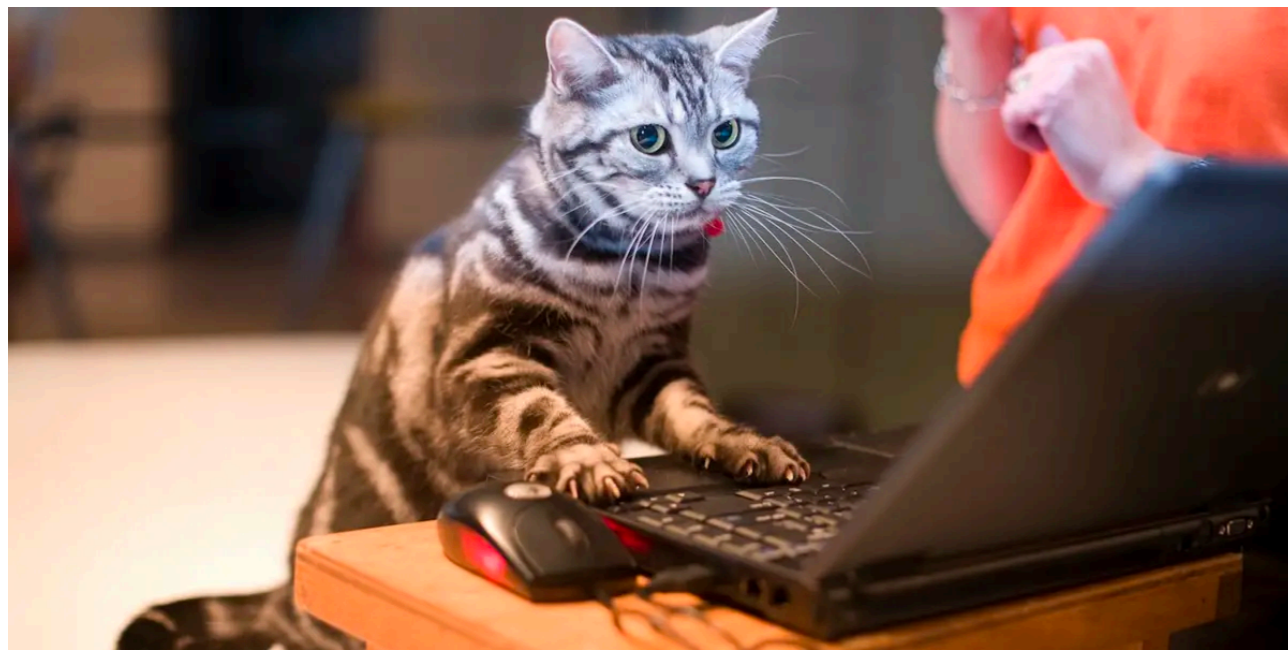
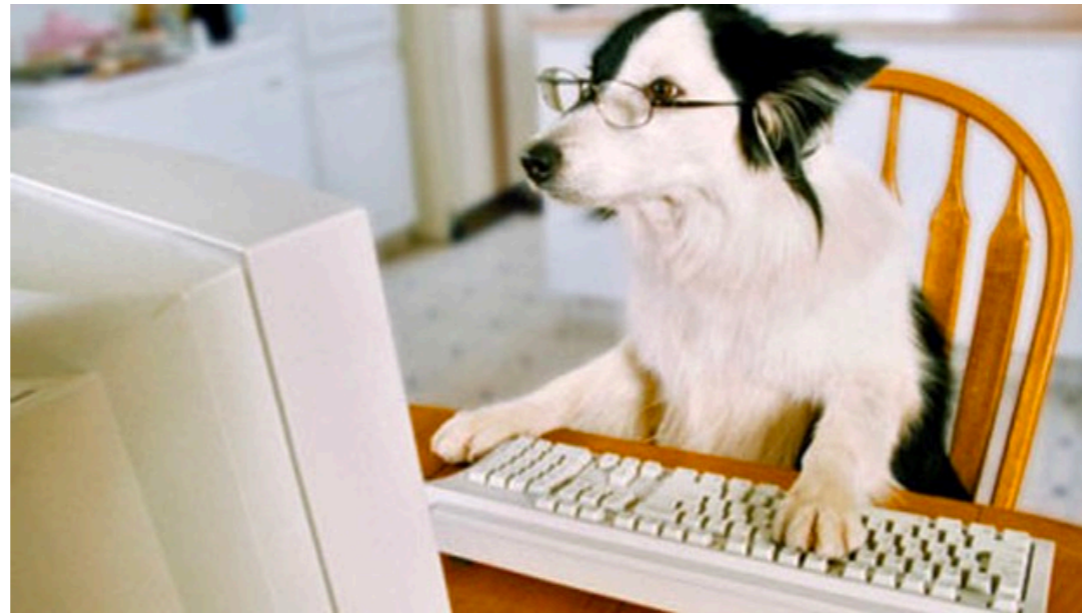




ML@MIT

P. Harris, E. Katsavounidis, M. Coughlin



Despite differences in language, there is a common theme

ML@MIT

P. Harris, E. Katsavounidis, M. Coughlin

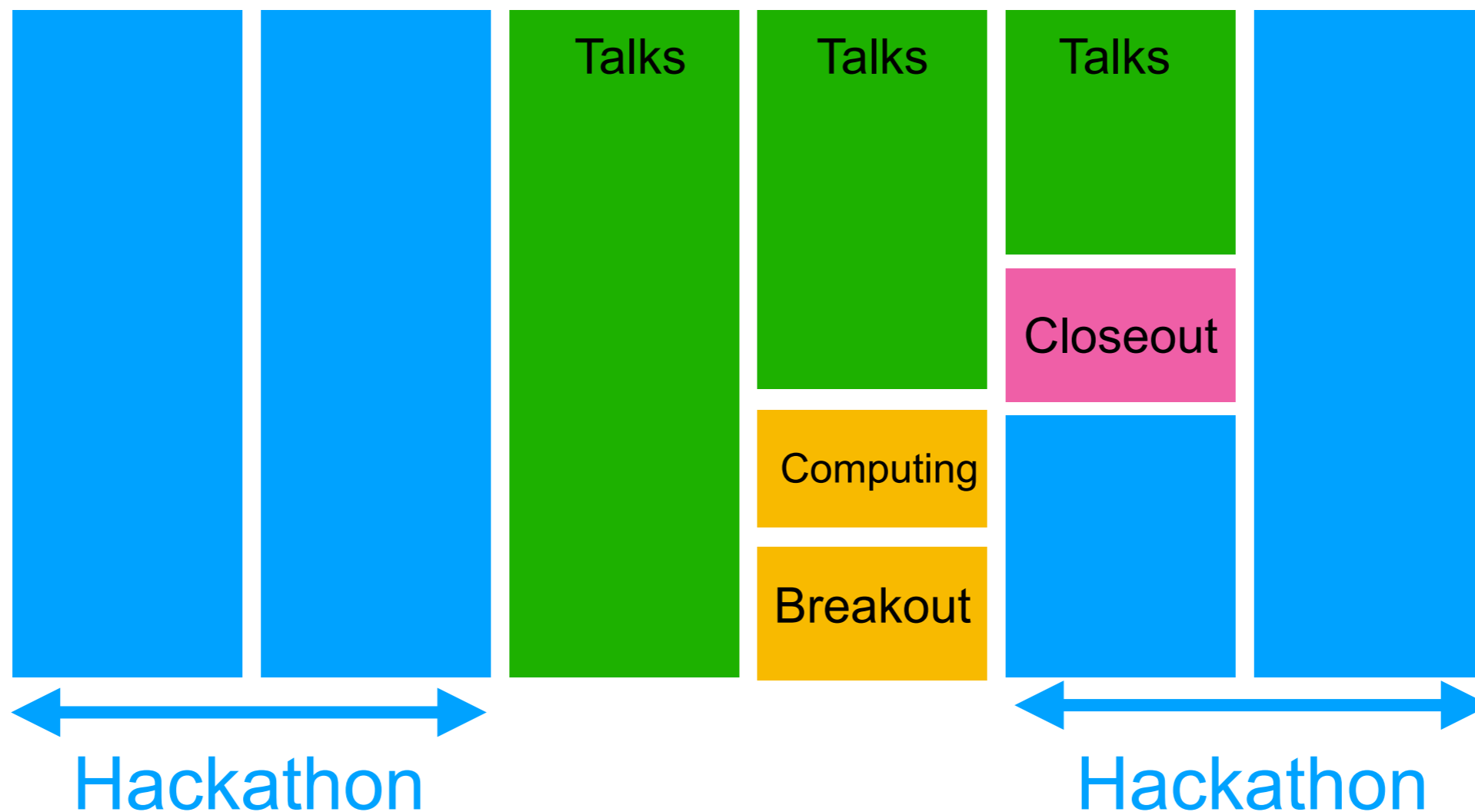
Welcome!

- We are excited that you have all come to the meeting
 - We hope to have a lot of interesting discussions
- We thank you for taking the time to come to this meeting
- This meeting is the result of ongoing discussions
 - Many of you are involved in our efforts
 - ▶ Astronomy, Gravitational Waves, Particle Physics
- Despite coming from many different venues
 - We all share a common goal and vision
 - We often share computers and code

Agenda

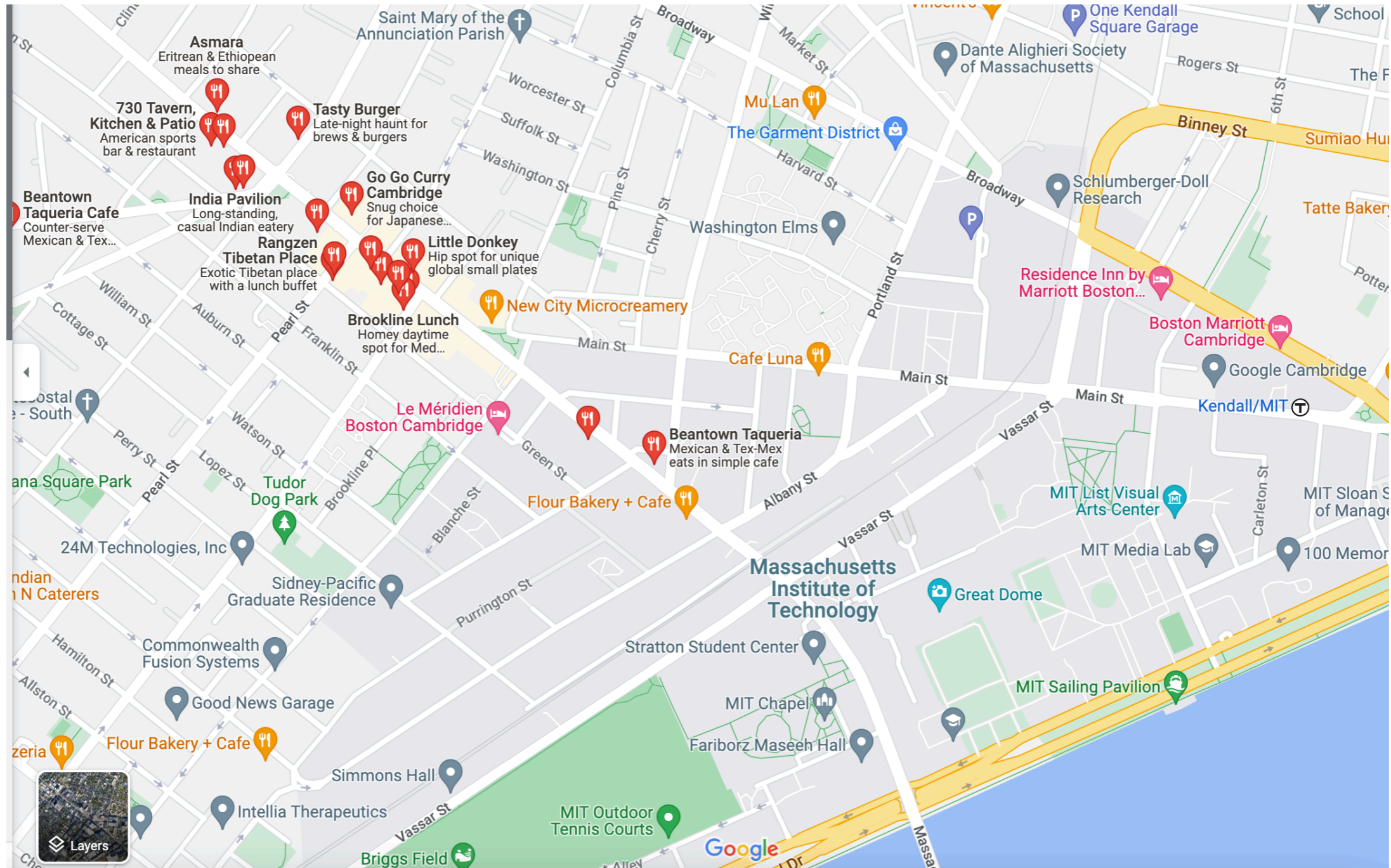
Full Agenda is available here :

<https://indico.cern.ch/event/1224718/timetable/#20230130>

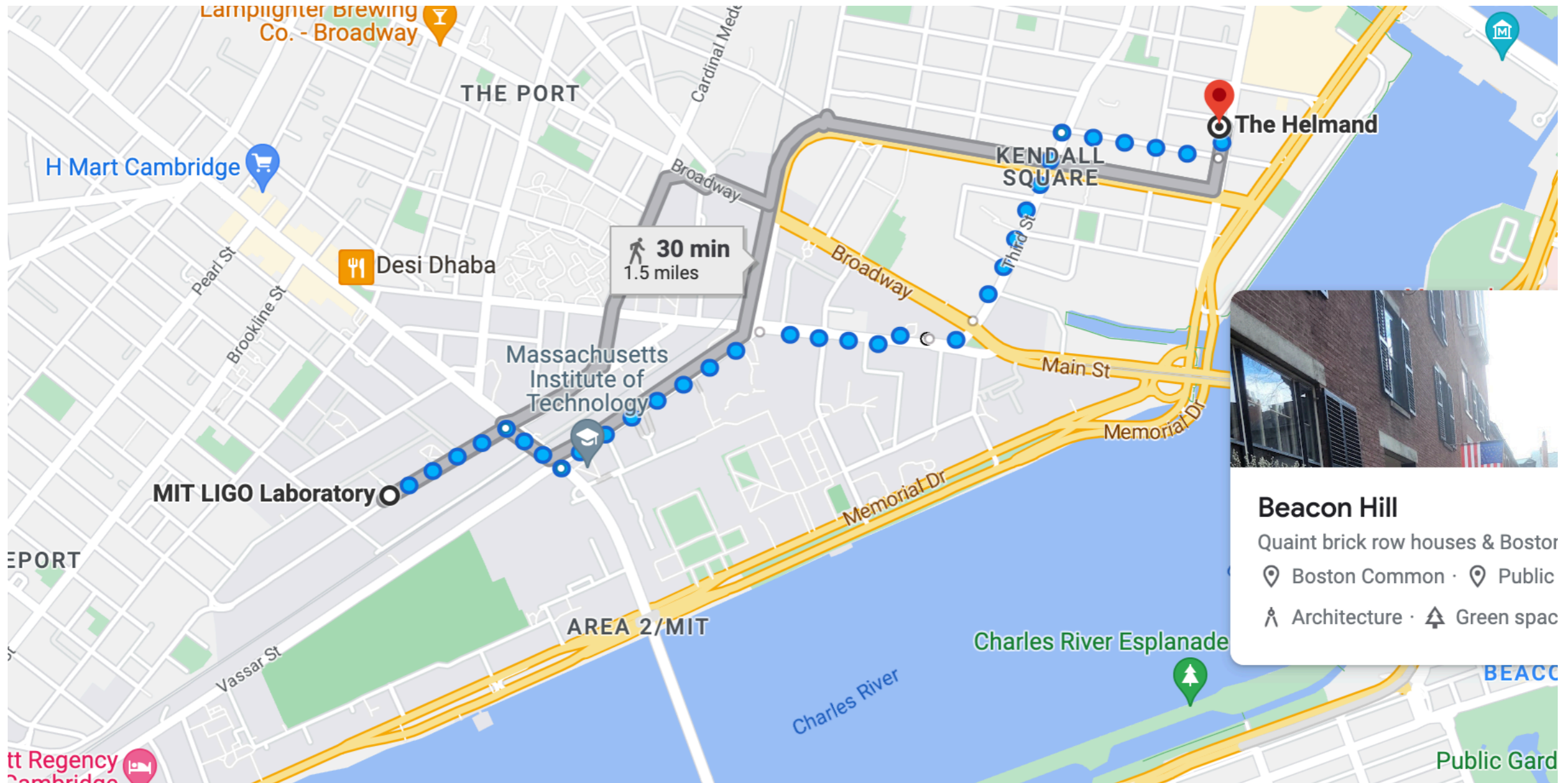


Many of you are following Hackathon
Its focus on the GW inference engine

Cambridge

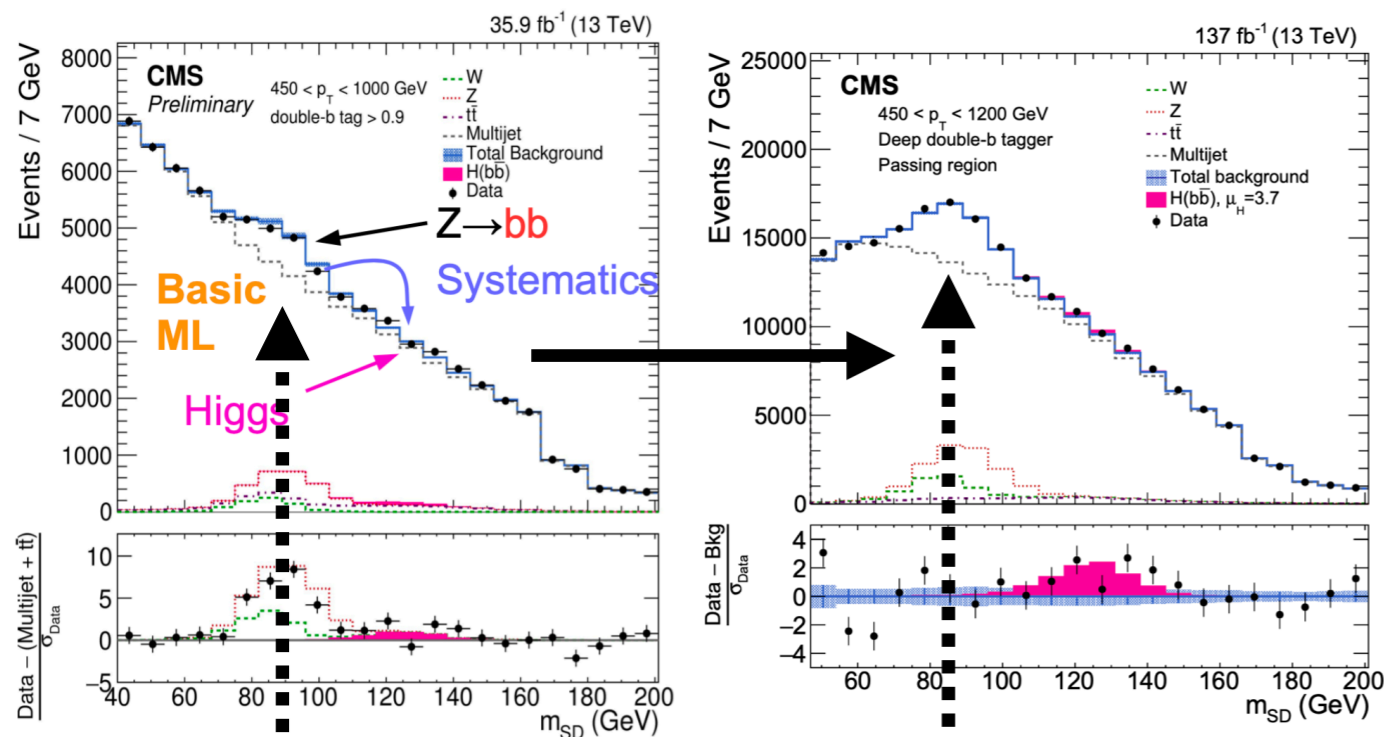


Dinner Location



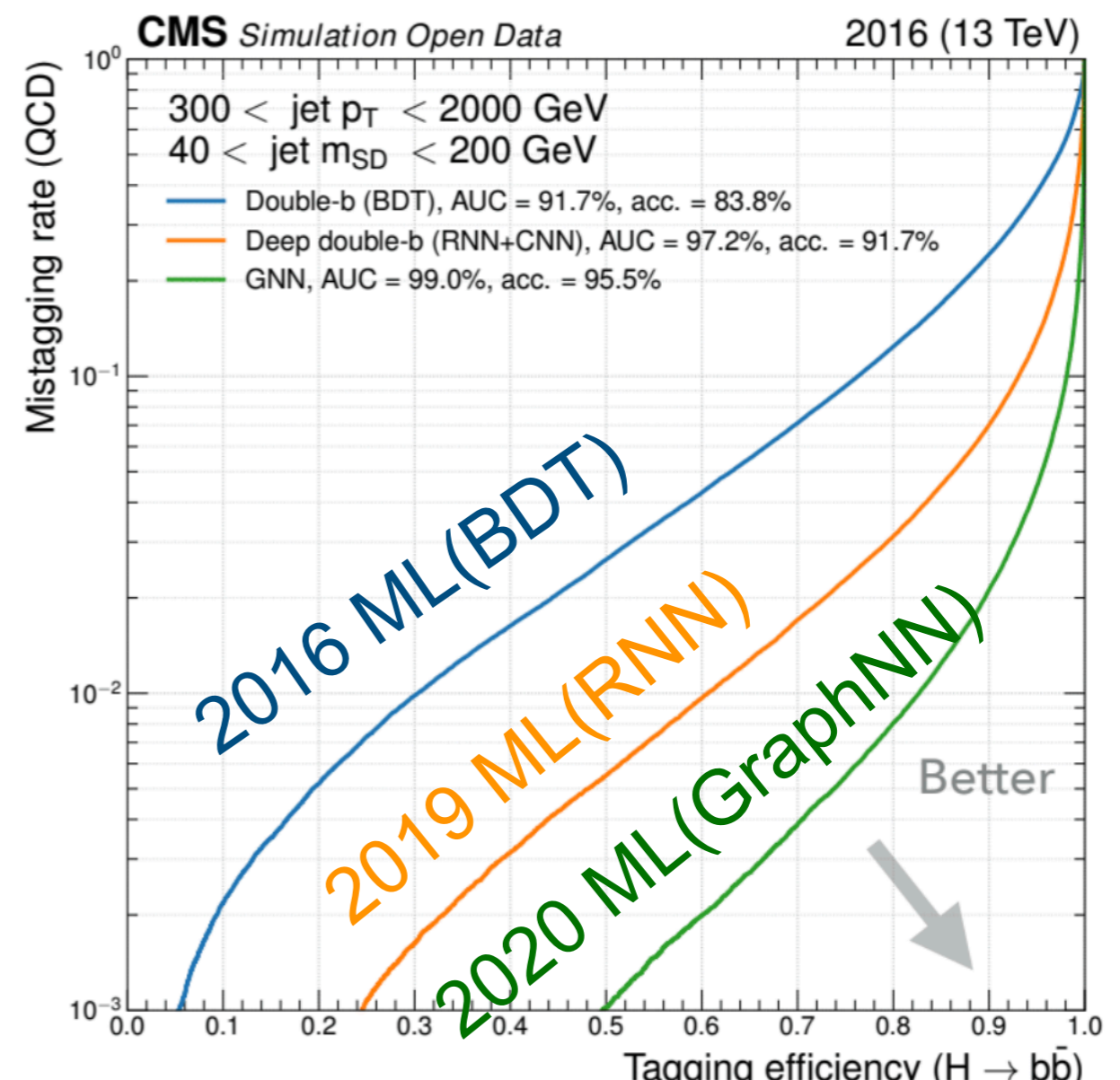
An Angle on AI revolution

- Things are starting to change in the way we compute
 - ML algorithms have the ability to go beyond algorithms
 - ▶ This is also b/c GPUs have helped to parallelize computation



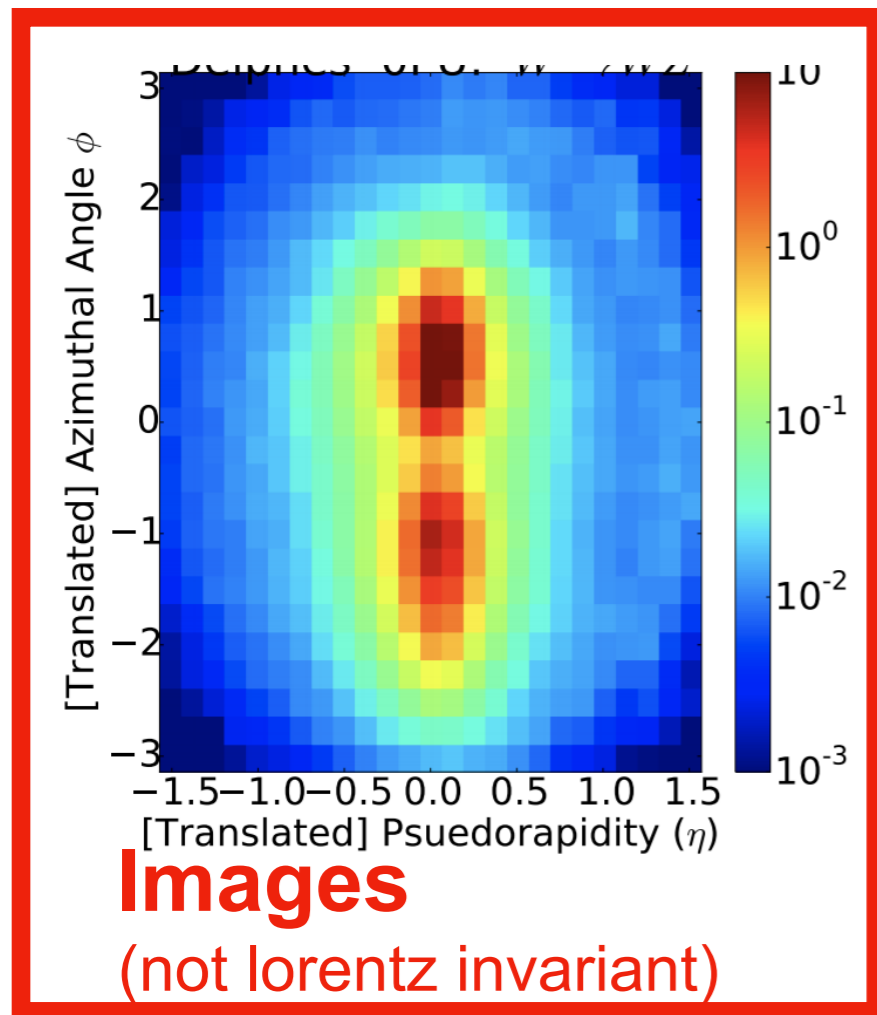
Small ML
Small Peak

Big ML
Big Peak

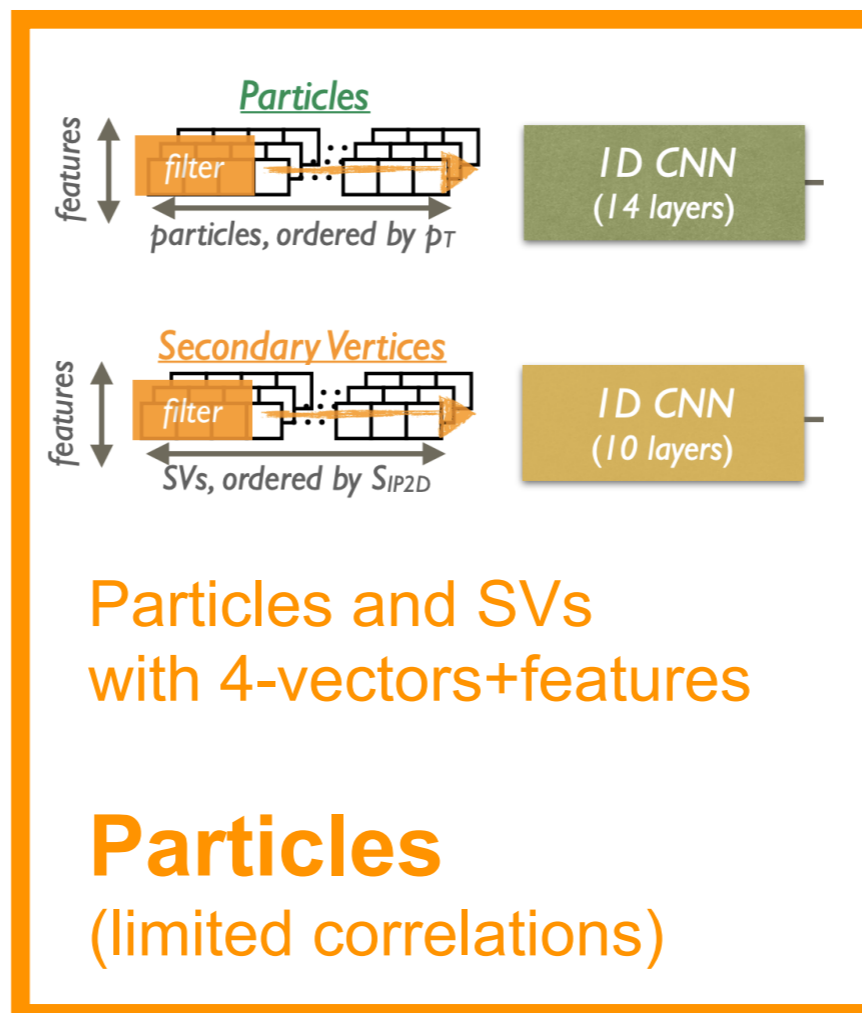


Deep Learning Progression

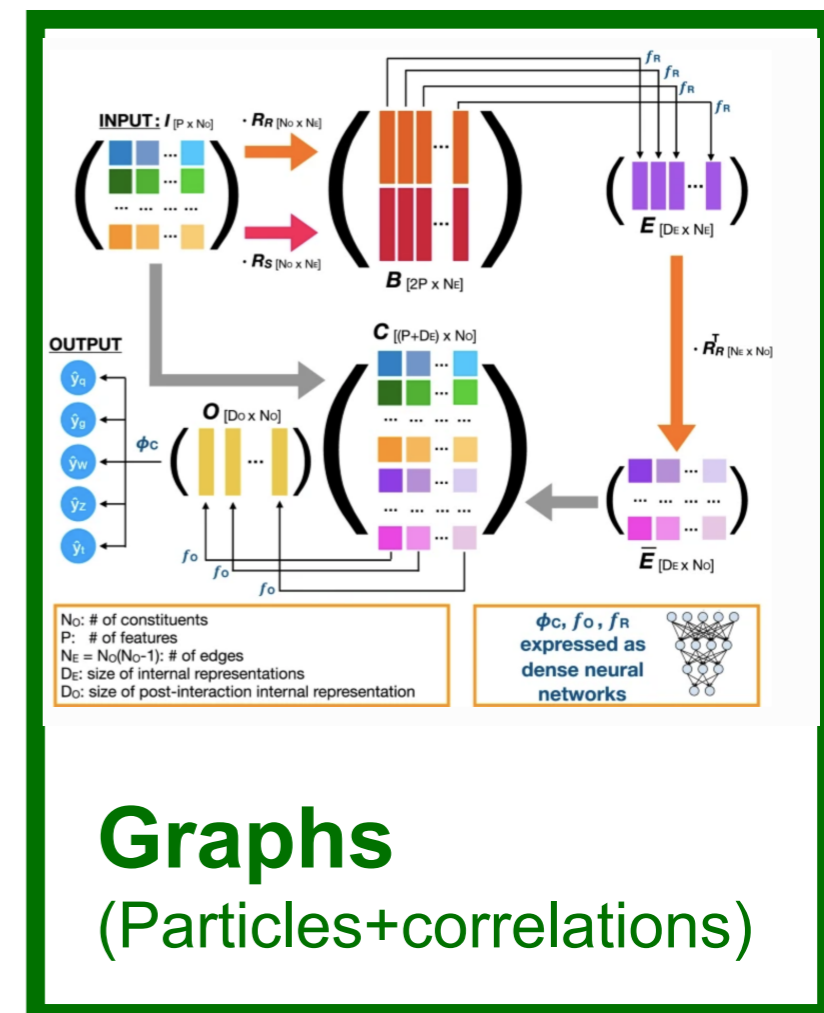
2016



2018



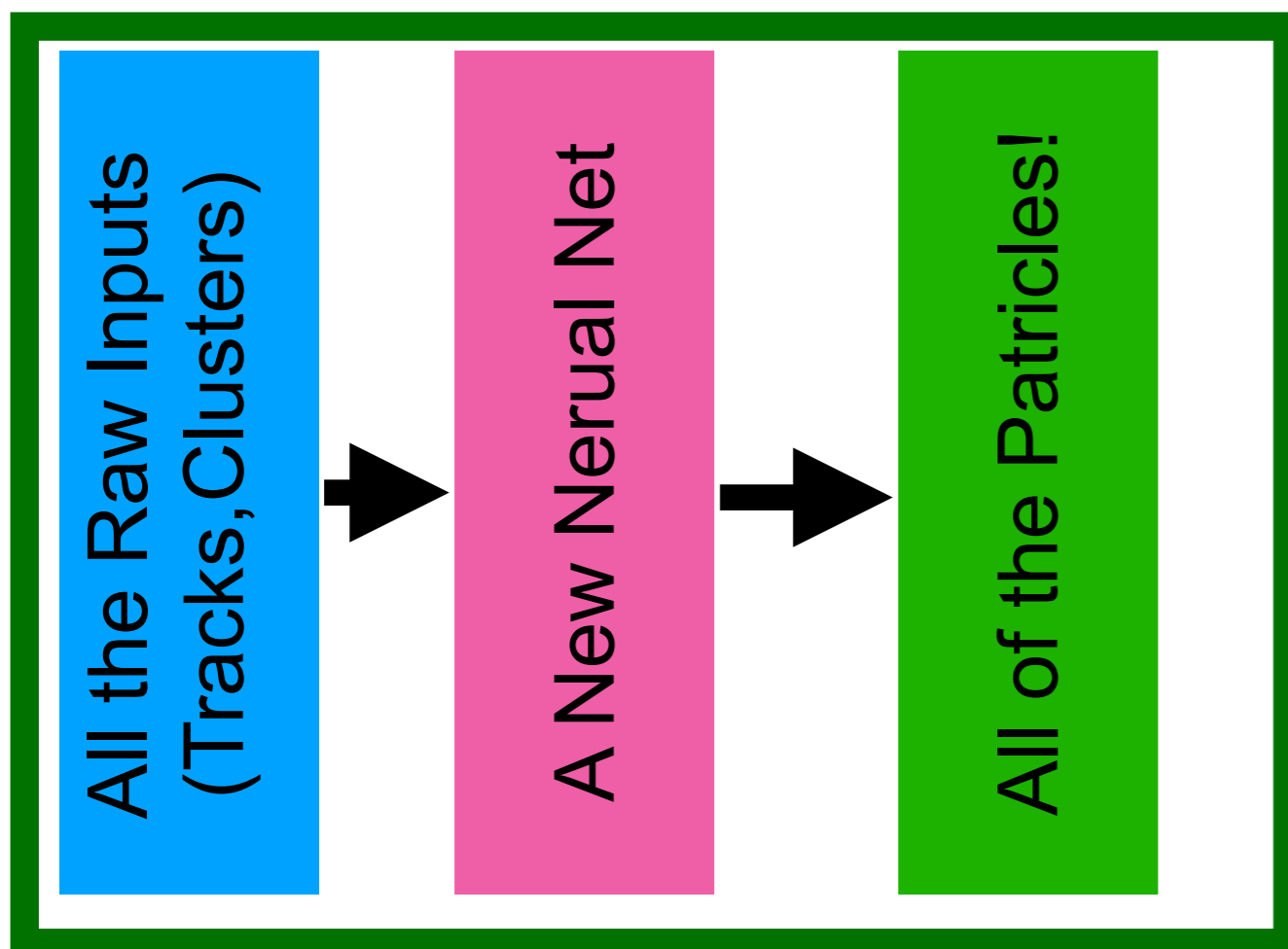
2020



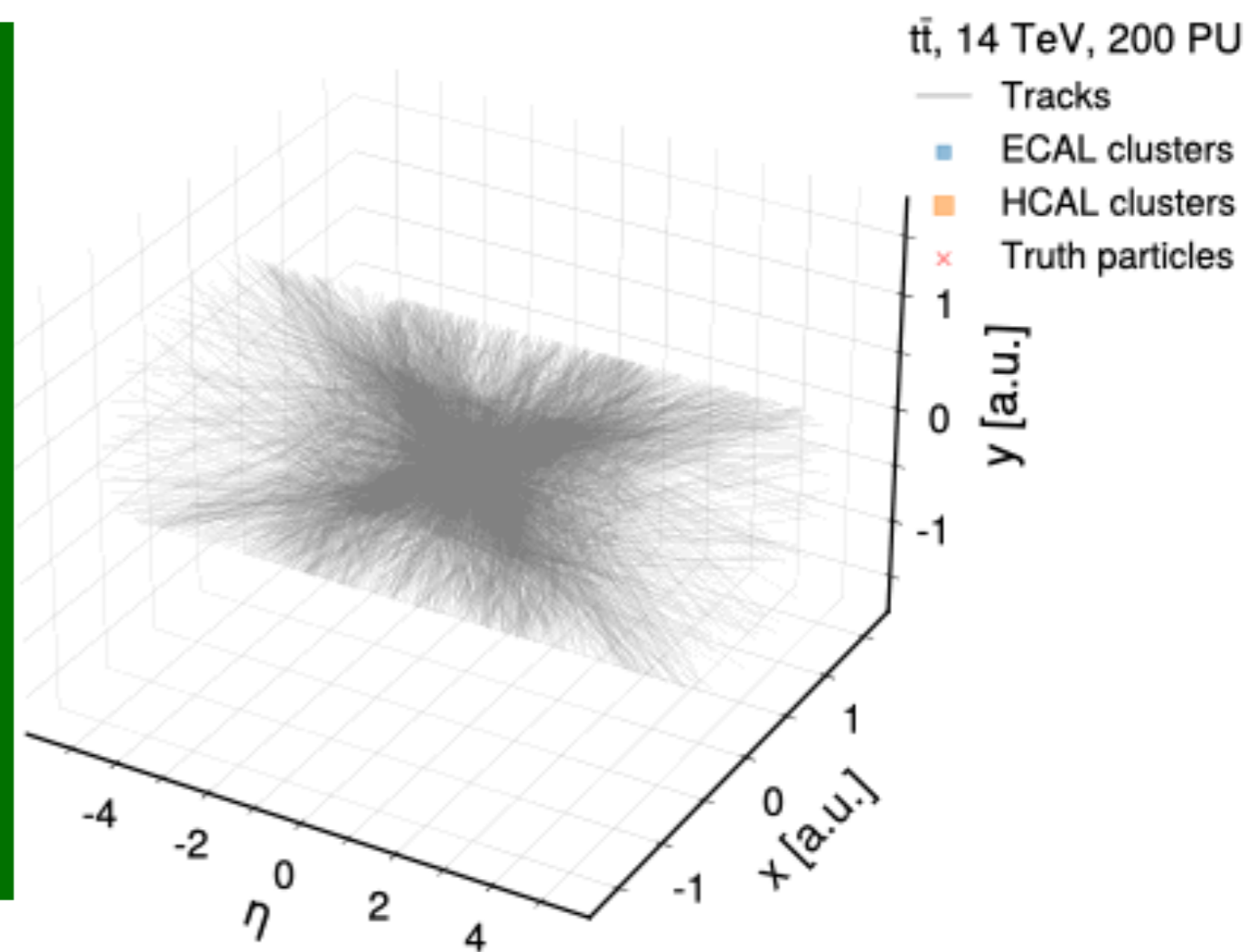
Progressively moving towards use of more info

What does this mean?

- Inevitable that our algorithms will become progressively larger

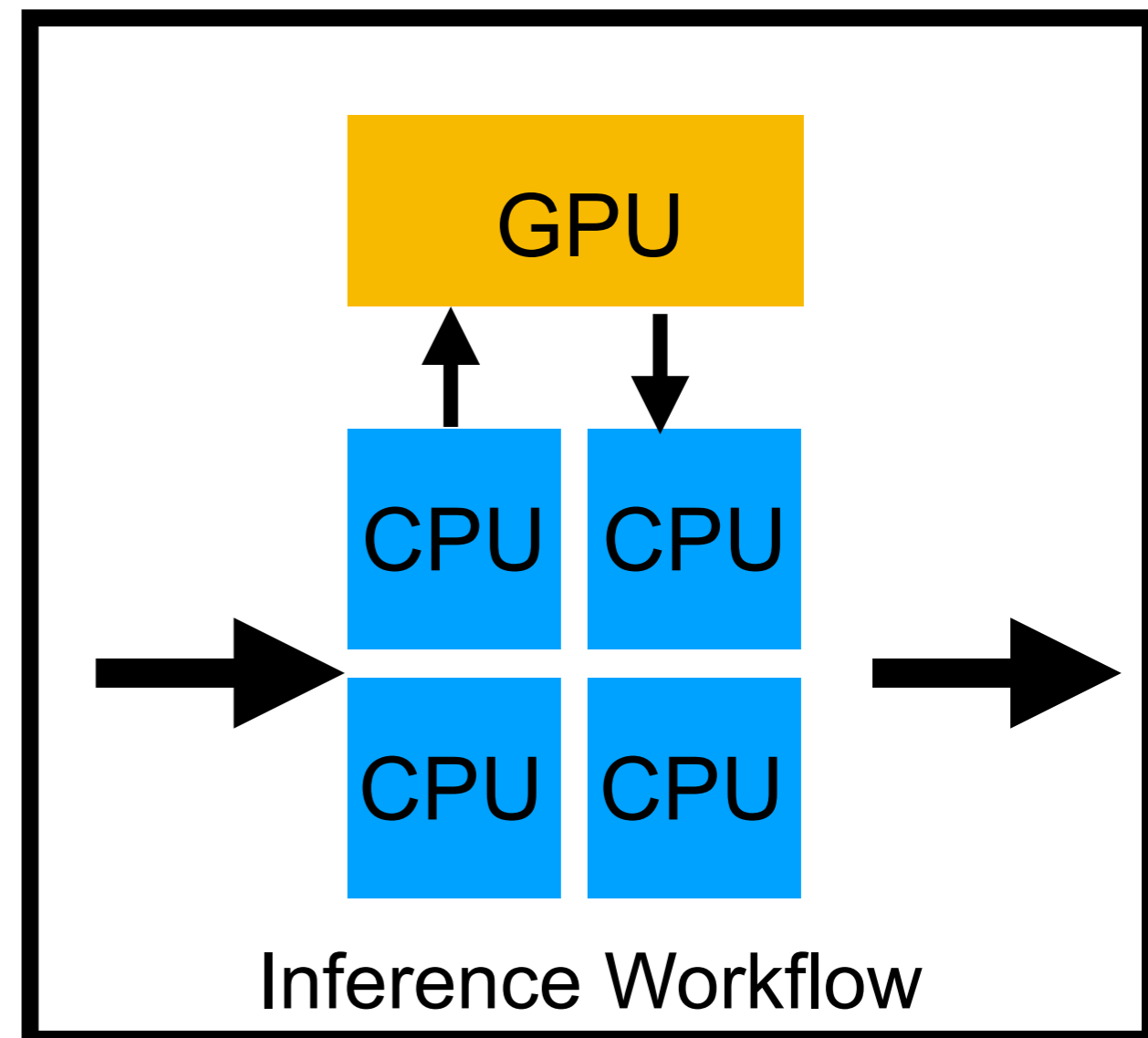
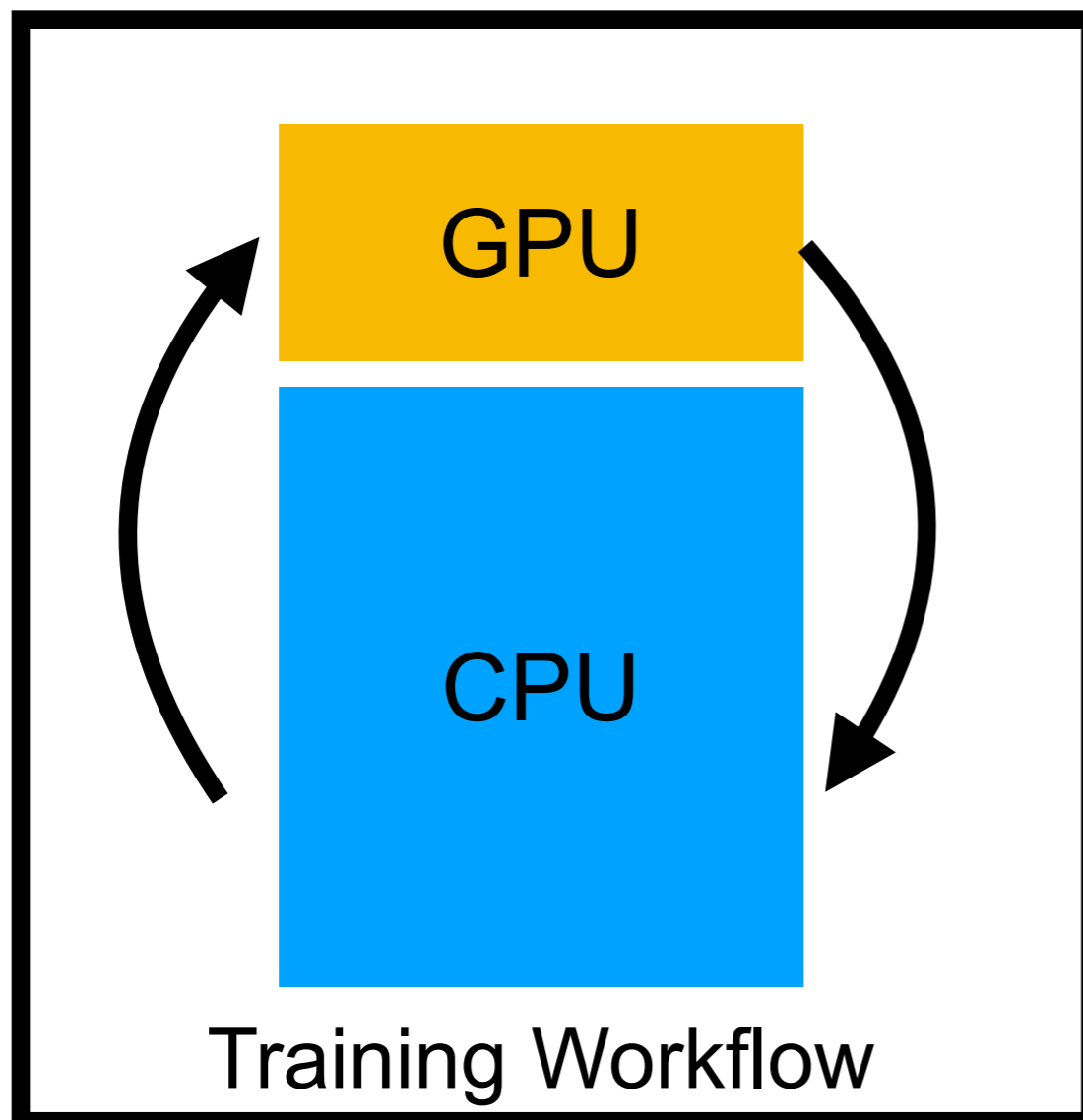


All particles in on fell swoop



Algorithm Needs

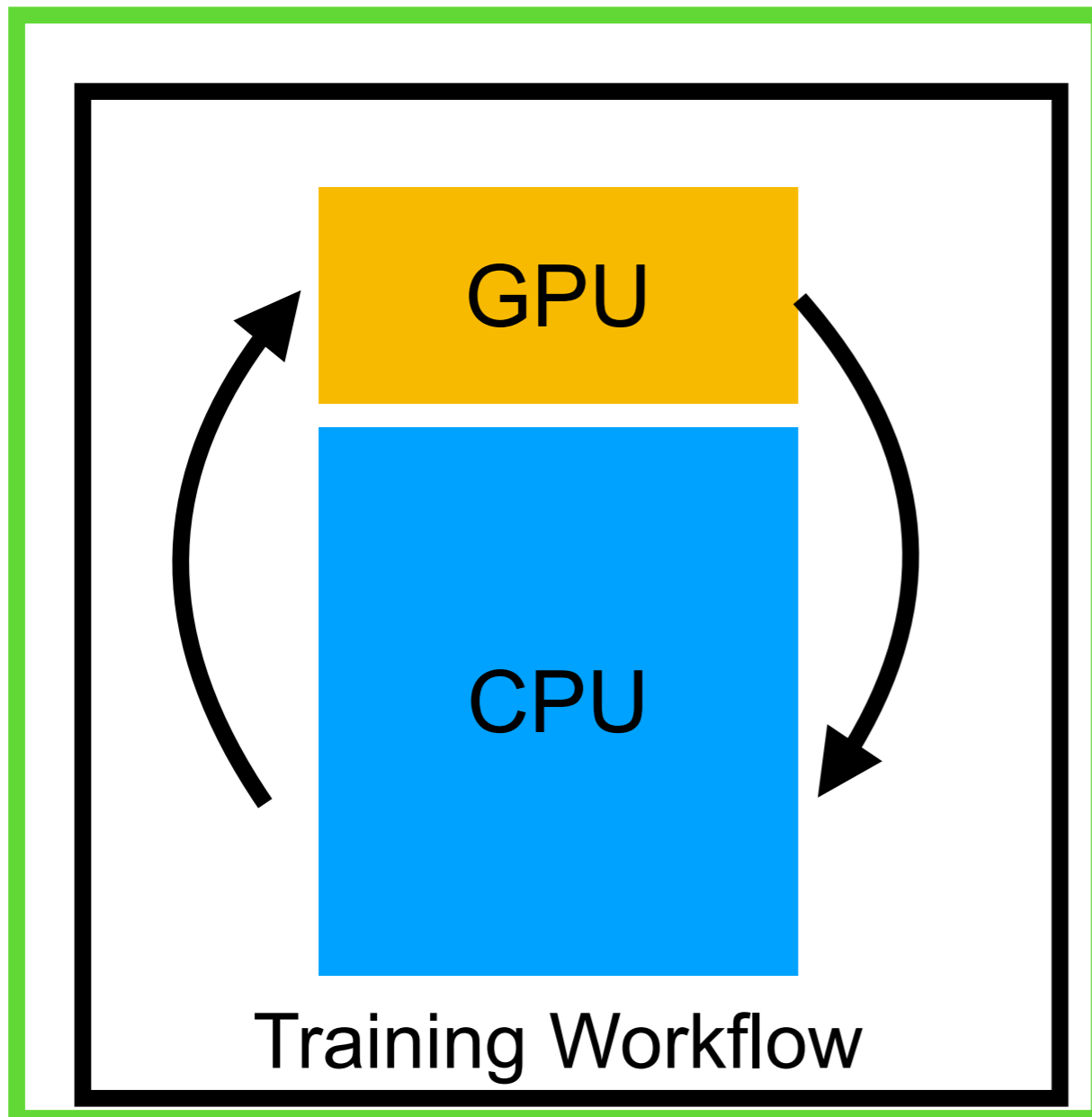
- With the development of AI algorithms we need two things
 - Training and Testing
 - Processing power to run on the data



Algorithm Needs

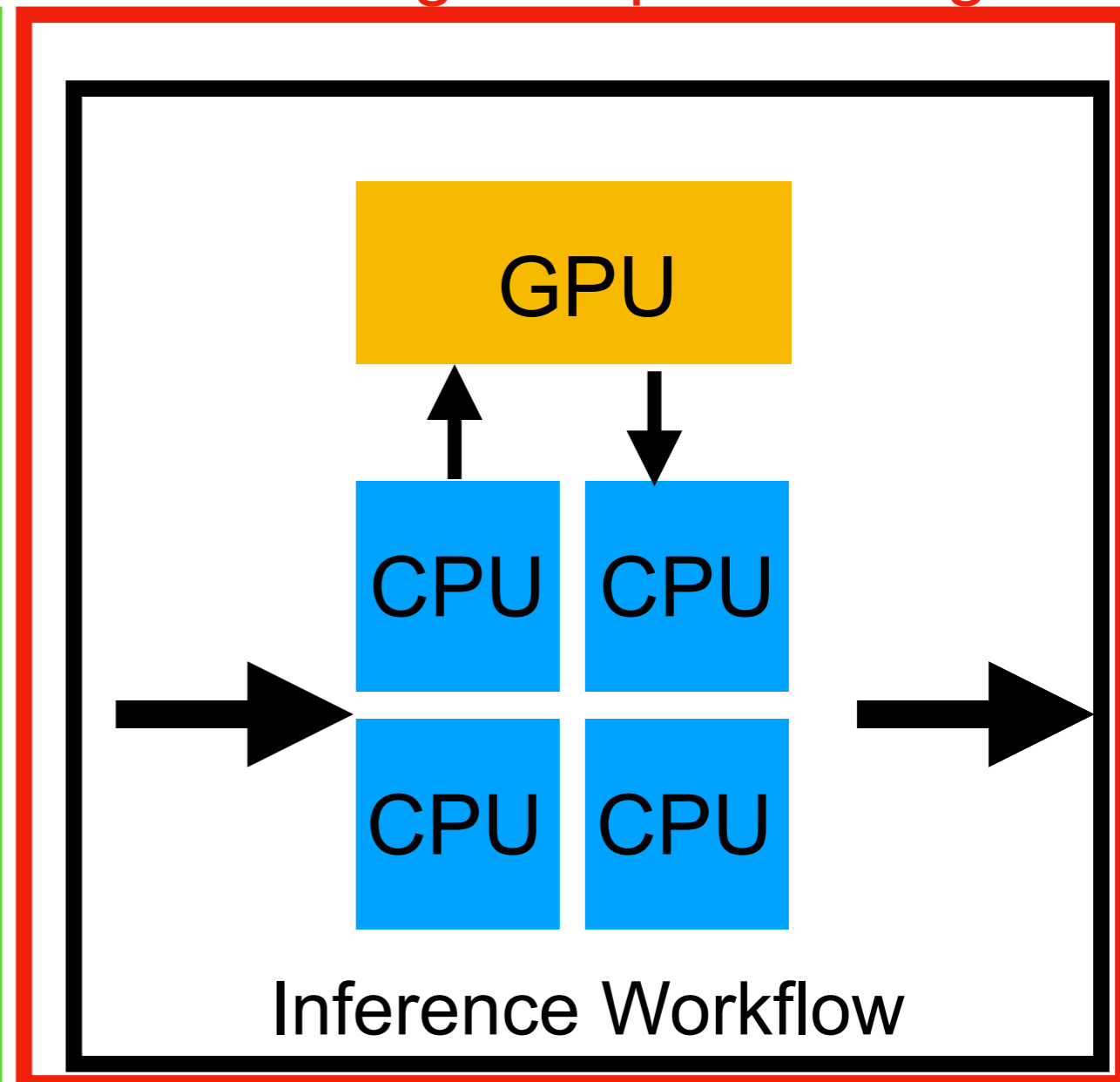
Solved

Big HPCs dump as many GPUs as they possibly can in a room
Aim for the maximum compute



What we need

Requires Dynamic allocation to balance GPUs and CPUs focus is on dealing with processing



Algorithm Needs

Solved

Big HPCs dump as many GPUs as they possibly can in a room
Aim for the maximum compute



Training Workflow

What we need

Requires Dynamic allocation to balance GPUs and CPUs focus is on dealing with processing



Inference Workflow

Why are we here?

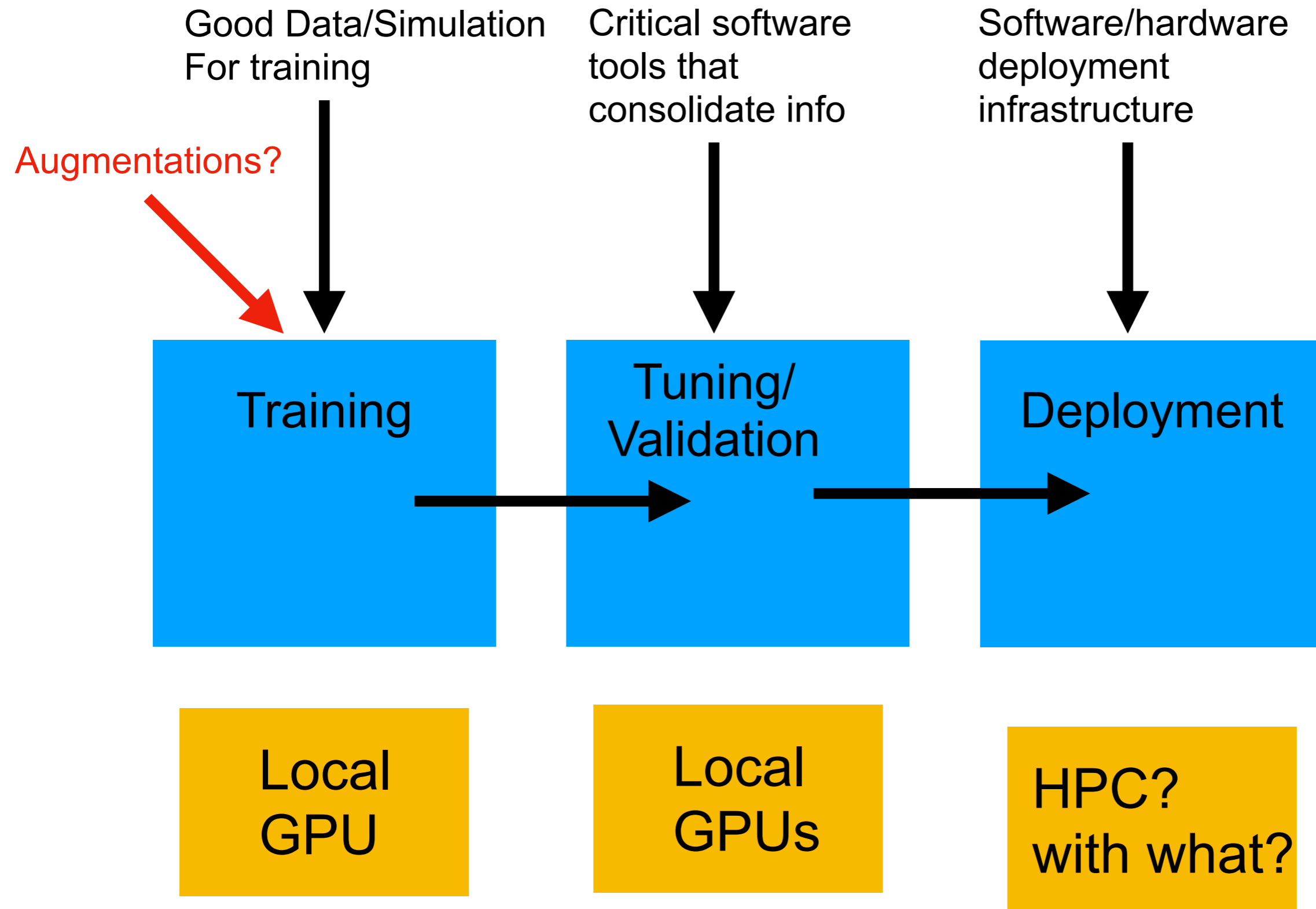
- We would like to highlight commonalities across domains
 - **Computing demands**
 - ▶ Looking for computing infrastructure for ML science deployment
 - ▶ We can assemble a list of common hardware(+tools)
 - **Software Stack**
 - ▶ With all ML algorithms aim for a set of core software tools
 - ▶ Need for good tools to validate and deploy algorithms
 - **ML Problems**
 - ▶ Across the domains similar ML problems exist
 - ▶ Highlighting the similarity is critical

Why are we here?

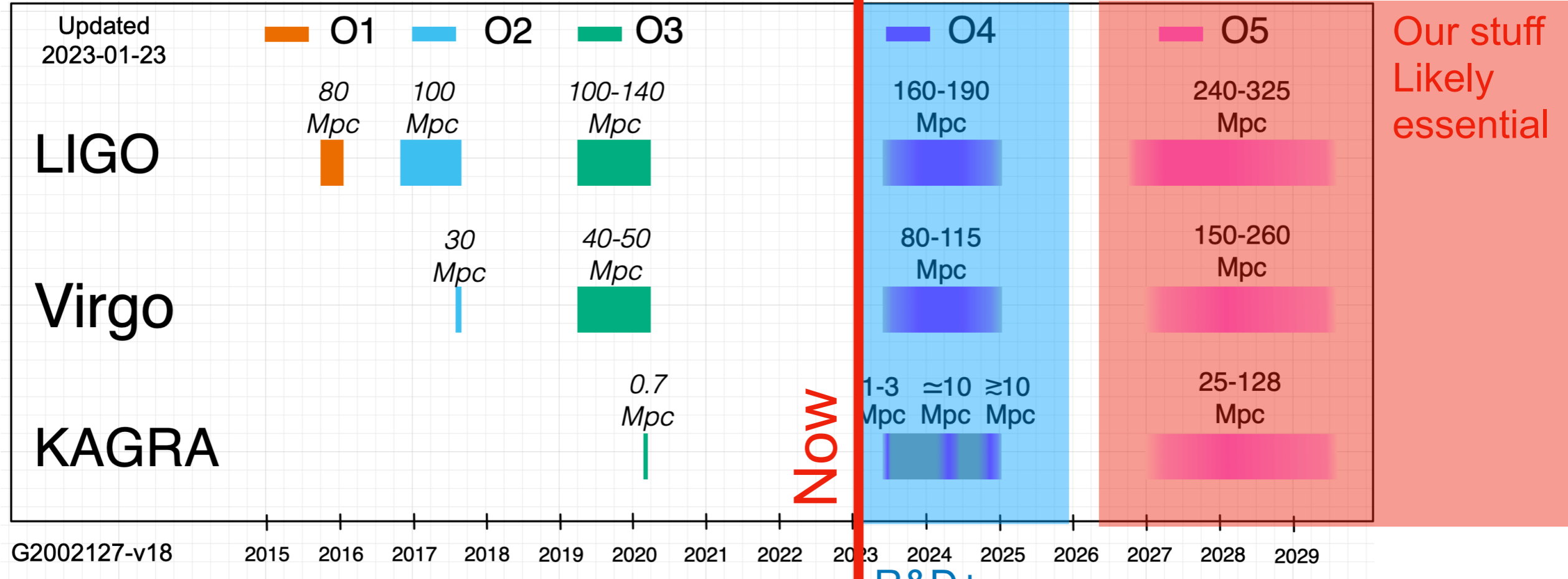
- We would like to highlight commonalities across
 - **Computing demands**
 - ▶ Looking for computing infrastructure for large scale deployment
 - ▶ We can assemble a list of software(+tools)
 - **Software Stack**
 - ▶ With all ML algorithms for a set of core software tools
 - ▶ Need for validate and deploy algorithms
 - **ML**
 - ▶ In some domains similar ML problems exist
 - ▶ Highlighting the similarity is critical

We want to write a white paper highlighting these issues

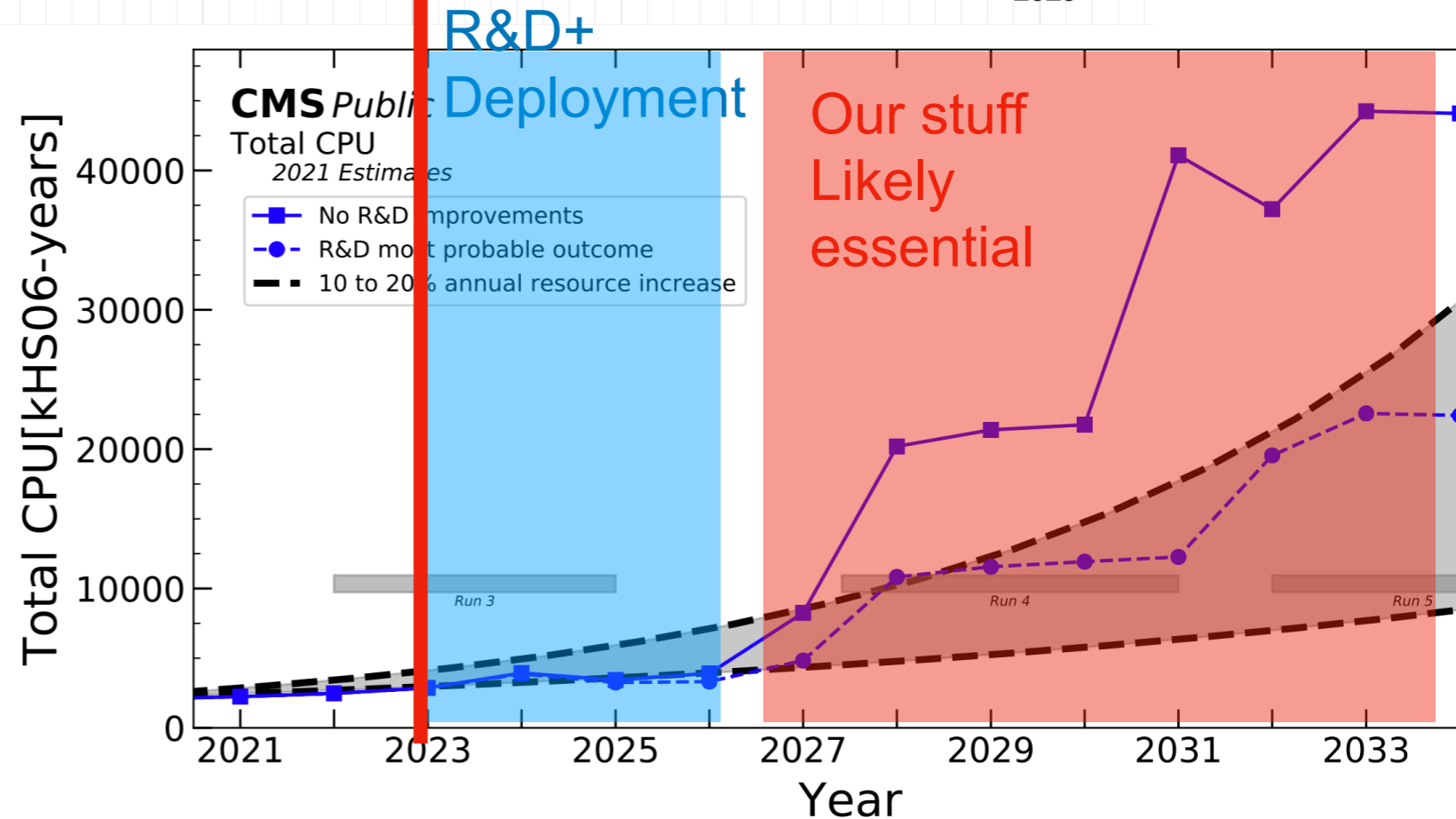
Anatomy of an Algo



Timelines ¹⁶

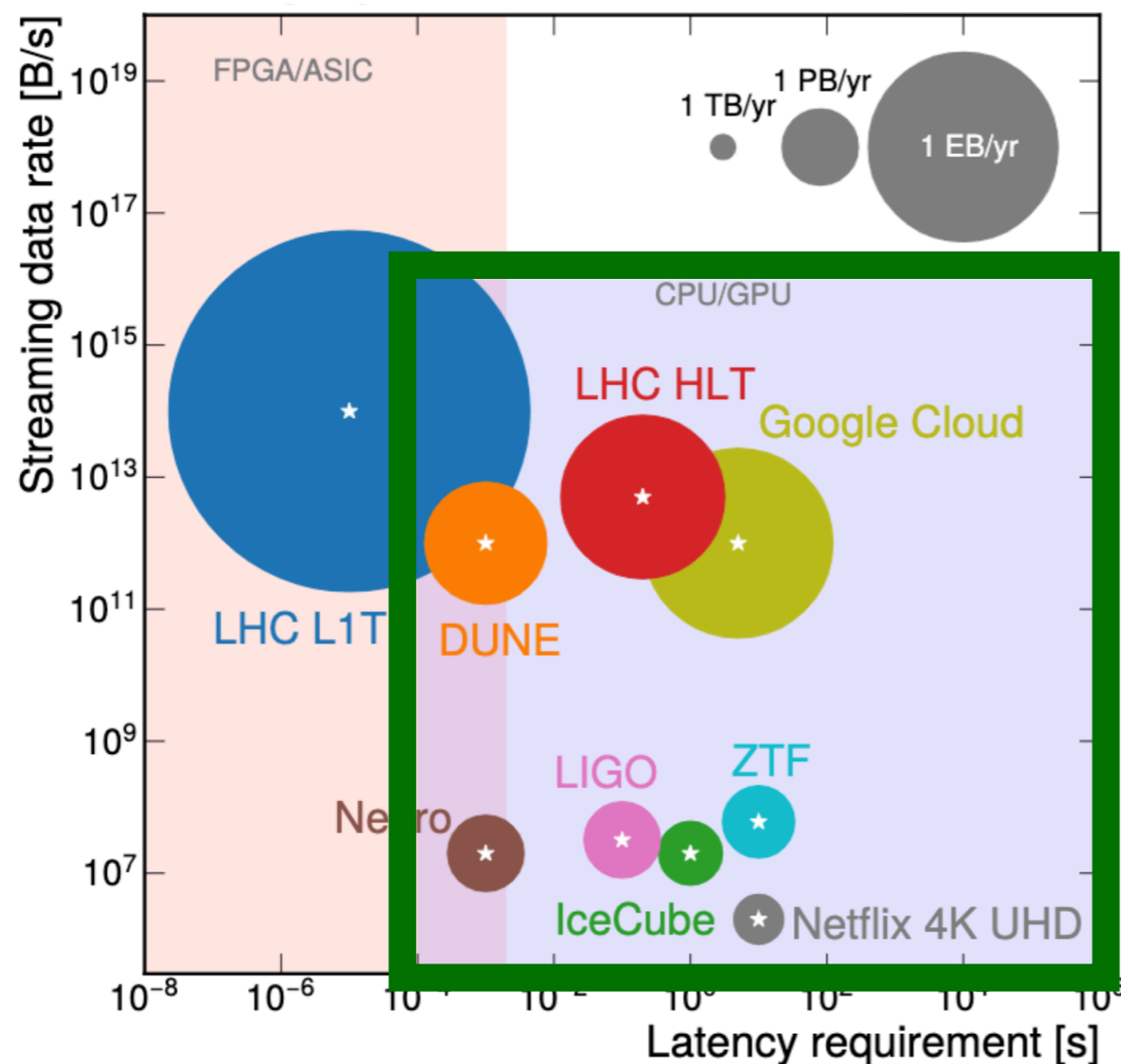


DUNE timeline and various astro timelines (Rubin/LSST)
Should also figure in our overall schedule



What computes are here?

- Within the FastML Community there is a broad range
 - We often try to characterize this range by customization
 - Low Latency and Low Power need more customization

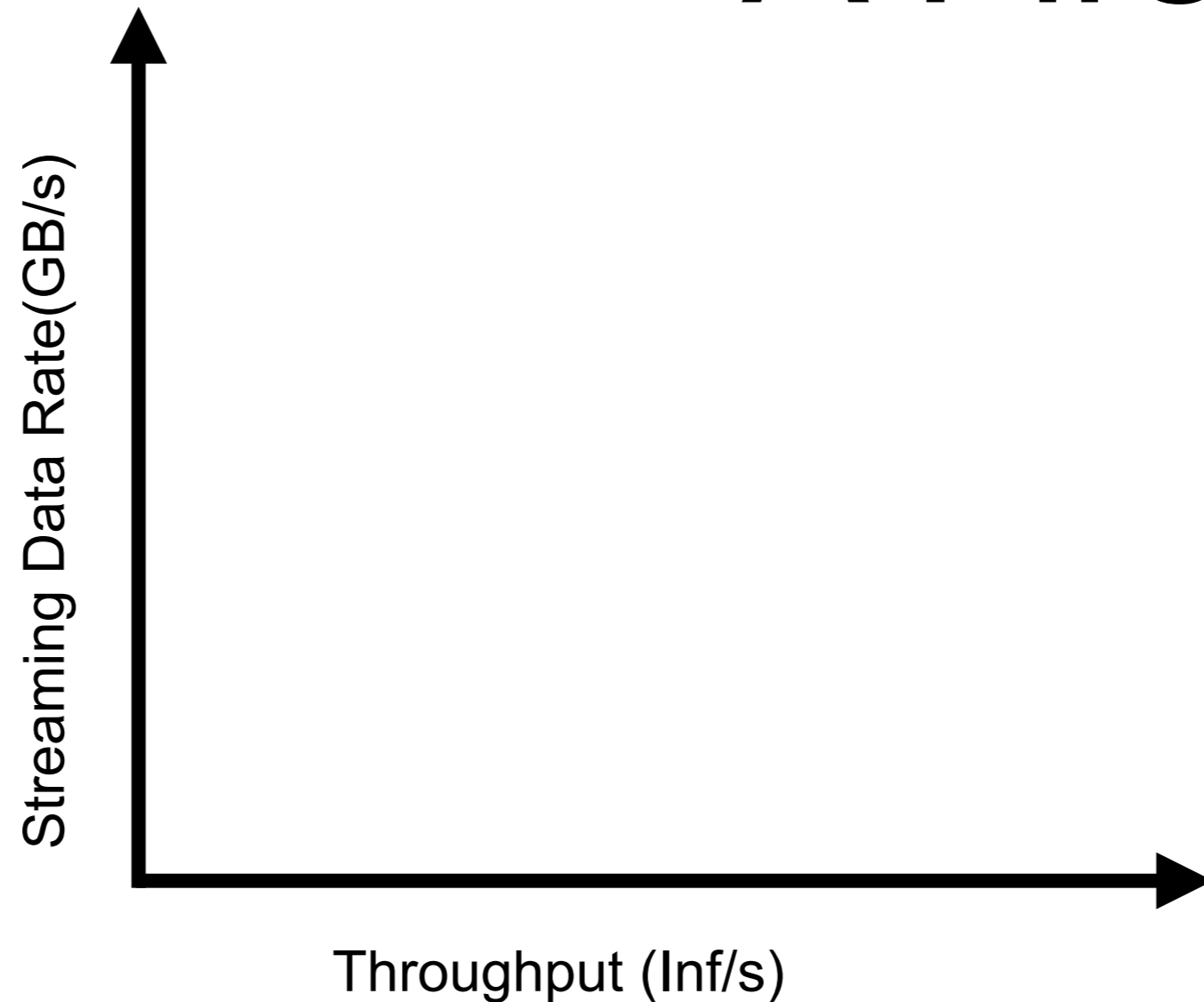


This is our focus here
 We want to understand the high throughput component

Visualizing Computing

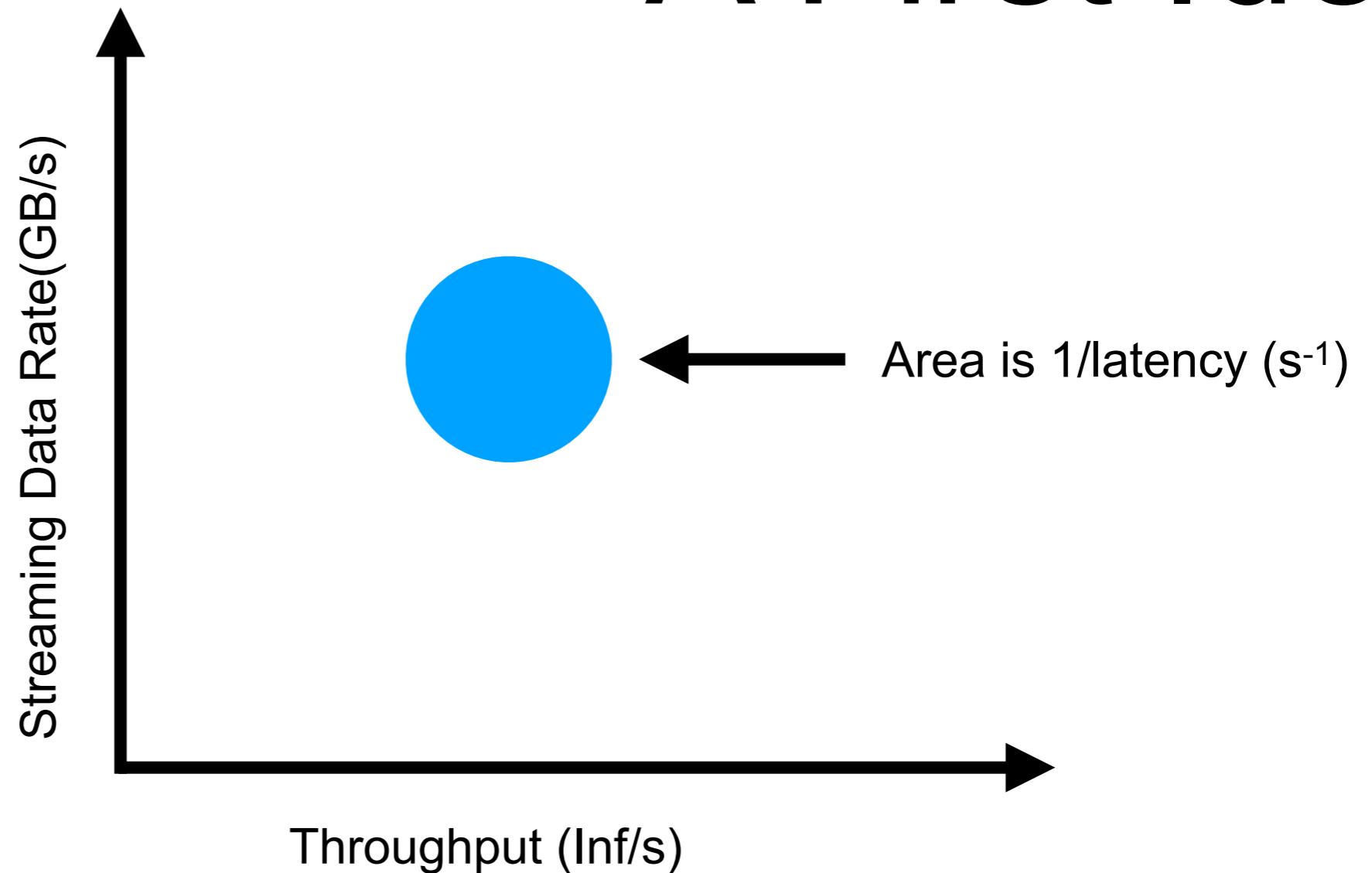
- All of us in the room require at least one thing in common
 - Computers
 - Also, with GPUs/Coprocessors to accelerate things
- As part of this workshop we would like to create a graphic
 - This graph illustrates the computing demands
 - We hope this graphic can be used as a motivator
- The A3D3 graphic has gotten a lot of traction
 - Highlighting the specific challenges for this conference helps
 - Would like to share this with HPCs as a motivator

A First Idea



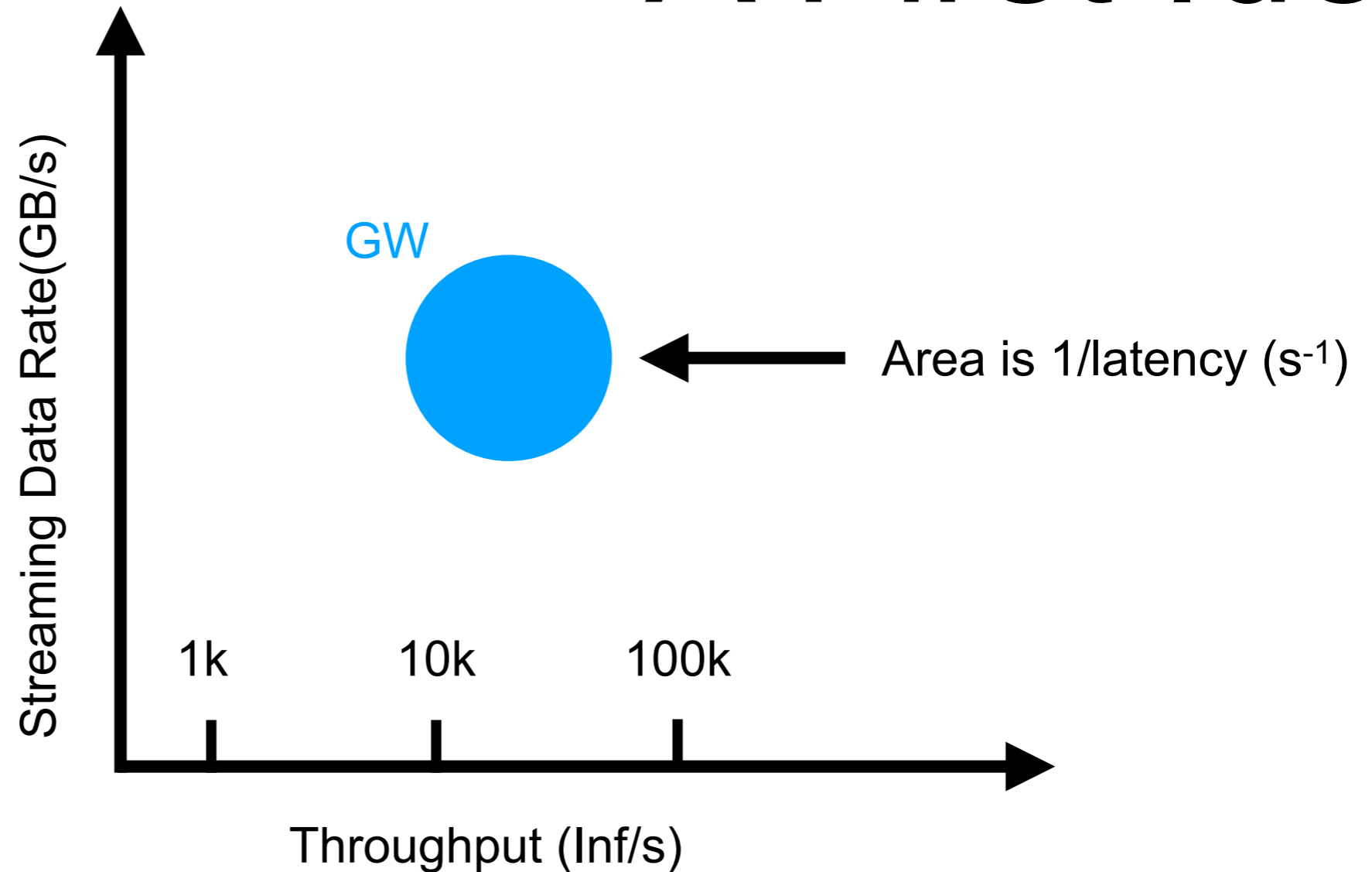
- This work is not at the cutting edge of latency
 - But speed and overall processing time is important
- We are typically concerned about throughput

A First Idea

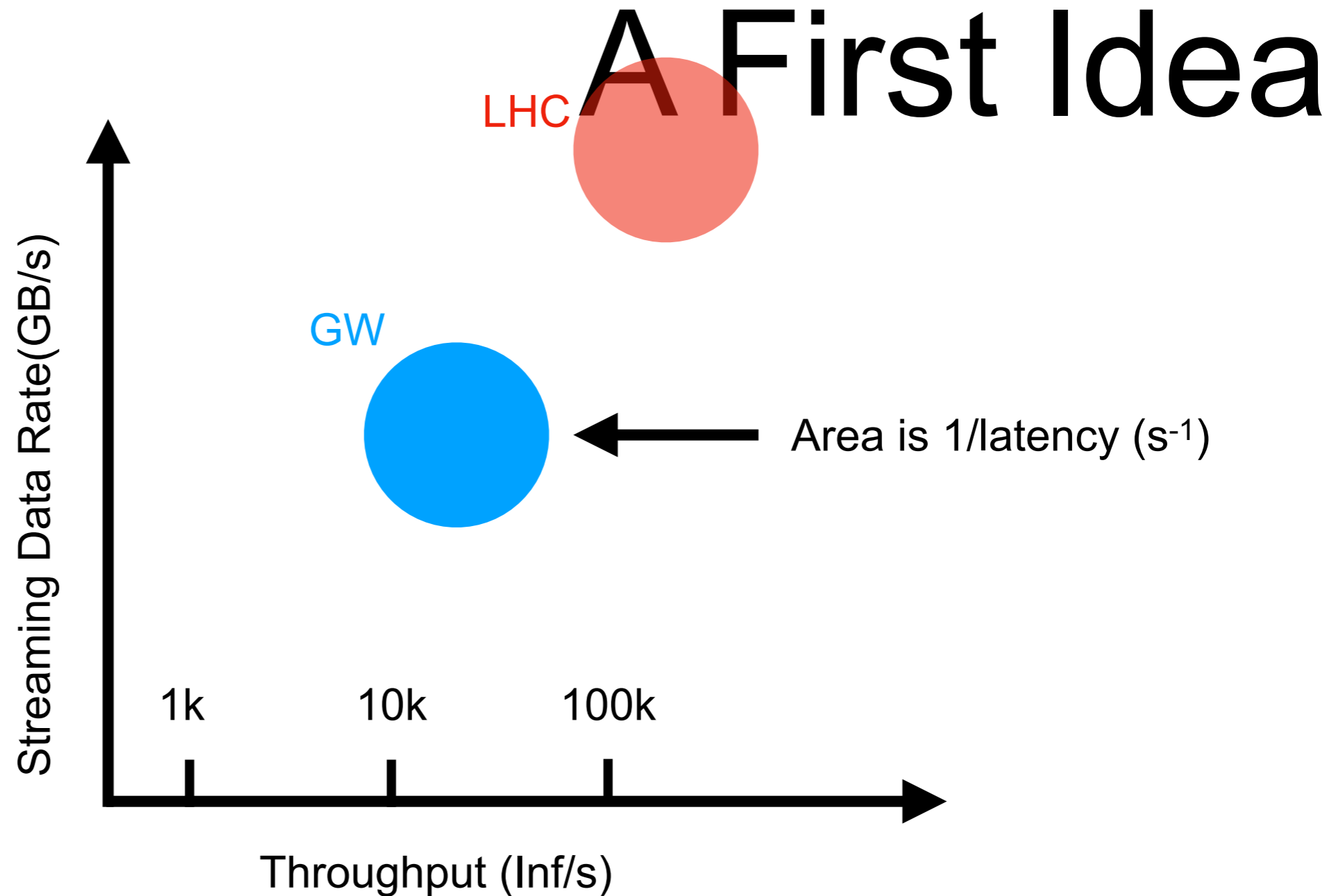


- This work is not at the cutting edge of latency
 - But speed and overall processing time is important
- We are typically concerned about throughput

A First Idea



- LIGO basic specs: 100k channels at 16000 Hz
 - Data Rate is 6.4 GB/s
 - Aiming for a 16000 inferences/s
 - Latency of one second



- LHC High Level Trigger :
 - Data Rate is 1.2 TB/s
 - Aiming for a 100000 events per second
 - Latency of one second

Standardization of Software

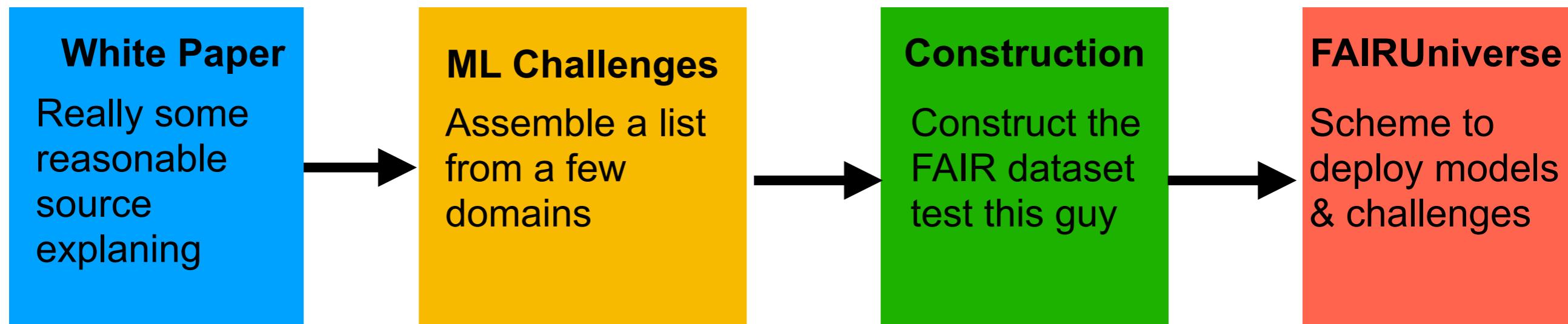
- There is an opportunity to standardize software infrastructure
 - A computing need that is common amongst domains
 - ▶ Astro: Event brokering
 - ▶ LIGO : Low latency alerts
 - ▶ Neutrino/LHC : Data Reconstruction
 - Some domains require different computing to others
- There is also a need for common ML problems and strategies
 - Deployment of effective deep learning problems
 - Anomaly detection
 - Data Augmentation/Generation for effective learning

Getting the right models

- Last part of the document, we would like to highlight ML
 - Classifying the style of problems helps alignment
 - ▶ Supervised learning & its challenges
 - ▶ Data Generation/augmentation with ML
 - ▶ Anomaly detection
 - ▶ Semi-supervision
- For these range of models we would like to highlight a few
 - Ultimately, we would like to **elevate to an ML challenge**

ML Challenges

- Through the HDR community
 - We are working to organize a set of ML Challenges
 - Aiming to align this work with two other communities
 - MLCommons scientific (through ML tiny)
 - FAIRUniverse grant aimed at supporting



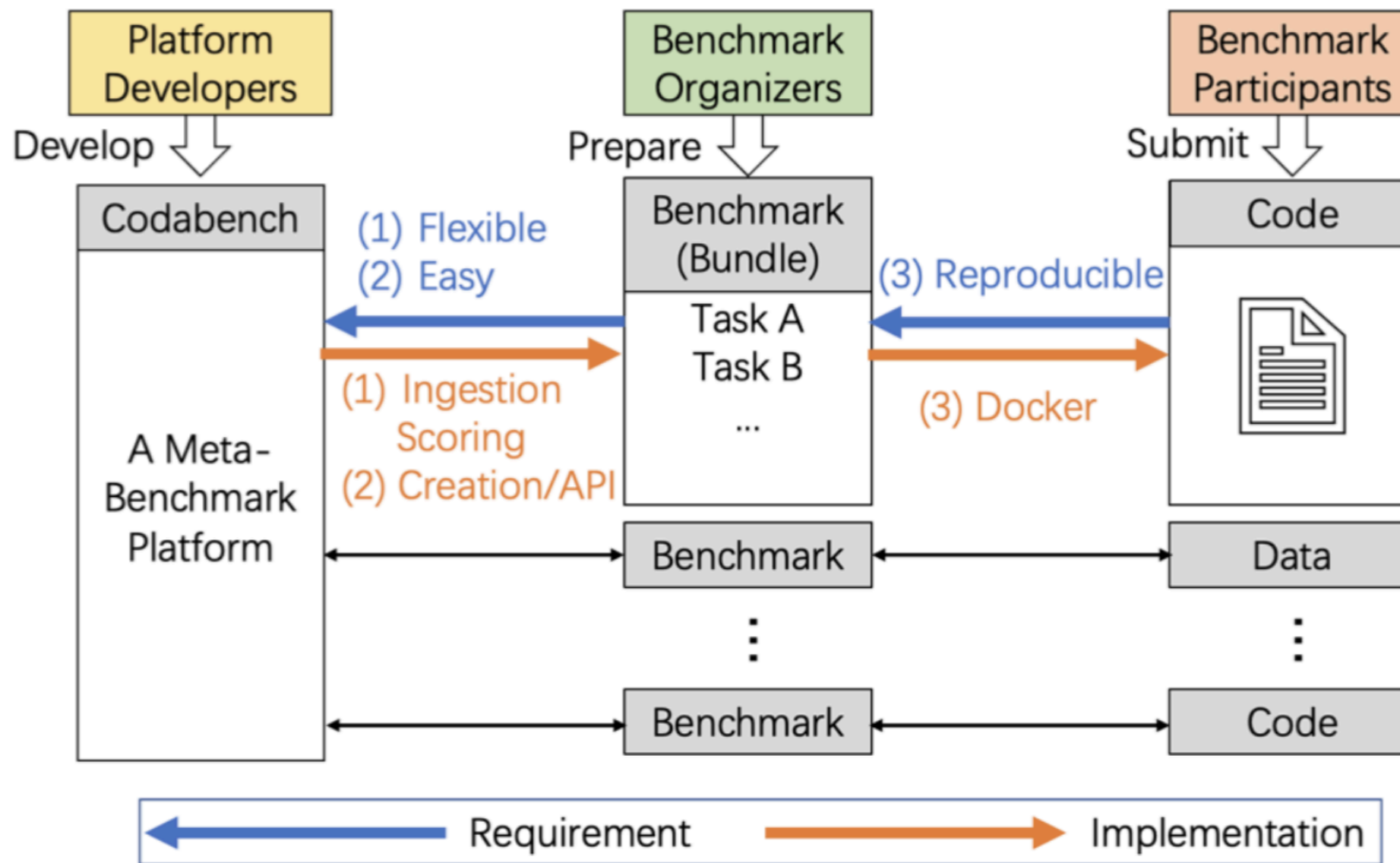
- Annual Bootcamp at UW to award results & have a tutorial

FAIRUniverse has established Infrastructure

Codabench and “Fair Universe” Platform

Based on

<https://www.codabench.org/>

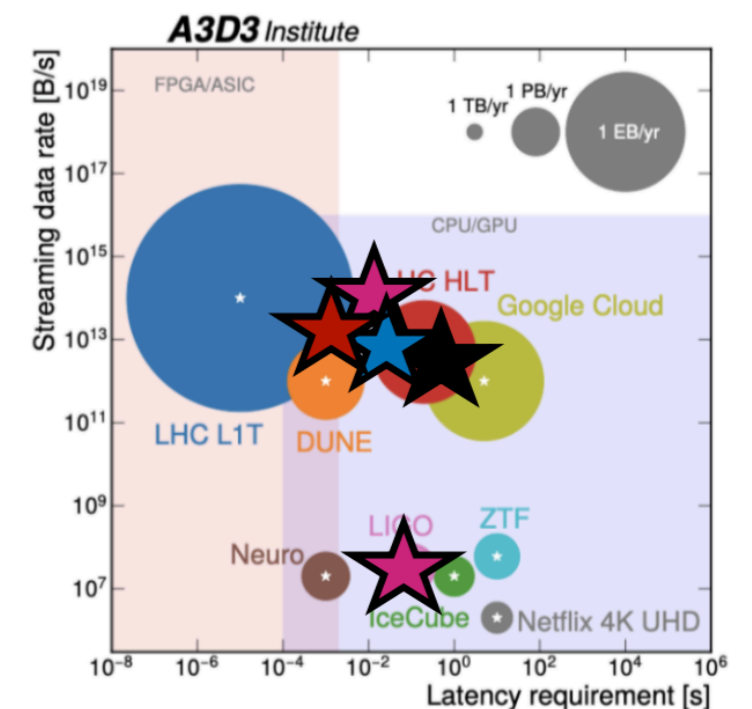
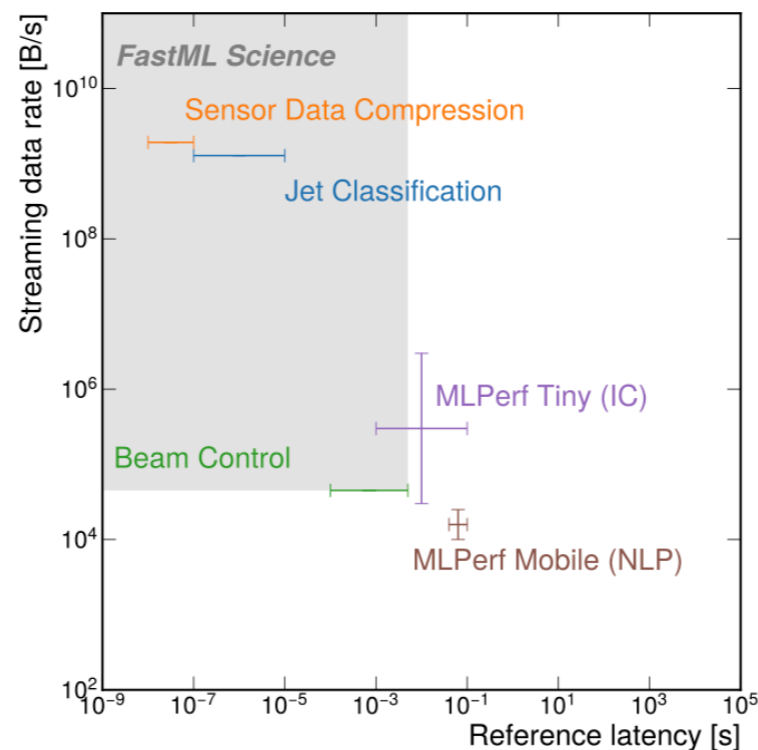


<https://docs.google.com/presentation/d/>

[1hqnlvmMgPgVfm7GzDjb6vJfgafI3PRInd9SX1H0GoFA/edit?usp=sharing](https://docs.google.com/presentation/d/1hqnlvmMgPgVfm7GzDjb6vJfgafI3PRInd9SX1H0GoFA/edit?usp=sharing)

Idea for ML Challenges

- There is one underway Icecube Kaggle Challenge
- Dylan's talk from FastML lists some HEP benchmark motivations
 - LHC tracking as a new benchmark
 - LIGO DeepClean as another benchmark
- More complicated challenges
 - Can we make a data generation challenge, or scheduling

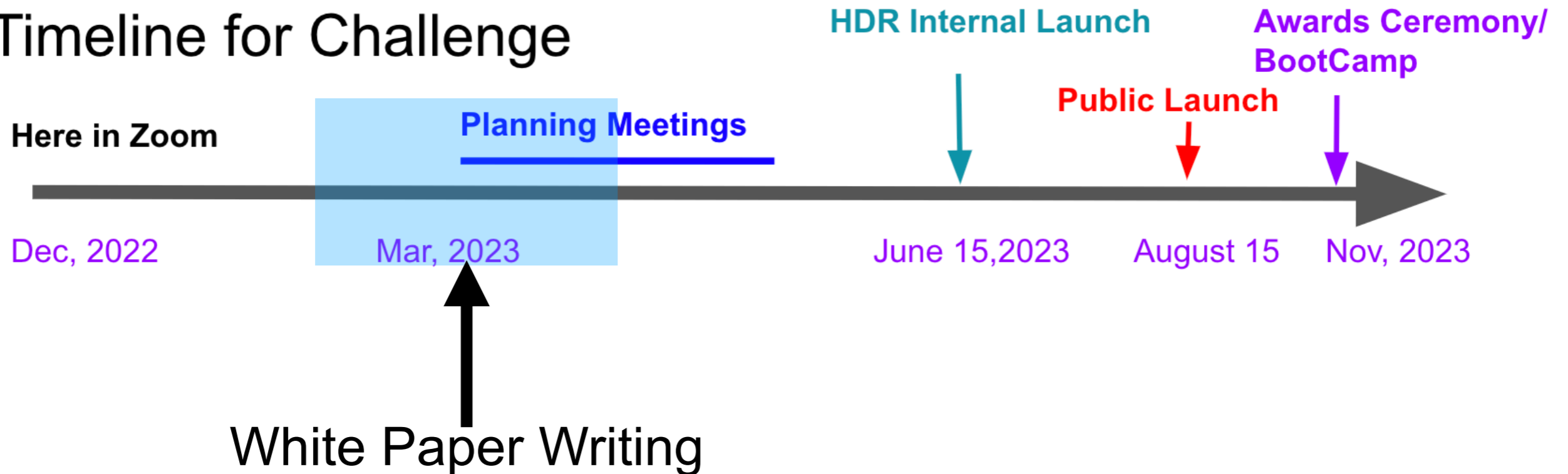


A Point to Highlight

- The best way for us to collaborate across domains
 - Making easy-to-use curated datasets or ML problems
 - We have the people in house to really test these datasets
- This is also a way to tie the different domains together
 - We can use this white paper to start testing out our challenges
 - ▶ Preparation of datasets
 - ▶ Release of models
- Can we get a dataset/model from each scientific domain
 - Also do we have the right benchmarks to do this?

Roadmaps

Timeline for Challenge



- This white paper can help us to assemble some models
 - We would like to assemble a few pages on these
 - ▶ Illustration of the problem (paragraph)
 - ▶ Illustration of the dataset preparation
 - ▶ Impact on the field

Conclusions

- Welcome! Enjoy your time here in Cambridge
 - We would like to write a white paper
 - We have some discussion time at the end of the conference
- Outline for the White paper (**Lets keep it short!**)
 - **Discussion of computing tools and software**
 - ▶ Path to aligning these across domains
 - **List of critical models in the field**
 - ▶ What makes these models
 - **One plot to rule them all and bind these sections**
- A roadmap for future computing can help us move this forward

White Paper

- <https://www.overleaf.com/3629142192jgpsrnqvzccd>

Backup

Possible Idea

Neural Benchmarks
Public Challenge

MMA Benchmarks
Public Challenge

HEP Benchmarks
Public Challenge

~~FAIR4HEP~~
A3D3

A3D3 Internal Challenge

To go from internal to public
We should seek Corporate
Sponsors

Have enough people in our
institute to build a pipeline

Public Challenge

Standardization of Software

- Alert brokering and Image processing
- ML Problems
- Anomaly detection
- Data Augmentation Generation