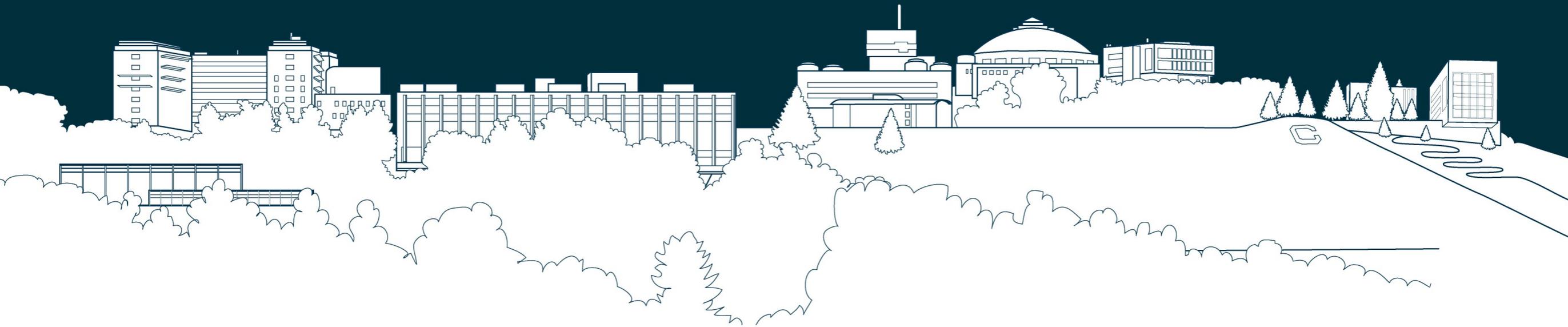


**Berkeley Lab Physics Division**

# **Machine Learning for Fundamental Physics**

**November 5, 2020**



<https://www.physics.lbl.gov/machinelearning/>

# Plan for today



- Welcome and Expectations
- Introduction round table
- Scope of group and results from survey
- Future meetings

# Groundwork for the meeting



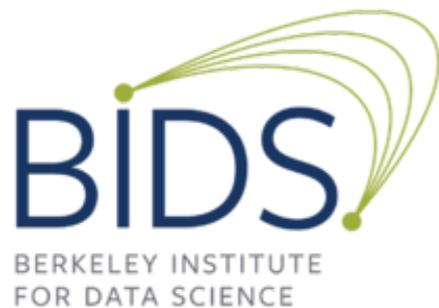
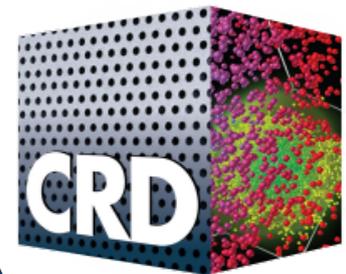
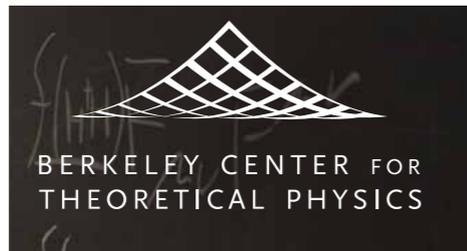
Berkeley Lab Physics Division Machine Learning Group  
*Group Meeting Code of Conduct*  
Updated Oct. 28, 2020

(flip over to the link and briefly go through the document)

# Machine Learning for Fund. Physics



## Why here and why now?



# Questions in fundamental physics

5

**Theoretical** and **experimental** questions motivate a deep exploration **of the fundamental structure of nature**

Dark matter

Hierarchy problem

Strong CP

Flavor puzzles

Baryogenesis

Dark energy

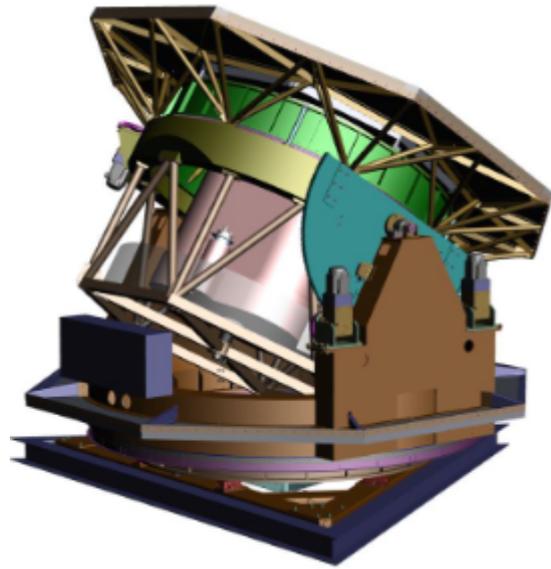
We have performed thousands of hypothesis tests & have **no significant evidence** for physics beyond the Standard Model

**We will need new tools to explore our data in new ways!**

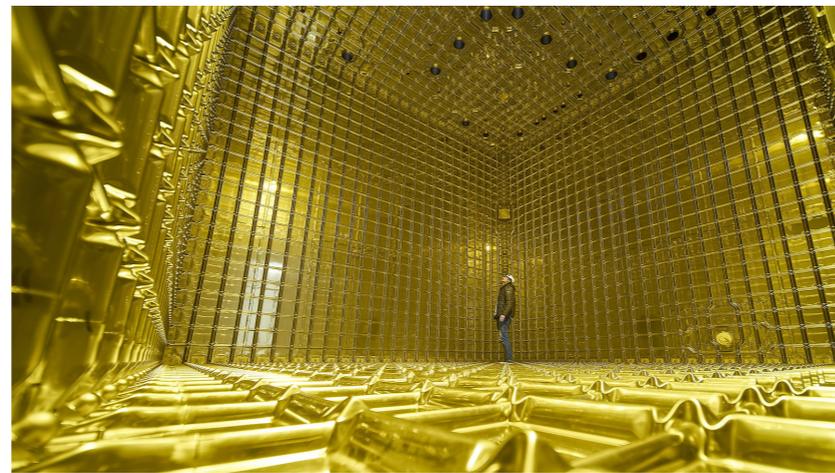
# New tools: detectors



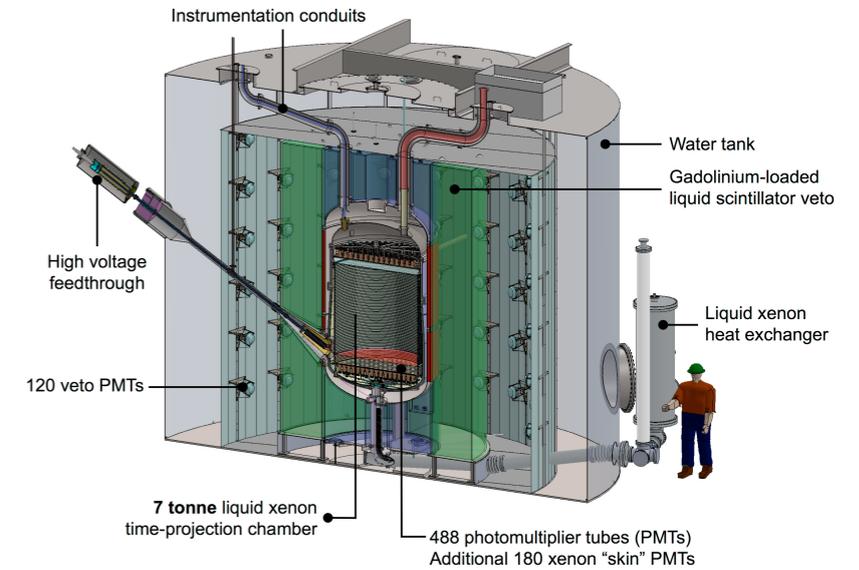
## Astro/Cosmo surveys



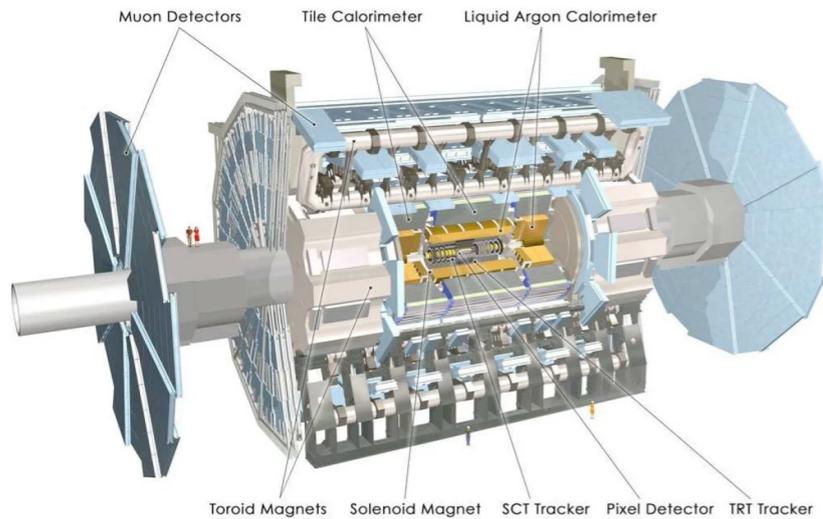
## Neutrino experiments



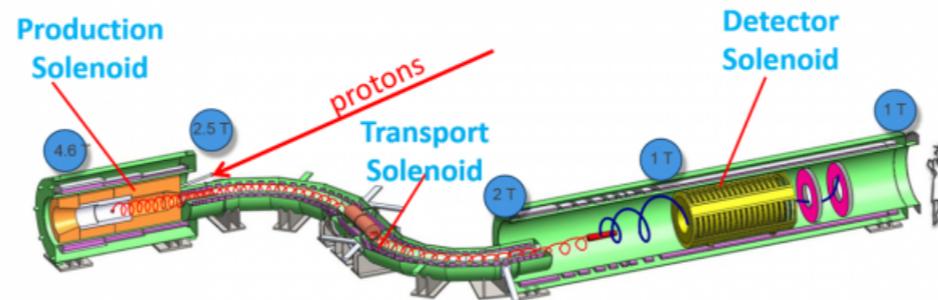
## Dark Matter direct detection



## Large Hadron Collider



## Beam dump / Fixed target experiments



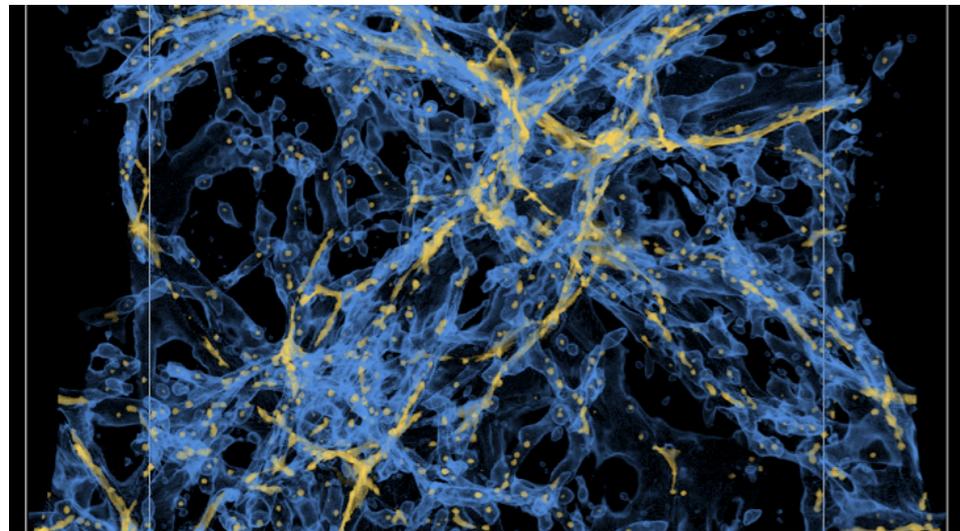
**+ others !**

(Note Berkeley leadership in all of these areas!)

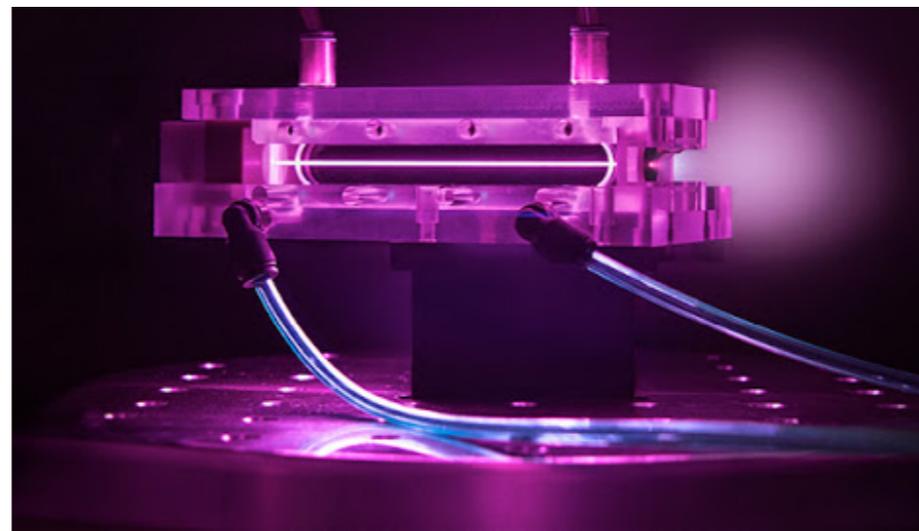
# New tools: methodology



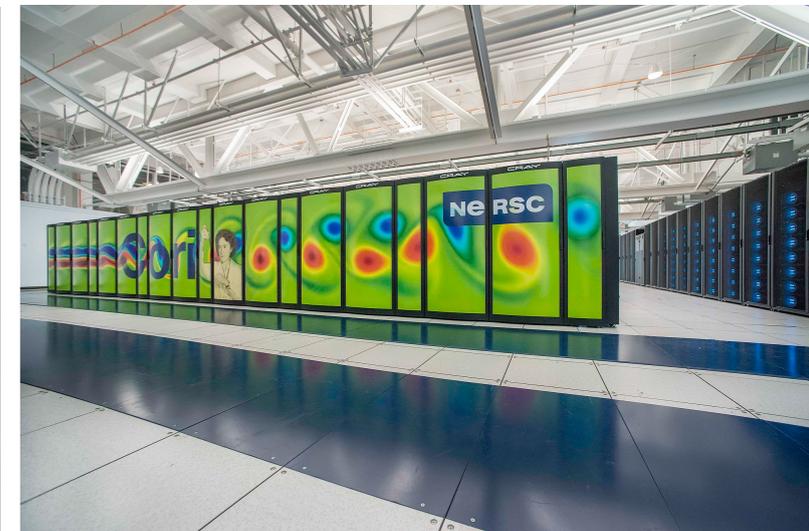
N-body simulations



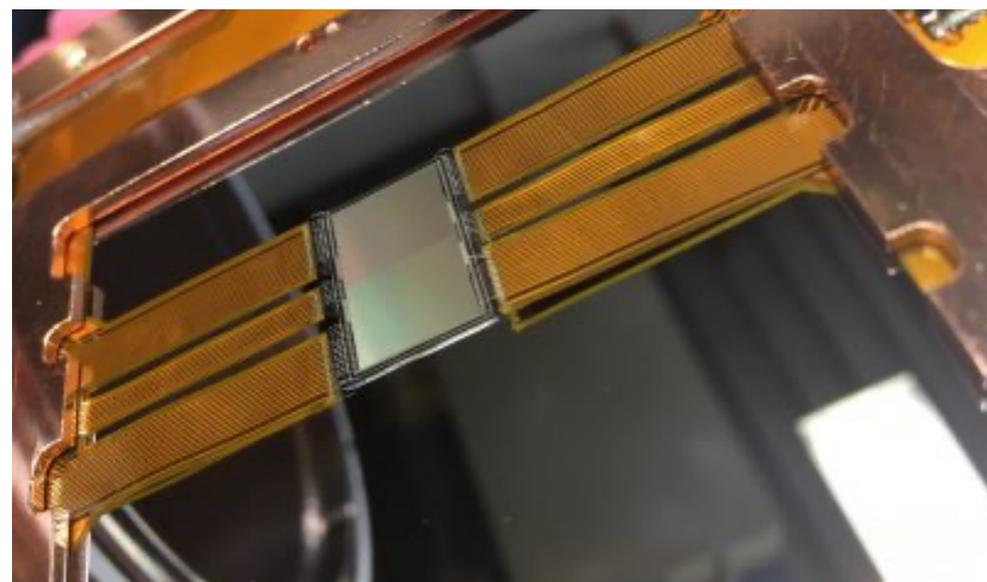
Advanced accelerators



Supercomputers



Advanced detectors



Theory Calculations

$$\mathcal{L} = \frac{4G_F}{\sqrt{2}} V_{cb} c_{XY} (\bar{c} \Gamma_X b) (\bar{\ell} \Gamma_Y \nu)$$

$$\frac{d\Gamma^{\{s\}}}{d\mathcal{PS}} = \sum_{\alpha, i, \beta, j} c_\alpha c_\beta^\dagger F_i F_j^\dagger(\{q\}) \mathcal{W}_{\alpha i \beta j}$$

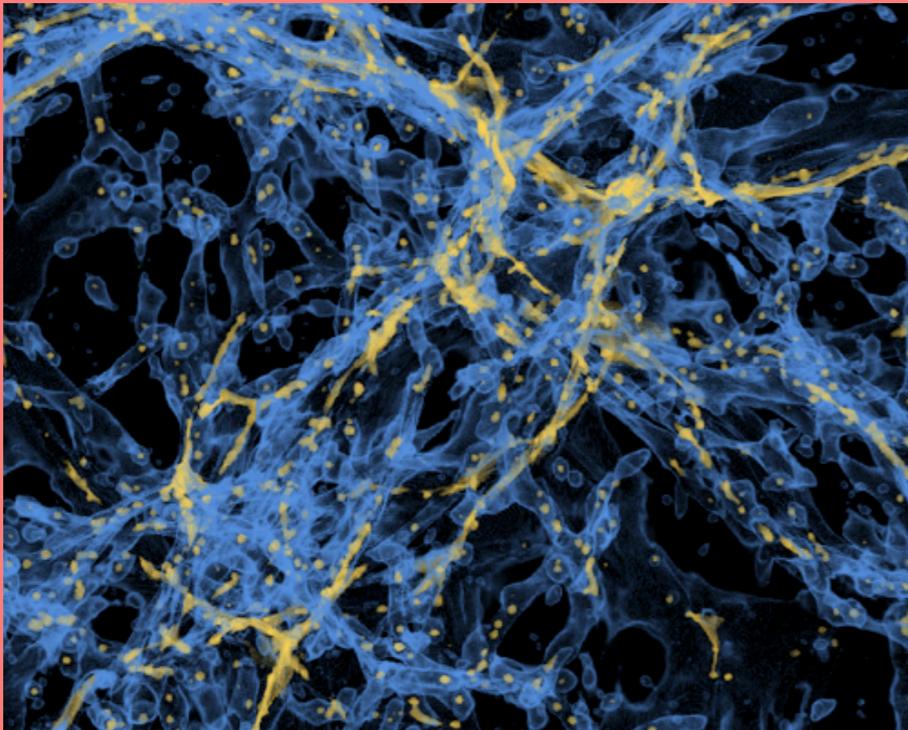
**+ others !**

(Note Berkeley leadership in all of these areas!)

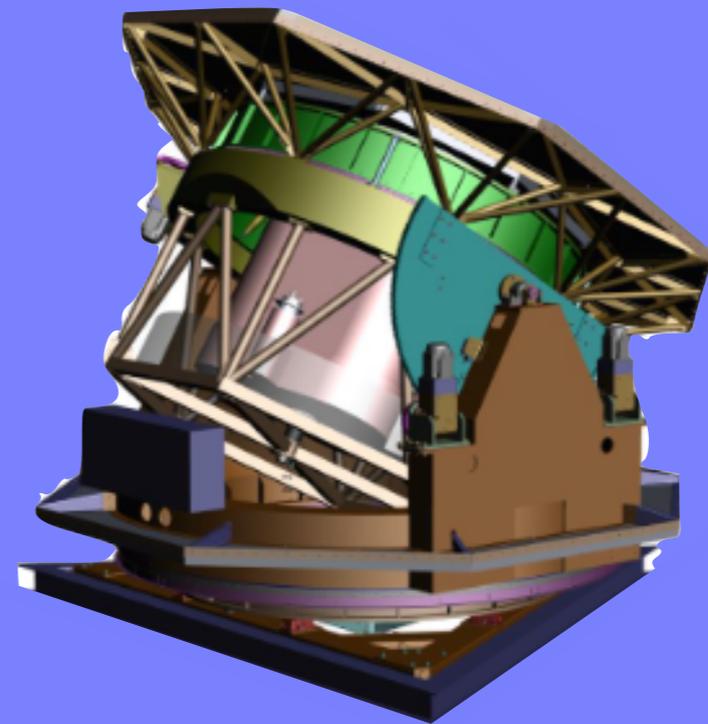
# A *hyper* challenge



Key **challenge** and **opportunity**: *hypervariate phase space*  
& *hyper spectral data*



**Methodology**

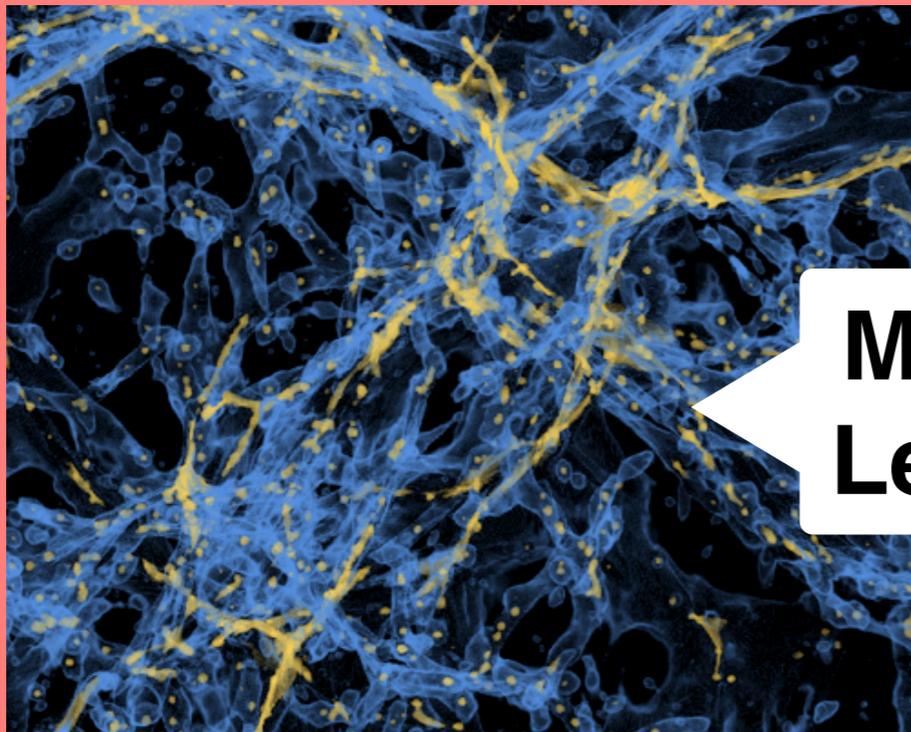


**Detectors**

# A *hyper* challenge

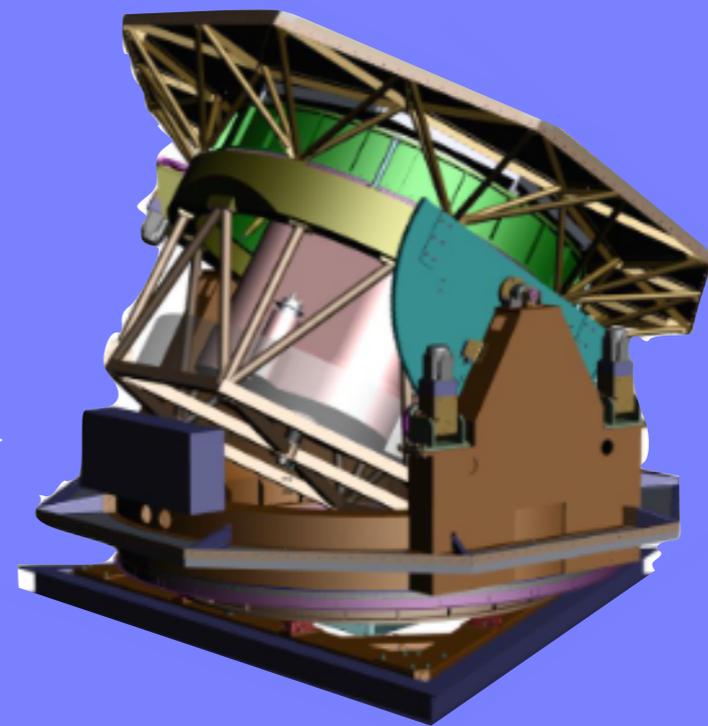


Key **challenge** and **opportunity**: *hypervariate phase space*  
& *hyper spectral data*

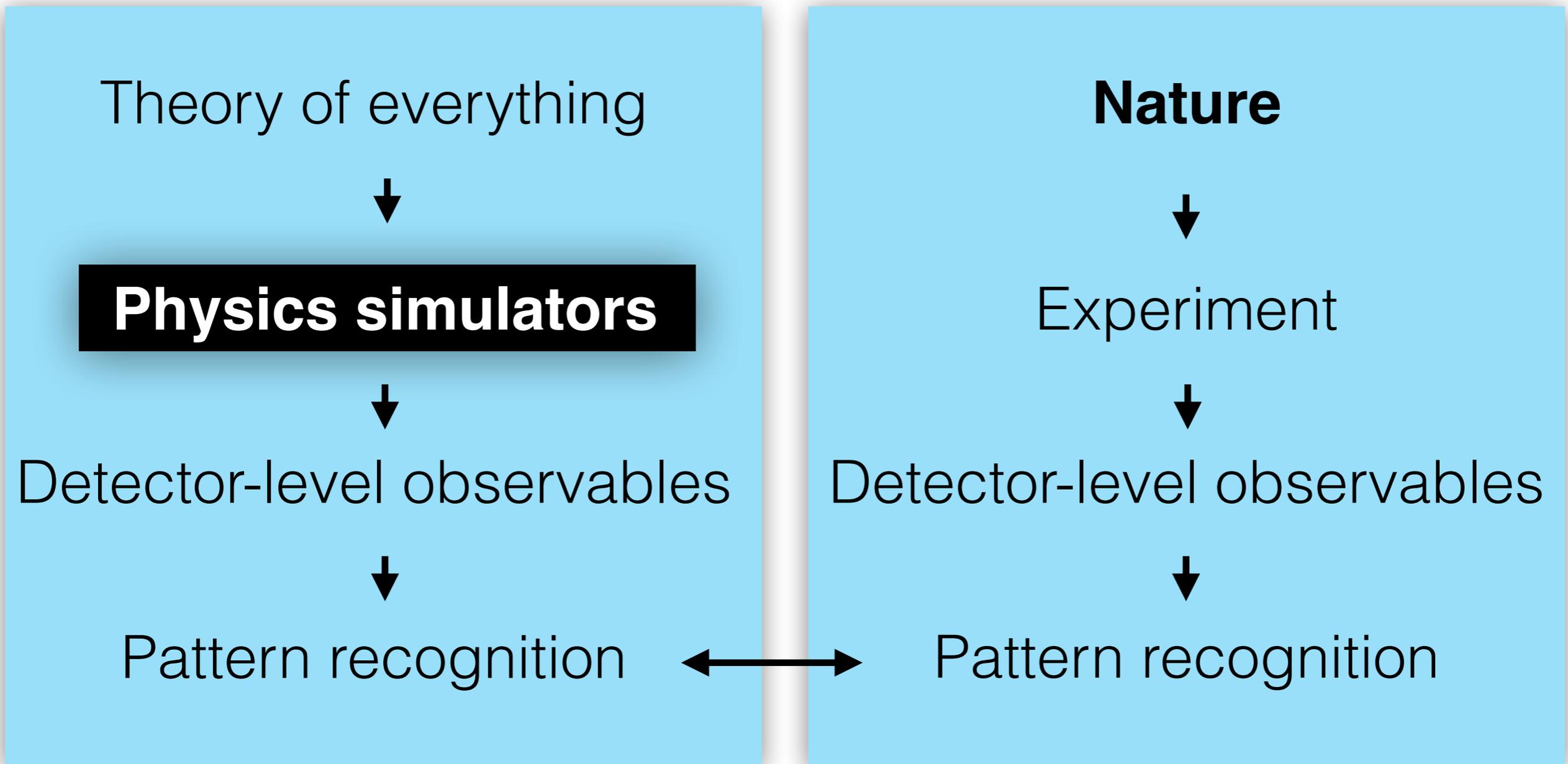


**Methodology**

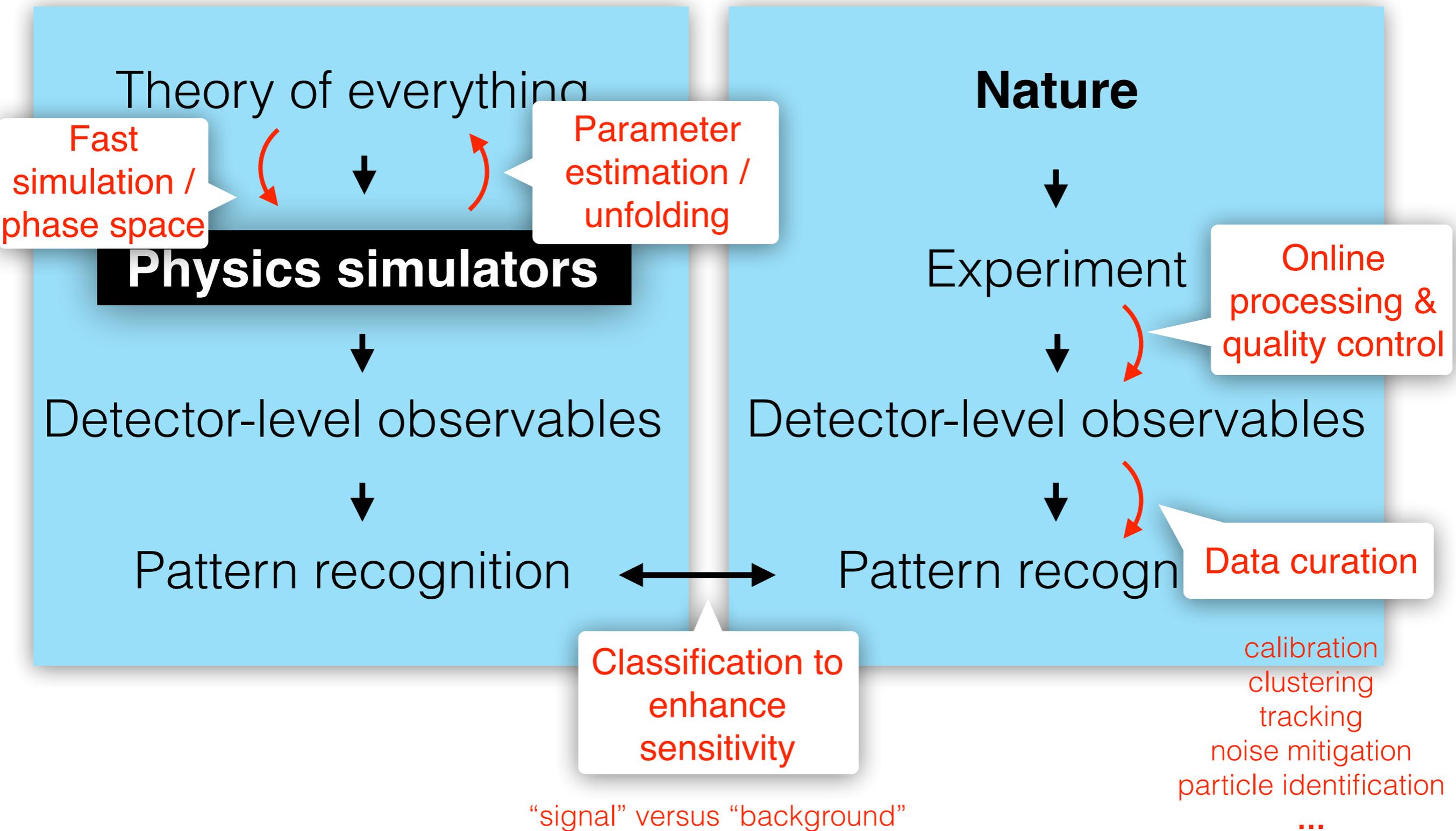
**Machine Learning**



**Detectors**

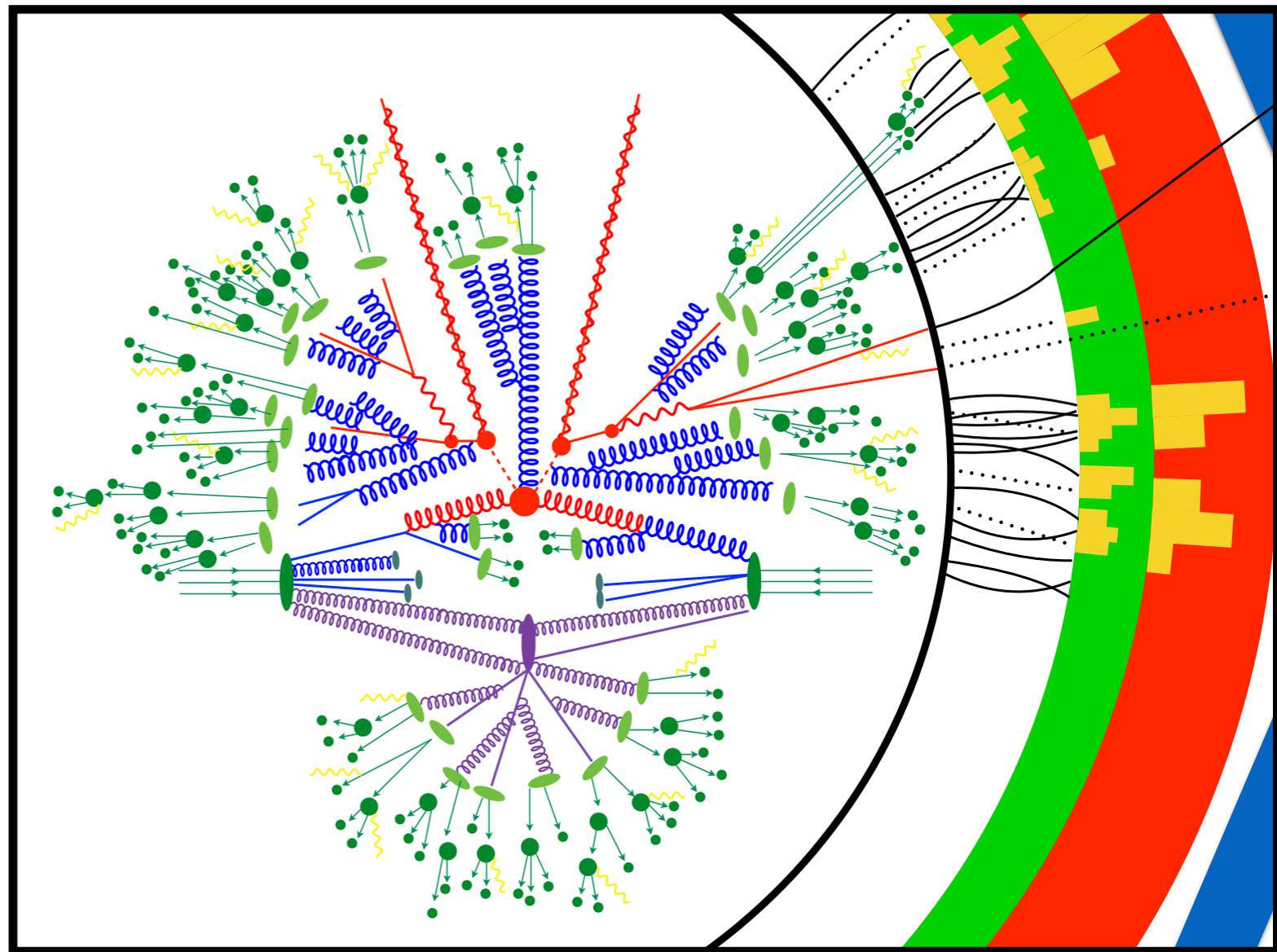


# Data analysis in fund. physics + ML



It is generally useful to connect domain experts using similar tools, but we also have **unique challenges** (+opportunities!)

- Largest scientific datasets
- High-fidelity, first-principles simulations
- Stringent requirements on uncertainties
- Demands on algorithm latency
- ...



# A new cross-cutting ML group



13

***Vision:*** *To advance the potential for discovery and interdisciplinary collaboration by approaching fundamental physics challenges through the lens of modern machine learning.*

**Mission:** The Physics Division Machine Learning group is a cross-cutting effort that connects researchers developing, adapting, and deploying artificial intelligence (AI) and machine learning (ML) solutions to fundamental physics challenges across the HEP frontiers, including theory. While most of the ML group members will have a primary affiliation with other areas of the division, there will be unique efforts within the group to develop methods with significant interdisciplinary potential. We have strong connections and collaborations with researchers in the Computational Research Division, the National Energy Research Scientific Computing Center (NERSC), and the Berkeley Institute of Data Science (BIDS).

# A new cross-cutting ML group



14

A variety of cross-cutting themes form the core of our research program:

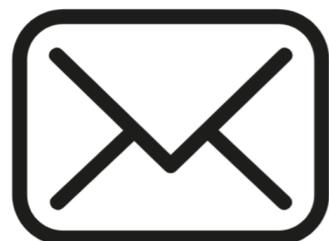
- Anomaly Detection
- Likelihood-free Inference
- Generative Models and Simulation
- Pattern Recognition, Calibration and Noise Mitigation
- Label-free learning / Simulation Agnostic Approaches
- Physics-aware Learning
- Uncertainty Quantification and Interpretability
- Hardware Interface

# A new cross-cutting ML group



15

The rest of the slides will go over how we will implement this mission!



[hep-ml@lbl.gov](mailto:hep-ml@lbl.gov)

Hopefully you are all signed up already



**slack:** [lblepml](#)

*N.B. need to use LBL or Berkeley email to sign up*

You are all welcome to join! Feel free to ask technical questions, start collaborations, etc.



[Google calendar](#)

Please let me know about any relevant seminars / workshops and I will add them to the calendar.

A technical aspect we can help with is to make it with software and computing. We are compiling some resources [here](#).

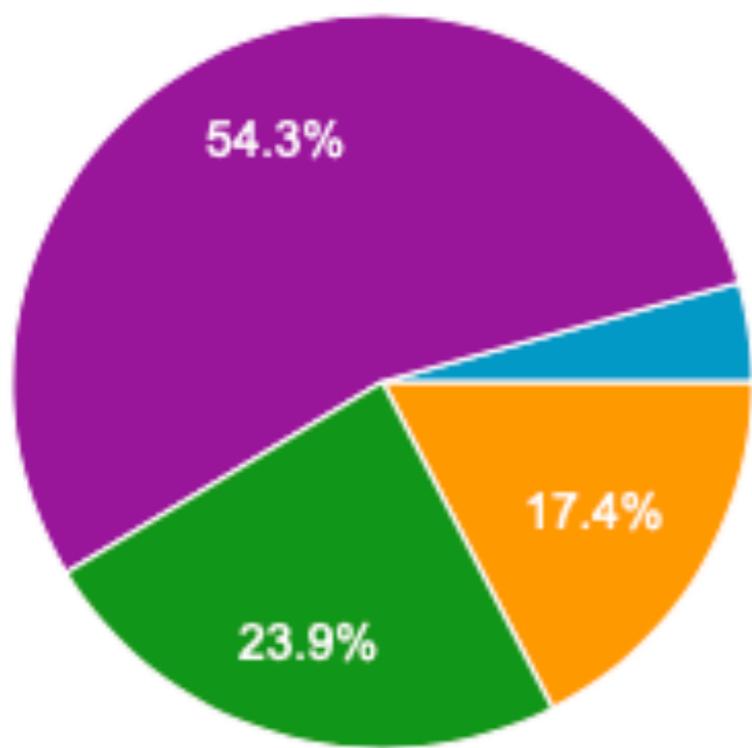
(Please let me know if you have suggestions!)

Additionally, if you do not have an account on Cori, we have an ML group account (m3705) that you can use while you become familiar with the computing environment.

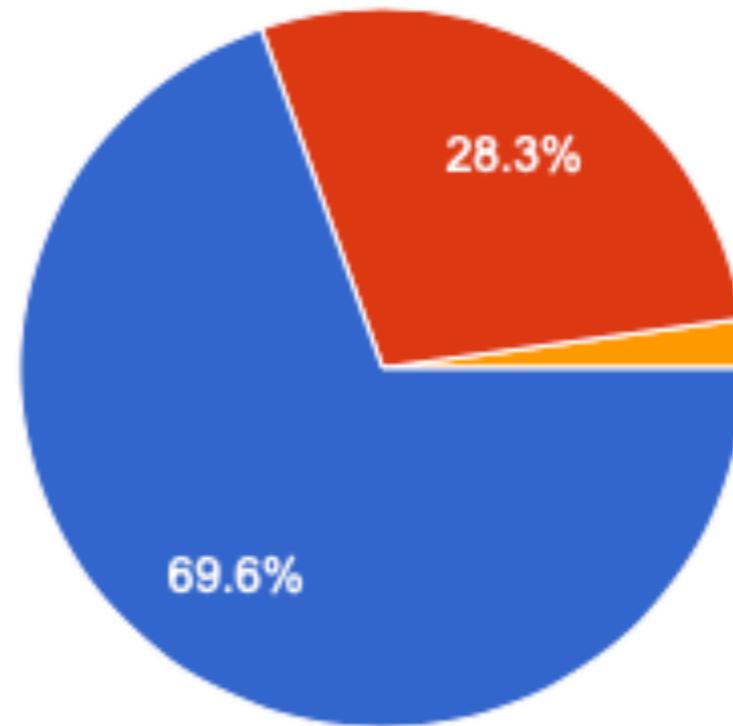


(Please ask me via email/  
slack for details)

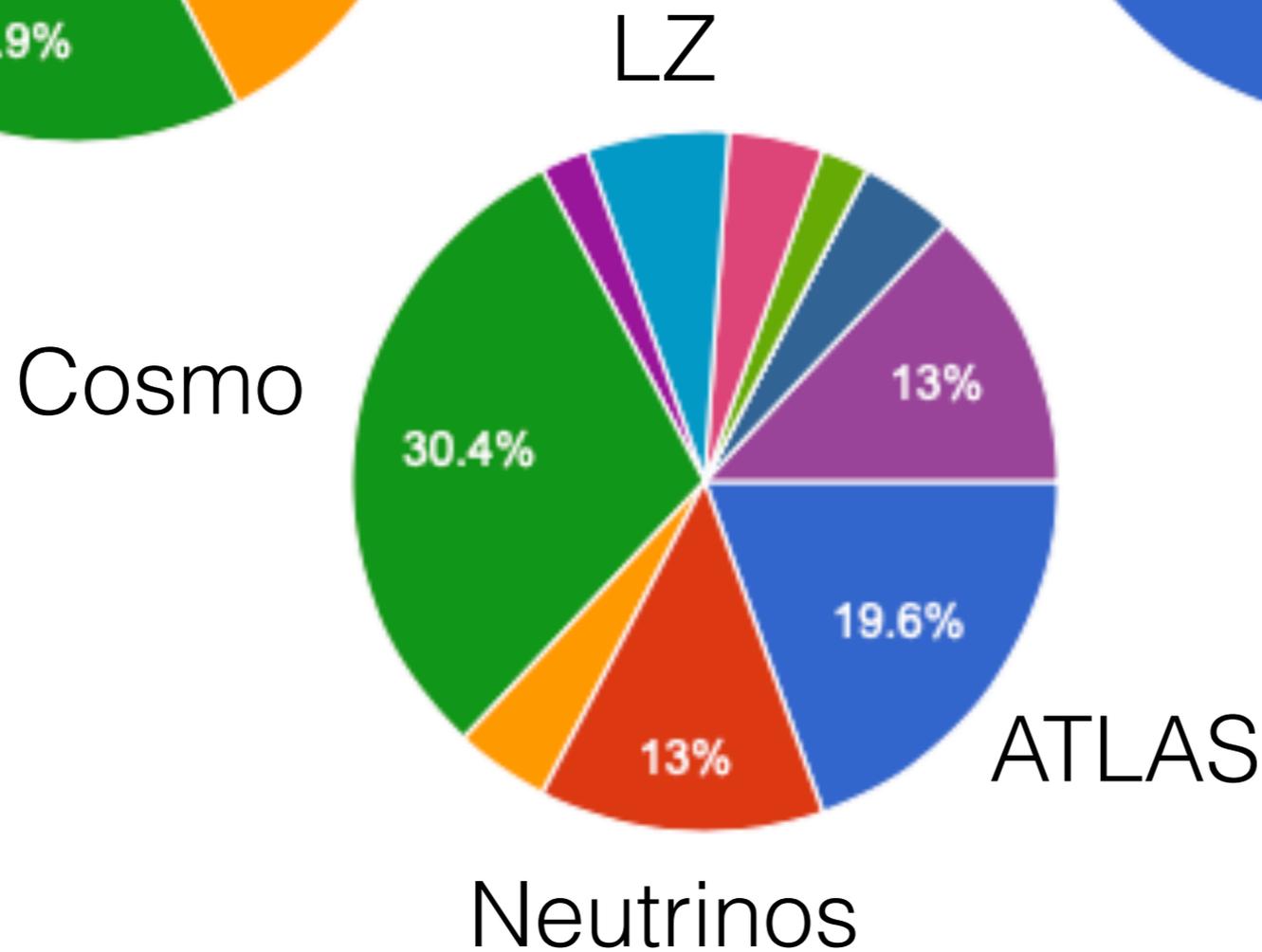
# Survey [50 responses]



- Undergraduate
- Post-baccalaureate
- Graduate student
- Postdoc
- Staff/Faculty
- Other



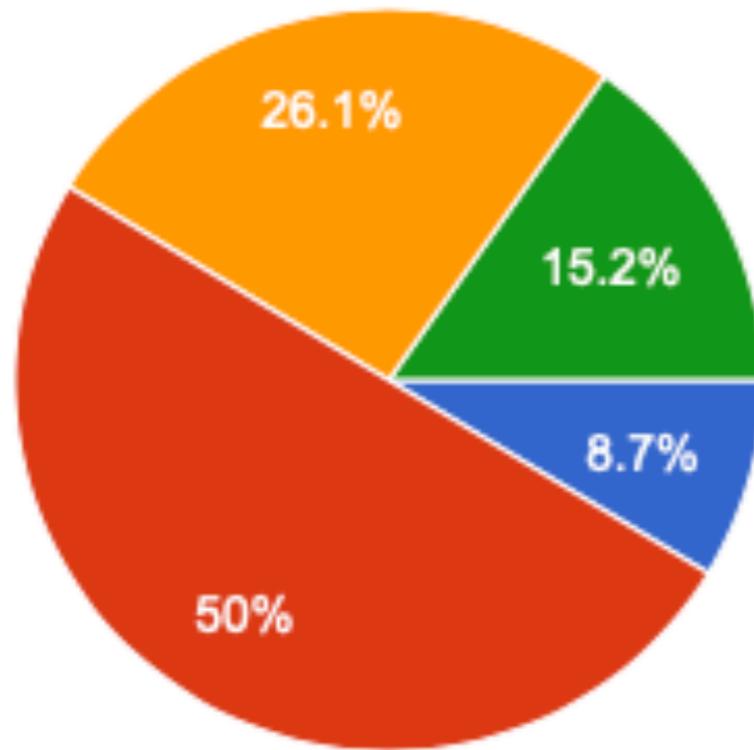
- LBNL
- UC Berkeley
- Other



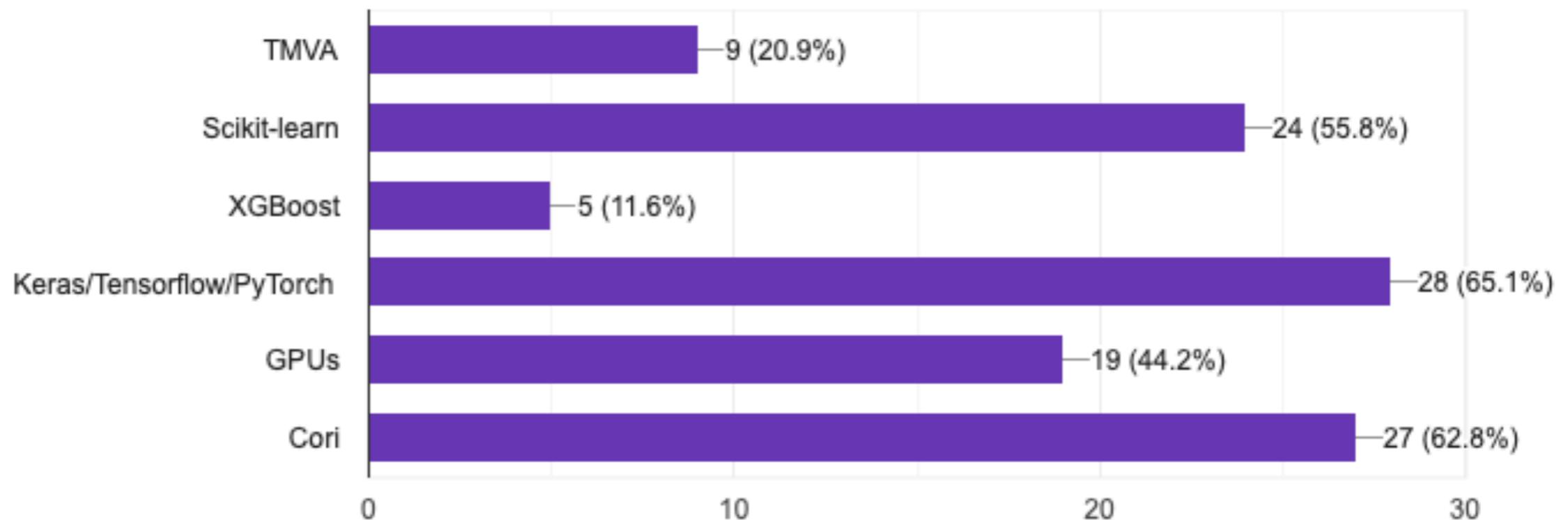
- ATLAS
- DUNE/Daya Bay
- Mu2e
- DESI/Cosmology
- CMB-S4
- LZ/Dark Matter
- Theory
- QIS
- Instrumentation
- PDG
- Other

# Survey [50 responses]

19

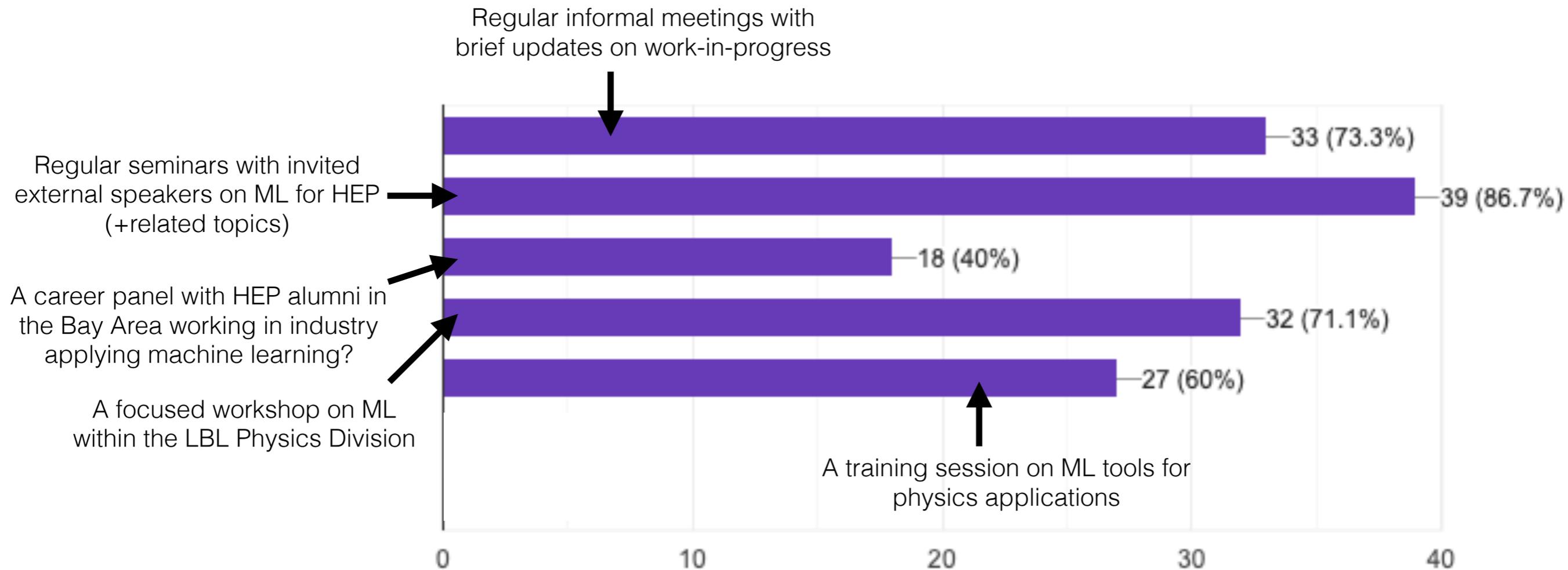


- I do not use machine learning for my research
- A small fraction of my research involves machine learning
- A significant (but less than 50%) of my research involves machine learning
- A majority of my research involves machine learning



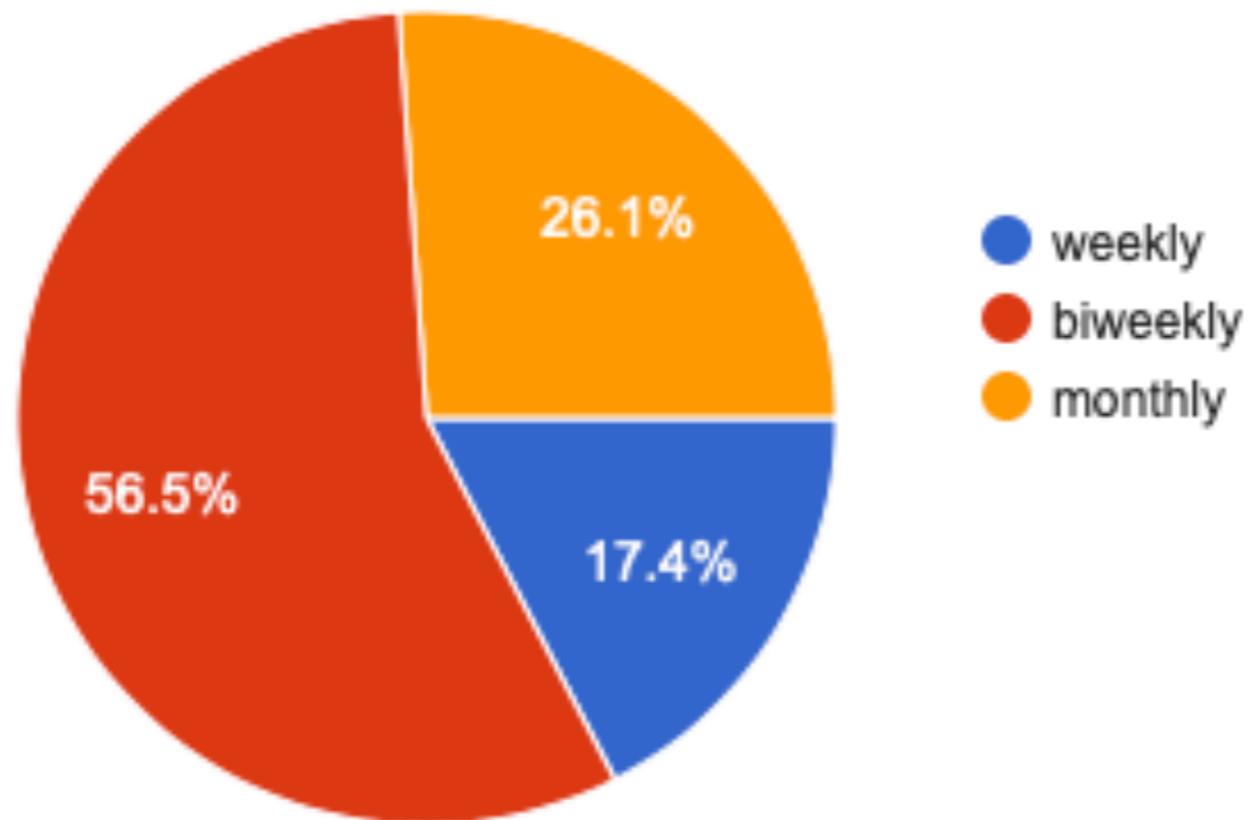
# Group activities

20



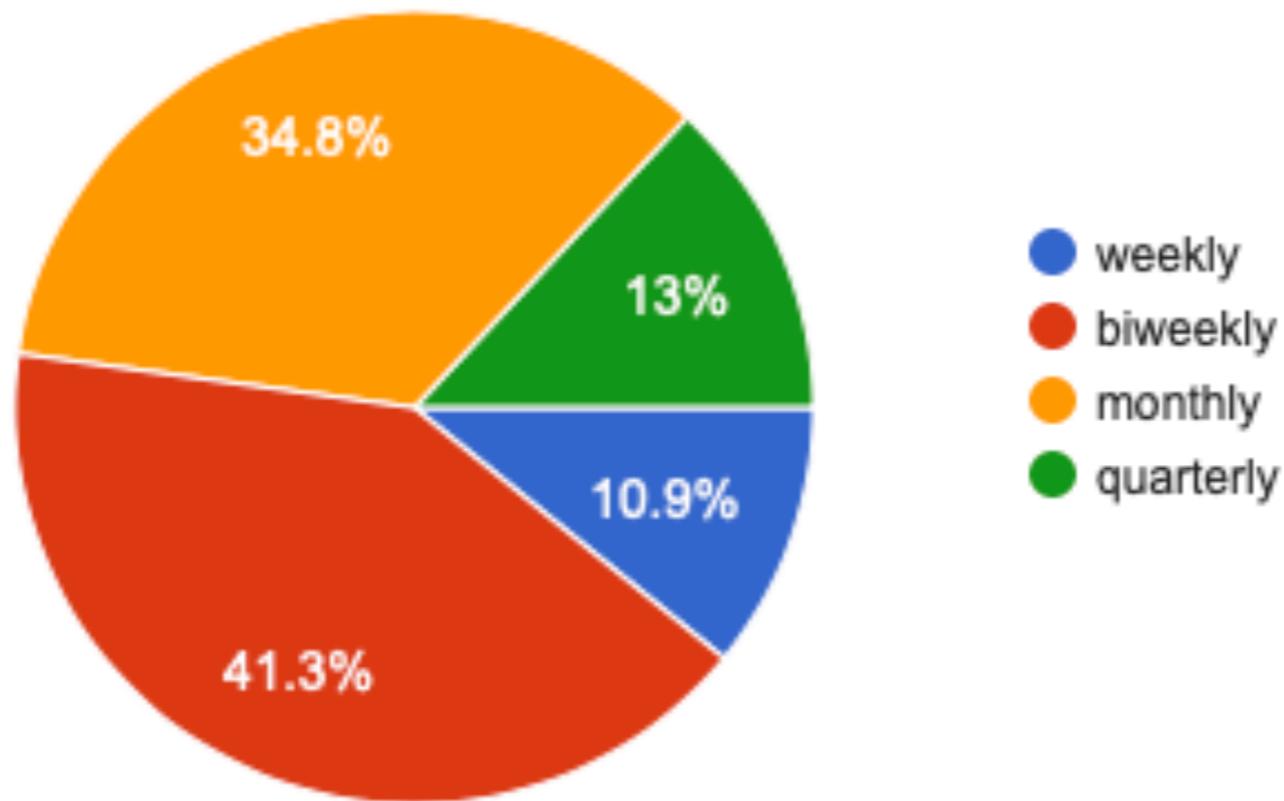
# Group meetings

What is the most useful frequency of informal meetings?



Based on this feedback, we will hold these meetings biweekly!

What is the most useful frequency of invited seminar speakers?



It may not be ideal to have another regular seminar, but I think through a combination of existing series and special ad-hoc events, we can achieve seminars with period ~ 2-4 weeks

What is the most useful frequency of invited seminar speakers?

## Jesse Thaler (MIT) “The Hidden Geometry of Particle Collisions”

November 5 @ 4:00 pm - 5:00 pm

### ABSTRACT:

In this talk, I explain how various concepts and techniques in quantum field theory and collider physics can be naturally translated into a new geometric language. Using the energy mover’s distance, which quantifies the minimal amount of “work” required to rearrange one event into another, we can define a distance between pairs of collider events. This distance can then be used to triangulate the “space” of collider events and rigorously define various geometric objects. Many well-known collider observables, jet algorithms, and pileup mitigation schemes have a simple geometric interpretation, as does the important concept of infrared and collinear safety. Intriguingly, these ideas can be lifted from a distance between events into a distance between theories, with potential relevance for visualizing and interpreting data from the LHC.

<http://physics.lbl.gov/rpm/index.php/events/>

Physics Division Seminar  
(Today!!)

It may not be ideal through a combination of special ad-hoc events, we can achieve seminars with period ~ 2-4 weeks

## Machine Learning and Science Forum

Date: **Monday, November 9, 2020**

Time: **11:00 AM - 12:00 PM Pacific Time**

Location: *Participate remotely [using this Zoom link](#)*

### **First-class machine learning model for Science: Graph Neural Network**

**Xiangyang Ju**, Lawrence Berkeley National Laboratory

This talk will dive deep into the graph neural network for science with a biased focus on exemplary applications from High Energy Physics. Data collected from the detectors at the Large Hadron Collider (LHC) are usually dynamic in sizes, sparse in density and high dimensions in representations. Such data cannot be fully represented by feature vectors, event images or point clouds without scarifying the information loss. However, graphs can naturally represent such science data because of its unlimited expressive power. In addition, the fast advancing geometrical learning generalizes convolutional and recurrent neural networks to datasets with arbitrary geometry and sparsity. It boosts the exploration of deeper hidden graph attributes, the learning of the global and local relational information in the graph. It can even form parameterized message-passing through which information is propagated across the graph, ultimately learning sophisticated graph attributes. The GNN architecture is unique, so are its computational characteristics.

The talk will use the workflow developed by the Exa.TrkX collaboration for reconstructing charged particles' trajectories for the High-Luminosity LHC as an example to demonstrate the power of GNN. Other HEP projects based on GNN will be briefly discussed as well. In addition, the talk will profile GNN's computational features and discuss the implication in terms of real-time data processing. The talk ends with an outlook of GNN for HEP and for science.

Berkeley Institute for  
Data Science Seminars  
(Next Monday!!)

new geometric  
er, we can define a  
ne various geometric  
does the important  
eories, with potential

but I think  
al ad-hoc  
4 weeks

# Seminars



## Berkeley Machine Learning Events

Today ◀ ▶ November 2020 ▾

Print Week Month Agenda ▾

Sun	Mon	Tue	Wed	Thu	Fri	Sat
Nov 1	2	3	4	5	6	7
			1pm Physics Divisior 4pm Jesse Thaler RP			
8	9	10	11	12	13	14
11am Karp Lecture - 11am ML and Scienc						
15	16	17	18	19	20	21
			1pm Physics Divisior			
22	23	24	25	26	27	28
11am ML and Scienc	9am ML for Planetar					
29	30	Dec 1	2	3	4	5
			1pm Physics Divisior			

Events shown in time zone: Pacific Time - Los Angeles

+ Google Calendar

<https://www.physics.lbl.gov/machinelearning/seminars/>

# Seminars

26

## Berkeley Machine Learning Events

Today ◀ ▶ November 2020 ▾

Print Week Month Agenda ▾

Sun	Mon	Tue	Wed	Thu	Fri	Sat
Nov 1	2	3	4	5	6	7
				1pm Physics Divisor 4pm Jesse Thaler RF		
8	11am 11am					
15						
22	11am					
29	30	Dec 1	2	3	4	5
			1pm Physics Divisor			

Events shown in time zone: Pacific Time - Los Angeles

 GoogleCalendar

If you have a suggestion for a speaker, please let me know and I will look into inviting them! It is especially “easy” now that all visits are virtual.

~50% of respondents would be interested in an ML101 training and ~80% would be interested in a training on how to do ML on Cori.

There is already a lot of great material that exists so I will find a way to connect our group with existing resources and will also look into a dedicated training (volunteers are more than welcome!).

*Other ideas: optimizing ML on Perlmutter, interpretability and uncertainty quantification, parallel computing with GPUs, the art-versus-science of building complex ML workflows, how to interpret loss curves, biases from training data, physics-aware learning,...*

Unlike most other areas of fundamental physics, there are many researchers who take their skills and apply them in industry.

The Bay Area is the worldwide capital of ML for industry and so many of these people are nearby.

Sometime soon I will try to organize an event (e.g. a career panel). While I know many people in the area, please send me names of people you know!

Members of the Physics Division (including affiliates) are invited to join the ML group's directory. Everyone is invited to participate in the group's activities, and we encourage division members who are actively developing, adapting, or deploying machine learning in their research to join the directory so we can be most effective in connecting researchers.

The screenshot shows the Berkeley Lab Physics Division website. At the top left is the Berkeley Lab logo and the text "Physics Division". Below this is a navigation menu with links for "Directories", "ATLAS", "Lepton Flavor", "Cosmology", "Dark Matter", "Theory", "QIS Comp", "Machine Learning", and "Detector R&D". There are also links for "PDG" and "Division Office". The main content area is titled "Machine Learning Home" and "Machine Learning Directory". Below this is a sub-menu with tabs for "Scientific Staff", "Term Staff/ Postdoctoral", "Students", "Engineering/ Technical/ Admin", "Affiliates", and "All Members". The "Scientific Staff" tab is selected. Below the sub-menu is a table with columns for "Name", "Groups", "Phone", and "EMail". The first row of the table shows a profile for Benjamin Nachman, with a small photo of him, his name "Nachman, Benjamin", his groups "ATLAS, QIS, Machine Learning", his phone number "(510)486-4606", and his email "BPNachman@lbl.gov".

Name	Groups	Phone	EMail
 Nachman, Benjamin	ATLAS, QIS, Machine Learning	(510)486-4606	BPNachman@lbl.gov

Join here: <https://forms.gle/oYAEExwH6rf2fK4CeA>

# Future meetings



30

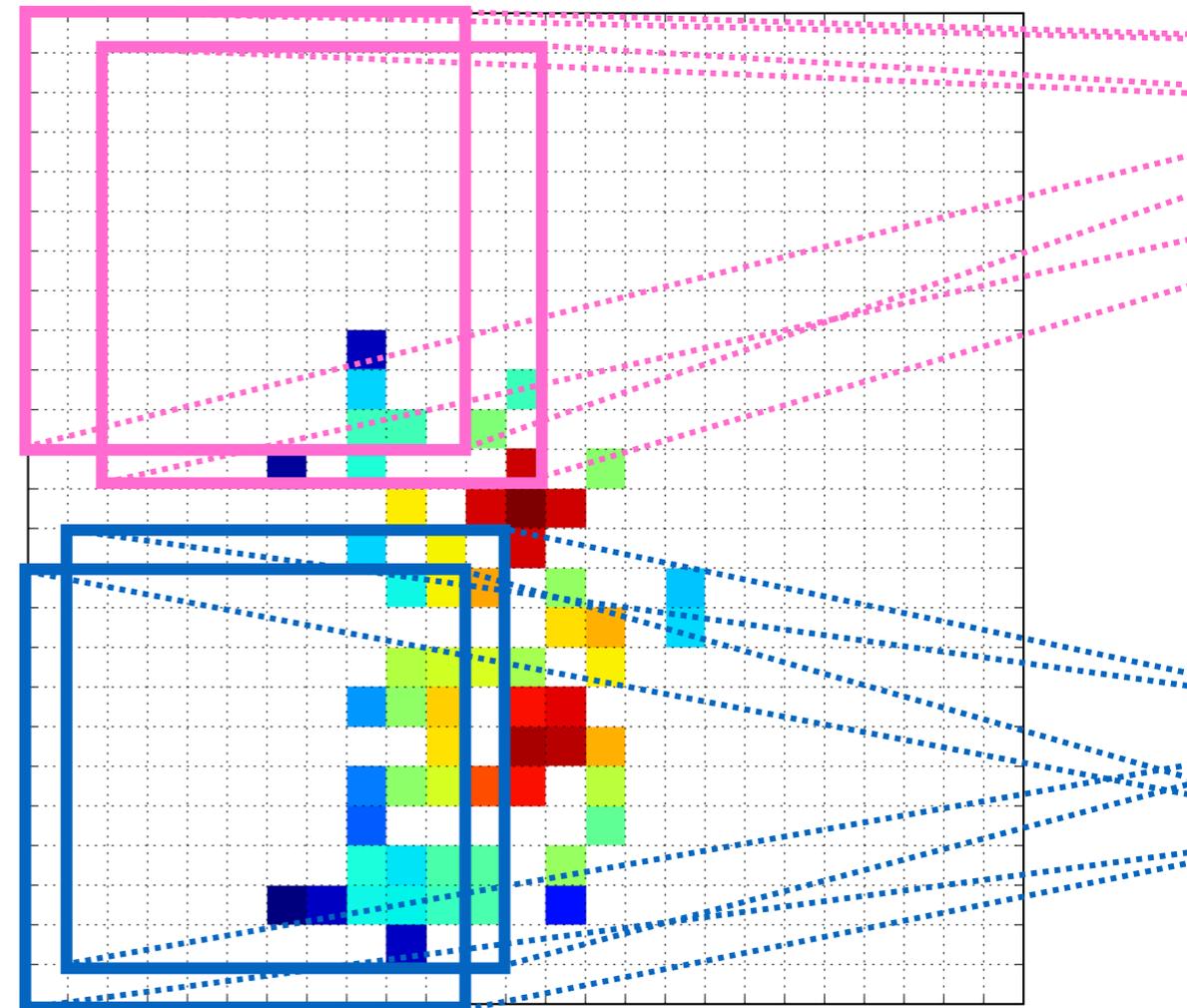
I was envisioning (feedback welcome!) to have a brief news roundtable followed by ~2 informal talks about work-in-progress (~15 min or so). [Here is a sign up sheet](#) for the meetings for the rest of the year.

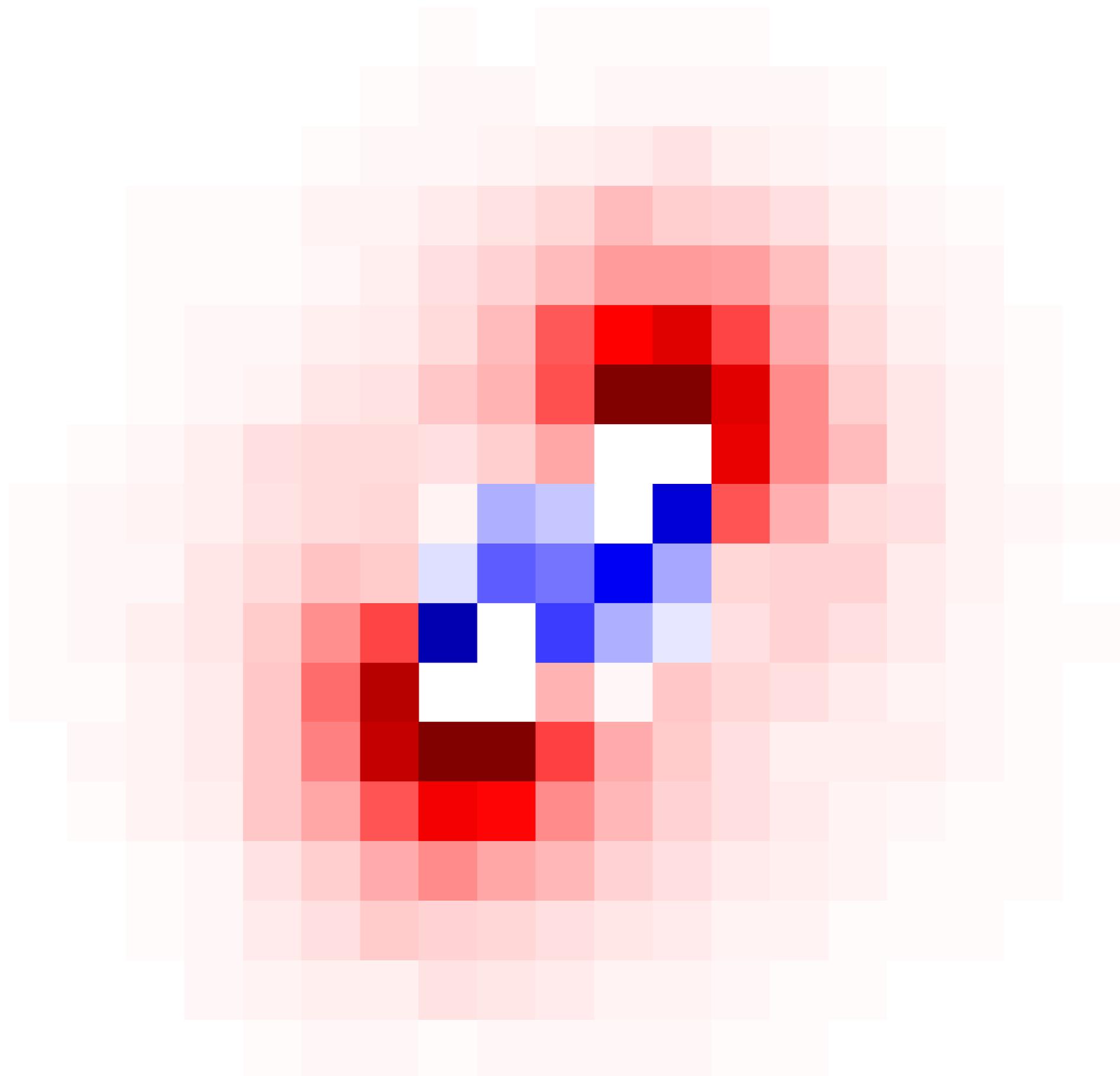
There was also interest in a workshop of some kind. It would be great to hear your feedback and if anyone is interested in volunteering to help organize, please let me know.

The Physics Division (+ friends) is well-positioned to be the leader of machine learning for fundamental physics

This new group is designed to connect and empower researchers in this area

We are really just getting started, so your ideas and contributions are essential!





Fin.