tinyML® Talks

Enabling Ultra-low Power Machine Learning at the Edge

"How to design a power frugal hardware for AI - the bio-inspiration path"

Alexandre Valentian - CEA

September 23, 2021

TINY ML

www.tinyML.org

TINY ML TALKS webcast

tinyML Talks Sponsors and Strategic Partners

AONdevIces

tinyML Strategic Partner

arm

tinyML Strategic Partner

Deeplite

EDGE IMPULSE

tinyML Strategic Partner

emza visual sense

tinyML Strategic Partner

GREENWAVES TECHNOLOGIES

tinyML Strategic Partner

LatentAI
Adaptive AI for a Smarter Edge

tinyML Strategic Partner

HOTG

tinyML Strategic Partner

imagimob

tinyML Strategic Partner

maxim integrated™ | NOW PART OF ANALOG DEVICES

Qeexo

tinyML Strategic Partner

Qualcomm

tinyML Strategic Partner

RealityAI®

tinyML Strategic Partner

seeed studio
The IoT Hardware Enabler

tinyML Strategic Partner

SensiML™

tinyML Strategic Partner

SynSense

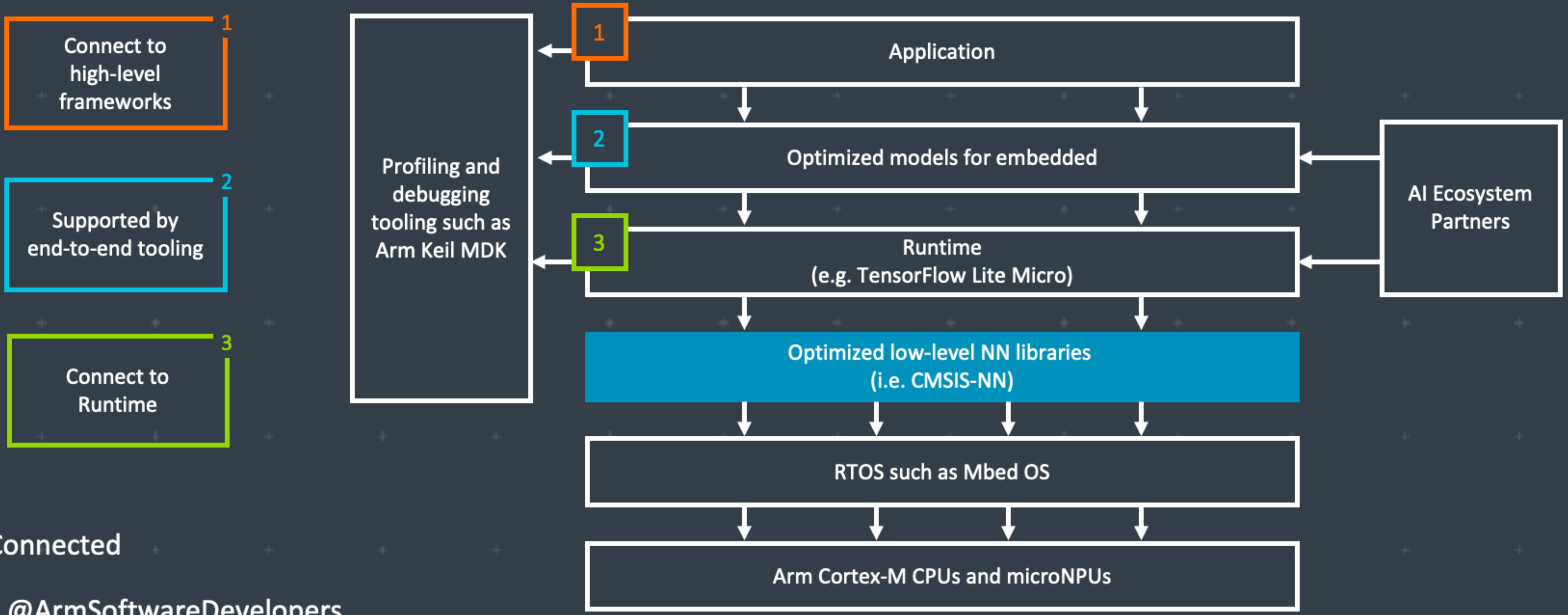
tinyML Strategic Partner

SYNTIANT

tinyML Strategic Partner

Additional Sponsorships available – contact Olga@tinyML.org for info

Arm: The Software and Hardware Foundation for tinyML


Connect to high-level frameworks
1
Supported by end-to-end tooling
2
Connect to Runtime
3
Profiling and debugging tooling such as Arm Keil MDK
1
Application
2
Optimized models for embedded
3
Runtime
(e.g. TensorFlow Lite Micro)
Optimized low-level NN libraries
(i.e. CMSIS-NN)
RTOS such as Mbed OS
Arm Cortex-M CPUs and microNPUs
AI Ecosystem Partners

Stay Connected

@ArmSoftwareDevelopers

@ArmSoftwareDev

Resources: developer.arm.com/solutions/machine-learning-on-arm

© 2020 Arm Limited (or its affiliates)

Deeplite

WE USE AI TO MAKE OTHER AI FASTER, SMALLER AND MORE POWER EFFICIENT

Automatically compress SOTA models like MobileNet to <200KB with **little to no drop in accuracy** for inference on resource-limited MCUs

Reduce model optimization trial & error from weeks to days using Deeplite's **design space exploration**

APPLY NOW

Deploy more models to your device without sacrificing performance or battery life with our **easy-to-use software**

BECOME BETA USER **bit.ly/testdeeplite**

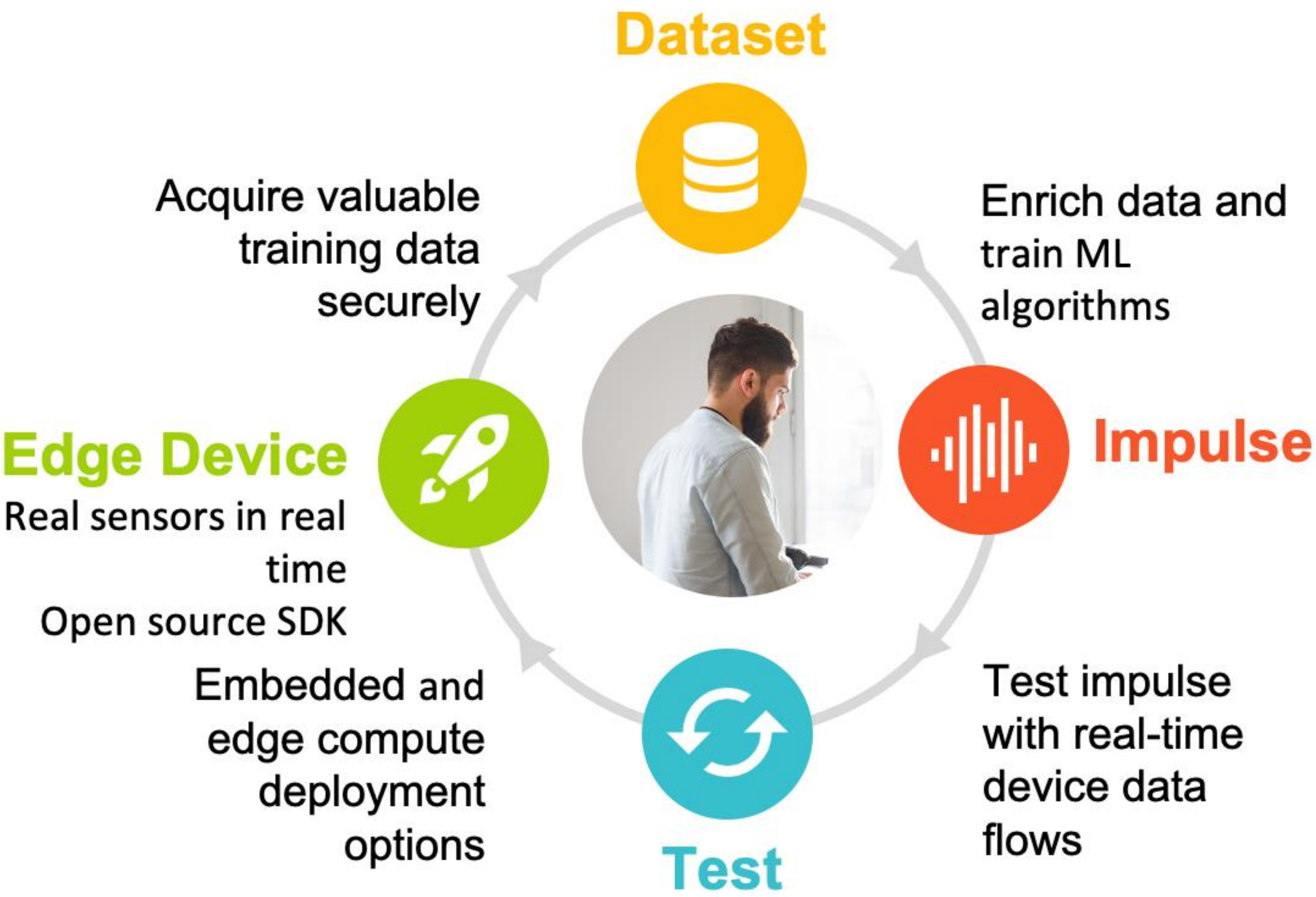
mobilityXlab arm AI 100 CBINSIGHTS

TinyML for all developers

C++ library

Arduino library

WebAssembly



Dataset

Acquire valuable training data securely

Enrich data and train ML algorithms

Impulse

Test impulse with real-time device data flows

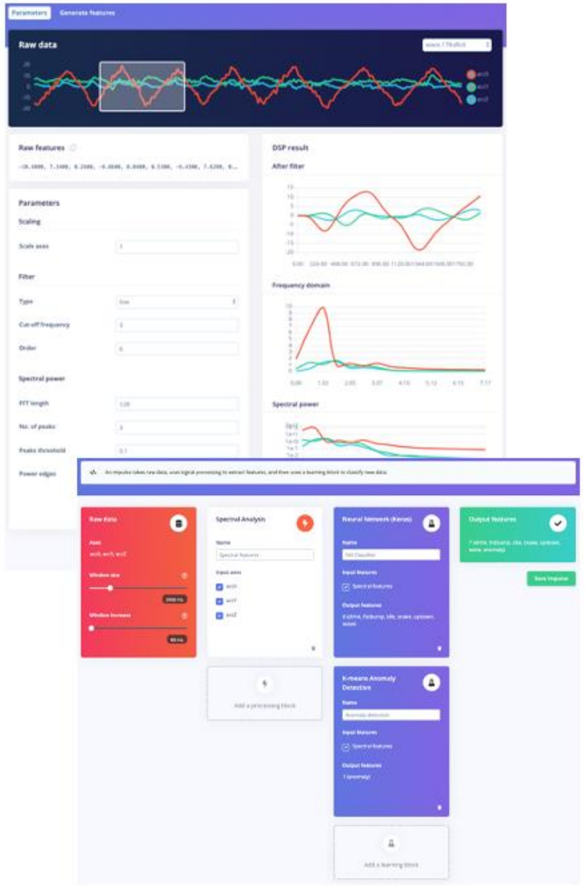
Test

Embedded and edge compute deployment options

Edge Device

Real sensors in real time

Open source SDK



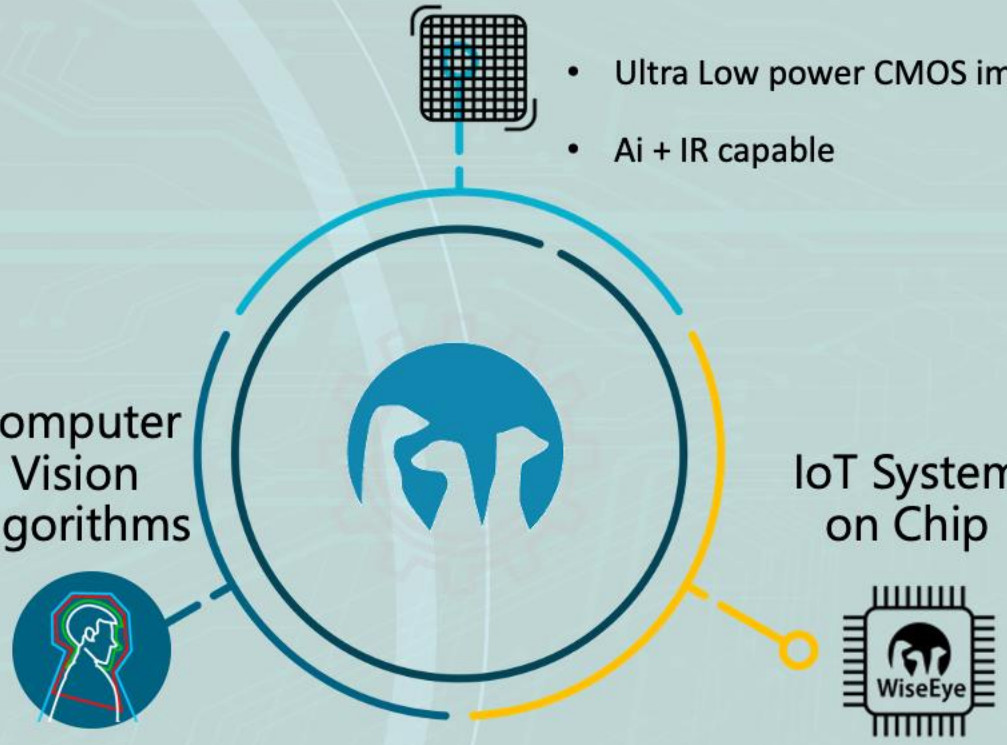
An impulse takes raw data, uses signal processing to extract features, and then uses a learning block to classify new data.

www.edgeimpulse.com

emza
visual sense

The Eye in IoT

Edge AI Visual Sensors



CMOS Imaging Sensor

- Ultra Low power CMOS imager
- Ai + IR capable

Computer Vision Algorithms

- Machine Learning algorithm
- <1MB memory footprint
- Microcontrollers computing power
- Trained algorithm
- Processing of low-res images
- Human detection and other classifiers

IoT System on Chip

- Machine Learning edge computing silicon
- <1mW always-on power consumption
- Computer Vision hardware accelerators

info@emza-vs.com

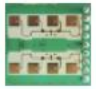
Enabling the next generation of Sensor and Hearable products to process rich data with energy efficiency

Visible Image

Sound

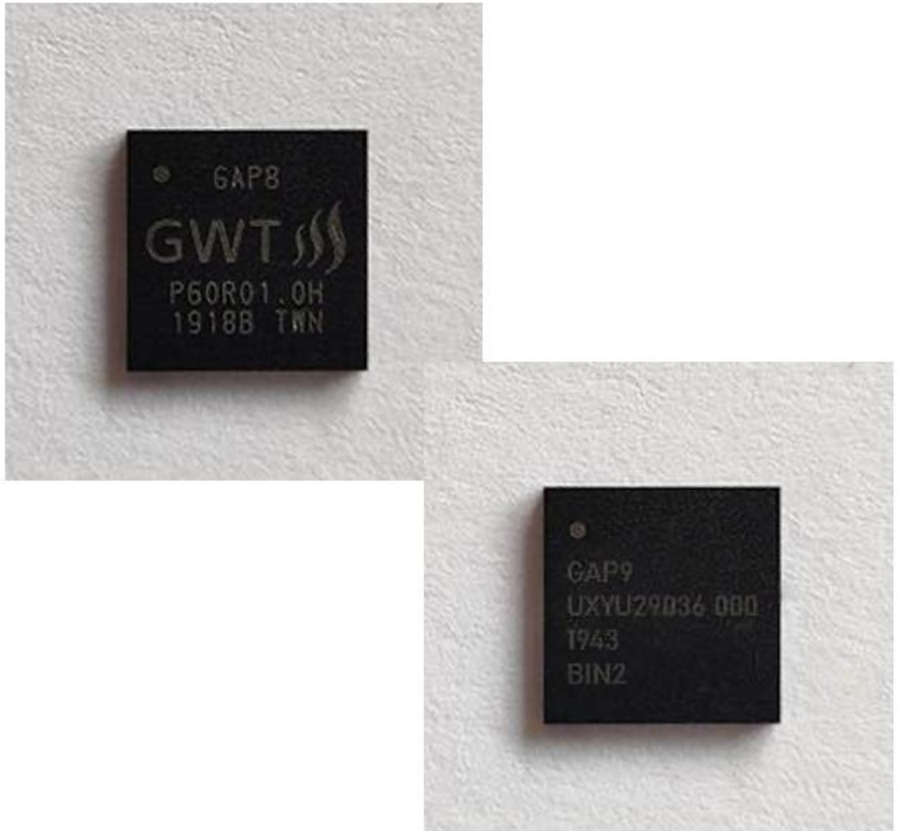
IR Image

Radar



Bio-sensor

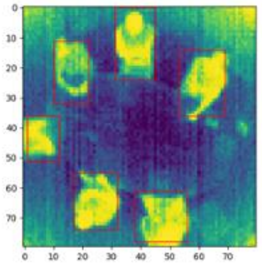
Gyro/Accel



Wearables / Hearables

Battery-powered consumer electronics

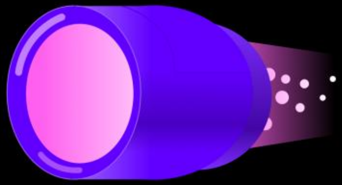
IoT Sensors

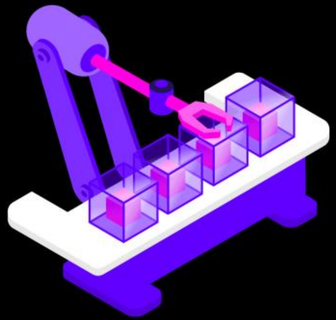


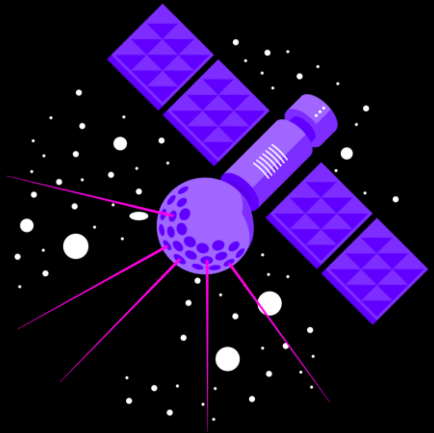
GREENWAVES
TECHNOLOGIES

Distributed infrastructure for TinyML apps

HOTG
Decoupling intelligence







Develop at warp speed

Automate deployments

Device orchestration

HOTG is building the distributed infrastructure to pave the way for AI enabled edge applications

LatentAI

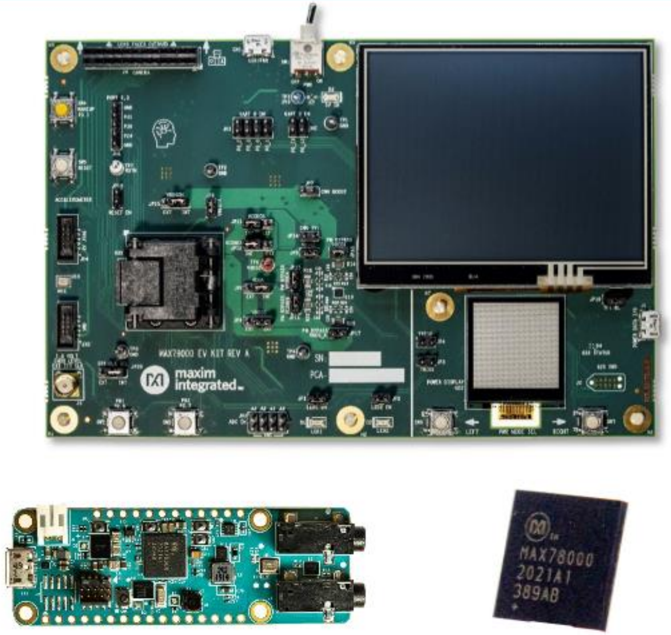
Adaptive AI for the Intelligent Edge

Latentai.com

maxim integrated™

Maxim Integrated: Enabling Edge Intelligence

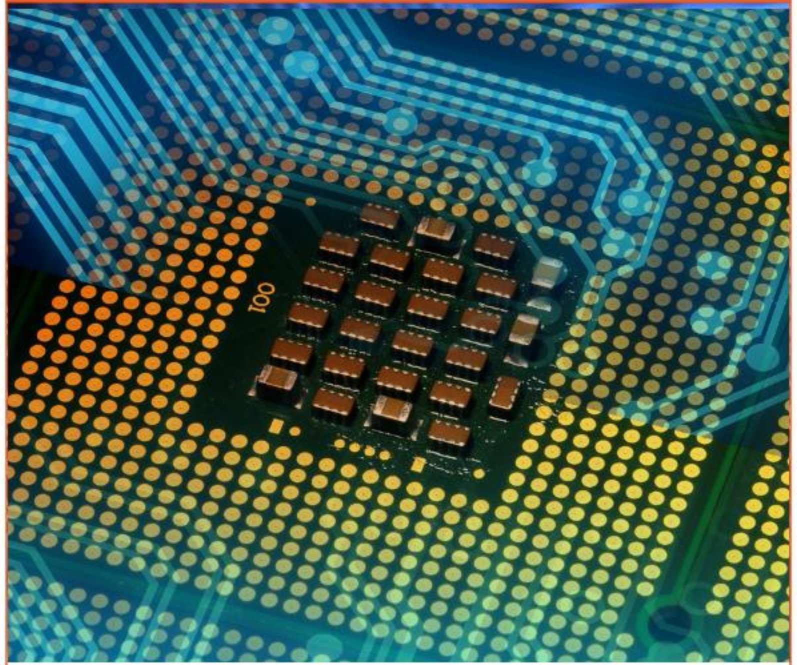
Advanced AI Acceleration IC



The new MAX78000 implements AI inferences at low energy levels, enabling complex audio and video inferencing to run on small batteries. Now the edge can see and hear like never before.

www.maximintegrated.com/MAX78000

Low Power Cortex M4 Micros



Large (3MB flash + 1MB SRAM) and small (256KB flash + 96KB SRAM, 1.6mm x 1.6mm) Cortex M4 microcontrollers enable algorithms and neural networks to run at wearable power levels.

www.maximintegrated.com/microcontrollers

Sensors and Signal Conditioning

Health sensors measure PPG and ECG signals critical to understanding vital signs. Signal chain products enable measuring even the most sensitive signals.

www.maximintegrated.com/sensors

Qeexo AutoML

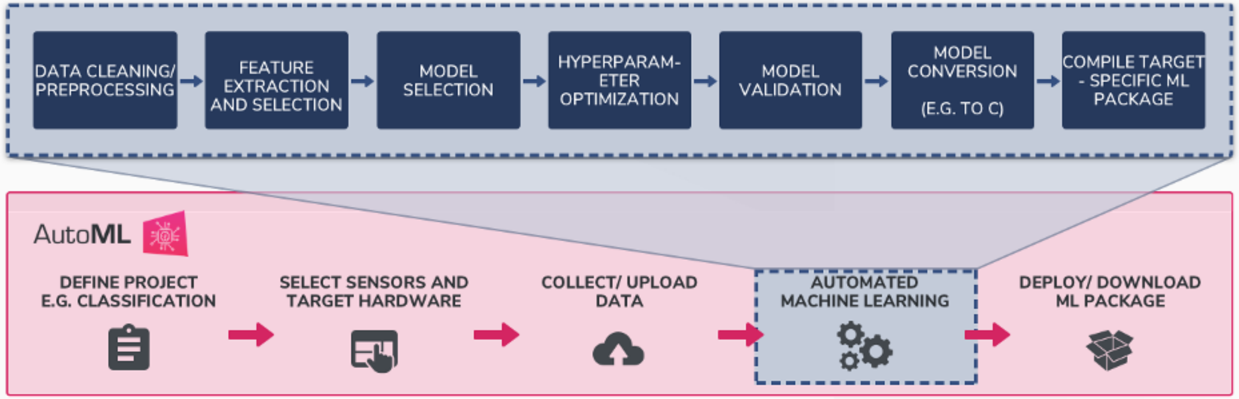
Automated Machine Learning Platform that builds tinyML solutions for the Edge using sensor data

CES INNOVATION AWARDS 2021 HONOREE

Key Features

- Supports 17 ML methods:
  - Multi-class algorithms: GBM, XGBoost, Random Forest, Logistic Regression, Gaussian Naive Bayes, Decision Tree, Polynomial SVM, RBF SVM, SVM, CNN, RNN, CRNN, ANN
  - Single-class algorithms: Local Outlier Factor, One Class SVM, One Class Random Forest, Isolation Forest
- Labels, records, validates, and visualizes time-series sensor data
- On-device inference optimized for low latency, low power consumption, and small memory footprint applications
- Supports Arm® Cortex™- M0 to M4 class MCUs

End-to-End Machine Learning Platform



DATA CLEANING/ PREPROCESSING → FEATURE EXTRACTION AND SELECTION → MODEL SELECTION → HYPERPARAMETER OPTIMIZATION → MODEL VALIDATION → MODEL CONVERSION (E.G. TO C) → COMPILE TARGET - SPECIFIC ML PACKAGE

AutoML

DEFINE PROJECT E.G. CLASSIFICATION → SELECT SENSORS AND TARGET HARDWARE → COLLECT/ UPLOAD DATA → AUTOMATED MACHINE LEARNING → DEPLOY/ DOWNLOAD ML PACKAGE

For more information, visit: www.qeexo.com

Target Markets/Applications

- Industrial Predictive Maintenance
- Smart Home
- Wearables
- Automotive
- Mobile
- IoT

Qualcomm
AI research

Advancing AI research to make efficient AI ubiquitous

Power efficiency

Model design, compression, quantization, algorithms, efficient hardware, software tool

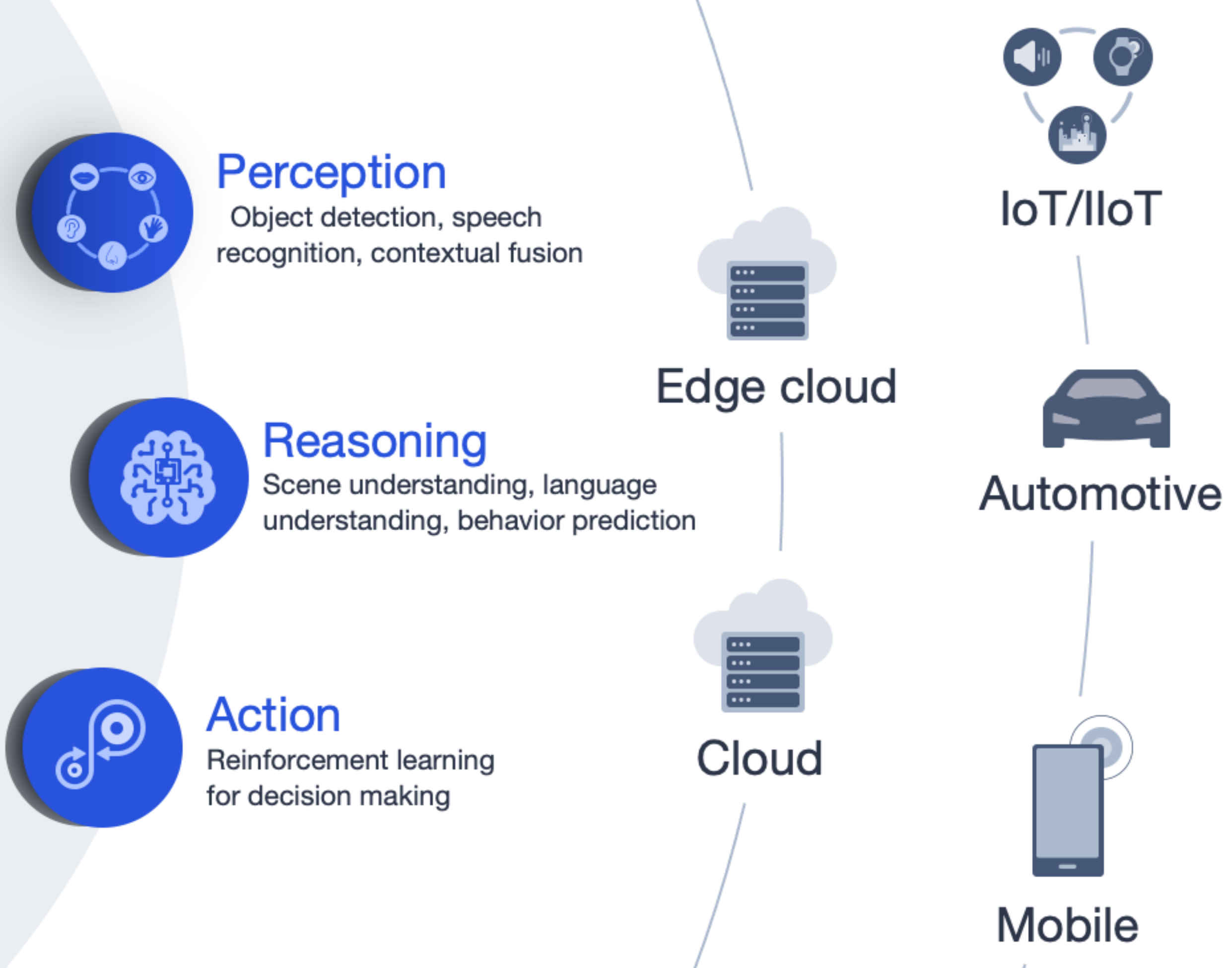
Personalization

Continuous learning, contextual, always-on, privacy-preserved, distributed learning

Efficient learning

Robust learning through minimal data, unsupervised learning, on-device learning

A platform to scale AI across the industry



Perception
Object detection, speech recognition, contextual fusion

Reasoning
Scene understanding, language understanding, behavior prediction

Action
Reinforcement learning for decision making

Qualcomm AI Research is an initiative of Qualcomm Technologies, Inc.

Reality AI®

Add Advanced Sensing to your Product with Edge AI / TinyML

https://reality.ai ✉ info@reality.ai @SensorAI Reality AI

Pre-built Edge AI sensing modules, plus tools to build your own

Reality AI solutions

Prebuilt sound recognition models for indoor and outdoor use cases

Solution for industrial anomaly detection

Pre-built automotive solution that lets cars “see with sound”

Reality AI Tools® software

Build prototypes, then turn them into real products

Explain ML models and relate the function to the physics

Optimize the hardware, including sensor selection and placement

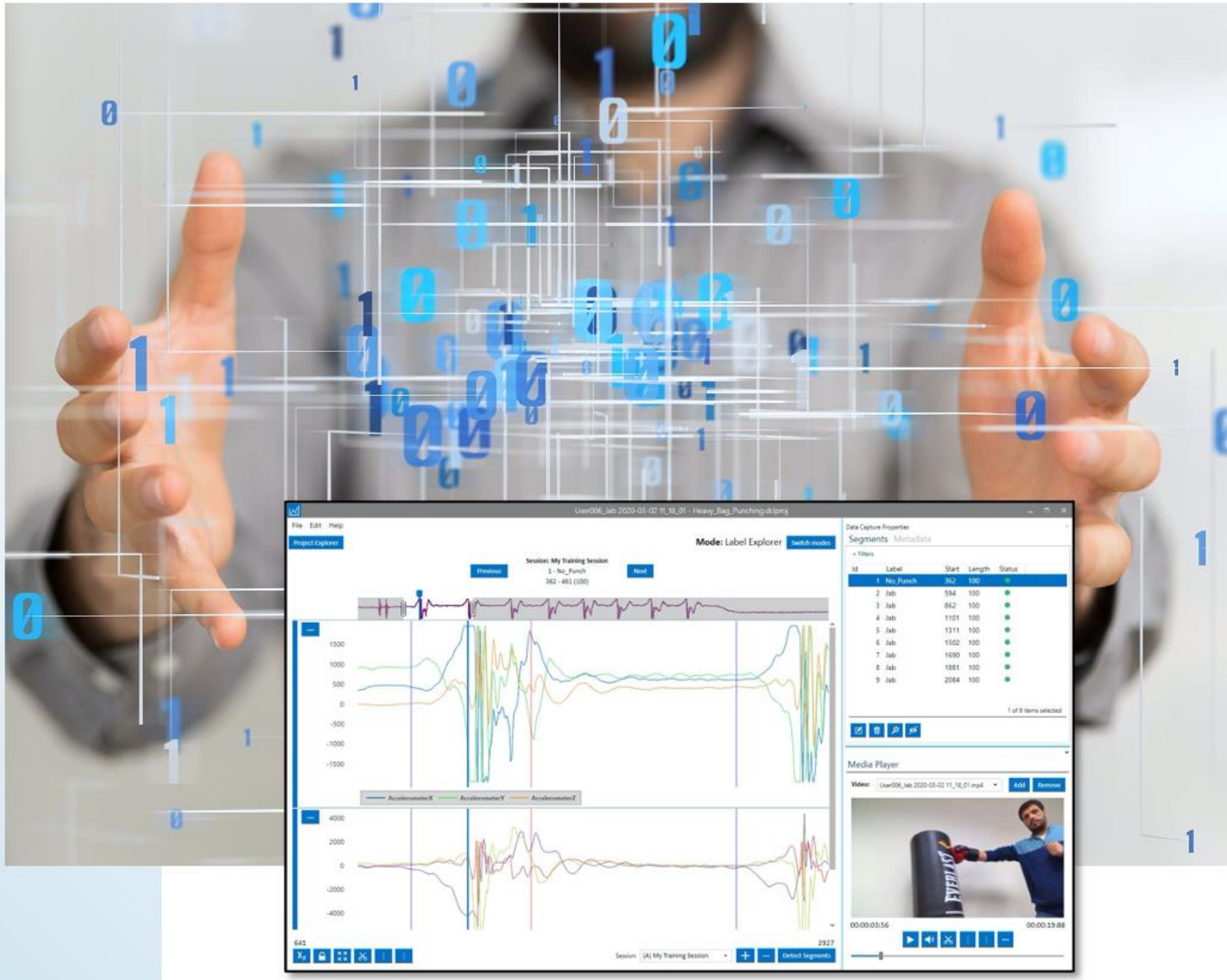
SensiML™

Build Smart IoT Sensor Devices From Data

SensiML pioneered TinyML software tools that auto generate AI code for the intelligent edge.

- End-to-end AI workflow
- Multi-user auto-labeling of time-series data
- Code transparency and customization at each step in the pipeline

We enable the creation of production-grade smart sensor devices.



Session: My Training Session
1 - No_Punch
362 - 461 (100)

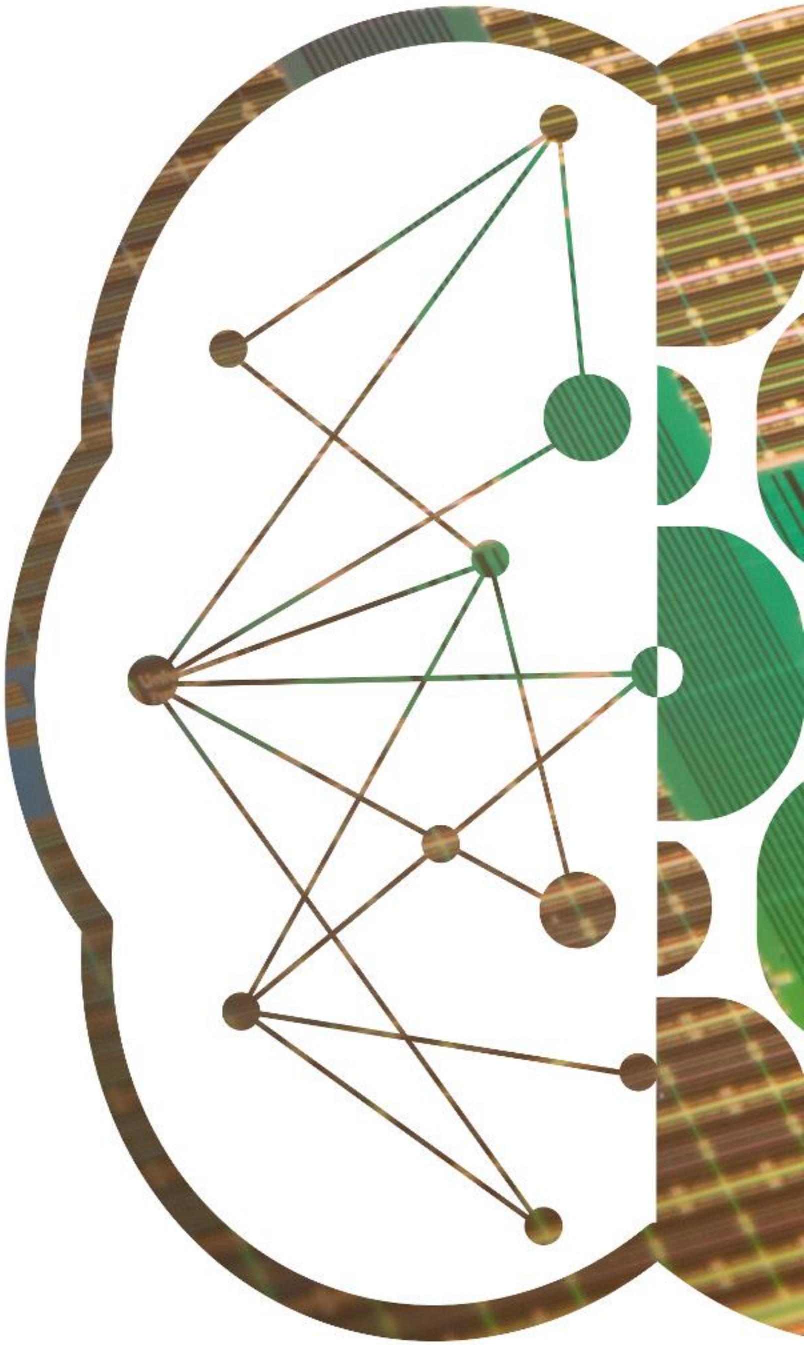
Id	Label	Start	Length	Status
1	No_Punch	362	100	●
2	Jab	594	100	●
3	Jab	862	100	●
4	Jab	1101	100	●
5	Jab	1311	100	●
6	Jab	1502	100	●
7	Jab	1690	100	●
8	Jab	1881	100	●
9	Jab	2084	100	●

sensiml.com

SynSense

SynSense builds **sensing and inference** hardware for **ultra-low-power** (sub-mW) **embedded, mobile and edge** devices. We design systems for **real-time always-on smart sensing**, for audio, vision, IMUs, bio-signals and more.

https://SynSense.ai



SYNTIANT

End-to-End Deep Learning Solutions

for

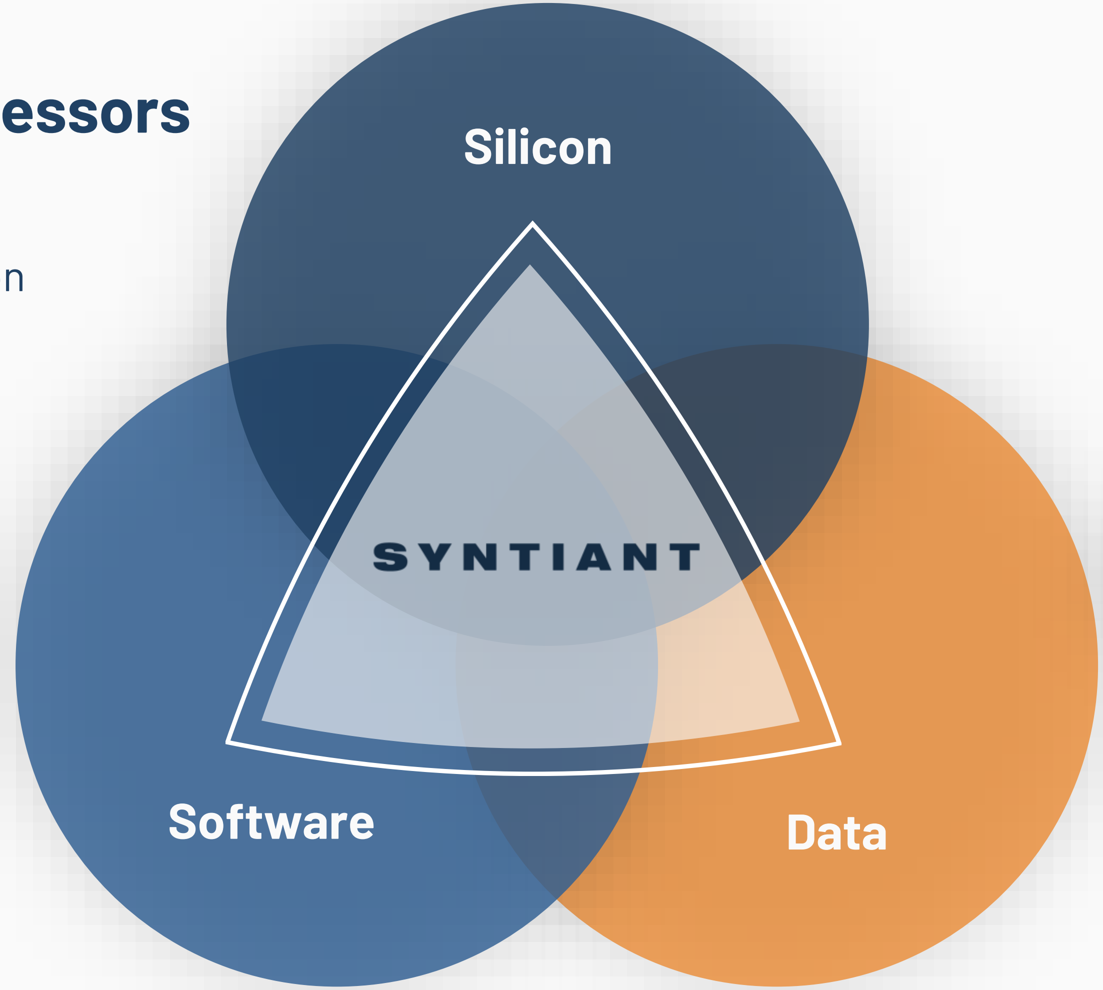
TinyML & Edge AI

Neural Decision Processors

- At-Memory Compute
- Sustained High MAC Utilization
- Native Neural Network Processing

ML Training Pipeline

- Enables Production Quality Deep Learning Deployments



Data Platform

- Reduces Data Collection Time and Cost
- Increases Model Performance

partners@syntiant.com

www.syntiant.com

TINY ML ASIA

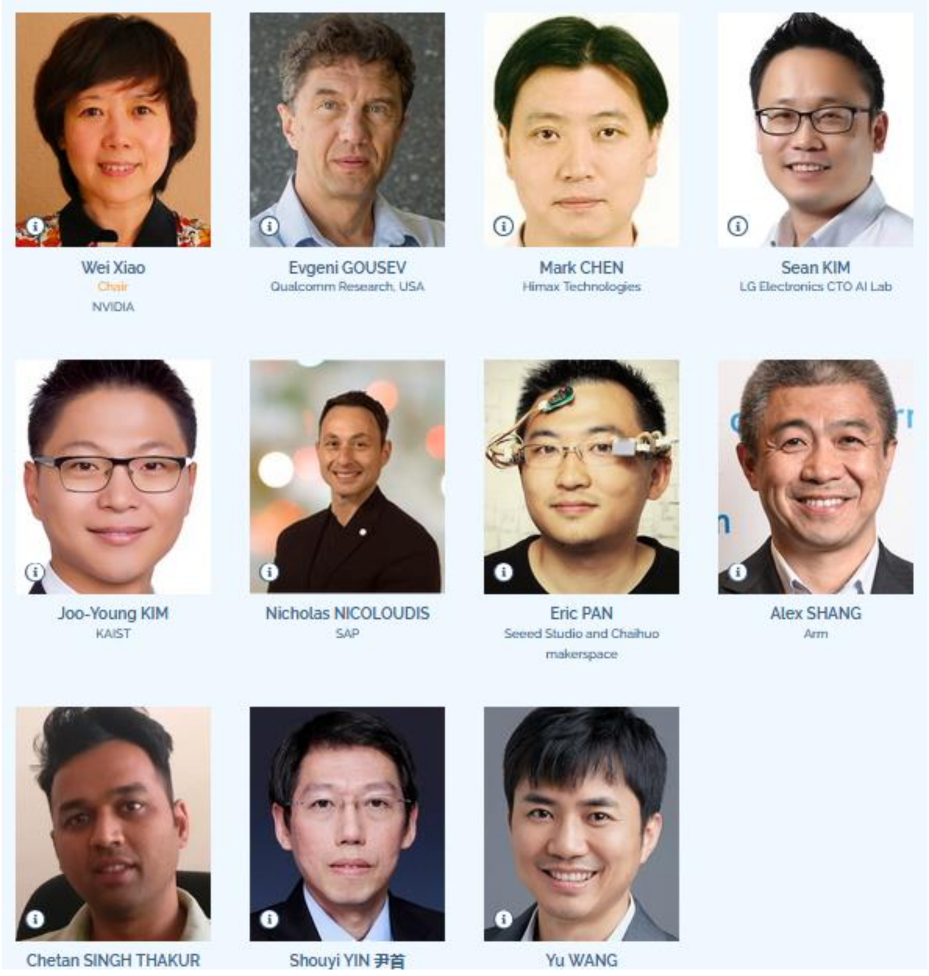
LIVE ONLINE **November 2-5, 2021**

(9-11:30 am China Standard time)

https://www.tinyml.org/event/asia-2021/

Register today!

Technical Programm Committee



Wei Xiao
Chair
NVIDIA

Evgeni GOUSEV
Qualcomm Research, USA

Mark CHEN
Himax Technologies

Sean KIM
LG Electronics CTO AI Lab

Joo-Young KIM
KAIST

Nicholas NICOLOUDIS
SAP

Eric PAN
Seeed Studio and Chaihuo makerspace

Alex SHANG
Arm

Chetan SINGH THAKUR

Shouyi YIN 尹首

Yu WANG

Free event courtesy of our sponsors and strategic partners

arm · EDGE IMPULSE · emza visual sense · GREENWAVES TECHNOLOGIES

HOTG · imagimob · LatentAI Adaptive AI for a Smarter Edge · Qualcomm

Qeexo · Reality AI · seeed studio The IoT Hardware Enabler · SensiML

SynSense · SYNTIANT

More sponsorships are available: sponsorships@tinyML.org

TINY ML

TINY ML FOUNDATION

collaboration with hackster.io

AN AVNET COMMUNITY

Focus on:
(i) developing new use cases/apps for tinyML vision; and (ii) promoting tinyML tech & companies in the developer community

Submissions accepted until September 17th, 2021
Winners announced on October 5th, 2021 ($6k value)
Sponsorships available: *sponsorships@tinyML.org*

https://www.hackster.io/contests/tinyml-vision

arm
EDGE IMPULSE
Himax
Qualcomm
intel

LATTICE SEMICONDUCTOR
PixArt Imaging Inc. SEE. FEEL. TOUCH.
SynSense
LUXonis

Next tinyML Talks

Date	Presenter	Topic / Title
Friday, September 24	**Peter Ing,** Edge Impulse	An Introduction to TinyML for all backgrounds with hands on introduction to Edge Impulse

Webcast start time is 8 am Pacific time

Please contact talks@tinyml.org if you are interested in presenting

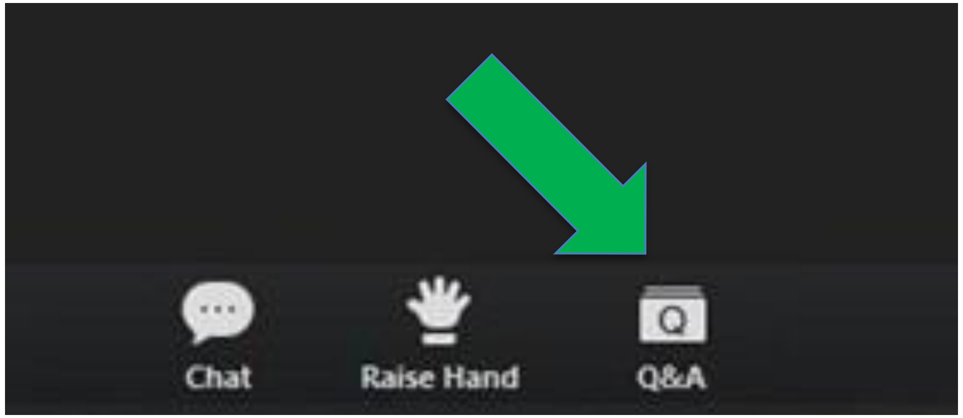
Reminders

Slides & Videos will be posted tomorrow

tinyml.org/forums

youtube.com/tinyml

Please use the Q&A window for your questions



TINY ML TALKS webcast

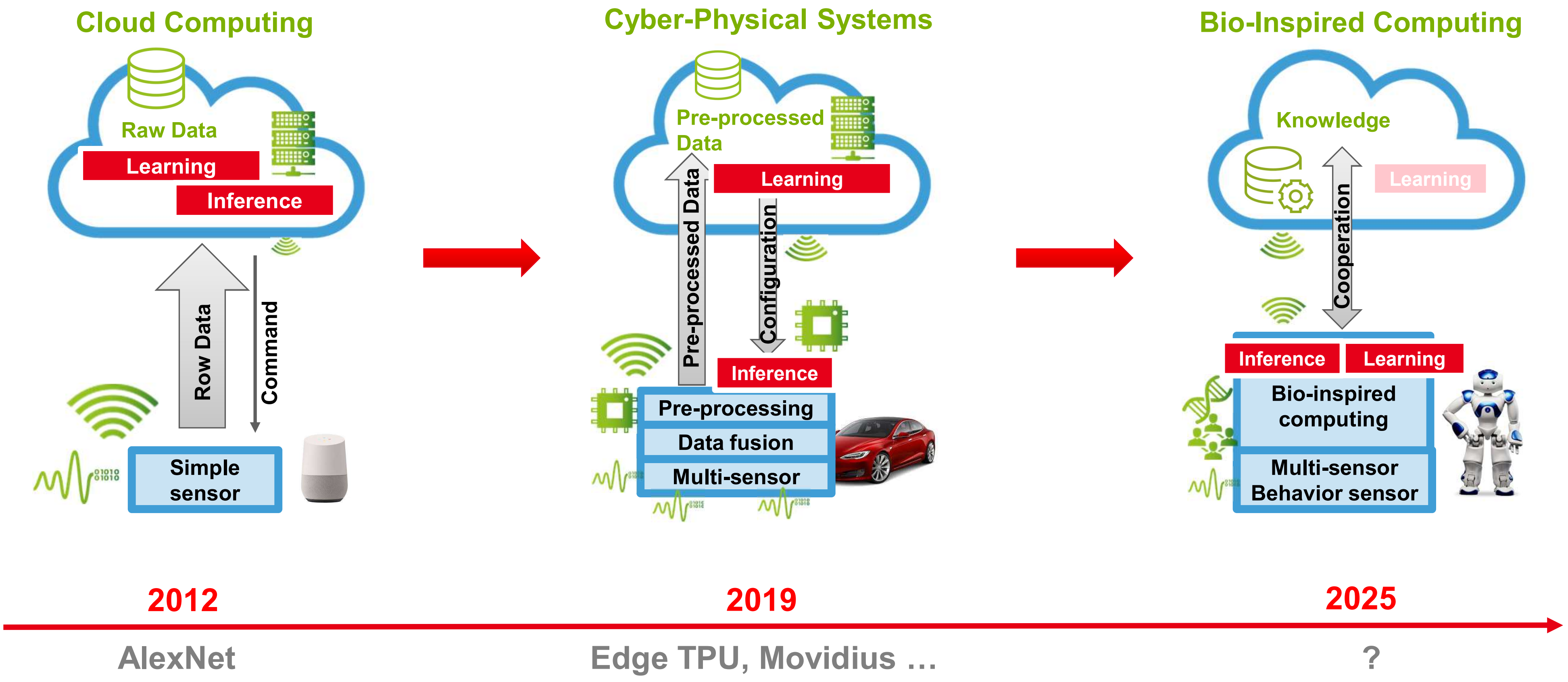
Alexandre Valentian

After an MSc and a PhD in microelectronics, Alexandre Valentian joined CEA LETI in 2005. His past research activities included design technology co-optimization, promoting the FDSOI technology (notably through his participation in the SOI Academy), 2.5D/3D integration technologies and non-volatile memory technology. He is currently pursuing the development of bio-inspired circuits for AI, combining memory technology, information encoding and dedicated learning methods. Since 2020, he heads the Systems-on-Chip and Advanced Technologies (LSTA) laboratory. Dr Valentian has authored or co-authored 80 conference and journal papers.

HOW TO DESIGN A POWER FRUGAL HARDWARE FOR AI – THE BIO-INSPIRATION PATH

Alexandre VALENTIAN

TRENDS IN AI COMPUTING



Cloud Computing

Raw Data

Learning

Inference

Row Data

Command

Simple sensor

Cyber-Physical Systems

Pre-processed Data

Learning

Pre-processed Data

Configuration

Inference

Pre-processing

Data fusion

Multi-sensor

Bio-Inspired Computing

Knowledge

Learning

Cooperation

Inference

Learning

Bio-inspired computing

Multi-sensor Behavior sensor

2012

AlexNet

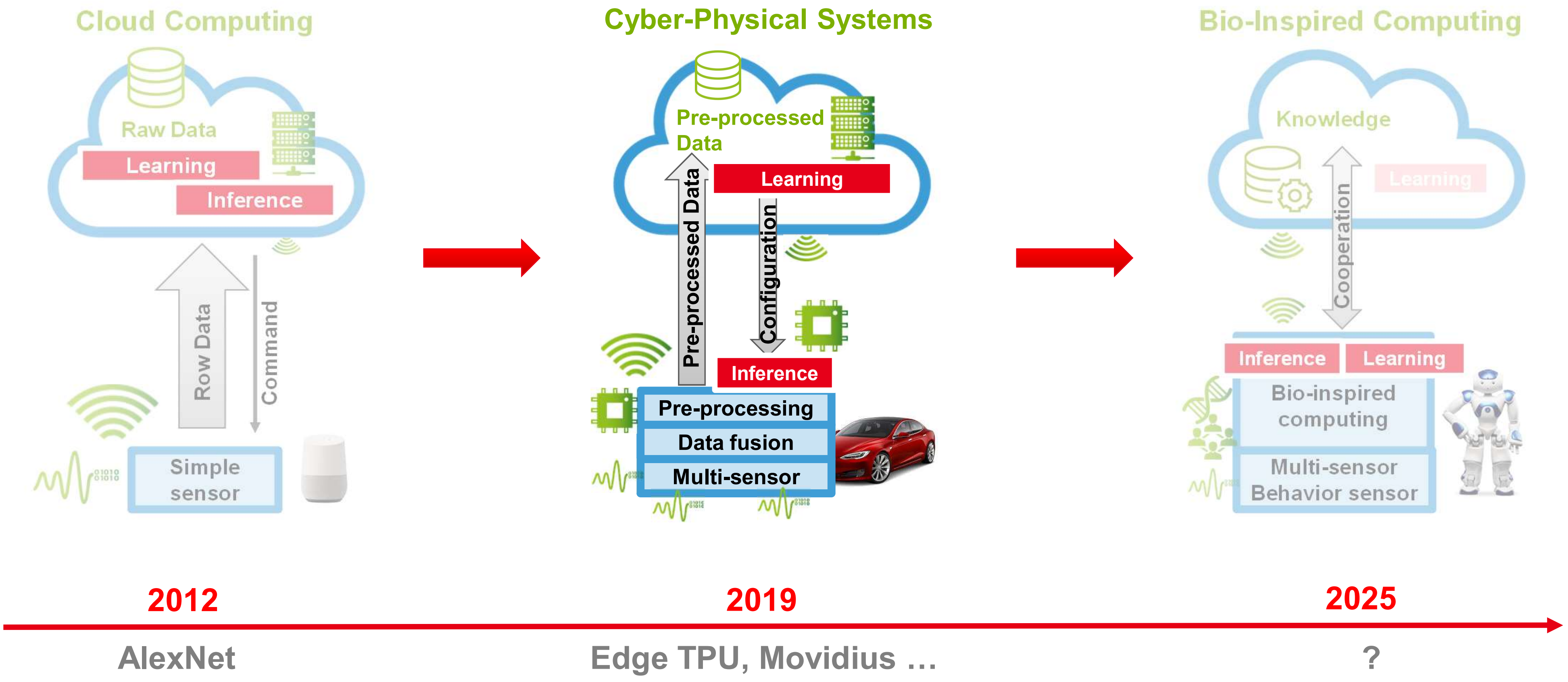
2019

Edge TPU, Movidius …

2025

?

TRENDS IN AI COMPUTING



Cloud Computing

Raw Data

Learning

Inference

Row Data

Command

Simple sensor

Cyber-Physical Systems

Pre-processed Data

Learning

Pre-processed Data

Configuration

Inference

Pre-processing

Data fusion

Multi-sensor

Bio-Inspired Computing

Knowledge

Learning

Cooperation

Inference

Learning

Bio-inspired computing

Multi-sensor Behavior sensor

2012

AlexNet

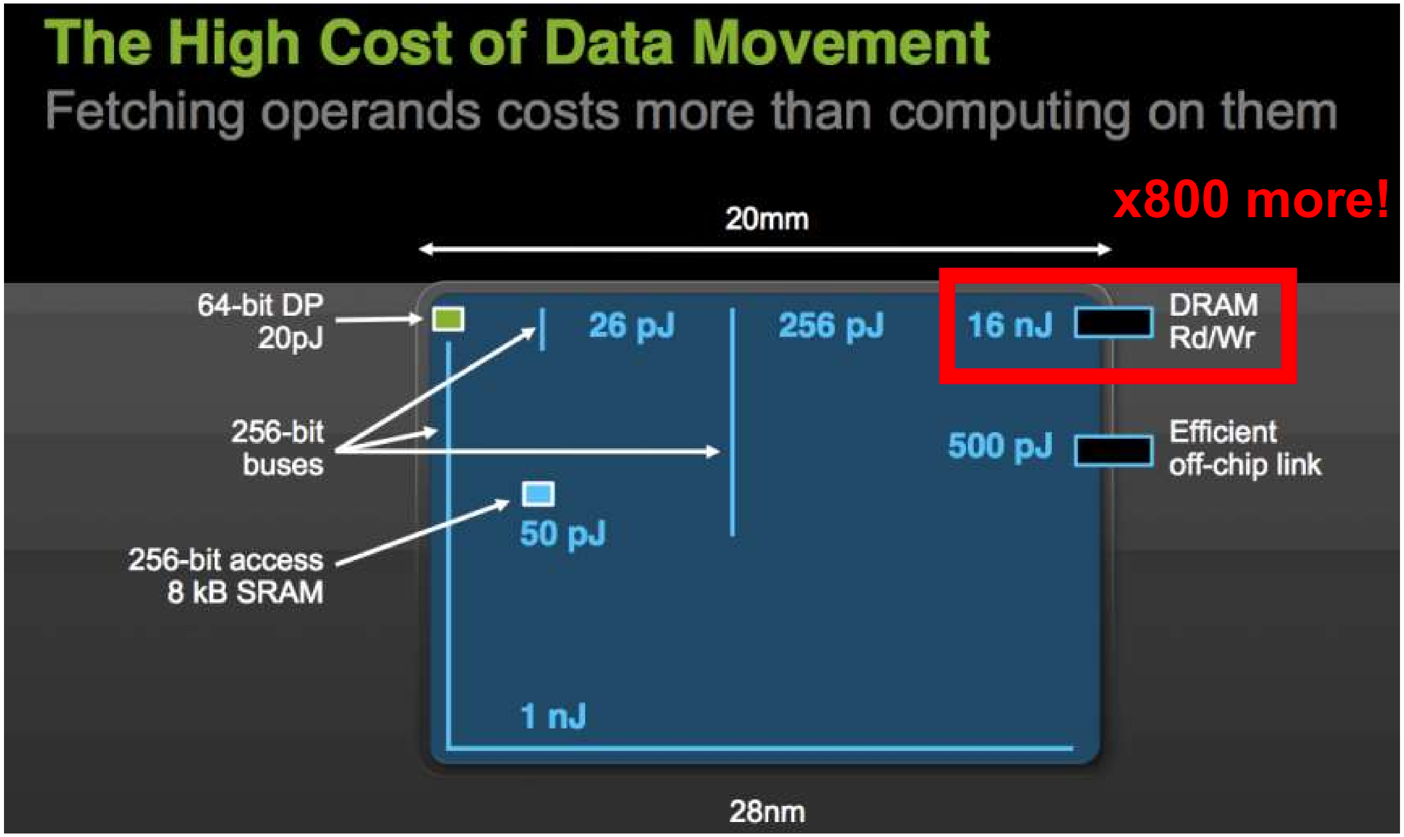
2019

Edge TPU, Movidius …

2025

?

SOLVING THE ENERGY CHALLENGE: COST OF MOVING DATA



The High Cost of Data Movement

Fetching operands costs more than computing on them

x800 more!

Bill Dally, *“To ExaScale and Beyond”*, 2010

TRENDS IN EDGE COMPUTING

Increased computing efficiency

Weight quantization

Reduced bit accuracy

- Smaller memory footprint
- Lighter operations

Variable bit precision

Handling higher bit accuracy when needed

- For higher inference precision

Sparsity

Skip MAC operations

- When weight or intermediate result is 0

Increased storage efficiency

Near memory computing

Avoid external memory accesses

Weights

- Embedded Non-Volatile Memory

Intermediate results

- SRAM or Embedded DRAM

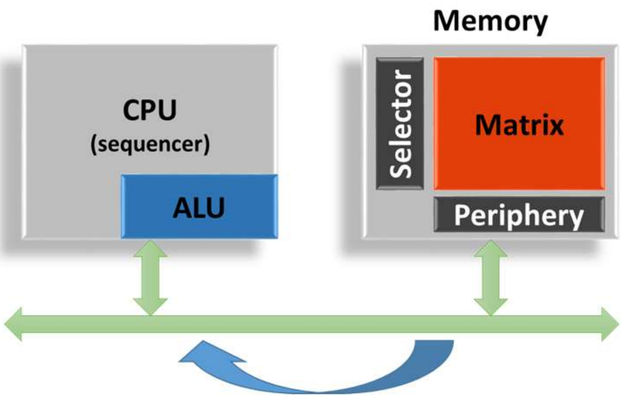
In-Memory computing

SRAM or Embedded NVM

Digital or analog

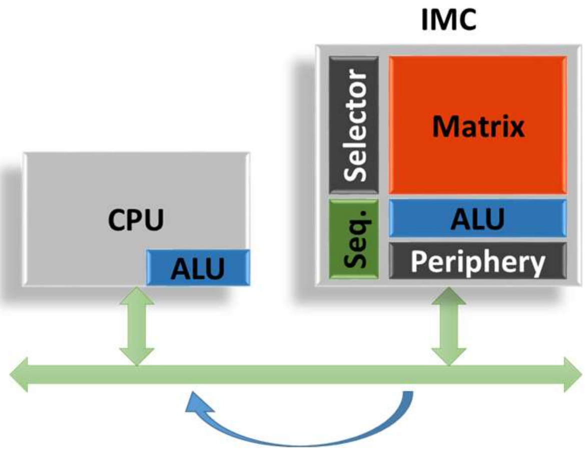
ACTIVATIONS – IN-MEMORY COMPUTING

Von Neumann architecture



High data transfer

In-Memory Computing (IMC) architecture



Low data transfer

Matrix Multiplication

	Clock Cycle	Energy
Matrix Size	Scalar vs. CSRAM	Scalar vs. CSRAM
8x8 8-bit integer (512 bits)	x256	X9

WEIGHTS – NON-VOLATILE MEMORIES

RRAM technology compatible with advanced logic

- Scalable to sub-20nm

Multilevel cell *

- From 1 bit to 4 bit and beyond

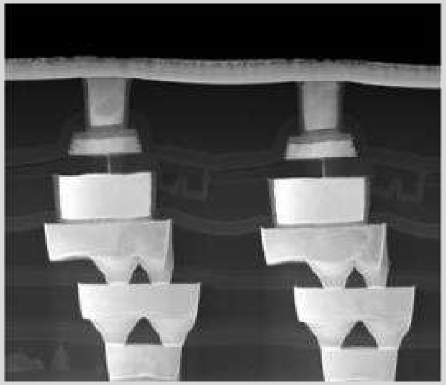
* *[T. Wu, ISSCC 2019]*

Roadmap for increasing embedded cell density

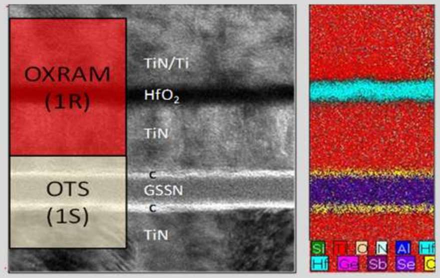
- From 40F² down to 4F² thanks to new selector technology

Examples with technology available today

- ResNet50 (74 MB of weights) → 15 mm² of memory
- YoloV3 (101MB of weights) → 20 mm² of memory

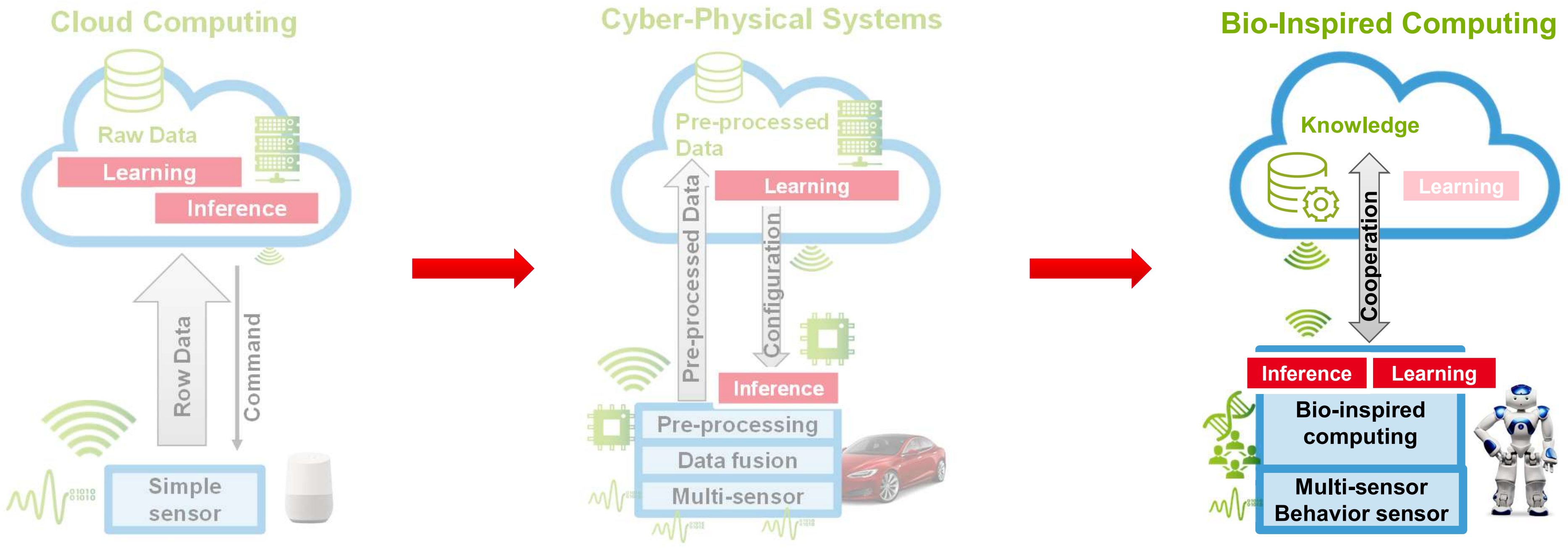


28nm RRAM integration



Selector and RRAM integration
[IEDM 2019]

TRENDS IN AI COMPUTING



2012	2019	2025
AlexNet	Edge TPU, Movidius …	?

CHALLENGE OF ONLINE LEARNING

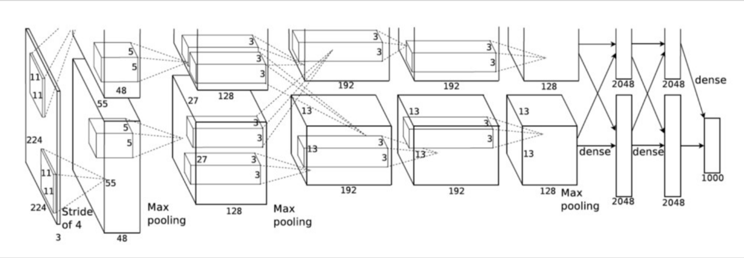
Back-propagation algorithm

- **Necessitates to keep all intermediate results (activations)**
- **With a batch size of more than one**
  - To not cycle too much the non-volatile memories

This requires a tremendous amount of activation memory

- **Example – YoloV3**
  - A batch of 20 images requires 800MB of memory

Feedforward



Back-propagation

TECHNOLOGY SOLUTION – 3D INTEGRATION

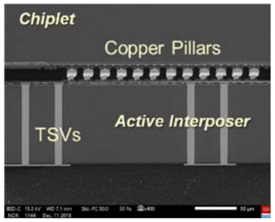
Advantages

- Increasing computing & memory capacity

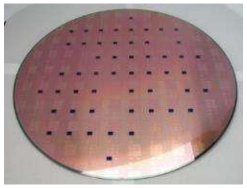
Trends

- « Denser Integration » : tight memory ⬄ logic computing paradigm
- « Chipletization » : Generic computing templates, Heterogeneous technologies

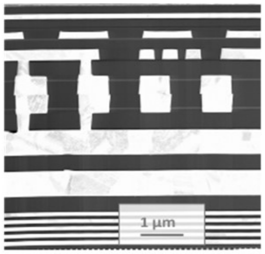
Roadmap



TSV & µ-bumps, pitch 20 µm
[P. Coudrain, ECTC 2019]
[P. Vivet, ISSCC 2020]



Hybrid Bonding
CoW, pitch 3-5 µm
Including K-G-D
[A. Jouve, 3DIC 2019]



Hybrid Bonding
WoW, pitch 1.44 µm
[A. Jourdan, IEDM 2018]

20 µm 5 µm 1 µm

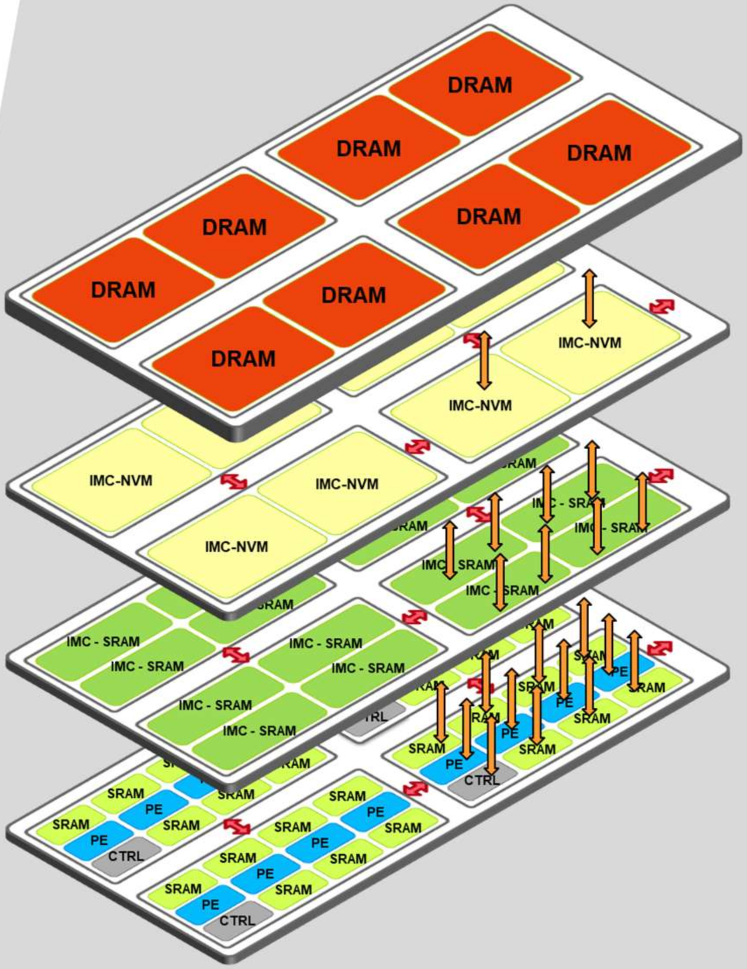
DISTRIBUTED MEMORY-CENTRIC EDGE AI COMPUTING ARCHITECTURE

Memory-Centric architecture

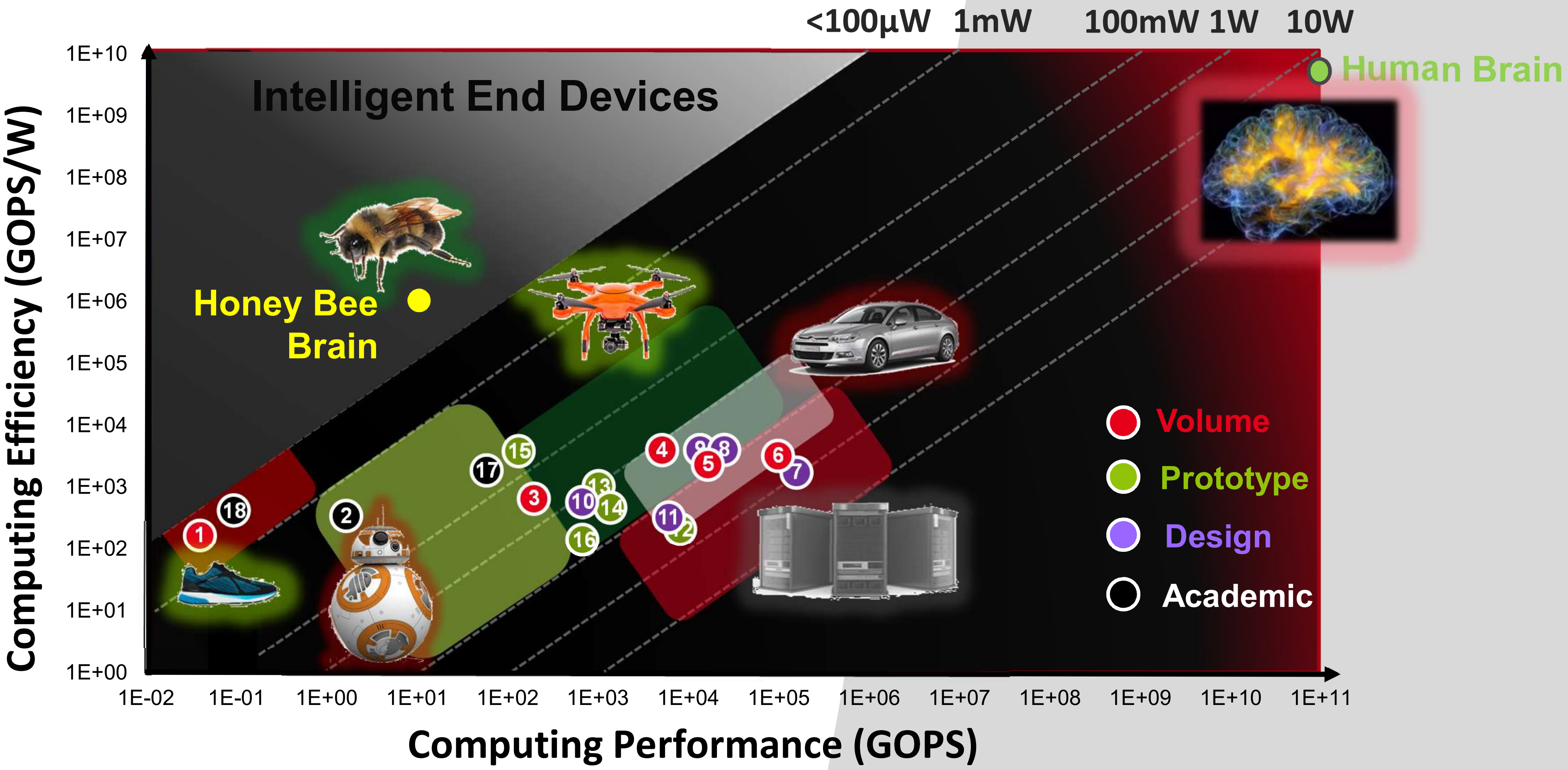
- No more global buffers
- No more power-hungry caches
- Fully distributed memory and control
- Energy efficient use of memory using 3D technology

Edge AI architecture, using

- Generic PE engines
- Vertically and horizontally connected computing clusters
- In-Memory Computing tiles (IMC)
- Dense NVM for storage
- DRAM for online learning



ENERGY EFFICIENCY IS FAR FROM BIOLOGICAL SYSTEMS

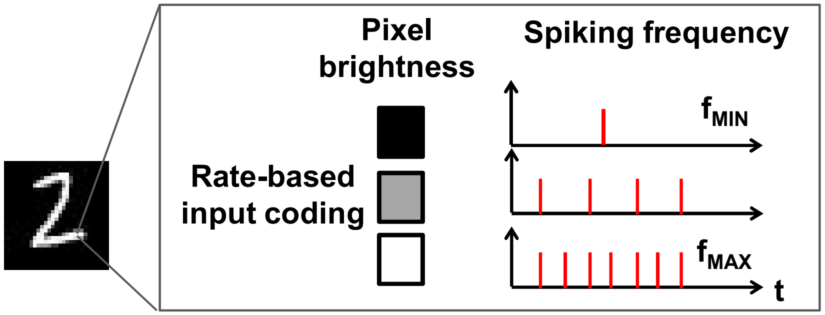


BRAIN-INSPIRED SOLUTIONS MIGHT BE THE KEY

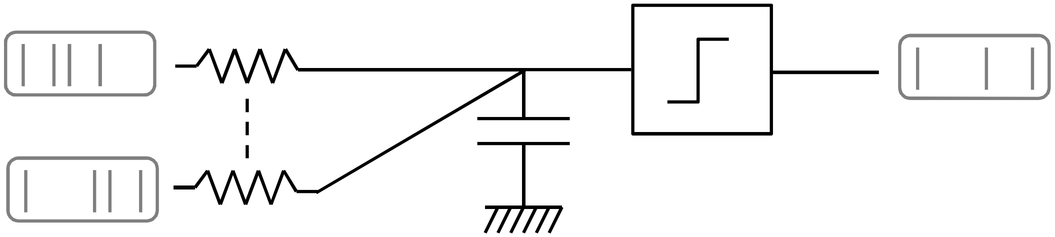
Human brain		*Brain inspired*
Massively parallel - 10^{11} neurons and 10^{15} synapses	→	**High density storage, close to neurons**
Doing processing using memory elements	→	**Computational storage**
Analog computation - Neuron soma = synaptic current integrator	→	**Analog neuron**
Digital communication - Spikes = unary events, very robust to noise	→	**Spike coding**

PROOF-OF-CONCEPT CIRCUIT

- **Spike coding**
- **RRAM synapses**
  - Weighted input thanks to Ohm's law
- **Analog neurons**
  - Inputs summation thanks to Kirchhoff's law



Frequency coding of pixel intensity

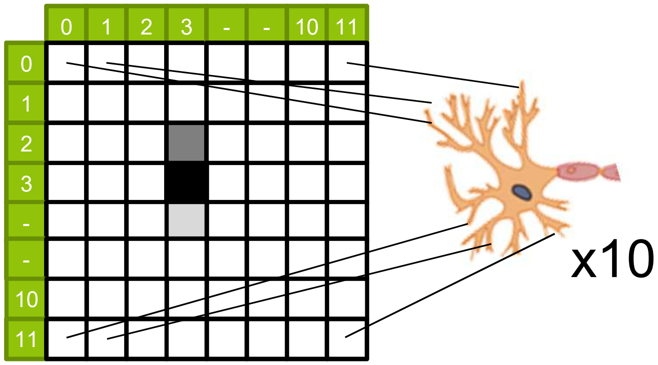


Simplified schematic view

NEURAL NETWORK TOPOLOGY

- **Fully-connected neural network topology**
  - 10 output neurons: 1 neuron / class
  - Each neuron is connected to the entire image: 144 synapses

Image 12*12



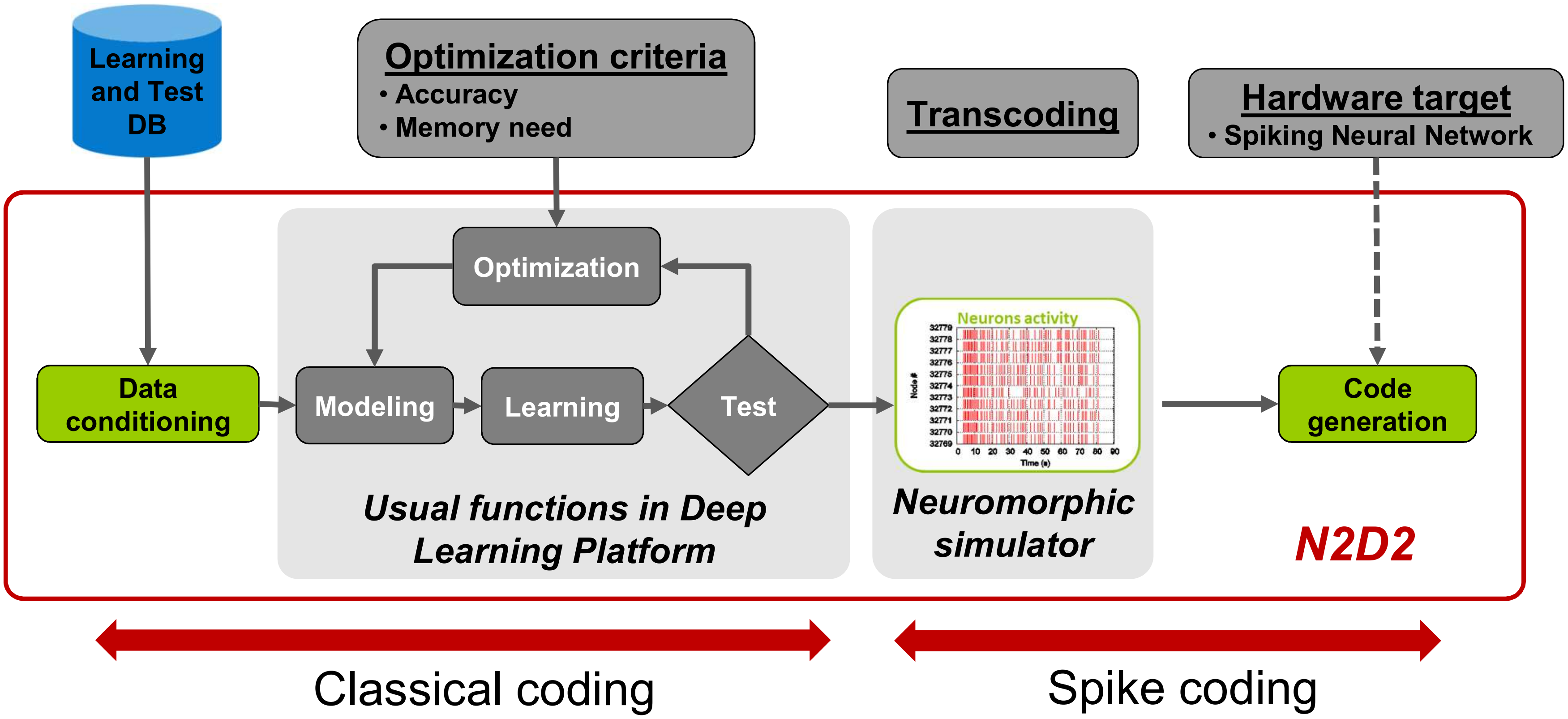
Fully-connected topology

LEARNING STRATEGY

- **Bio-inspired unsupervised learning rules**
  - Such as the Spike Timing Dependent Plasticity one
- **Give poorer results than the Gradient Descent algorithm**

- **Decision was made to do offline learning**
  - In the classical coding domain
- **And then to transcode into spikes**

LEARNING FRAMEWORK – N2D2



Learning and Test DB

Optimization criteria
- Accuracy
- Memory need

Transcoding

Hardware target
- Spiking Neural Network

Data conditioning → Modeling → Learning → Test → Optimization → Modeling

Code generation

Usual functions in Deep Learning Platform

Neuromorphic simulator

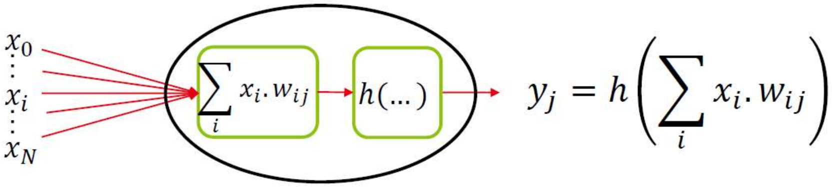
N2D2

Classical coding

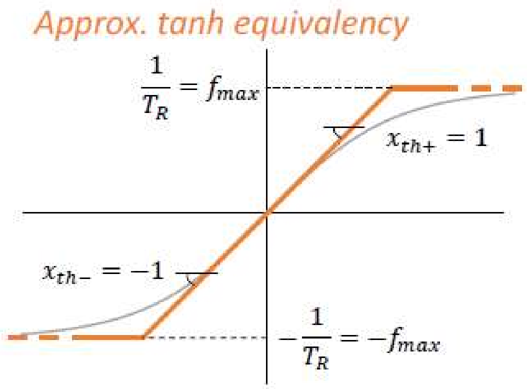
Spike coding

MATHEMATICAL EQUIVALENCE

- **“Classical” neural network model**
  - Multiply-Accumulate (MAC)
  - Non-linear operation (TANH)



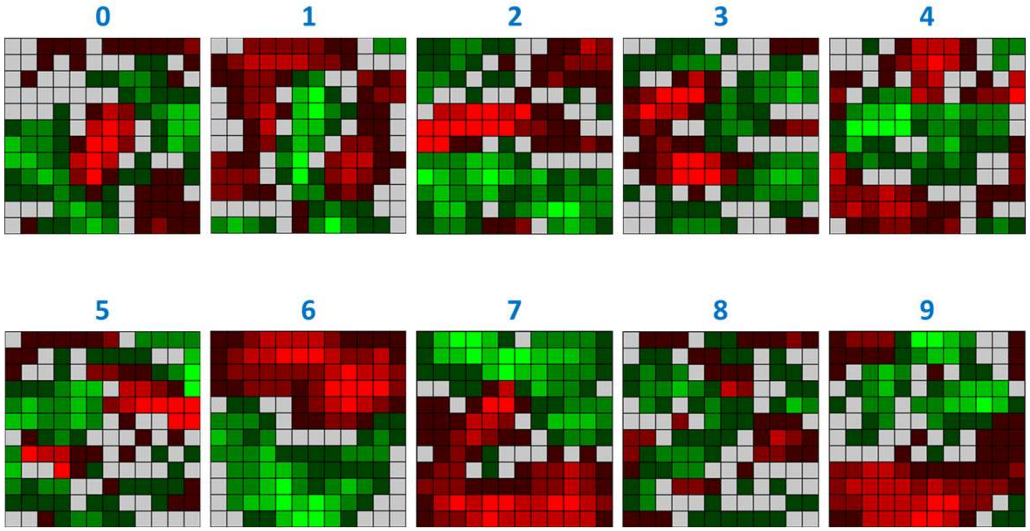
- **“Spiking”, rate-based equivalent model**
  - Integrate & Fire (IF) neuron model
  - Two thresholds, positive and negative
  - Refractory period



LEARNED NEURONS RECEPTIVE FIELDS

- **Excitatory synapses are represented in green**
  - The greener, the higher

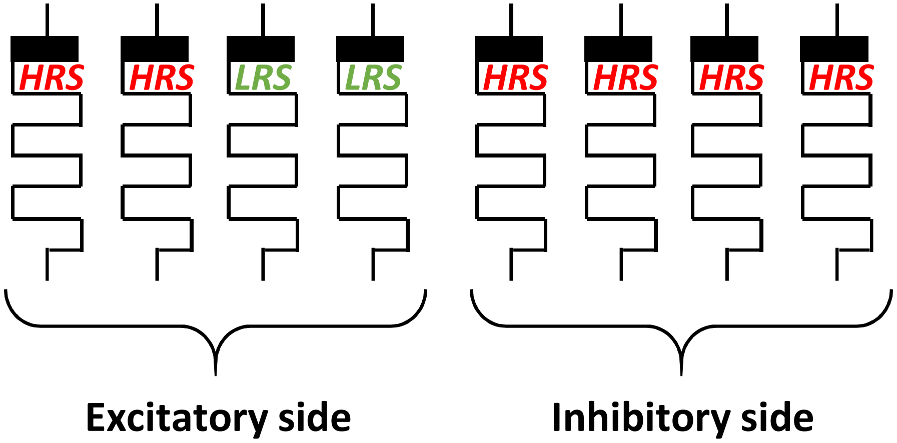
- **Inhibitory synapses are represented in red**
  - The more red, the higher



SYNAPSE IMPLEMENTATION

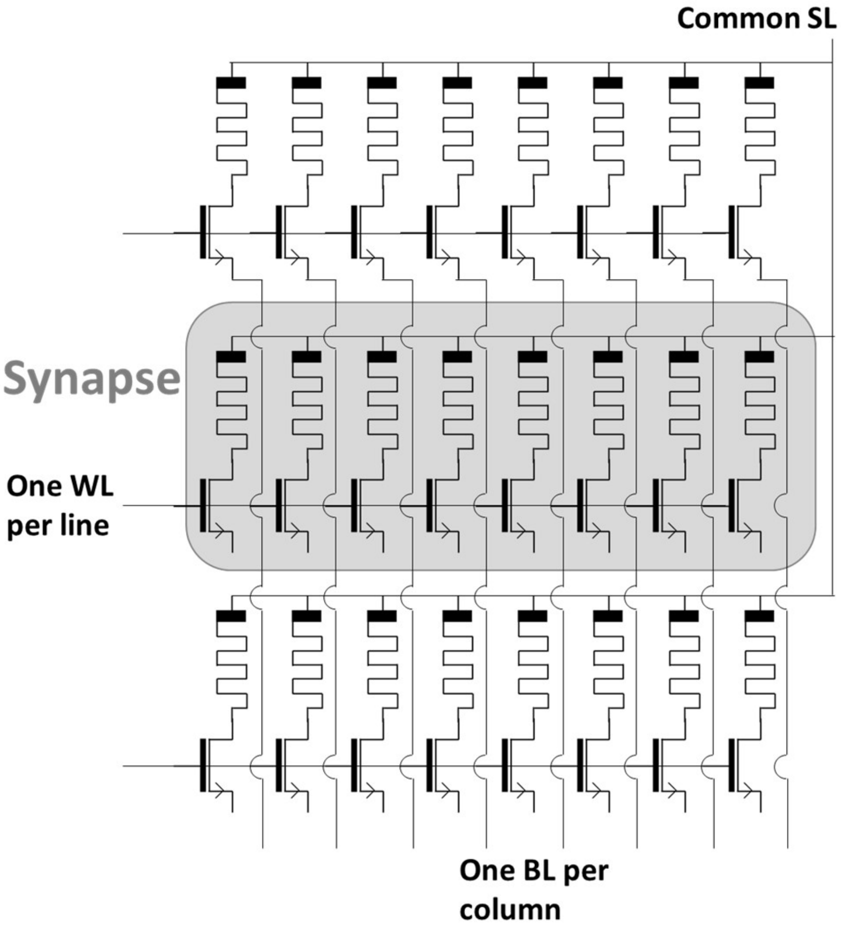
- **RRAMs are used in binary mode (LRS and HRS states)**

- **Four RRAMs encode a positive weight and four others a negative weight**
  - Nine synaptic weights are thus available : -4, -3, -2, -1 ,0, 1, 2, 3, 4



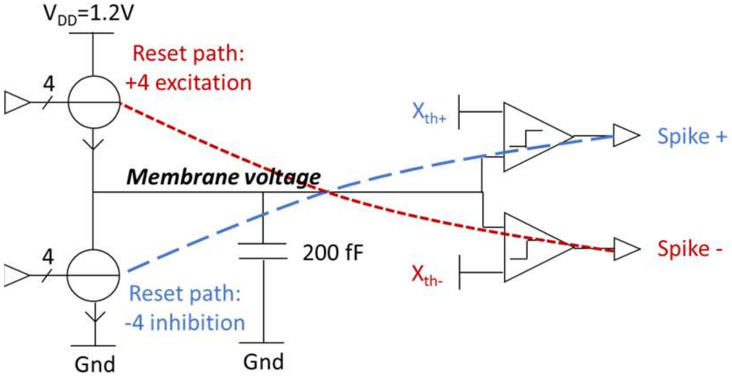
SYNAPTIC MATRIX

- **Synapses are arranged in a matrix, for sharing**
  - Word Line
  - Source Line
  - and Bit Line drivers

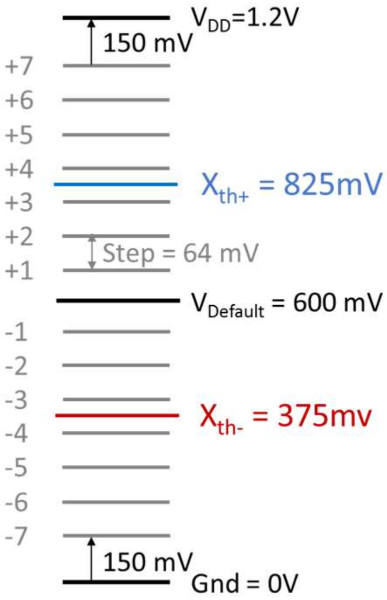


NEURON DESIGN

- **Goal: ensure mathematical equivalence to TANH model**
  - Two thresholds (positive and negative)
  - Peculiar Reset of the membrane voltage



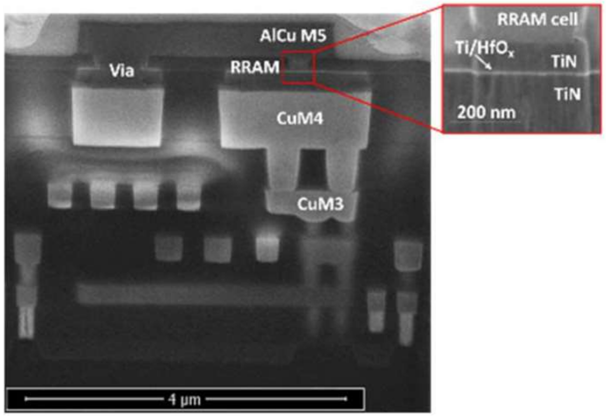
Neuron schematic, with reset paths for ensuring model equivalence



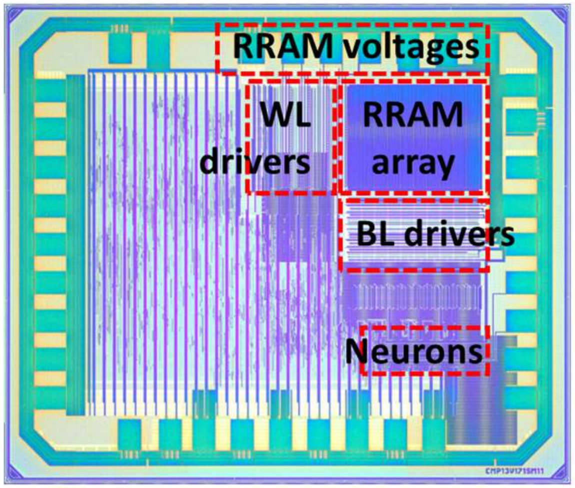
Voltage levels in membrane

FABRICATION DETAILS

- BULK 130nm base wafers
- RRAM Post-process between M4 and M5



Cross-section



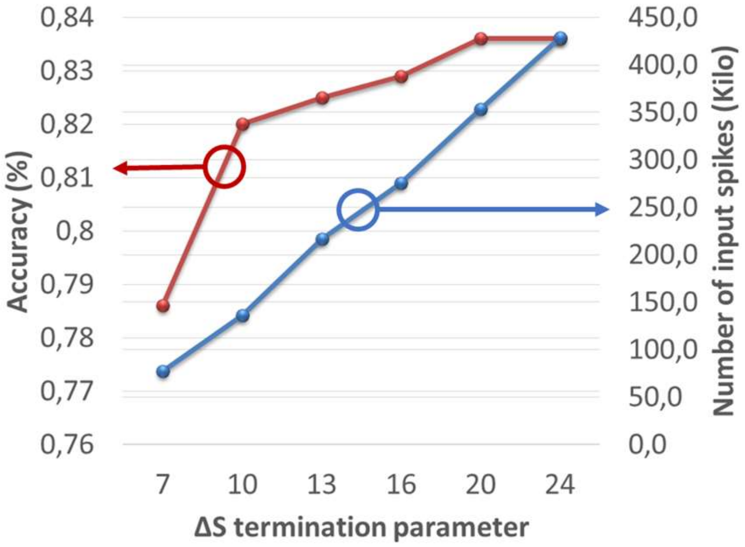
Chip micrograph

CMOS process	130 nm
Cu interconnect	5
RRAM number	11,5 K
RRAM configuration	1T-1R
RRAM size	0.5µm x 0.5µm

MEASUREMENT RESULTS

- **Classification accuracy = 84%**
  - Compared to 88% in simulation

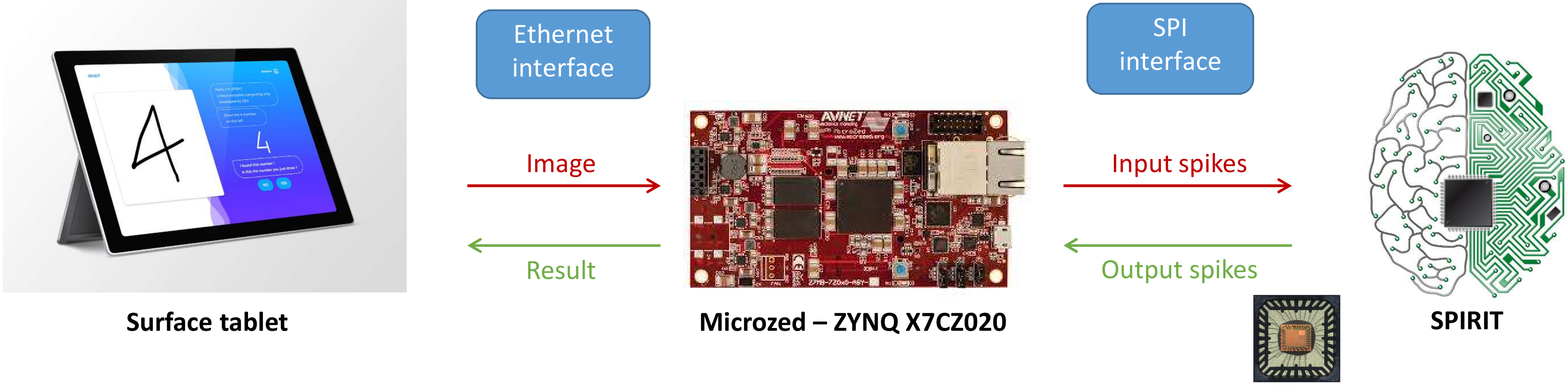
- **Energy**
  - 180pJ / synaptic event
    - 3,6pJ at RRAM + neuron level
  - 136 spikes, on average, to classify an image
    - 24,5nJ / image
  - Energy gain 5X
    - Compared to classical coding



Accuracy / Activity tradeoff

DEMONSTRATION

- **This spiking SNN is used in a live demo**
- **High energy efficiency: 24,5nJ / digit classification**
  - Less than 1 spike / synaptic connection



Ethernet interface

Image

Result

SPI interface

Input spikes

Output spikes

Surface tablet

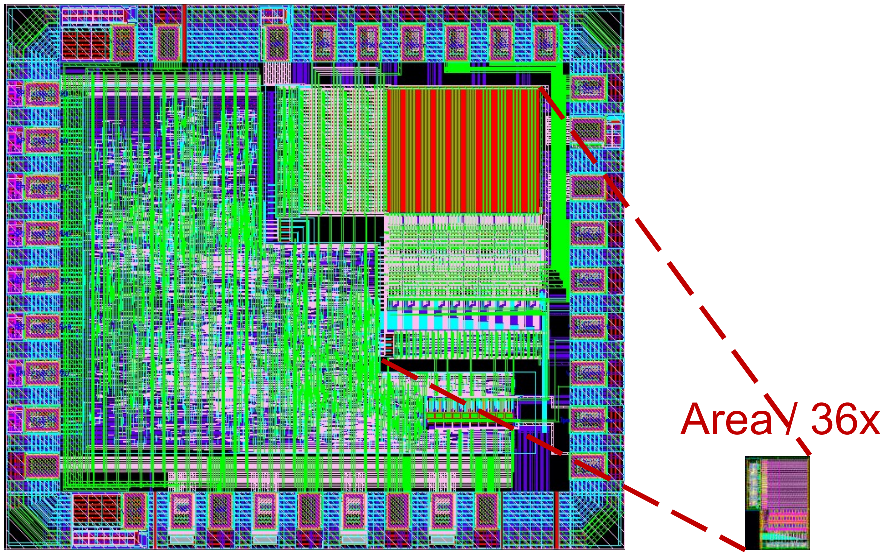
Microzed – ZYNQ X7CZ020

SPIRIT

A. Valentian, et. Al., “Fully Integrated Spiking Neural Network with Analog Neurons and RRAM Synapses,” *IEEE International Electron Devices Meeting (IEDM)*, San Francisco, CA, USA, December 2019

IMPROVEMENT BY MOVING TO 28NM FDSOI

- **Area of RRAM matrix**
  - Divided by 36X
- **Area of neurons**
  - Divided by 17X
- **Energy per event**
  - Divided by 10X



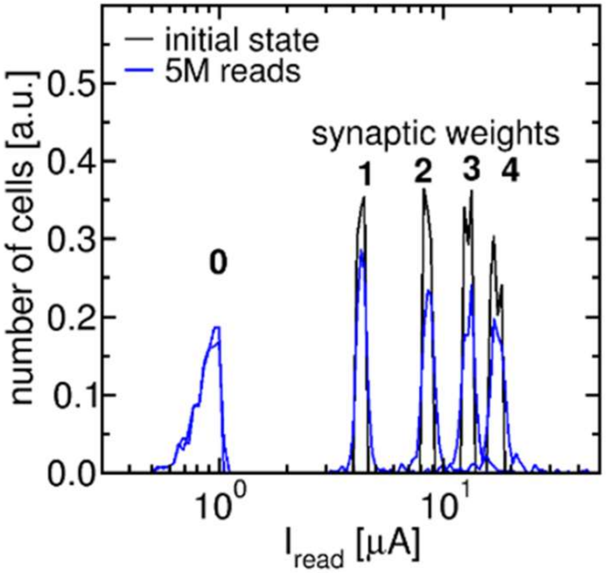
130nm Bulk

28nm FDSOI

RRAM TECHNOLOGICAL PATH FOR IMPROVEMENT

- **Multiple level cells**
  - Enable to increase synapse density by 4X

- **Synapse implementation**
  - One Single Level Cell (SLC) for the Sign
  - One MLC for the weight value



Multiple valued synapse

Comparison to the state-of-the-art

	Science 2014 [4]	Micro 2018 [5]	VLSI 2018 [6]	VLSI 2018 [6]	This work	This work scaled	This work scaled + multivalued
Technology	28nm	14nm	40nm	180nm	130nm	28nm	28nm
Coding	Spike	Spike	Formal	Formal	Spike	Spike	Spike
Weight storage	SRAM	SRAM	RRAM	RRAM	RRAM	RRAM	RRAM
Synapses	256M	130M	4M	2M	13.5K	13.5K	-
Synapses/mm^2	195K	2000K	1480K	160K	16K	575K	**2300K**
Power	63mW	-	9.9mW	15.8mW	1.5mW	-	-
Energy/syn. event	27pJ	105pJ	N/A	N/A	180pJ	**17,1pJ**	-

PATH TOWARDS DEEPER NETWORKS

- **Variability issue needs to be tackled**
  - Cost 2% of classification accuracy
    - On a shallow analog network
  - Would be way to high for a deep network

- **Two solutions arise**
  - Online retraining, for coping with variability
    - Not an “industrial” solution: would be too time consuming, thus expensive

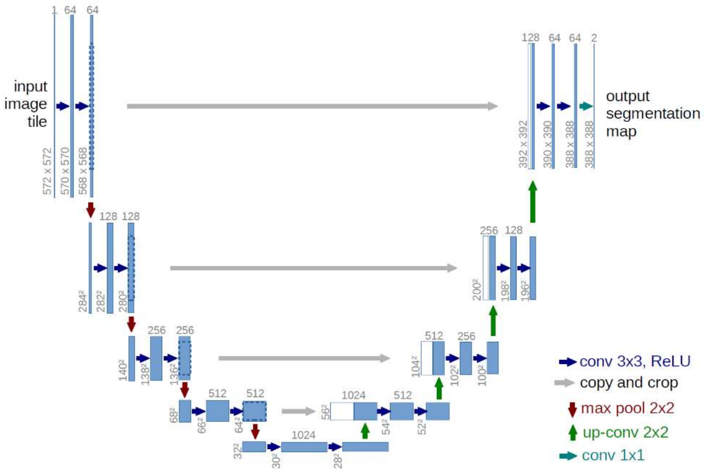
  - Digital implementation
    - Enables consistency with cycle-accurate simulation

REQUIREMENTS DEFINITION

- **Need to perform**
  - Detection
  - Classification
  - Segmentation

- **Points mostly towards Convolution network**
  - UNet network
    - Convolution + Deconvolution

- **Conv layers represent the majority of the computation workload**
  - Focus of this implementation



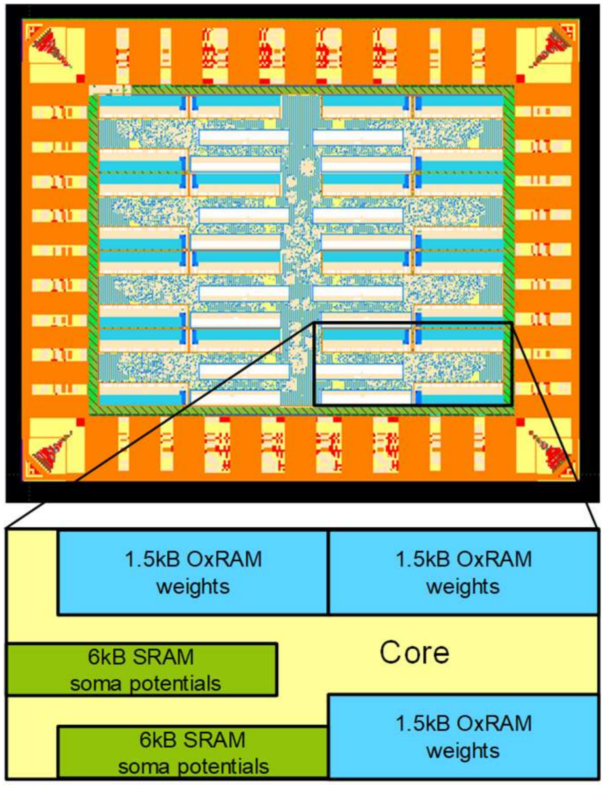
UNet topology

IMPLEMENTED CIRCUIT

- **Key parameters of interest**
  - Technology: 28nm FDSOI + RRAM
  - Area: 3mm²
  - 8 Convolutional SNN cores
  - 131k neurons, 73k weights, 75M synapses

- **Total computing power**
  - 25.6 GOPS (synaptic operations per second)
  - 128 Processing Engines @ 200 MHz

- **Energy efficiency**
  - 1pJ per synaptic event



Multicore architecture for spiking convolution operations

CONCLUSION

Trends in Edge AI applications

- Inference first
- Then lifelong local learning

Main challenge is to reduce data movement

This can be solved thanks to a combination of architecture and technology

- Combination of In-Memory Computing
- Non-volatile memory for synaptic weights
- 3D technology for heterogeneous integration

Brain-inspired solutions might just be the Key for high energy efficiency solutions

Contact information:
Dr Alexandre VALENTIAN

email: alexandre.valentian@cea.fr

THANK YOU FOR YOUR ATTENTION

FROM RESEARCH TO INDUSTRY

cea tech

Copyright Notice

This multimedia file is copyright © 2021 by tinyML Foundation. All rights reserved. It may not be duplicated or distributed in any form without prior written approval.

tinyML® is a registered trademark of the tinyML Foundation.

www.tinyml.org

Copyright Notice

This presentation in this publication was presented as a tinyML® Talks webcast. The content reflects the opinion of the author(s) and their respective companies. The inclusion of presentations in this publication does not constitute an endorsement by tinyML Foundation or the sponsors.

There is no copyright protection claimed by this publication. However, each presentation is the work of the authors and their respective companies and may contain copyrighted material. As such, it is strongly encouraged that any use reflect proper acknowledgement to the appropriate source. Any questions regarding the use of any materials presented should be directed to the author(s) or their companies.

tinyML is a registered trademark of the tinyML Foundation.

www.tinyML.org