



The Spiking Neural Processor:

Mixed-Signal MCU for power constrained tiny ML applications

Rui Teixeira

rui.teixeira@innatera.com

- **Ultra-low power intelligence for sensors**
- Spun out of the **Delft University of Technology** in 2018
- Over 80 employees, offices in the **Netherlands** and **India**, staff across the **EU**
- Over **€25m** in funding from leading European deep-tech investors

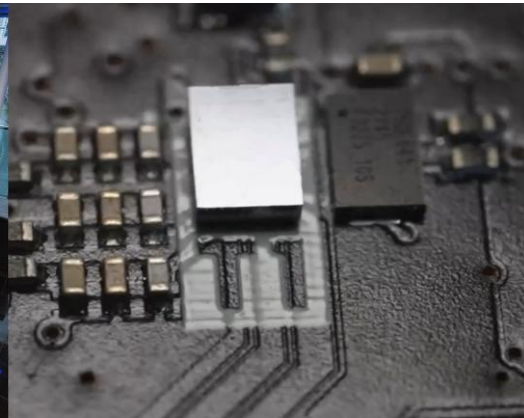
European
Innovation
Council



INVESTNL

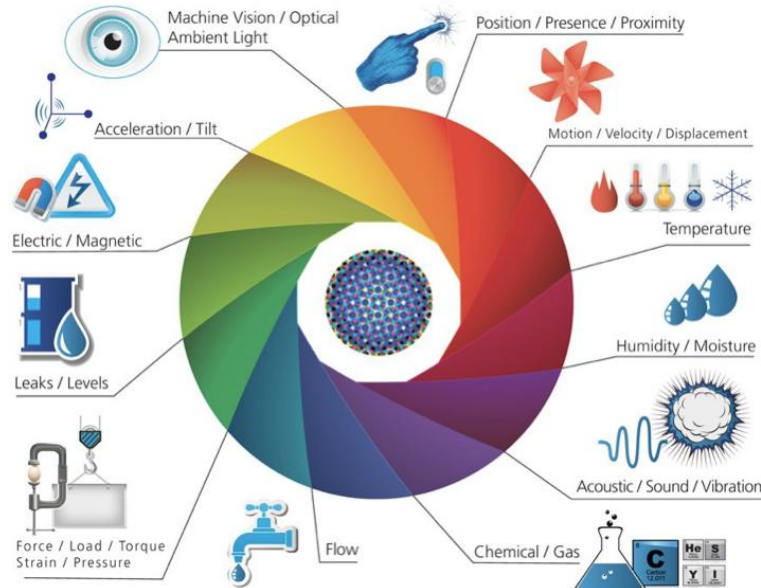
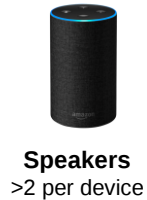
MIG
CAPITAL

INNAVEST



Sensors are everywhere

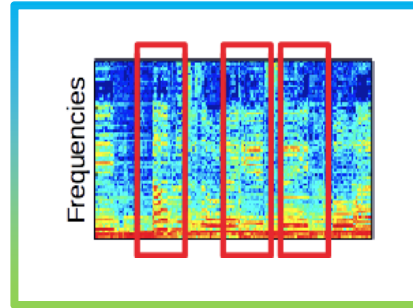
Estimated 4 billion new sensor-driven devices each year*



Every sensor needs a processor



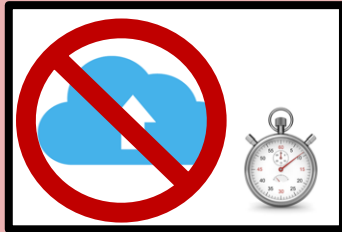
Find patterns



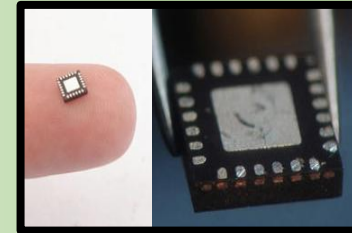
Characterize them



Identify them

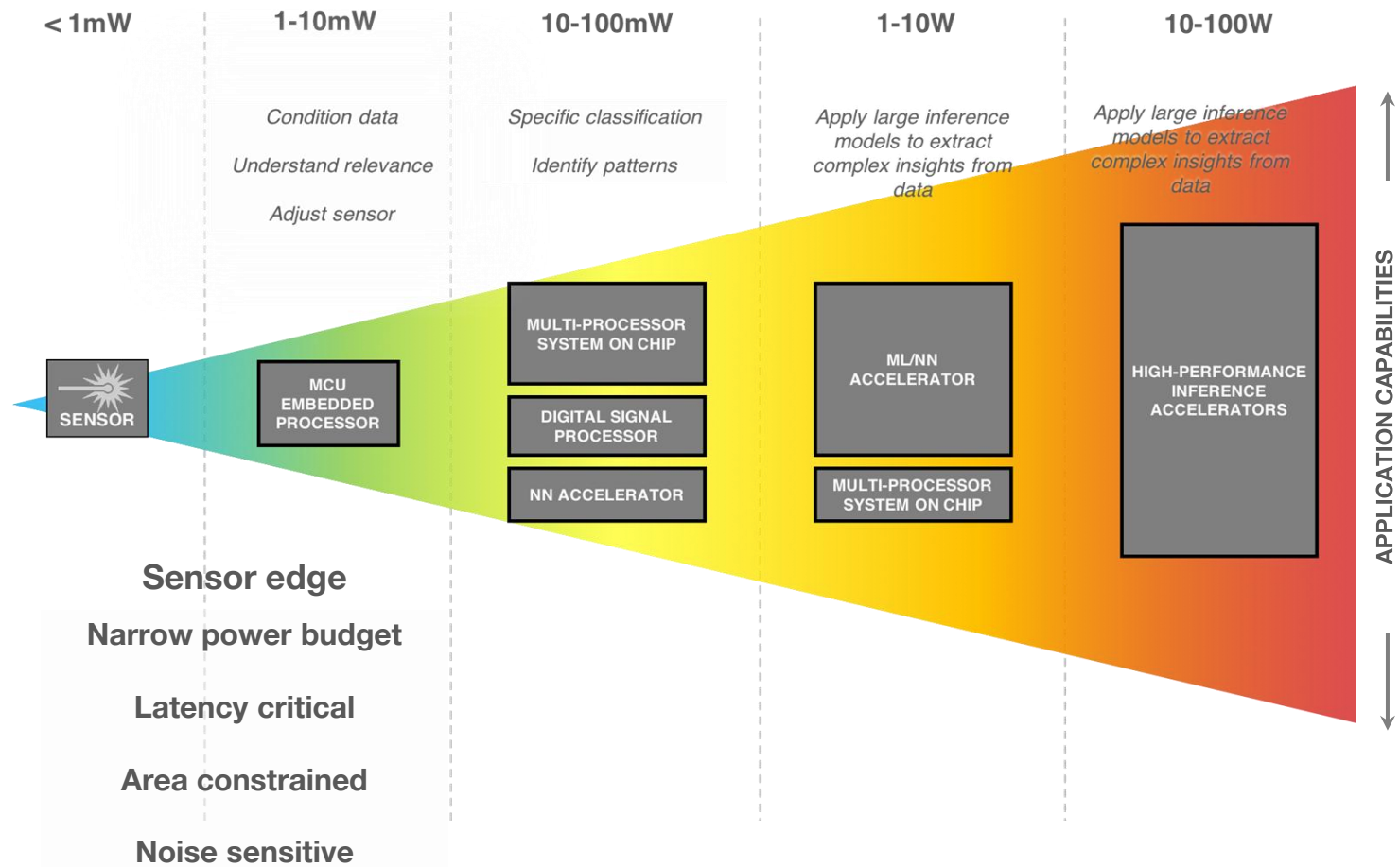


Cloud-based processing
not feasible

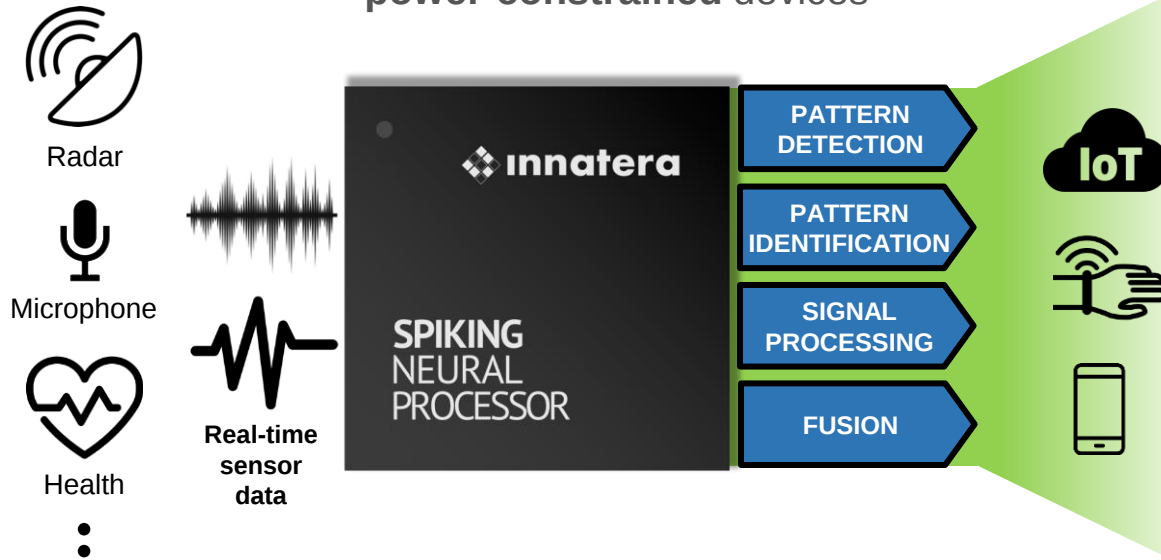


Processing close to
the sensor

Bringing intelligence to the sensor edge



Brain-inspired processor for **turn-key intelligence in**
power-constrained devices

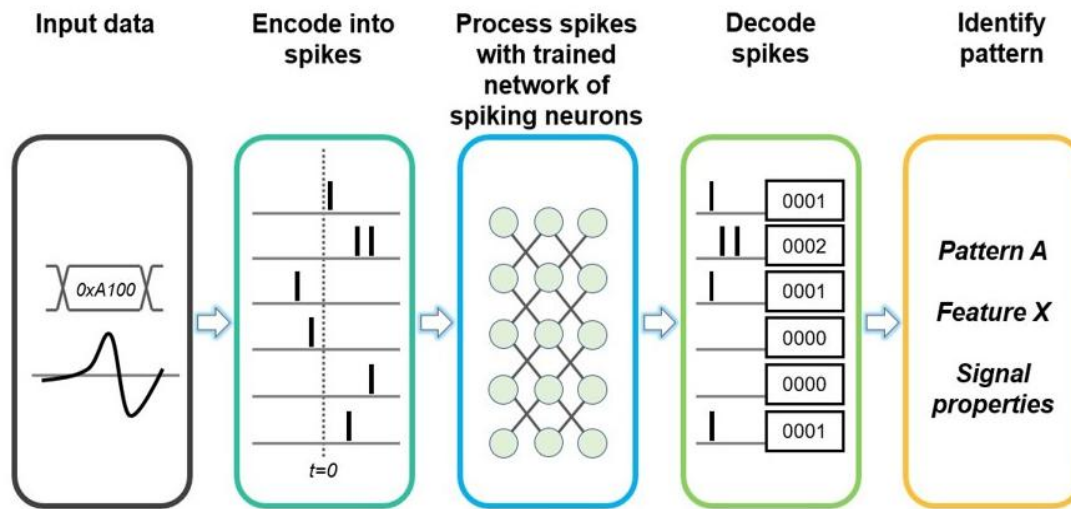


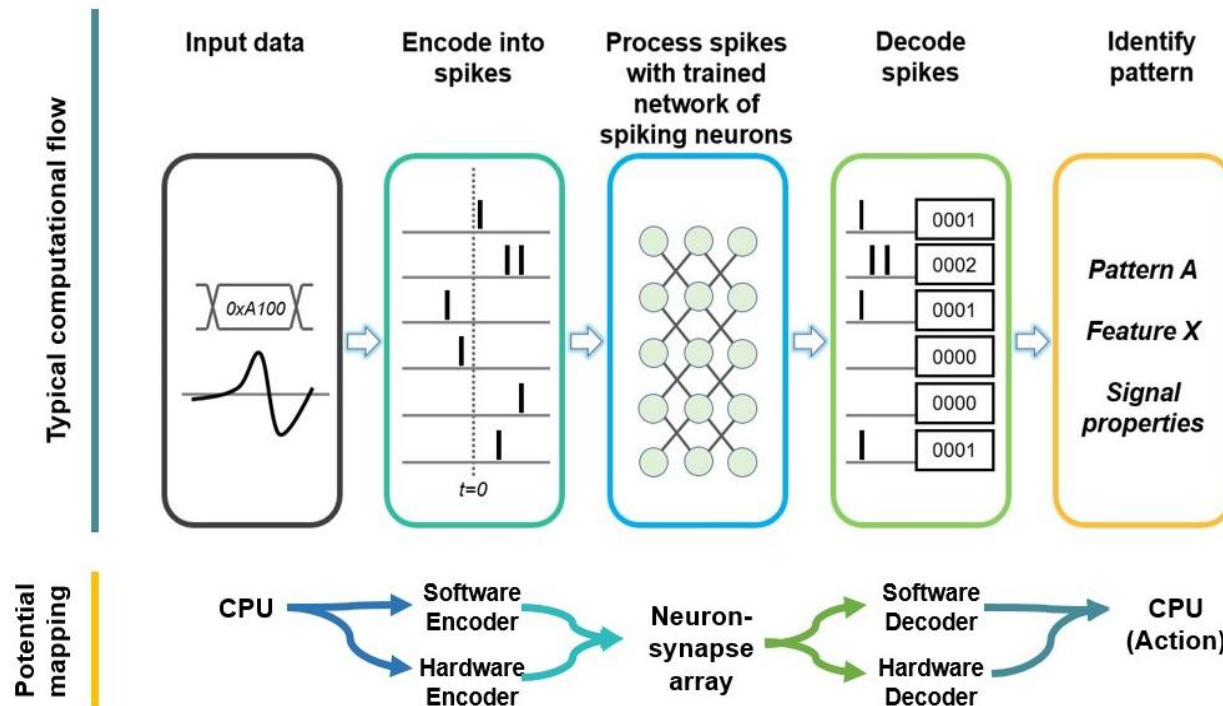
Always-on event-driven processing

Sub-millisecond processing latency

Sub and few-mW power dissipation for the sensor-edge

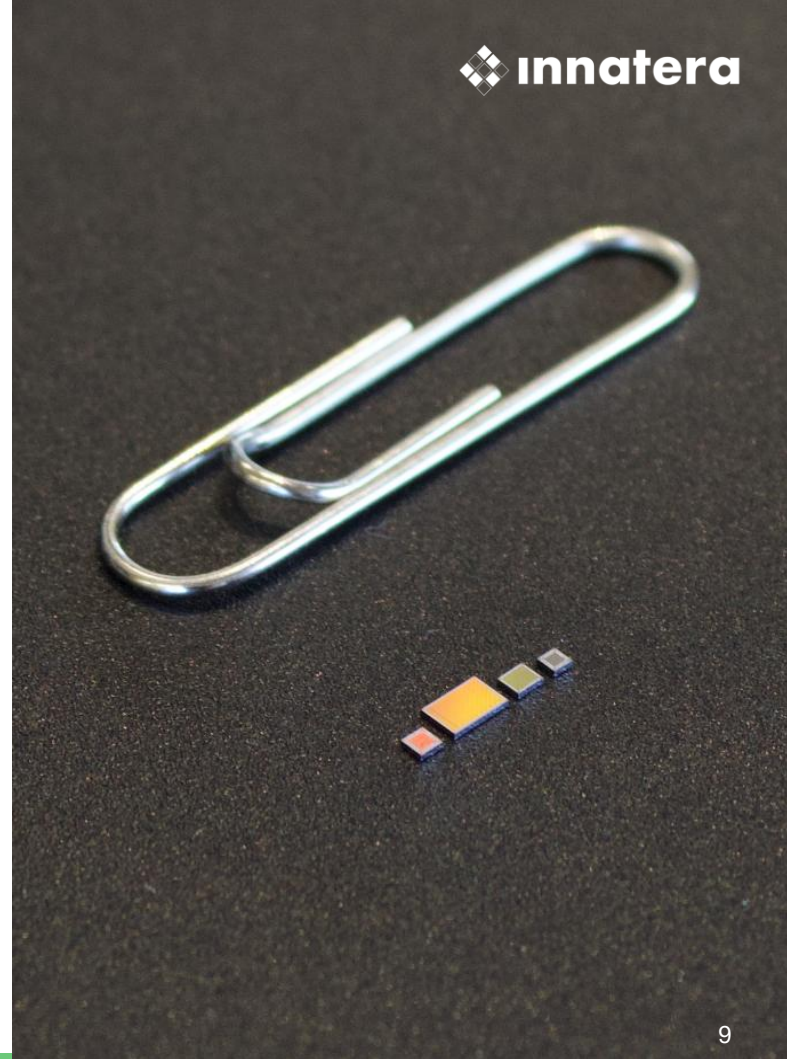
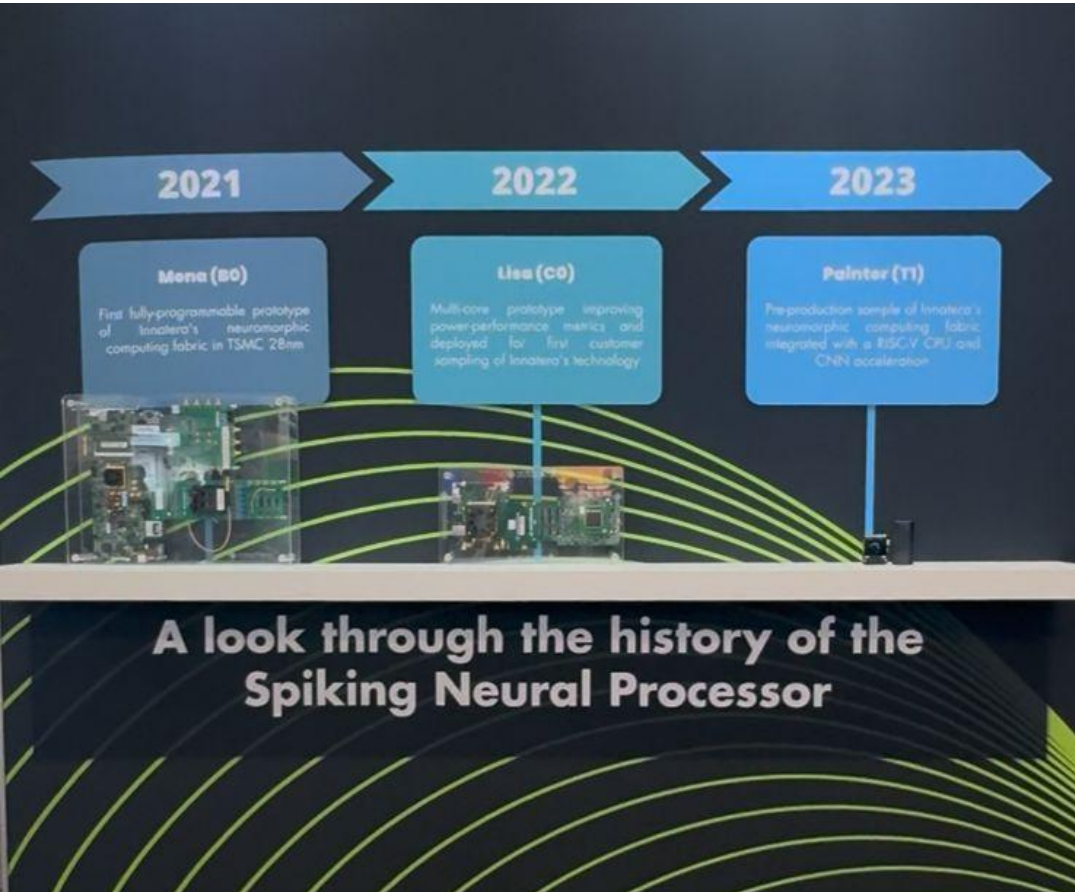
Processing sensor data with Spiking Neural Networks





The SNP is a flexible platform to efficiently compose heterogeneous applications

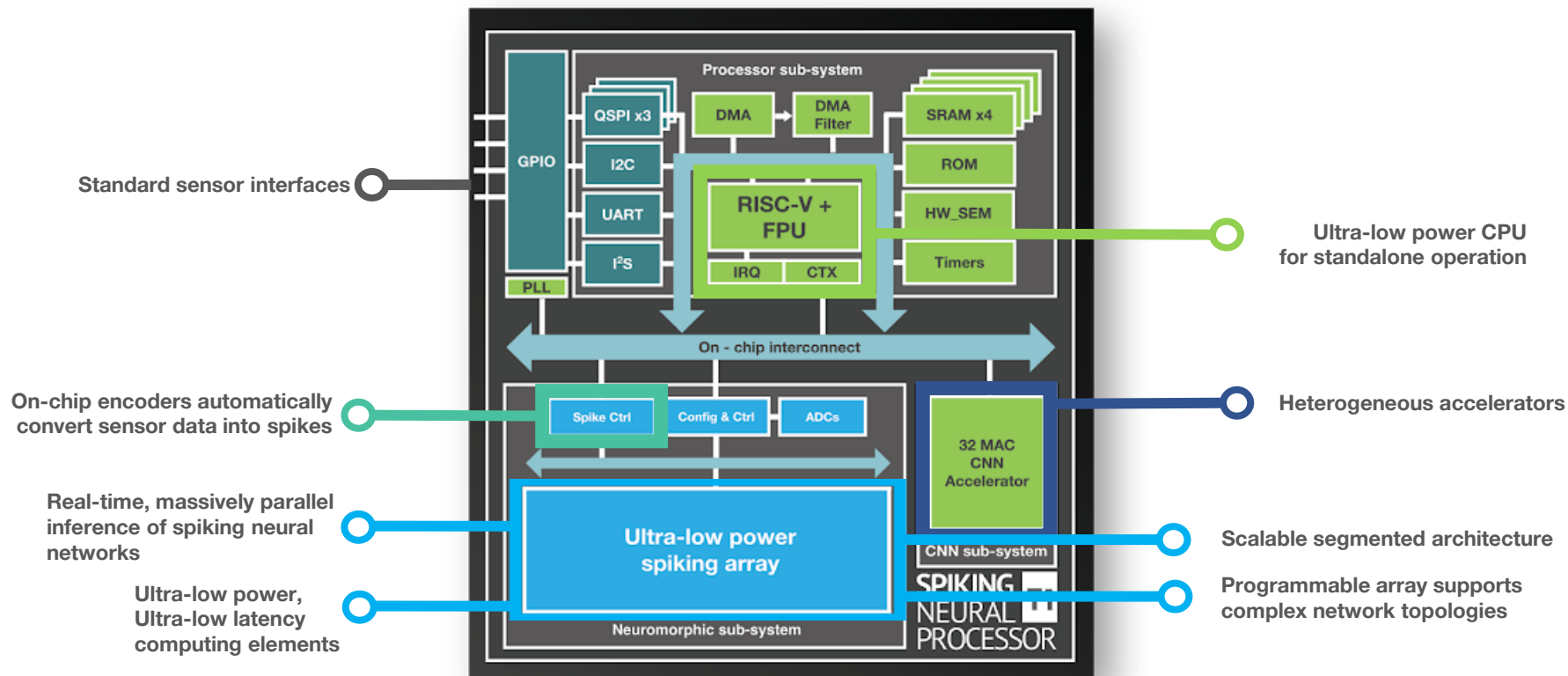
Validation and improvement over multiple generations – from table to pocket size



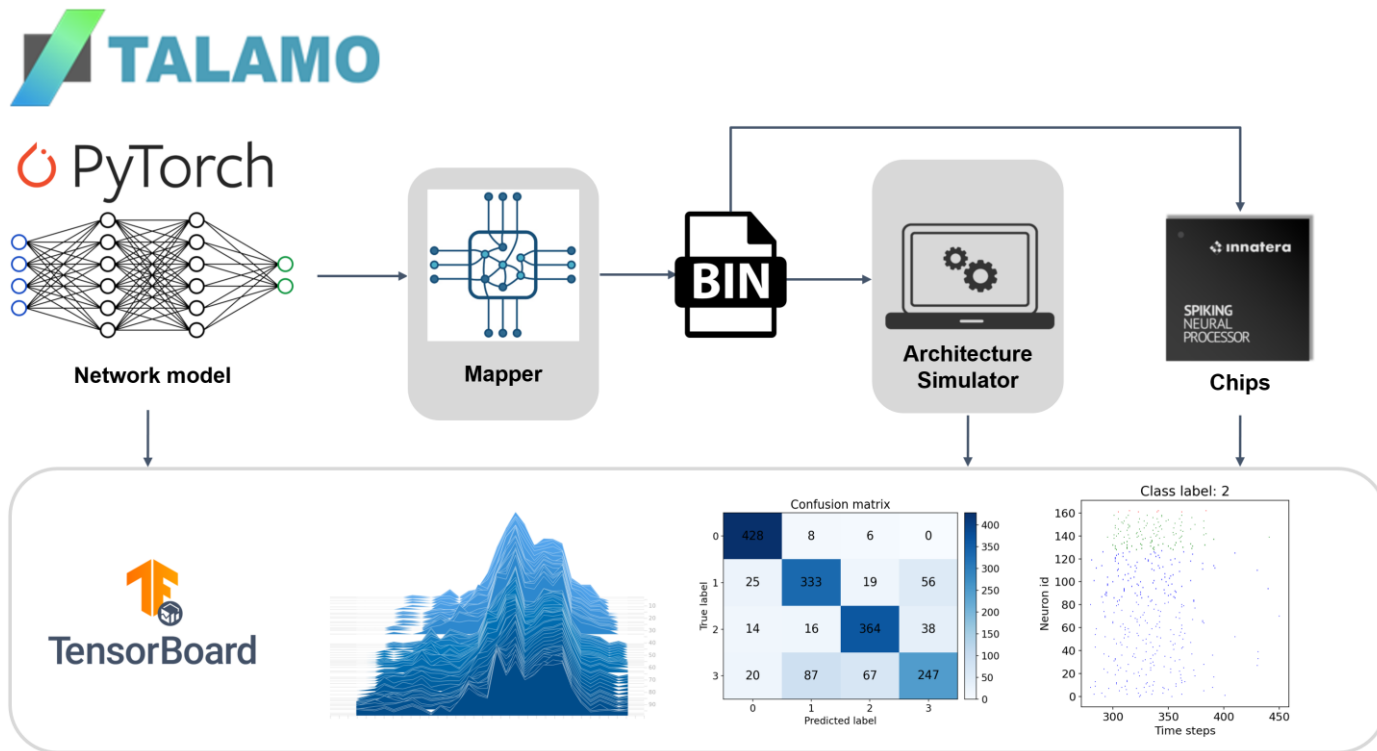
Hardware-to-Software stack

Developing experience

Spiking Neural Processor T1 (SNP T1)



Innatera's powerful Software Development Kit - Talamo



Simple

Easy to use, familiar workflow

Turn-key

Easily build and deploy models to hardware

Standard

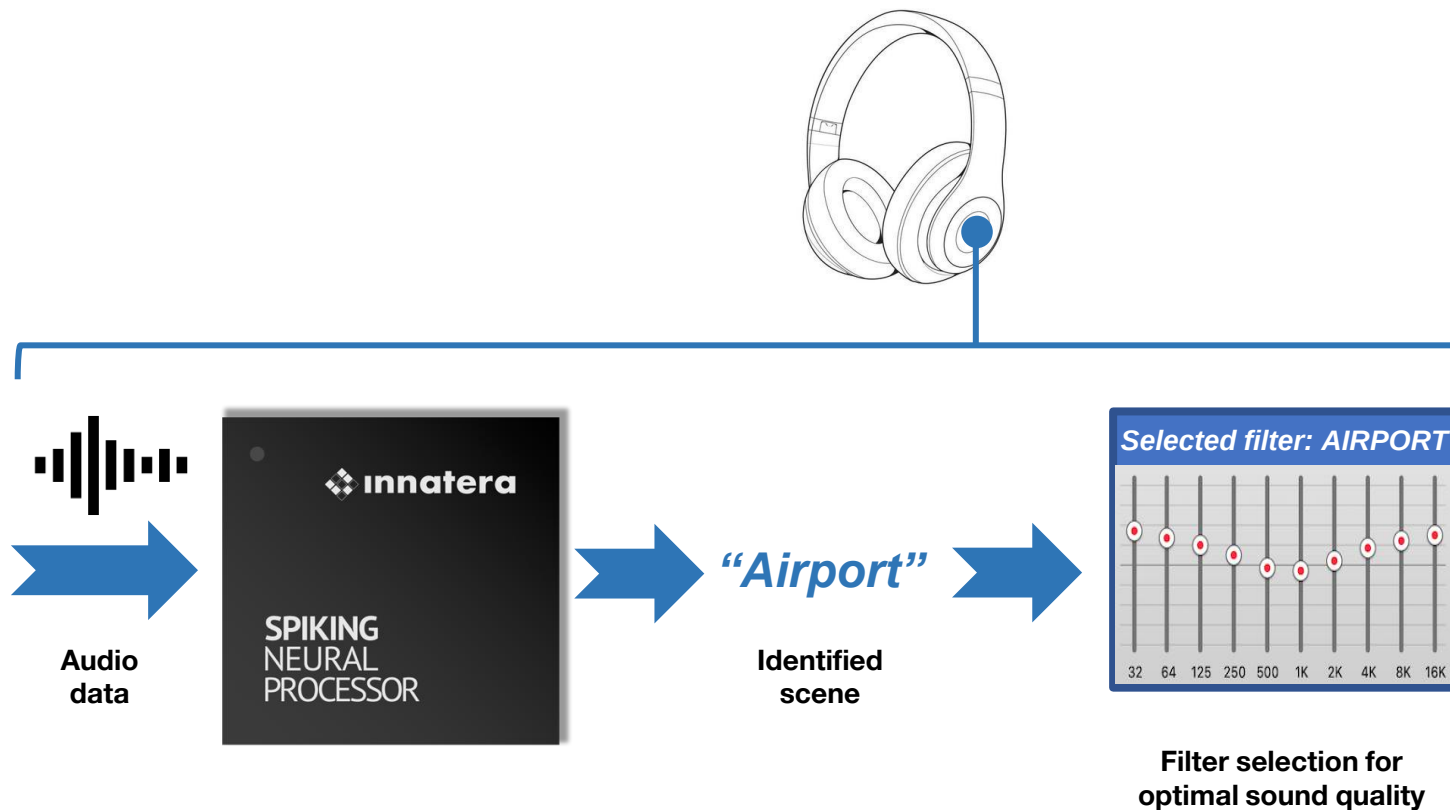
Native integration with PyTorch, TensorBoard

Fast

Rapid simulation and deployment



Audio scenes



Advanced
strategy

Custom feature
extraction

Custom

Ensembles
Feedback

Time to first spike
Custom

Audio

Preprocessing

Spike
encoding

Spiking
neural
network

Decoder

Identified
scene

Initial
strategy

Mel-frequency
cepstral coefficient

Filter bank
FFT
CNN

Integrate and Fire

Rate
Thermometer

Two-layer
feedforward

Multi-layer feedforward
Recurrent layers

Rate softmax

Innatera's powerful Software Development Kit - Talamo



```
pipe = talamo.pipeline.elements.Pipeline(  
    steps=[  
        MFCC(n_mfcc=num_features, n_fft=512, hop_length=512, sampling_rate=load_sr),  
        talamo.encoders.IFEncoder(num_encoder_channels=num_features,  
                                   threshold_initializer=talamo.initializers.RandomUniform(32, 64),  
                                   encoder_timestep=time_step),  
        talamo.snn.containers.Snn(snn_model),  
        talamo.decoders.MaxRateDecoder(),  
    ])
```

Spike Data Processing

Encoders

Decoders

Spike data
utilities

Network Building

Neuron

Layer

Synapse

Container

Training

Surrogate
gradients

Quantization

Initializers

Regularization

Utilities

Tensorboard

Callbacks

Checkpoint

Metrics

Application Creation

Pipeline API

Built-in
libraries

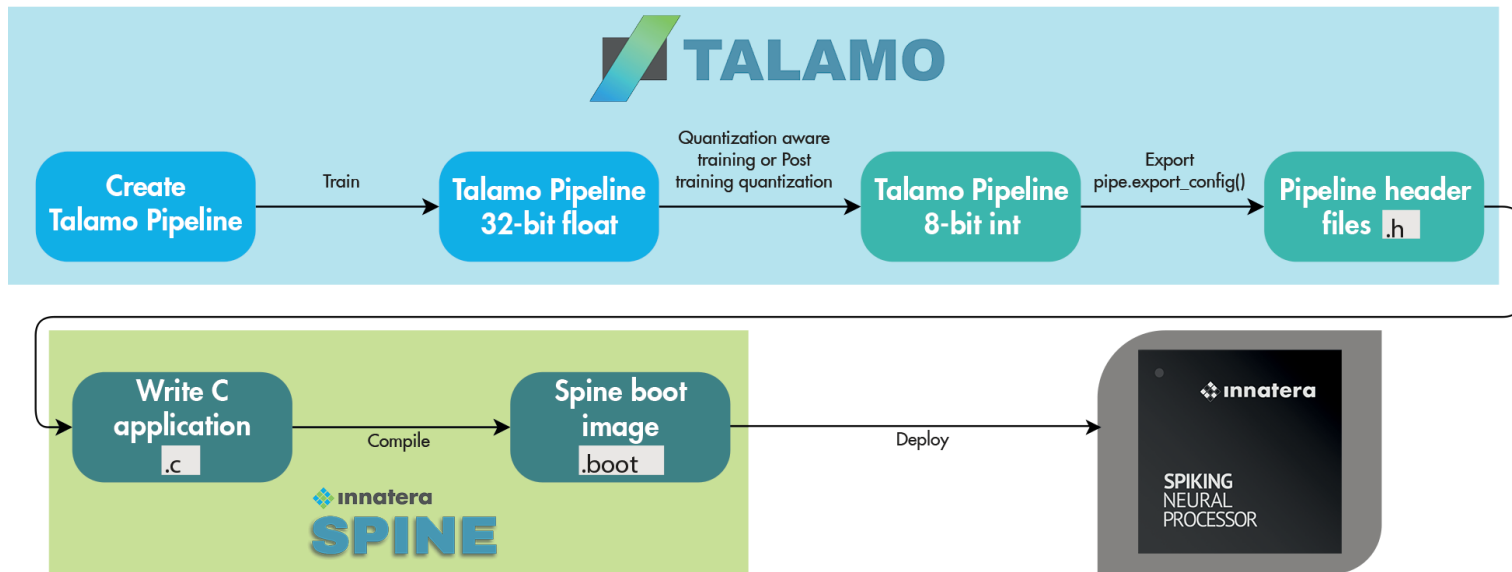
Device Access

Mapper

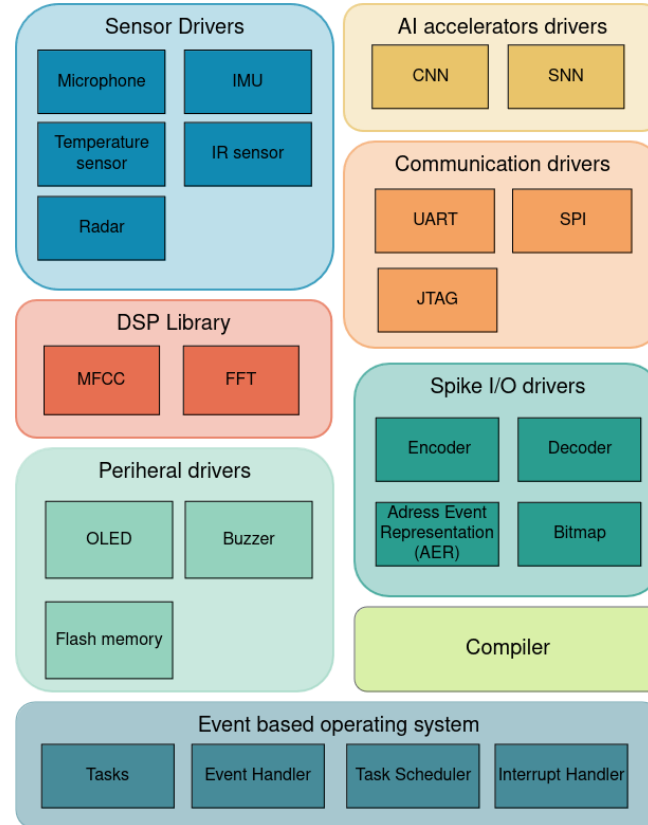
Hardware

Simulator

Application workflow – Full embedded application

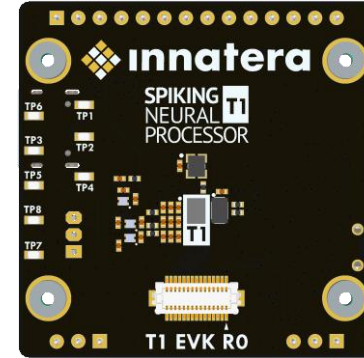


- Embedded development kit
- Written in C
- Contains:
 - Chip drivers
 - Sensor drivers
 - Peripheral drivers
 - Compiler
 - DSP library



Outreach

Early sampling to customers and partners

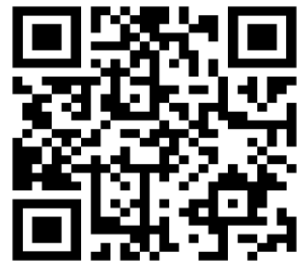


University Access Program



Neuromorphic WG mission statement

- Foster academic-industrial partnerships, support open-source neuromorphic initiatives, and provide a platform for knowledge transfer, educational initiatives, and funding.
- Facilitate collaboration through forums, consortia, and long-term support for emerging neuromorphic computing developments



Summary

Spiking Neural Processor T1

An ultra-low power neuromorphic SoC for tiny ML / Edge AI



On-chip RISC-V CPU

32-bit RISC-V core, 384 KB embedded SRAM



Analog-mixed signal spiking neural networks

Enables fast sub-mW pattern recognition and signal processing based on event-driven spiking neural networks.



Tiny CNN accelerator

32 MAC CNN accelerator



Diverse Interfaces

QSPI, I2C, UART, JTAG, GPIO, front end ADC



Compact

2.16mm x 3mm, 35-pin WLCSP package



T1 Evaluation kit

Familiar development kit with a sensor shield, a screen and a breadboard



Comprehensive platform

Ready to use platform for application prototyping and profiling



Accessible

Built in software for developing applications



Wide interfacing options

Standard interfaces supporting a range of sensors



Talamo SDK

Powerful software environment for application development and deployment



PyTorch front-end

Develop and train models with the industry-standard PyTorch ML framework, use powerful measurement and visualization capabilities of Tensorboard



Comprehensive

Wide-range of models, network topologies, libraries for application development



Applications

The SNP T1 platform supports a wide range of applications



Presence detection

Radar
Infra-red



Audio interfaces

Scene classification
Keyword spotting



Touch-free interfaces

Gestures



Activity recognition

IMU HAR



Let's make sense together.

Innatera Nanosystems BV
Lange Kleiweg 40,
2288 GK Rijswijk
The Netherlands

info@innatera.com

www.innatera.com