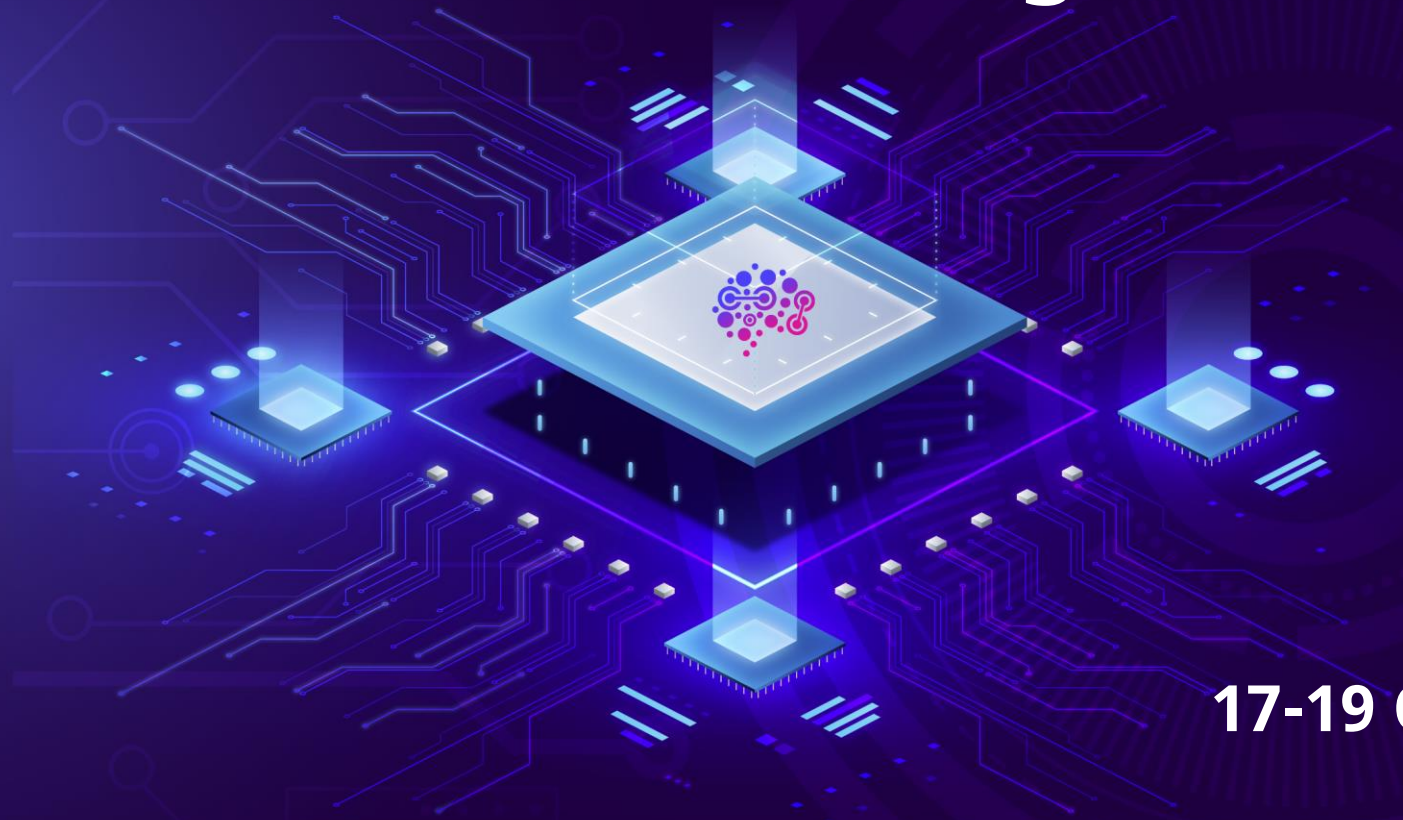
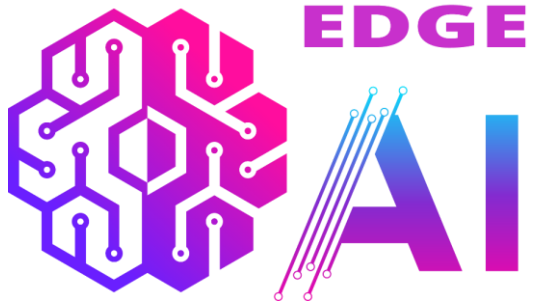
European Conference on EDGE AI Technologies and Applications - EEAI



Advancing emerging edge AI technologies and driving next generation intelligent applications.

17-19 October 2023, Athens, Greece

European Conference on EDGE AI Technologies and Applications - EEAI



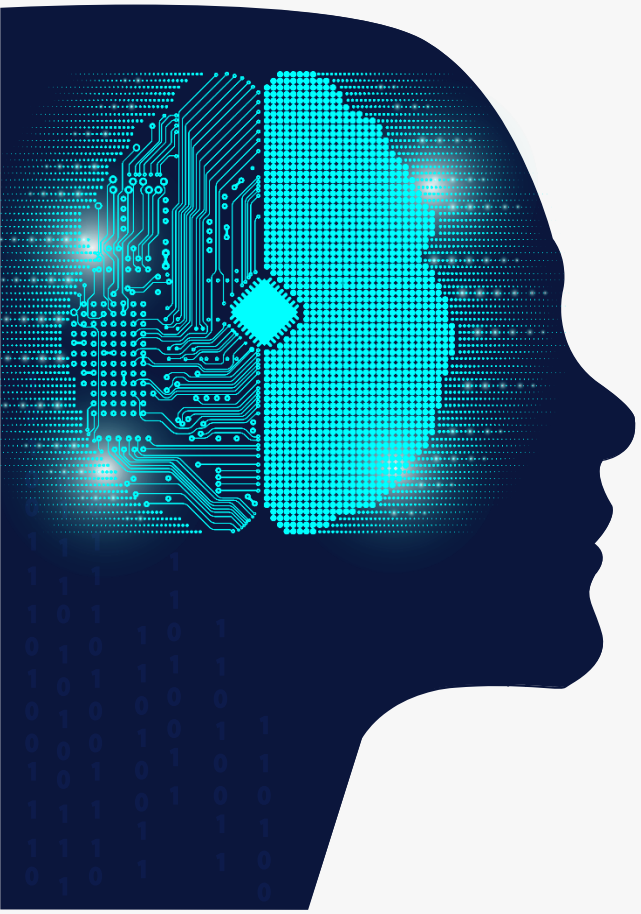
Shrinking Embedded Machine Learning Applications: FMCW Radar Gesture Sensing on Cortex-M0+ under 10KB and 10mW.

Jonas Daugalas, Infineon Technologies AG, Germany
Stephan Schoenfeldt, Jianyu Zhao, Maximilian Strobel

17-19 October 2023 Athens, Greece

Copyright © 2023

Outline



- Introduction
- State-of-the-art
- Research Goal
- Methods
- *Implementation + Results*
  - ⇧ x4
- Summary of Results
- Conclusion

Copyright © 2023

Introduction

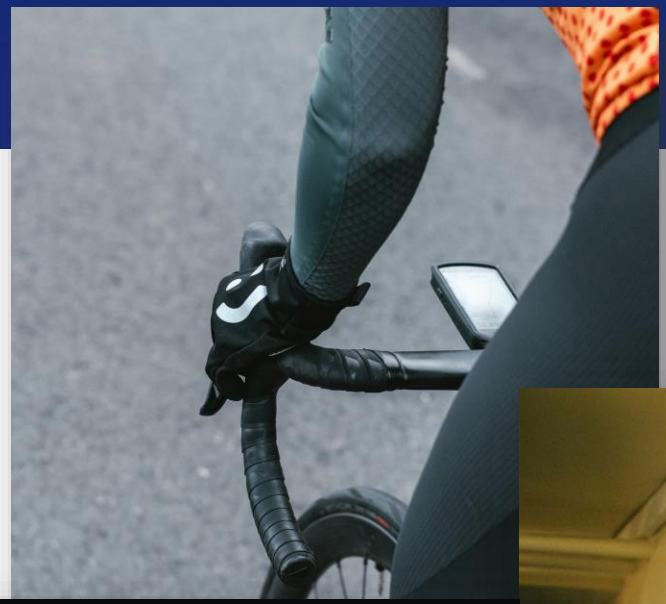
Contactless Gesture Sensing

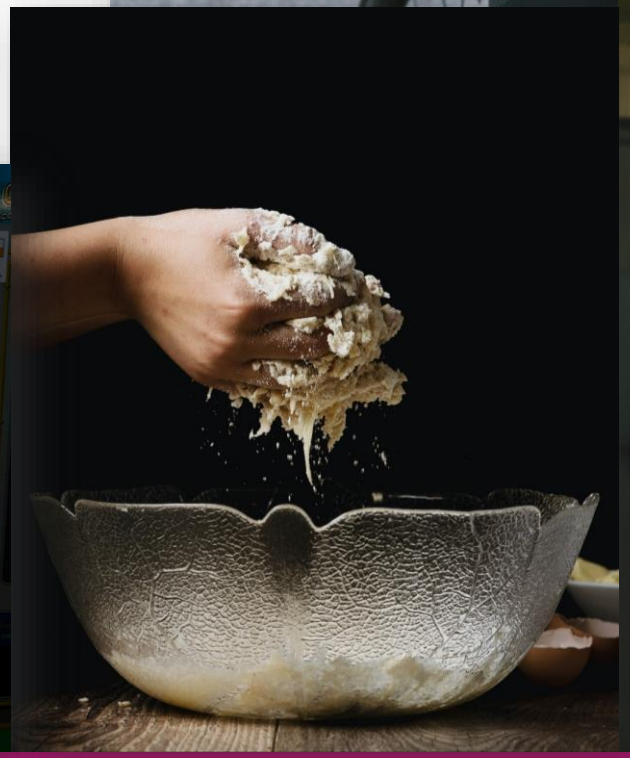
+

Radar

+

Edge





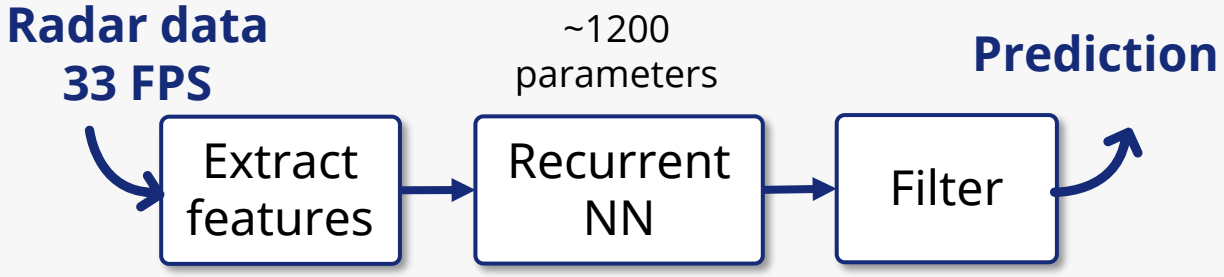
Copyright © 2023

State-of-the-art

5 hand gestures at ~1 meter

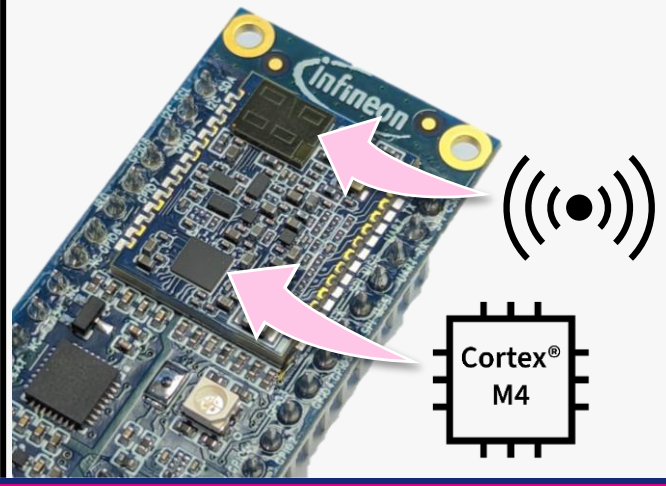
Push *Swipe Left* *Swipe Right* *Swipe Down* *Swipe Up*

Algorithm



M. Strobel, S. Schoenfeldt, and J. Daugalas, "Gesture Recognition for FMCW Radar on the Edge". *2023 IEEE Radio Wireless Week (RWW).*

Deployment

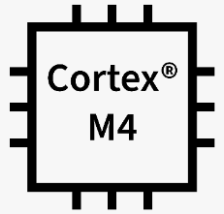


- 115 KB total RAM
- 195 KB total Flash
- 30 % CPU @ 150 MHz
- ARM Cortex-M4
- ~75 mW power

Copyright © 2023

Research Goal

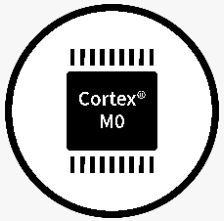
Great but … can we do better?



- 115 KB total RAM
- 195 KB total Flash
- 30 % CPU @ 150 MHz
- ARM Cortex-M4
- ~75 mW power

- Less compute
- Less memory
- Less power
- Cheaper

… gesture sensing with Cortex-M0?



- < 16 KB total RAM
- < 128 KB total Flash
- CPU @ 48 MHz
- ARM Cortex-M0+

Copyright © 2023

Methods

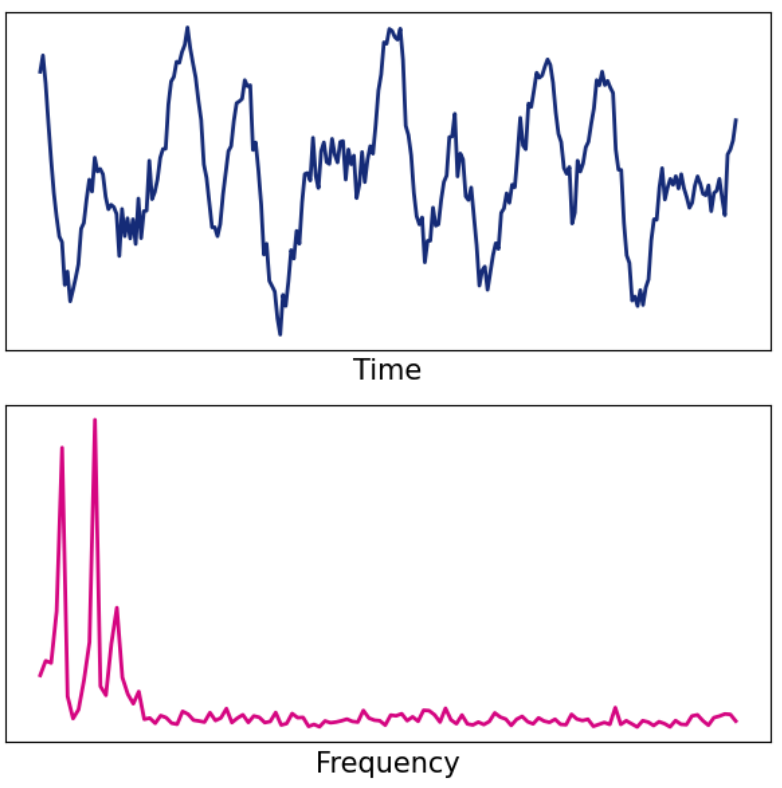
Optimization vectors

- Radar Preprocessing Algorithm
- Preprocessing Optimization
- Neural Network Inference Engines
- Neural Network Optimization

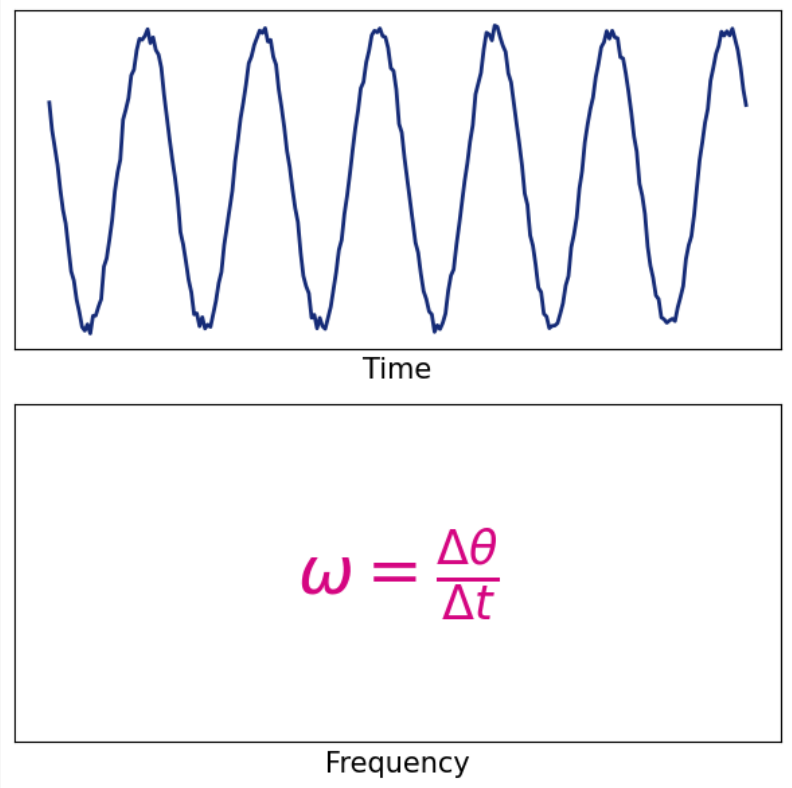
Copyright © 2023

Radar Preprocessing Algorithm: background

FMCW Radar processing relies on frequency analysis.



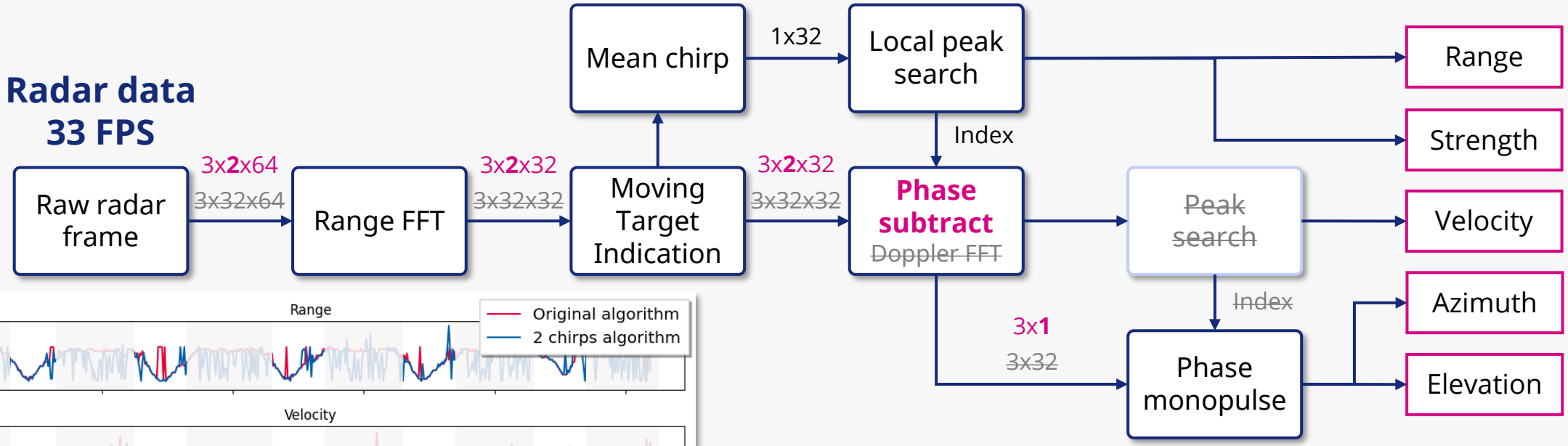
In close-range Gesture Sensing the „velocity signal“ is often dominated by a single frequency.

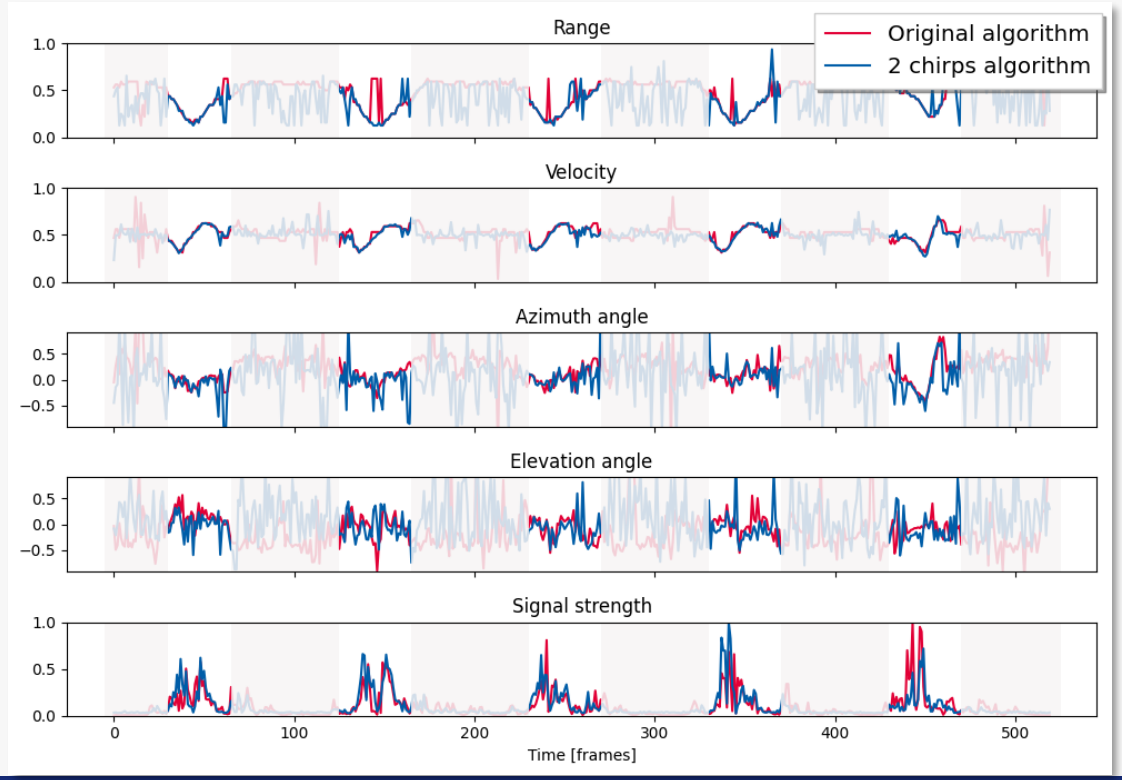


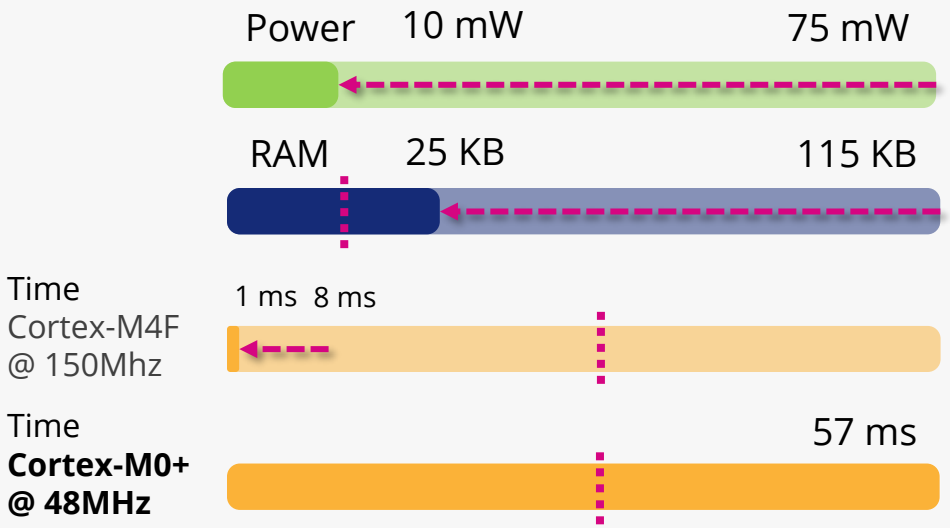
Copyright © 2023

Radar Preprocessing Algorithm

Radar data
33 FPS







Copyright © 2023

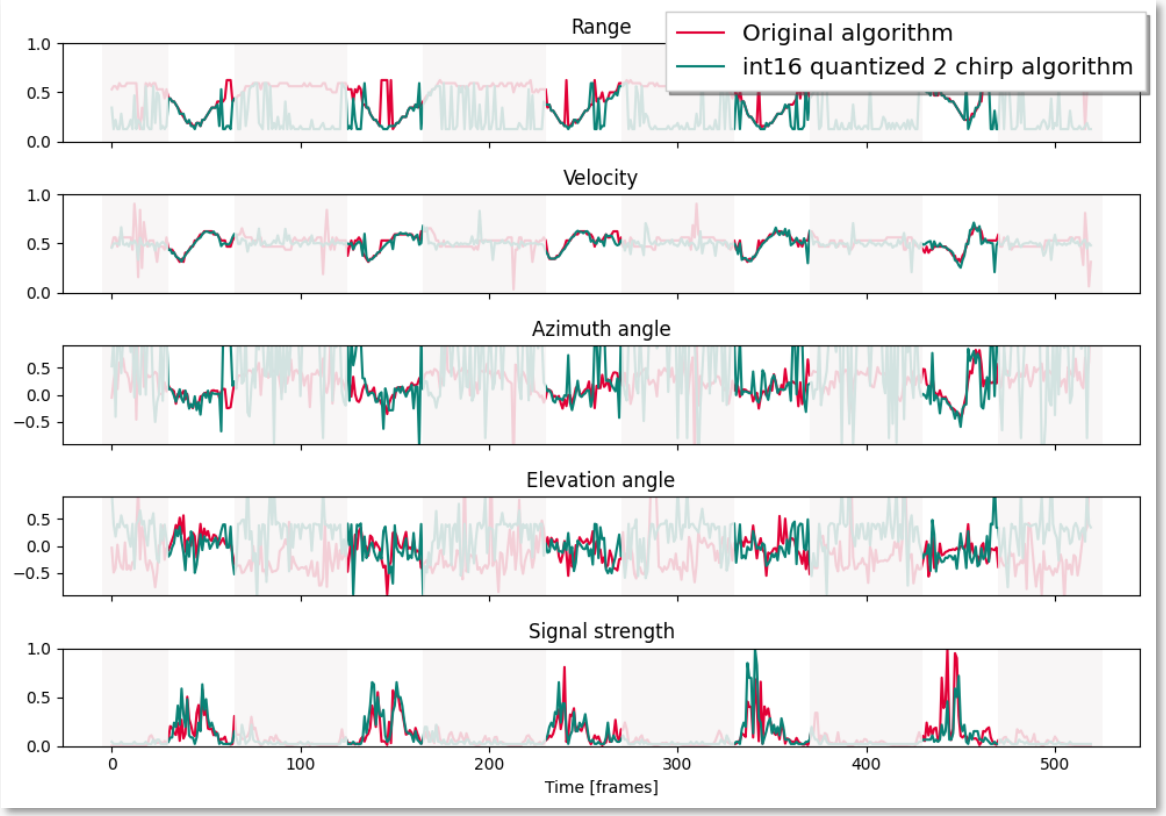
Preprocessing Quantization

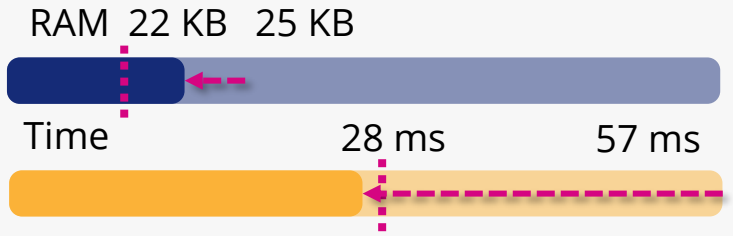
```c
void ifx_mean_removal_f32(float32_t* v, uint32_t len)
{
    assert(v != NULL);

    float32_t mean;
    arm_mean_f32(v, len, &mean);
    arm_offset_f32(v, -mean, v, len);
}
```

```c
void ifx_mean_removal_q15(q15_t* v, uint32_t len)
{
    assert(v != NULL);

    q31_t mean;
    arm_mean_q15(v, len, &mean);
    arm_offset_q15(v, -mean, v, len);
}
```





Copyright © 2023

Neural Network Inference Engines

TensorFlow Lite Micro (**TFLM**)

- Port of TensorFlow Lite designed to run ML models on devices with limited memory.
- Mature and extensively tested on Cortex-M cores.
- Memory hungry.

Interpreterless TensorFlow Lite Micro (**ITFLM**)

- Infineon mod of TFLM, removing the interpreter.
- Builds on a strong foundation of TFLM.
- Improved memory efficiency.
- Supports only Cortex-M0(+) and Cortex-M4.

Apache TVM **microTVM**

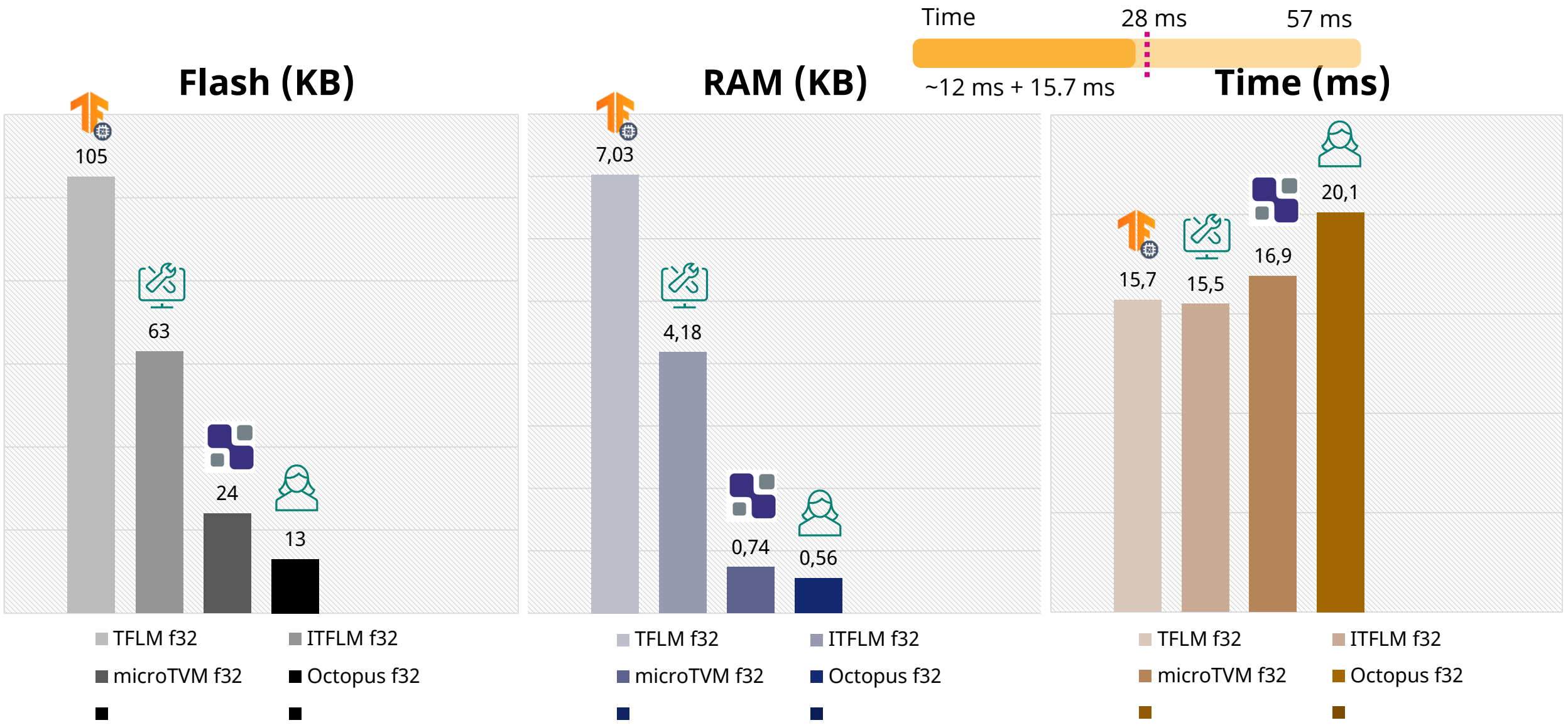
- Apache TVM deep learning compiler extension that targets model deployment to microcontrollers.
- Designed for minimal overhead and portability.
- Rapidly evolving.

Octopus

- Infineon Python and C library for quantization and code generation of neural networks.
- Designed to minimize memory requirements and produce readable and customizable code.
- Limited NN architecture support.

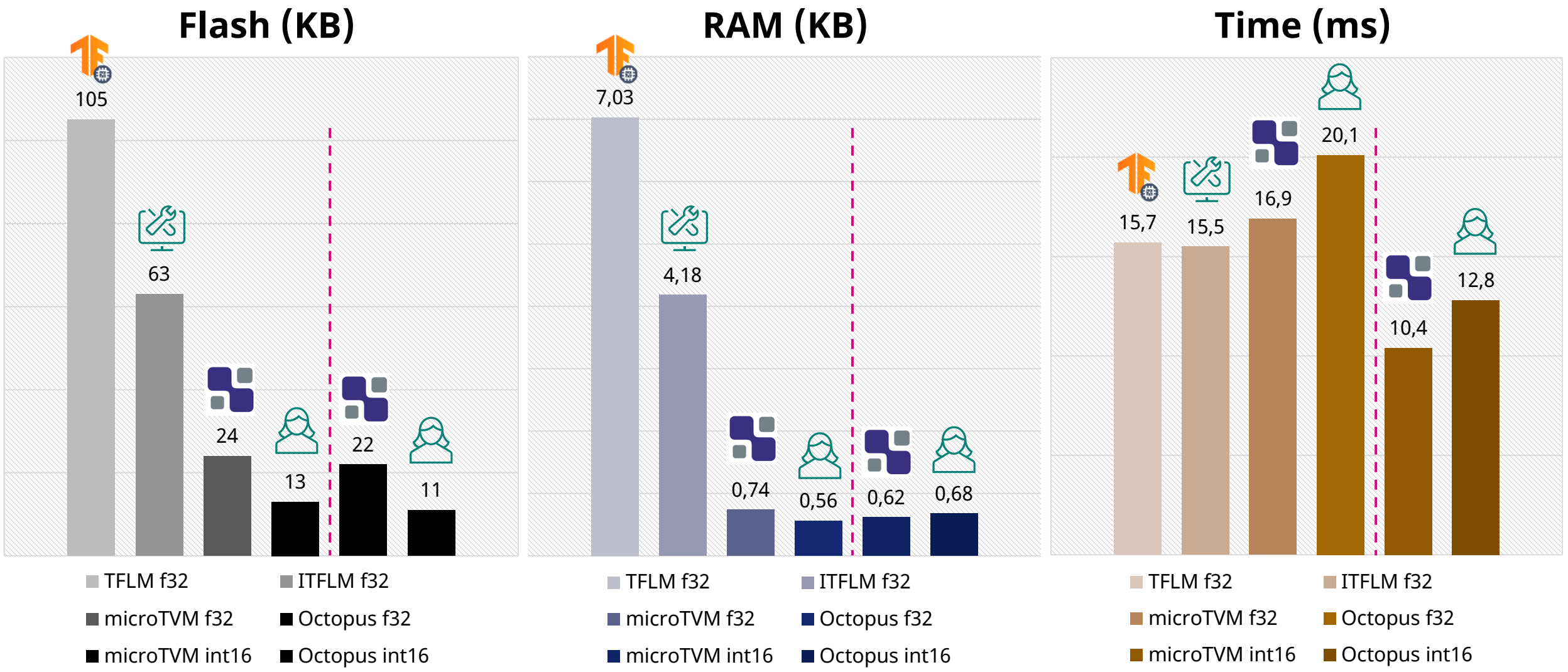
Copyright © 2023

Comparing Neural Network Inference Engines



Copyright © 2023

Neural Network Quantization



Copyright © 2023

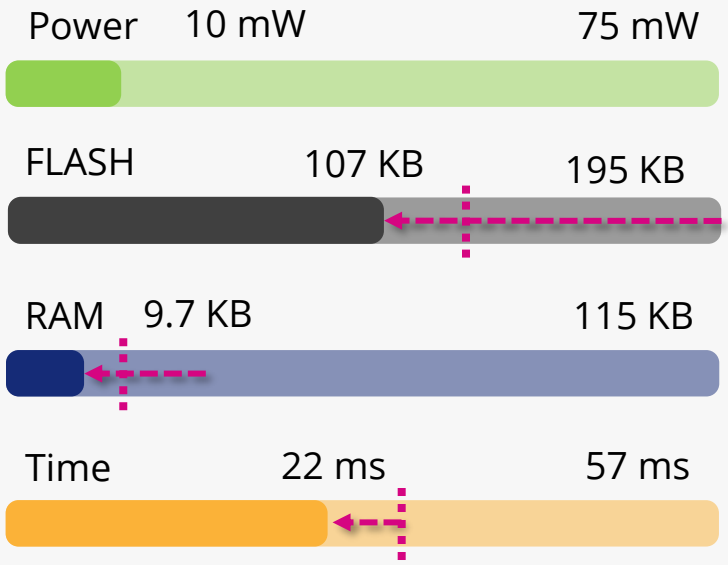
Summary of Results

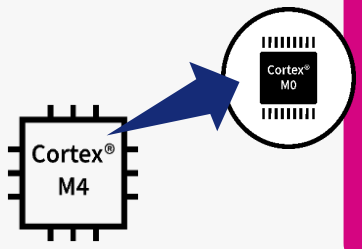
RADAR PREPROCESSING ALGORITHM

PREPROCESSING OPTIMIZATION

NEURAL NETWORK INFERENCE ENGINES

NEURAL NETWORK OPTIMIZATION

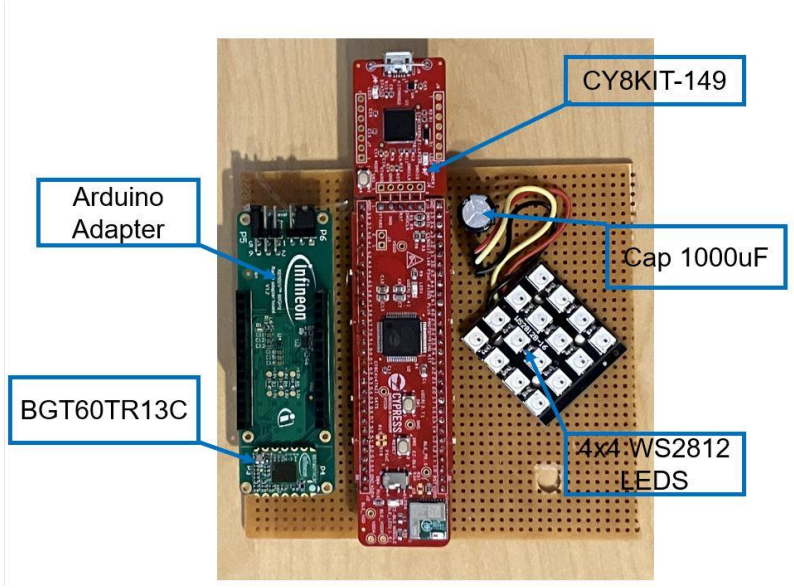


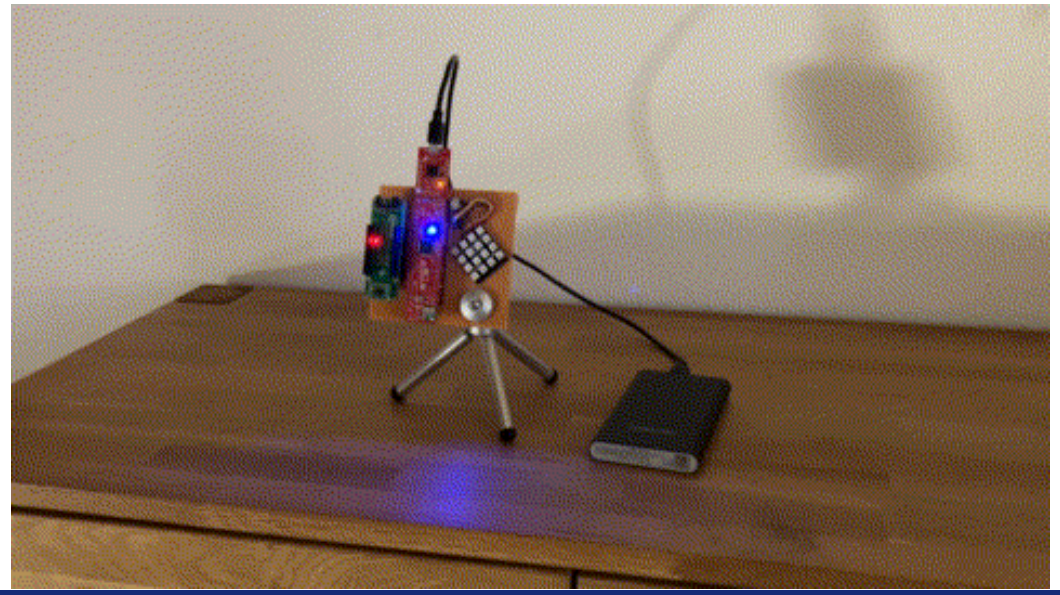


- < 16 kB total RAM
- < 128 kB total Flash
- CPU @ 48 MHz
- ARM Cortex-M0+

Copyright © 2023

Conclusion



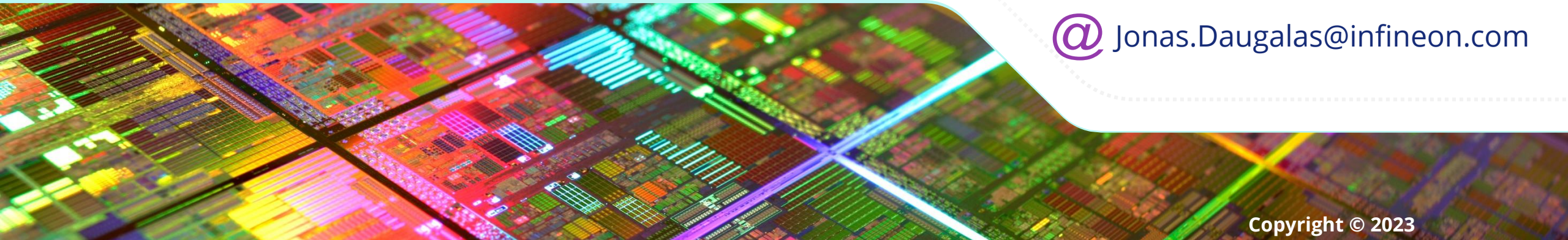


- Squeezing Machine Learning applications requires cross-domain expertise:
  - Machine learning
  - Embedded software
  - Problem domain
- microTVM strikes a good balance of performance, resource usage, and manual effort requirement for applications like embedded Radar Gesture Sensing.
- Now you can also make a highly efficient Radar Gesture Sensing solution. Probably the leanest of its kind so far.

Copyright © 2023

EDGE AI

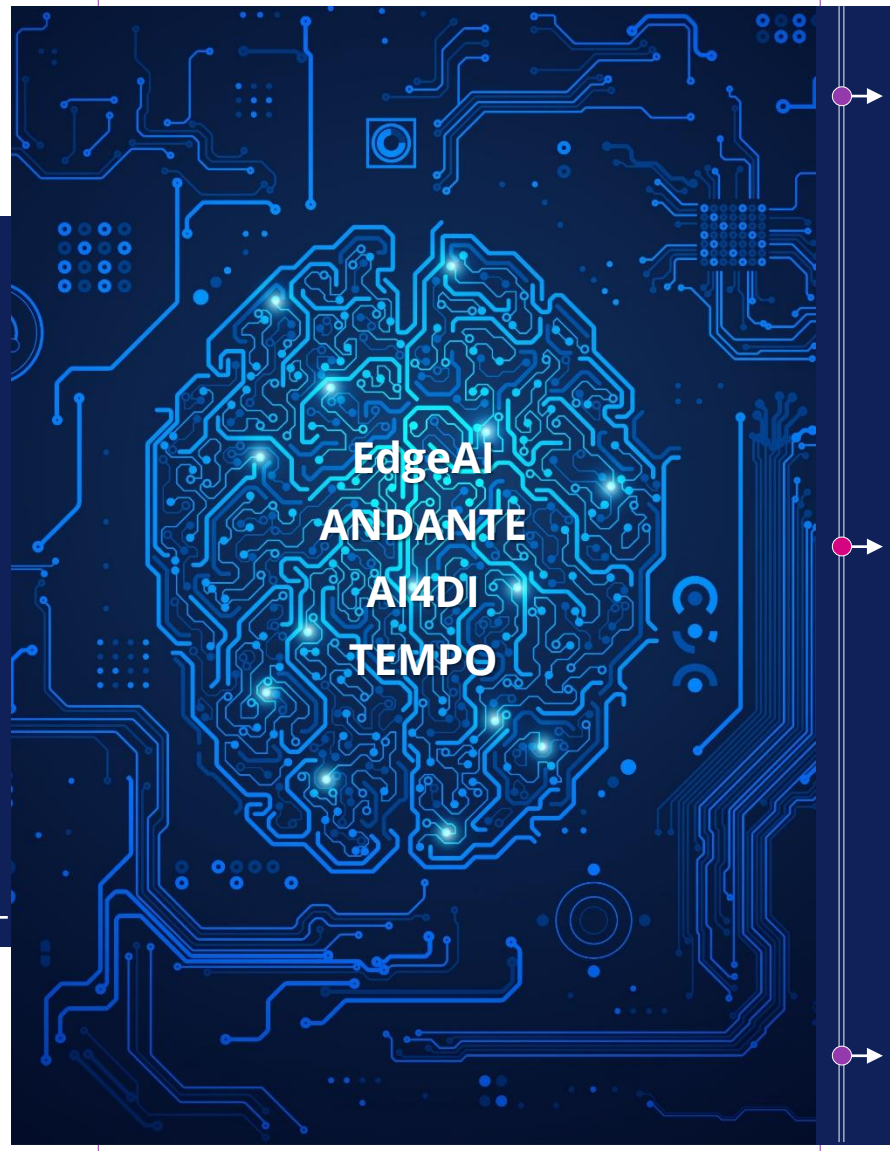
Let’s talk



@ Jonas.Daugalas@infineon.com

Copyright © 2023

Event Organisers



EdgeAI (Edge AI Technologies for Optimised Performance Embedded Processing) develops new electronic components and systems, processing architectures, connectivity, software, algorithms, and middleware through the combination of microelectronics, edge AI, embedded systems, and edge computing. www.edge-ai-tech.eu

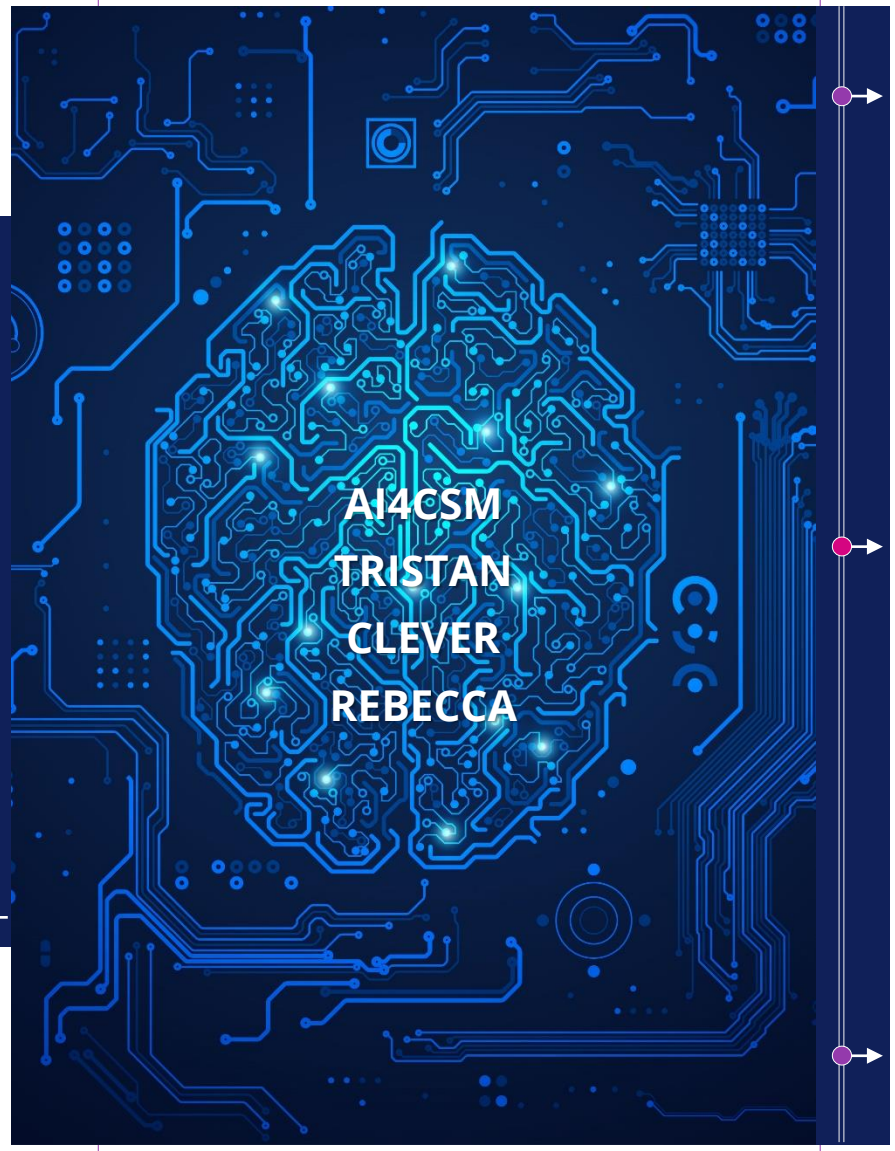
The ANDANTE project has received funding from the ECSEL Joint Undertaking (JU) under grant agreement No 876925. The JU receives support from the European Union's Horizon 2020 research and innovation programme and Belgium, France, Germany, The Netherlands, Portugal, Spain, Switzerland. www.andante-ai.eu

The TEMPO project has received funding from the ECSEL Joint Undertaking (JU) under grant agreement No 826655. The JU receives support from the European Union's Horizon 2020 research and innovation programme and Belgium, France, Germany, The Netherlands, Switzerland. www.tempo-ecsel.eu

The AI4DI project has received funding from the ECSEL Joint Undertaking (JU) under grant agreement No 826060. The JU receives support from the European Union's Horizon 2020 research and innovation programme and the national authorities. www.ai4di.eu

Copyright © 2023

Event Organisers



AI4CSM (Automotive Intelligence for Connected Shared Mobility) develops advanced electronic components and systems (ECS) and architectures for future mass-market ECAS vehicles based on sensor fusion and perception platforms, efficient propulsion and energy modules, advanced connectivity and trustworthy AI techniques and methods. www.ai4csm.eu

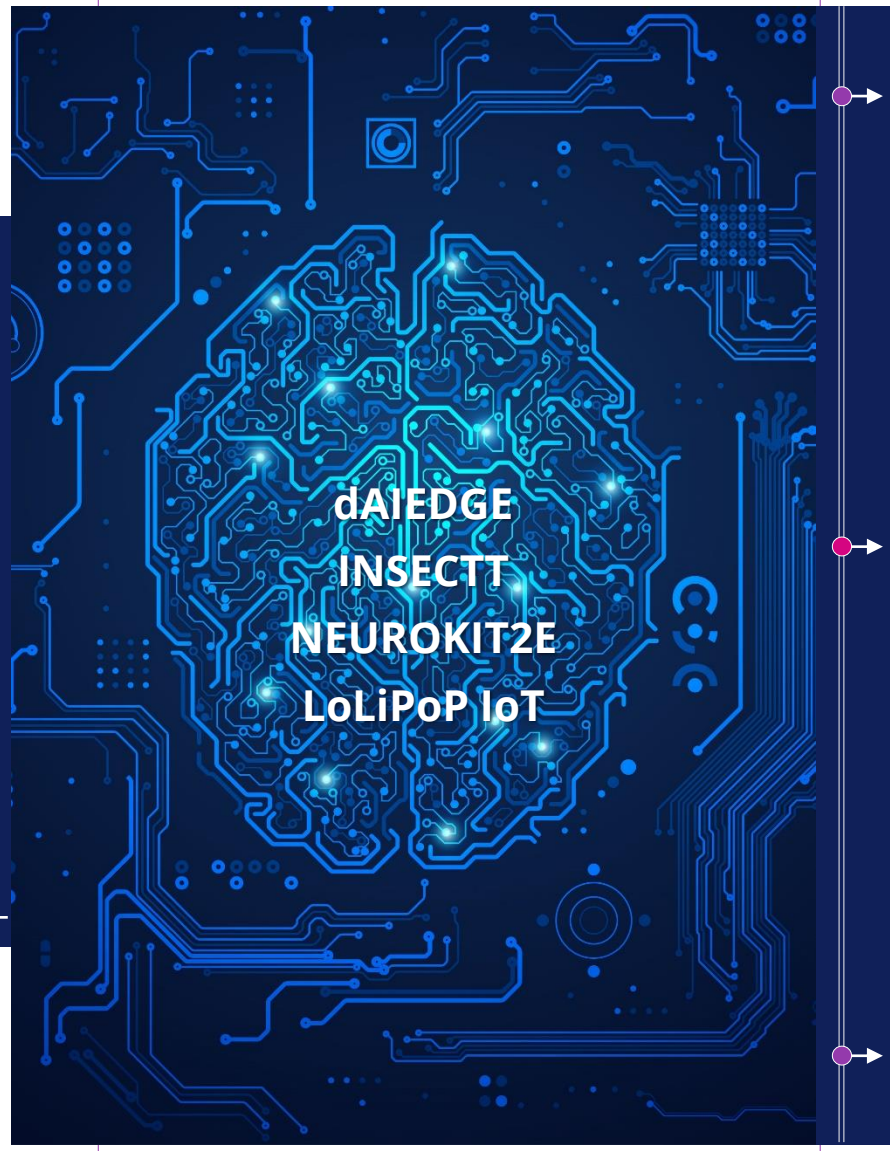
TRISTAN aims to expand, mature, and industrialize the European RISC-V ecosystem to compete with commercial options by leveraging the Open-Source community to gain productivity and quality. A European strategy for RISC-V designs will be defined, creating a repository of industrial rate building blocks for SoC designs in various application domains. www.tristan-project.eu

CLEVER (Collaborative edge-cLoud continuum and Embedded AI for a Visionary industry of thE futuRe) proposes innovations in hardware accelerators, design stack, and middleware software that revolutionize the ability of edge computing platforms to operate federated, leveraging sparse resources that are coordinated to create a powerful swarm of resources. www.cleverproject.eu

REBECCA (Reconfigurable Heterogeneous Highly Parallel Processing Platform for safe and secure AI) aims to democratize the development of edge AI systems and create a complete hardware and software stack centered around a RISC-V CPU, which offer higher performance, energy efficiency, safety, and security than existing systems. www.rebecca-chip.eu

Copyright © 2023

Event Organisers



dAIEDGE

dAIEDGE is the European Network of Excellence for distributed, trustworthy, efficient, and scalable AI at the Edge and promotes the application, development, and deployment of Artificial Intelligence (AI) on edge computing platforms.

InSecTT

INSECTT (Intelligent Secure Trustable Things) provide intelligent, secure, and trustworthy systems for industrial applications to provide comprehensive cost-efficient solutions of intelligent, end-to-end secure, trustworthy connectivity and interoperability to bring the Internet of Things and Artificial Intelligence together. www.insectt.eu

NEUROKIT2E

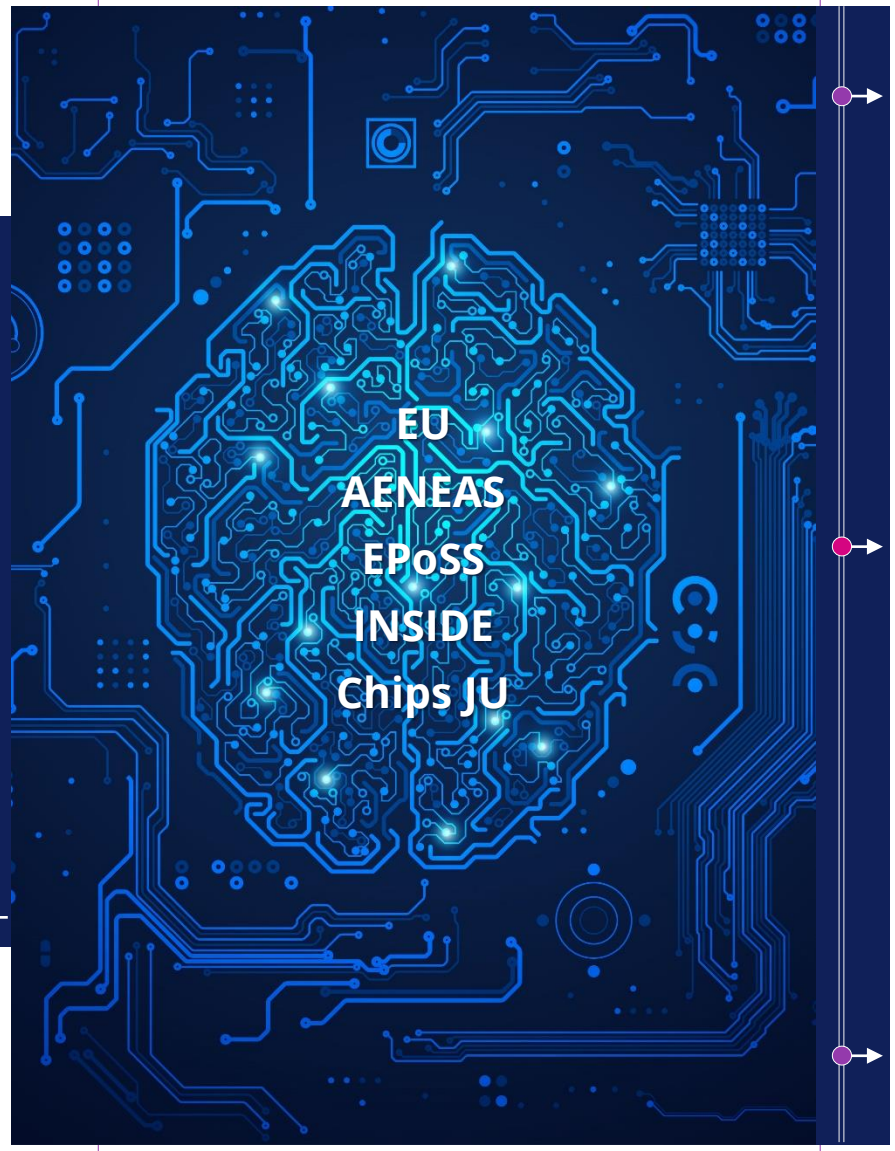
NEUROKIT2E (Open source deep learning platform dedicated to Embedded hardware and Europe) proposes a Deep Learning Platform for Embedded Hardware around an established European value chain (AI HW/SW). The solutions developed support neural network design, optimisation, and implementation on constrained HW.

LoLiPoP-IoT

The objectives of LoLiPoP IoT (Long Life Power Platforms for Internet of Things) are to develop energy harvesting-based innovative Long Life Power Platforms that enable retrofit of wireless sensor network edge devices for asset tracking, condition and performance monitoring. www.lolipop-iot.eu

Copyright © 2023

Supporting Organizations



AENEAS standing for Association for European NanoElectronics ActivitieS, is an industrial Association, established in 2006, providing unparalleled networking opportunities, policy influence & supported access to funding to all types RD&I participants in the field of micro and nanoelectronics enabled components and systems. https://aeneas-office.org/

The European Technology Platform on Smart Systems Integration is an industry-driven policy initiative, defining research, development and innovation needs as well as policy requirements related to Smart Systems Integration and integrated Micro- and Nanosystems. The main objective is to develop a vision and to set up a Strategic Research Agenda. www.smart-systems-integration.org

Inside Industry Association is the European Technology Platform for research, design and innovation on Intelligent Digital Systems and their applications. The Association is a membership organisation for the European research and innovation actors with more than 200 members and associates from all over Europe. www.inside-association.eu

Chips Joint Undertaking supports research, development, innovation, and future manufacturing capacities in the European semiconductor ecosystem. Launched as part of the Chips for Europe Initiative, it confronts semiconductor shortages and strengthens Europe's digital autonomy, engaging a significant EU, national/regional and private industry funding of nearly €11 billion. www.kdt-ju.europa.eu

Copyright © 2023