CERN openlab

Bartłomiej Borzyszkowski

Intel Poland, CMS Experiment at CERN

ML Gdańsk, 10th May 2021

Neuromorphic Computing
in High Energy Physics

Presentation plan

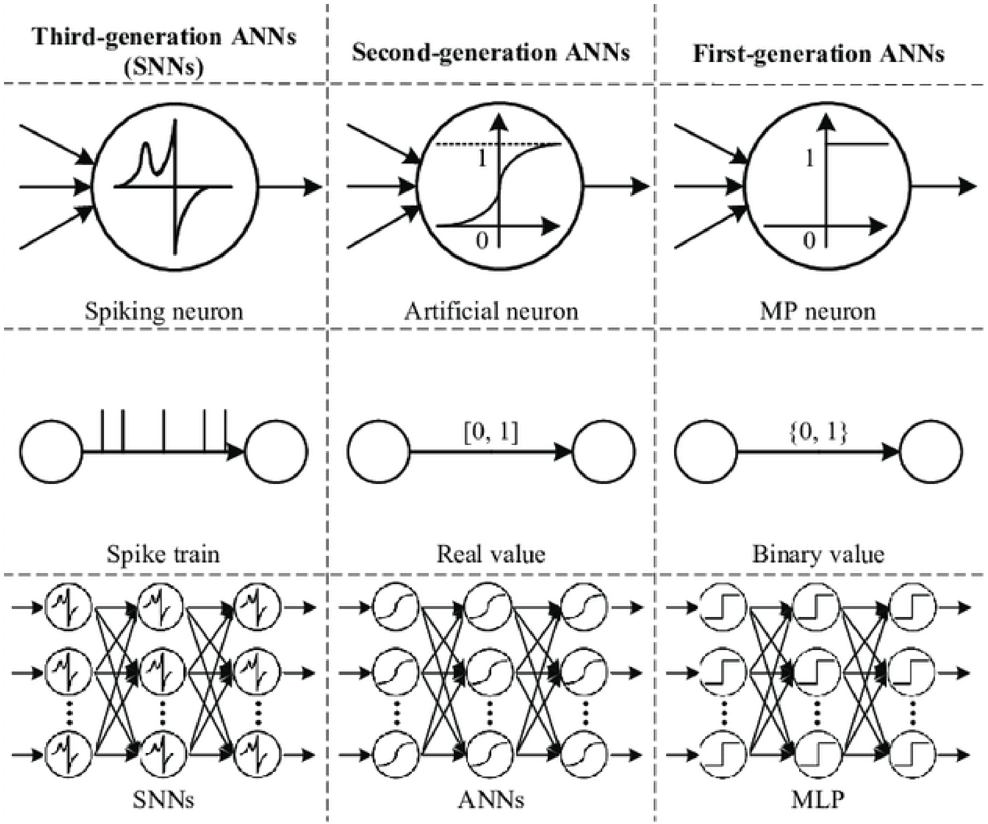
- Third generation of artificial intelligence (spiking neural networks)
- Principles of neuromorphic computing

- Overview of experiments at CERN
- Neuromorphic Jet Tagging

- Overview of applications at LIGO
- Neuromorphic Gravitational-Wave Detection

- Research opportunities and my experience at CERN
- Questions and further discussion

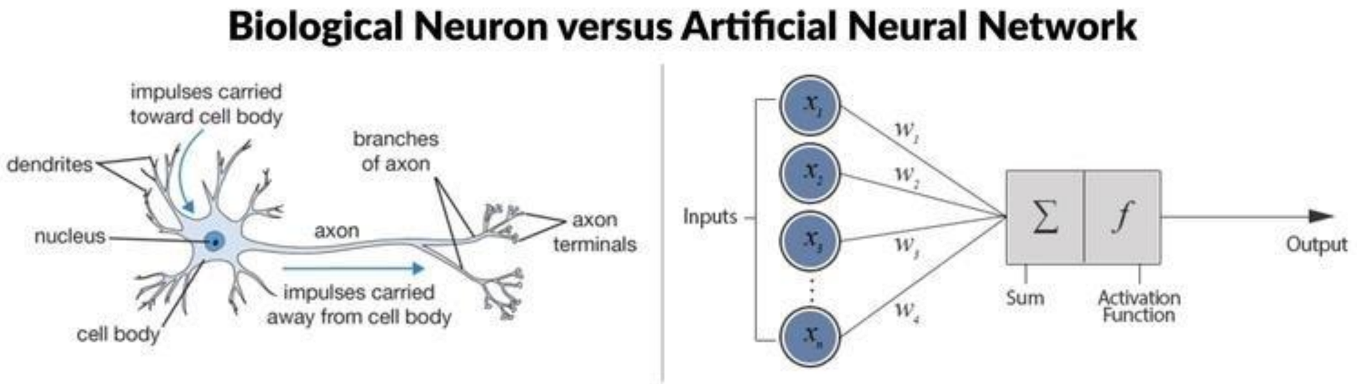
Three generations of neural networks



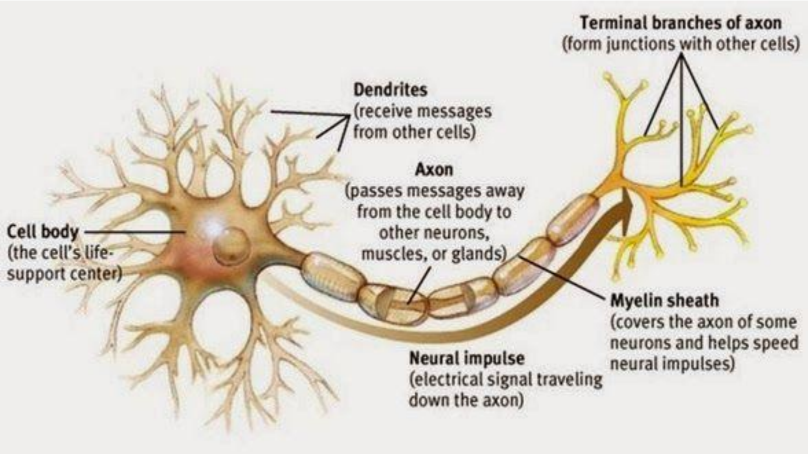
CERN openlab

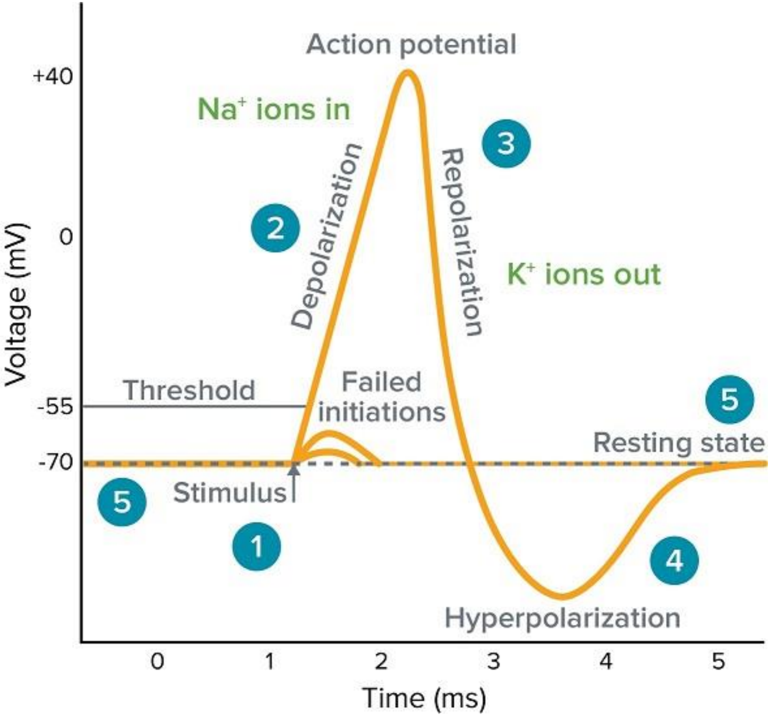
Artificial Neural Networks

- Spectacular successes of AI in recent years
- Inaccurate approximation of the actual intelligence in nature
- Neurons operate on a common clock cycle



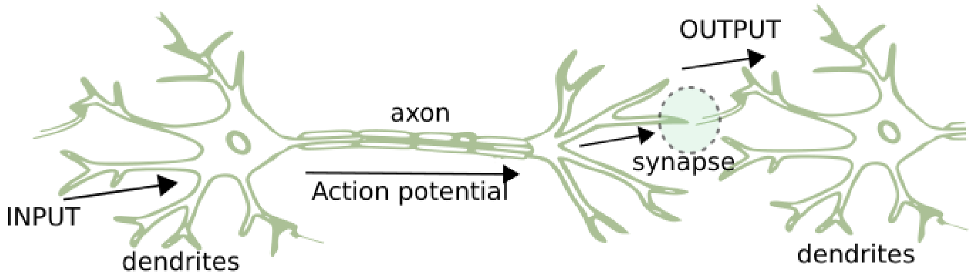
Real neurons in nature

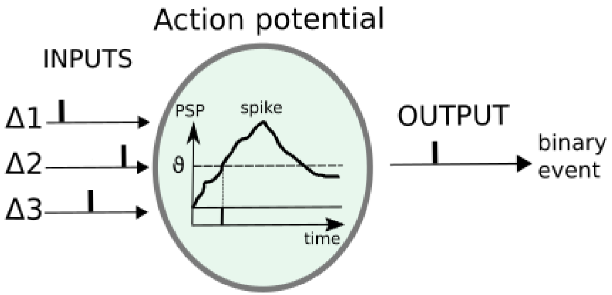




Spiking Neural Networks

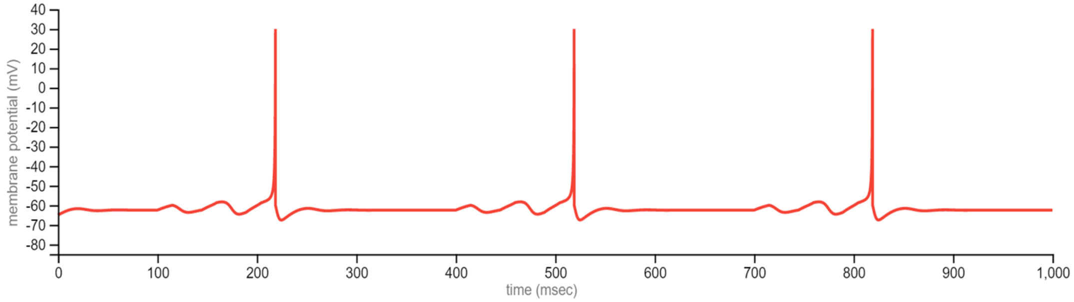
- Inspired by information processing in biology
- Data encoded by spike trains (combinations of spikes)
- Neurons process the data continuously in time



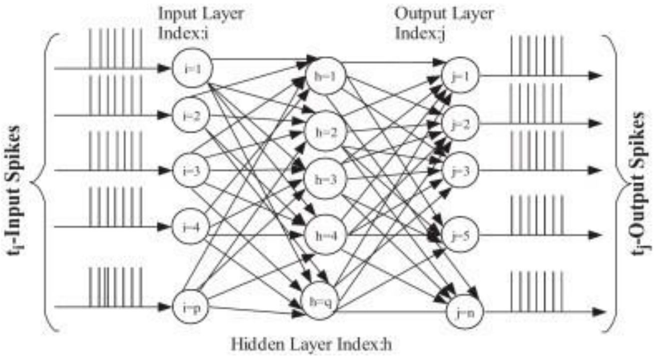


Key differences

- Time dependency
- Asynchronous output (no clock mechanism)
- Event-driven processing (computationally efficient)



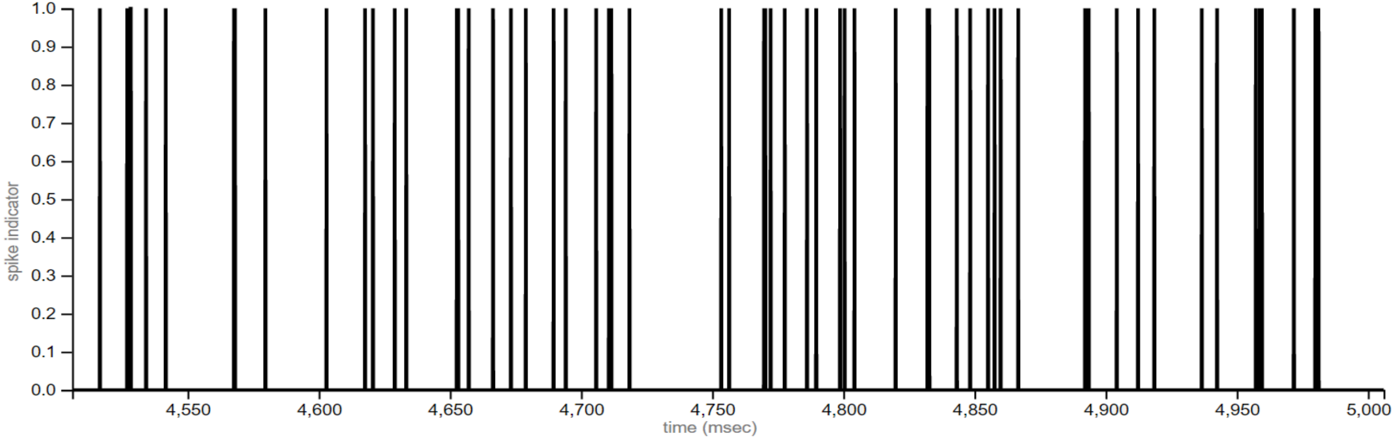
Membrane potential of a single neuron



Architectuire of SNN

Spikes and information encoding

- Spikes are binary signals in time
- A spike train is a sequence of recorded times at which a neuron fires an action potential. This is how the information in SNNs are encoded

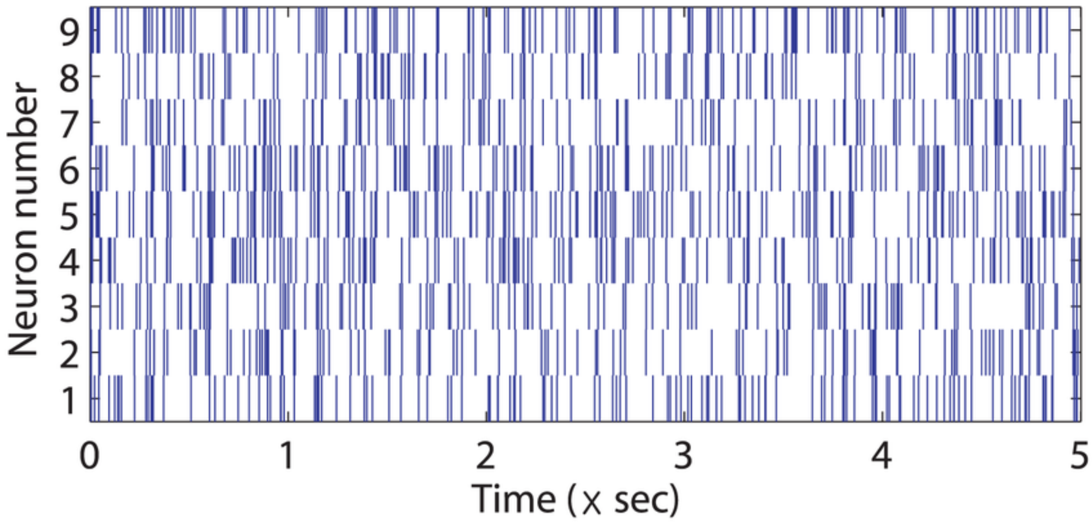


Binary output of a single neron (series of spikes form the spike trains)

Source: http://jackterwilliger.com/biological-neural-networks-part-i-spiking-neurons/

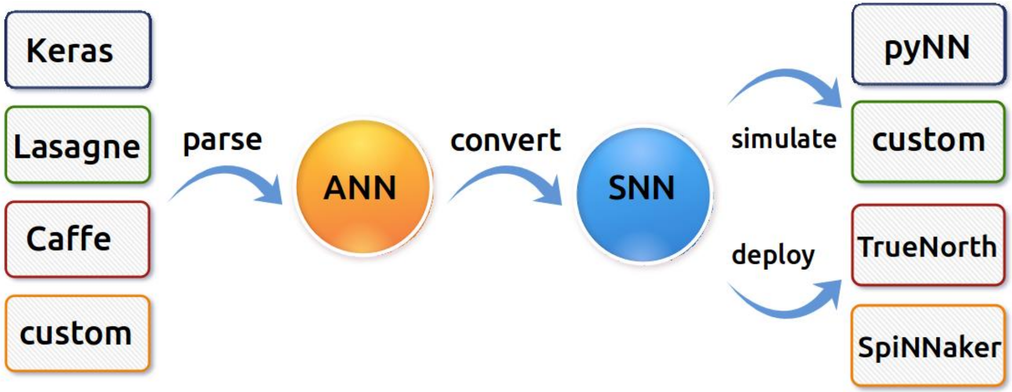
Difficulties in training

- Non-differentiable nature of spiking neurons (SGD is impossible)
- Asynchonicity (difficulties to encode the information)



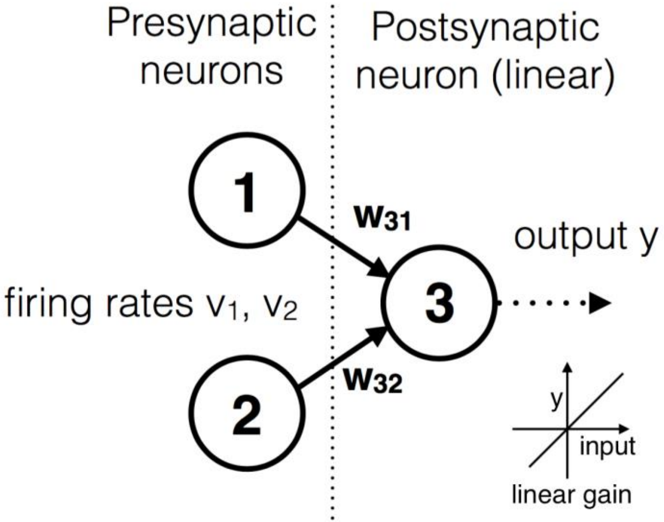
Approaches to train the SNN

- Conversion of a pre-trained ANN into SNN
- Approximated backpropagation
- Local learning rules (inspired by biology)



STDP and Hebbian Learning Rule

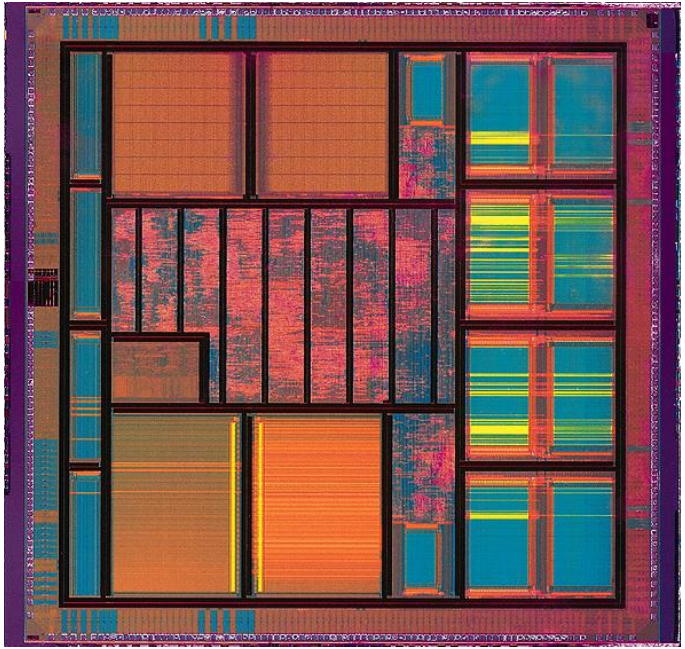
- Directly inspired by biology (spike-timing-dependent plasticity)
- Synapses increase their efficiency if the synapse persistently takes part in firing the postsynaptic target neuron
- Online learning (no backpropagation)
- Often used as a post-conversion



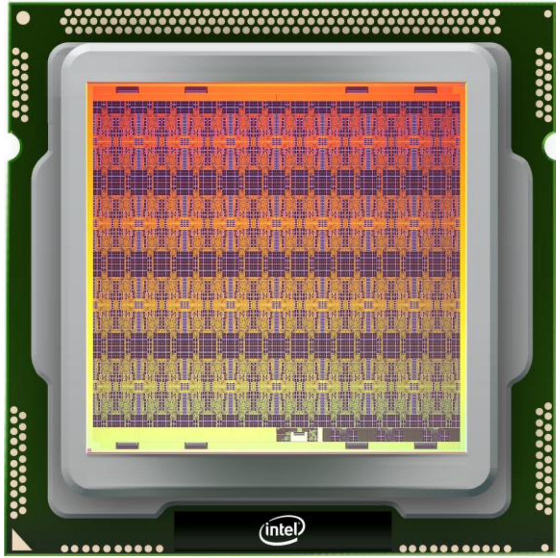
Neuromorphic engineering

- SpiNNaker (University of Manchester / Human Brain Project)
- True North (IBM)
- Neurogrid (Stanford)
- Loihi chip (Intel)

FPGA-based emulators; VLSI systems

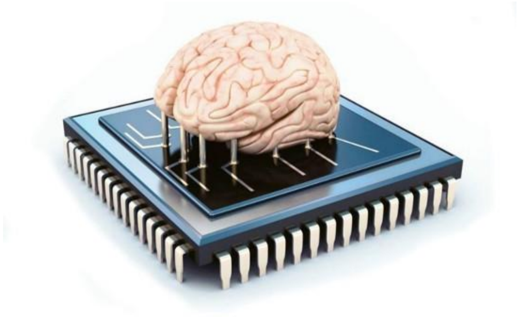


Intel Loihi chip



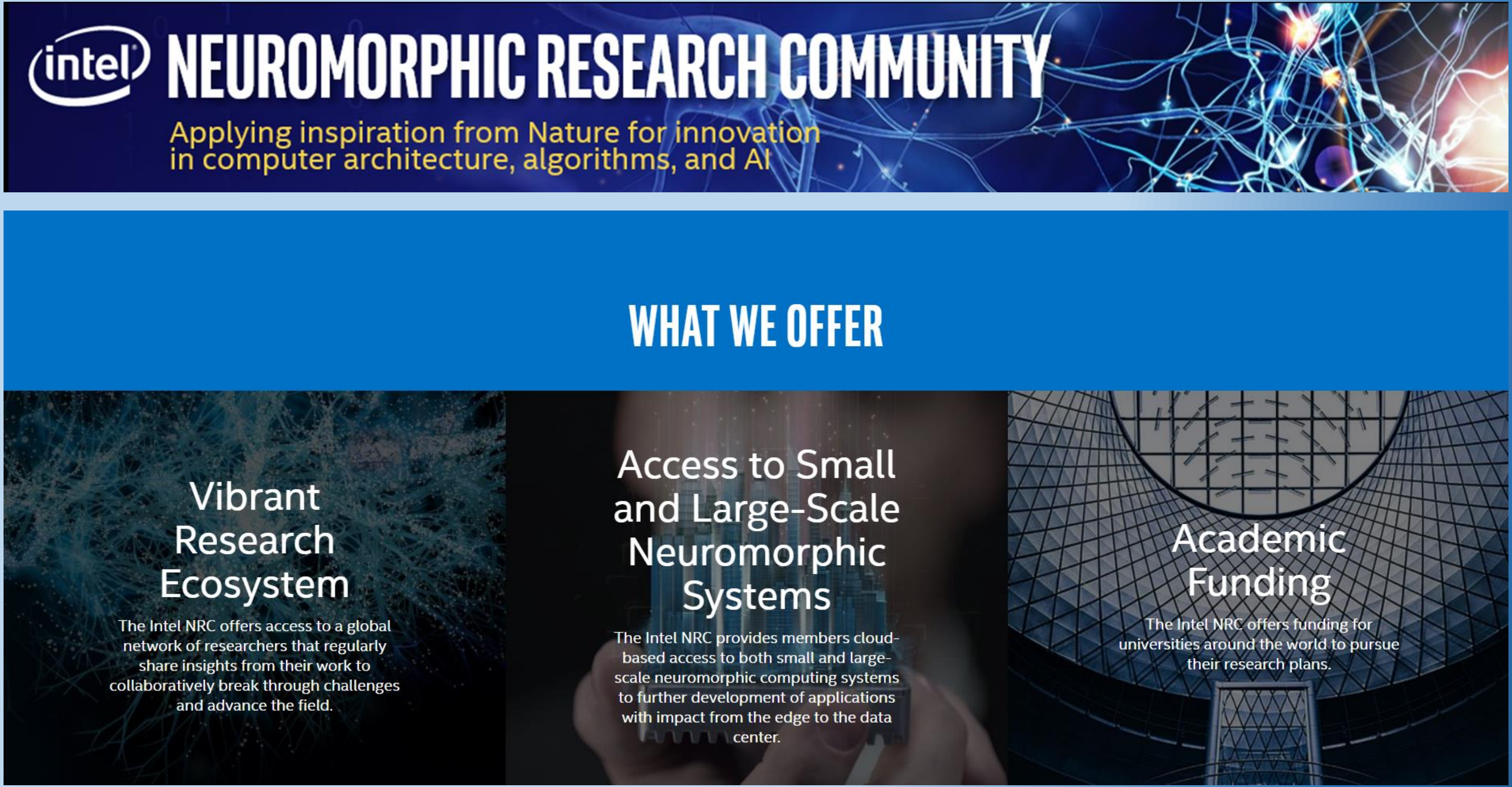
- Dedicated neuromorphic processor
- Supports training and inference of SNNs
- Efficient asynchronous computing

state-of-the-art in neuromorphic hardware



Presented by Intel Labs in 2018

Source: https://www.intel.com/content/www/us/en/research/neuromorphic-computing.html



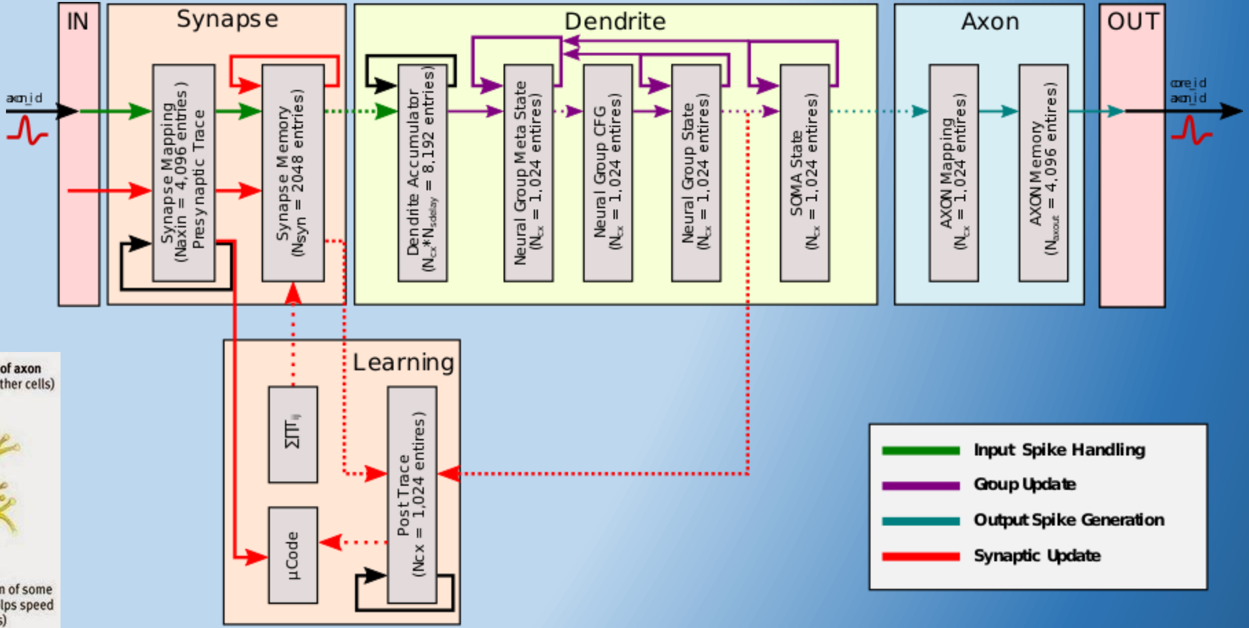
Source: INRC: https://www.intel.com/content/www/us/en/research/neuromorphic-community.html

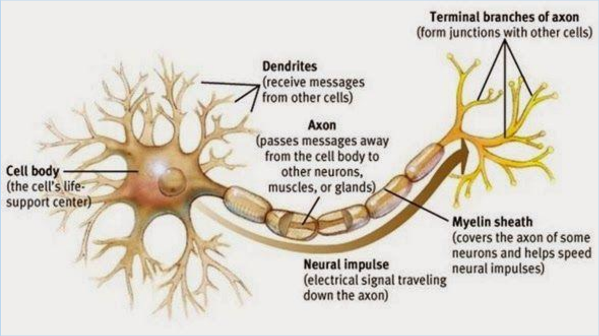
Leading Together

The INRC is a global network of more than **75 research groups** who are committed to delivering on the promise of neuromorphic computing to make the technology a commercial reality.

Source: INRC: https://www.intel.com/content/www/us/en/research/neuromorphic-community.html

Loihi neuromorphic core



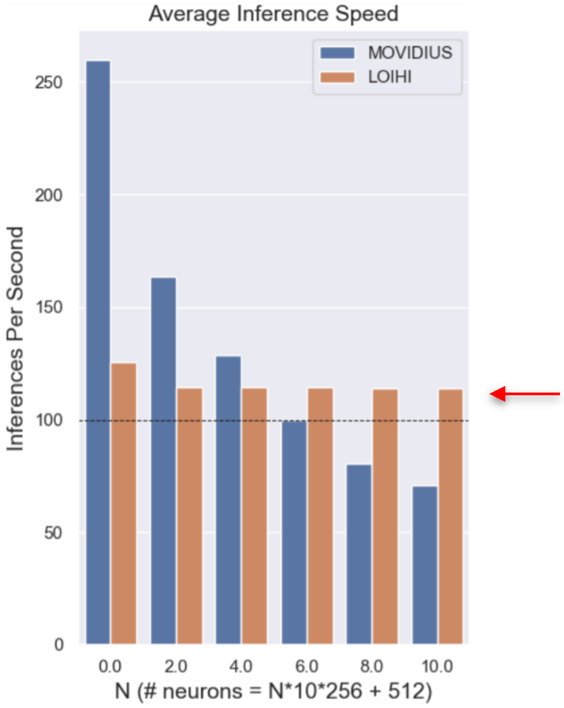


https://en.wikichip.org/wiki/intel/loihi

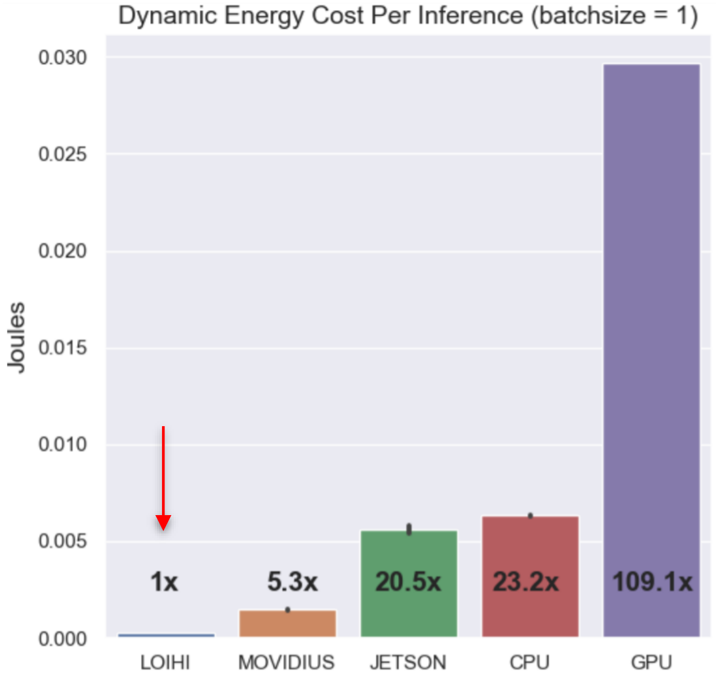
Hardware overview

Source: https://intel-ncl.atlassian.net/wiki/spaces/INRC/pages/2752539/Loihi+Systems+Overview

Selected Benchmarks



Speed comparison
(Loihi vs. Movidius)



Power Consumption
(different hardware types)

Source: https://arxiv.org/abs/1812.01739, arXiv:1812.01739
Peter Blouw, Xuan Choo, Eric Hunsberger, Chris Eliasmith - Benchmarking Keyword Spotting Efficiency on Neuromorphic Hardware

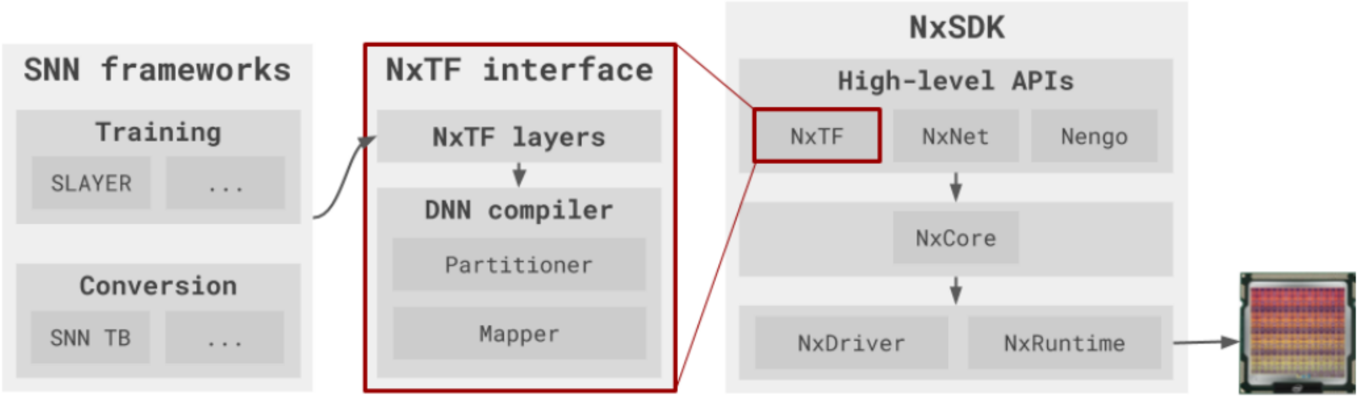
Software examples

NxSDK – dedicated software by Intel Labs

Nengo – supporting Loihi backend

SNNToolbox – conversion and simulations

Neuromorphic computing software is being intensively developed



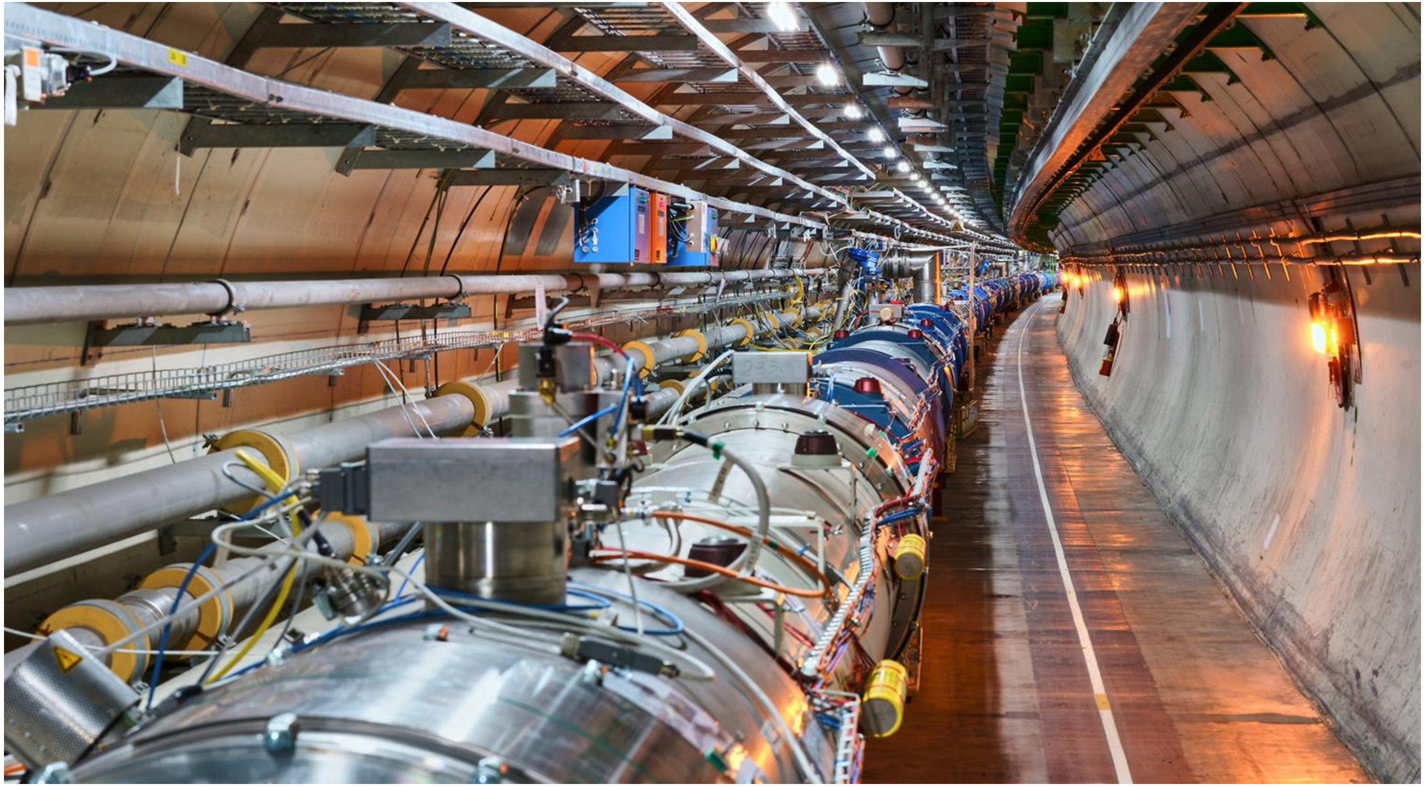
Source: https://snntoolbox.readthedocs.io/en/latest/guide/intro.html, https://www.nengo.ai/nengo-loihi/

Overview of experiments at CERN

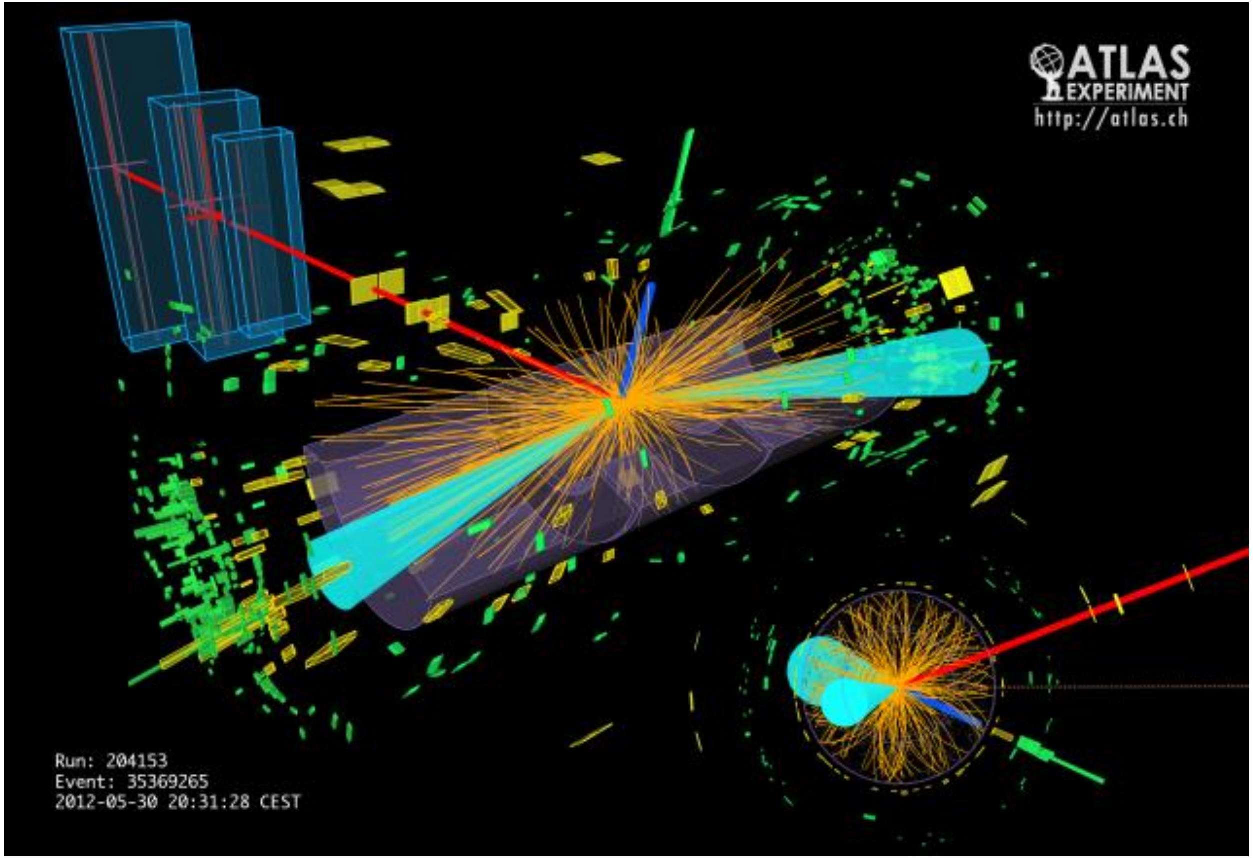
the European Organization for Nuclear Research

LHC - the Large Hadron Collider

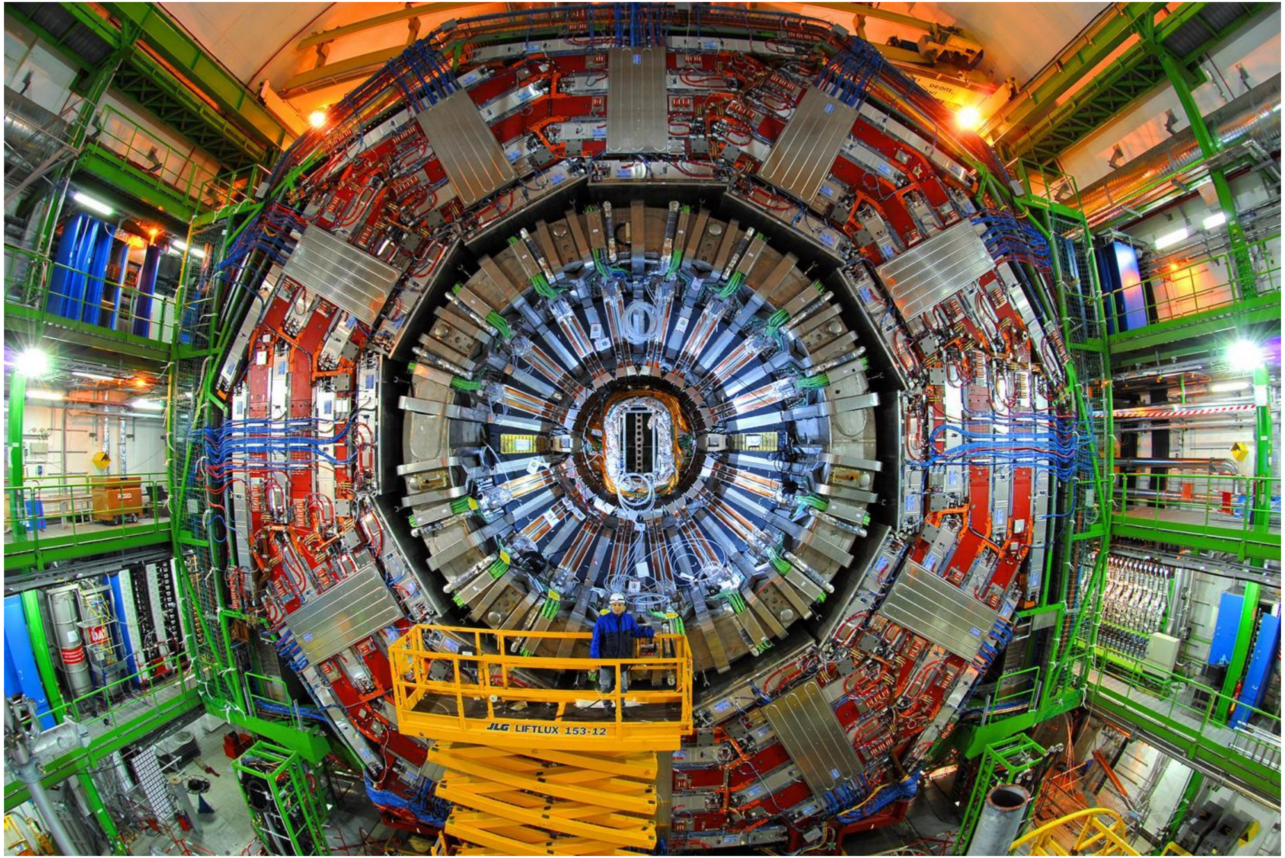
LHC - the Large Hadron Collider



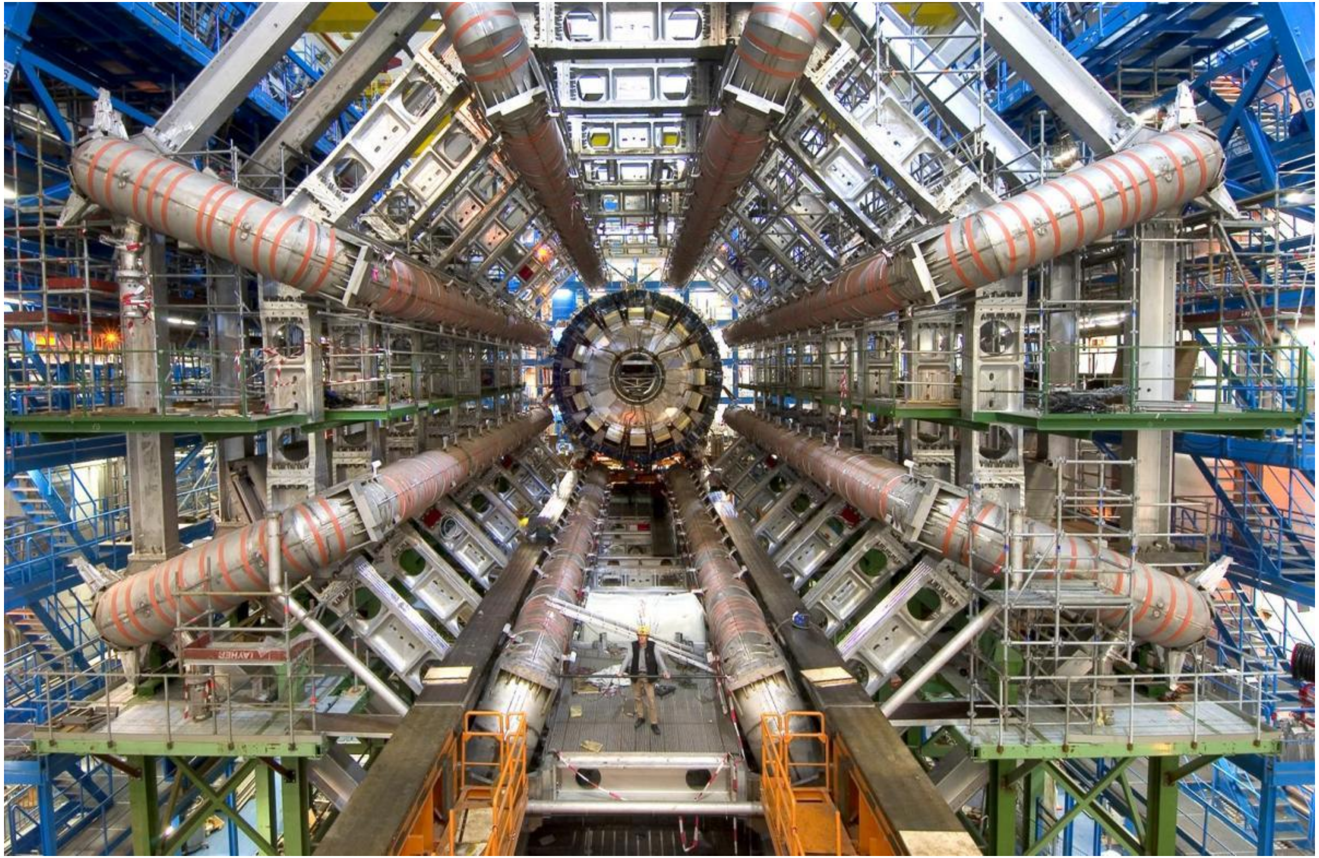
Collisions every 25ns



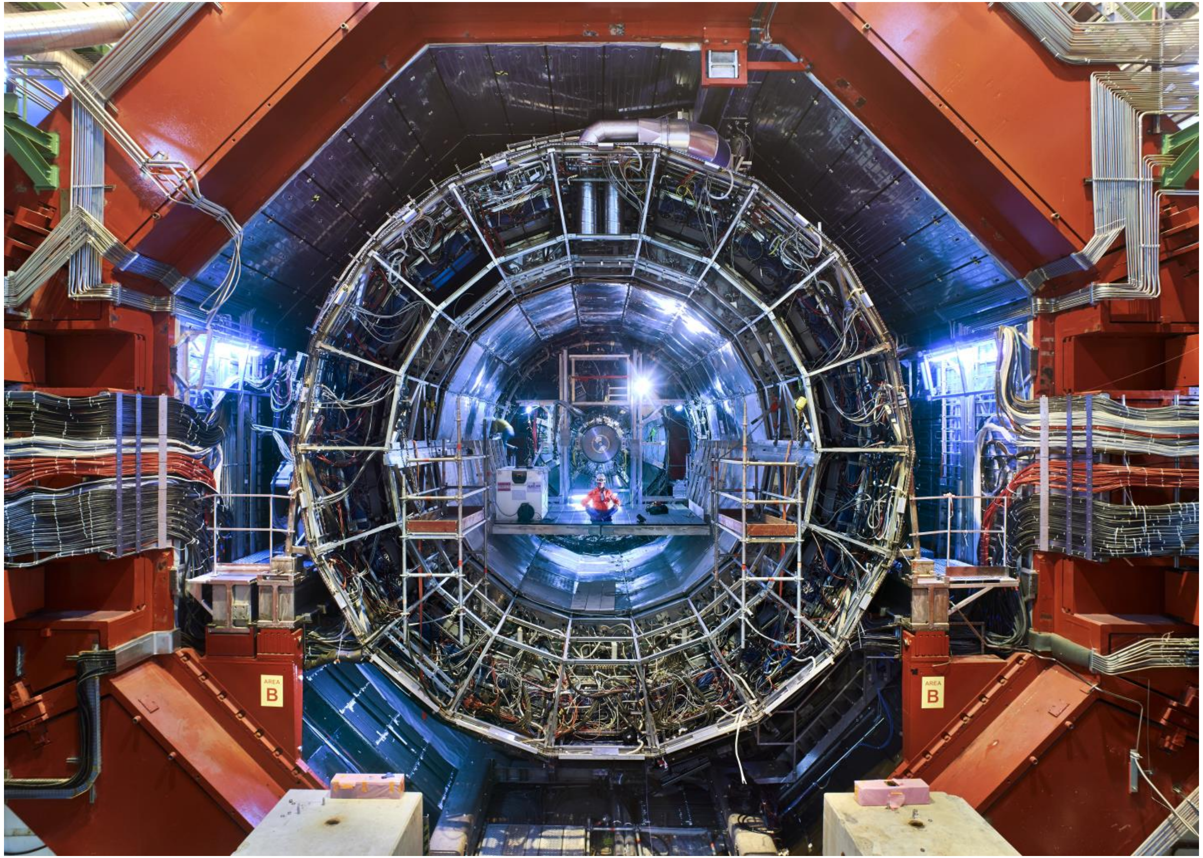
CMS



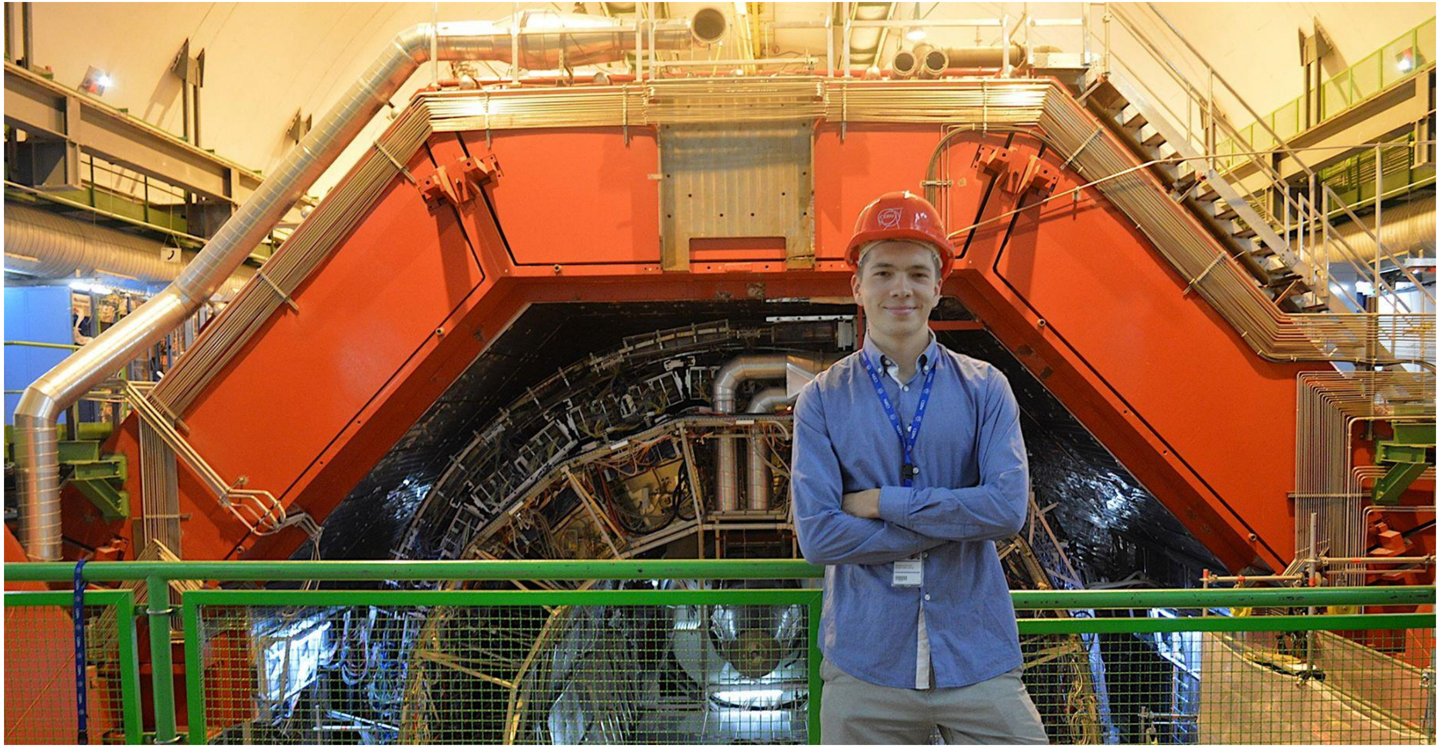
ATLAS



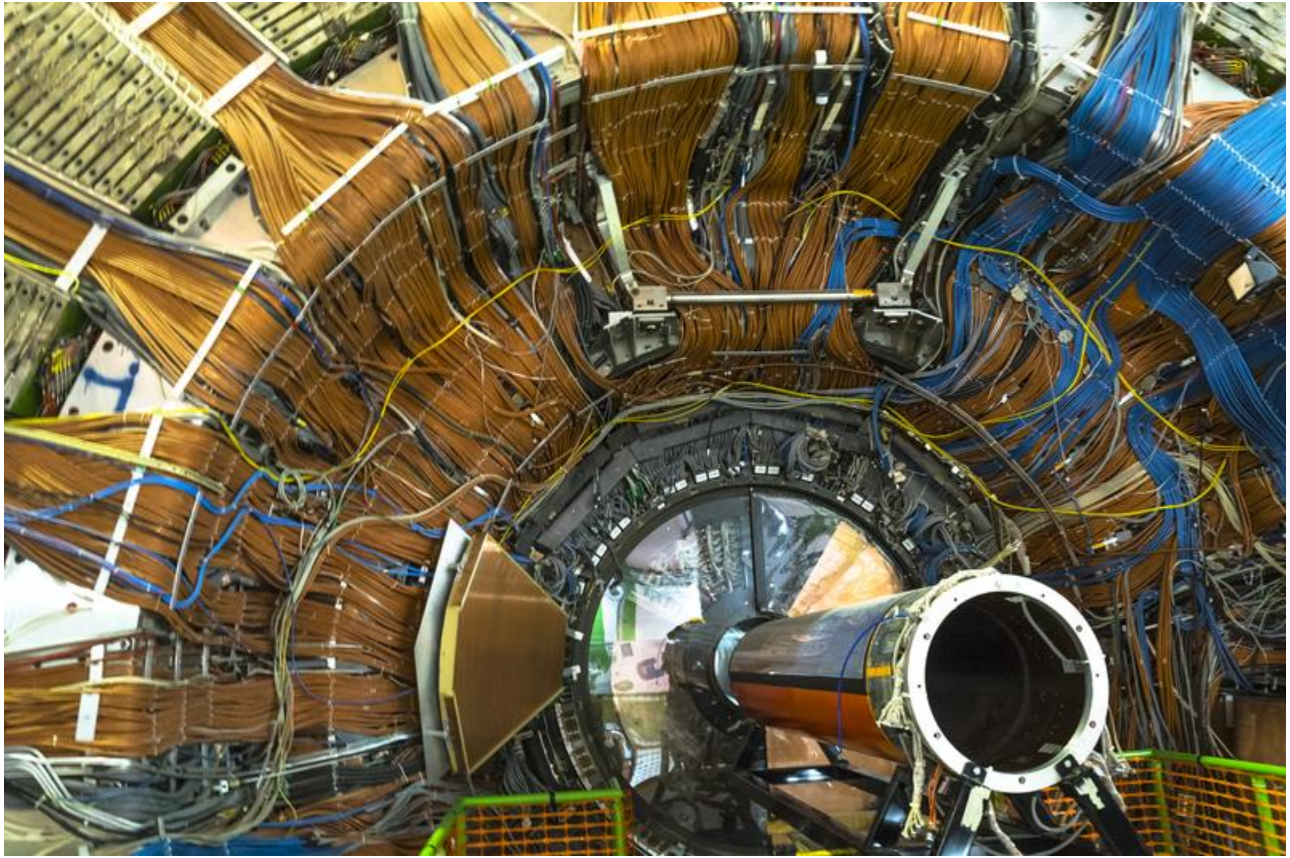
ALICE



Research at ALICE



LHCb



CERN openlab

Accelerating Science since 1954

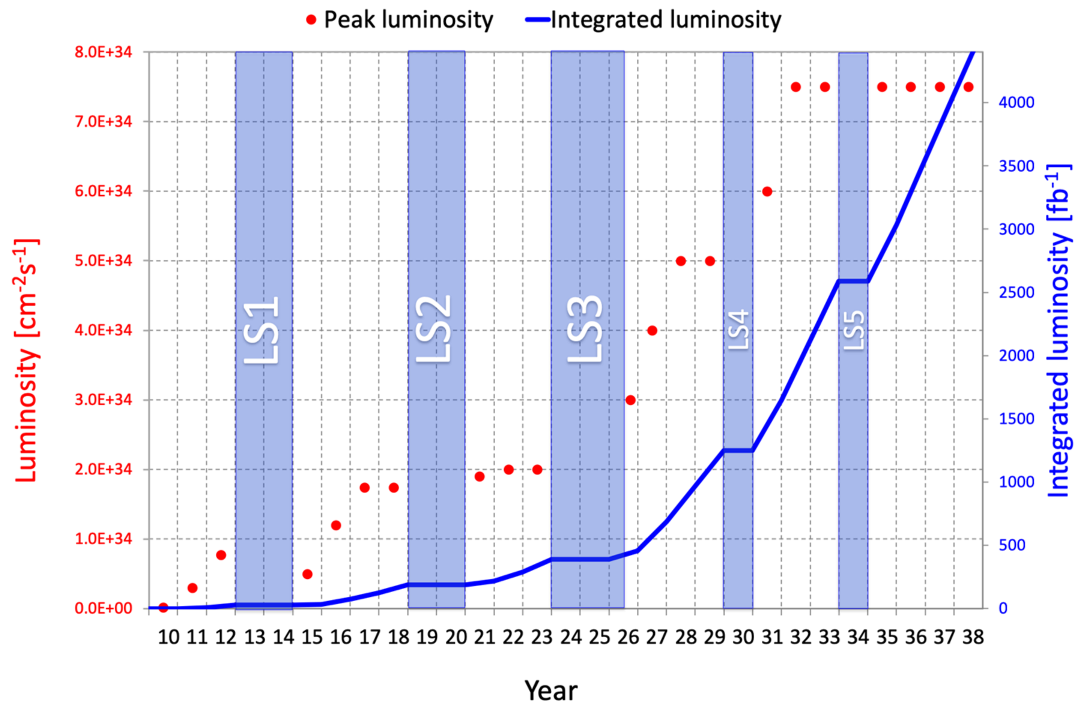
- 10 Nobel Prizes for scientists connected with CERN
- Over 2600 staff members and 8000 researchers from over 500 institutions
- 23 member countries (including Poland)
- Experimental explanation of the Standard Model (Higgs Boson discovery)
- Birthplace of the World Wide Web (1989)

CERN openlab

CERN openlab collaboration

- Collaboration with leading ICT companies
- Testing of products in CERN's demanding environment
- Applications of cutting-edge technology in HEP
- Impressive research opportunities

High-Luminosity LHC



Scheduled to start in 2026 with

10 times more data generated

How could we process it?

Let's apply AI

Source: https://lhc-commissioning.web.cern.ch/

Fast Inference of **ML** on **FPGAs** for **HEP** Trigger Systems

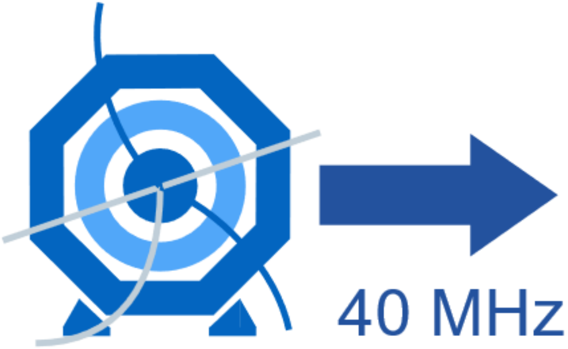
by Hamza Javed

Supervisors : Maurizio Pierini, Jennifer Ngadiuba, Vladimir Loncar

arxiv.org/abs/1804.06913
hls-fpga-machine-learning.github.io/hls4ml/

Problem

- 1 collision every 25 nanoseconds
- 99.99975% of data has to be rejected



Source: https://indico.cern.ch/event/830002/contributions/3523507/

Utilization of SNNs at CERN

- Data processing at High Luminosity LHC
- Signal-to-noise discrimination
- Classification of particles basing on the records

Triggering system of upgraded detectors

Source: http://hilumilhc.web.cern.ch/

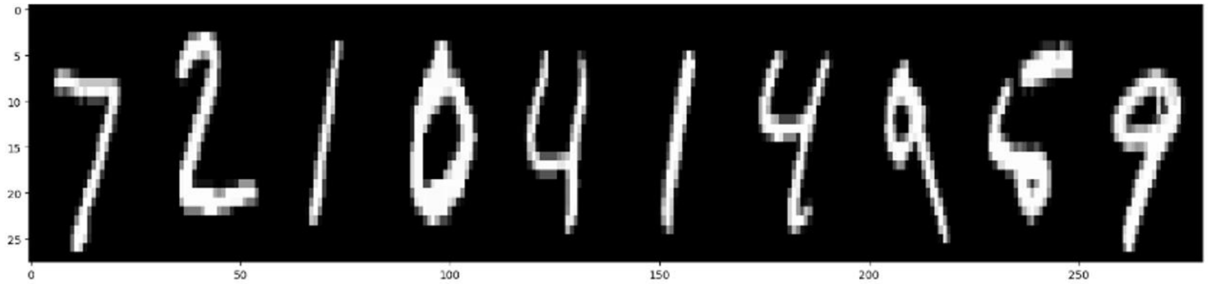
MNIST Classification

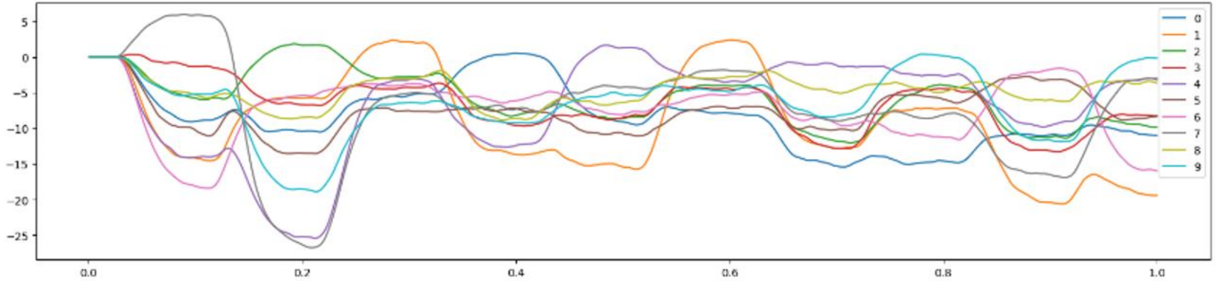
<u>SNN Simulation</u>:

2,79% error
(Simulated with Nengo-DL)

<u>SNN Deployment</u>:

2,00% error
(Deployed on the Loihi chip with Nengo)





Source: https://www.nengo.ai/nengo-loihi/examples/mnist_convnet.html

Jet Tagging Task at CMS Experiment

- 16 features (detector input)
- 5 particles (output classes):

  - gluon (**g**)

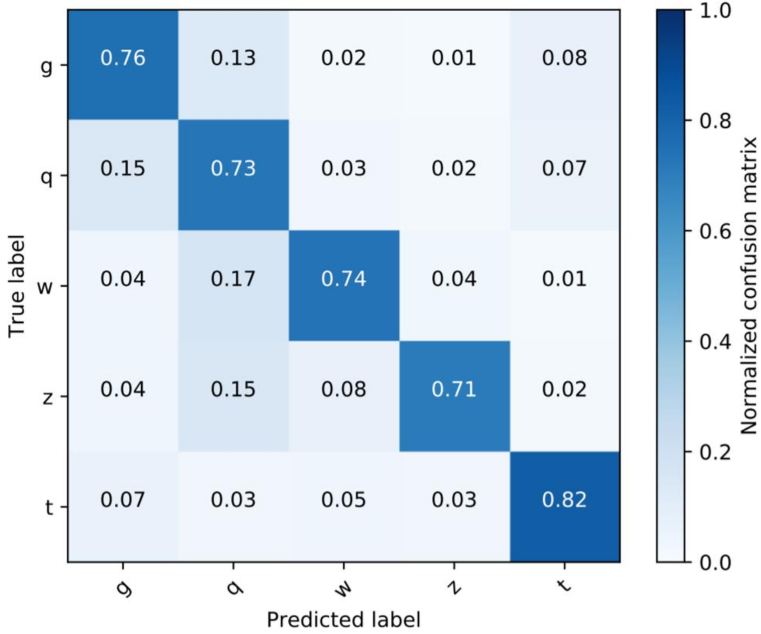
  - quark (**q**)

  - W boson (**w**)

  - Z boson (**z**)

  - top quark (**t**)

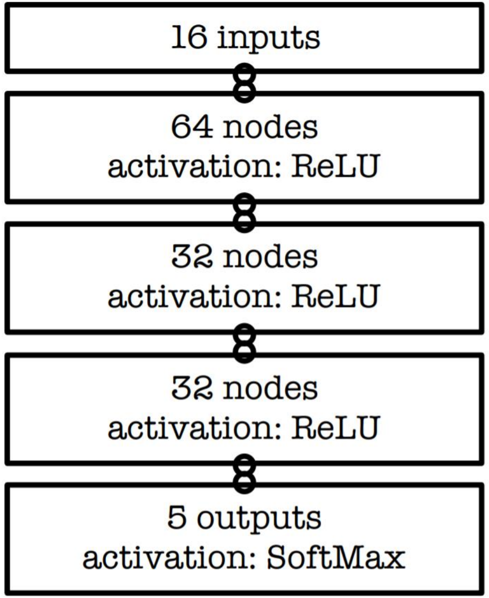
Current results with DNNs:



Source: https://arxiv.org/abs/1804.06913, arXiv:1804.06913
Javier Duarte et al. Fast inference of deep neural networks in FPGAs for particle physics

Jet Tagging – Spiking Neural Networks

ANN model architecture:



DNN accuracy:

- **75,20%**

(Trained in Keras on conventional hardware)

SNN accuracy (simulation):

- **69,72%**

(Trained in Keras, simulated in SNN Toolbox)

SNN accuracy (hardware):

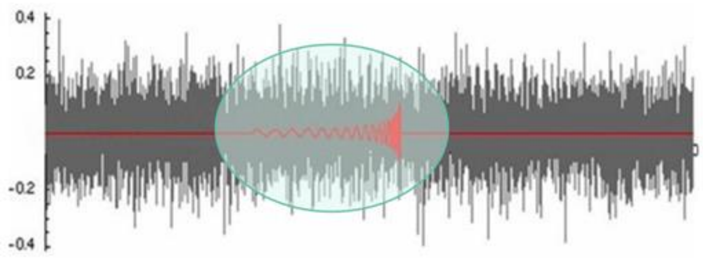
- **69,80%**

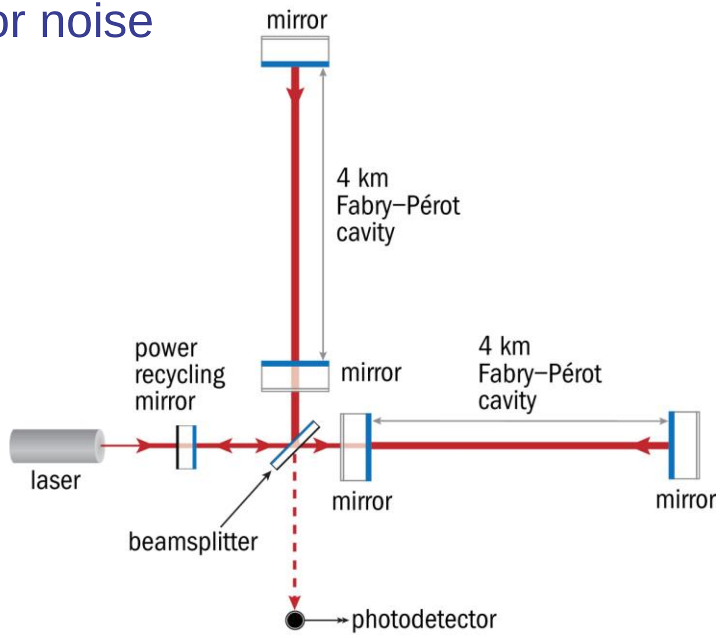
(Trained in Nengo, deployed on the Loihi chip)

github.com/borzyszkowski/SNN-CMS
DOI: 10.5281/zenodo.3755310

Anomaly detection in time series at LIGO

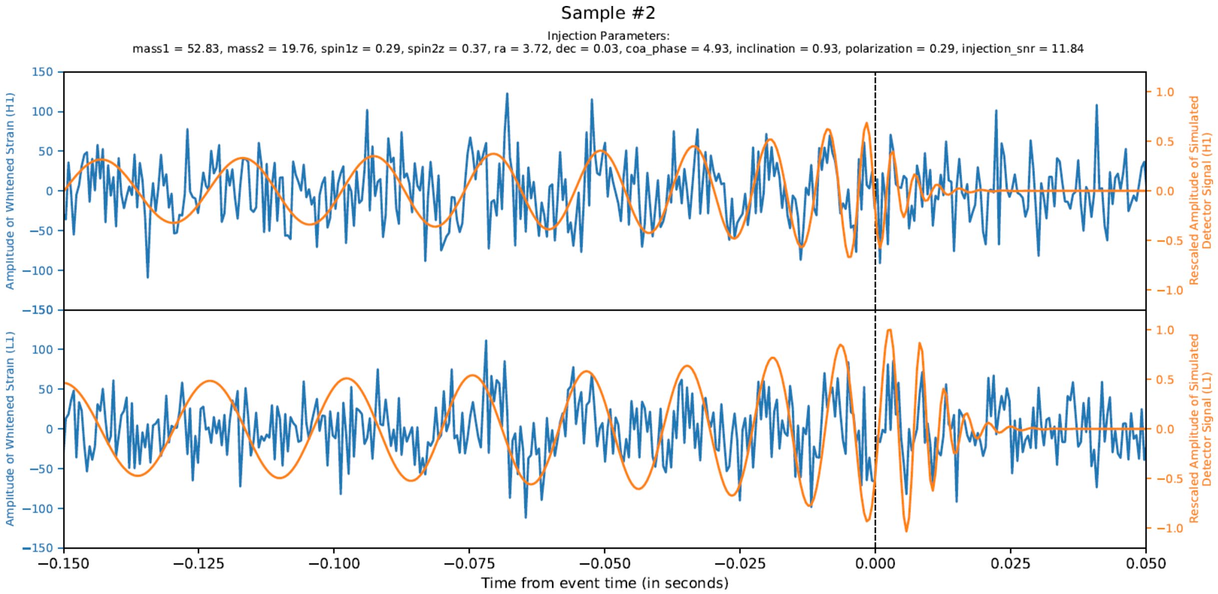
- Transient signals in gravitational waves
- Classification of glitches from the detector noise
- LIGO/Virgo data

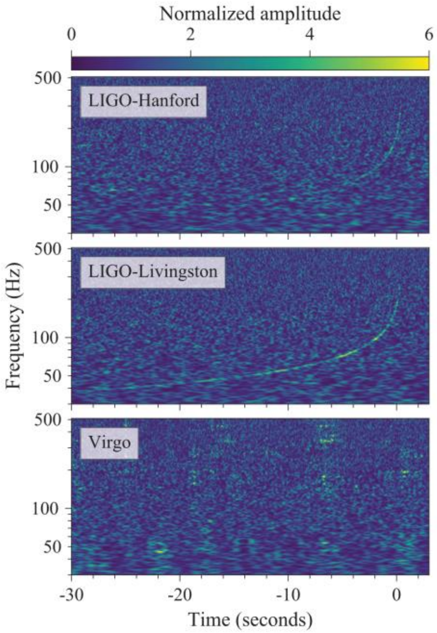




Source: Elena Cuoco - Real Time Classifier for transient signals in Gravitational Waves, From raw data to classified triggers

LIGO data





Source: github.com/timothygebhard/ggwd, https://www.gw-openscience.org/data/

Currently used methods

Matched Filtering

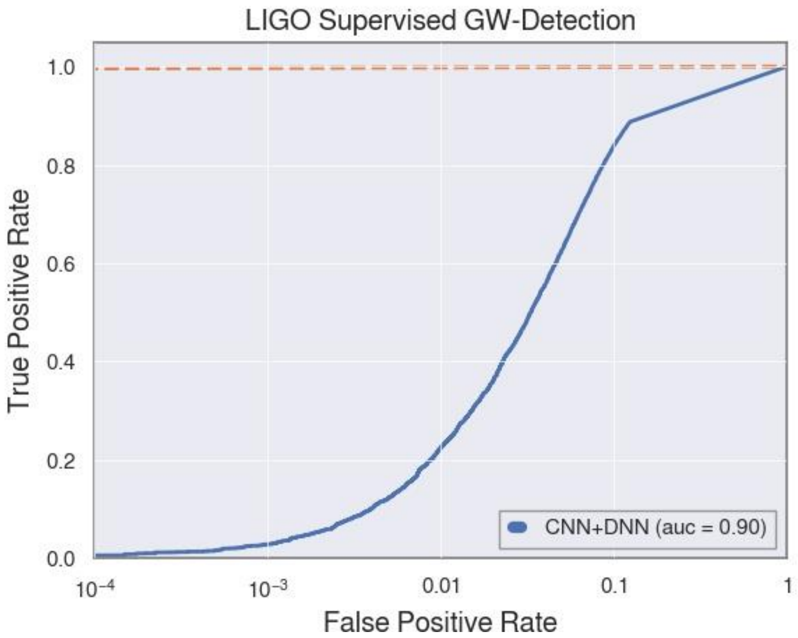
- **Current method** used by LIGO
- Compares incoming GW data to bank of simulated waveforms
- Can only identify GWs that are available in GW banks (no exotic events)

Deep Filtering

- Convolutional Neural Networks (CNNs)
- Take time-series inputs, can determine detections and estimate parameters of events
- Still can miss events that aren't included in training set

Source: https://arxiv.org/pdf/1711.03121.pdf, https://arxiv.org/pdf/2005.06534.pdf ,

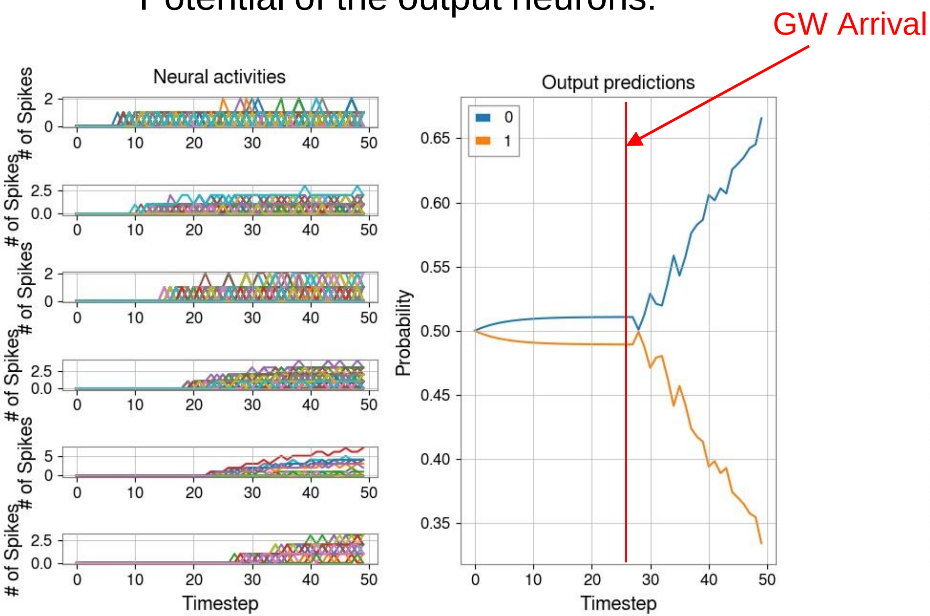
Supervised GW detection



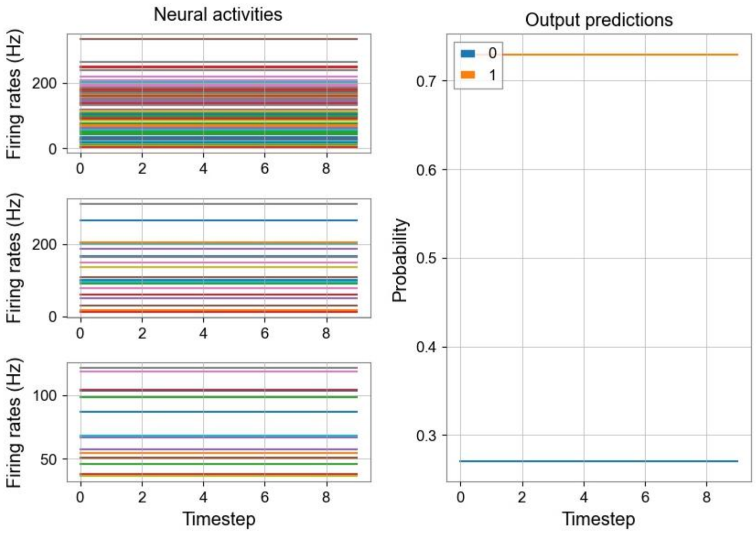
	Input	vector (size: 8192)
1	Reshape	matrix (size: 1 × 8192)
2	Convolution	matrix (size: 64 × 8177)
3	Pooling	matrix (size: 64 × 2044)
4	ReLU	matrix (size: 64 × 2044)
5	Convolution	matrix (size: 128 × 2014)
6	Pooling	matrix (size: 128 × 503)
7	ReLU	matrix (size: 128 × 503)
8	Convolution	matrix (size: 256 × 473)
9	Pooling	matrix (size: 256 × 118)
10	ReLU	matrix (size: 256 × 118)
11	Convolution	matrix (size: 512 × 56)
12	Pooling	matrix (size: 512 × 14)
13	ReLU	matrix (size: 512 × 14)
14	Flatten	vector (size: 7168)
15	Linear Layer	vector (size: 128)
16	ReLU	vector (size: 128)
17	Linear Layer	vector (size: 64)
18	ReLU	vector (size: 64)
19	Linear Layer	vector (size: 2)
	Output	vector (size: 2)

Neuromorphic GW Detection

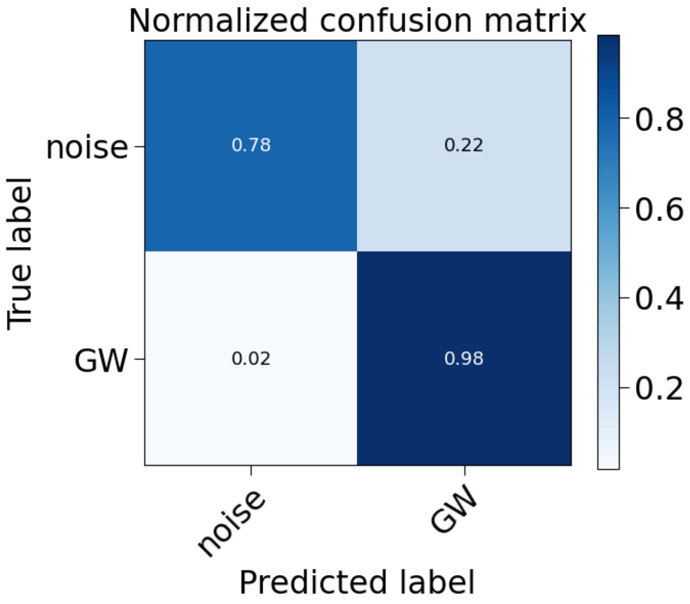
Potential of the output neurons:

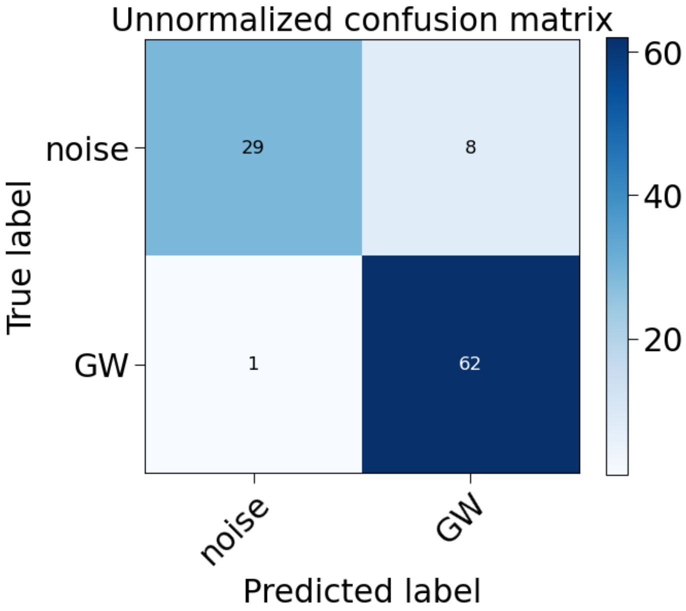


Binary spike (classification):



Preliminary results





Future opportunities

- Triggering system at CERN
- Direct (on-chip) training
- Spiking RNNs
- … and much more

Project supervised by:

- Maurizio Pierini (CERN)
- Jean-Roch Vlimant (Caltech)

CMS Experiment

Research at LIGO was done together with

- Eric Moreno (MIT)

Special thanks to Intel and CERN openlab for supporting the project

Beyond the science

Mountains

40 visions of creativity

CERN openlab


AERO
NYC 1987

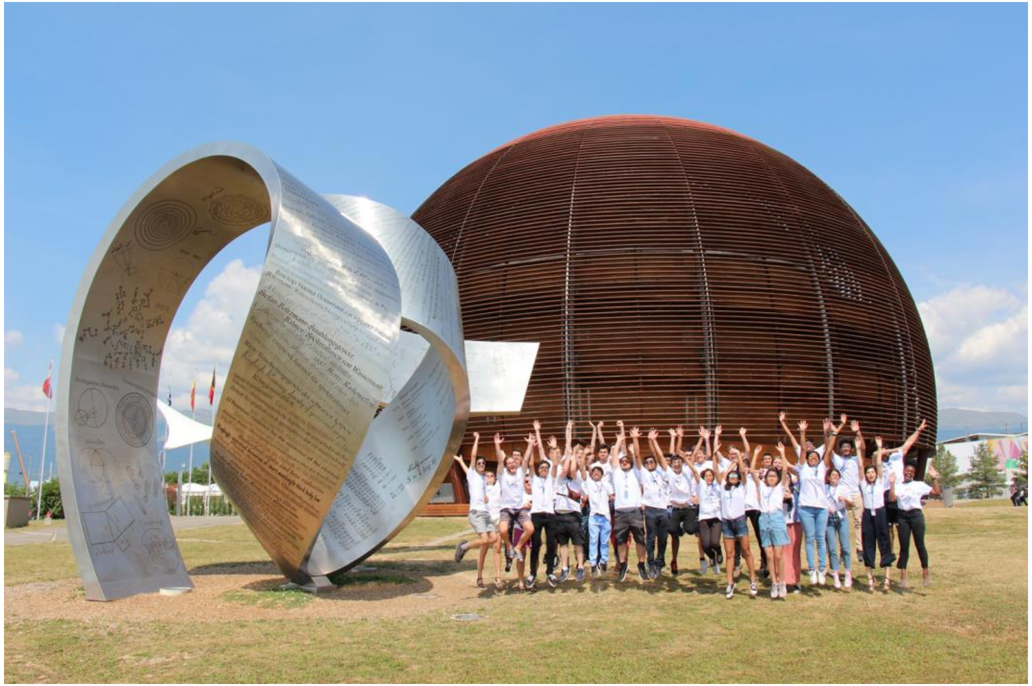
Join research at CERN!

- Summer Student Programme
(STEM students, member states)
(up to 13 weeks, ~300 students)

- openlab Summer Student Programme
(IT students, all over the world)
(9 weeks, ~40 students)

- Technical Student Programme
(STEM students, member states)
(up to 12 months, ~120 students)

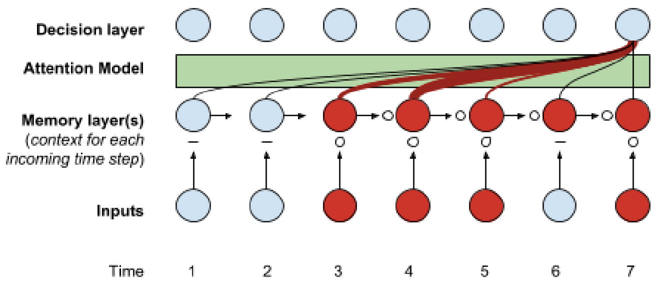
- Summer Programs application
deadline: 31 January 2022



https://careers.cern/students

Thank you for your attention!

Attention Mechanism



Discussion and Questions