20th AGATA Week, 16-20 September 2019

Data Analysis, Past, Present & Future

O. Stézowski

➥ It contains : Status / GRETINA in AGATA / Machine Learning

A slide from the past …

Data Analysis – Team #3

ROOT as a framework for AGATA

O. Stézowski IPN Lyon

AGATA Week 15-19 September 2003 Legnaro

A slide from the past …

Propositions made during this talk …

Framework approach … not obvious back in 2003 !

- ☛ See software as hardware
- ☛ it means collaborative works, infrastructure for software, compatibility etc …

It has required the choice of some technologies

- ☛ **C++ / ROOT**

NOTE: The talk mentioned also parallel processing, needs of computing power …

A slide from the past …

Propositions made during this talk …

Framework approach … not obvious back in 2003 ! ✔

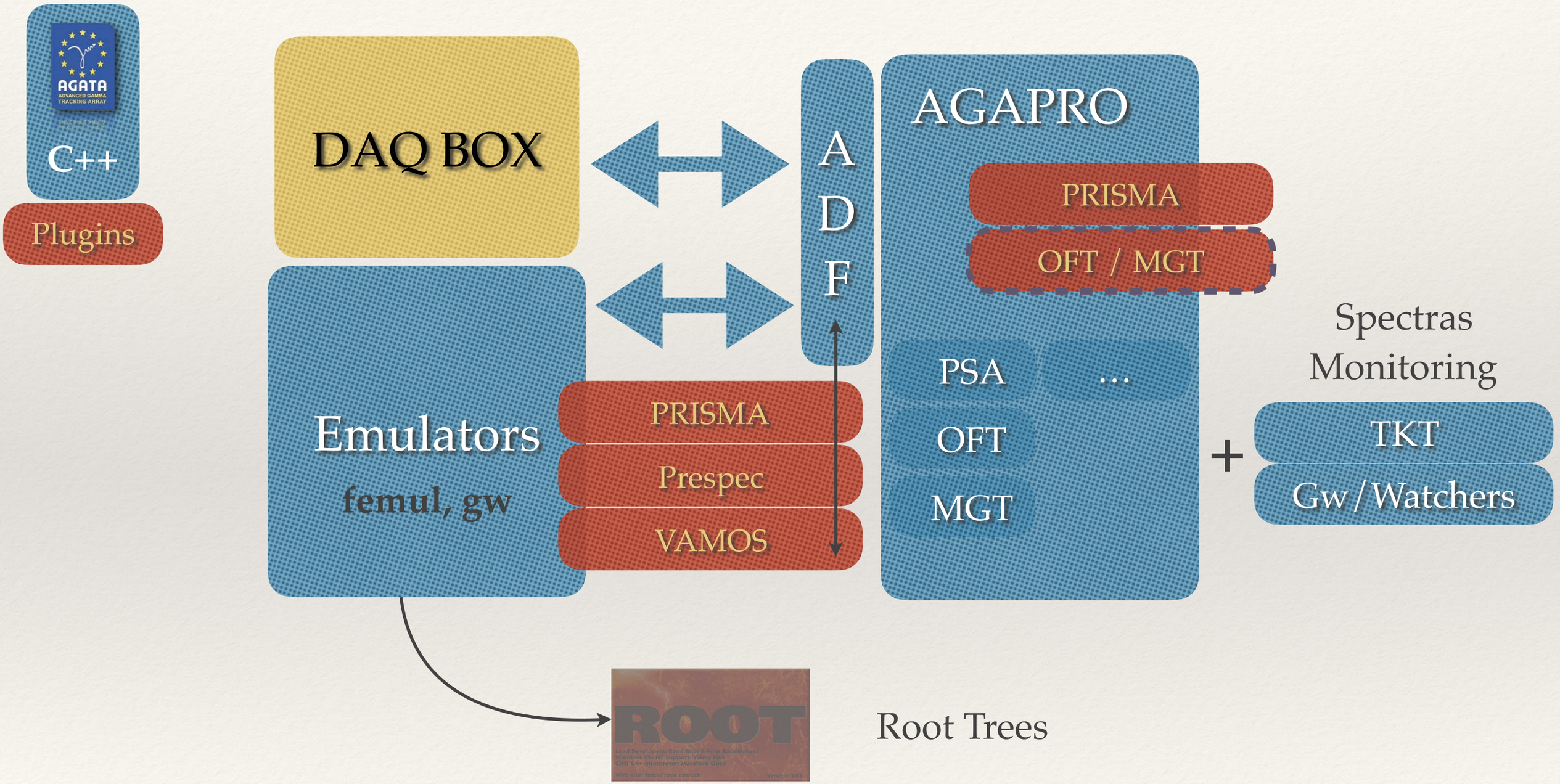
- ☛ See software as hardware
- ☛ it means collaborative works, infrastructure for software, compatibility etc …

It has required the choice of some technologies

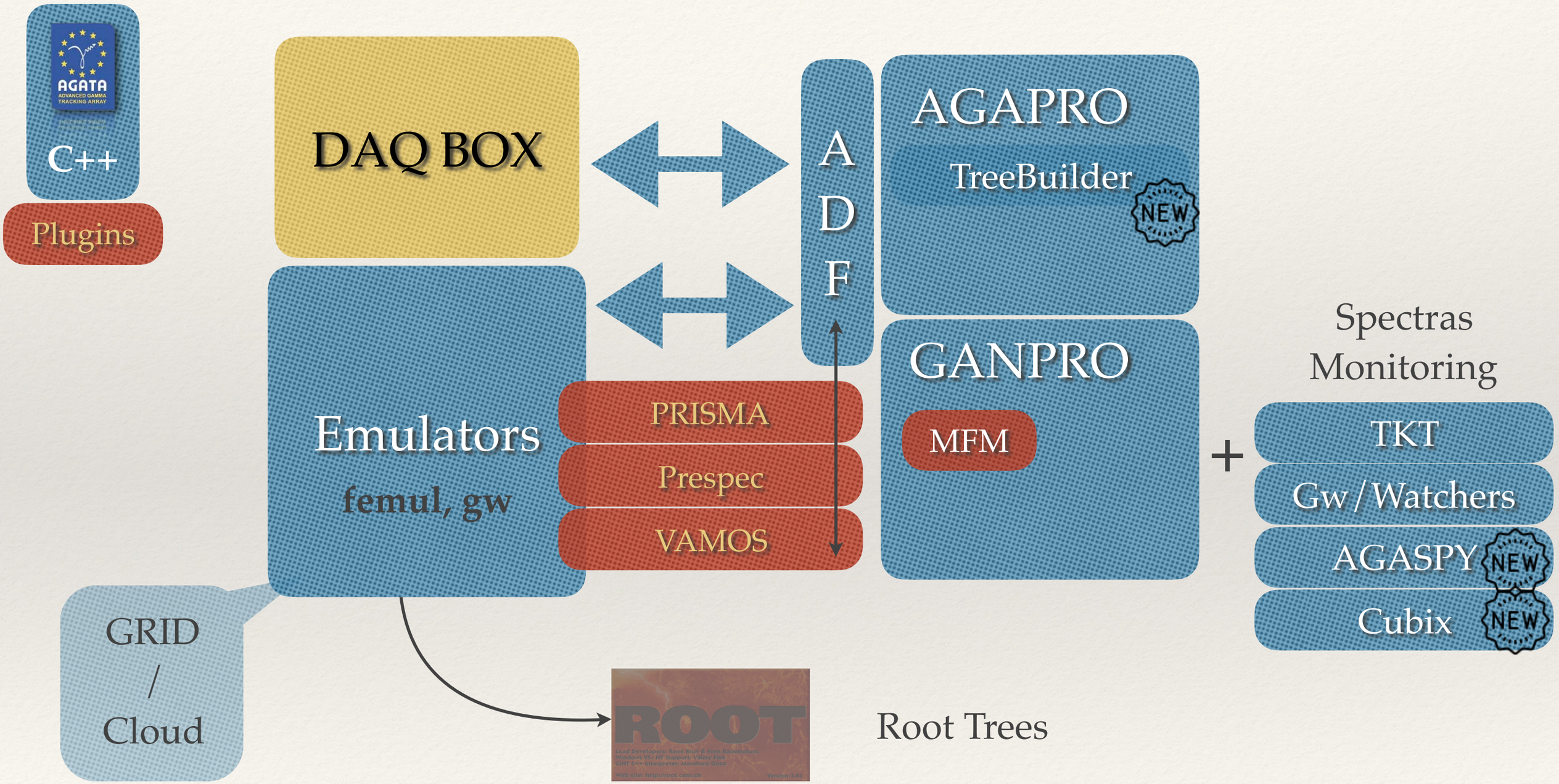
- ☛ C++ ✔ / **ROOT** ✔

NOTE: The talk mentioned also parallel processing, needs of computing power … ≈

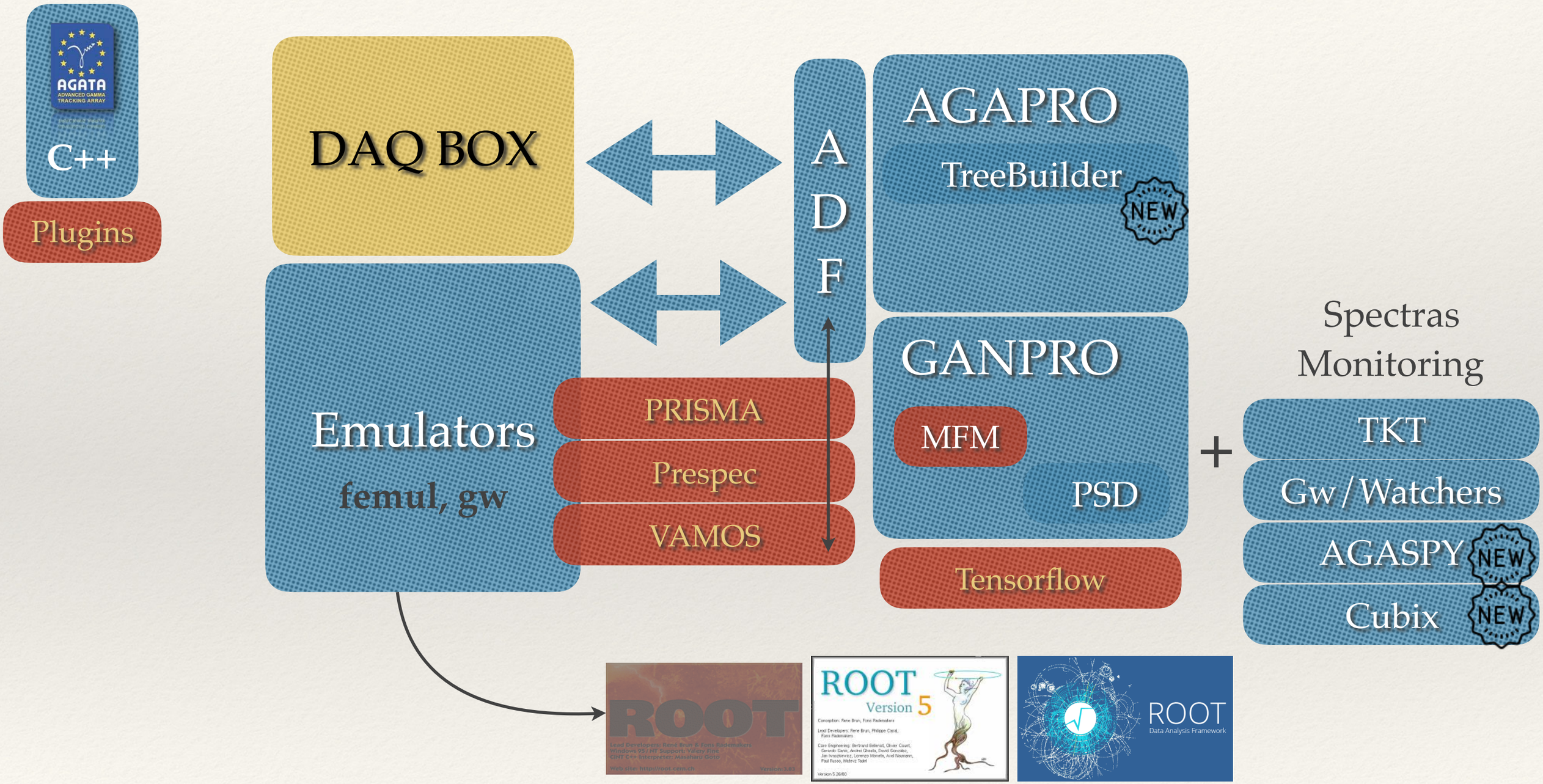
Past / Present - Framework approach



Past / Present - Framework approach



Past / Present - Framework approach



3 different sites LNL - GSI - GANIL ! **ROOT Migration** **svn ☛ git Migration**

Past / Present - Framework approach

Software tracked fro almost 10 years

Collaborative developments

IPNL_GAMMA › agapro › Commits › **b58fc1cd**

Commit b58fc1cd authored 9 years ago by **dino**

Browse files | Options

```
git-svn-id: svn://gal-serv.lnl.infn.it/agata/trunk/narval_emulator@736 170316e4-aea8-4b27-aad4-0380ec0519c9
```

parent f52e8da4 master ···

No related merge requests found

Cmake has built system almost from the beginning
☛ migration to 'modern' make has started …

Changes 1

Showing 1 changed file with 0 additions and 0 deletions

Hide whitespace changes | Inline | Side-by-side

filters/Ancillary/includeVME/DANTE.h → filters/Ancillary/includeVME/Dante.h

View file @ b58fc1cd

File moved

Past / Present - Framework approach

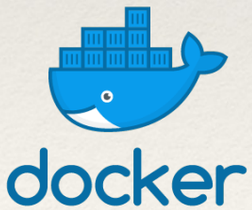
Last developments regarding software infrastructure:
modifications of the code trigs

Automatic Check of the Quality of the code :

☛ it allows to identified possible bugs, suggestion of more efficient code

Continuous integration :

☛ AGAPRO, GANPRO, Gw, femul, ReplayLLP, ReplayGLP …

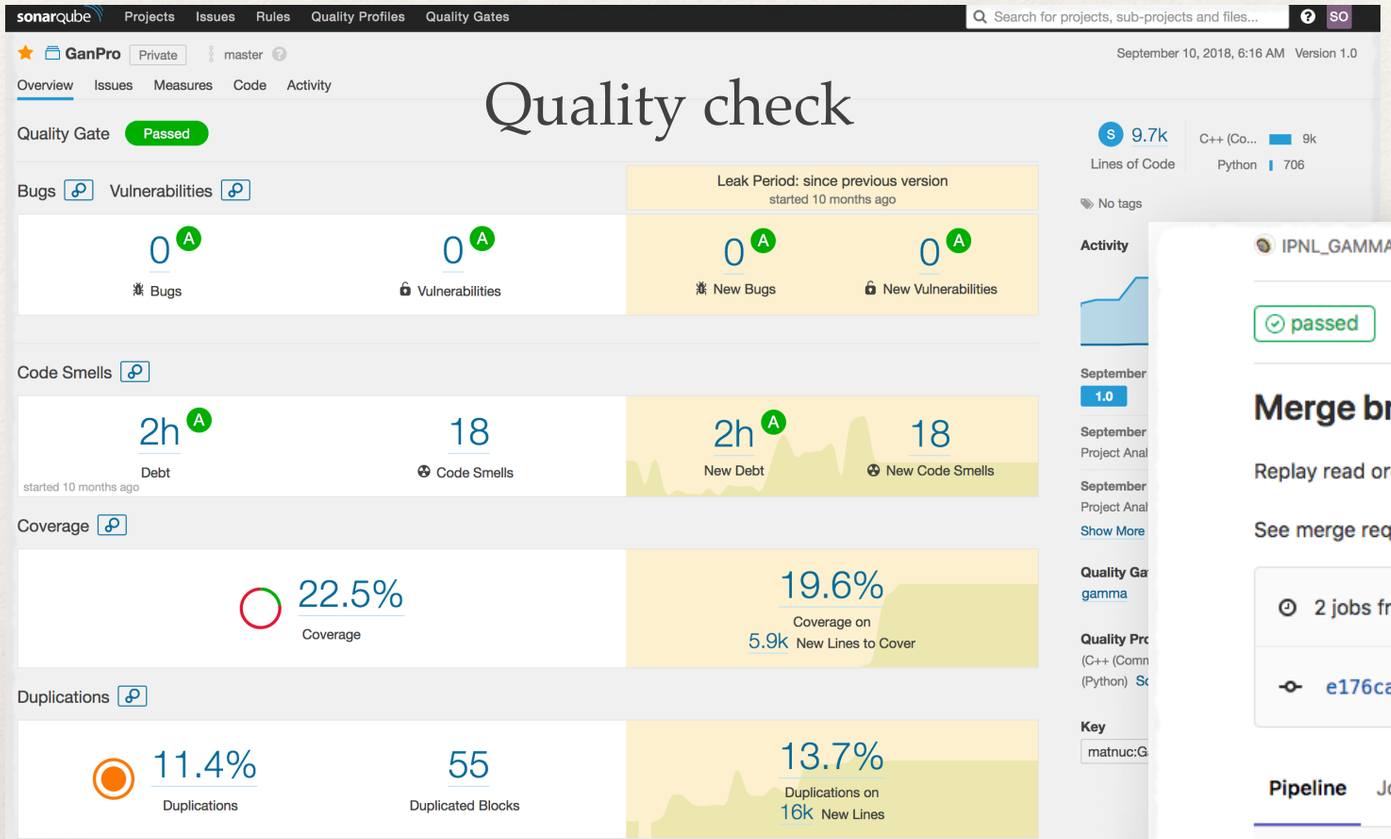


For those tests, **Containers** are used for that

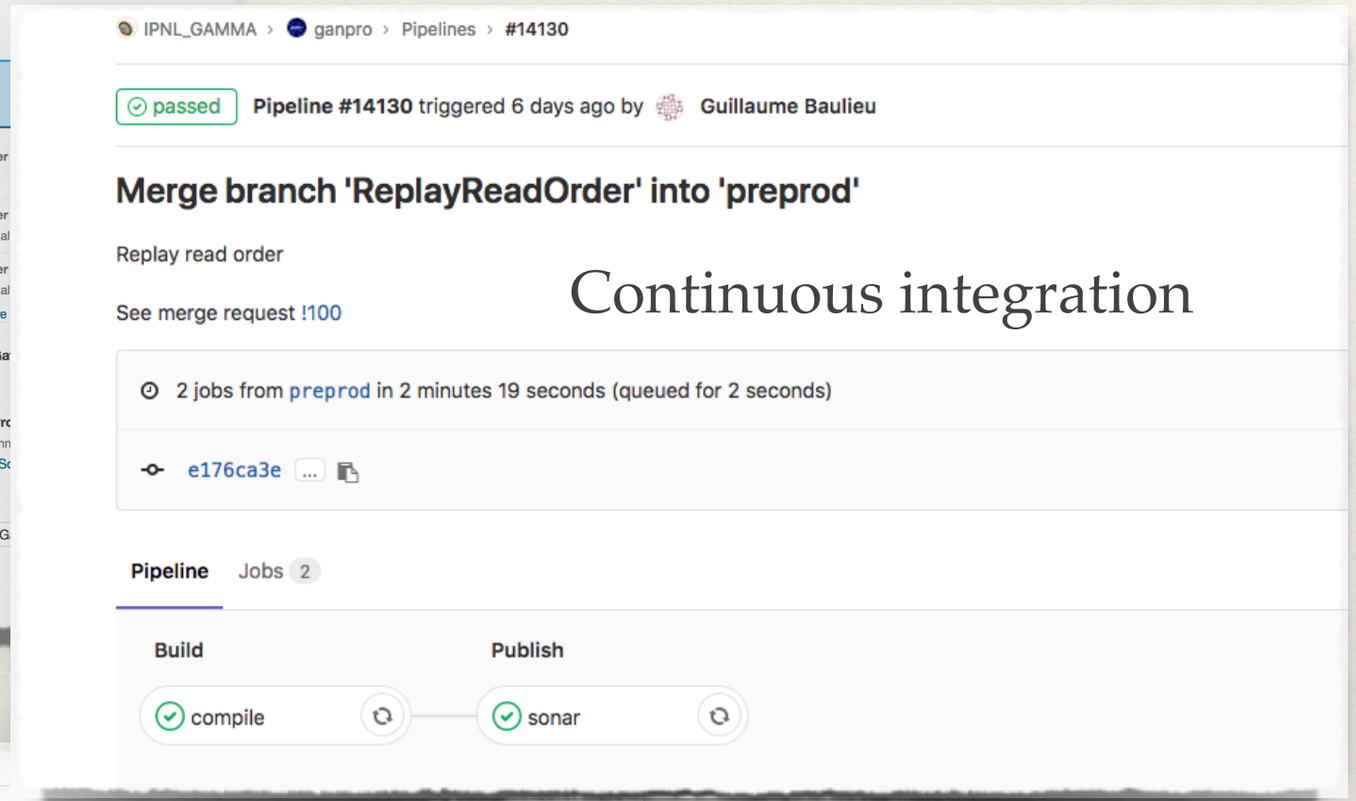
☛ they contains all the code compiled !

☛ they could be used also to distribute a full working environment

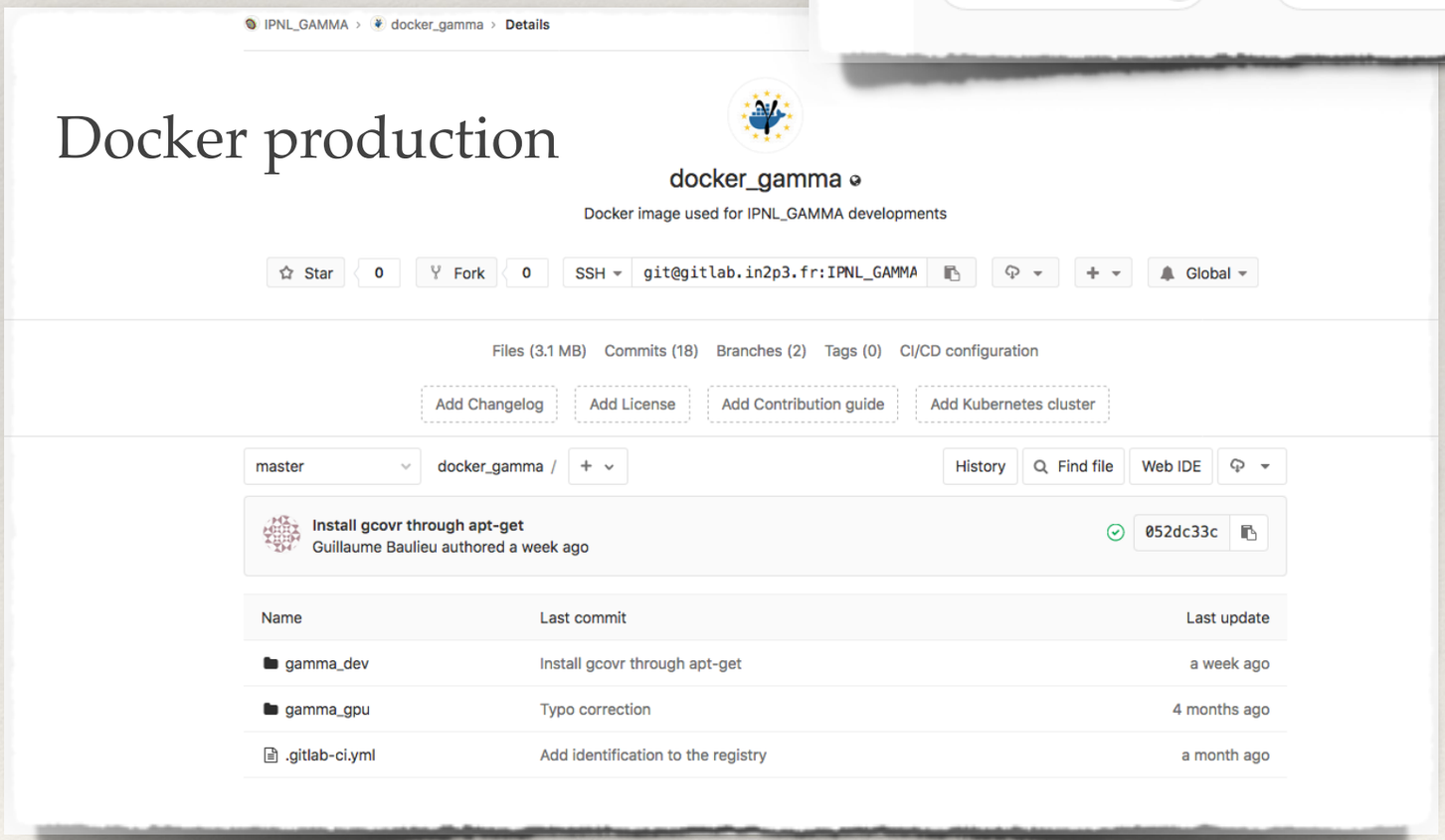
Quality check



Continuous integration



Docker production



Past / Present - Framework approach

What about processing GRETINA Data using our Framework ???

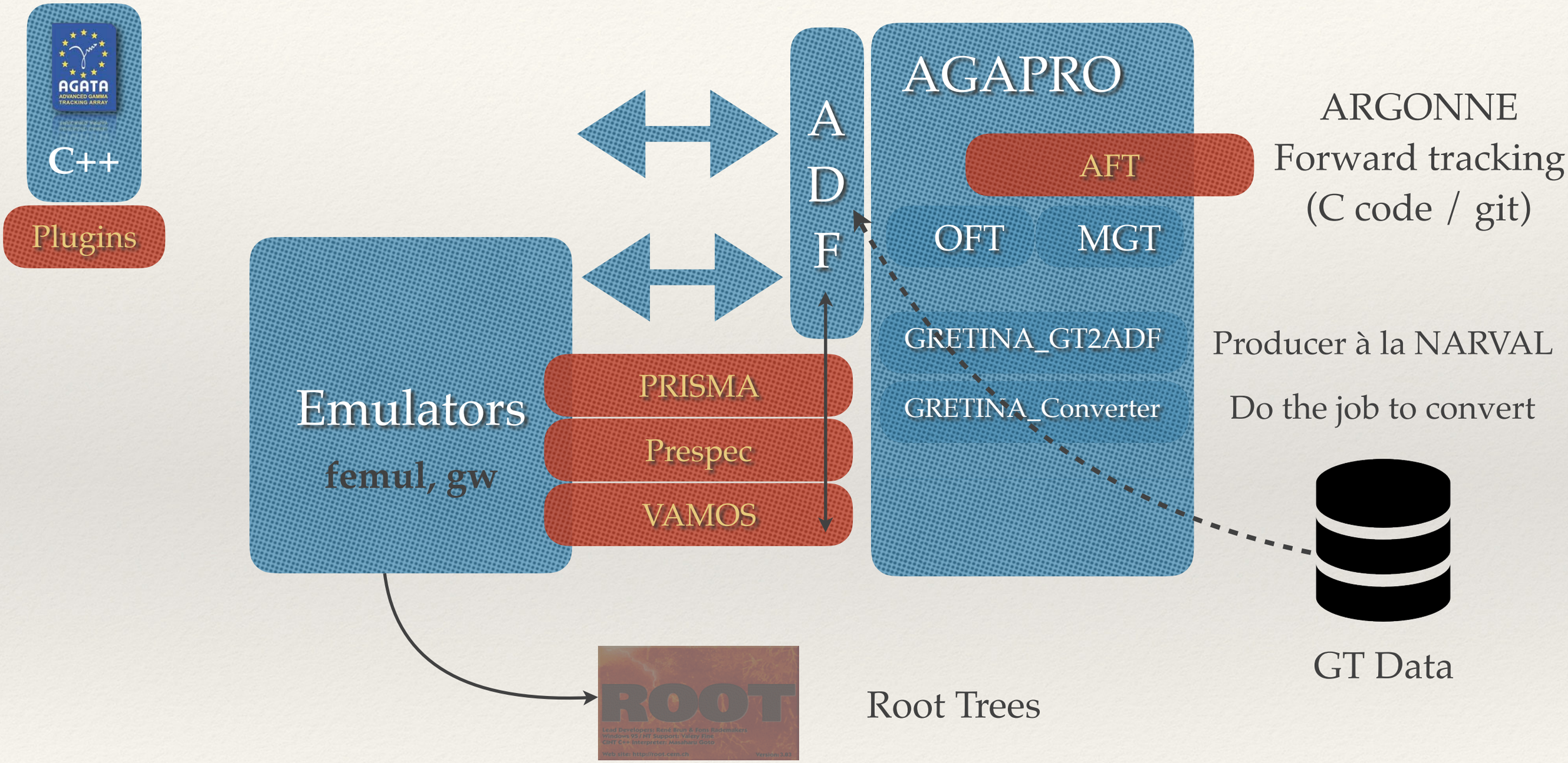
☛ it should not be that difficult, same 'kind' of crystals however 3 Triple 4 Quad

☛ But the data processing logic is not the same …

☛ slow process started, stopped, started again … etc

What is required first is to read GRETINA files and convert data into ADF Frames

Past / Present - Framework approach



Still some quality / quantity checks to be done on those developments before prod.

The most difficult task / time consuming remains to be done : PSA on GRETINA Data !

Past / Present - and Future ?

Is the AGATA Data Processing Framework robust for the future ?

Some thoughts to try and answer …

- C++ is moving … cxx11 … cxx14, cxx17 … cxx20
  ↳ Probably some pieces of code could be improved …

- ROOT is moving to ROOT7 … a lot of changes in the interface
  ↳ next run @ CERN, huge increase of data.
  ↳ current HEP models has to changed !

- Commercials drive the future !
  ↳ Ex : Tensorflow
  ↳ Amazon computing clouds, Internet of Things, Machine learning (IA)

- A world of containers … python is used a lot !

Present / Future

Machine Learning, at lot to learn !

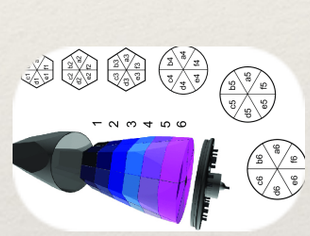
Our approach to learn machine learning NEDA ☛ AGATA

NEDA

1 Detector / 1 Signal

Pulse Shape Discrimination (γ/n)

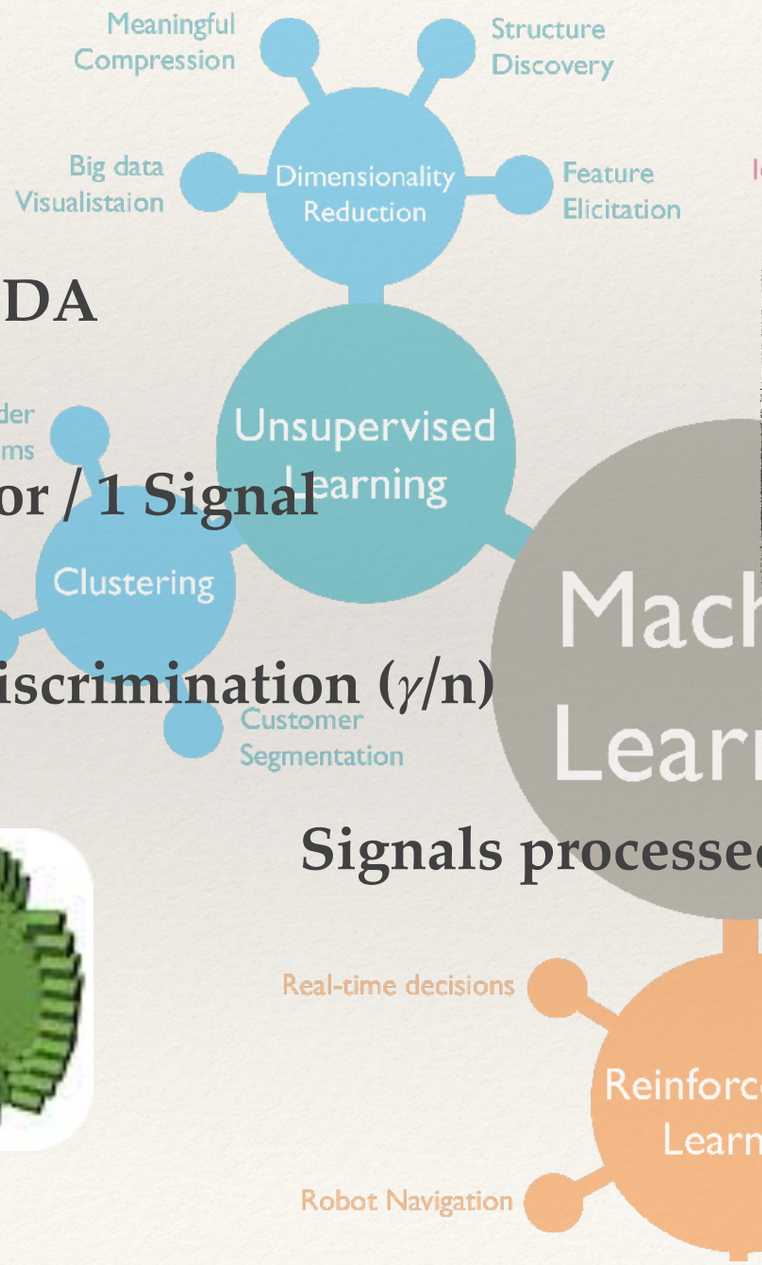
AGATA



1 Detector / 37 Signals

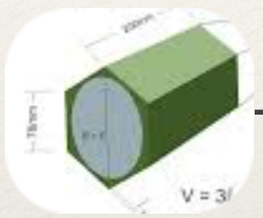
Pulse Shape Analysis (e,x,y,z)

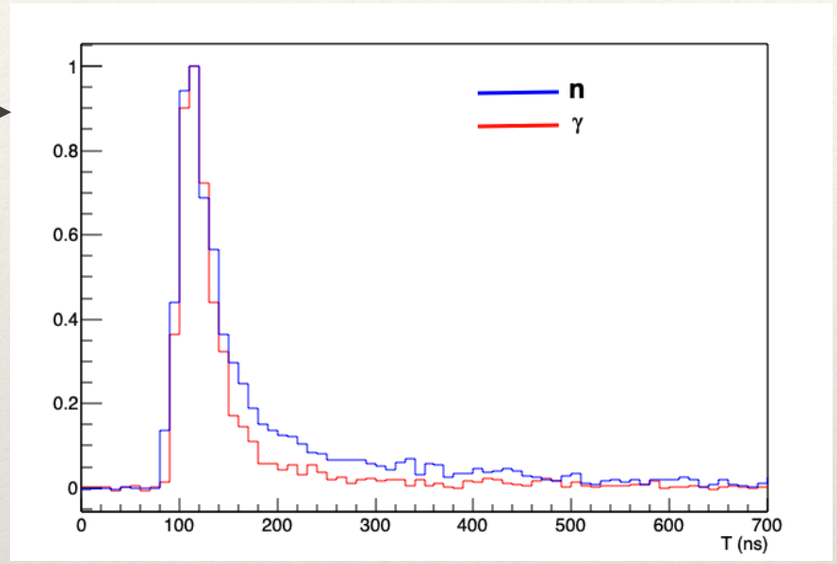
Signals processed online/offline



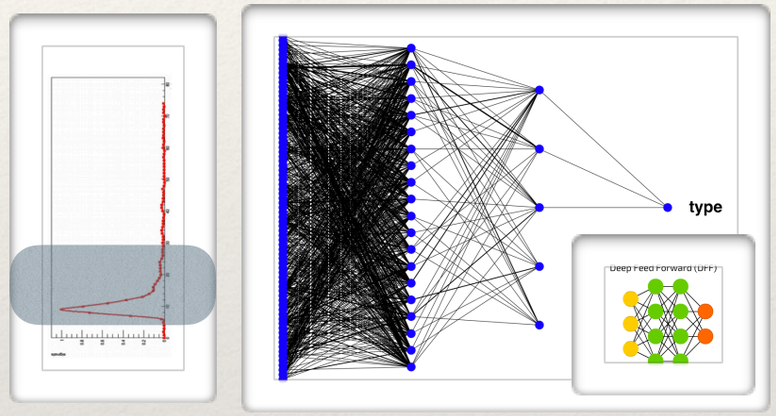
Present / Future

Pulse Shape Discrimination in NEDA





R&D NEDA, PSD with Neural network



Implementation with ROOT (monothread / CPU)
Best discrimination for low energy
Ronchi et al., A 610 (2009) 534–539

Signal parametrisation

$$s(t) = \mathbf{A}\,[\exp(-t/\mathbf{td1}) - \exp(-t/\mathbf{tr}) + \mathbf{R}^*(\exp(-t/\mathbf{td2}) - \exp(-t/\mathbf{tr})] \text{ si } t > \mathbf{T0}$$

A amplitude
td1, **td2**, **tr** ‘identical’ γ & **n**
T0 depend of signals alignments
R different between γ & **n**

Present / Future

Our first work has been to run NN PSD online / offline

GANPRO

PSD

- **We have moved from ROOT to Tensorflow/keras (python / C++)**
  Python interface for training, C++ interface for inference
  The library deals with hardware, transparent to users (multi-core/CPU, GPU)
  Facteur 50 gained [on CPU] , online inférence !

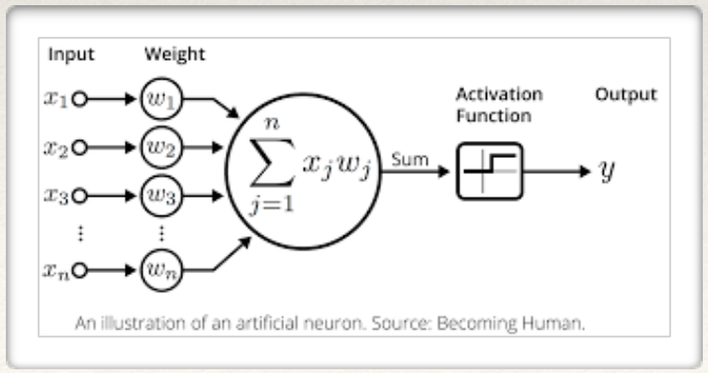
Tensorflow

TDC is an input of the network

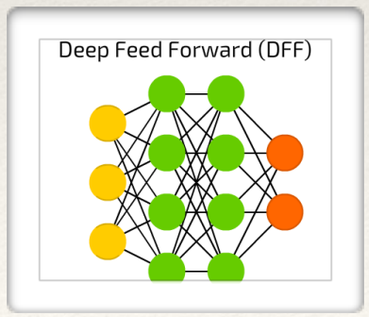
We have decided to study other NN architectures

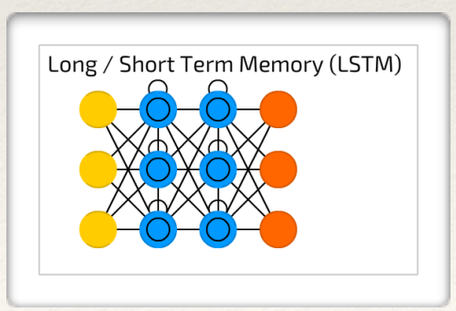
Three types of networks has been compared :

Multi Layer Perceptron (MLP), Long Short Term Memory (LSTM), Convolutional Neural Network (CNN)

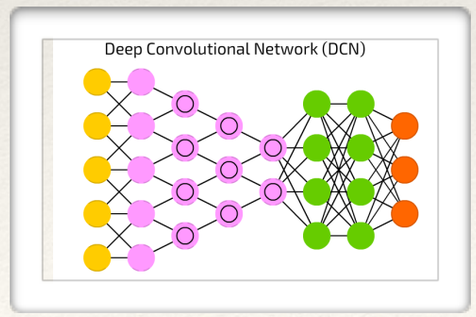


One neuron





👍 **time series**



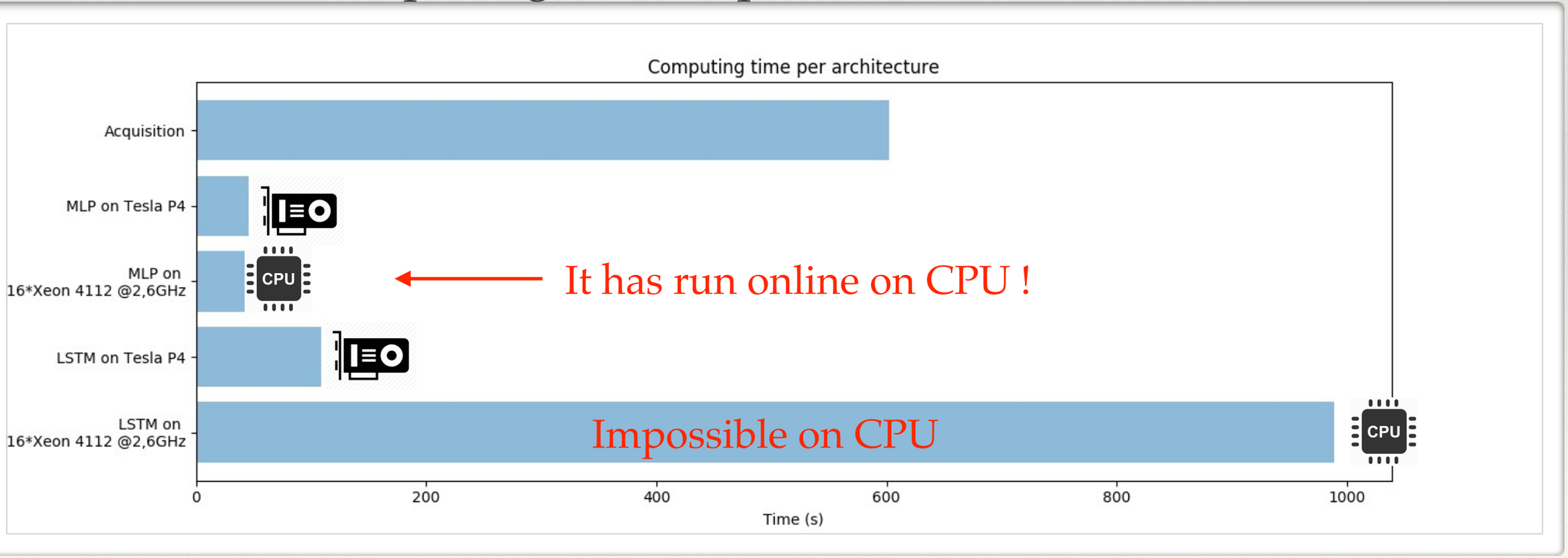
👍 **pattern**

Present / Future

Network configurations

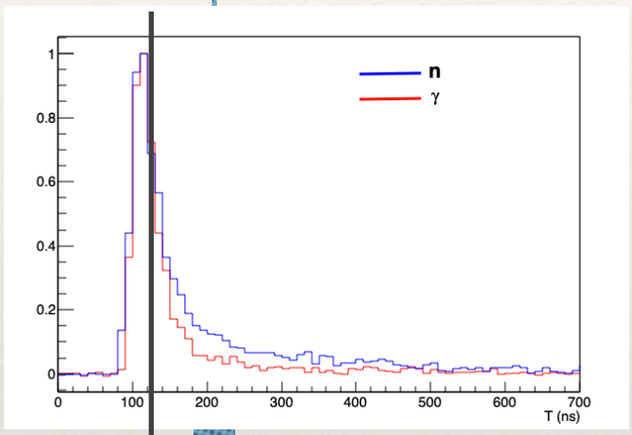
Network type	Structure	Activation functions	Number of trainable parameters
MLP	3 Dense layers (75 x 10 x 4 x 2)	Relu x ReLu x SoftMax	814
LSTM	75 x 1 LSTM layer (50 hidden units) x 1 Dense layer (50 x 2)	SoftMax	10 502
Convolution	75 x 3 (Conv1D+Max_Pooling) layers x 2 Dense layers (100 x 20 x 2)	ReLu x ReLu x ReLu x ReLu x SoftMax	7 042

Computing time required for inference

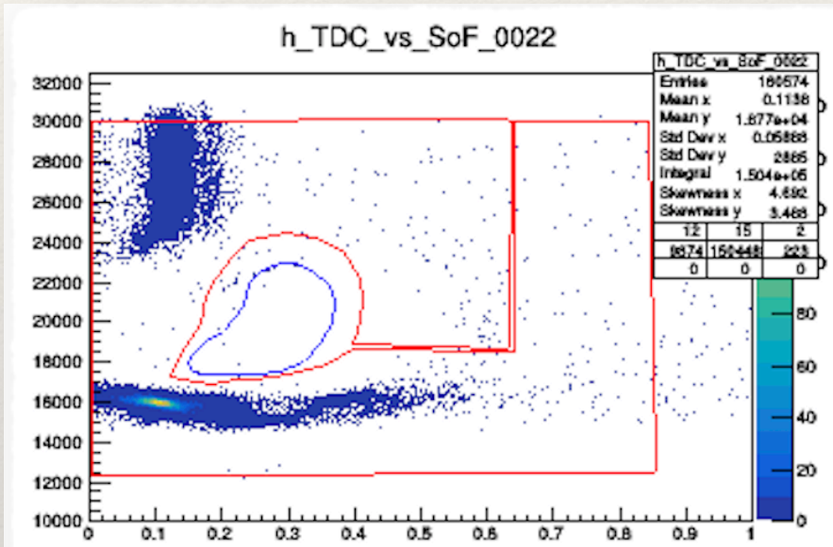


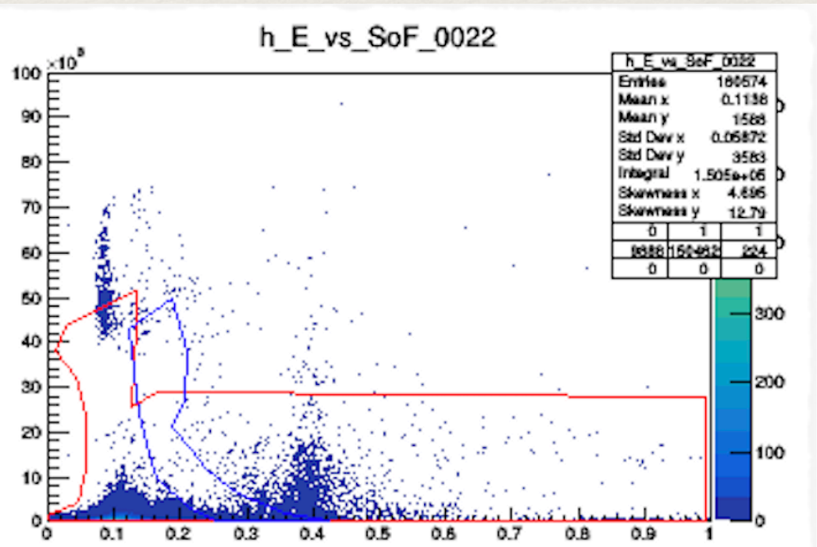
Present / Future

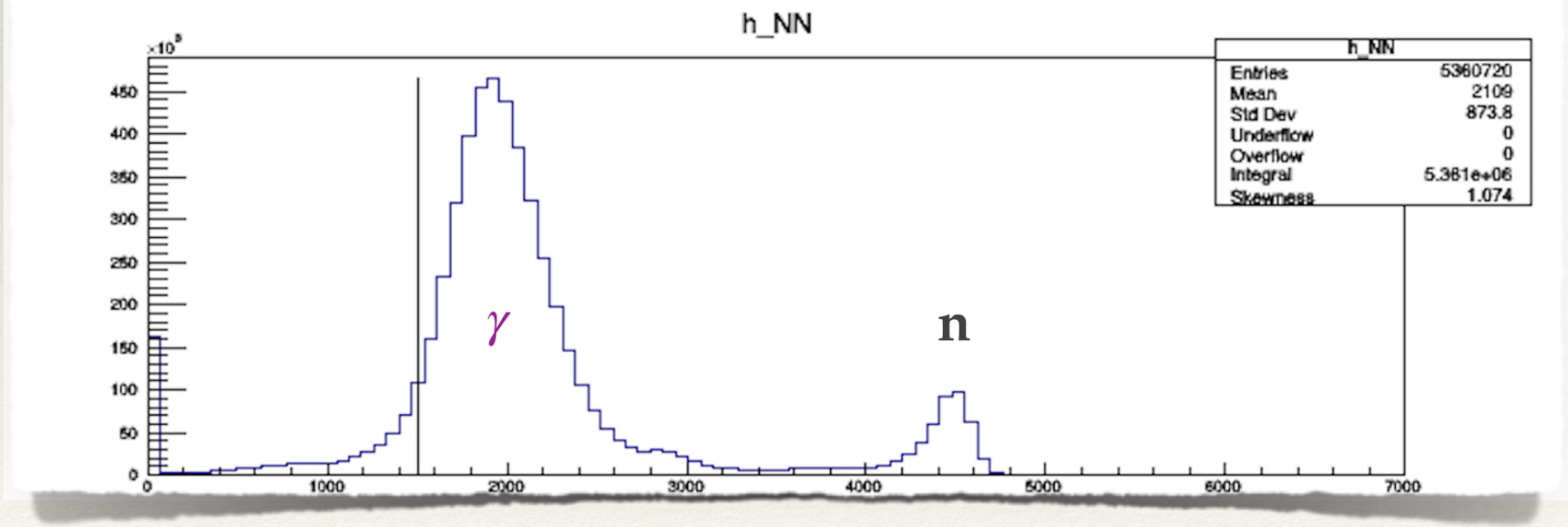
Training of the networks using 2 2D cuts on SoF/TDC, A/SoF



Training done using the python of Tensorflow



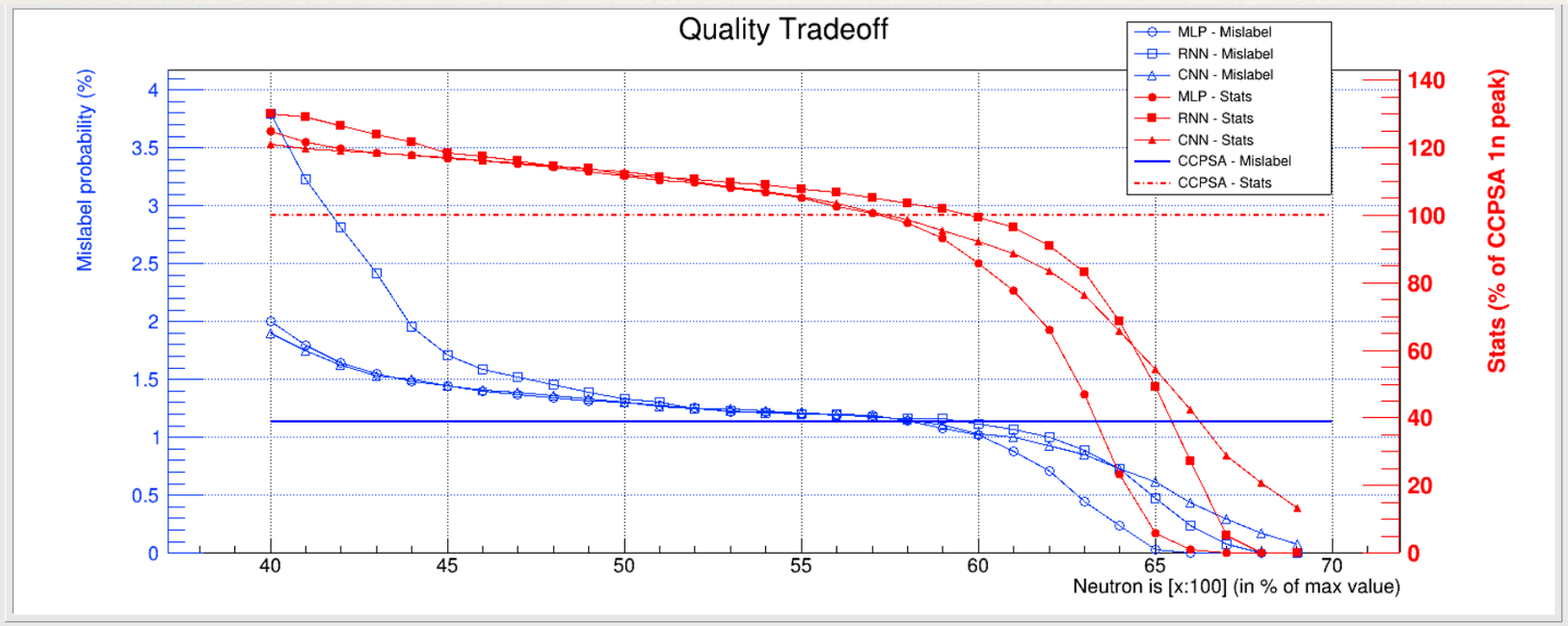




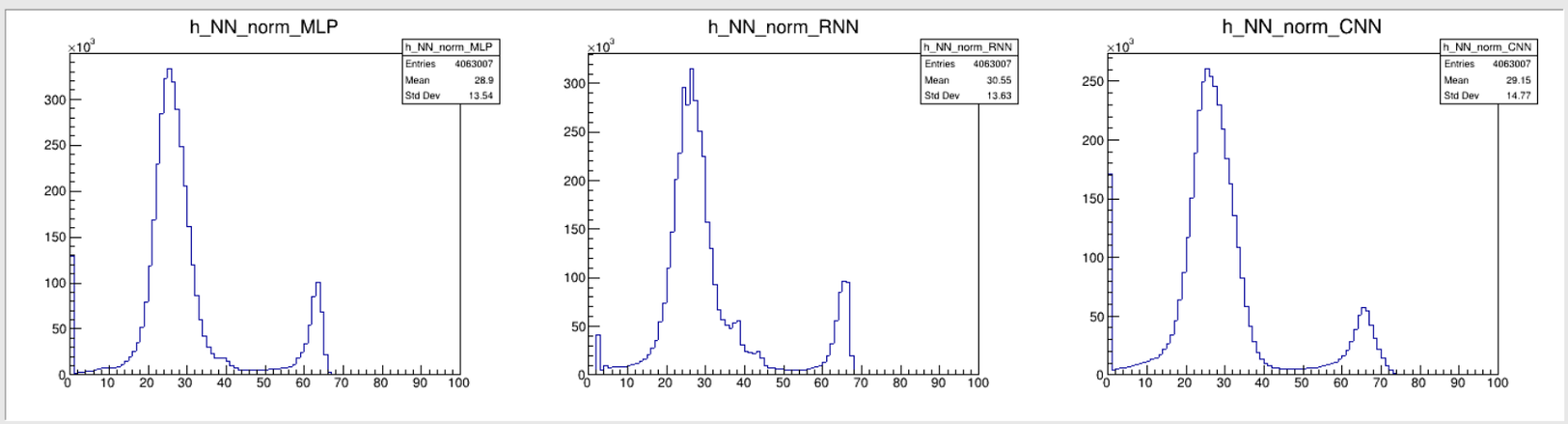
Present / Future

We have AGATA/NEDA/DIAMANT Data,

↳ AGATA γ spectra to evaluate wrong n discriminations in NEDA



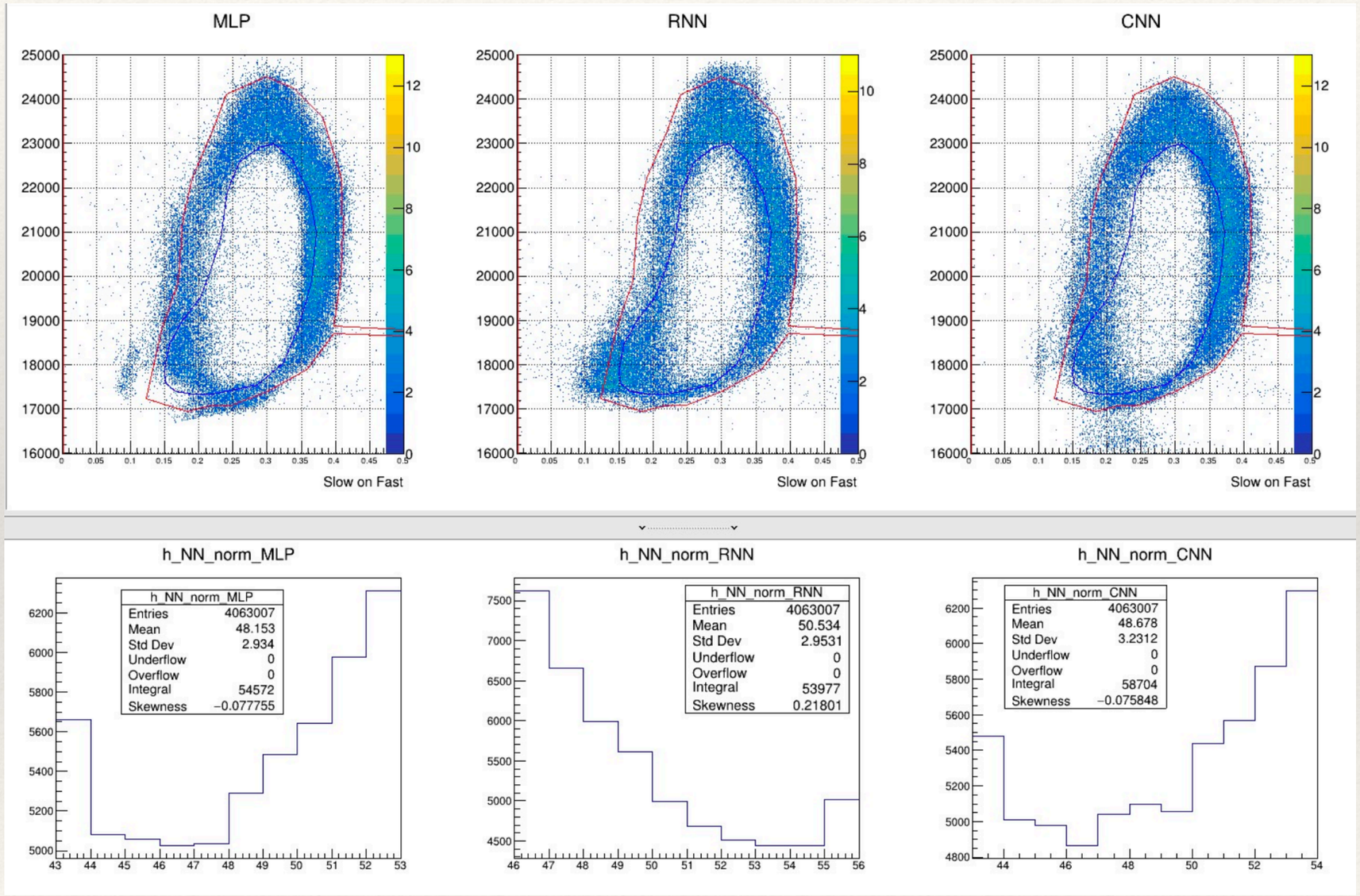
Mislabel probability



Neural Network answer to is γ or **n** ?

Present / Future

How networks extrapolate on data out of the cuts used for training ?



☛ We are working on the qualification of those sub-events using γ spectra

Present / Future

We have moved to simulations to check for strengths/weaknesses of the different NN

↪ labels on γ or **n** are 100 % sure !

Function used to generate signals

s(t) = **A** [exp(-t/**td1**) - exp(-t/**tr**) + **R***(exp(-t/**td2**) - exp(-t/**tr**)] si t > **T0**

Study 1 : sensibility to **T0**

Training done with gaussian distribution for T0, $\sigma = 2$

Test done with gaussian distribution for T0, $\sigma = 20$

Study 2 : using NN to tag pileup signals

ΔT between two signals, random distribution

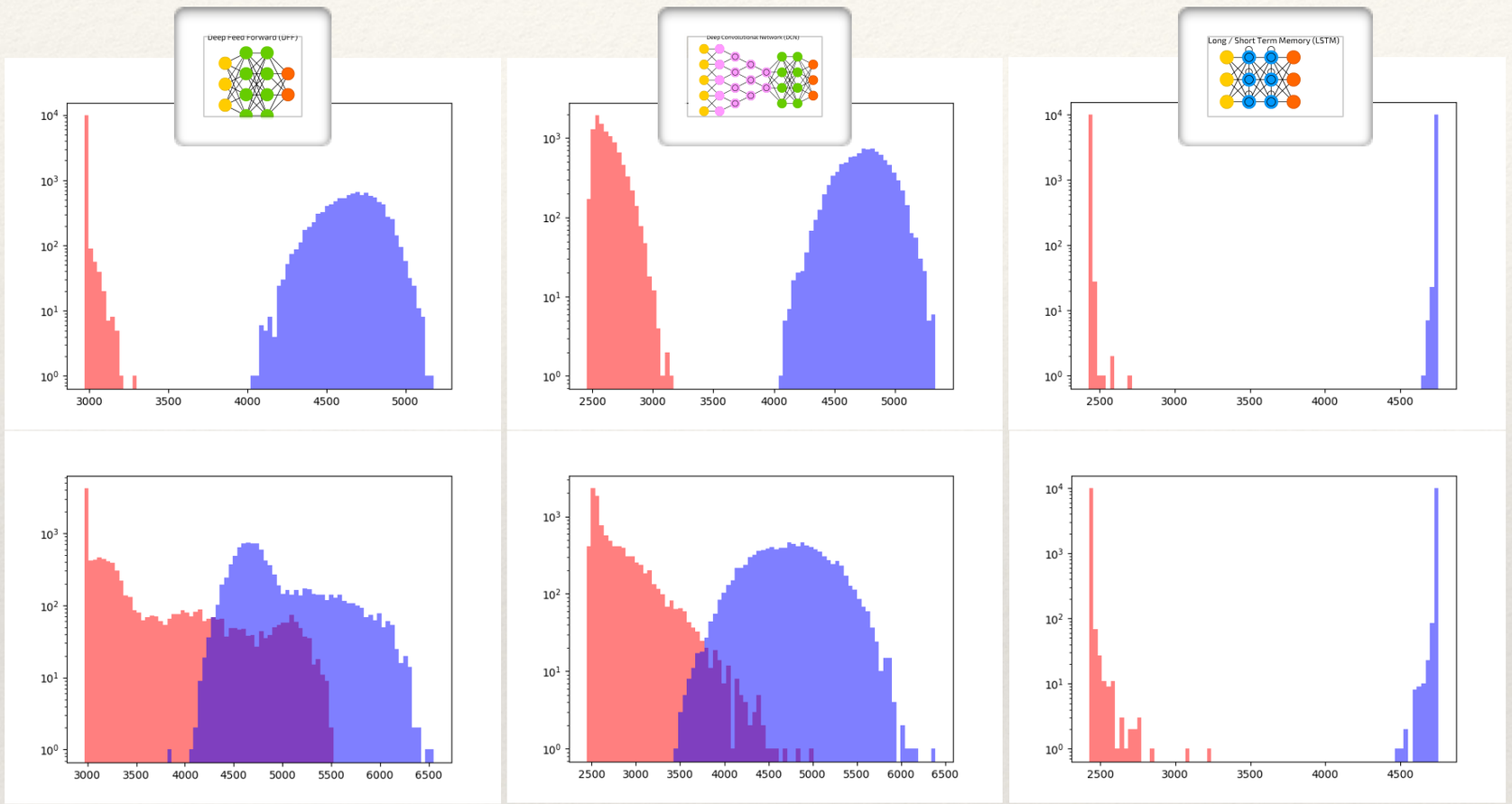
Almost same networks, just more categories, more outputs

Network type	Structure	Activation functions	Number of trainable parameters
MLP	3 Dense layers (75 x 10 x 4 x 2)	Relu x ReLu x SoftMax	814
LSTM	75 x 1 LSTM layer (50 hidden units) x 1 Dense layer (50 x 2)	SoftMax	10 502
Convolution	75 x 3 (Conv1D+Max_Pooling) layers x 2 Dense layers (100 x 20 x 2)	ReLu x ReLu x ReLu x ReLu x SoftMax	7 042

Present / Future

Study 1 : sensibility to **T0**

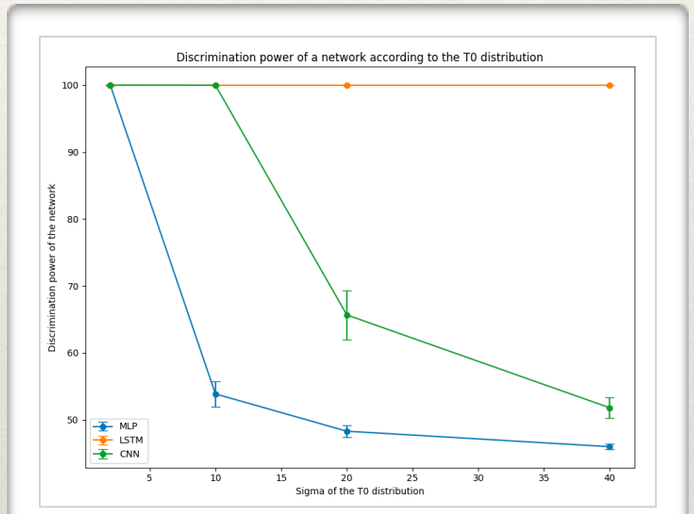
$\sigma = 2$

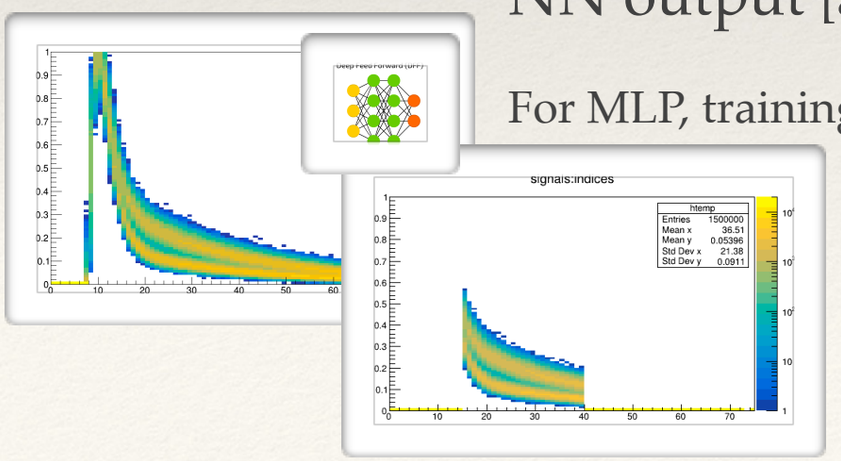


$\sigma = 20$

NN output [as coded in the data flow]

LSTM the most robuste !





For MLP, training with full signals, test with partial signals

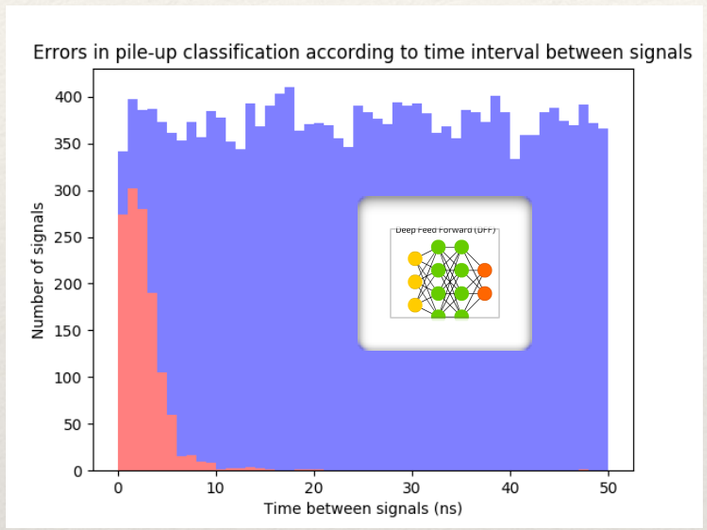
- ☛ the network works fine with partial signals !
- ☛ Important to have well 'calibrated'* signals for that kind of NN

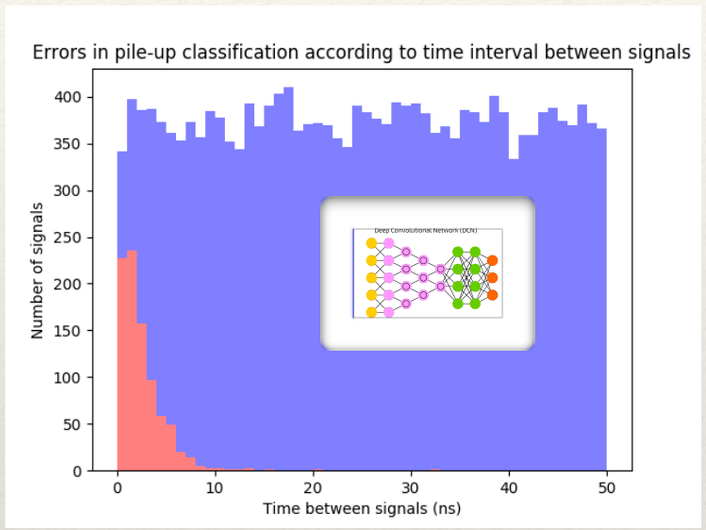
* Feature extraction in machine learning language

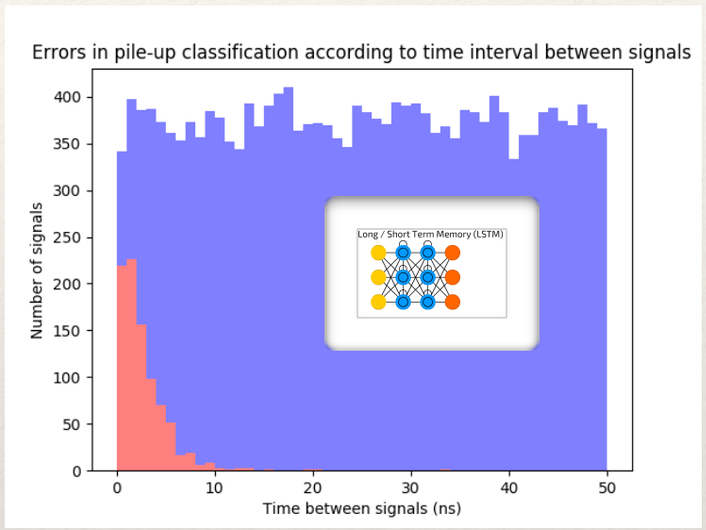
Present / Future

Study 2: Pileup identification

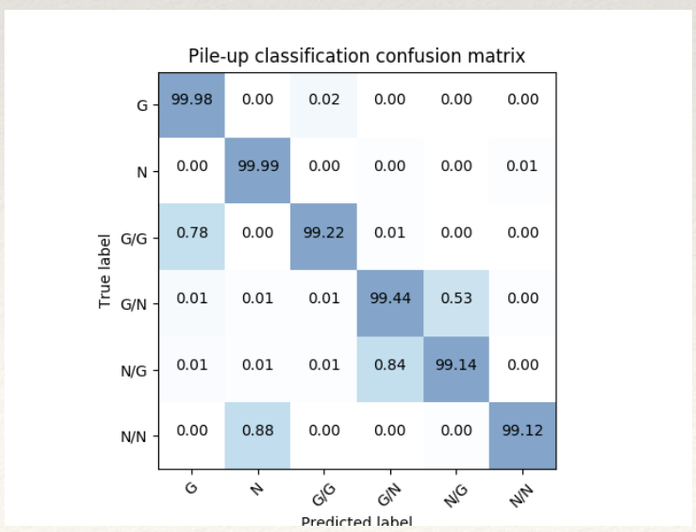
Error as fonction of the time between signals

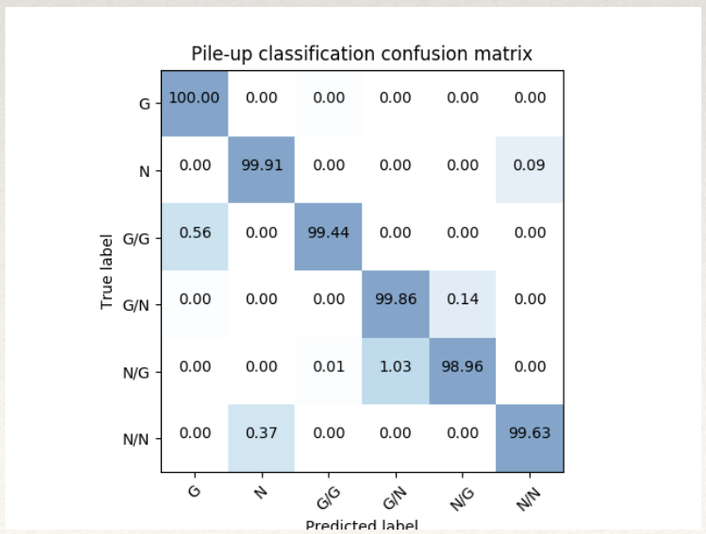


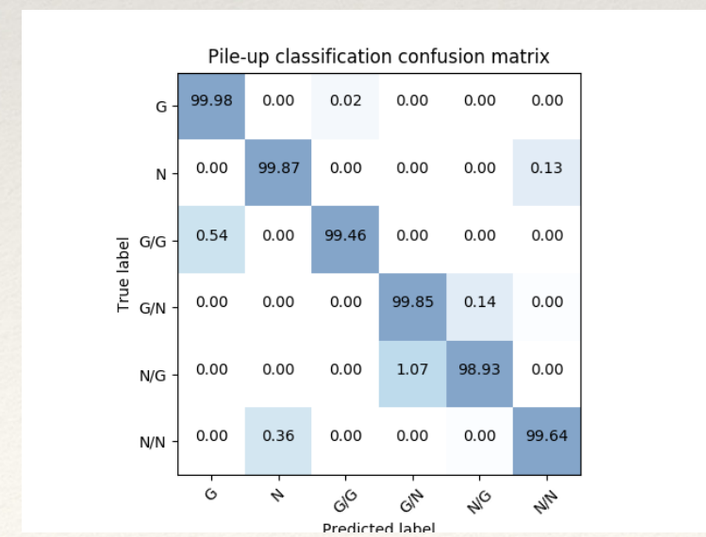




Confusion matrix



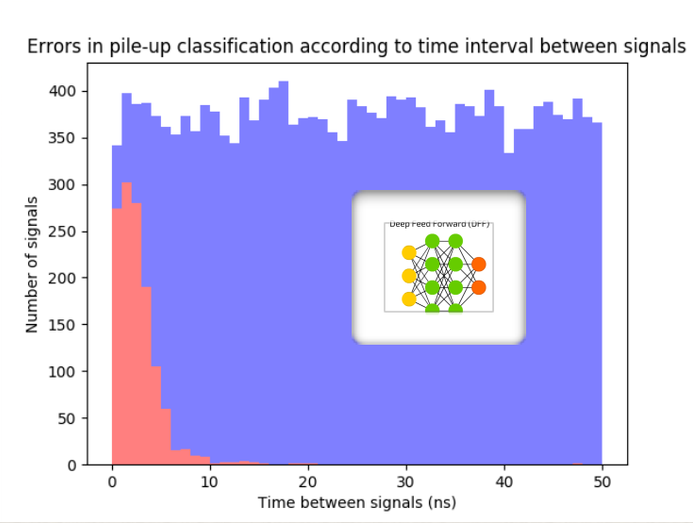


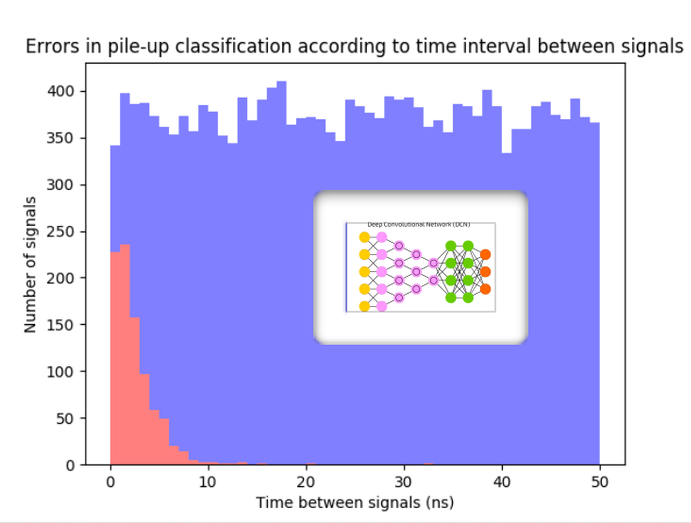


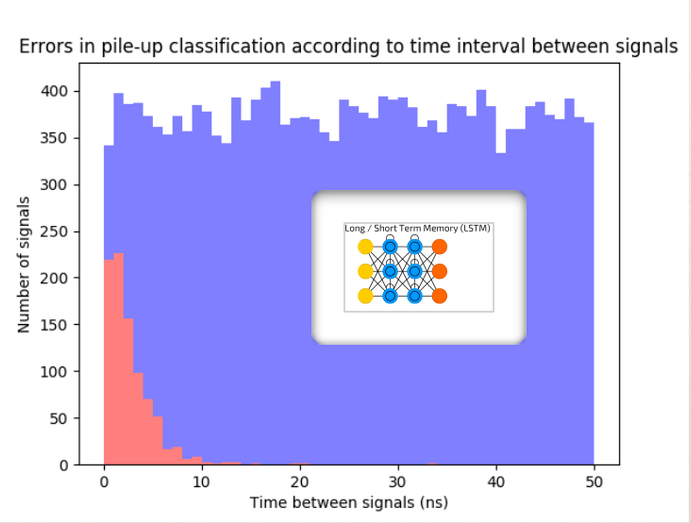
Present / Future

Study 2: Pileup identification

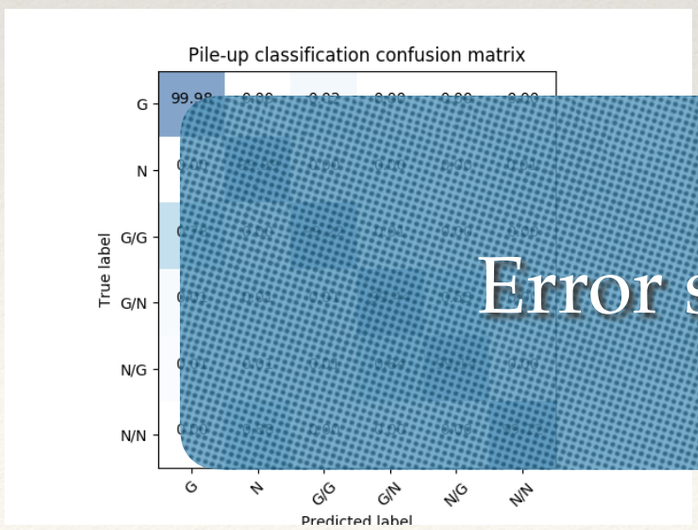
Error as fonction of the time between signals

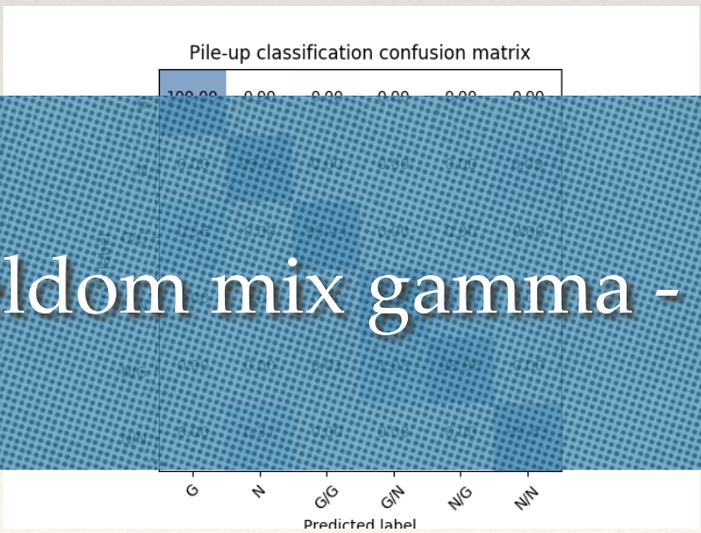


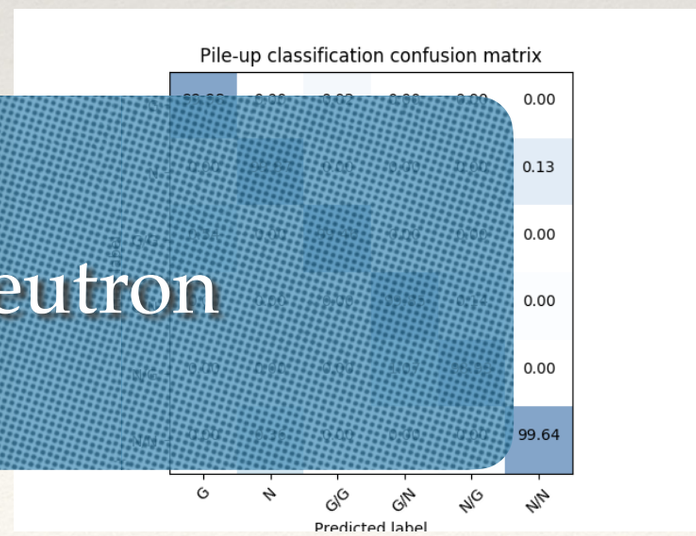




Confusion matrix

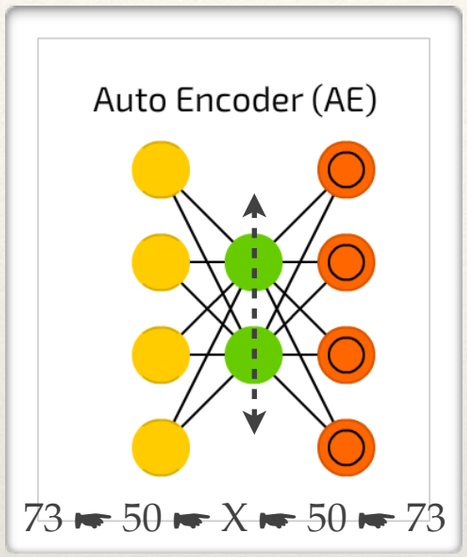




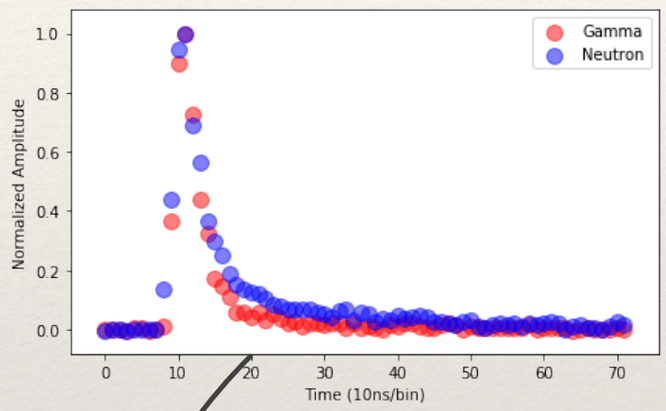


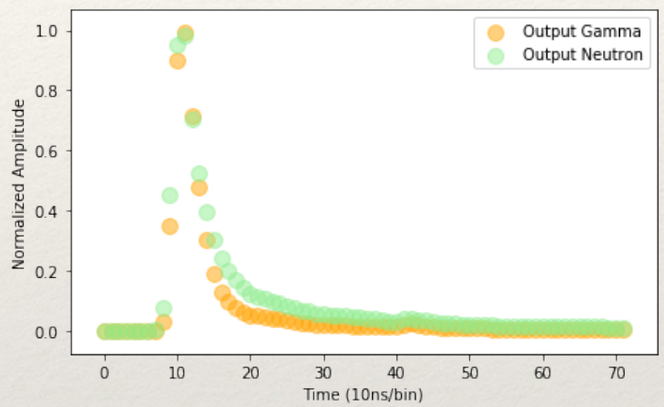
Error seldom mix gamma - neutron

Present / Future



Auto-encoder, unsupervised learning ☛ self learning !
➥ avoid problems of having labelled training data …



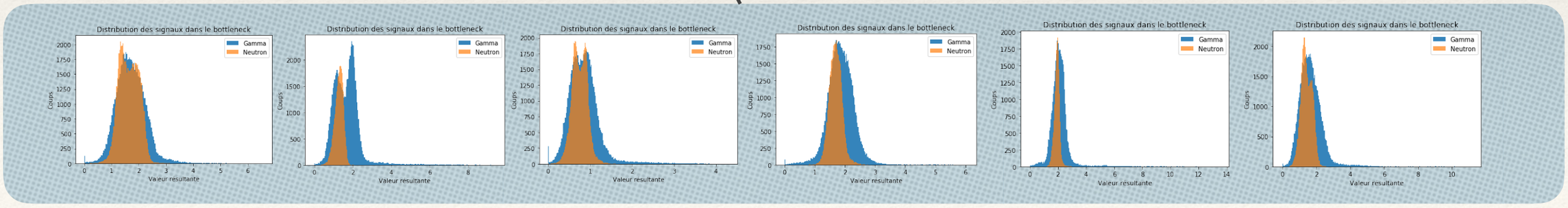


Objectives :
Denoising
Data reduction !

We found @ least 4 neurones needed in the bottleneck,
Is it link to
$s(t) = \mathbf{A}\,[\exp(-t/\mathbf{td1}) - \exp(-t/\mathbf{tr}) + \mathbf{R}^*(\exp(-t/\mathbf{td2}) - \exp(-t/\mathbf{tr})]$?

Identification of anomalies ?
Simulation of signals ?

Distribution for 6 central neurons



Conclusions / perspectives

The Data Processing / Analysis Framework is **set** almost **from the beginning**
It has **grown** and **seen several ('minor') migrations**. It has :

- ☛ been used @ three different centers
- ☛ been more controlled using **continuous integration** processes
- ☛ followed ROOT evolution … What about ROOT7 ?
- ☛ Moved from svn to **git**
- ☛ Been used with several third party libraries
  - ➥ PRISMA … VAMOS … AFT … Tensorflow

GRETINA Data could be **processed** through AGATA Processing
  - ➥ should help future developments

New challenges are there :

- ☛ More and more detectors in the array !
- ☛ **Machine Learning** technology [python heavily used !]
- ☛ heterogeneous architectures, containerised applications …

The Data Processing/Analysis Framework is likely to go through 'major' changes

Conclusions / perspectives

The Data Processing / Analysis Framework is **set** almost **from the beginning**
It has **grown** and **seen several ('minor') migrations**. It has :

- ☛ been used @ three different centers
- ☛ been more controlled using **continuous integration** pro
- ☛ followed ROOT evolution … What about ROOT7
- ☛ Moved from svn to **git**
- ☛ Been used with several third party lib
  - ➥ PRISMA … VAMOS … borflow

GRETINA Data could be **process** DATA Processing
  - ➥ should he lopments

New challenges
- ☛ More and more detectors in the array !
- ☛ **Machine Learning** technology [python heavily used !]
- ☛ heterogeneous architectures, containerised applications …

Many thanks to all the people involved !!!

The Data Processing/Analysis Framework is likely to go through 'major' changes