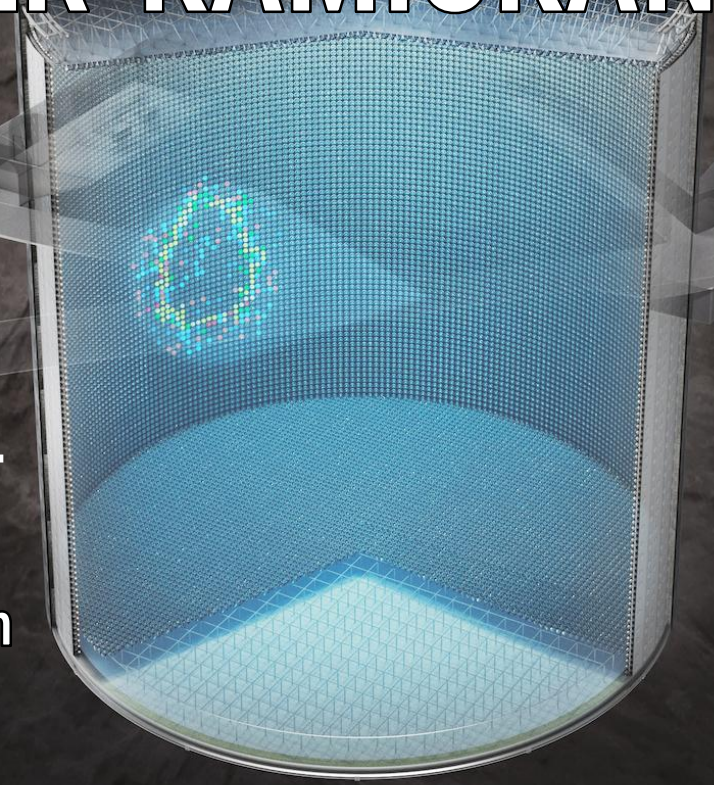
TOWARD A FAST AND PRECISE RECONSTRUCTION OF NEUTRINO EVENTS IN HYPER-KAMIOKANDE



LORENZO PERISSE

for Hyper-Kamiokande collaboration

NNN26, OCTOBER 29th 2026

NIKKEN SEKKEI

Outline

1. **Hyper-Kamiokande introduction**

2. **Likelihood algorithms in HK**
   a. Low energy algorithms: BONSAI and LEAF
   b. High energy algorithms: fiTQun and CORAL
   c. Strategies to improve performances

3. **Machine Learning development in HK**
   a. Convolution Neural Networks
   b. Graph Neural Networks
   c. Performance comparison with fiTQun

4. **Conclusion**

Outline

1. **Hyper-Kamiokande introduction**

2. **Likelihood algorithms in HK**
   a. Low energy algorithms: BONSAI and LEAF
   b. High energy algorithms: fiTQun and CORAL
   c. Strategies to improve performances

3. **Machine Learning development in HK**
   a. Convolution Neural Networks
   b. Graph Neural Networks
   c. Performance comparison with fiTQun

4. **Conclusion**

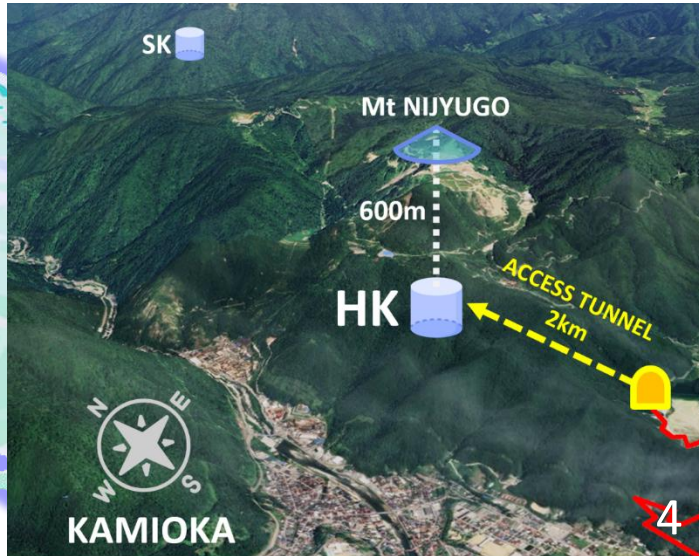


Broad & ambitious program covering many ν sources, proton decay, DM search, …



See Talk by **Christophe BRONNER** *Latest status and Prospect of HK*





260 kt (FV 188 kt ~8 x SK)

1600 m w.e. → 50Hz cosmic μ

- Shallower than SK (2700 m w.e., 2Hz cosmic μ)
- After OD cut, in ID: ~3.8 μ/day, ~80 atm. ν/day
- [1m from wall] cut → μ contamination x10^{-3}
- HK expects spallation per unit volume ~6-13x SK

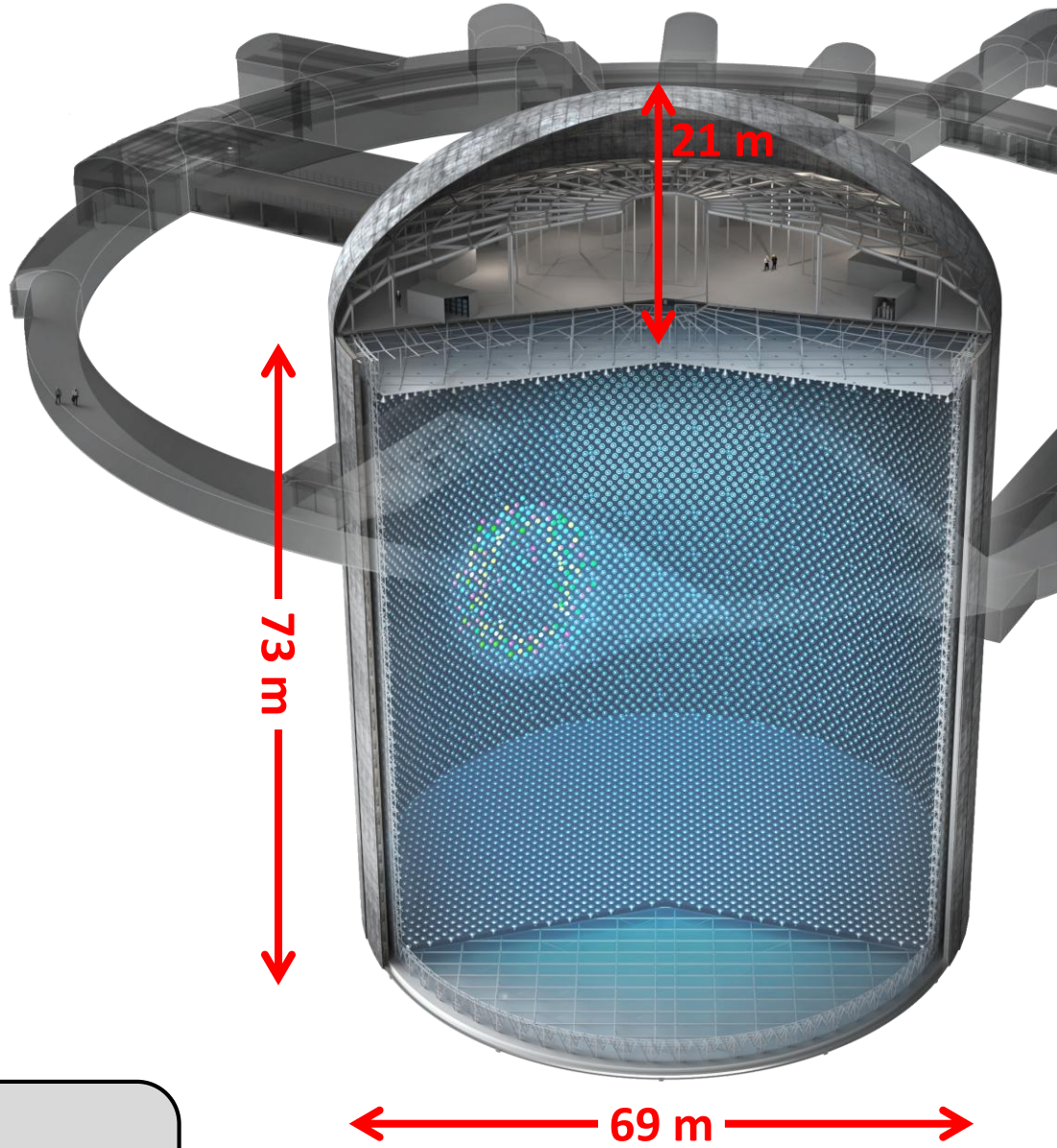
⇒ **Need improved vertex reco. to reduce μ & correlate with remaining spallation**

20 000 PMT 50cm + ~800 mPMT

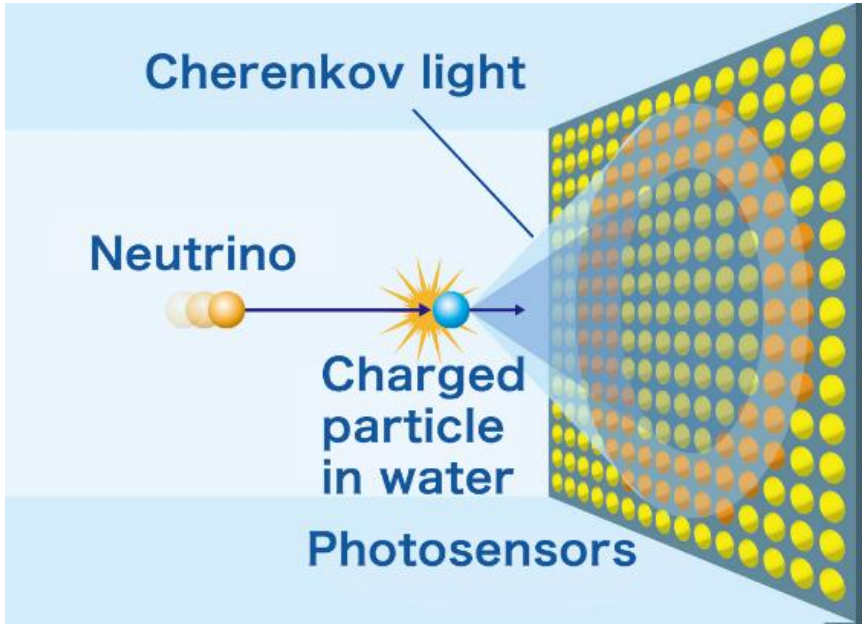


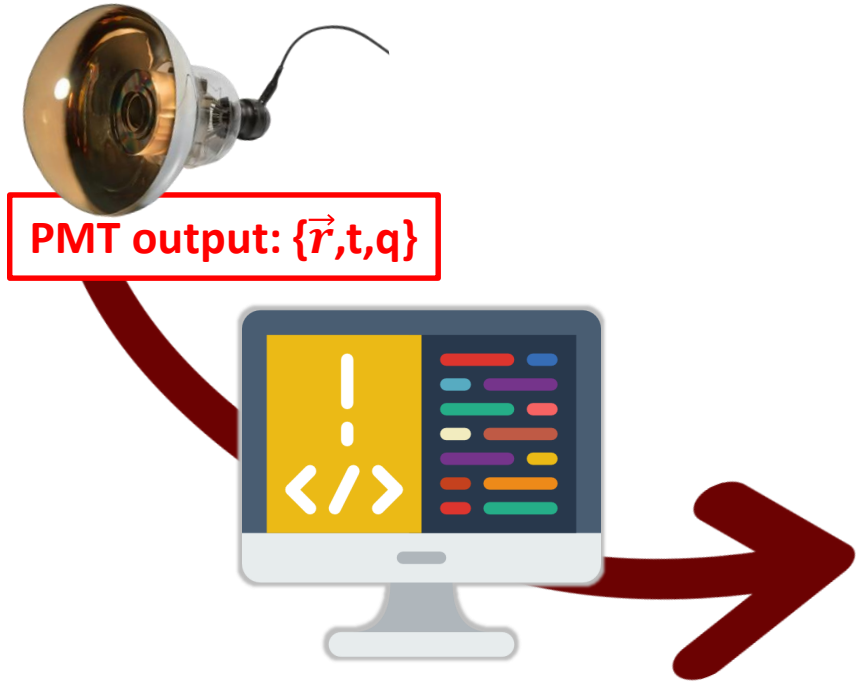
- LowE trigger rate ~10 kHz
- Algo. scale with # PMT & FV

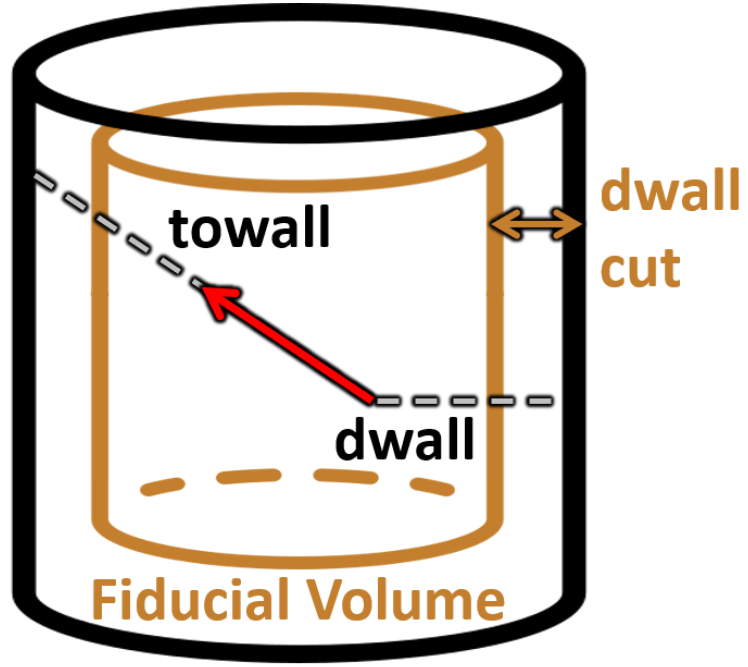
⇒ ↗ **algo performance for best CPU/GPU usage**



See Talk by **Robert KRALIK**
HK detector techologies for next gen ν physics







Variables of interest for *ν* physics

- **Vertex**
  - Limited perf. near wall
- **Direction**
  - Limited perf. low towall
- **Energy**
  - Sub/Multi GeV
- **# of rings**
  - 1R / MR
- **PID**
  - e / μ / π

Cuts (e.g. FV, angle, …)

Subsamples for Oscillation Analysis

FV cut	SK	T2K	HK
dwall	>50cm	>80cm	not yet defined
towall	sample dependant	>170cm	>200cm default

**Values above are provided for illustrative purpose, the final FV cut can vary depending on the considered analysis*

Outline

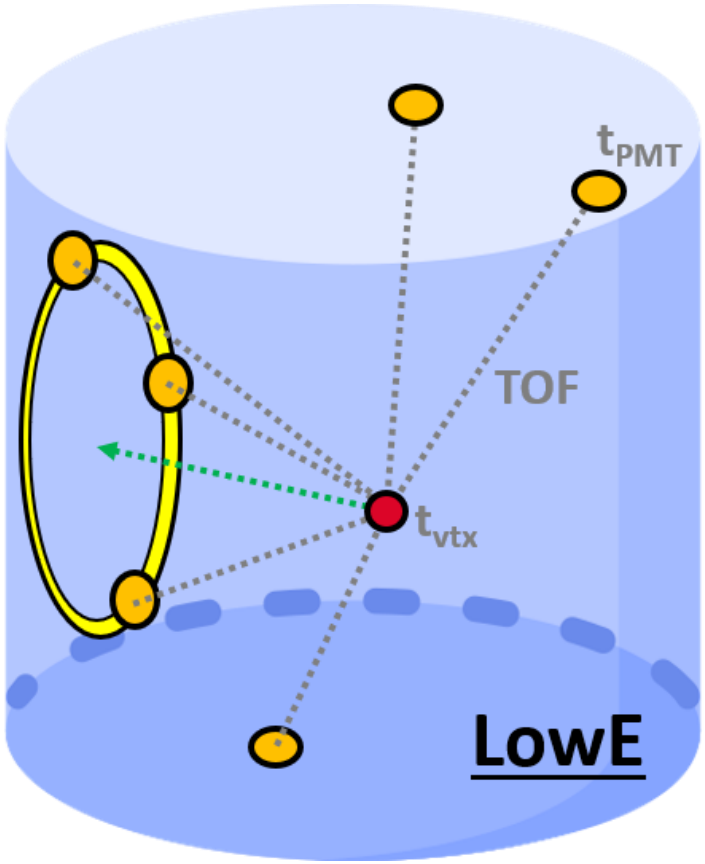
1. Hyper-Kamiokande introduction

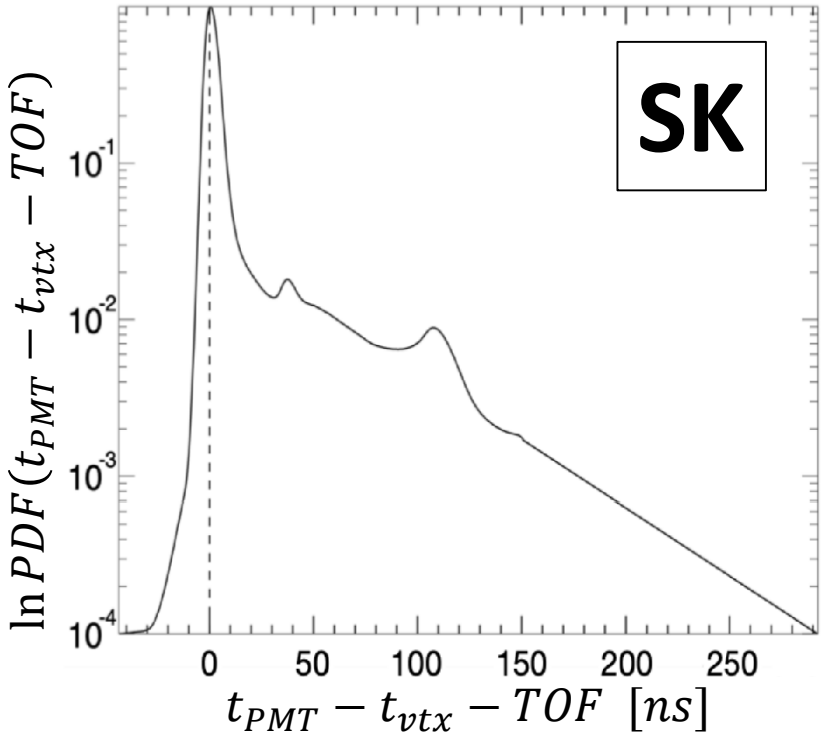
2. **Likelihood algorithms in HK**
   a. Low energy algorithms: BONSAI and LEAF
   b. High energy algorithms: fiTQun and CORAL
   c. Strategies to improve performances

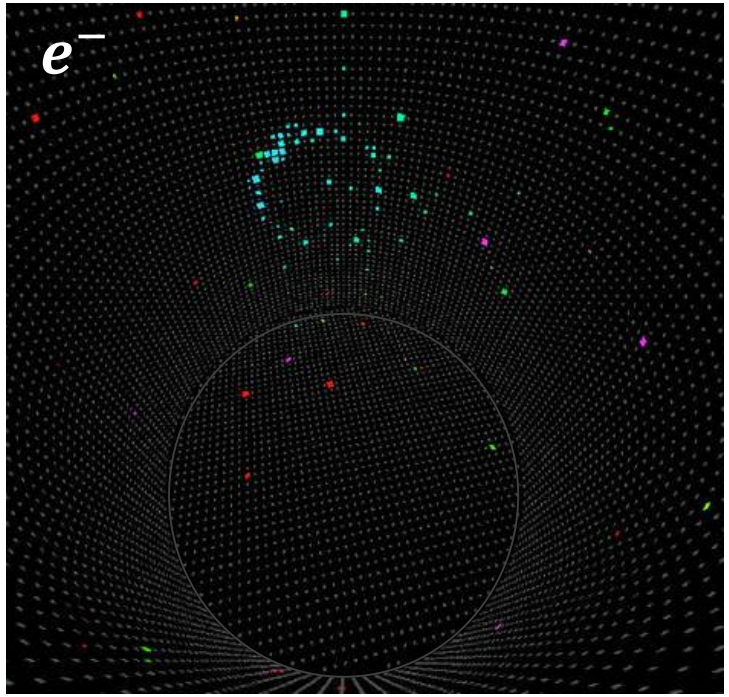
3. Machine Learning development in HK
   a. Convolution Neural Networks
   b. Graph Neural Networks
   c. Performance comparison with fiTQun

4. Conclusion

BONSAI / LEAF







Prefit vertex → $\mathcal{L}$(vtx) → $\mathcal{L}$(dir) → Energy

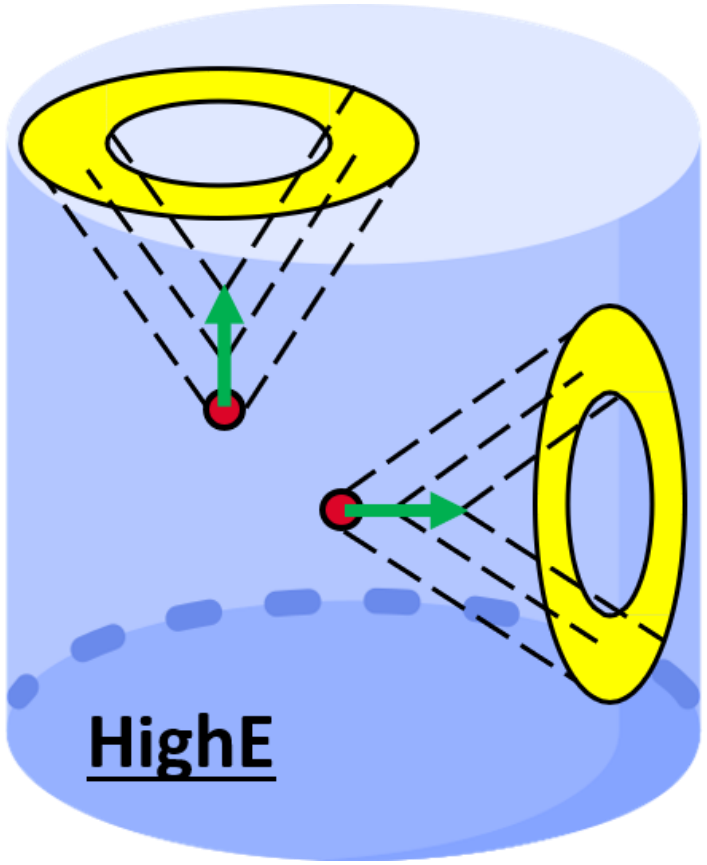
- Time residuals t_{PMT} - t_{vtx} – TOF
- $\mathcal{L}(\text{vtx}) = -\sum_{\text{PMT}} \ln PDF(t_{res})$
- Single particle topology

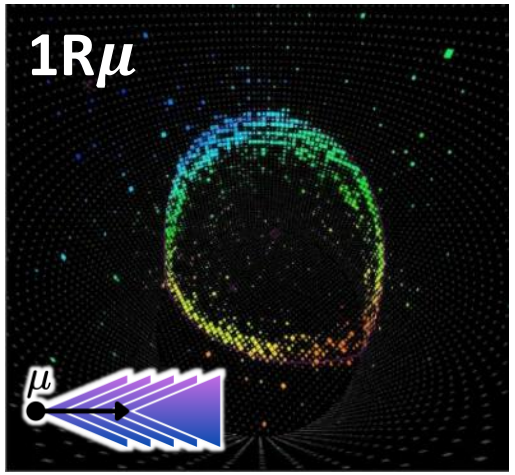
- Dark noise impact
  - Trigger rate limitation
  - $\mathcal{L}$ reco far from true

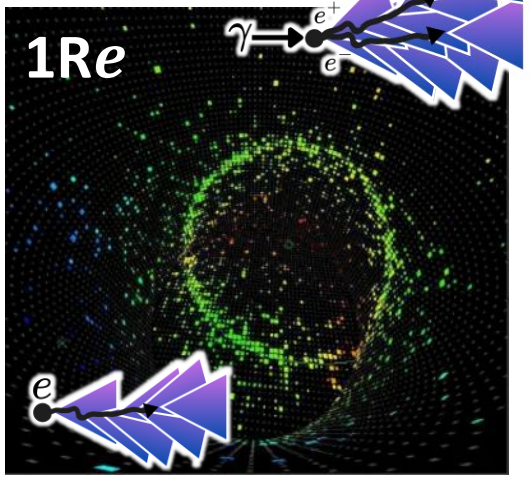
$\mathcal{O}$(1 s) per event

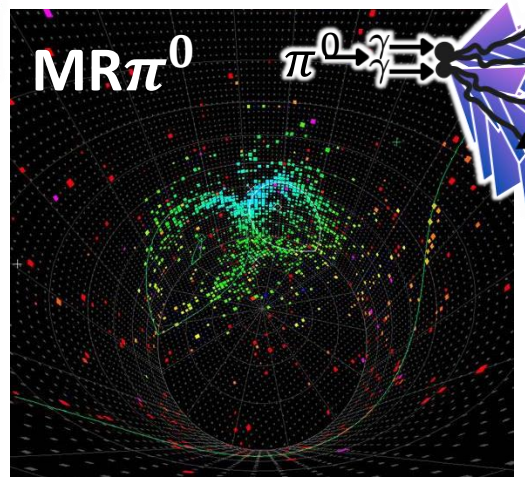
CPU-time

FITQUN / CORAL









Prefit vertex → Subeventing → 1R fits $\mathcal{L}$(x) → MR fits

- $\mathcal{L}(\mathrm{x}) = \prod^{\mathrm{unhit}} P(\mathrm{unhit}|\mathrm{x}) \prod^{\mathrm{hit}} P(hit|\mathrm{x})\, f_q(q|\mathrm{x})\, f_t(t|\mathrm{x})$
- Likelihood fit over x (7 par.) using all PMTs (hit q,t & unhit)
- $\mathcal{L}$(x) interpolates MC-based tuning tables at run time
- Sequential fit $e/\mu/\pi^+$ and incremental # of rings
- $\mathcal{L}$(x) ratios to determine PID & 1R/MR topology

$\mathcal{O}$(1 min) per event

CPU-time

SPEEDING-UP LIKELIHOOD ALGORITHMS

- BONSAI, fiTQun → excellent reconstruction in SK for 20+ years
  - fiTQun e^-/μ^- separation mis-ID < 1% & momentum resolution <4 % for e^-
- Nominal version ported to HK as reference C++ algorithms

⇒ Currently ready for day 1 analysis

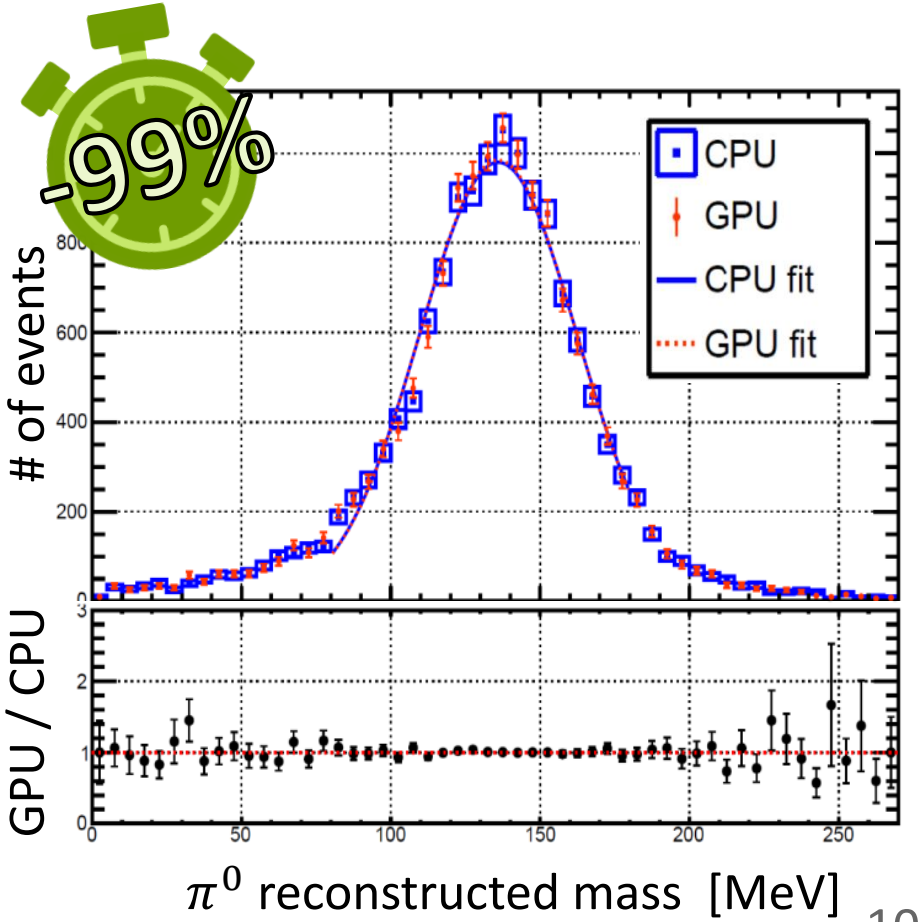


3 strategies to speed-up process time



fiTQun on GPU, by N. Calabria

- Unmodified physics performances
- 1R and MR fits succesfully ported
- GPU process time ≥ 100x CPU



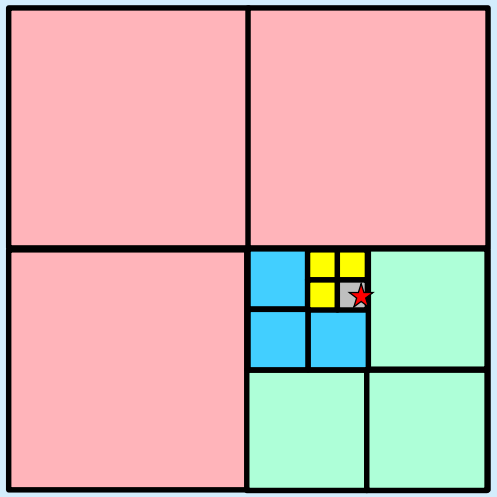
SPEEDING-UP LIKELIHOOD ALGORITHMS

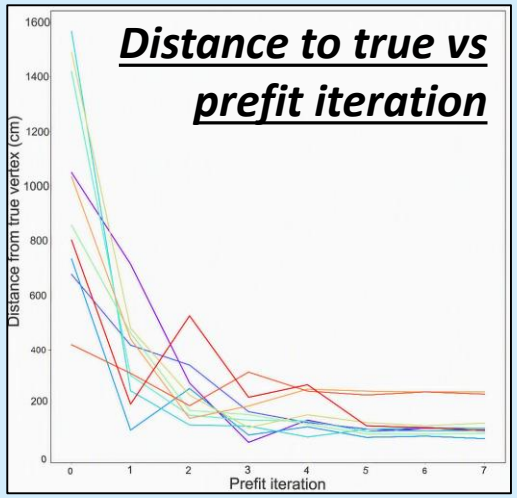
B Prefit search, by C. Sammons

fiTQun prefit

-94%

- Current: grid 6x6x6x40 = 8640 steps
  - Process time ~1s
- New: grid 4x4x4 then subgrid best quadrant, repeat 8 times = 512 steps in total
  - Prefit time cut by 94%, similar resolution



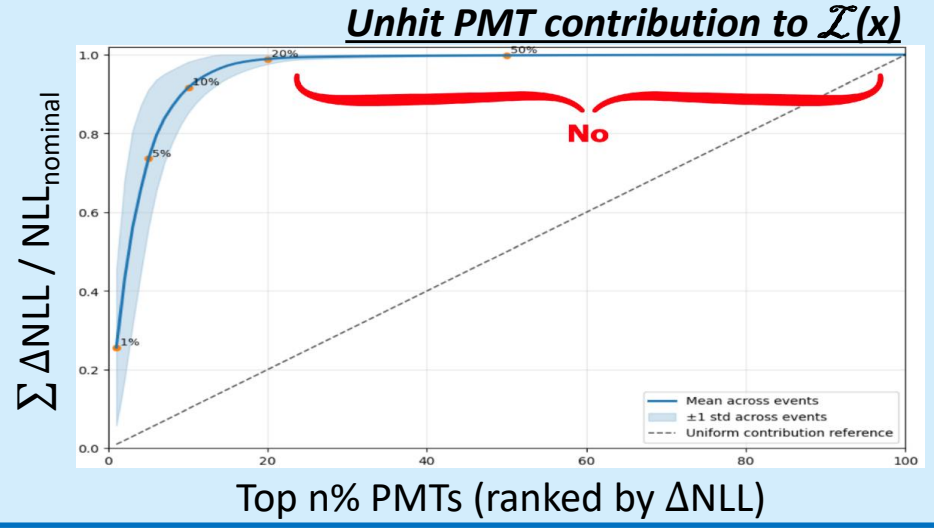


C Unhit PMTs, by B. Carew

fiTQun/CORAL

-87%

- CORAL = fiTQun ported to python
  - Just-In-Time compilation speed-up
- fiTQun's $\mathcal{L}$(x) uses all hit & unhit PMTs
  - Only 20% unhit contribute to 100% $\mathcal{L}$(x)
- Full fit time cut by 87%, similar resolution



Outline

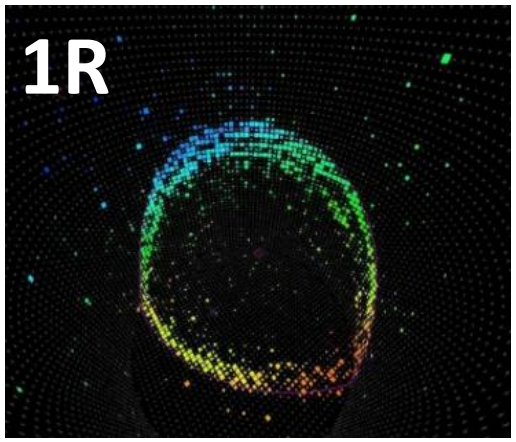
1. Hyper-Kamiokande introduction

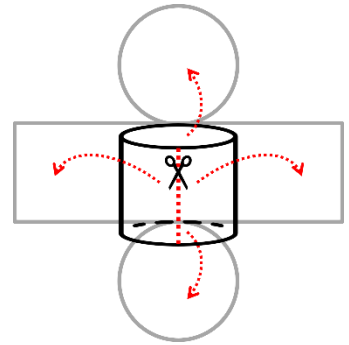
2. Likelihood algorithms in HK
   a. Low energy algorithms: BONSAI and LEAF
   b. High energy algorithms: fiTQun and CORAL
   c. Strategies to improve performances

3. **Machine Learning development in HK**
   a. Convolution Neural Networks
   b. Graph Neural Networks
   c. Performance comparison with fiTQun

4. Conclusion

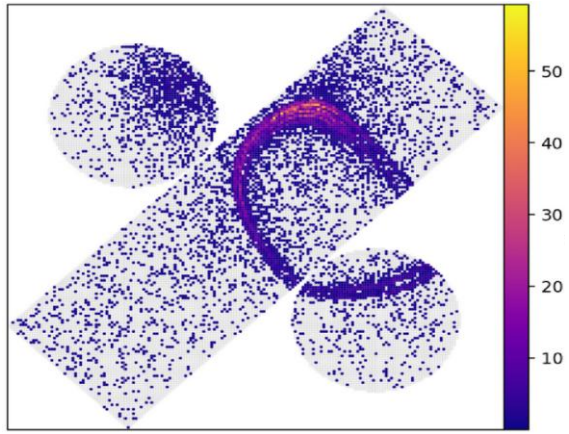
CONVOLUTIONAL NEURAL NETWORKS (CNN)





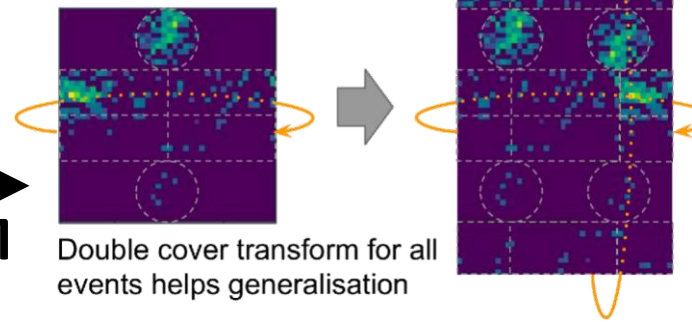
UNFOLD

PMT := 2 channel (q,t) pixel



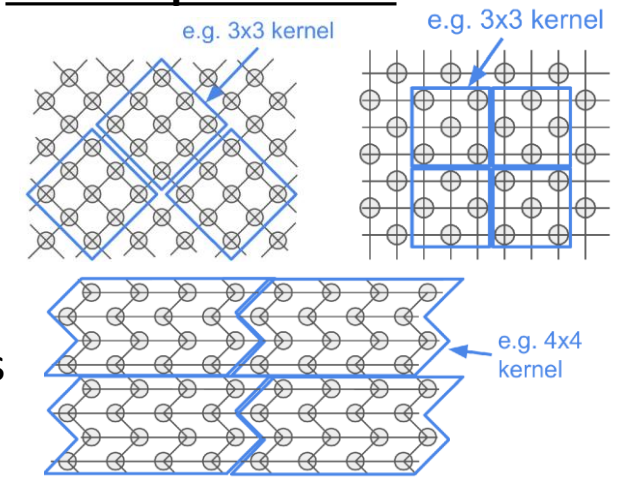
TRANSFORM

Boundary condition



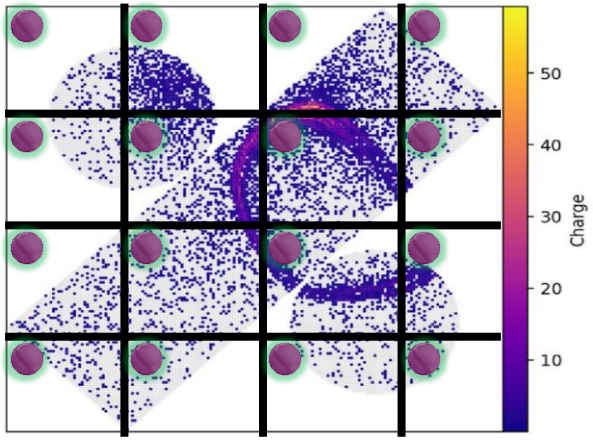
Double cover transform for all events helps generalisation

Kernel optimization



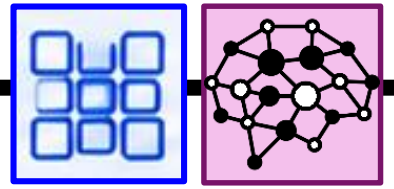
PATCHING

- Image → N patches
- 1 token 🟣 per patch
- Learnable position embedding on tokens
- Tokens → transformer encoder



Shuoyu Chen
ViT

CNN + Transformer



PREDICTIONS

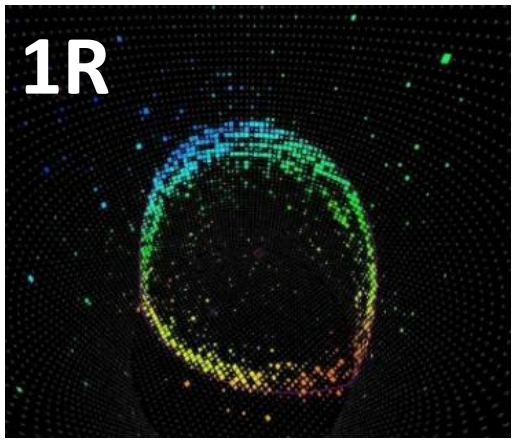
CNN



Andrew Atta
ResNet-1

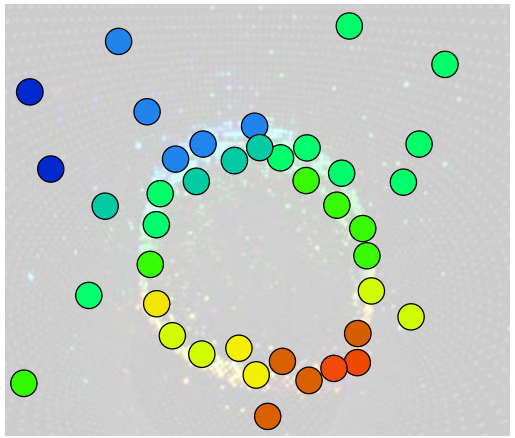
Nick Prouse
ResNet-2

GRAPH NEURAL NETWORKS (GNN)



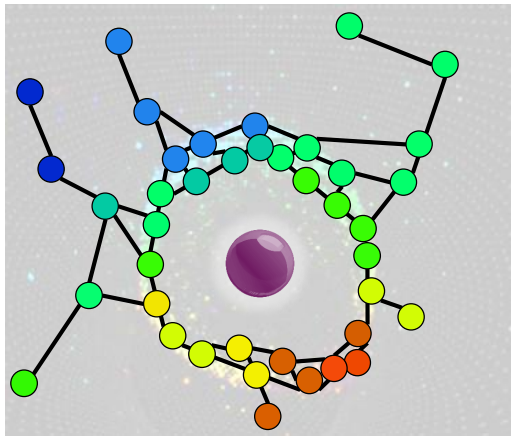
POINT CLOUD FOR FREE

PMT → 5D nodes ($\vec{r}$,q,t)

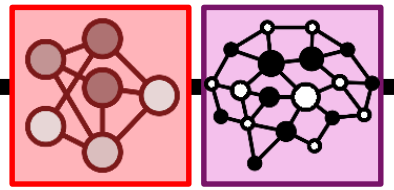


GENERATE EDGES

Node proximity ($\vec{r}$,t,kD projection,…)



GNN + Transformer

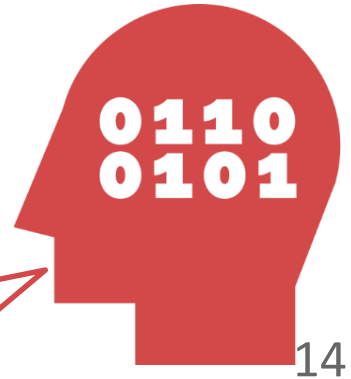


- N tokens per event to aggregate node info
- Tokens → transformer encoder

PREDICTIONS

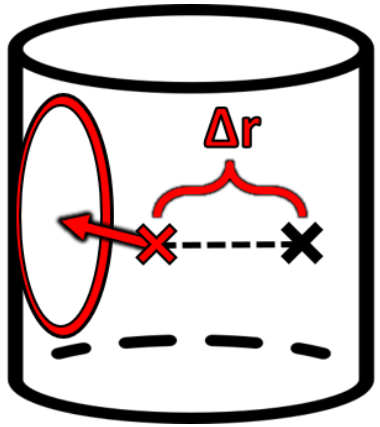
Naomi Foster	Lorenzo Perisse	Erwan Le Blevec
CheRP	**CLSGAT**	**GAT**

Machine learning based inference techniques greatly improved in last decades
1M events → fiTQun +10k CPU hours / **ML 100 CPU hours or 10 GPU minutes**

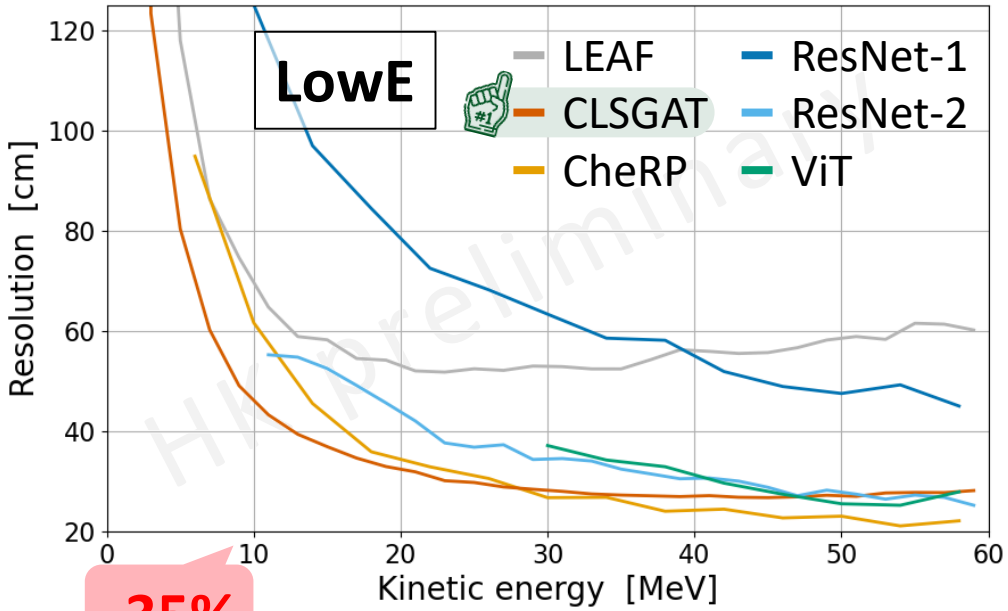


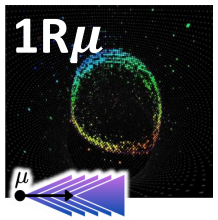
VERTEX RESOLUTION

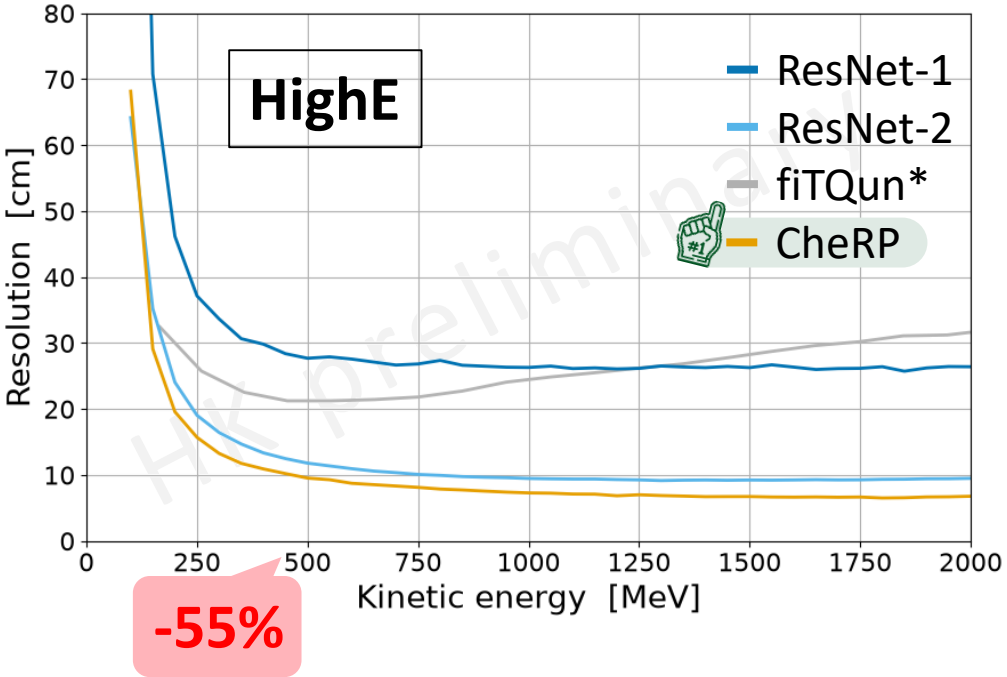
Δr CI 68%

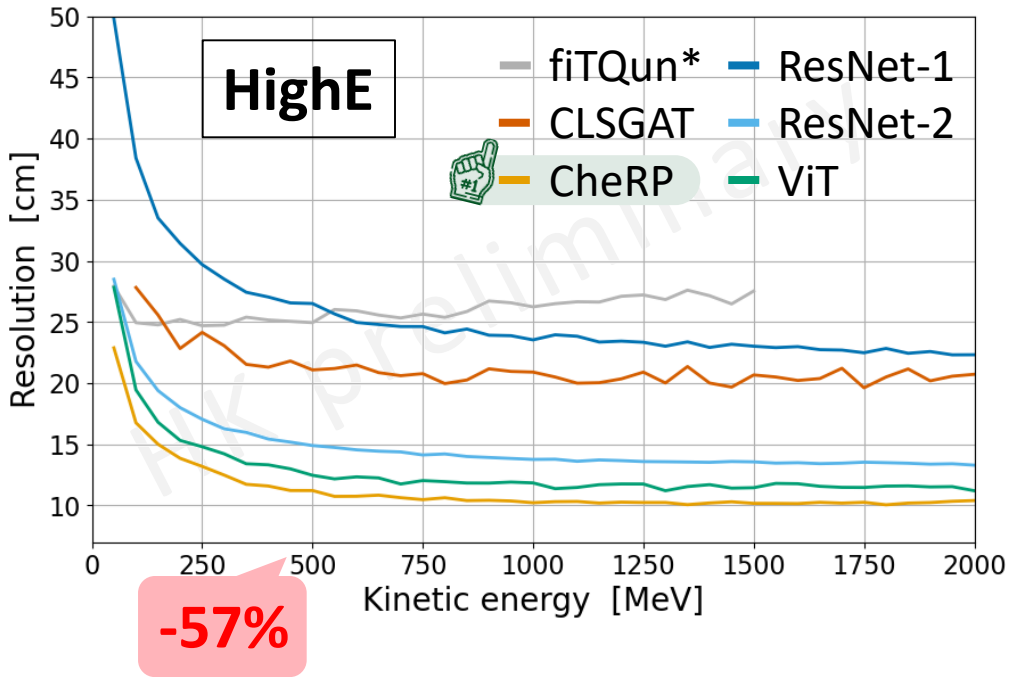


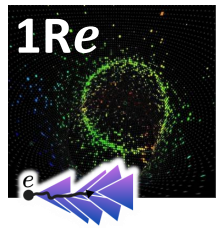
*fiTQun results presented here are preliminary.







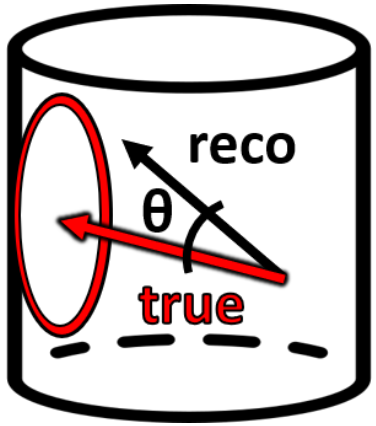


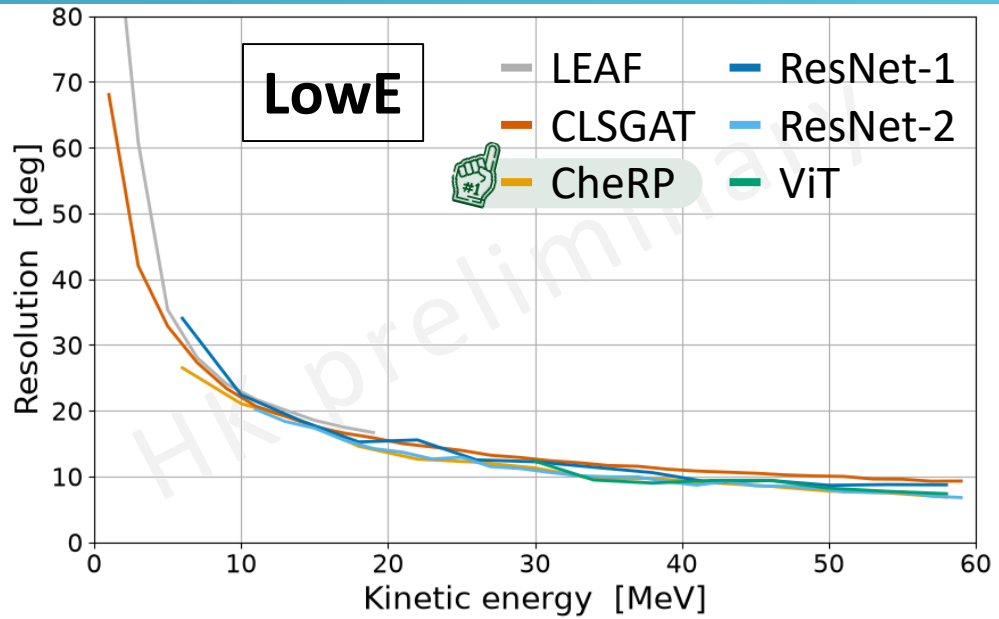


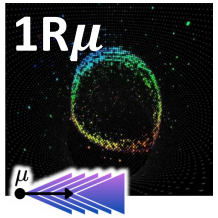
DIRECTION RESOLUTION

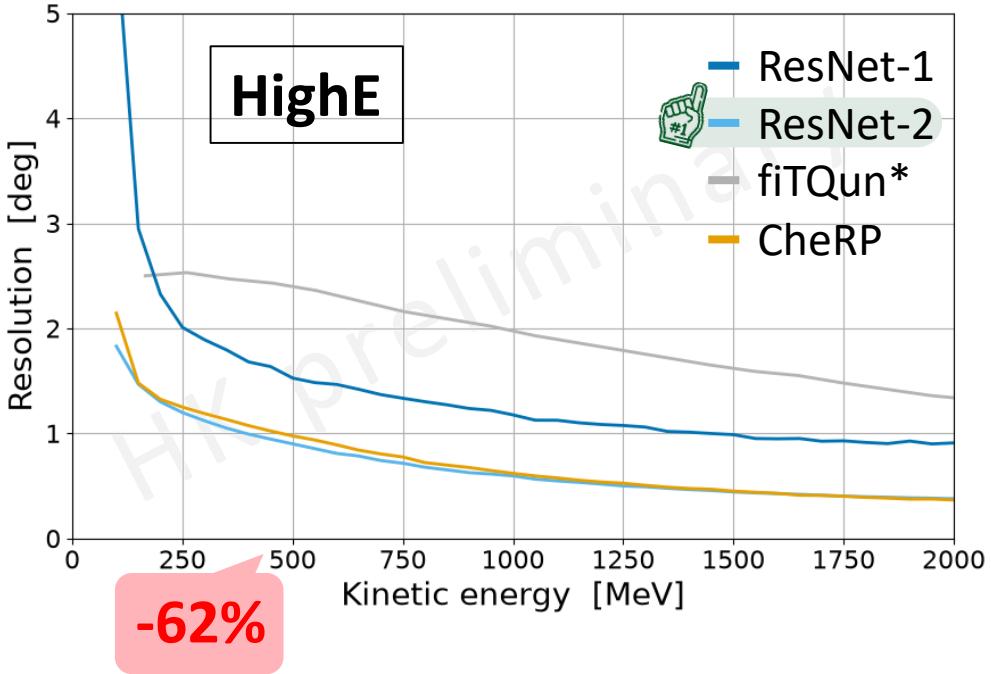
θ CI 68%

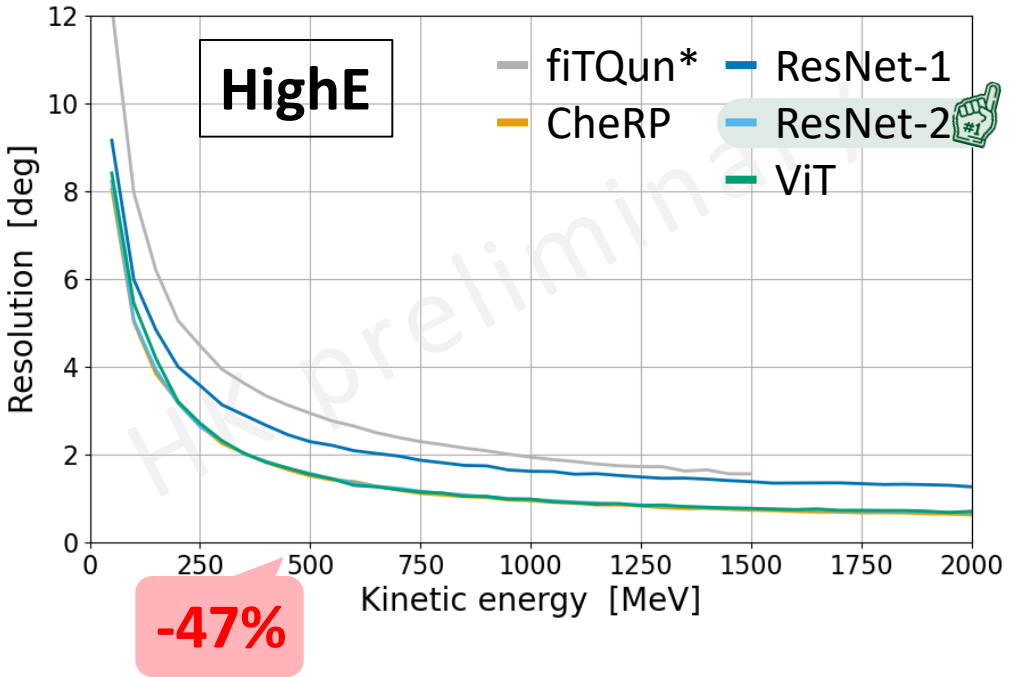
*fiTQun results presented here are preliminary.

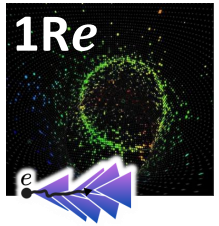












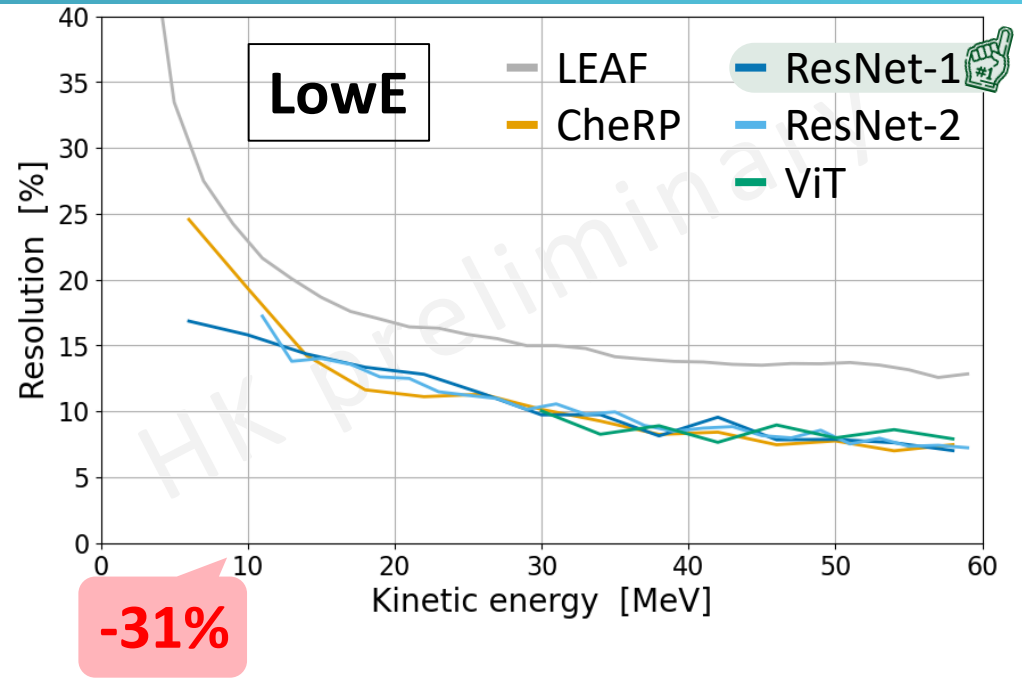
ENERGY RESOLUTION

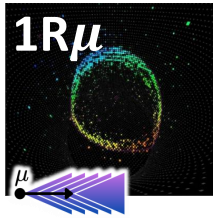
$|\Delta E|/E$ CI 68%

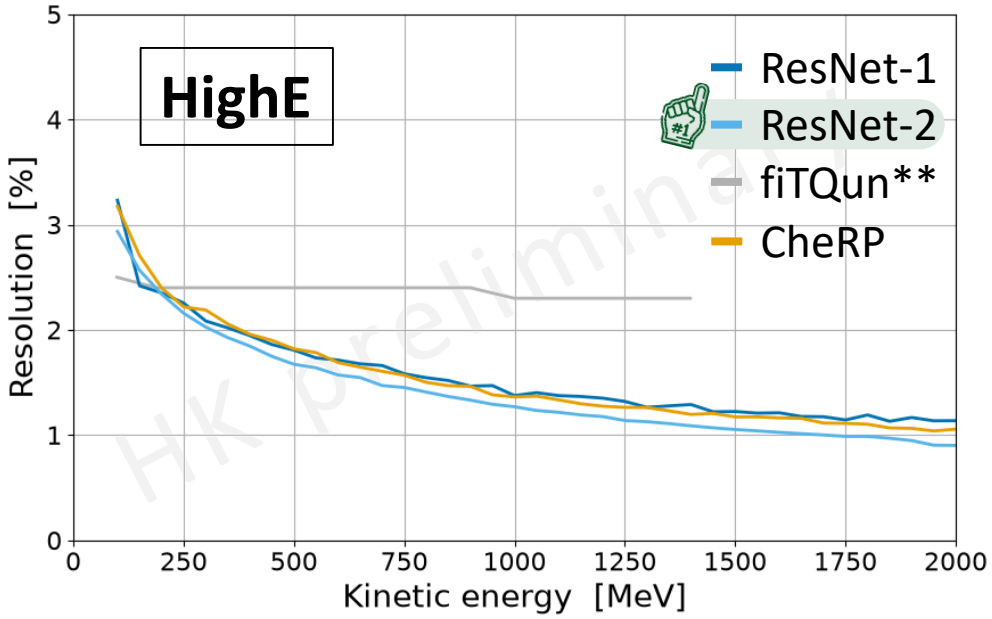


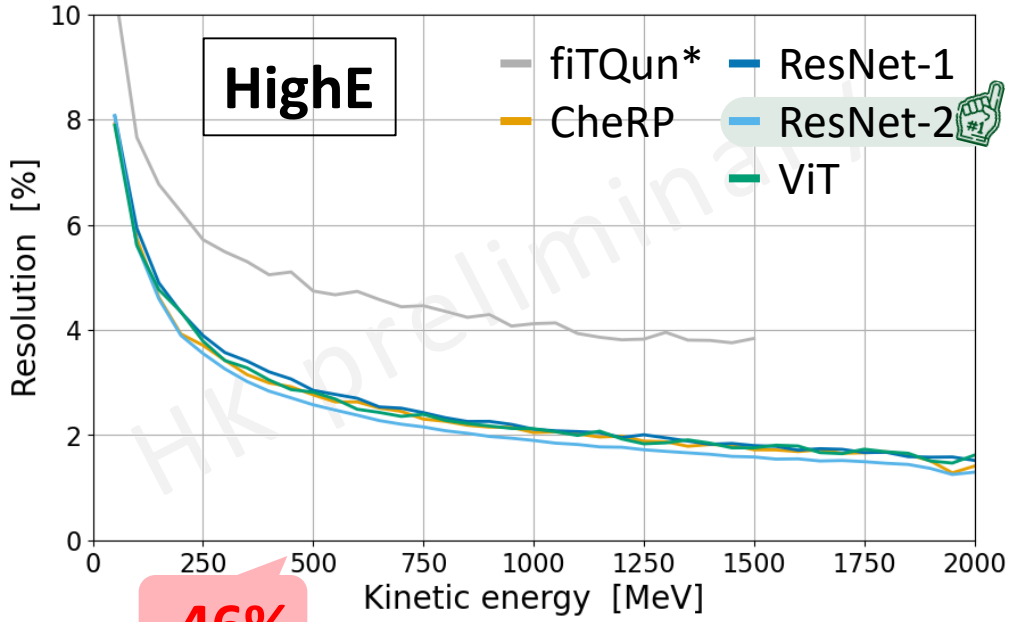
*fiTQun results presented here are preliminary.

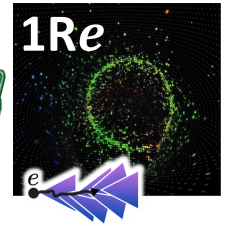
**fiTQun results for Super-K









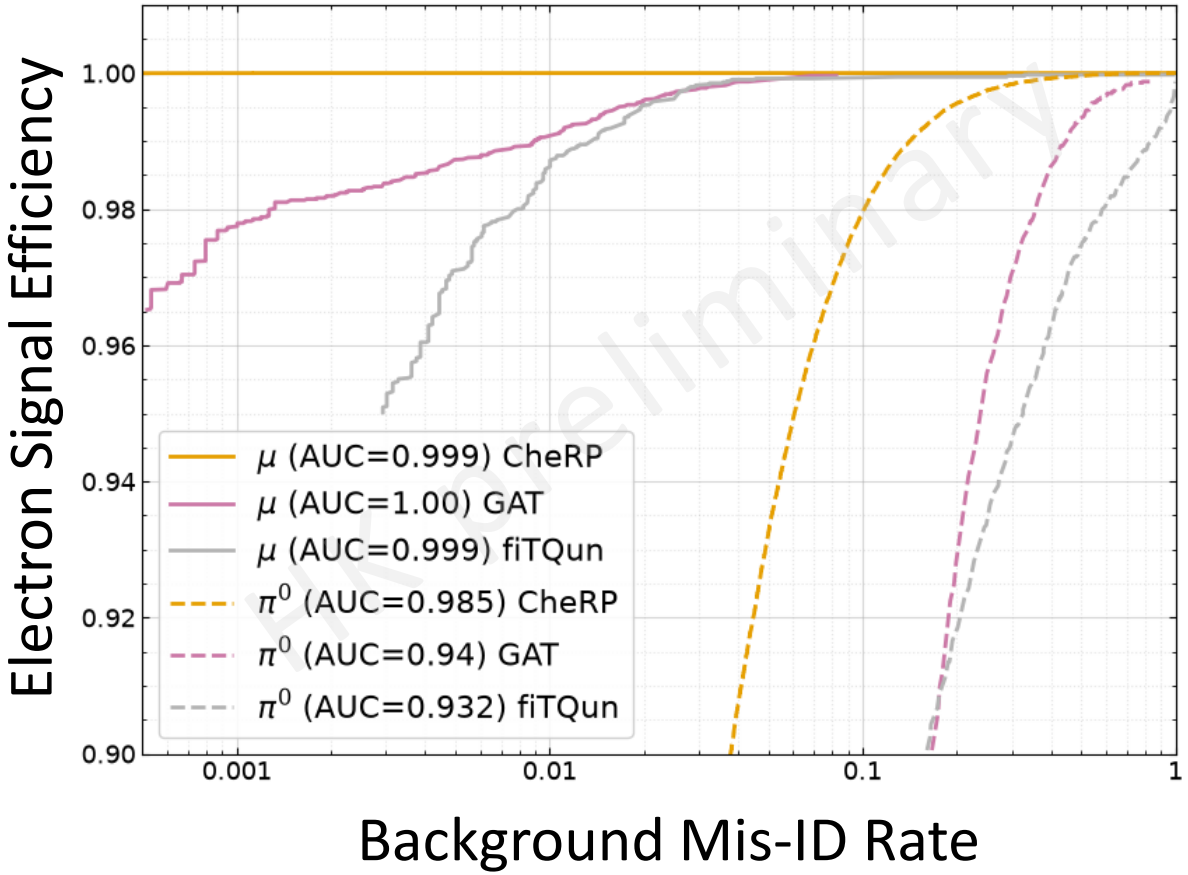


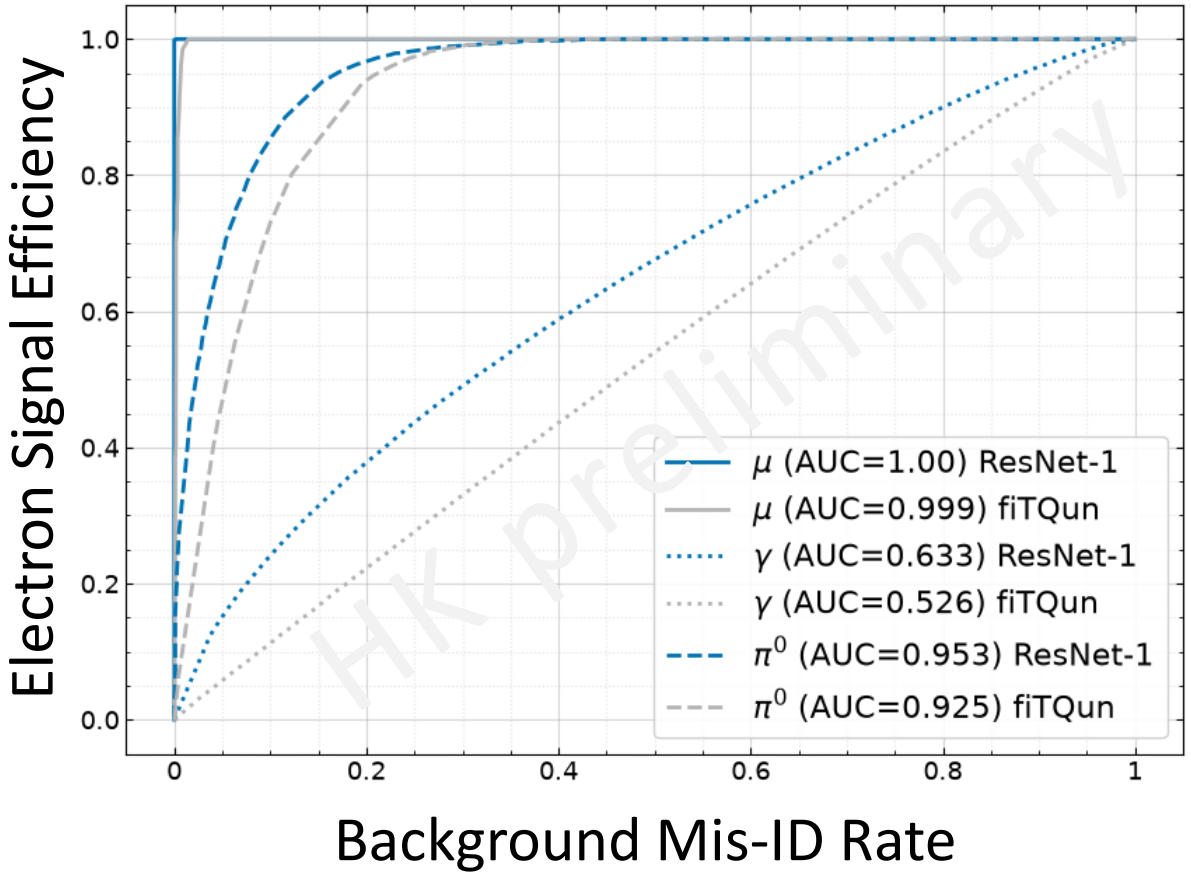
PARTICLE CLASSIFICATION

Erwan Le Blevec **GAT**

Naomi Foster **CheRP**

Andrew Atta **ResNet-1**



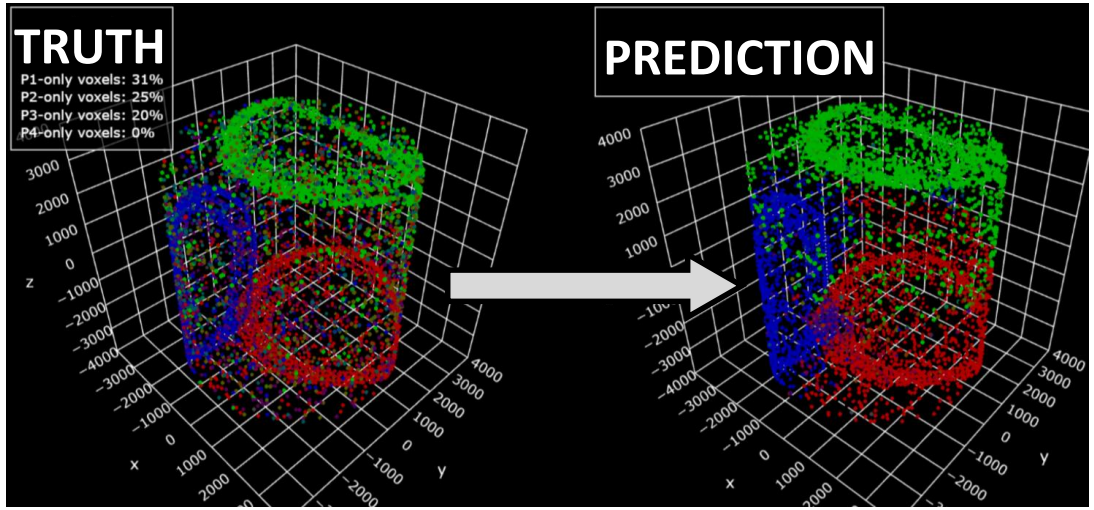


⇒ ML performances ≫ likelihood for all regression tasks (1R) & classification

Can ML be used to fully segment MR event → set of 1R events ?

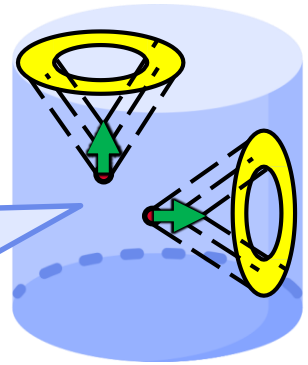
RING SEGMENTATION

- Predict per-PMT ring-contribution propa



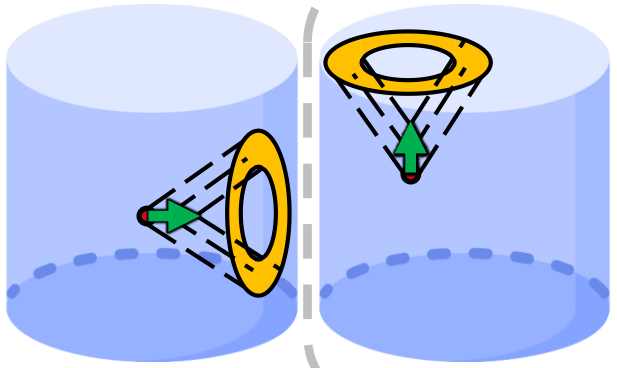
Metrics	Coral MR	Coral + ML seg
Ring found	92.5%	95.5%
MR fit duration	600s	80s
Momentum residual CI68% [MeV]	112	82
Momentum residual fraction CI68% [%]	13.9	10.0
Direction CI68% [deg]	2.3	2.3
Vertex – single vtx fit CI68% [cm]	11.6	10.8

- 10k events
- 1R to 4R
- Mix e^-/μ^-



Mathias El Baz
Ring Seg.

Ring Seg.



CORAL



⇒ ML segmentation improves 1R reconstruction with fiTQun/CORAL

Outline

1. Hyper-Kamiokande overview

2. Likelihood algorithms in HK
   a. Low energy algorithms: BONSAI and LEAF
   b. High energy algorithms: fiTQun and CORAL
   c. Strategies to improve performances

3. Machine Learning development in HK
   a. Convolution Neural Networks
   b. Graph Neural Networks
   c. Performance comparison with fiTQun

4. **Conclusion**

Conclusion & Perspective

Hyper-Kamiokande = next generation neutrino detector with massively increased statistics (x8)

Traditional likelihood-based fitter pipeline ported from SK to HK (BONSAI/LEAF & fiTQun/CORAL)

Robust and precise, ready for seamless start of physics analysis on day 1

- Increased neutrino interaction rate → developing strategies to speed-up process time
- Better reconstruction → improve statistical σ (e.g. increase FV close to the wall)
  → improve systematic σ (e.g. improve multi-ring event discrimination)

Very active ML development in HK with excellent performances

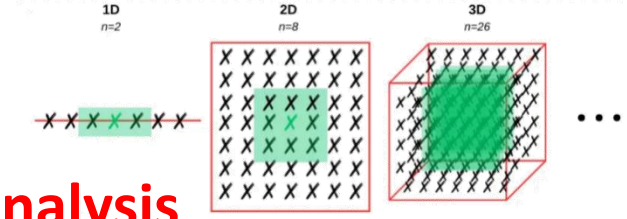
- CNN & GNN >> NLL methods for all tested tasks & metrics, **up to 60% improvement !**
- x100 faster on CPU, x100k faster on GPU → <1 sec/event at HighE
- Trade-off = multiplicity of ML models VS 2 likelihood algorithms

Short term : speed-up NLL ; perform all tasks with every models
Mid term : ring segmentation for modular NLL/ML pipeline
Long term : Differentiable Detector Sim (CIDeR-ML/LUCiD)



BACKUP

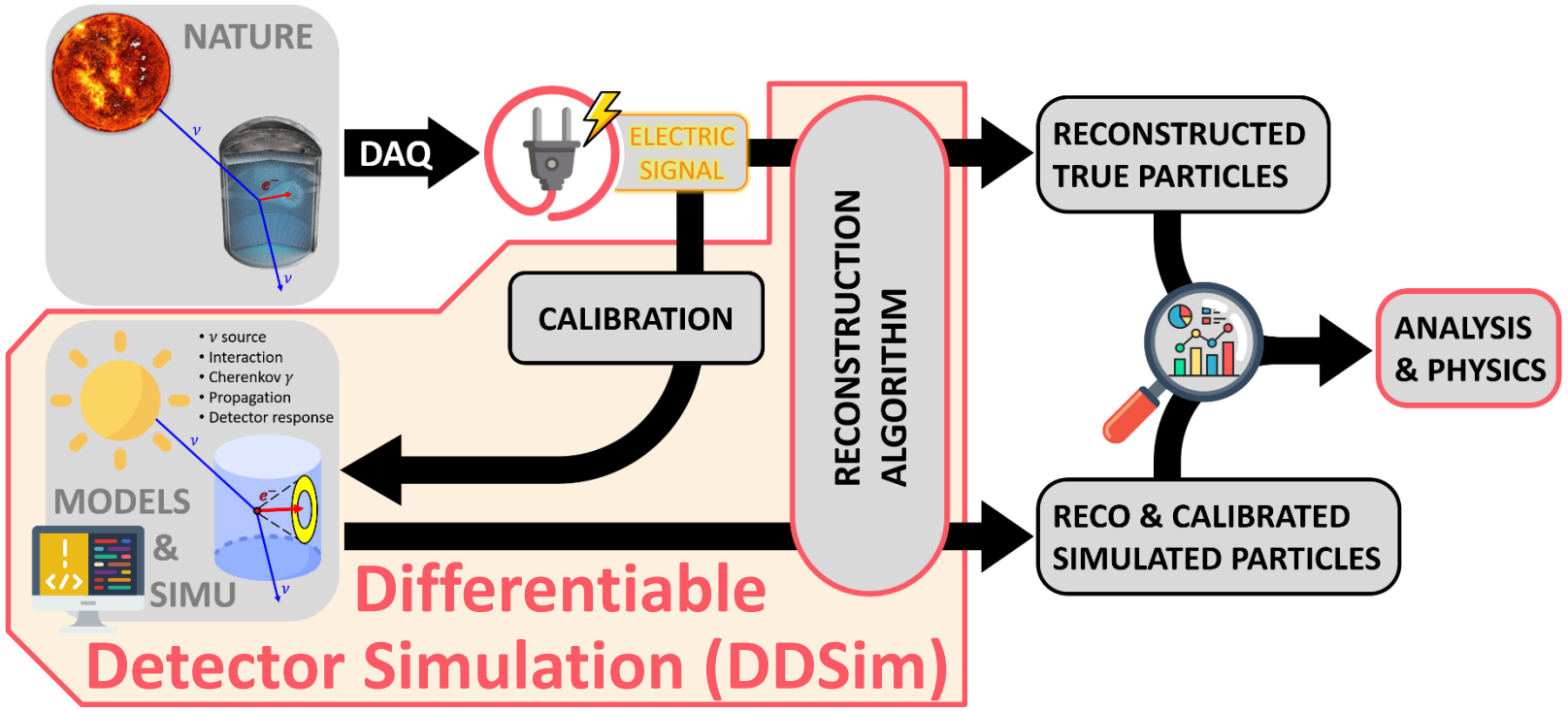
- Many detector params: geometry, Cherenkov physics, water properties, PMT and wall reflectivity, residual magnetic fields, PMT and electronic response, …

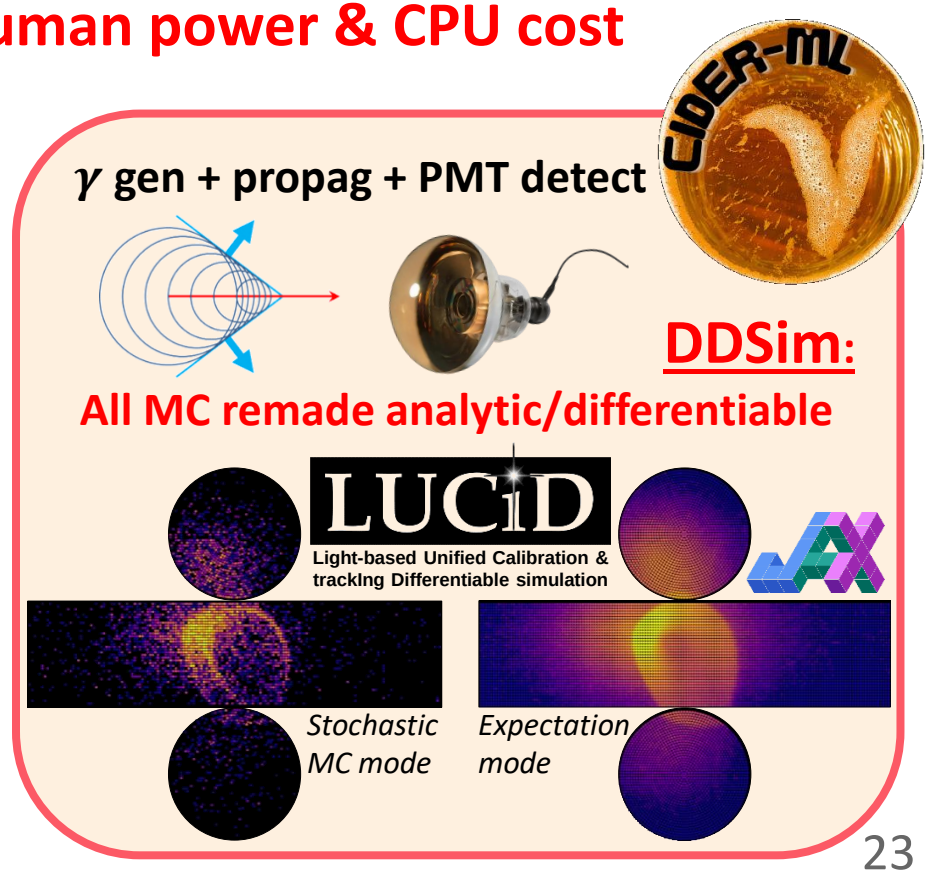


⇒ **Cannot reasonably optimize all parameters with conventional MC analysis**

- Detector condition ⟶ Calibration ⟶ MC ⟶ Eval detector systematics using particle control sample
  ⟶ Tune reconstruction algorithms

⇒ **Not convenient and takes time: regular iteration, high human power & CPU cost**





- Originally developed for Super-K detector
  - Based on algorithm of MiniBooNE: https://arxiv.org/abs/0902.2222
- Uses information from unhit PMTs + time & charge of hit PMTs:

$$L(\mathbf{x}) = \prod_{j}^{unhit} P_j\left(unhit|\mathbf{x}\right) \prod_{i}^{hit} P_i\left(hit|\mathbf{x}\right) f_q\left(q_i|\mathbf{x}\right) f_t\left(t_i|\mathbf{x}\right)$$

- L: Likelihood to maximise
- $\mathbf{x}$: Candidate event hypothesis
- $P_j\left(unhit|\mathbf{x}\right)$: Probability of no hit at PMT
- $P_i\left(hit|\mathbf{x}\right)$: Probability of hit at PMT
- $f_q\left(q_i|\mathbf{x}\right)$: Hit charge probability density
- $f_t\left(t_i|\mathbf{x}\right)$: Hit time probability density

- Probabilities calculated based on direct + scattered + reflected light
  - Approximations / assumptions are used to simplify these calculations
- Likelihood ratios used to distinguish particle types and single-ring / multi-ring event topology hypotheses

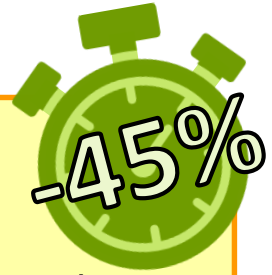
$\mathbf{x} = (x, y, z, t, p, \theta, \phi)$ track parameters

SPEEDING-UP LIKELIHOOD ALGORITHMS

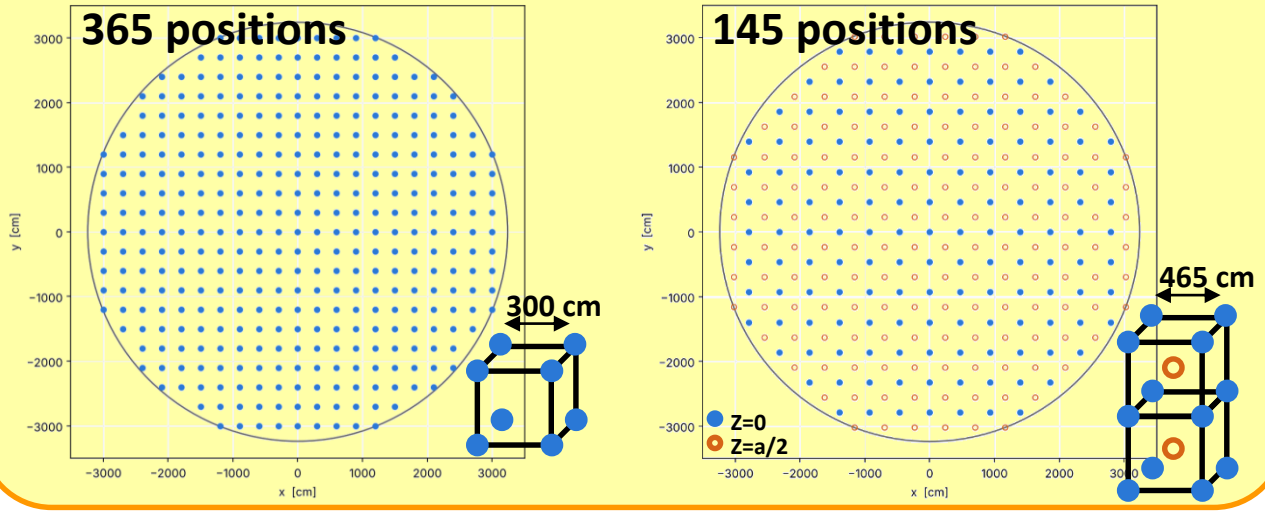
B Prefit optimized search, by L. Perisse & C. Sammons

- Prefit search → final likelihood minimization in a box around best vertex candidate(s)
- Currently prefit search ~750-800 ms

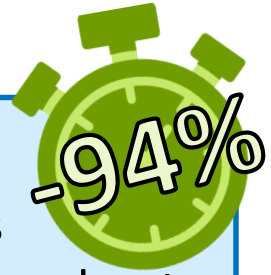
LEAF prefit (LowE)



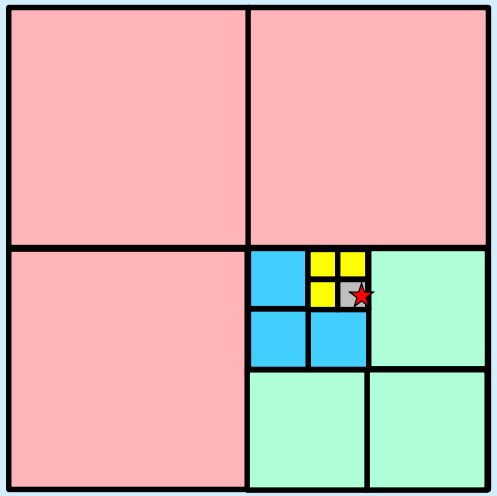
- Coarse grid search with fewer candidates
- Minimize worst case distance vtx candidate ↔ true
  - Δt~10 ns for direct PMT hits to count in NLL
  - Resolution with new grid ~ former grid



fiTQun prefit (HighE)



- Current: grid 6x6x6x40 = 8640 steps
- New: grid 4x4x4 then subgrid best quadrant, repeat 8 times = 512 steps in total
  - Prefit resolution similar with both grids



Distance to true vs prefit iteration

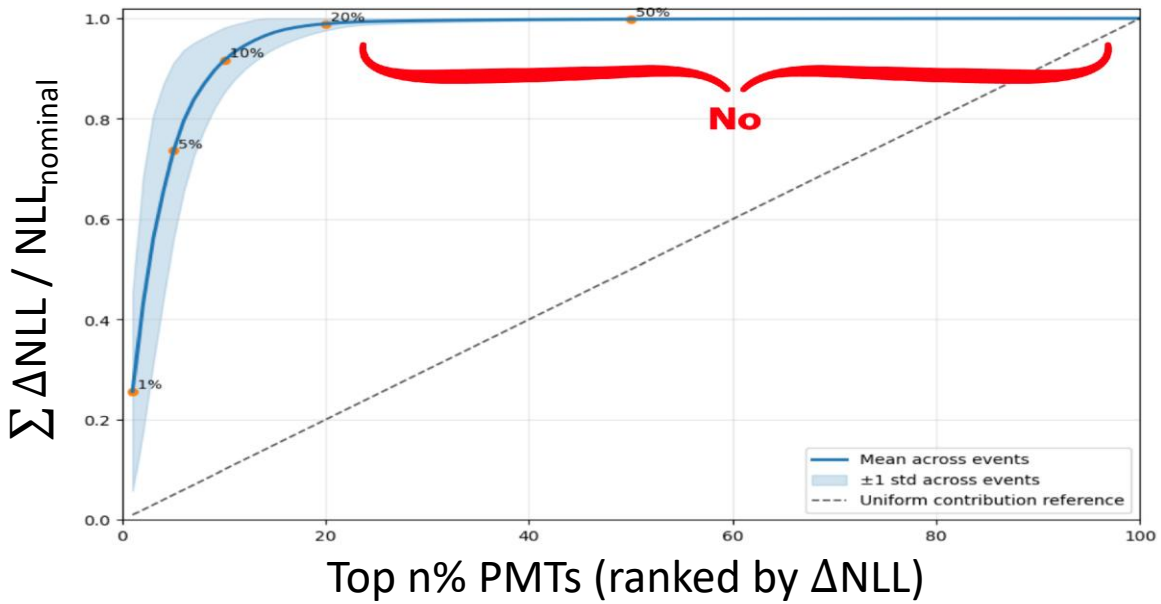
⇒ Prefit time -45%

SPEEDING-UP LIKELIHOOD ALGORITHMS

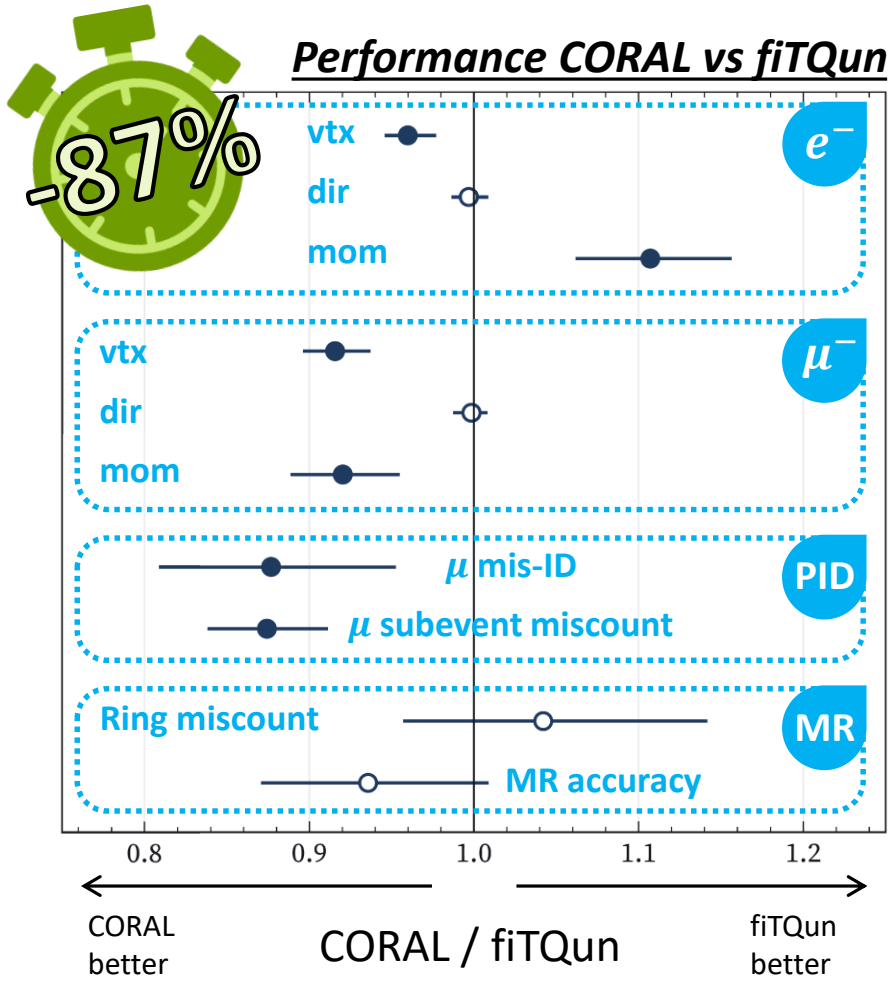
C Studying the impact of unhit PMTs with CORAL, by B. Carew

- CORAL = fiTQun ported to python
  - Just-In-Time compilation speed-up
- fiTQun's $\mathcal{L}$(x) uses all hit & unhit PMTs
  - Only 20% unhit contribute to 100% $\mathcal{L}$(x)

Unhit PMT contribution to $\mathcal{L}$(x)



Performance CORAL vs fiTQun



LEAF GRID SEARCH

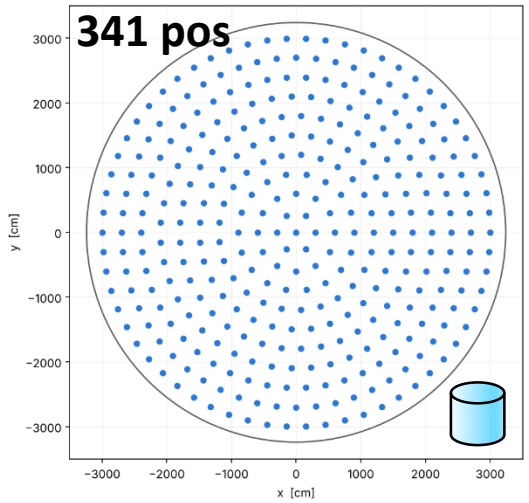
→ **Coarse grid search = most of process time**

- Prefit vertex ~80% of $\mathcal{O}$(1s)/event in HK

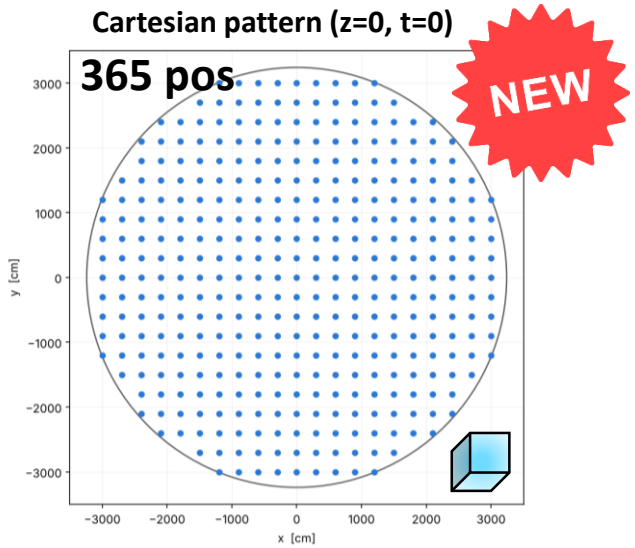
Coarse grid search: Scores all candidates $t_{res} \in [-3,+4]$ ns, refines top 20

- Final vertex fit reach ±600 cm, ±28 ns from any candidate
- Ranking of candidate is the pitfall → nearest candidate not making top 20
- +1m away from nearest candidate → ~4.6 ns of t_{res} smearing, missing peak
- Quality of a grid = how far true vertex sits from nearest candidate for a given number of candidates, considering typical and worst cases

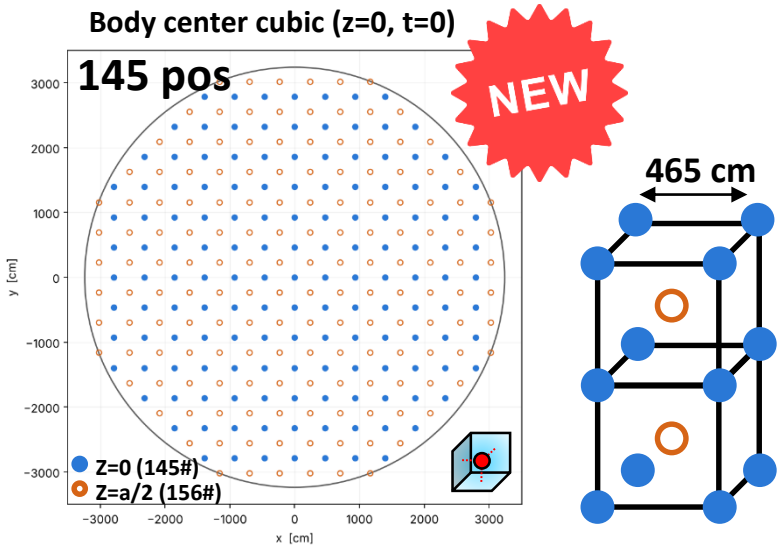
Cylindrical pattern
(z=0, t=0, LEAF built-in)



Cartesian pattern (z=0, t=0)

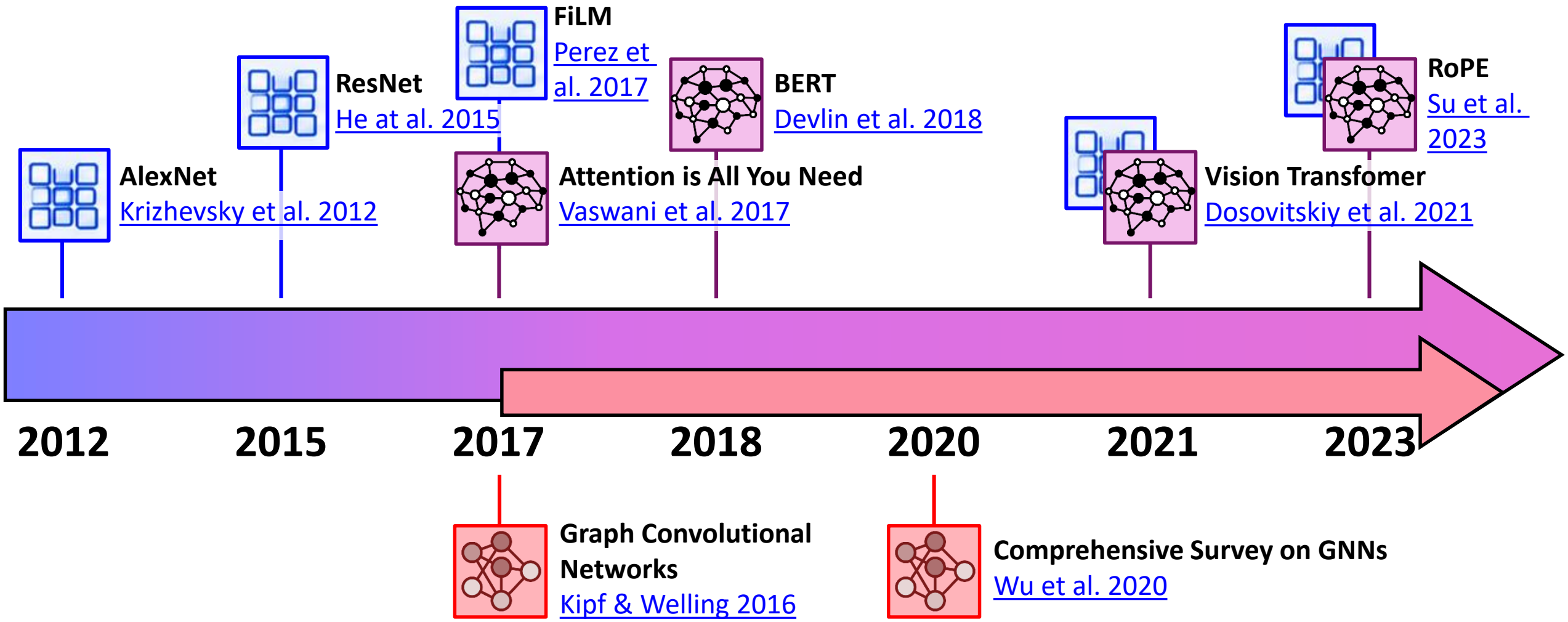


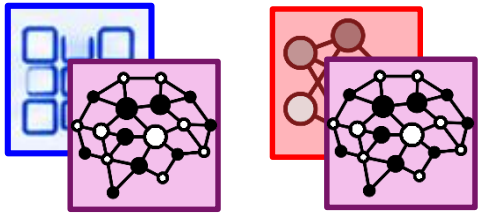
Body center cubic (z=0, t=0)



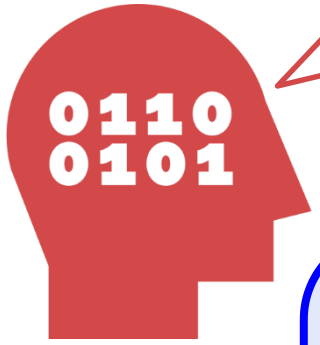
	Pattern	Candidate	Median	99%	Worst dist	*Distance true vtx to nearest candidate in inner volume (>300m from wall)*
	Cylindrical	7502	149 cm	224 cm	257 cm	Rings don't line up, uneven spacing → less optimized case
	Cartesian	7665	148 cm	227 cm	258 cm	Uniform spacing, cost ~cylindrical, weak spot at cube center
	BCC (step 300)	4359	182 cm	244 cm	259 cm	Less candidt, same worst-case, higher median → weakens ranking
	BCC (step 250)	7213	152 cm	203 cm	216 cm	Cost ~cartesian, minimizes worst distance from true to candidate

Chronology of CNN, GNN & Transformer





Very active development in Hyper-K for HighE and LowE



Machine learning based inference techniques greatly improved in last decades
1M events → fiTQun +10k CPU hours / **ML 100 CPU hours or 10 GPU minutes**

WatChMaL

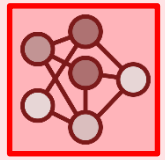
Water Cherenkov Machine Learning



CNN oriented

CAVERNS

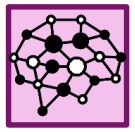
CP Asymmetry discovery with an Enhanced Reconstruction algorithm for Neurinos in Super and Hyper-Kamiokande



GNN oriented

LowE

HighE



Transformer

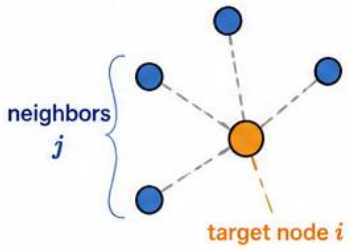
Gilberto Rodriguez — IWCD	Felix Cormier — SK	Alberto Saborido — WCTE	Lorenzo Périssé — HK
Kieren Joseph — WCTE	Shuoyu Chen — HK	Mathieu Ferey — WCTE	Naomi Foster — HK
John Nugent — WCTE	Andrew Atta — HK	Viet Nguyen — WCTE	Mathias El Baz — HK
	Nick Prouse — HK	Erwan Le Blévec — HK	Zhuojun Hu — HK

THE ATTENTION MECHANISM

- Attention computes a weighted combination of values from a set of neighbors.
- The weights (attention scores) are learned from the features of the target node and its neighbors.
- More relevant neighbors → higher weight. Irrelevant neighbors → lower weight.
- Weights are normalized with a softmax so they sum to 1.

1 INPUT

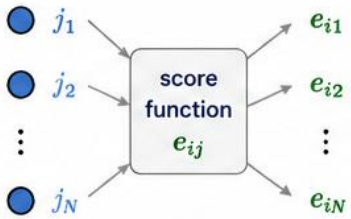
Target node i attends to its neighbors j.



Each node has a feature vector h

2 SCORE

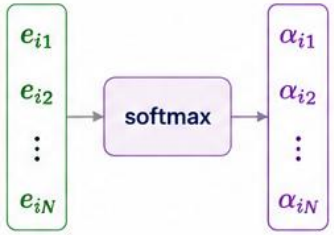
Compute a compatibility score between i and each neighbor j.



e_{ij} measures how relevant neighbor j is to node i

3 ATTENTION WEIGHTS

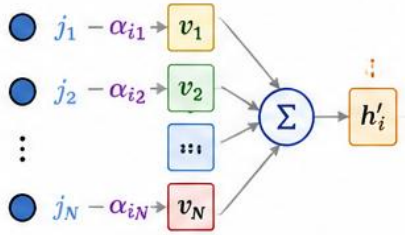
Normalize scores with softmax to obtain attention weights.



$\alpha_{ij} \geq 0$ and $\sum_j \alpha_{ij} = 1$

4 AGGREGATE

Take the weighted sum of neighbors' values (features).



Output: updated representation for node i

INTUITION

Attention acts like a data-driven importance score.

The model learns what to focus on for the task.

Noisy or irrelevant neighbors get low weights.

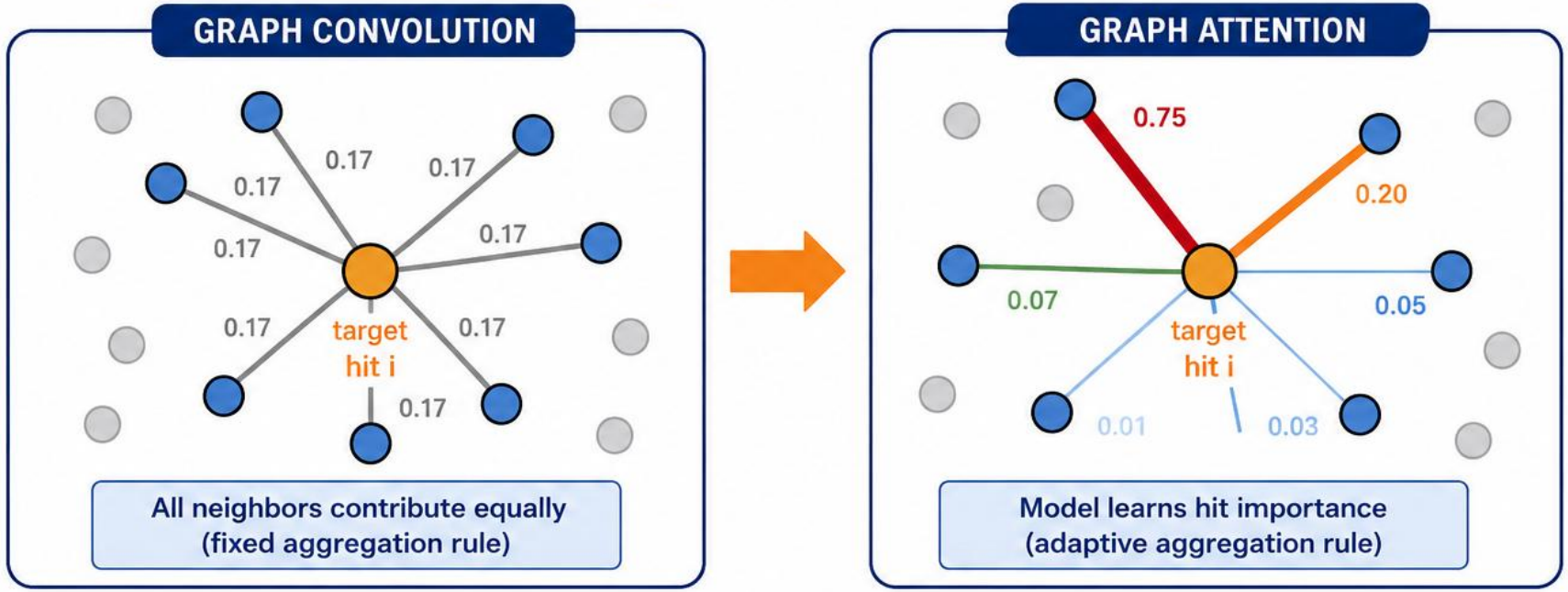
Common formulation

$$e_{ij} = a(\underbrace{W_q h_i}_{\text{query (from } i)}, \underbrace{W_k h_j}_{\text{key (from } j)}) \qquad \alpha_{ij} = \frac{exp(e_{ij})}{\sum_{k \in \mathcal{N}(i)} exp(e_{ik})} \qquad h'_i = \sum_{j \in \mathcal{N}(i)} \alpha_{ij}, \underbrace{W_v h_j}_{\text{value (from } j)}$$

Attention lets the model adaptively focus on the most informative neighbors for each node.

ATTENTION MECHANISM IN GRAPHFORMER

- Graph convolution treats all neighboring hits similarly
- Some PMT hits contain more useful information than others
- Attention learns which hits are important for the reconstruction task
- Learned weights depend on the event topology



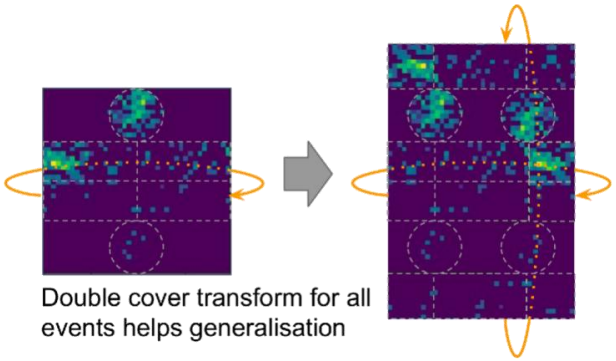
KEY ADVANTAGES

- Adaptive neighbor weighting
- Noise suppression
- Better handling of sparse detector data
- Foundation of the GraphFormer architecture

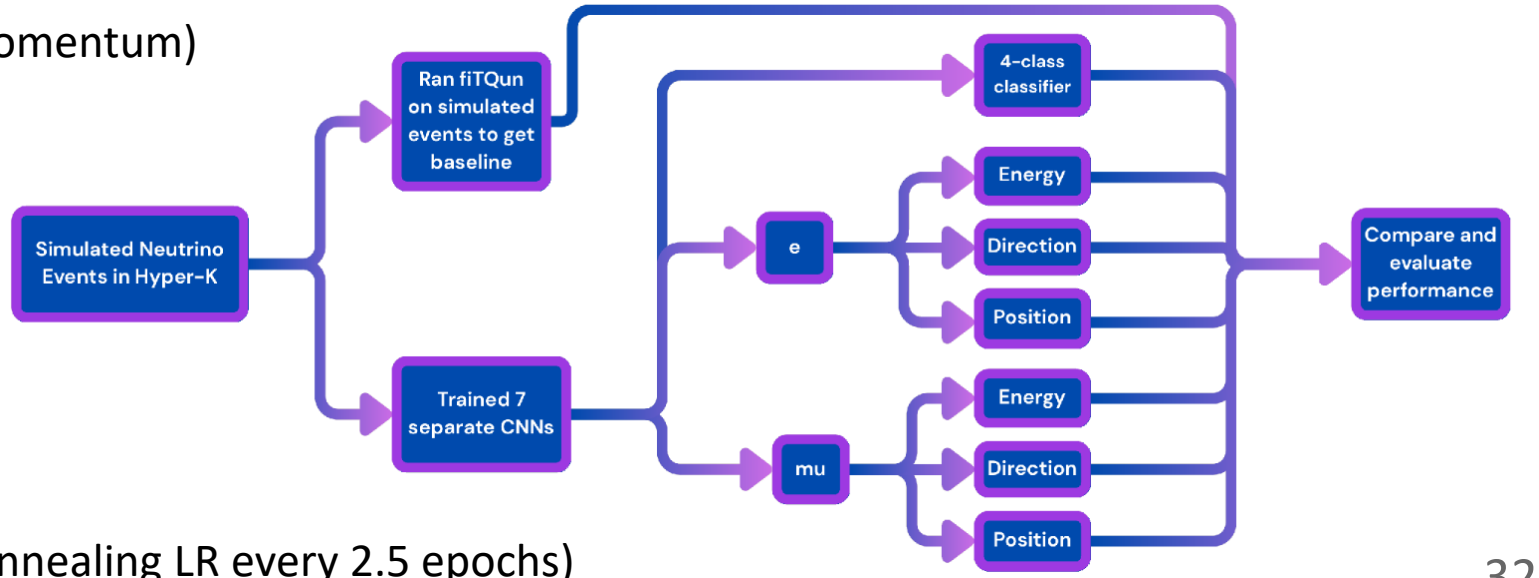
Attention replaces fixed aggregation rules with learned event-dependent weights

RESNET MODELS

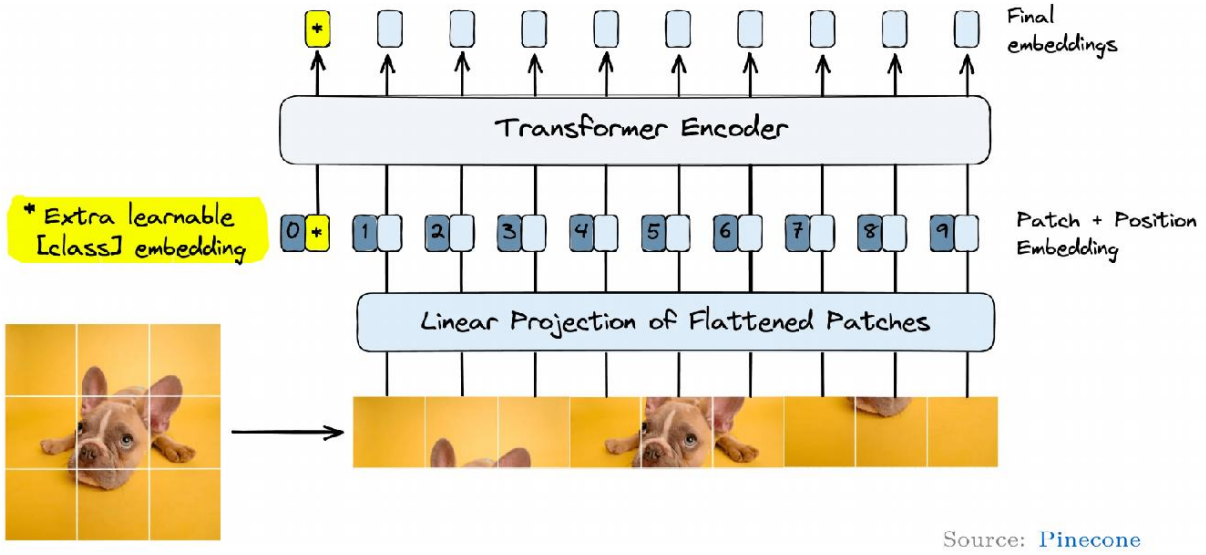
- Simulate HK using Geant4-based tool WCSim
  - Interaction processes + Cherenkov light
- 40M events = 10M e^- + 10M μ^- + 10M γ + 10M π^0
- Uniform vertex distribution in the tank
- Isotropic direction distribution
- Uniform 0-2 GeV kinetic energy distribution
  - Above Cherenkov threshold for each particle
- Only fully contained (FC) events are kept
- Muon decay disabled
- 7 models are trained: 6 regression, single 4-class classifier
- Regession loss function: Huber (relative for momentum)
- ResNet-152 architecture
- 190x189 input images
  - 2 channels: PMT hit time and charge
- Trained on 4 A100 GPUs
  - 20 epochs for regression
  - 8 epochs for classification
  - 512 batch size
  - AdamW optimiser
  - Learning rate from 1e-3 to 1e-6 (cosine annealing LR every 2.5 epochs)


Double cover transform for all events helps generalisation

Cuts	FC and nhits>200
Energy	0-2GeV
PID	μ, e, γ, π^0
Muon Decay?	no decay
Training Split	96:2:2
Training Set	~8,569,149
Validation Set	~178,449
Testing Set	~178,595



ViT ARCHITECTURE



Treat an image as a sequence of tokens.

Positional embedding:

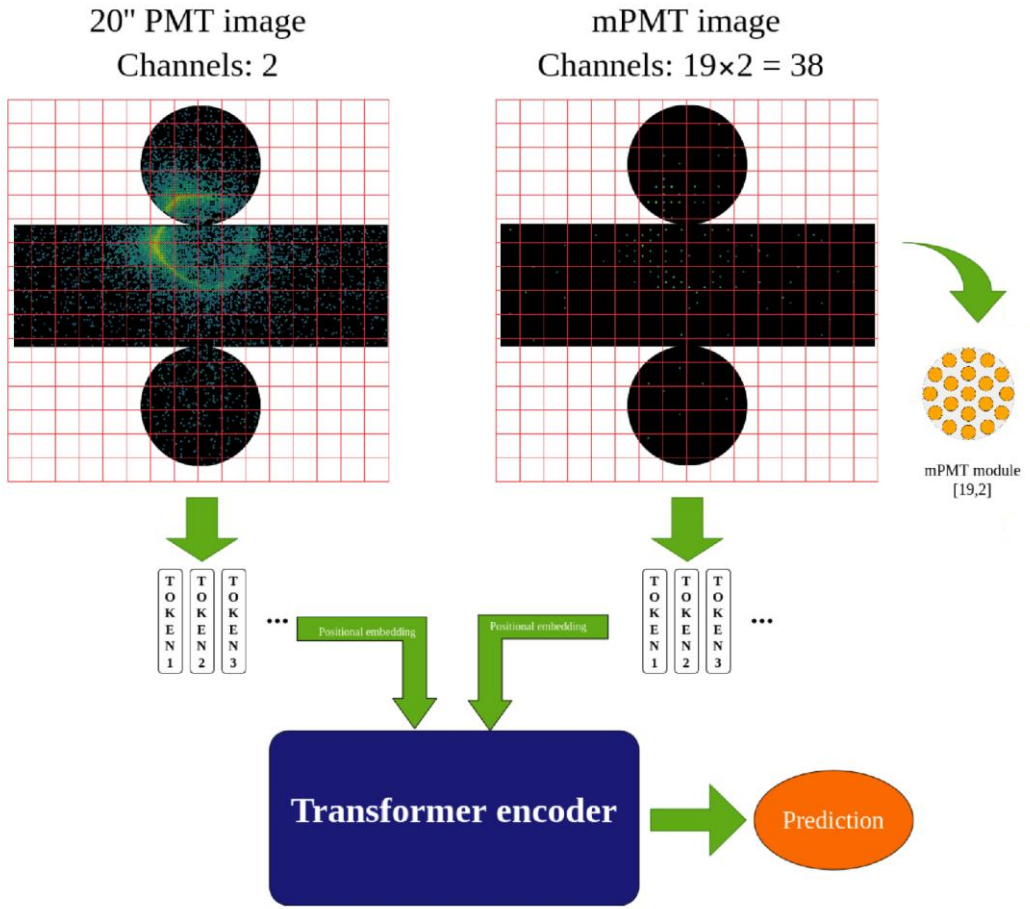
- When an image is divided into patches, the model loses information about where each patch comes from.
- To recover this information, ViT adds a learnable positional embedding to each token:

$$z_i = x_i + E_{\text{pos},i}$$

where x_i is the patch token and $E_{\text{pos},i}$ encodes its spatial location.

Transformer encoder:

- A stack of layers combining multi-head self-attention and feed-forward neural networks.
- Allows tokens to exchange information across the entire image.



Convert the 20” PMTs and mPMTs hits into tokens separately.

Process them together in the transformer encoder.

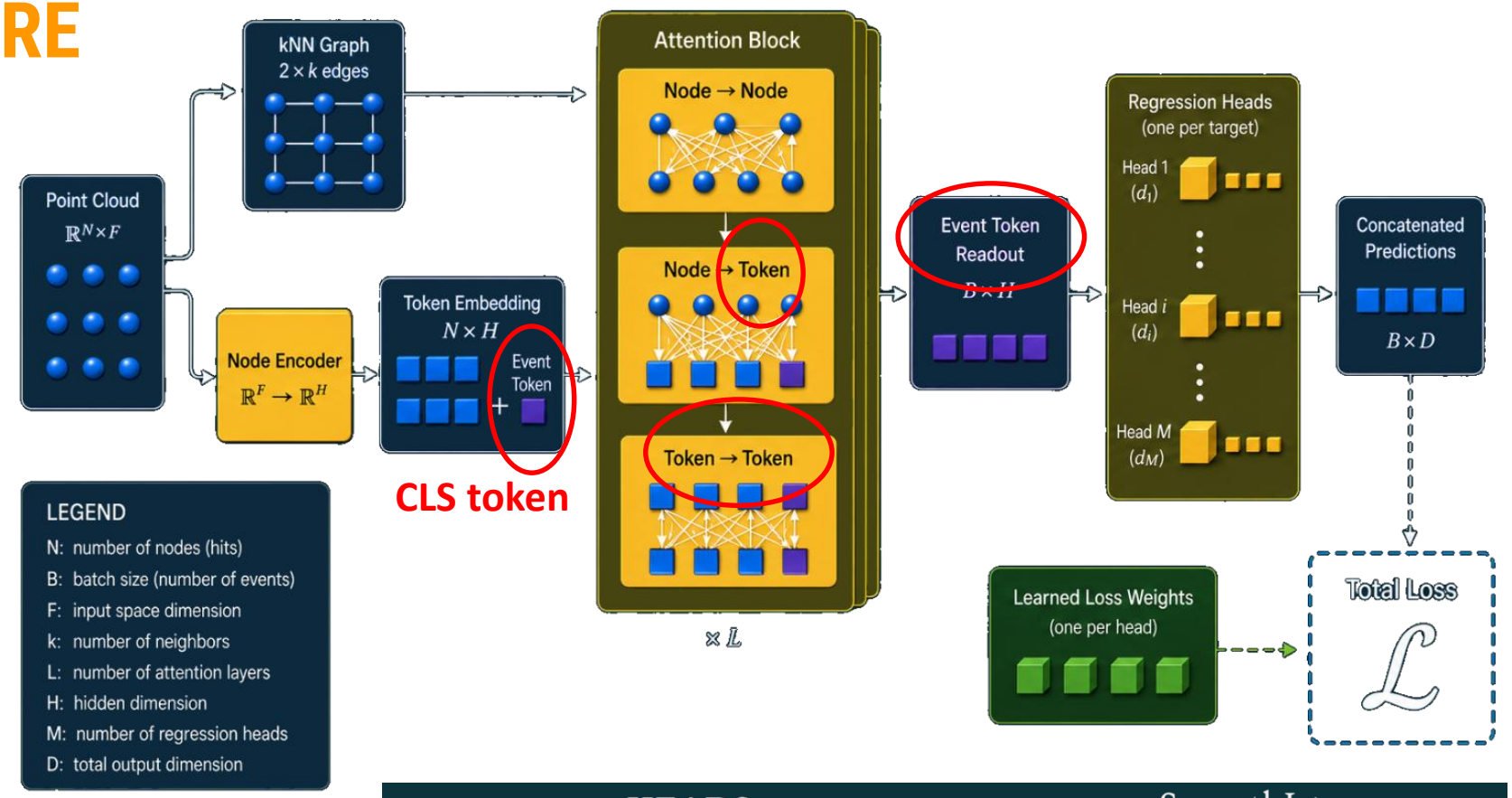
CLSGAT ARCHITECTURE

Electrons up to 60 MeV

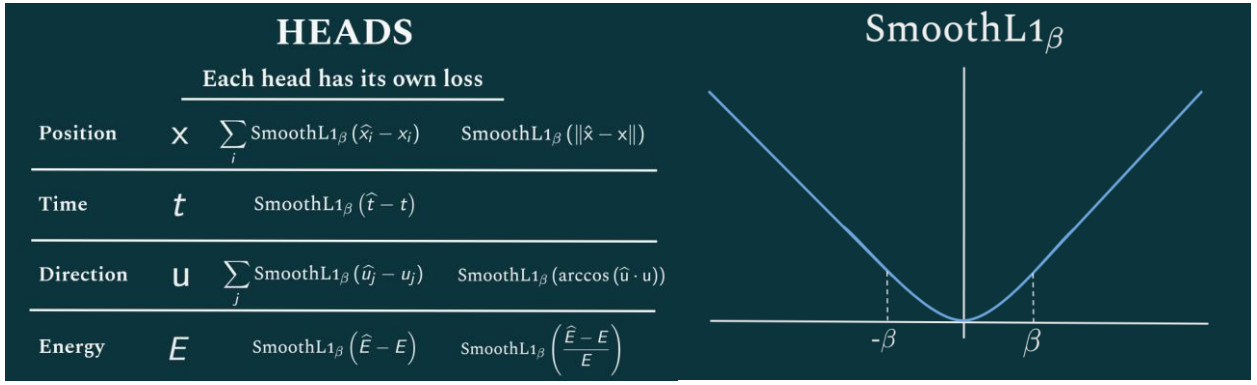
- WCSim 1.12.29
- Uniform in ID, isotropic
- 4.2 kHz dark noise
- 400ns time window maximizing nhits
- Train: 400k events
- Val: 100k events
- Test: 500k events

Electrons up to 20 MeV

- WCSim 1.12.16
- Uniform in ID, isotropic
- 4.2 kHz dark noise
- 400ns time window maximizing nhits
- Train: 460k events
- Val: 110k events
- Test: 350k events



- CLSGAT: CLS token + Graph Attention Transformer
- CLS (classification) token: global node added to a graph's node sequence to aggregate overall network information, function equivalently to a virtual node, and serve as the final embedding for whole-graph prediction tasks



CheRP ARCHITECTURE

Model Parameters

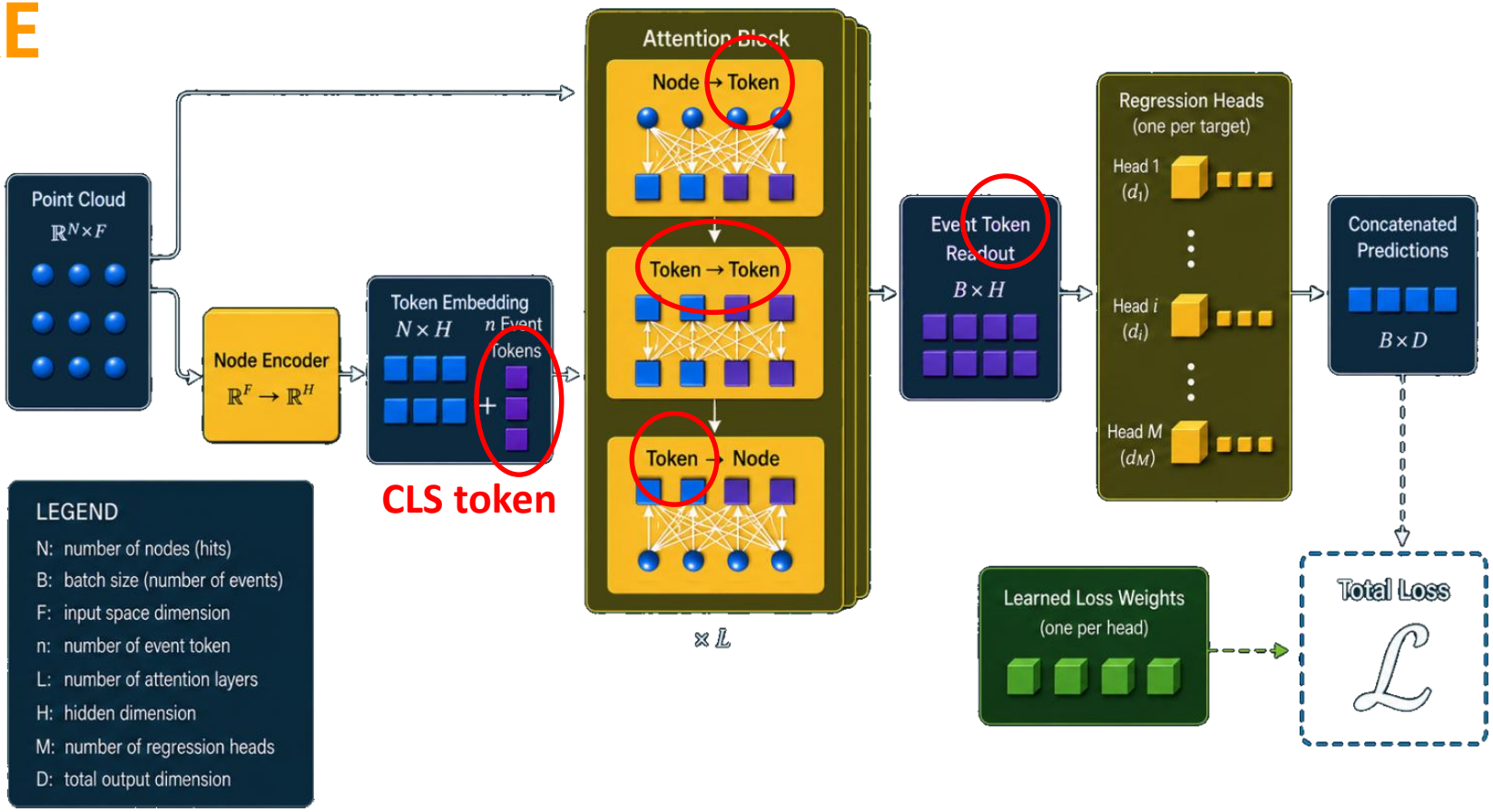
- 72 token dimension
- 8 attention heads
- 6 tokens
- 4 perceiver layers

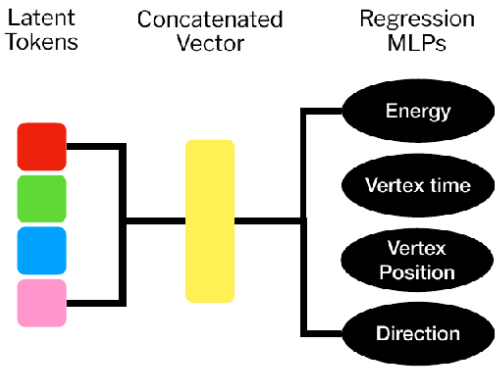
Half a million total parameters!

Training Data

Using MC production macros with WCSim v1.12.29:

- 5 million electron events.
- Uniform energy 0 – 2 GeV.
- Isotropic in direction, uniform throughout the ID.
- Events not contained within the ID cut.
- Cutting events with zero hits.





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

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

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

- CheRP : Cherenkov Ring Perceiver
- Directly works with pointcloud
  - Graph edges not needed

RING SEGMENTATION ARCHITECTURE

3D Voxels $\in \mathbb{R}^{V\times 5}$

Per voxel features= (q, x, y, z, t)
(PMT position, digi charge and time)
Label=(Ring 1, … , Ring N , Background)

↓

Sparse U-Net Encoder

↓

Embedding : $e_{vk} \in \mathbb{R}^{V\times D}$

↓

Transformer head

↓

Predictions : $p_{vk} \in \mathbb{R}^{V\times(N+1)}$

Ground truth :

$$y_{vk} = [0,1]^{V\times(N+1)}, \sum_k y_{vk} = 1$$

Predictions:

$$p_{vk} = [0,1]^{V\times(N+1)}, \sum_k p_{vk} = 1$$

Cost function :

$$\mathrm{Cost}(y_{vk}, p_{vk'}) = \mathrm{CE}(y_{vk}, p_{vk'}) + \beta\left(1 - \mathrm{Dice}(y_{vk}, p_{vk'})\right)$$

$$\mathrm{CE}(y_{vk}, p_{vk'}) = -y_{vk}\log(p_{vk'})$$

Matching function :

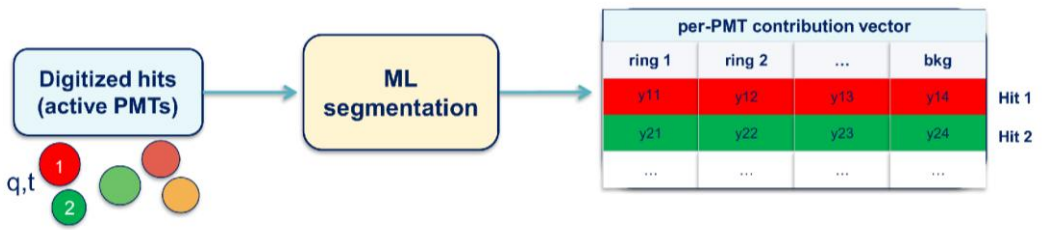
$\mathrm{m} \in \wp_N : [\![0, N]\!] \to [\![0, N]\!]$ (permutation matches label to predicted ring)

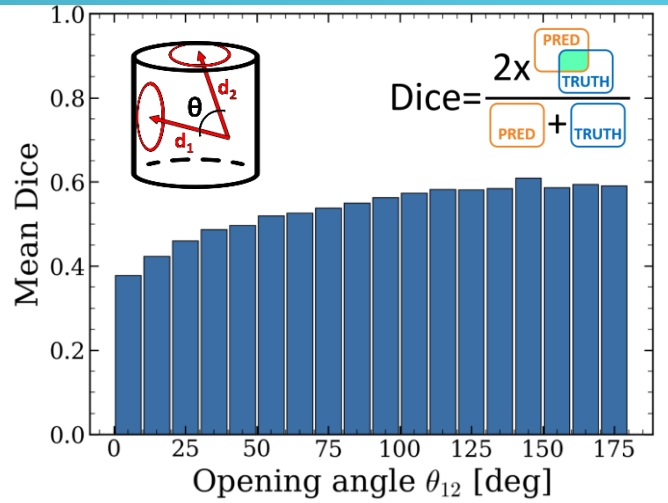
$$m = \mathrm{argmin}_{\sigma\in\wp_N}\left(\sum_v\sum_k \mathrm{Cost}\left(y_{vk}, p_{v\sigma(k)}\right)\right)$$

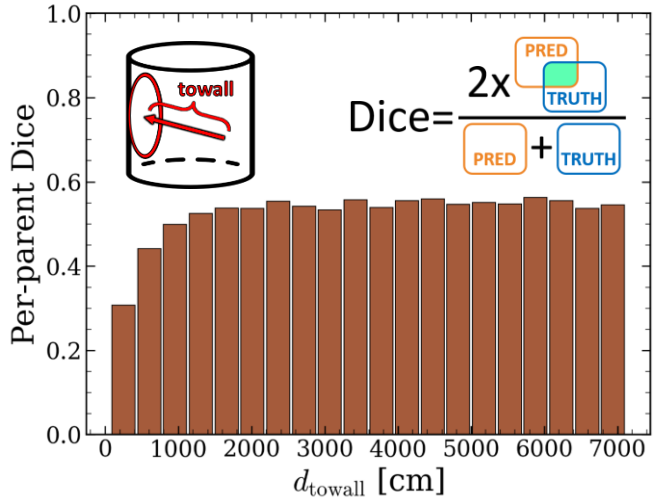
Loss function:

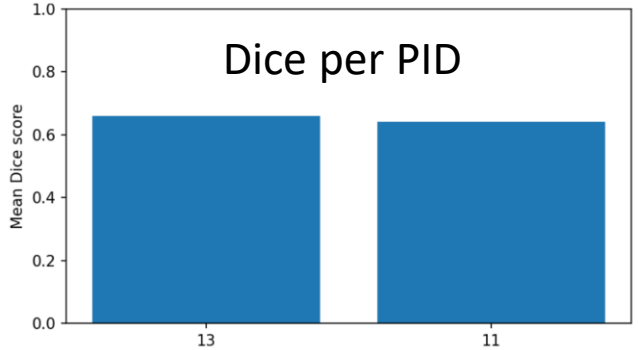
$$\mathrm{Loss}(y, p) = \sum_{vk} \mathrm{Cost}\left(y_{vk}, p_{vm(k)}\right)$$

$$y_{vk} = [0,1]^{V\times(N+1)}, \sum_k y_{vk} = 1, \forall v \text{ active PMT}$$





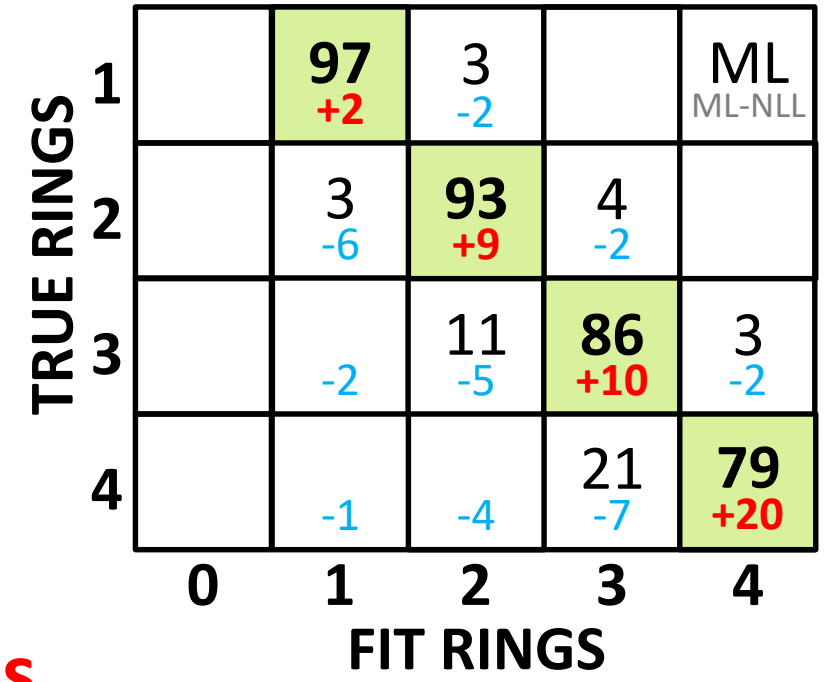


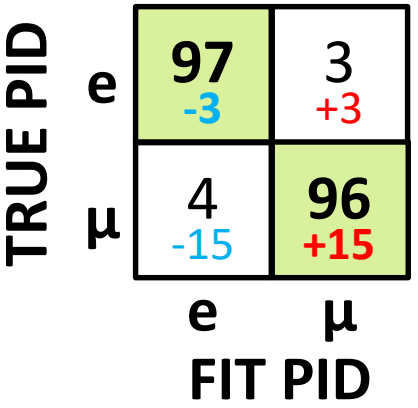


RING SEGMENTATION

- Model = Sparse U-Net encoder + Attention head
- Input = Pointcloud of hit PMTs
- Output = per-PMT ring-contribution propability

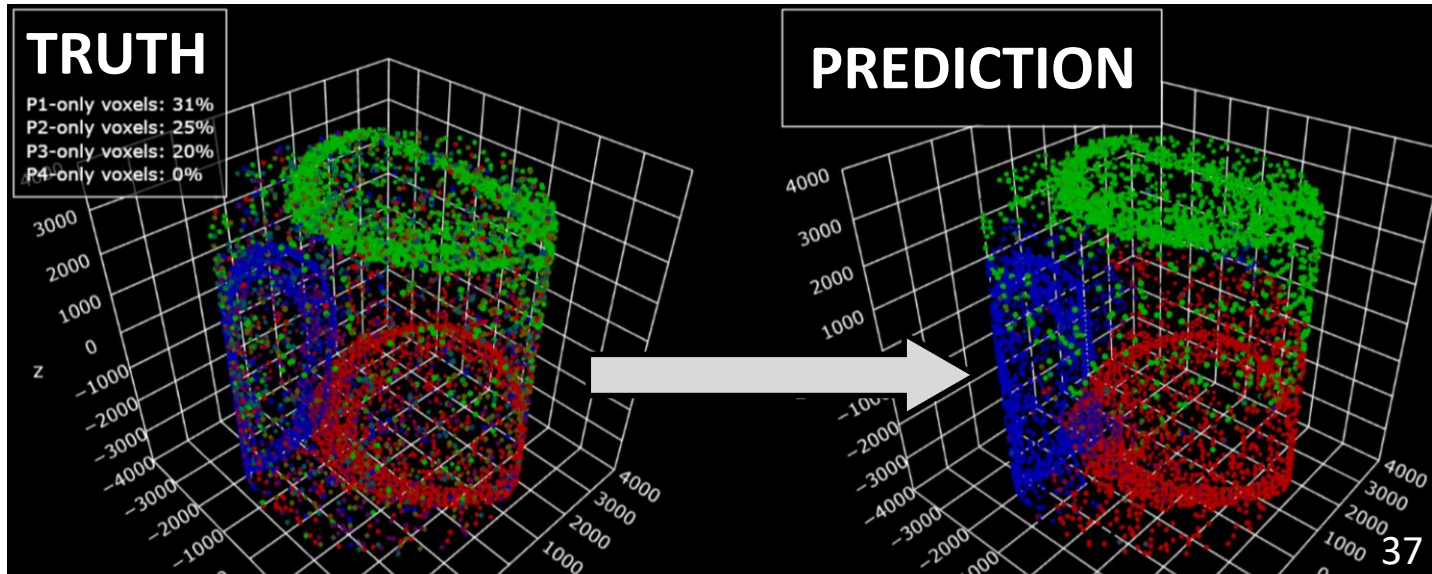
⇒ ML segmentation improves PID reconstruction





Mathias El Baz
Ring Seg.

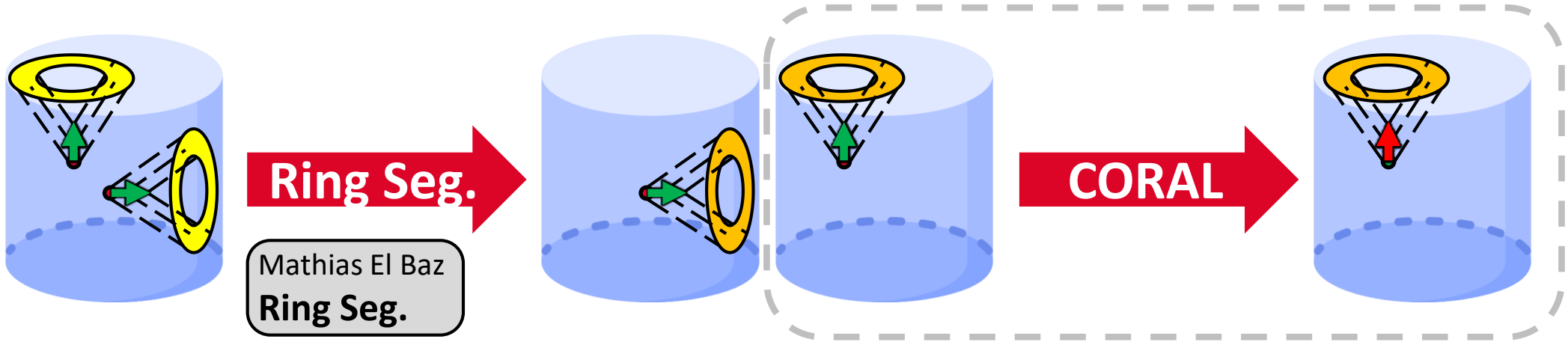
OUTPUT	Ring #1	Ring #2	Ring #3	Ring #4	Bkgd
PMT #1	p11	p12	p13	p14	p15
PMT #2	p21	p22	p23	p24	p25
…	…	…	…	…	…



RING SEGMENTATION

- Dataset multi-ring
  - 10k MR events
    - 1R / 2R / 3R / 4R
    - Mix of e^- and μ^- (50/50)
  - Segment 10k MR → 25k 1R
- Apply CORAL on same dataset

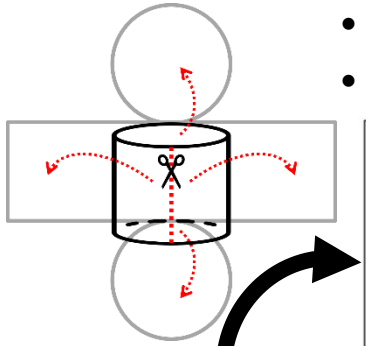
Metrics	Coral MR	Coral + ML seg
Ring found	92.5%	95.5%
MR fit duration	600s	80s
Momentum residual CI68% [MeV]	112	82
Momentum residual fraction CI68% [%]	13.9	10.0
Direction CI68% [deg]	2.3	2.3
Vertex – single vtx fit CI68% [cm]	11.6	10.8

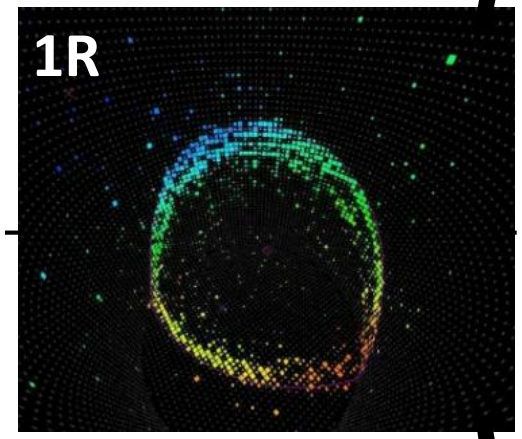


Key principles of HK models

CNN



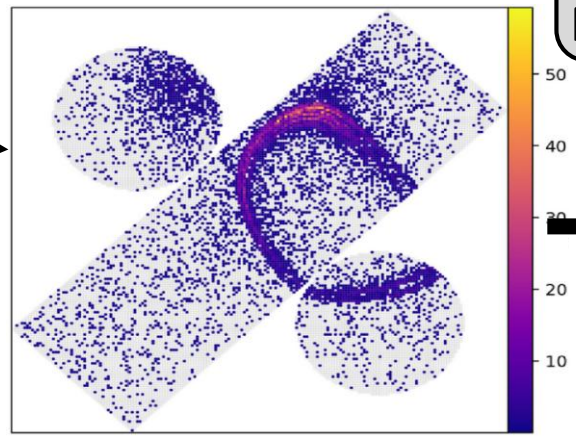


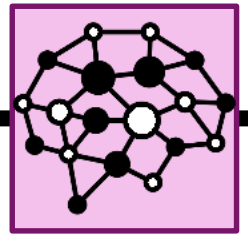


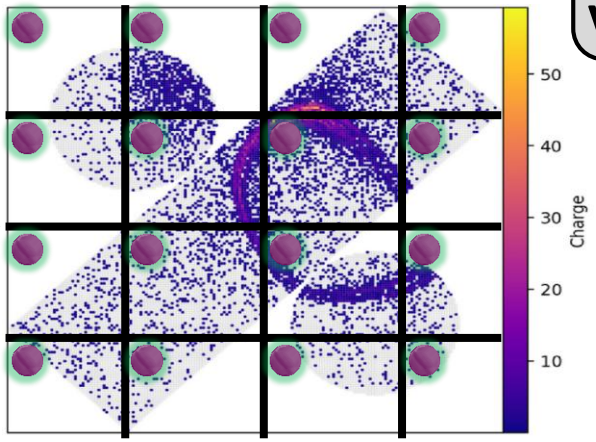
- Based on ResNet-152
- 2 channel PMT input (q,t)
- Double cover transform
- Loss function

Andrew Atta
ResNet-1

Nick Prouse
ResNet-2



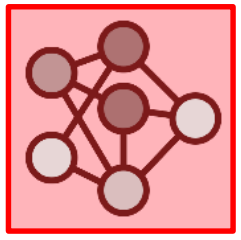


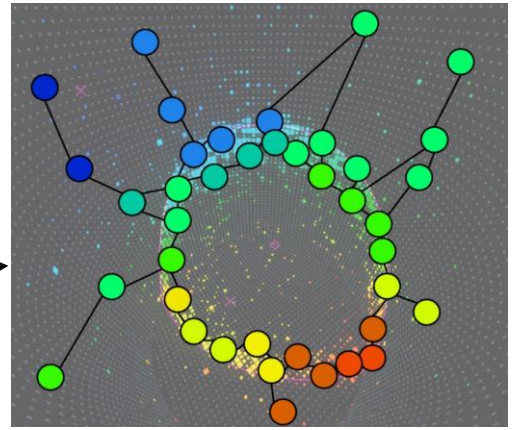


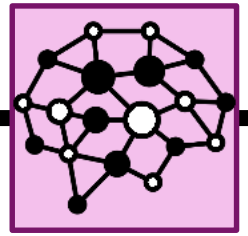
- 2 channel image → patches → token
- Learnable position embedding on tokens
- Tokens → transformer encoder

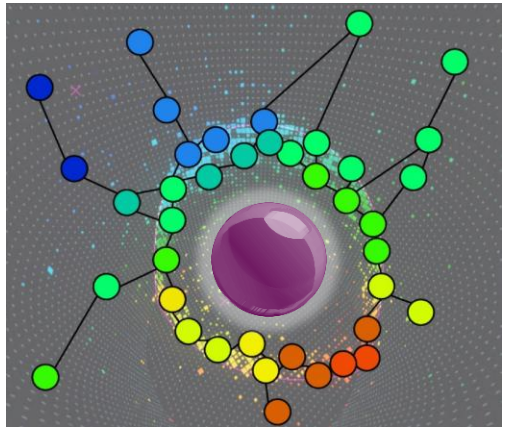
Shuoyu Chen
ViT

GNN









- Point cloud: 5D nodes ($\vec{r}$,q,t)
- Edge = node proximity
- Loss function

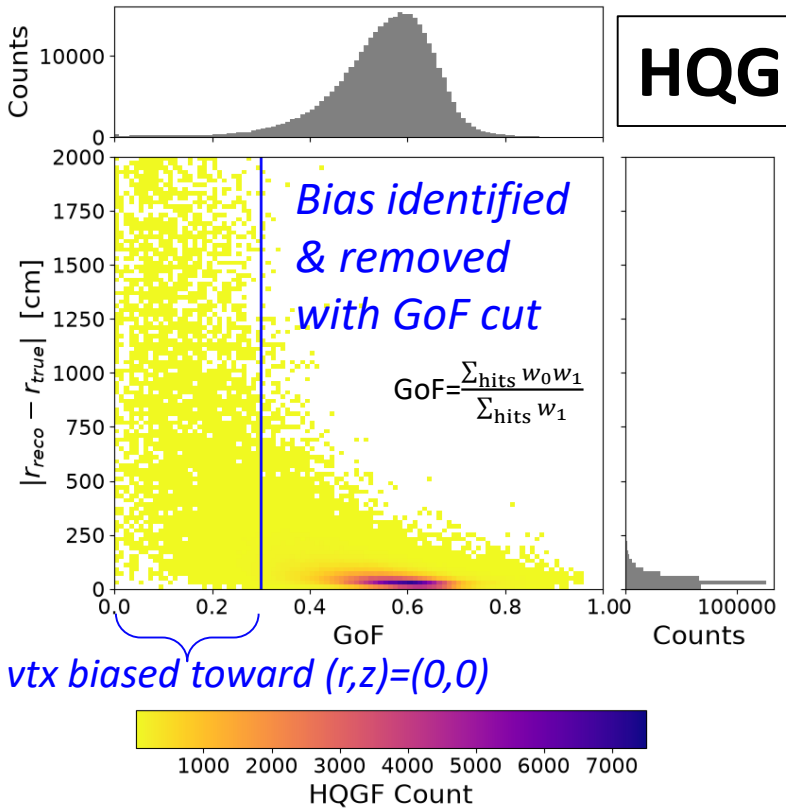
- Attention token/node
- Hit quality classification
- Tokens → transformer encoder

Naomi Foster
CheRP

Lorenzo Perisse
CLSGAT

Erwan Le Blevec
GAT

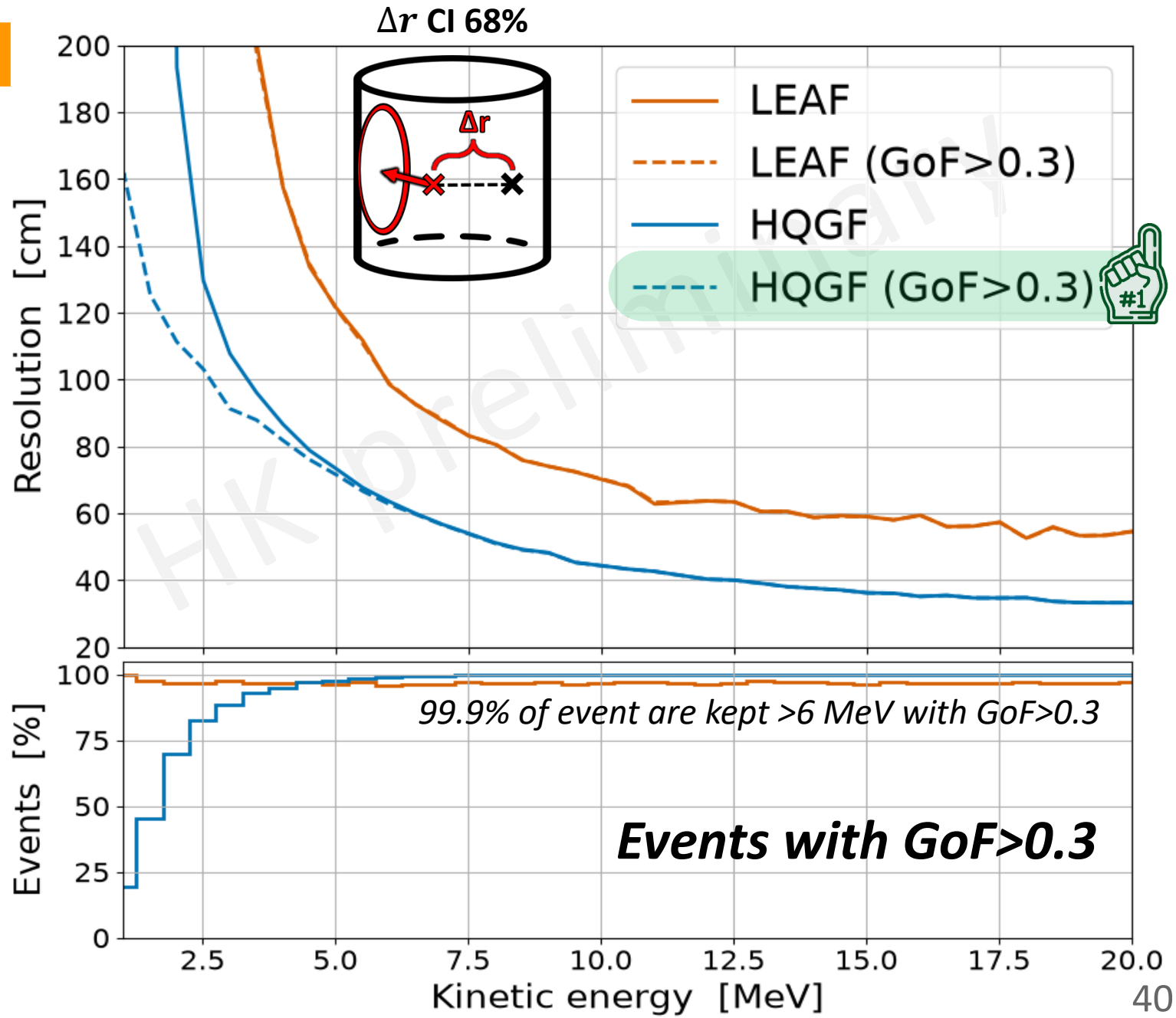
VERTEX RESOLUTION



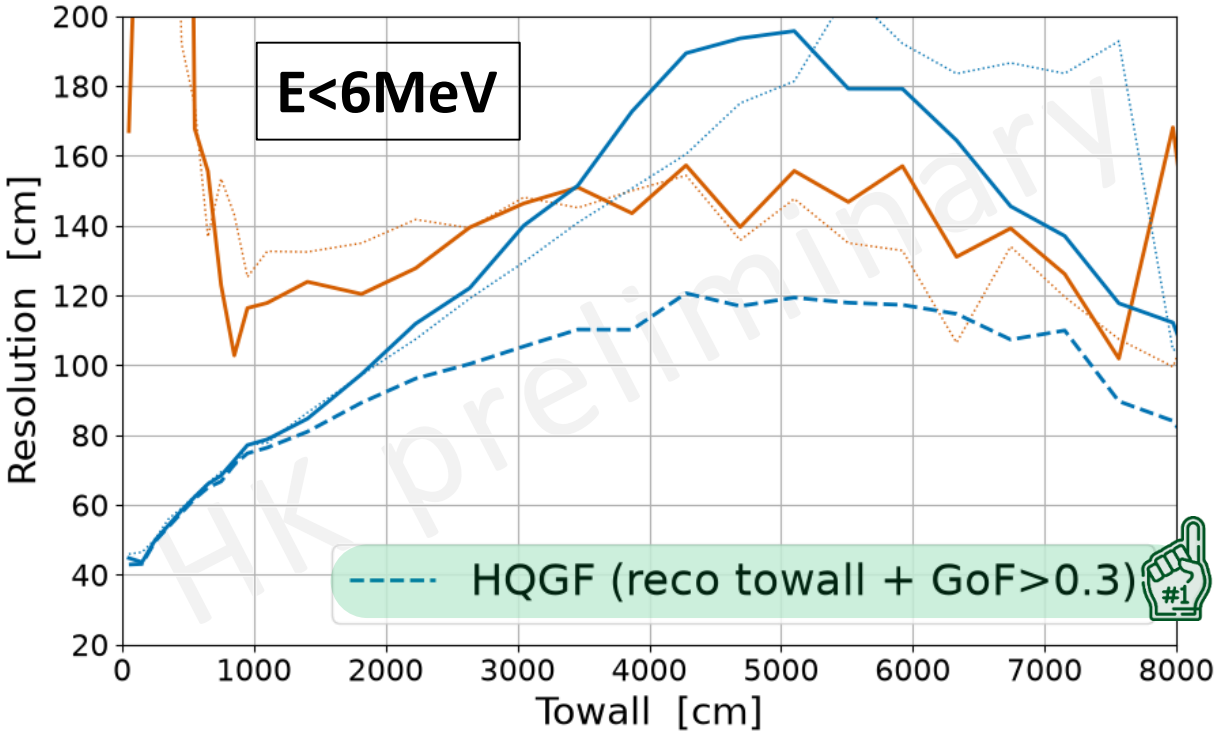
Vertex resolution

110cm @2 MeV

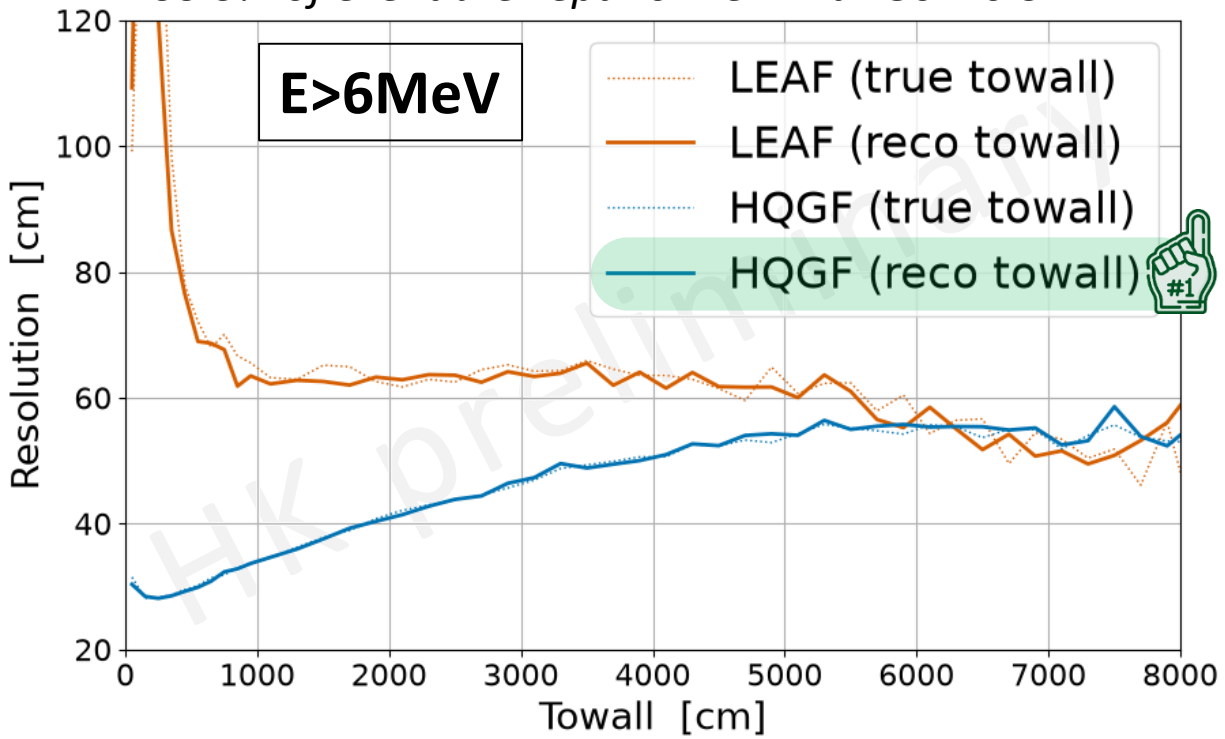
44cm @10 MeV

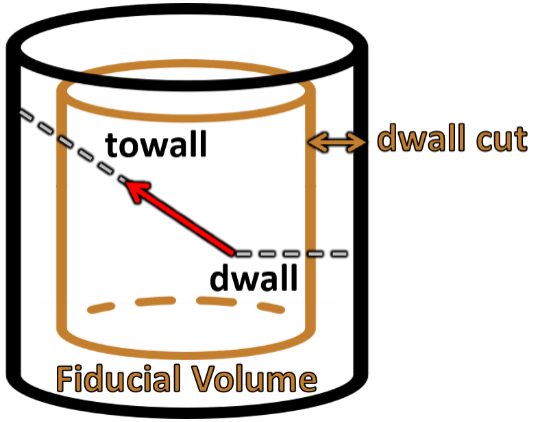


TOWALL DEPENDENCY



99.9% of event are kept >6 MeV with GoF>0.3

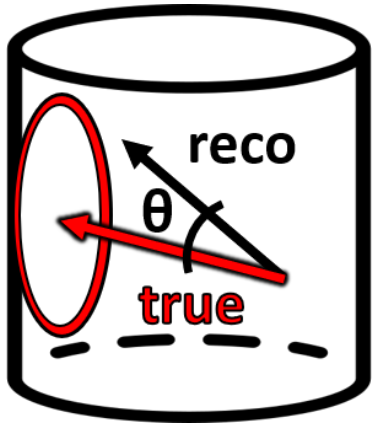




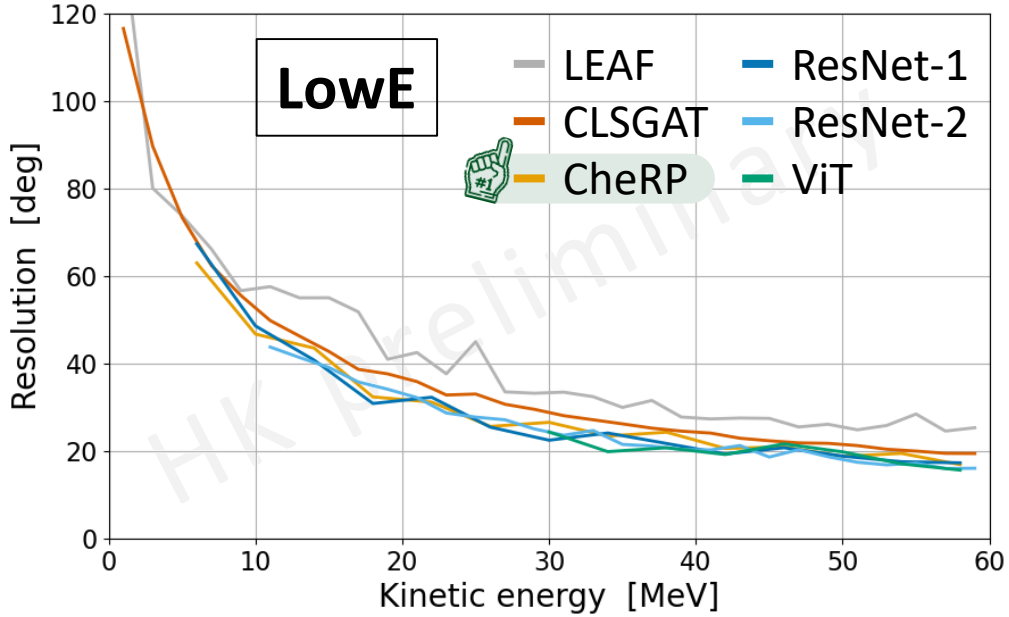
⇒ Up to 60cm improvement for E<6MeV with GoF>0.3

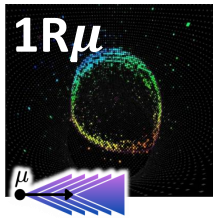
DIRECTION @CL95%

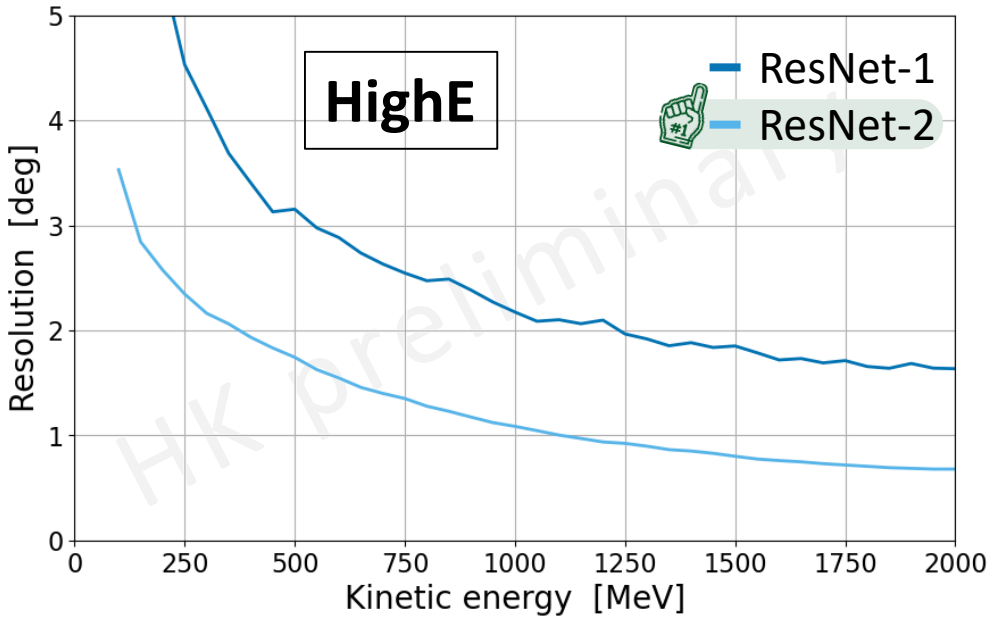
θ CI 95%

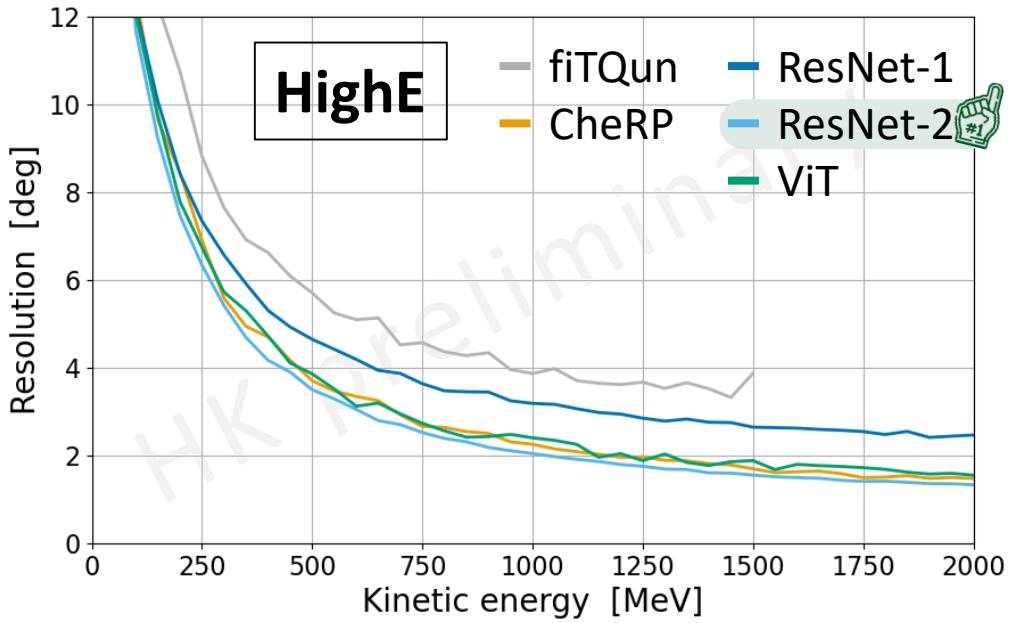


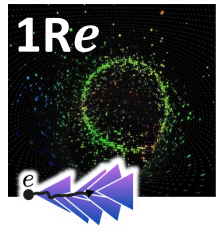
*fiTQun results presented here are preliminary.





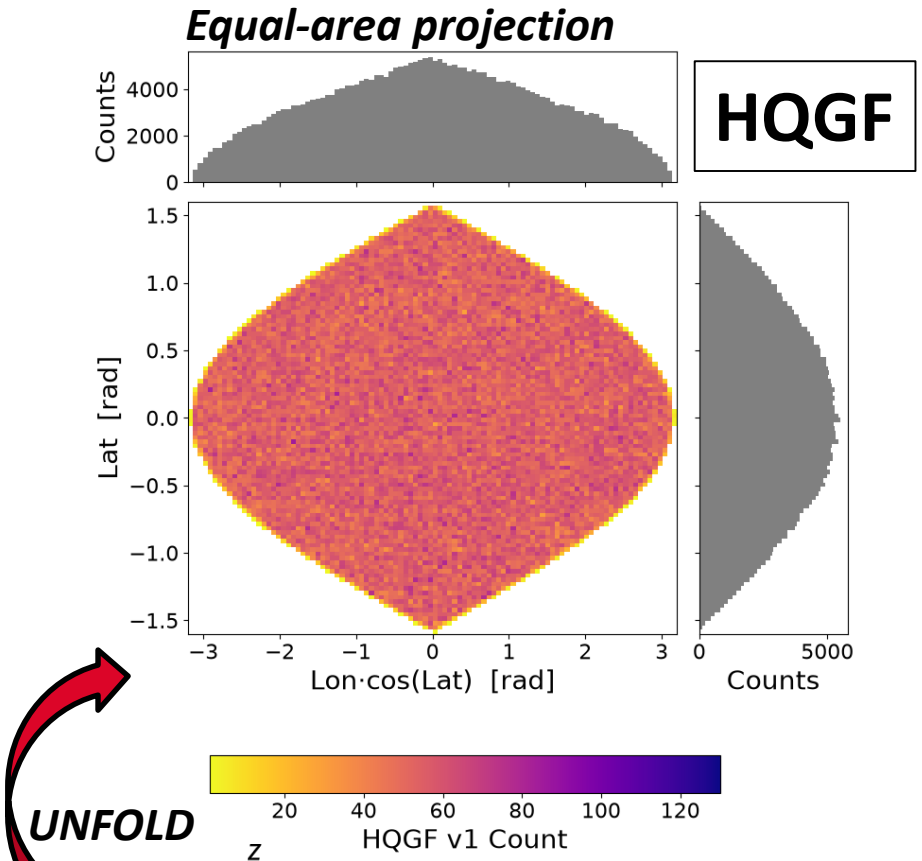


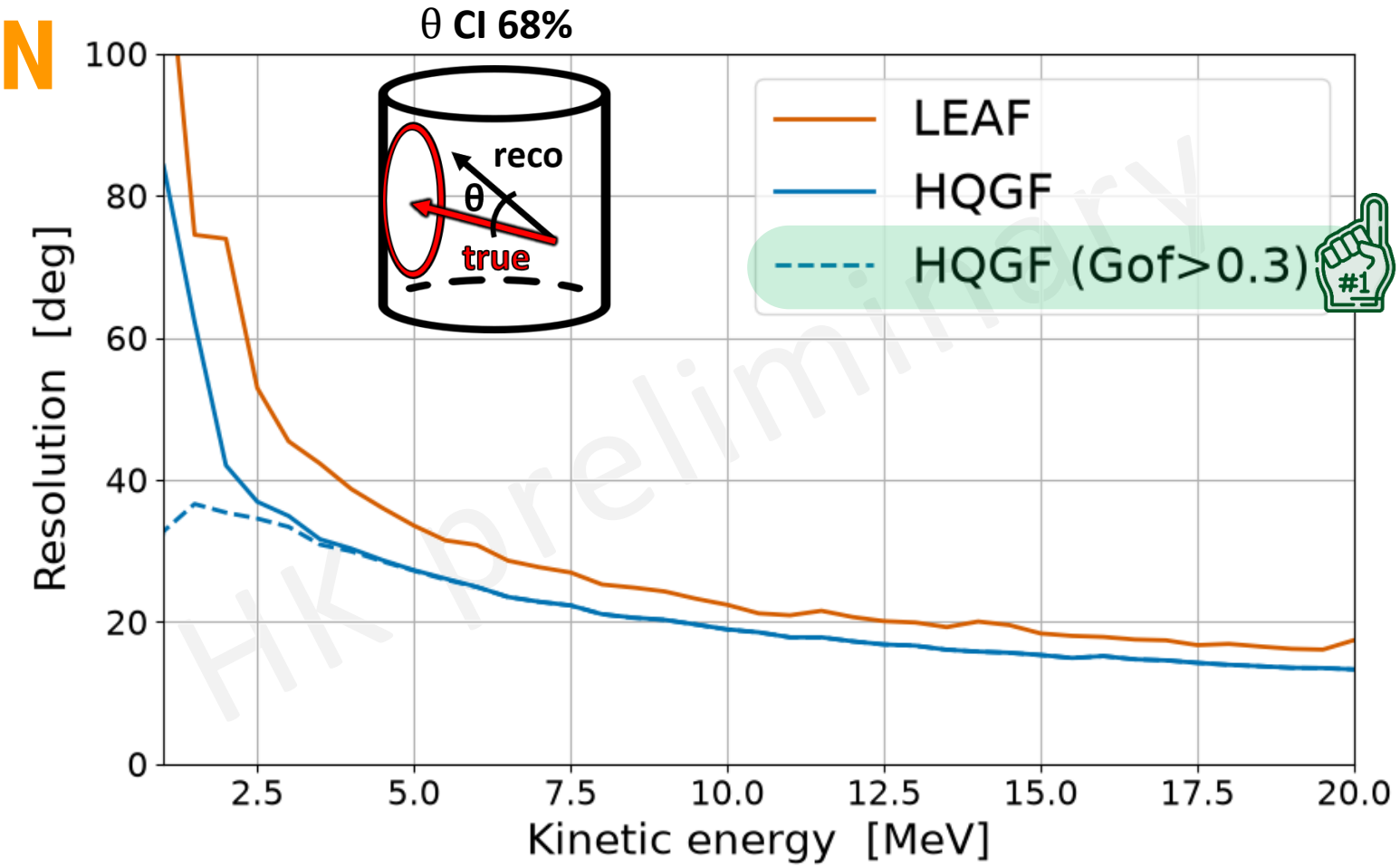




⇒ **ResNet-2 better for direction reconstruction of electron events when looking at CI95%**

DIRECTION RESOLUTION





Direction resolution

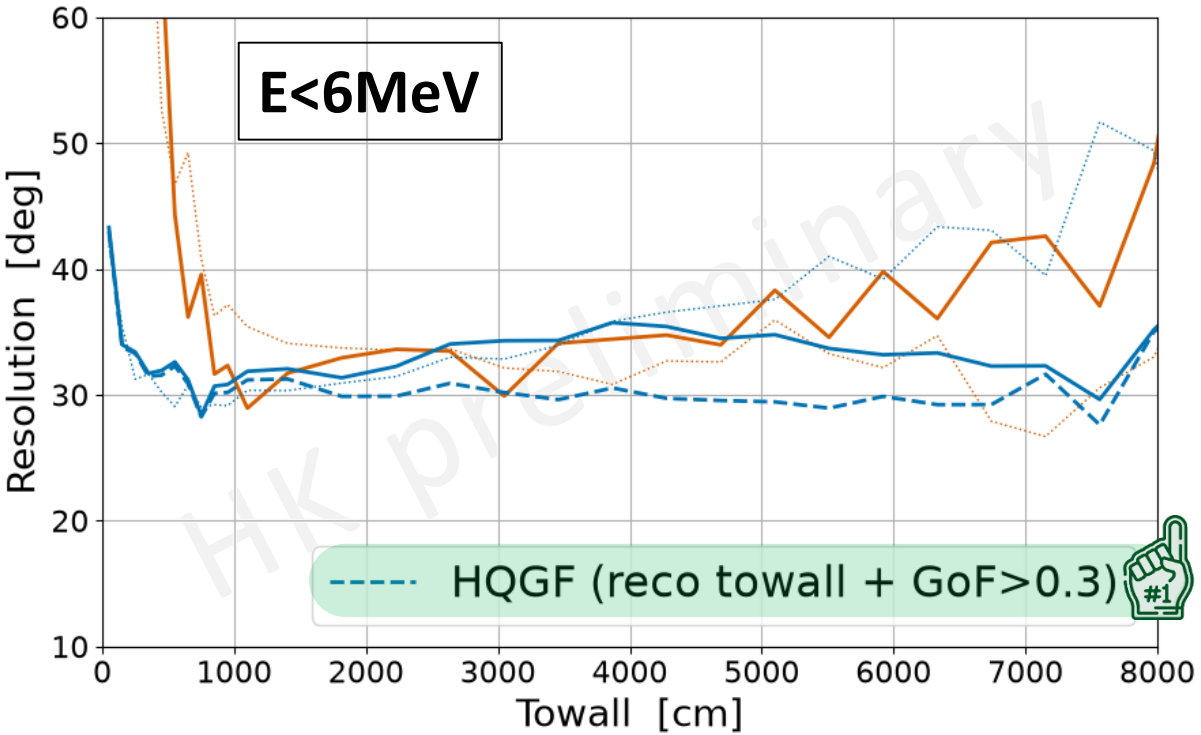
35° @2 MeV

19° @10 MeV

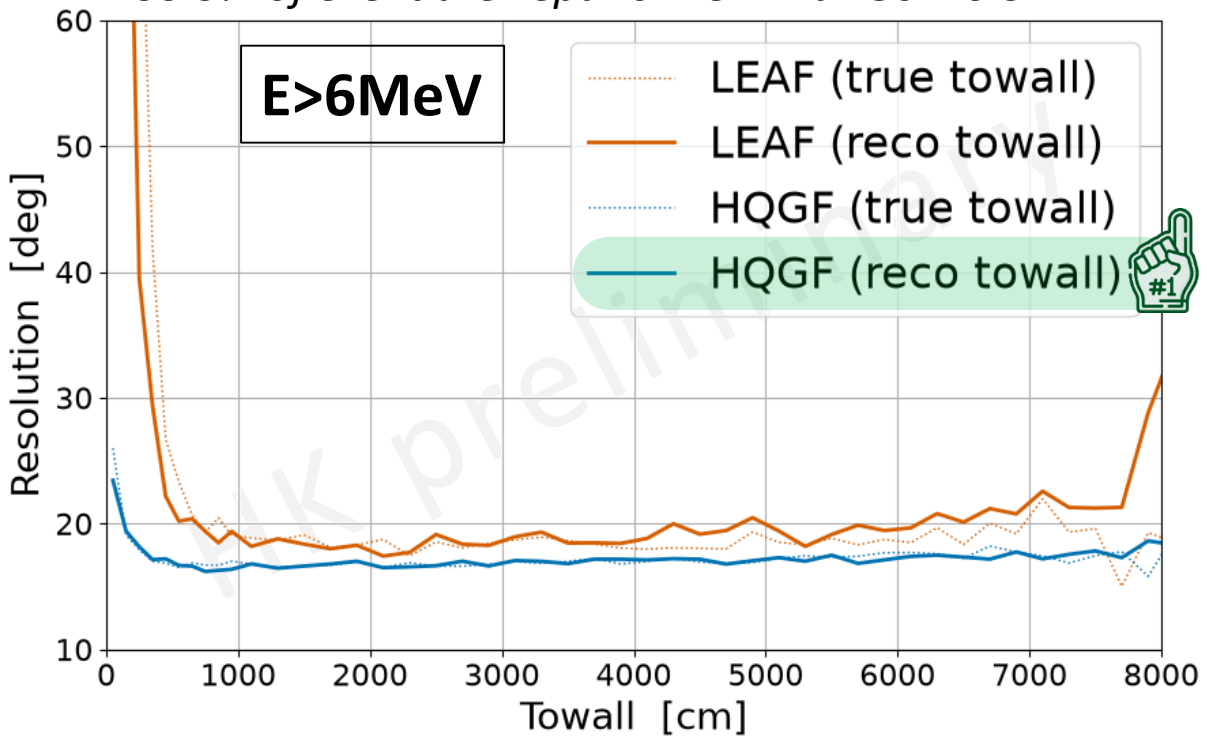
⇒ Reconstruction improved ∀E

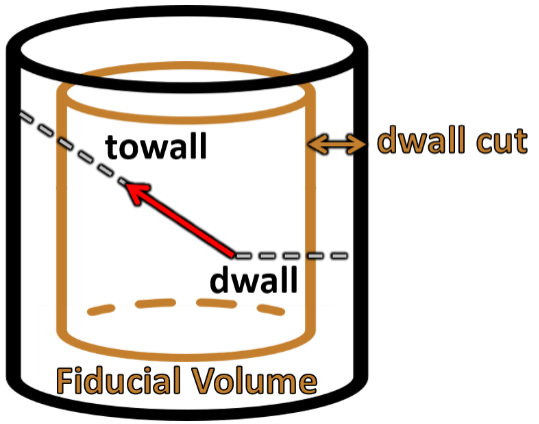
GoF cut keeps 70% of events @2 MeV

TOWALL DEPENDENCY



99.9% of event are kept >6 MeV with GoF>0.3





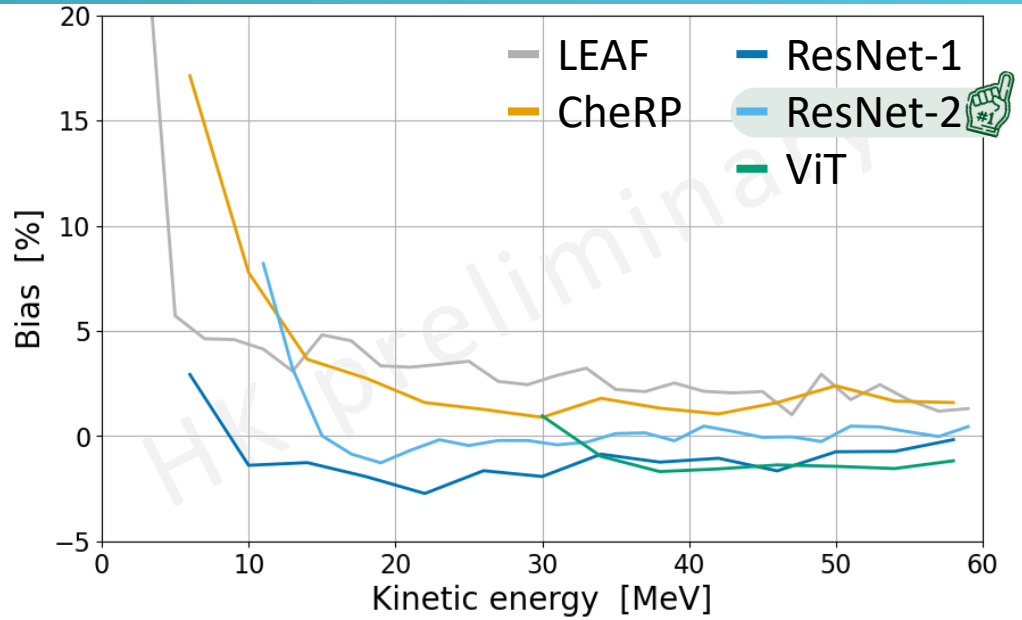
⇒ Up to 10° improvement for E<6MeV with GoF>0.3

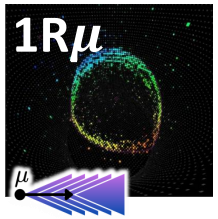
ENERGY BIAS

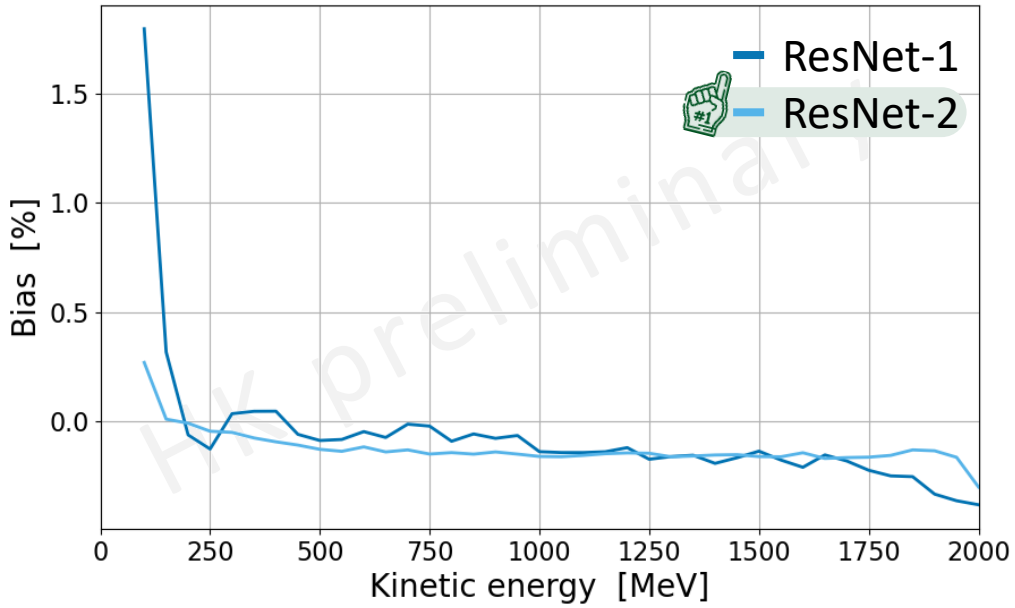
$\Delta E/E$ mean

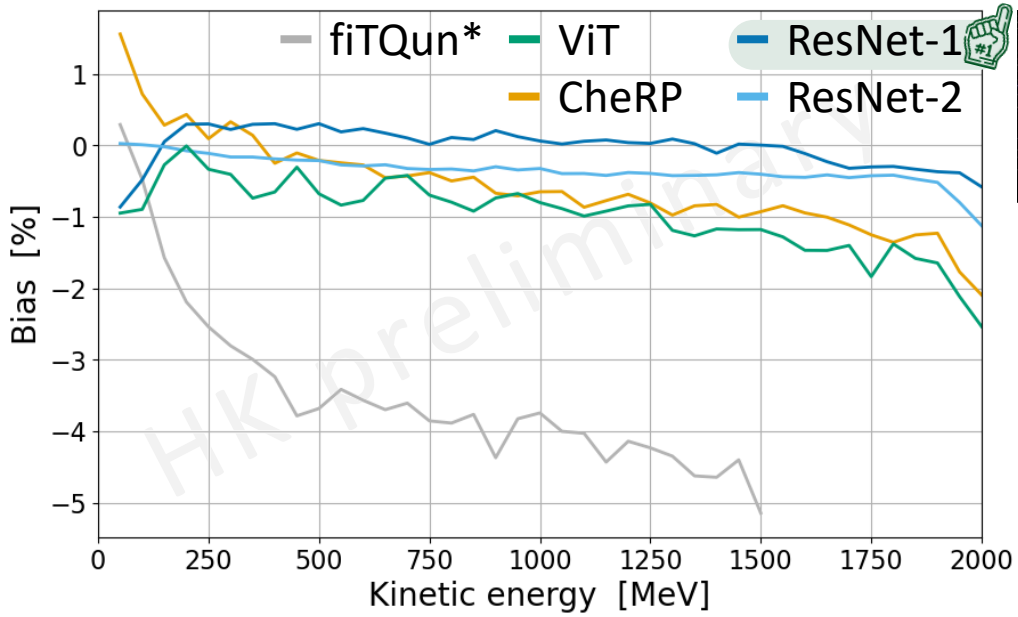


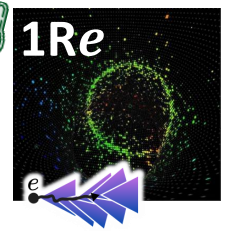
*fiTQun results presented here are preliminary.









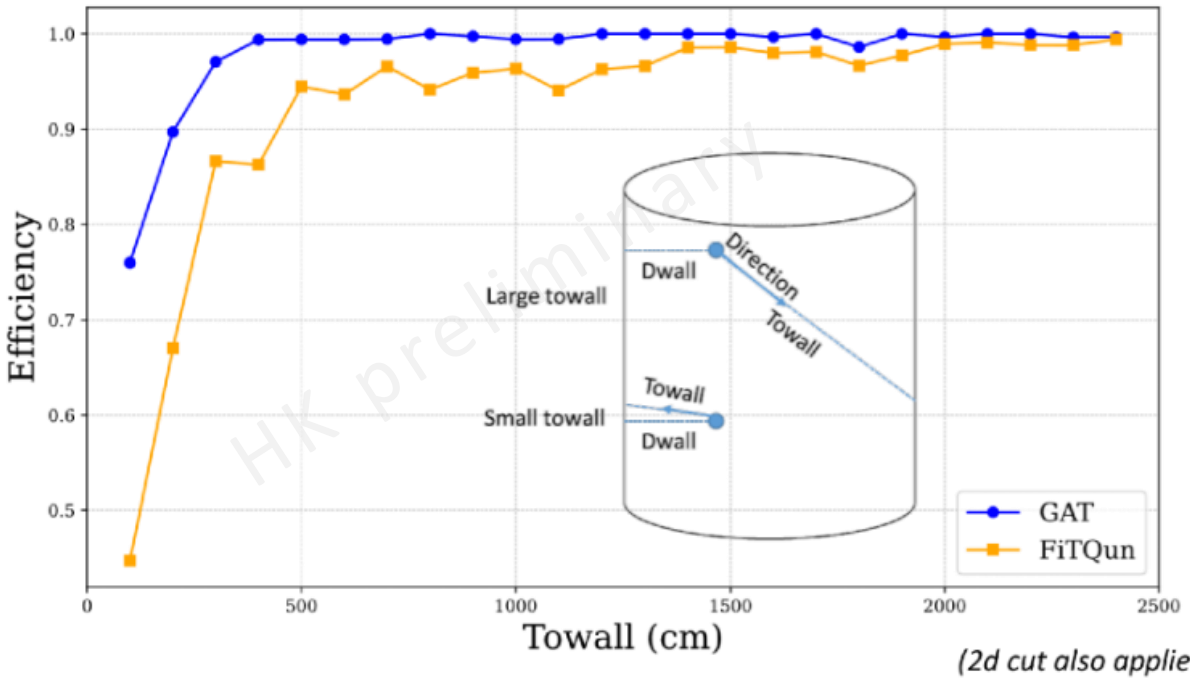


⇒ **ResNet-1 & 2 better for momentum reconstruction when looking at remaining bias**

PARTICLE CLASSIFICATION

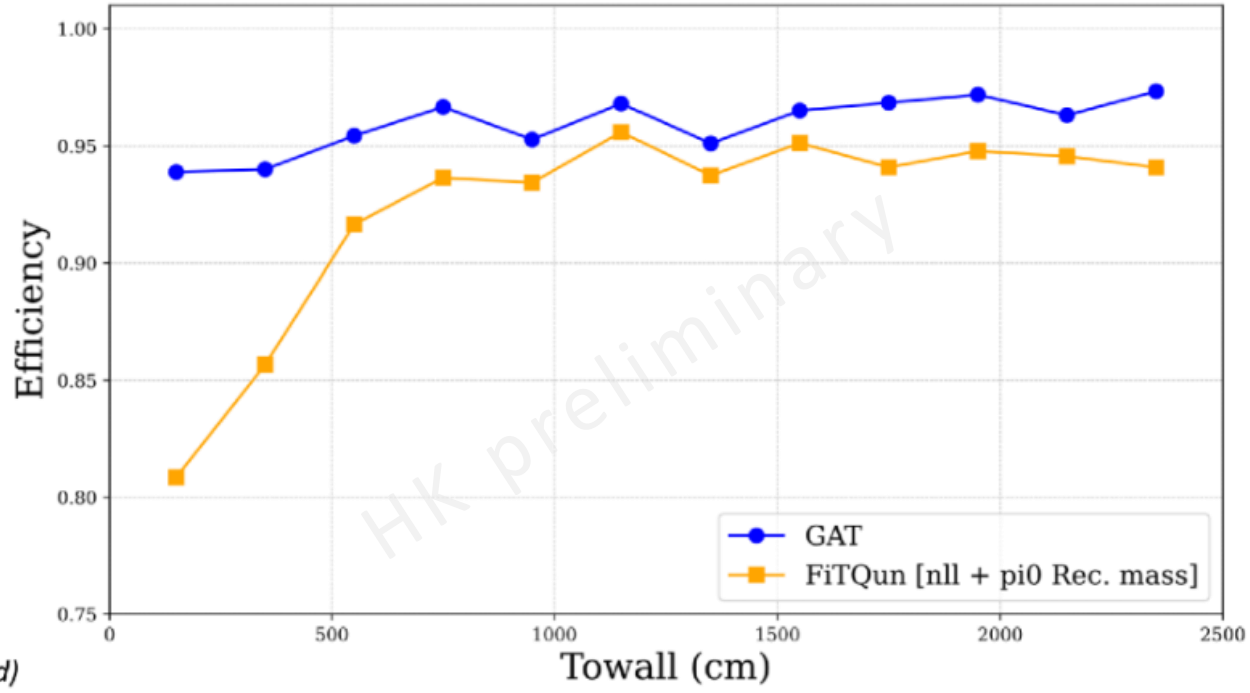
Erwan Le Blevec
GAT

Electron & Muon



(2d cut also applied)

Electron & π^0



HYPER-K SYSTEMATIC REQUIREMENT

Systematic Source	Required Precision	For Which Measurement	Detector	Achievable Precision
$\sigma(\nu_e)/\sigma(\nu_\mu)$	3-5%	CP Violation, δ_{CP} precision at $\sin(\delta_{cp})\sim0$, θ_{23} precision at $\sin(\theta_{23})\sim0.5$	IWCD	3.5-5%
$\sigma(\bar{\nu}_e)/\sigma(\bar{\nu}_\mu)$	3-5%	CP Violation, δ_{CP} precision at $\sin(\delta_{cp})\sim0$, θ_{23} precision at $\sin(\theta_{23})\sim0.5$	IWCD	4-7%
Wrong-sign background normalization	9%	CP Violation, δ_{CP} precision at $\sin(\delta_{cp})\sim0$	ND280	TBD (expect <9%)
Intrinsic ν_e, $\bar{\nu}_e$ and NC backgrounds	3-4%	CP Violation, δ_{CP} precision at $\sin(\delta_{cp})\sim0$	IWCD	2.3% (neutrino)
Normalization of non-QE with E_ν>0.7 GeV	5%	θ_{23} precision at $\sin(\theta_{23})\neq0.5$	IWCD	5% (neutrino)
Normalization of non-QE with all energies	5%	δ_{CP} precision at $\sin(\delta_{cp})\sim0$ Δm^2_{32} precision	IWCD, ND280*	5% (IWCD neutrino) <4% (N280 neutrino) <7% (ND280 antineutrino)
Beam Direction	0.6 mrad (4 MeV shift)	δ_{CP} precision at $\sin(\delta_{cp})\sim0$ Δm^2_{32} precision	INGRID	<0.3 mrad (<2 MeV)
Removal (binding) energy	4 MeV*	δ_{CP} precision at $\sin(\delta_{cp})\sim0$ Δm^2_{32} precision	IWCD, ND280	2.6 MeV (IWCD on O) ~1 MeV (ND280 on C)**
High angle measurement (cosθ<0.2)	4%	CP Violation, δ_{CP} precision at $\sin(\delta_{cp})\sim0$	IWCD, ND280	<4% statistical precision in both detectors
Beam rate monitoring	~1% per day	General monitoring of beam quality	INGRID	<0.5% per day for neutrinos and antineutrinos
Neutron Multiplicity	TBD	Atmospheric neutrino Nucleon decay	IWCD, ND280	<5% IWCD <4% ND280
$\mu\pi^0$ cross section & neutron multiplicity	TBD	$e\pi^0$ proton decay	IWCD	TBD

⇒ IWCD is critical to control many important systematic σ in HK

***Complementary approaches in IWCD and ND280.**
IWCD relies on flux model in linear combination method, but minimizes Xsec model dependence.
ND280 fits transverse variables to constrain Xsec model.

***Energy scale in detectors must be calibrated to 0.5% to achieve this level**

⇒ HK total CPV syst. σ should be reduced <3%