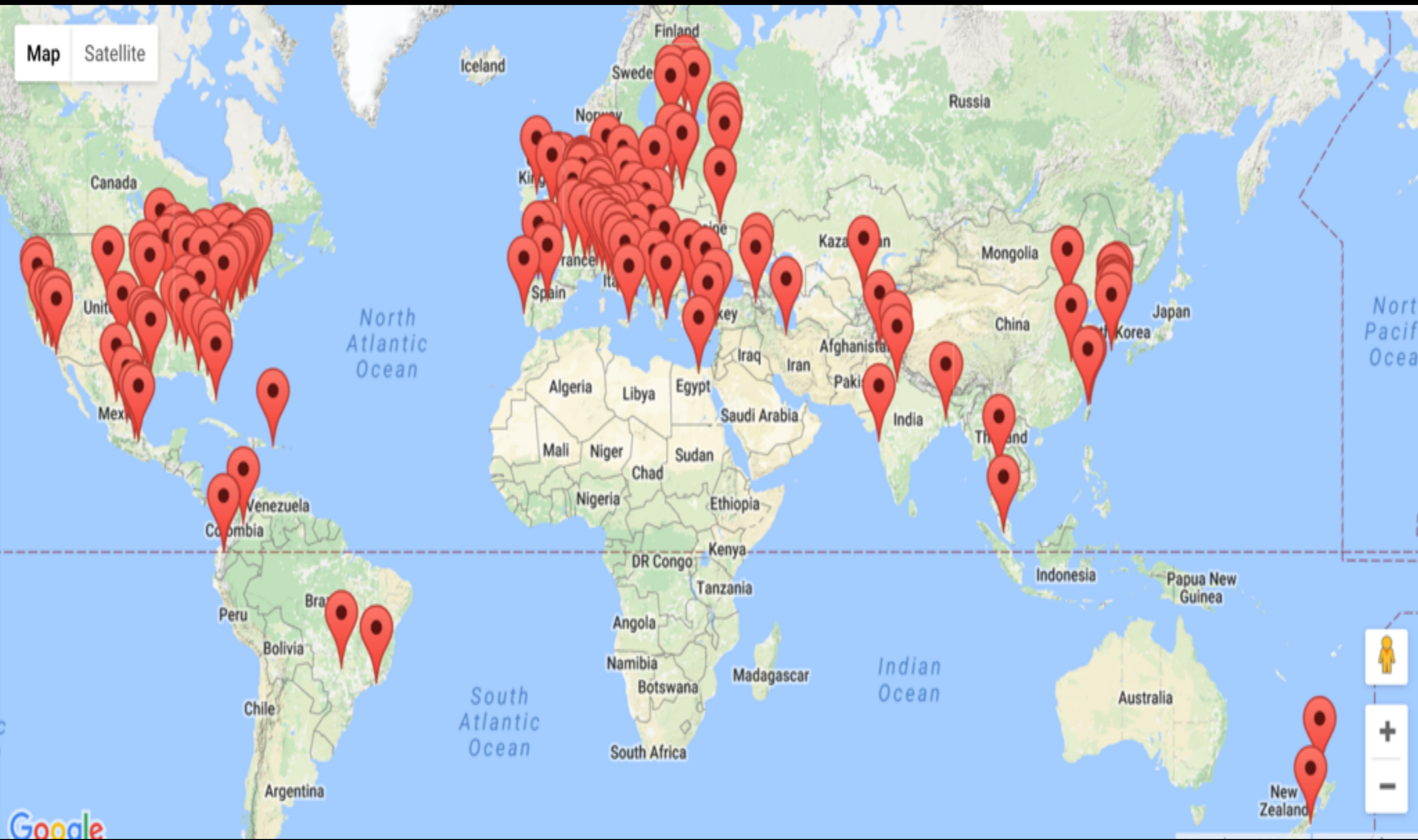
Physics Analysis Training Model at the CMS Experiment

Sudhir Malik
Physics Department
University of Puerto Rico Mayagüez

- Sun never sets on CMS collaboration
- 40 countries

- Sun never sets on CMS collaboration
- 40 countries
- 200 institutes



- Sun never sets on CMS collaboration
- 40 countries
- 200 institutes
- 4000 physicists

New Paradigm for LHC Users

- Collaboration size ~ 5 times the Tevatron experiments
  - Requires huge resources, money, manpower
  - Long life span of the experiment ~ 30 years
  - Enormous data rate - 10,000 copies/sec of *Encyclopaedia Britannica*
- Users located worldwide
  - Significant numbers of scientists not co-located at CERN
    - Located individual institutions or grouped regionally
  - Possible financial and logistic constraints to not to be at CERN
  - Larger distances and large time differences
- Highly distributed environment for
  - Computing (Grid)
  - Physics analysis

Challenges

- Engage the collaboration discovery potential and maximize physics output
- How to quickly come up to a level to contribute to physics
  - CMS has a complex computing environment
  - The tools to do physics are non-trivial
  - Most of time the problem is not lack of information but to find and access it
- Users come from different backgrounds
  - Language and culture
  - Varied Physics and Computing skills
  - Know-how & facilities

Needs Organized Training

To meet the challenge CMS Experiment has invested early on in training its Users

CMSDAS Context: LHC & CMS

The LHC is a global phenomenon; the public is captivated, our global collaborations are seen as a model for the world. This is wonderful; and an opportunity for the entire field of particle physics.

CMS is striving to nurture and increase engagement of the myriad talents of CMS, in the development of physics, service, upgrade, education of those new to CMS and the career development of younger members.

CMSDAS: investing in the future

CMS will likely remain a major productive scientific collaboration for at least two decades. We will be colleagues for the majority of our careers.

Given this longevity, it is especially important to continue to develop a nurturing and sustainable collaboration where all can reach their full potential. CMS aspires to be a harmonious family.

The future of the collaboration depends on its youngest members. It is critical that CMS provide effective collaboration-scale opportunities to aid the development of young physicists in all aspects of experimental physics: training in data analysis, software and hardware, and to provide career development paths.

A PRIME EXAMPLE OF COLLABORATION SCALE EDUCATION IS THE CMS DATA ANALYSIS SCHOOL (CMSDAS)

- 5 day school
  - 10% lectures
  - 90% hands on
- Analysis of collision data
- >1000 students trained
- 1st school 2010 at Fermilab LPC
- 16 schools held in 7 years
  - Fermilab, CERN, Pisa, Bari, DESY, Taipei, Seoul, Kolkata
- Genesis: JTERM physics school
  - at FNAL LPC (2006-2009) 90%
  - lectures (prior to data taking)

http://lpc.fnal.gov/programs/schools-workshops/cmsdas.shtml

"Tell me and I forget,
show me and I remember,
involve me and I understand."

CMSDAS is a Collaboration between the students and teachers

CMSDAS Format

Pre-school exercises
t = -1 month

1-week hands on physics analysis exercises

Lectures
(0.5 days)

Short Exercises
(2 days)
Objects & techniques

Long Exercises
(2.5 days)
Physics measurements

(All exercsies use CMS collision data)

Pre-school exercises

(1 month before school)

Remote participation

Four sets of exercises on espace

learn basics of:

CMS Software, account access, access to code & data, run grid jobs, PAT basics, python, ROOT, distribution fitting, GitHub

Students use personal laptop (as they will @ school

Answer submission via web-forms

Answers checked and feedback given to students

Students hit the ground running @ CMSDAS

Lectures (0.5 days)

Context of CMS/LHC

LHC the machine

CMS detector

CMS physics,

Analysis design

CMS software, analysis tools

08:00 - 09:00 Registration and Tutorial Preparation *1h0'* (outside One West (WH))

09:00 - 12:40 Plenary

Location: WH One West

Material: EVO url

09:00 **WELCOME** *10'*

Speakers: Pier Oddone (Director of Fermilab), Young-Kee Kim (Associate Director of Fermilab)

09:10 **Introduction and Goals of the CMS Data Analysis School** *30'*

Speakers: Ian Shipsey, Cavanaugh Rick (UIC/FNAL), Sudhir Malik

Material: Slides

09:40 **The Really Big Picture** *50'*

Speaker: Joe Lykken (Fermilab)

Material: Slides

10:30 **Break - Group Photo in Atrium Followed by Coffee Break** *30'* (outside One West)

11:00 **The LHC in Context and the Current Status** *40'*

Speaker: Prebys Eric

Material: Slides

11:40 **Introduction to CMS Software, Computing Model and Analysis Tools** *30'*

Speaker: Sudhir Malik

Material: Slides

12:10 **What Constitutes a Full Analysis - CMS Exa**

Speaker: Salvatore Rappoccio

Material: Slides

y, 11 January 2012

08:45 - 10:45 Plenary

Location: WH One West

Material: EVO url

08:45 **The CMS Detector Present and Future** *1h0'*

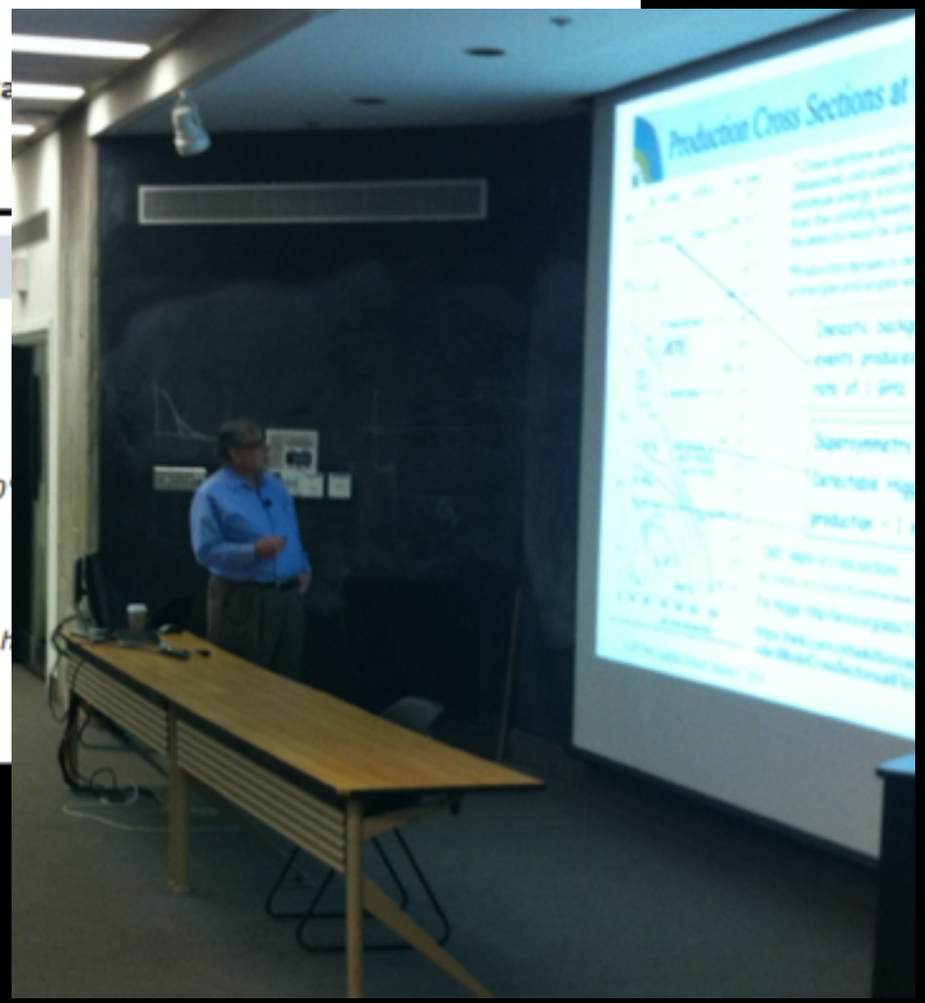
Speaker: Joel Butler

Material: Slides

09:45 **Physics in CMS - Now and in the Future** *1h*

Speaker: Chris Hill

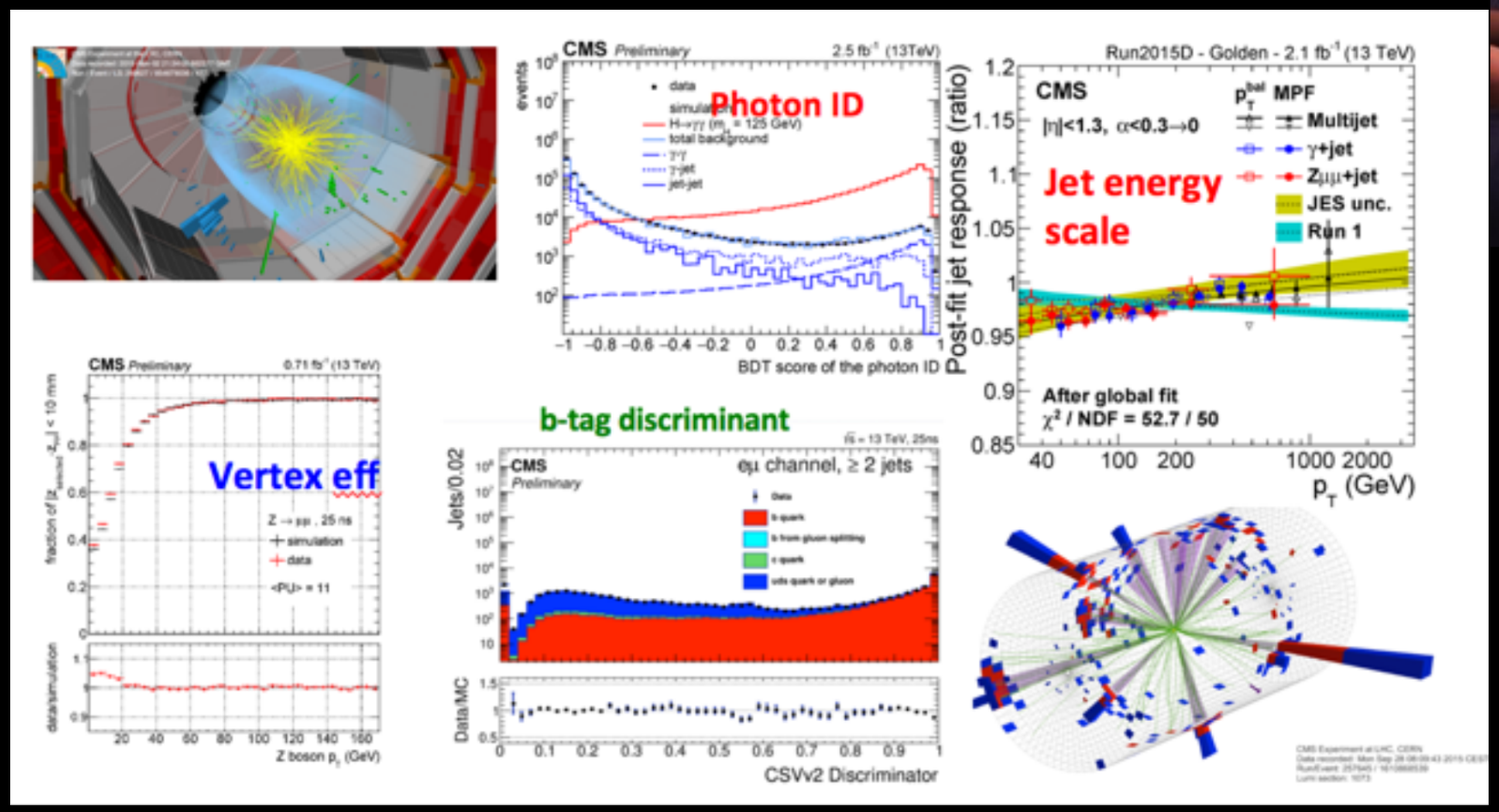
Material: Slides



Short Exercises (2 days)

11 topics: Roostats, Generators Tracking & vertexing, Electrons, Muons, Photons, Jets, b-tagging, particle flow, pile-up, event visualisation

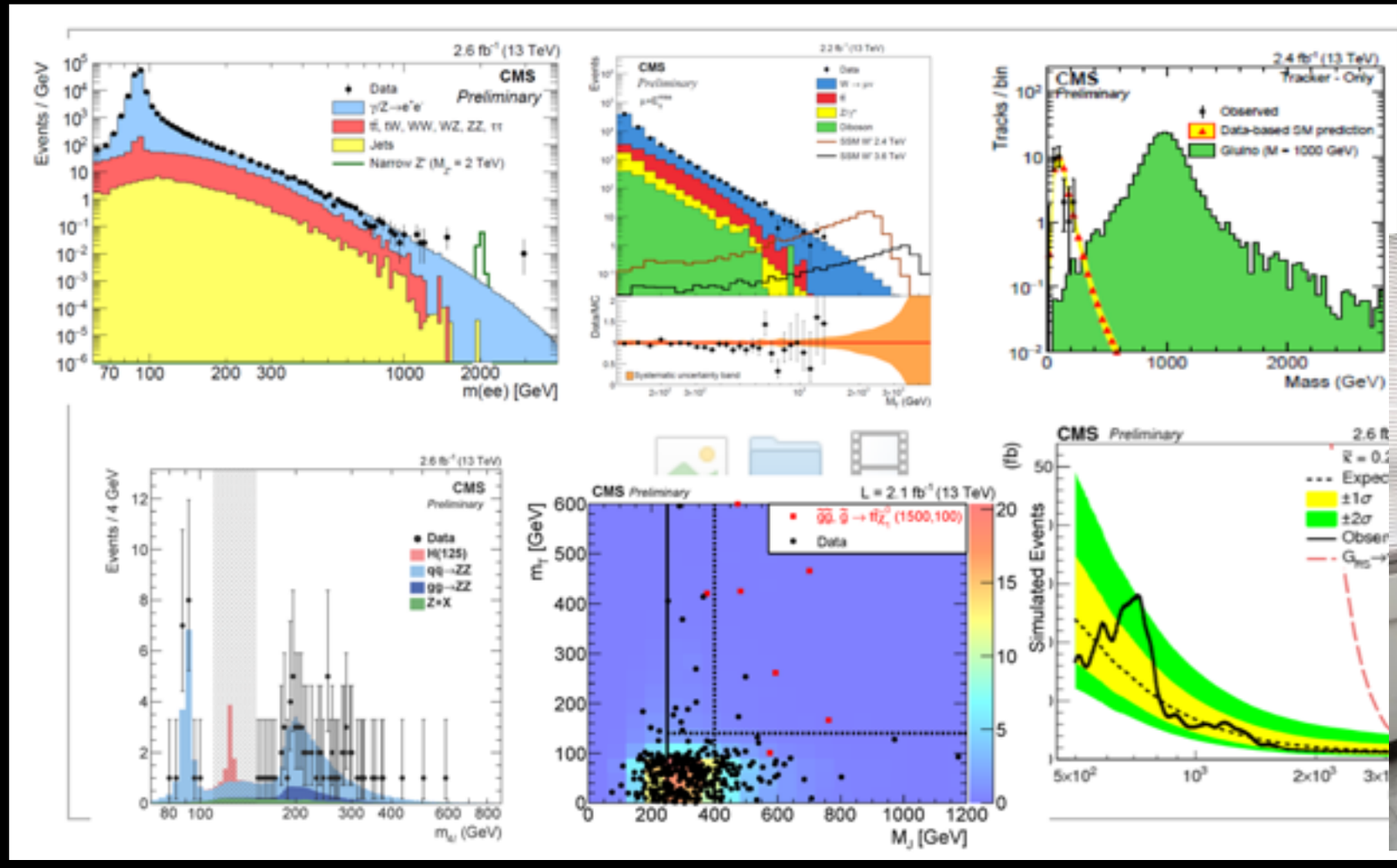
Students sign up for 6 short exercises prior to school: webform
Exercises chosen depend on both student interest & satisfying prerequisites for the physics measurements in the second half of the school
Each exercise ~2hrs
Each exercise on a CMS twiki, answer submission via espace
Students work individually in a class-like setting
Typical class size 6-8 with 3 teachers



Long Exercises (2.5 days)

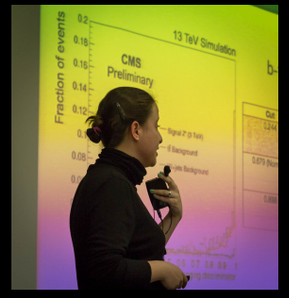
students perform one at the shool, signing-up pre-school via webform 6-8 choices per school

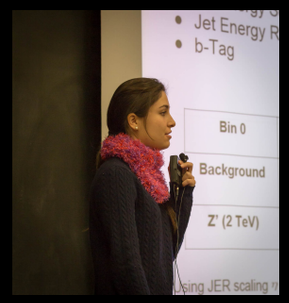
For each long exercise students work in teams that number 5-10 with 3-4 teachers
Teachers are leaders of the corresponding CMS measurements
A full measurement is performed: search or cross section,
Typically including:
the determination of trigger & object efficiencies and acceptance.
Fits to the data to extract signal and background yields
Methods of data driven background determination
Measurement of a range of systematic uncertainties

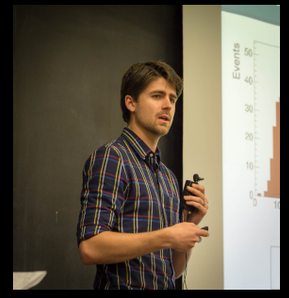


Mini-symposium

The student groups present their work and compete for the "first prize" judged by panel of senior CMS physicists







Feedback

Survey post-school: what students say:

It was a very good experience and I do not regret having travelled all this distance to attend. It was really worth it and I also had a great boost in motivation for my analysis back home.

It allowed me to see that there are a lot more tools out there that I was not taking advantage of…….

An opportunity to learn the new software from the experts while meeting fellow beginners along the way.

The encouragement to learn more.

intense, fun, bar is set quite high

This school is probably the best CMS event of the year !

a Lot of Other Hands on Training

Physics Analysis Toolkit (PAT) Tutorial

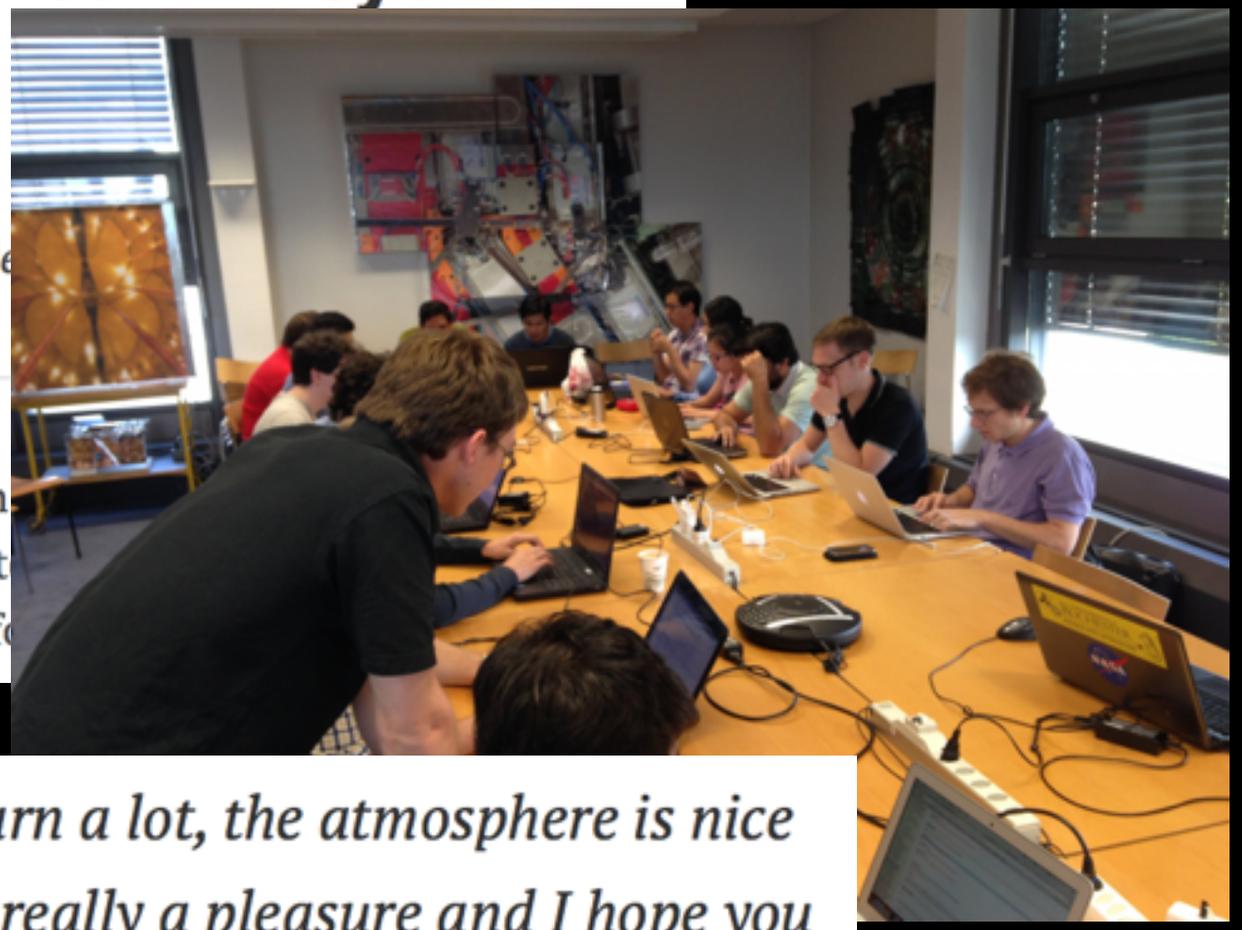
5-day Tutorials 15 in total since 2009, at CERN
trained > 500 students

The gateway to CMS physics analysis

Achintya Rao on Thu, 31/07/2014 - 11:50

Guest post by Sudhir Malik, University of Puerto Rico at Mayagüe

The CMS Physics Analysis Toolkit (PAT) tutorial, the fifteenth
to 5 July, has come a long way. Started January 2009, these tut
CMS physicists to use CMSSW, the CMS software framework f



The PAT tutorial was really helpful. We learn a lot, the atmosphere is nice and the team you build is excellent. It was really a pleasure and I hope you will continue to provide such opportunities to the collaboration.

http://cylindricalonion.web.cern.ch/blog/201407/gateway-cms-physics-analysis

FNAL Software School

4-8 August 2014
Fermilab, LPC
US/Central timezone

Search

Day 0 pre course exercise: introduction

Short description: The exercise involves compiling and running an example programs that generates detector data and a second program that reads the generated data. The students should then modify a Modules to histogram track, hit and strip information. The students must complete the exercise before of class.

Input: StripSet, GenHitSet, and GenTrackSet which contain the generated track data and the simulate the detector to that data.

Output: Histograms of data object proper

Day 1 course exercise: Cluster and hit findi

Description: This exercise involves cluster f StripSet object.

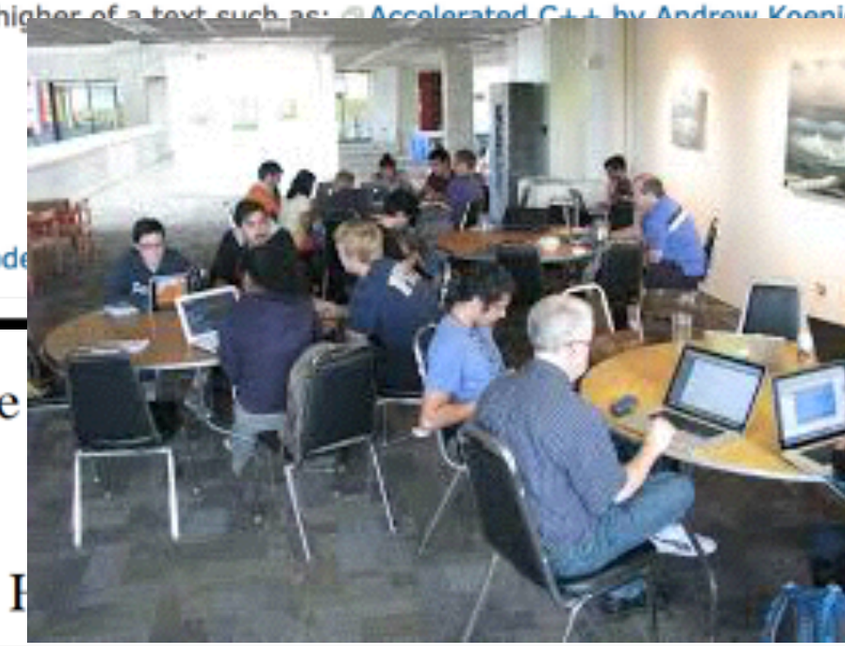
a) Cluster finding
Cluster finding algorithm: The students will above threshold adc value on each layer of t possible more than one charged track, when

Course prerequisites

- Students should have experience in C++ programing.
- Students should have familiarity with the process of c
- Students should have familiarity with the physics of charged particle trajectories in magnetic fields
- Students must complete the pre course exercise. (available starting July 28th)
- Students should have knowledge of C++ at the level or higher of a text such as: Accelerated C++ by Andrew Koenig and Barbara Moo

Documents and course software

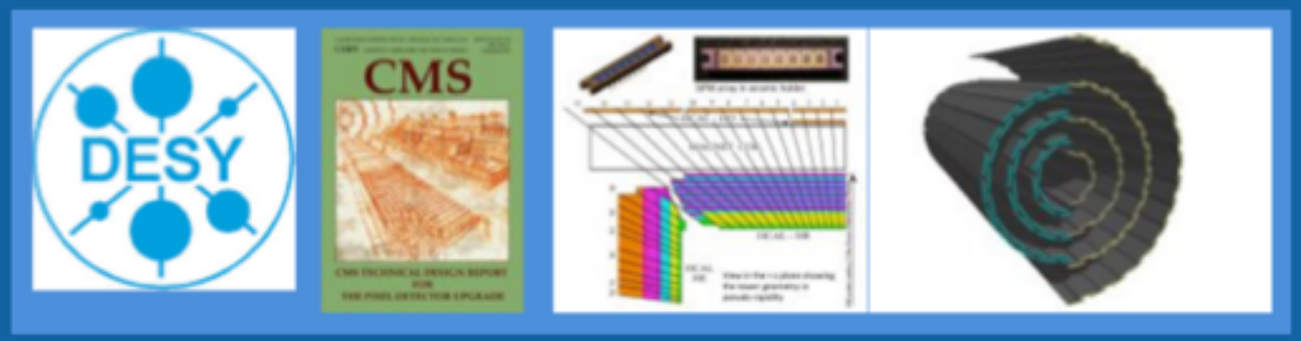
- Course Syllabus.
- Coding Standards document for the FNAL C++ course.
- Software school code repository: GitHub FNALComp code



Coding: The students use and modify HitCompareModule to produce s.

a) Hit performance metrics:
Input: The generated and reconstructed Hit Sets are used as input to H

https://cdcvs.fnal.gov/redmine/projects/fnal-soft-school-summer-2014



CUPS - CMS Upgrade School

17-21 November 2014
DESY
Europe/Zurich timezone

Search...

CUPS is a hands on learning experience. It will introduce students, post docs, and new faculty and scientists to our detector and how to care for it, and to help to design the detector for Phase II. In this inaugural edition of CUPs participants will have an opportunity to gain a deeper understanding of the physics motivation for the HL-LHC and the CMS upgrade. There will be in depth presentations of the Tracker and Muon Systems including the detector technologies (silicon sensors, and gaseous detectors) and the associated electronics and mechanics, an introduction to DAQ and μTCA, and system layout.

Participants will have the opportunity to understand, analyze, and work with:
- Test-beam data (resolution, efficiency, etc.)
- Sensor characterization data (charge collection efficiency, etc.)
- Thermal and mechanical lab measurements on test structures
- A portable telescope for muon detectors
- Tuning operational parameters of gas detectors
- A test DAQ system (including setting it up)
- Design a tracker or muon system

Hands-on Advance Tutorial Session Only at Fermilab LPC

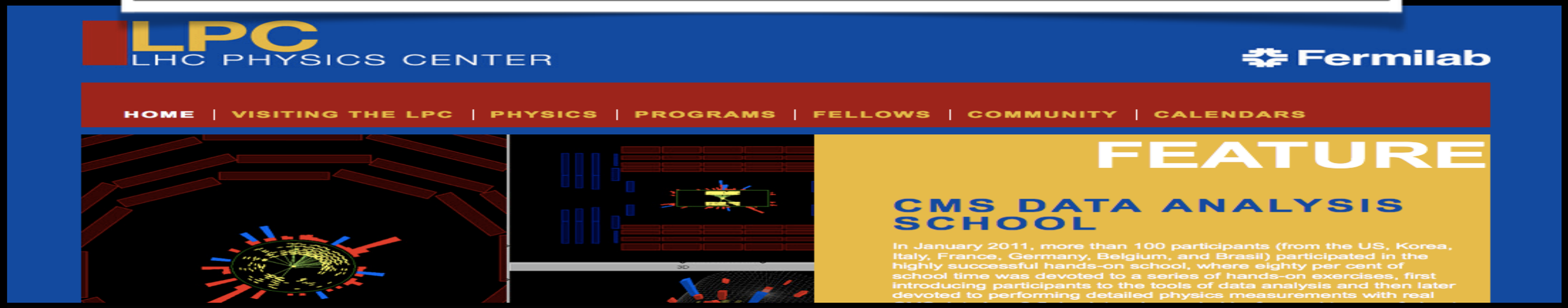
2017 HATS@LPC (and lectures)

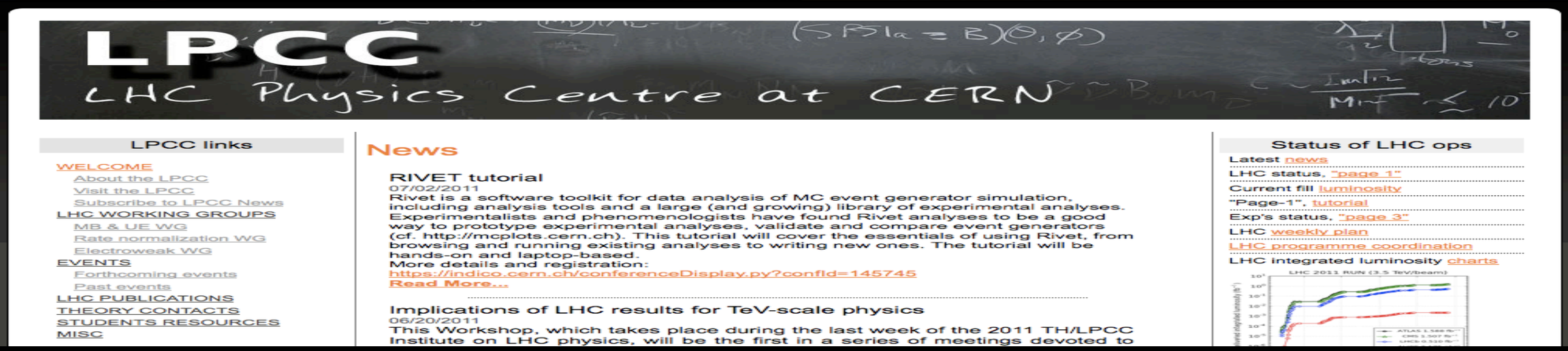
- HCAL HATS@LPC (May 22, 2017) - Indico Agenda
- Machine Learning HATS@LPC (Session II) (May 16, 2017) - Indico Agenda
- Machine Learning HATS@LPC (Session I) (May 15, 2017) - Indico Agenda
- HEP analysis in the Python ecosystem HATS@LPC (Apr 27, 2017) - Indico Agenda
- Histogrammar HATS@LPC (Apr 26, 2017) - Indico Agenda
- PyRoot and rootpy HATS@LPC (Apr 19, 2017) - Indico Agenda

2016 HATS@LPC (and lectures)

- BTagging HATS@LPC (Aug 2, 2016) - Indico Agenda
- Tracking and Vertexing HATS@LPC (July 27, 2016) - Indico Agenda
- Jets Energy Corrections and Pile-Up Mitigation HATS@LPC (June 30, 2016) - Indico Agenda
- Jets Algorithms and Substructure HATS@LPC (June 29, 2016) - Indico Agenda
- Muon HATS@LPC (June 28, 2016) - Indico Agenda
- Tau HATS@LPC (June 27, 2016) - Indico Agenda
- Electron/Photon HATS@LPC (Session II) (June 10, 2016) - Indico Agenda
- Multivariate Analysis Techniques HATS@LPC (June 9, 2016) - Indico Agenda
- CMS 101 (Lecture) (June 9, 2016) - Indico Agenda
- Electron/Photon HATS@LPC (Session I) (June 8, 2016) - Indico Agenda
- The LHC (Lecture) (June 8, 2016) - Indico Agenda
- Trigger HATS@LPC (June 7, 2016) - Indico Agenda
- Git/GitHub HATS@LPC (June 6, 2016) - Indico Agenda
- HCAL HATS@LPC (May 25, 2016) - Indico Agenda

LHC Physic Centers
Training Hubs and Centers of Excellence





Conclusion

- As CMS / LHC Experiments / HEP enters challenging times in future in terms of Software, Computing, Upgrade and with big and diverse collaborations, one cannot emphasize more the importance of training the user community
- The training will not only help meet challenges but also prepare the users for careers outside academia
- CMS training model has been very well noticed appreciated by our funding agency and can serve as good role model
- CMS training Model has succeeded due to support and willingness from hundreds of its users in training its younger generation
- Collaborative effort is the key to the success
- S2I2 (Scientific Software Innovation Institute) efforts can empower current and future workforce (HEP User community is part of it) by training them essential skills to use and develop software