DUNE AI/ML Kick-Off Workshop 2025

April 29, 2025

Anomaly Detection & Smart TPC

Sunny Seo
Fermilab

U.S. DEPARTMENT of ENERGY

Fermi National Accelerator Laboratory is managed by FermiForward for the U.S. Department of Energy Office of Science

When DUNE reaches Stat. = Syst. Uncertainty?

in CPV measurement

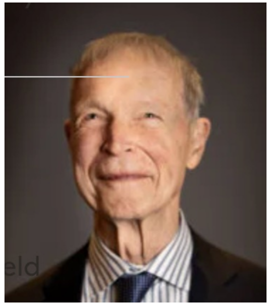
In which exposure?

- AI/ML could reduce **resources** to reach the entrance to the syst. era.
- AI/NL could be also used to improve systematic uncertainty.

→ AI/ML will lead us to earlier discoveries.

Nobel Prizes 2024 with AI/ML

- Physics: *"for foundational discoveries and inventions that enable machine learning with artificial neural networks"*



John J. Hopfield

Geoffrey Hinton

- Chemistry:

The Nobel Prize in Chemistry 2024 is about **proteins**, life's ingenious chemical tools. **David Baker** has succeeded with the almost impossible feat of building entirely new kinds of proteins. **Demis Hassabis** and **John Jumper** have **developed an AI model to solve a 50-year-old problem**: predicting proteins' complex structures. These discoveries hold enormous potential.

Roadmap for DUNE AI/ML (my personal view)

AI/ML Application from start (trigger level) to finish (analysis)

- Anomaly detection (raw data)
- Data volume reduction (raw data, trigger)
- Signal processing
- Pattern recognition (reco, analysis)

01

Anomaly Detection

Searching for Anomalies

We'd like to use **"unbiased data"** to detect anomalies.

<u>Anomalies</u> could be due to:

- New physics (very rare processes) → the best scenario
- Known physics: random & rare (SNB) → one of our missions
- Detector malfunction
- Statistical fluctuation
- Systematic effect

→ Detector Monitoring

Using AI/ML, we would like to detect anomalies at real/near-time.
The AI/ML model can reduce data volume as well.

Anomaly Detection in CMS

arXiv:2309.10157

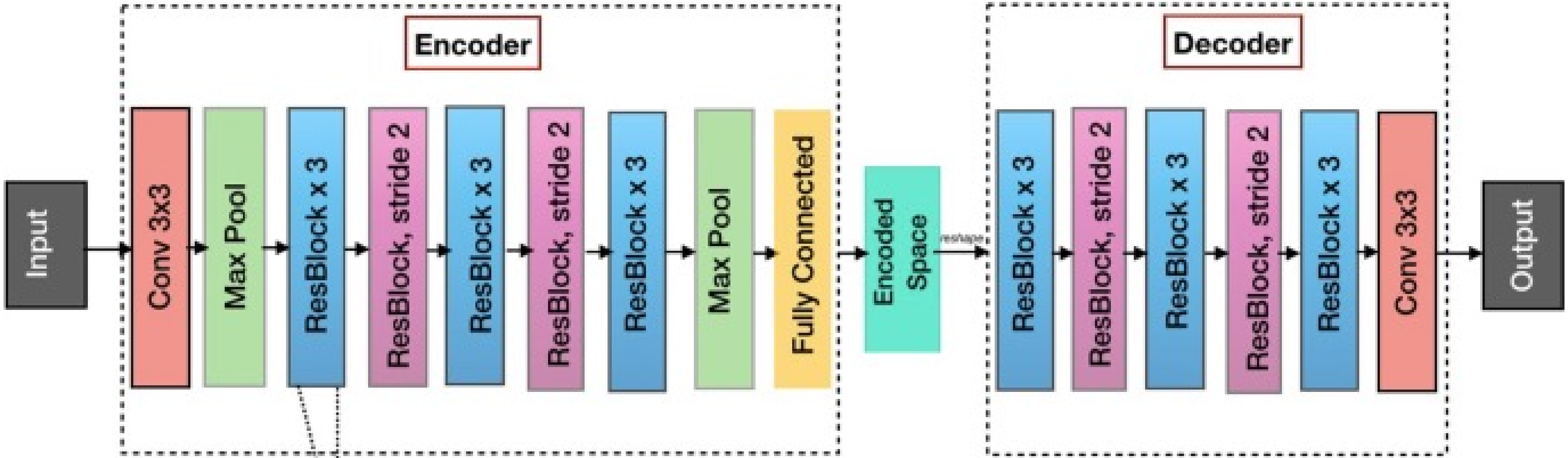
❑ CMS EM Calorimeter:

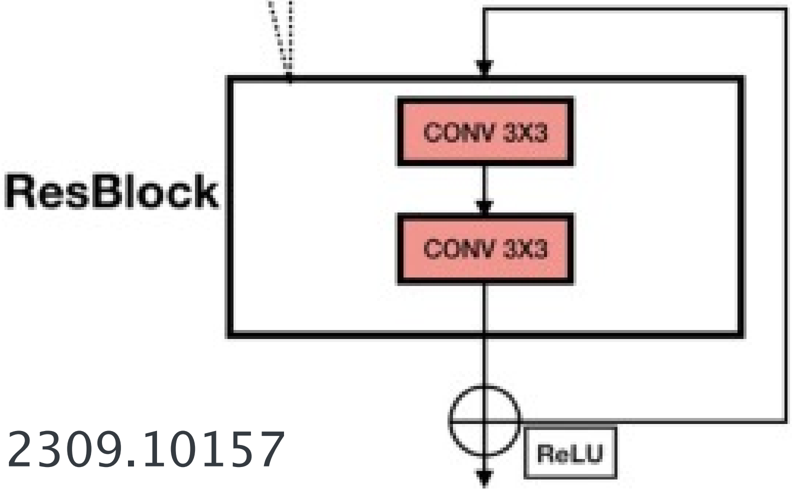
-- **Autoencoder**-based anomaly detection system using **semi-supervised** ML techniques

→ The network is trained exclusively on the certified good physics data set.

→ Identify anomalies without needing prior examples of anomalous data

•The autoencoder takes care of **spatial variations** in detector response and the **time-dependent nature of anomalies**, improving overall detection performance.





arXiv:2309.10157

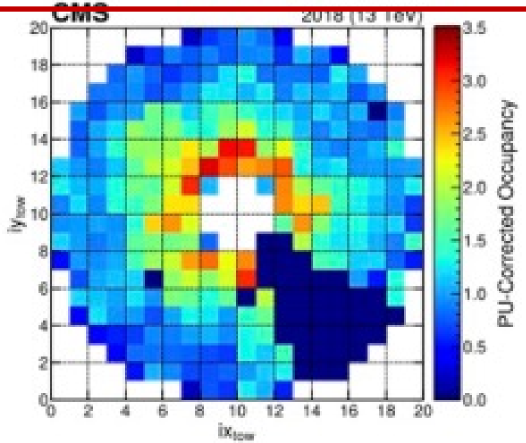
Loss function

input *output*

$$\boxed{\mathcal{L}(x, x') = ||(x - x')||^2}$$

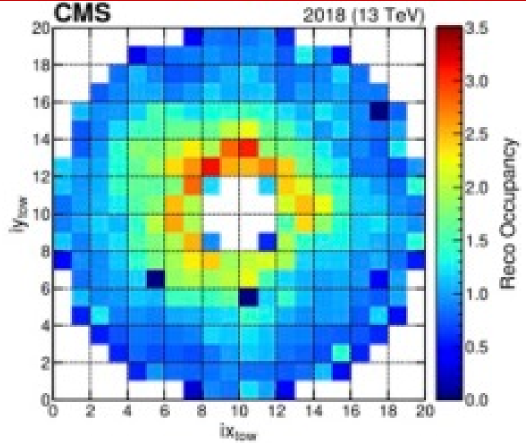
Autoencoder Architecture @CMS Anomaly Detection

Input



Input occupancy histogram with anomaly:
missing sector

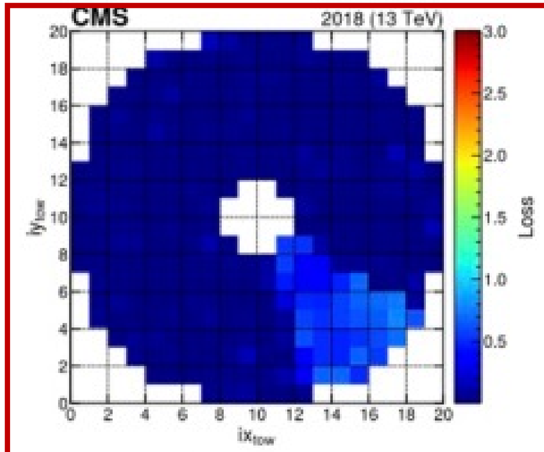
AE
Endcap



Reco-
output

AE-reconstructed image:
anomaly not reconstructed

Spatial & time
correction

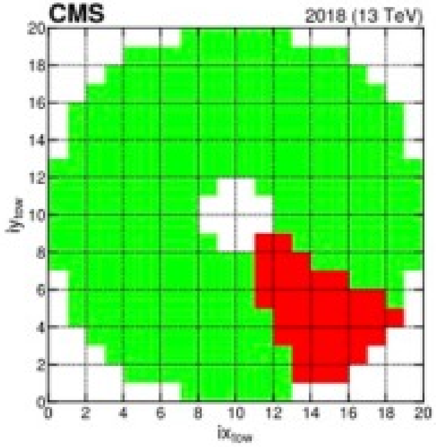


Loss
Map

Loss map:
anomalous showing high loss

Set threshold
for flagging anomaly

Flag
Anomaly



Final quality output:
anomalous towers: **red**
good towers: **green**

CMS ECAL Barrel Fake Anomaly Scenarios

arXiv:2309.10157

	FDR for 99% anomaly detection		
	Missing Supermodule	Zero Occupancy Tower	Hot Tower
Baseline			
Baseline no correction	14%	90%	5.2%
Baseline after time correction	5.9%	80%	$< 0.01\%$
Autoencoder			
AE no correction	3.6%	51%	2.8%
AE after spatial correction	3.1%	49%	2.9%
AE after spatial and time corrections	0.13%	4.1%	$< 0.01\%$

Benefits of Autoencoder-based Anomaly Detection System

- Real-time monitoring → immediate feedback
- Enhanced detection capabilities
- Low false discovery rate
- Automation of anomaly detection

➢ We can adopt the same/similar method in DUNE TPC & PDS for anomaly detection.

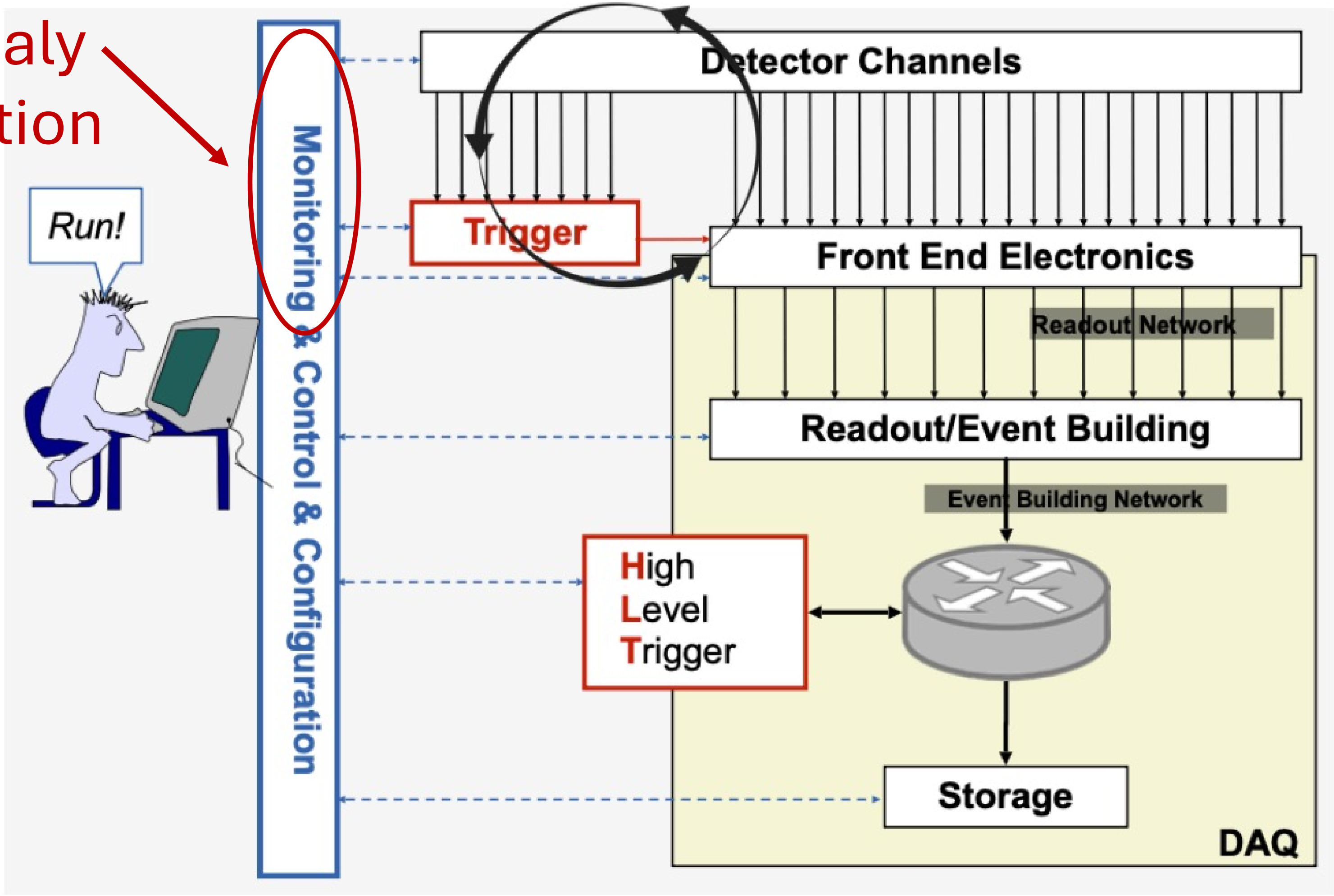
➢ SuperNova Burst can be detected in “real-time” as an anomaly.

02

Smart TPC

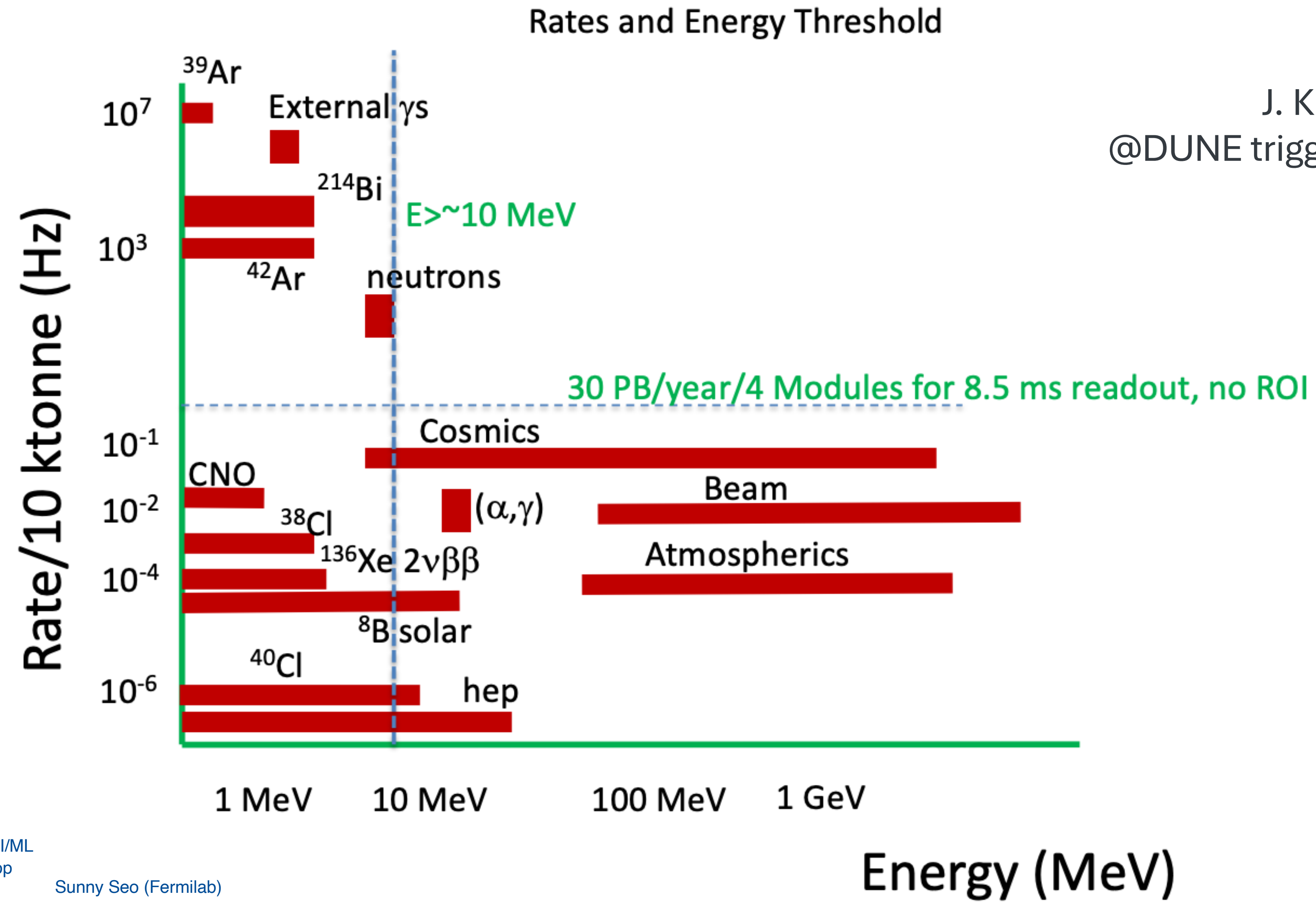
DAQ Trigger & Monitoring

Anomaly detection

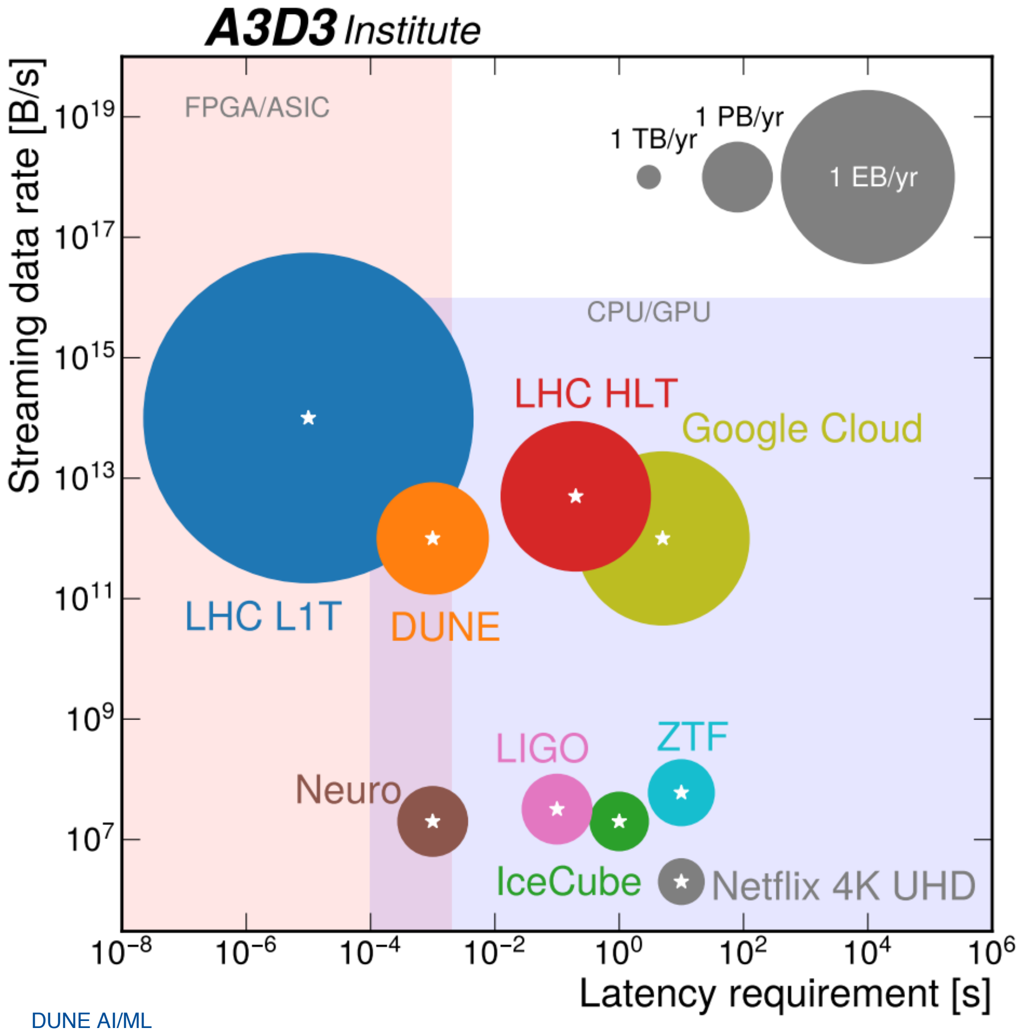


Trigger Design Constraints

J. Klein's talk
@DUNE trigger workshop, 2025



Motivation for Data Volume Reduction



→ We expect high raw data volumes for (HL-)LHC and DUNE.

❑ DUNE: arXiv:2203.16255

- ~ TB/sec → ~32 EB/year
- 100 % livetime (for Supernova)
- > ~ 20 years of operation

❑ DUNE's cap for computing: ~ 30 PB/year

→ 3 orders of magnitude reduction is needed.

Data Volume Summary Table

J. Klein's talk
@DUNE Trigger Workshop, 2025

Under the assumption we read out 2x drift time * all channels with no zero suppression, no ROI imposition, and no compression:

Source	Annual Data Volume/ 10 kt	Assumptions
Beam interactions	10 TB	700 νs+700 dirt μs; **everything** read out for 8.5 ms (no ZS); 10 MeV threshold in coincidence with beam time; include cosmics
Cosmics (+atmospherics)	10.5 PB	**All channels** in 8.5 ms window around HE event; 4000/day/10 kt
Radiologicals	1 PB	E > 10 MeV so rate < 10% cosmics; **no ROI imposition**
Supernova Bursts		Fake rate of 1/month; 100 s storage/burst
Front-End cals	200 TB	Worst case of measuring every single ADC bin with 100 measurements/point; four times/year
Radioactive source cals	100 TB	**Source rate < 5 Hz (or pre-scaled to that);** only one APA/CRP readout; PDS is negligible; full readout window per tag; no ZS
Laser cals	200 TB	1×10^6 total laser pulses; **tight ZS** for both induction and collection; ½ of all wires in TPC illuminated
Random Triggers	100 TB	Same as cosmics scheme; rate is 40/day
Trigger Primitives	38 PB	All planes; 26 B/primitive; noise rate = ^{39}Ar rate;

~50 PB /year/module

Still, we should reduce data volume!
(~one order of magnitude)

Proposed Solution

- AI/ML “*in-situ*” → Reducing computing resources
  - Reducing raw data volume by applying AI/ML “*in-situ*”
  - Use **all subdetector info “*in-situ*”** to increase S/B

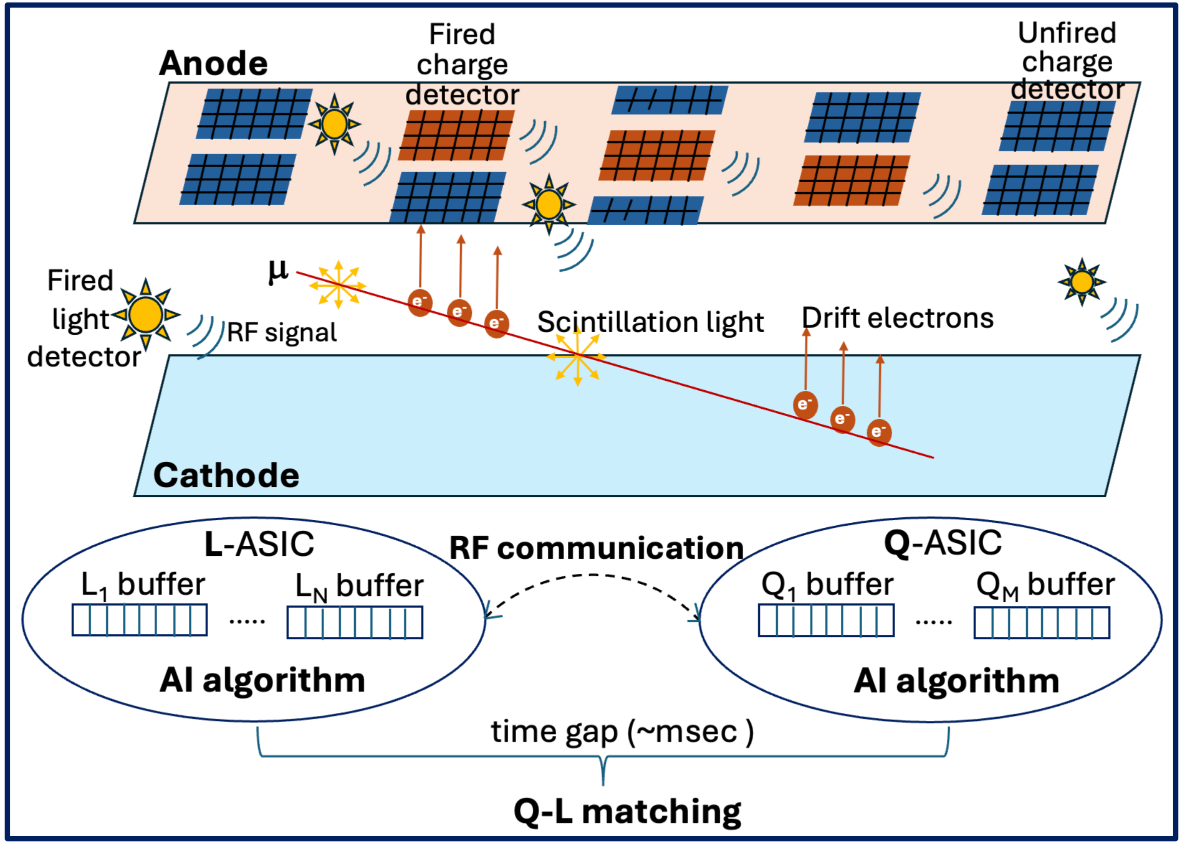
Unique & Challenging

- RF network → Reducing use of materials
  - Essential for subdetector communication
  - Enabling cable-free detector

- Cryo dSiPM → Saving project budget
  - ~10% that of commercial SiPM

AI/ML “*in-situ*” in LArTPC

Smart TPC: Charge (Q) + Light (L) matching “*in-situ*”



- ❑ Q-L matching “*in-situ*”
- ❑ Pattern recognition “*in-situ*”
  - Time
  - Space
  - Deposited charge & light

→ RF network among all subdetector components are essential.

Test of RF Device(s) at LAr

❑ RF communication between Q & L detectors is essential for raw data reduction.

- We will start with a commercial RF device(s) to test it at LArTPC.
  ex) LoRa
  low power, cheap

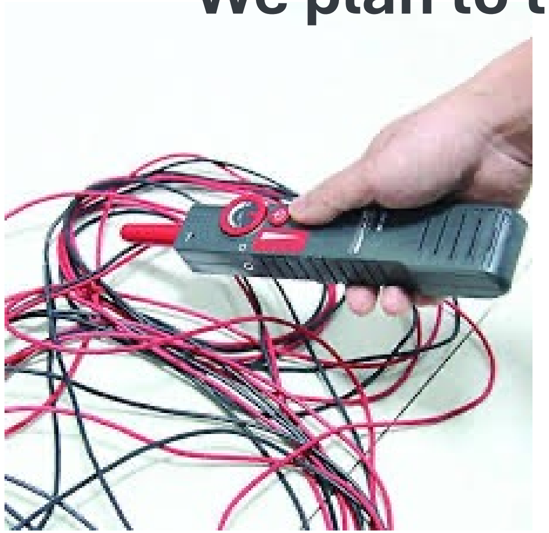
- Through tests in LAr,
  the device could be optimized to work properly in Cryo temperature.

RF Communication

❑ Not only RF communication between sub-detectors, we can consider sending out raw data over RF network.

→ This will allow us to build a cable-free detector.

We plan to test this during our R&D.



Here, imagine DUNE detector w/ & w/o cables.

Alternative Light Pixel Detector

SPAD: Single Photon Avalanche Diode

❑ Cryo dSiPM

→ Currently under-development

by Fermilab, EPFL, and Global Foundry

- Pros:

  -- Cheaper (~10% of commercial SiPMs)

- Cons:

  -- High dark count rate (→ need to measure it in LAr)

  -- Need VUV wavelength shifter

Key Technologies in this R&D

- **AI/ML on Chip (ASICs)**
  → Reduce raw data volume "*in-situ*"

- **RF Network**
  -- Communication between Q & L detector
    → essential for Q-L matching "*in-situ*"
  -- Raw Data transfer
    → cable-free detector

- **Digital CryoSPAD**
  -- Alternative to commercial SiPMs
  -- not required for this R&D (but helps reduce budget)

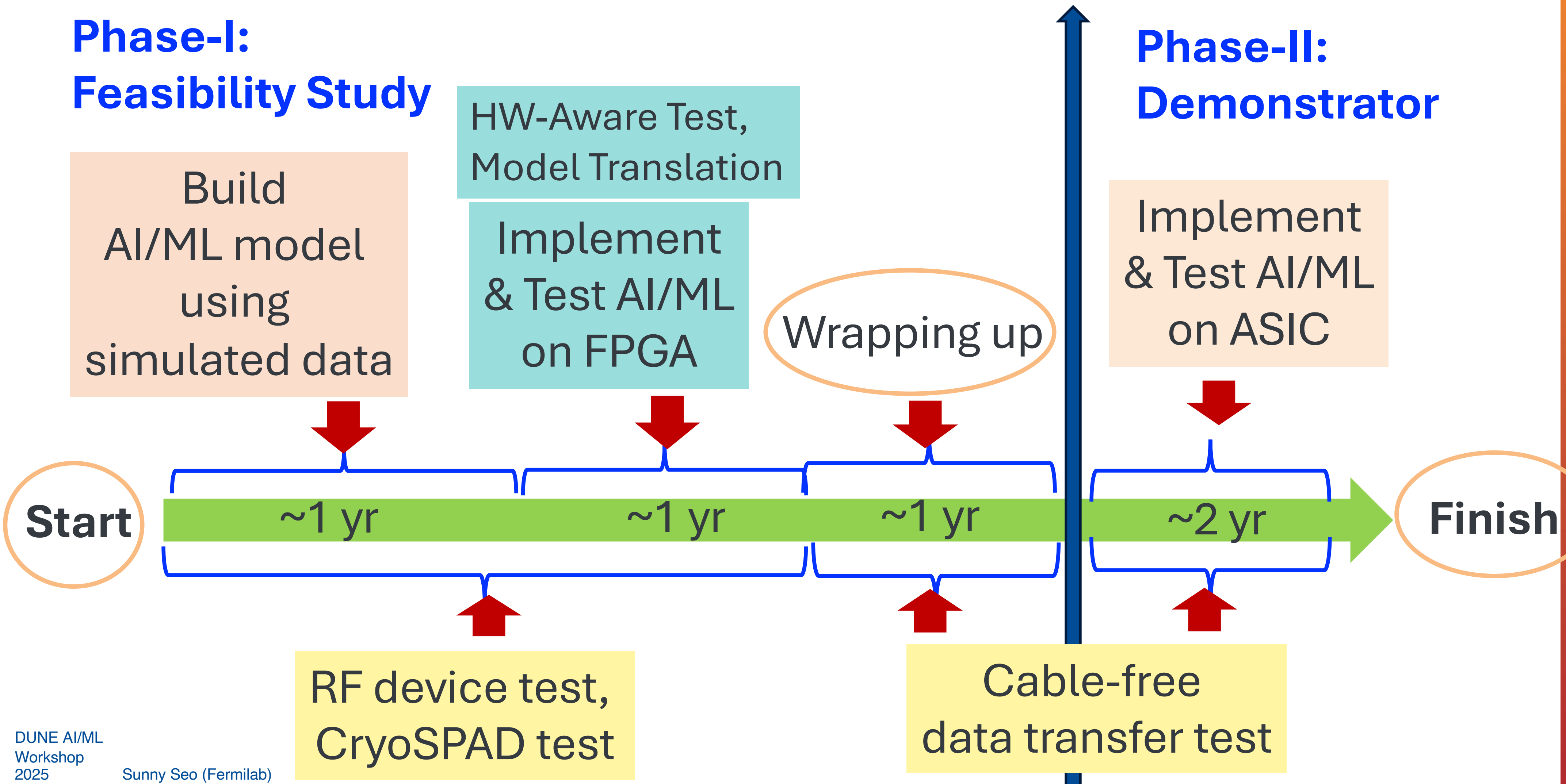
Application to any other experiments

AI/ML Model to Hardware

Model Building	HW-Aware Test	Model Translation	HW Implementation
	To optimize the model for given HW constraints (memory, speed, power,..)	Co-design model (physicist, AI expert, HW engineer)	
AI/ML model	Model Optimization (for resources)	HLS4ML (SW → HW)	AI/ML on FPGA, ASIC
+ Simulated data	+ Quantization + Pruning	+ Optimization + Profiling + Tuning	

HLS =high level synthesis

Timeline of “Smart-Pixel TPC” R&D



AI/ML Start-to-Finish

Trigger level

Data Processing
(noise filtering, calib)

Reconstruction

Data Analysis

❑ Real-Time Anomaly Detection

* New physics search
* Checking data quality
* Monitoring detector

❑ Q-L Matching "*in-situ*"

* Increase S/B ratio at the beginning
* Reduce data volume

Roadmap for DUNE AI/ML (my personal view)

Employ a "**Foundation Model**" for LArTPC experiments

- ❑ A **foundation model**, also known as **large X model (LxM)**, is a machine learning or deep learning model that is trained on vast datasets so it can be applied across a wide range of use cases.

- ❑ Foundation model is itself incomplete but it serves as the common basis from which many task-specific models are built via adaptation.

- ❑ The term **'foundation'** to connote the significance of architectural **stability, safety and security**: poorly constructed foundations are a recipe for disaster and well-executed foundations are a reliable bedrock for future applications."

Common Triggers → Triggers/Physics
→ Data Processing → Reco → Analysis

Summary & Conclusion

- ❑ AI/ML has become an essential tool for modern science, as proven by Nobel 2024.

- ❑ DUNE, as an imaging detector, is the best place to apply AI/ML.

- ❑ DUNE has been applying AI/ML mostly to reconstruction, but it needs to extend it to the entire process of data taking.

- ❑ Applying AI/ML to raw data and the trigger level would enhance DUNE's potential for its discoveries.

→ If you are interested in this, let's work together. Let me know!