USCMS PURSUE program (Program for Undergraduate Research SUmmer Experience)

Tribute to dear Meenakshi

on behalf of all Organizers, Funders, Mentors, Facilitators, Speakers, and Supporters of USCMS PURSUE

See also USCMS PURSUE Snowmass paper - https://arxiv.org/abs/2209.10109

Goals and Overview

- Annual program, first edition in 2022
- Reflects USCMS commitment to diversity and inclusion
- Maps to the USCMS DEI Action Plan “Identity formation and community engagement”.
- Offers undergraduate students an opportunity to perform a project under the mentorship of scientists working at the frontier of Physics.
- Students learn and use computational tools and data-science methods to know more about fundamental particles and their interactions by analyzing data obtained from the CMS
- Our aim is to strengthen our research by building an inclusive and diverse community with a wide range of perspectives
- In addition, the research internships will possibly help and encourage the students to persist in a STEM major through college and train them in skills needed for a future career in the STEM workforce.
- This is recruitment to our PhD programs

Composition of Program

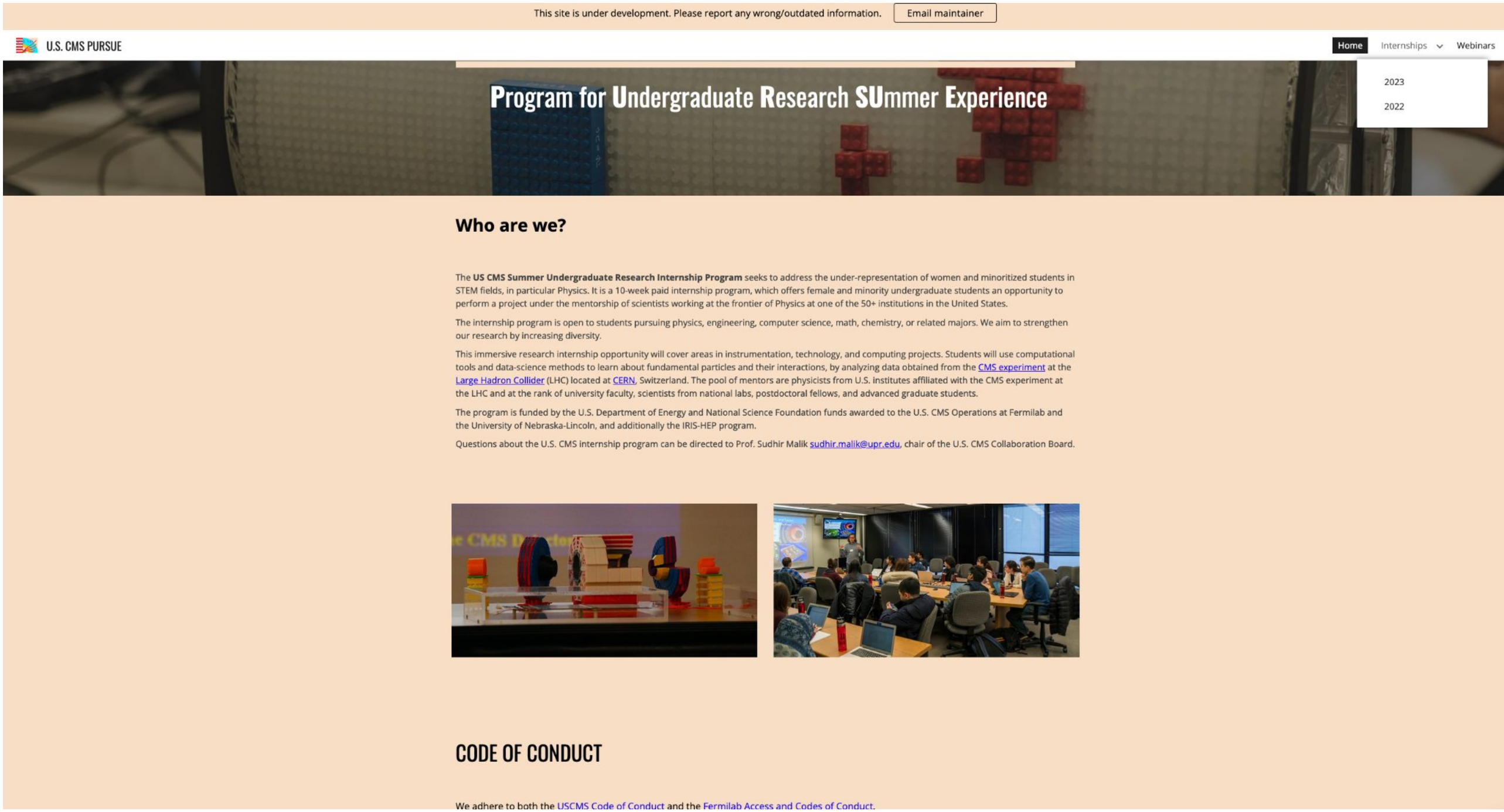
- **10 weeks (June - Aug)**
  - First edition in 2022 - virtual
  - This summer (2023) - in person
- **In 2022**
  - 16 students, predominantly from no HEP MSI, 3 from non MSI/R1
  - 9 females, including 2 white, 2 Asians
  - All 7 male students are minority
- **Over 100 applications received in 2022 (this year ~160 applicants)**
  - Interest / need is very high
  - Extensive rubrics to select students
- **Students have minimal background in physics (let alone particle physics**
- **Deliverables**
  - posters/ presentations
- **LPC helped in critical stuff like getting accounts, letting students use the LPC Tier-3 analysts facility. Some DRs- led the projects**

Program Structure

<table>
<tr><td colspan="5">U.S. CMS PURSUE (Program for Undergraduate Research SUmmer Experience)</td></tr>
<tr><td rowspan="3">Preparation</td><td colspan="4">Announcement/Call for the School
Application Period: January 31, 2022- February 28, 2022
Program Dates: Between June 6 – August 22, 2022
Acceptance Dates: March 31, 2022</td></tr>
<tr><td colspan="4">Internal Preparation (Projects assignments, mentor selection, schedule, speakers)</td></tr>
<tr><td colspan="4">Computer accounts setup/on-boarding</td></tr>
<tr><td rowspan="2">Internship</td><td>3-weeks</td><td>Hands-on Software and Computing Skills</td><td rowspan="2">Lectures
(2 per week)
○ Physics concepts
○ Detector techniques</td><td rowspan="2">Networking & Professional Development meetings
(1 per week)
○ Meet USCMS
■ Students
■ Scientists
■ Professors
■ Engineers/Techs
○ Job skills
○ Resume writing
○ DEI topics</td></tr>
<tr><td>7-weeks</td><td>Physics Analysis Projects</td></tr>
</table>

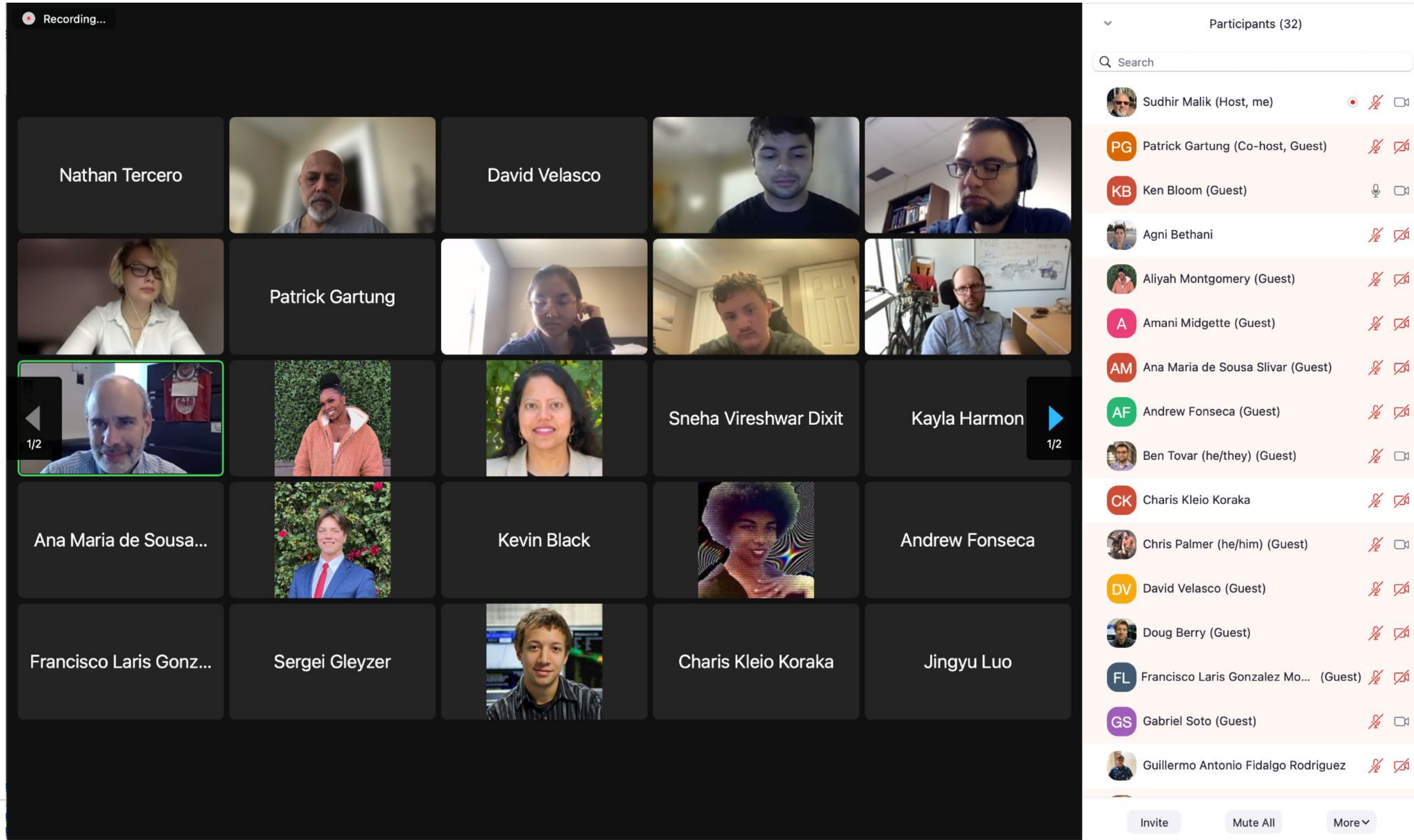
Website

Website:https://sites.google.com/upr.edu/uscms-pursue/home?authuser=0



USCMS - PURSUE Day 1 (2022)

Zoom Screenshot 6 June 2022



Incredible learning:Modules/Projects

- Hands-on Software skills in HEP, ~3 weeks
  - Unix, Git, Python, ROOT, Machine Learning, C++
- Research projects ~ 7 weeks
  - *EMTF (EndCap Muon Track Finder) BDT Training for Run-3*
  - *Building a GUI in Python for MaPSA testing*
  - *A New NANOAOD Era For 2015 CMS Open Data*
  - *Predicting Job Queue Time Using Machine Learning*
  - *Introducing GPUs For Electron And Photon Reconstruction*
  - *Searching For The Double Higgs Production*
  - *Software That Helps CMS Move Forward*
  - *Performance of Gas Electron Multiplier (GEM) Detector*
  - *Measuring The Efficiency Of A Higgs Decay To Tau Tau*
  - *Performance Benchmarks For Analysis Grand Challenge*
  - *Searching For The Extreme Events In Multilepton Data From the LHC*
  - *Comparison of Profiling Results for Run 3 and High Pileup LHC Simulation and Reconstruction*
  - *Setting Limits On The Toponium Cross Section*
  - *Dimensionality Reduction for Supersymmetry Dataset*
  - *Performance of reconstruction and identification of τ leptons decaying to hadrons and νΤ in pp (proton-proton) collision*
  - *Identifying b-quarks in the CMS detector*
- **26 talks** on several topics **by USCMS experts**
- **Posters** and **presentations** on research **projects by interns**

Poster example - Sneha Dixit

SNEHA VIRESHWAR DIXIT
sdixit2@washcoll.edu
MENTORS - NICK SMITH, OKSANA SHADURA

A NEW NANOAOD ERA FOR 2015 CMS OPEN DATA

CMS OPENDATA = PRIMARY DATA + SIMULATIONS RELEASED PUBLICLY BY THE CMS EXPERIMENT @ CERN, GENEVA

CMS OPENDATA IS USED BY: THEORISTS, GENERAL PUBLIC, NON CERN SCIENTISTS, STUDENTS

SCIENTISTS NEED CMS OPEN DATA THAT IS FAST AND EASY TO USE, BUT DATA FROM BEFORE 2016 IS IN A SLOWER DATATYPE

'NANOAOD' FILES ARE SMALLER + EASIER TO USE, BUT ALL CMS OPENDATA FROM BEFORE 2016 IS IN THE OLDER MINIAOD FORMAT

WORKFLOW

MINIAOD INPUT FILE (.root) → GENERATE CONFIG FILE (.py) → RUN CONFIG FILE → NANOAOD OUTPUT FILE (.root)

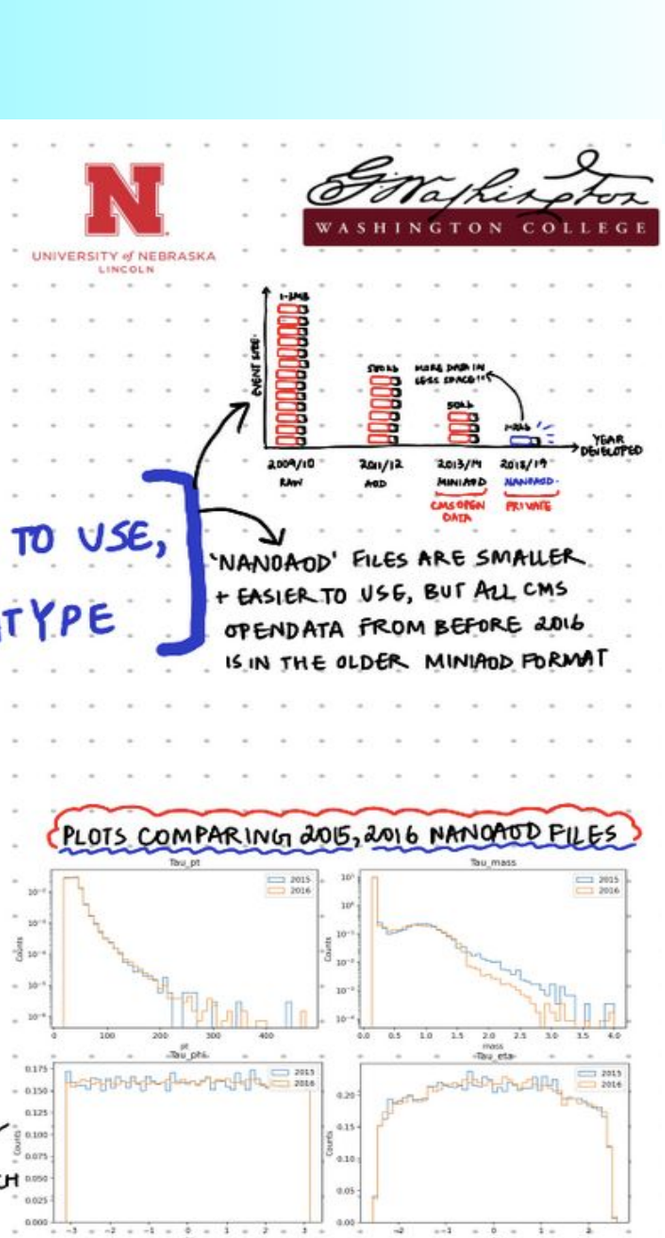
RUNNING CONFIG FILE PROCESSES EVENTS AND GENERATES NANOAOD FILE.

MODIFYING THE CONFIG FILE + CREATING A NEW ERA LETS US CONVERT 2015 MINIAOD TO NANOAOD

CONVERTING 2015 MINIAOD TO NANOAOD

SO, WE WROTE CODE TO CONVERT IT TO A FASTER DATATYPE

PLOTS COMPARING 2015, 2016 NANOAOD FILES



PLOTS LOOK SIMILAR, WHICH IS GOOD

MADE W/ COFFEA CASA

THESE 2 ARE UNEXPECTED VARIABLES HAVE CLUSTERED AROUND DEFAULT VALUES.

WHY DO WE USE ERAS?

- CUSTOMISES CONFIG FILE FOR DIFFERENT SCENARIOS
- USED TO MODIFY SETTINGS W/ SIMPLIFIED SYNTAX
- CHANGES HOW CONFIG FILE RUNS

```
era1.toModify(module, setting=blah)
era2.toModify(module, setting=blah)
```

```
(era1|era2).toModify(module, setting=blah)
```

```
_new_sequence = sequence.copy()
_new_sequence.remove(module1)
_new_sequence.remove(module2)
era.toReplaceWith(sequence, _new_sequence)
```

```
era.toReplaceWith(sequence,
sequence.copyAndExclude([module1, module2]))
```

GENERATING CONFIG FILE W/ CMSDRIVER

INPUT FILES USED

2015 MINIAOD FILE (DRELL-YAN SIMULATION)

root://eospublic.cern.ch//eos/opendata/cms/mc/RunIIFall15MiniAODv2/DYJetsToLL_M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/MINIAODSIM/PU25ns_Data2015v1_76X_mcRun2_asymptotic_v12_ext1-v1/10000/004544C8-6DD8-E511-9764-0026189438F6.root

2016 MINIAOD FILE (DRELL-YAN SIMULATION)

root://xcache//store/mc/RunIISummer16MiniAODv3/DYJetsToLL_M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/MINIAODSIM/PUMoriond17_94X_mcRun2_asymptotic_v3_ext2-v2/100000/2016.root

AIM - CREATE 2015 ERA WITHOUT AFFECTING 2016 ERA + COMPARE OUTPUT FILES TO ENSURE 2015 ERA GENERATES PROPER OUTPUT

NEED TO CONSULT OBJECT EXPERTS FOR FURTHER PROGRESS

MIGHT RELEASE CONFIG FILES AS IS

PULL REQUEST ON CMSSW 10_6_X

https://github.com/cms-sw/cmssw/pull/39040

2022 Mentors, Facilitators, Speakers, Supporters

Lothar and Ken - unwavering support and helping launch the program USCMS Operations funds

Tulika, Sudhir, Meenakshi - overall organizers

Guillermo Fidalgo - assisting the students and organizers everyday ,taking care of many daily details

Marco Mambelli, Bruno Coimbra, Guillermo Fidalgo, Julie Hogan, Charis Koraka for leading the computing trainings and professional development workshops

Bo Jayatilaka, Kevin Black and Marguerite Tonjes for LPC support, setting up computing accounts and environment

Don Lincoln for appearing as a featured “graduation” speaker

Mentors and facilitators	**Darin Acosta, Douglas Berry, Agni Bethani, Bruno Coimbra, Pieter Everaerts, Patrick Gartung, Sergei Gleyzer, Ulrich Heintz, Andy Jung, Charis Koraka, Carl Lundstedt, Jingyu Luo, Marco Mambelli, Scarlet Nordberg, Isabel Ojalvo, Christopher Palmer, John Rotter, Oksana Shadura, Benjamin Tovar, Fengwangdong Zhang**
Guest speakers and poster reviewers	**Todd Adams, Agni Bethani, Kevin Black, Johan Bonilla, Javier Duarte, Susan Dittmer, Jay Dittmann, Georgios Krintiras, Amit Lath, Daniel Li, Devin Mahon, Guenakh Mitselmakher, Abdollah Mohammadi, Alexx Perloff, Joseph Reichert, Indara Suarez, Farrah Simpson, Jieun Yoo**

Internship Certificate

CERTIFICATE OF COMPLETION

U.S.CMS PURSUE Internship

June 4 to August 22, 2022

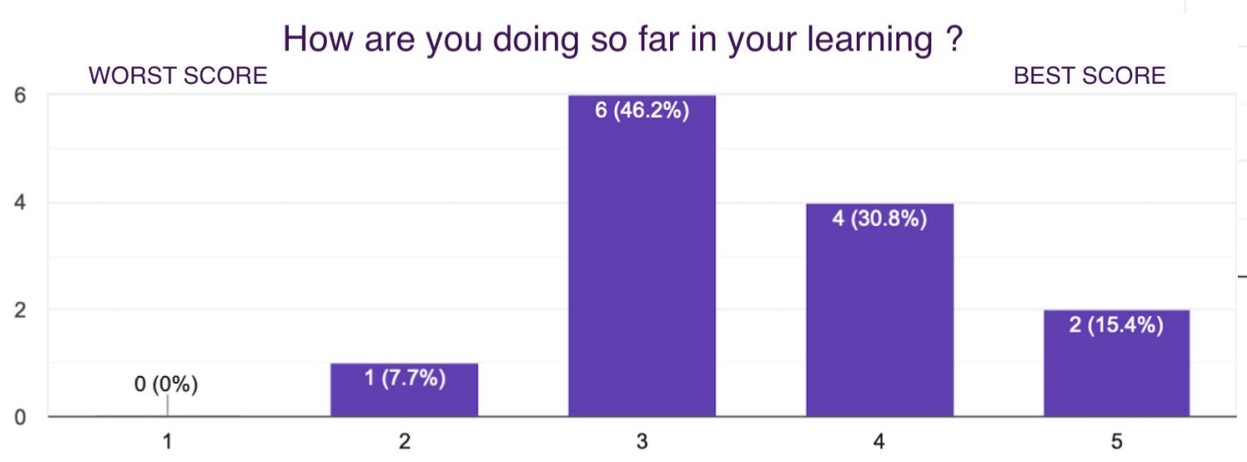
Student Name

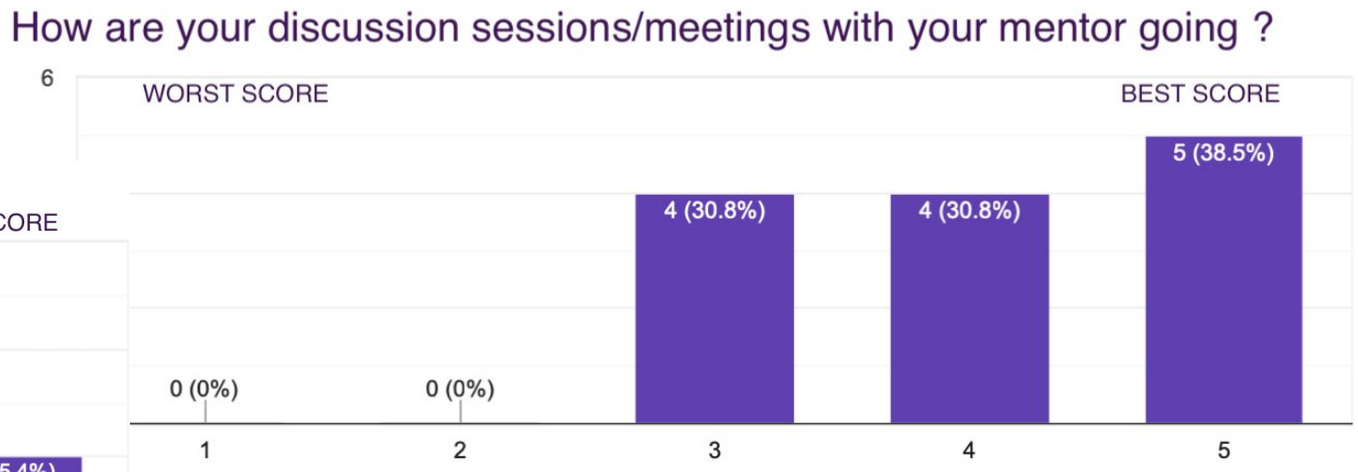
Title of Project

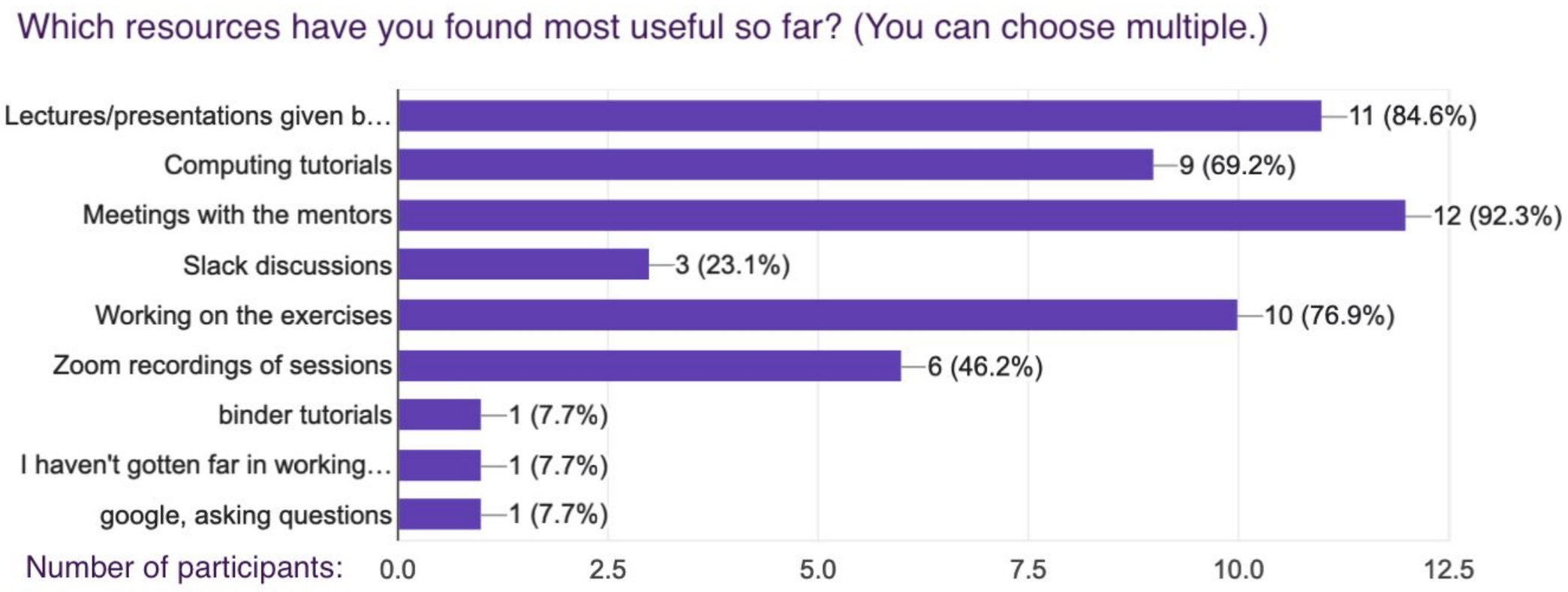
Organizing Committee

Dr. Lothar Bauerdick (Fermilab)	Prof. Kenneth Bloom (University of Nebraska)	Prof. Tulika Bose (University of Wisconsin-Madison)	Prof. Sudhir Malik (Univ. of Puerto-Rico, Mayaguez)	Prof. Meenakshi Narain (Brown University)
		TulikaBose	Sudhir Malik	Meenakshi Narain

Midterm Program Evaluation

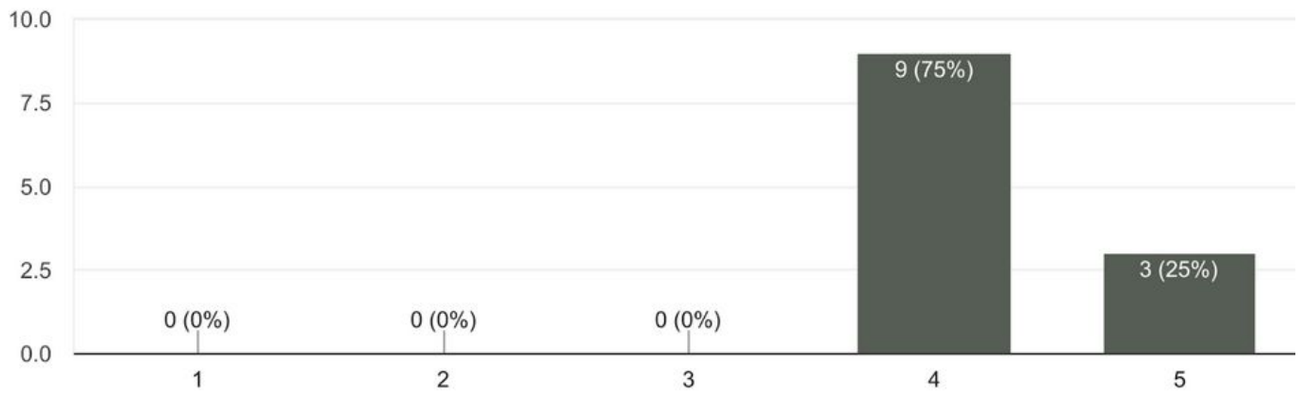




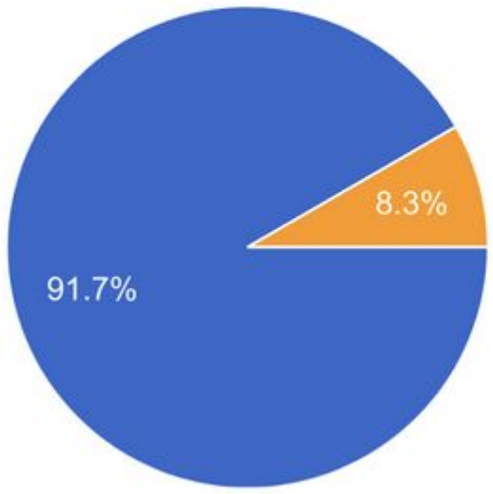


End of Program Evaluation

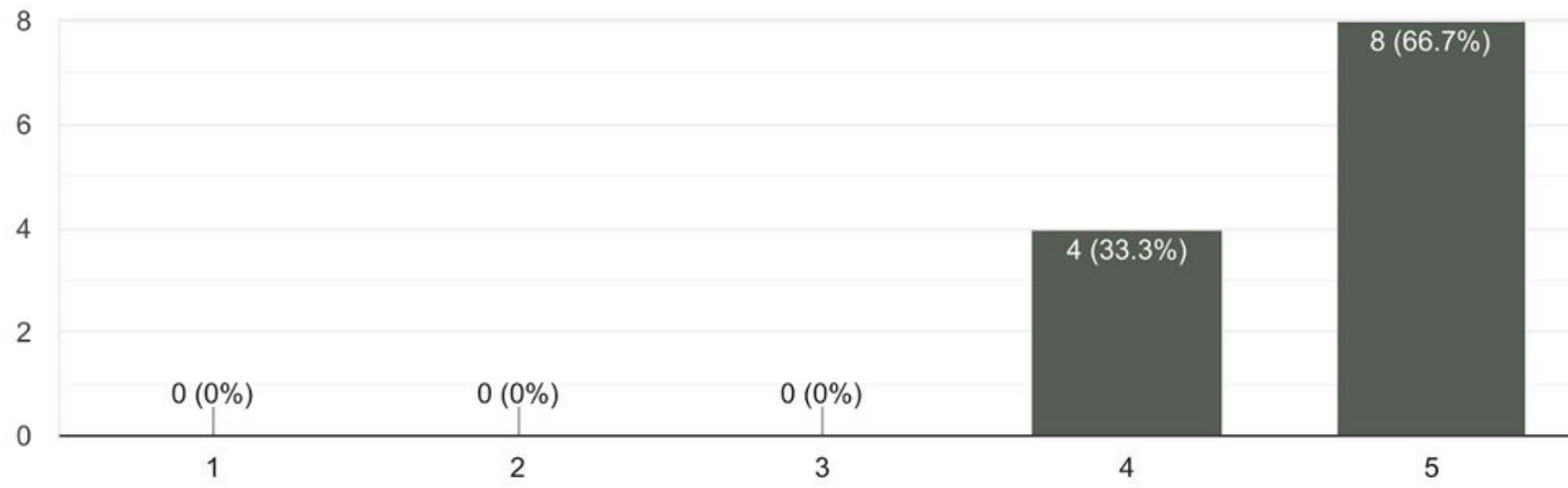
How would you rate your overall learning during the program?



Do you think this program has enhanced your skills towards a STEM job?

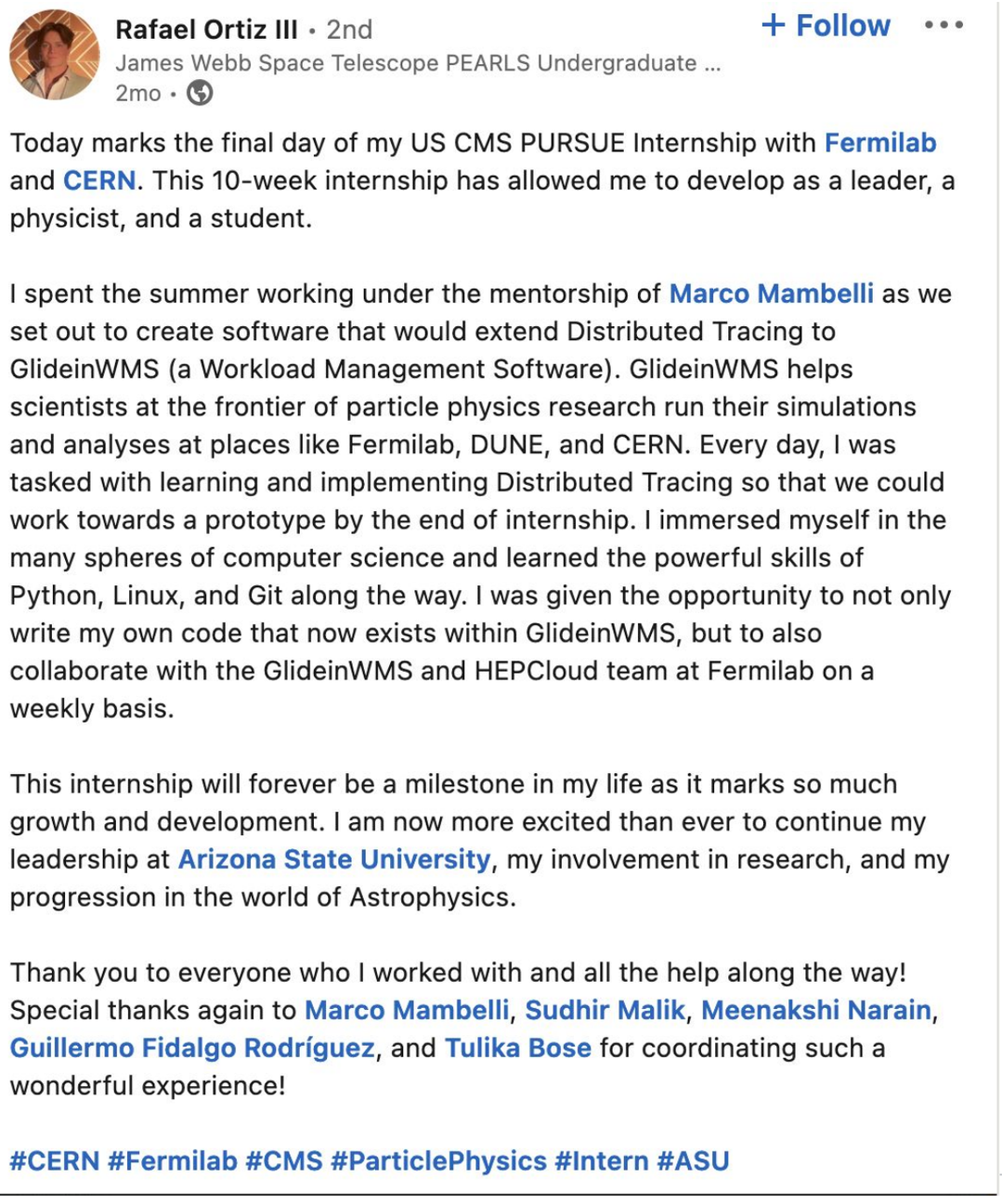


As a result of this my motivation to continue on a future STEM career is



Student testimonials

https://www.linkedin.com/posts/rafaelortiziii_cern-fermilab-cms-activity-6967545135278287872-IRpF?utm_source=share&utm_medium=member_desktop



Rafael Ortiz III • 2nd
James Webb Space Telescope PEARLS Undergraduate ...
2mo •

+ Follow

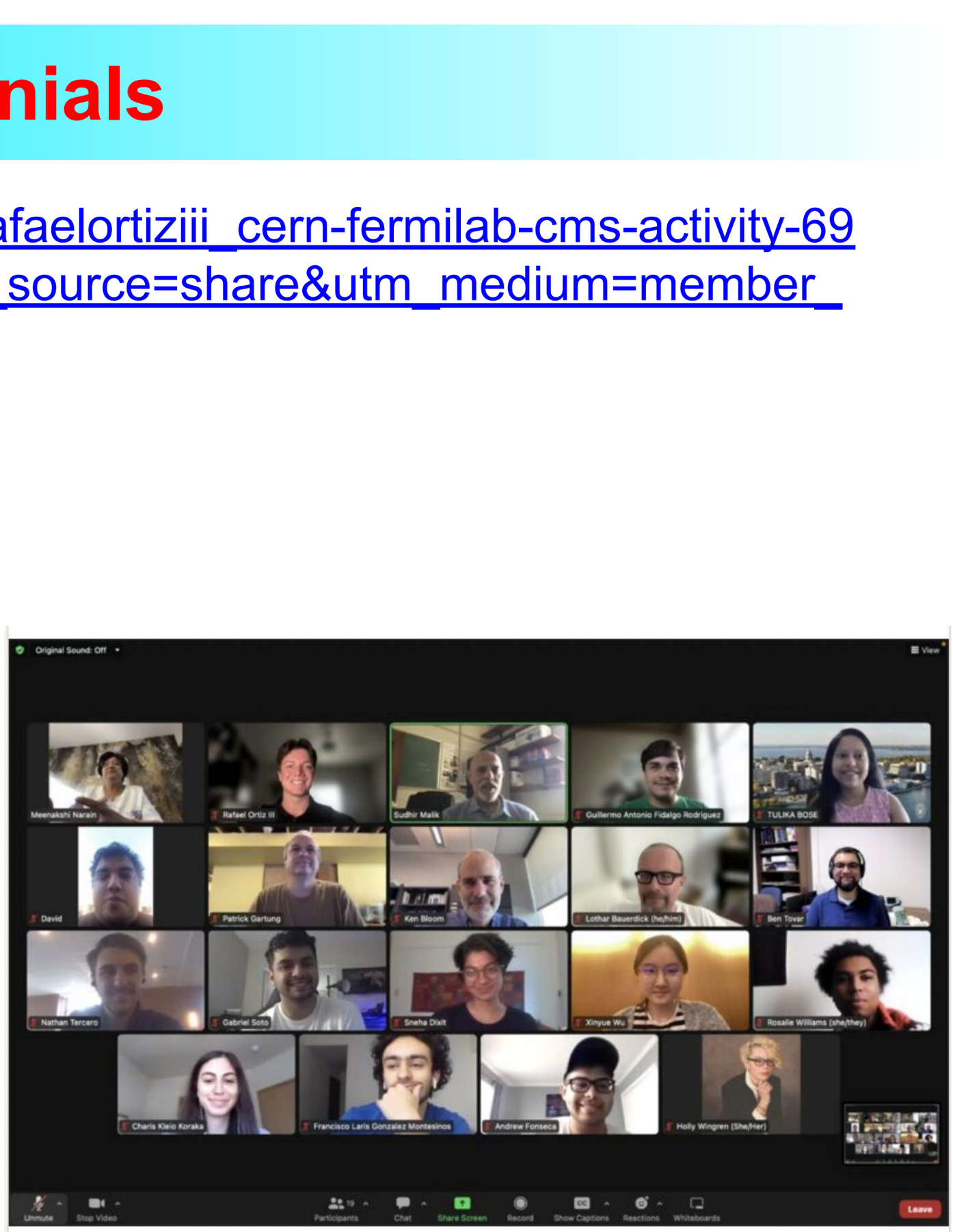
Today marks the final day of my US CMS PURSUE Internship with **Fermilab** and **CERN**. This 10-week internship has allowed me to develop as a leader, a physicist, and a student.

I spent the summer working under the mentorship of **Marco Mambelli** as we set out to create software that would extend Distributed Tracing to GlideinWMS (a Workload Management Software). GlideinWMS helps scientists at the frontier of particle physics research run their simulations and analyses at places like Fermilab, DUNE, and CERN. Every day, I was tasked with learning and implementing Distributed Tracing so that we could work towards a prototype by the end of internship. I immersed myself in the many spheres of computer science and learned the powerful skills of Python, Linux, and Git along the way. I was given the opportunity to not only write my own code that now exists within GlideinWMS, but to also collaborate with the GlideinWMS and HEPCloud team at Fermilab on a weekly basis.

This internship will forever be a milestone in my life as it marks so much growth and development. I am now more excited than ever to continue my leadership at **Arizona State University**, my involvement in research, and my progression in the world of Astrophysics.

Thank you to everyone who I worked with and all the help along the way! Special thanks again to **Marco Mambelli**, **Sudhir Malik**, **Meenakshi Narain**, **Guillermo Fidalgo Rodríguez**, and **Tulika Bose** for coordinating such a wonderful experience!

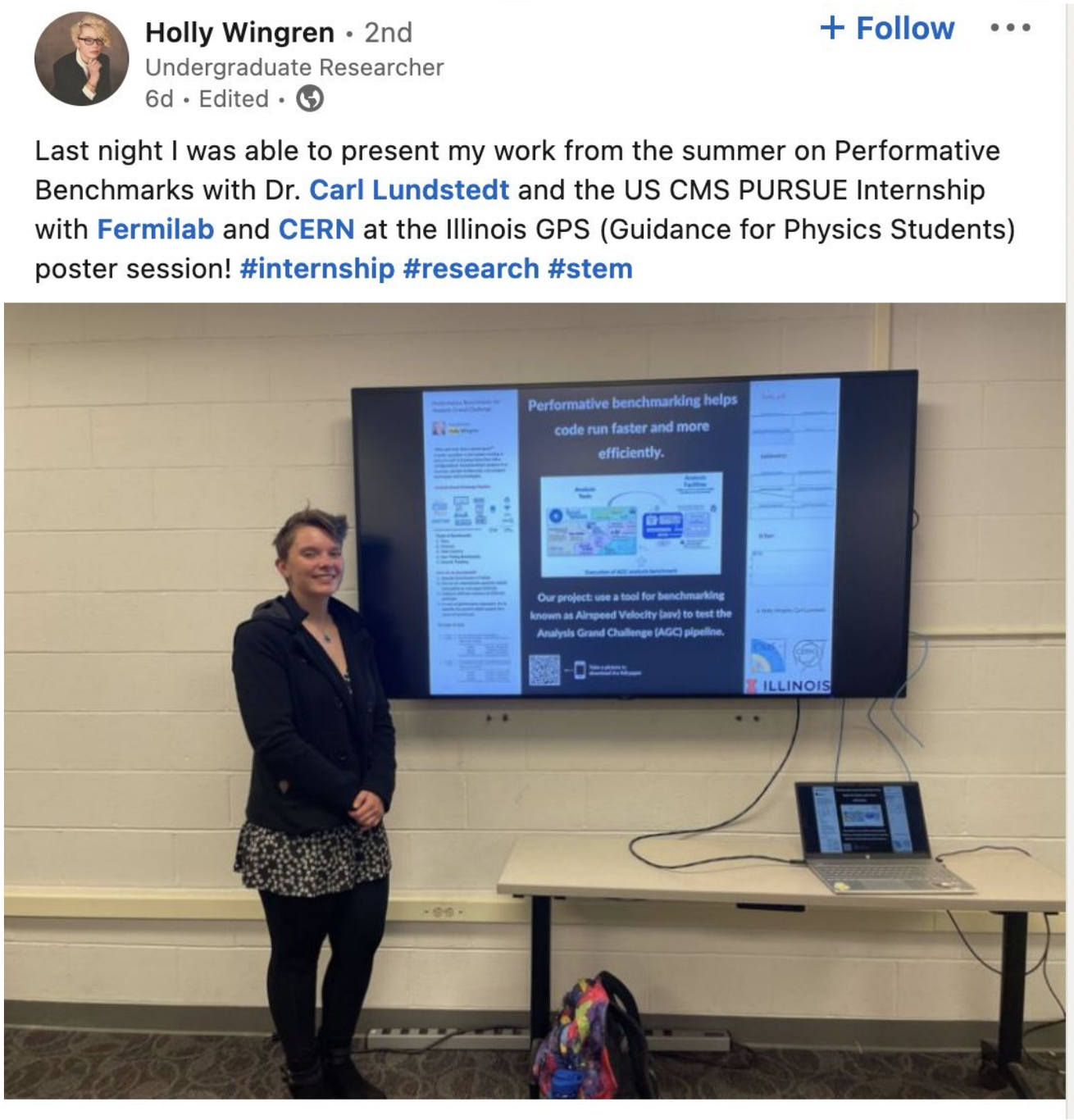
#CERN #Fermilab #CMS #ParticlePhysics #Intern #ASU



32 8 comments

Student Testimonial

https://www.linkedin.com/posts/holly-wingren_internship-research-stem-activity-6986783889314779139-xAto?utm_source=share&utm_medium=member_desktop



Mentor Testimonials

- At an external Venue:

- Oksana's slides during IRIS-HEP retreat

  Mentors: Oksana and Nick Smith
  Intern: Sneha Dixit

  - **US CMS fellow project**: Sneha Dixit (supervised by Nick Smith) enabled NanoAOD generation for 2015 CMS data
    - Will allow us to switch input data from custom ntuples to NanoAODs

Lessons Learnt

- Organize a focus group to understand how to
  - restructure the program.
  - to optimize the design of the syllabus/weekly work
- Ask students about their interest prior to matching with mentors
- Find out about computing skills
- Streamline a few activities across the research activities
- Train the Mentors prior to the program
  - invest of effort needed in the program
  - talk about expected skill levels of participants
  - how to keep in touch with the mentees on the daily basis
  - how to promote networking within the local groups and in USCMS
- More..

Recruitment

- Involve alumni / planned webinars
- Start early this year
- If possible, involve hardware projects
- Help program alumni apply to PhD programs or jobs in STEM
- **This effort is directly coupled with recruitment to our PhD programs**

For 2023 edition

- Strategy was multifold
  - Meenakshi planned the activities listed below to reach a broad and large pool of applicants, and encourage them to apply to our program.
- Contacted department chairs (2000 emails were sent) of to advertise the program at their institutions
- Organized hour long webinars
- Advertise at affinity workshops and conferences including NSBP, NSHP, SACNAS, ERN, and CUWiP.
- One-day virtual visits by the organizing team to selected institutions (could not do)

Poster and Booth at NSBP conference

November 6- 9, 2022, in Charlottesville, Virginia

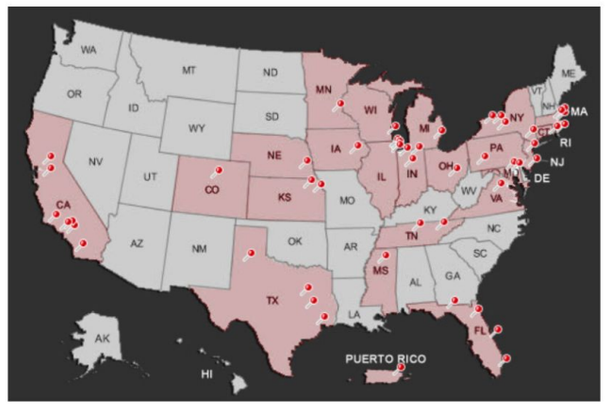
U.S. CMS PROGRAM FOR UNDERGRADUATE RESEARCH SUMMER EXPERIENCE

What will you PURSUE?

Learn how to unravel mysteries of our universe from scientists on CMS experiment that discovered Higgs Boson leading to 2013 Nobel prize in Physics

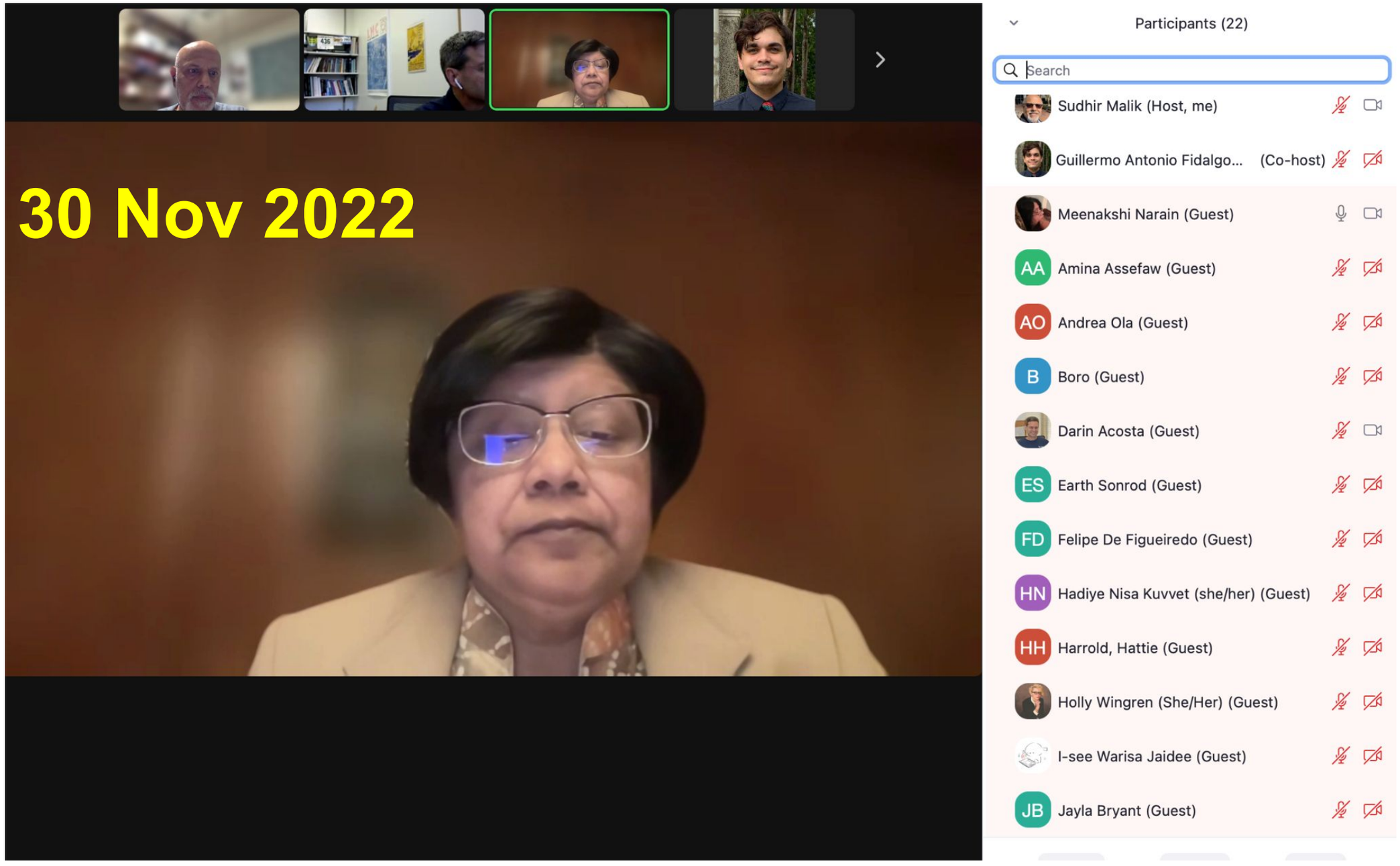
Opportunity to collaborate and be mentored by physicists from 50+ institutions

Learn how to perform cutting edge physics analysis with state-of-the-art skills in computing, software and instrumentation



Contact:
sudhir.malik@upr.edu

Webinars



- 30 Nov and 9 Dec 2022
- 50 registrants, lots of emails to us of further queries post that

DOE HEP-RENEW Funding for 3-years

- Applied for DOE HEP-RENEW for this internship
- Had one month to submit proposal before even Snowmass finished
  - Would not be possible without Meenakshi's charismatic push

Tulika Bose - University of Wisconsin- Madison
Sudhir Malik - University of Puerto Rico Mayaguez
Meenakshi Narain - Brown University
Santanu Banerjee - Tougaloo College (a historically black college)

In proposal we called the internship as
"**RENEW-HEP: U.S. CMS SPRINT - A Scholar Program for Research INTernship**"
https://science.osti.gov/-/media/funding/pdf/Awards-Lists/FY22-HEP-RENEW.pdf

THANK YOU

THANK YOU Meenakshi for enabling us to do things that seemed impossible and always finding solutions

and as Sarah said -
No one of us can replace Meenakshi, but maybe if we all work together, we can make her dream a reality