Machine Learning Event Reconstruction for the Hyper-Kamiokande Far Detector

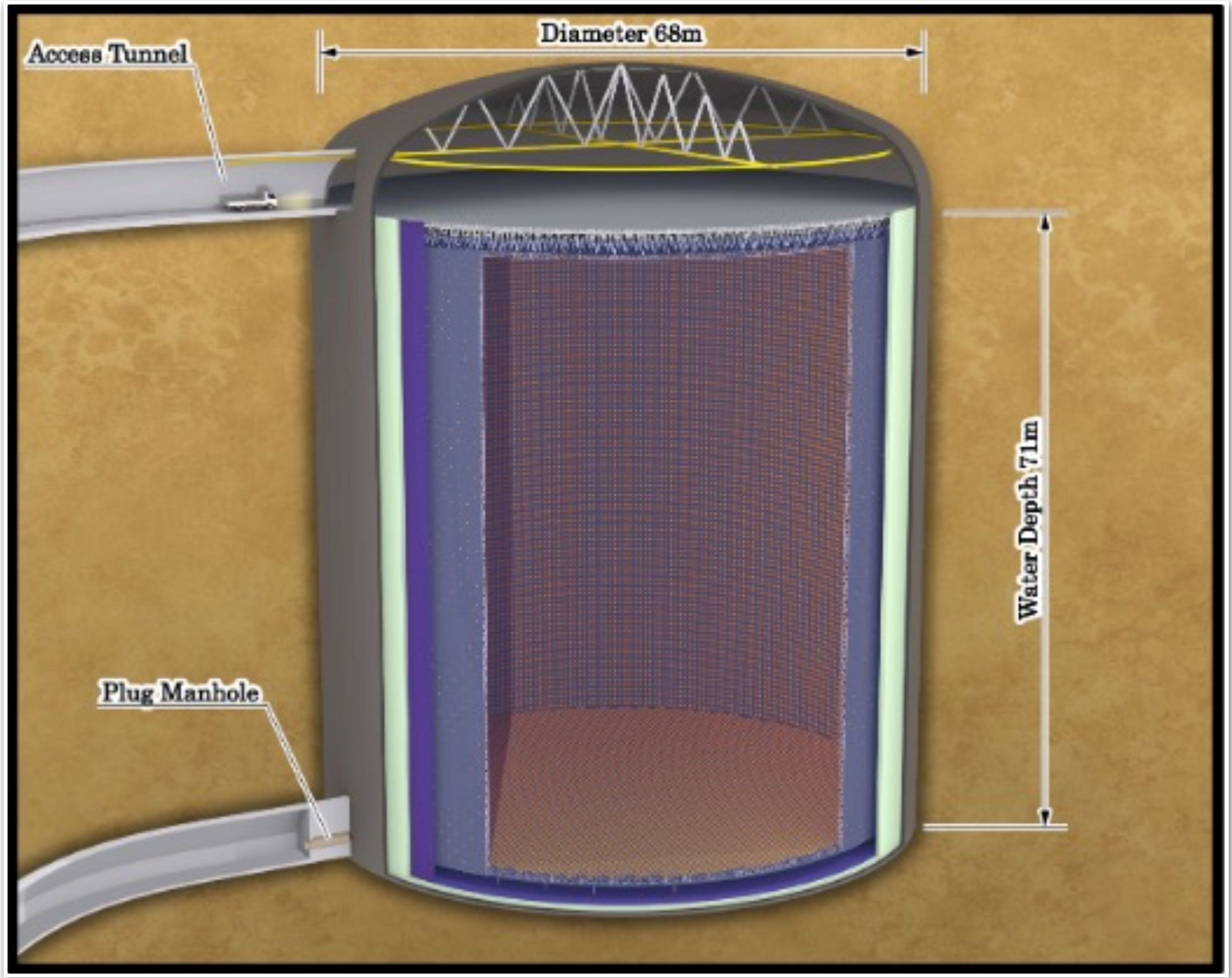
Naomi Foster

Supervised by
Neil McCauley, Sam Jenkins and Jon Coleman

Hyper-Kamiokande

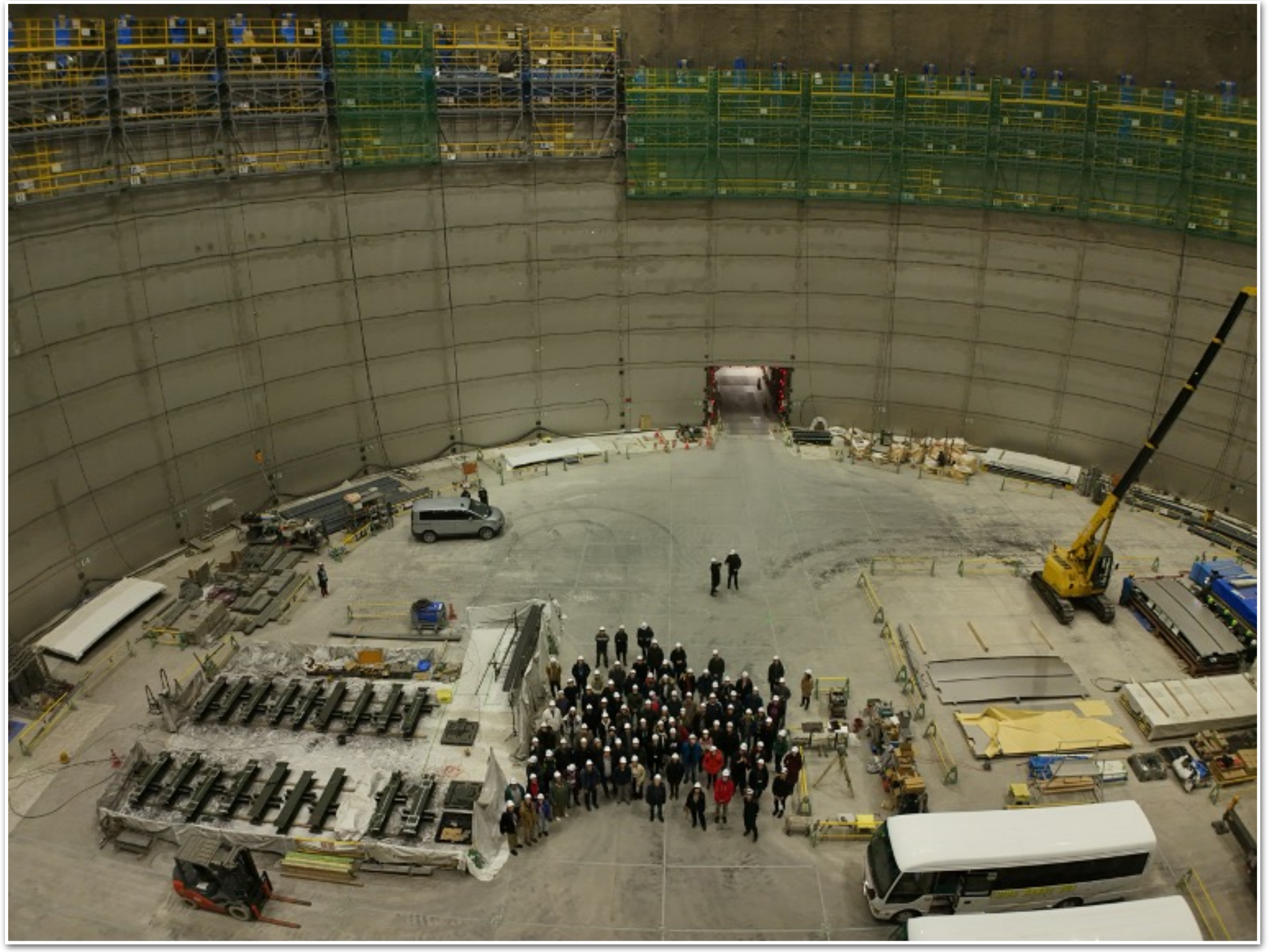
UNIVERSITY OF LIVERPOOL

Hyper-Kamiokande Far Detector



Water Cherenkov detector for neutrino detection.

- 250 ktons ultra-pure water.
- 1300 m w.e rock overburden.
- ~ 20,000 20" photomultiplier tubes.
- 8.4 times the fiducial volume compared to Super-Kamiokande, with the same effective photocathode coverage.



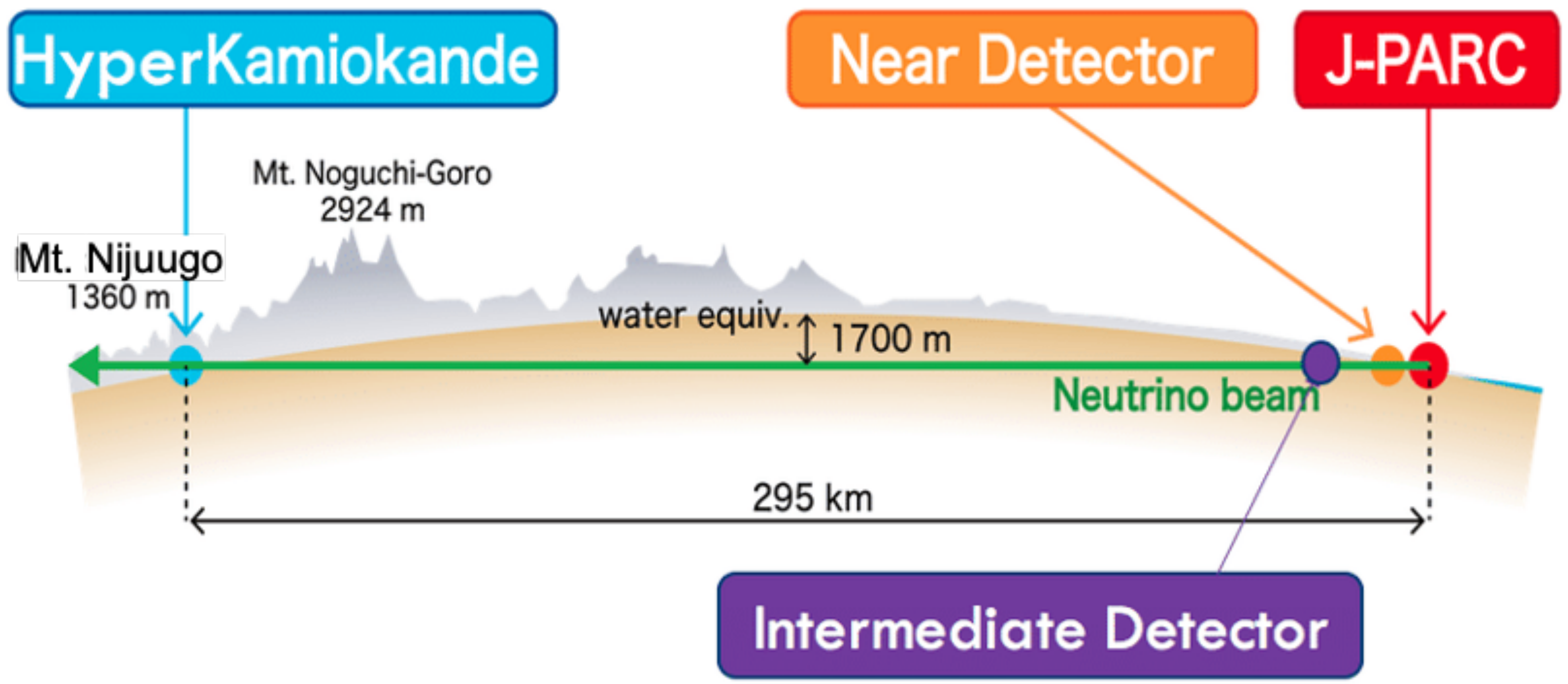
Planned completion in 2028, current status seen in above photo of our visit during the February collaboration meeting.

Accelerator Neutrinos and Measuring CP Violation

At J-PARC a 1.3 MW beam of very pure muon (anti)neutrinos will be produced and measured by a suite of near detectors.

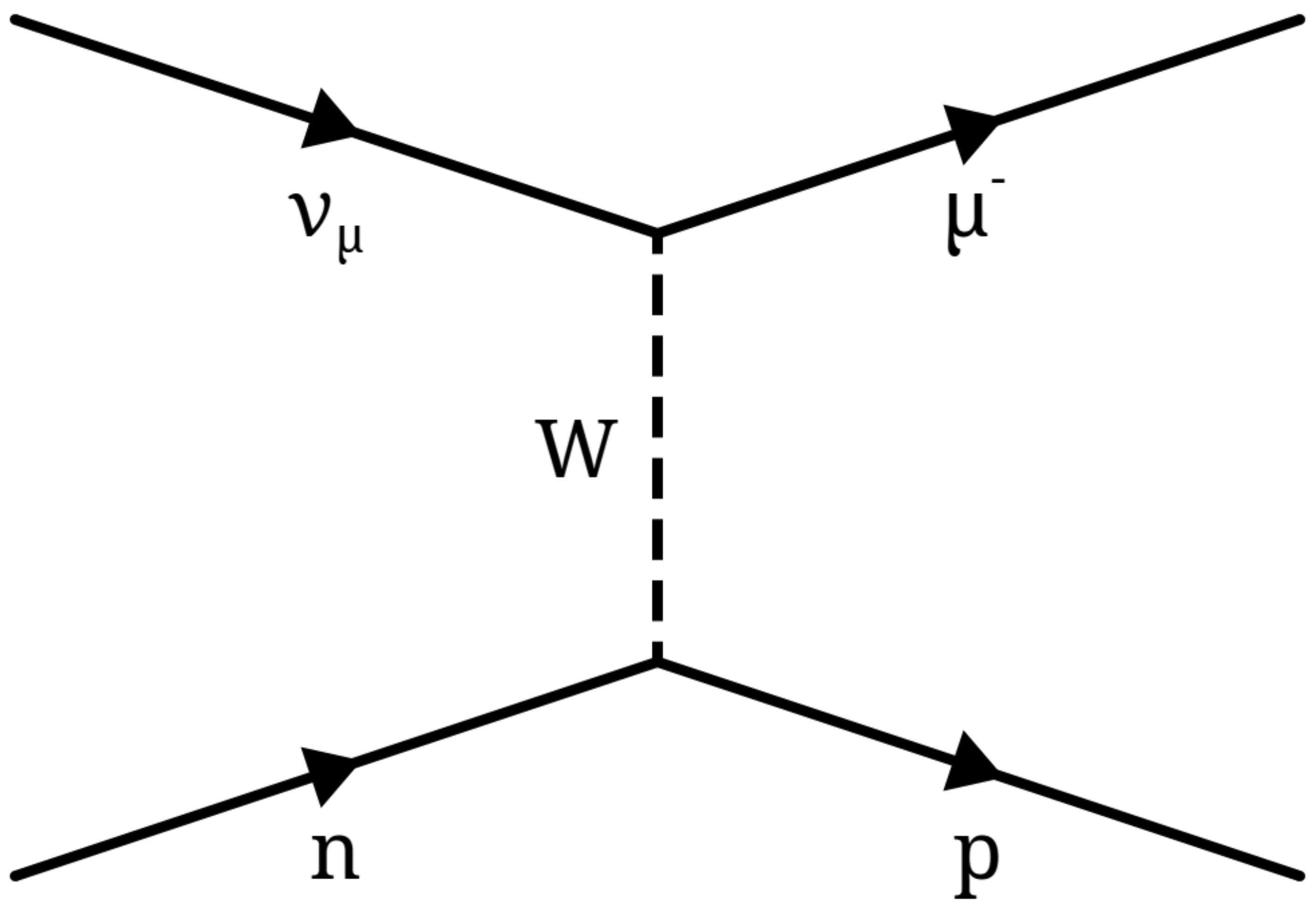
Electron (anti)neutrino appearance will then be measured at the HK far detector 295 km away and 2.5° off-axis.

Comparing appearance between neutrinos and antineutrinos probes CP violation in the lepton sector, where HK's large statistics will provide the best sensitivity yet to δ_{CP}.



$$A_{CP} = \frac{P(\nu_\mu \to \nu_e) - P(\bar{\nu}_\mu \to \bar{\nu}_e)}{P(\nu_\mu \to \nu_e) + P(\bar{\nu}_\mu \to \bar{\nu}_e)}$$

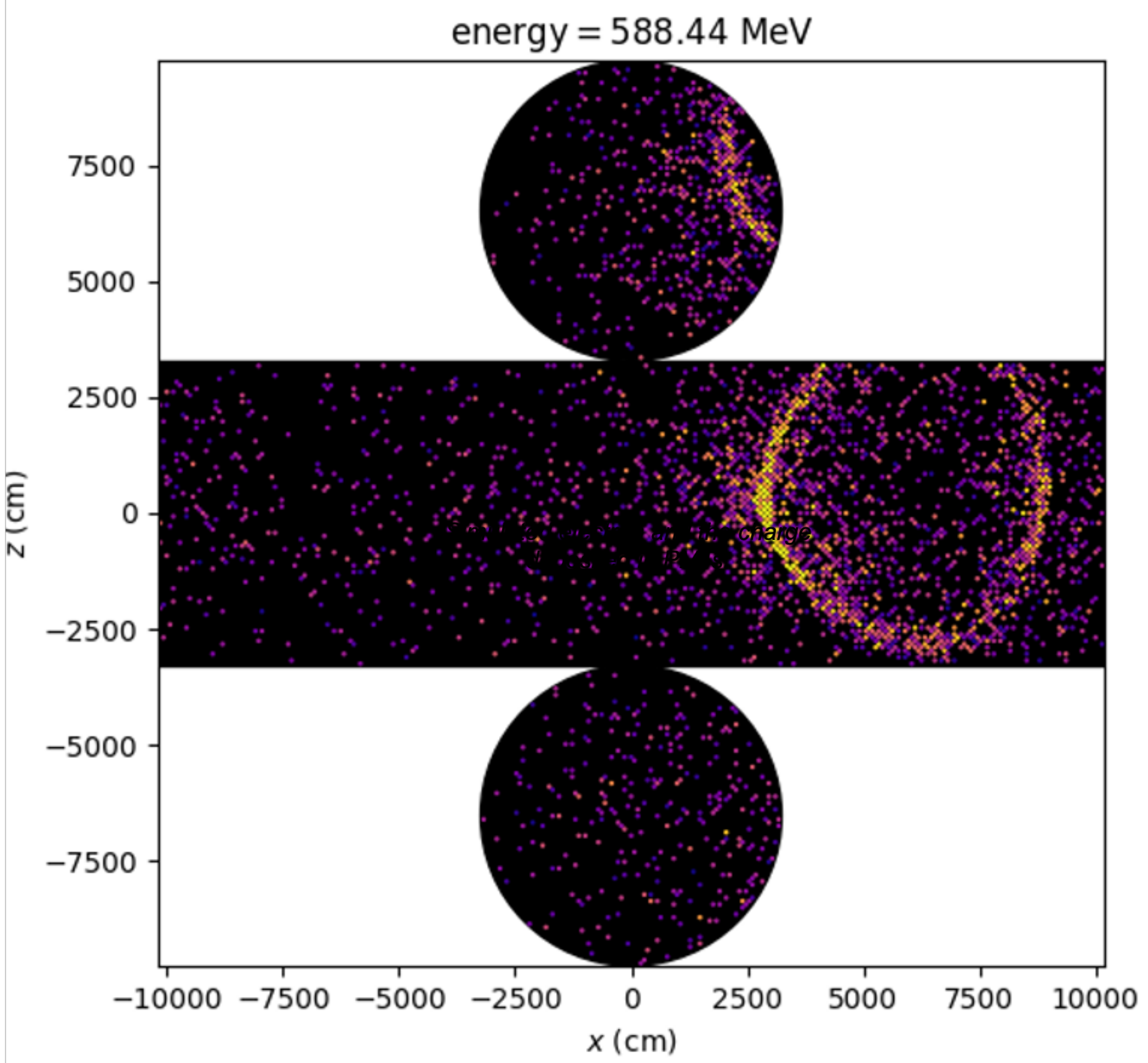
Detecting Neutrinos



Charged Current Quasi-Elastic Interaction

The charged lepton is relativistic and produces Cherenkov light in a distinctive ring pattern.

At higher energies (> 1 GeV) this gets more complicated and we have a mess of multiple rings.



Simulated electron and the charge deposited in PMTs.

Single Ring Reconstruction

Basics

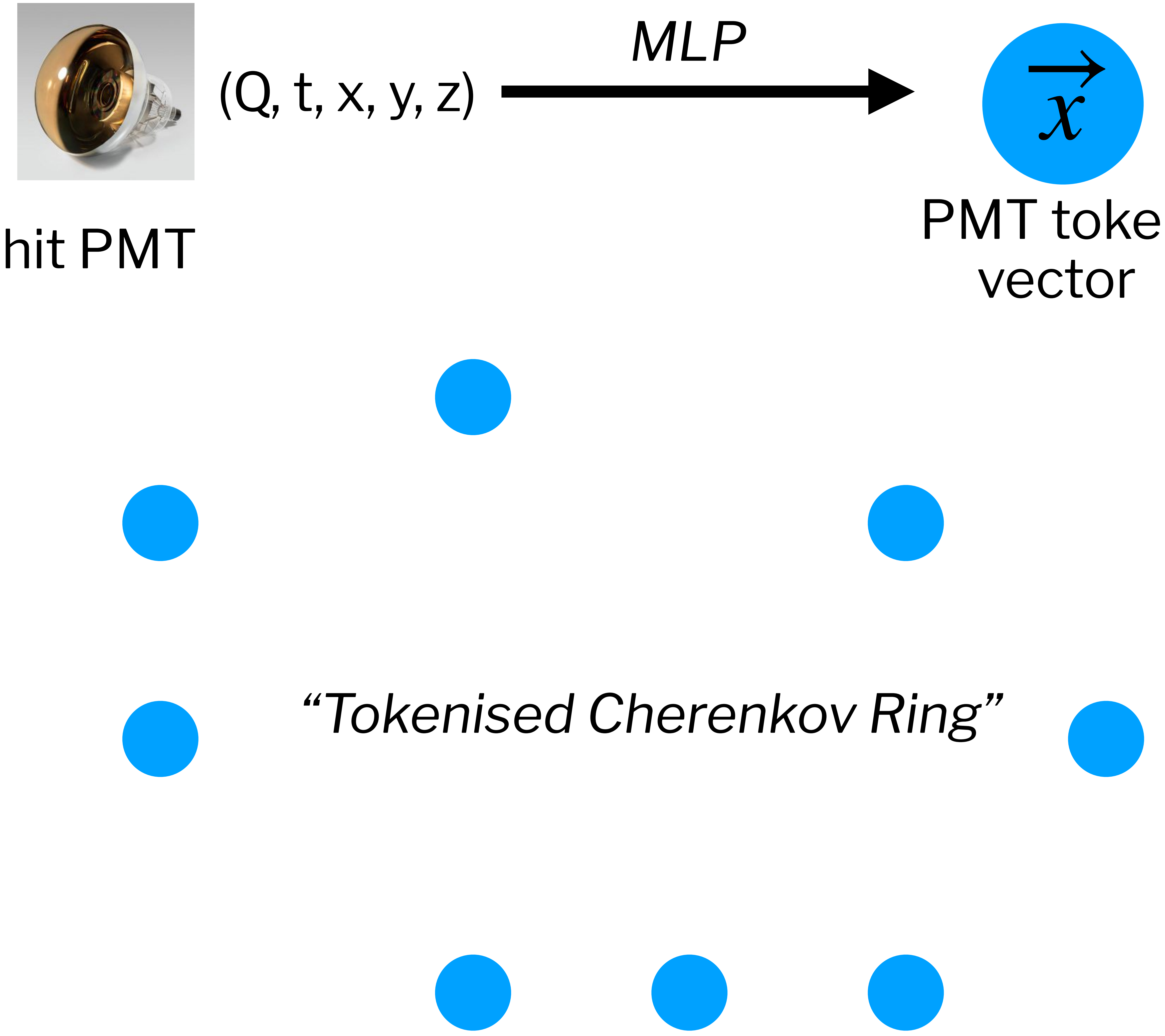
- **Particle Identification –** *Muons have a 'cleaner' ring than electrons.*
- **Energy –** *Based on total charge deposited in all hit PMTs from a given particle.*
- **Vertex –** *Optimise time residual across all hits.*
- **Direction –** *Charge weighted average across hit positions.*

The current reconstruction software, fiTQun, uses maximum likelihood fitting of each PMT's hit probability via a PDF. This process takes 10–60 s per ring.

$$\mathcal{L}(\mathbf{x}) = \prod_j^{\text{unhit}} P_j(\text{unhit}|\mu_j) \prod_i^{\text{hit}} \{1 - P_i(\text{unhit}|\mu_i)\} f_q(q_i|\mu_i) f_t(t_i|\mathbf{x})$$

While manageable for live data, this is prohibitively slow for large MC datasets. A machine learning based approach offers much faster reconstruction, with potential gains in precision as a bonus.

Transformer Multi-Head Attention



$$Q_i = x_i W_Q, \; K_i = x_i W_K, \quad (1)$$

$$V_i = x_i W_V$$

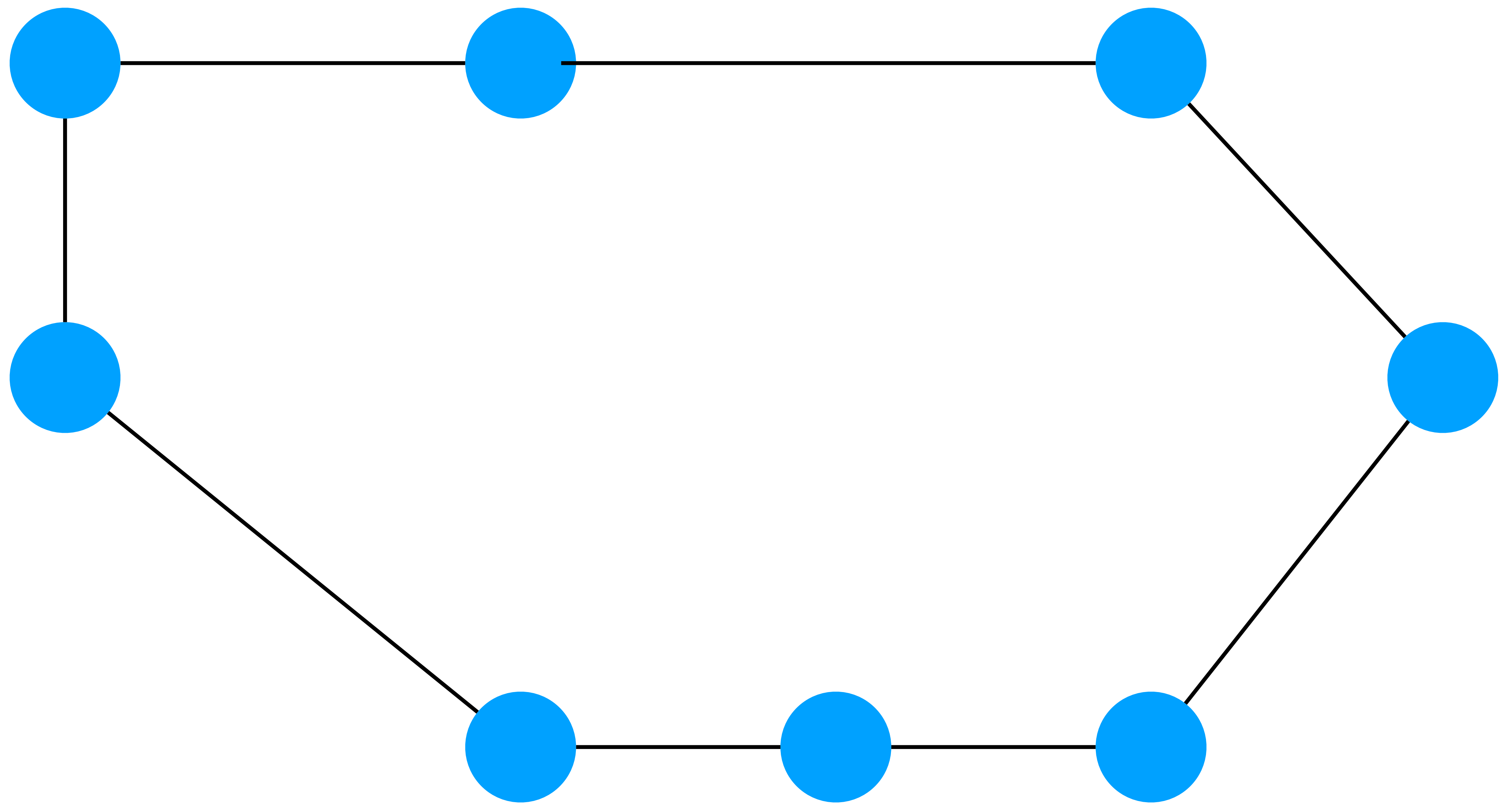
$$score_{i,j} = \frac{Q_i \cdot K_j}{\sqrt{d_k}} \; (2)$$

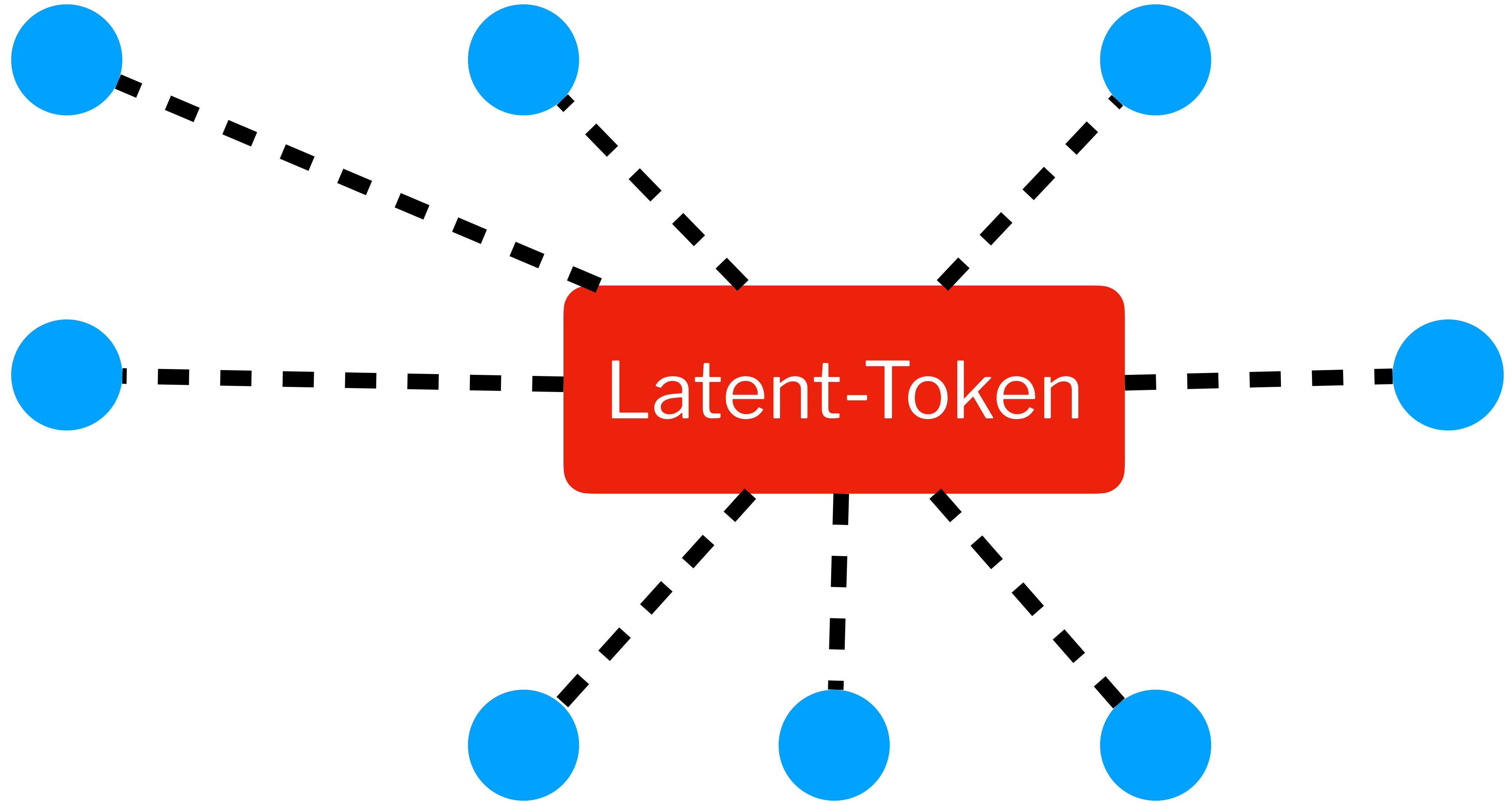
$$\alpha_{ij} = softmax(score_{i,j}) \; (3)$$

$$output_i = \sum_j^N \alpha_{ij} V_j \; (4)$$

$$x_i^{(\ell)} = MLP(x_i | output_i) + x_i \; (5)$$

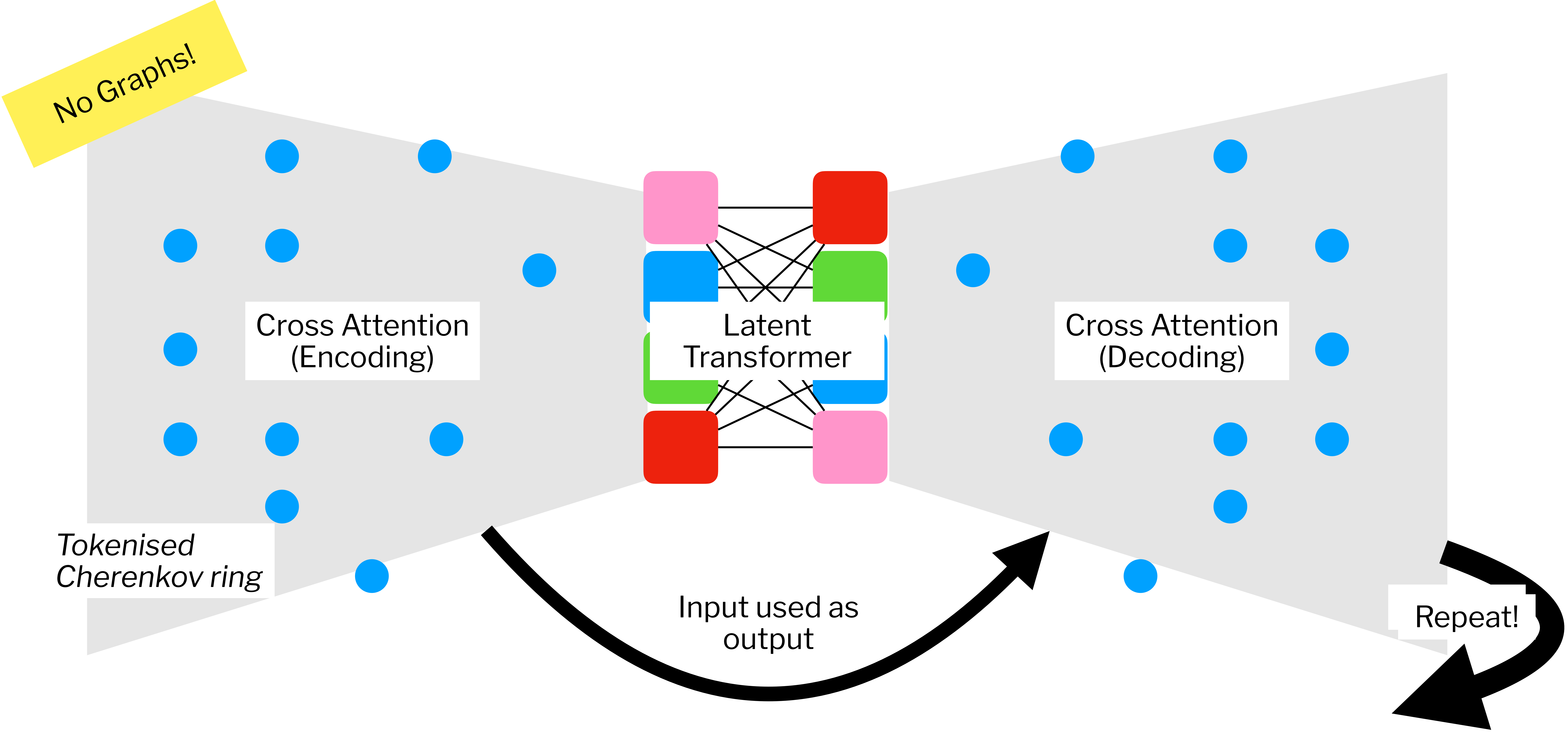
Linear Transformers





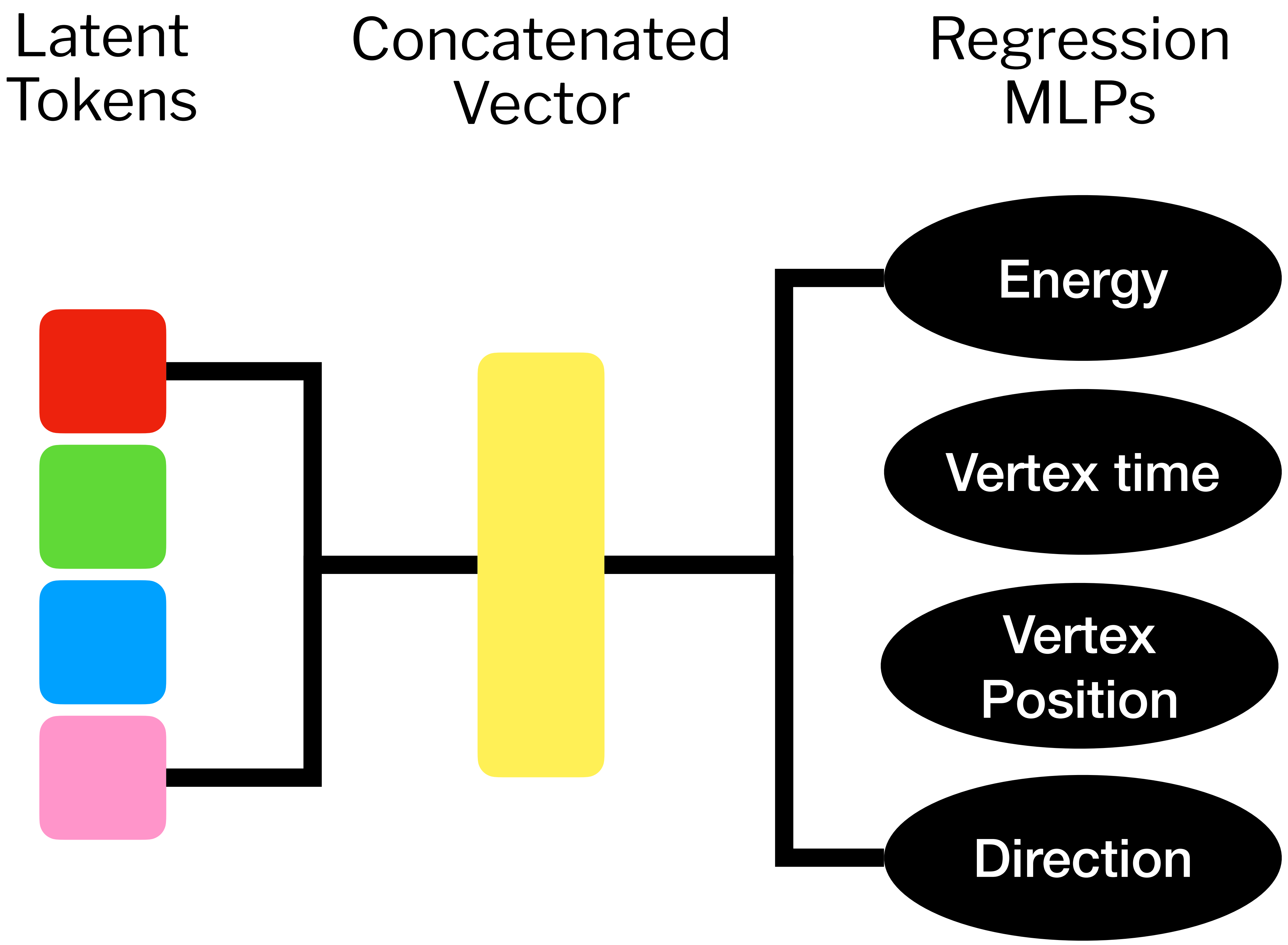
- Dense attention memory scales a as $O(n^2)$, this is infeasible for ~3000 hit PMTs per ring.
- Sparse attention $O(kn)$ is needed.
- Solutions: graph-based local connectivity, or cross-attention to a smaller set of latent tokens

CheRP - Cherenkov Ring Perceiver



* A. Jaegle, S. Borgeaud, J.B. Alayrac, et al., "Perceiver IO: A General Architecture for Structured Inputs & Outputs," arXiv:2107.14795 [cs.LG] (2021)

Multi-Task Regression



Multi-task loss function that learns a weighting of the different heads.

$$\sigma_i = e^{W_i}$$

$$\frac{L(\Delta E)}{\sigma_1} + \frac{L(\Delta t)}{\sigma_2} + \frac{L(\Delta x)}{\sigma_3} + \frac{L(\Delta \theta)}{\sigma_4} + \sum_{i=1}^{4} \log(1 + \sigma_i)$$

Performance - Kinematic Regression of Electrons

Model Parameters

- 72 token dimension
- 8 attention heads
- 6 tokens
- 4 perceiver layers

Half a million total parameters!

	Direction (°)	Vertex (cm)	Energy (%)
fiTQun (I used)	2.9	25.0	4.4
fiTQun*	2.0	20.0	4.0
CheRP	1.7	11.7	2.6

In development version with new scattered light table

Resolutions for 500 MeV electrons
(Where resolution is defined as the 68th percentile of the absolute difference between reco and true.)

Training Data

Using MC production macros with WCSim v1.12.29:

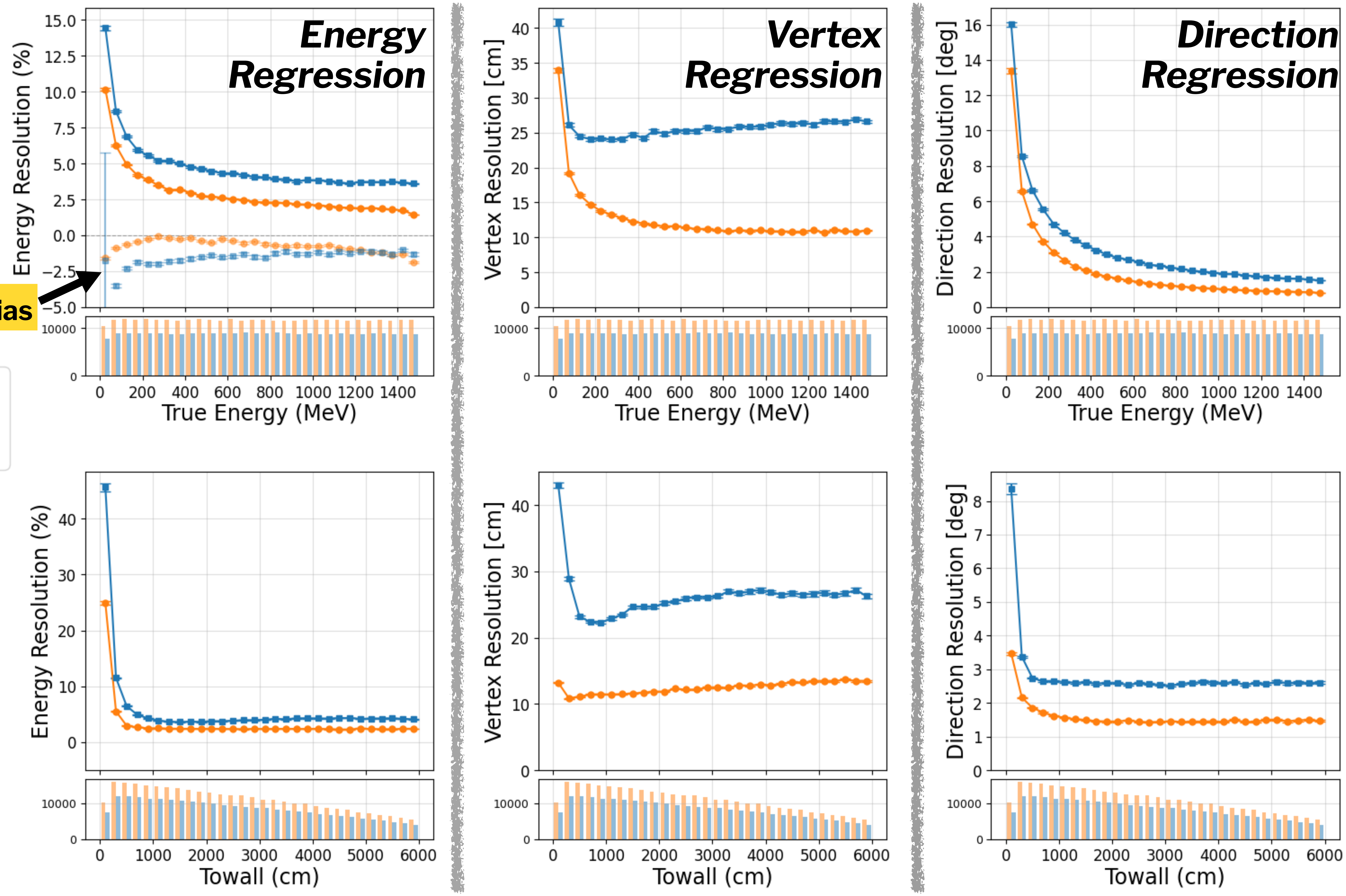
- 5 million electron events.
- Uniform energy 0 – 1.5 GeV.
- Isotropic in direction, uniform throughout the ID.
- Events not contained within the ID cut.
- Cutting events with zero hits.

**Trained using WatChMaL/CAVERNS a PyTorch-based software.*

Fitting an event takes ~ms!

True Energy

Energy of the simulated electrons.



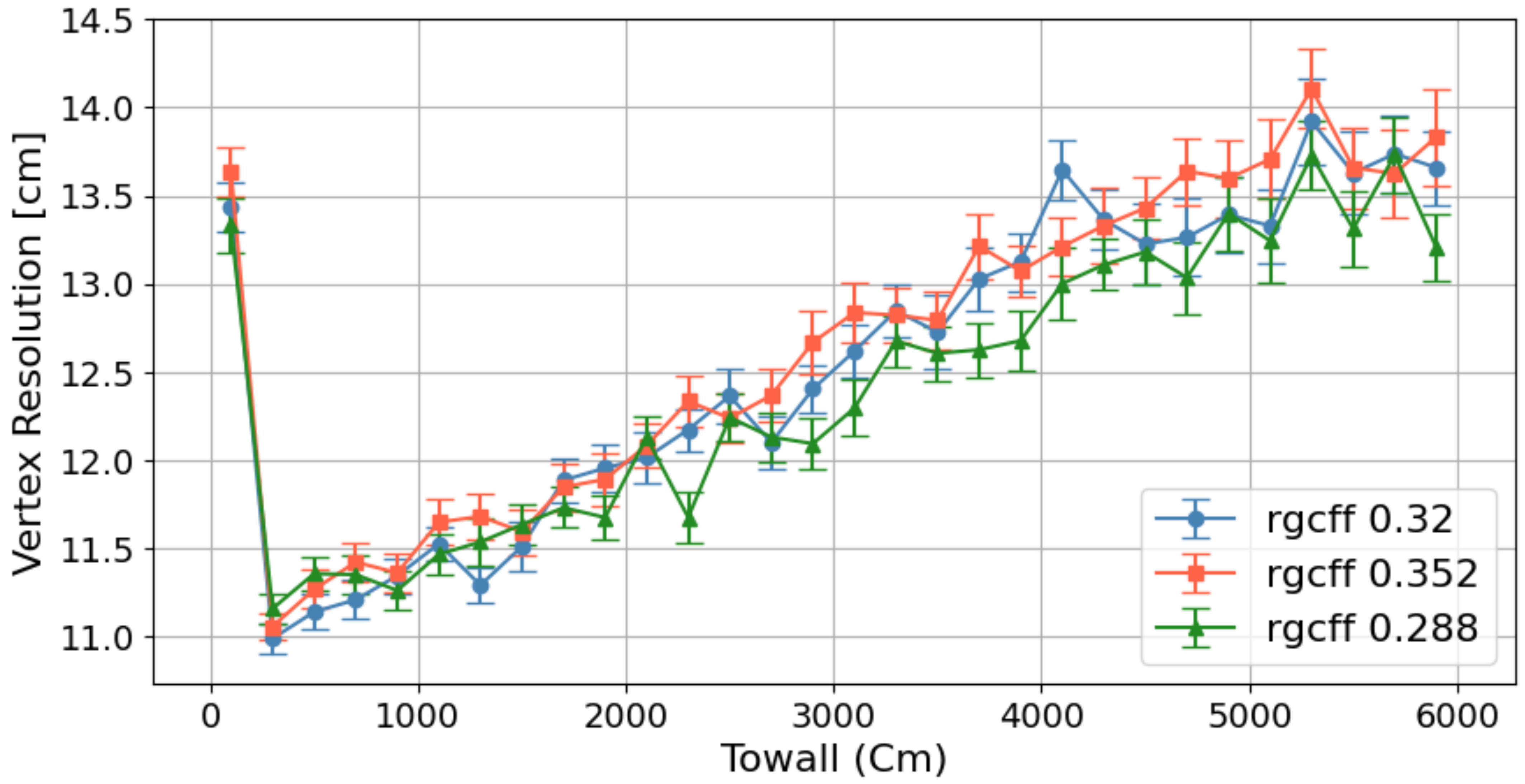
Towall

Distance from vertex to inner wall, along particle direction.

Is it Learning a Robust Fit?

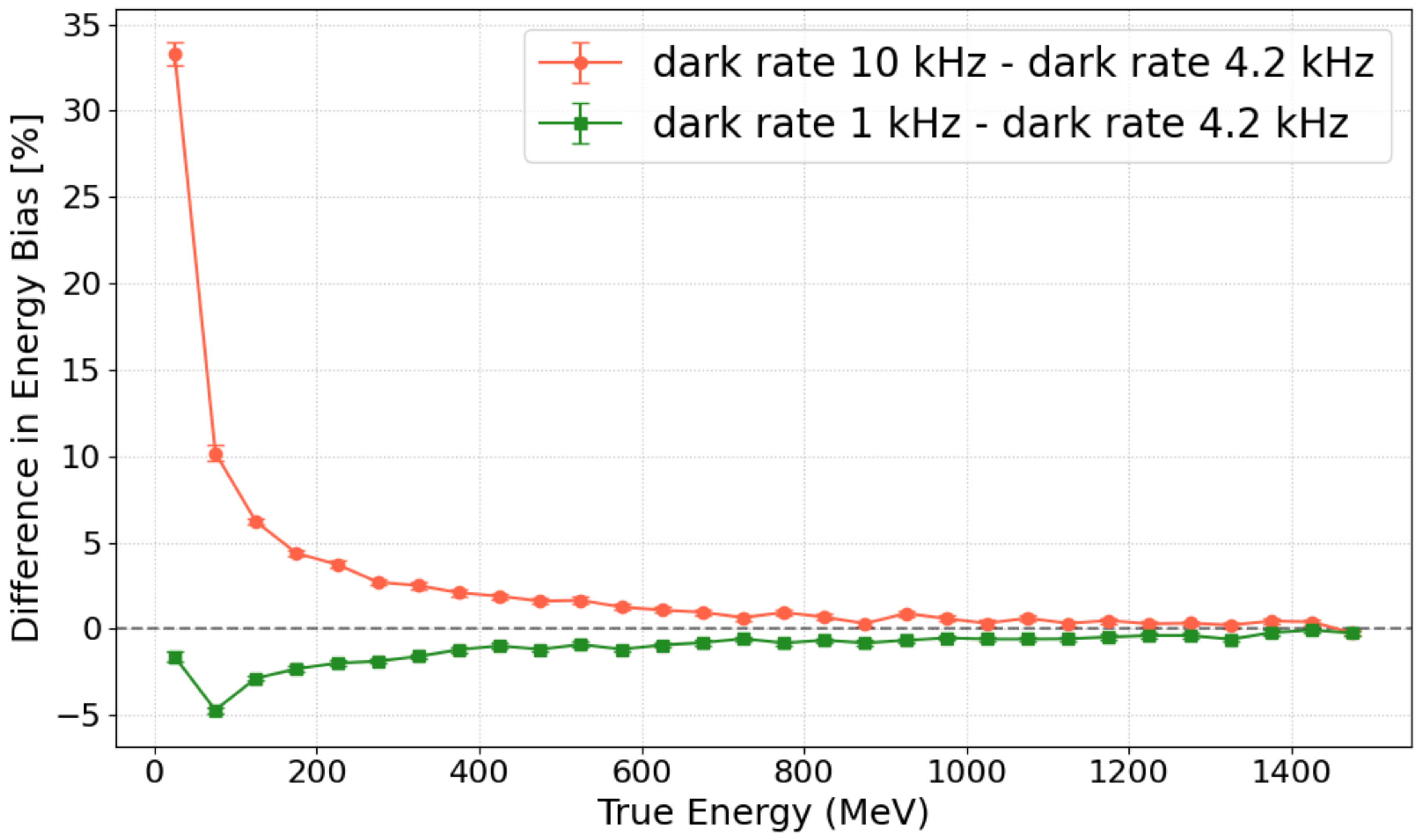
Training on fixed MC parameters and then testing across varied samples to probe model robustness.

Yes! For direction and vertex fitting, the model is very robust to adjusted MC parameters.



Varied cathode reflectivity has minimal impact on vertex reconstruction.

Less so… The model doesn't handle changes in dark rate well and this has a larger than desired impact on energy scale.



Difference in energy bias from reference for different dark rates.

Investigating training with varied dark rate next!

Summary/Next Steps

- Perceivers are linear transformers with powerful applications for physics, CheRP is my custom implementation built on previous development with graph transformers.
- CheRP out performs fiTQun in terms of resolution and computation time on electron kinematic regression.
- Work is beginning to prepare CheRP to be used for real data, investigating how it behaves and how to robustly train models.
- Will start looking at Particle identification and fit some muons.

Backup

Graph Transformer

An issue arose specifically around vertex and direction regression.

The common way to get a fixed representation for final regression via MLP was to use a pooling.

A classification (CLS) token is a non-physical token which connects to all the PMT tokens in the graph and acts an information “sponge”.

This CLS token replaces pooling and performs better at vertex and direction.