Products and applications enabled by tinyML

March 28 – 29, 2023

www.tinyML.org
Tiny spiking AI for the sensor-edge

Petrut Bogdan
Neuromorphic Architect
petrut.bogdan@innatera.com
Outline

• Innatera
• Spiking Neural Processor
• Talamo SDK
• Ultra-low power edge applications
Made in Delft

• Ultra-low power intelligence for sensors

• Spun out of the Delft University of Technology in 2018

• 57 employees, offices in the Netherlands and India

• Funded by deep-tech investors Matterwave Ventures and MIG Capital
Sensor-edge constraints:
- average power < 1mW
- code size < 10 KB
- latency < 100 ms
- bill of materials
Spiking Neural Processor

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

- Radar
- Microphone
- Health
- Real-time sensor data

**Always-on sensing applications**

- Millisecond-scale processing latency envelope
- Milli- and sub-milliwatt power envelope
Processing sensor data with Spiking Neural Networks

Spiking Neural Network models can be up to 100x smaller than conventional Artificial / Deep Neural Networks
Spiking Neural Processor (SNP)

On-chip encoders automatically convert sensor data into spikes

Standard sensor interfaces

Ultra-low power CPU for standalone operation

Data-to-Spike Encoders

Real-time, massively parallel inference of spiking neural networks

Ultra-low power, ultra-low latency computing elements

Programmable array supports complex network topologies

Scalable segmented architecture

The only processor needed for always-on sensing applications
Innatera’s powerful Software Development Kit - Talamo

PyTorch

- Powerful PyTorch-based SNN Compiler
- Bin programming file
- Native integration with PyTorch, TensorBoard
- Rapid simulation and deployment

Simple
- Easy to use, familiar workflow

Turn-key
- Easily build and deploy models to hardware

Standard
- Native integration with PyTorch, TensorBoard
Talmo’s PyTorch API - spiking neural networks made easy

Includes everything required to build and train models

Identical to the PyTorch API – easy to adopt and use

Requires no knowledge of spiking neural networks
This notebook defines the workflow solving the MNIST classification task using Talamo
Always-on audio scene classification in hearables

Audio scenes

Audio data

Audio NEURAL PROCESSOR

Identified scene

“Airport”

Filter selection for optimal sound quality

Selected filter: AIRPORT
Solution pipeline

Audio → Preprocessing → Spike encoding → Spiking neural network → Decoder → Identified scene

- Mel-frequency cepstral coefficient
- Poisson rate
- Two-layer feedforward
- Rate softmax
Power (total peak) 1.06 mW

Accuracy ~85%

Inference latency ~1 ms / 1 s

Model size ~3kB

Audio scene classification (DCASE 2020)
Summary

• Edge AI is:
  • Power constrained
  • Latency constrained
  • Code-size constrained
  • BOM constrained

• The Spiking Neural Processor delivers audio scene classification in
  • a power budget of ~1mW
  • an inference latency of ~1ms
  • with a code size of ~3kB

• Talamo SDK – simplifies model development with a standard, well-understood work flow

• Easy to adopt, build, and deploy sensor-edge solutions with unprecedented power-performance

The future of TinyML is Neuromorphic!
Let’s make sense together.

Innatera Nanosystems BV
Patrijsweg 20
Rijswijk 2289EX
The Netherlands

info@innatera.com

www.innatera.com
Copyright Notice

This presentation in this publication was presented at the tinyML® Summit (March 28 - 29, 2023). 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