Fair Universe: Unbiased Data Benchmark Ecosystem for Physics

Wahid Bhimji,
NERSC, Berkeley Lab
wbhimji@lbl.gov
Paolo Calafiura, Ragansu Chakkappai, Yuan-Tang Chou, Sascha Diefenbacher, Steven Farrell, Aishik Ghosh, Isabelle Guyon, Chris Harris, Shih-Chieh Hsu, Elham Khoda, Benjamin Nachman, Benjamin Thorne, Peter Nugent, Mathis Reymond, David Rousseau, Ihsan Ullah, Daniel Whiteson

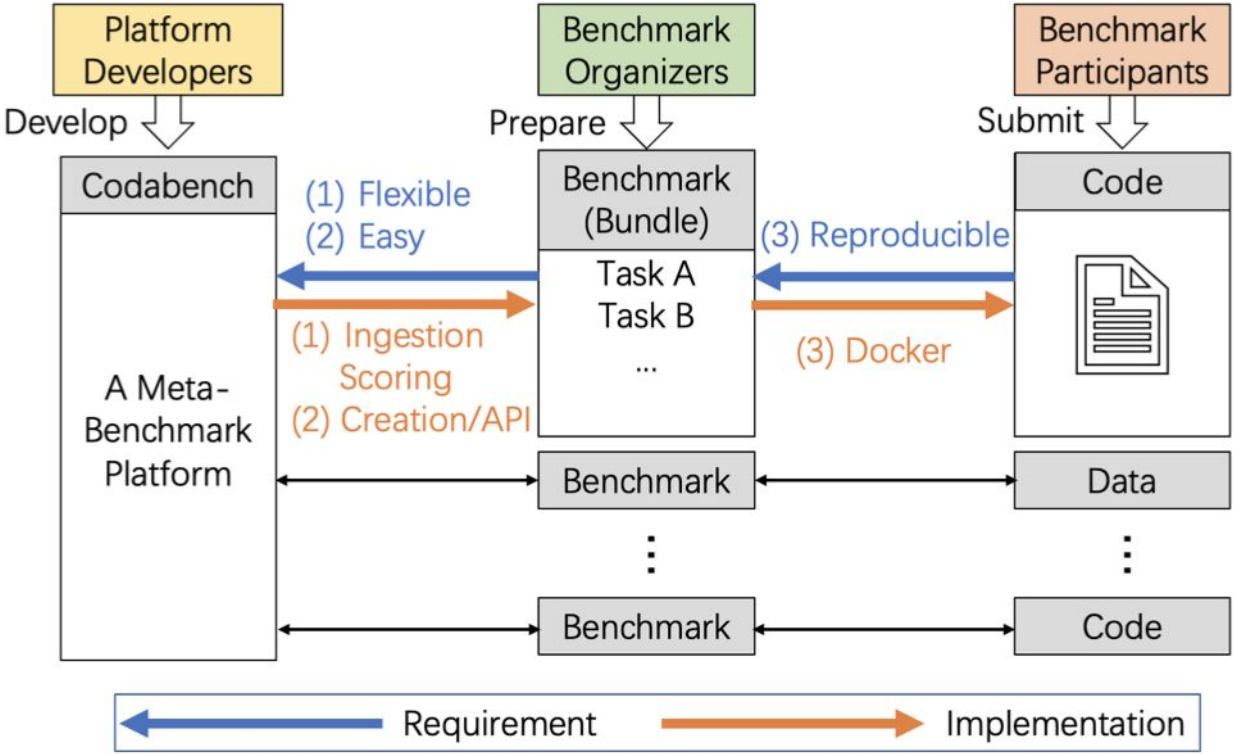
Background on Fair Universe Project

- 3 year US Dept. of Energy, AI for HEP project. Aims to:
  - Provide an open, **large-compute-scale AI ecosystem** for sharing datasets, training large models, fine-tuning those models, and **hosting challenges and benchmarks.**
  - **Organize a challenge series**, progressively rolling in tasks of increasing difficulty, based on novel datasets.
  - Tasks will focus on **measuring and minimizing the effects of systematic uncertainties** in HEP (particle physics and cosmology).
- Broad team in HEP, ML and computing involved in several previous challenges and benchmarks for HEP (e.g. HiggsML and TrackML) and wider (e.g NeurIPS competition track, MLPerf HPC); as well as Uncertainty aware learning in HEP

Large-compute-scale AI ecosystem … hosting challenges and benchmarks.

Codabench/“Fair Universe” Platform

Based on
https://www.codabench.org/



codabench.org

Search Competitions

Codabench

Benchmarks | Resources | Queue Management | Login | Sign-up

Announcement

Welcome to **Codabench**!

Join the Google group to connect with the community!

Popular Benchmarks

	Track 1: Pedestrian Attribute Recognition - WACV'24 As a part of the WACV'2024 Pedestrian Attribute Recognition and Person Retrieval Challenge... *Organized by: julioj*	September 9, 2023 **50** Participants
AutoML Cup 2023	**AutoML Cup Phase 1** AutoML Cup Phase 1 *Organized by: automlcup*	June 6, 2023 **49** Participants
	(ended) Auto-Survey Challenge'23 Auto-Survey Challenge'23 *Organized by: fnachaleam*	July 7, 2023 **36** Participants
	SNAKE #1 SaNitization Algorithms under attacK *Organized by: louisbeziaud*	May 25, 2023 **22** Participants

Featured Benchmarks

	WACV 2024 - Grand Challenge - AV-Sync Error Measurement Regression prediction of temporal offset between audio and video *Organized by: highamdh*	October 11, 2023 **7** Participants
AutoML Cup 2023	**AutoML Cup Phase 2** AutoML Cup Phase 2 *Organized by: spencrr*	July 15, 2023 **20** Participants
	Cross-Domain MetaDL Any-Way Any-Shot Learning Competition with Novel Datasets from Practical Domains *Organized by: pavao*	November 15, 2022 **9** Participants
	RescueNet - Semi-Supervised Semantic Segmentation Semi-Supervised Semantic Segmentation of RescueNet dataset into 10 defined classes *Organized by: binalab*	October 2, 2023 **6** Participants

Get Started

Participate

Find benchmarks that pique your interest! A benchmark allows you to test new algorithms against reference datasets OR (inverted benchmark) submit challenging data to reference algorithms.

Organize

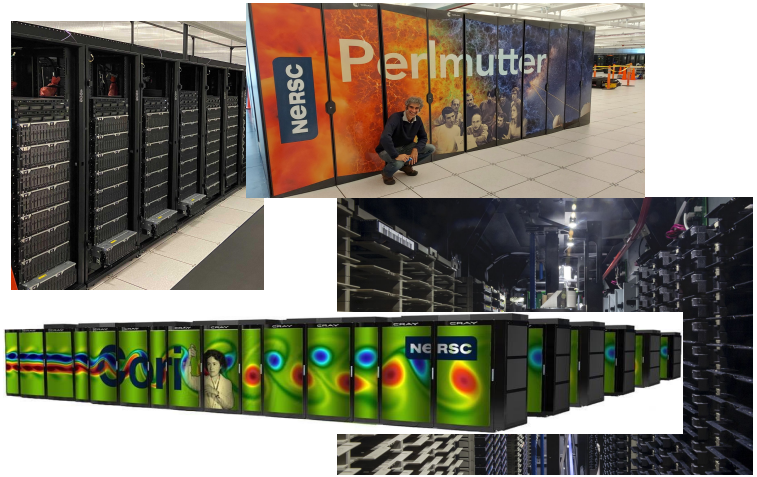
Organize a benchmark on Codabench. Start with our tutorial.

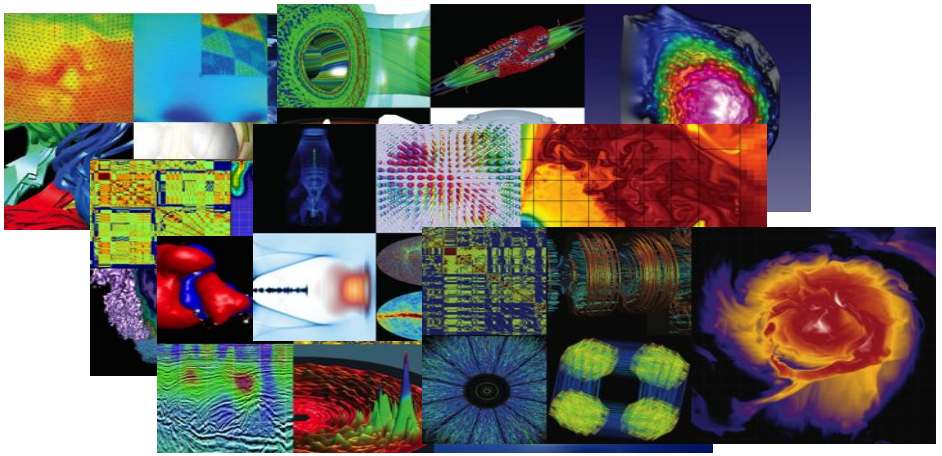
Contribute

Interested in joining the development team? Join us on Github or contact us directly.

BERKELEY

NERSC: Mission HPC for the Dept. of Energy Office of Science





Large compute and data systems

- Perlmutter: ~7k A100 GPUs
- 128PB Community Filesystem ....

Broad science user base

- > 9,000 users,
- 1000 projects,

FAIR Universe Platform - Backed by NERSC!

TOP 500 The List.
#7, 93.8PF Peak

>5 *TB/s*

35 **PB All-Flash Scratch**

HPE Slingshot 11 interconnect
4 NICs/GPU node,
1 NIC/CPU node

1,792 GPU-accelerated nodes
4 NVIDIA A100 GPUs+1 AMD "Milan" CPU
448 TB (CPU) + 320 TB (GPU) memory
3,072 CPU-only nodes
2 AMD "Milan" CPUs
1,536 TB CPU memory

1.6 TB/s

Coming 2026/27

NERSC-10

~10x Perlmutter capability

Platform Storage System (PSS)

Quality of Service Storage System (QSS)

Workflow Environment
Management Environment

Off Platform Storage

HPSS Tape Archive ~300 PB — *50 GB/s*

Community File System 130 PB — *100 GB/s*

/home 450 TB — *5 GB/s*

DTNs, Gateways

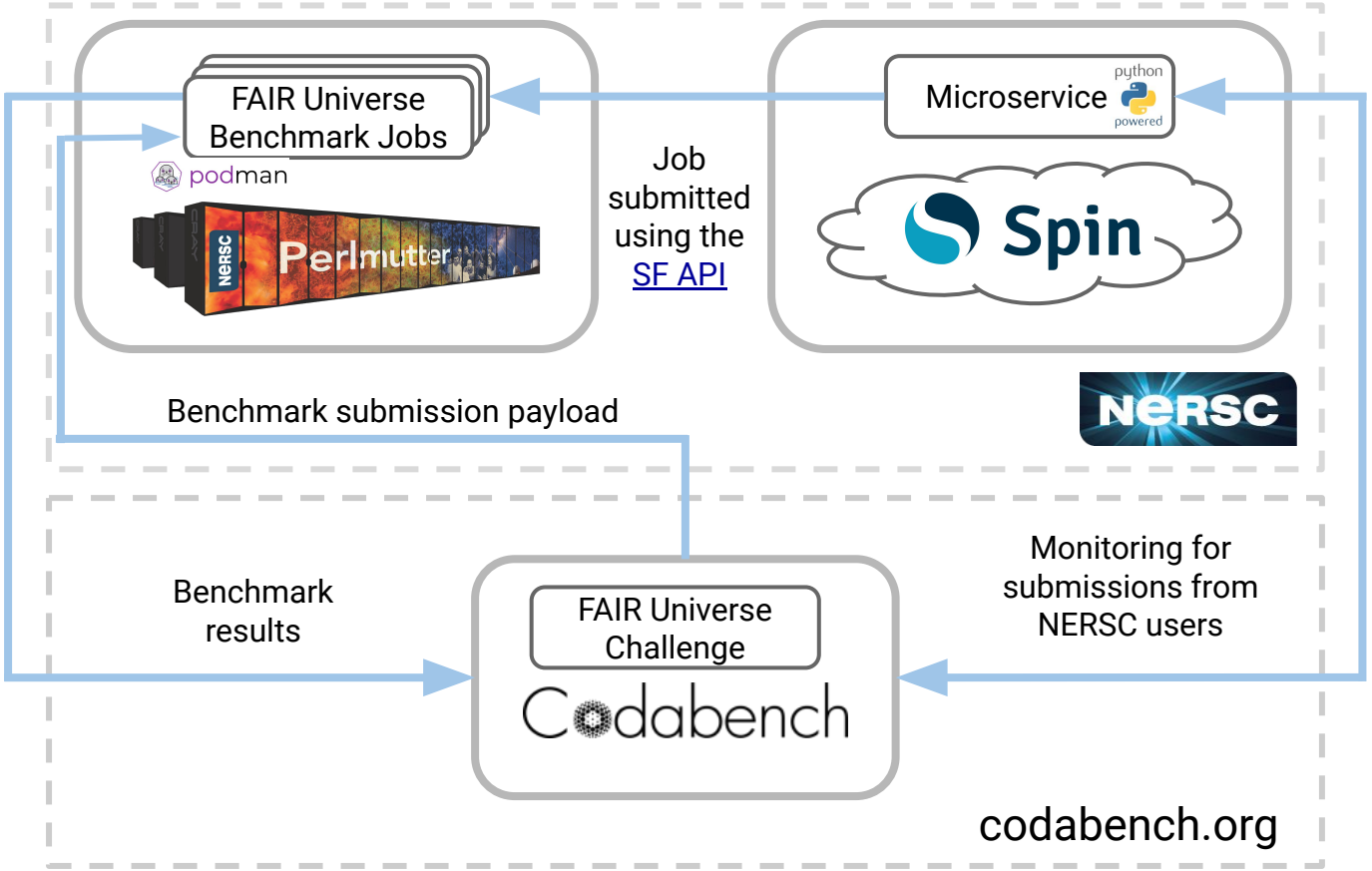
Spin edge services

Ethernet
Science Friendly
Security
Production Monitoring
Power Efficiency
LAN

ESnet ENERGY SCIENCES NETWORK
2 x 400 Gb/s
2 x 100 Gb/s

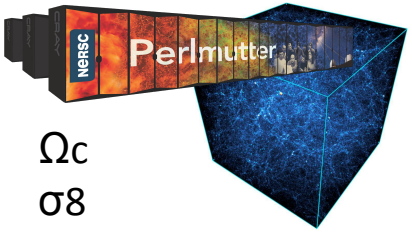
Experimental Facility | ASCR Facility | Home Institution | Cloud | Edge

Fair Universe Platform: Current Codabench/NERSC integration



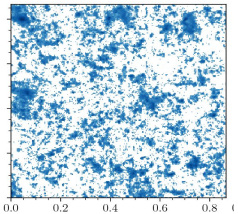
Measuring and minimizing the effects of systematic uncertainties in HEP

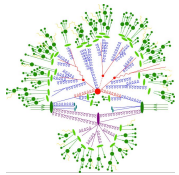
Bias and uncertainty in the fundamental sciences

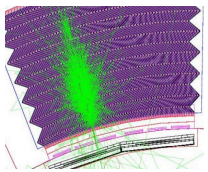


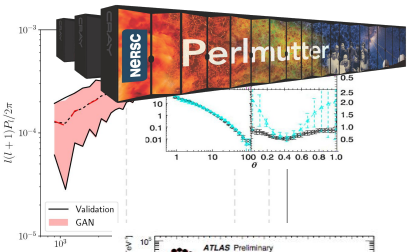
Ω_c
$\sigma 8$

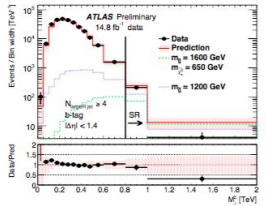
θ_{12}
m_H …

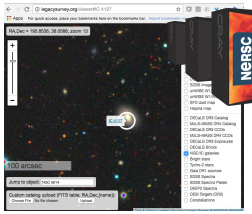


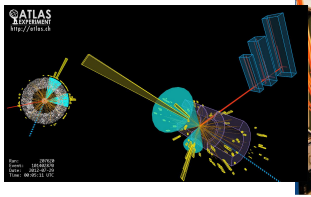


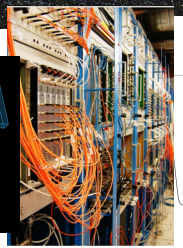


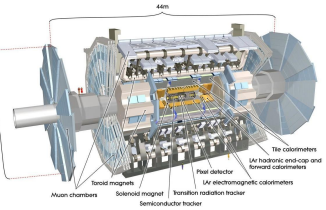












Theory into Simulations

- High-resolution with detailed physics and instrument/ detector simulation

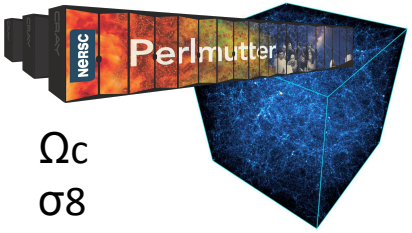
Summary statistics:

- E.g. 2pt /3pt correlation: spatial distribution
- E.g. Masses of reconstructed particles

Exp/Obs reconstruction

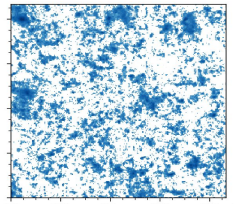
- Derive position of galaxies/stars and properties for catalogs
- Reconstruct particle properties

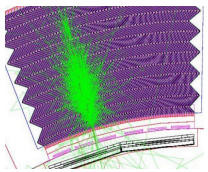
Bias and uncertainty in the fundamental sciences

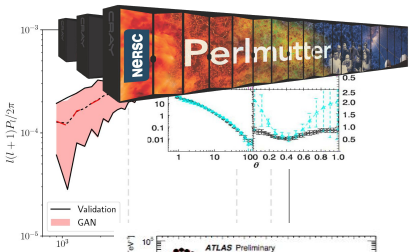


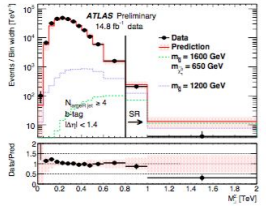
Ω_c
$\sigma 8$

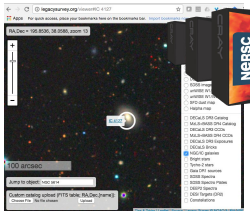
θ_{12}
m_H …

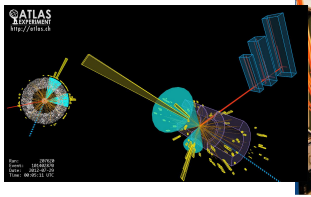


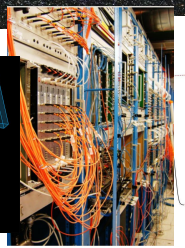


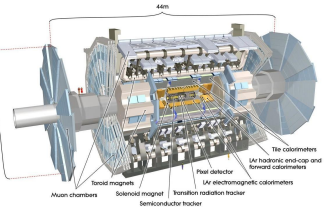












Theory into Simulations

- Estimate Systematic Uncertainties (Z)

Exp/Obs reconstruction

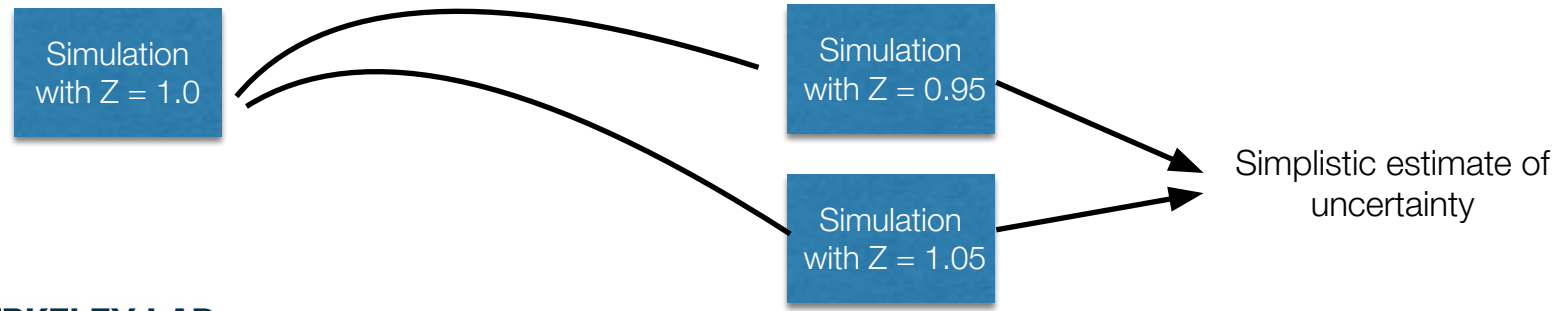
- Detector state Z=?

Differences between simulation and data can bias measurements

Bias and uncertainty in ML in the fundamental sciences

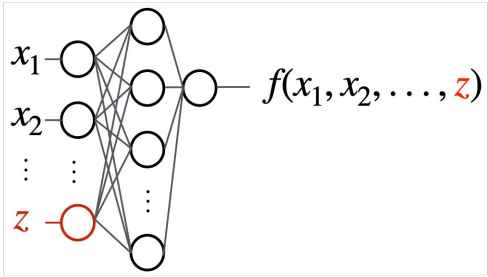
- ML methods in HEP are often trained based on simulation which has estimated systematic uncertainties (“Z”)
- These are then applied in data with the different detector state Z=?

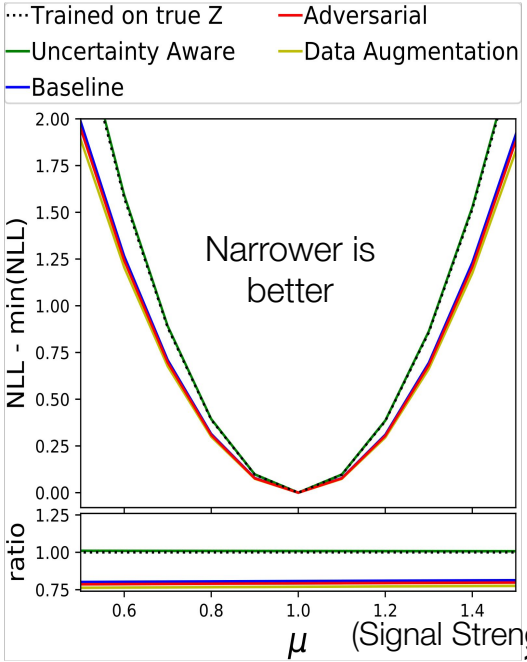
- Common baseline approach: Train classifier on nominal data (e.g. Z=1) and estimate uncertainties with alternate simulations. Shift Z and look at impact or perform full profile likelihood



Increasingly sophisticated approaches

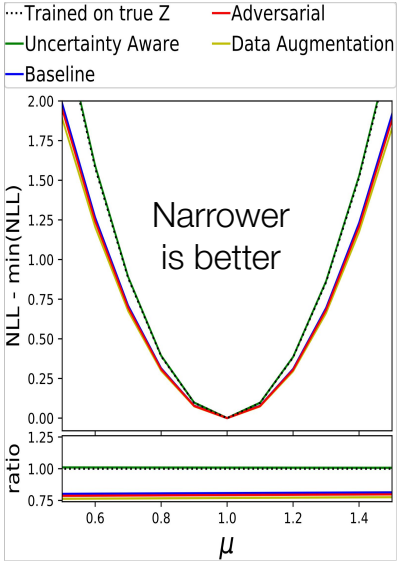
- Several focussed on decorrelation, e.g. augmentation; adversarial training; tangent propagation etc.
- “Uncertainty-aware” approach of Ghosh, Nachman, Whiteson PhysRevD.104.056026
  - Parameterize classifier using Z
  - Measured on “Toy” 2D Gaussian Dataset and dataset from *HiggsML Challenge* modified to include systematic on tau-energy scale
  - Performs as well as classifier trained on true Z
- Other novel approaches e.g. (not comprehensive)
  - Inferno: arxiv:1806.04743
  - Direct profile-likelihood: e.g. arxiv:2203.13079

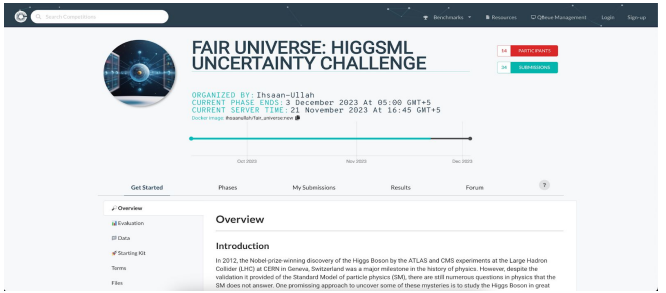




Progress requires new datasets, metrics, and platform

- Uncertainty-aware paper demonstrated on single systematic uncertainty, with limited data
- Original HiggsML dataset too small for ambitious approaches (systematic uncertainty small compared to statistical uncertainty)
- How to scale methods to many values of Z? (training difficulty increases, profiling approach used is expensive)
- Can faster methods allow for directly evaluating profile likelihood?
- Need for novel metrics to evaluate uncertainty determination for such methods

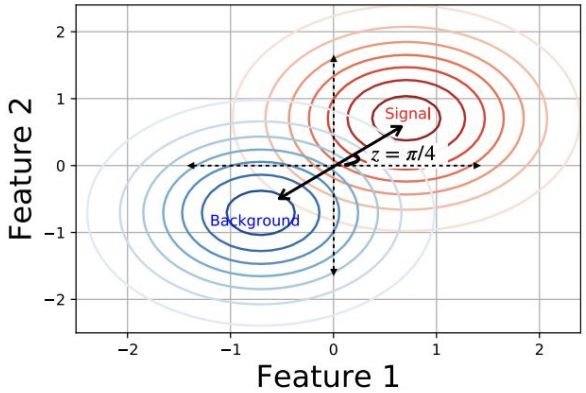


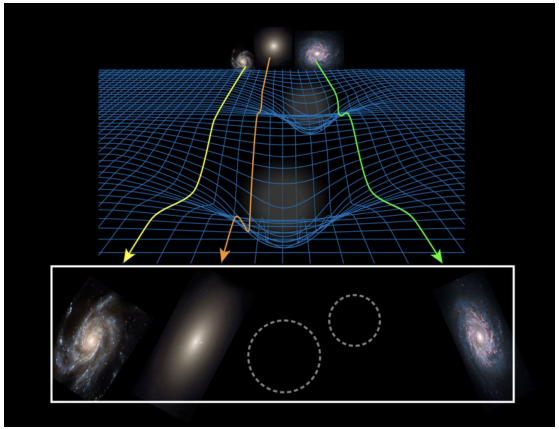


Organize a challenge series, progressively rolling in tasks of increasing difficulty, based on novel datasets

We plan phased challenges and datasets:

- "Toy Problem" - Shifted 2D Gaussian challenge to evaluate the platform and introduce problem
- ➢ *HiggsML Uncertainty Challenge prototype* to refine approach, metrics and scoring: available this week
  - Introduced this Wed 29th Nov pm
- Full HiggsML Uncertainty Challenge in 2024
- We also plan a Cosmology (weak lensing) challenge
  - To be organised by the "Fair Universe" team with collaborators - large dataset available
- As well as other challenges, not organized by us but using related platforms, tools or approaches:
  - E.g. See Wed morning talks at this workshop

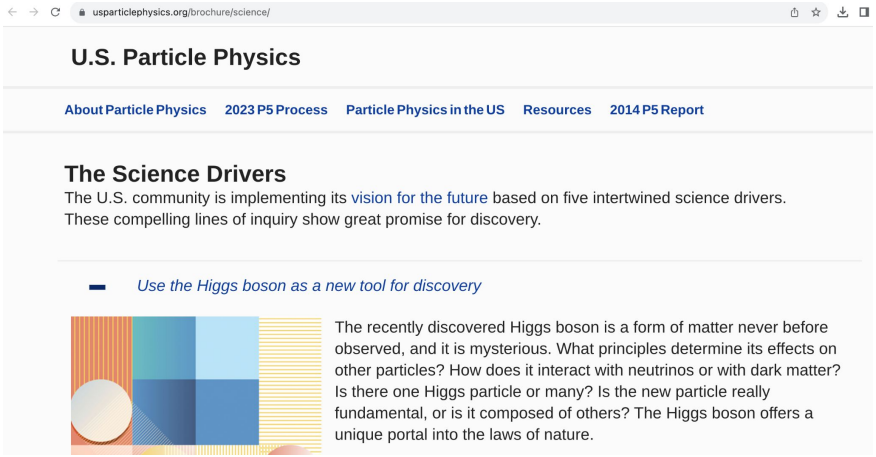


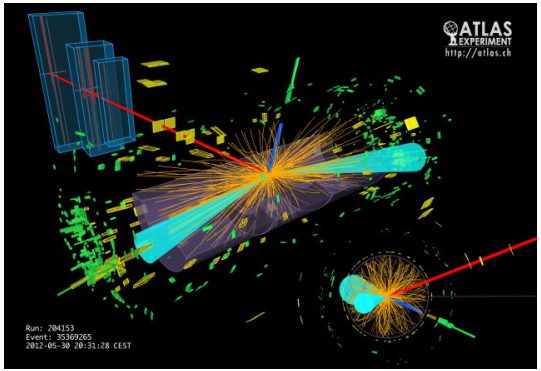


Fair Universe: HiggsML Uncertainty Challenge - motivation

- Top science driver: “Use the Higgs Boson as a new tool for discovery”
  - One avenue is detailed study of Higgs decaying to “Tau”s
  - Need to separate these Higgs decays from more prevalent “backgrounds”
- Tau leptons decay to other particles observed in LHC detectors
  - Rates for these ‘channels’ vary over orders of magnitude
  - Including uncertainties on detector observations is crucial

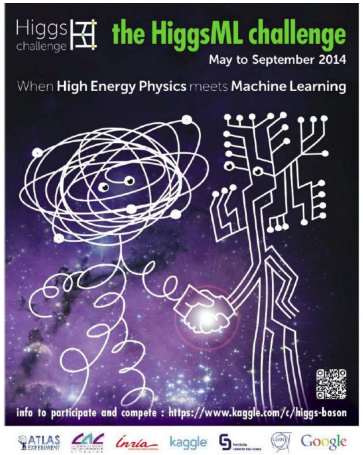
https://www.usparticlephysics.org/brochure/science/





Fair Universe: HiggsML Uncertainty Challenge

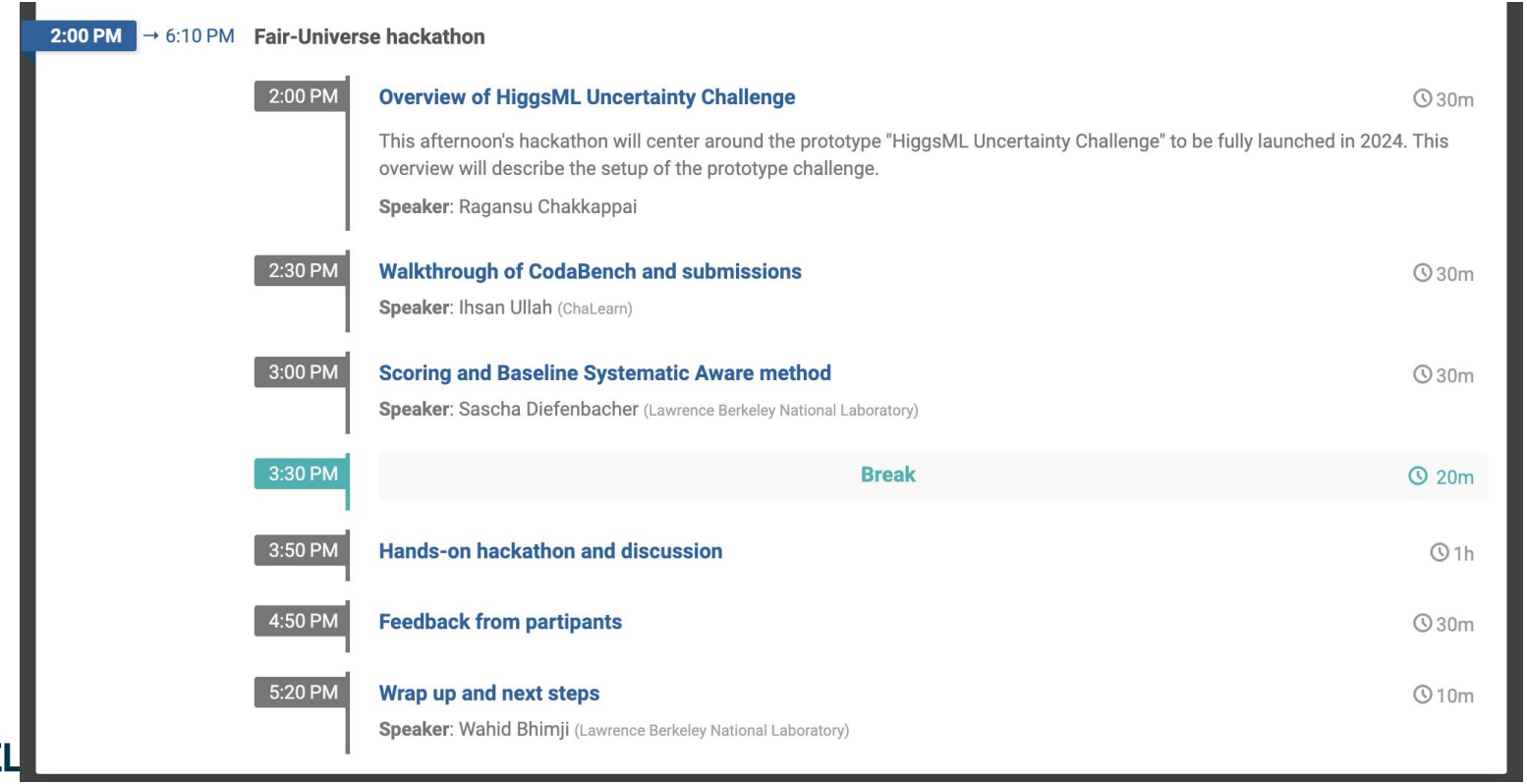
- Extension of previous HiggsML challenge from 2014 (which was a classification problem for Higgs decaying to Tau leptons in an ATLAS simulation based on momenta of decay particles and derived quantities)
- We have a larger dataset, and include systematic variations, initially, for prototype, on “Tau Energy Scale”:
  - Plan to expand to other systematics (e.g. Jet Energy Scale, MET, backgrounds or simulators - TBD)
  - Systematics implemented both using in Delphes detector simulation and post-hoc scripts to compare
- Participants submit methods that go beyond classification: predict signal strength (μ) and an uncertainty interval
  - Metric/score on precision and accuracy (see following presentation by Sasha Diefenbacher)



Join us at the kick-off event

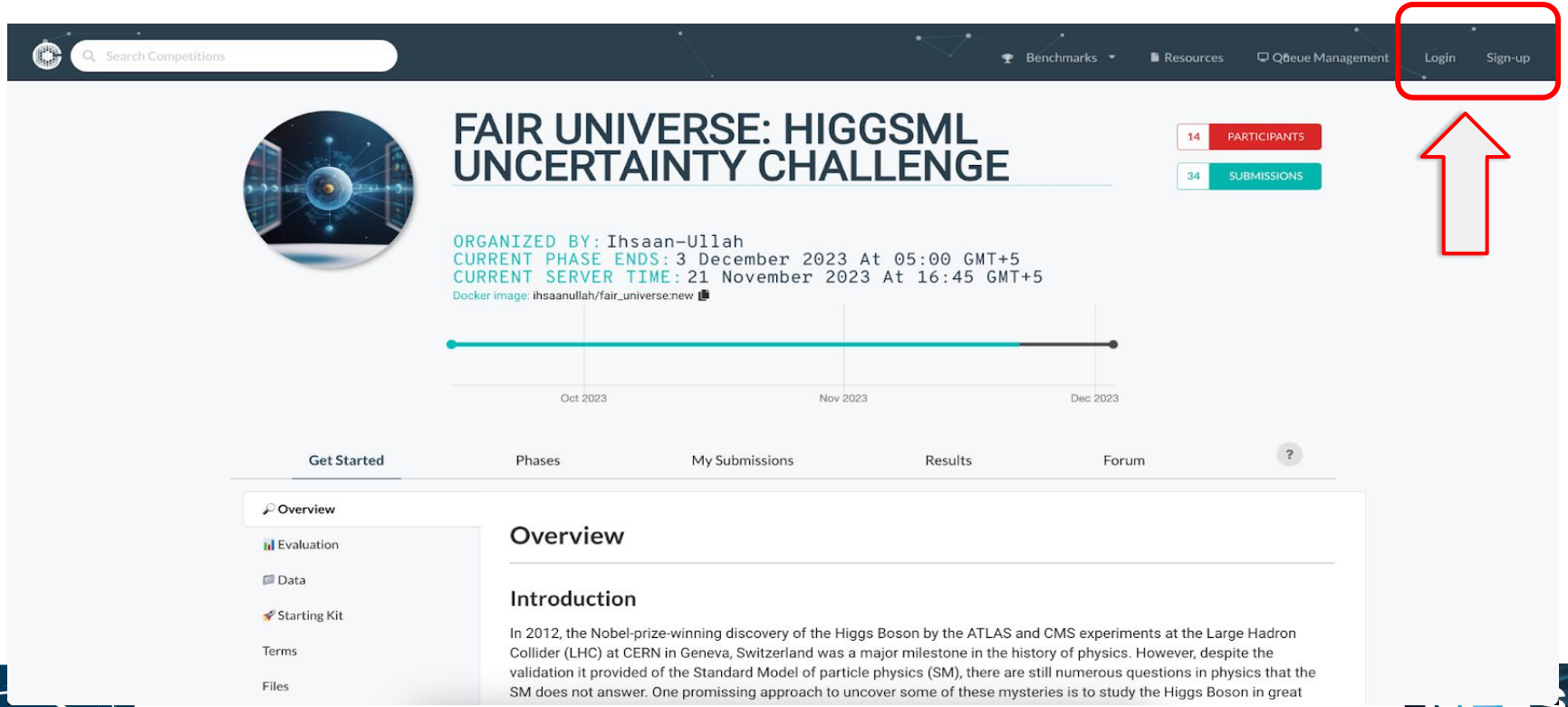
https://indico.in2p3.fr/event/30589/timetable/#b-26323-fair-universe-hackatho

#fair-universe-hackathon channel on AIUPHYS2023 slack workspace



Access to prototype challenge: CodaBench

Link to challenge will be added to Indico session



See more detailed [walkthrough slides]

BERKELEY LAB

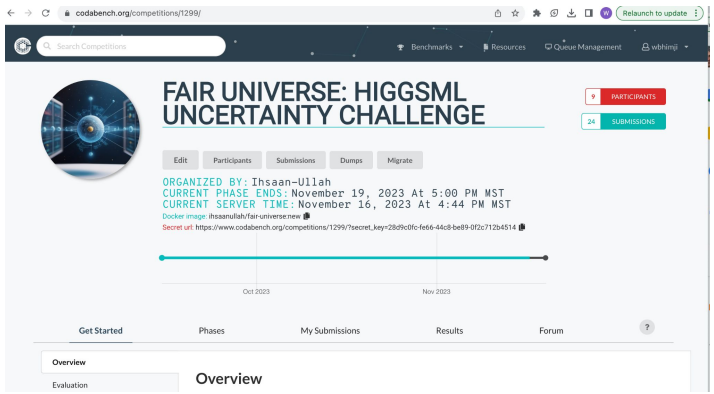
NERSC

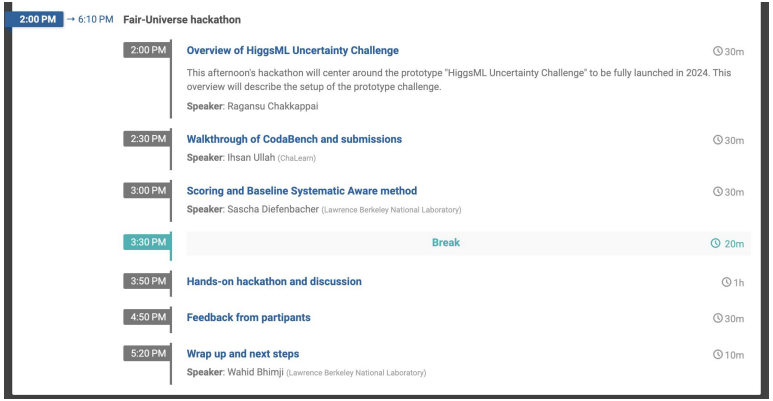
U.S. DEPARTMENT OF ENERGY | Office of Science

Setting up for challenge: NERSC Accounts

- Submissions will run on CodaBench backend by default
- If you want these to run NERSC resources (for example for methods that need a GPU or for possibly faster turnarounds in general) then can request training account valid for the week
- **To get a training account**
  - Fill the training account request form at https://iris.nersc.gov/train
  - Use training code to be provided in indico / slack
  - **Use the same email for this form as for your CodaBench account**
  - **Let us know on slack and we will enable you**

Join us on Wednesday to find out more!





https://indico.in2p3.fr/event/30589/timetable/#b-26323-fair-universe-hackatho

#fair-universe-hackathon channel on AIUPHYS2023 slack workspace

For ongoing information Google Group: Fair-Universe-Announcements

Collaborations, questions, comments: wbhimji@lbl.gov