LABORATÓRIO DE INSTRUMENTAÇÃO
E FÍSICA EXPERIMENTAL DE PARTÍCULAS
partículas e tecnologia

Competence Center on Simulation and Big Data

Jornadas Científicas do LIP, Évora, 17th Feb 2018

Nuno Castro
(on behalf of the Competence Center)
nfcastro@lip.pt

Universidade do Minho
Escola de Ciências

IF INVESTIGADOR FCT

FCT Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR

IF/00050/2013/CP1172/CT0002

Competence center
for Simulation and Big Data

- Aims to gather and expand LIP's competences on these areas
  - achieve critical mass
  - train students and researchers
  - explore the synergies between different LIP groups
  - provide services and consulting inside and outside LIP

- This effort has started ~1 year ago
  - establishing communication channels and discussion forums
  - survey of the LIP competences on these areas
  - planning of the future

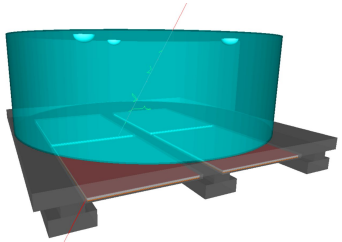
Simulation

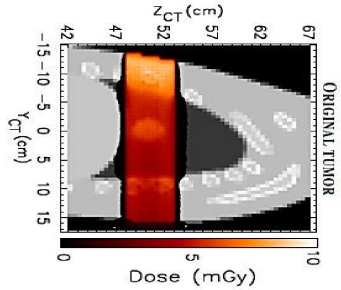
- coordination: Bernardo Tomé and Patrícia Gonçalves
- mailing list: geant4@lip.pt

  - GEANT4 has long been used in LIP: from biomedical applications to detector performance studies or space radiation effects

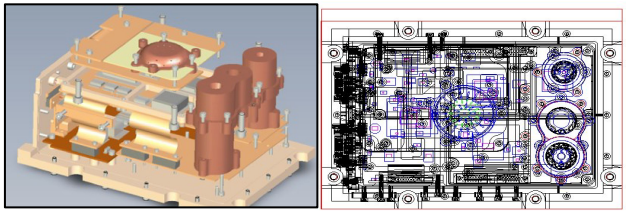
    - critical activity for different LIP groups

  - LIP is a member of the GEANT4 collaboration for more than 10 years, accumulating an important expertise, both from the user and developer points of view;



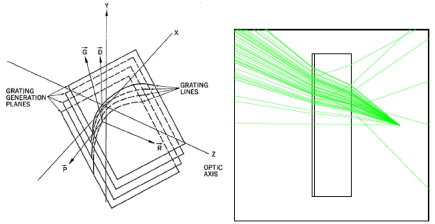


Simulation
Survey of competences



GUIMesh – CAD to GDML conversion

- A survey of the GEANT4 competences at LIP was undertaken
- Know-how beyond applications development
- Expertise in several GEANT4 kernel categories
- LIP members are teaching courses with some emphasis in GEANT4 (MSc and PhD @ Lisboa and Coimbra Universities)

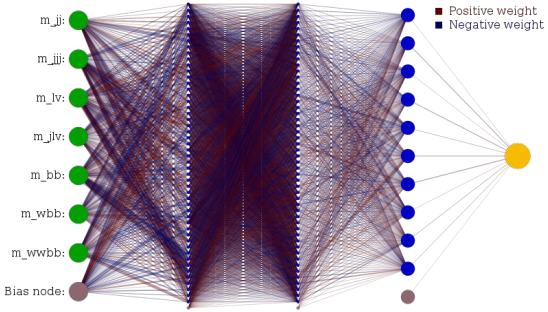


Simulation of diffraction gratings

Simulation
Next steps

- Coordinate the existing expertise in order to foster the interaction with the community
- LIP GEANT4 workshop/school/tutorial towards summer... to be better defined
- Possible participation at *Física 2018*
- Identify the potential and strategy(ies) to create links with the *outside world*
- Increase our contribution to (and visibility within) the GEANT4 collaboration

Big Data



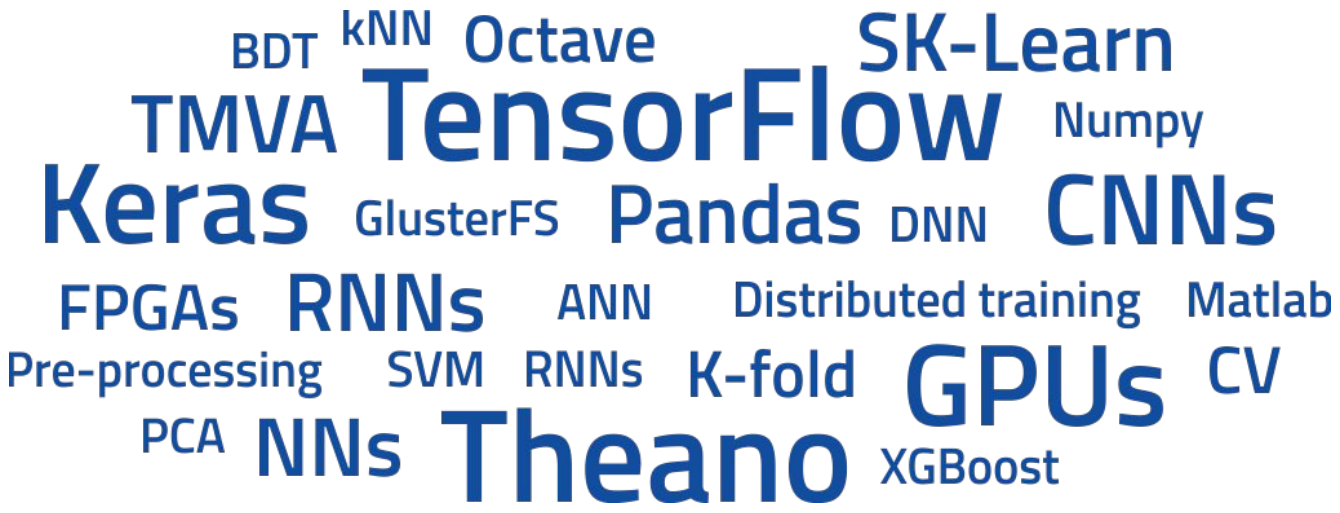
- coordination: Nuno Castro and Guilherme Milhano
- mailing list: big-data@lip.pt

  - LIP has been involved in the analysis of extremely large amounts of data produced by different experiments in HEP, for a long time
  - Expertise on the implementation and development of elaborate multivariate techniques aiming at a vast range of applications
  - Competence in efficient data processing to better use the available computing resources

Big Data and machine learning

Tools and interests of the team

- Survey to understand what we know (or would like to know)

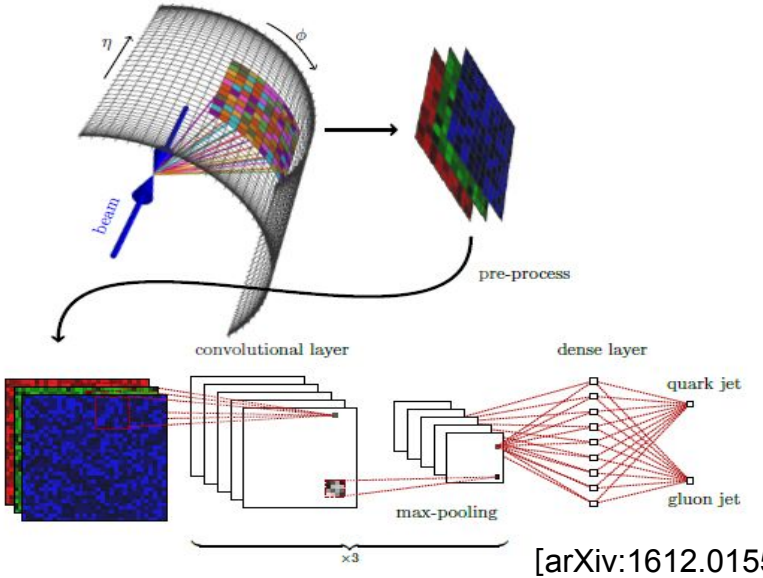


https://docs.google.com/document/d/1Cd0vLKFSb886OuMb583Vo9YPdS_G_8JL9q8N0HHPaN4/edit?usp=sharing

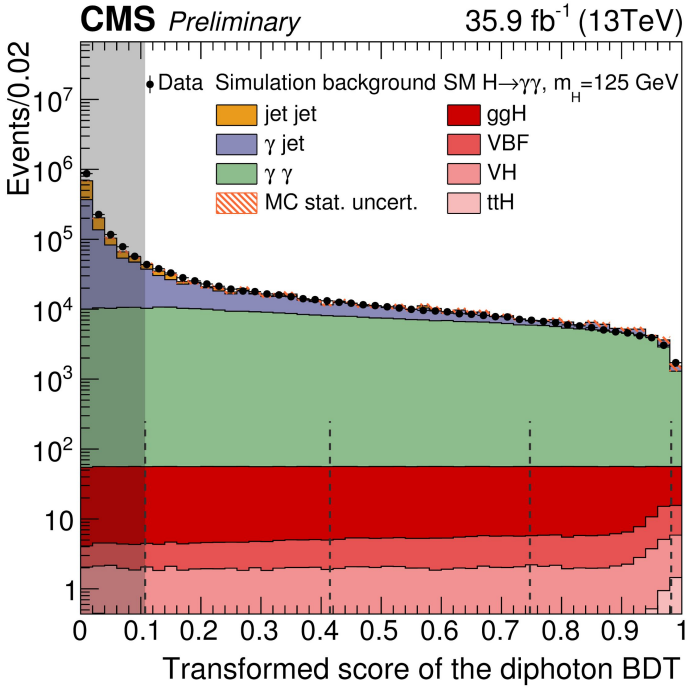
Big Data - machine learning

Regular informal meetings

- Emphasis on the journal club style:
  - discussion on new techniques:



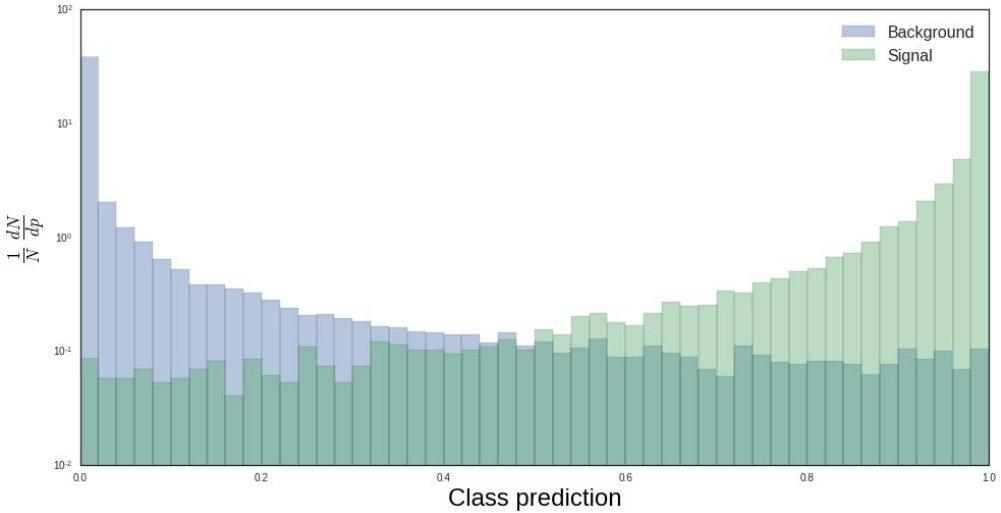
[arXiv:1612.01551]

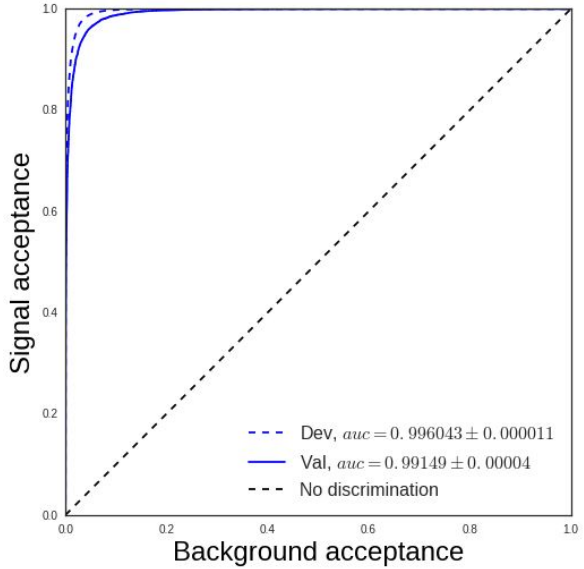


[https://cds.cern.ch/record/2264515]

Big Data - machine learning
Regular informal meetings

- Emphasis on the journal club style:
  - Keras tutorial by Giles Strong





[https://github.com/GilesStrong/ML_Tutorials]

Big Data - machine learning
Regular informal meetings

- Emphasis on the journal club style:
  - Docker and udocker tutorial by Mário David

```
$ git clone git@github.com:LIP-Computing/bigdata_tutorial.git
$ cd bigdata_tutorial/
$ mkdir dockerfiles
```

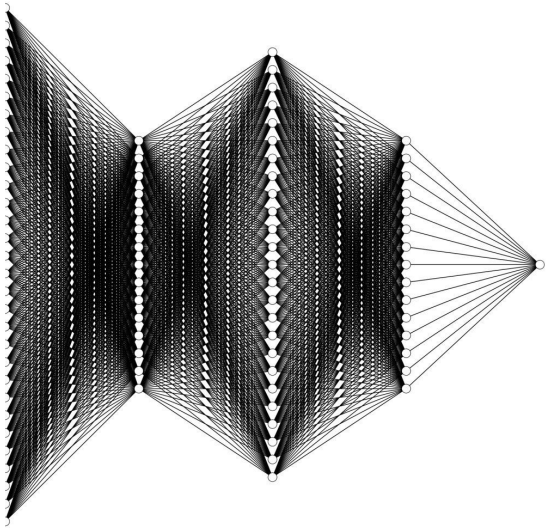
```
$ git add dockerfiles/docker-keras-theano
$ git status
On branch master
Your branch is up-to-date with 'origin/master'.
Changes to be committed:
  (use "git reset HEAD <file>..." to unstage)

        new file:   dockerfiles/docker-keras-theano

$ git commit -m 'add dockerfile'
[master 454cc81] add dockerfile
 1 file changed, 25 insertions(+)
 create mode 100644 dockerfiles/docker-keras-theano
$ git push
X11 forwarding request failed on channel 0
Counting objects: 4, done.
Delta compression using up to 6 threads.
Compressing objects: 100% (3/3), done.
Writing objects: 100% (4/4), 696 bytes | 0 bytes/s, done.
Total 4 (delta 0), reused 0 (delta 0)
To git@github.com:LIP-Computing/bigdata_tutorial.git
   baf3ebf..454cc81  master -> master
```

Big Data - machine learning
Computational resources

- It is important to ensure computing resources (hardware and software)
  - flexibility
  - training
- GPU grant from NVIDIA:
  - a Titan XP board was donated
  - a dedicated server is being prepared and will become available soon
    - contact us (hcarvalho@lip.pt) if you want an account

Big Data - machine learning
Computational resources

- Access to accelerator devices (e.g. CPUs) is critical for the training of complex (deep) neural networks



```
# set the matplotlib backend so figures can be saved in the background
# (uncomment the lines below if you are using a headless server)
# import matplotlib
# matplotlib.use("Agg")

# import the necessary packages
from pyimagesearch.minigooglenet import MiniGoogLeNet
from sklearn.preprocessing import LabelBinarizer
from keras.preprocessing.image import ImageDataGenerator
from keras.callbacks import LearningRateScheduler
from keras.utils.training_utils import multi_gpu_model
from keras.optimizers import SGD
from keras.datasets import cifar10
import matplotlib.pyplot as plt
import tensorflow as tf
import numpy as np
import argparse
```

Organizing Committee:

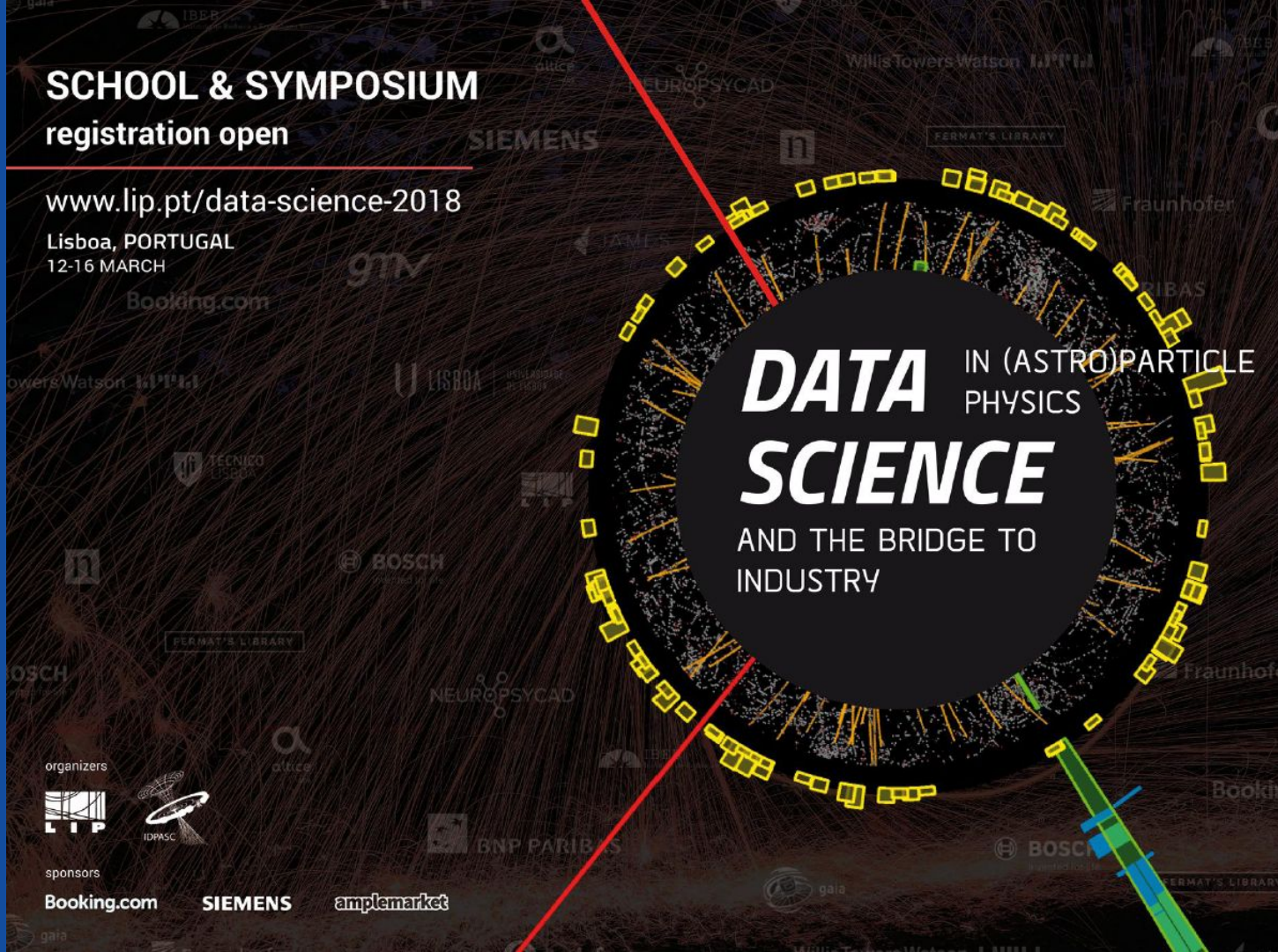
Bernardo Tomé (co-chair)
Catarina Espírito Santo
Daniel Galaviz
Guilherme Milhano
Lorenzo Cazon (chair)
Mário Pimenta
Nuno Castro
Ricardo Gonçalo
Ruben Conceição (co-chair)

Design and Development:

Carlos Manuel

Secretariat:

Natália Antunes



Work is ongoing...

- The competence center has started its activities
- Strategic to LIP
  - Training
  - Synergies
  - Funding
- **We need you!**

JORNADAS CIENTÍFICAS

Competence Center on
Simulation and Big Data

big-data@lip.pt
geant4@lip.pt