Machine Learning for Boosted Jet Classification in High Energy Physics

J. C. RUIZ VARGAS, RAPHAEL COBE, S. STANZANI

CENTER OF SCIENTIFIC COMPUTING – SAO PAULO STATE UNIVERSITY

Intel – Unesp CoE for Machine Learning

Objective

- Establish a Center of Excellence in ML
- Tackle challenging projects related to ML

Activities

- R&D, consulting services
  - Industry and academia
- Training sessions in Data Science and ML

Partners

- São Paulo Research and Analysis Center
  - www.sprace.org.br

Outline

Why High Energy Physics?

- From Atoms to Quarks
- Quantum Chromodynamics
- The CERN's Large Hadron Collider

Machine Learning

- General Strategy
- Convolutional Neural Nets

Boosted Jet Classification

- Data Simulation
- Preprocessing of Jet images
- HPC nodes at NCC-Unesp
- Performance on Intel® Xeon Phi™
- Results and Outlook

Acknowledgments

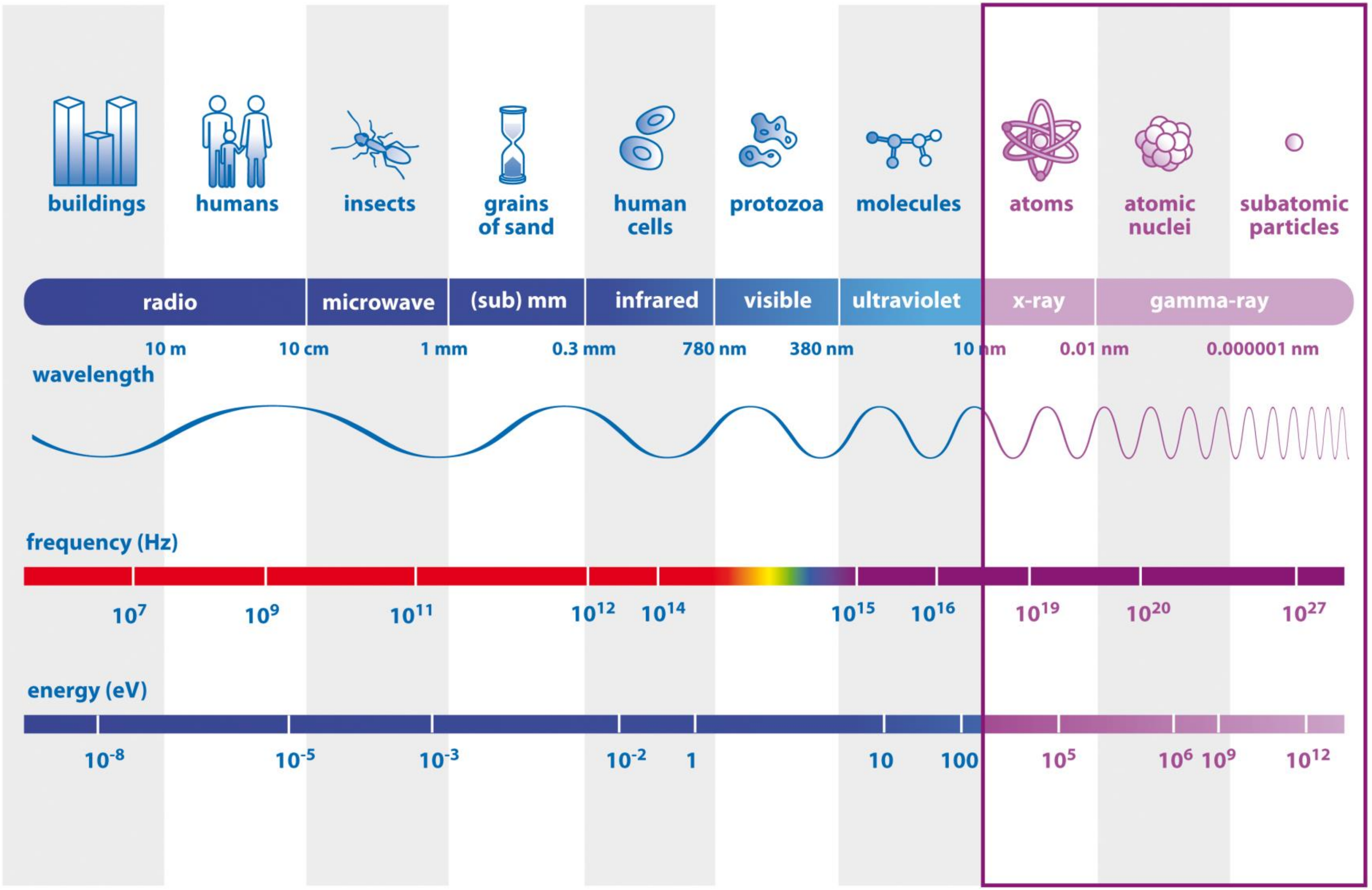
Why High Energy Physics?

According to Quantum Mechanics, subatomic particles behave like waves.

The higher the energy of the particle, the smaller the length probed by the particle's wave.

Energy units: electron volt (eV)

$1\ \text{MeV} = 10^6\ \text{eV}$
$1\ \text{GeV} = 10^9\ \text{eV}$
$1\ \text{TeV} = 10^{12}\ \text{eV}$

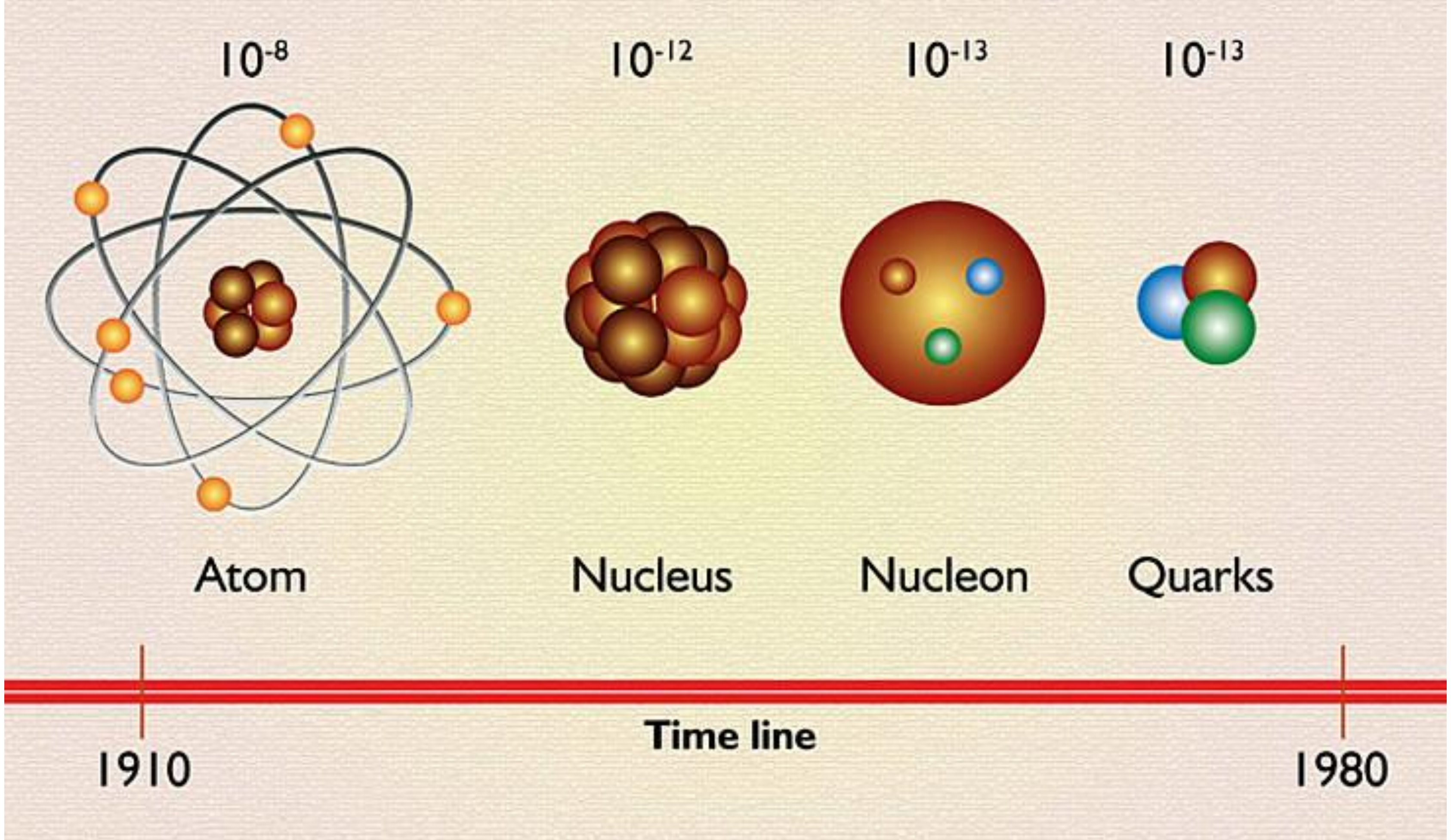


From Atoms to Quarks

Atoms consist of a nucleus and electrons surround it.

Quarks are the fundamental constituents of nucleons.

Protons are made out of three quarks.

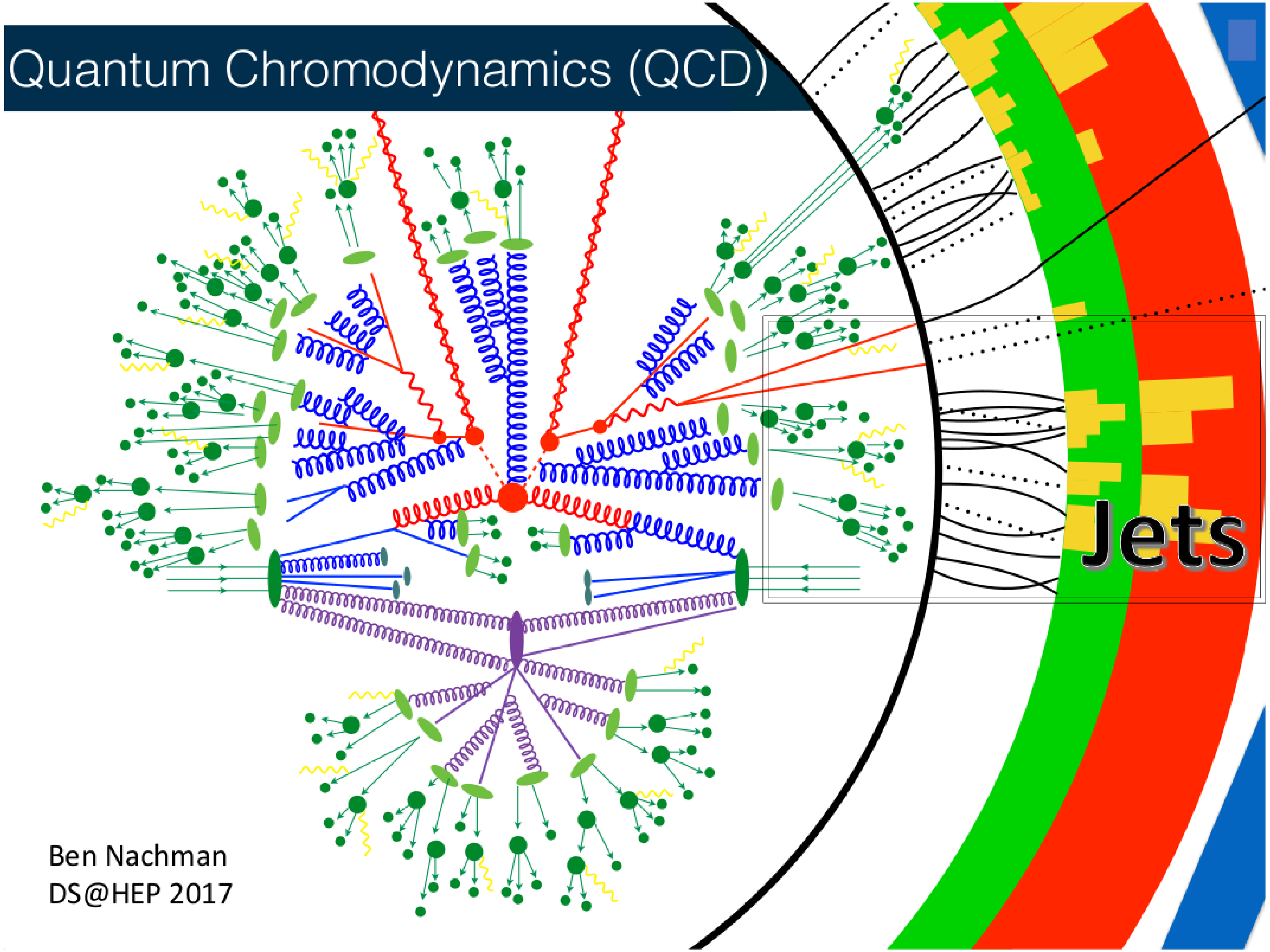


Quantum Chromo-Dynamics

Theory to describe interactions between quarks.

The experimental signature of a quark is called a “Jet”.

The adjective “boosted” means high energy in the system of reference of the laboratory.



Boosted Jet Classification

Signal (s)

- High energy jets coming from W/Z processes

Background (b)

- Similar jets coming from QCD processes

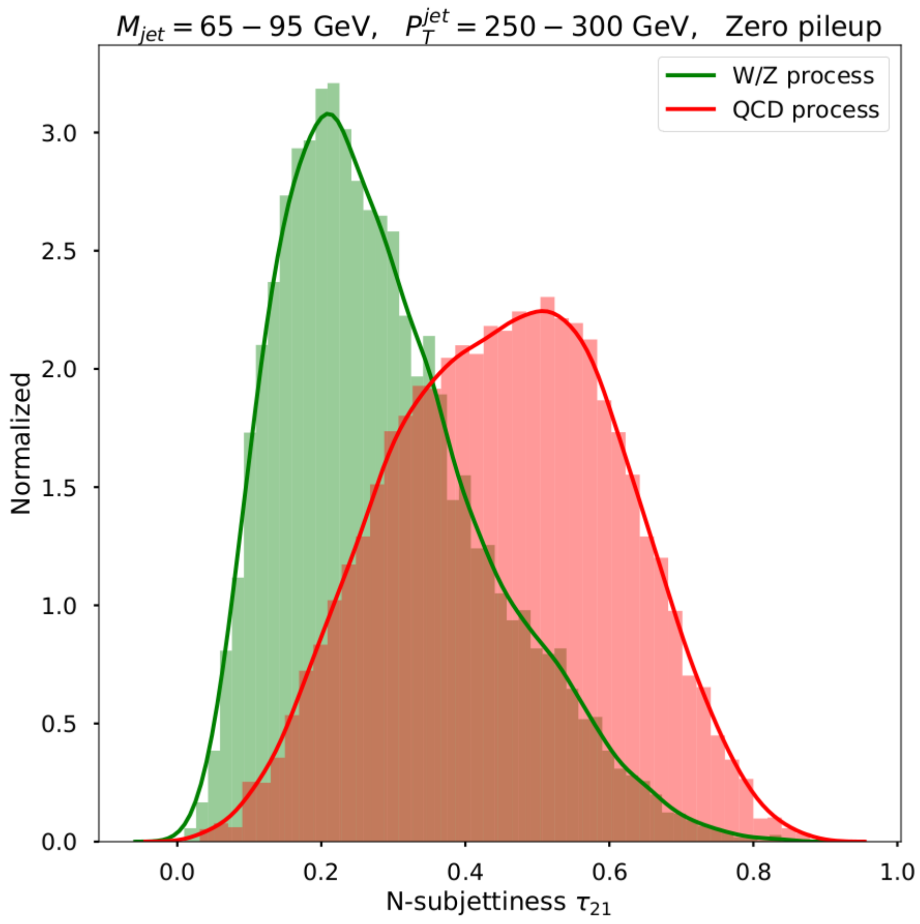
Description of the problem

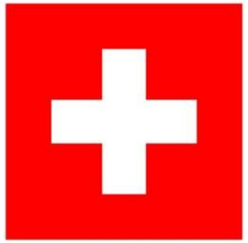
- Train a classifier $g : \mathbb{R}^d \rightarrow \{\mathrm{b}, \mathrm{s}\}$ on data

$$\mathcal{D} = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)\}$$

$\mathbf{x}_i$ is a *d*-dimensional training example

y_i is the target label



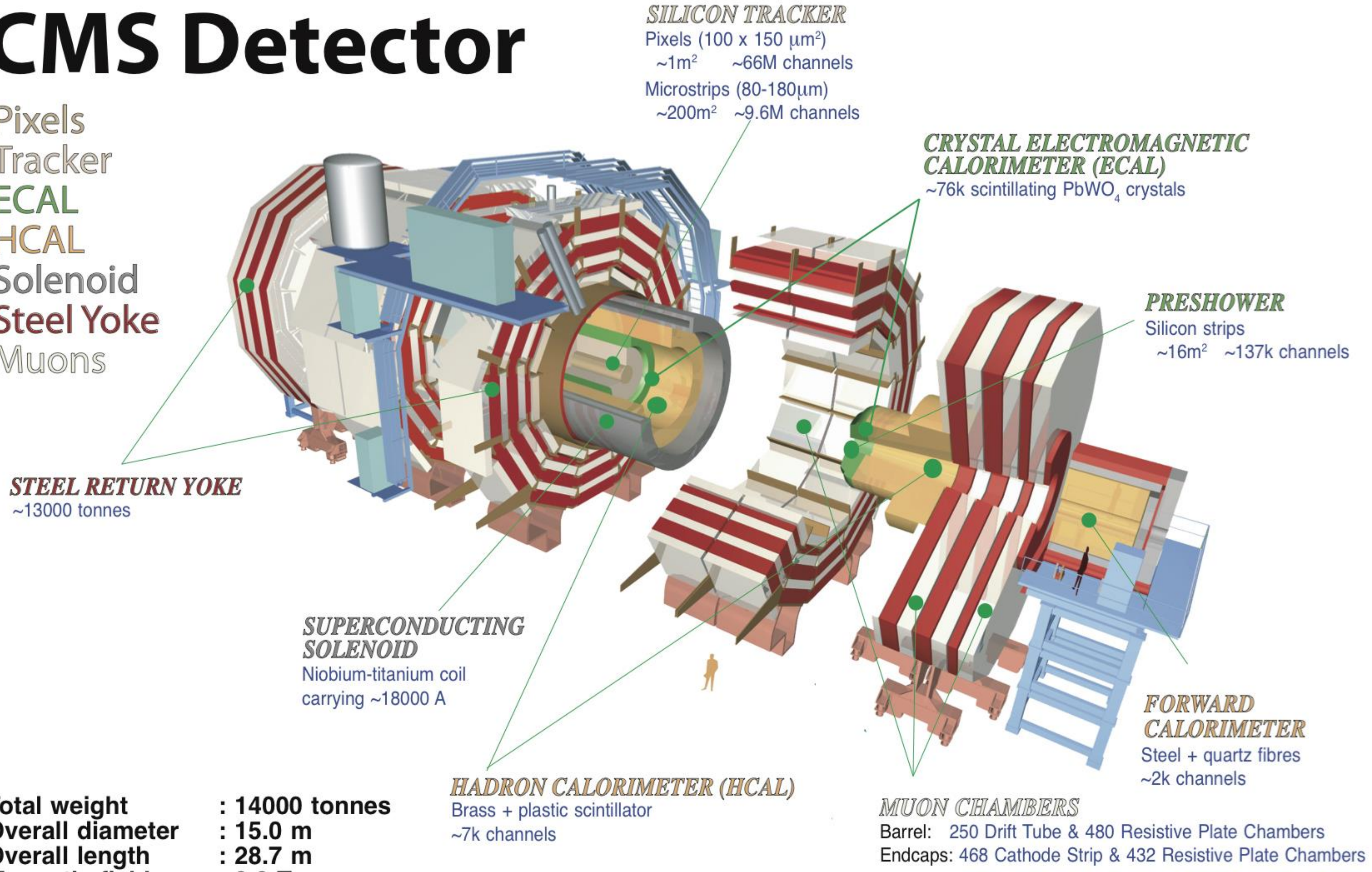


SWITZERLAND

FRANCE

CMS Detector

Pixels
Tracker
ECAL
HCAL
Solenoid
Steel Yoke
Muons



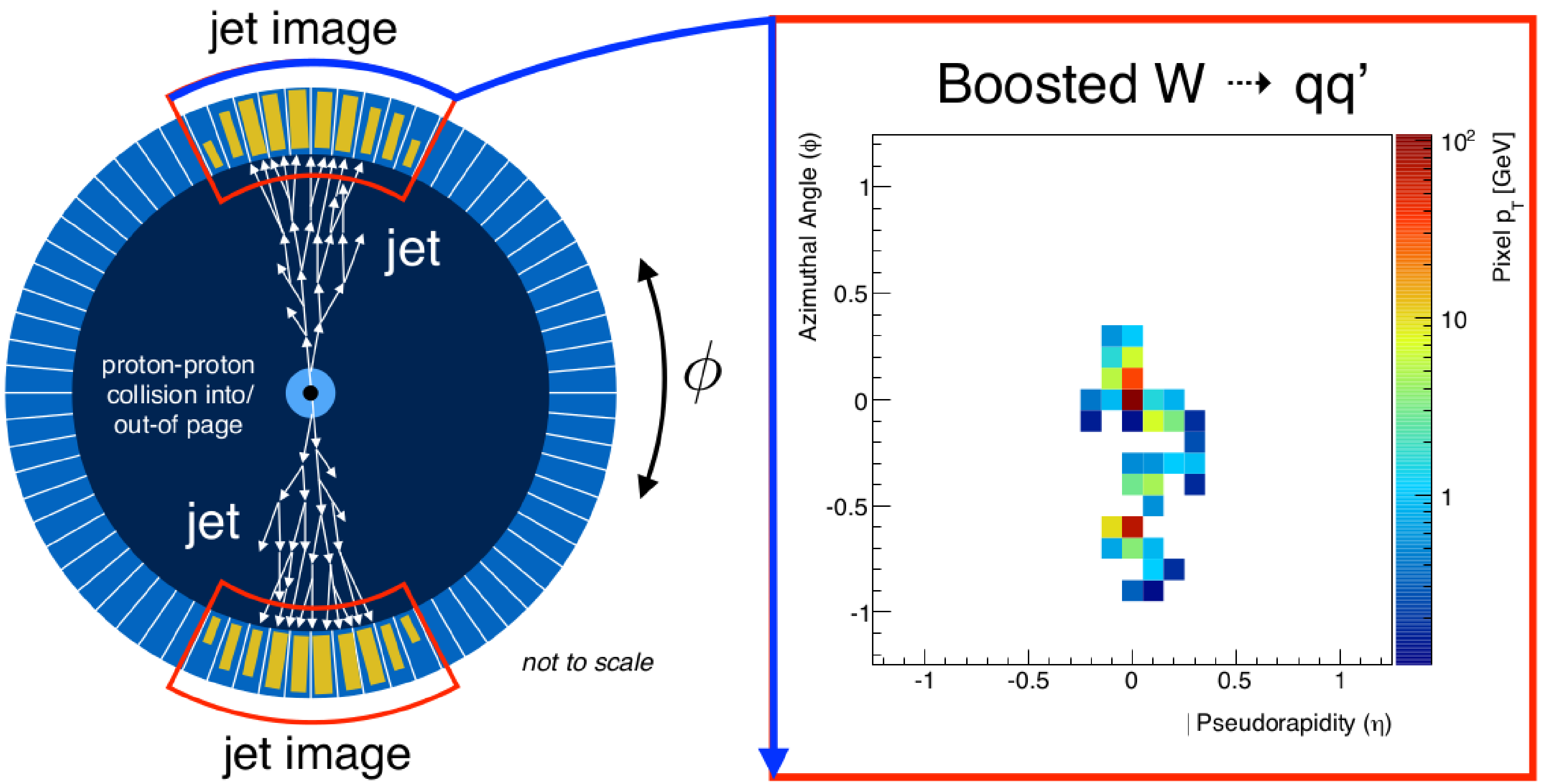
Total weight	**: 14000 tonnes**
Overall diameter	**: 15.0 m**
Overall length	**: 28.7 m**
Magnetic field	**: 3.8 T**

Data Simulation and Preprocessing of Jet Images

Pythia 8 event generator: simulates proton-proton collisions with the same conditions of the LHC.

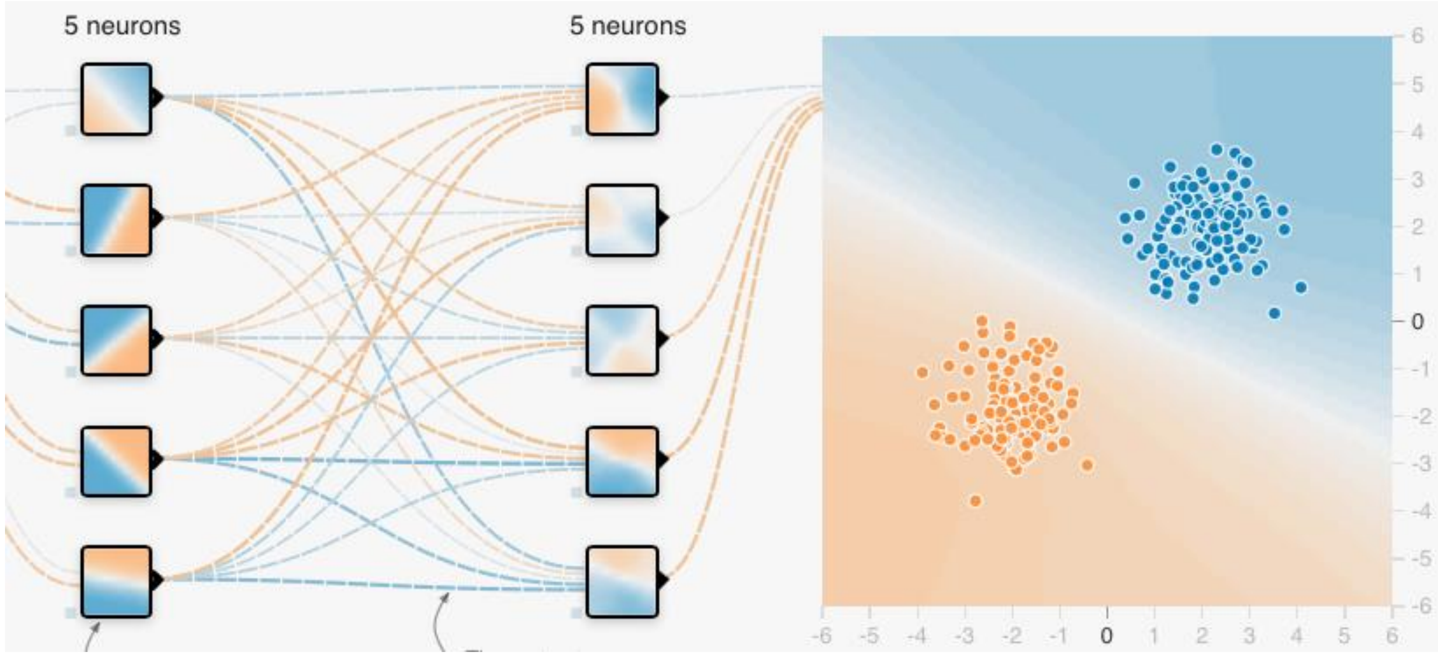
FastJet library for jet clusterization.
http://fastjet.fr

ROOT data analysis framework.
http://root.cern.ch

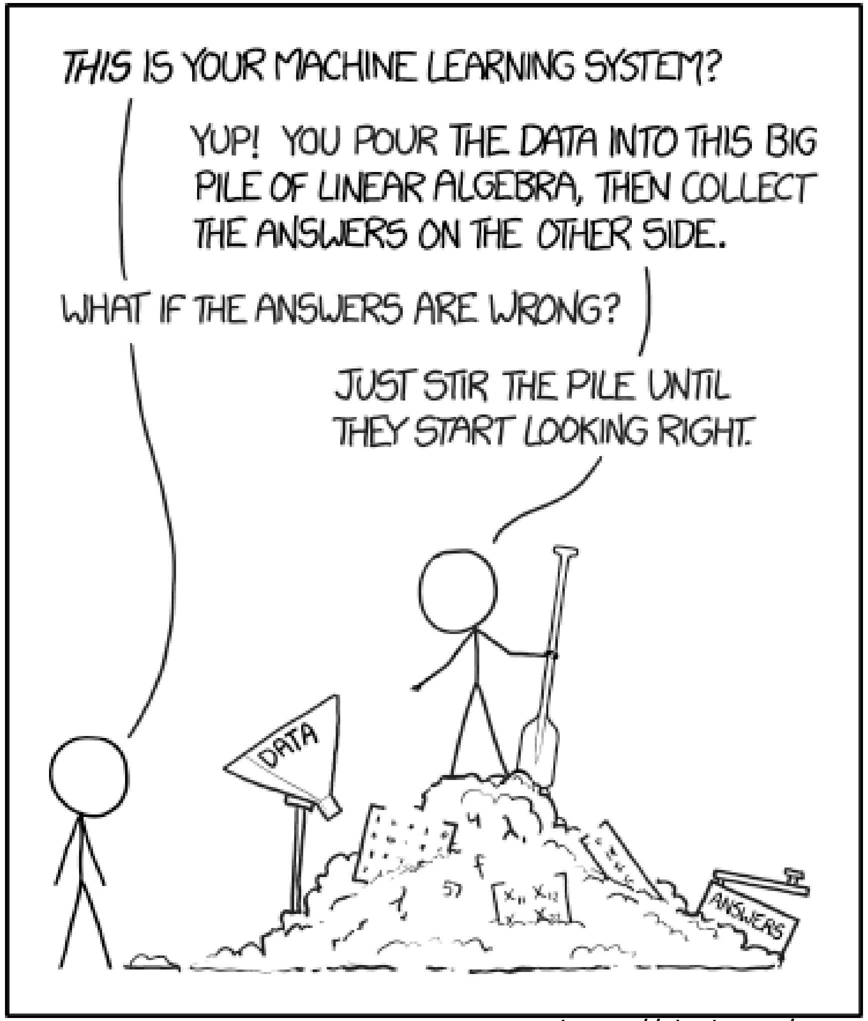


Ben Nachman, DS@HEP 2017

What is Machine Learning?



TensorFlow



https://xkcd.com/1838

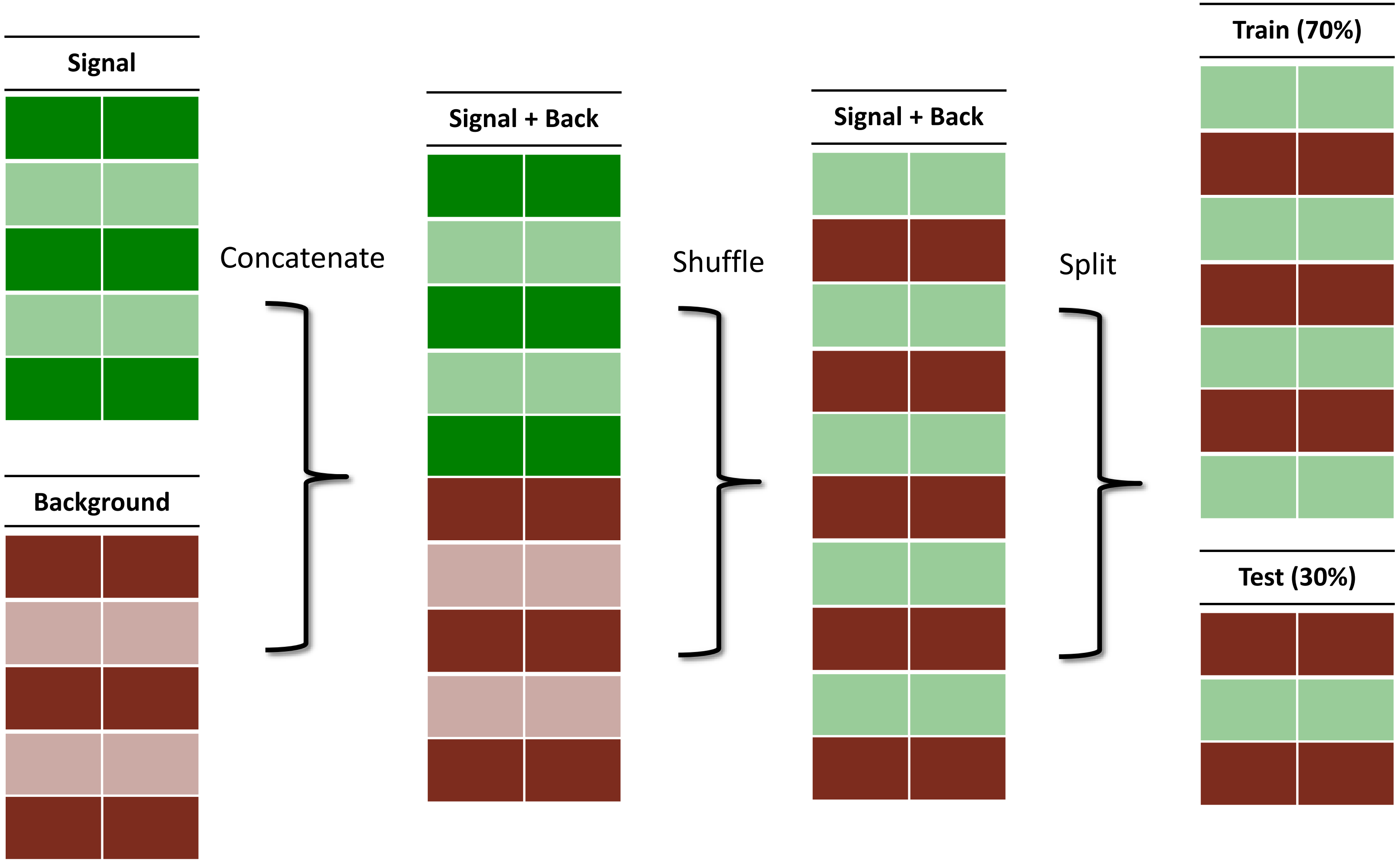
Model Selection and Evaluation

ML models

- Logistic regression
- Multilayer perceptron
- Convolutional Neural Net

Model evaluation

- Cross validation
- Training set (70 %)
- Test set (30 %)



Training Artificial Neural Nets

Input data

Raw data	CSV format

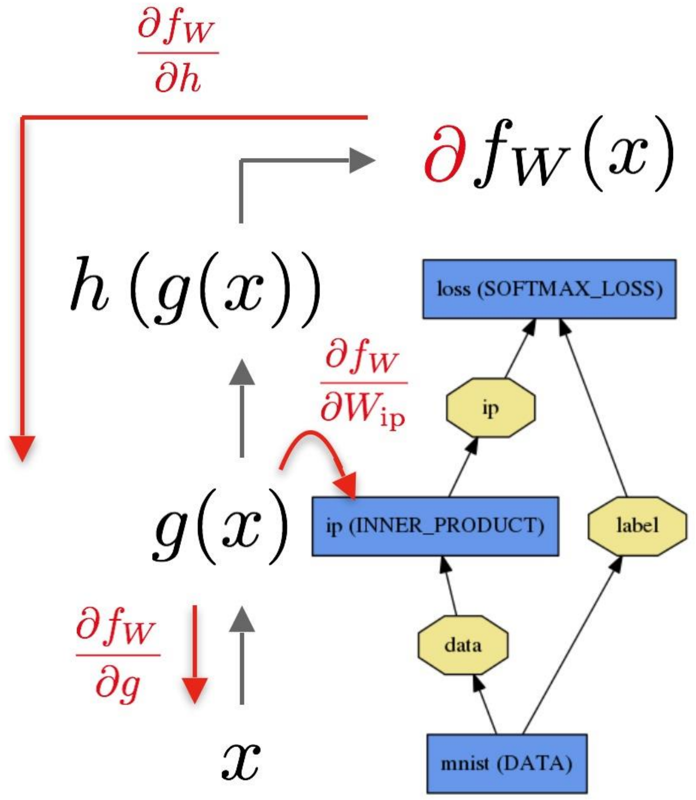
Create deep network

Framework	TensorFlow

Output classification

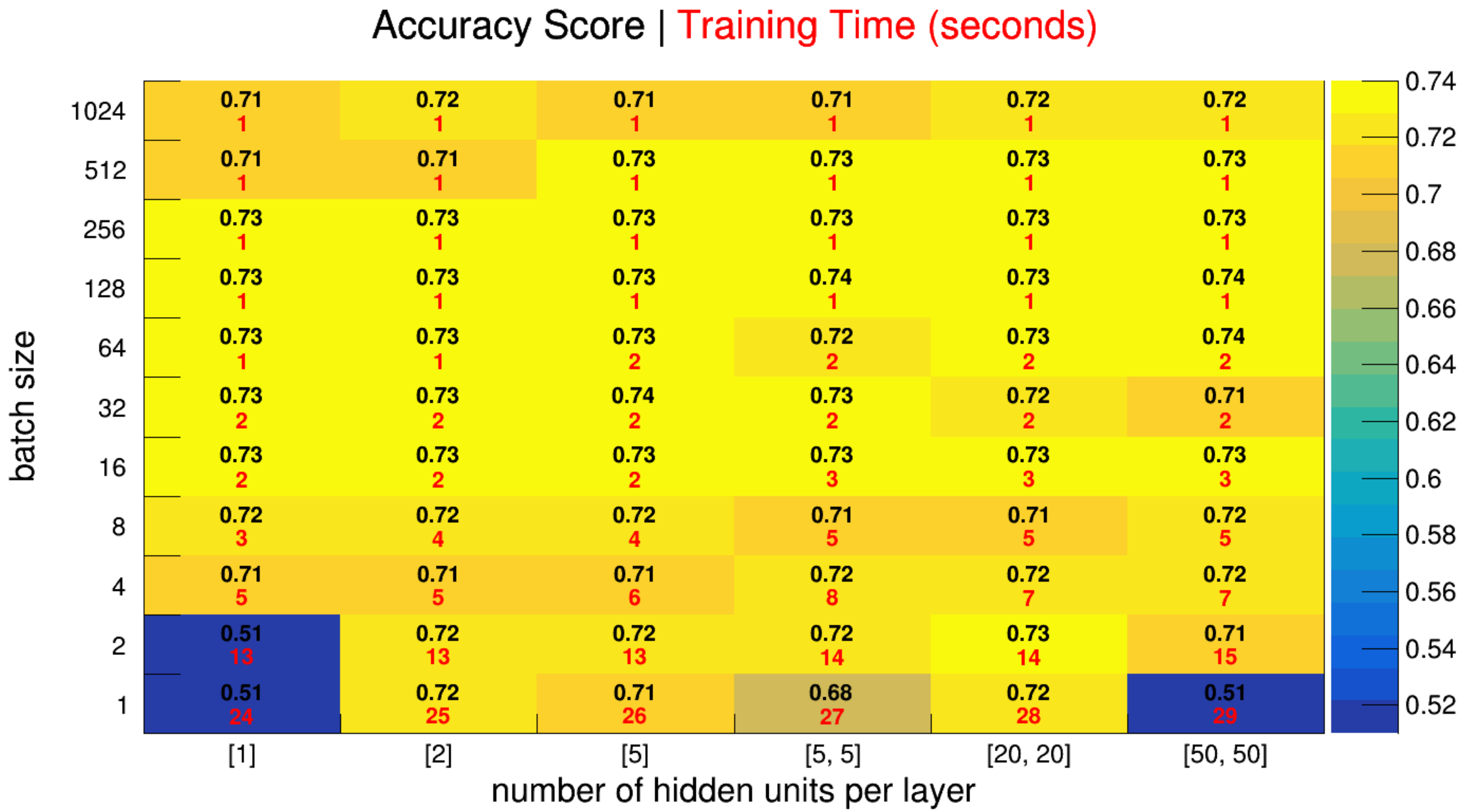
Model evaluation	Accuracy score

Back propagation



http://caffe.berkeleyvision.org

Hyper-parameter tuning



batch size	[1]	[2]	[5]	[5, 5]	[20, 20]	[50, 50]
1024	0.71 / 1	0.72 / 1	0.71 / 1	0.71 / 1	0.72 / 1	0.72 / 1
512	0.71 / 1	0.71 / 1	0.73 / 1	0.73 / 1	0.73 / 1	0.73 / 1
256	0.73 / 1	0.73 / 1	0.73 / 1	0.73 / 1	0.73 / 1	0.73 / 1
128	0.73 / 1	0.73 / 1	0.73 / 1	0.74 / 1	0.73 / 1	0.74 / 1
64	0.73 / 1	0.73 / 1	0.73 / 2	0.72 / 2	0.73 / 2	0.74 / 2
32	0.73 / 2	0.73 / 2	0.74 / 2	0.73 / 2	0.72 / 2	0.71 / 2
16	0.73 / 2	0.73 / 2	0.73 / 2	0.73 / 3	0.73 / 3	0.73 / 3
8	0.72 / 3	0.72 / 4	0.72 / 4	0.71 / 5	0.71 / 5	0.72 / 5
4	0.71 / 5	0.71 / 5	0.71 / 6	0.72 / 8	0.72 / 7	0.72 / 7
2	0.51 / 13	0.72 / 13	0.72 / 13	0.72 / 14	0.73 / 14	0.71 / 15
1	0.51 / 24	0.72 / 25	0.71 / 26	0.68 / 27	0.72 / 28	0.51 / 29

Convolutional Neural Net Architecture

SPRACE Net

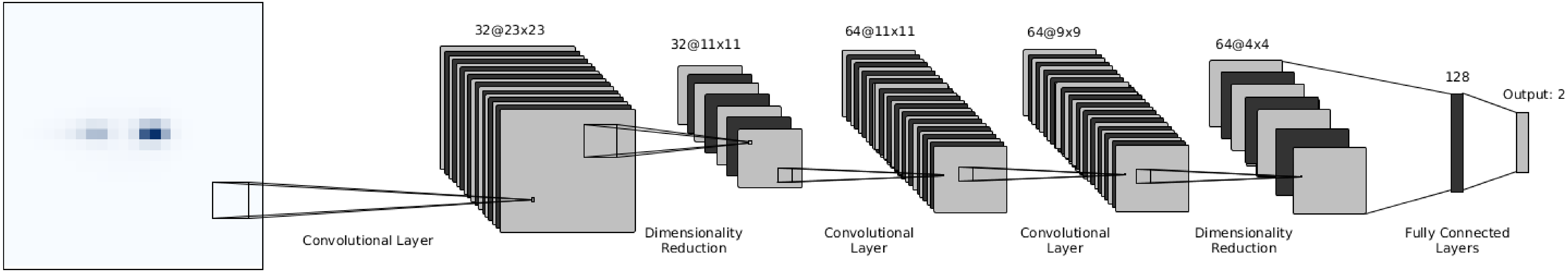
3 convolutional layers

Dropout of 25% to control over fitting

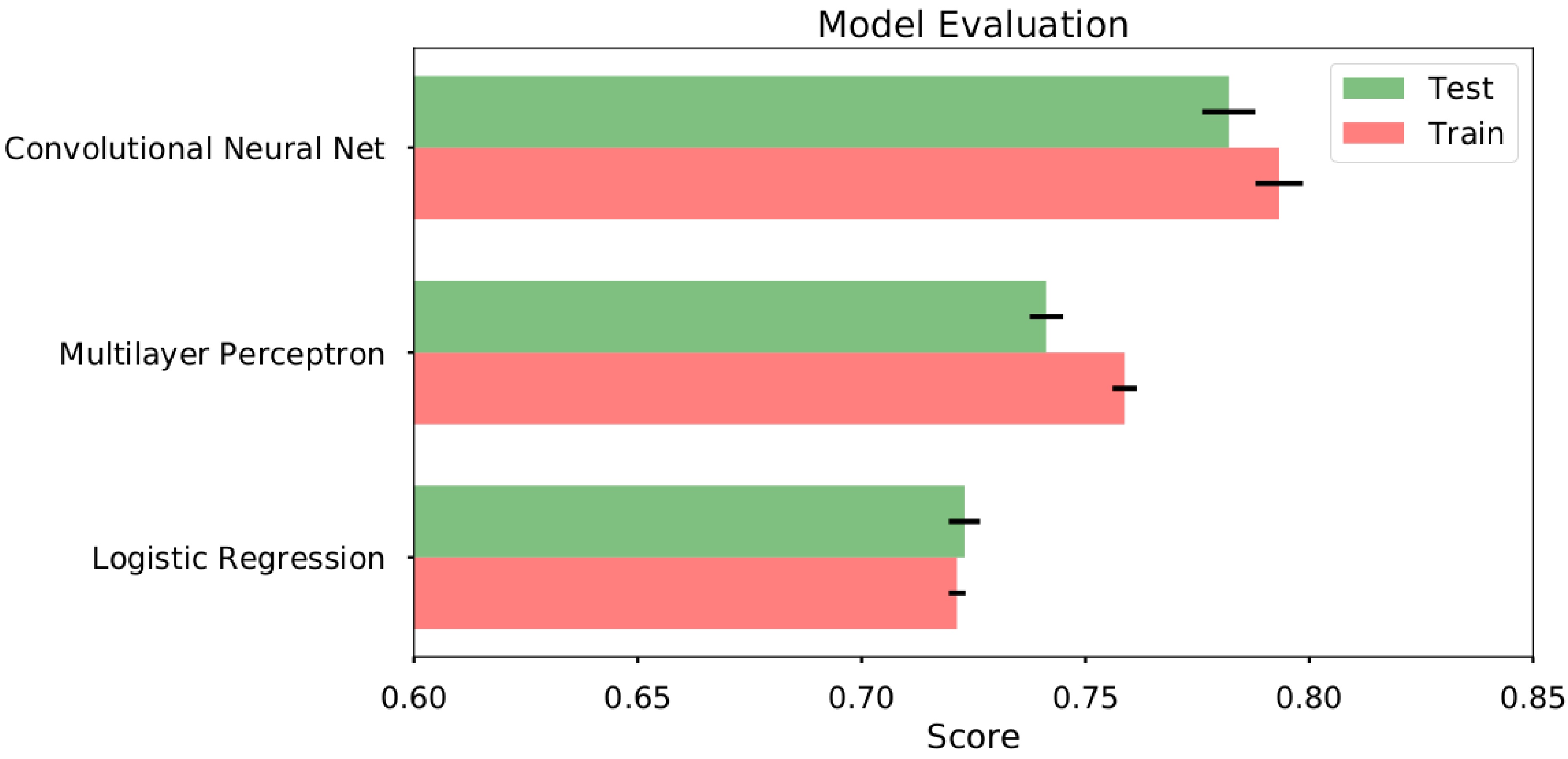
32 filters in the first convolutional layer

Dimensionality reduction using MaxPooling

Total parameters: **187.202**



Classification Score



HPC nodes at NCC-Unesp

Computational node		phi01	phi02	phi03
Excerpt: Processors Central Intel ® Xeon ®	How Many Processors	2	2	2
	Identification	E5-2670	E5-2699v3	E5-2699v3
	Physical cores per processor	8	18	18
	Frequency	2, 6 GHz	2, 3 GHz	2, 3 GHz
	Central Memory	64 GB	128 GB	128 GB
Accelerated Snippet: Processors Intel ® Xeon ® Phi ™	How Many Processors	2	5	4
	Identification	3120A	5110P	7120P
	Physical cores per processor	57	60	61
	Frequency	1, 1 GHz	1, 3.0	1, 2 GHz
	Memory per processor	6 GB	8 GB	16 GB
I/O	SSD Memory	-	1, 2 TB	1, 2 TB
	SATA Drive	4 TB	4 TB	4 TB
Network connection	Ethernet	2 x Gigabit	2 x Gigabit	2 x Gigabit
	InfiniBand	-	40 Gb/s QDR	40 Gb/s QDR

https://software.intel.com/pt-br/articles/tutorial-para-uso-dos-n-s-acelerados-por-intel-xeon-phi-no-ncc-unesp

Performance on Intel® Xeon Phi™

- Server phi02

Batch size	5	**50**	500	5000
Training time (s)	328 ± 1	74 ± 1	57 ± 1	51 ± 1
Accuracy score	0.775	**0.783**	0.776	0.734

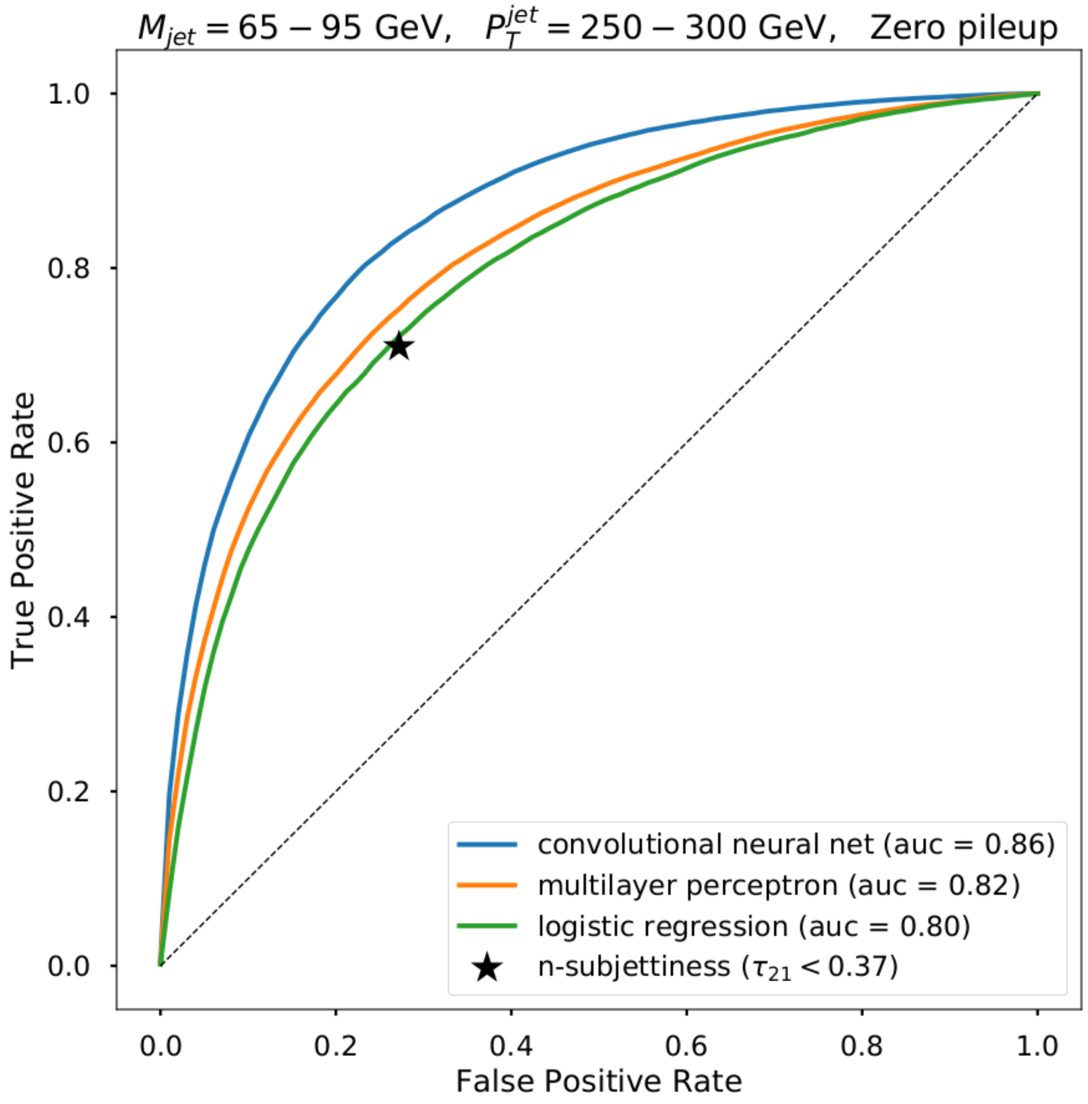
- Server phi07

Batch size	5	**50**	500	5000
Training time (s)	1960 ± 11	425 ± 8	191 ± 5	124 ± 4
Accuracy score	0.776	**0.784**	0.775	0.734

Results and Outlook

Our results confirm the good performance of convolutional neural networks to handle the problem of classification of jet images, demonstrated by the area under the ROC curve (auc = 0.86).

Deep learning applications in the field of high energy physics must improve data analysis techniques in the coming years.



Acknowledgments

intel®

SPRACE

FAPESP
FUNDAÇÃO DE AMPARO À PESQUISA
DO ESTADO DE SÃO PAULO

References

J. Cogan, M. Kagan, *Jet-images: computer vision inspired techniques for jet tagging*, JHEP 02 (2015) 118

Thank You

Backup

CONTROL DISTRIBUTIONS

Jet Algorithm

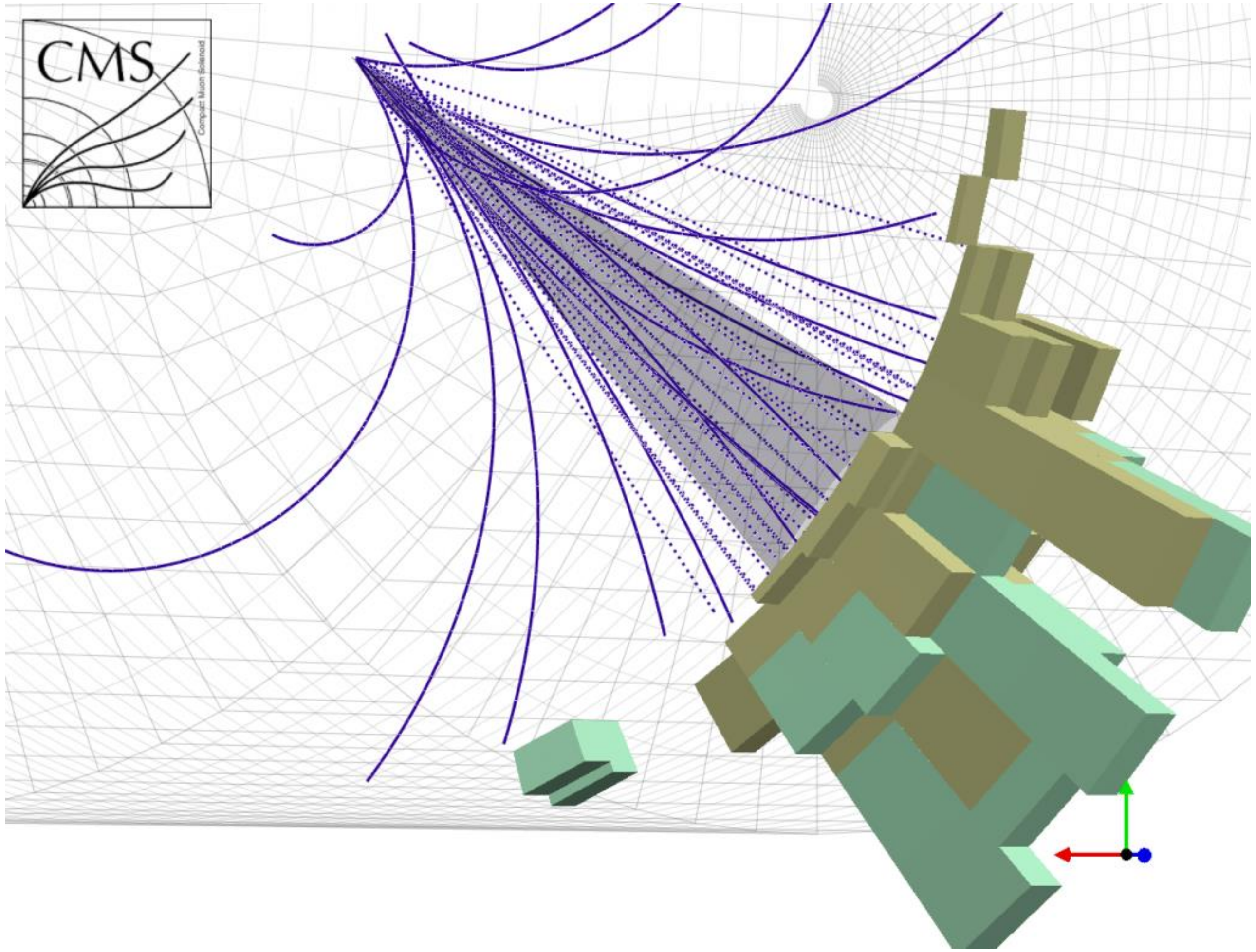
Collimated spray of hadrons

- Local deposits of energy

Anti-k_T algorithm

- Size parameter R
  - R = 0.4 for “standard” hadronic jets
  - R = 0.8 for “boosted” jets
- Collinear, infra-red safe algorithm
- Iteratively combine particles according to the distance d_{ij}

$$d_{ij} = \min(k_{ti}^{2p}, k_{tj}^{2p}) \frac{(\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2}{R^2}$$

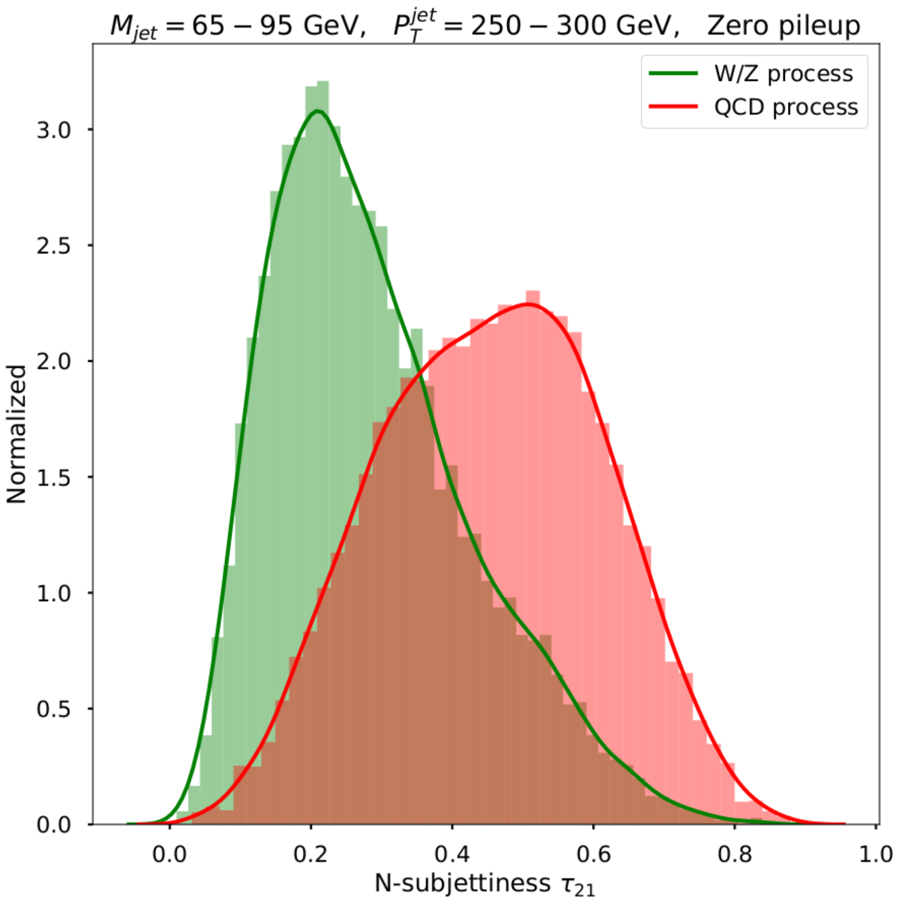


N-subjettiness

Substructure variable made out of jet constituents.

Quantifies the capability of clustering the jet constituents in exactly N subjets.

N-subjettiness $tau_{21} < 0.37$ provides optimal descrimination between signal and background.



Performance on Intel® Xeon Phi™

- Server phi02

	OpenBLAS	MKL
Run time	0h 12min 26s	0h 12min 17s

- Server phi07

	OpenBLAS	MKL
Run time	1h 16min 01s	1h 10min 41s