# Use of Machine Learning tools: plans from IFCA

COMCHA meeting
5th June 2018

A. Calderón, C. Fernández, P. Fernández, I. Heredia, L. Lloret,
J.Marco de Lucas, P. Martínez, J. Piedra, M. Señas, N. Trevisani

IFCA
Instituto de Física de Cantabria

CMS

UC
UNIVERSIDAD DE CANTABRIA

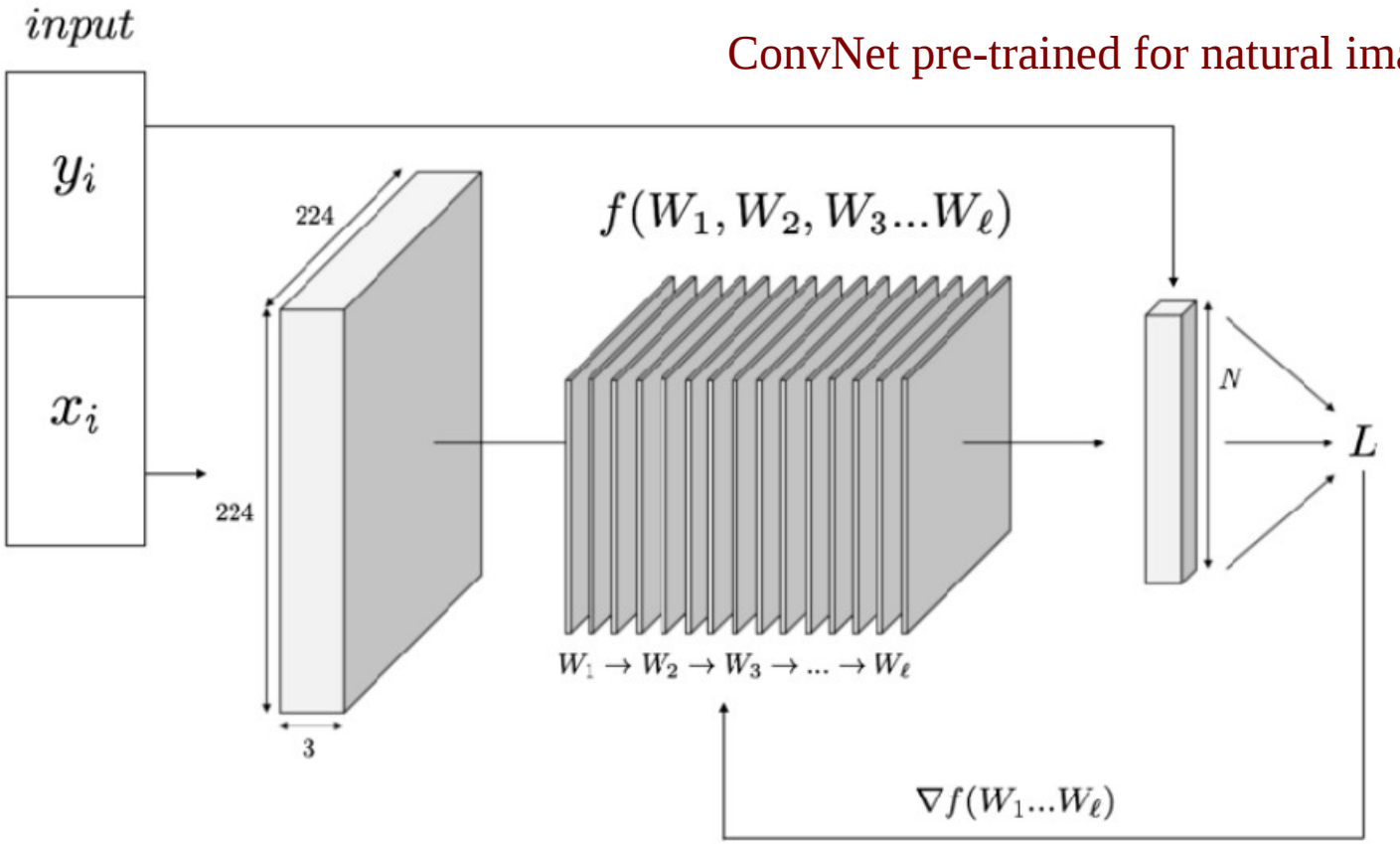
# Introduction and context

- IFCA has a strong group of advanced computing with focus on ML techniques.
  - About 10 people including physicists and software engineers.
  - This has to be added up to the High Energy Physics group.
- Administration of the IFCA Tier2 + HEP Tier3 and other resources:
  - Altamira super computer.
  - Currently 10 new GPUs + 1 old GPU.
  - Cloud resources.
- Also strong involvement in the Open Data group within CMS.
- Several years of expertise in Deep Learning techniques mainly for image classification.
  - Biodiversity. (Mobile application). **arXiv:1706.03736**.
  - High Energy Physics classification. **arXiv: 1708.07034**
  - Working now in medical diagnosis.
- Excellent environment to explore and promote ML tools in the context of HEP.

# HEP example: event image classification

- Final degree project developed by Celia Fernández. Presented at the CMS ML forum.
- Use of a Convolutional Network to classify event images from different processes.

ConvNet pre-trained for natural images.

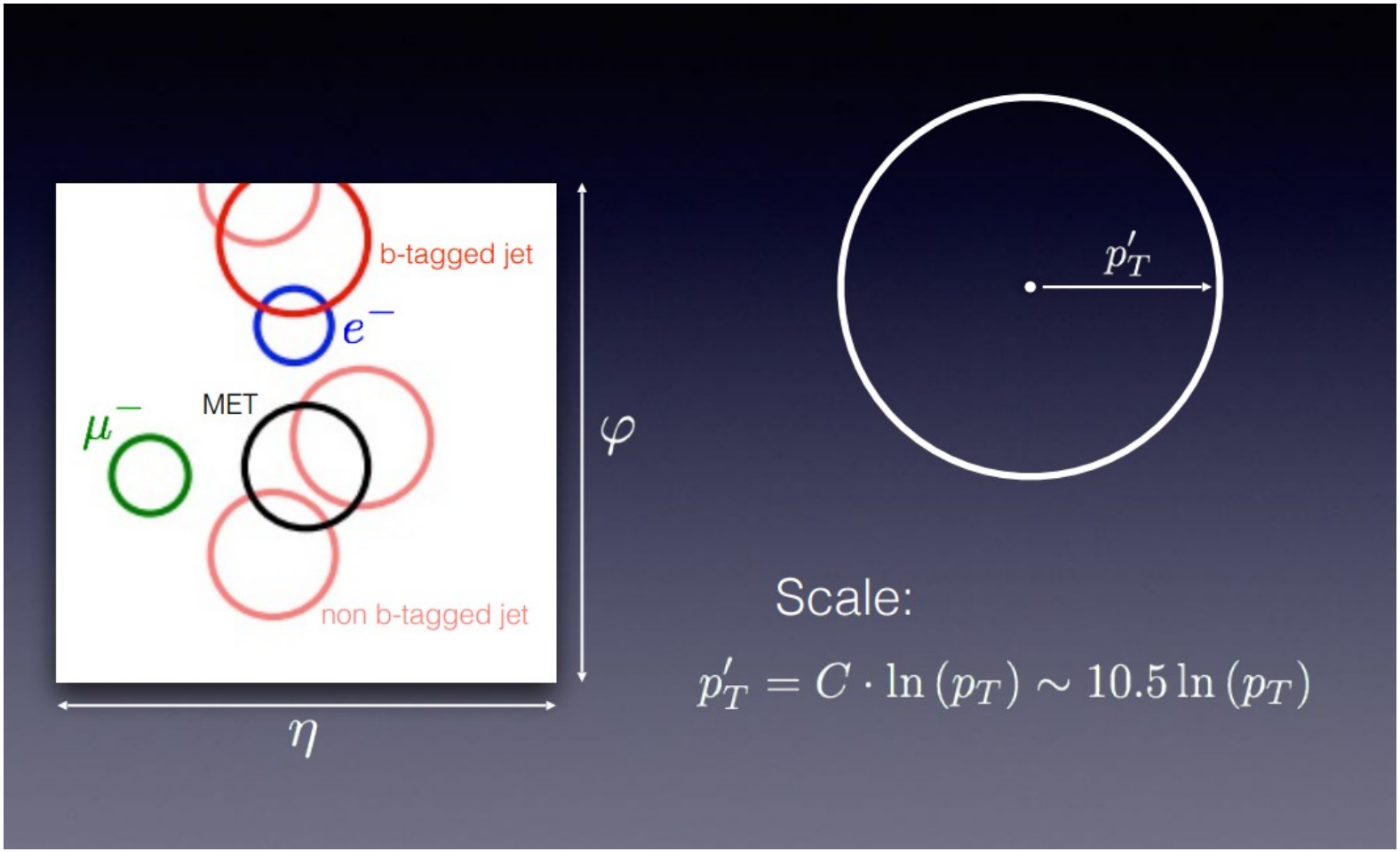



https://indico.cern.ch/event/646782/

**arXiv: 1708.07034**

# HEP example: event image classification

- HEP events are translated into pictures according to a simple geometrical mapping.

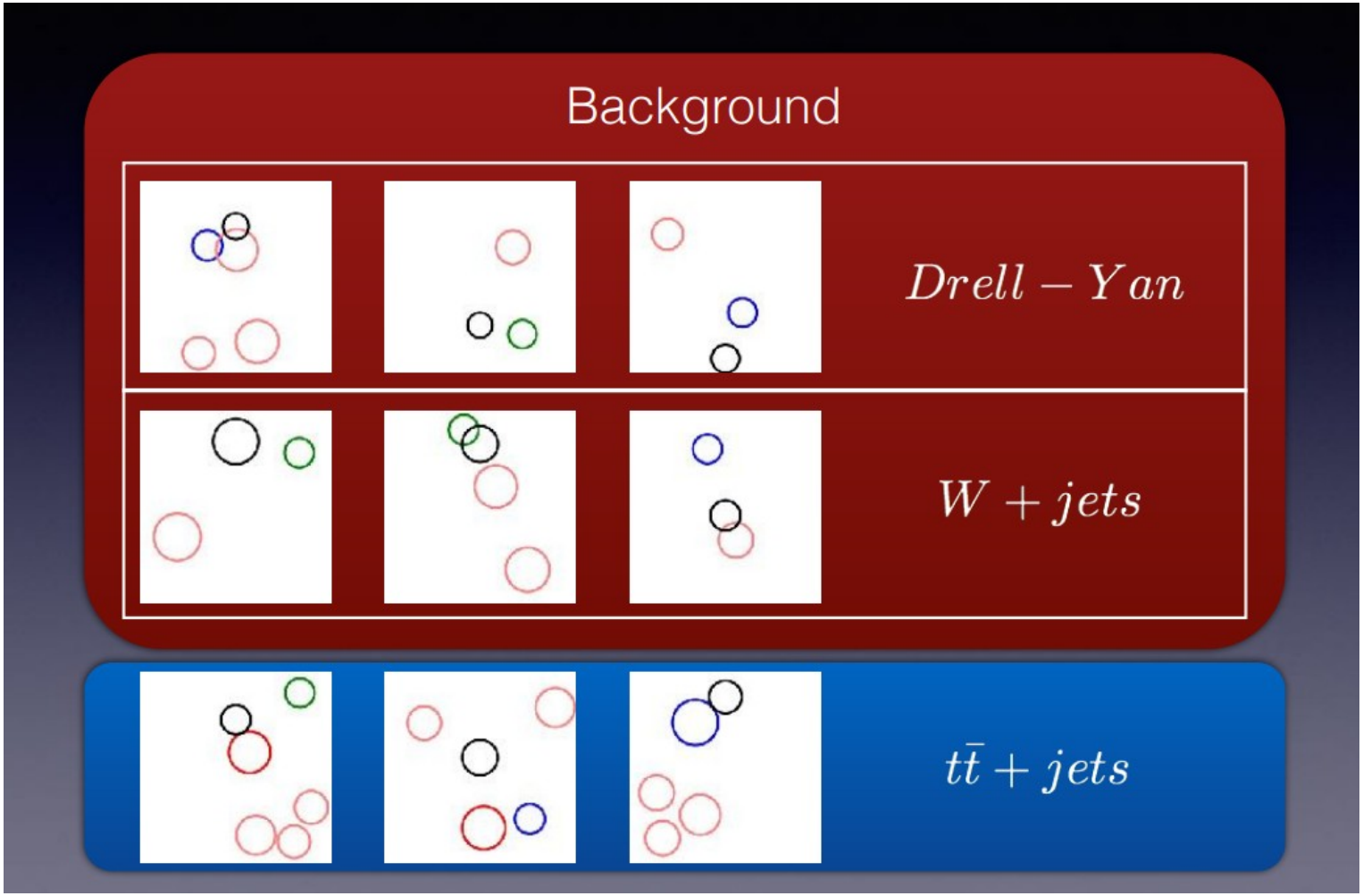



https://indico.cern.ch/event/646782/

**arXiv: 1708.07034**

# HEP example: event image classification

- Using these images to treat the background discrimination as an image classification.

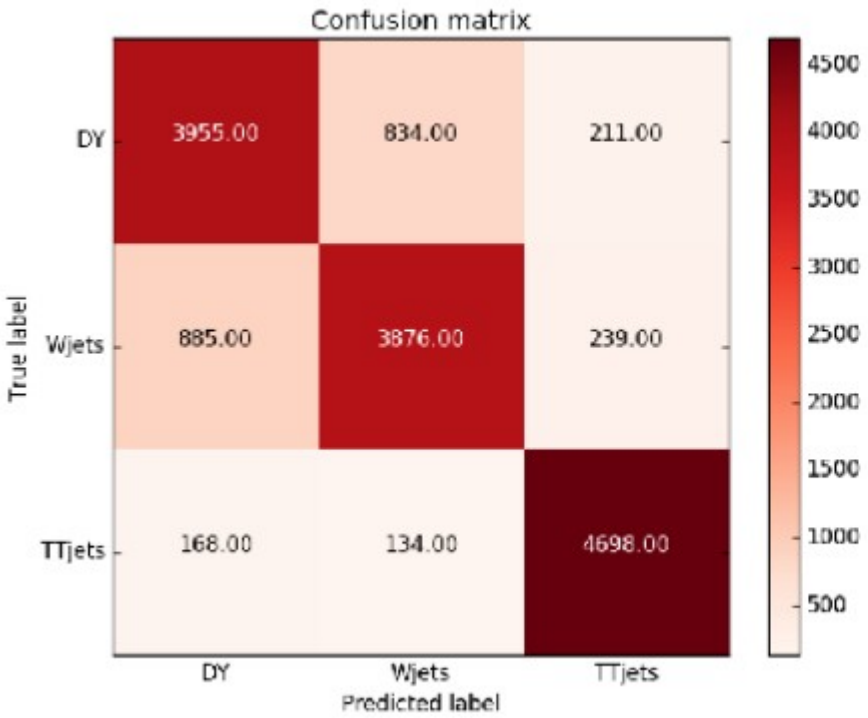


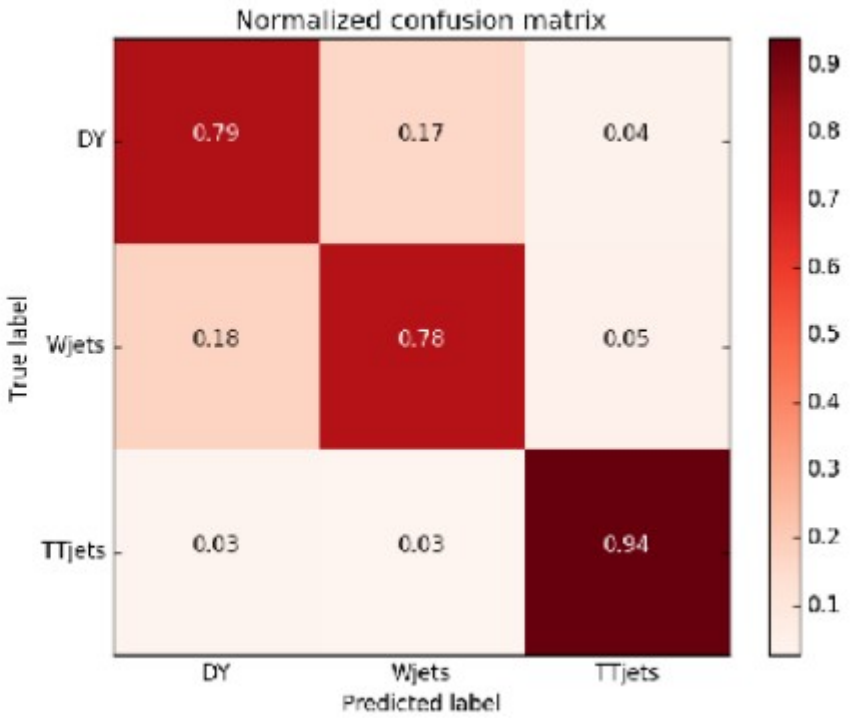

https://indico.cern.ch/event/646782/

**arXiv: 1708.07034**

# HEP example: event image classification

- Using these images to treat the background discrimination as an image classification.




- Probability of choosing the right category: 84% (98% within the two first choices)

https://indico.cern.ch/event/646782/

**arXiv: 1708.07034**

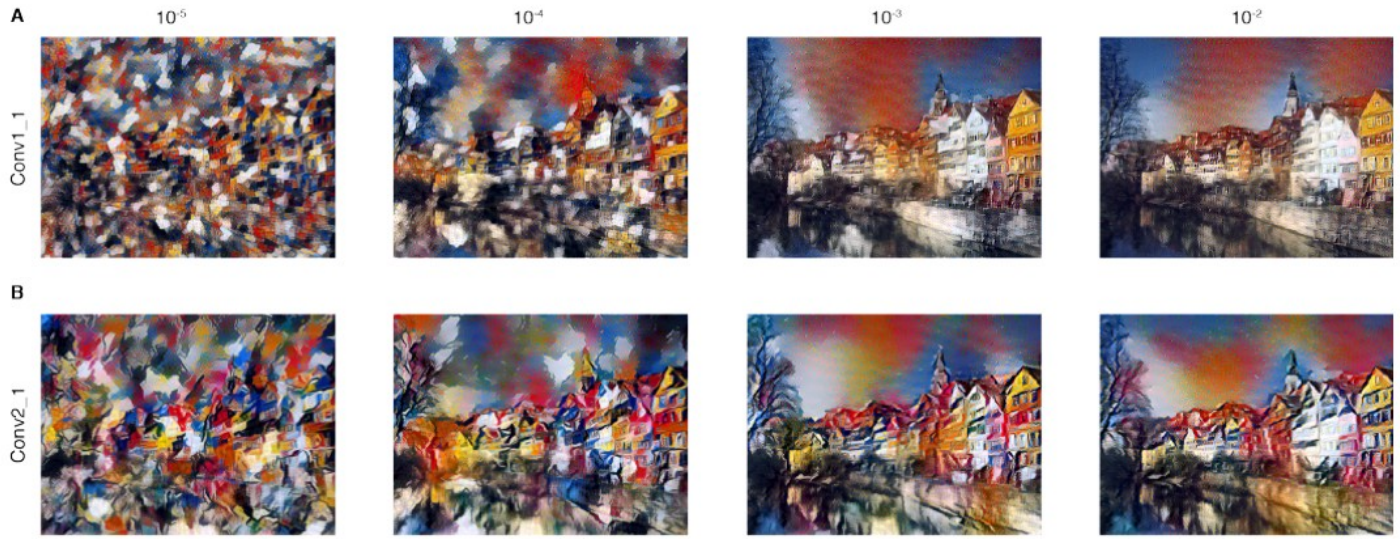
# HEP example: realistic MC generation

- This is currently a Final Degree Project (Mario Señas) to be presented in September.
- Learning local correlation patterns to produce realistic MonteCarlo simulations.


**Based on arXiv:1508.06576**

# HEP example: realistic MC generation

- The “style”of an artist is encapsulated in the local correlations among pixels.
- The content produces larger-scale correlations among distant pixels.
- Convolutional networks learn local correlations in the first layers/filters.
- Idea is to consider MC simulation as a real image and data as an “artist/detector” vision.
  - Detector effects in real data introduce local correlations in the observables.
  - We train with data to learn these local features and then apply to real MC.




$$\mathcal{L}_{total}(\vec{p},\vec{a},\vec{x}) = \alpha\mathcal{L}_{content}(\vec{p},\vec{x}) + \beta\mathcal{L}_{style}(\vec{a},\vec{x})$$

**Based on arXiv:1508.06576**

# HEP example: realistic MC generation

- In order to map processes into pixel maps where correlations make sense:
  - We make two categories of observables: simple and composed.
  - Both of them are used in a 2D table. Size of each pair map into a pixel colour.

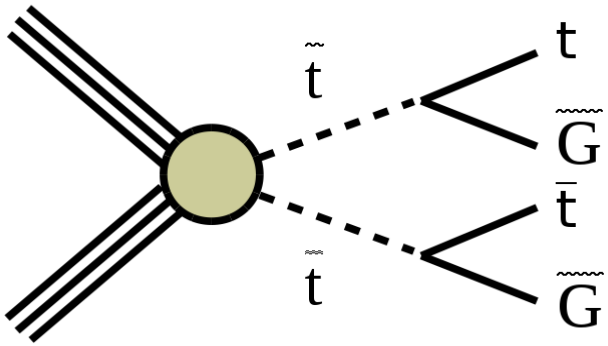
| | MET | N Jet | MT2 | HT | MT | ... | ... | ... | ST |
|---|---|---|---|---|---|---|---|---|---|
| Jet1 pt | | | | | | | | | |
| Jet1 phi | | | | | | | | | |
| Jet1 eta | | | | | | | | | |
| Jet2 pt | | | | | | | | | |
| Jet2 phi | | | | | | | | | |
| ... | | | | | | | | | |
| Lep1 pt | | | | | | | | | |
| ... | | | | | | | | | |
| Lep2 eta | | | | | | | | | |

Work ongoing...

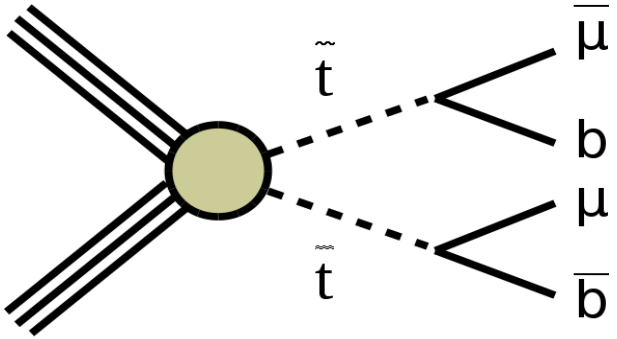
**Based on arXiv:1508.06576**

# Plans for displaced vertex/track reconstruction

- IFCA group has an increasing interest in searches including displaced tracks/leptons.
  - Expertise both in SUSY and EXO searches in conventional prompt-signatures.
  - Also expertise in b-tagging algorithms.
- Starting now involvement within the new timing detectors for the CMS upgrade.
- Already contributing in the context of the Yellow Report for the European Strategy.
- Two PhD students (Celia Fernández, Pedro Fernández) with focus on this activity.



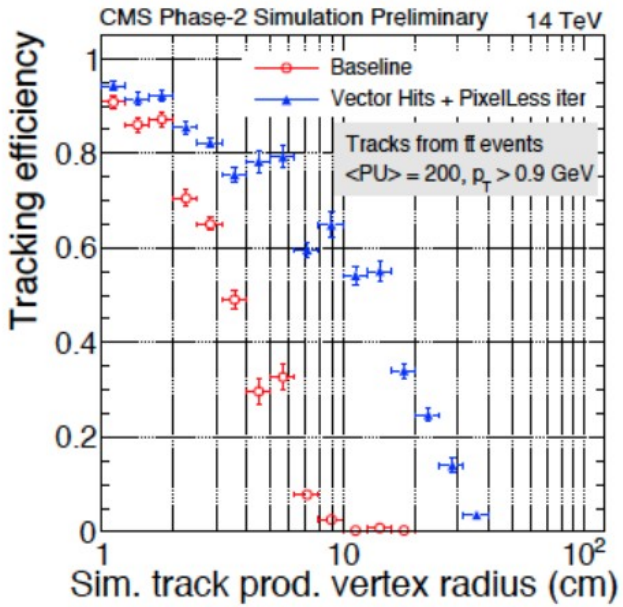

GMSB stop

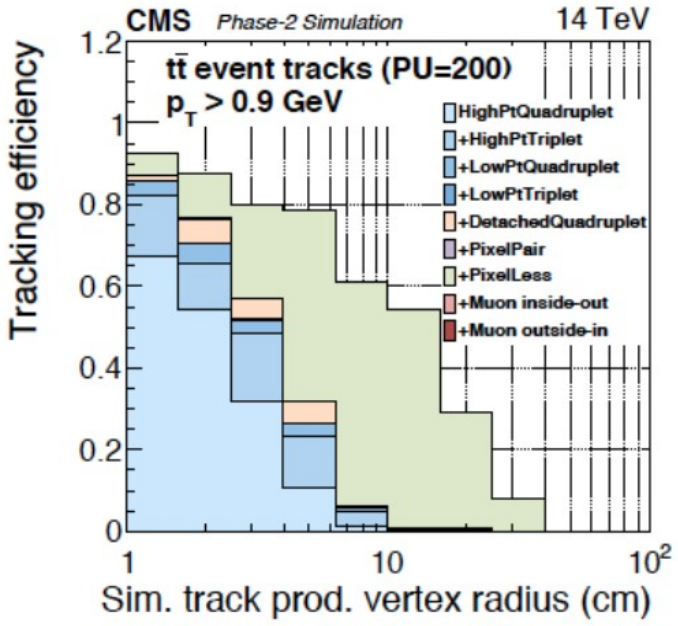



RPV stop

# Plans for displaced vertex/track reconstruction

- Vertex and displaced tracks have very low efficiency (even lower for HL-LHC)
- Plan to work in the application of Deep Algorithms to reconstruct/identify these tracks.
  - Both including the timing and only with spatial information.
- Two PhD students (Celia Fernández, Pedro Fernández) with focus on this activity.
- Using Deep CSV (convolutional network for b-tag identification) as a starting point.
- **Exploration and first steps being carried out right now.**



# Conclusions

- IFCA HEP group is increasingly getting into the use of Machine Learning tools.
- Great synergies with the Advanced Computing group with expertise on the subject.
- Computing resources also available at IFCA to perform studies/work.
- Several exploration examples have been already developed.
- Facing now the first large project to reconstruct displaced tracks w/wo timing information.