M.Dallavalle, D.Bonacorsi, R.Travaglini, L.Guiducci,
A.Perrotta, A.Montanari, C.Baldanza, F.Odorici

Bologna, 13/11/2018

Proponiamo di formare un

gruppo di attività sul Machine Learning

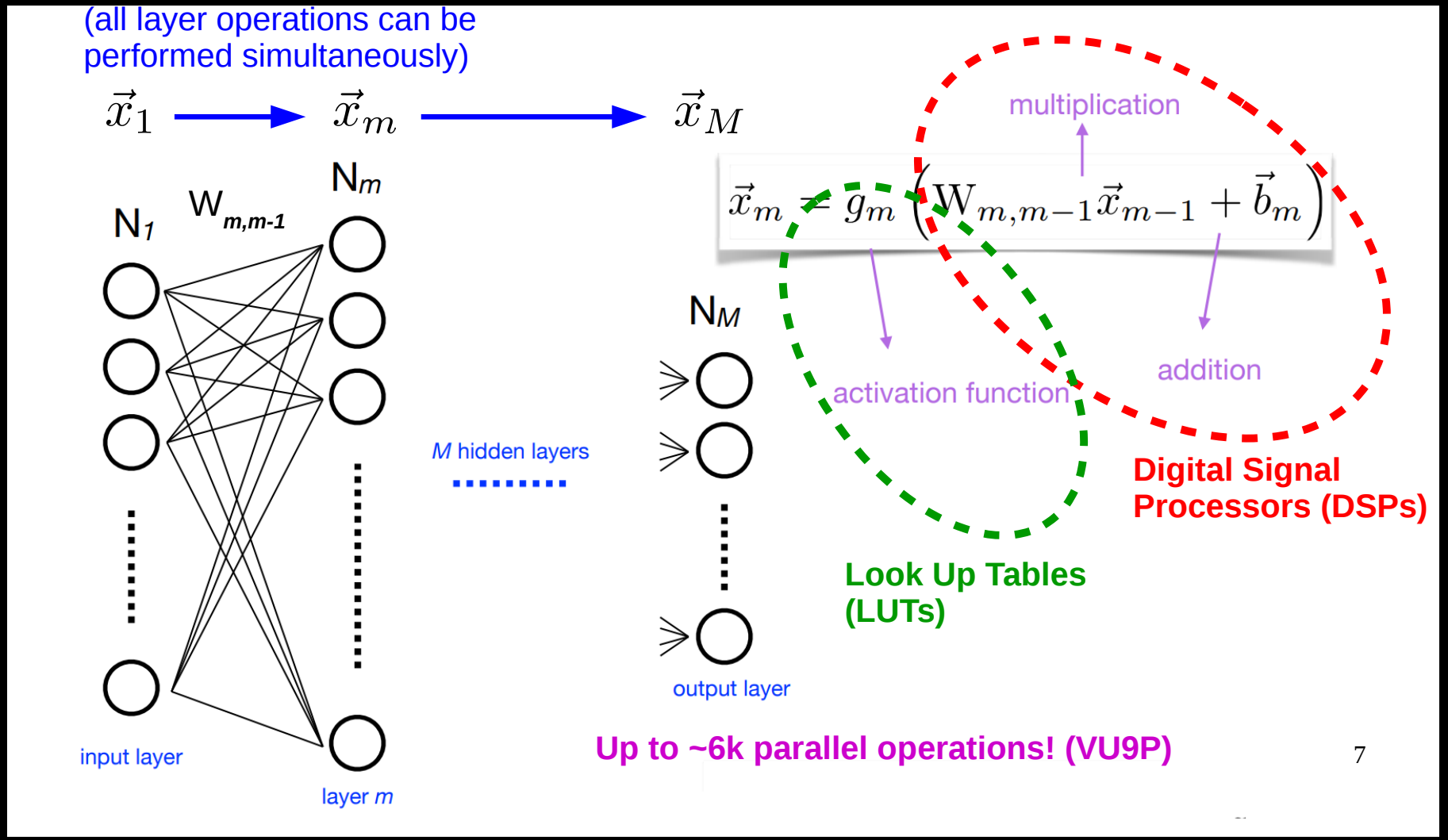
L'iniziativa è aperta a tutti.
Il target è un know-how di sezione, condiviso.

La nostra enfasi è sull'implementazione nell'elettronica dei rivelatori (*) in HE, ma …

molti sono gli utilizzi al di fuori del nostro specifico contesto:
diagnostica medica, automotive, QC catene di produzione, olografia digitale, etc …

(*) Slow Controls, Quality Control, Pattern Recognition, Event Selection, Trigger,…

Machine Learning è Reti Neurali NN (ma non solo)



NN Profondo (Deep) = molti layers intermedi

Alcuni tipi di architetture: MLP, CNN, LSTM, BDT

DNNs in Industry

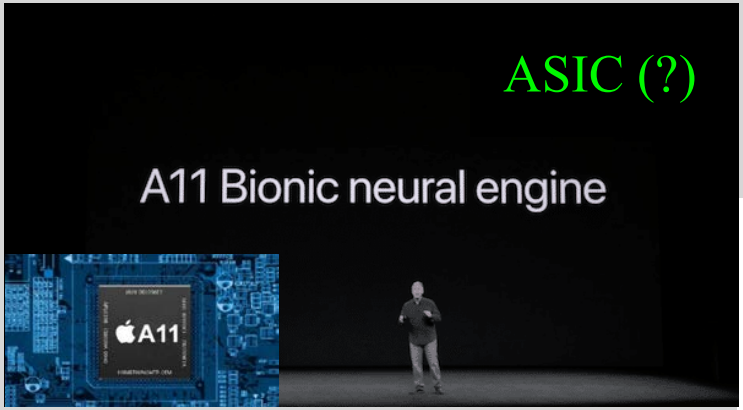
- Image recognition uses convolution neural networks (CNNs)
- Examples: ResNet, GoogLeNet, VGGNet, etc.
- Useful papers: arXiv:1512.03385, arXiv:1605.07678
- These networks generate a large set of features from input (2D pixel grid)
- Features can be used by another fully-connected network for classification
  - Featurizing is accelerated by FPGA, classification proceeds on CPU

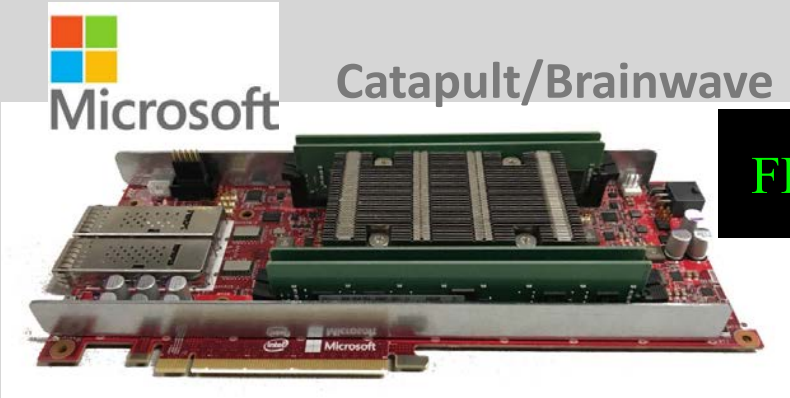
Brainwave supports:
- ResNet50
- ResNet152
- DenseNet121
- VGGNet16

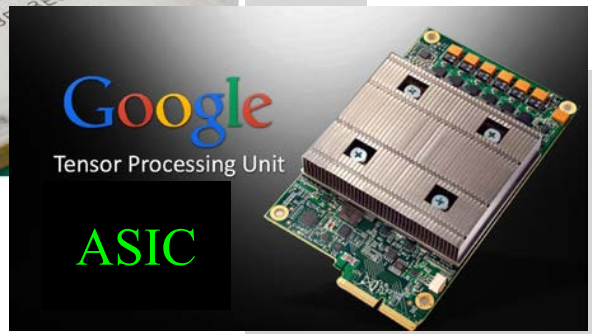
- Brainwave also allows weight tuning
  - Retrain supported architecture, provide new weights

Co-processors: An Industry Trend

Specialized co-processor hardware for machine learning inference







INTEL® FPGA ACCELERATION HUB

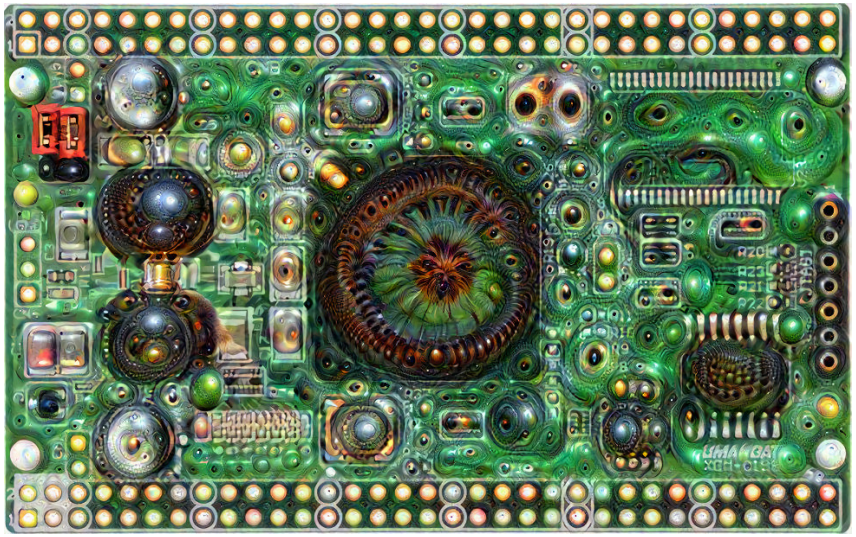
The Intel® Xeon® Acceleration Stack for FPGAs is a robust framework enabling data center applications to leverage an FPGA's potential to increase

Universität Zürich[UZH]

Physik-Institut

How to do ultrafast Deep Neural Network inference on FPGAs

6. February 2019
Physik Institut - Universität Zürich



A key component in autonomous vehicle and low-latency triggering systems, learn how FPGAs do
real-time DNN inference in this hands-on course. Topics include:

- Model compression and quantization
- High-level synthesis
- Firmware implementation
- Model acceleration on cloud FPGAs

Lecturers:
Dr. Jennifer Ngadiuba (CERN)
Dr. Dylan Rankin (MIT)

Registration and further info at
indico.cern.ch/e/FPGA4HEP

Organizers:
Thea Årrestad (UZH)
Jennifer Ngadiuba (CERN)
Dylan Rankin (MIT)
Maurizio Pierini (CERN)
Ben Kilminster (UZH)

HLS4ML è in sviluppo al CERN

hls-fpga-machine-learning.github.io/hls4ml/

Type to search

- hls4ml
- Status and Features
- Setup
  - Dependencies
  - Quick Start
  - Configuration
- Concepts
- Release Notes
- Reference and Contributors
- Code Repository

Published with GitBook

A package for machine learning inference in FPGAs. We create firmware implementations of machine learning algorithms using high level synthesis language (HLS). We translate traditional open-source machine learning package models into HLS that can be configured for your use-case!

The project is currently in development, so please let us know if you are interested, your experiences with the package, and if you would like new features to be added.

contact: hls4ml.help@gmail.com

Project status

For the latest status including current and planned features, see the Status and Features page.

L’attività richiede partecipanti
con conoscenze differenziate e complementari:

— SW degli algoritmi ML, Deep ML, caratterizzazione, disponibilità sul mercato

— progettazione del training,

— implementazione FW, in particolare su FPGA

— architettura dell’HW, per reti estese

— sistema di test, compatibilità con i grandi esperimenti e con il mondo industriale/commerciale. Pensiamo all’ambiente ATCA. Attrezzatura di sezione, condivisa.

Per organizzare l'attività è utile, all'inizio, focalizzare su uno use-case.

Per CMS pensiamo ad un trigger con muoni nel barrel, in modo da confrontare con quello corrente, che conosciamo in dettaglio.

Altro esempio:
separazione dei jet in eventi adronici boostati (top)

La nostra sezione ha esperienze pregresse sui NN:
1. MA16 1992-1998 chip Intel— Odorico, Baldanza, D'Antone, Malferrari, Mazzanti, Odorici
2. ML in HEP: community white paper (2018): https://arxiv.org/abs/1807.02876
3. ML at the energy and intensity frontiers of PP (Nature, 2018): bit.ly/ML-DBonacorsi
4. Progress in ML Studies for the CMS Computing Infrastructure (2017): https://pos.sissa.it/293/023

TESI:
4. nel gruppo CMS, 4 tesi (sia L che LM) su applicazione di ML a problemi di Computing e Muon Trigger
5. docenza (D. Bonacorsi) al corso "Applied ML" al Dottorato UniBo/Golinelli in "Data Science and Computation"

NON SOLO PER LHC:

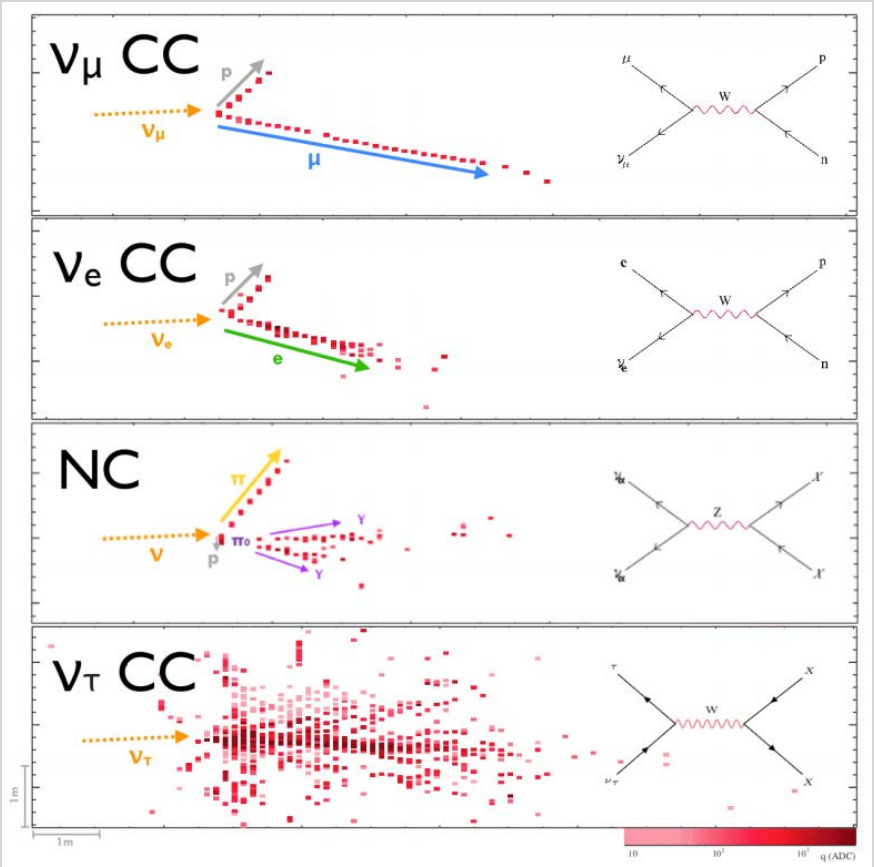
Image Recognition for Neutrinos

Key challenges:

- Discriminating between muons and charged pions (muons produce longer tracks, less interaction with nuclei)
- Discrimination between electron and photons (photons can travel a short distance before showering)

CNNs used for neutrinos:

- arXiv:1604.01444 (NOvA)
- arXiv:1611.05531 (MicroBooNE)



STATO (13/11/2018):

R.Travaglini, C.Baldanza hanno proposto un HW

Anche CNAF interessati

G.Bruni, G.Maron, A.Perrotta, V.Vagnoni sostengono

Crate localizzabile nei locali TIER1

HW disponibile un paio di mesi dall'ordine (Feb 2019?)

Avvio raccolta delle adesioni e formazione del gruppo

Primo task: istruirsi en attendant Godot (HW):

- dividersi in gruppi di studio secondo gli interessi,
- preparare relazioni, seminari per spiegare agli altri