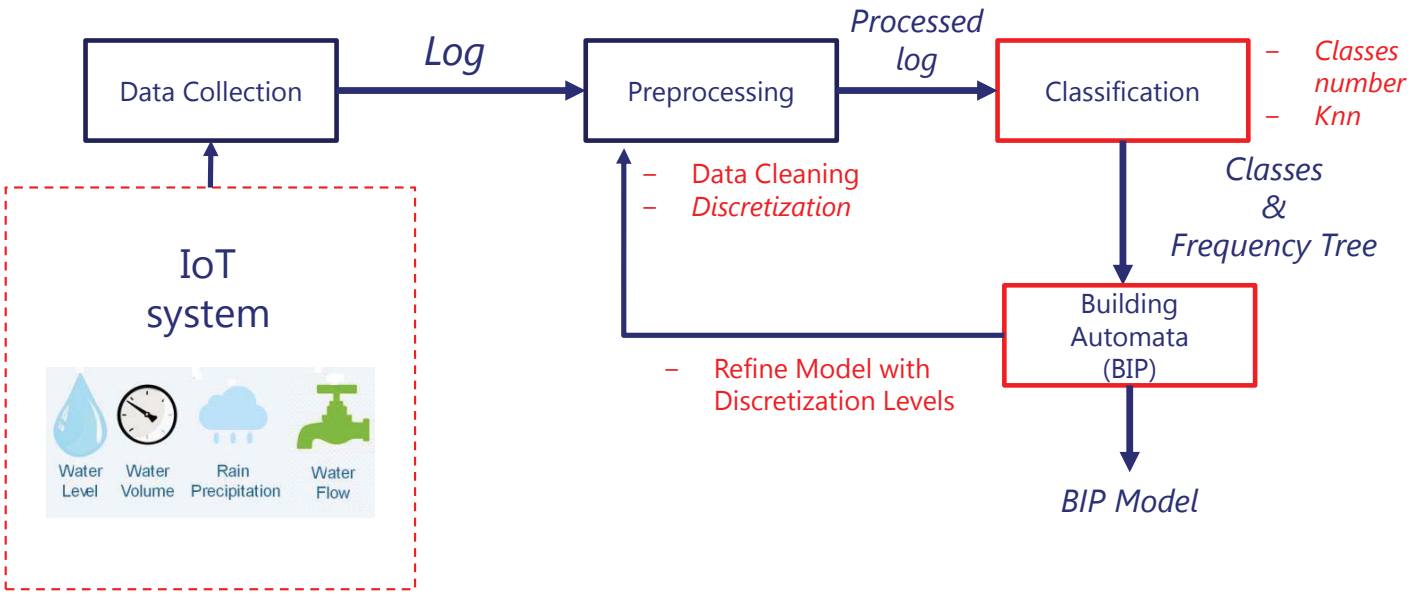
PRB-UPD:

$$\frac{p\in P,\ var(p)\in X,\ Dom(X^p)=Dom(X^{p'}),\ Dom(X^d)=Dom(X^{d'}),\ [\![g(X^d)]\!]=\top}{(s:X^p\cup X^d)\xrightarrow{p}(s':X^{p'}\cup X^{d'})}$$

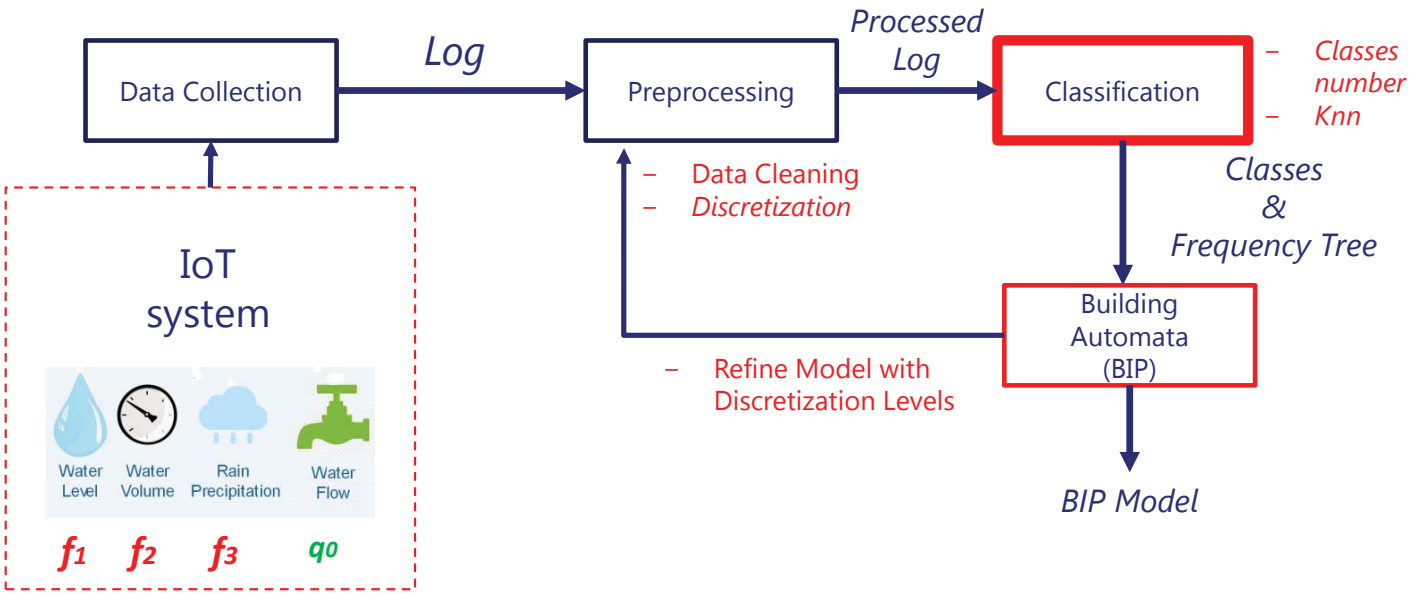
FROM DATA TO AUTOMATA : A HYBRID APPROACH

MACHINE LEARNING MEETS FORMAL METHODS BRAIN-IoT Generation and verification of learned stochastic automata

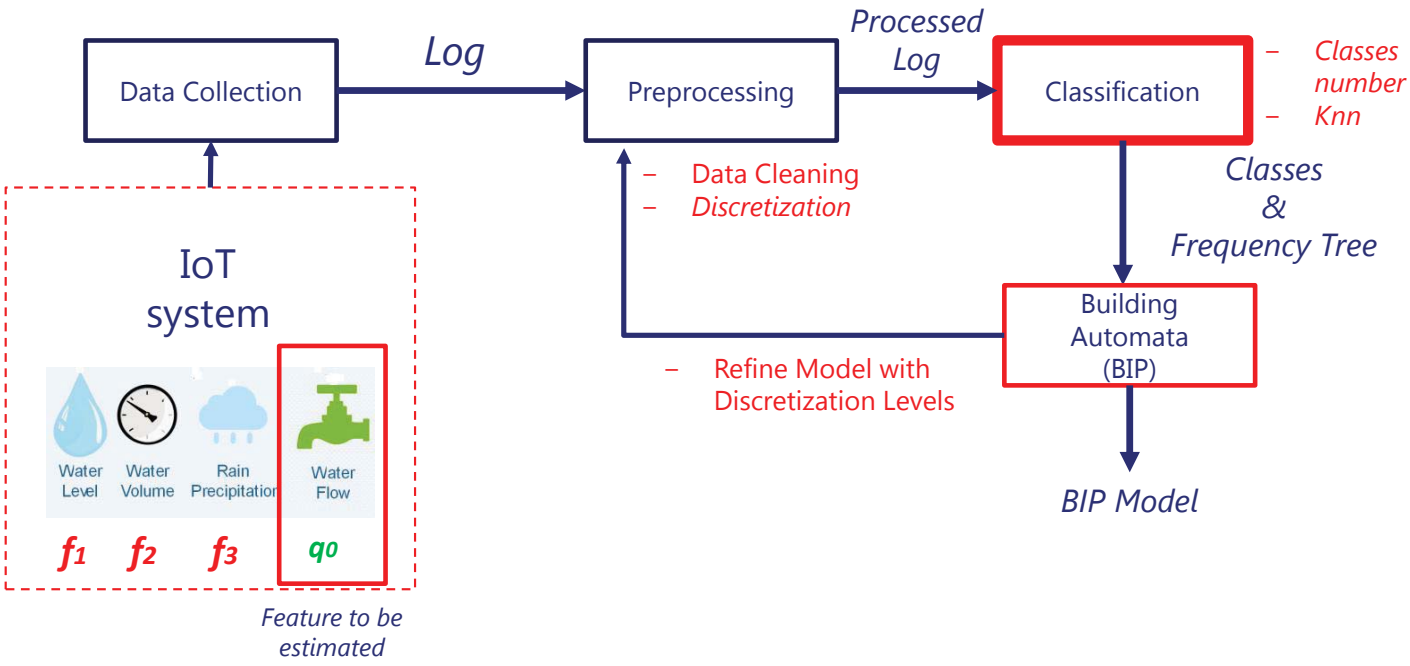
HYBRID APPROACH



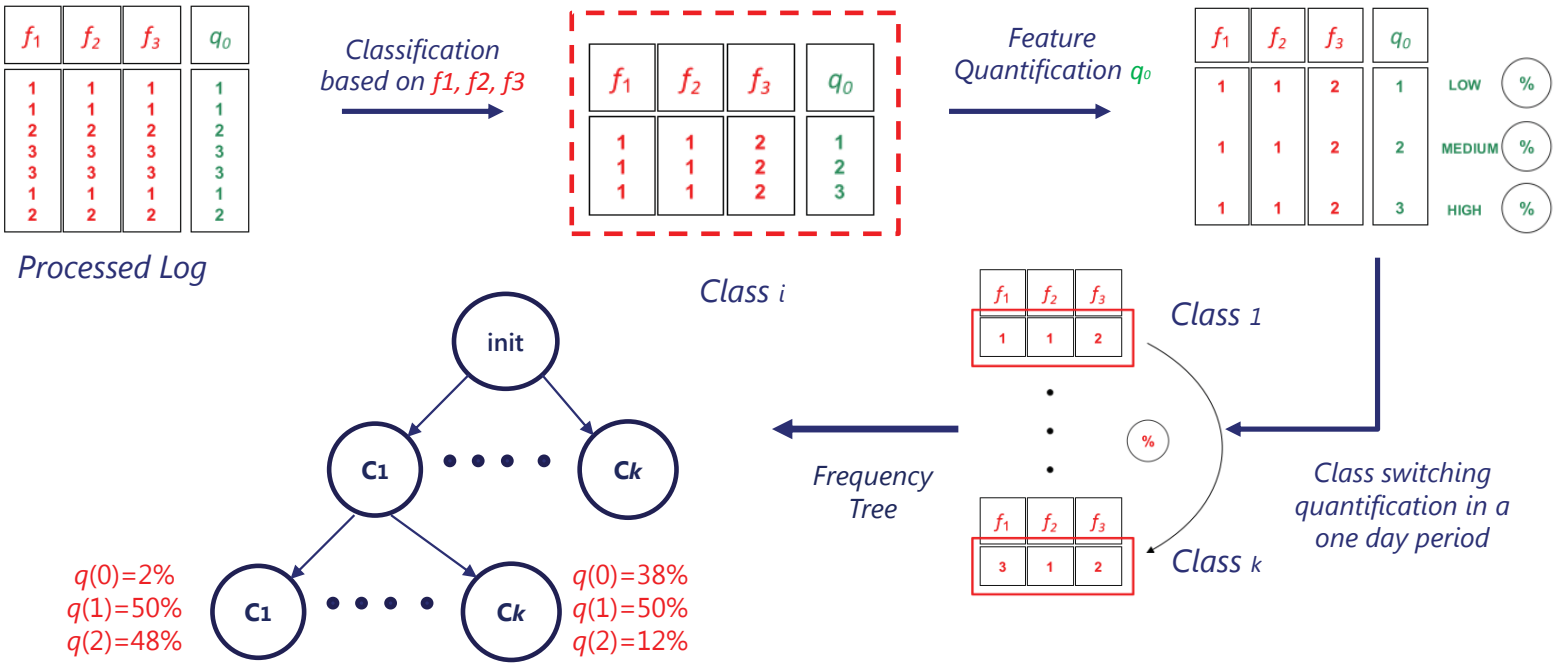
HYBRID APPROACH



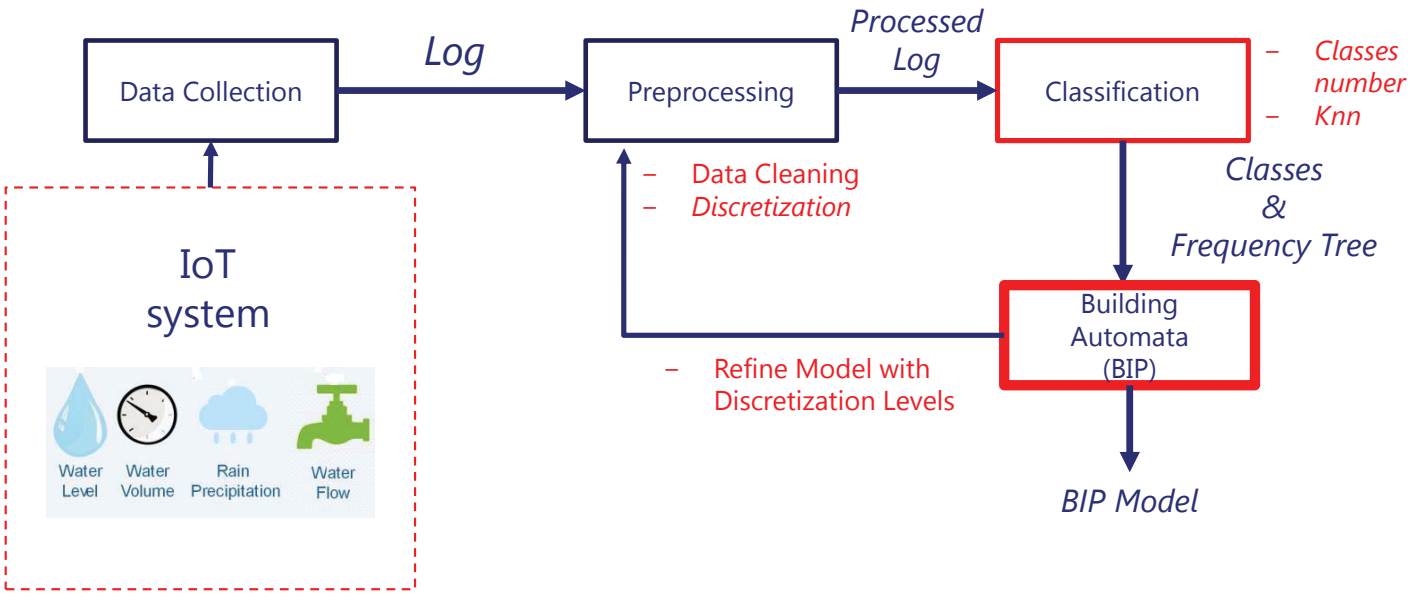
HYBRID APPROACH



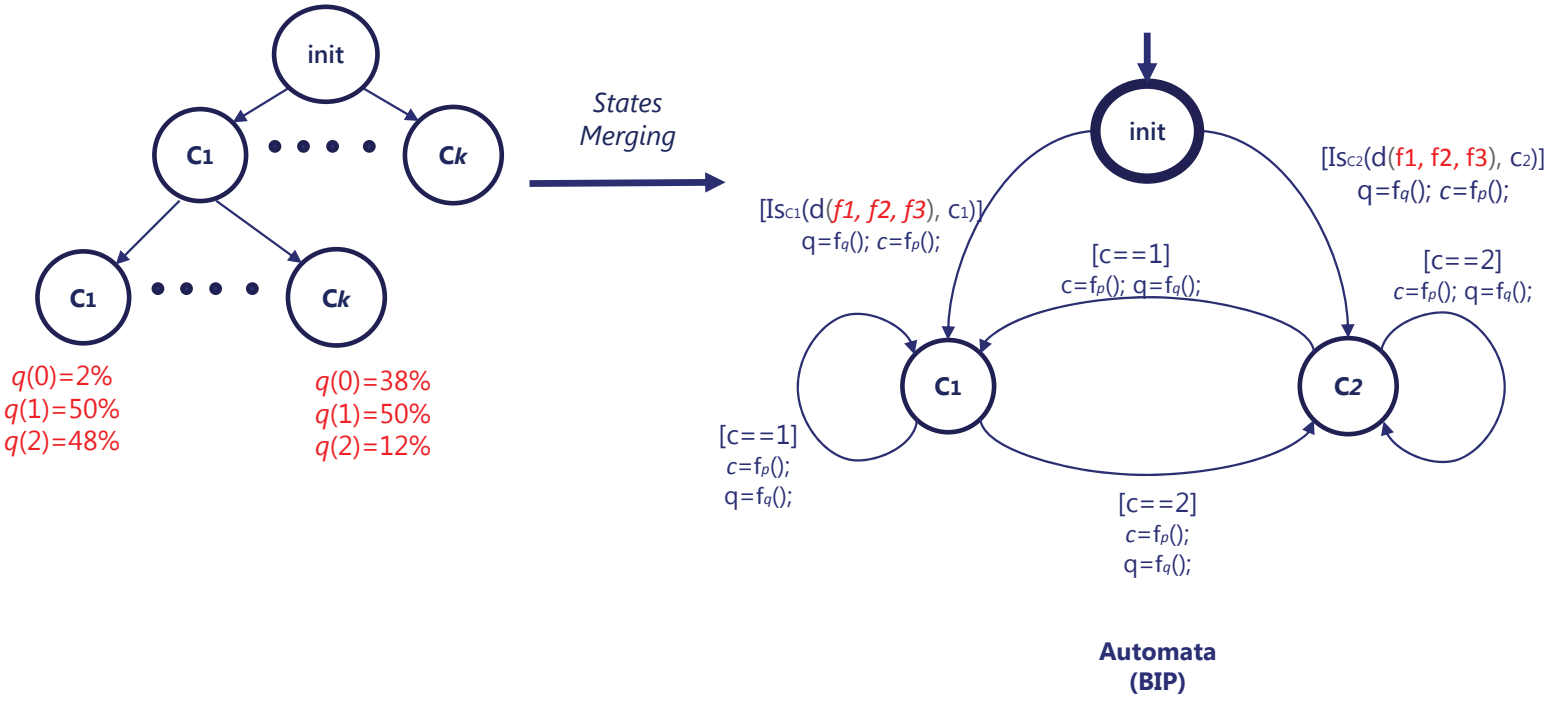
CLASSIFICATION



HYBRID APPROACH



BUILDING AUTOMATA

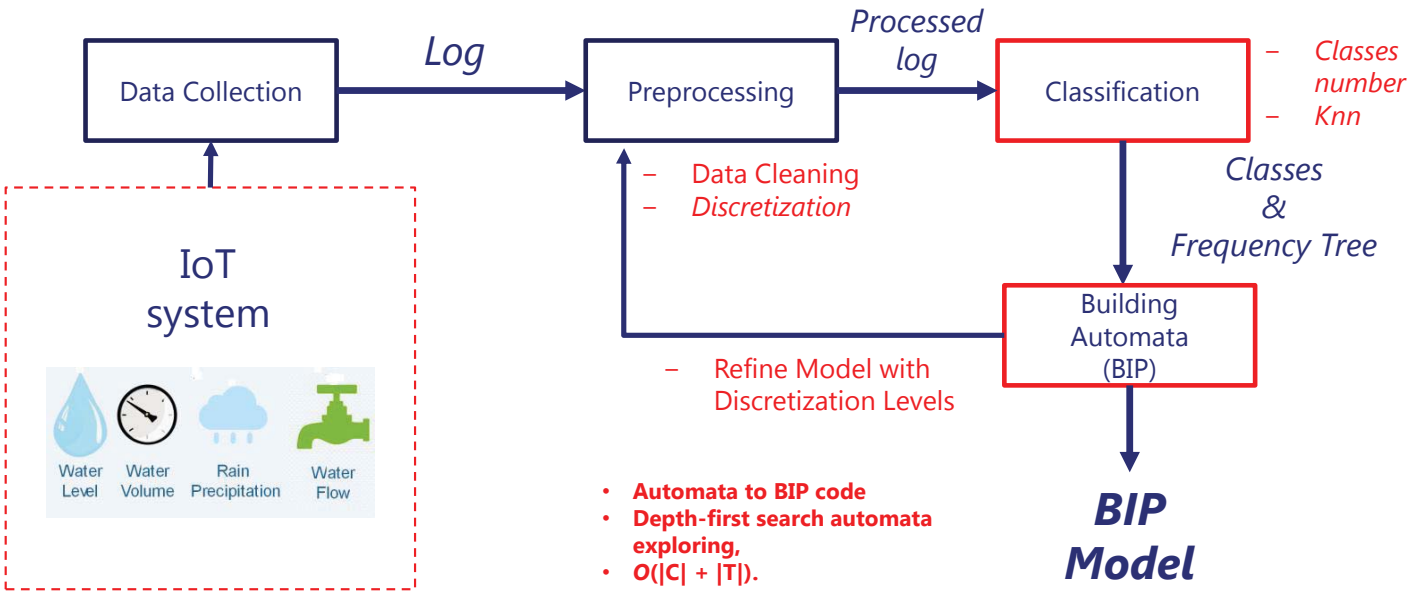


BUILDING AUTOMATA

Definition 4: [Learned Stochastic Component]. *A learned stochastic component is an atomic component extended with probabilistic variables* ***B***$= \langle$ ***C***, ***P***, ***T***, ***i***, $\mathcal{V} \rangle$ *, where:*

- ***C*** $= \{i\} \cup \{c_0, \ldots, c_k\}$ is a set of states where i is the initial state,
- ***P*** *is a set of ports,*
- $\mathcal{V} = \mathcal{V}^d \cup \mathcal{V}^q \cup \mathcal{V}^p$, where $\mathcal{V}^q$ and $\mathcal{V}^d$ is a set of deterministic variables and $\mathcal{V}^p$ are probabilistic variables,
- ***T*** *is a set of transitions of the form* $t = (c, prt, f, c')$ *where* $c, c' \in C$, $prt \in P$, g *is a guard over Eval*$(\mathcal{V})$*, and* f *is a pair* (f_q, f_p) *where* f_q *and* f_p *are deterministic update functions on features occurrence* $\mathcal{V}^q$ *and classes switching* $\mathcal{V}^p$*,*
- ***i*** is the initial state.

HYBRID APPROACH



BUILDING AUTOMATA-BASED MODEL

Model Checking	Pure Simulation	Statistical Model Checking
• Exhaustive • Guarantees • State space explosion	• Partial • No guarantees • Fast	• Partial • Bounded error • Fast

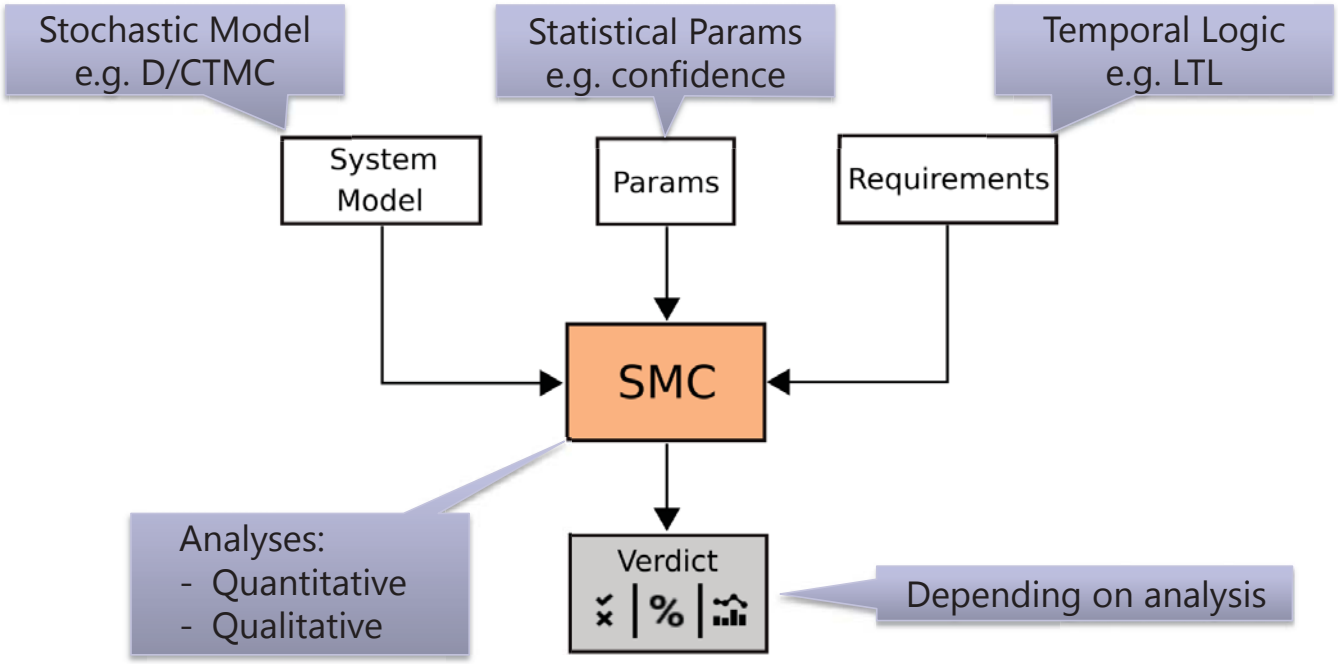
✓ **SMC** is a tradeoff between analysis **speed** and result **guaranties**.

SMC IN A NUTSHELL

INTRODUCTION TO SMC

MACHINE LEARNING MEETS FORMAL METHODS BRAIN-IoT Generation and verification of learned stochastic automata

SMC SETTING

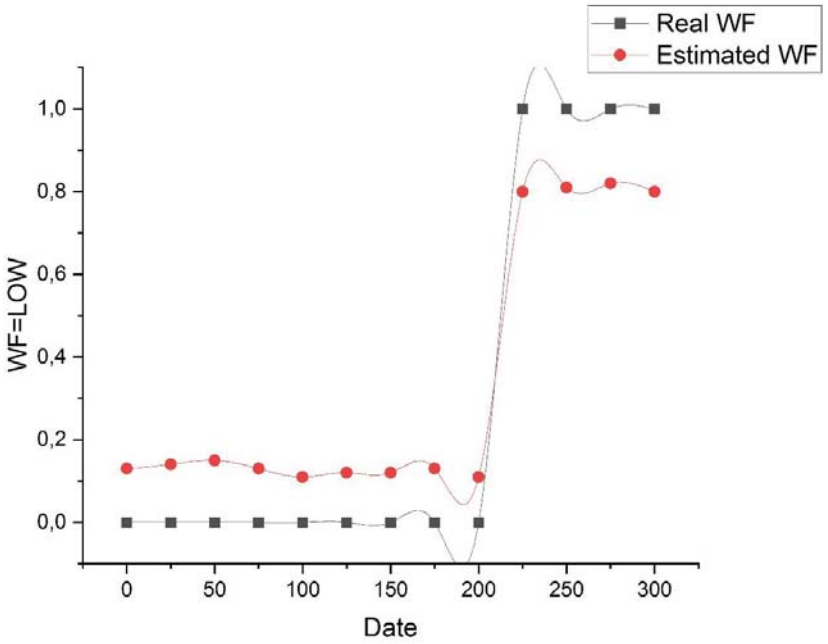

Stochastic Model
e.g. D/CTMC
Statistical Params
e.g. confidence
Temporal Logic
e.g. LTL
System Model
Params
Requirements
SMC
Analyses:
- Quantitative
- Qualitative
Verdict
Depending on analysis

VERIFICATION

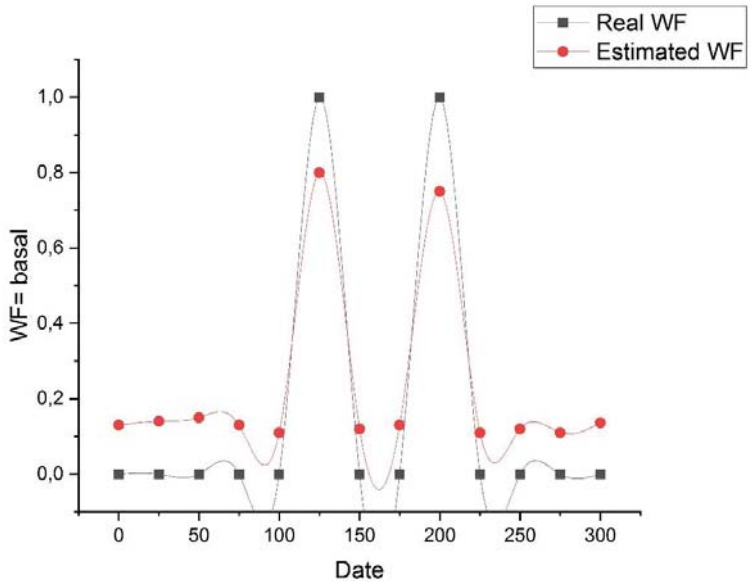
- Property : What is the probability of experiencing variation on water flow for $low, basal, high$ levels ?

$$P_{=?} = [F^{1000} f_{wf} = v_{wf} \;\&\&\; days \leq T] \text{ ; T=1 : 274 : 7;}$$
$$v_{wf} \in \{low, basal, high\}$$

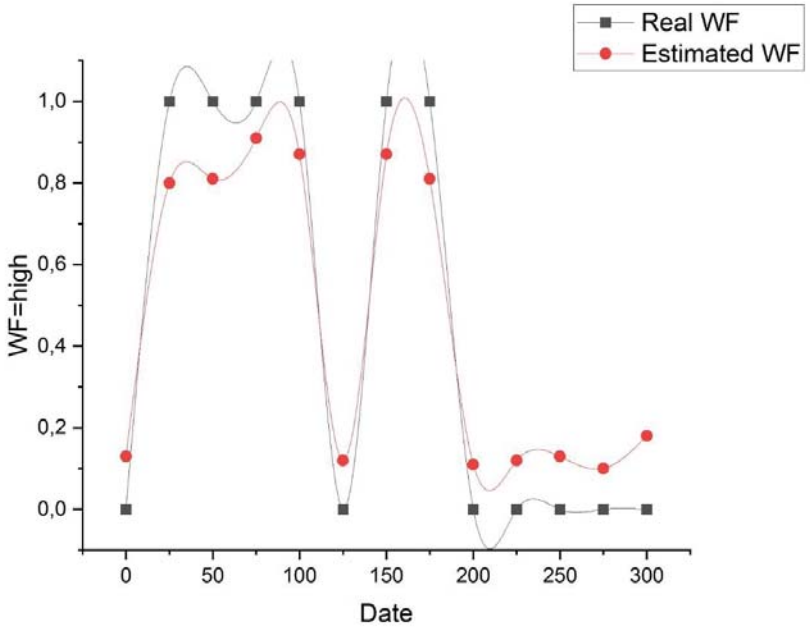
$$v_{wf} = low$$


Real WF
Estimated WF
WF=LOW
1,0
0,8
0,6
0,4
0,2
0,0
0
50
100
150
200
250
300
Date

$$v_{wf} = basal$$


Real WF
Estimated WF
WF= basal
1,0
0,8
0,6
0,4
0,2
0,0
0
50
100
150
200
250
300
Date

$$v_{wf} = high$$


Real WF
Estimated WF
WF=high
1,0
0,8
0,6
0,4
0,2
0,0
0
50
100
150
200
250
300
Date

BIP: ONLINE RESOURCES

- **Available online!** Link to web page:
  - http://www-verimag.imag.fr/BIP-SMC-A-Statistical-Model-Checking.html
- **Open source!** Link to Git repository:
  - https://gricad-gitlab.univ-grenoble-alpes.fr/verimag/bip/sbip2
- Distributed with **user manual** and **video tutorial!**
  - Link to user manual: http://www-verimag.imag.fr/www-verimag.imag.fr/~nouri/bip-smc/sbip-2.2.1-user-manual.pdf
  - Link to video tutorial: https://www.youtube.com/watch?v=MvNfZrvlVAs

CONCLUSION

- Hybrid approach to build and check BIP models using *k*-NN an BIP SMC,
- Data correlation is performed to check the veracity of the estimated values.
- → Accuracy issue will be resolved by transformation soundness,
- → Large data set need to be processed to check the feasibility of the approach,
- →Building security patterns from learned models to check the quality of the captured data.

CONTACT

BAOUYA ABDELHAKIM
RESEARCH ASSOCIATE
UNIVERSITY GRENOBLE ALPES
abdelhakim.baouya@univ-grenoble-alpes.fr

EU Projects : Brain-IoT, CPS4EU and, FOCETA

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 780089.

BRAIN-IoT
model-Based fRamework for dependable sensing and Actuation in INtelligent decentralized IoT systems

SELFSUSTAINED CROSS-BORDER CUSTOMIZED
CYBERPHYSICAL SYSTEM EXPERIMENTS
FOR CAPACITY BUILDING AMONG EUROPEAN STAKEHOLDERS

Principles of performance effective node design for smart systems

Prof. dr Radovan Stojanović
University of Montenegro and MECOnet

SS-CPSIoT2022, Budva, Montenegro

Co-funded by the Horizon 2020 programme of the European Union
DT-ICT-01-2019
Smart Anything Everywhere Area 2
www.smart4all-project.eu
Grant Agreement: 872614

Content

- Instead introduction
- Design considerations
- Examples (hardware and software)
- Conclusion

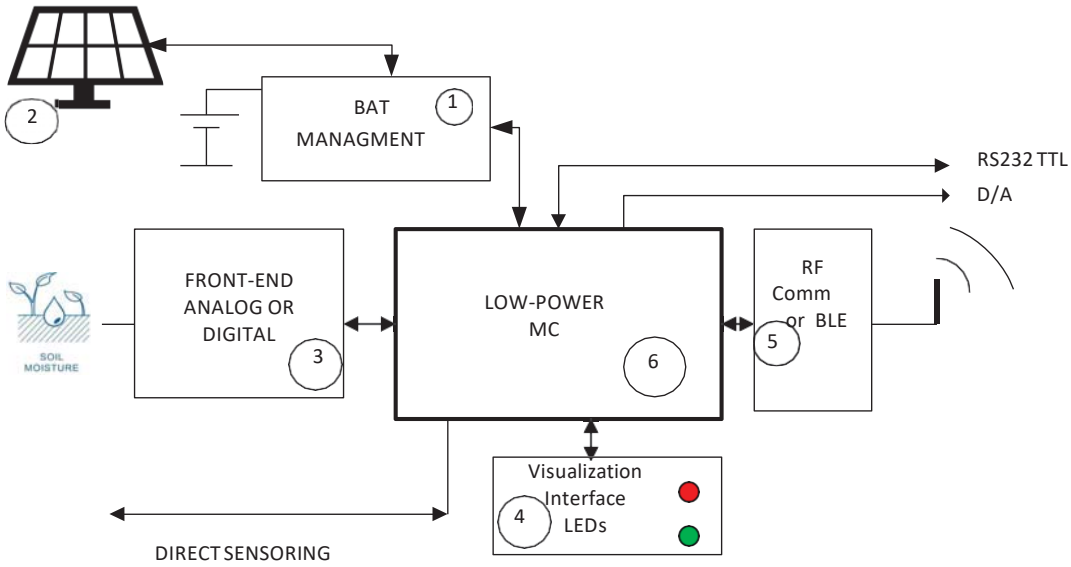
Instead introduction

- Resolve the dilemma. Do You do this out of hobby or business?
- A hobby is a useful thing for a person. Sometimes can bring a grant (project). Someone can develop a business based on it.
- Business philosophy is different? How can I sort out the problem in the simplest and most economical way?
- **"Sometimes when you turn a hobby into a job, it becomes work".** Jeff Bennett
- "Whatever you like to do, make it a hobby and whatever the world likes to do, make it a business", Warren Buffett

"Shooting sparrows with cannons"

Design considerations

- Smart node architecture (general)



www.smart4all-project.eu

- Each system component is separate design story, should be cost and performance effective and carefully designed.

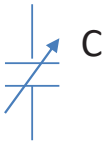
Design considerations

- **(1) and (2). Powering**. Standard powering, battery, rechargeable battery, solar, wind, earth, low-dropout regulator (**LDO regulator**), **DC-DC regulator** , linear or switching, **quiescent current**, **current capacity**, **low-power, low-cost**…etc are parameters to be considered. As example, The MCU may have low power consumption, but if the regulator has significant quiescent current, then the whole system will not be too power effective.
- Sometimes we can take energy from nothing, sun, wind movement, even from earth. Example. Simplest smart system for soil moisture measurement**.** No battery, 3 in 1. Analog metering. Use a soil as battery and analog circuits as an actuator.

www.smart4all-project.eu

Design considerations

- (3) **Analog front-end.** Convert non-electric value to its digital equivalent to be readable with MC. It can be done in several ways.
- Example, soil moisture measurement. Convert soil moisture in C or R, voltage, current, time, frequency etc...

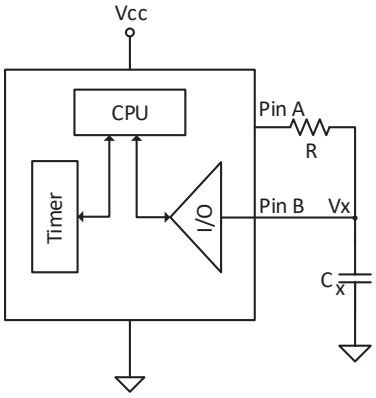


C

www.smart4all-project.eu

Design considerations

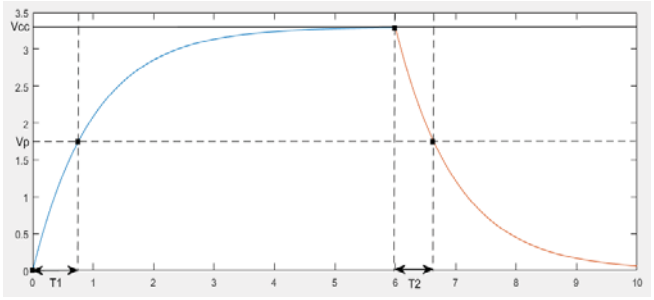
- **Analog-front-end**, how to convert an analog signal to its digital equivalent?. **Decay time**. Passive circuits, active circuits, complex, simple, direct interfacing to MCU. This part need analog knowledge and experience. Sometimes, MCU have built-in analog active components as operational amplifiers and analog comparators, that can be very useful.
- **Direct interfacing.** RC principle, charging discharging. Resistor-Capacitor (RC) decay circuit. The time it takes for the voltage to decay to a threshold (Vth) is a function of the capacitance Cx. Good resolution. No need active components and A/D converter. In the simplest algorithms T1 and T2 sensitive on R, Cx, Vth and vice-versa. **But, there is a way to make Cx non-dependent on them**.



$$Cx = -\frac{1}{R}\frac{B + \sqrt{B^2 - 4AC}}{2A}$$

$$A = -[2 + \ln(2)]\ln(2)$$

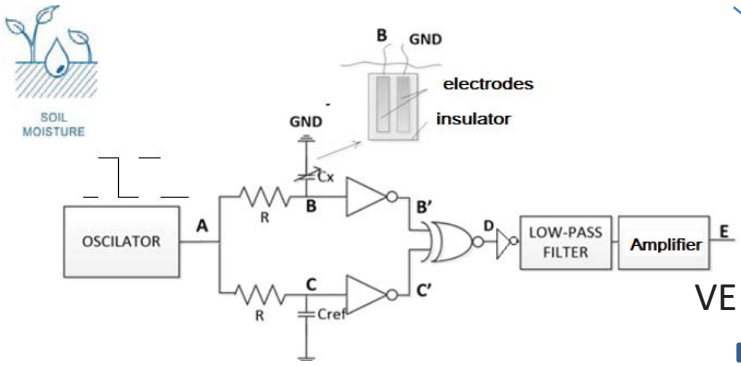
$$C = -\frac{{T_1}^2 + {T_2}^2}{2}$$

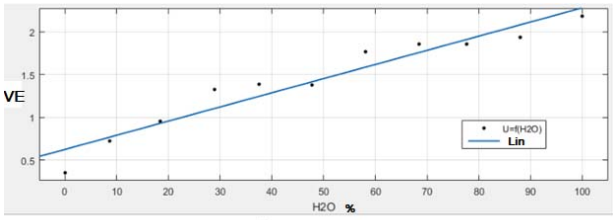


www.smart4all-project.eu

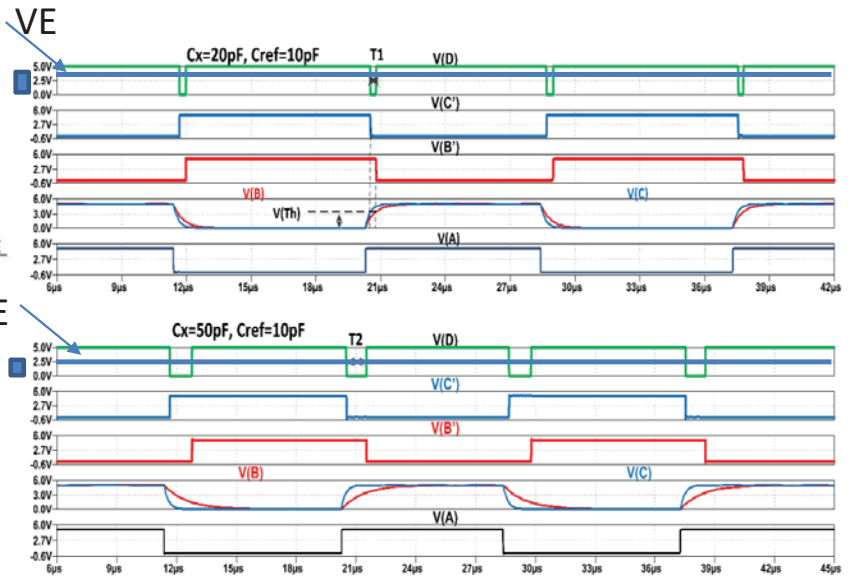
Design considerations

- **Analog front end. Difference of delays (times) principle**. Converting difference of delays in two branches R-C and R-Cx to PWM signal and then to voltage.





$f(x) = 0{,}5465 * x + 1{,}426;\ SKG = 0{,}1572$



www.smart4all-project.eu

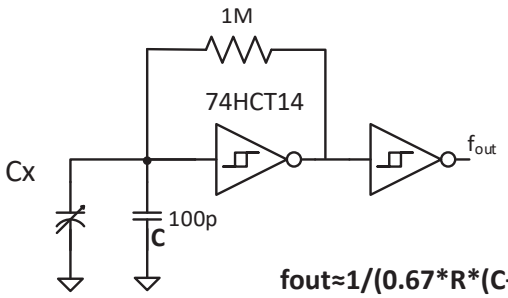
Design considerations

- Code for soil moisture reading, as voltage equivalent (VE from above circuit)

```
#include <SoftwareSerial.h> // Software Serial Communication
#include <tinysnore.h>     // Sleeping library for ATtiny

#define SENSOR_pin 1
#define POWER_pin 0
#define RX_pin 4
#define TX_pin 3

SoftwareSerial SwSerial(RX_pin,TX_pin);
int readA = 0;
float moisture;
byte i;

void setup(){
  pinMode(TX_pin, OUTPUT);
  pinMode(RX_pin, INPUT);
  pinMode(SENSOR_pin, INPUT);
  pinMode(POWER_pin, OUTPUT);
  SwSerial.begin(9600);
  delay(1000);
} // end setup

void loop(){
  digitalWrite(POWER_pin, HIGH);  // Power sensor on
  delay(5000);   // Wait

  readA = 0;
  for(i=0; i<5; i++)
                    {                    readA += analogRead(SENSOR_pin);
                                         delay(20); }
  digitalWrite(POWER_pin, LOW);  // Power sensor off
  readA = readA / 5; // Average from 5 measurement

  moisture = (float) readA * 3.3 / 1023.0;  // Convert to voltage scale
  moisture = (moisture - 0.27)/0.0185;  // subtract voltage offset and convert voltage
// to moisture, by 1st order curve linearisation
  if(moisture > 100) moisture = 100; // Out of Limits
  if(moisture < 0) moisture = 0;         //
  SwSerial.println(moisture);
  snore(5000); // Go to deep sleep
} // end loop
```

Design considerations

- **Analog front end.** Frequency based measurement, Cx change with moisture, i.e. fout by Cx. One Schmitt's trigger 74HCT14 can support 6 sensors.

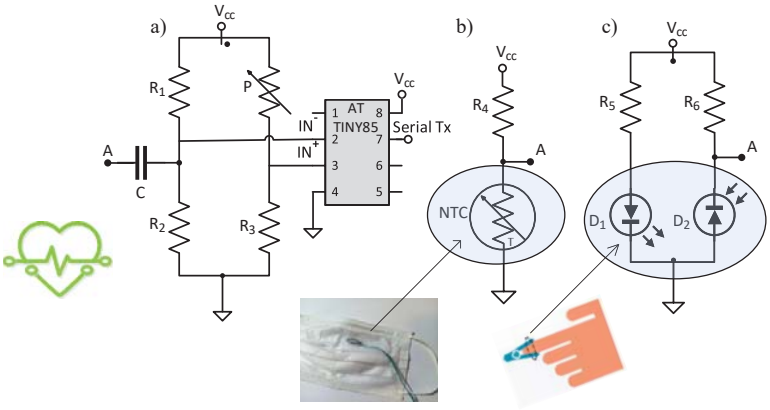


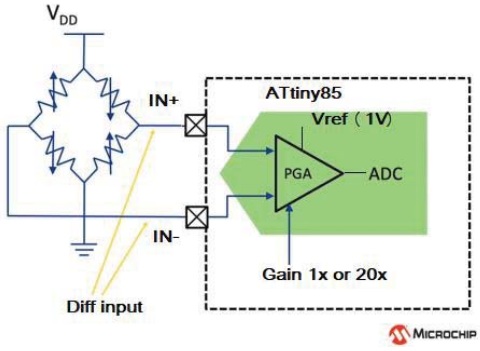
fout≈1/(0.67*R*(C+Cx))

F [kHz] \ C [pF]	22pF	100pF
f_{out} - air	44	14
f_{out} - water	20	10,3

Design considerations

Analog front-end. R and Photo Current (Voltage) Conversion. Converting NTC resistor value or photofluctations (photo voltage or photo current) to voltage, or differential voltage, VIN+-VIN-. Examples. Detection Respiration Rate (RR) and Heart Rate (HR) by simple analog front-end, directly connected to MCU. By using internal differential amplifiers of MCU, lowering ADC Vref=1.024V, selecting PGA=20, we can achieve resolution of 1mV/20=50uV that enough even to tie directly thermocouple.

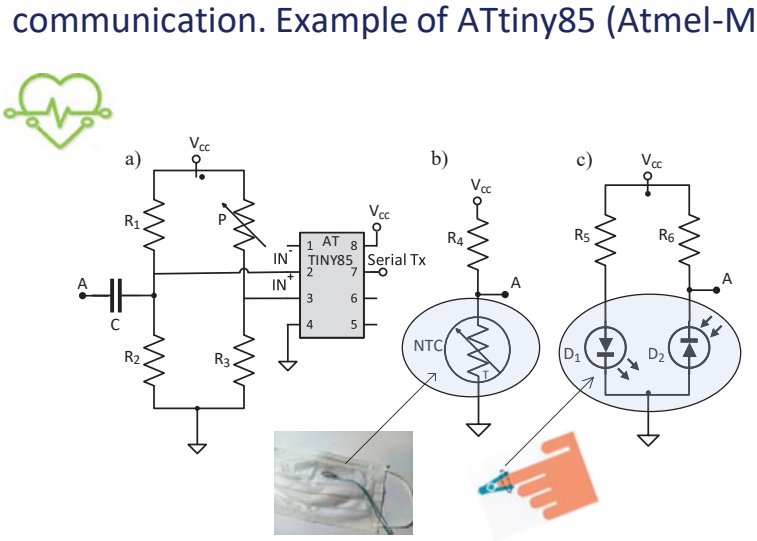


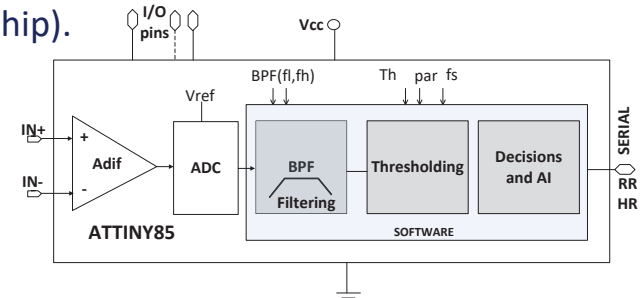


www.smart4all-project.eu

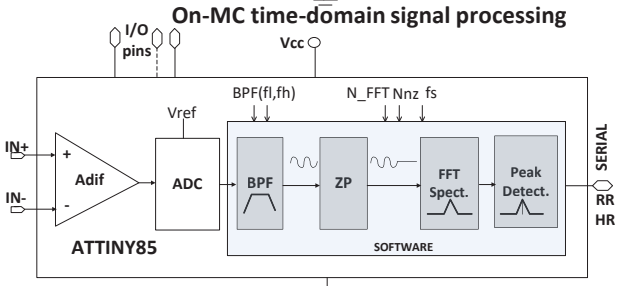
Design considerations

(6) Low power 8-bits RISC microcontroller (MC) processor unit. MC from Atmel, Microchip, STMicroelectronics, NXP. On-MC implementation of all tasks: signal management and acquisition, signal processing in time and frequency domain, low power management and communication. Example of ATtiny85 (Atmel-Microchip).





On-MC time-domain signal processing



www.smart4all-project.eu

On-MC frequency-domain signal processing

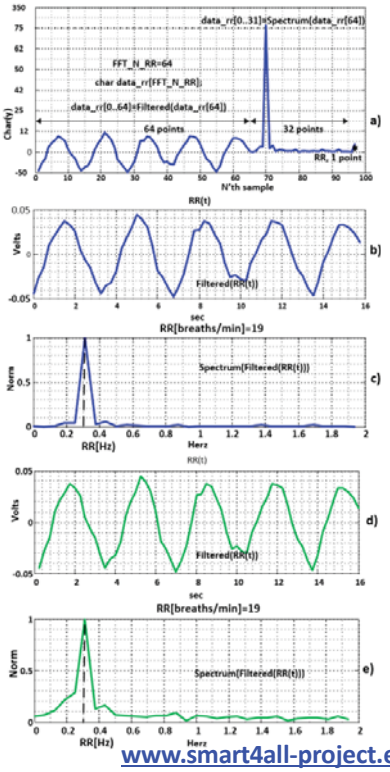
Design considerations

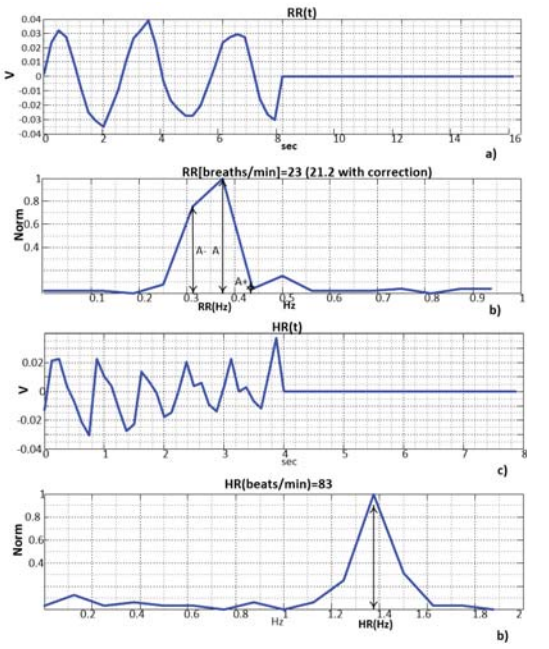
- On-MC signal processing.
- All parameters should be on-place calculated, filter border frequencies, time responses, Spectrums, adaptive thresholds.
- Strongly integer (fixed point) based, highly optimized signal processing.
- Take care about sampling frequency and interrupt based sampling generation.
- The software process should be synchronized in real time.

www.smart4all-project.eu

Design considerations

- Calculation RR and HR from FFT spectrum a) the same vector is used for time and spectral data. b) time signal of respiration signal RR(t), c) its FFT spectrum, calculated by microcontroller. d) and e) the same spectrum calculated by MATLAB.
- Zero padding technique applied to respiration a), and heart rate c) signals, with adequate FFT spectrums, b) and c). Locations of the peaks, RR, and its correction, Rrcorrected, using values from b)





$$RR_{corrected} = RR - \mu \qquad \mu = -\frac{A^+ - A^-}{2 * A - A^- - A^+}$$

www.smart4all-project.eu

Design considerations

- **Code for determination RR and HR from FFT spectrum, based on fixed point FFT and zero crossing**

```
void loop()
{
  float y_lp, y_lp_hp, RRR = 0;
  if (sample_ready == true)
  {
    cli();
    sample = ReadInternal(); //Sample
    y_lp = low_pass(...); //LPF
    y_lp_hp = high_pass(....); //HPF
    y_lp_hp = int(y_lp_hp);
    y_lp_hp *= -1; // Invert
    if (y_lp_hp >= +127.00) y_lp_hp = 127.00; //limit on 8bits arithmetic
    if (y_lp_hp <= -127.00) y_lp_hp = -127.00;
   data_rr[rr] = char(int8_t(y_lp_hp)); //convert to char
    rr++;
    //When over number of the points do FFT
    if (rr > FAST)
    {
      for (int i = 0; i < FFT_N_RR; i++) im[i] = 0; /fill imaginary with zeros
      fix_fft(data_rr, im, 7, 0); //implement fix FFT, 7 is 7<<1 for 128
      maxi = 0;
      maxiin = 0;
      for (int i = 0; i < FFT_N_RR / 2; i++) // find max point in spectrum;
      {
 dat = (int8_t(data_rr[i]) * int8_t(data_rr[i])) + (int8_t(im[i]) *
int8_t(im[i]));
        if (dat > maxi) {
          maxi = dat;  // find max amplitude
          maxiin = i; // find index corresponds to max
        }
      //Do CORRECTION
      }
      RRR = 60 * ((float(maxiin) * float(fs)) / float(FFT_N_RR));
      if (maxi > 20) {
        SwSerial.println((int)RRR); //Send RRR
      }
      else
      {
        SwSerial.println("-100"); //Send error
      }
      for (int i = 0; i < FFT_N_RR; i++) data_rr[i] = 0; //fill with zero
      rr = 0;
    }
    if (rr > FAST) rr = 0;
    sample_ready = false;
    sei();
  }
} // loop
```

```
#include "fix_fft.h" // include library for FIX FFT
#define F_CPU 1000000UL
#define pinRx PB5
//.....
SoftwareSerial SwSerial(....); // Software serial
(RX_pin,TX_pin)
bool sample_ready = false; int sample; const byte FFT_N_RR =
128;
char im[FFT_N_RR], data_rr[FFT_N_RR]; //FFT Spectrum
const byte FAST = 64; //Number of Nzp points
const byte fs = 8;  // Sampling frequency
int dat = 0; //Spectrum calculation
byte rr = 0; //Sample counter
int8_t maxiin = 0;
long maxi = 0;
//....
void setup() { //..... }
```

```
void comp_match() // Timer interrupt setting { //.....}
ISR (TIMER1_COMPA_vect) // Timer interrupt sampling function { //
....  }
void diff_amp() // Seting diff amplifier
{
  ADMUX = 0b10000000; ADMUX |= 0b00000111;  // + diff_input on
PB4, - diff input on PB3, Gain 20x
  ADCSRA = 0b10000011;   ADCSRB = 0x00;
}
int ReadInternal() //Read A/D converted
{
  static int Read;
  uint8_t low, high;
  ADCSRA |= 0b01000000;
  while (ADCSRA & 1 << ADSC); // wait untill coversion be completed
(ADSC=0);
  low = ADCL;  high = ADCH;
  Read = high << 8 | low;
  return Read;
}
```

www.smart4all-project.eu

Design considerations

- **Performances of optimized code** design of HR and RR estimation on low-memory, low-cost 8-bits RISC microcontroller, as it is ATtiny85. As seen, 128 points FFT can be implemented using 433 bytes of RAM. 32 non-zero points are taken in 8sec, allowing 2 breaths/min resolution for 4Hz sampling frequency. The MCU on 1MHz is consuming about 1.32mA.

Signal	N_{FFT} (points)	N_{nz} (points)	f_s (Hz)	CT (sec)	Err (breaths or beats/min)	MU (bytes of RAM)	f_c (MHz)	I_c (mA)	MC
RR	64	64	4	16	3.75	307	1	1.32	ATTINY85
	64	32	4	8	3.75	307	1	1.32	
	128	128	4	32	1.875	433	1	1.32	
	128	64	4	16	1.875	433	1	1.32	
	128	32	4	8	1.875	433	1	1.32	
HR	64	64	8	8	7.5	307	1	9.6	
	64	32	8	4	7.5	307	1	9.6	
	128	128	8	16	3.75	421	1	9.6	
	128	64	8	8	3.75	421	1	9.6	
	128	32	8	4	3.75	421	1	9.6	
	256	256	8	32	1.875	N/A			ATMEGA32U

www.smart4all-project.eu

Design considerations

- Code. RC (LP) i CR (HP) code implementation of 1st order filters. The only input parameter are sampling frequency (fs) and bordering frequencies (fc_, fc_h). Coefficients calculation and call filters.

```
const int fs=200; //sampling frequency

//filter variables
const int fc_l=5; //corner frequency HP
float alfa=0; //coefficient LP
float y_old_lp=0; //previous value y LP

const int fc_h=15; //corner frequency LP
float beta=0; //coefficient HP
float y_old_hp=0; //previous value x HP
float x_old_hp=0; //previous value y HP

void setup()
{
.....
  alfa=calculate_alfa((float)(fc_h), fs); //calculate_alfa
  beta=calculate_beta((float)(fc_l), fs); //calculate beta
.....
}
//coefficient alfa for LP filter
float calculate_alfa(float fc, float fs)
{
float alfa;
alfa=(2*PI*fc/fs)/((2*PI*fc/fs)+1);
return alfa;
}

//coefficient beta in HP filter
float calculate_beta(float fc, float fs)
{
float beta;
beta=1/((2*PI*fc/fs)+1);
return beta;
}
```

Calculation of coefficients

```
//LP filter of 1st order
float low_pass1(float alfa, float x)
{
float y=0;
y=alfa*x+(1.0-alfa)*y_old_lp;
y_old_lp=y;
return y;
}
//HP filter of 1st order
float high_pass1(float beta, float x)
{
float y=0;
y=beta*y_old_hp+beta*(x-x_old_hp);
y_old_hp=y;
x_old_hp=x;
return y;
}

void loop()
{
.....
sample= analogRead(AD0);
y1=high_pass1(beta, float(sample);//HPF 5Hz
y2=low_pass1(alfa,y1); // LPF 15Hz
......
}
```

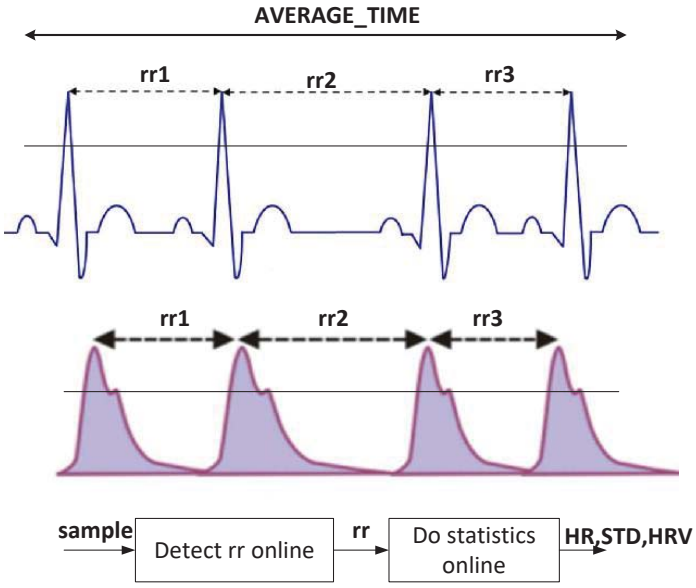
www.smart4all-project.eu

Design considerations

- Code. Iterative calculation of statistical parameters. Less memory, on-line calculation. Example of processing RR intervals of PPG and ECG signals for purpose of Heart Rate Variability (HRV) calculation. Stress detection.

```
//DO STATISTICS FOR HR AND STDEV
short int stat_count=0;
long int par_sum_rr=0; //partial sum for mean value
long int std_sum=0;  //partial sum for std value
short int HR_AVE=0; //HR in AVERAGE_TIME
int STD=0; // STD in AVERAGE_TIME
int arithmia=0; //arithmia counter
void do_statistics(int rr)
{
 float B;
 if(rr>1500 || rr<500) // Arrhythmias detected
{
  HR_AVE=0;
  arithmia++;
  par_sum_rr=0;
  std_sum=0;
  stat_count=0;
  STD=0;
}
else
{
 stat_count++;  //statistics counter
 if(stat_count>1)
 {
 par_sum_rr=par_sum_rr+long(rr); //partial sum for mean value
 std_sum=std_sum+long(rr)*long(rr);  //partial sum for std value
                                  if(stat_count>=AVERAGE_TIME)
                                  {
                                   B=float(par_sum_rr)/float(AVERAGE_TIME-1); //mean value
                                  HR_AVE=short(60000/B); // HR from mean value
                                  B=(B*B); // mean*mean
        std_sum=std_sum/(AVERAGE_TIME-1);
        STD=round(sqrt(float(float(std_sum)-B))); //formula for standard deviation
                                  stat_count=0;
                                  par_sum_rr=0;
                                  std_sum=0;
                                  } } }
}
```

$$s^2 = \frac{\sum x^2}{n} - \bar{x}^2 \text{ instead } s^2 = \frac{\sum (x-\bar{x})^2}{n}$$



www.smart4all-project.eu

Design considerations

- **Code. IIR filter implementation in floating and fixed point aritmetics**

```
//IIR FILTER FLOAT IMPLEMENTATION
//a(1)*y(n) = b(1)*x(n) + b(2)*x(n-1) + ... + b(nb+1)*x(n-nb)
  //       - a(2)*y(n-1) - ... - a(na+1)*y(n-na)
  // calling    yout=iir_filtar(xin, a_c, b_c, n);
#define N 4
double y[N+1]={0,0,0,0,0};
double x[N+1]={0,0,0,0,0};

// a(1)*y(n)=b(1)*x(n)+b(2)*x(n-1)-a(2)*y(n-1)
double a_c[]={1.0000, -0.9975}; //floating coefficients
double b_c[]={ 0.0013,  0.0013};

double iir_filtar(double p, double *a_coef, double *b_coef, int N_order)
{
  int i;
  x[0]=p;
  y[0]=*b_coef*x[0];
  for(i=1; i<=N_order; i++)
  y[0]=y[0]+(*(b_coef+i)*x[i]);
  for(i=1; i<=N_order; i++)
  y[0]=y[0]-(*(a_coef+i)*y[i]);
  for(i=N; i>0; i--) //Circular
  {
  y[i]=y[i-1];
  x[i]=x[i-1];
  }
  return(y[0]); }
```

```
//IIR FILTER INTEGER IMPLEMENTATION
long a_co[]={1, -199};  //integer coefficients
long b_co[]={29,  29};
long yi[N+1]={0,0,0,0,0};
long xi[N+1]={0,0,0,0,0};

long iir_filtar_int(long p, long *a_coef, long *b_coef, int N_order)
{
  short i;
  xi[0]=p;
  yi[0]=(*b_coef*xi[0])>>8;

  for(i=1; i<=N_order; i++){
  yi[0]=yi[0]-((*(a_coef+i)*yi[i])>>8);
  }


  for(i=N; i>0; i--) //Circular
  {
  yi[i]=yi[i-1];
  xi[i]=xi[i-1];
  }
  return(yi[0]);
}
```

```
p=(long)(x<<8);   //Calling integer IIR filter
 yk=iir_filtar_int(p,a_co,b_co, 1);
```

www.smart4all-project.eu

Design considerations

- **Code. DC Removal, DC Tracking. IIR notch filter coefficients calculation. Positive slope, Smoothing**

```
//DC REMOVAL
float al=0.995;
float yn_1=0;
float xn_1=0;

float DC_removal(float x)
{
                    float y;
                    y=al*yn_1+x-
xn_1;
                    yn_1=y;
                    xn_1=x;
                    return(y);

}
```

$$H(z)=\frac{1-2\cos\omega_0 z^{-1}+z^{-2}}{1-2r\cos\omega_0 z^{-1}+r^2 z^{-2}}$$

```
//DC TRACKING
int32_t ydc_old=0;
int DC_Tracking(int x)
{
int32_t ydc;
  ydc= ydc_old+((((int32_t) x << 16) - ydc_old) >> 9);
  ydc_old=ydc;
  return (ydc>>16);
}
```

```
//IIR NOTCH FILTER WITH //COEFFICIENTS CALCULATION
fs=1000;
f0=50;  // REMOVE 50Hz flicker
b0=1;
b1=-2*cos(2*pi*f0/fs);
b2=1; r=0.999;
a0=1;
a1=-2*r*cos(2*pi*f0/fs)
a2=r*r;
```

```
//a(0)*y(n) = b(0)*x(n) + b(1)*x(n-1) + b(2)*x(n-2) ...
  //        - a(1)*y(n-1) - a(2)*y(n-2)
```

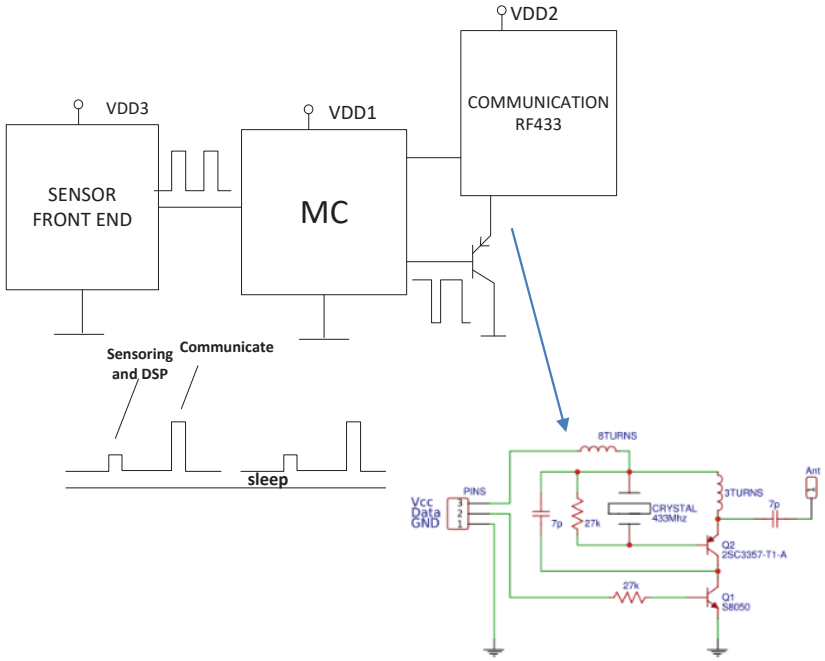
```
//POSITIVE SLOPE calculation
int16_t x_old_slope_fix=0;
int16_t slope_fix(int16_t x)
{
int16_t slope=0;
slope=x-x_old_slope_fix;
if(slope<=0) slope=0;
x_old_slope_fix=x;
return slope;
}
```

```
//SMOOTHING, CIRCULAR BUFFERING
float average_sum(float x)
{
short i;
float filterout=0.0;
// Direct-Form FIR
del[0] = x; // input for filter
filterout = del[0]; // Set up filter sum
for (i = LENGTH-1; i > 0; i--){ // Get sum of products
filterout += del[i];
del[i] = del[i-1]; // Renew input array
}
return (filterout);
}
```

www.smart4all-project.eu

Design considerations

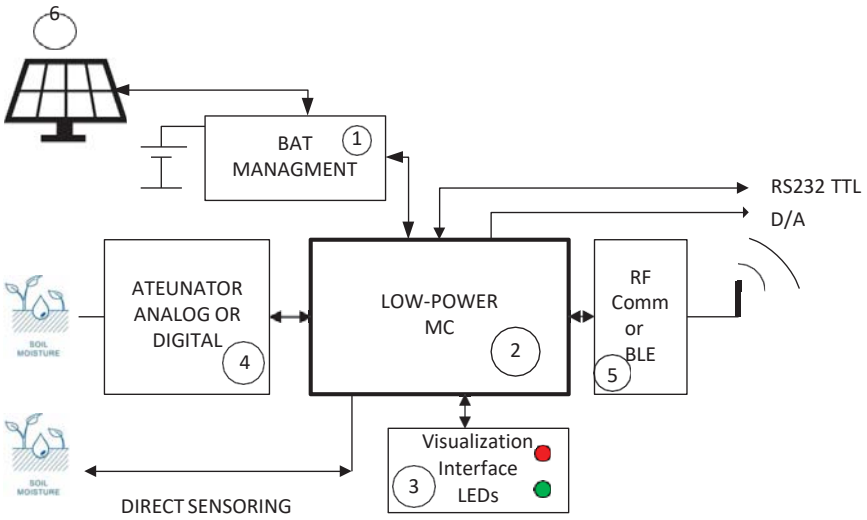
- **(5) Communication and low powering**
- Communication is the most powering part. Then, if feasible, it should be as rare as possible from time to time, at short intervals
- Use power consumption saving modes on MC.
- Take care about timing of sensoring, DSP, and communication. Good experience with RF433MHz communication, one way. The communication is modulated switched on-off by Tx signal (Data). To increase communication power VDD2>VDD1, VDD3.

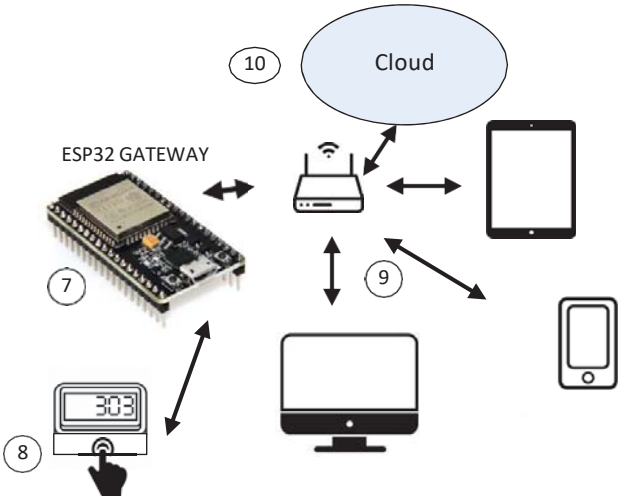


www.smart4all-project.eu

Integration

- Home(office) integration. Low-power node and ESP32 gateway





Integration

- Out-door, on-field-cloud scenario. Precision agriculture example.

23

Conclusions

- We discussed some of cost and performance effective hardware and software designs for smart nodes.
- It uses OFF-THE-SHELF components as 8-bits microcontrollers of general purposes.
- The explanation has been done through real examples in both hardware and software from fields of health wearables, precision agriculture etc…
- To design effective nodes we need wide knowledge, not only from programming.
- In majority of cases, when design it is not necessary to **"Shoot sparrows with cannons".**

www.smart4all-project.eu

Acknowledgment

The research within presentation has been supported by EU H2020, SMART4ALL, Grant Agreement: 872614. We are thankful.

stox@ucg.ac.me

Your feedback

Thanks. Questions, comments?

stox@ucg.ac.me

www.smart4all-project.eu

Data Processing Pipelines on small satellites and drones: challenges and solutions

Milica Orlandić
Department of Electronic Systems
NTNU – Norwegian University of Science and Technology

Global Context - Megatrends

1 Sustainability

2 World Interdependency

3 Smart Systems

4 Digital Transformation

5 Industry 4.0

6 Complexity Growth

Incose Systems Engineering vision 2035

Global Context - Megatrends

1

Sustainability

Global Context

1

Sustainability

United Nations Sustainable Development Goals: https://sdgs.un.org/goals

ENVIRONMENTAL SUSTAINABILITY BECOMES A HIGH PRIORITY

Global Context - Megatrends

SMART SYSTEMS PROLIFERATE

Smart elements employing AI, automation and autonomy features, and advanced sensors for system functional behaviors as well as system self-diagnosis and repair, will be commonplace.

Smart systems will be commonplace in diverse fields, such as agriculture, urban complexes, homes, appliances, health, financial services, energy, telecommunications, private and public transportation, and national security.

3

Smart Systems

Intelligence will move closer to devices and away from central control.

Global Context - Megatrends

1 Sustainability

2 World Interdependency

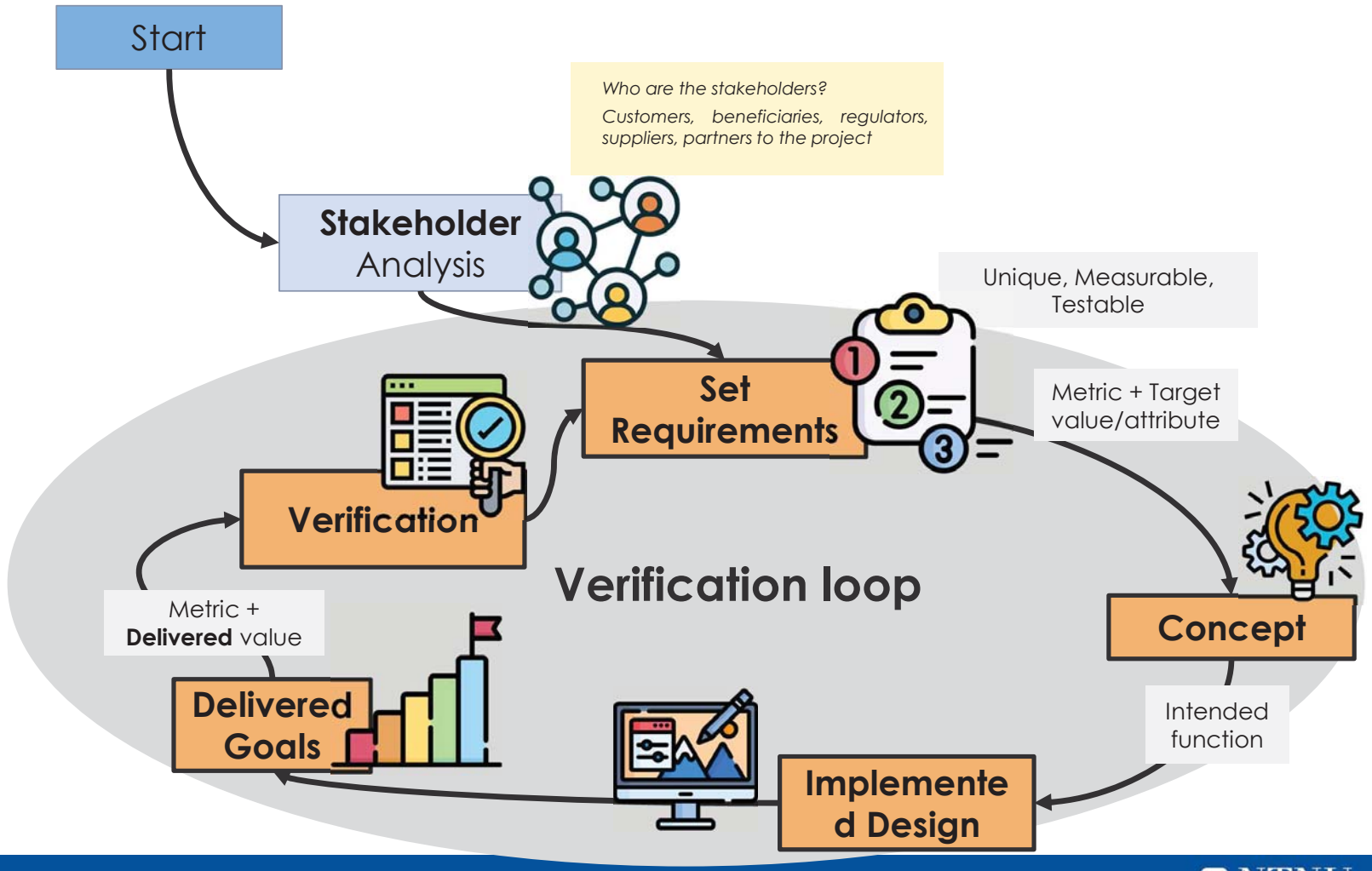
3 Smart Systems

4 Digital Transformation

5 Industry 4.0

6 Complexity Growth

Systems Engineering Process



Stakeholder Expectrations


simple


timely


safe


secure


predictable


smart


sustainable


maintainable


scalable


affordable

Incose Systems Engineering vision 2035

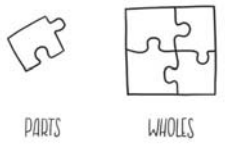
Requirements & Constraints Definition

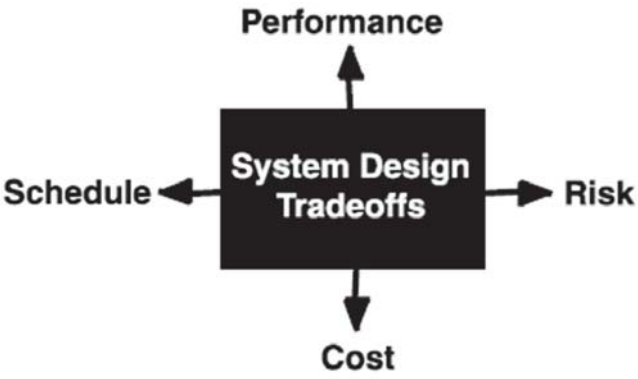
Requirements
describe the necessary functions and features of the system we are to conceive, design, implement and operate.

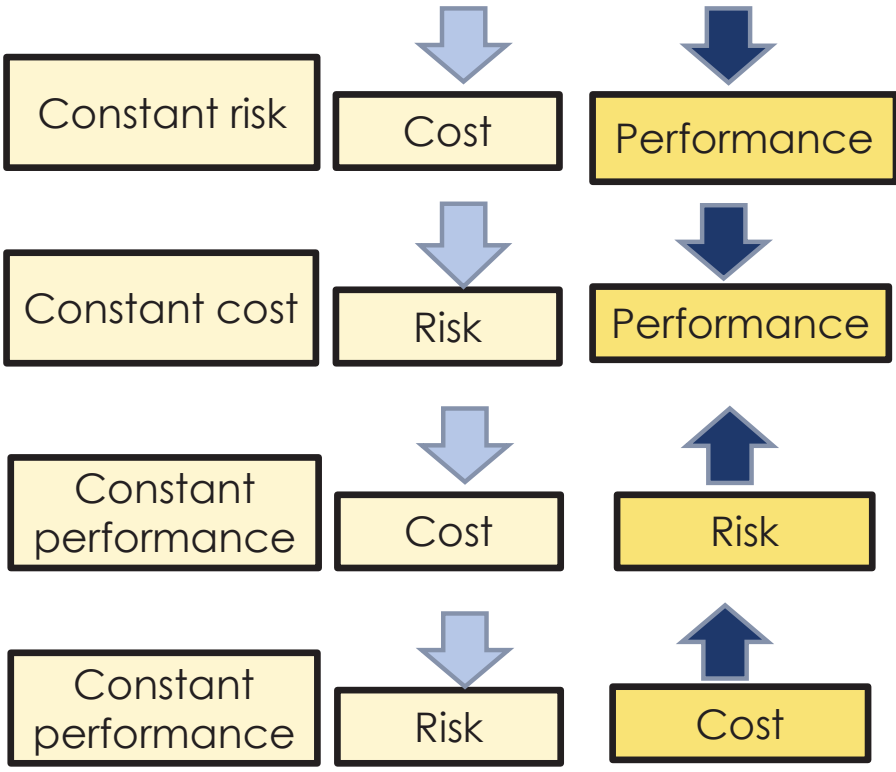
Constraints
show the limits within which the system should be realized

Requirements specify **WHAT** the system shall/should do,
Requirements Specification - rigorous modeling of requirements to provide formal definitions for various aspects of the system

System trade-offs

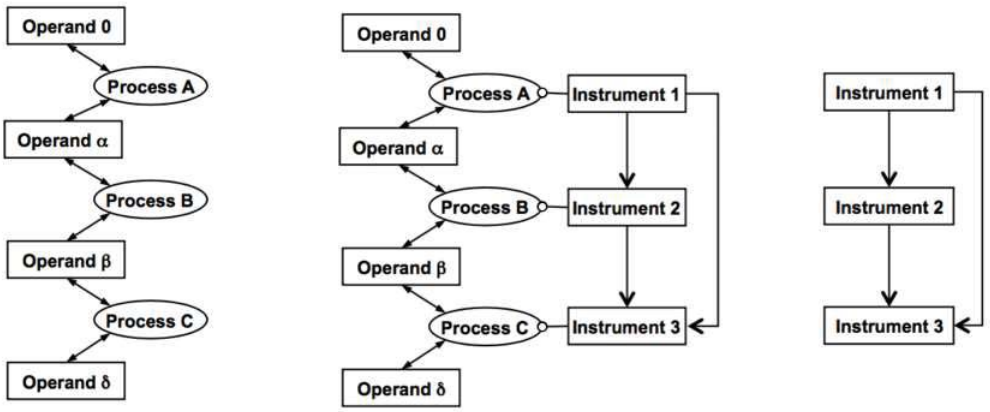






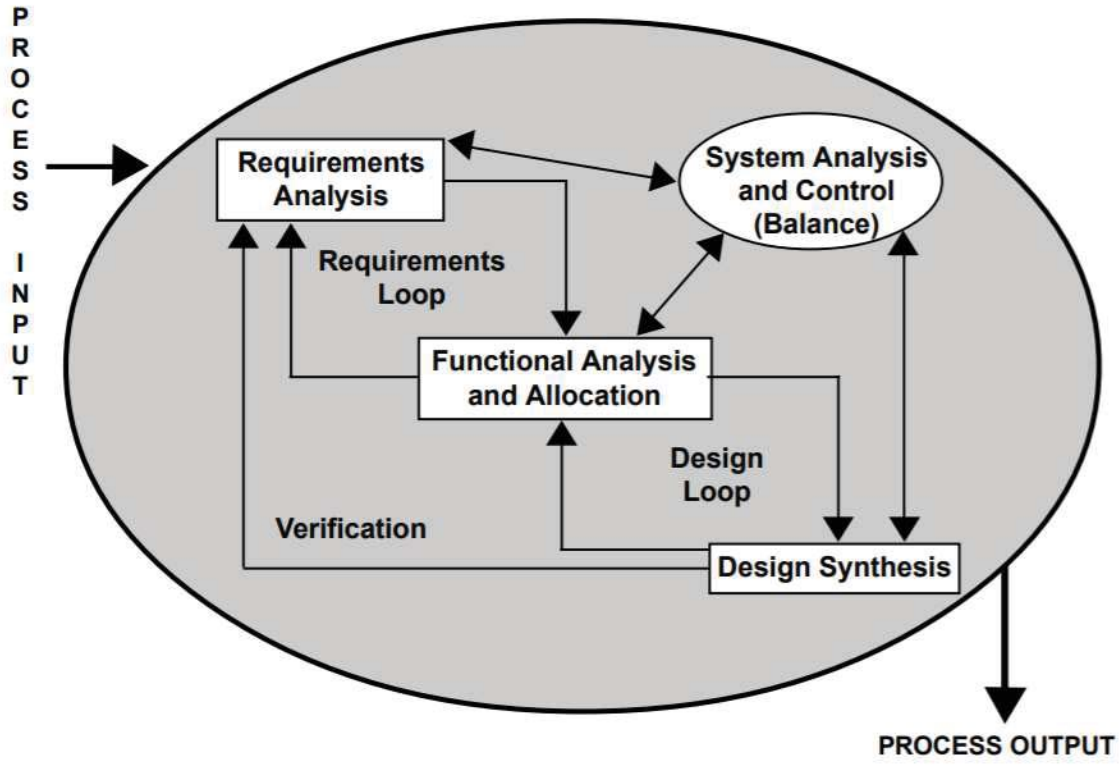
Architecture

The embodiment of concept, and the allocation of physical/informational **function** (process) to **elements of form** (objects) and definition of structural interfaces among the objects



Functional Architecture ***System Architecture*** ***Formal Structure***

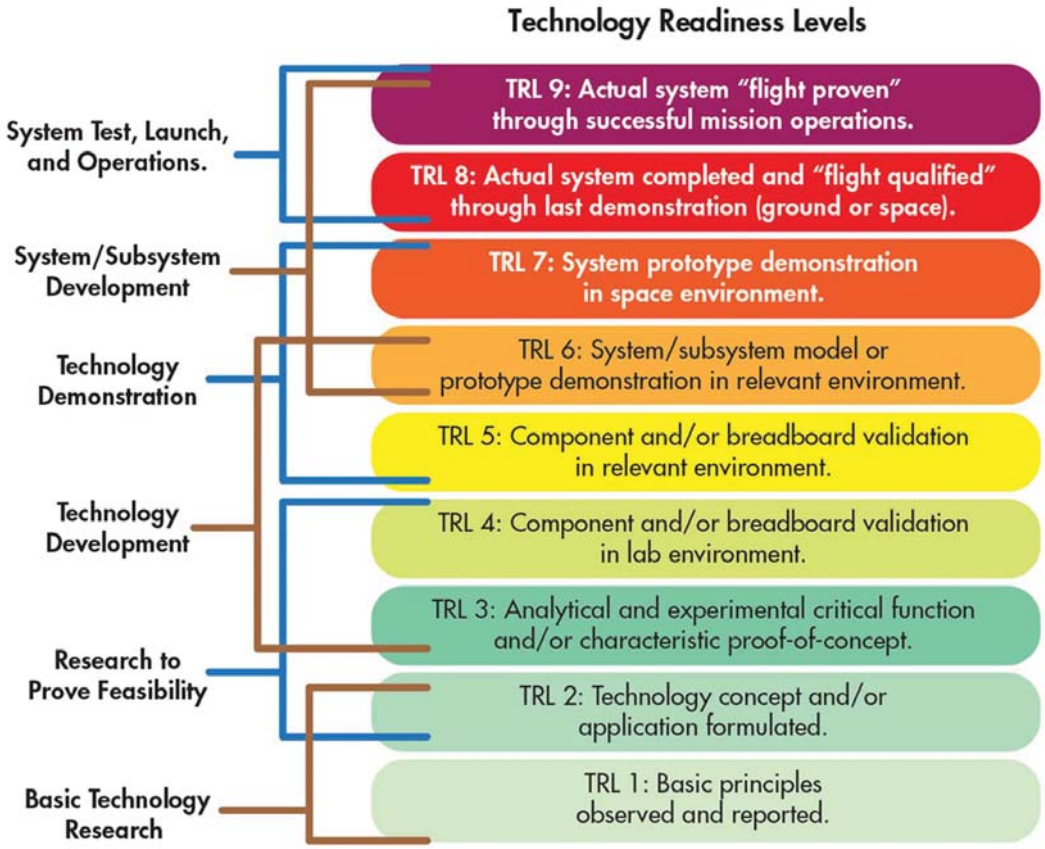
System Design Process



NASA Systems Engineering Handbook Revision 2

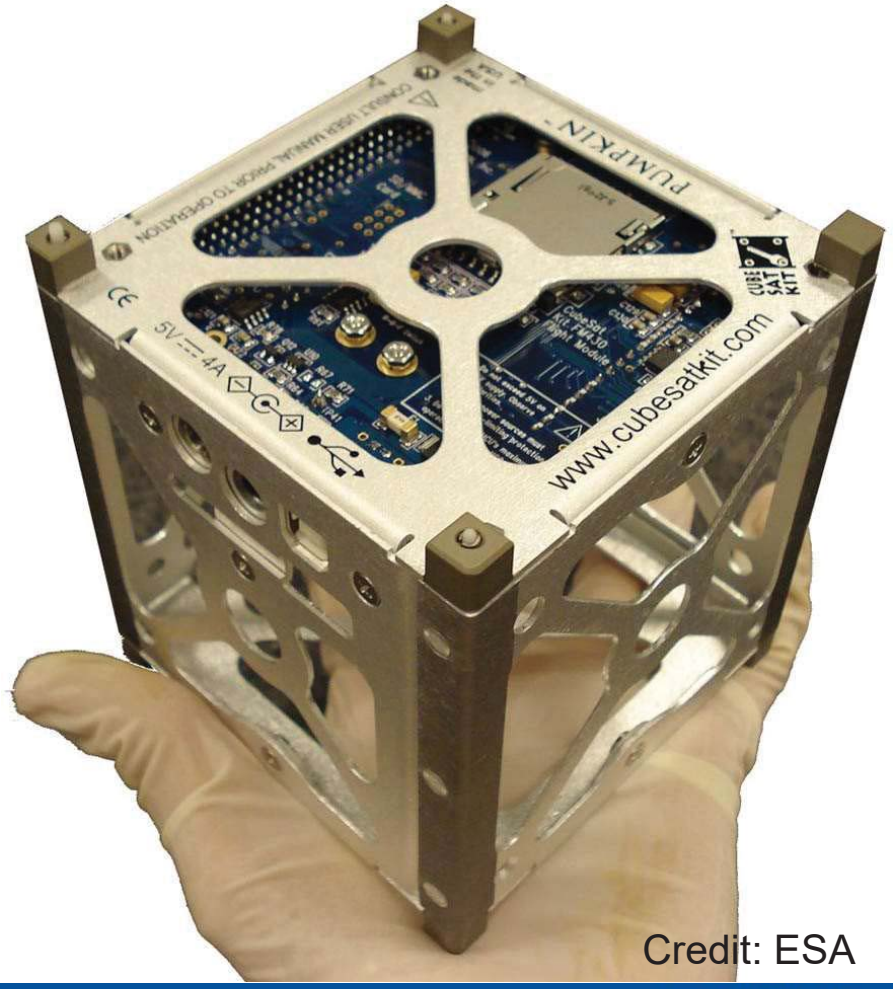
Technology Assessment

NASA Technology Readiness Levels (TRLs)



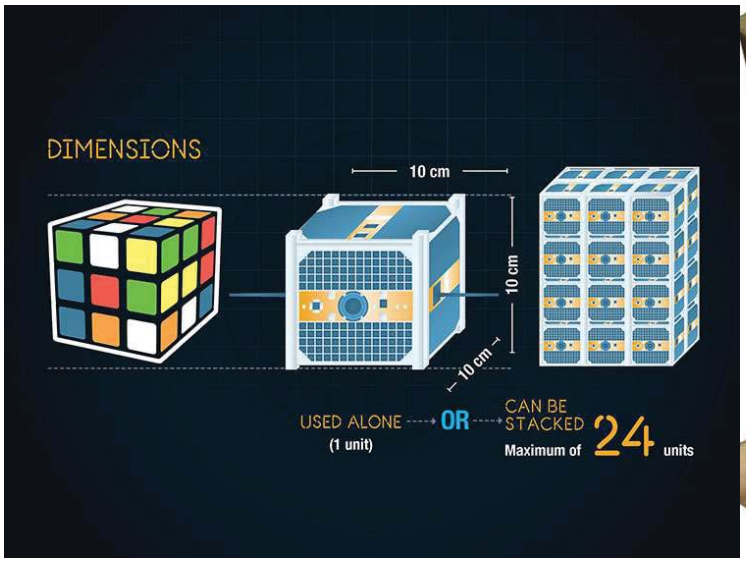
Small Satellite Characteristics

Smallsat <300 kg
Minisat 100-500 kg
Microsat 10-100 kg
Nanosat 1-10 kg
Picosat 0.1-1 kg

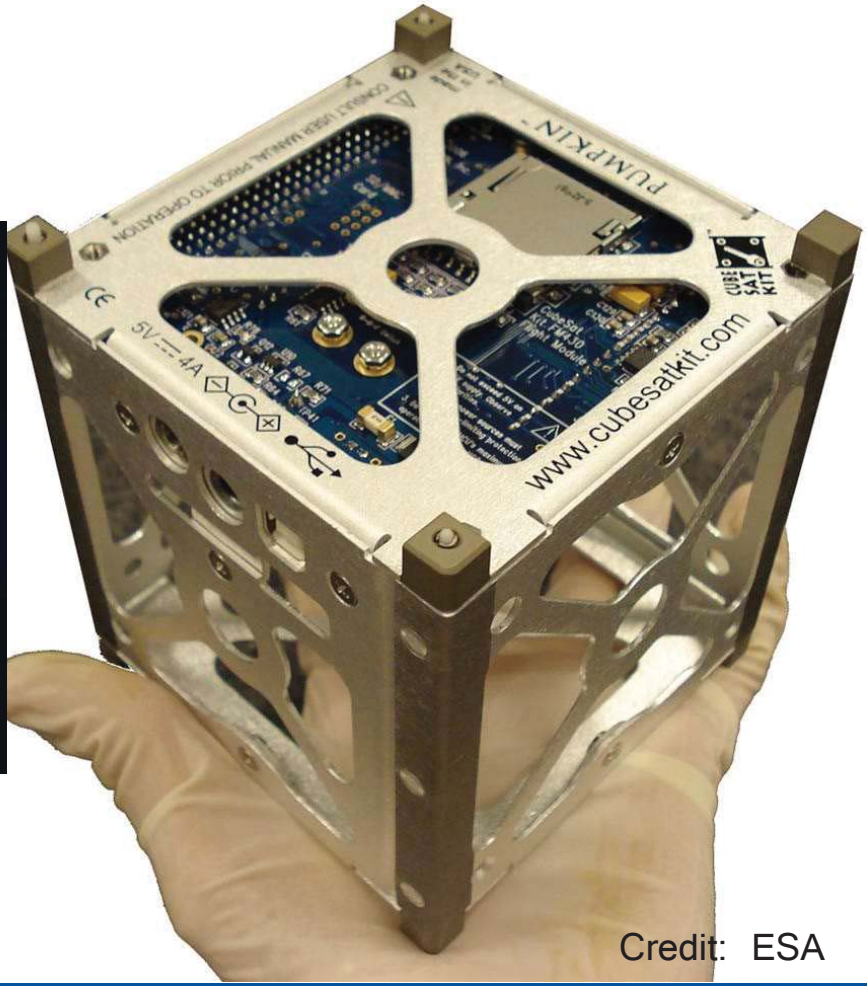


Credit: ESA

CubeSat



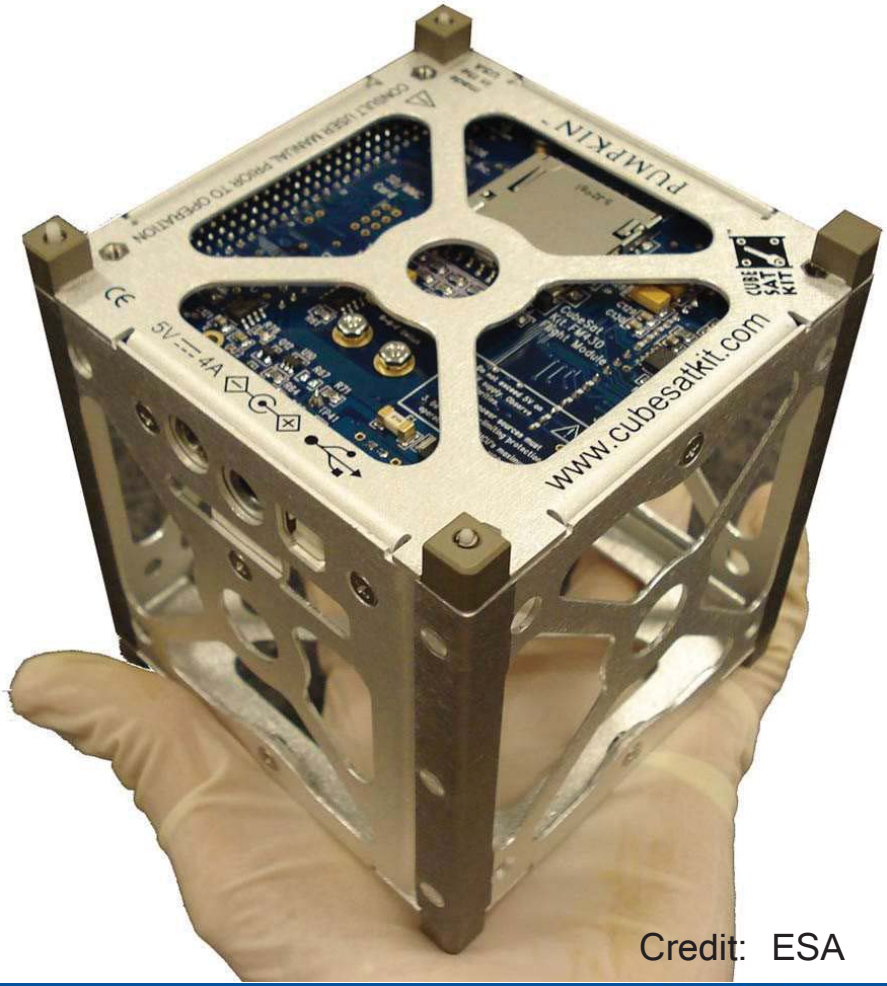
Credit: Canadian Space Agency



Credit: ESA

Small Satellite Characteristics

- Capable platforms
- Rapid infusion of technology
- Low cost, components of the shelf (COTS)
- Flexibility
- Distributed functions
- Observation strategies
- Trend: increased on-board processing, hybrid and reconfigurable computing

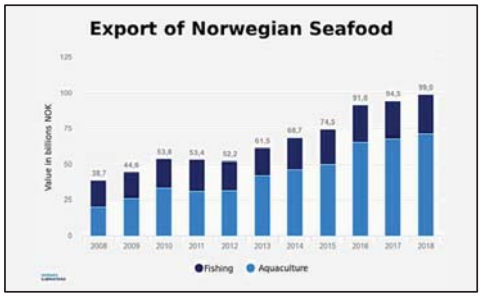


Credit: ESA

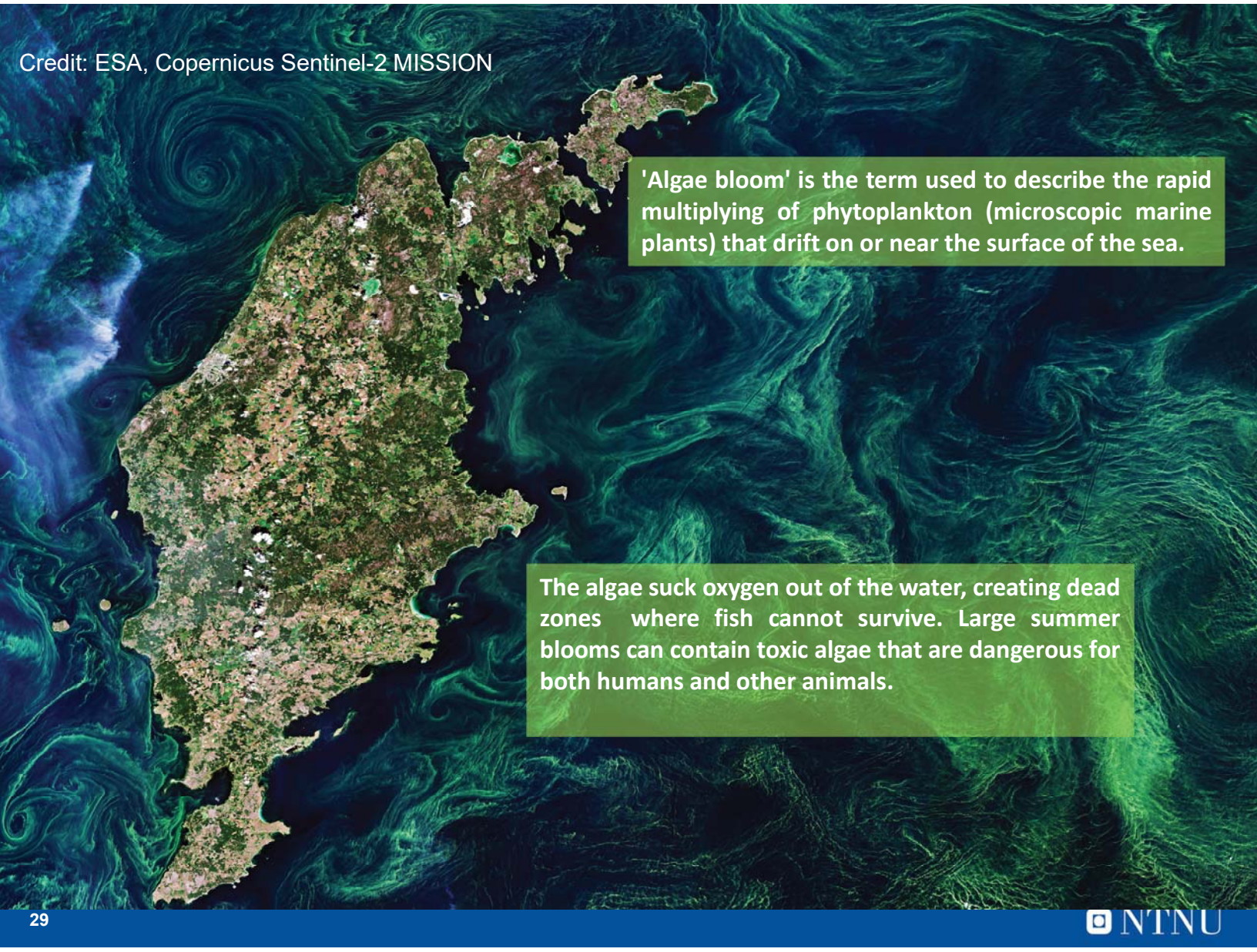
Background - Norway + Ocean

Coastal zones are rich in fish that visit the Norwegian Sea from the North Atlantic or from the Barents Sea

Norway is responsible for ⅓ of the salmon production in the world, and of all the Seafood produced in Norway 95% is exported



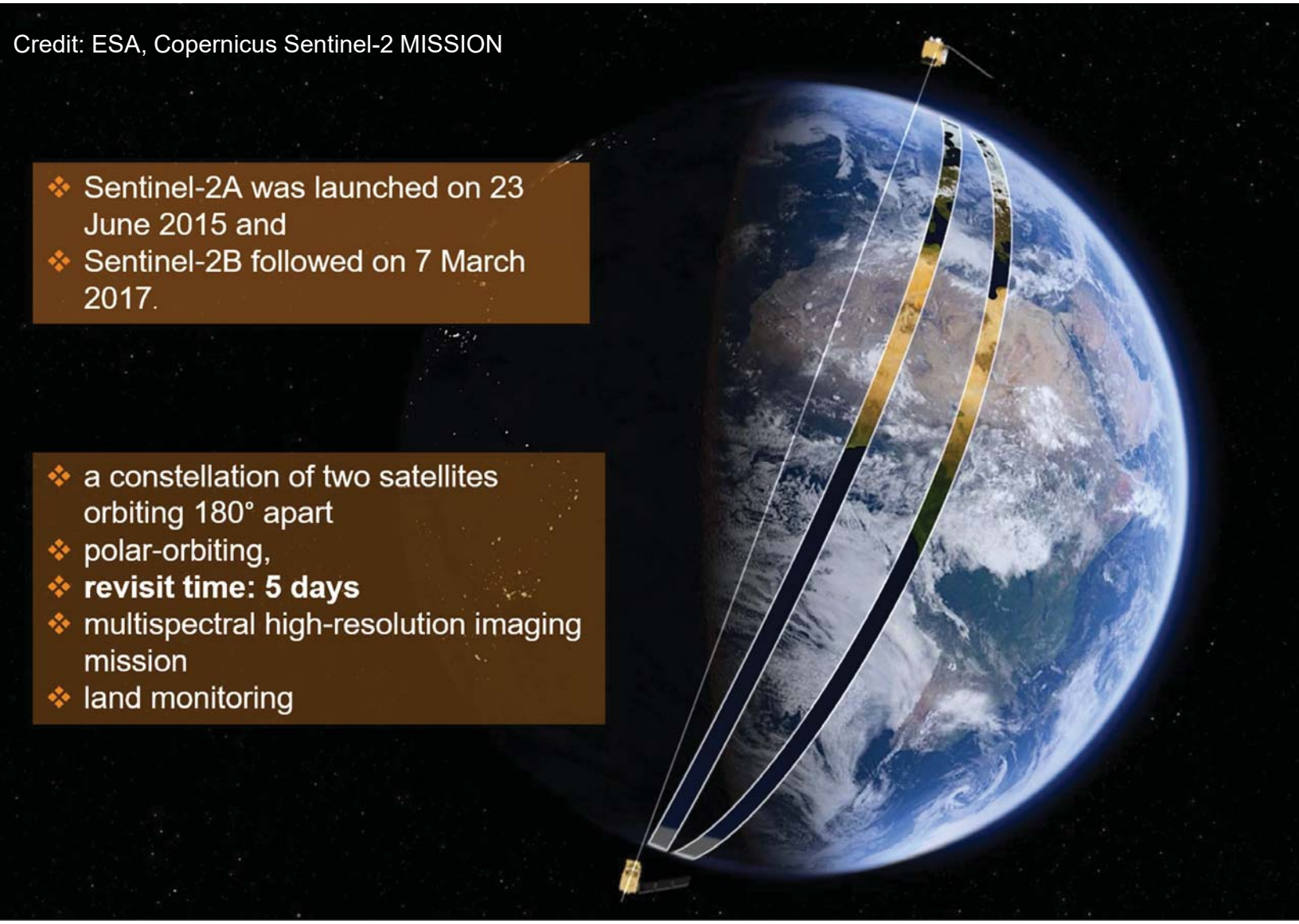
The Norwegian continental shelf is 4 times the the Norwegian mainland, and ⅓ of the area of Europe is the Norwegian continental shelf


Credit: ESA, Copernicus Sentinel-2 MISSION
'Algae bloom' is the term used to describe the rapid multiplying of phytoplankton (microscopic marine plants) that drift on or near the surface of the sea.
The algae suck oxygen out of the water, creating dead zones where fish cannot survive. Large summer blooms can contain toxic algae that are dangerous for both humans and other animals.
29
NTNU


Credit: ESA, Copernicus Sentinel-2 MISSION
Satellite data can track the growth and spread of harmful algae blooms in order to alert and mitigate against damaging impacts for tourism and fishing industries.
Without *in situ* measurements, it is difficult to distinguish the type of algae that covers the sea as many different types of algae grow in these waters.

The key recommendations for ocean color remote sensing

1) Image spatial resolution should be better than **100 m/pixel**;
2) Spectral resolution should be better than **10 nm**;
3) SNR at Top of Atmosphere (ToA) should be greater than **100**;
4) Data should be delivered to end users in less than **12 hours** in general and less than **3 hours** for HABs; and
5) Revisit times to target should be **at least 3 per day**.


Credit: ESA, Copernicus Sentinel-2 MISSION
Sentinel-2A was launched on 23 June 2015 and
Sentinel-2B followed on 7 March 2017.
a constellation of two satellites orbiting 180° apart
polar-orbiting,
revisit time: 5 days
multispectral high-resolution imaging mission
land monitoring

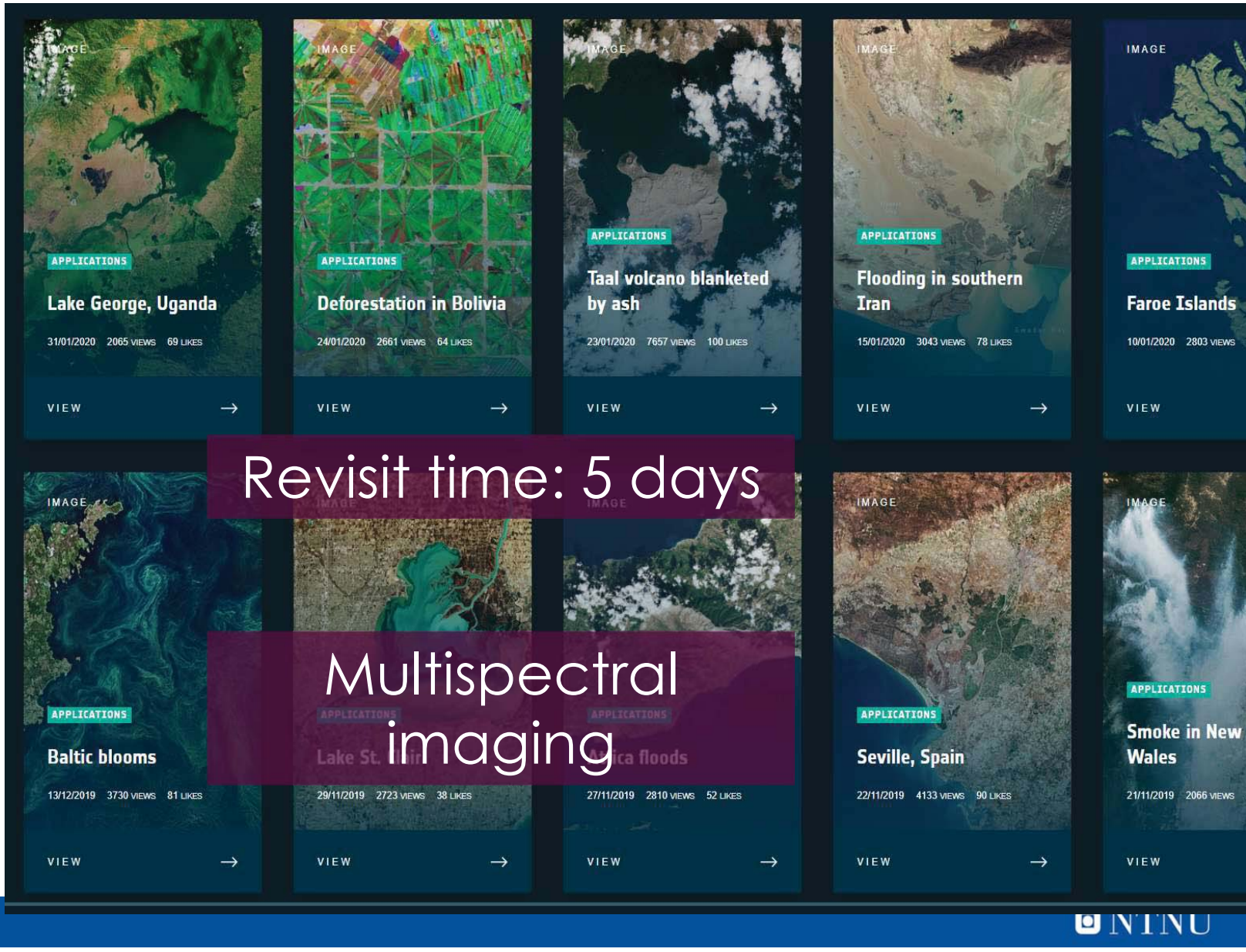
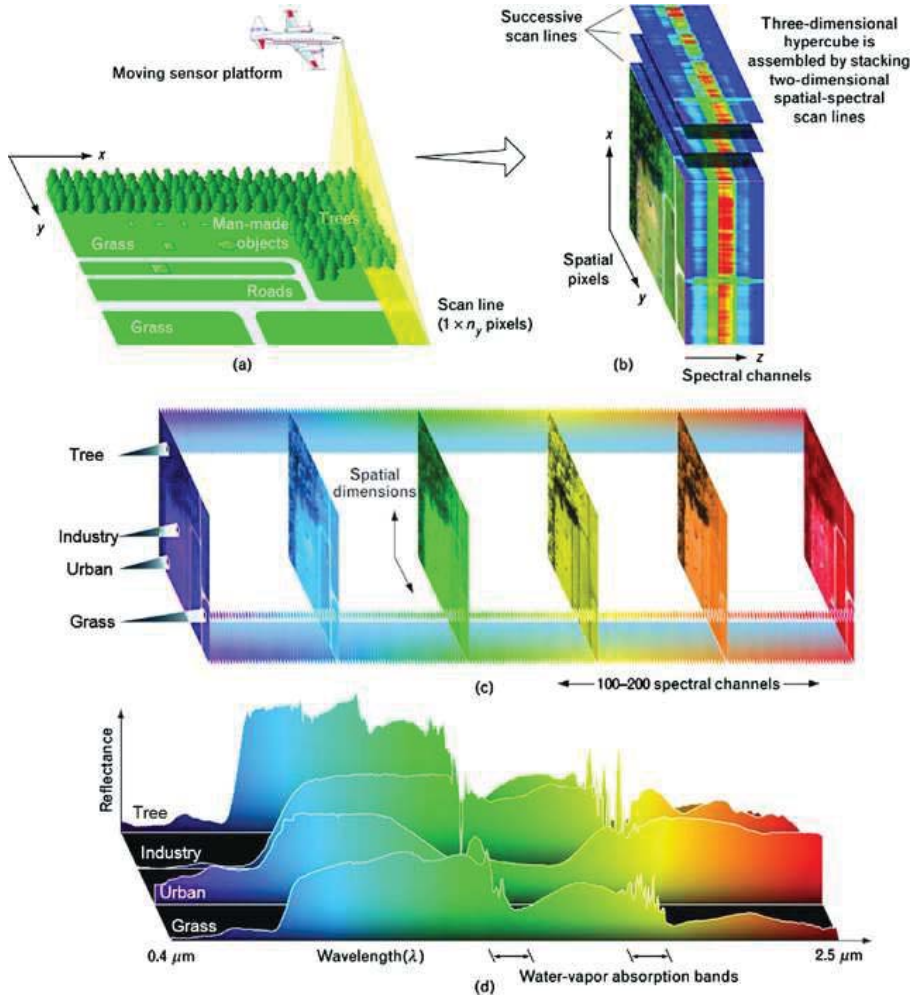


IMAGE
APPLICATIONS
Lake George, Uganda
31/01/2020 2065 VIEWS 69 LIKES
VIEW
IMAGE
APPLICATIONS
Deforestation in Bolivia
24/01/2020 2661 VIEWS 64 LIKES
VIEW
IMAGE
APPLICATIONS
Taal volcano blanketed by ash
23/01/2020 7657 VIEWS 100 LIKES
VIEW
IMAGE
APPLICATIONS
Flooding in southern Iran
15/01/2020 3043 VIEWS 78 LIKES
VIEW
IMAGE
APPLICATIONS
Faroe Islands
10/01/2020 2803 VIEWS
VIEW
Revisit time: 5 days
IMAGE
APPLICATIONS
Baltic blooms
13/12/2019 3730 VIEWS 81 LIKES
VIEW
Multispectral imaging
IMAGE
APPLICATIONS
Lake St.
29/11/2019 2723 VIEWS 38 LIKES
VIEW
IMAGE
APPLICATIONS
ica floods
27/11/2019 2810 VIEWS 52 LIKES
VIEW
IMAGE
APPLICATIONS
Seville, Spain
22/11/2019 4133 VIEWS 90 LIKES
VIEW
IMAGE
APPLICATIONS
Smoke in New Wales
21/11/2019 2066 VIEWS
VIEW
NTNU

Observational pyramid

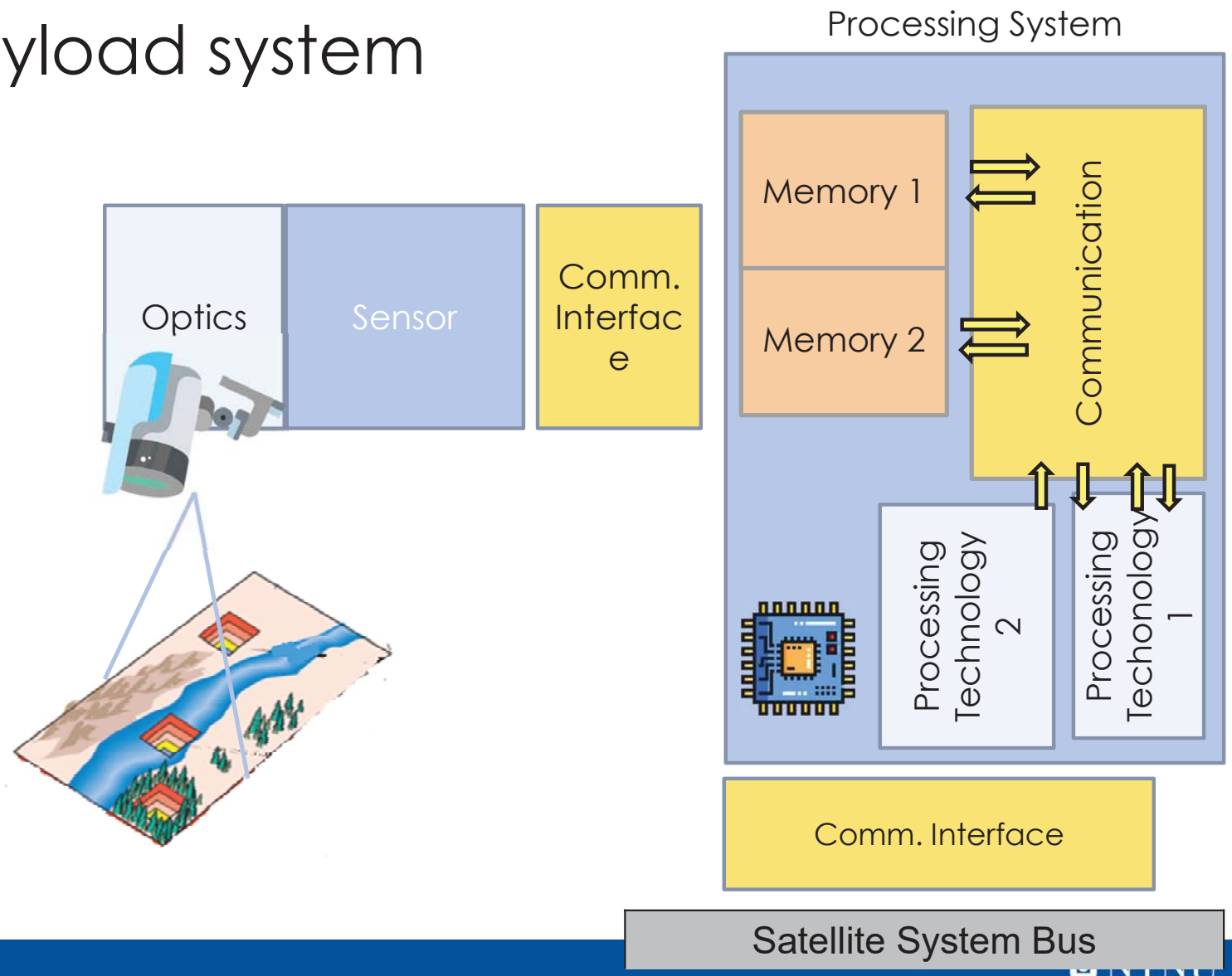
Hyperspectral imaging



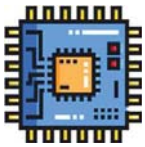
Payload Design / Trade study

- Orbit Selection
- Spacecraft autonomy
- Mission specific flight software
- Data management
- Technology trades
- Operational trades
- Risk versus return trades

Payload system



Payload system



Wavelength Range

Orbit

FoV

Optics

Noise/Signal Ratio

Frame Rate

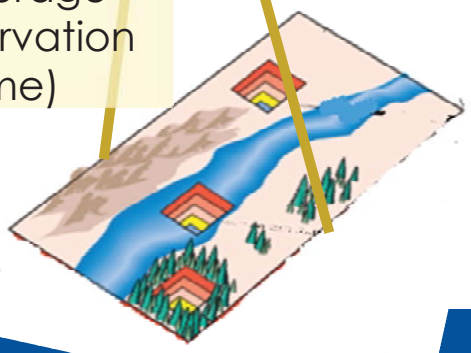
Attitude Control

Spectral Resolution

Attitude Determination

Radiometric Resolution

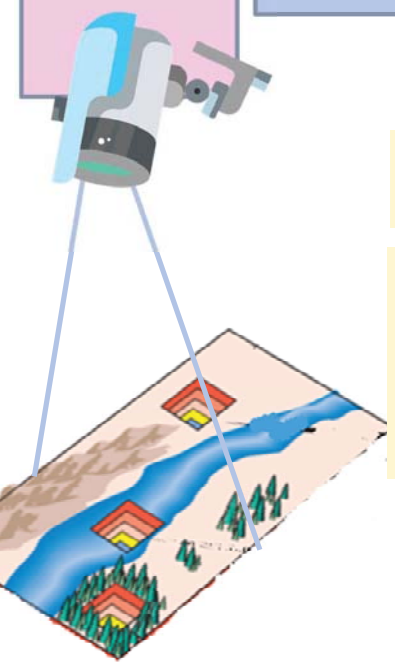
Coverage (Observation Time)

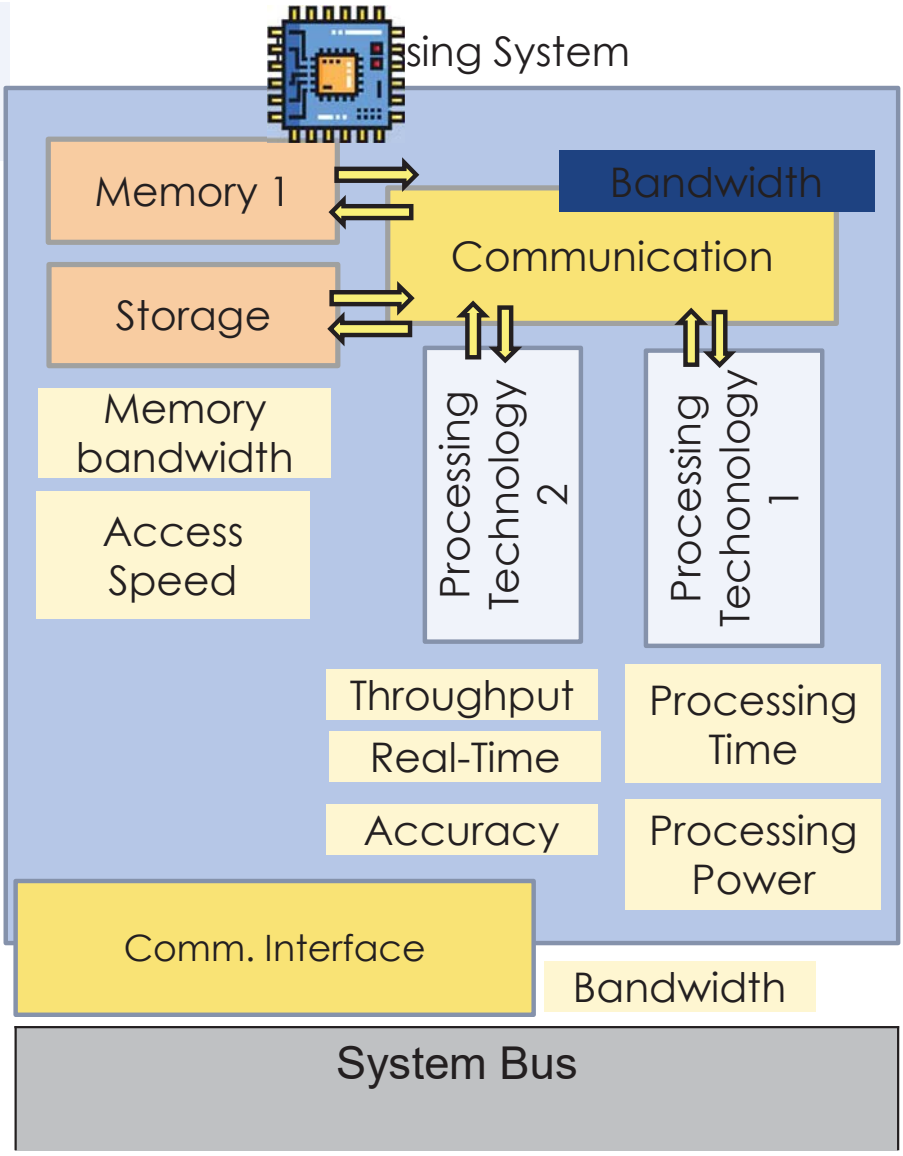


Temporal Resolution

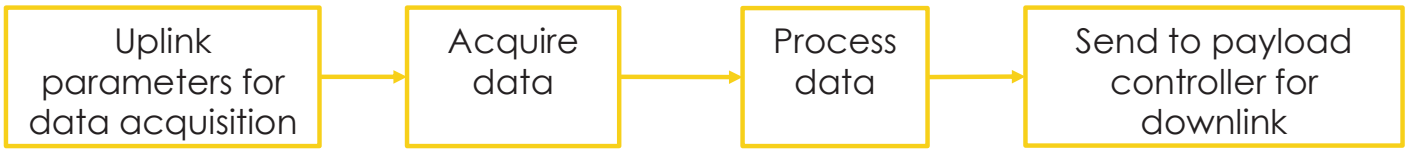
Spatial Resolution

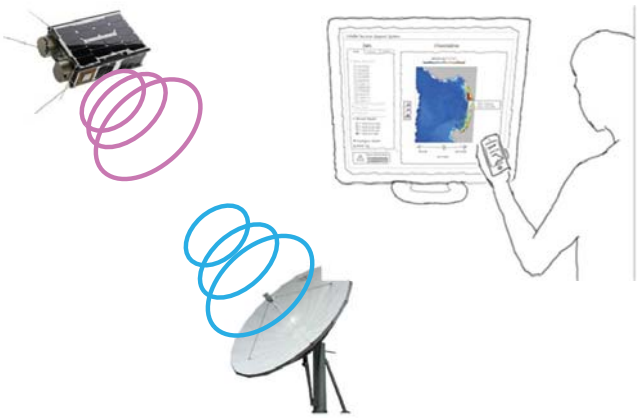
Payload system





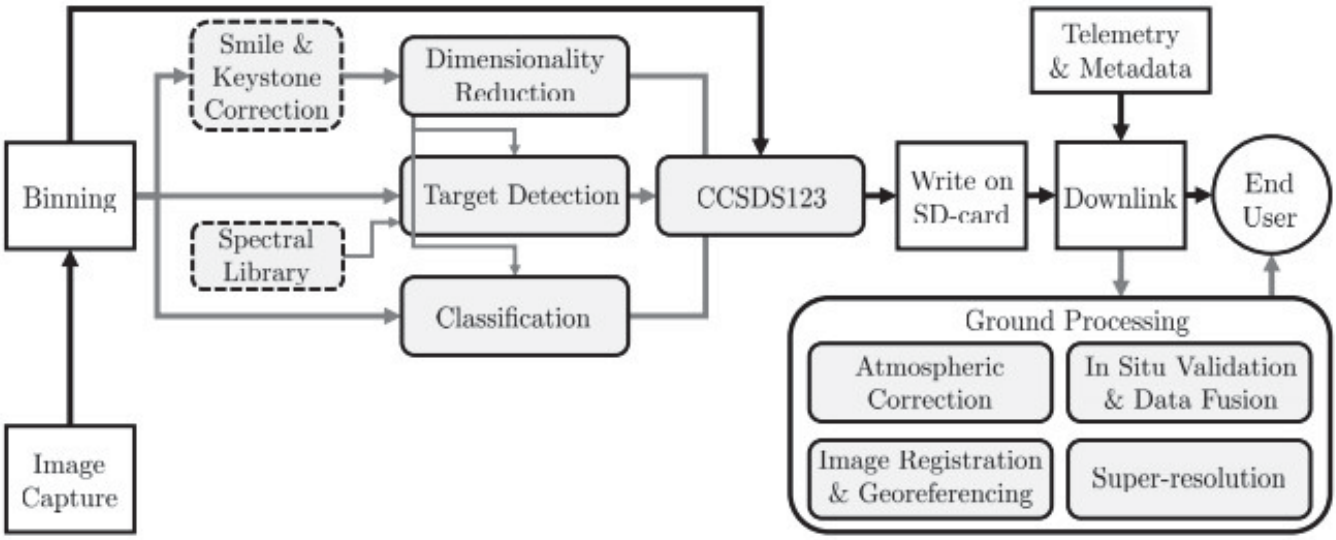
An image processing pipeline





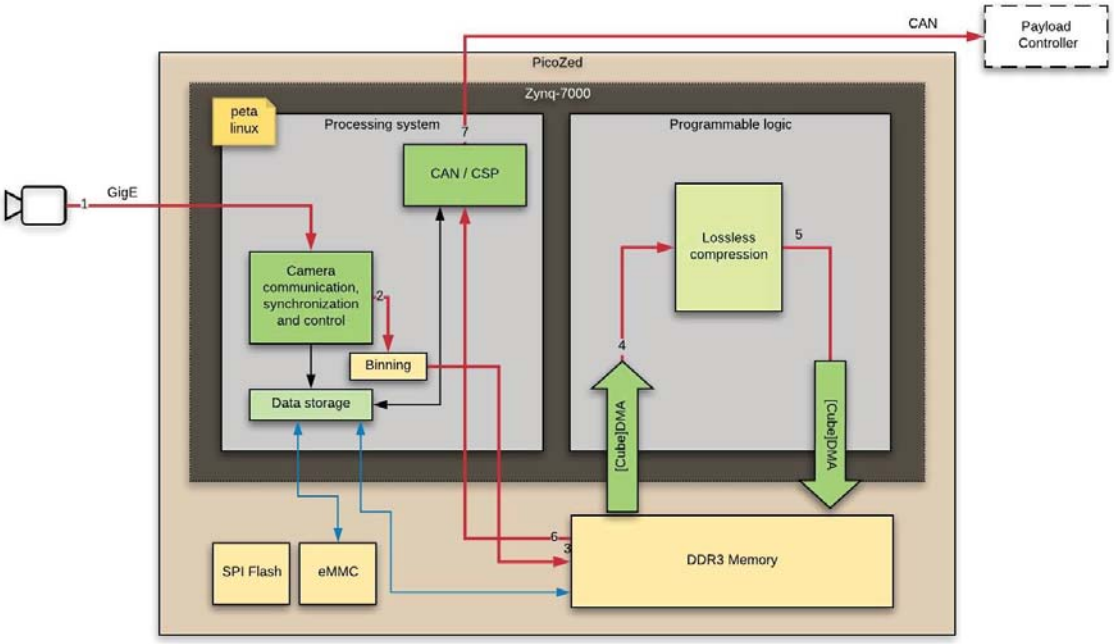
- Large data sizes (GBs)
- Timing is critical
- All processing in about X minutes
- Downlink bandwidth

Onboard processing pipeline

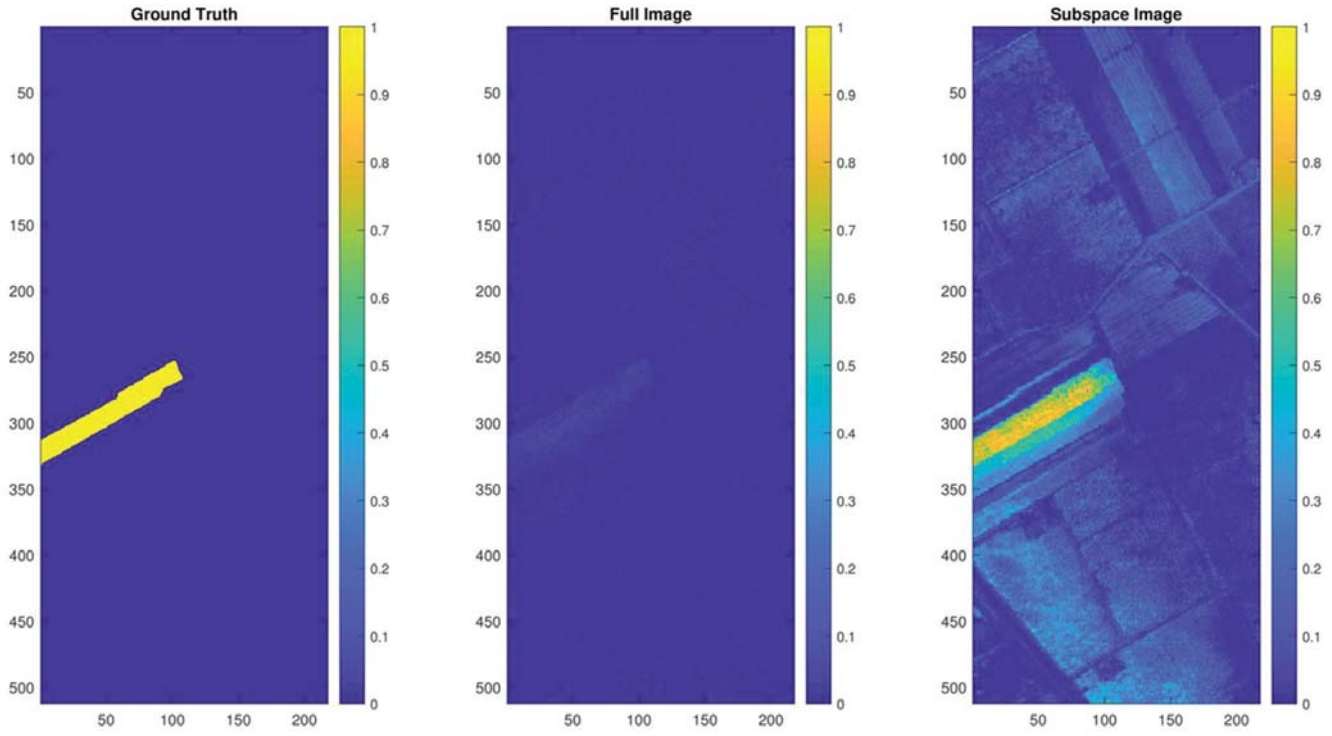


M. E. Grøtte *et al*., "Ocean Color Hyperspectral Remote Sensing With High Resolution and Low Latency—The HYPSO-1 CubeSat Mission," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1-19, 2022, Art no. 1000619, doi: 10.1109/TGRS.2021.3080175.

Hyperspectral onboard processing architecture – HYPSO 1



A three color-composite of three spectral bands taken from the hyperspectral dataset. Image shows the coast of New Zealand recorded with *Hyperspectral Imager for the Coastal Ocean* (HICO) spectrometer.



D. Bošković, M. Orlandić, S. Bakken and T. A. Johansen, HW/SW Implementation of Hyperspectral Target Detection Algorithm, 8th Mediterranean Conference on Embedded Computing – MECO, Montenegro, 2019
S. Bakken, M. Orlandic, T. A. Johansen, The effect of dimensionality reduction on signature-based target detection for hyperspectral imaging, CubeSats and SmallSats for Remote Sensing III, SPIE Optical Engineering + Applications, San Diego, 2019;

SELFSUSTAINED CROSS-BORDER CUSTO**M**IZED
CYBERPHYSIC**A**L SYSTEM EXPE**R**IMEN**T**S
FOR C**A**PACITY BUI**L**DING AMONG EUROPEAN STAKEHO**L**DERS

SMART4ALL

An extensive network of Digital Innovation Hubs for boosting technology and business development in South, Eastern and Central Europe

Nikolaos Voros, *Professor*, SMART4ALL Coordinator
Christos Antonopoulos, *Assistant Professor*, SMART4ALL Technical Manager
Georgios Keramidas, *Assistant Professor*, SMART4ALL Technical Manager

Co-funded by the Horizon 2020 programme of the European Union

DT-ICT-01-2019
Smart Anything Everywhere Area 2

www.smart4all-project.eu
Grant Agreement: 872614

SMART4ALL ID Card

Funded by the Horizon 2020 Framework Programme of the European Union

DT-ICT-01-2019: *Smart Anything low energy computing Everywhere* – *Area 2: Customized powering CPS and the IoT*

Project information

SMART4ALL

Grant agreement ID: 872614

Status
Ongoing project

Start date
1 January 2020

End date
31 December 2023

Funded under:
H2020-EU.2.1.1.

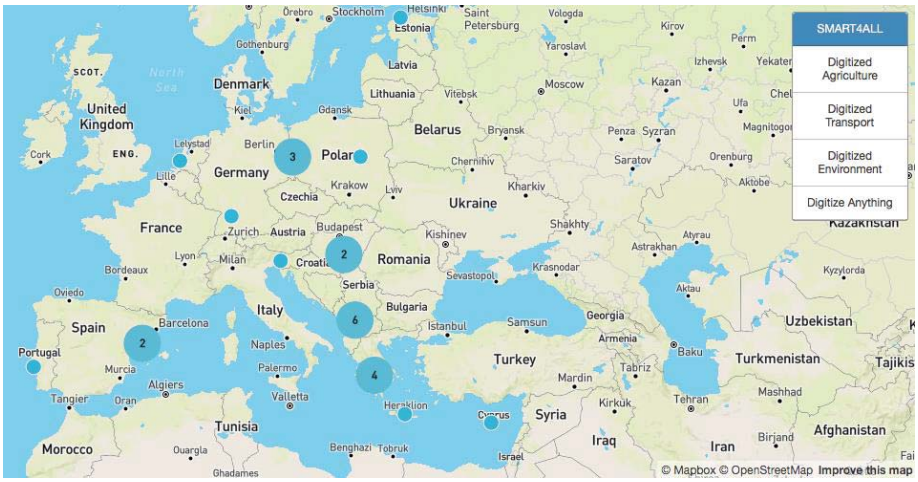
Overall budget:
€ 8 660 872,50

EU contribution
€ 7 997 647,50

Coordinated by:
UNIVERSITY OF PELOPONNESE
Greece

SMART4ALL Consortium

- 25 partners originated from **South**, **Eastern** & **Central Europe**
- The consortium is composed of universities, research institutes, investors, networking organizations, SMEs, Innovation Hubs and NGOs



Diversity in Partners Expertise	Geographic Diversity (15 countries)	
Universities: 10	Albania: 1	Montenegro: 1
Research institutes (R&D): 3	North Macedonia: 2	Poland: 2
Research institutes (BizDev): 1	Cyprus: 1	Portugal: 1
SMEs (R&D): 6	Estonia: 1	Serbia: 1
SMEs (BizDev): 2	Germany: 3	Slovenia: 1
Innovation Hubs: 1	Greece: 5	Spain: 1
NGOs: 1	Hungary: 1	The Netherlands: 1
	Kosovo: 2	

SMART4ALL Partners

University of Peloponese
Greece

German Aerospace Center
Germany

Brandenburg University of Technology
Germany

Patras Science Park
Greece

Faculty of Technical Sciences University of Novi Sad
Serbia

AVN Innovative Technology Solutions Limited
Cyprus

MECOnet d.o.o.
Montenegro

PRAXI Network
Greece

Sensing & Control Systems S.L.
Spain

Delft University of Technology
The Netherlands

FundingBox Accelarator
Poland

Metropolitan Tirana University
Albania

Tallinn University of Technology
Estonia

South East European University
North Macedonia

Marseco
North Macedonia

University "Ukshin Hoti" Prizren
Kosovo

Budapest University of Technology and Economics
Hungary

Rezos Brands
Greece

FastTrack Ventures
Portugal

Leibniz Institute for Agricultural Engineering and Bioeconomy
Germany

DATAPROGNET SH.P.K.
Kosovo

Vocational Training Centre MARGARITA
Greece

RED PITAYA D.D.
Slovenia

POZNAN UNIVERSITY OF TECHNOLOGY – POZNAN
Polland

Polytechnic University of Valencia
Spain

What are SMART4ALL goals?

To build **cross-border experiments** that **transfer knowledge** and **technology** between **academia** and **industry** in **Customised Low-Energy Computing** (CLEC) for **Cyber-Physical Systems** (CPS) and the **Internet of Things** (IoT)

To **accelerate digital transformation** and **increase digital skills** in underrepresented geographical areas –especially **SEE**– in the 4 SMART4ALL verticals

5

SMART4ALL Management Team

Nikolaos Voros

Project Coordinator

University of Peloponese

Greece

Michael Huebner

Project Sub-coordinator

Brandenburg University of Technology

Germany

Christos Antonopoulos

Technical Manager

University of Peloponese

Greece

Georgios Keramidas

Technical Manager

Aristotle University of Thessaloniki

Greece

SMART4ALL Officers

George Dimitriou
Innovation Officer
FORTH, PRAXI Network
Greece

Sónia Magalhães
Business Development Officer
FASTTRACK Action
Portugal

Dimitris Tourlidas
Ethics Officer
Vocational Training Centre MARGARITA
Greece

Tanya Politi
Communication Officer
Patras Science Park
Greece

Vicky Tomara
Communication Officer
Patras Science Park
Greece

Katerina Labrakopoulou
Communication Officer
Patras Science Park
Greece

SMART4ALL in a Nutshell

- **Vision:** To build capacity amongst European stakeholders by joining different cultures, different policies, different geographical areas and different application domains
- **How:** Through the development of selfsustained, cross-border experiments that transfer knowledge and technology between *academia* and *industry*
- **Technological Areas:** Customized low energy computing (CLEC) & IoT
- **Application Areas:** Digitized Environment, Digitized Agriculture, Digitized Anything and Digitized Transport

SMART4ALL PAE Types

PAE
Pathfinder Application Experiments

KTEs
Knowledge Transfer Experiments

Funding: 8K Euros
Type: Internship

FTTEs
Focused Technology Transfer Experiments

Funding: 80K Euros
Type: Project

CTTEs
Cross-domain Technology Transfer Experiments

Funding: 80K Euros
Type: Project

- **Internal**: 21 FTTE (vertical) will be developed by SMART4ALL
- **External:** 67 KTEs/FTEs/CCTEs will be developed via open call funding

Who can apply?

ACADEMIC

Universities and other **Academic Institutions**

INDUSTRIAL

SME and Slightly Bigger Companies (< 500 employees & < EUR 100M turnover)

System Integrators and **Technology Providers** specialised in technology transfer or system integration to End-users (provided they can be categorised in one of the two previous types of beneficiaries)

Must be a **consortium of 2 entities** and led by the **Industrial** partner

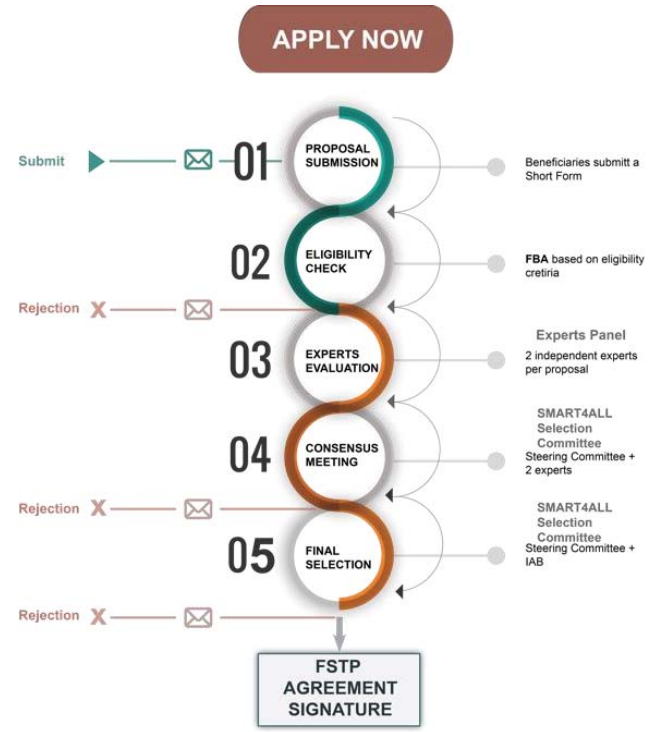
Must be **cross-border** from 2 different **eligible** countries

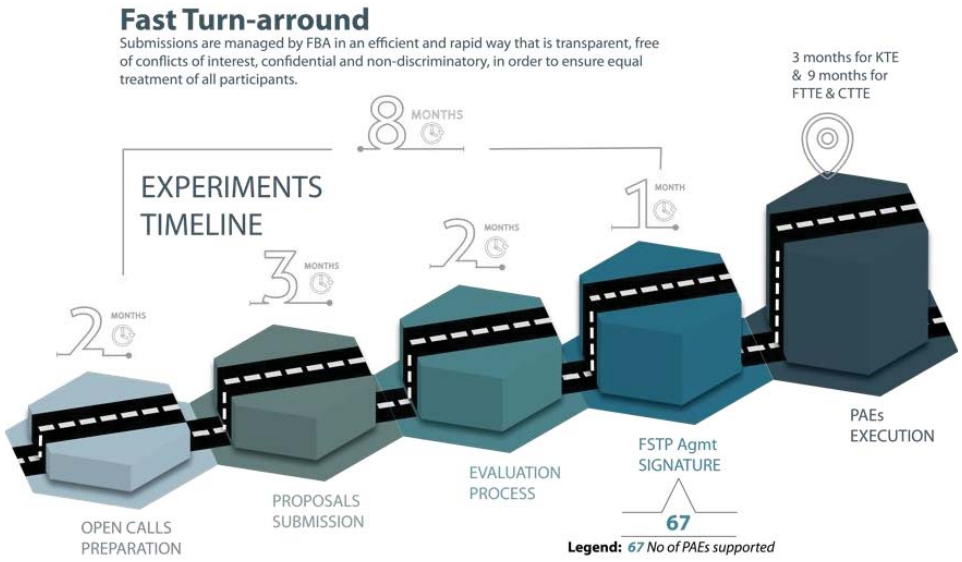
10

SMART4ALL PAE Cut Off Dates

PAE Type		Call 1	Call 2	Call 3
Knowledge Transfer Experiments (KTE)	Call Announcement	Apr 15th, 2020	Mar 2021	Mar 2022
	Submission Deadline:	Jul 15th, 2020	May 2021	June 2022
Focused Technology Transfer Experiments (FTTE)	Call Announcement	Jul 1st, 2020	Jun 2021	Jun 2022
	Submission Deadline:	Sep 30th, 2020	Aug 2021	Aug 202
Cross Domain Technology Transfer Experiments (CTTE)	Call Announcement	Dec 2020	Sep 2021	Sep 2022
	Submission Deadline:	Feb 2021	Nov 2021	Nov 2022

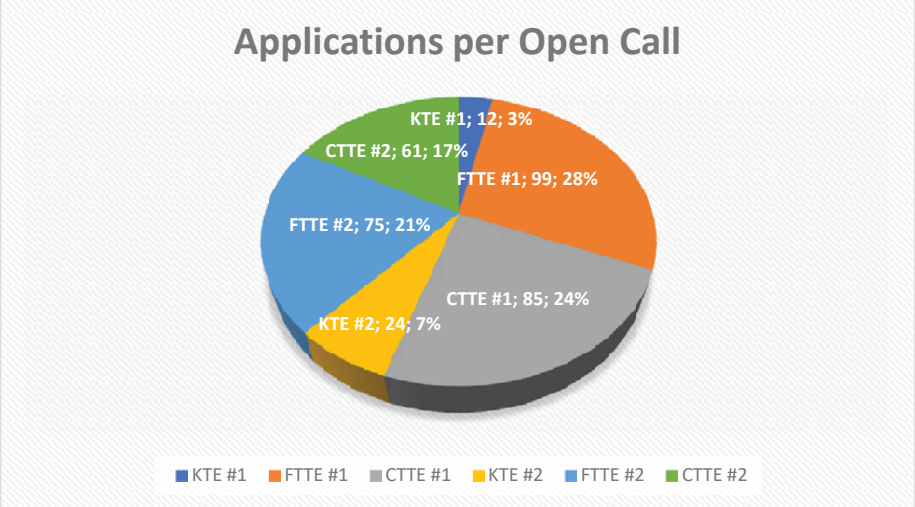
SMART4ALL PAE Application Stages





Open Call Statistics (4 open calls)

Applications per call



356 proposals

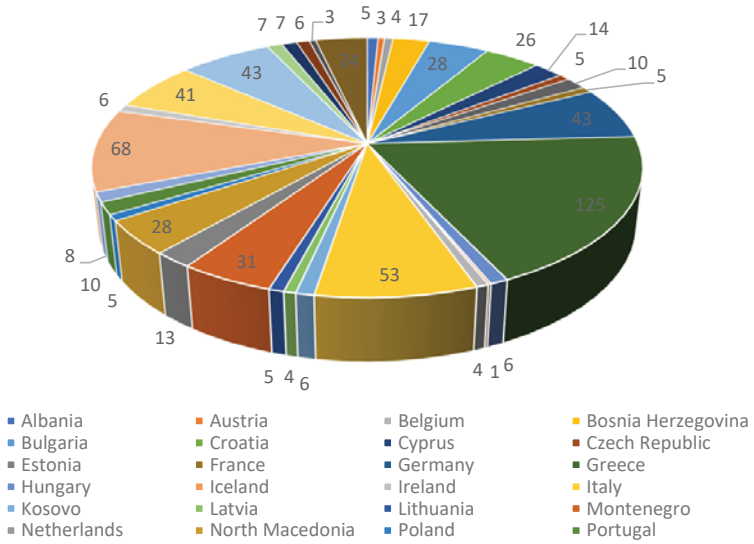
4 February 2021

www.smart4all-project.eu

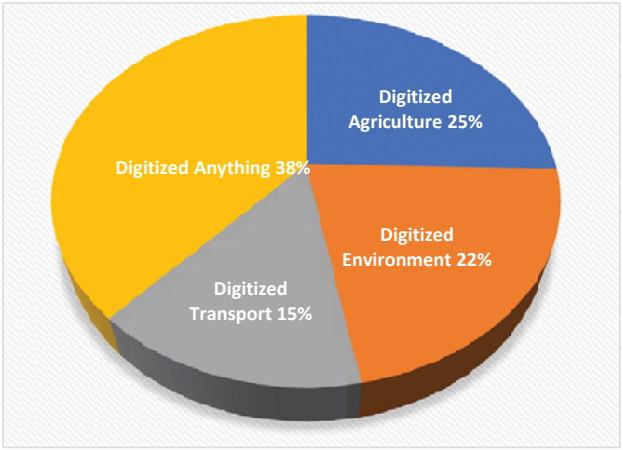
Open Call Statistics (4 open calls)

Applications by Country and Vertical

COUNTRIES (35 in total)



- Albania
- Austria
- Belgium
- Bosnia Herzegovina
- Bulgaria
- Croatia
- Cyprus
- Czech Republic
- Estonia
- France
- Germany
- Greece
- Hungary
- Iceland
- Ireland
- Italy
- Kosovo
- Latvia
- Lithuania
- Montenegro
- Netherlands
- North Macedonia
- Poland
- Portugal



VERTICALS

Application Submission

How to apply?

https://smart4all.fundingbox.com

Find our **Guides & Documents** here: https://smart4all.fundingbox.com

If you cannot find the answer you need in the **FAQ**, you can submit your question(s) through Q&A Space at SMART4ALL online community **Helpdesk**: https://spaces.fundingbox.com/spaces/smart4all-helpdesk or email: helpdesk@smart4all-project.eu

15

SMART4ALL Task Forces

Digitized Transport Task Force (DTTF)	Digitized Argriculture Task Force (DAgTF)
Leader: DLR Members: TalTech, PUT	Leader: ATB Members: REZOS BRANDS, FTN
Digitized Environment Task Force (DETF)	**Digitized Anything Task Force (DATF)**
Leader: S&C Members: PSP, MTU	Leader: BME Members: UPZ, MARGARITA.VTC, UPV
SMEs Task Force (STF): SEEU (Leader), AVN, MAR, DPN, RP	

- The roles of Task Forces are to:
  - ✓ **Promote** the SMART4ALL vision in each thematic pillar
  - ✓ **Build** the SMART4ALL ecosystem
  - ✓ **Manage** the corresponding business generation activities
- **Overall:** to customize SMART4ALL activities and formulate the MaaS services to each thematic pillar
- **Role:** to provide feedback to the project from the start-up and SME point of view.

SMART4ALL Task Force Leaders

Umut Durak

Task Force Leader for Digitized Transport

German Aerospace Center

Germany

Cornelia Weltzien

Task Force Leader for Digitized Agriculture

Leibniz Institute for Agricultural Engineering and Bioeconomy

Germany

Alberto Fernández

Task Force Leader for Digitized Environment

Sensing & Control

Spain

Tamas Kovacshazy

Task Force Leader for Digitized Anything

Budapest University of Technology and Economics

Hungary

Adrian Besimi

Task Force Leader for SMEs

South East European University

North Macedonia

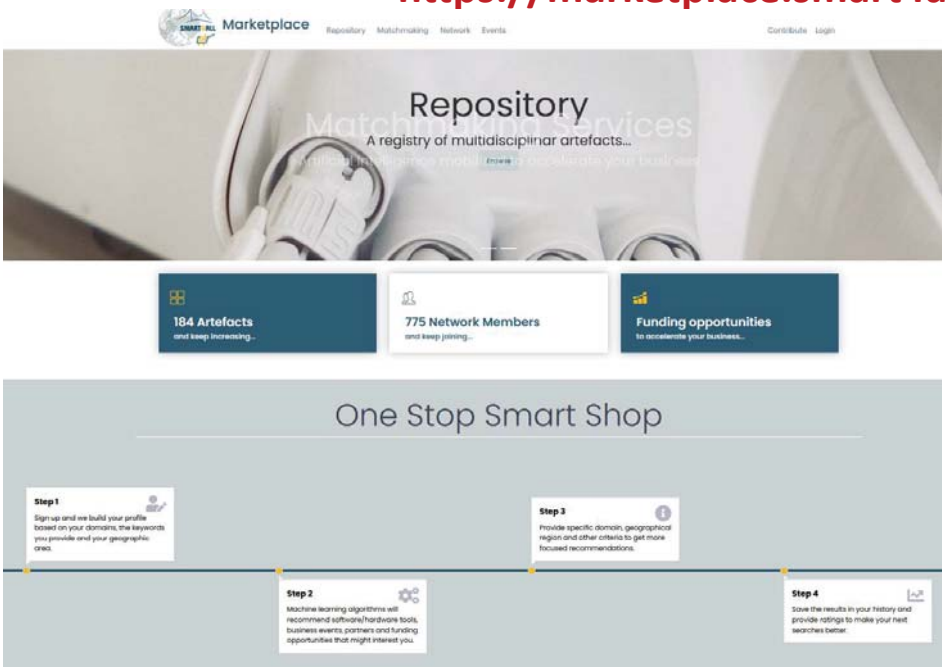
SMART4ALL Marketplace

Marketplace as a Service Concept is one the cornerstones/flagships of Smart4ALL

- **What is it ?** A novel one-stop-smart-shop for experts and non-expert third parties seeking (open-source mainly) ICT technologies
- **What is its main target ?** reduce the development time of an startup /SME/mid-cap that is doing business in one of the four SMART4ALL thematic areas
- **Unique selling point:** AI based match-making and collaboration activities will be hosted in the MaaS

The MarketPlace SW infrastructure

https://marketplace.smart4all-project.eu



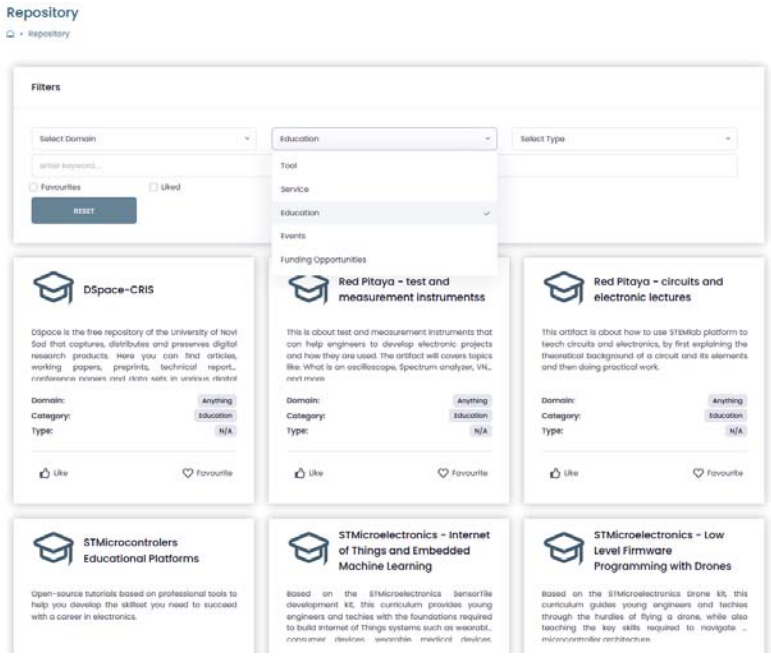
Marketplace follows the **microservices** paradigm where all the functionality is given as a set of **loosely coupled services powered by containerization technology**

MaaS includes:

- Open and proprietary cloud services
- Open and proprietary computing and communication platforms
- Open-source tools and middleware frameworks
- Open-source design tools Training (model-based) open course

SMART4ALL Tools – Repository

https://marketplace.smart4all-project.eu

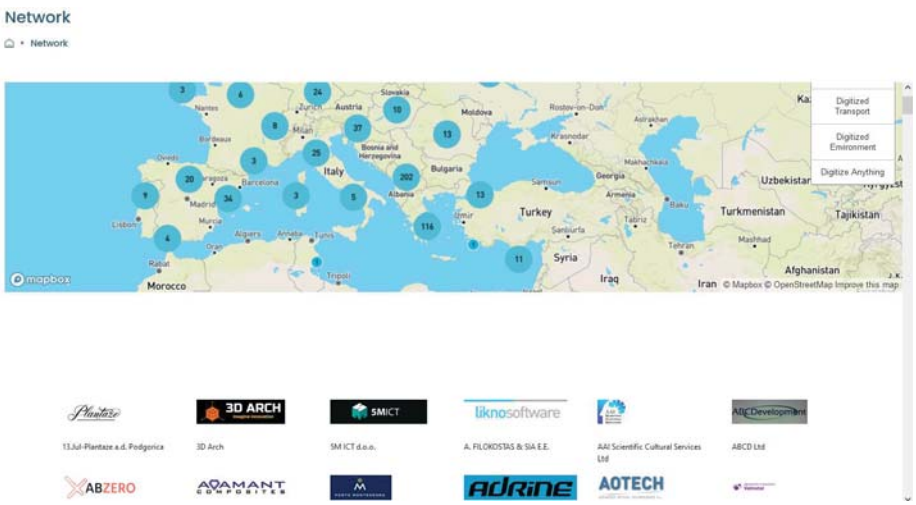


Repository goal:

- Offer multifaceted artefacts (either directly or through links)
- Advertise and promote the entities behind the artefacts
- Increase the self-sustainability of the artefacts and the contributors
- Matchmake the artefacts to the right collaborators and funding opportunities

SMART4ALL Tools – Network

https://marketplace.smart4all-project.eu



Browse through SMART4ALL network by:

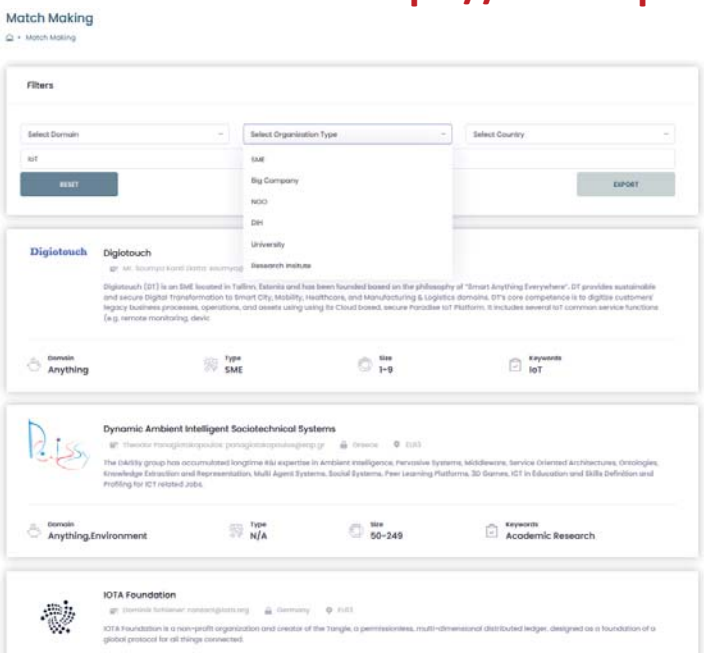
- Vertical
- Location
- Name

SMART4ALL Tools – Events

https://marketplace.smart4all-project.eu

SMART4ALL Tools – Matchmaking Service

https://marketplace.smart4all-project.eu



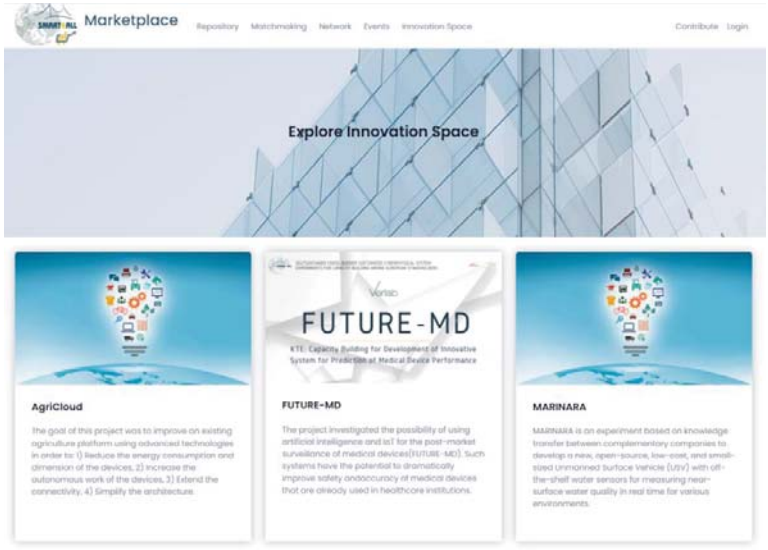
Matchmaking goal:

- Based on filters and keywords suggest suitable candidates from SMART4ALL network
- Offer all the necessary information
- Include at the matchmaking process artefacts
- Use AI to continuously increase accuracy and efficiency

➢ All types of artefacts are considered when matchmaking is triggered

SMART4ALL Innovation Space – Marketplace

https://marketplace.smart4all-project.eu/innovation

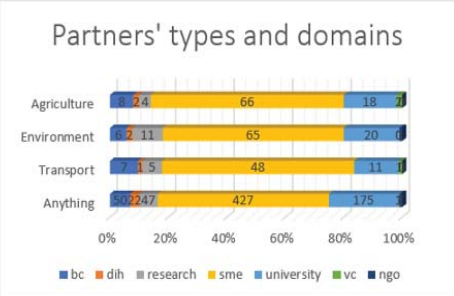


Innovation Space goal:

- Showcasing SMART4ALL beneficiaries
- Highlighting the innovation points of each experiment
- Outlining the market needs that each SMART4ALL project answers to
- Presenting images, videos and infographics of the completed projects
- Offering inspiring examples of cross-border collaboration and synergies on the 4 SMART4ALL application areas

SMART4ALL Tools – Statistics

- A network of 868 members including 490 SMEs, 14 DIHs, 50 RTOs, 160 Universities, 9 NGOs and 62 LEs across Europe
- In the 4-month period from December 2021-March 2022,
  - the marketplace was used 430 times
  - matchmaking services was used 2000 times
  - stable growth throughout the period
- Currently exploring interconnecting with other Marketplaces to increase impact



Top-15 countries with most of the Marketplace's members

Network members and their type & domain

Keywords	Services	Academic Research	Software/ Algorithms	IoT	Hardware	ML/AI	Networks	Cyber-physical Systems	Embedded Computing	Research	Augmented Reality	Civil organization	Food technology	Computing	Signal processing
#	285	140	154	84	59	53	33	24	23	20	12	10	9	10	2

Current Marketplace partners by keyword

SMART4ALL High Performance Computing Center

SMART4ALL offers to the members of its network/ecosystem, via a state-of-the-art **High Performance Computing Center**, the following bouquet of services :

- Software-as-a-Service (SaaS)
- Hardware-as-a-Service (HaaS)
- Scalable architecture to meet high workloads and provide 24/7 availability
- 24/7 support by technology experts
- High speed network interconnection via GRNET backbone
- Open source software employed from virtualization to application layer

SMART4ALL **High Performance Computing Center** is hosted and maintained by

ECE Patras
UNIVERSITY OF PELOPONNESE

ESDA LAB
Embedded System Design & Applications Laboratory

SMART4ALL Infrastructure

ESDA Lab Data Center

8 Rack Mounted High Performance Servers
Total CPUs: **164**
Total Memory: **1.216 GBytes**
Type 1 Hypervisors: VMWARE ESXi / Proxmox Virtualization Technologies

Next Generation Firewalls (HGFW) - Hardware / High Availability Cluster
Double hardware firewalls

Storage Area Network (SAN) - Fibre Channel (FC @ 16 Gbps) Infrastructure
Capacity: **20 TBytes**

Network Attached Storage (NAS @ 1 Gbps) Infrastructure
Capacity: **30 Tbytes**

Specialized High Performance Computing Nodes
FPGA and GPU Cluster Provision

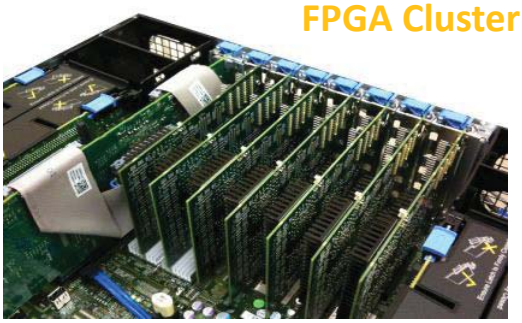
ESDA LAB
Embedded System Design & Applications Laboratory

ECE Patras
UNIVERSITY OF PELOPONNESE

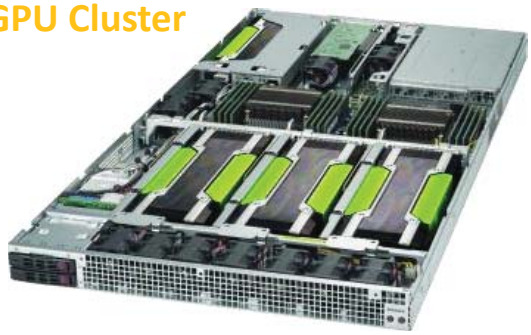
SMART4ALL High Performance Computing Center

High Performance Embedded FPGAs & High-End GPUs

FPGA Cluster



GPU Cluster

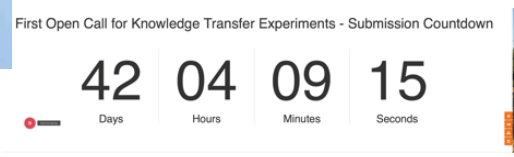


A cluster of high-performance embedded FPGAs and high-end GPUs will be available and offered through the SaaS frameworks

SMART4ALL Tools – Project Website

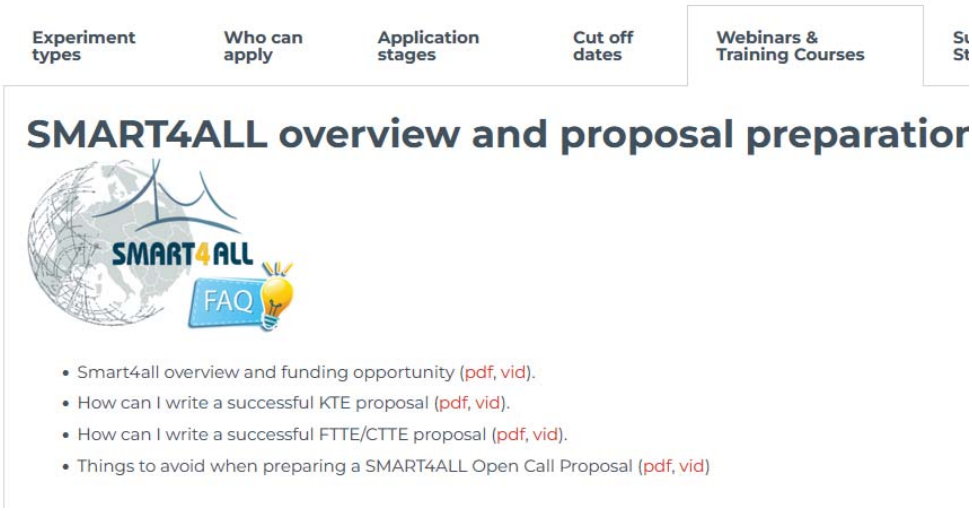
https://www.smart4all-project.eu

- Project news
- Access to SMART4ALL services
- Access to SMART4ALL Network in SEE
- Open call announcement
- Apply for SMART4ALL Open Calls



SMART4ALL Tools – Project Website

https://www.smart4all-project.eu

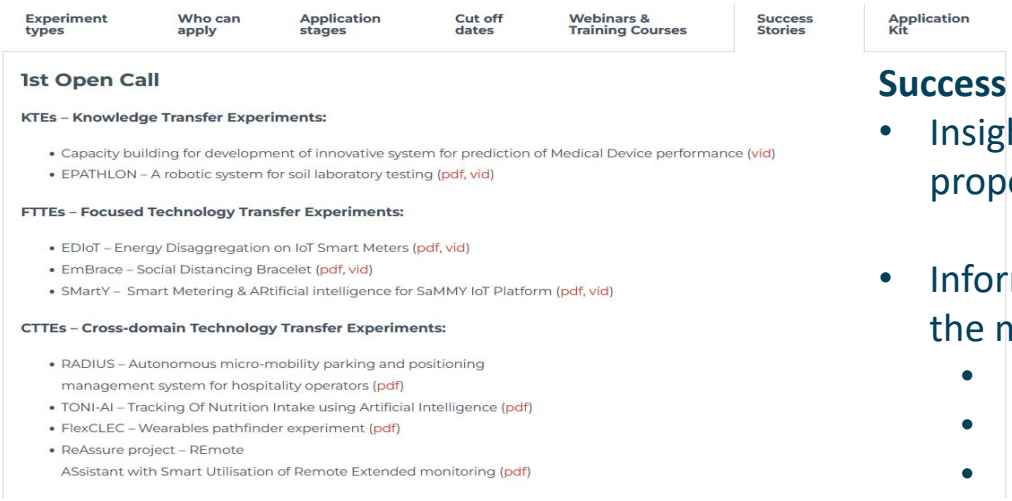


Webinars & Training (pdf & vid)

- Important information regarding Open Calls
- Useful insights on how to prepare a competitive proposal
- Mistakes and shortcoming to avoid based on reviewers' comments

SMART4ALL Tools – Project Website

https://www.smart4all-project.eu



Success Stories

- Insights of successfully prepared proposals
- Information on how to address the main issues of
  - Excellence
  - Workplan
  - Impact

SMART4ALL Dissemination Services (1)

- **Project and Open Calls dissemination**

- **Online through SMART4ALL channels/social media**

LinkedIn: 731 followers

Twitter: 354 followers

Facebook: 716 followers

YouTube: 50 subscribers

MailChimp: 679 subscribers

SMART4ALL Dissemination Services (2) - Webinars

- **International & regional webinars, local satellite events**

7 international webinars → 621 registered participants from 34 countries

Webinar statistics per country

SMART4ALL Dissemination Services (3)
Collaboration with other initiatives

- SAE Initiative
- DIHnet
- HUBCAP
- I4MS
- HiPEAC:

- 4 articles showcasing SMART4ALL-funded Focused Technology Transfer Experiments were featured on 2 issues of the HiPEAC magazine (October 2021 & January 2022)

SMART4ALL Dissemination Services (4) – Major events

- **SMART4ALL 1st Joint workshop with other DIHs and SAE initiatives in the context of MECO2021**

✓**10 european programs**

✓**13 European DIHs**

✓**80+ participants**

✓**38 speakers**

✓**10 SMART4ALL-funded success stories**

✓**5 scientific papers on cutting-edge technologies (AI, IoT, cloud etc.)**

✓**keynote speech of Dr. Hui Cao, Head of Policy and Strategy of Huawei's EU Office**

SMART4ALL Dissemination Services (5) – Major events

- **SMART4ALL special session and pitching event on the 6th SEEDA-CECNSM 2021**

✓40 attendees from different countries

✓7 PAEs presentations on the special session

✓5 PAEs participated on the pitching event

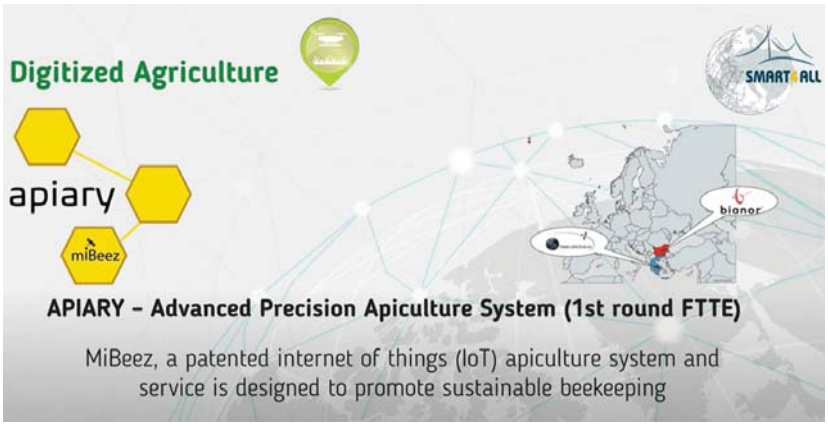
✓4 evaluators

✓1 winner

Cross-domain Technology Transfer Experiment

✓1 keynote speech by Cisco Senior Account manager, Mr. Nikolaos Lambrogeorgos

SMART4ALL success stories – Funded experiments



SMART4ALL success stories – Funded experiments

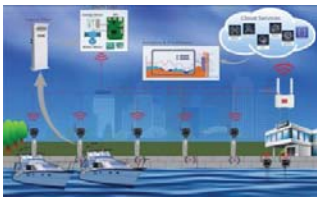
SMART4ALL success stories – Funded experiments

SMART4ALL success stories – Funded experiments

Special highlights and beneficiaries' achievements

- **In the context of Neurofeedback VR, a 2nd round KTE, an experimental platform to investigate the effects of the temporal cortex on the cardiovascular system was developed and one of the partners, Metacognis Institute, submitted a patent application to the Intellectual Property Office in Serbia**

SMartY project (1st round FTTE) won the Golden Award on the 2022 IoT Greek Awards in the Maritime/Cargo Handling category for developing a program which utilises cutting edge technologies and cloud computing to provide marina administrators & yachters automated water and electricity consumption end-to-end electronic services.



A visualized success story

FUTURE-MD

Capacity building for development of innovative system for prediction of Medical Device performance

Digitized Anything

Innovation Space • FUTURE-MD

Join now SMART4ALL Network !

Join now SMART4ALL Network in South, Eastern and Central Europe by:

- Employing SMART4ALL marketplace: https://smart4all-project.eu/marketplace
- Subscribing to SMART4ALL newsletter: https://smart4all-project.eu/joinus
- Applying to SMART4ALL Open Calls: https://smart4all-project.eu/opencalls-apply-now
- Get all information from SMART4ALL Site: https://www.smart4all-project.eu

SMART4ALL Social Media

Follow SMART4ALL on:

https://www.linkedin.com/groups/12369183

https://twitter.com/Smart_4All

https://www.facebook.com/SMART4ALL.Project

https://www.youtube.com/channel/UCwmSl9LCl2vNBO-3k75dvJA

… and stay up to date for the latest SMART4ALL news, activities & funding opportunities !

SMART4ALL Contact persons

Project Coordinator

Prof. Nikolaos Voros

University of Peloponnese, Greece

voros@uop.gr

Central Europe Contact

Prof. Michael Huebner

Brandenburg University of Technology, Germany

huebner@b-tu.de

East Europe Contact

Prof. Tamás Kovácsházy

Budapest University of Technology and Economics, Hungary

khazy@mit.bme.hu

Balkan Contacts

Prof. Radovan Stojanovic

Mediterranean Excellence in Computing and Ontology, Montenegro

stox@ac.me

Prof. Betim Cico

Metropolitan Tirana University, Albania

betim.cico@gmail.com

South Europe Contacts

Prof. Fransisco Blanes

Polytechnic University of Valencia, Spain

pblanes@ai2.upv.es

Mr. George Dimitriou (M.B.A.)

PRAXI Network, Greece

dimitriou@praxinetwork.gr

Thank you for your attention – Questions?

An IoT-enabled Smart Grid: Definitions, Characteristics, Challenges, and Future Directions

Abeer Akkad
Dept. of Computer Science
University of Southampton
Southampton, UK

Gary Wills
Dept. of Computer Science
University of Southampton
Southampton, UK

Abdolbaghi Rezazadeh
Dept. of Computer Science
University of Southampton
Southampton, UK

Abstract—The IoT-enabled Smart Grid (SG) could be viewed as a large Cyber-Physical-System (CPS). The Smart Grid is considered to be part of the vital critical infrastructure for many communities worldwide. Globally, the energy market is considered the biggest market for any country to grow economically and therefore Smart Grids are one of the biggest applications of IoT. The evolution of an Internet of Things-enabled Smart Grid affords better automation, communication, monitoring, and control of electricity consumption. It is now essential to supply and transmit the data required, to achieve better sensing, more accurate control, wider information communication and sharing, and more rational decision-making. However, the rapid growth in connected entities, accompanied by the increased demand for electricity, has resulted in several challenges to be addressed. This research helps better depicts and understand the challenges of Smart Grid implementation. Identifying the challenges is considered a pre-requisite pillar to understanding whether the applied implementation and design approach can be employed to foster the IoT-enabled Smart Grid. To the best of the author's knowledge, no previous study addresses sufficiently and comprehensively the challenges of the IoT-enabled Smart Grid.

Keywords— Smart Grid, Smart energy, IoT, Internet of Things, IoT - enabled Smart Grid, Cyber-Physical-System, CPS.

I. INTRODUCTION

As cities transform into smart cities, it is essential that they have sustainable green energy, and the implementation of the IoT-enabled Smart Grid is considered a way to achieve this goal. According to a study conducted by McKinsey Global Institute, the IoT will have a significant economic contribution from $3.9 to $11.1 trillion per year by 2025 (Manyika *et al.*, 2015) influencing (homes, factories, retail environments, offices, worksites, human health, outside environments, cities, and vehicles). The electric utility industry is currently developing an IoT-enabled Smart Grid (SG) which is envisioned as the largest installation of an IoT system, with billions of smart objects and things, such as smart meters, smart appliances, and other sensors (Reka and Dragicevic, 2018)(Reka and Dragicevic, 2018). This huge number of connected devices besides the increasing demand for electric energy results in many significant challenges that may face the Smart Grid. Although the Smart Grid could address the drawbacks of the traditional power system, it also contained some challenges of security, big data processing, cost, centralization, scalability, interoperability, heterogeneity, and latency. This research discusses the existent challenges of the IoT-enabled Smart Grid which could help practitioners and engineers in the energy sector for better implementation of the IoT-enabled Smart Grid.

This paper is organised as follows: Section II defines the Smart Grid, its importance, characteristics, and components. Then, it describes the link between IoT modules and Smart Grid. In section III the challenges of the IoT-enabled Smart Grid are investigated. Section IV looks at the future work directions in the context of IoT and Smart Grid. Then, the work is concluded in section V.

II. BACKGROUND

This part of the paper offers an overview of the IoT-enabled Smart Grid definitions, characteristics, components, as well as its benefits. Moreover, the role of IoT in the Smart Grid is explained.

A. Definition of Smart Grid

The many definitions of SG vary between organisations and studies, as shown in Table 2-1, and there is no agreement; however, the common concept is that SG revolves around an information communication infrastructure. For instance, in the definition by the largest standardisation authority, IEEE, the SG describes a new age of electricity that features the use of ICT in the generation, delivery, and consumption of electricity and the electric system (IEEE, 2018). Likewise, the viewpoint of the Ontario Independent Electricity System Operator (IESO), the leader in SG, is that it involves the use of ICT in optimising all power system operations for the benefit of the consumer and the environment (Singer, 2009). Both definitions focus on the SG component, which is specifically the communication infrastructure, whereas others focus on the outcomes that benefit from SG. For instance, the Energy Independence and Security Act of 2007 (EISA, 2007) produced the first official definition of SG (US Public Law, 2007; Al Khuffash, 2018), as given in a report to the US Congress: "*The modernization of the Nation's electricity transmission and distribution system to maintain a reliable and secure electricity infrastructure that can meet future demand growth and to achieve a set of requirements that together characterize the SG*" (US Public Law, 2007;

U.S. Department of energy, 2018). It is worth pointing out that this definition describes SG from the perspective of its benefits. Within the IEEE and EISA definitions, SG domains are prominent, including electricity generation, transmission, distribution, and consumption (US Public Law, 2007; IEEE, 2018).

In the context of information technologies, other definitions focused on how information could be transferred through the SG. The bi-directional flow has given rise to the term "prosumers" in SG (Dalipi and Yayilgan, 2016), meaning customers who generate energy for the grid, as stressed by the European Union's viewpoint as well as the UK Institution of Engineering and Technology (IET, 2013), also shown in Table 2-1. The IET's definition of SG is based on that of the ETP (IET, 2013). From an environmental perspective, both Singer (2009) and the Electric Power Research Institute (2005) mention green energy and the environmental impact of SG in their definitions as the most important advantages of SG due to their contribution to a reduction in the CO_2 footprint (EPRI, 2005; Singer, 2009).

From the above, the SG can be defined as the integration of ICT into the existing electrical network, consisting of renewable sources and involving its multiple domains (generation, transmission, distribution, and consumption) in the efficient automation and real-time demand management of a reliable, sustainable, bi-directional, and economic green electrical energy.

Table 1. A summary of some Smart Grid definitions.

Organisation	Definition
IEEE	Smart Grid describes a new age of electricity that features the use of Communications and Information Technology (CIT) in the generation, delivery, and consumption of the electrical system. (IEEE, 2018)
DOE/EISA (US Dept of Energy)	The modernization of the Nation's electricity transmission and distribution system to maintain a reliable and secure electricity infrastructure that can meet future demand growth and to achieve a set of requirements that together characterize a Smart Grid. (U.S. Department of energy, 2018)
IESO (Independent Electricity System Operator)	Smart Grid is the employment of ICT in optimizing all power system operations for the benefit of the consumer and the environment. (Singer, 2009)
ETP (European Union)	Smart Grid is developed by the European Technology Platform, and it means the smart integration of all operations from the connected producer, consumers, and prosumers to supply sustainable, and secure power energy. (ETP, 2006)
EPRI (Electric Power Research Institute)	A Smart Grid is one that incorporates information and communications technology into every aspect of electricity generation, delivery, and consumption in order to minimize environmental impact, enhance markets, improve reliability and service, reduce costs, and improve efficiency. (EPRI, 2005)

There are two flows in the IoT-enabled Smart Grid:

- **Electricity flow** is the classic flow in a conventional electrical grid from generating stations to consumers, while in SG this flow is bi-directional (Bekara, 2014).
- **Information flow** is a bi-directional flow between utilities and all components of the SG, including smart meters, sensors, actuators, smart appliances, and electric vehicles. Consequently, this flow is a real-time Big Data flow, owing to the increase in the number of connected devices on the SG (Bekara, 2014).

B. Why IoT-enabled Smart Grid?

Decarbonisation has become a goal worldwide for all countries, to address climate change and limit global warming by reducing CO_2 emissions (Colak, 2016; Reka and Dragicevic, 2018). Over time, the exponential growth of demand and the variety of loads have become a burden on the electricity grid (Al Khuffash, 2018). For instance, electric vehicles require charging, and the United Kingdom is investing £30 million in supporting this charging infrastructure (Jenkins *et al.*, 2015). Consequently, there is a strong probability that the grid will be overwhelmed by increasing demand for electricity. Then, costs will rise due to operational expense and latency (Al Khuffash, 2018). Thus, Colak (2016) emphasised that the rapid growth in demand for electricity, and the variety of loads, must be managed and planned efficiently to secure cost containment or reduction.

In addition, transmission and distribution lines experience both losses and unauthorised consumption (Colak, 2016; Al Khuffash, 2018), so inevitably there has arisen a need to be smarter with the electricity grid to manage and monitor consumption effectively and to ensure power availability. The electricity supply could be managed efficiently by increased standardisation of the information system between utilities and consumers in the SG (DeBlasio and Tom, 2008; Sortomme *et al.*, 2011; Colak, 2016). In a SG, more monitoring and control could regulate power generation to respond to demand. By contrast, in a conventional power grid, the traditional meter readings provide insufficient information on grid conditions and consumption, with no real-time energy information (Al Khuffash, 2018).

Consequently, consumers have no data on their usage, which, in turn, leads to rising costs. In addition, the centralised architecture of the conventional grid may represent a burden on its productivity. Therefore, the SG is considered to have the potential to solve the drawbacks of the old infrastructure on a conventional grid (Ghasempour, 2016).

Both Ghasempour (2019) and Al Khuffash (2018) argued that SG is essential to enable and integrate all the renewable energy sources in the system, such as solar, hydro, and wind. The SG can both handle the variability, and counterbalance the constant fluctuations in wind and solar.

From the above, the reasons why cities need an IoT-enabled SG can be summarised as: the SG ensures controllable automation, the integration of renewables, sustainable green energy solutions, and real-time awareness (IET, 2013; Colak, 2016; Kaur and Kalra, 2016). As a result, the development of the SG is looked on as the infrastructure to overcome the challenges of rising carbon emissions, rapid growth in demand, overloading, latency, transmission losses and outage, real-time information inadequacy, a centralised old architecture, and integrating the multiple forms of green energy.

C. *IoT-enabled SG Characteristics*

The following 10 points characterise the IoT-enabled SG. This step will contribute to developing a proper model by assuring the functionality need to be accomplished in each characteristic. The characteristics issued by NIST and commonly used in the sector (NIST, 2014; U.S. Department of energy, 2018):

1. To increase the usage of ICT to enhance the reliability and efficiency of the power grid.
2. To optimise the operations and resources of the grid, with full cybersecurity.
3. To integrate distributed resources and generation, such as those of renewable resources.
4. To incorporate the demand response, demand-side resources, and energy-efficiency resources.
5. To deploy smart technologies such as real-time, automated, interactive technologies that improve the physical operation of appliances and consumer devices for metering, communicating, and reporting grid status.
6. To integrate smart appliances and consumer devices.
7. To integrate the advanced electricity storage and peak sharing technologies, including plug-in electric and hybrid electric vehicles and thermal-storage air conditioning.
8. To provide timely information and control services to consumers.
9. To standardise the communication and interoperability of appliances and devices connected to the power grid with the grid's infrastructure.
10. To reduce unnecessary obstacles to the adoption of SG technologies, practices, and services.

D. *IoT Modules and SG*

This section presents the background of the Internet of things (IoT) and IoT devices. To address the research goals, it discusses the main IoT modules of IoT devices. IEEE defines IoT as the integration of things, which are equipped with sensors, via the internet. The ITU Telecommunication Standardisation Sector (ITU-T) considers the IoT system as an infrastructure for information systems that connect physical and virtual entities. Cisco (2013) gave a definition for IoT as 'the Internet of Everything', with the ability to gather people, data, and things to construct a network capable of exchanging information (Cisco, 2013). Hewlett-Packard states that IoT is a system in which every object is connected to the internet. Shakerighadi et al. (2018) defined IoT as infrastructure, including sensors, communication systems, information systems, and objects connected to the internet, which are essentially standardised.

According to Rathke and Sassone (2010), an IoT device consists of the five main modules, shown in Figure 2-2: (1) a sensing module, (2) a processing module, (3) an actuation module, (4) a communication module, and (5) an energy module. These are supported by storage and applications.

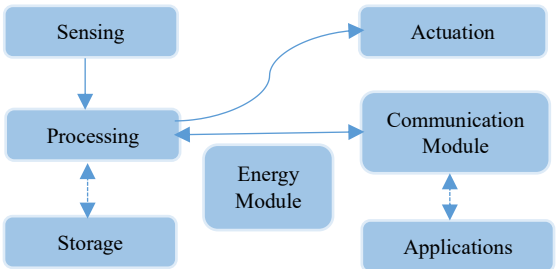


Figure 1: Main modules of an IoT device (Rathke & Sassone, 2010)

Mugunthan and Vijayakumar (2019) supported the claim that IoT technologies have afforded SG with the cloud, 5G, mobile wireless networks, application programming interfaces (APIs), machine learning, AI, predictive analytics, and Big Data management.

In SG, in the context of IoT, each device is connected to the internet. To facilitate communicating information and receiving control commands via the internet protocols, each must have a unique IP address (Al-Ali and Aburukba, 2015; Saleem *et al.*, 2019). Under the IP addressing schemas, IoT can offer **monitoring and control capabilities** for SG, as discussed by Kaur and Kalra (2016). This monitoring aspect can cover the generation plan, distribution, storage, and consumption to achieve efficiency management, demand management, renewable energy needed measurement, and CO_2 emissions administration. Therefore, IoT devices contribute to the reduction of wasted energy and the accurate estimation of required energy.

Further, those devices exchange data in bi-directional flow via the SG communication layer, using several communication protocols, such as Wi-Fi, Zigbee, WiMax, LET, and GPRS. IoT standardises communication, reducing the number of these protocols relating to the SG components

(Al-Ali and Aburukba, 2015). Both Saleem et al. (2019), and Al-Ali and Aburukba (2015), emphasised that IoT technologies enable SG to **communicate** across all its multiple subsystems of generation, transmission, distribution, and consumption. Al-Ali and Aburukba (2015) stated that each device can exchange data and commands from the control centres and utilities.

Both Kaur and Kalra (2016) and Al-Ali and Aburukba (2015) suggested that all objects in a SG can be represented as IoT devices distributed throughout the residential network, substations, and utilities. For instance, these devices could be:

- Smart home appliances with electric vehicle charging
- Substation devices (smart meters, actuators, circuit breakers, transformers, switches, routers, concentrators, voltage regulators, capacitors, or cameras)
- Renewable energy sources
- Utility and control data centres (servers or testing devices)

The conventional power grid relies on SCADA systems, which are built with a centralised architecture (Yang, 2019). Utilising IoT in such systems will increase their scalability, efficiency, and availability; however, there are serious risks. Similar to devices, on the control centre side, SCADA has its own IP address. Classical SCADA systems are renowned for having no proper security controls, in part because they were never designed to be open to the internet. With the integration of complex new architectures such as IoT, therefore, in deploying SCADA systems on the internet, security is an issue (Sajid *et al.*, 2016), especially since several types of malware have recently targeted SCADA systems owing to this lack of built-in security (Pour *et al.*, 2017; Kimani *et al.*, 2019). Centralisation and security issues are discussed in the challenges section.

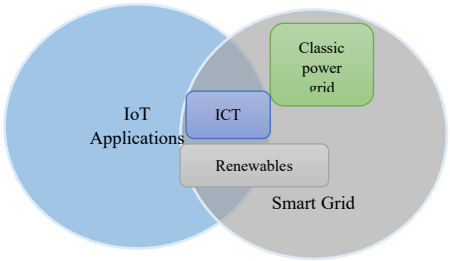


Figure II-2: SG and IoT

In short, SG is considered as one of the biggest applications of IoT (Bekara, 2014; Reka and Dragicevic, 2018; Al-Turjman and Abujubbeh, 2019), as shown in Figure 2-3. In other words, IoT technology is subsumed under the wider umbrella of SG. However, many studies considered IoT as a technology separate from the Smart Grid. From the cyber-physical systems point of view, this research considers IoT as part of the SG itself, enabling all those features that are discussed here.

According to the IoT Security Foundation (IoTSF, 2020), the term **Consumer IoT (CIoT)** concerns consumer usage, while **Industrial IoT (IIoT)** is about industrial purposes, including manufacturing, supply chains, monitoring, and controlling. Consumer IoT and Industrial IoT are discussed from the security viewpoint in the next section.

E. Smart Grid components

By combining the views of the US DOE (2018), Al Khuffash (2018), and Bekara (2014), it can be seen that SG comprises two major elements: smart devices and Advanced Metering Infrastructure (AMI) (Bekara, 2014; Al Khuffash, 2018), although some studies consider smart meters to be a part of AMI (Mohassel *et al.*, 2014; Mrabet *et al.*, 2018).

1) Smart Devices

represent the physical infrastructure of the SG and include smart meters, smart appliances, sensors, phasors, measuring units, and circuit breakers. Smart meters are digital meters consisting of a microprocessor and local memory, and they represent the fundamental blocks with which to build a SG (Rahman, 2009). They measure and collect energy consumption data with a timestamp, which is crucial to delivering electricity in a reliable manner. Also, on the utility side, smart meters transmit data in real-time to the AMI. The smart meters are installed on the consumer side and at other locations around the SG, and report information annually, monthly, daily, hourly, or even each second, for the purpose of management and control. Smart meters record other information, such as voltage and current for both consumers and utilities, due to their two-way capability. From the consumers' perspective, smart meters raise consumption consciousness by informing them of their average usage, advising them of peak demand times, and alerting them when a specific usage limit is reached. Therefore, smart meters can contribute to an energy-efficient economy and energy conservation to manage the rapid growth in demand (Bekara, 2014; Al Khuffash, 2018; Ghasempour, 2019).

From the utility perspective, smart meters enable monitoring and the detection of power theft. They provide failure/shortage notifications, as well as real-time overviews on grid status to support decision-making on electricity generation, distribution, load balancing, and scheduling. Moreover, they assure a swift response to any controlling commands, including shortage management, software upgrade, on/off turns, and pricing systems. They enhance the planning process by capturing the information so that, with sophisticated analysis, utilities can predict future usage and demand patterns (Flick and Morehouse, 2011; Bekara, 2014; Al Khuffash, 2018; Ghasempour, 2019).

2) Advanced Metering Infrastructure (AMI)

like a smart device, enables two-way communication between smart meters and utilities. Before the AMI, automatic meter readings (AMR) allowed only unidirectional communication, from smart meters to utilities (Ghasempour, 2019; Martins *et al.*, 2019). AMI collects, analyses, measures, and stores the energy data sent by the consumer's smart meter to the utility's information management systems. The AMI transmits requests, command signals, notifications, recommendations, pricing information, and software updates from the utilities back to the consumer's smart meter (Bekara,

2014; Mohassel *et al.*, 2014; Al Khuffash, 2018; Ghasempour, 2019). It consists of three elements: (i) a smart meter; (ii) the AMI headend; and (iii) concentrators or collectors.

On the utility side, the **AMI headend** is an AMI server that includes meter data management system (MDMS). The communication with the smart meters is established using communication protocols such as Zigbee and Z-wave (Mrabet *et al.*, 2018).

3) Communications network

The communications network aims to enable data sharing and exchange between IoT smart devices and the utility side (U.S. Department of energy, 2018). It includes the network itself and transmission and distribution devices such as switches, voltage regulators, capacitors, and transformers (PTI, 2011; Al Khuffash, 2018). The network collects information from smart meters and transmission and distribution devices to aid in diagnosing and monitoring network status, thereby providing supply distribution (Gungor *et al.*, 2010). The communication network is standardised by IoT to reduce the number of protocols that have to be used to communicate. Al-Ali and Aburukba (2015) proposed the 6LowPAN communication protocol as the backbone of the IoT communication layer in SG. SG employs ICT with a centralised architecture (Al-Omar *et al.*, 2012; Al-Ali and Aburukba, 2015; Yang, 2019). According to the DOE (2018), ICT is what makes the grid smart. As a CPS, SG uses ICT to monitor, manage, and control its processes and physical assets, including substations, transformers, circuit breakers, smart meters, and cables (Khan *et al.*, 2017). Thus, ICT is the most important characteristic of a SG, and it is the key factor in designing IoT systems.

4) The supervisory control and data acquisition system (SCADA)

SCADA is the central system that controls the power grid. SCADA is situated on the control centre side, and is composed of three elements (Mrabet *et al.*, 2018):

- Remote terminal unit (RTU): a device consisting of three elements used, respectively, for data acquisition, instruction execution for the Master Terminal Unit (MTU), and communication.
- Master terminal unit (MTU): a device that controls the RTU.
- Human-Machine Interface (HMI): a graphic interface for the SCADA system.

5) Information systems

Information systems are essential for processing, computing, analysing, and accessing the data collected from digital devices in the SG. Information systems of SG can be classified into the following systems (U.S. Department of energy, 2018), according to their location (Wang *et al.*, 2019):

- On the generation side, such as Supervisory Information System (SIS) and Demand Response Management (DRM).
- On the transmission side, such as Energy Management System (EMS), Electricity Operation System, and Decision-Making System.
- On the distribution side, such as Data Management System (DMS).
- On the Utility side, such as Customer Information System (CIS).
- On the SCADA side, such as Substations Automation System (SAS).

III. CHALLENGES IN THE IOT- ENABLED SMART GRID

As such, before implementing IoT-enabled Smart Grid, it is vital to research potential challenges and risks that could be faced. Many researchers have considered Smart Grids as the largest part of an IoT framework with billions of smart objects and entities. . Therefore, this section is giving a subset of challenges that filter the challenges of IoT infrastructure with the ones that apply to SG. The challenges that are inherent in the use of IoT.

As a result of the growth in the number of objects connected in the SG, **Big Data processing** becomes an issue (Sagiroglu *et al.*, 2017). AMI in the SG produces Big Data that needs to be handled, stored, and analysed efficiently (Shakerighadi *et al.*, 2018). Smart metering and ICT deployment lead to generating big energy data in terms of volume, velocity, and variety (Hu and Vasilakos, 2016). These data can be exploited to obtain insight, make decisions, predict future consumption patterns, and the required distribution of power supplies (Shakerighadi *et al.*, 2018; Ghasempour, 2019). With sophisticated data analytics, superior monitoring and control can be achieved by the SG. In this context, Big Data could consume huge amounts of energy and other resources when information is collected, transferred, and handled by IoT devices. The SG should thus be designed to deal with the collection of Big Data (Ghasempour, 2019).

The internet is used in SGs for monitoring and control purposes, exposing to attack the information from sensors, smart meters, and other smart devices. Any tampering with the data collected in and from smart meters may cause serious financial loss. Thus, the SG's exposure to the internet could make it vulnerable, rendering its **security** another challenge (Bekara, 2014; Arasteh *et al.*, 2016; Risteska Stojkoska and Trivodaliev, 2017; Ghasempour, 2019).

Ghasempour (2019) and Mahmood et al. (2016) held the view that the implementation of SG should consider the **constrained nature of IoT devices** in computation power and storage capabilities. This, in turn, requires proper security algorithms to meet the limited ability of the IoT devices, so that they are capable of running them (Ghasempour, 2019).

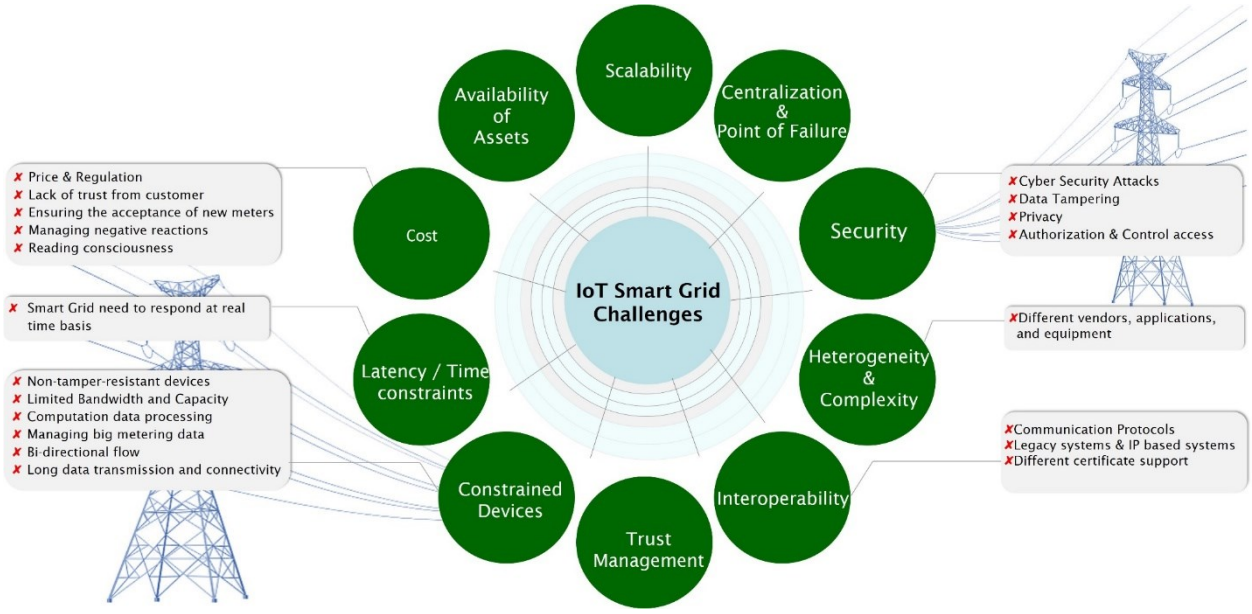


Figure 3. IOT - enabled Smart Grid challenges

The SG is considered one of the biggest applications of IoT, thus **security** is the greatest challenge that it faces, inherent in the use of IoT devices, as there are many security concerns around IoT technologies. As a cyber-physical system, it is argued that security represents a serious challenge for IoT-enabled SG. All studies are similarly concerned about the SG's security (Arasteh *et al.*, 2016; Risteska Stojkoska and Trivodaliev, 2017; Bedi *et al.*, 2018; Reka and Dragicevic, 2018; Shakerighadi *et al.*, 2018; Ganguly *et al.*, 2019; Kimani *et al.*, 2019). As argued by Ghasempour (2019), an attacker could extract private information about prosumers and their consumption. Data values in smart meters could be manipulated. **Trust management and social factors** between parties such as consumers, substations, and utility companies could be violated through IoT devices, so the SG's confidentiality, integrity, and reliability could be affected negatively. As a consequence, the CIA triad of security may be compromised. The mobile nature of IoT devices in SG, such as electric vehicles, and connection stability, are major issues in SG security (Bekara, 2014; Shakerighadi *et al.*, 2018; Mugunthan and Vijayakumar, 2019). Specifically, this involves the identification of IoT devices over the internet, which allows identity spoofing attacks to hack the SG.

Shakerighadi et al. (2018) suggest that supplying sensors with energy may pose a challenge in terms of **cost**. The situation is complicated by the bi-directional flow of data, capacity and bandwidth limitations, smart meter constraints, and the long-distance transmission of data. Costs can be assigned to smart devices, software, staff training, regulations, and managing customer acceptance of smart meters (Bekara, 2014; Shakerighadi *et al.*, 2018; Connor and Fitch-Roy, 2019). From a performance point of view, the SG collects Big Data on a real-time basis from a variety of devices, presenting a significant challenge, since real-time analysis is computationally expensive, and this needs to be taken into consideration, as discussed by Shakerighadi et al. (2018) and Bekara (2014). One explanation is that electricity varies in current, voltage, and frequency; consequently, power fluctuations may cause a power overload or shortage. Similarly, the **centralised** architecture of the SG is a challenge to performance, causing a single point of failure in a power system (Atlam and Wills, 2019). If the node processing the information is attacked and thus unavailable, the whole power system becomes unavailable; however, a decentralised architecture may support power distribution and enhance the system's bandwidth (Al Khuffash, 2018).

An increase in connected smart devices, with their constrained nature, leads to another challenge, that is **scalability,** which causes a bottleneck in SG. Multiple requests may not be processed synchronously, thus increased communication latency could occur and a noticeable delay in serving consumers could be experienced (Mahmood *et al.*, 2016). Since the SG is used to connect many cities in a country, there is a need for a scalable system (Bekara, 2014). Scalability is the adaptability of the SG to expand incrementally, aiming to meet the prospective future rapid growth in electricity demand and to assure clustering and load balancing techniques (Bekara, 2014; Al-Turjman and Abujubbeh, 2019). Al-Turjman and Abujubbeh (2019) considered that scalability plays an essential role in enhancing the power grid's reliability and quality since it affects the availability of this vital asset. Thus, scalability affects security. Furthermore, SG consists of devices from various vendors, applications, services, protocols, and communication stacks, introducing **heterogeneity and complexity** challenges (Bekara, 2014; Arasteh *et al.*, 2016; Bedi *et al.*, 2018). The SG comprises subsystems of power systems, control systems, and communication systems. Furthermore, integrating these subsystems involves issues of information management that need to be considered, since the SG system is a system of systems (Shakerighadi *et al.*, 2018).

From a communication point of view, the challenge of **interoperability** relates to heterogeneity. SG exchanges information among many IoT devices and gateways of varying specifications. By combining the views of Bekara (2014), Risteska Stojkoska and Trivodaliev (2017), Shakerighadi et al. (2018), and Ghasempour (2019), it can be seen that interoperability may be attributed to the heterogeneity of protocols and communication stack, as discussed above. There are many separate standards for IoT devices with no unified standardisation efforts in the SG, causing interoperability issues for IoT devices (Ghasempour, 2019). For example, legacy systems cannot communicate with IP-based systems in the SG due to the lack of support for some protocols, such as TCP/IP (Bekara, 2014; Risteska Stojkoska and Trivodaliev, 2017; Shakerighadi *et al.*, 2018).

From the above, it can be concluded that the security of the SG system is affected by issues of heterogeneity, complexity, interoperability, and constrained devices. Furthermore, it is argued that centralisation may affect the availability of the system. Therefore, it is significant to remember each of these challenges alongside the others in SG implementation: for example, to maintain security it is important to consider correlated aspects.

IV. Future Directions

According to the challenges of Smart Grid that are identified in the last section, a list of Smart Grid requirements could be listed. As described above, it is argued that automation systems such as SCADA were designed without any regard for security (F. A. Aloul, 2012). Moreover, Modbus, which exchanges SCADA information to control industrial processes, was not intended for critical security environments such as SG (F. A. Aloul, 2012). Thus, securing the information system in SG must be assigned the highest priority, since power assets represent critical national infrastructure that may attract terrorists and state actors. Any damage, such as security attacks on the power grid, could cause chaos across whole cities. Electric Power Research Institute (ERPI) reported security is the main concern in IoT-enabled SG worldwide. Compromising the security of smart meters will mislead estimations, and an incorrect consumption estimation would lead to large financial losses (Mahmood, Ashraf Chaudhry, Naqvi, Shon, & Farooq Ahmad, 2016). Security is also important to protect the privacy of consumers and the utility.

Threat modelling is the process of analysis that allows security experts to discover the potential vulnerabilities to be addressed (Swiderski & Snyder, 2004). Threat modelling for an IoT-enabled SG during system design identifies the necessary controls and countermeasures (Khan et al., 2017): potential attacks need to be identified at the system design stage by the security designer, not the attacker (Myagmar et al., 2005).

In developing a threat model, security designers are concerned with defining threats (Myagmar et al., 2005). Security requirements can be mapped to threats to show the effect of each threat and the required security criteria of the system. It is argued that security requirements for the system can be defined clearly once the threats are identified.

V. Conclusion

In this paper, the link between IoT modules and the Smart Grid is highlighted. Firstly, a comprehensive overview is given of IoT-enabled Smart Grid. Then, the challenges of the IoT- enabled Smart Grid are discussed which will help in directing future research. Finally, further threat analysis will be undertaken in future work to identify the controls required to build a secure information system for the IoT -enabled Smart Grid. This research could serve and direct the research objectives in the future.

VI. Acknowledgements

I sincerely acknowledge the awards of the King Abdulaziz University scholarship and the Saudi Arabian Cultural Bureau in London (SACB) for allowing the research to be funded and undertaken.

References

Al-Ali, A.R. and Aburukba, R. (2015) 'Advanced role of internet of things in the smart grid technology', Journal of Computer and Communications, pp. 229–233. doi:http://dx.doi.org/10.4236/jcc.2015.35029 Role.

Al-Omar, B., Al-Ali, A.R., Ahmed, R. and Landolsi, T. (2012) 'The Role of Information and Communication Technologies in', Journal of Emerging Trends in Computing and Information Sciences, 3(5), pp. 707–716.

Al-Turjman, F. and Abujubbeh, M. (2019) 'IoT-enabled smart grid via SM: An overview', Future Generation Computer Systems, 96, pp. 579–590. doi:10.1016/j.future.2019.02.012.

Arasteh, H., Hosseinnezhad, V., Loia, V., Tommasetti, A., Troisi, O., Shafie-Khah, M. and Siano, P. (2016) 'Iot-based smart cities: A survey', EEEIC 2016 - International Conference on Environment and Electrical Engineering, pp. 1–6. doi:10.1109/EEEIC.2016.7555867.

Atlam, H.F. and Wills, G.B. (2019) Technical aspects of blockchain and IoT. 1st edn, Advances in Computers. 1st edn. Elsevier Inc. doi:10.1016/bs.adcom.2018.10.006.

Banerjee, M., Lee, J. and Choo, K.K.R. (2018) 'A blockchain future for internet of things security: a position paper', Digital Communications and Networks, 4(3), pp. 149–160. doi:10.1016/j.dcan.2017.10.006.

Bedi, G., Venayagamoorthy, G.K., Singh, R., Brooks, R.R. and Wang, K.C. (2018) 'Review of Internet of Things (IoT) in Electric Power and Energy Systems', IEEE Internet of Things Journal, 5(2), pp. 847–870. doi:10.1109/JIOT.2018.2802704.

Bekara, C. (2014) 'Security issues and challenges for the IoT-based smart grid', Procedia Computer Science, 34, pp. 532–537. doi:10.1016/j.procs.2014.07.064.

Cisco (2013) The Internet of Everything Global Public Sector Economic Analysis. Available at:

https://www.cisco.com/c/dam/en_us/about/business-insights/docs/ioe-value-at-stake-public-sector-analysis-faq.pdf (Accessed: 26 January 2020).

Colak, I. (2016) 'Introduction to smart grid', 2016 International Smart Grid Workshop and Certificate Program, ISGWCP 2016, pp. 1–5. doi:10.1109/ISGWCP.2016.7548265.

Connor, P. and Fitch-Roy, O. (2019) The Socio-Economic Challenges of Smart Grids, Pathways to a Smarter Power System. Elsevier Ltd. doi:10.1016/B978-0-08-102592-5.00014-4.

Dalipi, F. and Yayilgan, S.Y. (2016) 'Security and privacy considerations for IoT application on smart grids: Survey and research challenges', Proceedings - 2016 4th International Conference on Future Internet of Things and Cloud Workshops, W-FiCloud 2016, pp. 63–68. doi:10.1109/W-FiCloud.2016.28.

DeBlasio, R. and Tom, C. (2008) 'Standards for the smart grid', 2008 IEEE Energy 2030 Conference, ENERGY 2008, (November), pp. 1–7. doi:10.1109/ENERGY.2008.4780988.

EPRI (2005) EPRI | SmartGrid Resource Center. Available at: https://smartgrid.epri.com/ (Accessed: 7 January 2020).

ETP (2006) Smart Grids European Technology Platform-EARPA. Available at: https://www.earpa.eu/earpa/39/etp_smartgrids.html.

Flick, T. and Morehouse, J. (2011) 'Threats and Impacts: Consumers', pp. 19–33. doi:10.1016/B978-1-59749-570-7.00002-9.

Ganguly, P., Nasipuri, M. and Dutta, S. (2019) 'Challenges of the Existing Security Measures Deployed in the Smart Grid Framework', Proceedings of 2019 the 7th International Conference on Smart Energy Grid Engineering, SEGE 2019, pp. 1–5. doi:10.1109/SEGE.2019.8859917.

Ghasempour, A. (2016) 'Optimized Advanced Metering Infrastructure Architecture of Smart Grid based on Total Cost , Energy , and Delay', 2016 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT), pp. 1–6. doi:10.1109/ISGT.2016.7781250.

Ghasempour, A. (2019) 'Internet of things in smart grid: Architecture, applications, services, key technologies, and challenges', Inventions, 4(1). doi:10.3390/inventions4010022.

Gungor, V.C., Lu, B., Member, S., Hancke, G.P. and Member, S. (2010) 'Opportunities and Challenges of Wireless Sensor Networks in Smart Grid', 57(10), pp. 3557–3564.

Hu, J. and Vasilakos, A. V. (2016) 'Energy Big Data Analytics and Security: Challenges and Opportunities', IEEE Transactions on Smart Grid, 7(5), pp. 2423–2436. doi:10.1109/TSG.2016.2563461.

IEEE (2018) About - IEEE Smart Grid. Available at: https://smartgrid.ieee.org/about-ieee-smart-grid (Accessed: 4 December 2019).

IET (2013) What is a Smart Grid?

IoTSF (2020) Best Practice Guidelines 2 – IoT Security Foundation. Available at: https://www.iotsecurityfoundation.org/best-practice-guidelines/ (Accessed: 23 July 2020).

Jenkins, N., Long, C. and Wu, J. (2015) 'Smart Grid — Review An Overview of the Smart Grid in Great Britain', Engineering, 1(4), pp. 413–421. doi:10.15302/J-ENG-2015112.

Kaur, M. and Kalra, S. (2016) 'A Review on IOT Based Smart Grid', International Journal of Energy, Information and Communications, 7(3), pp. 11–22. doi:10.14257/ijeic.2016.7.3.02.

Khan, R., Mclaughlin, K., Laverty, D. and Sezer, S. (2017) 'STRIDE-based Threat Modeling for Cyber-Physical Systems', pp. 0–5.

Al Khuffash, K. (2018) Smart grids—Overview and background information, Application of Smart Grid Technologies. Elsevier Inc. doi:10.1016/b978-0-12-803128-5.00001-5.

Kimani, K., Oduol, V. and Langat, K. (2019) 'Cyber security challenges for IoT-based smart grid networks', International Journal of Critical Infrastructure Protection, 25, pp. 36–49. doi:10.1016/j.ijcip.2019.01.001.

Mahmood, K., Ashraf Chaudhry, S., Naqvi, H., Shon, T. and Farooq Ahmad, H. (2016) 'A lightweight message authentication scheme for Smart Grid communications in power sector', Computers and Electrical Engineering, 52, pp. 114–124. doi:10.1016/j.compeleceng.2016.02.017.

Manyika, J., Chui, M., Bisson, P., Woetzel, J., Dobbs, R., Bughin, J. and Aharon, D. (2015) Unlocking the potential of the Internet of Things | McKinsey & Company, McKinsey.

Martins, J.F., Pronto, A.G., Delgado-Gomes, V. and Sanduleac, M. (2019) 'Smart Meters and Advanced Metering Infrastructure', in Pathways to a Smarter Power System, pp. 89–114. doi:10.1016/b978-0-08-102592-5.00004-1.

Mohassel, R.R., Fung, A., Mohammadi, F. and Raahemifar, K. (2014) 'A survey on Advanced Metering Infrastructure', International Journal of Electrical Power and Energy Systems, 63, pp. 473–484. doi:10.1016/j.ijepes.2014.06.025.

Mollah, M.B., Zhao, J., Niyato, D., Lam, K.-Y., Zhang, X., Ghias, A.M.Y.M., Koh, L.H. and Yang, L. (2020) 'Blockchain for Future Smart Grid: A Comprehensive Survey', IEEE Internet of Things Journal, X(vi), pp. 1–1. doi:10.1109/jiot.2020.2993601.

Mrabet, Z. El, Kaabouch, N., Ghazi, Hassan El and Ghazi, Hamid El (2018) 'Cyber-security in smart grid: Survey and challenges', Computers and Electrical Engineering, 67, pp. 469–482. doi:10.1016/j.compeleceng.2018.01.015.

Mugunthan, S.R. and Vijayakumar, D.T. (2019) 'REVIEW ON IOT BASED SMART GRID ARCHITECTURE IMPLEMENTATIONS', 01(01), pp. 12–20.

NIST (2014) NIST Special Publication 1108R3 NIST Framework and Roadmap for Smart Grid Interoperability Standards, Release 3.0, NIST Special Publication. doi:10.6028/NIST.SP.1108r3.

Pour, M.M., Anzalchi, A. and Sarwat, A. (2017) 'A review on cyber security issues and mitigation methods in smart grid systems', Conference Proceedings - IEEE SOUTHEASTCON, pp. 17–20. doi:10.1109/SECON.2017.7925278.

PTI (2011) Smart Grid 101 for Local Governments.

Rahman, S. (2009) 'Smart grid expectations: What will make it a reality', IEEE Power and Energy Magazine, 7(5). doi:10.1109/MPE.2009.933415.

Reka, S.S. and Dragicevic, T. (2018) 'Future e ff ectual role of energy delivery : A comprehensive review of Internet of Things and smart grid', Renewable and Sustainable Energy Reviews, 91(April), pp. 90–108. doi:10.1016/j.rser.2018.03.089.

Risteska Stojkoska, B.L. and Trivodaliev, K. V. (2017) 'A review of Internet of Things for smart home: Challenges and solutions', Journal of Cleaner Production, 140, pp. 1454–1464. doi:10.1016/j.jclepro.2016.10.006.

Sagiroglu, S., Terzi, R., Canbay, Y. and Colak, I. (2017) 'Big data issues in smart grid systems', 2016 IEEE International Conference on Renewable Energy Research and Applications, ICRERA 2016, 5, pp. 1007–1012. doi:10.1109/ICRERA.2016.7884486.

Sajid, A., Abbas, H. and Saleem, K. (2016) 'Cloud-Assisted IoT-Based SCADA Systems Security: A Review of the State of the Art and Future Challenges', IEEE Access, 4, pp. 1375–1384. doi:10.1109/ACCESS.2016.2549047.

Saleem, Y., Crespi, N., Rehmani, M.H. and Copeland, R. (2019) 'Internet of Things-Aided Smart Grid: Technologies, Architectures, Applications, Prototypes, and Future Research Directions', IEEE Access, 7, pp. 62962–63003. doi:10.1109/ACCESS.2019.2913984.

Shakerighadi, B., Anvari-Moghaddam, A., Vasquez, J.C. and Guerrero, J.M. (2018) 'Internet of things for modern energy systems: State-of-the-art, challenges, and open issues', Energies, 11(5). doi:10.3390/en11051252.

Singer, J. (2009) Enabling Tomorrow's Electricity System: Report of the Ontario Smart Grid Forum.

Sortomme, E., Hindi, M.M., MacPherson, S.D.J. and Venkata, S.S. (2011) 'Coordinated charging of plug-in hybrid electric vehicles to minimize distribution system losses', IEEE Transactions on Smart Grid, 2(1), pp. 186–193. doi:10.1109/TSG.2010.2090913.

U.S. Department of energy (2018) Smart Grid System Report: 2018 Report to Congress.

US Public Law (2007) Energy independence and security act of 2007, US Government Printing Office.

Wang, Q., Tai, W., Tang, Y. and Ni, M. (2019) 'Review of the false data injection attack against the cyber-physical power system', IET Cyber-Physical Systems: Theory and Applications, 4(2), pp. 101–107. doi:10.1049/iet-cps.2018.5022.

Yang, Q. (2019) 'Internet of things application in smart grid: A brief overview of challenges, opportunities, and future trends', in Smart Power Distribution Systems. doi:10.1016/b978-0-12-812154-2.00013-4.

CPS&IoT'2022 Summer School on Cyber-Physical Systems and Internet-of-Things

Budva, Montenegro, June 7-11, 2022

Schedule

Day 1, Tuesday 7 June:

09:00-09:15 Event Chairs and Special Guests
Title: Opening Ceremony of the CPS&IoT'2022 Summer School, and MECO'2022 and CPS&IoT'2022 Conferences
09:15-10:15 Luca Benini, Integrated Systems Lab, ETH Zürich, CH
Keynote: PULP: Extreme Energy Efficiency for Extreme Edge AI Acceleration
10.15-10.30 Break
10.30-11.00 Lech Jóźwiak, TU/e, NL
Title: Introduction to the CPS&IoT'2021 Summer School
11.00-12.30 Lech Jóźwiak, TU/e, NL
Title: Green CPS and IoT for Green World
12.30-14.00 Lunch Break
14.00-15.30 Mario Kovač, FER, HR (To be Confirmed)
Title: European Processor Initiative Technology for Exascale Era
15.30-17.00 Gianluca Bellocchi, Alessandro Capotondi, Andrea Marongiu - UNIMORE and Francesca Palumbo, Daniel Madroñal Quintin UNISS, IT
Title: Accelerator-Rich FPGA Architecture Exploration via a Programmable and Reconfigurable Overlay
17.00-17.30 Break
17.30-19.00 Reda Nouacer and Morayo Adedjouma, CEA, LIST, FR
Title: From Embedded-Systems towards swarms: opportunities and challenges
21.00 Gala Dinner

Day 2, Wednesday 8 June:

09.00-10.00 Letizia Jaccheri, Fac. Information Techn. and Electrical Eng., NTNU, NO
Title: Keynote: Software for a Better Society
10.00-11.00 Roberto Giorgi, Dept. of Information Engineering, University of Siena, IT
Title: Keynote: Extending Performance and Reliability via Modular FPGA Clusters
11.00-11.20 Break
11.20-12.50 Filippo Cugini, CNIT, IT, Pavel Burget, CVUT, CZ, and Martin Ron, Factorio, CZ
Title: Edge computing: the BRAINE solution
12.50-14.10 Lunch Break
14.10-15.40 Axel Jantsch, TU-WIEN, AT and Zhonghai Lu, KTH, SE
Title: Embedded Machine Learning
15.40-17.10 Muhammad Shafique and Muhammad Abdullah Hanif, NYU-AD, UAE
Title: Embedded Machine Learning for the Edge: From Algorithms to Architectures
17.10-17.30 Break
17.30-19.00 Eugenio Villar, Hector Posadas, Raul Gomez, Jose María Gandara, TEISA/UNICAN, ES
Title: Modeling, design and Implementation of drone-based services (Hands-on Tutorial)

Day 3, Thursday 9 June:

09.00-10.00 Dimitrios Serphanos, U. Patras and CTI, Stavros Koubias, U. Patras and ISI, Christos Koulamas, ISI, GR
Title: Keynote: Synthesis of Runtime Monitors for Safe and Secure Industrial Systems
10.00-10.30 Break
10.30-13.00 Dominique Blouin, Telecom Paris, and Anish Bhobe, Institut Polytechnique de Paris, FR
Title: Embedded systems modeling, analysis and automatic code generation with AADL and RAMSES (Hand-on Tutorial)
13.00-14.30 Lunch Break
14.30-16.00 Rupert Schlick, AIT, AT, and Thomas Bauer, Fraunhofer IESE, DE
Title: How to design and tailer a perfect fitting verification and validation process for your CPS&IoT project?
16.00-17.00 Peter Mörtl, Virtual Vehicle, AT
Title: Framework to facilitate Trustworthiness of Smart Systems for End Users
17.00-17.30 Break
17.30-19.00 Ramiro Samano Robles, ISEP, PT
Title: Reference architecture for trusted AIoT systems: certification, standardization and regulation

Day 4, Friday 10 June:

09.00-10.30 Christoph Schmittner, AIT, AT
Title: Cybersecurity Engineering and Management
10.30-12.00 Samir Ouchani, Lineact CESI, FR
Title: Secure and Reliable Smart Cyber Physical Systems
12.00-13.30 Lunch Break
13.30-15.00 Abdelhakim Baouya, University Grenoble-Alpes, FR
Title: Artificial Intelligence meets Formal Methods: Generation and verification of learned stochastic automata
15.00-15.45 Radovan Stojanović, University of Montenegro and MECOnet, ME
Title: Principles of performance effective nodes design for smart systems
15.45-16.30 Milica Orlandić, NTNU - Norwegian University of Science and Technology
Title: Data Processing Pipelines on small satellites and drones: challenges and solutions
16.30-17.00 Break
17.00-18.30 Nikolaos Voros et all, University of Peloponnese, GR
Title: The achievements of SMART4ALL project in Customized Low-Energy Computing for CPS
18.30-19.00 Closing of the CPS&IoT'2022 Summer School
+ Free participation in sessions of the CPS&IoT'2022 Conference and MECO'2022 Conference
Summer School participants are expected to come with their own laptops. Internet access will be guaranteed.

Day 5, Saturday 11 June: Excursion possible (excursion fee is not included in the summer school fee)

3rd Summer School on Cyber Physical Systems and Internet of Things - SS-CPSIoT'2022
3rd Generation (Students and Teachers)
Budva, Montenegro, 07-11.06.2022

5.6.2022	SS-CPSIoT2022, ALL			
#	**Students**	**Email**	**Country**	**Affiliation**
1	Abdullah Bin Masood	a.masood@cyens.org.cy	Cyprus	CYENS - Centre of Excellence
2	Shamoona Imtiaz	shamoona.imtiaz@mdh.se	Sweden	Mälardalen University
3	Guilherme Moura Reis Correia Gil	guilhermegil@ua.pt	Portugal	University of Aveiro
4	Priscilla Benedetti	priscilla.benedetti@vub.be	Italy	University of Perugia, Vrije Universiteit Brussel
5	Saurabh Band	sband@uni-bremen.de	Germany	University of Bremen
6	Francesco Terrosi	francesco.terrosi@unifi.it	Italy	Università degli Studi di Firenze
7	Khadija Shaheen	shaheen.khadija@ntnu.no	Norway	Norwegian University of Science and Technology
8	Zenepe Satka	zenepe.satka@mdh.se	Sweden	Mälardalen University
9	Aldin Berisa	aldin.berisa@mdh.se	Sweden	Mälardalens University
10	Naomi Stricker	nstricker@ethz.ch	Switzerland	Computer Engineering Group, ETH Zurich
11	Emil Njor	emil.njor@gmail.com	Denmark	The Technical University of Denmark
12	Maryam Zahid	maryam.zahid@mdh.se	Sweden	Malardalens University
13	Stijn Bellis	stijn.bellis@uantwerpen.be	Belgium	University of Antwerp, Cosys-Lab
14	Marina Subotin	marina.bulat@uns.ac.rs	Serbia	Fakultet tehnickih nauka
15	Erling Jellum	erling.r.jellum@ntnu.no	Norway	Norwegian University of Science and Technology
17	Hussein Marah	hussein.marah@ymail.com	Belgium	Department of Computer Science | University of Antwerp
18	Muhammad Naeem	mnaeem@cs.aau.dk	Pakistan	Denmark
19	Alexey Serdyuk	alexey.serdyuk@kit.edu	Germany	Karlsruhe Institute of Technology
20	Genti Rustemi	grustemi@umt.edu.al	Albania	Metropolitan University
21	Đorđije Bošković	dordije.boskovic@ntnu.no	Norway	Norges teknisk-naturvitenskapelige universitet NTNU
22	Quentin Picard	quentin.picard@cea.fr	France	CEA LIST, University of Paris-Saclay
23	Stevan Đurašković	sdjuraskovic@ufl.edu	Montenegro	University of Florida
24	Burak Karaduman	Burak.Karaduman@uantwerpen.be	Belgium	University of Antwerp
25	Jurgen Soom	jurgen.soom@taltech.ee	Estonia	Tallinn University of Technology
26	Jurgen Becker	becker@kit.edu	Germany	Karlsruhe Institute of Technology
27	Fatima Majed	fatimakmajed@gmail.com	Lebanon	University Paris Nanterre University Paris Nanterre
28	Gabor Fekete	feketegabor@edu.bme.hu	Hungary	Budapest University of Technology and Economics
29	Károly János Sipos	siposkaroly@edu.bme.hu	Hungary	Budapest University of Technology and Economics
30	Adrian Besimi	a.besimi@seeu.edu.mk	North Macedonia	South East European University
31	Florije Ismaili	f.ismaili@seeu.edu.mk	North Macedonia	South East European University
32	Dimitrios Kontargyris	d.kontargiris@esdalab.ece.uop.gr	Greece	University of Peloponnese
33	Georgia Kaisari	g.kaisari@esdalab.ece.uop.gr	Greece	University of Peloponnese
34	Angelos Karaiskos	a.karaiskos@esdalab.ece.uop.gr	Greece	University of Peloponnese
35	Eleftherios Pappas	e.pappas@esdalab.ece.uop.gr	Greece	University of Peloponnese
36	Michalis Nanos	ece2012smart-ic@go.uop.gr	Greece	University of Peloponnese
37	Alexandros Spournias	a.spournias@esda-lab.gr	Greece	University of Peloponnese
38	Ramandeep Kaur	ks.ramandeep@gmail.com	India	Gujranwala Guru Nanak Institute of Management & Technology (GGNIMT)
39	Erling Jellum	erling.r.jellum@ntnu.no	Norway	Norwegian University of Science and Technology
40	Abeer Akkad	asaa1n18@soton.ac.uk	UK	University of Southampton
41	Jovan Djurkovic	dj.jovan94@gmail.com	Montenegro	MECOnet
#	**Presenters**	**Email**	**Country**	**Affiliation**
1	Luca Benini	lbenini@iis.ee.ethz.ch	Switzerland	ETH Zurich, Integrated Systems Lab
2	Lech Jozwiak	l.jozwiak@chello.nl	Netherlands	Eindhoven University of Technology
3	Mario Kovač	Mario.Kovac@fer.hr	Croatia	University of Zagreb, Dept. of Elec. and Comp. Engineering
4	Josip Knezović	Josip.Knezovic@fer.hr	Croatia	University of Zagreb, Dept. of Elec. and Comp. Engineering
5	Gianluca Bellocchi		Italy	University of Modena and Reggio Emilia
6	Alessandro Capotondi	alessandro.capotondi@unimore.it	Italy	University of Modena and Reggio Emilia
7	Andrea Marongiu		Italy	University of Modena and Reggio Emilia
8	Daniel Madroñal Quintin	dmadronalquin@uniss.it	Italy	University of Sassari
9	Francesca Palumbo		Italy	University of Sassari
10	Reda Nouacer	reda.nouacer@cea.fr	France	French Alternative Energies and Atomic Energy Commission, CEA
11	Morayo Adedjouma		France	French Alternative Energies and Atomic Energy Commission, CEA
12	Letizia Jaccheri	letizia.jaccheri@ntnu.no	Norway	Norwegian University of Science and Technology, Fac. Information Techn. and Electrical Eng.
13	Roberto Giorgi	giorgi@unisi.it	Italy	University of Siena, Dept. of Information Engineering
14	Filippo Cugini	filippo.cugini@cnit.it	Italy	National Inter-University Consortium for Telecommunications
15	Pavel Burget	smrcka@vutbr.cz	Czech Republic	Czech Technical University in Prague
16	Martin Ron		Czech Republic	Factorio, Wube Software
17	Axel Jantsch	axel.jantsch@tuwien.ac.at	Austria	Vienna University of Technology
18	Zhonghai Lu	zhonghai@kth.se	Sweden	KTH Royal Institute of Technology in Stockholm
19	Muhammad Shafique	ms12713@nyu.edu	United Arab Emirates	New York University - Abu Dhabi
20	Muhammad Abdullah Hanif	mh6117@nyu.edu	United Arab Emirates	New York University - Abu Dhabi
21	Eugenio Villar	evillar@teisa.unican.es	Spain	University of Cantabria, TEISA
22	Hector Posadas		Spain	University of Cantabria, TEISA
23	Raul Gomez		Spain	University of Cantabria, TEISA
24	Jose María Gandara		Spain	University of Cantabria, TEISA
25	Dimitrios Serphanos	serpanos@ece.upatras.gr	Greece	Univeristy of Patras, Computer Technology Institute and Press
26	Stavros Koubias	koubias@ece.upatras.gr	Greece	University Of Patras, Industrial Systems Institute
27	Christos Koulamas	koulamas@isi.gr	Greece	Industrial Systems Institute
28	Dominique Blouin	dominique.blouin@telecom-paris.fr	France	Telecom Paris
29	Anish Bhobe	anish.bhobe@ip-paris.fr	France	The Polytechnic Institute of Paris
30	Rupert Schlick	rupert.schlick@ait.ac.at	Austria	Academy of Information Technology
31	Thomas Bauer	thomas.bauer@iese.fraunhofer.de	Germany	Fraunhofer Institute for Experimental Software Engineering
32	Peter Mörtl	Peter.Moertl@v2c2.at	Austria	Virtual vehicle
33	Ramiro Samano Robles	rasro@isep.ipp.pt	Portugal	Polytechnic of Porto - School of Engineering
34	Christoph Schmittner	Christoph.Schmittner@ait.ac.at	Austria	Academy of Information Technology
35	Samir Ouchani	souchani@cesi.fr	France	Lineact CESI, Innovation and research laboratory
36	Abdelhakim Baouya	abdelhakim.baouya@univ-grenoble-alpes.fr	France	University Grenoble-Alpes
37	Radovan Stojanović	stox@ucg.ac.me	Montenegro	University of Montenegro and MECOnet
38	Milica Orlandić	milica.orlandic@ntnu.no	Norway	Norwegian University of Science and Technology
39	Nikolaos Voros	voros@esdalab.ece.uop.gr	Greece	University of Peloponnese


CPS&IoT
Certificate of Attendance
SMART4ALL
THIS ACKNOWLEDGES THAT
Marko Markovic
Montenegrin Association for New Technologies – MANT, Montenegro
has successfully attended
The 3rd Summer School on
Cyber Physical Systems and Internet of Things (SS-CPSIoT'2022)
(3 ECTS)
in Budva, Montenegro, June 07-11 2022
On behalf of the organizers:
Prof. dr. Lech Jozwiak
Prof. dr. Radovan Stojanović
Prof. dr. Betim Cico
Prof. dr. Budimir Lutovac

Author Index

Abdullah Hanif Muhammad,558
Adedjouma Morayo, 287

Baouya Abdelhakim, 1013
Bauer Thomas, 719
Bellocchi Gianluca, 189
Benini Luca, 4
Bhobe Anish, 667
Blouin Dominique, 667
Burget Pavel, 410

Capotondi Alessandro, 189
Cugini Filippo, 410

Giorgi Roberto, 371
Gomez Raul, 592

Jaccheri Letizia, 323
Jantsch Axel, 471
Jozwiak Lech, 56

Knezović Josip, 158
Koubias Stavros, 637
Kovač Mario, 158

Lu Zhonghai, 471

Madroñal Quintin Daniel, 189
María Gandara Jose, 592
Marongiu Andrea, 189
Mörtl Peter, 777

Nouacer Reda, 287

Orlandić Milica, 1076
Ouchani Samir, 946

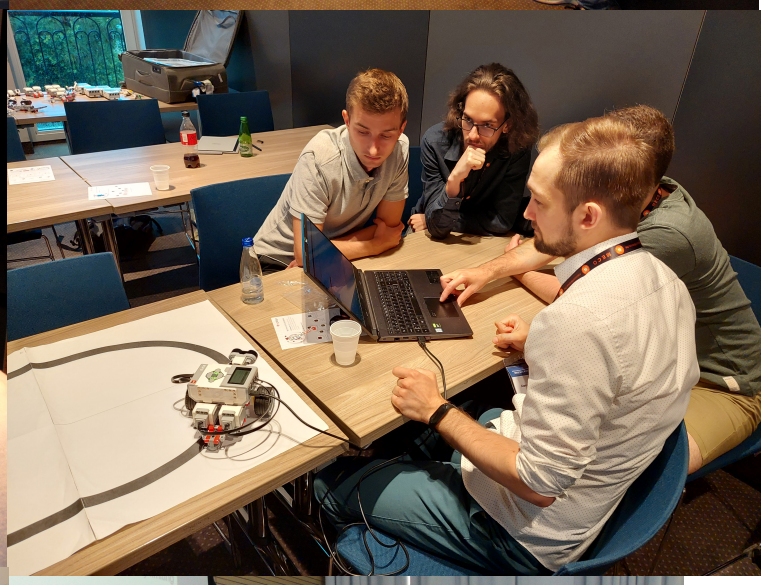
Palumbo Francesca, 189
Posadas Hector, 592

Ron Martin, 410

Samano Robles Ramiro, 206
Schlick Rupert, 719
Schmittner Christoph, 875
Serphanos Dimitrios, 637
Shafique Muhammad, 558
Stojanović Radovan, 1050

Villar Eugenio, 592
Voros Nikolaos, 1109

PHOTO GALLERY



CPS&IoT'2022 3rd Summer School on Cyber-Physical Systems and Internet-of-Things

Budva, Montenegro, June 7-11, 2022

Citation:
Author/s, "Title of contribution-pressentation", in Proceedings of the 3rd Summer School on Cyber-Physical Systems and Internet-of Things, Editors: Lech Jozwiak, Radovan Stojanovic and Nikolaos Voros, Vol. III, June 2022, pp. xx-yy, DOI: https://doi.org/10.5281/zenodo.6698645

Technical editors:

Prof. dr Budimir Lutovac, University of Montenegro

Stevan Djuraskovic, University of Florida

Publishers:

MECOnet Institute (Mediterranean Excellence in Computing and Ontology), www.meconet.me

Montenegrin Association for New Technologies - MANT, www.mant.me

Year/Place: Montenegro, Podgorica, 2022