



# AI at Fermilab

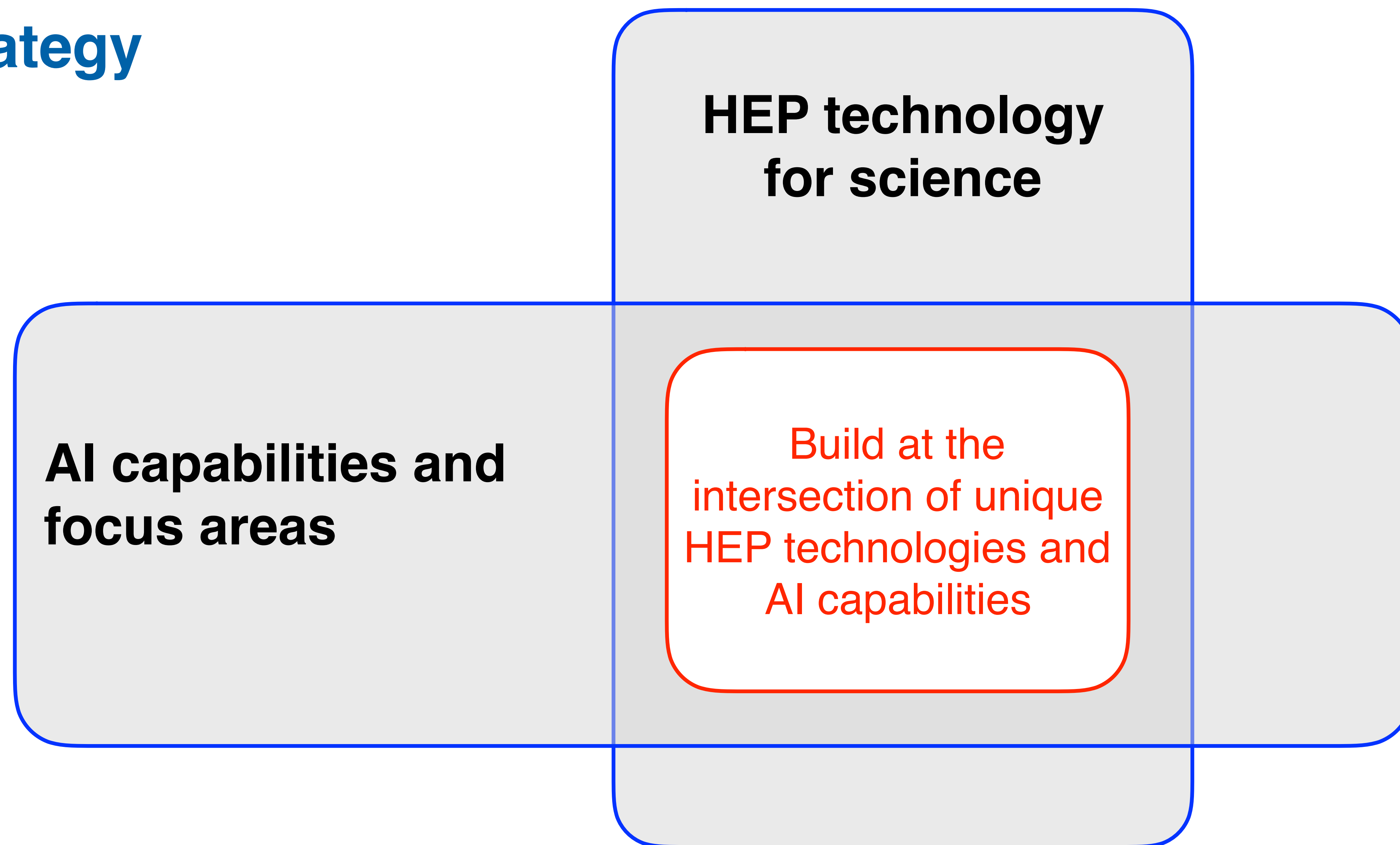
**Nhan Tran for the AI Project**

January 14, 2020

Fermilab PAC Meeting



# Strategy



# Fermilab & HEP in the AI ecosystem

- **Exciting applications in fundamental particle physics**

- a few examples:

- Neutrino applications in NOvA
- Higgs @ CMS
- CMB @ SPT

Driven by the experimental collaborations

Fermilab scientists involved in research are driving the highlighted applications

- Intersection of **HEP & AI technology provides opportunities for innovation!**

- Processing and simulating massive datasets
- Requirements for real science (uncertainty quantification, models with embedded physics,...)
- Large, integrated operations and data management
- Real-time, edge/sensor systems

Fermilab developing AI technologies for HEP and beyond

# Outline

**Fermilab & HEP in the AI Ecosystem**  
scientific applications

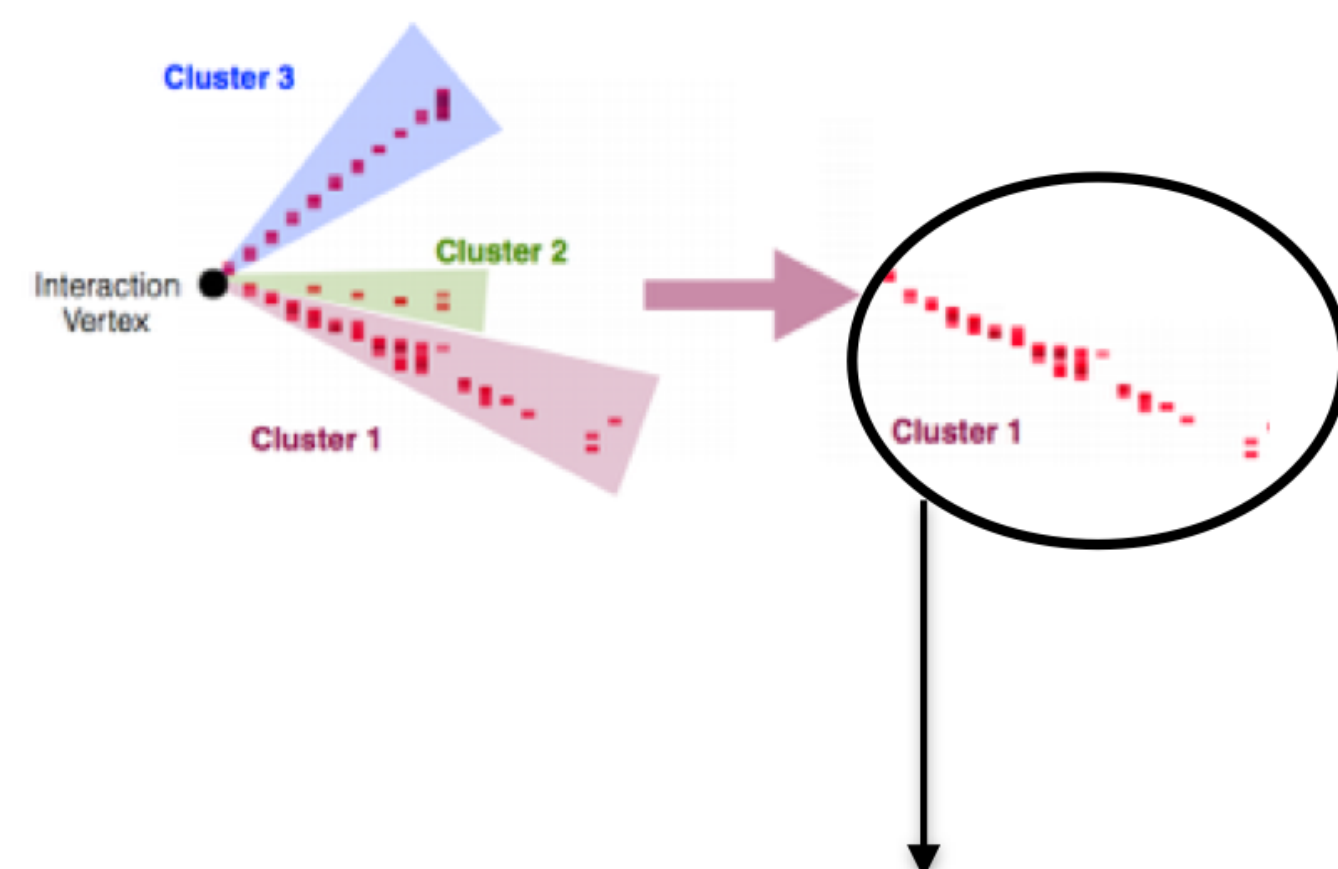
**AI capabilities and focus areas**  
capabilities developed for HEP

**Who are we?**  
Building a community

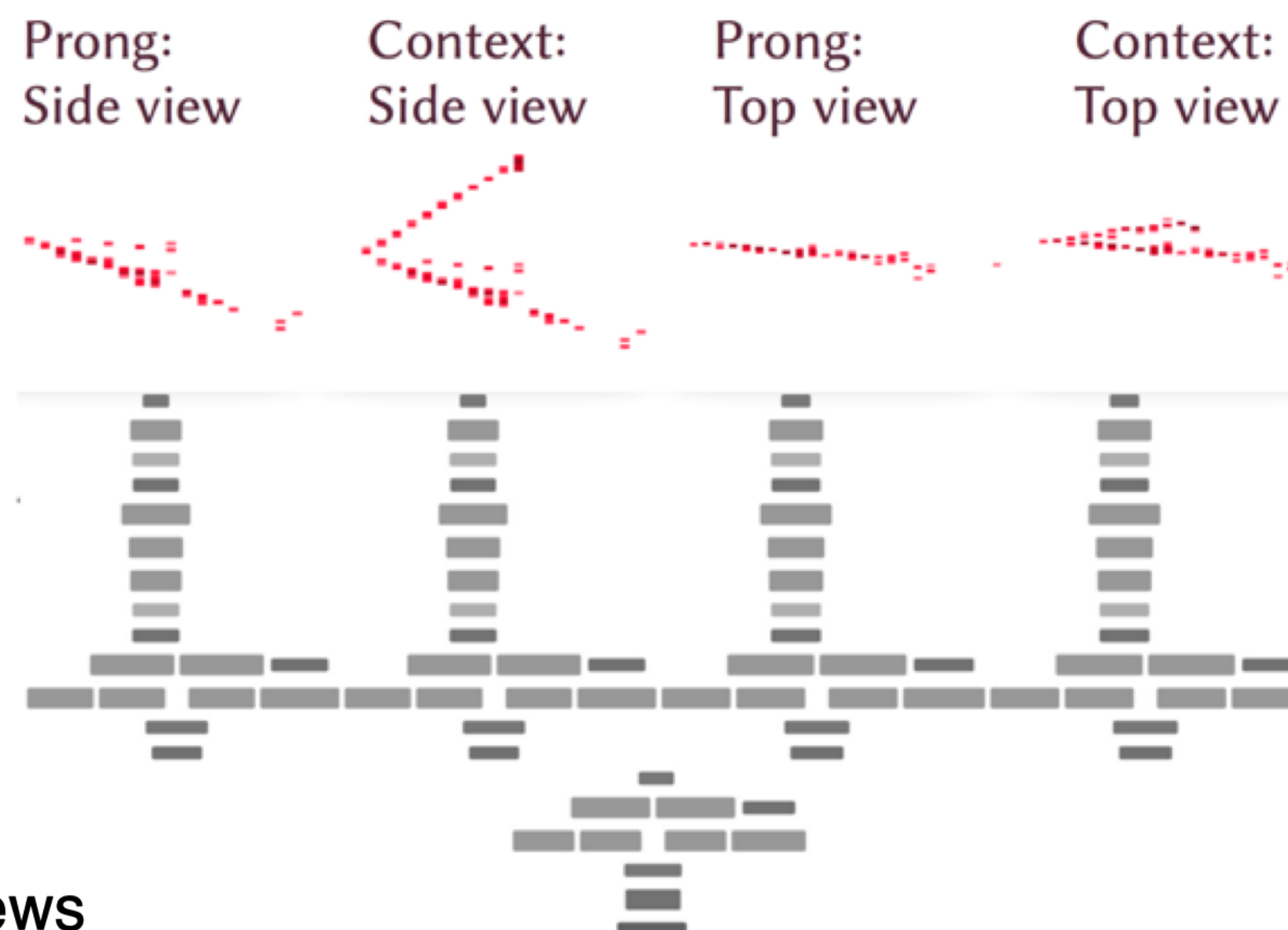


# Example: NOvA Reconstruction

Single particles are separated using geometric reconstruction methods.



Classify particles using full event topology from both views as well as reconstructed cluster information (**4 views**)

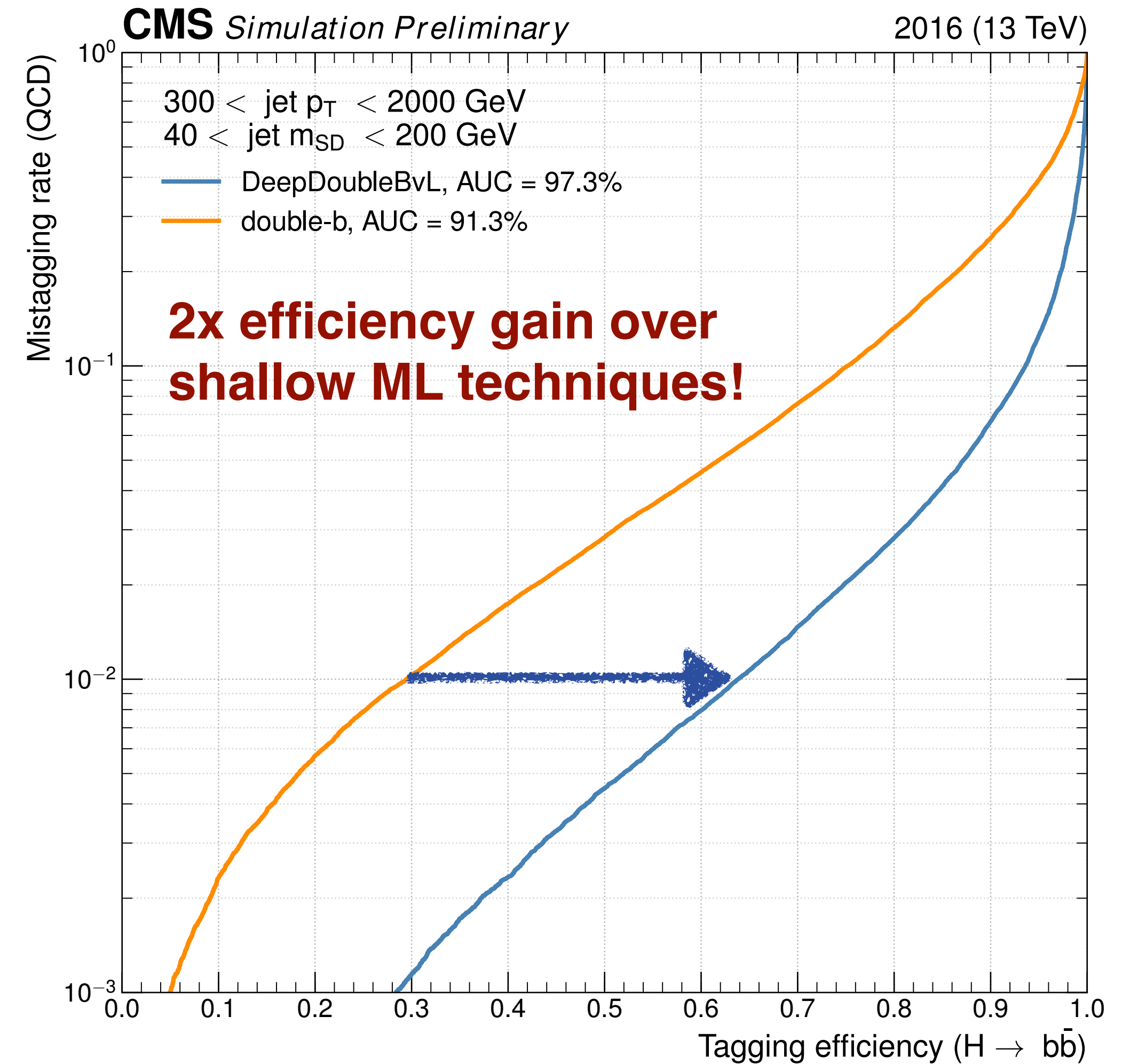
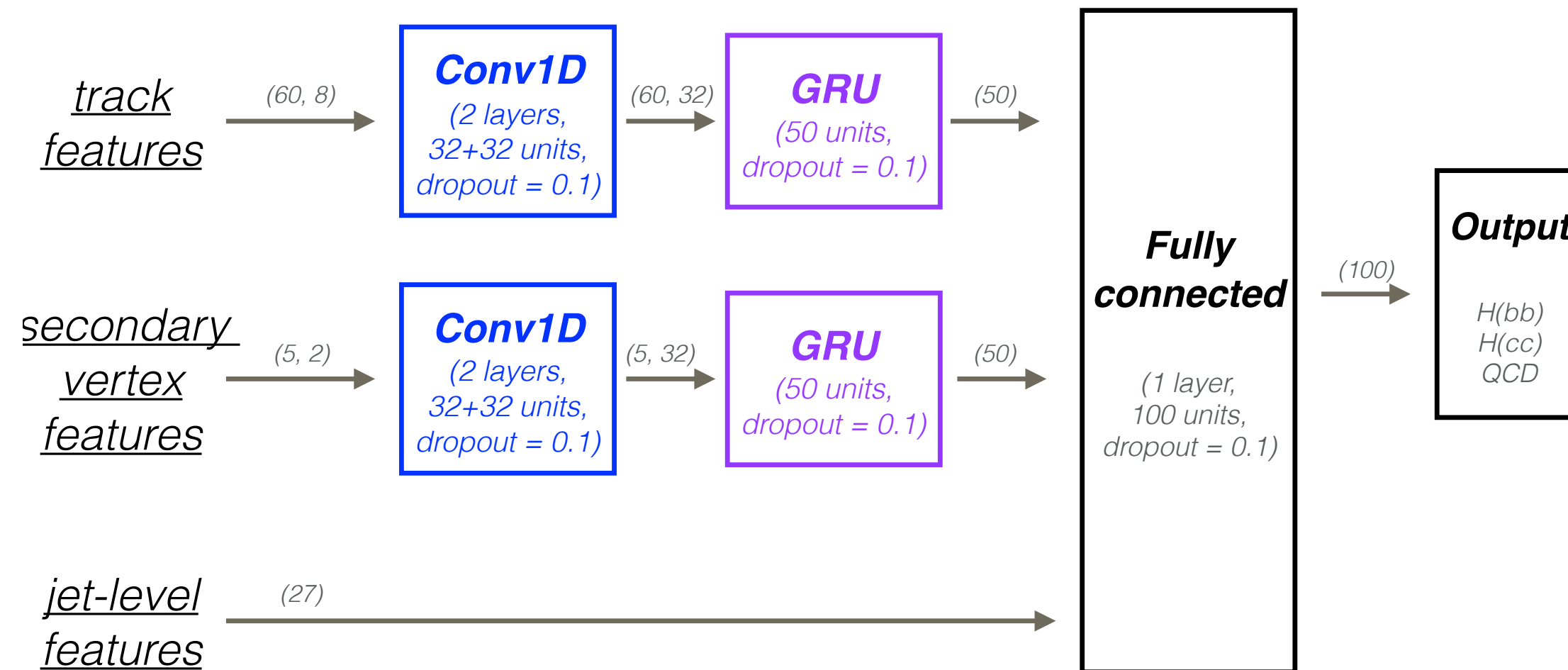


A. Himmel, E. Niner, F. Pshihas et al.  
<https://arxiv.org/abs/1604.01444>

1st deployed in oscillation analysis  
<https://arxiv.org/abs/1703.03328>

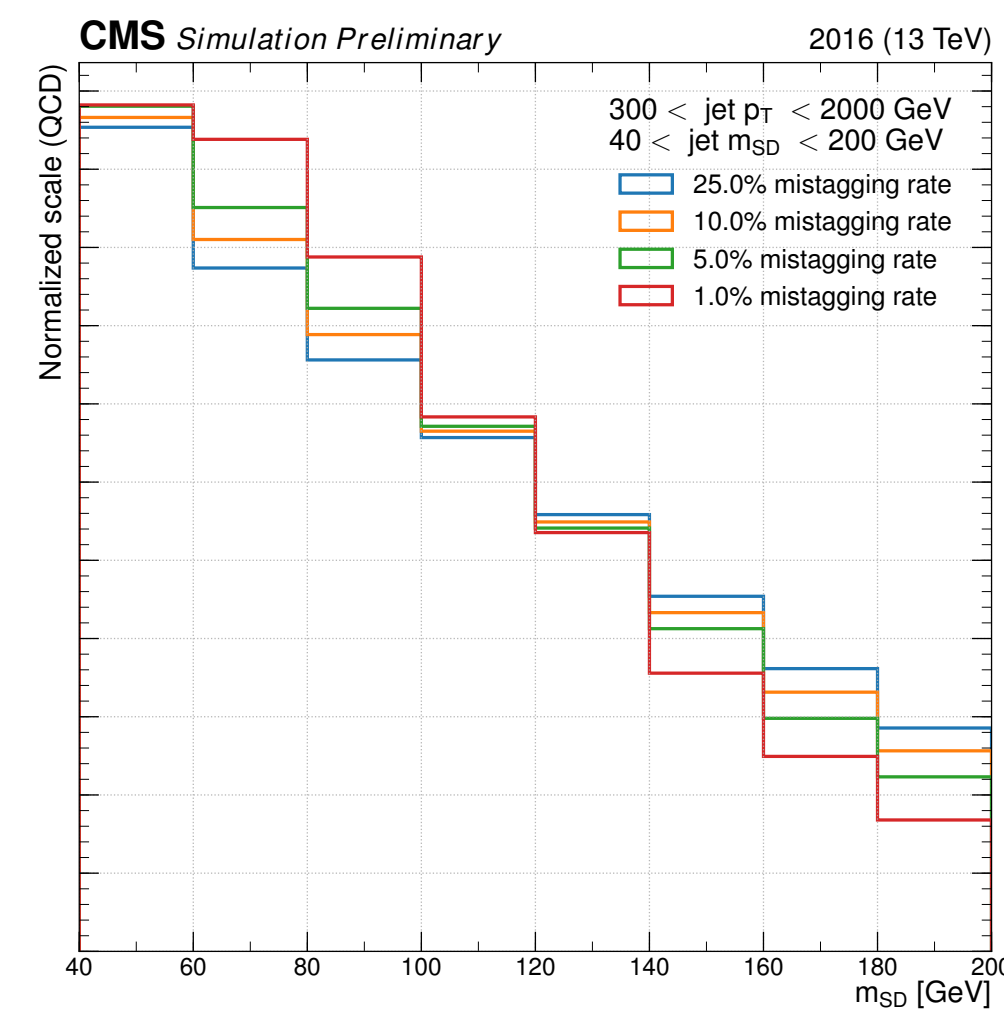
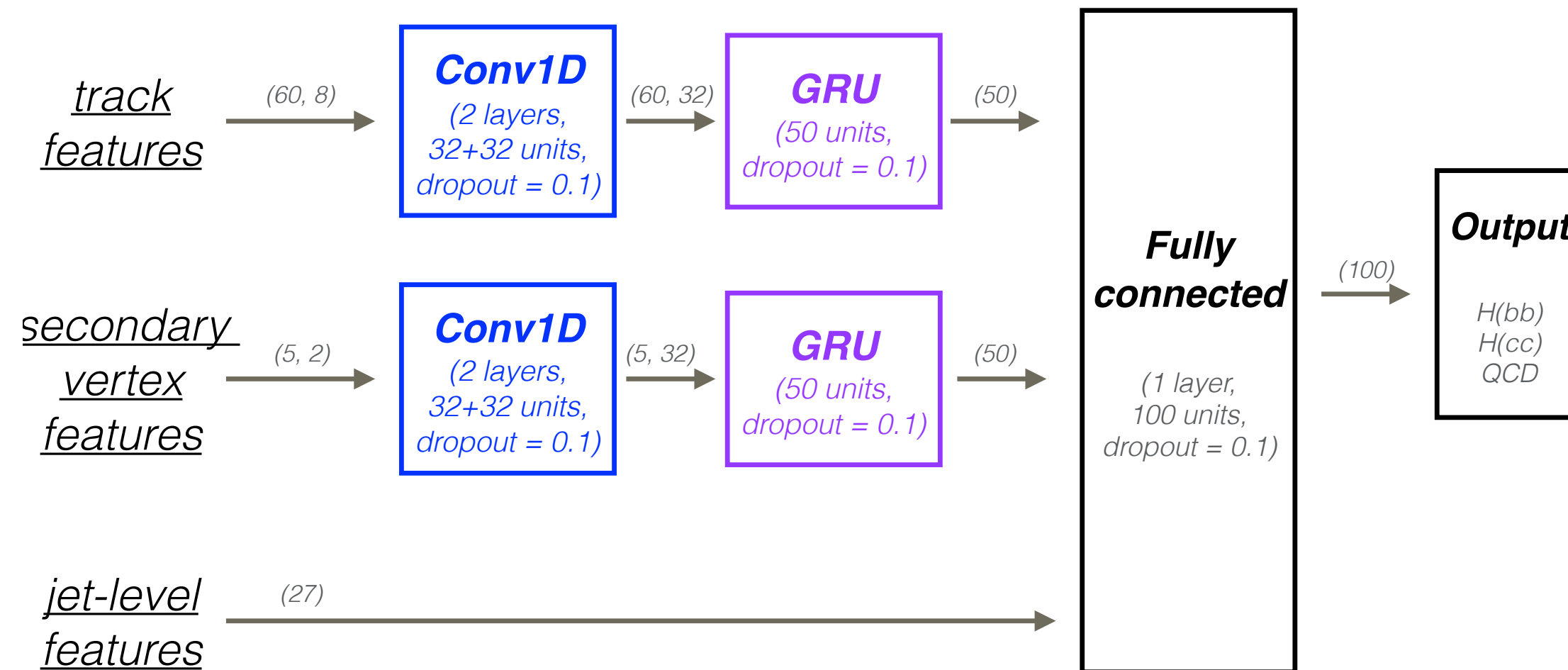
Performance improvement equivalent to **4.2 kilotons of additional detector mass** with traditional particle identification algorithms.

# Example: Identifying the Higgs

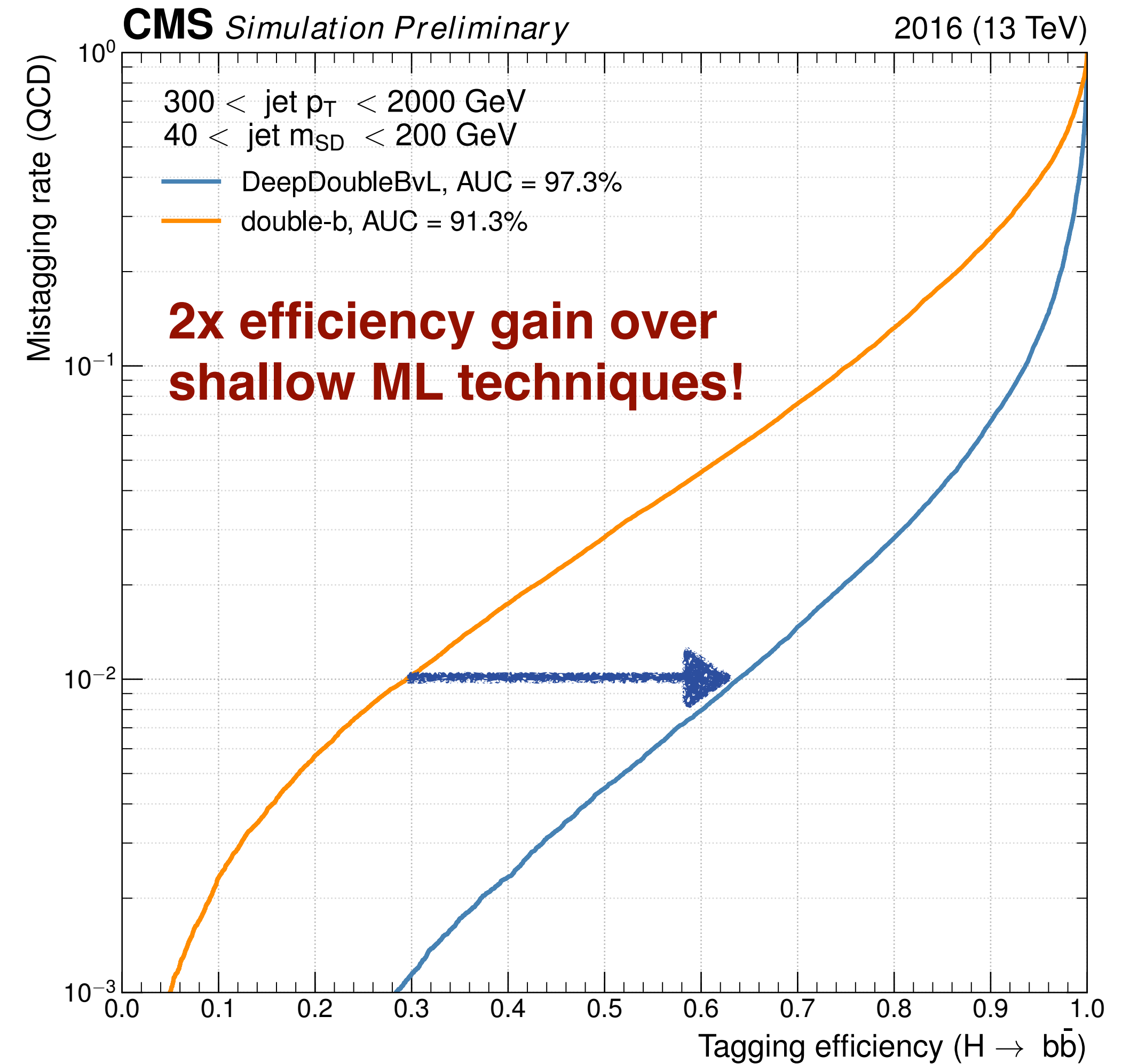




# Example: Identifying the Higgs



**decorrelation:**  
teach the network how to *not* learn certain physical features;  
important for controlling systematic uncertainties

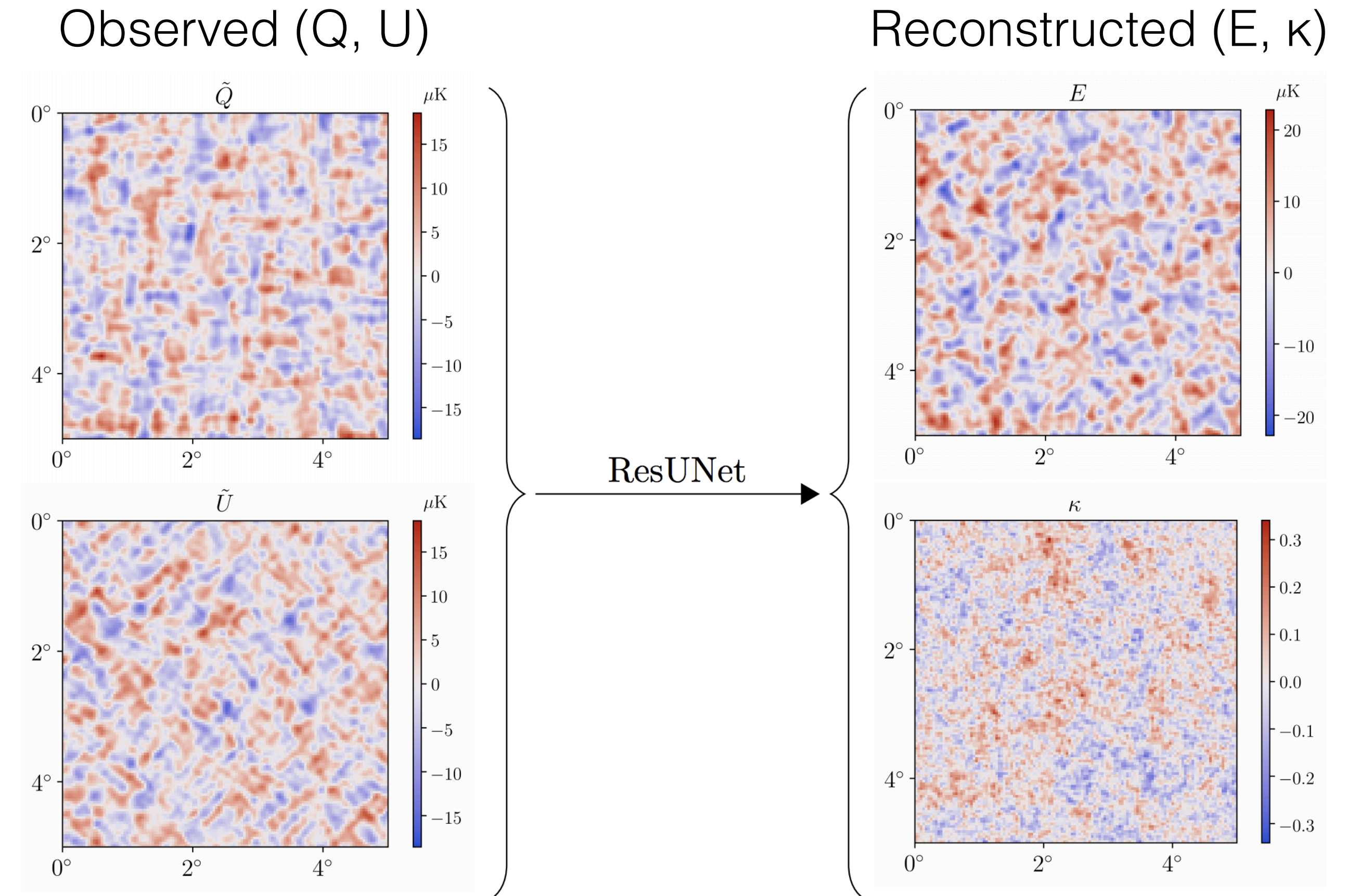




# Read between the layers: decompose microwave maps



- South Pole Telescope (SPT): Polarized cosmic microwave background maps
- Earliest gravitational wave signatures that have very low signal
- Applicable for CMB-S4 next generation experiments



- Noise and other foregrounds obfuscate primordial GW signatures
- Pioneered use of **Residual UNets** to separate lensing signals ( $\kappa$ ) from CMB polarization map ( $E$ )



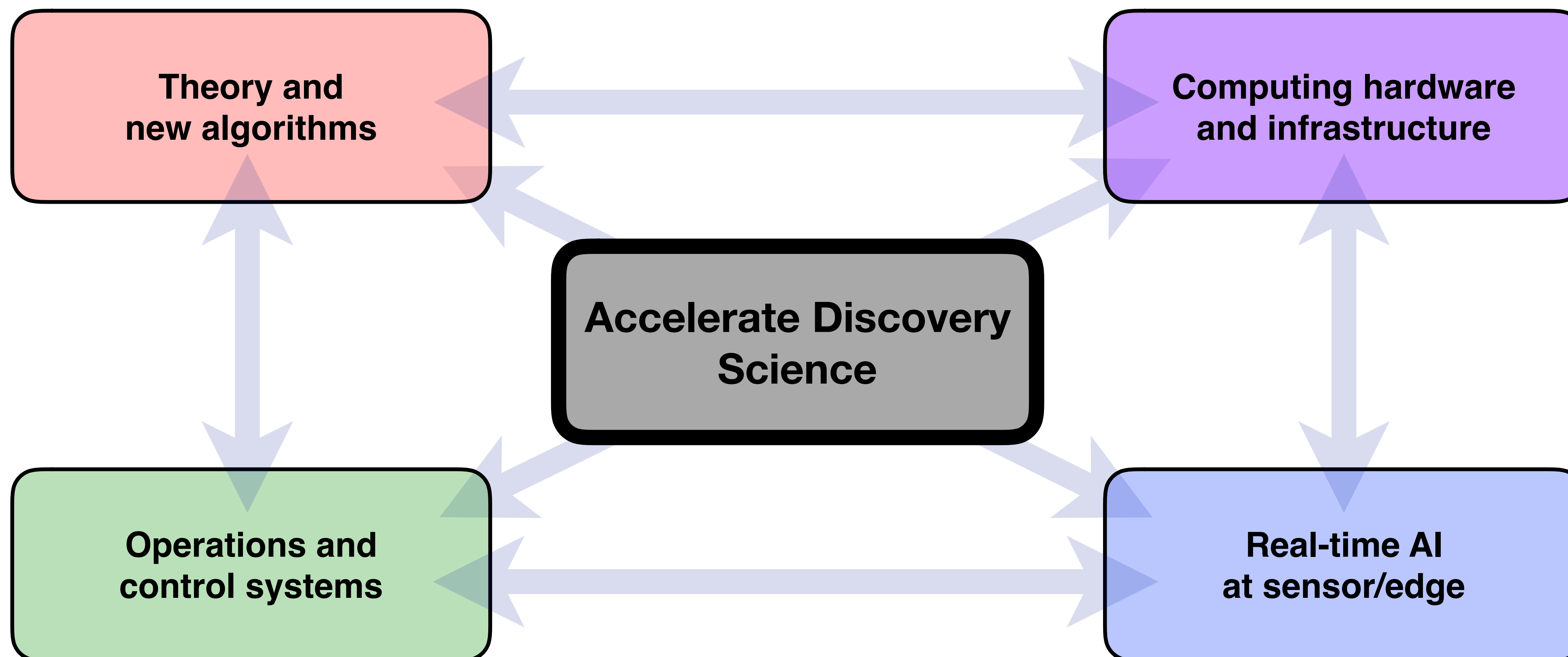
# Outline

**Fermilab & HEP in the AI Ecosystem**  
scientific applications

**AI capabilities and focus areas**  
capabilities developed for HEP

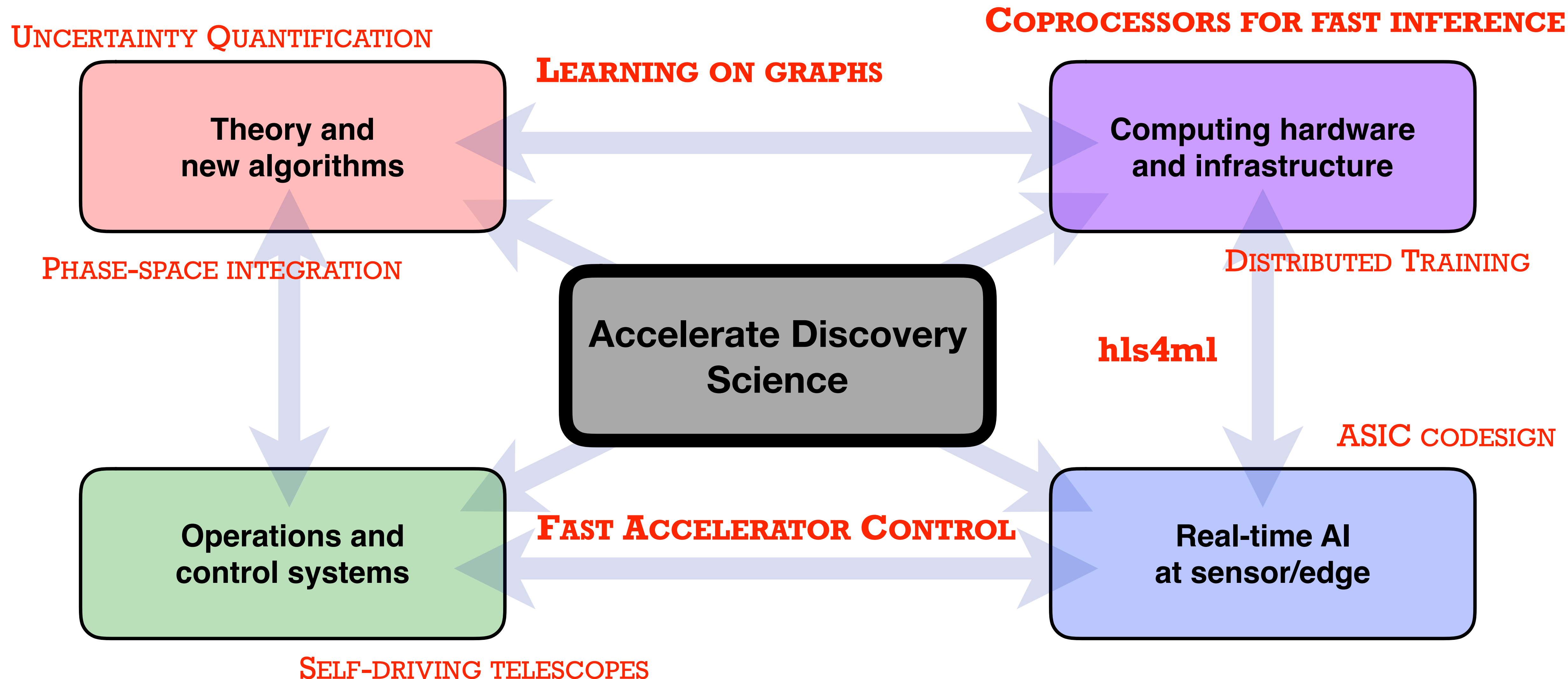
**Who are we?**  
Building a community

# Fermilab AI Capabilities





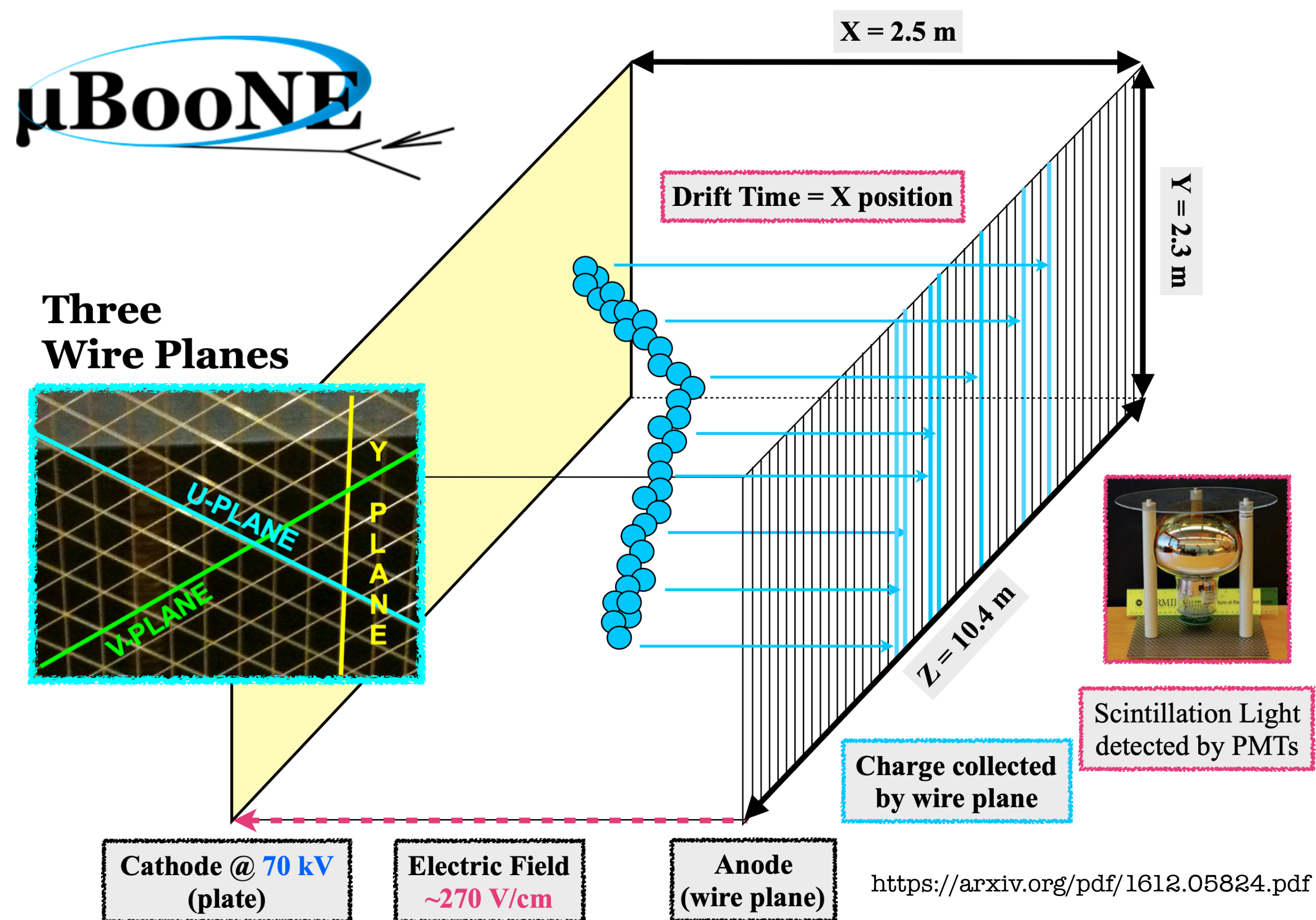
# Fermilab AI Capabilities



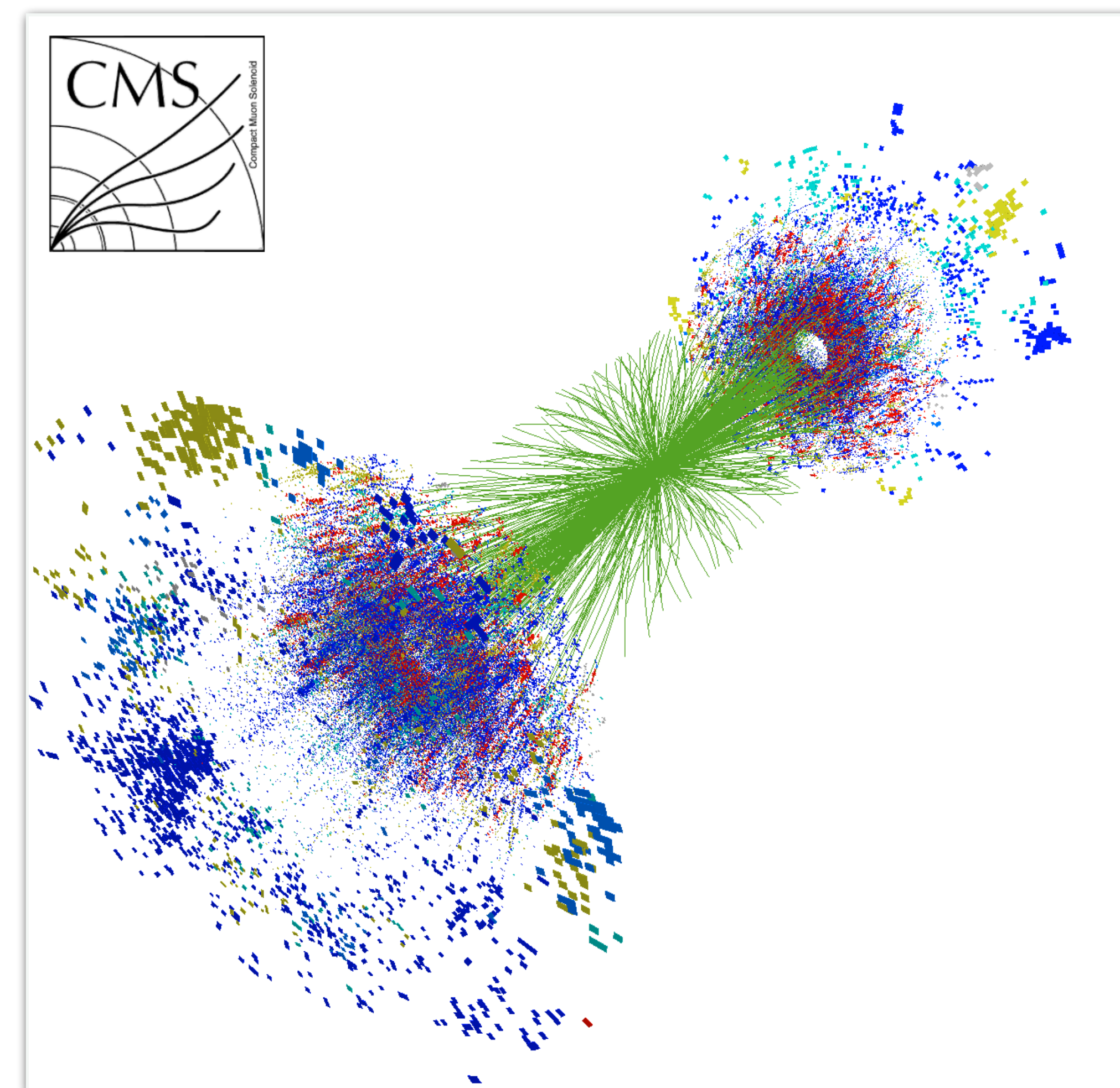
# Beyond images

Theory and  
new algorithms

Sparse, multi-modal, high-dimensional



$$\vec{v} = \{x, y, z, E, t, id, \delta, \dots\}$$

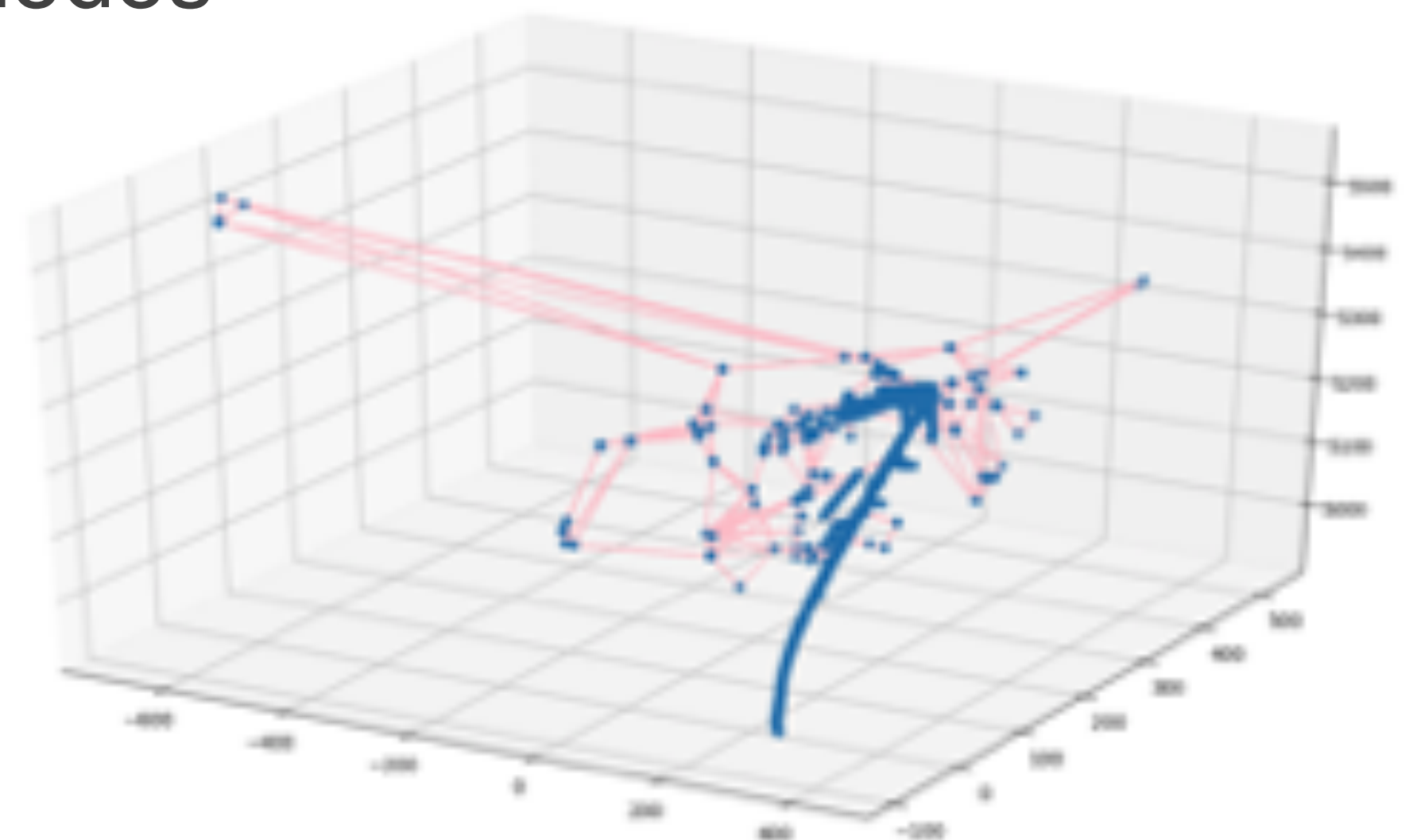




# Beyond images

Theory and  
new algorithms

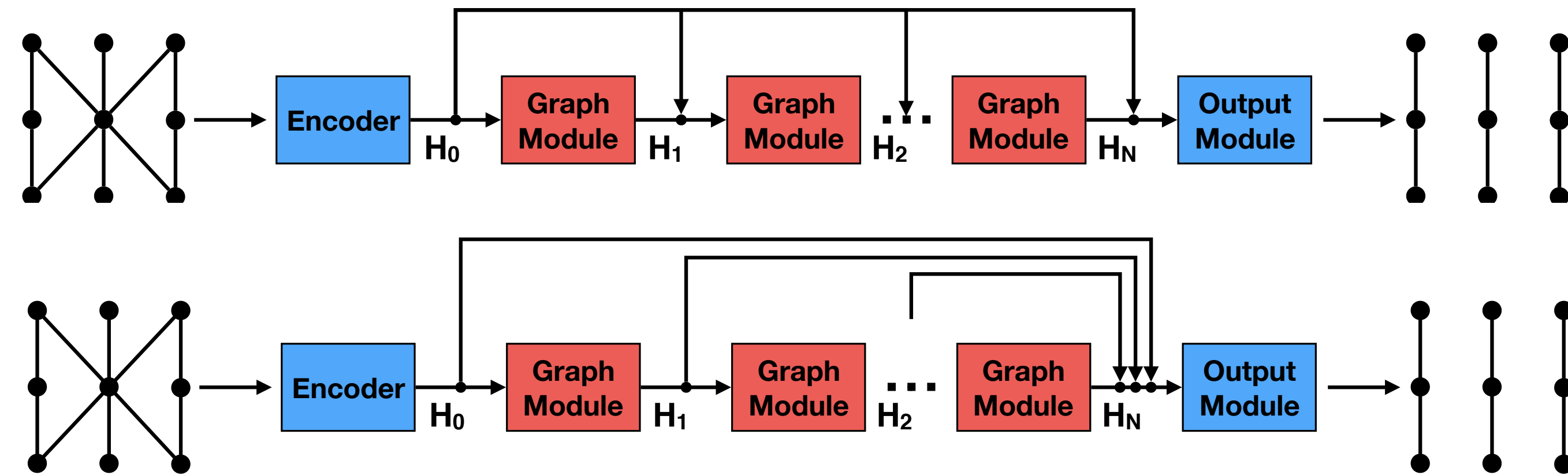
- Multiple activities into **learning new representations of detector data** for different physics applications
- Explore neural network architectures based on **point clouds and graphs**; n-dimensional inputs in non-Euclidean space
- Promising first results for multiple applications
  - Learn the strength of connections (edges) between nodes
  - Charged particle tracking [1]
  - Calorimetry for irregular geometries [2]



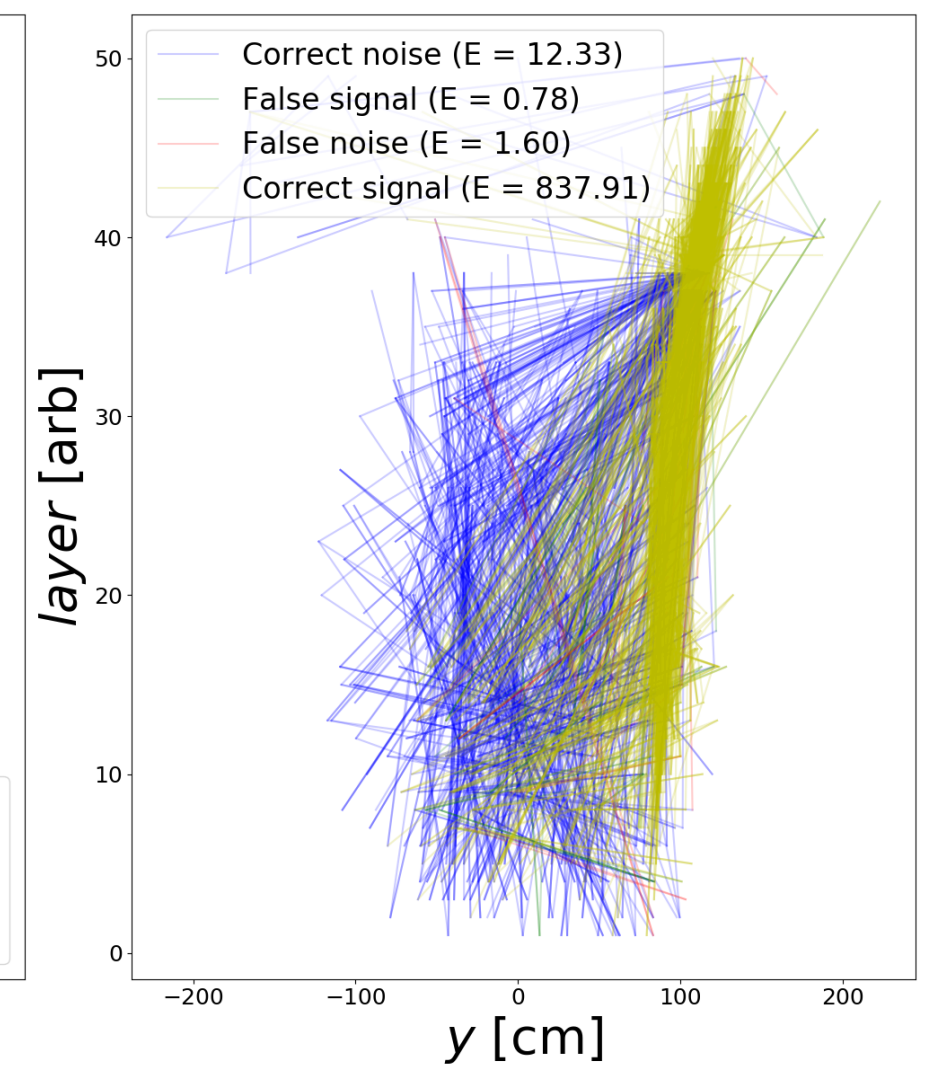
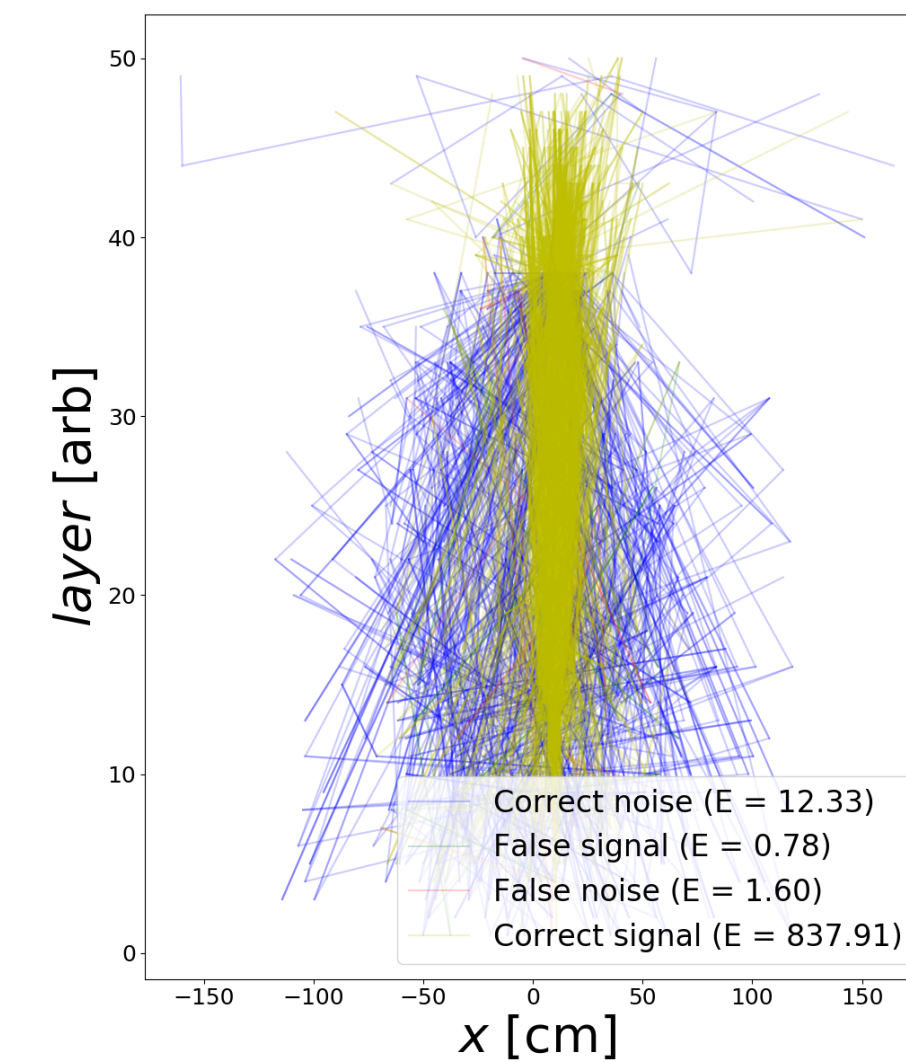
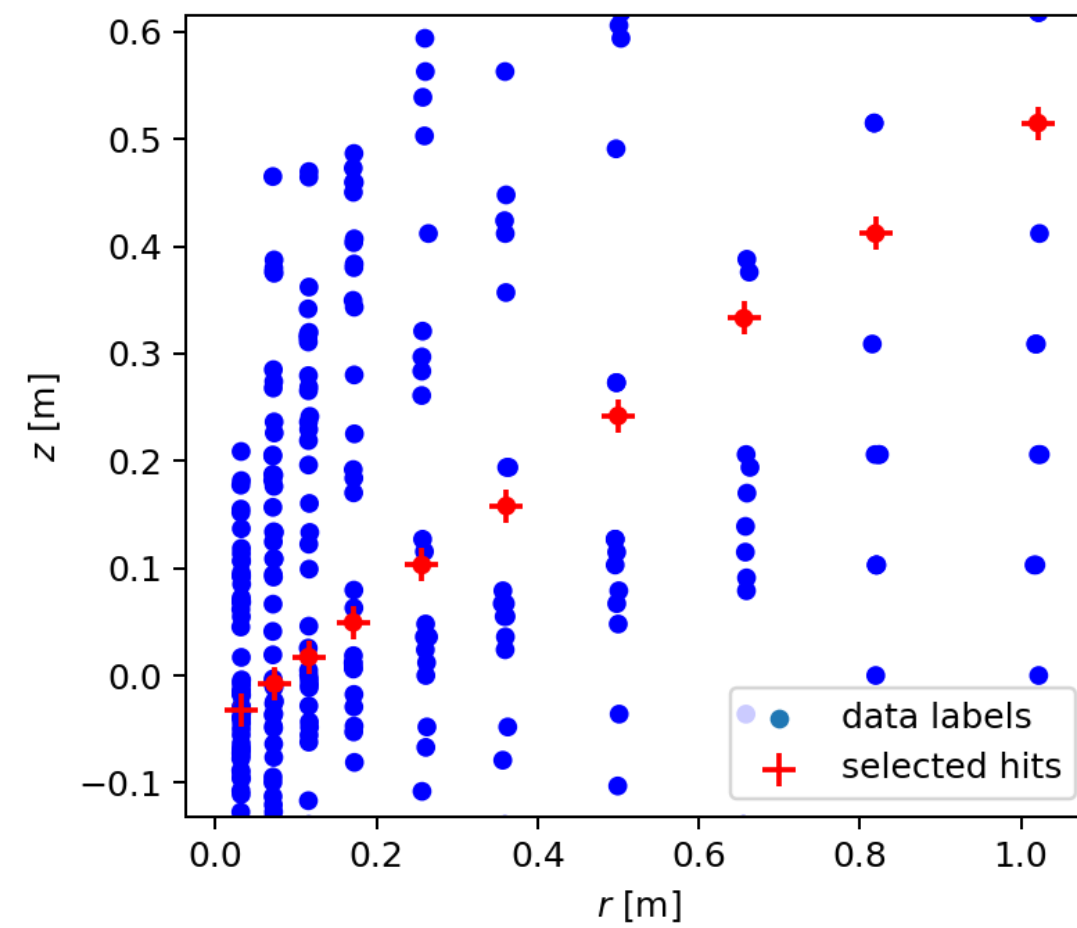
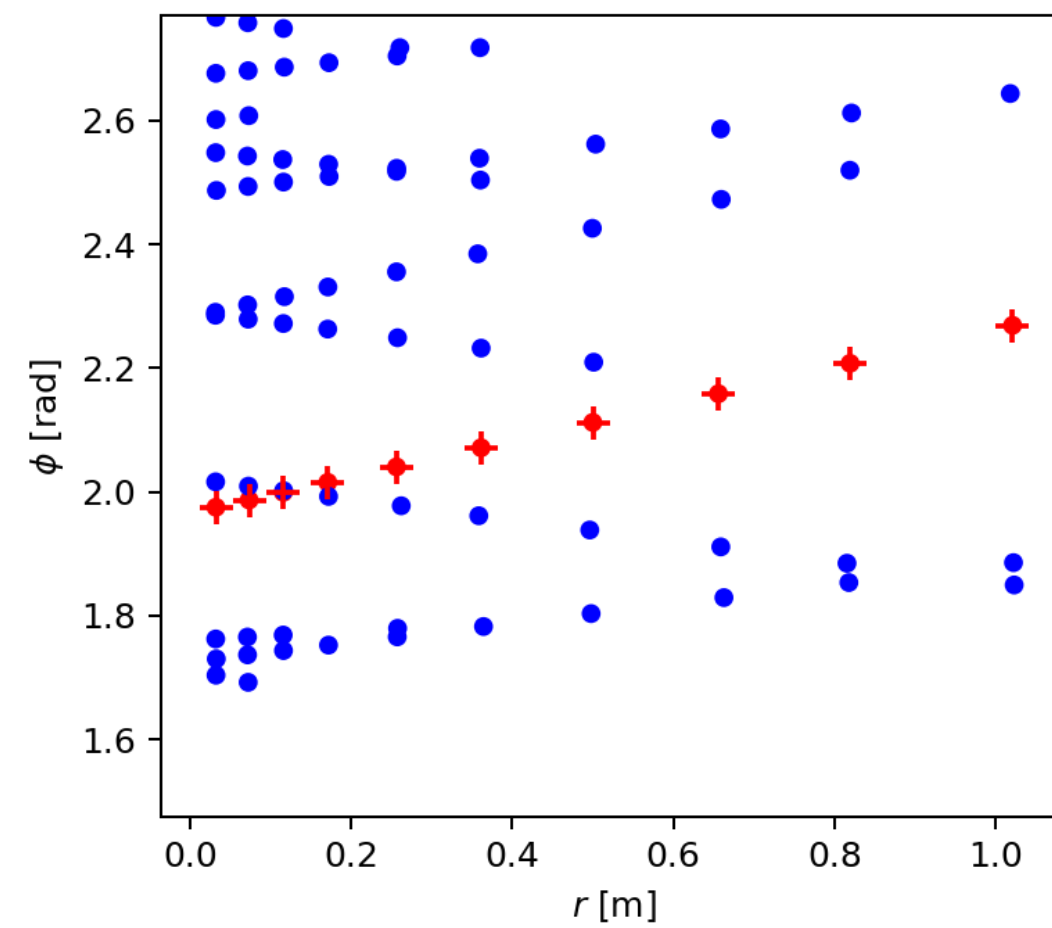
# Beyond images

Theory and  
new algorithms

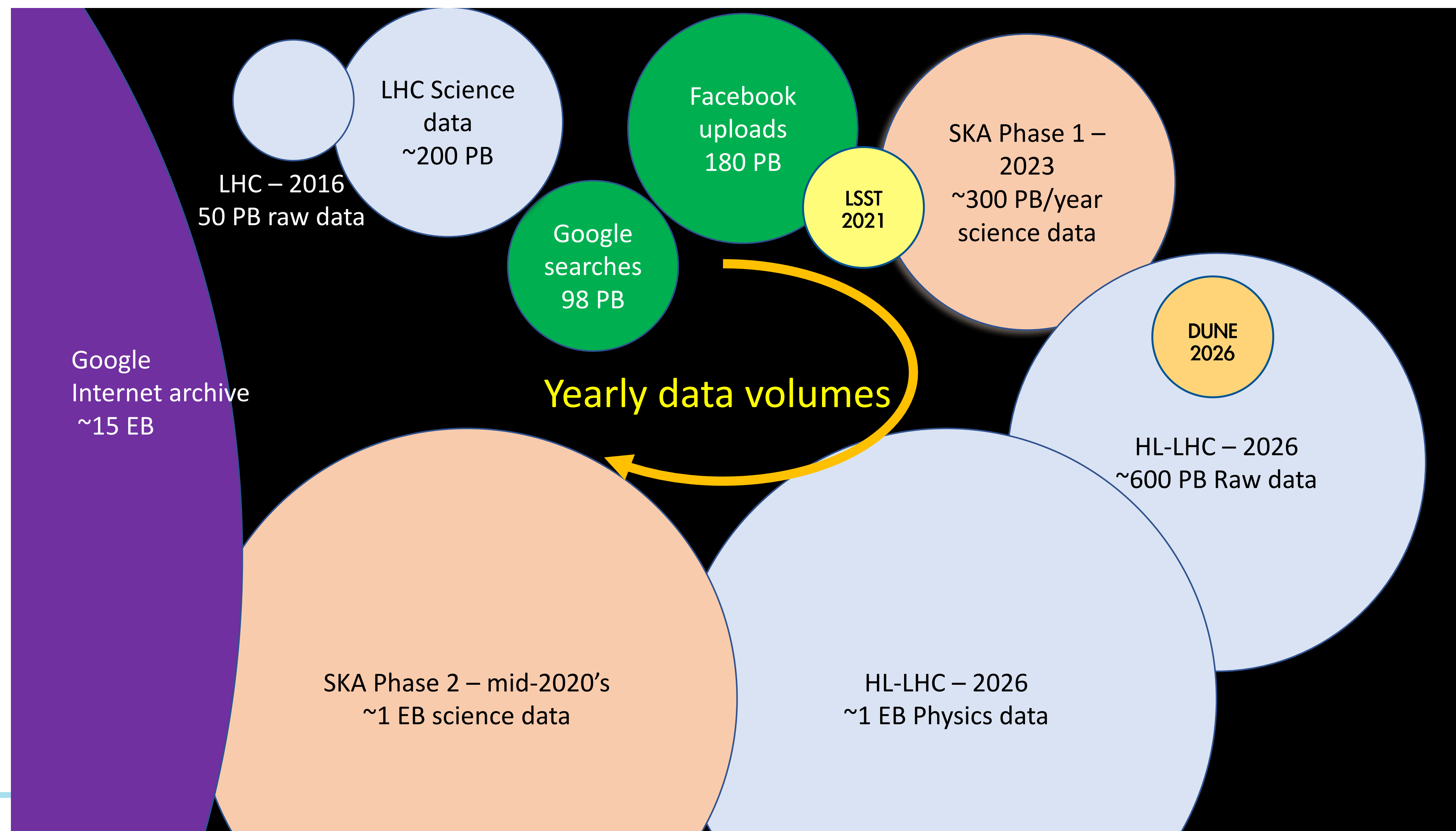
Tracking



Clustering



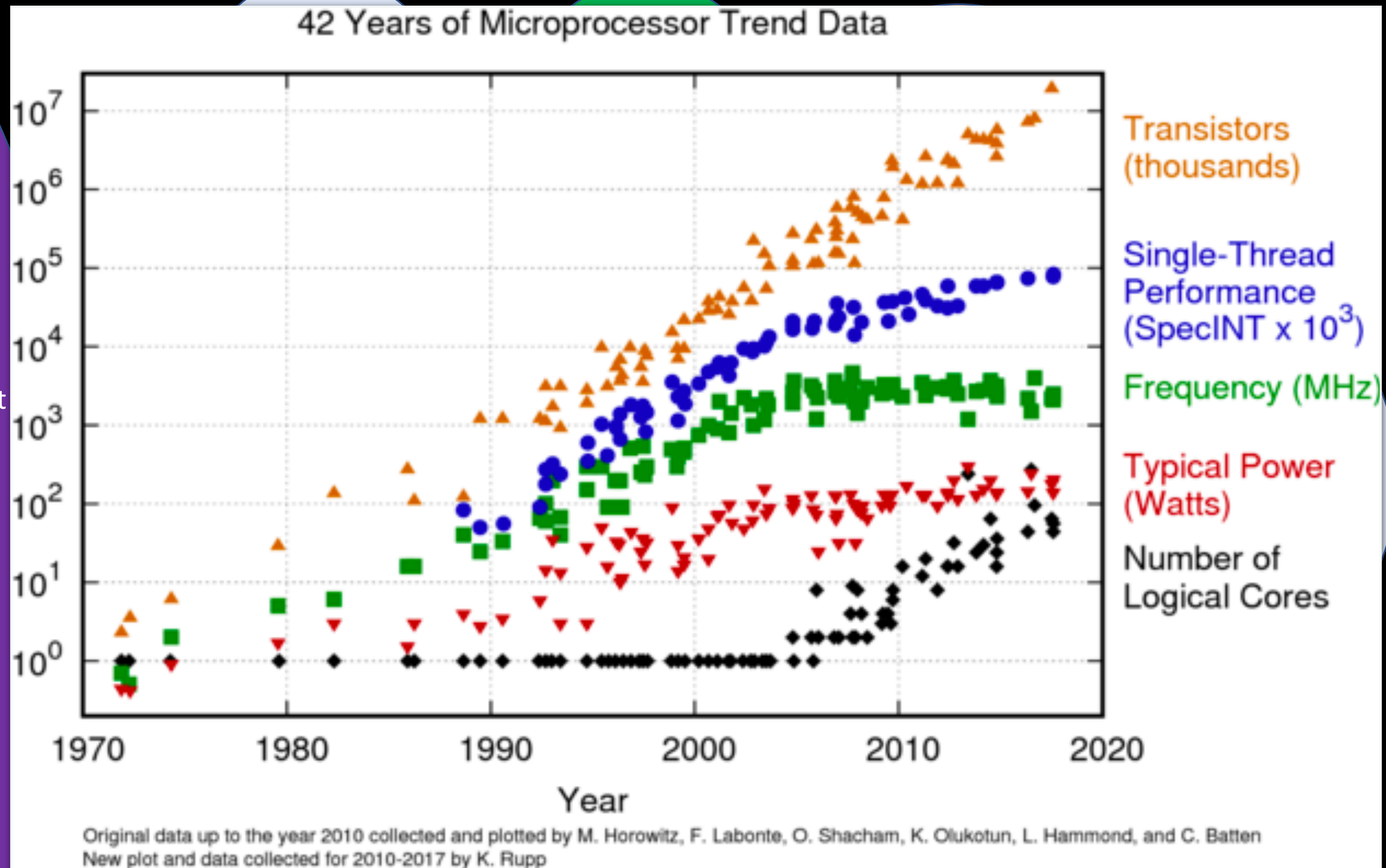
# Big datasets





# Big datasets

Google  
Internet  
~15 EB



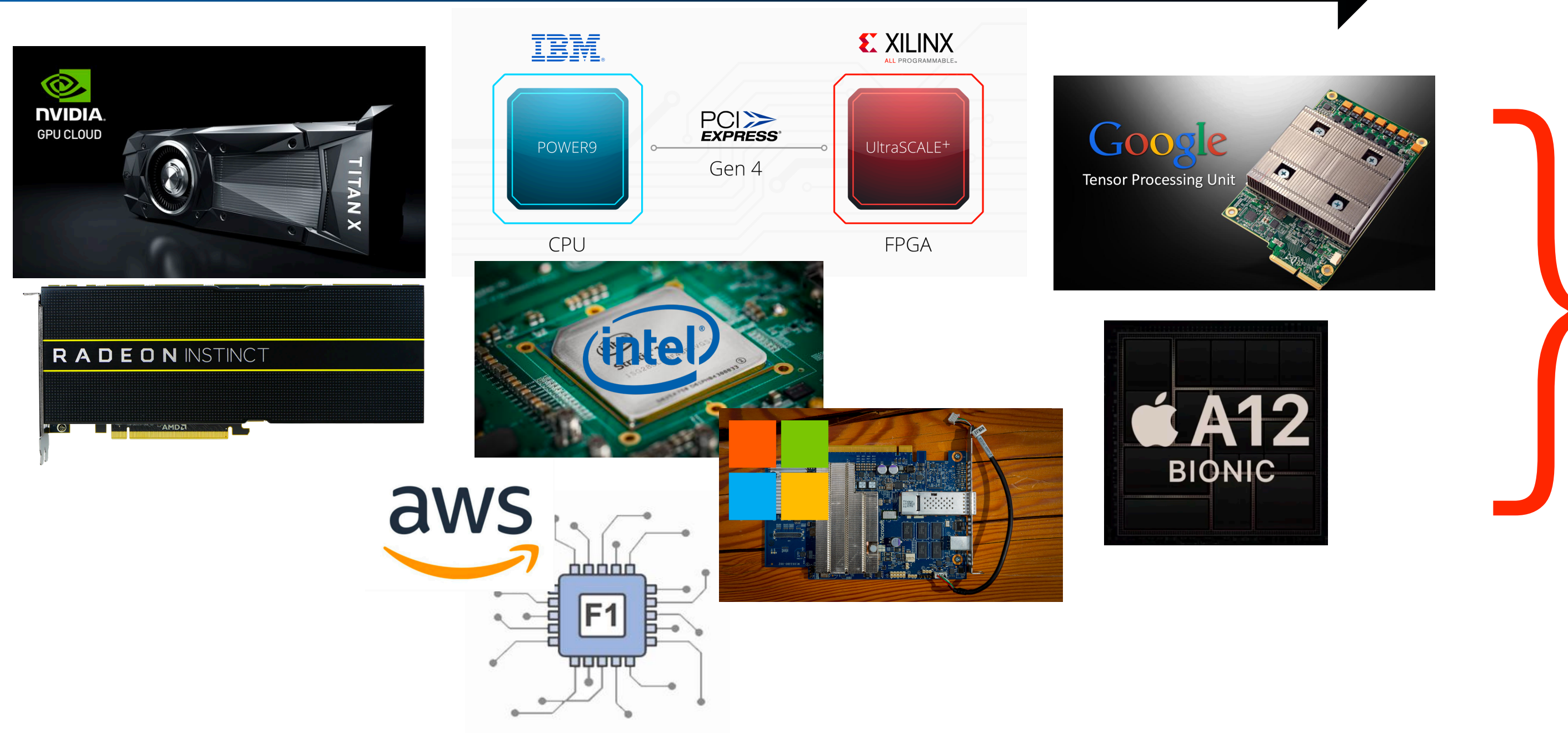


# Computing infrastructure and hardware

Computing hardware and infrastructure



Advances in heterogeneous computing driven by **machine learning and big data explosion**





# Complex and massive datasets

Computing hardware  
and infrastructure

- Big science requires *both* high-performance and high-throughput compute
  - Translation: accelerated computing technologies for **training** and **inference**
- Example, proof-of-concept: CMS requires  $> 10x$  more compute for HL-LHC
  - In collaboration with Microsoft and many university partners,  
**FPGA acceleration of machine learning inference in the cloud and the edge**



Microsoft  
Visit [microsoft story](#)

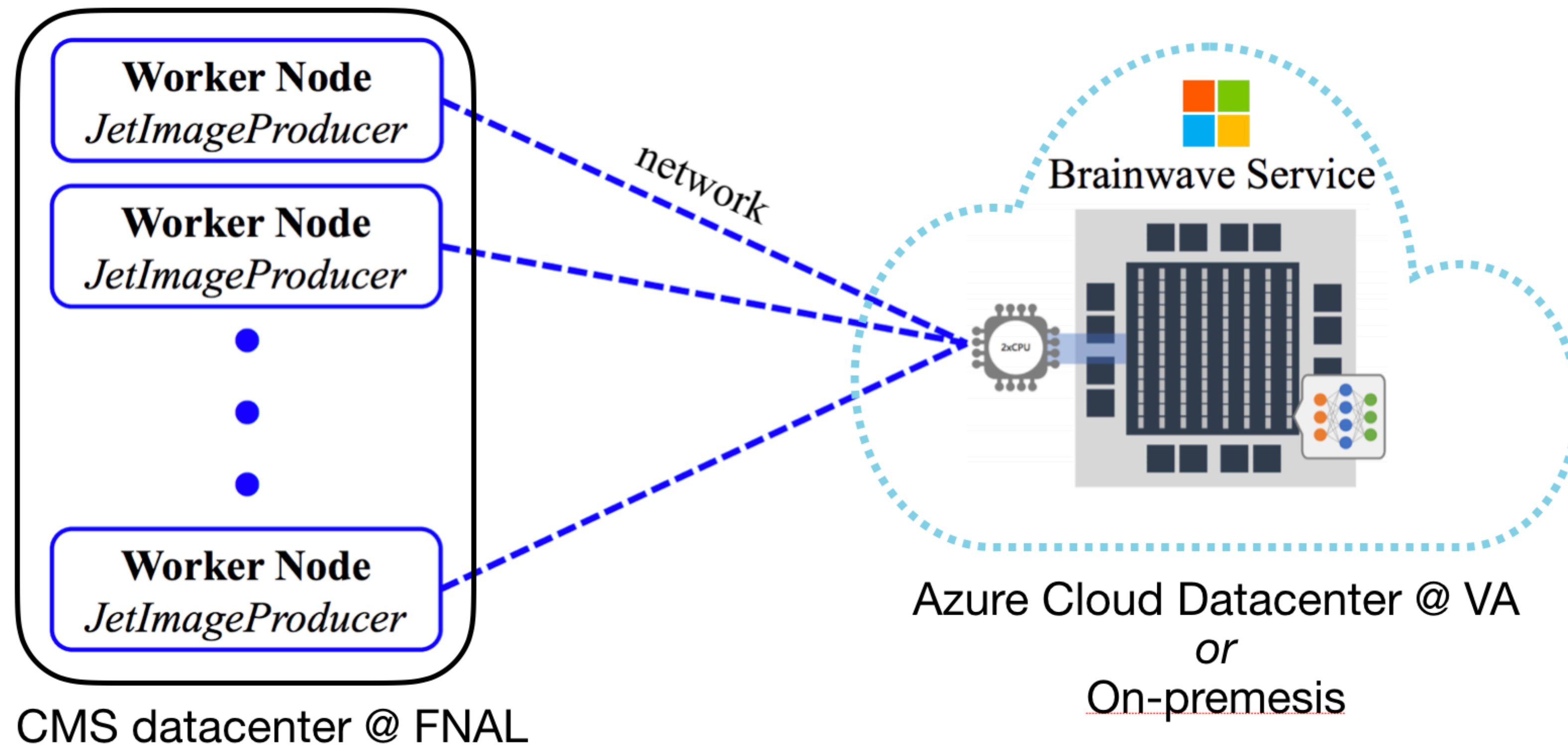




# Complex and massive datasets

Computing hardware  
and infrastructure

- Study found **30x (175x)** speed-up for cloud (edge) inference of ResNet50 over experimental software framework



**Non-disruptive integration**  
into HEP computing model;  
deploying as a service can be  
**more cost-effective**

Exploring various heterogeneous  
hardware and applications  
(*LHC, neutrinos, cosmology*)



# Real-time AI at sensor/edge ★

Real-time AI  
at sensor/edge

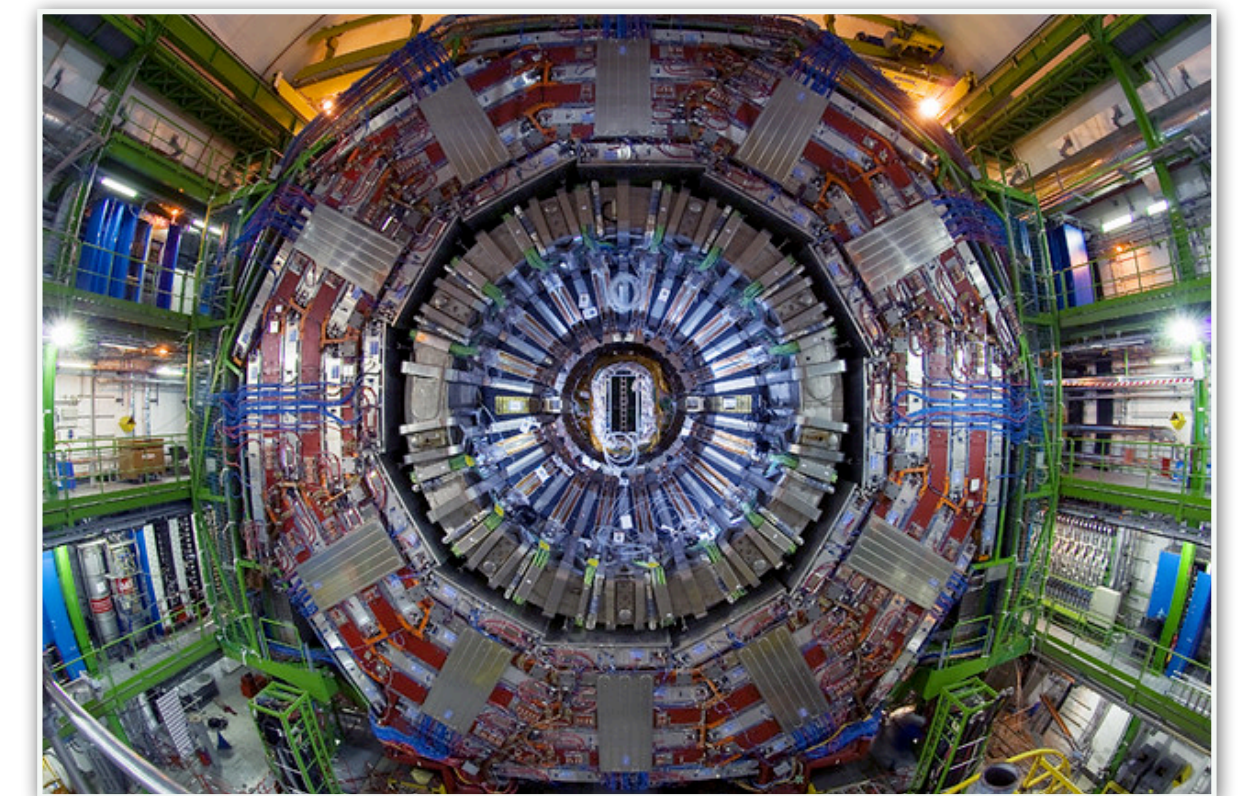
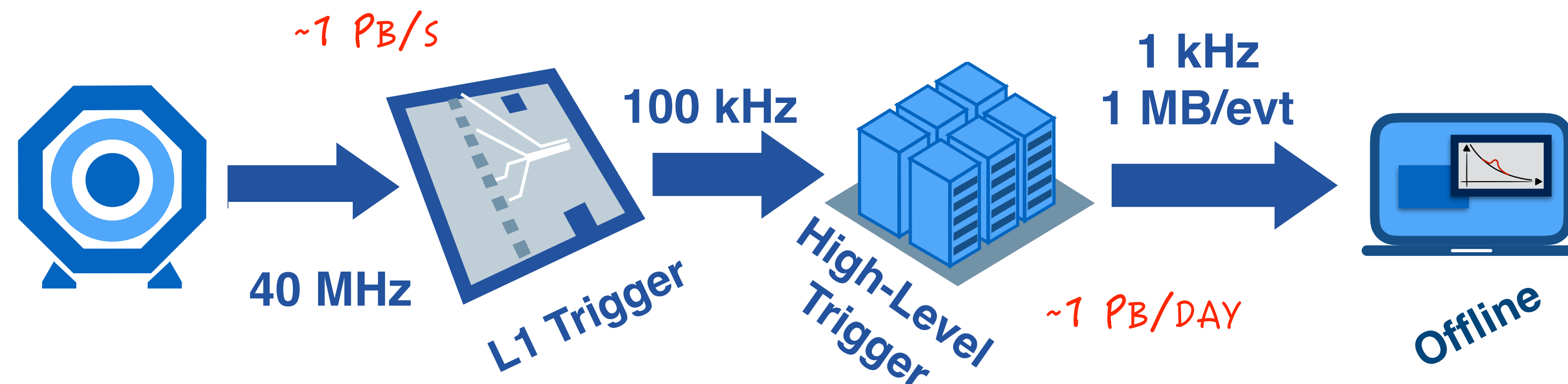
- Resource-constrained AI
- Low-latency, low-power, high bandwidth
- Cryogenics, high-radiation

## LHC at CERN

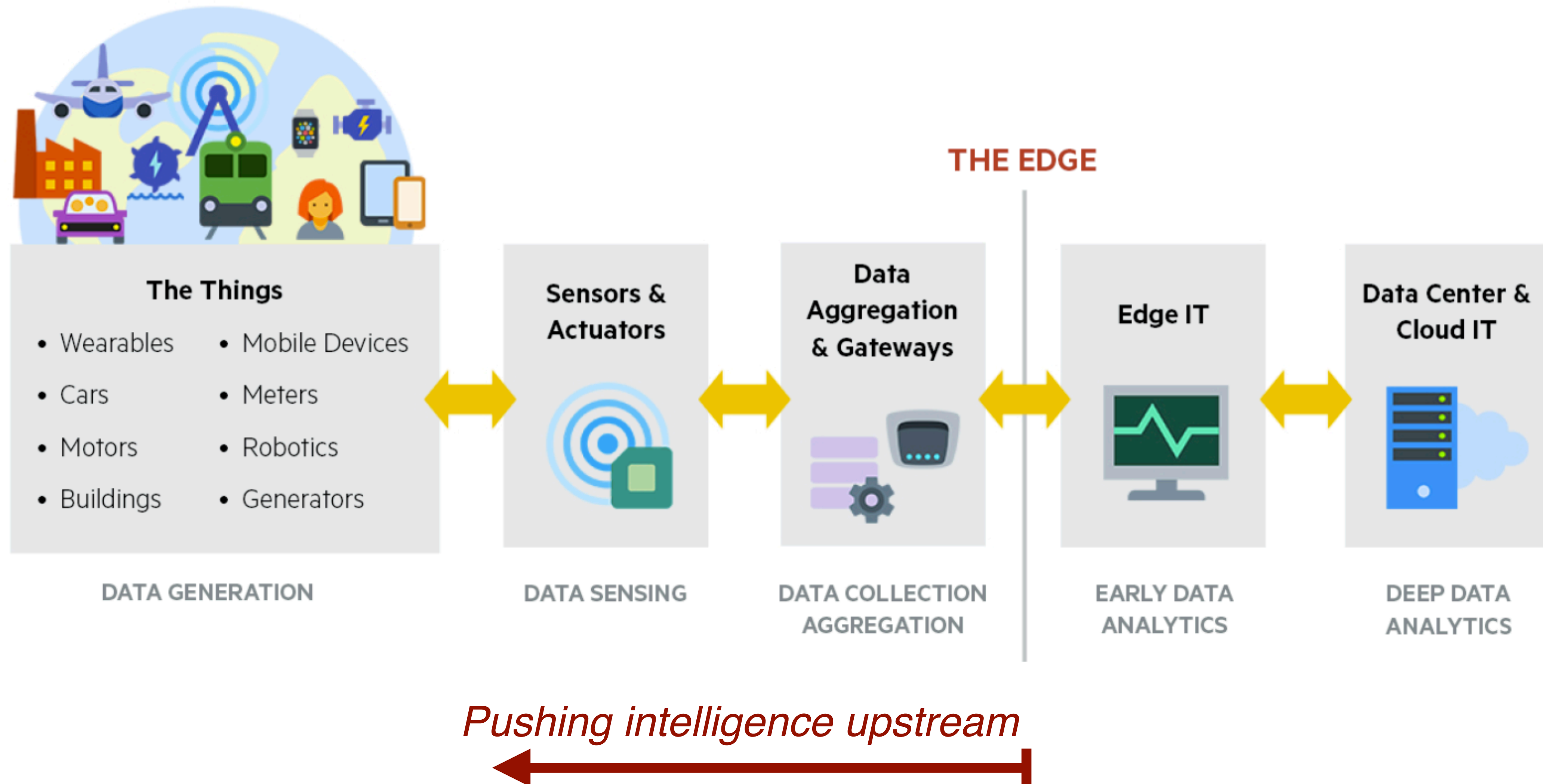
40 MHz collision rate, ~20 hrs/day

## Compact Muon Solenoid (CMS)

→ > 1 billion channels



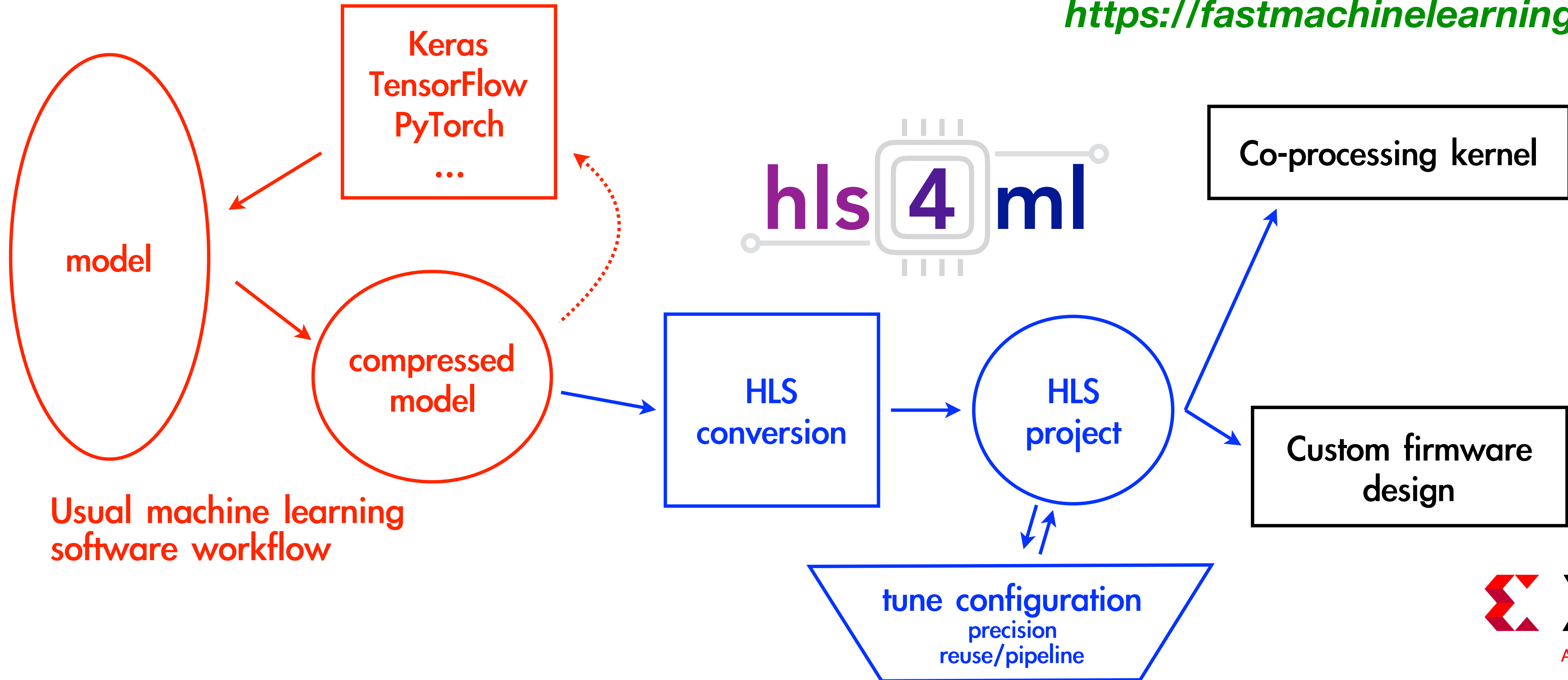




# AI on chip

- **hls4ml** — open-source automated translation tool, ML models to firmware

<https://fastmachinelearning.org/hls4ml>



*featured Xilinx case study!*

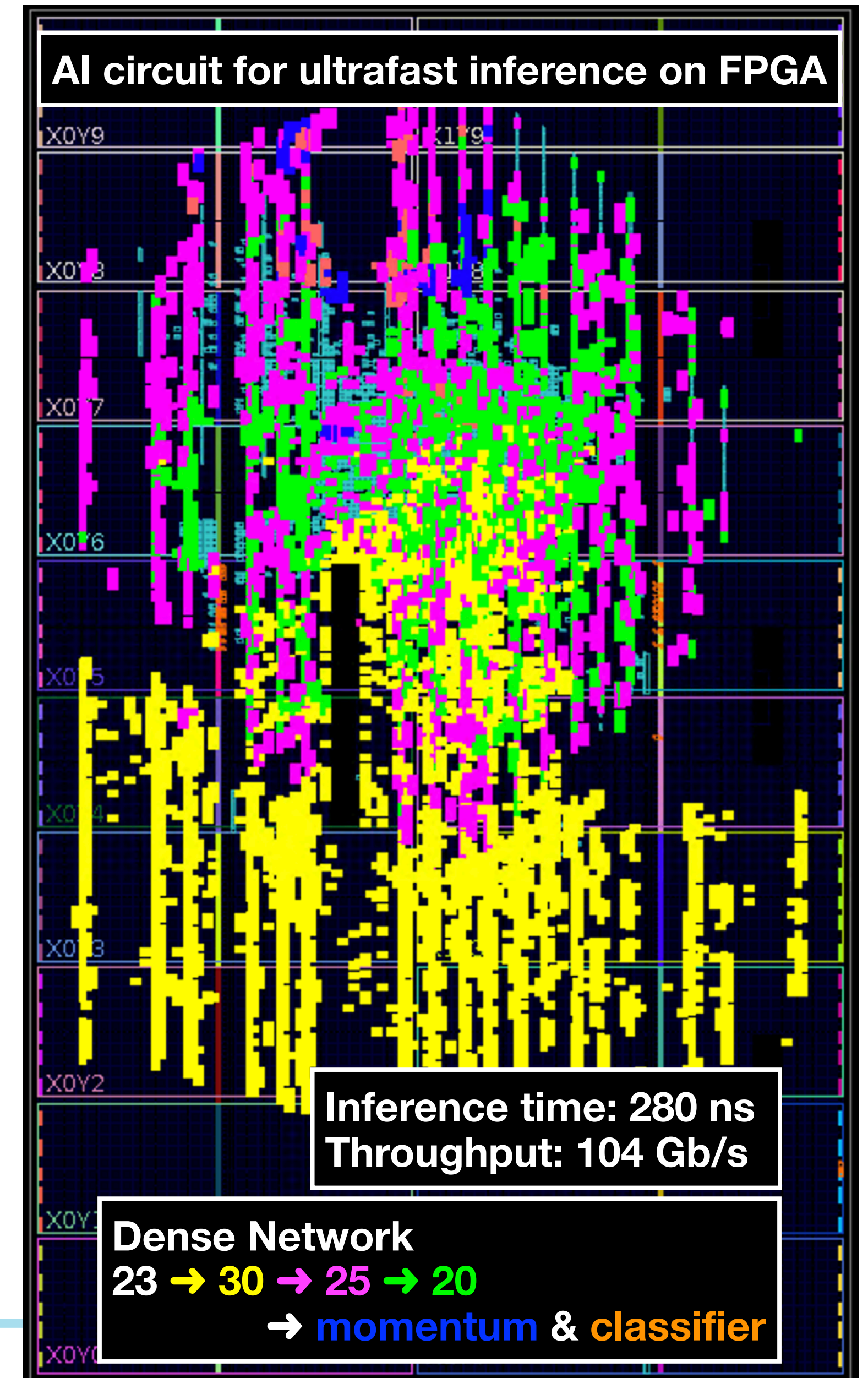


# Edge/sensor AI

Real-time AI  
at sensor/edge

[<https://arxiv.org/abs/1804.06913>]

- **All FPGA design**
  - Flexible: many algorithm types for layers of processing
- **Application** and adoption growing across the LHC
  - Firmware in hours instead of weeks/months
- **Growing interest with many on-going developments**
  - CNNs, Graphs, RNNs, auto-encoders, binary/ternary
  - Alternate HLS (Intel, Mentor, Cadence)
  - Co-processors, multi-FPGA
  - Intelligent ASICs
  - Other physics domains and beyond!





# Edge/sensor AI

Real-time AI  
at sensor/edge

[<https://arxiv.org/abs/1804.06913>]

- **All FPGA design**
  - Flexible: many algorithm types for layers of processing
- **Application** and adoption growing across the LHC
  - Firmware in hours instead of weeks/months
- **Growing interest with many on-going developments**
  - CNNs, Graphs, RNNs, auto-encoders, binary/ternary
  - Alternate HLS (Intel, Mentor, Cadence)
  - Co-processors, multi-FPGA
  - Intelligent ASICs
  - Other physics domains and beyond!

## Fast Machine Learning

September 10-13, 2019 at Fermilab

Sept. 10-11  
IRIS-HEP Blueprint Meeting

Sept. 12-13  
Developer Bootcamp




*Accelerating ML in science:*

- Ultrafast on-detector inference and real-time systems
- Acceleration as-a-service
- Hardware platforms
- Coprocessor technologies (CPU/GPU/TPU/FPGAs)
- Distributed learning

**Local Organization:**  
 Gabriele Benelli (Brown U.)  
 Javier Duarte (Fermilab)  
 Lindsey Gray (Fermilab)  
 Mia Liu (Fermilab)  
 Kevin Pedro (Fermilab)  
 Alexx Perloff (CU Boulder)  
 Zhenbin Wu (U. Illinois Chicago)

**Scientific Organization:**  
 Phil Harris (MIT)  
 Burt Holzman (Fermilab)  
 Shih-Chieh Hsu (U. Washington)  
 Sergo Jindariani (Fermilab)  
 Maurizio Pierini (CERN)  
 Mark Neubauer (U. Illinois Urbana-Champaign)  
 Nhan Tran (Fermilab)

<https://indico.cern.ch/e/FML>





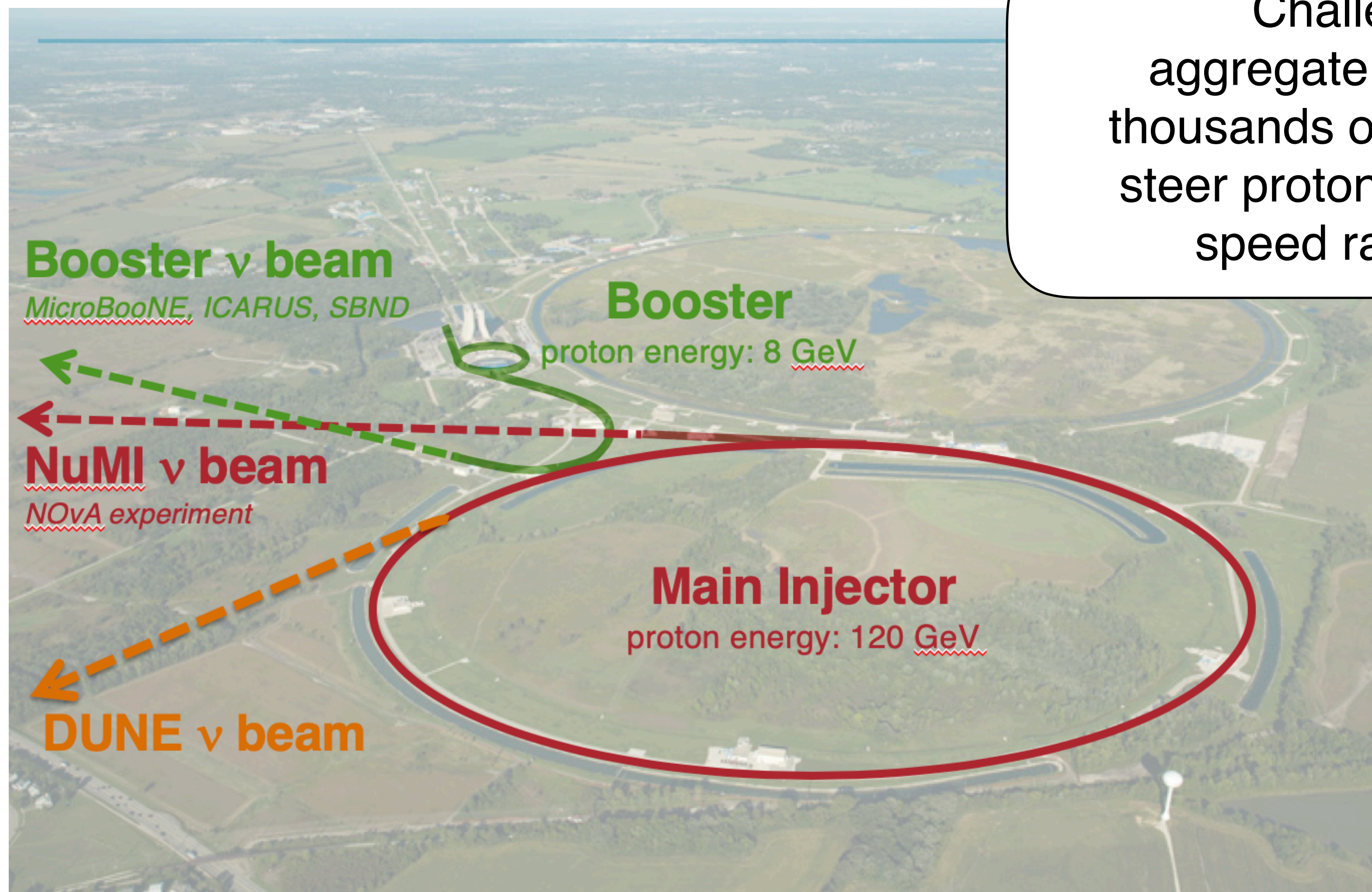





# Operations and control systems

Operations and  
control systems

Challenge:  
aggregate data from  
thousands of sensors to  
steer protons on a light  
speed racetrack

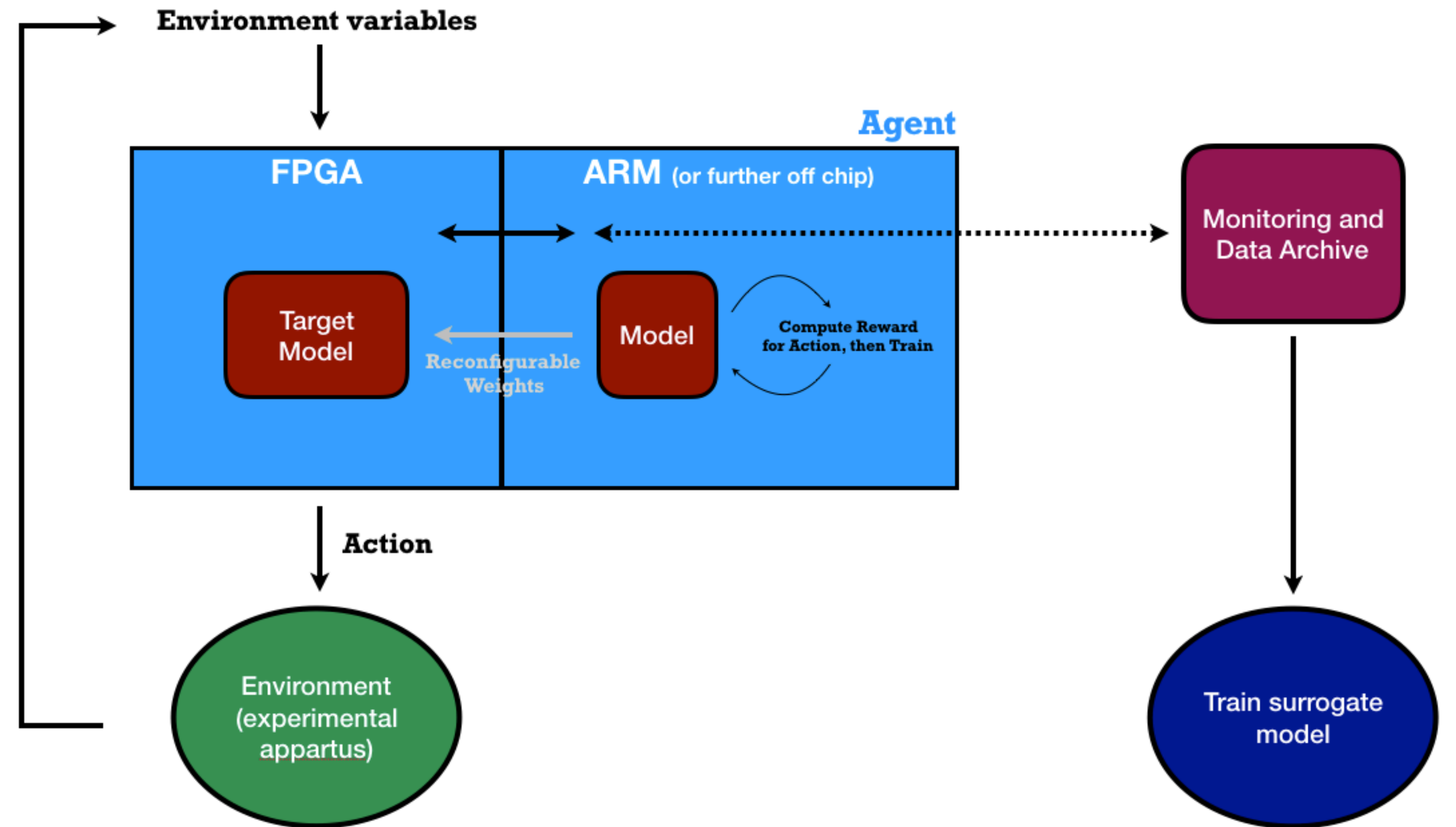




# Accelerator controls with reinforcement learning

Operations and control systems

- Goal to reduce proton beam losses in Booster Accelerator
- Develop **reinforcement learning** algorithm to deployed on FPGA board to control the magnet power supplies (GMPS) — deploy the hls4ml tool
- Single crate control system; project **lays the foundation** for a more ambitious future program.





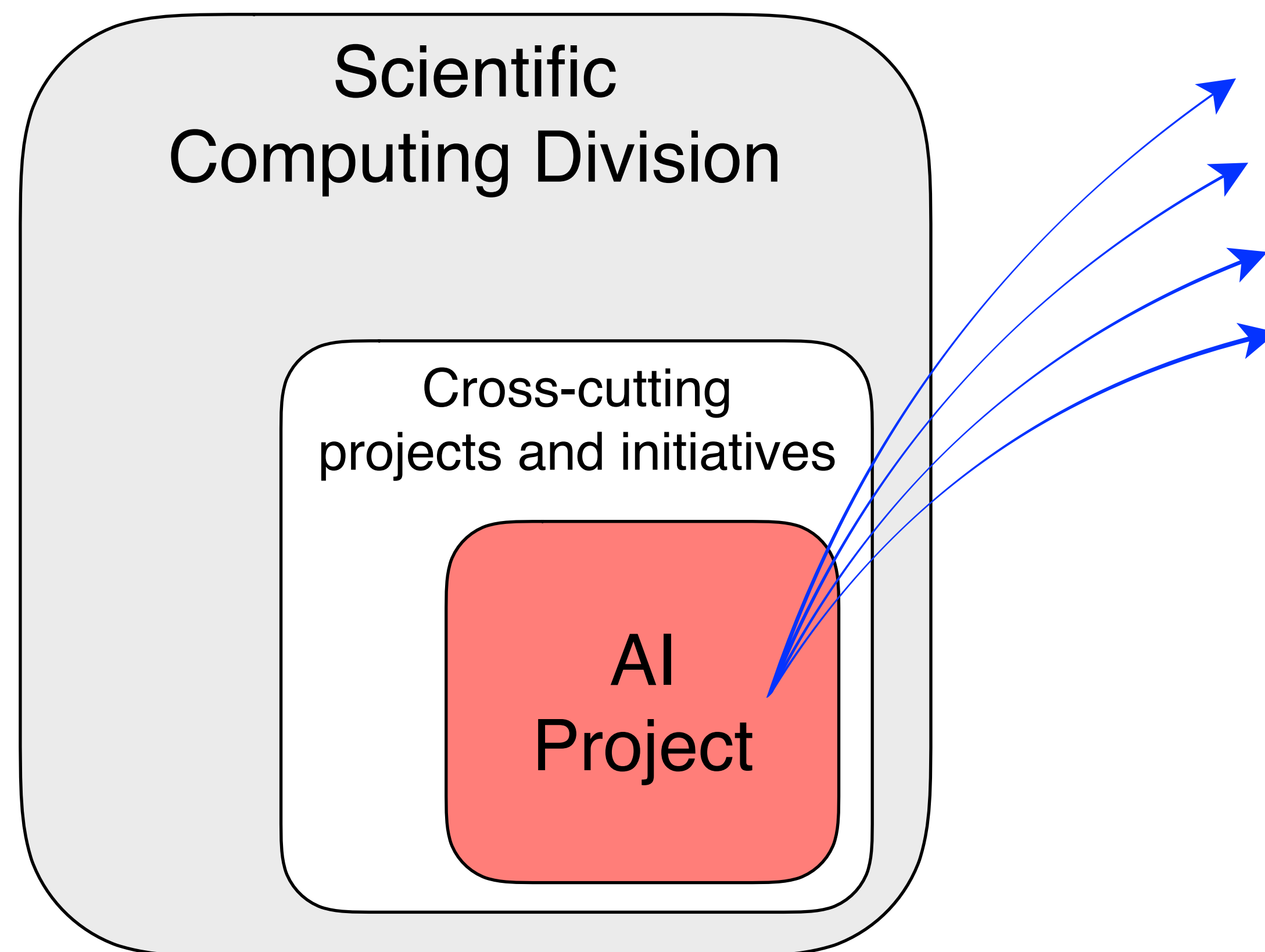
# Outline

**Fermilab & HEP in the AI Ecosystem**  
scientific applications

**AI capabilities and focus areas**  
capabilities developed for HEP

**Who are we?**  
Building a community

# The AI initiative



**Lab-wide initiative**  
 Formal home in SCD,  
 but engaging the entire laboratory

## Artificial Intelligence



Artificial intelligence has the potential to be a transformative technology that benefits nearly all aspects of society. At Fermilab, we are committed to artificial intelligence research and development investments in order to enhance the scientific mission of particle physics.

The unique challenges at the heart of high-energy physics research present opportunities for advancing artificial intelligence technologies. From massive and rich data sets to building and operating some of the world's most complex detector and accelerator systems, the technologies we are developing have potential connections to a broad domain of cutting-edge AI research.

### Fermilab's Artificial Intelligence Project aims to

- Accelerate science with the goal of solving the mysteries of matter, energy, space and time and technologies
- Develop AI capabilities within the national ecosystem that build on high-energy physics challenges
- Build community around cross-cutting problems in order to share the work of Fermilab and the high-energy physics community's AI work with the world

### Project team

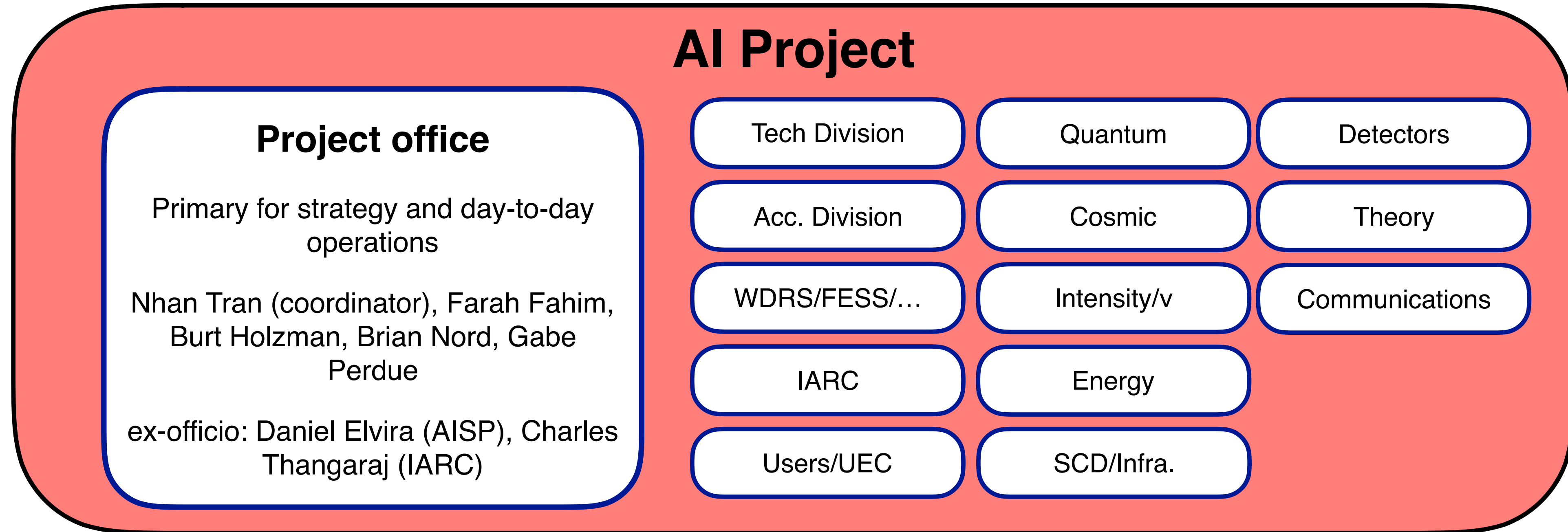
- Farah Fahim
- Burt Holzman
- Brian Nord
- Gabriel Perdue
- Nhan Tran, project lead
- Domain AI experts who serve as liaisons from across Fermilab

Email the project team

*ai.fnal.gov*



# The AI initiative



## **Liaisons:** link across the laboratory

communicate interests and needs of focus area to AI project and focus area participants  
 providing input to overall AI project strategy  
 organize materials, inputs for AI-related funding calls and communications.

# Community building

- Mid-to-long term: **build the community** and focus on **workforce development**
- Seminars, tutorials, hackathons
  - Planning for an AI Jamboree in February (coinciding with engineering week); chance for cross-pollination for experts and enthusiasts across lab, “idea incubators”
- **Engage broader AI & HEP community**
  - Local example: UC/ANL/FNAL joint computational seminar  
<https://indico.fnal.gov/event/22307/>
  - **Existing and growing collaborations with laboratories, universities, industry**
    - many of today’s examples are multi-institutional



# Outlook

## Fermilab & HEP in the AI Ecosystem

### AI capabilities and focus areas

