Event Builder for EICRecon

Streaming Readout Workshop

August 19, 2026

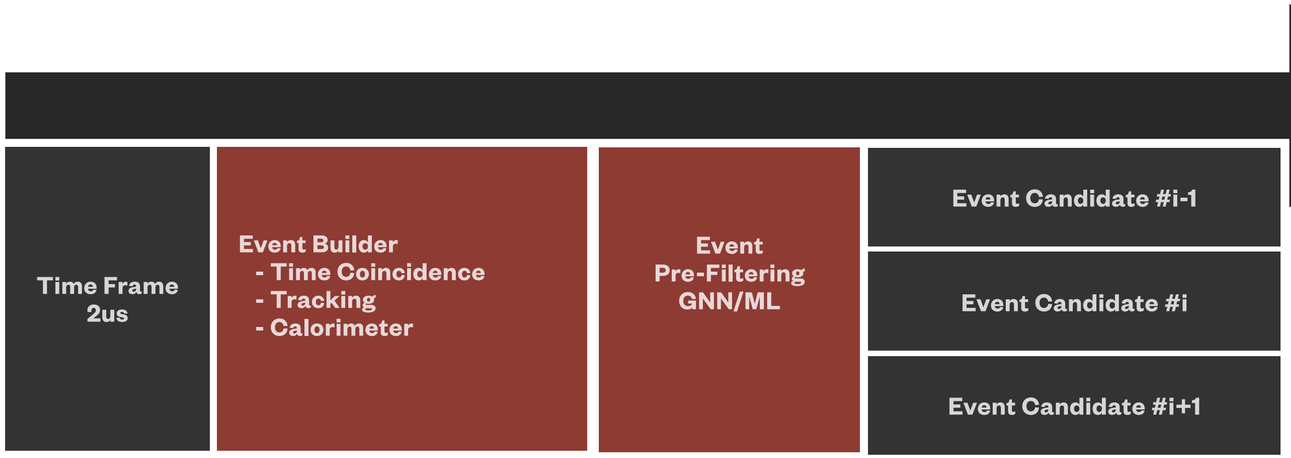
Oliver Partridge, Summer Master Student

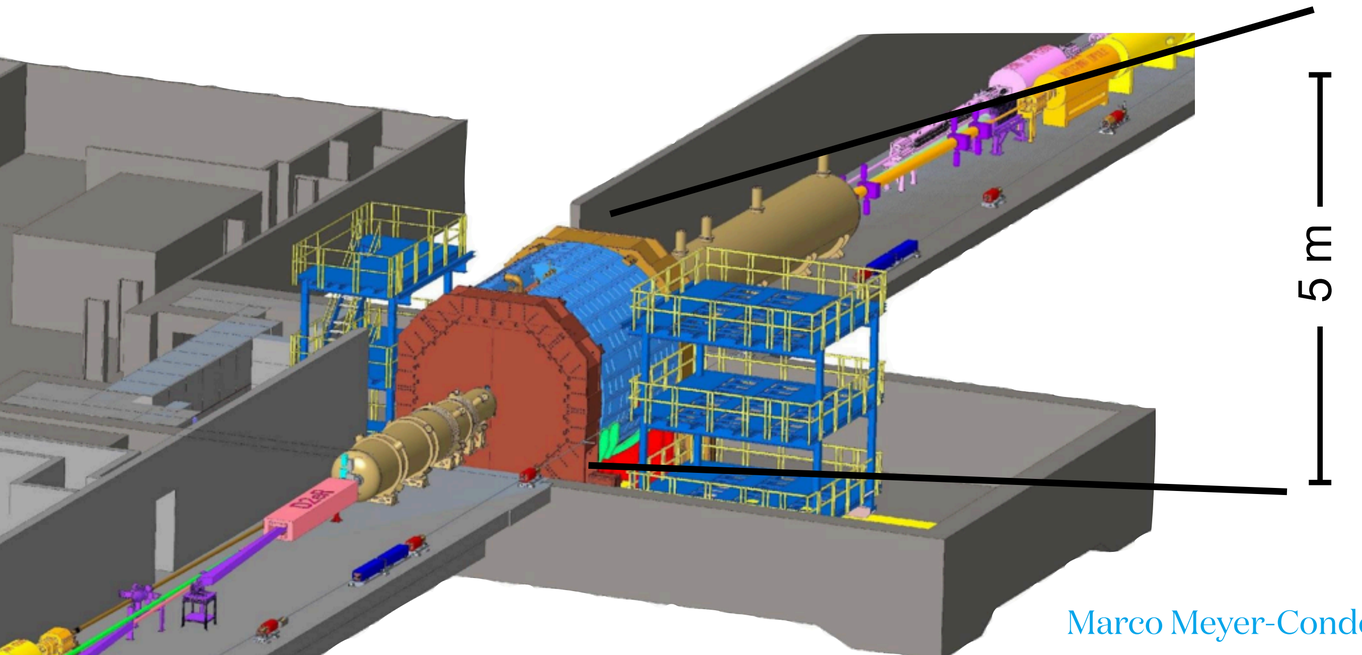
Marco Meyer-Conde, Assistant Professor

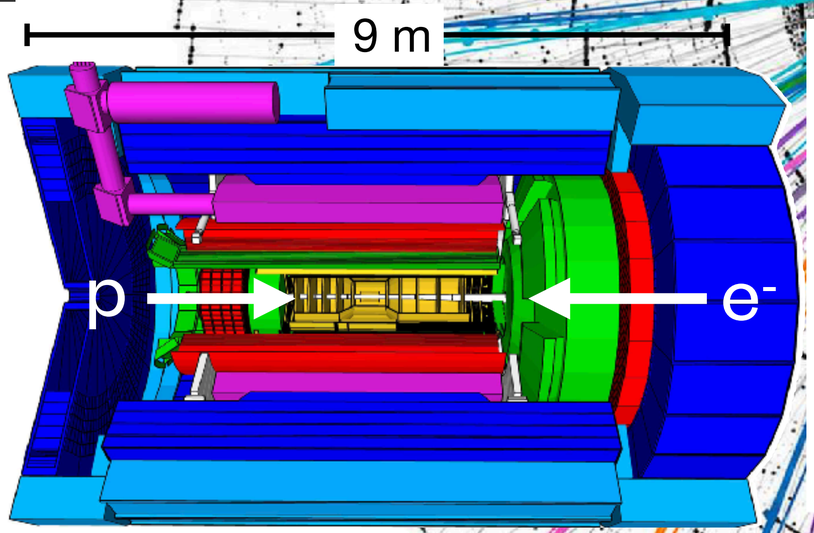
Advanced Research Laboratories, Tokyo City University

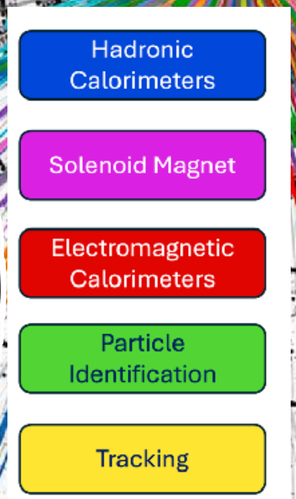
Electron-Ion Collider

Streaming Readout — Event Builder & Graph Neural Network









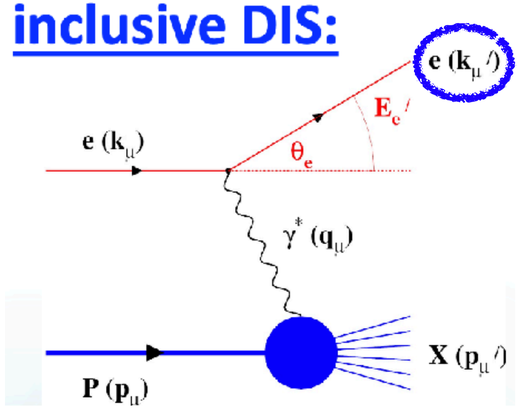
Marco Meyer-Conde/Oliver Pa

Img from DANIEL MURNANE

Coverage of ePIC Detector

Rapidity VS. Q^2

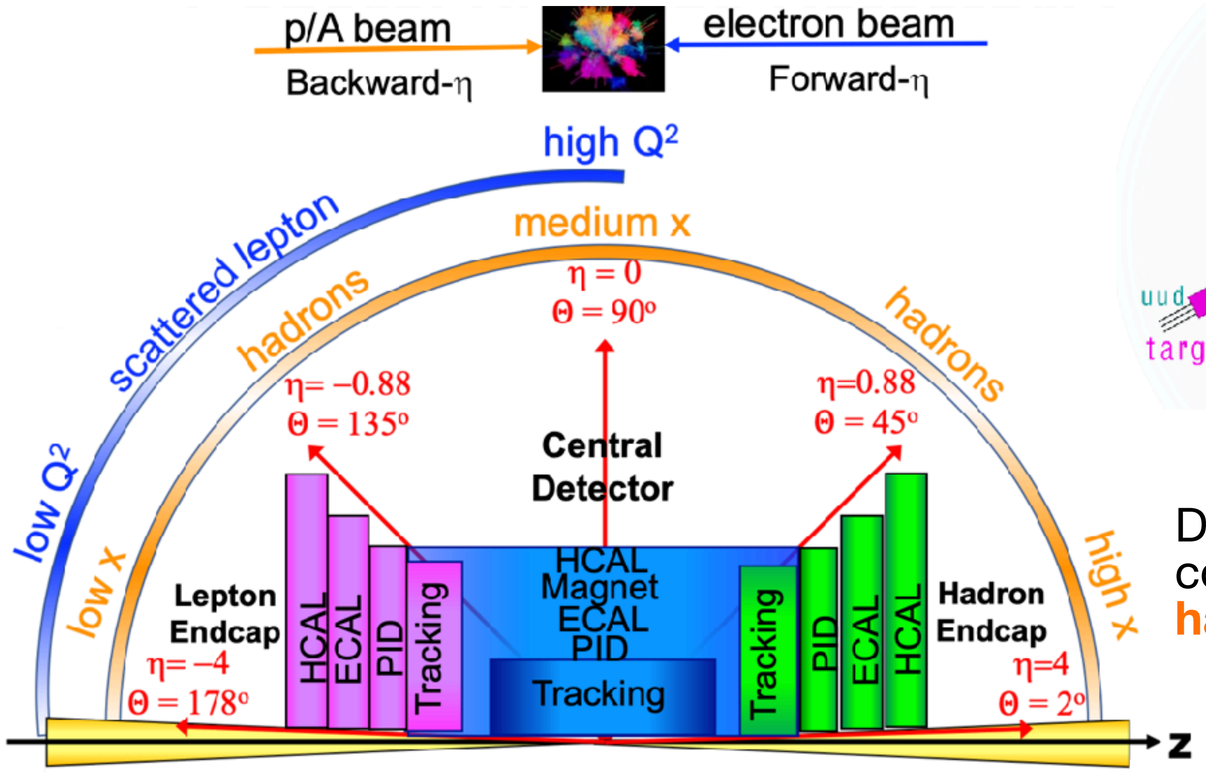
inclusive DIS:



Detect only the **scattered lepton** in the detector

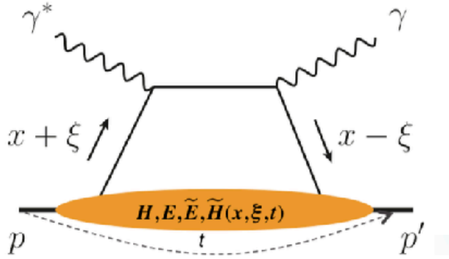
Very low Q^2 scattered lepton Bethe-Heitler photons for luminosity

Luminosity Detector
Low Q^2-Tagger

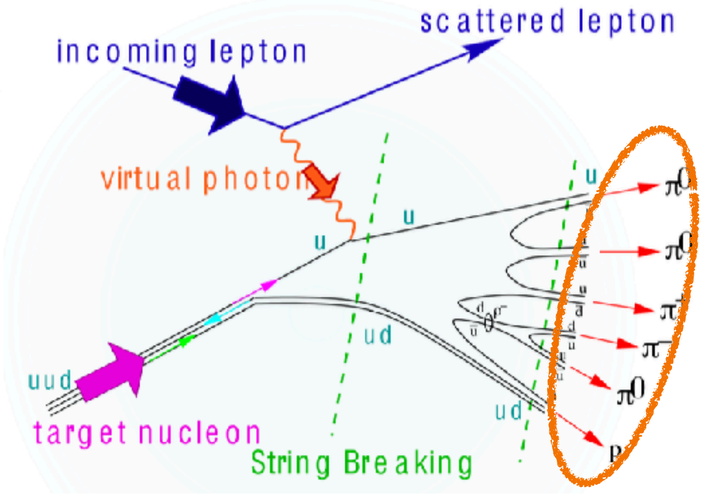


exclusive DIS

Detect the scattered lepton, identify produced hadron/jets and **target remnants**



semi-inclusive DIS



Detect the scattered lepton in coincidence with **identified hadrons/jets**

Particles from nuclear breakup and from diffractive reactions

ZDC

Forward Tracking

EICRecon Software

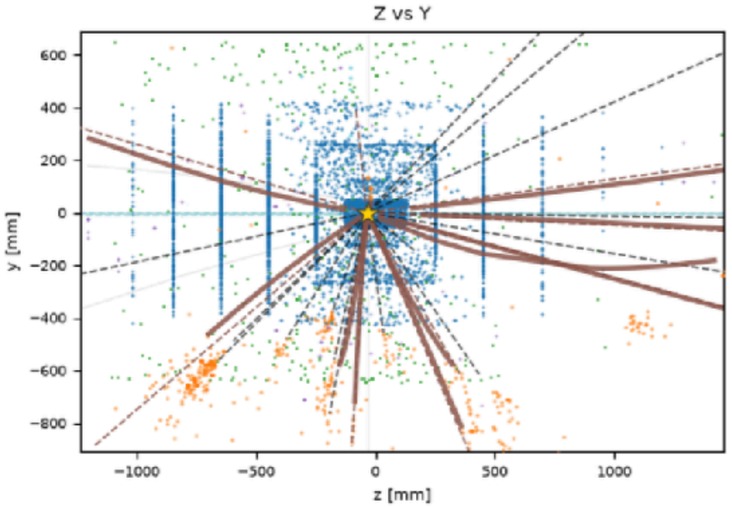
What problem are we trying to solve?

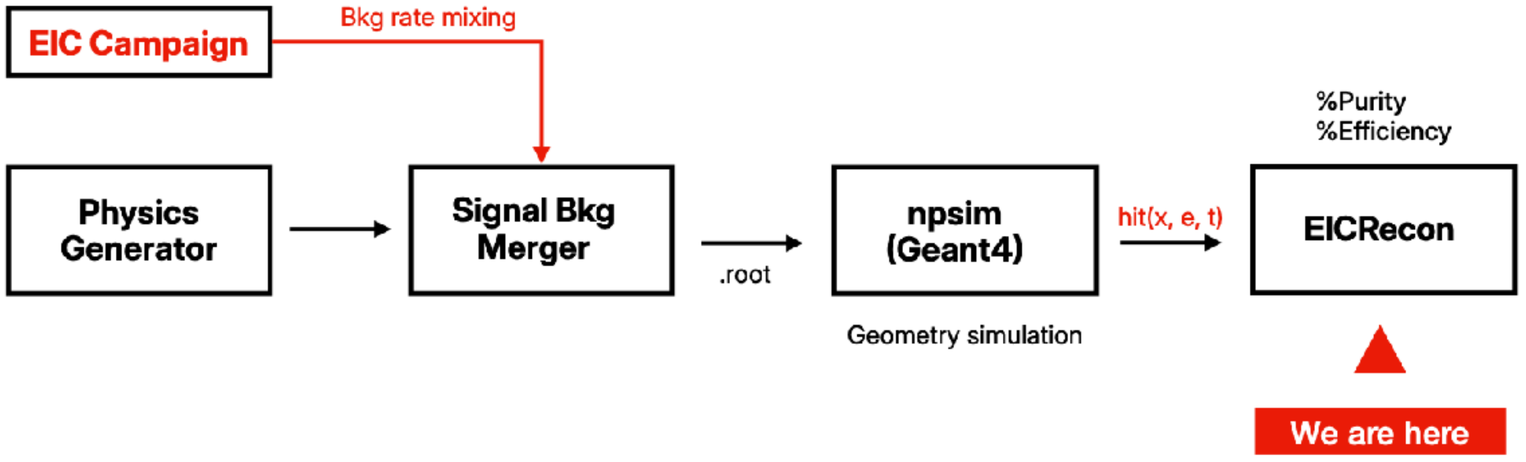
No traditional trigger detection —> continuous readout in DAQ pipeline

ACTS Reconstructions are heavy on compute

EventBuilder Provides Event Candidate with 100% efficiency but low purity

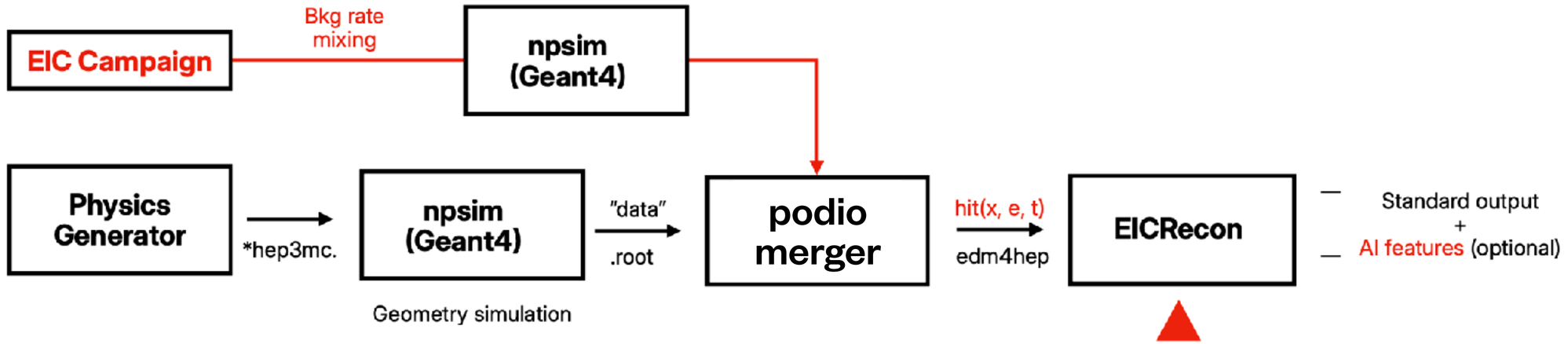
The model SRO-GNN's goal is to improve that purity to avoid unnecessary ACTS Combinatorial Kalman Filter reconstruction





Implementation in EICRecon

From TimeFrame to PhysicsEvent



Important considerations:

- High rate background
- In-Acceptance
- Piled-up Events
- Time-structured spills (beam bunch) currently flatly distributed

```
eicrecon \
  -Peventbuilder=true -Pplugins=eventbuilder \
  # eventbuilder:topology:* (finding-stage grid)
  # eventbuilder:trigger:* (adaptive fake-rate mode) --
  # eventbuilder:tracklet:* (vertex-pointing score, Sec. 4) --
  # eventbuilder:unfolder:* (per-candidate gate widths) --
  # eventbuilder:statistics:* (S/FAR null model / the Fano correction)
  # eventbuilder:findable:* (YR truth-acceptance cuts, offline use)

  [..]
```

Naming Convention

RawHit / RecHit / Cluster naming convention

Suffix	Meaning
*SimHits	Raw Geant4 truth hits (edm4hep::SimTrackerHit/SimCalorimeterHit), pre-digitization
*RawHitLinks	Same idea, via the newer templated podio::Link<A,B> mechanism between Rec/Raw Hits
*RawHits (via *RawHitDigi)	Digitized detector response — ADC counts, TDC bins — no physics calibration yet (edm4eic::RawTrackerHit/RawCalorimeterHit)
*RecHits	Calibrated reconstructed hits — position (x,y,z), time (ns), energy (GeV) (edm4eic::TrackerHit/CalorimeterHit)
*Clusters	Calorimeter-only: RecHits aggregated into a shower object (summed energy, weighted centroid, time; edm4eic::Cluster)
*ClusterLinks	Cluster ↔ truth particle

Implementation in EICRecon

TimeFrame → PhysicsEvent

(1) Timing Alignment: $t_i = | t_i' - r/c |$

(2) Trigger Candidate:
- Gated Fast hits (MPGD+TOF): $3\sigma_{MPGD+TOF}$

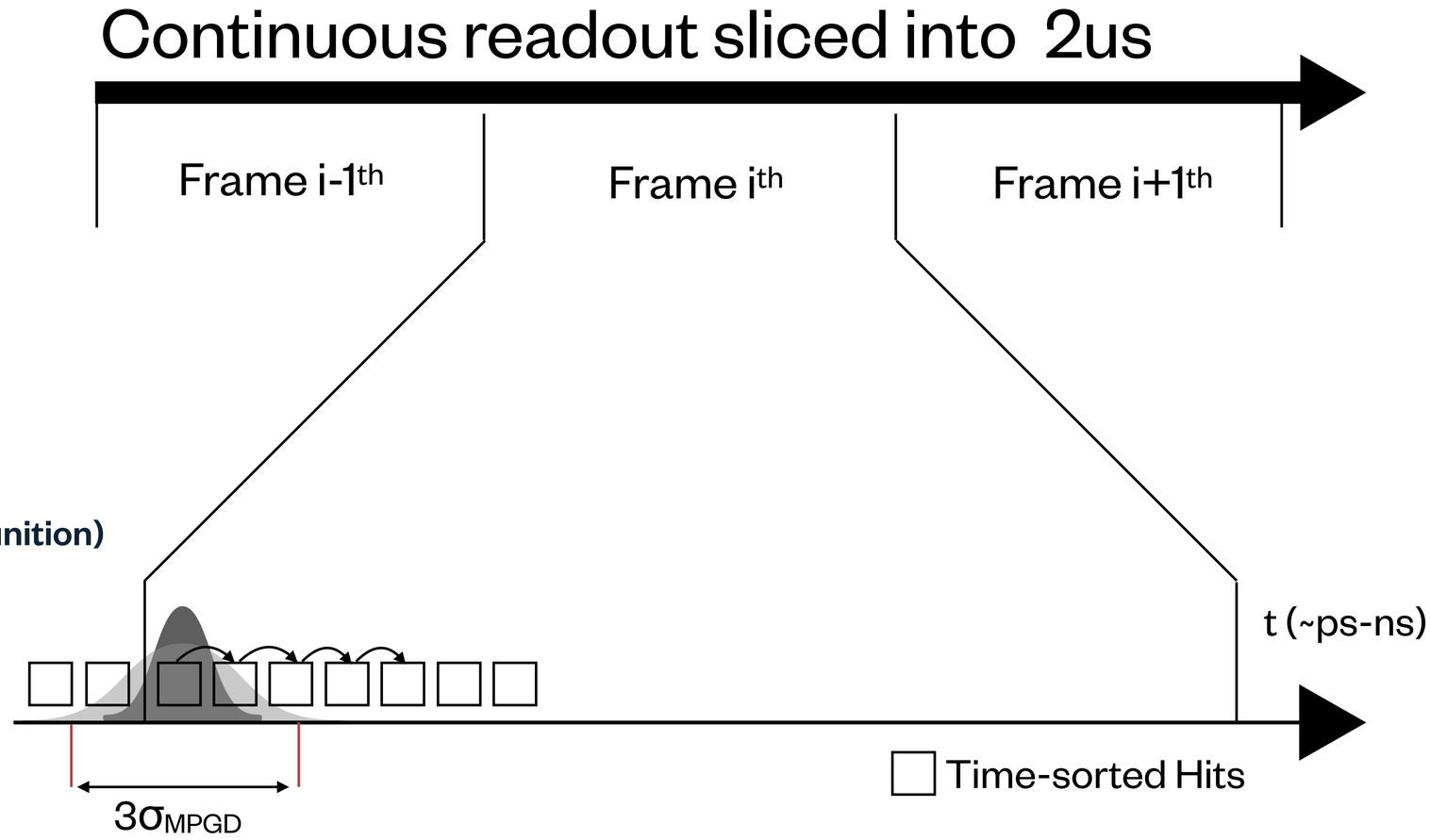
(3) Timing Coincidence:
- Per-Detector Gating: $3\sigma_i$

(σ_i being i-th detector res. => no pure time window definition)

(4) Tracklet Coincidence:
- Pointing toward IP

(5) Calorimeter Energy threshold
- Energy coincidence within sigma t0

(attempt to introduce time error VS. E_{hit})



Trigger Candidate

$$t_0 = \frac{\sum_i w_i t_i}{\sum_i w_i}, \qquad \sigma_{t_0} = \left(\sum_i \frac{1}{\sigma_i^2} \right)^{-1/2}$$

$$w_i = \frac{1}{\sigma_i^2},$$

- t_i: corrected timestamp of hit i,
- σ_i: time resolution of detector i.

Timing Coincidence

$$|t_i - t_0| < n\,\sigma_i$$

where:

- t_i is the corrected hit or cluster time,
- t_0 is the candidate event-time estimate,
- σ_i is the hit or cluster time error,
- n is the configured number of standard deviations.

Tracking Coincidence

Geometrical Alignment

MPGD Inner↔TOF

TOF ↔MPGD Outer

Energy Coincidence

$$|t_{clu} - t_0| < n\,\sigma_{t,cal}(E_{clu})$$

$$\sigma_{clu} = \frac{\sqrt{\sum_i \left(\sigma_{hit,i}\,E_{hit,i}\right)^2}}{\sum_i E_{hit,i}}$$

$$\sigma_{hit,i} = \sqrt{\frac{a^2}{E_{hit,i}} + b^2}$$

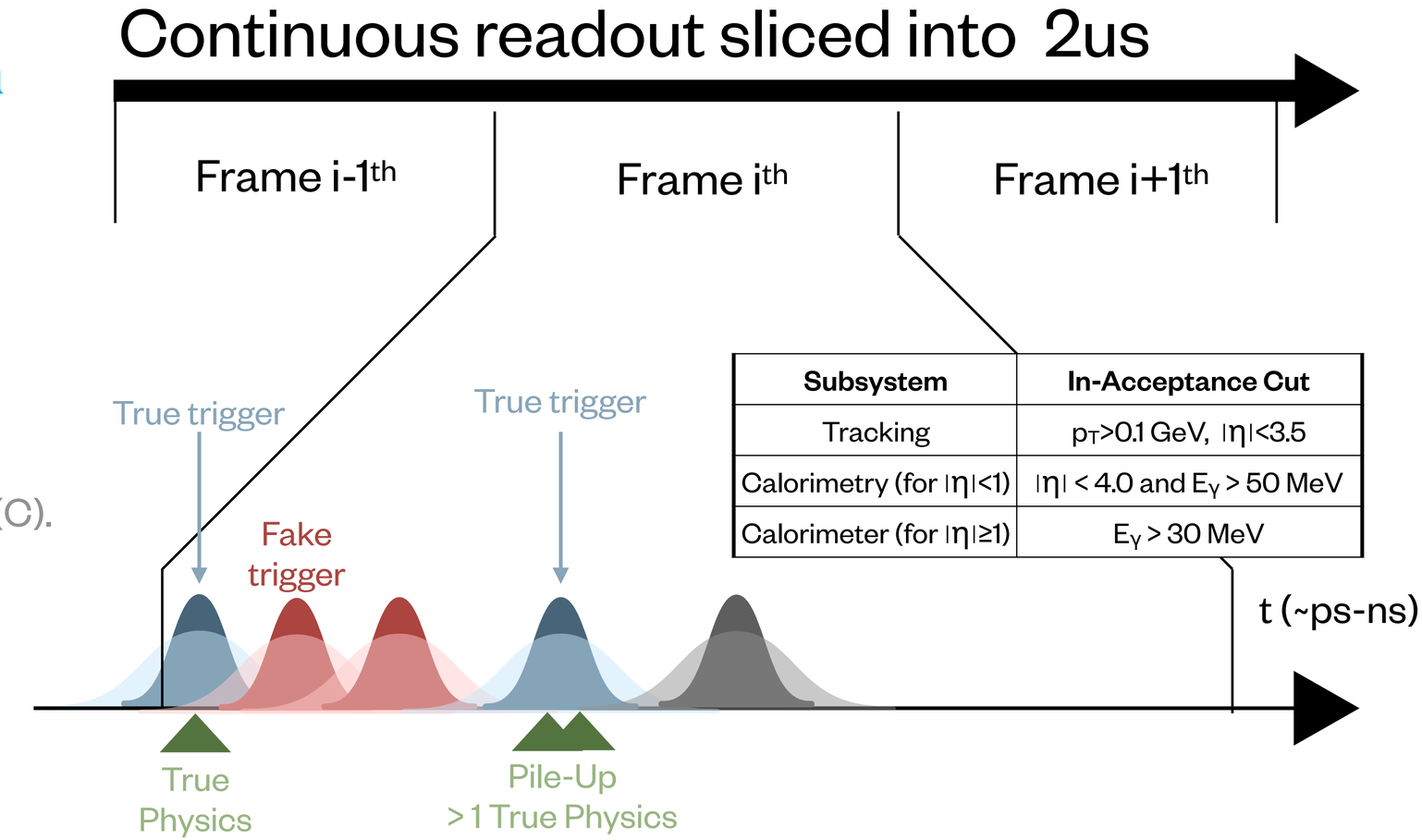
Implementation in EICRecon

Software Trigger Decision

(1) Trigger Design:
- Time Aligned Hits
- Exclusive MPGD+TOF hits
- Slicing window σ_{MPGD} (avoid over triggering)
  (A,B,C,D) => (A,B) within σ_{MPGD} ; Restart from (C).

(2) Event Candidate Filtering:
- Per-Detector Hit Filtering
- Time, Tracking, Calorimeter Coincidence

(3) Event Unfolding:
- All hits may be considered multiple times



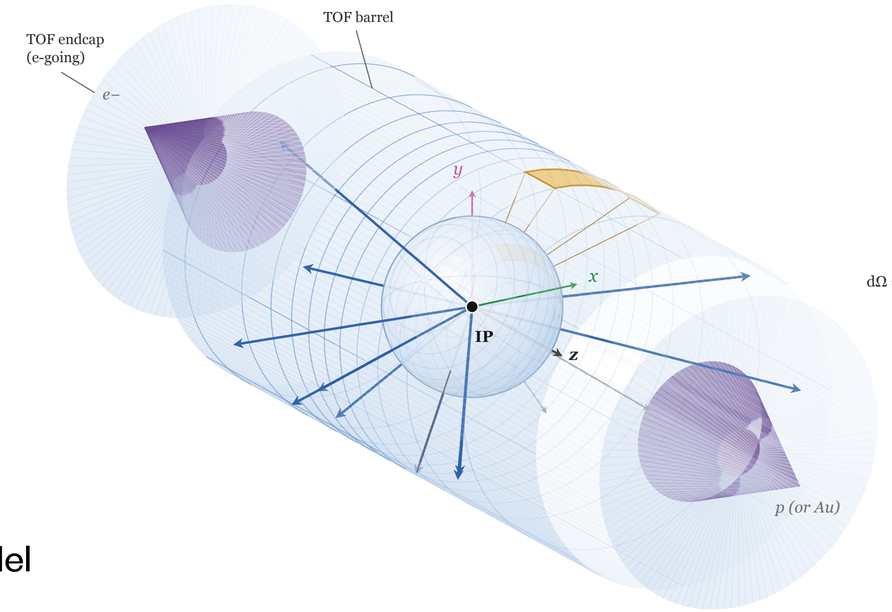
Subsystem	In-Acceptance Cut
Tracking	p_T>0.1 GeV, $\lvert\eta\rvert$<3.5
Calorimetry (for $\lvert\eta\rvert$<1)	$\lvert\eta\rvert$ < 4.0 and E_γ > 50 MeV
Calorimeter (for $\lvert\eta\rvert\geq$1)	E_γ > 30 MeV

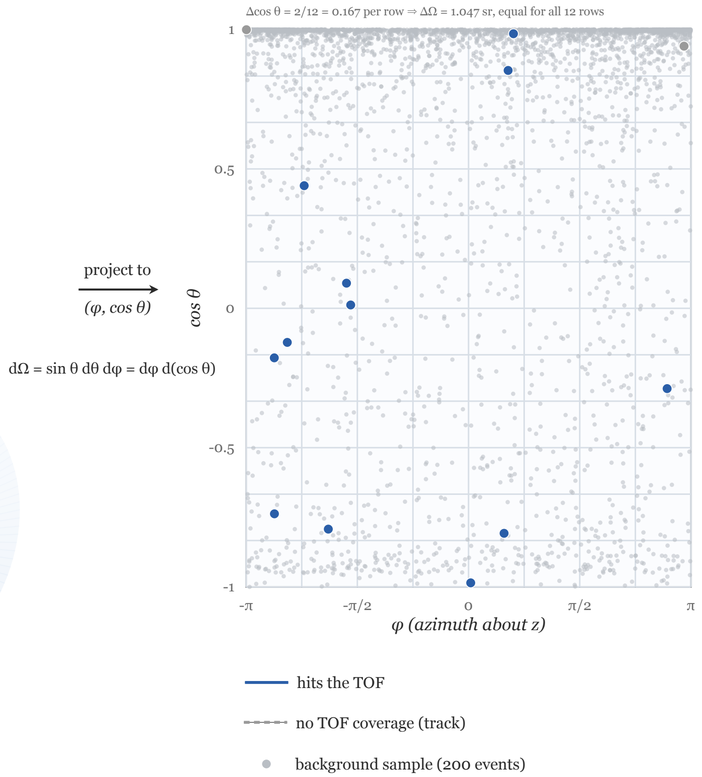
Timing Coincidence	Tracklet Pointing	Calorimeter Energy Threshold	Software Trigger
$\begin{cases} \%^{\text{timing}}_{\text{efficiency}} = 100\,\dfrac{N_{\text{real}}}{N_{\text{inj}}} \\ \%^{\text{timing}}_{\text{purity}} = 100\,\dfrac{N_{\text{real}}}{N_{\text{real}}+N_{\text{fake}}} \end{cases}$	$\begin{cases} \%^{\text{tracking}}_{\text{efficiency}} = 100\,\dfrac{N^{\text{trk}}_{\text{true}}}{N^{\text{trk}}_{\text{exp}}} \\ \%^{\text{tracking}}_{\text{purity}} = 100\,\dfrac{N^{\text{trk}}_{\text{true}}}{N^{\text{trk}}_{\text{true}}+N^{\text{trk}}_{\text{ghost}}} \end{cases}$	$\begin{cases} \%^{\text{caloing}}_{\text{efficiency}} = 100\,\dfrac{N^{\text{cal}}_{\text{true}}}{N^{\text{cal}}_{\text{exp}}} \\ \%^{\text{caloing}}_{\text{purity}} = 100\,\dfrac{N^{\text{cal}}_{\text{true}}}{N^{\text{cal}}_{\text{true}}+N^{\text{cal}}_{\text{ghost}}} \end{cases}$	$\begin{cases} \%^{\text{trigger}}_{\text{efficiency}} = 100\,\dfrac{N_{\text{valid}}}{N_{\text{real}}} \\ \%^{\text{trigger}}_{\text{purity}} = 100\,\dfrac{N_{\text{valid}}}{N_{\text{valid}}+N_{\text{fake}}} \end{cases}$

Software Trigger = Timing Coincidence & (Tracklet Alignment || Calorimeter Clustering)

Implementation in EICRecon

Statistical Event Significance





Probability Law: **primary physics is poissonian**

$X_i \sim \text{Poisson}(\mu_i)$

μ_i = background occupancy per bin

(Best option: $(N_{\cos\theta}, N_\phi) = (1,1)$ — no spatial binning)

Motivations

- Additional per-event information to feed the ML model
- Defines False Alarm Rate "FAR" and Significance "S" as a probabilistic score.

Definitions

$$\text{FAR} = \frac{N_{\text{fake}}}{N_{\text{frames}}} \times \frac{1}{T_{\text{frame}}}\ [\text{Hz}], \quad T_{\text{frame}} = 2\mu s \qquad S = -\,10\log_{10}(p_{\text{tail}}) \quad \text{with} \quad p_{\text{tail}} = P(X \geq c_{\max} \mid \mu_i)$$

Caveats

- Anisotropic background distribution
- Overdispersed "Poissonian" law (primary physics is poissonian —> hit distribution is multiple)
  Fano factor $F = \sigma^2/\mu > 1$ (correction to be applied)

EIC Campaign

Signal and Background Mixing

- Machine background (expressed in kHz): https://github.com/eic/simulation_campaign_datasets/tree/main/config_data
- Physics signals: https://github.com/eic/simulation_campaign_datasets

Source	Status	Vacuum	Gold Coating	File size	Effect
Synchrontron Radiation	2000	36.608 GHz	13.277 GHz	~78 MB	1–2 photons per event; unique templates reused 10 000×
Electron bremsstrahlung radiation	3000	3.177 MHz	3.177 MHz	~94 MB	Few-particle QED elastic scatter
Electron Touscheck scattering	4000	29.56 kHz	29.56 kHz	~525 MB	Intra-bunch scatter Slightly more complex final state
Electron Coulomb scattering processes	5000	233.5 kHz	233.5 kHz	~6.2 GB	Full DIS final state with hadronisation
Proton beam gas interactions	6000	31.9 kHz	31.9 kHz	~16.7 GB	Soft/hard hadronic production Many secondary tracks

EIC Campaign

How data get wrapped into a simulation stream?

- Bkg + Events #1...N using generatedStatus
- Event level information using Physics Event weight
- Frame level information using EventBuilderFrameInfo
- AI ground truth encoded using generatedStatus in *HitLinks for segmentation

Generated Status Range	Meaning
0-99999	bkg/machine-background (unchanged — 10×1000-wide sub-bands: SynRad, e-beam-gas/brems, Coulomb, Touschek, p-beam-gas, plus headroom for a future 6th+ source)
100000-199999	ncdisq1
200000-299999	ncdisq10
....	....

Issue being discussed on GitHub:
https://github.com/AIDASoft/DD4hep/issues/1657#issuecomment-5280917973

EIC Campaign

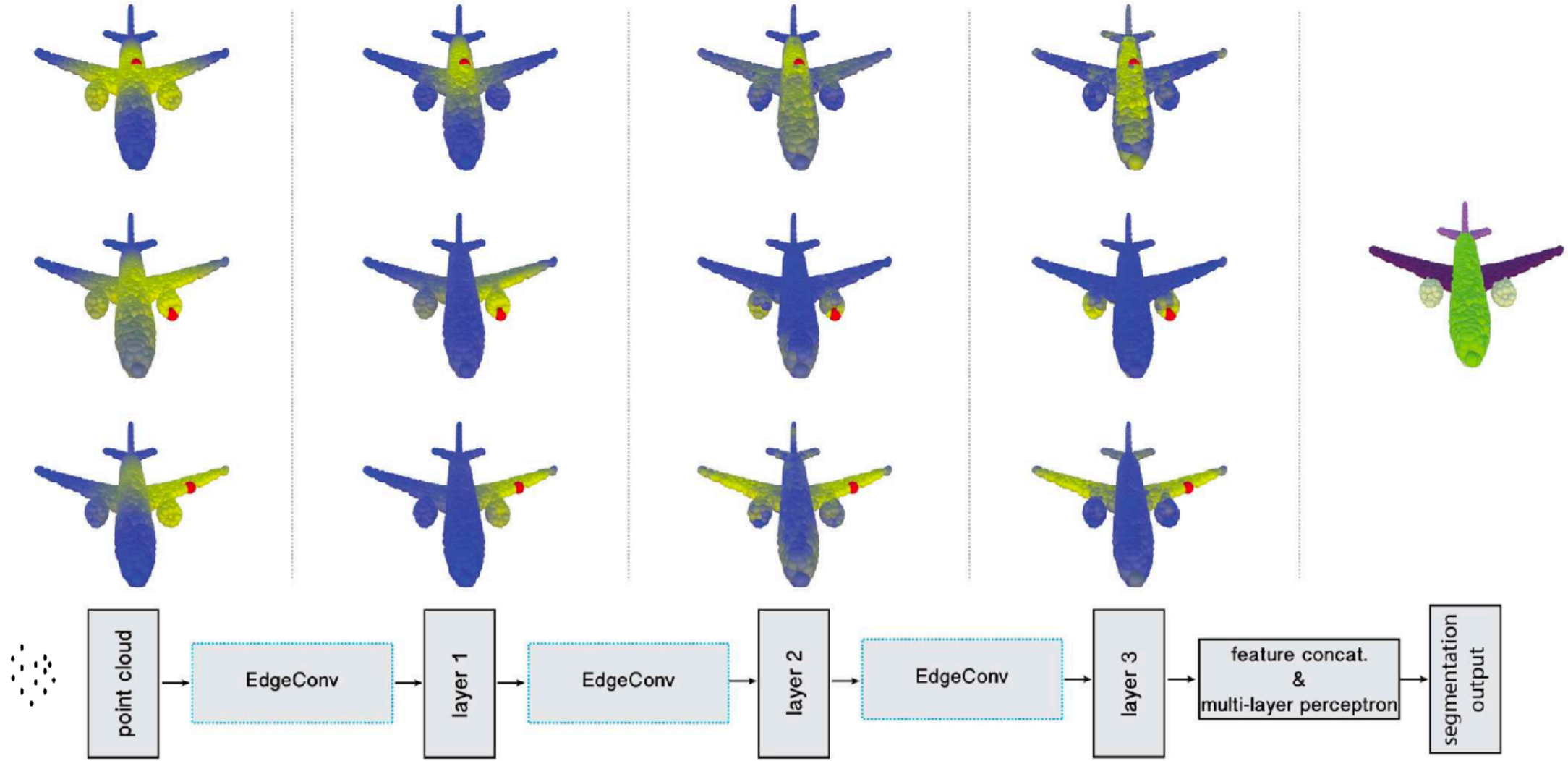
Machine Learning Classes

- **Goal of GNN:**
  - Classification (Bkg vs Phy.)
  - Segmentation (Attributed hits)
- **Definition of classes:**
  - NC/CC DIS
  - Exclusive Hadon
  - Exclusive Meson
  - Photo-production

#	Class	dir	Physics Process	Generator / source
0	Background	bkg	SynRad (2000), e-beam-gas/brems (3000), Coulomb (4000), Touschek (5000), p-beam-gas (6000)	dataprod / GETaLM / Xsuite / Pythia8
1	NeutralCurrentDISQ1	ncdisq1	NC DIS, Q2 ≥ 1	Pythia8
2	NeutralCurrentDISQ10	ncdisq10	NC DIS, Q2 ≥ 10	Pythia8
3	NeutralCurrentDISQ100	ncdisq100	NC DIS, Q2≥100	Pythia8
4	NeutralCurrentDISQ1000	ncdisq1000	NC DIS, Q2≥1000	Pythia8
5	ChargedCurrentDISQ100	ccdisq100	CC DIS, Q2≥100	Pythia8
6	ChargedCurrentDISQ1000	ccdisq1000	CC DIS, Q2≥1000	Pythia8
7	DiffractiveDIS	ddis	Diffractive DIS, intact proton	RAPGAP
8	DVCS	dvcs	Deeply Virtual Compton Scattering	DVCS generator
9	DDVCS	ddvcs	Double DVCS (ep→e ℓ + ℓ −p) — gated to BEAM=18x275	EpIC (18x275/10x130 only)
10	DVMP	dvmp	Deeply Virtual Meson Production	EpIC
11	DEMPQ3to10	dempq3to10	Deep Exclusive Meson Production, Q2 ∈ [3,10]	DEMPgen
12	DEMPQ10to20	dempq10to20	DEMP, Q2∈[10,20]	DEMPgen
13	DEMPQ20to35	dempq20to35	DEMP, Q2∈[20,35]	DEMPgen
14	TCS	tcs	Timelike Compton Scattering	TCS generator
15	DiffractiveJPsi	jpsi	Diffractive J/ψ electroproduction, ep	lAger
16	JPsiPhotoproduction	jpsiphoto	Coherent J/ψ photoproduction — gated to BEAM=18x275	eSTARlight (18x275 only)
17	UChannelRho	urho	u-channel ρ exchange	SLIGHT/GENEX
18	CoherentRho	cohrho	Coherent diffractive ρ-off eAu (nuclear rapidity gap)	Sartre
19	UChannelPi0	upi0	u-channel π0 exchange — gated to BEAM=18x275	(18x275 only)
20	Upsilon	upsilon	ϒ threshold production	eSTARlight-family
21	MesonStructureFn	mesonsf	K+Λ exclusive meson structure function	eicMesonSFGen
22	CoherentPhi	cohphi	Coherent diffractive ϕ→K+K− — off eAu (nuclear rapidity gap)	Sartre
23	Rho	rho	Exclusive ρ→π+π− electroproduction, ep, Q2∈[0,5] — *npsim SIGSEGVs, unresolved*	SLIGHT
24	Spectroscopy	spectroscopy	Exotic states X, Y, ψ(2S)	dedicated spectroscopy generators
25	Omega	omega	ω neutral-decay photoproduction — gated to BEAM=18x275	SLIGHT (18x275 only)
26	Photoproduction	photoprod	Min-bias photoproduction, Q2∈[0,1]	Pythia6-eic

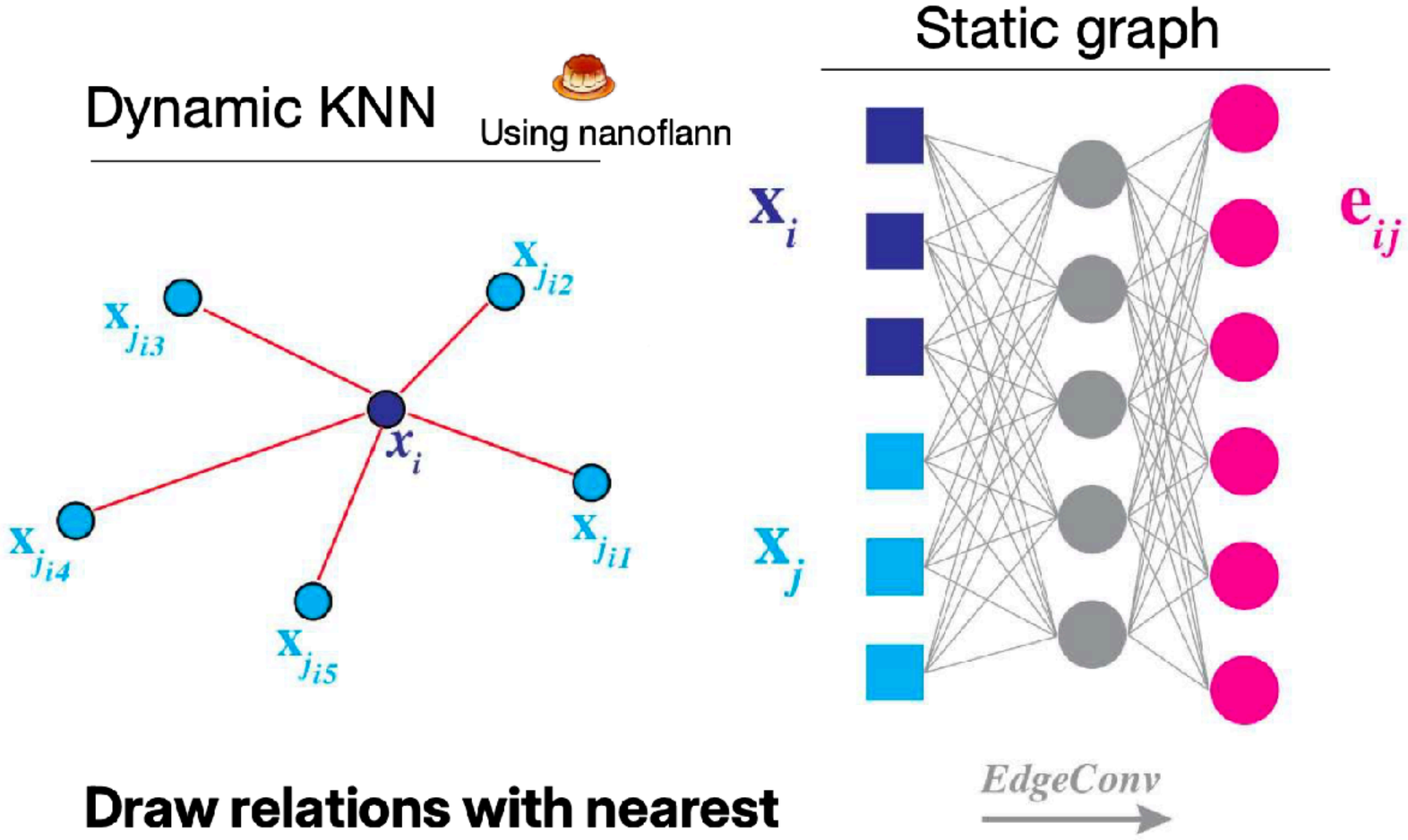
Point Cloud Model

Multi-scale Interaction Learned (DGCNN)



ONNX Implementation in EICRecon

Technical Subtlety

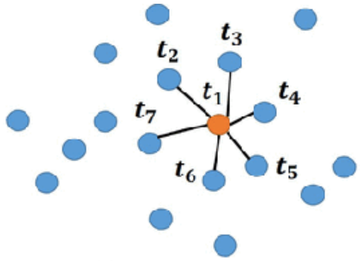


Draw relations with nearest neighbours

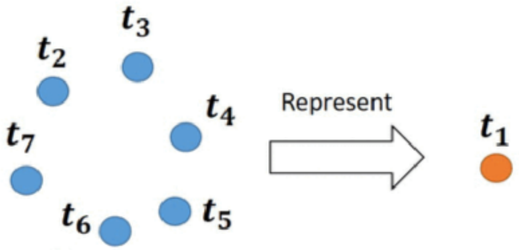
Represent each hit by its interaction with KNN

Dynamic Graph CNN

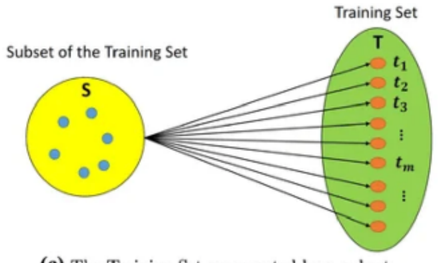
Event Pre-Filtering



(a) K-nearest neighbor



(b) A sample represented by the Top K samples



(c) The Training Set represented by a subset

arXiv:1801.07829

- Physics Process Classification & Segmentation: Multiscale Hit Interactions
  Avoid unnecessary **ACTS Combinatorial Kalman Filter** Reconstruction

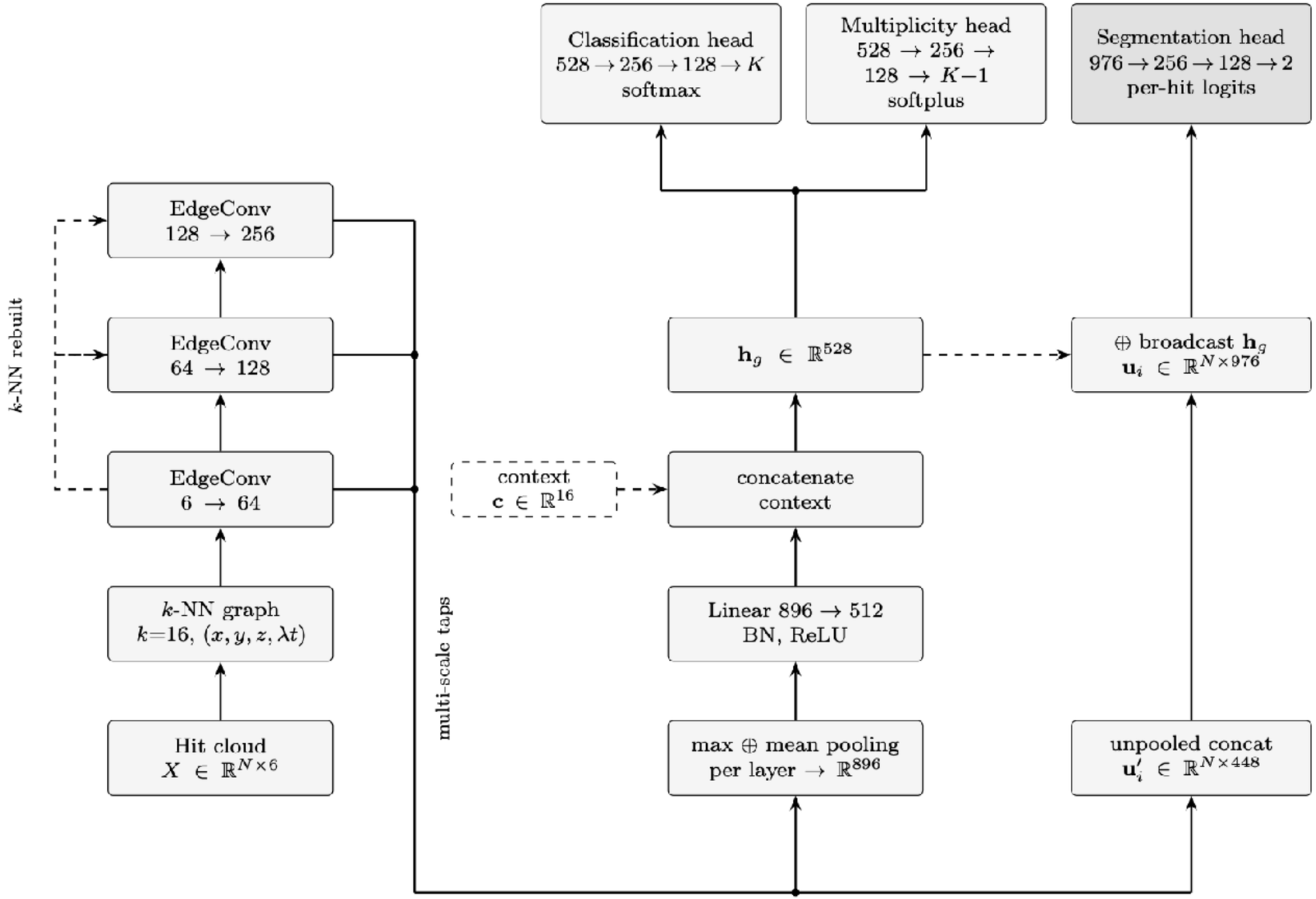
- ONNX is **only** capable of importing static graphs: Inference on CPU / Training on GPU
  GNN Layer ~ kNN (dynamic graph) + EdgeConv (static graph)

- **kNN :** k-NearestNeighbors is dynamic graph (flexible number of input hits)
  “nanoflann” lightweight header library: “KDTree Space + kNN algorithm” in C++

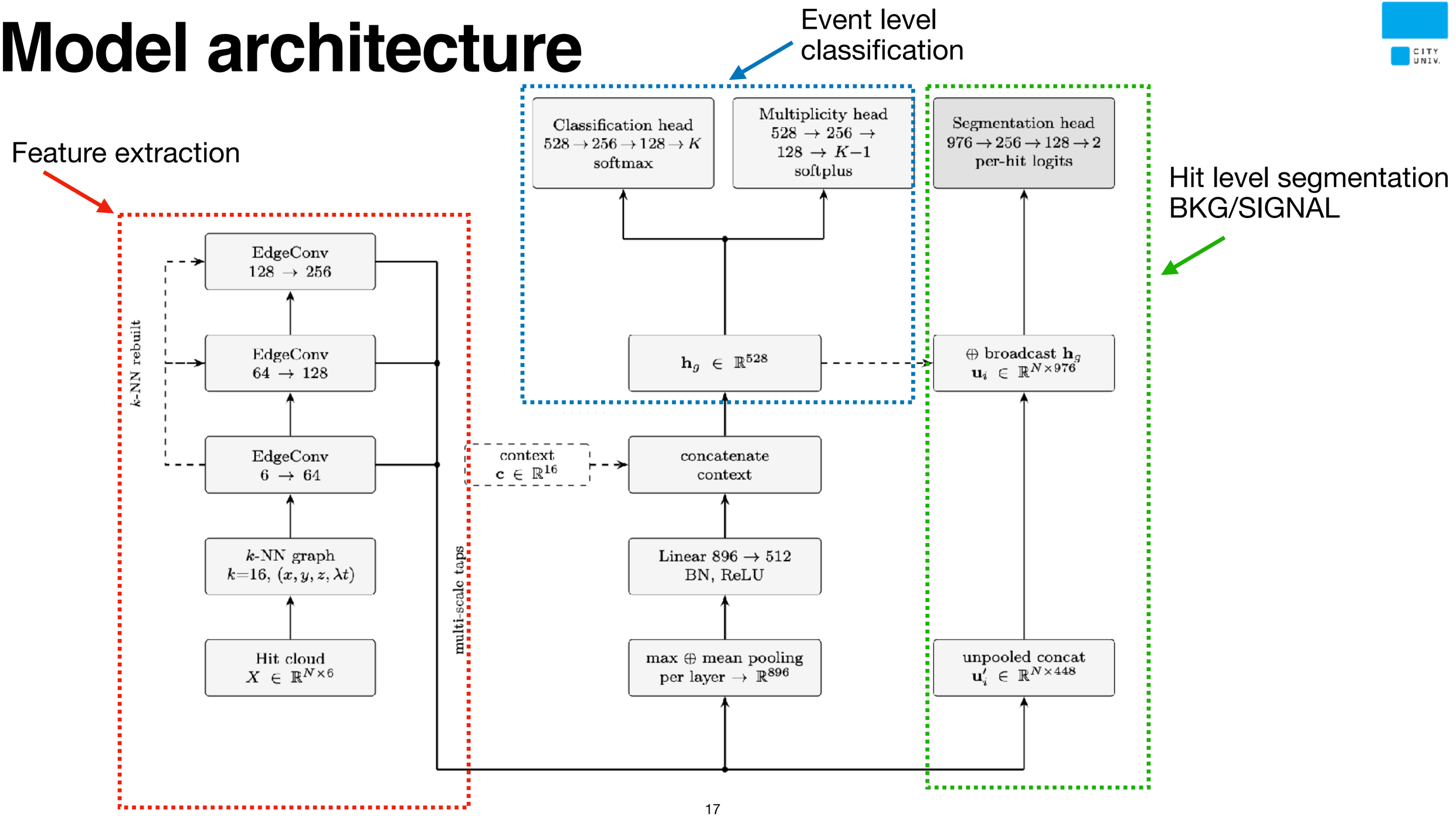
https://github.com/jlblancoc/nanoflann

- **EdgeConv:** hit-to-hit connection to be trained

Model architecture



Model architecture



Dataset

Input, ground truth and output, per head

Head	Input	Ground Truth	Output
cls_head	Pooled embedding hg (528-d)	Physics class (0–26), from EventBuilder (trigger_classes tail)	27-way softmax logits/probabilities
mult_head	hg.detach() (528-d)	26-dimensional binary presence vector, from trigger_classes_mask bitmask	26 raw logits (sigmoid → class presence)
seg_head	Per-hit multiscale features + broadcast hg.detach() (784-d)	Per-hit class (0–32), from RawHitLinks/Associations, inside generatorStatus	N × 33 per-hit seg logits

GNN Pre-filtering statistics being prepared here
Goal: improved purity after ML is applied
(NB: Trigger logic and check will also be better evaluated/improved before the use of AI)

Event Builder Performances

Each class is mixed with ”gold-coating” — 1 true physics per 2us

— datasets/gnn-20260721-eventbld/eicrecon/10x100-gold-coating classes

Class		eicrecon	(time,hits/tf)	(hits)	STAGE 1: TRIGGER LOGIC time %eff	time %pur	track %eff	track %pur	calo %eff	calo %pur	trigger %eff	trigger %pur	STAGE 2: ACTS track %eff	track %pur	calo %eff	calo %pur
bkg	2f	100tf 4,252ev	(0.5s 11,096h/tf)	–	0%	0%	–		–		–		–		–	
ncdisq1	1f	100tf 4,289ev	(0.7s 22,667h/tf)	(595h/ev)	101%	2%	95%	49%	100%	2%	100%	2%	83%	96%	46%	98%
ncdisq10	2f	131tf 5,609ev	(0.6s 14,845h/tf)	(613h/ev)	100%	2%	98%	49%	100%	2%	100%	2%	80%	94%	41%	98%
ncdisq100	1f	100tf 4,286ev	(0.8s 22,752h/tf)	(630h/ev)	100%	2%	100%	49%	100%	2%	100%	2%	79%	94%	40%	98%
ncdisq1000	1f	100tf 4,282ev	(0.9s 22,764h/tf)	(633h/ev)	100%	2%	100%	49%	100%	2%	100%	2%	78%	97%	29%	99%
ccdisq100	1f	100tf 4,292ev	(0.8s 22,759h/tf)	(616h/ev)	100%	2%	99%	48%	100%	2%	100%	2%	77%	95%	34%	99%
ccdisq1000	1f	100tf 4,293ev	(0.8s 22,707h/tf)	(616h/ev)	100%	2%	100%	48%	100%	2%	100%	2%	76%	96%	31%	99%
ddis	1f	100tf 4,286ev	(0.8s 22,557h/tf)	(564h/ev)	100%	2%	90%	48%	100%	2%	100%	2%	74%	100%	23%	97%
dvcs	1f	100tf 4,264ev	(0.7s 22,468h/tf)	(532h/ev)	100%	2%	67%	39%	100%	2%	100%	2%	97%	97%	7%	92%
ddvcs		–	–	–	–		–		–		–					
dvmp	1f	100tf 4,263ev	(0.7s 22,439h/tf)	(524h/ev)	100%	2%	72%	43%	100%	2%	100%	2%	100%	96%	31%	96%
dempq3to10	1f	100tf 4,268ev	(0.8s 22,437h/tf)	(541h/ev)	100%	2%	90%	54%	100%	2%	100%	2%	95%	94%	–	
dempq10to20	1f	100tf 4,252ev	(0.8s 22,395h/tf)	(538h/ev)	100%	2%	73%	47%	100%	2%	100%	2%	97%	98%	0%	97%
dempq20to35	1f	100tf 4,262ev	(0.7s 22,454h/tf)	(545h/ev)	100%	2%	92%	57%	100%	2%	100%	2%	98%	94%	0%	98%
tcs	1f	100tf 4,254ev	(0.7s 22,393h/tf)	(545h/ev)	100%	2%	87%	50%	100%	2%	100%	2%	88%	100%	–	
jpsi	1f	100tf 4,279ev	(0.7s 22,617h/tf)	(549h/ev)	100%	2%	94%	51%	100%	2%	100%	2%	92%	92%	11%	88%
jpsiphoto		–	–	–	–		–		–		–					
urho	1f	100tf 4,253ev	(0.7s 22,322h/tf)	(537h/ev)	100%	2%	25%	24%	100%	2%	100%	2%	100%	100%	–	
cohrho	1f	100tf 4,270ev	(0.7s 22,640h/tf)	(568h/ev)	100%	2%	92%	46%	100%	2%	100%	2%	86%	89%	–	
upi0		–	–	–	–		–		–		–					
upsilon	1f	100tf 4,264ev	(0.8s 22,368h/tf)	(543h/ev)	100%	2%	93%	53%	100%	2%	100%	2%	93%	97%	–	
mesonsf	1f	100tf 4,271ev	(0.7s 22,496h/tf)	(555h/ev)	100%	2%	97%	51%	100%	2%	100%	2%	98%	98%	0%	98%
cohphi	1f	100tf 4,269ev	(0.7s 22,571h/tf)	(552h/ev)	100%	2%	94%	48%	100%	2%	100%	2%	87%	98%	–	
rho	1f	0tf	–	–	–		–		–		–		–		–	
spectroscopy	1f	100tf 4,270ev	(0.7s 22,538h/tf)	(553h/ev)	100%	2%	98%	49%	100%	2%	100%	2%	83%	94%	–	
omega		–	–	–	–		–		–		–					
photoprod	2f	200tf 8,543ev	(0.7s 22,510h/tf)	(567h/ev)	100%	2%	85%	48%	100%	2%	100%	2%	71%	97%	46%	98%
TOTAL	26f	2,331tf 99,571ev	(0.7s 22,533h/tf)	(568h/ev)	100%	2%	88%	48%	100%	2%	100%	2%	81%	96%	36%	98%

Trigger Logic

Effective Reconstruction
(ACTS + Calo Island)

Event Builder Performances

Each class is mixed with ”gold-coating” — N true physics per 2us (~5MHz)

— datasets/gnn-20260721-eventbld-pileup/eicrecon/10x100-gold-coating classes —

Class			eicrecon	(time,hits/tf)	(hits)	STAGE 1: TRIGGER LOGIC time %eff %pur	STAGE 1: TRIGGER LOGIC track %eff %pur	STAGE 1: TRIGGER LOGIC calo %eff %pur	STAGE 1: TRIGGER LOGIC trigger %eff %pur	STAGE 2: ACTS track %eff %pur	STAGE 2: ACTS calo %eff %pur	pileup %eff (found/inj)
bkg			–	–	–	–	–	–	–		–	
ncdisq1	1f	0tf		–	–	–	–	–	–	–	–	–
ncdisq10	1f	100tf	4,377ev	(1.0s 24,182h/tf)	(621h/ev)	100% 3%	97% 34%	100% 3%	100% 3%	82% 85%	14% 49%	97.0% (766/790)
ncdisq100	1f	0tf		–	–	–	–	–	–	–	–	–
ncdisq1000	1f	100tf	4,388ev	(1.0s 24,365h/tf)	(661h/ev)	100% 3%	100% 37%	100% 3%	100% 3%	79% 86%	11% 50%	97.2% (806/829)
ccdisq100	1f	100tf	4,395ev	(1.0s 24,065h/tf)	(642h/ev)	100% 3%	99% 37%	100% 3%	100% 3%	76% 84%	11% 46%	97.7% (816/835)
ccdisq1000	1f	100tf	4,348ev	(1.4s 22,932h/tf)	(608h/ev)	100% 3%	99% 35%	100% 3%	100% 3%	75% 88%	10% 45%	97.5% (818/839)
ddis	1f	100tf	4,366ev	(1.4s 22,670h/tf)	(554h/ev)	100% 3%	91% 37%	100% 3%	100% 3%	78% 79%	4% 25%	96.8% (825/852)
dvcs	1f	100tf	4,387ev	(1.3s 22,825h/tf)	(524h/ev)	100% 3%	78% 33%	100% 3%	100% 3%	98% 64%	0% 28%	96.5% (821/851)
ddvcs			–	–	–	–	–	–	–		–	
dvmp	1f	100tf	4,369ev	(1.4s 22,256h/tf)	(528h/ev)	100% 3%	79% 34%	100% 3%	100% 3%	98% 71%	5% 30%	96.5% (807/836)
dempq3to10	1f	100tf	4,367ev	(1.4s 22,594h/tf)	(535h/ev)	100% 3%	97% 41%	100% 3%	100% 3%	96% 69%	–	96.4% (825/856)
dempq10to20	1f	0tf		–	–	–	–	–	–	–	–	–
dempq20to35	1f	100tf	4,402ev	(0.9s 24,126h/tf)	(554h/ev)	100% 3%	91% 35%	100% 3%	100% 3%	96% 67%	0% 35%	97.2% (869/894)
tcs	1f	0tf		–	–	–	–	–	–	–	–	–
jpsi	1f	0tf		–	–	–	–	–	–	–	–	–
jpsiphoto			–	–	–	–	–	–	–		–	
urho	1f	100tf	4,347ev	(1.0s 23,657h/tf)	(576h/ev)	100% 3%	40% 22%	100% 3%	100% 3%	100% 49%	–	96.4% (795/825)
cohrho	1f	100tf	4,384ev	(1.3s 23,112h/tf)	(569h/ev)	100% 3%	95% 37%	100% 3%	100% 3%	88% 71%	–	96.9% (873/901)
upi0			–	–	–	–	–	–	–		–	
upsilon	1f	100tf	4,350ev	(1.3s 22,609h/tf)	(540h/ev)	100% 3%	95% 38%	100% 3%	100% 3%	96% 73%	–	97.0% (770/794)
mesonsf	1f	0tf		–	–	–	–	–	–	–	–	–
cohphi	1f	100tf	4,362ev	(1.3s 22,445h/tf)	(553h/ev)	100% 3%	89% 34%	100% 3%	100% 3%	82% 72%	–	97.9% (806/823)
rho	1f	0tf		–	–	–	–	–	–	–	–	–
spectroscopy	1f	100tf	4,357ev	(0.9s 23,545h/tf)	(582h/ev)	100% 3%	97% 40%	100% 3%	100% 3%	84% 78%	–	98.2% (763/777)
omega			–	–	–	–	–	–	–		–	
photoprod	1f	0tf		–	–	–	–	–	–	–	–	–
TOTAL	22f	1,400tf	61,199ev	(1.2s 23,242h/tf)	(575h/ev)	100% 3%	89% 36%	100% 3%	100% 3%	81% 78%	10% 42%	97.1% (11360/11702)

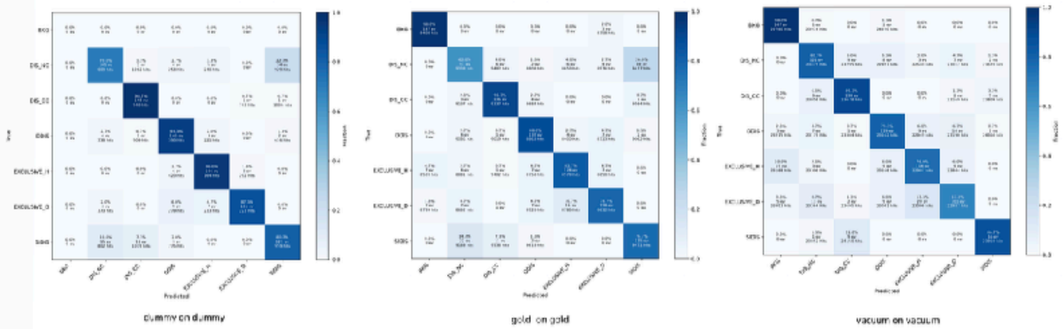
(f=files tf=2 µs timeframes ev=eventbuilder candidates, real+fake ()=mean h=hits (h/ev counts

In a nutshell: coincident_count(ncdisq10): 0 (74%), 1 (22%), 2 (3.3%), 3 (0.3%), 4 (0.0%), 5 (0.0%), 6 (0.0%), 7 (0.0%)

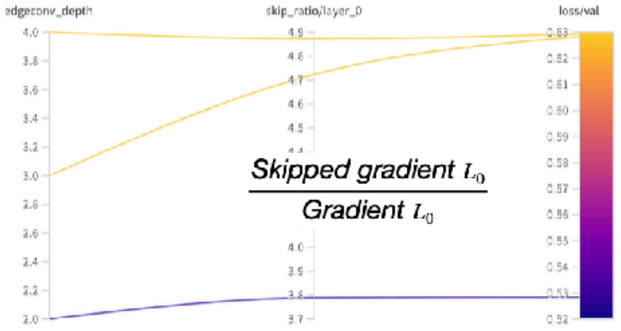
Slide Title

Slide Subtitle

Noise level curriculum learning

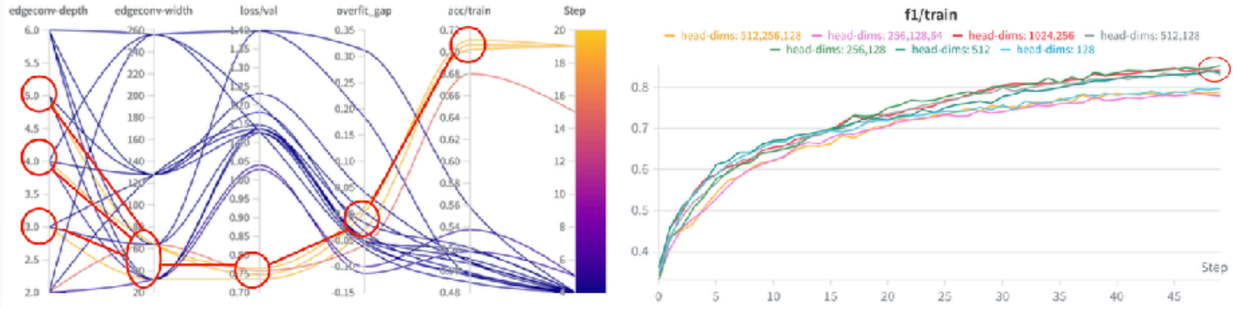


Measuring vanishing gradient



In place:
skip connections
BatchNorm/LN
multi-scale readout
ReLu

Architecture optimization W&B



DDP training

GPU	Memory Used	Memory Utilization	Compute Utilization
GPU 0	93.9 GB	95.9%	100%
GPU 1	86.6 GB	88.5%	100%
GPU 2	93.4 GB	95.4%	100%
GPU 3	83.8 GB	85.6%	100%

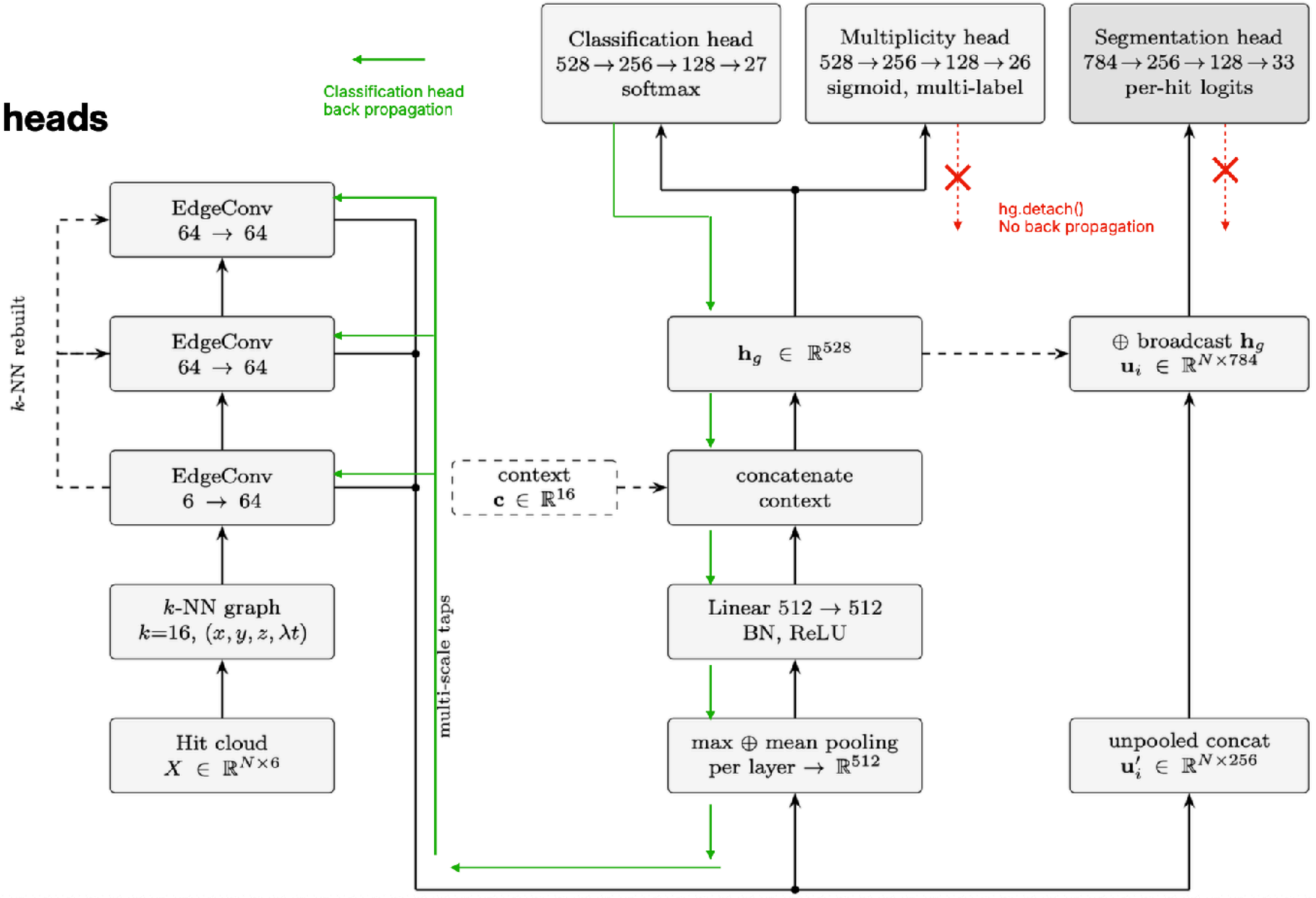
Gradient isolation

Gradient isolation between heads

Using .detach()
isolates Mlt and Seg heads
from the backbone

Gradient flow :
Cls head → backbone
Mlt/Seg → MLP only

Stable shared representation
of classification



GNN Architecture

Test Production

Data

~8000 simulated collisions events split into 75/15/15 training, validation and test

Optimization

Epochs 20-200
Optimizer Adam
Batch size

Loss function

Cross-entropy loss

Stability

Curriculum strategy
Dropout, weight decay, gradient clipping

Hardware/compute

Training runs on a 4-GPU workstation (NVIDIA Blackwell)

Hyperparameter selection

3-4 layer depth
Graduated width 64, 64, 128, 256

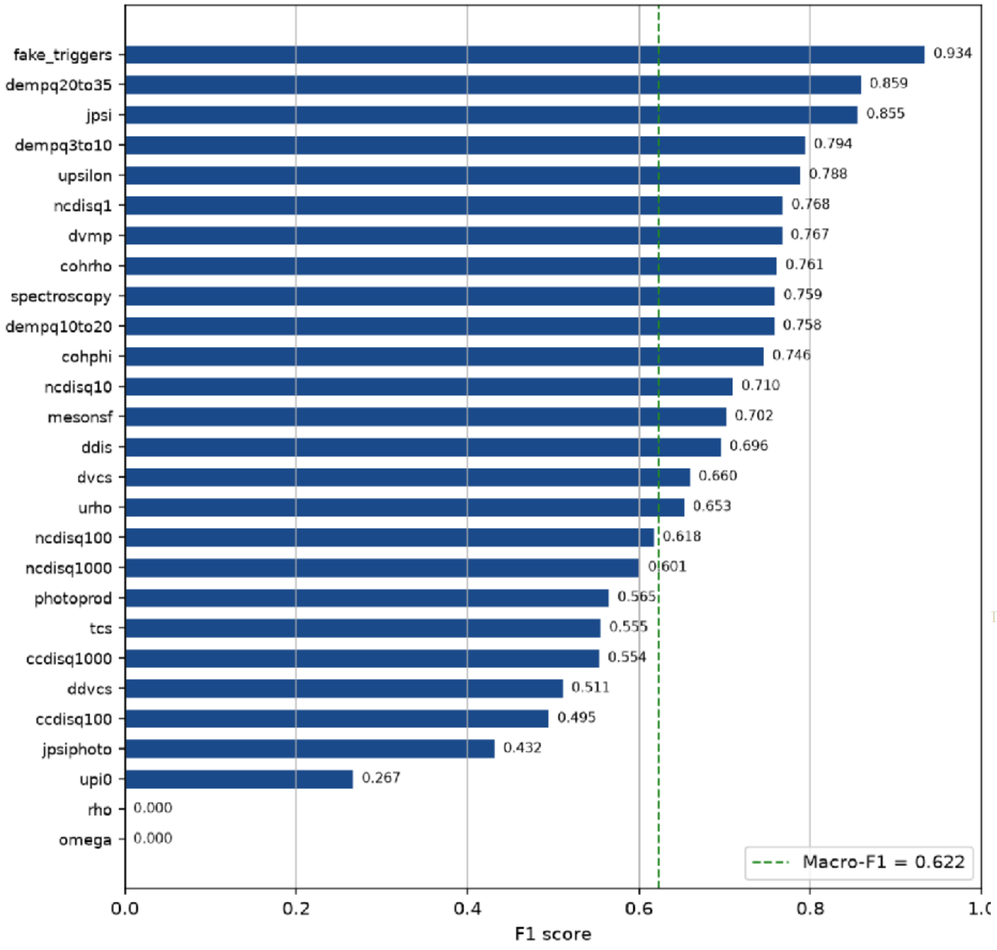
Goal

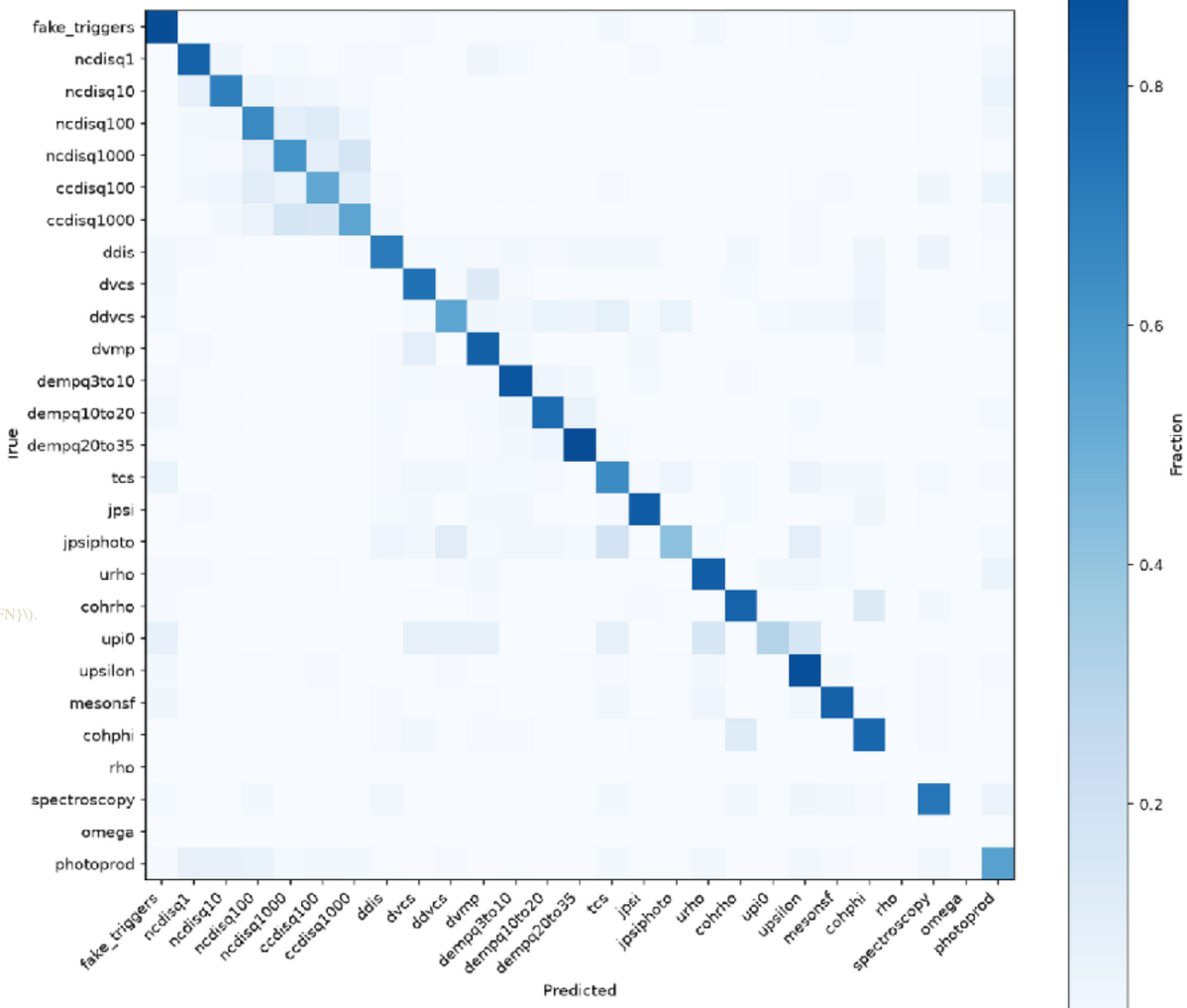
2 heads:
- Classification
- Multiplicity

Model dimensionning

Latest Dataset

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$



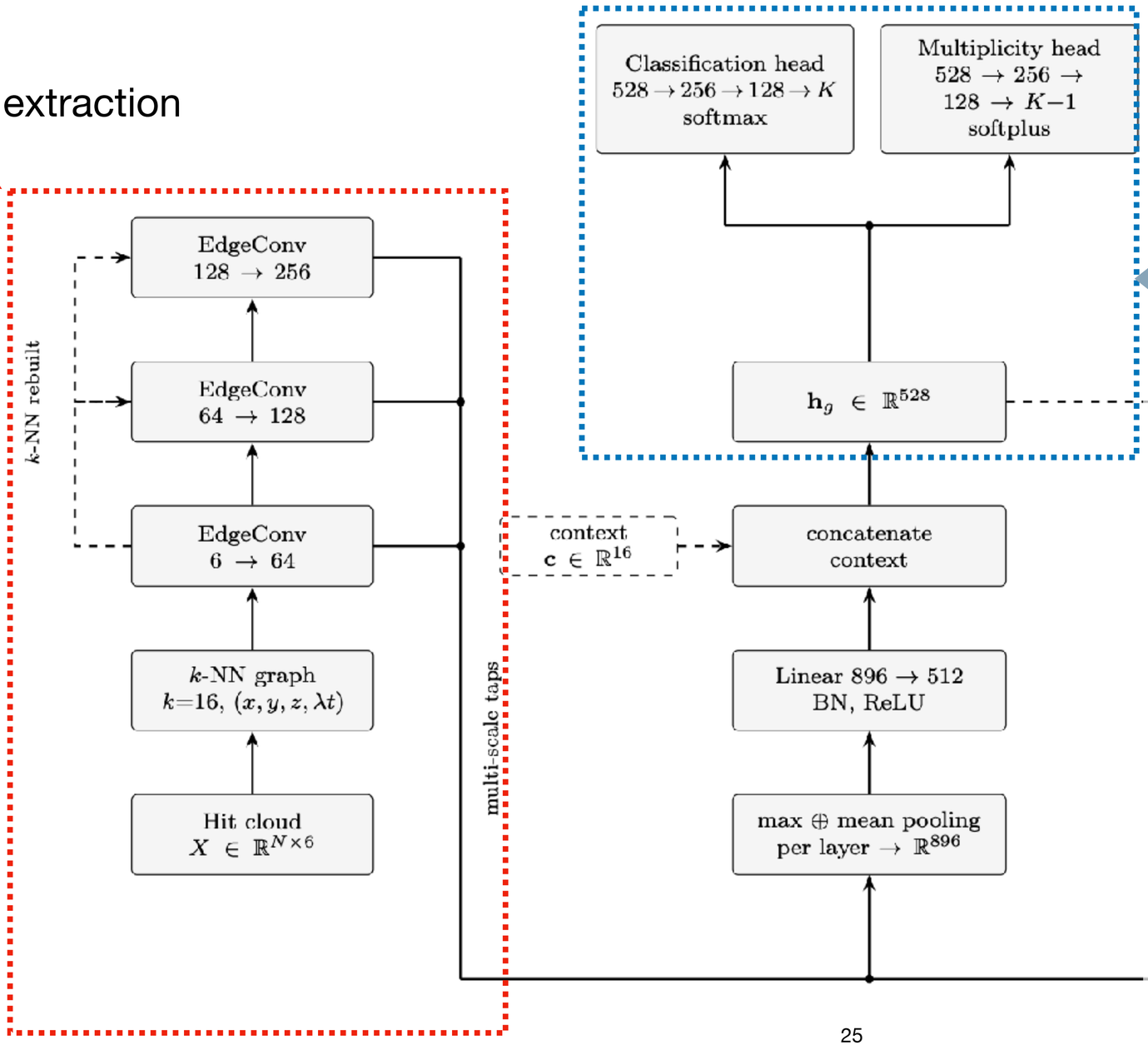


AI Terminology	Physics terms equivalent
$\text{Recall} = \frac{TP}{TP + FN}$	—> %Efficiency
$\text{Precision} = \frac{TP}{TP + FP}$	—> %Purity

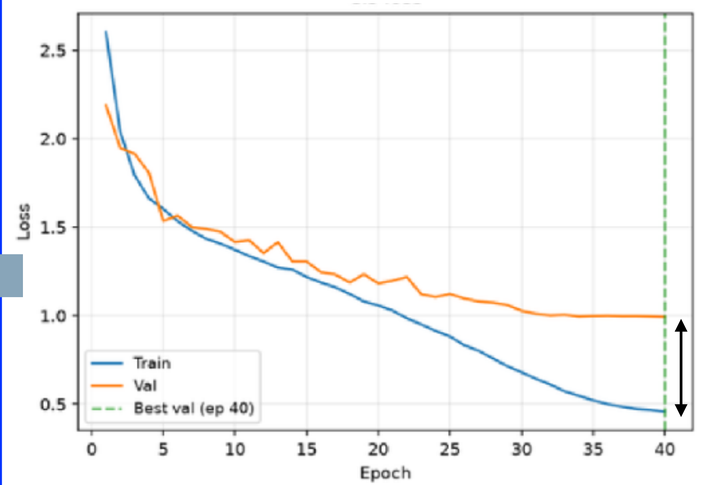
Model architecture

Event level classification

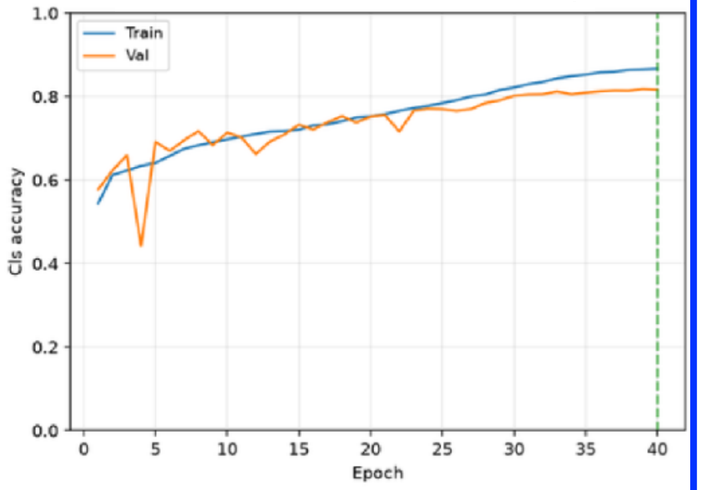
Feature extraction



Loss Estimate (Classification head only)



Classification head Accuracy



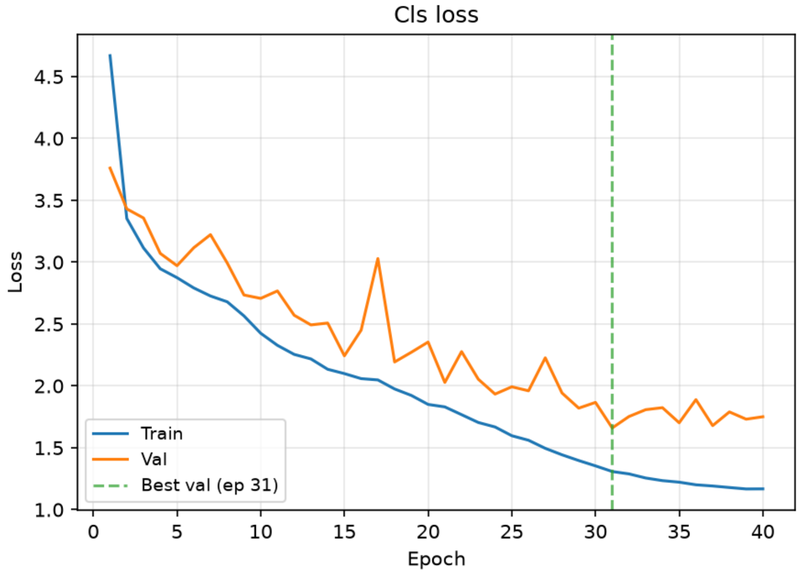
~85% with unoptimized training (!)

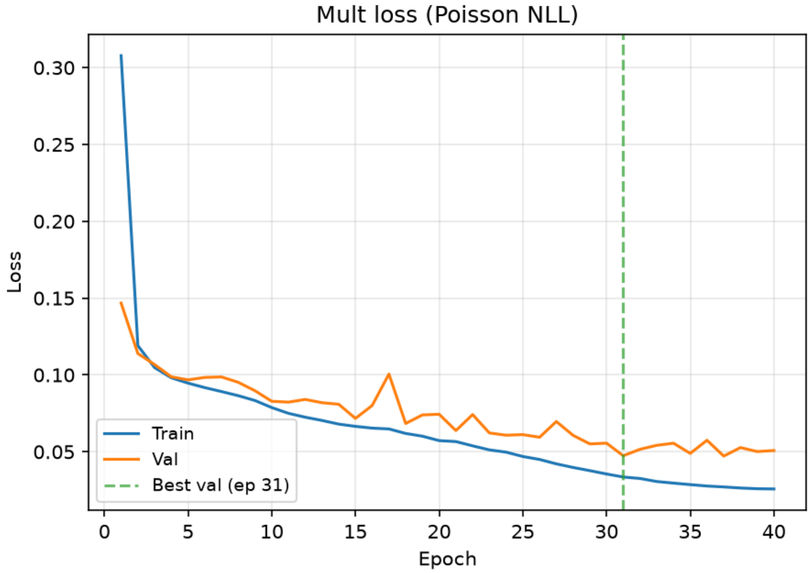
Dataset
generalizatio
being fixed

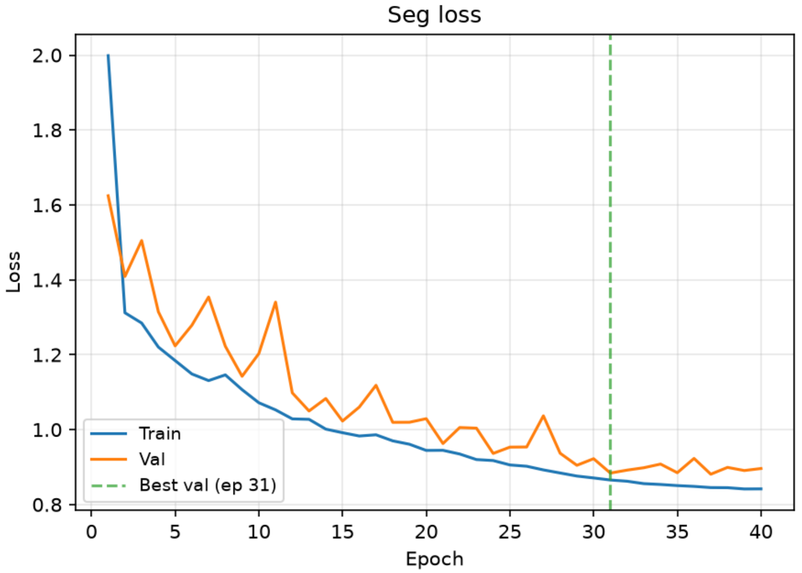
Results

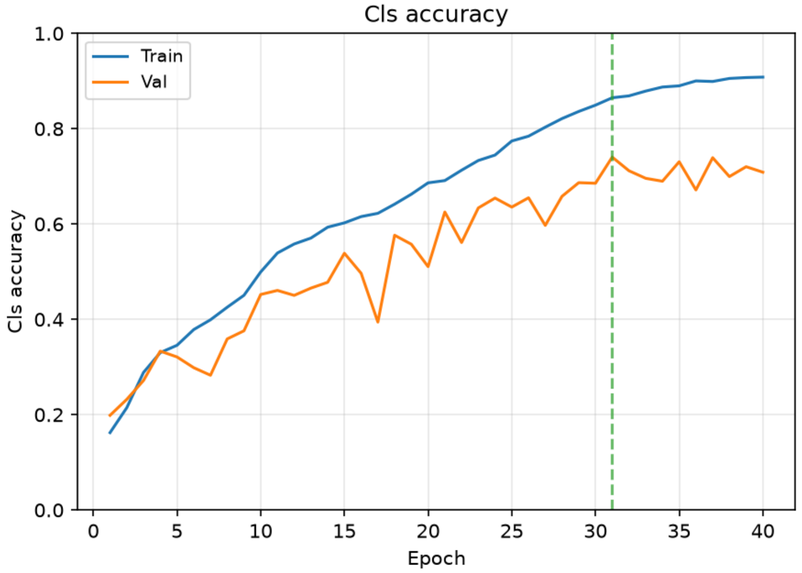
Training and intermediate results

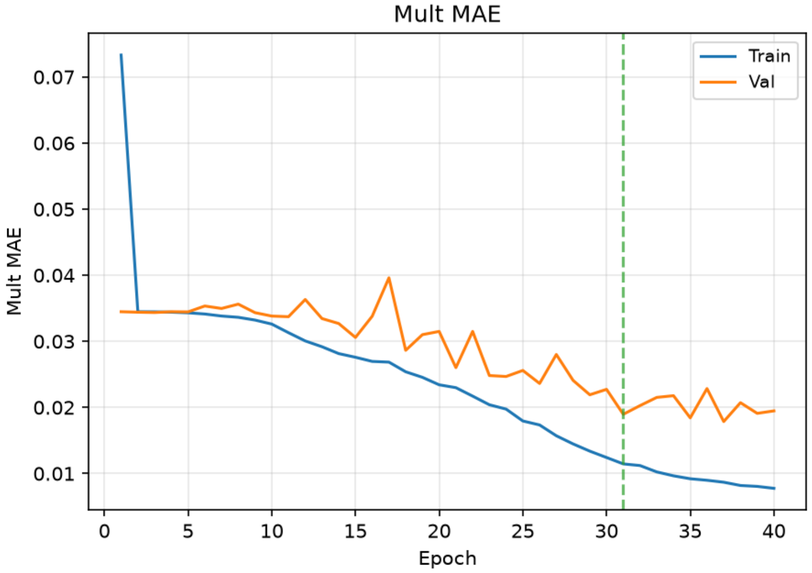
3 heads training

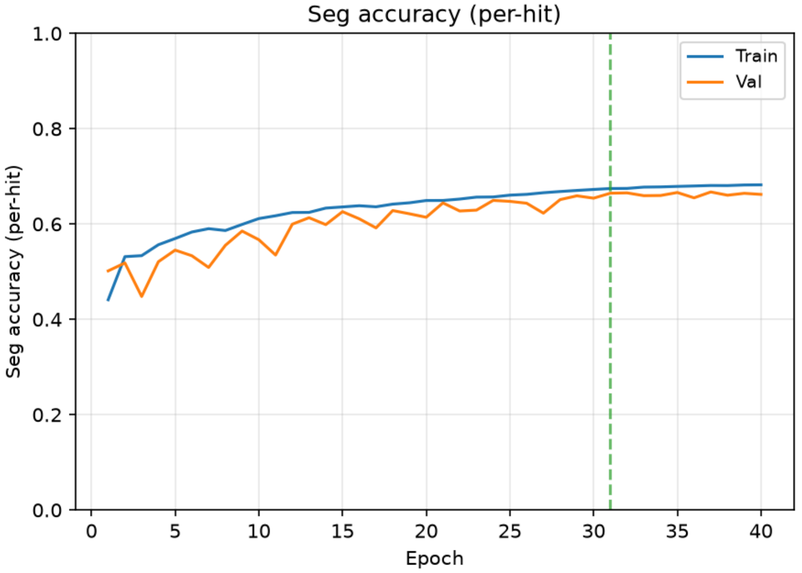










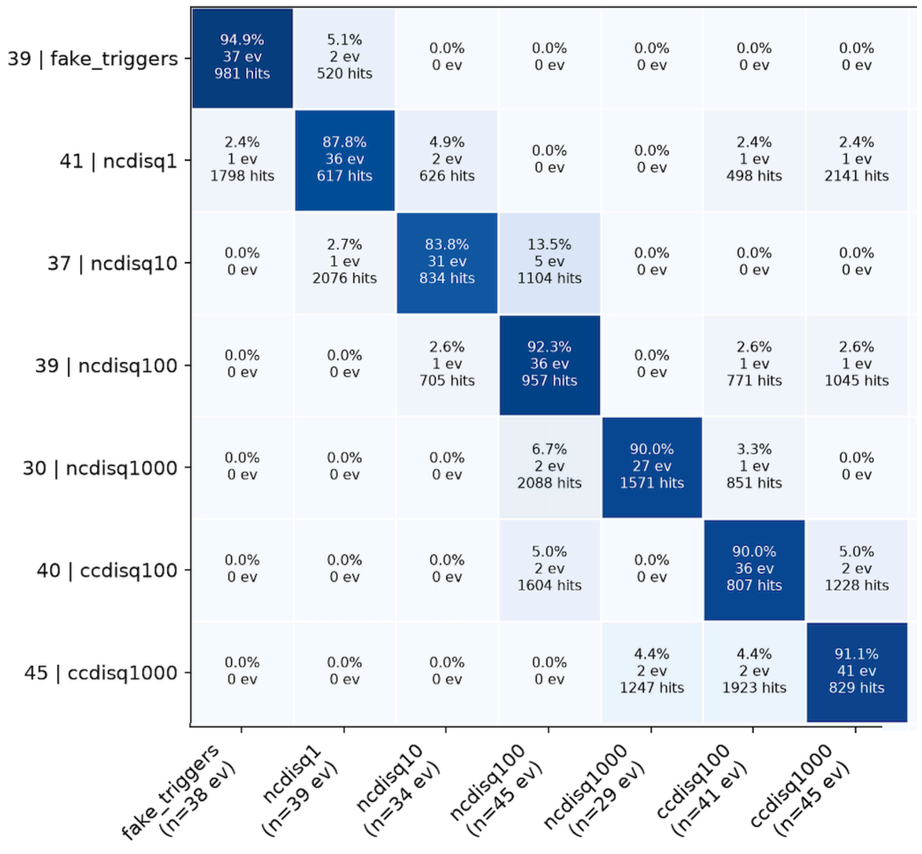


Results

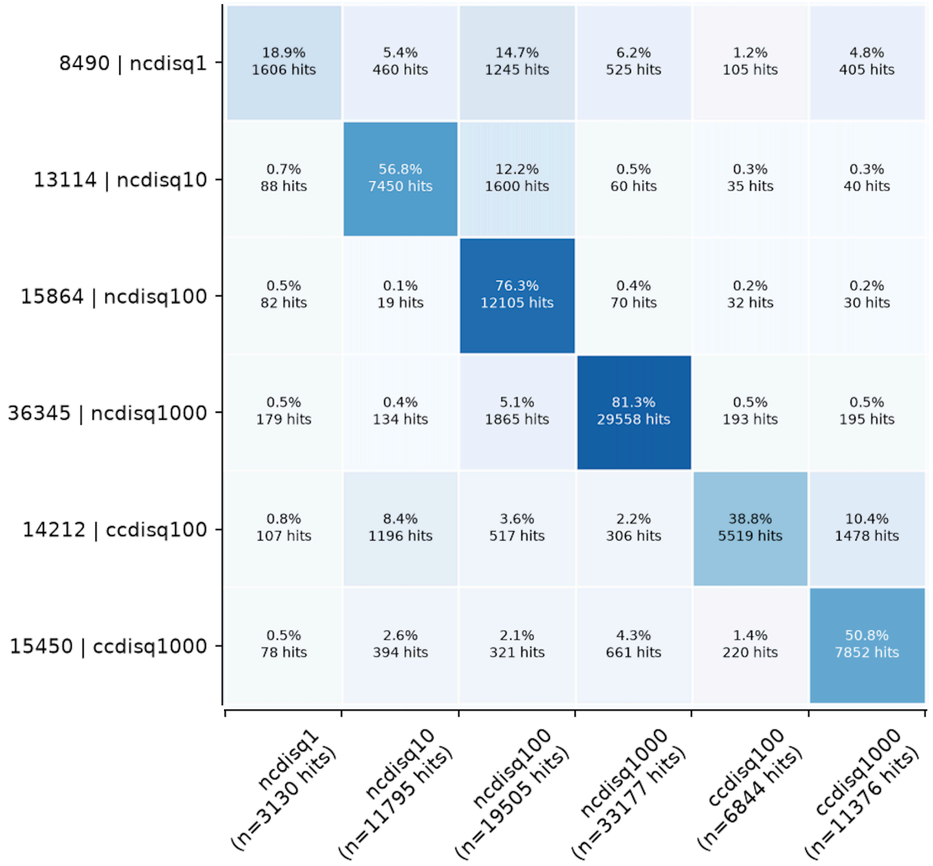
Training and intermediate results

3 heads training

Classification head



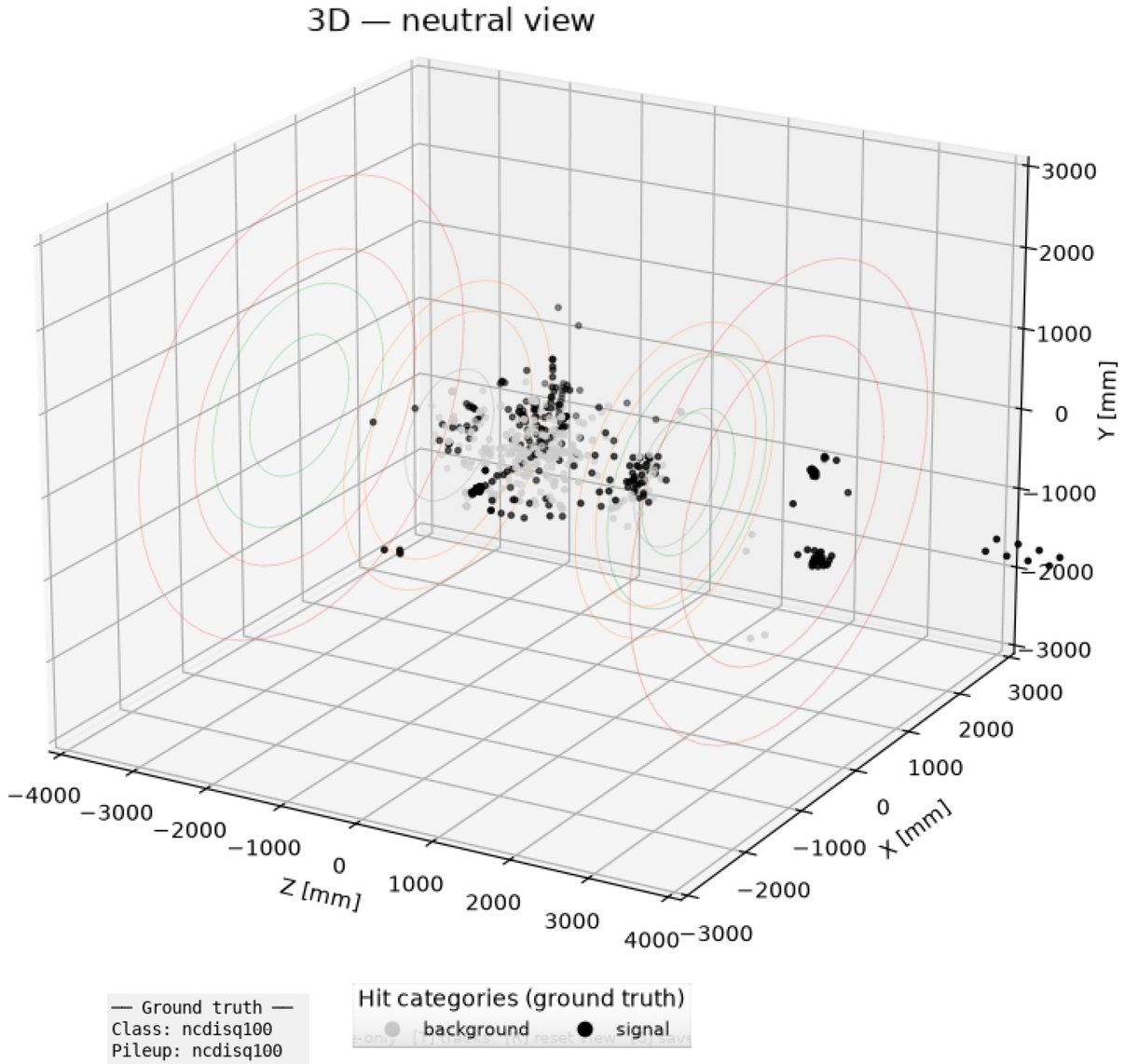
Segmentation head

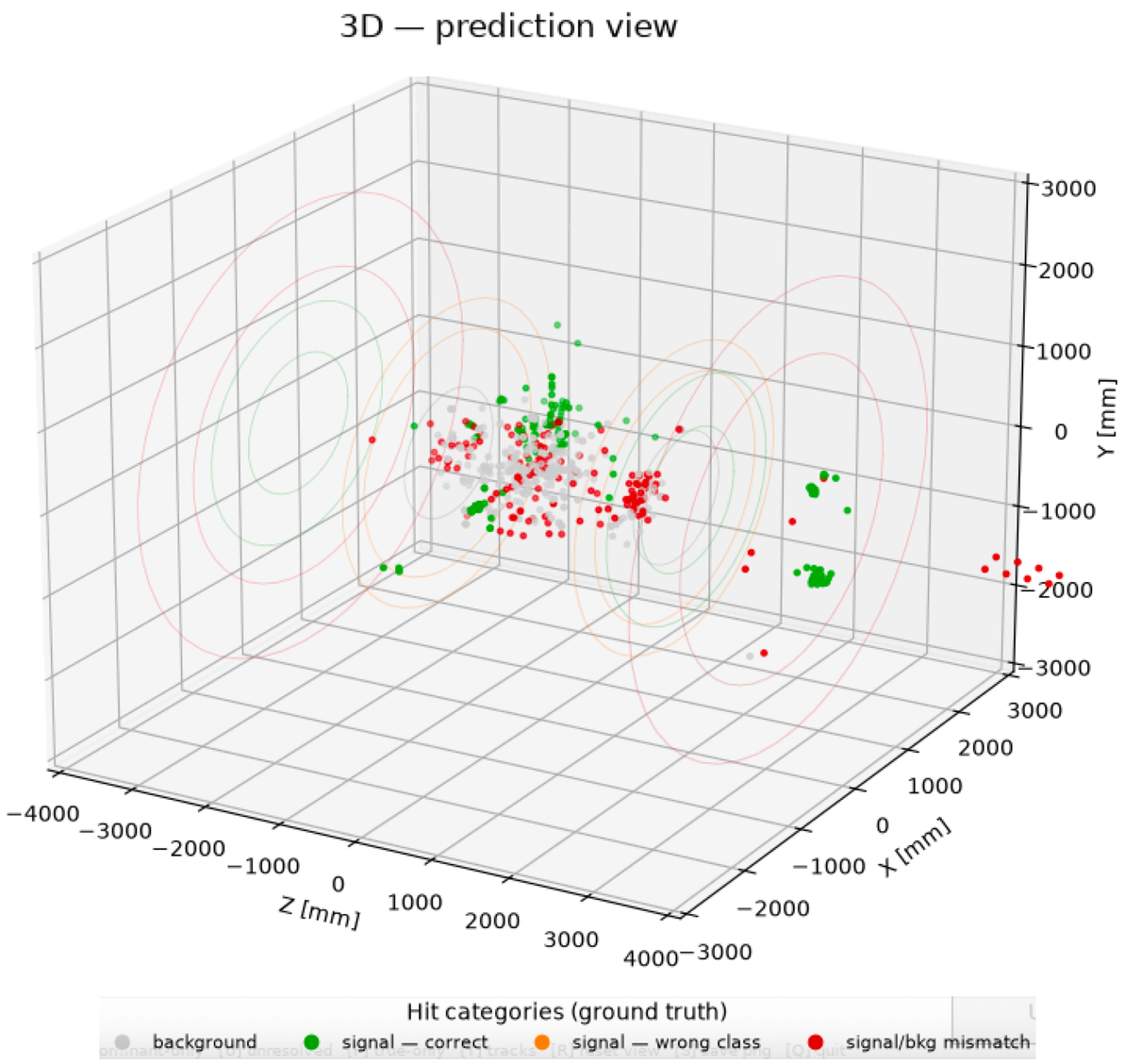


Results

True positive prediction

Event 47 | Hits : 913

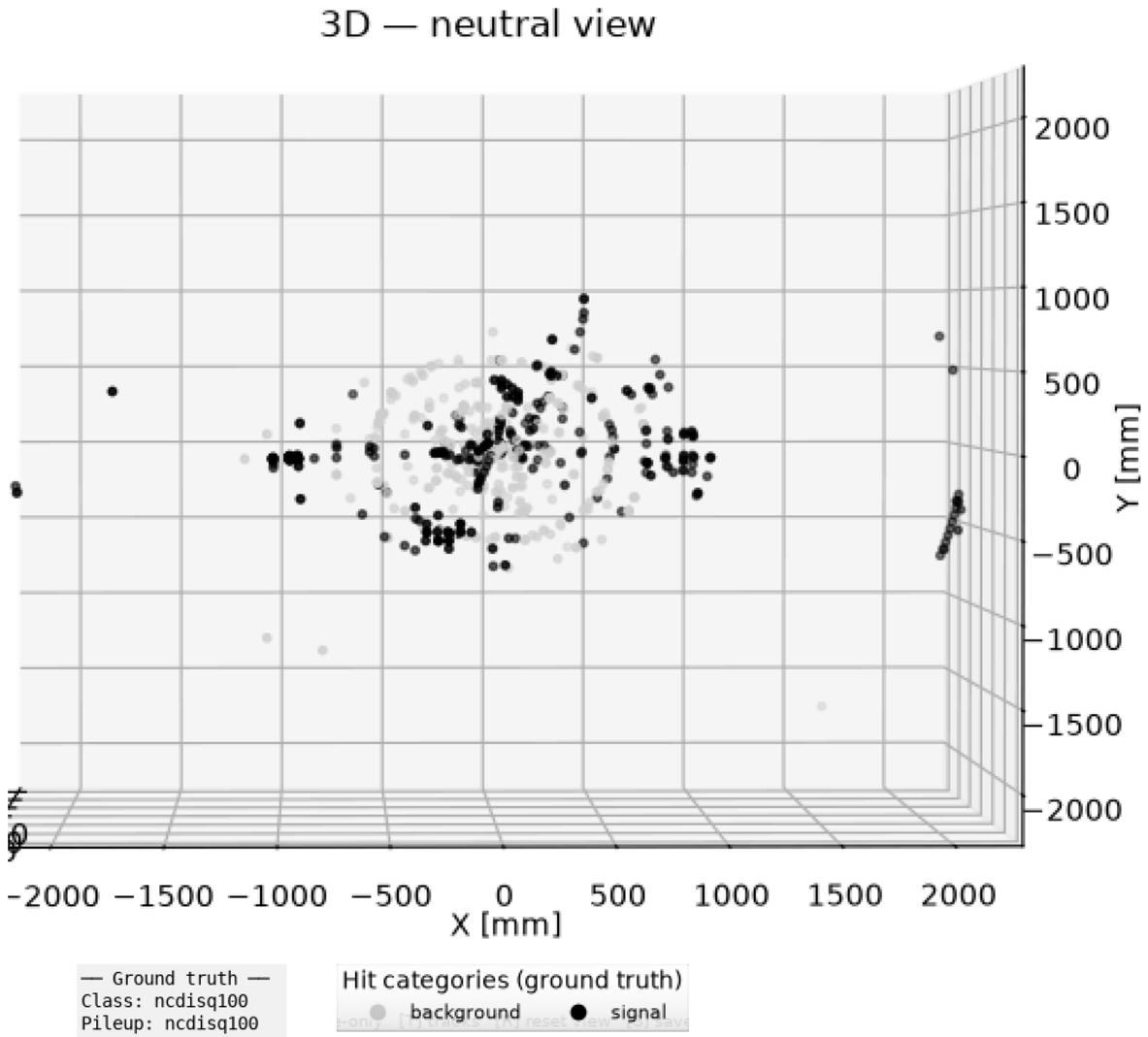


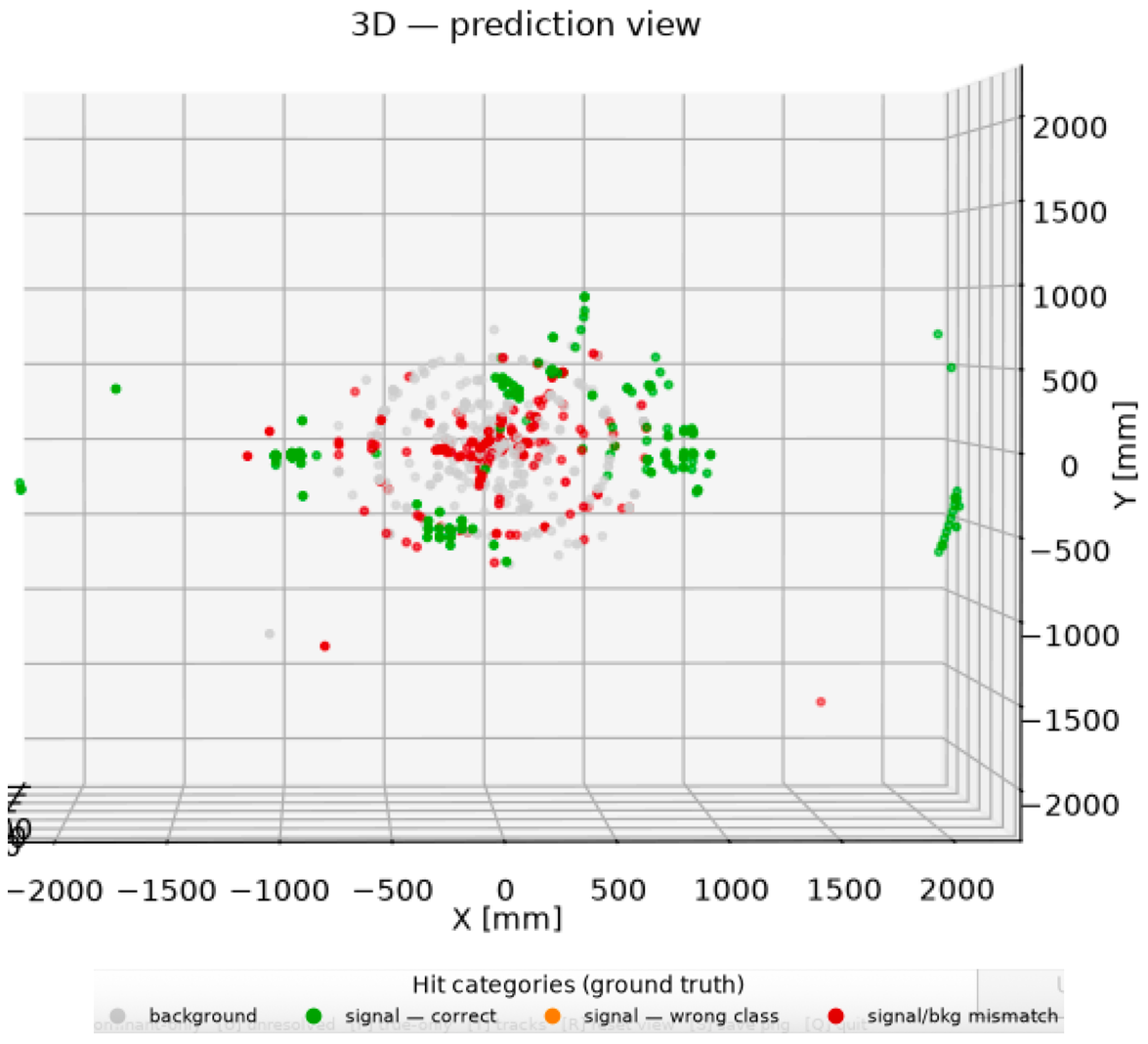


Results

True positive prediction

Event 47 | Hits : 913





Results

True positive prediction

Classification Head

Predicted class: ncisq100 == GT
with 86.2% confidence

Multiplicity Head

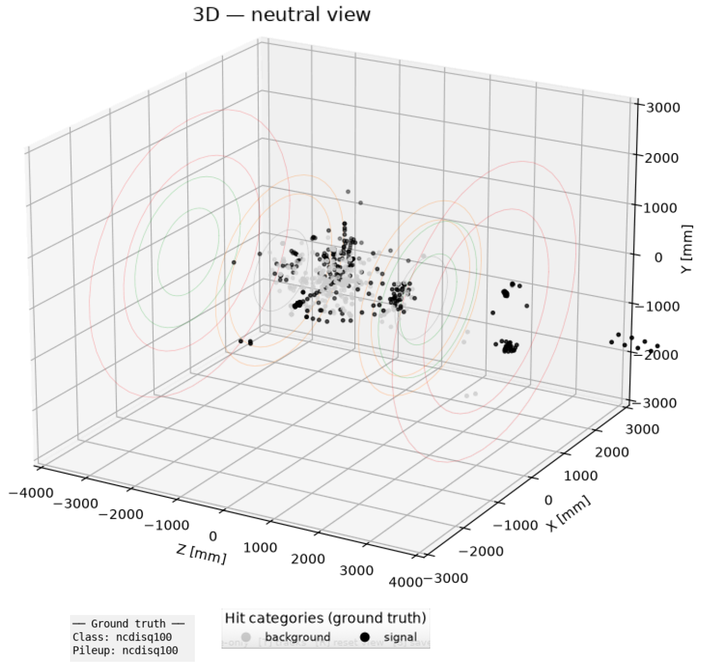
Predicted pileup: ncisq100 == GT

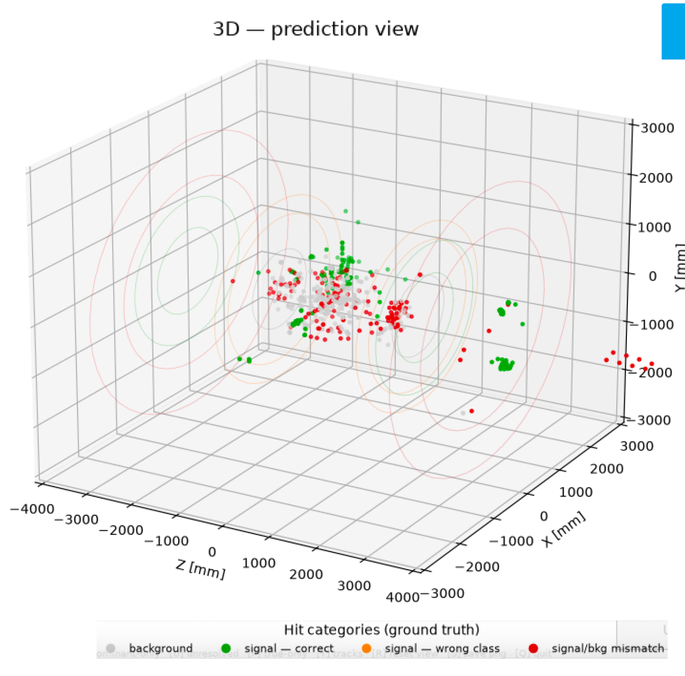
Segmentation Head

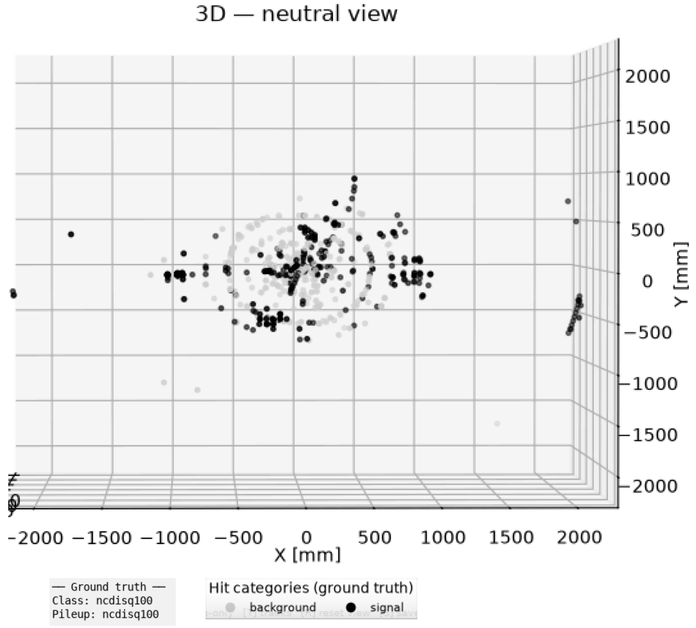
Global accuracy: 58.7%

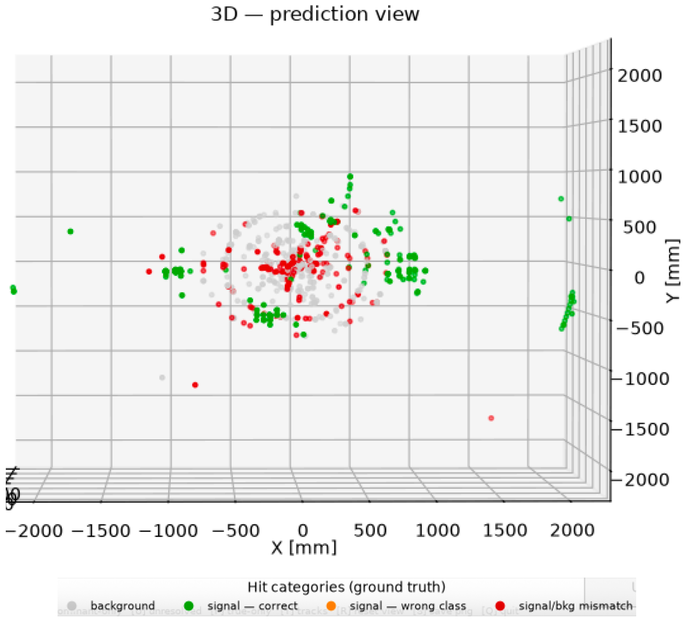
In identified event class:
Recall: 71.1%
Precision: 89.1%

Bkg/signal accuracy:
82%









Results

Unresolved class

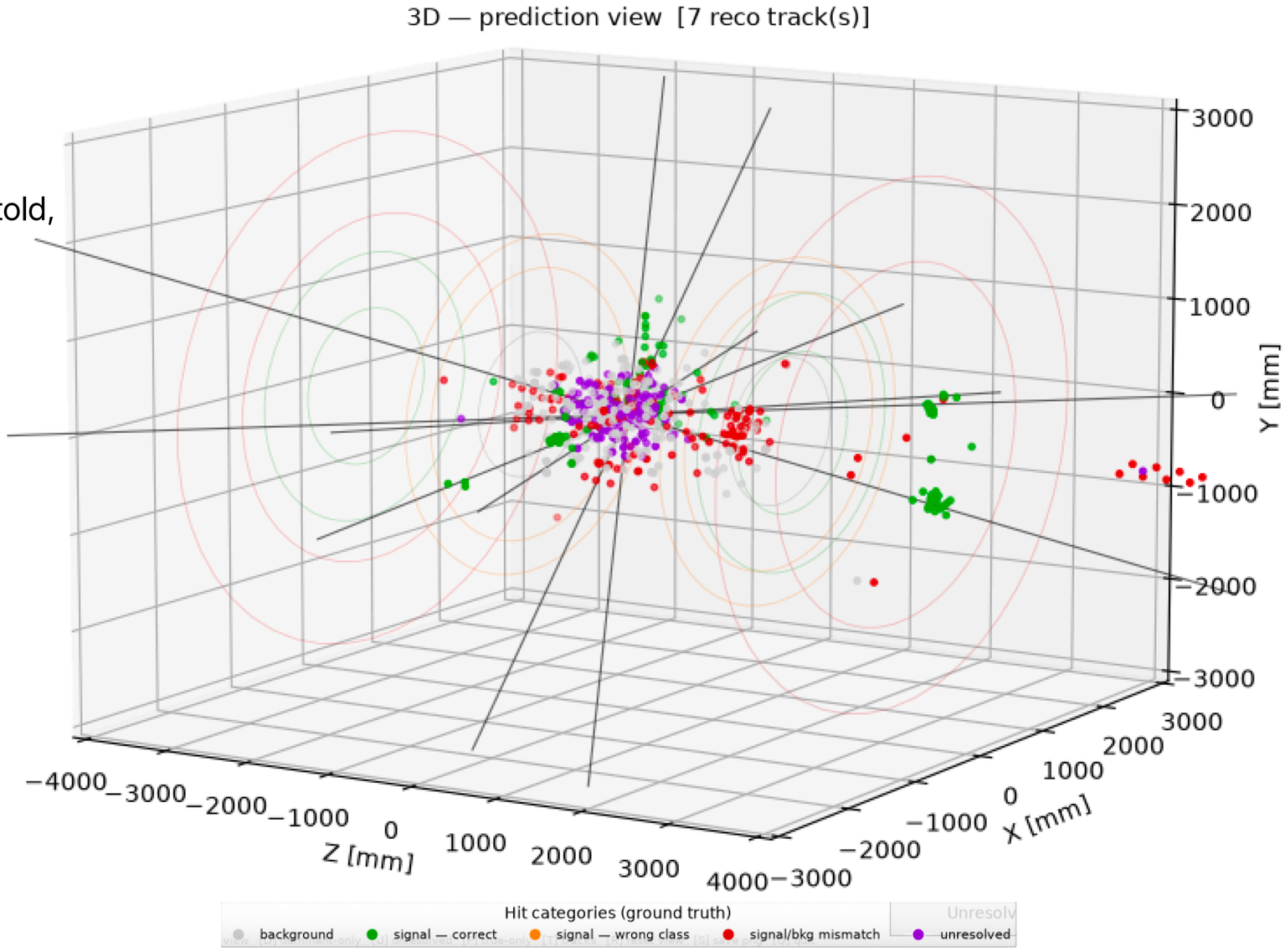
Unresolved class issue

"Unresolved" = a hit whose true origin the model can't be told, for one of two reasons:

- No truth link at all
- Undocumented generatorStatus gap [7000,10000]

Potential root cause:

Integer overflow or signed/unsigned conversion error in physics generation software



Issue being discussed on GitHub: https://github.com/AIDASoft/DD4hep/issues/1657#issuecomment-5280917973

Conclusion

Event Builder and GNN-Based Event Pre-Filtering

- End-to-end factory integrated in EICRecon and being implemented
- Eventbuilder including statistical relevance implemented and used by AI for event-level information.
- Robustness of event builder to be further improved
  NB: Orchestration in-place, but overall performances to be highly improved (!!)
- Next step on the AI part: the segmentation is being evaluated
- Provide some more comprehensive representation and tables to be confronted with ACTS benchmark
  (are the selected hits are good enough to be reconstructing via ACTS+CaloIsland)

Pending questions:

- Time-structured dataset in the future may be helpful and more representative
- timeError() for calorimeters to perform time coincidence.
- Issue being discussed on GitHub: https://github.com/AIDASoft/DD4hep/issues/1657#issuecomment-5280917