The CMS Data Analysis School experience

Nicola De Filippis

Dipartimento Interateneo di Fisica "M. Merlin", Politecnico and INFN Bari, Italy

on behalf of the CMS Collaboration

The CMSDAS are the official schools for learning about CMS Data Analysis; they are coordinated by the CMS Schools Committee.
Originated in 2010 at the LHC Physics Center (LPC) in Fermilab, CMS typically has three schools every year in Europe, Asia, and the US.
So far it has been conducted at Fermilab, Pisa, CERN, DESY, Bari, Taipei, Kolkata, and Daegu.

Goals of the CMSDAS

- teach master students, Ph.D. and young post-docs for the data analysis with the CMS software via hands-on tutorials

- train them about timing and competition in doing their work

- expose to them about how to write a paper and the steps towards a publication in a scientific journal

- Provide venues for discussions: educational, professional and social

Preparation steps

- coordination between the host institute and the international CMS advisory committee for the school foreseen

- the time of the school is chosen carefully to avoid overlap with or closely preceding major physics conferences

- host institute must ensure a computing infrastructure be able to deal with hundreds of analysis jobs running in parallel and to host few TB of data

- simultaneous availability of multiple conference rooms equipped with wi-fi / cable connection / blackboard

- preparatory exercises on computing and software to be executed and answered by the students before the beginning of the school

- twikis for short and long exercises to be prepared well before the beginning of the school → to be tested by the people from the host institute

- each student can choose only one "long" exercise and then they are mapped to at least 3 "short exercises" pointed by the facilitators for the assigned long exercise.
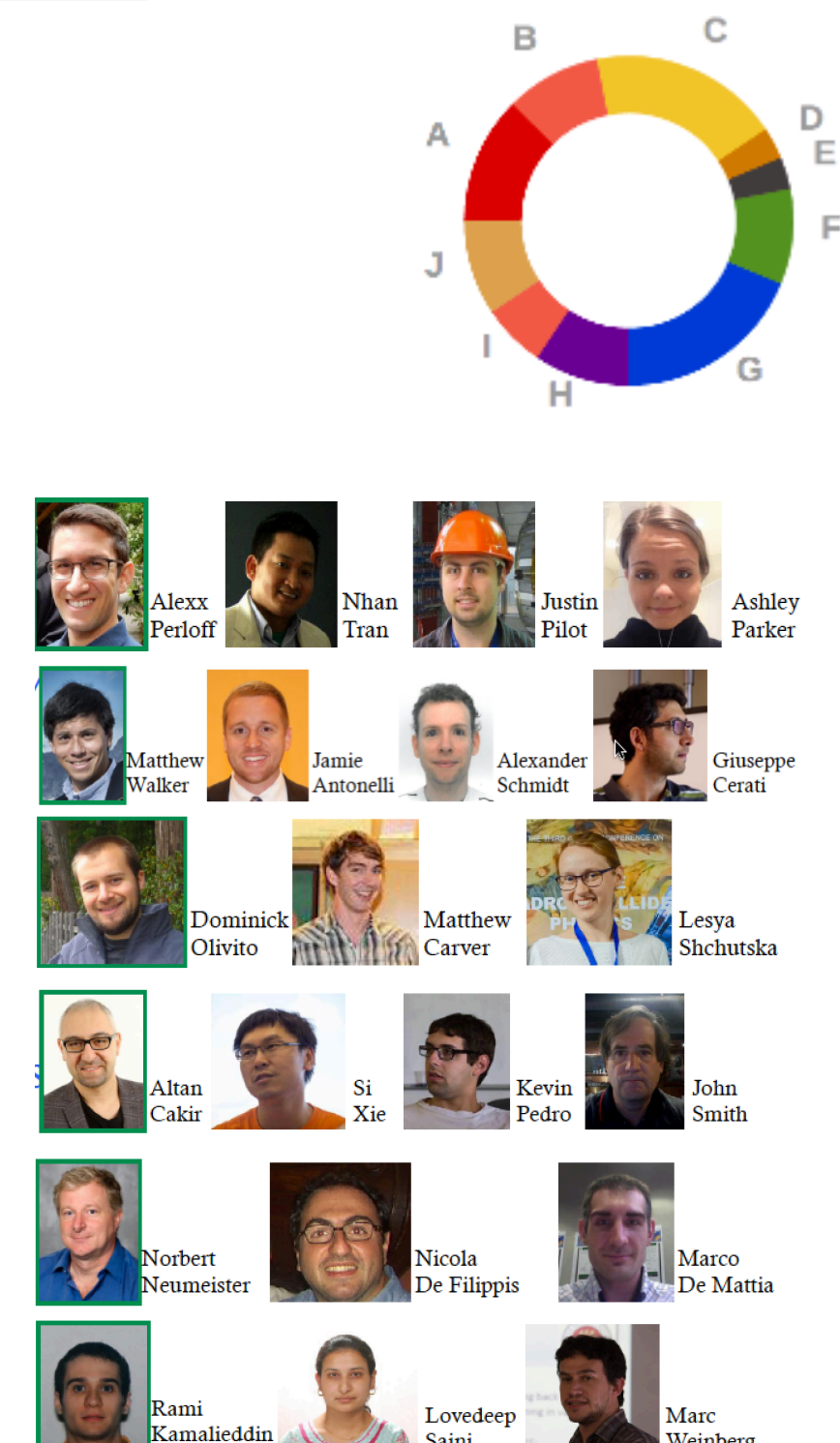
Operations

- Half of the first day devoted to plenary lectures on physics, detector and software tools

- two days of 9-12 "Short" exercises about objects reconstruction/identification/trigger/generators/statistics

- few hours for the writer publication training

- two and a half days of 8-10 "Long" exercises about physics analysis from official CMS physics groups (HIG, EXO, SUSY, SMP, TOP …)

- team of 6-8 students formed

- bunch of "facilitators" between the CMS experts selected to train the students

- last day "mini-symposium": competition between the analysis teams for the "Best Analysis Team" Prize

Plenary lectures

- The "big picture" in physics :
  - theory
  - CMS experimental physics results and challenges

- LHC status, results and plans

- CMS detector status, results and plans
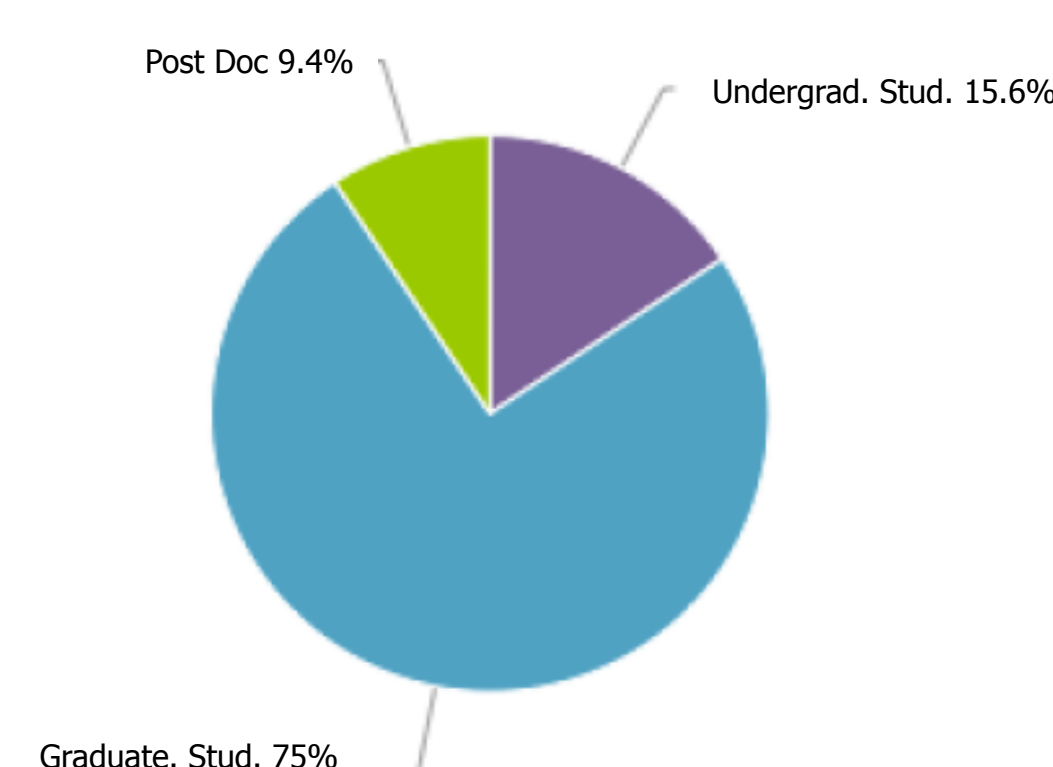
- CMS Software and computing tools

CMSDAS@LPC 2016

CMSDAS@Bari 2015

Short exercises

- A. b-tagging: 4 (12.50%)
- B. MVA analysis: 3 (9.38%)
- C. RooFit/RooStat: 6 (18.75%)
- D. MET: 1 (3.12%)
- E. Tracking/Vertexing: 1 (3.12%)
- F. Trigger: 3 (9.38%)
- G. Generators: 6 (18.75%)
- H. Muons: 3 (9.38%)
- I. Electrons: 2 (6.25%)

Results from surveys

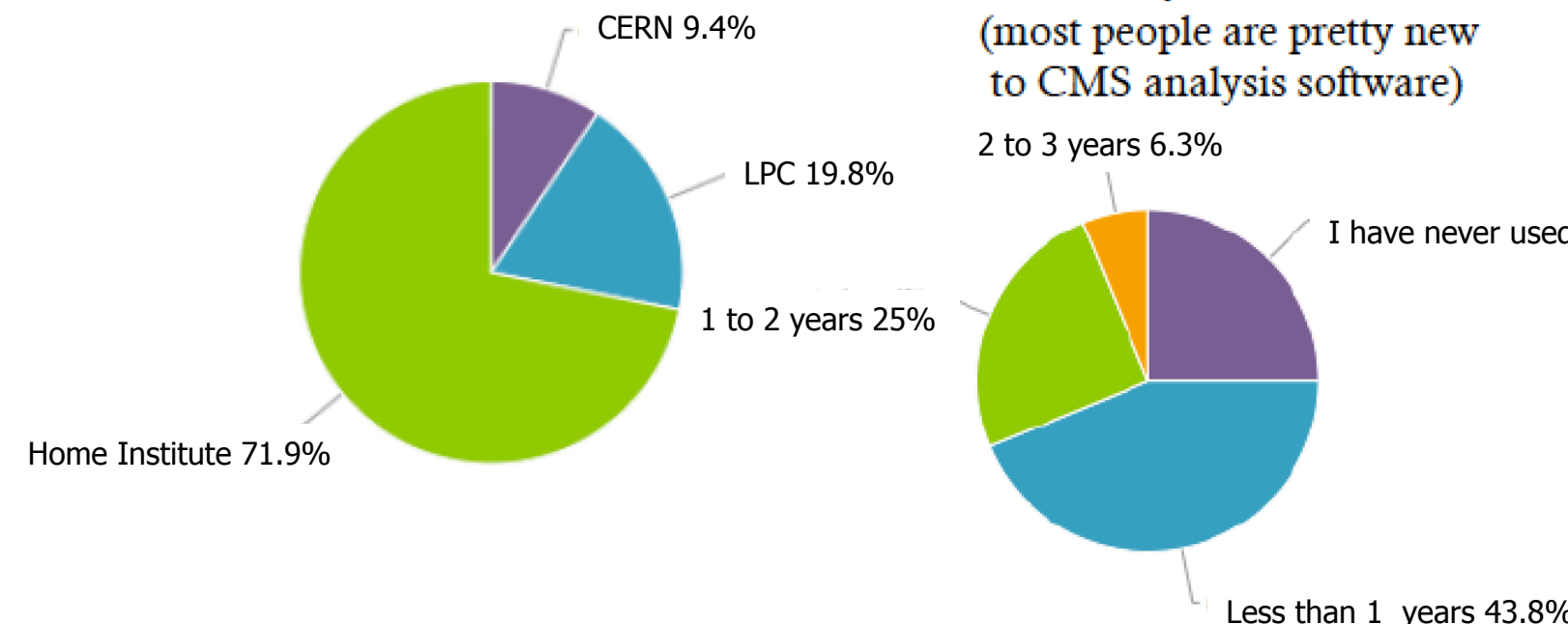Participation level (mostly grad students)
- Post Doc 9.4%
- Undergrad. Stud. 15.6%
- Graduate. Stud. 75%

Primary research location
- CERN 9.4%
- LPC 19.8%
- 1 to 2 years 25%
- Home Institute 71.9%

Years of experience using CMS analysis software: (most people are pretty new to CMS analysis software)
- 2 to 3 years 6.3%
- I have never used CMS software before
- Less than 1 years 43.8%
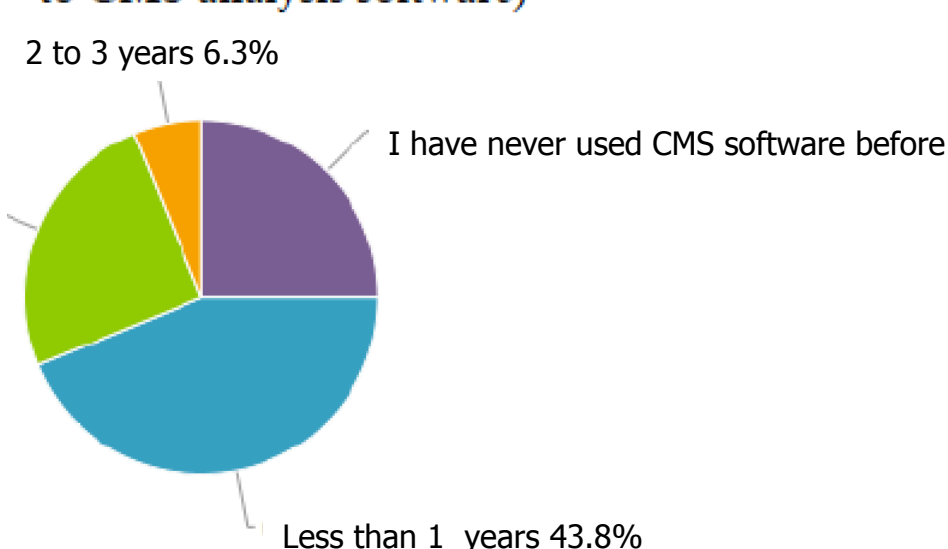
CMSDAS enabled to make new connections? 100% yes

CMSDAS has been a valuable experience? 100% yes

I would recommend CMSDAS to a colleague

I would attend another CMSDAS as a student

I would serve as a facilitator

Time duration of the school. Is that right ?
- A. too long: 0 (0.00%)
- B. too short: 12 (37.50%)
- C. just right: 20 (62.50%)

Long exercises
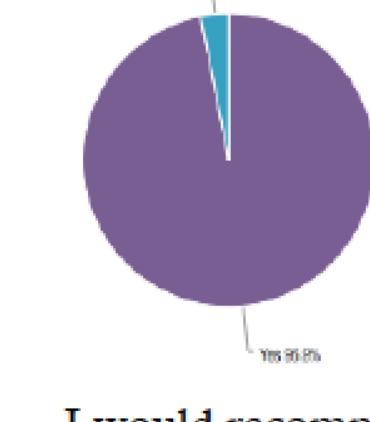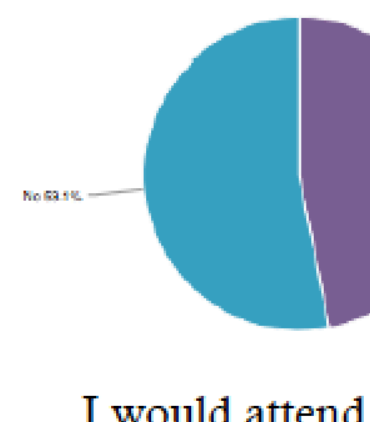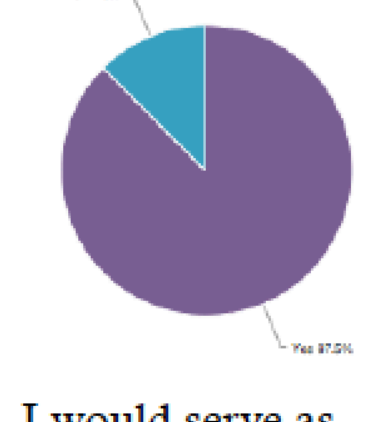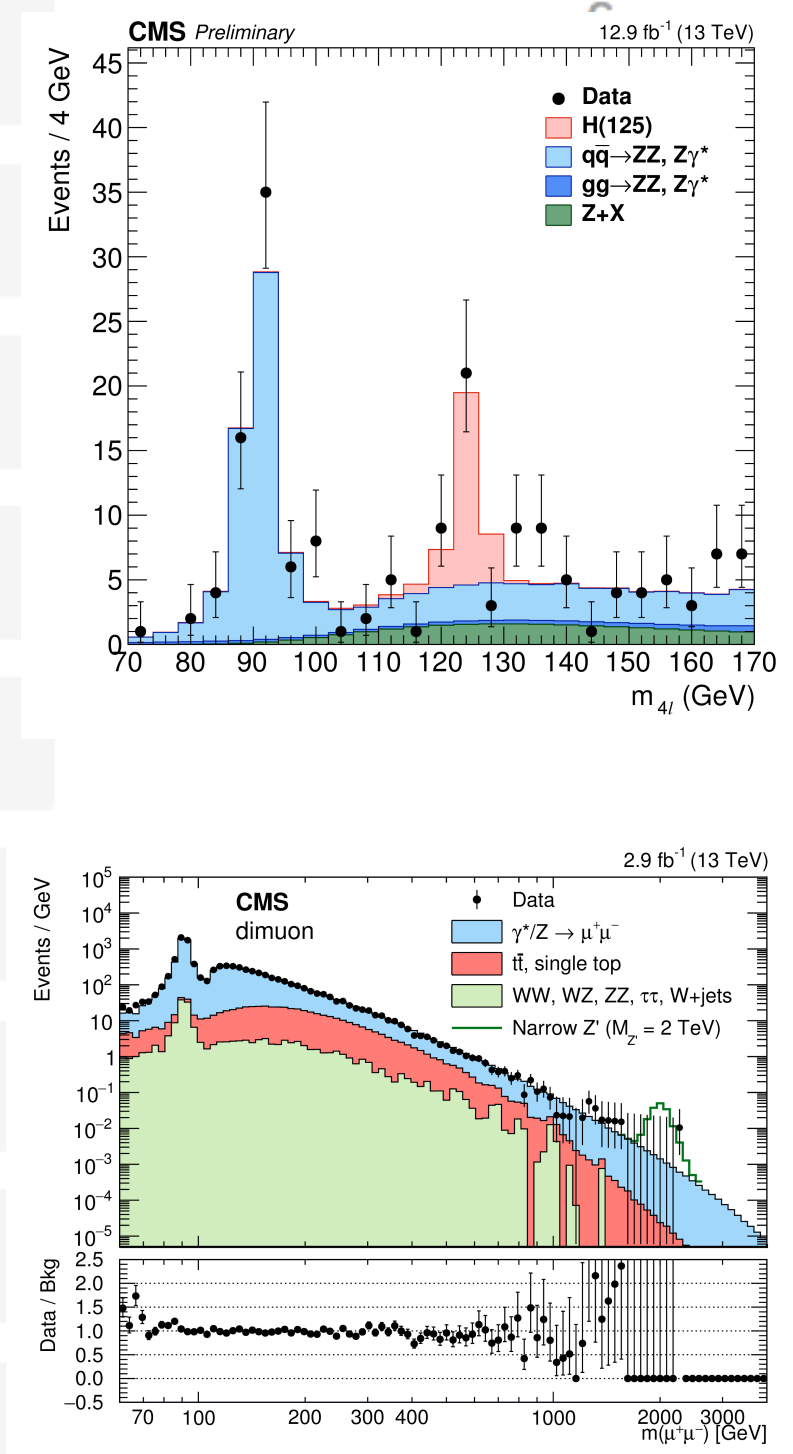
- A. H --> bbar: 3 (9.38%)
- B. X --> hh --> bbtautau: 4 (12.50%)
- C. ttH, H --> bb: 4 (12.50%)
- D. DM via H --> invisible: 1 (3.12%)
- E. Z' --> mumu: 4 (12.50%)
- F. ttbar cross section: 4 (12.50%)
- G. SUSY hadronic: 3 (9.38%)
- H. H --> ZZ --> 4l: 2 (6.25%)
- I. DM via monojet: 3 (9.38%)

Results and conclusions:

- The analyses teams at the school were able to:
  - reproduce the latest public results by the collaboration
  - **extend** the state of the art on CMS either by using a larger dataset or by modifying selection to improve sensitivity
- About **1000** users trained so far at **15** schools and more to follow worldwide
- After the CMSDAS training the students jumped immediately to start on their own physics analysis which otherwise it would take months of preparation

→ The CMSDAS training program is a **successful** model, proven to be a key for the preparation of new and young physicists

Best Analysis Team prize