上海交通大學
SHANGHAI JIAO TONG UNIVERSITY

SJTU

Long-lived Particles Search with Deep Learning at Lepton Collider

Xiang Chen

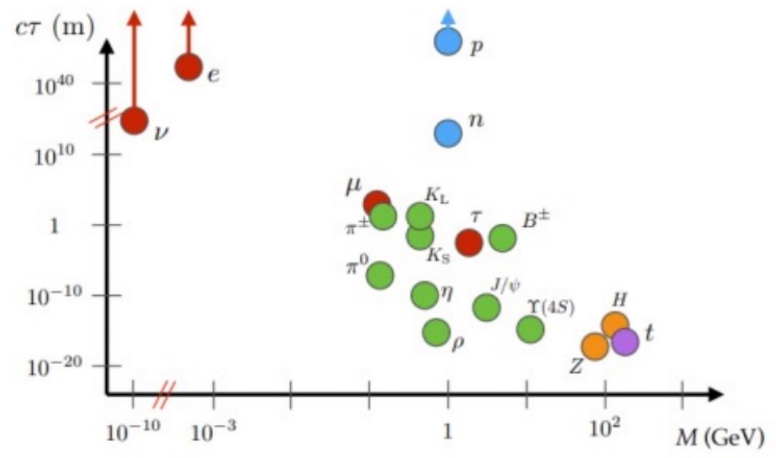
Shanghai Jiao Tong University

饮 水 思 源 • 爱 国 荣 校

New Physics Beyond SM- LLPs

Long-lived particles (LLPs) are important ways to new physics

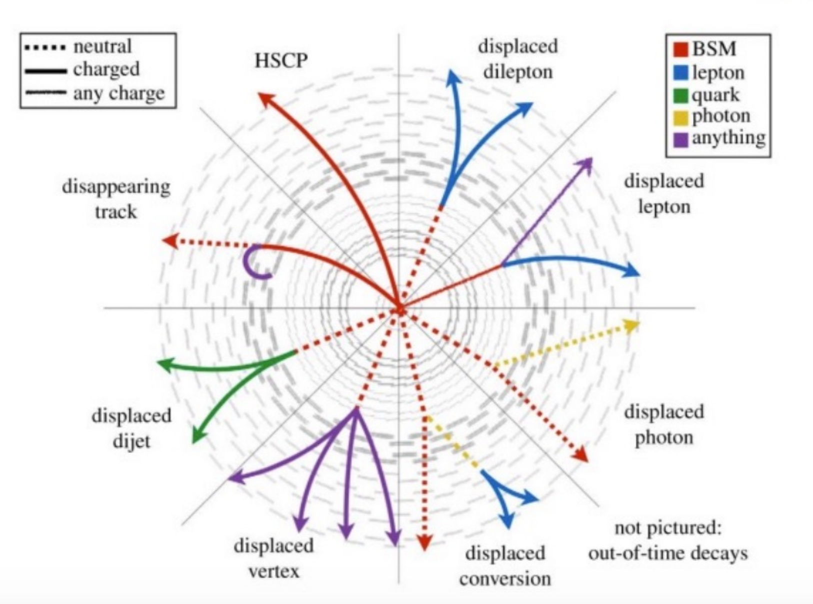
- Many particles in BSM models have a relatively long lifetime: weak coupling to SM particles, maybe new scalars, dark photons, ALP, SUSY….



JOURNAL OF PHYSICS G-NUCLEAR AND PARTICLE PHYSICS. 47(9) (2020)

LLP topology, a strong signature for detection:

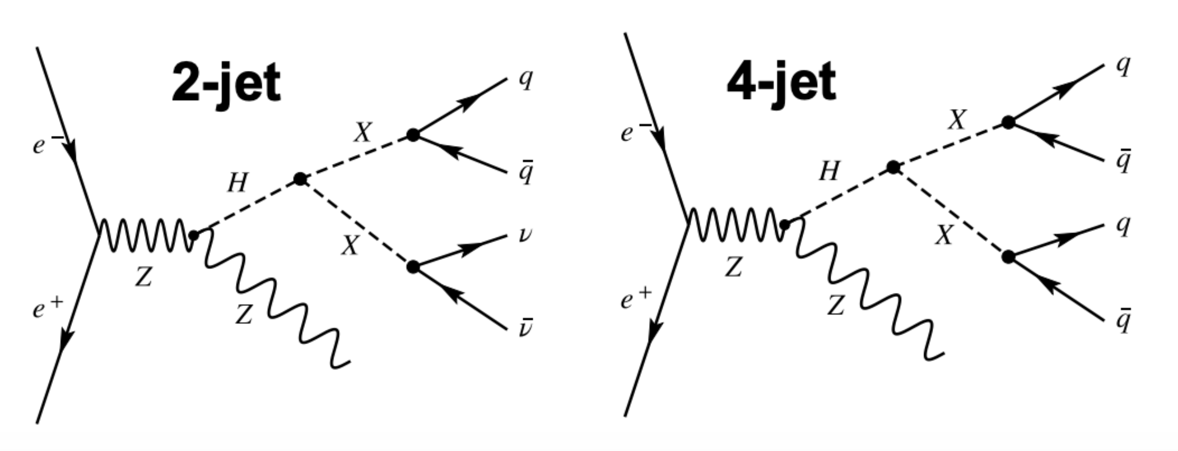
- Displaced vertex with a long distance from the main vertex
- Different performance for neutral particles: a burst of energy appearing of nowhere and far away from the collision point



Potential with Lepton Collider:

- The advantage of the lepton collider: clean environment
- Making use of deep learning techniques: Image recognition and pattern identification

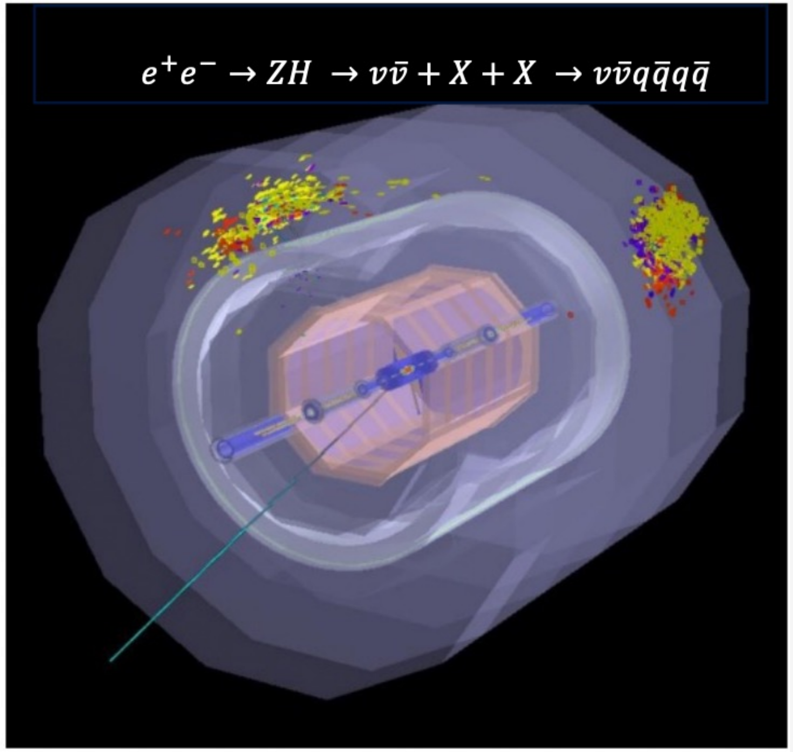
LLP at CEPC



We consider two LLP final state scenrios in CEPC: **2-jet and 4-jets final state**

- We use the **full simulation** sample using CEPC official software (v4) to an integrated luminosity of 20 ab^{-1}

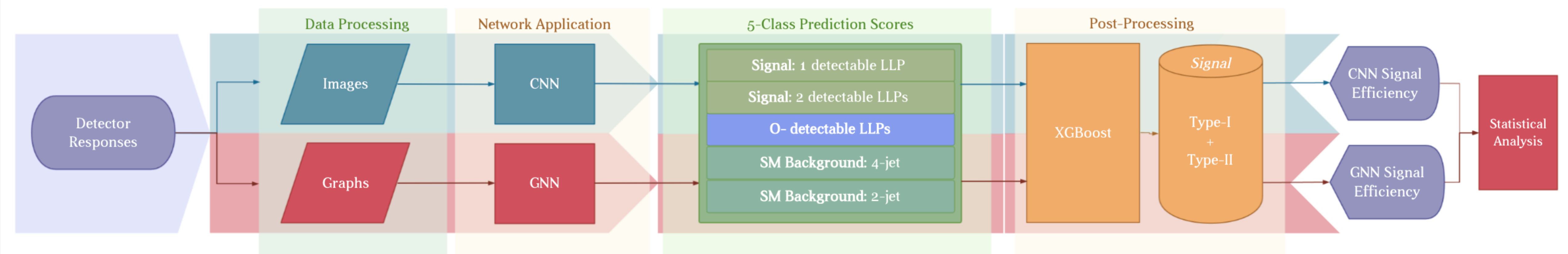
Acceptance (%)	Lifetime [ns]				
Mass [GeV]	0.001	0.1	1	10	100
1	100.00 ± 0.00	99.86 ± 0.01	48.76 ± 0.18	6.49 ± 0.09	0.67 ± 0.03
10	100.00 ± 0.00	100.00 ± 0.00	99.78 ± 0.01	46.80 ± 0.16	6.22 ± 0.08
50	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	99.31 ± 0.03	40.37 ± 0.16



- Displaced vertex of LLPs: $0 < r_{decay} < 6$ m
- Traditionally, need a special vertex reconstruction algothrim

Analysis Strategy

Advanced neural networks trained with low-level detector information:



- No need for vertex reconstruction and object reconstruction
- Detector response has been calibrated and detector resolution has been considered
- Universal treatment for all decay channels

Signal: LLP events

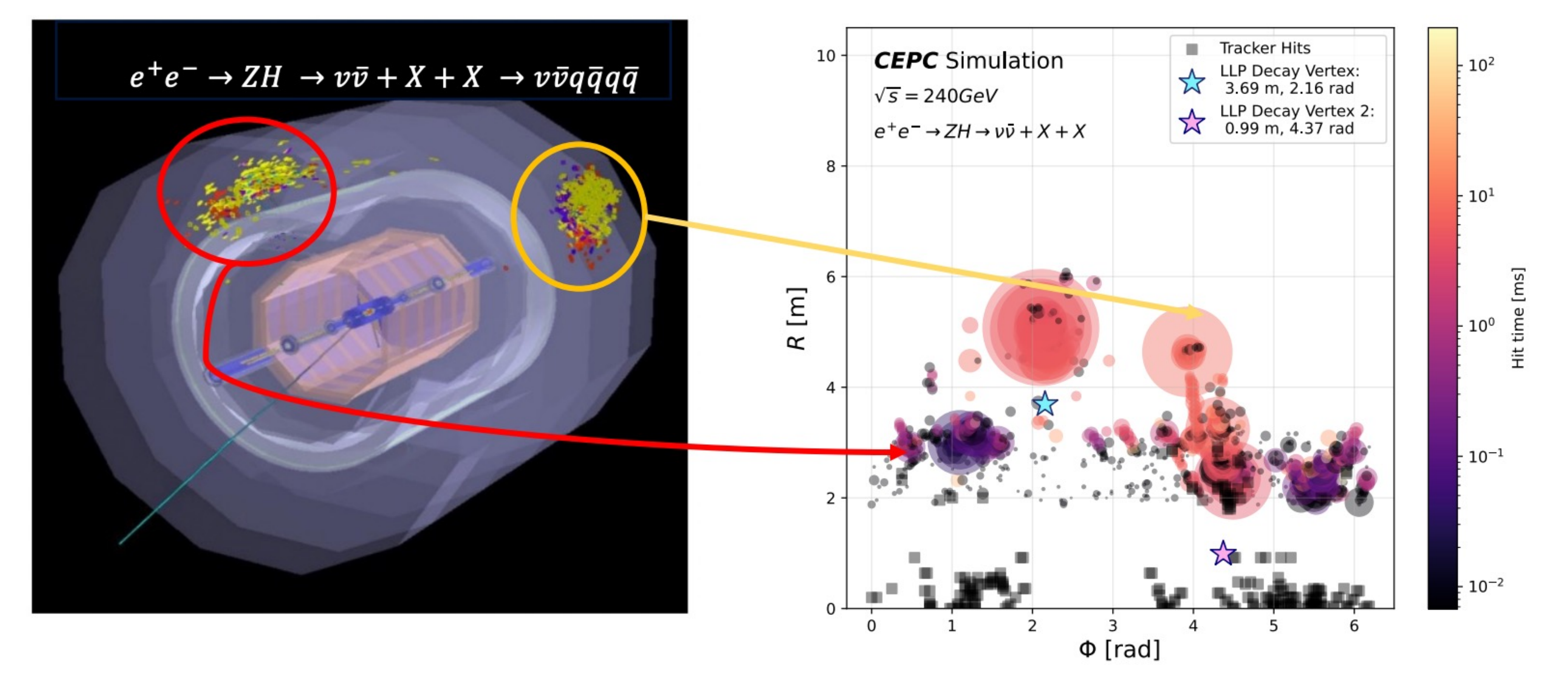
Two scenarios: 0-llp as signal and 0-llp as background

Background: SM process

Post-Processing: converting 5-class output to a 2-class classfication task

CNN: ResNet

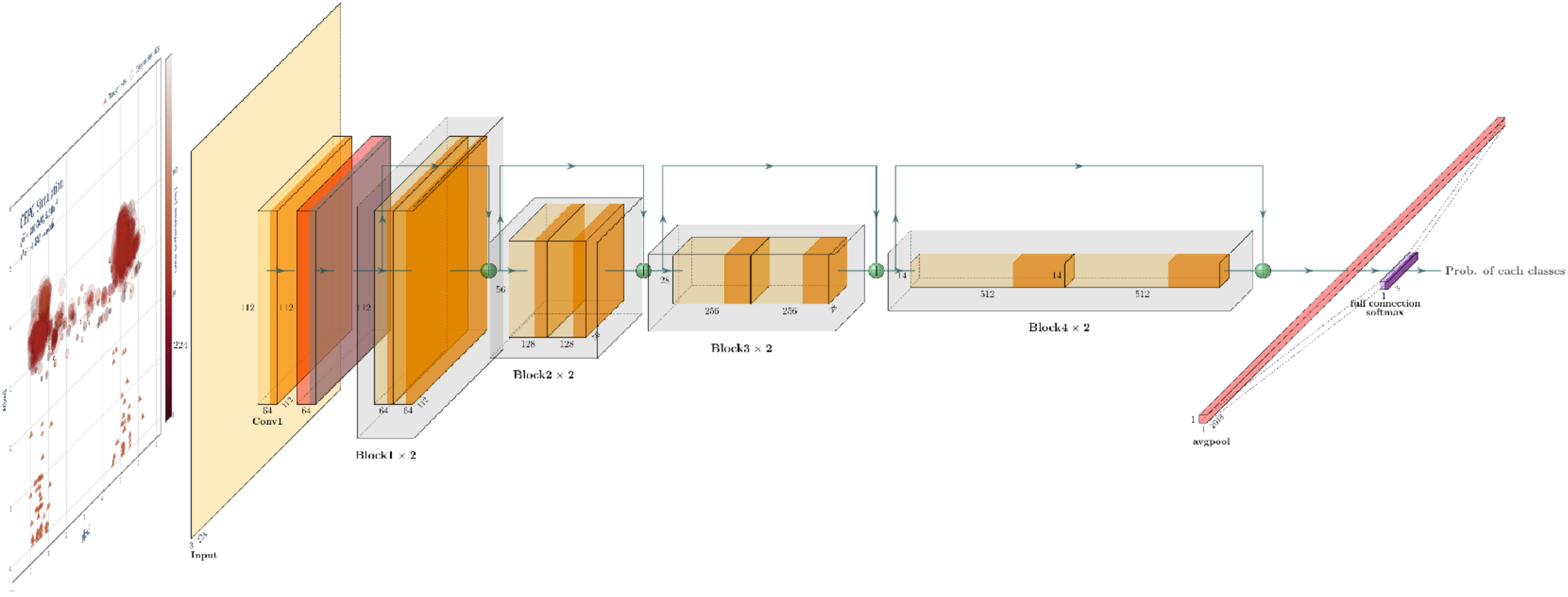
- Converting the detector information to 2D image
- Use ResNet18 model with the cross-entropy loss



Cross Entropy Loss: $loss = -[\omega_0 * y_0 \log(x_0) + \omega_1 * y_1 \log(x_1) + \omega_2 * y_2 \log(x_2)]$

Class 0: 2-fermion bkg $\omega_0 = 0.5$	Class 1: 4-fermion bkg $\omega_1 = 0.25$	Class 2: LLP Signal $\omega_2 = 0.25$

CNN: ResNet



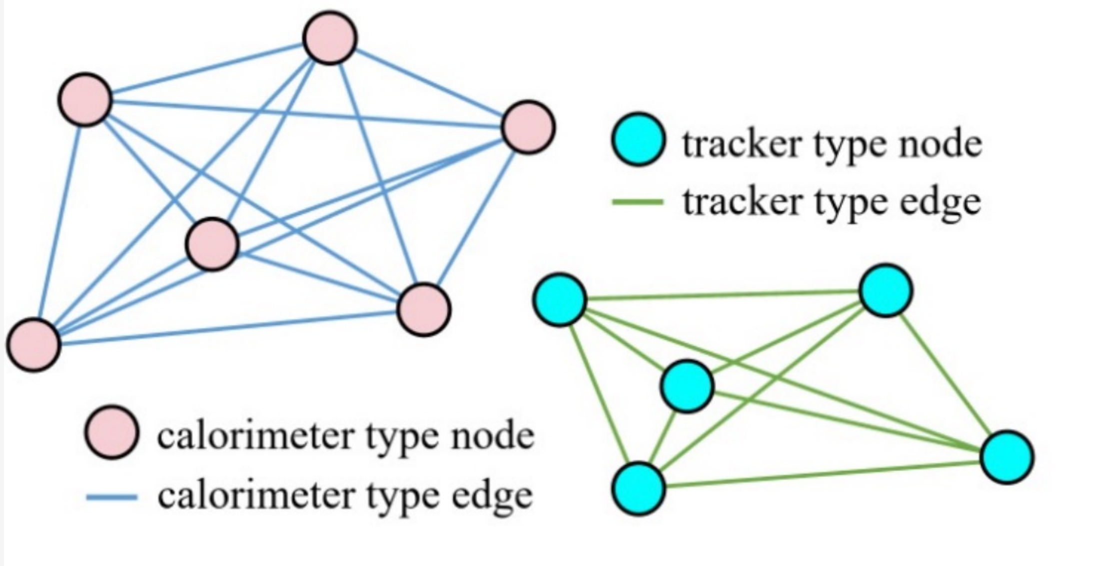
The structure of the ResNet18: 4 convolution blocks and 1 MLP layer
Output : 5–class score after softmax

GNN

The detector hits in the calorimeter and tracker are converted to point-cloud dataset

- Simple clustering is used to reduce graph complexity
- Nodes of the same detector type are fully connected

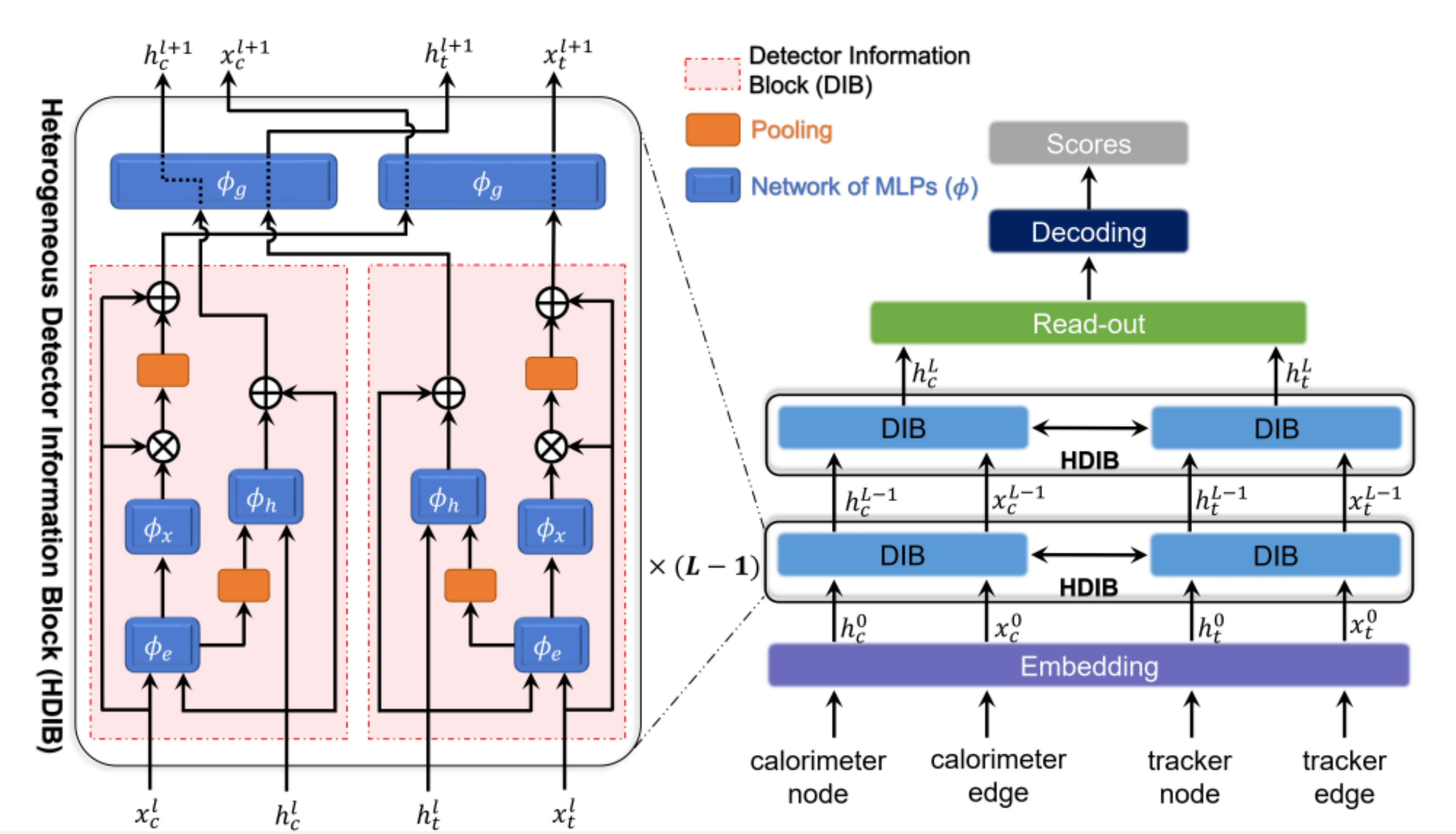
Features of nodes: calorimeter-type and tracker-type.
Features of edges: interaction between neighbor nodes



Features	Variable	Definition
calorimeter type node i	$\lvert x_i^\mu \rvert$	the space-time interval
	$\lvert p_i^\mu \rvert$	the invariant mass
	N_i	the number of hits
	η_i	$\frac{1}{2}\ln\frac{1+\frac{p_z}{p}}{1-\frac{p_z}{p}}$
	ϕ_i	$\arctan\frac{p_y}{p_x}$
	$\mathcal{R}_i$	$\sqrt{\eta^2+\phi^2}$
calorimeter type edge between node i and j	$x_i^\mu x_{j\mu}$, $p_i^\mu p_{j\mu}$, $x_i^\mu p_{j\mu}$, $p_i^\mu x_{j\mu}$ $\lvert x_i^\mu - x_j^\mu \rvert$, $\lvert p_i^\mu - p_j^\mu \rvert$, $\eta_i-\eta_j$, $\phi_i-\phi_j$, $\mathcal{R}_i-\mathcal{R}_j$	
tracker type node i	$\lvert r \rvert$	euclidean distance
	N_i	the number of hits
	η_i	$\frac{1}{2}\ln\frac{1+\frac{z}{r}}{1-\frac{z}{r}}$
	ϕ_i	$\arctan\frac{y}{x}$
	$\mathcal{R}_i$	$\sqrt{\eta^2+\phi^2}$
tracker type edge between node i and j	$\lvert r_i - r_j \rvert$, $r_i r_j$, $\eta_i-\eta_j$, $\phi_i-\phi_j$, $\mathcal{R}_i-\mathcal{R}_j$	

GNN

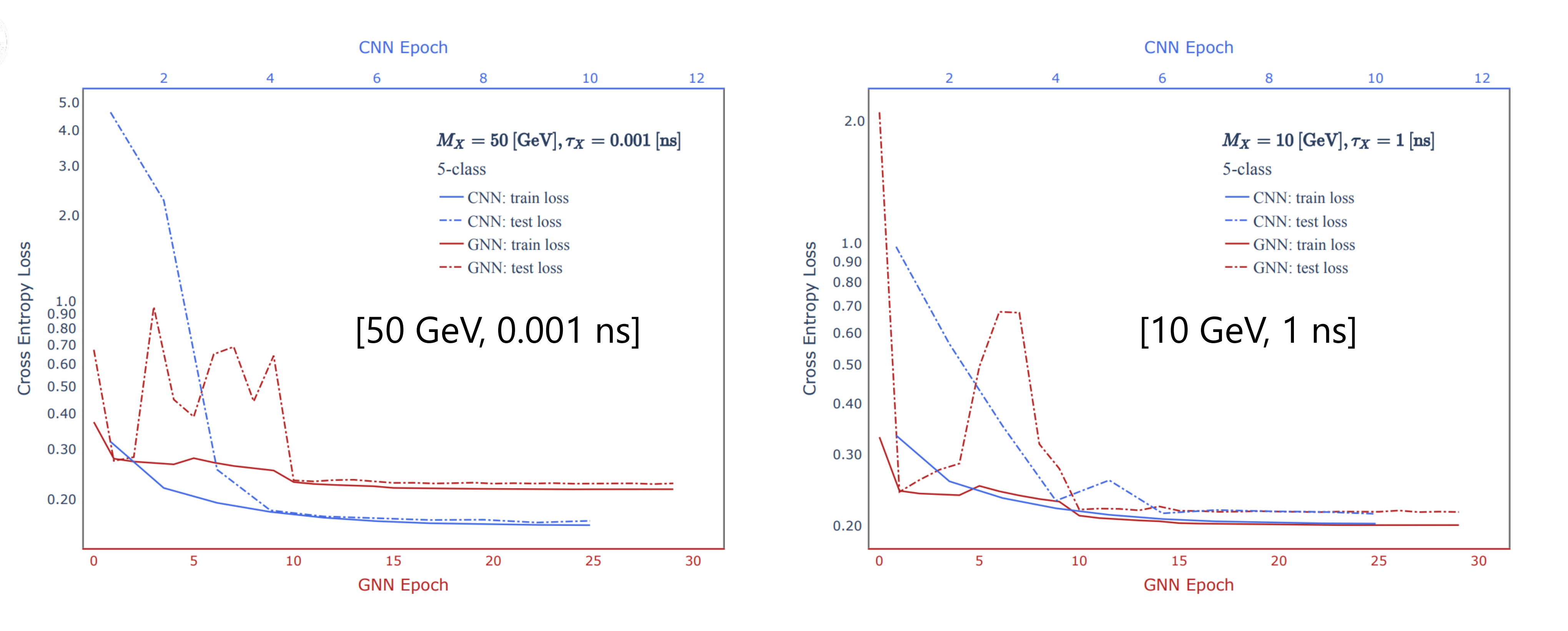
- DIB (detector information block): considering Lorentz equivariant in the graph network
- HDIB (Heterogeneous detector information block): The output of the DIB exchanges information in the MLP layer



Output : 5–class score after softmax

CNN & GNN Training Results

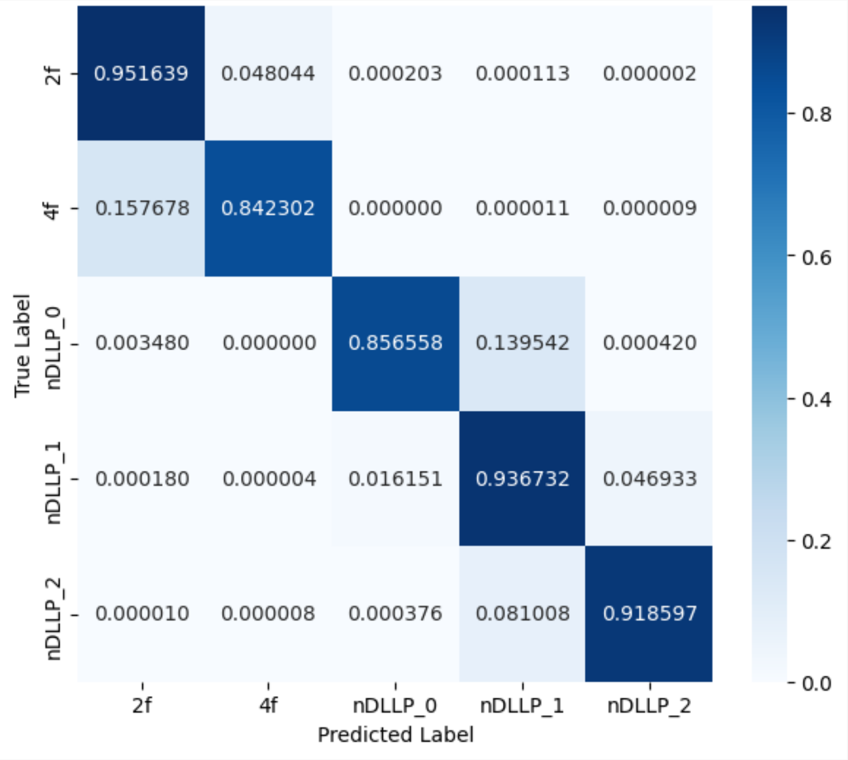
The training/ test loss looks reasonable for all mass and lifetime points



XGBoost

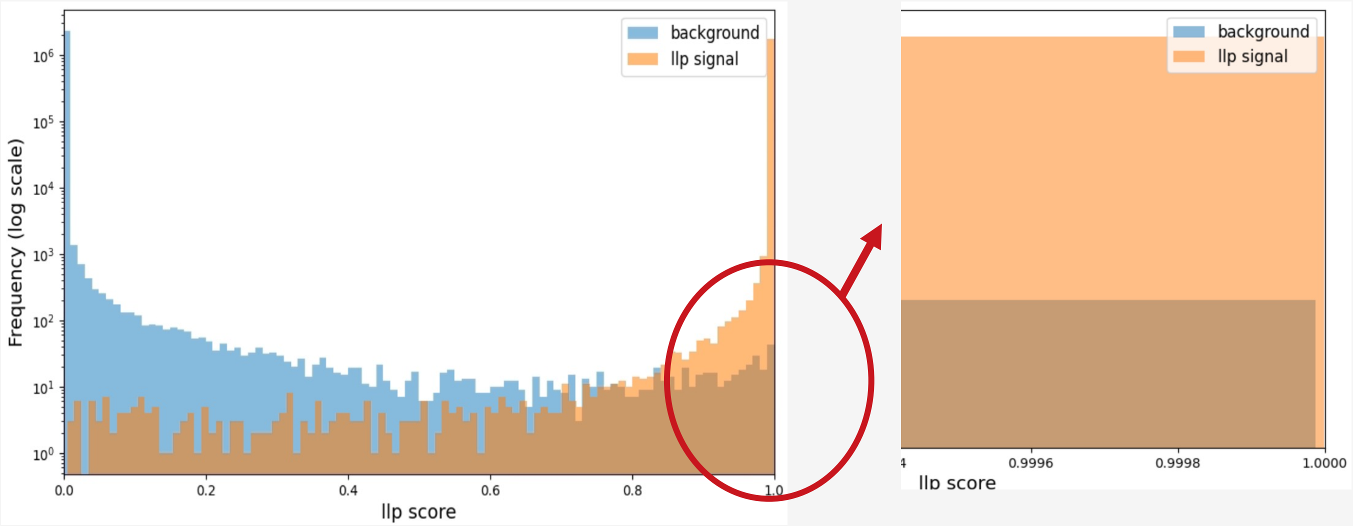
XGBoost is used to convert a 5-class classification task to 2-class classification task

Confusion matrix



Signal efficiency @ [50 GeV, 10 ns]: 95%

XGBoost



Signal efficiency @ [50 GeV, 10 ns]: 99%

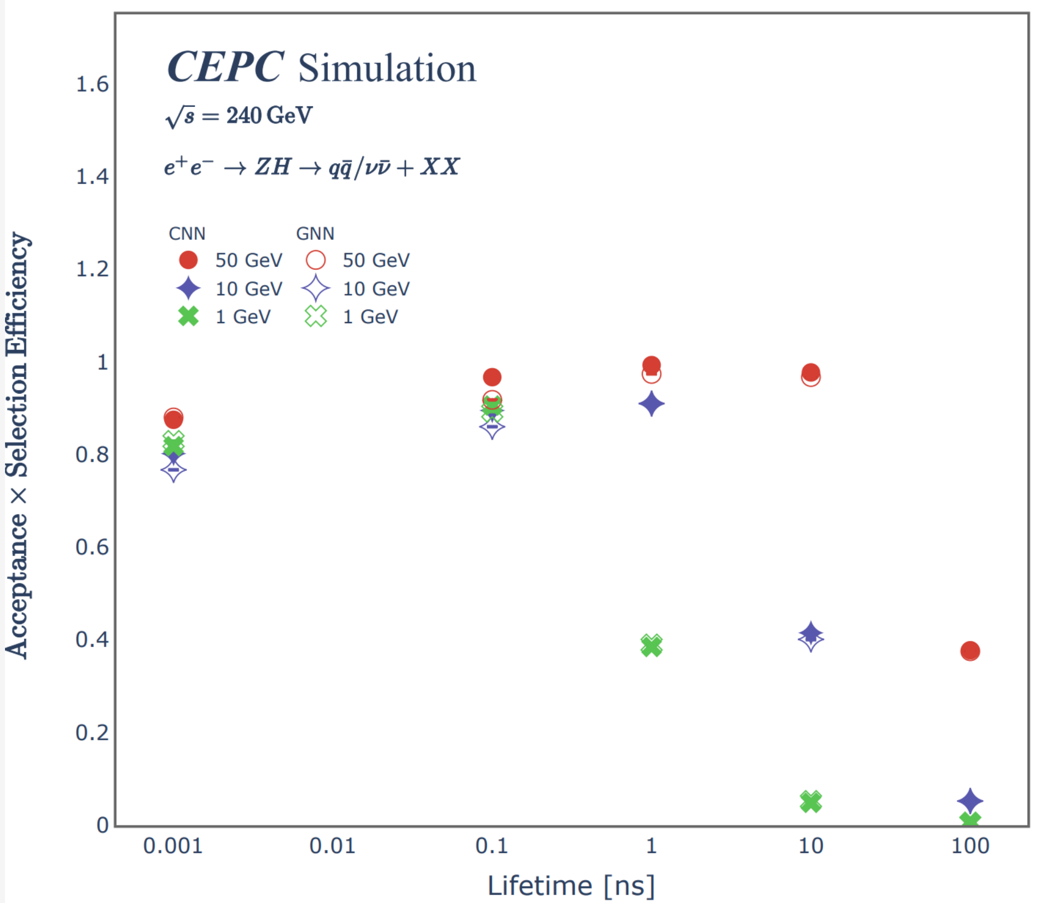
Background-free achieveable

Signal Efficiency

- Both CNN and GNN achieve high signal efficiencies while rejecting all SM backgrounds
  - Similar trend seen across different LLP mass and lifetime considerations.
- Systematics uncertainties of 2.0%
  - Luminosity and neutral network training uncertainties
  - Pile-up and cosmic rays background are neglected

Approach	Efficiency (%)	Lifetime [ns]				
	Mass [GeV]	0.001	0.1	1	10	100
CNN	1	81.8 ± 0.1	90.7 ± 0.1	78.9 ± 0.2	74.4 ± 0.6	76.5 ± 1.9
	10	80.2 ± 0.1	89.5 ± 0.1	91.2 ± 0.1	88.7 ± 0.1	83.6 ± 0.5
	50	87.5 ± 0.1	96.7 ± 0.1	$\boxed{99.3 \pm 0.0}$	98.4 ± 0.0	93.5 ± 0.1
GNN	1	82.9 ± 0.1	89.4 ± 0.1	79.9 ± 0.2	79.9 ± 0.6	80.2 ± 1.8
	10	76.7 ± 0.1	86.0 ± 0.1	91.2 ± 0.1	85.7 ± 0.2	83.7 ± 0.5
	50	88.0 ± 0.1	91.8 ± 0.1	97.4 ± 0.1	97.4 ± 0.1	93.0 ± 0.1

Best efficiency at 99% (50 GeV, 1ns)
0-llp is considered as signal



Comparison between Non-XGBoost and XGBoost

- XGBoost can reduce the uncertainty and improve the signal efficiency in both scenarios
  - Backgrounds are all rejected in both scenarios
- Training uncertainty is calculated by (standard deviation)/(mean value)
  - Each training dataset contains $5*10^5$ events

0-llp belongs to signal

	50 GeV 0.1 ns	50 GeV 10ns	1 GeV 10 ns
CNN	0.81 (9.9%)	0.95 (1.1%)	0.57 (13%)
CNN+XGBoost	0.96 (1.1%)	0.99 (0.33%)	0.69 (4.7%)

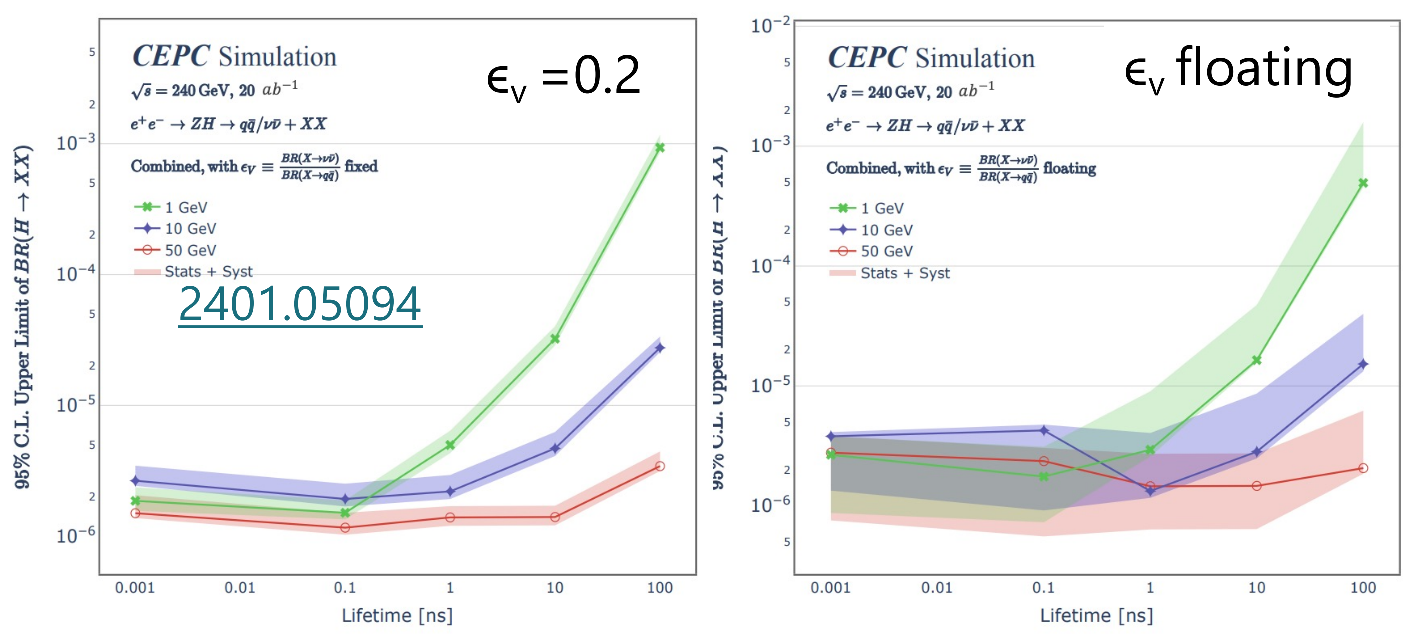
0-llp belongs to background: overlap with SM background or other BSM physics

	50 GeV 0.1 ns	50 GeV 10ns	1 GeV 10 ns
CNN	0.81 (9.9%)	0.74 (10%)	0.14 (27%)
CNN+XGBoost	0.96 (1.1%)	0.95 (1.2%)	0.19 (23%)

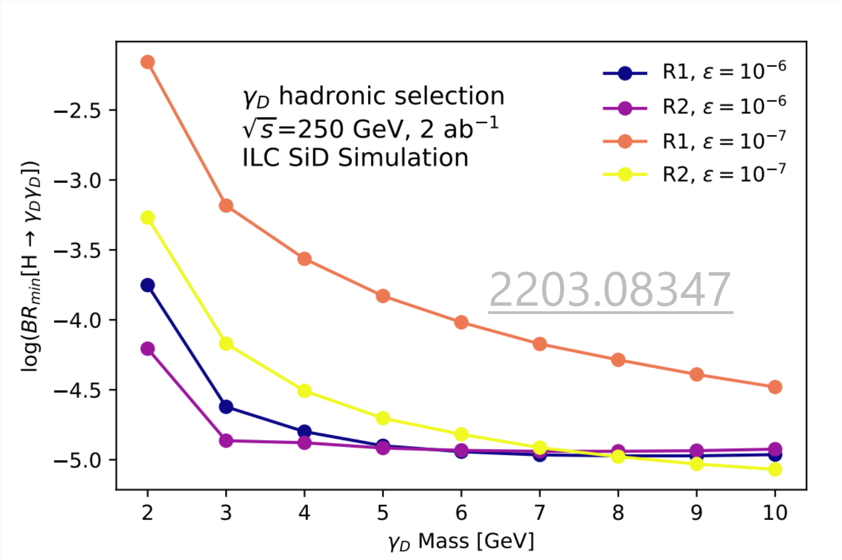
LLP Search Limits

- The best expected limit of BR(H → XX) achieves 10^{-6}
- Outperforming the current limit from ATLAS and CMS by 2 - 3 orders of magnitude
- An order of magnitude better than the ILC when the lifetime of LLP is over 1ns

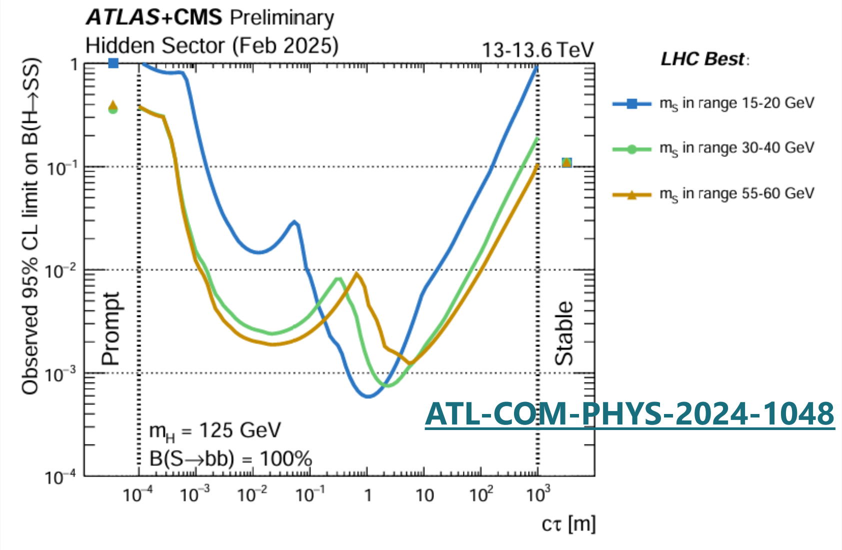
$$\epsilon_v = BR(X \to \bar{\nu}\nu)/BR(X \to \bar{q}q)$$



Best limit: ~10^{-6}



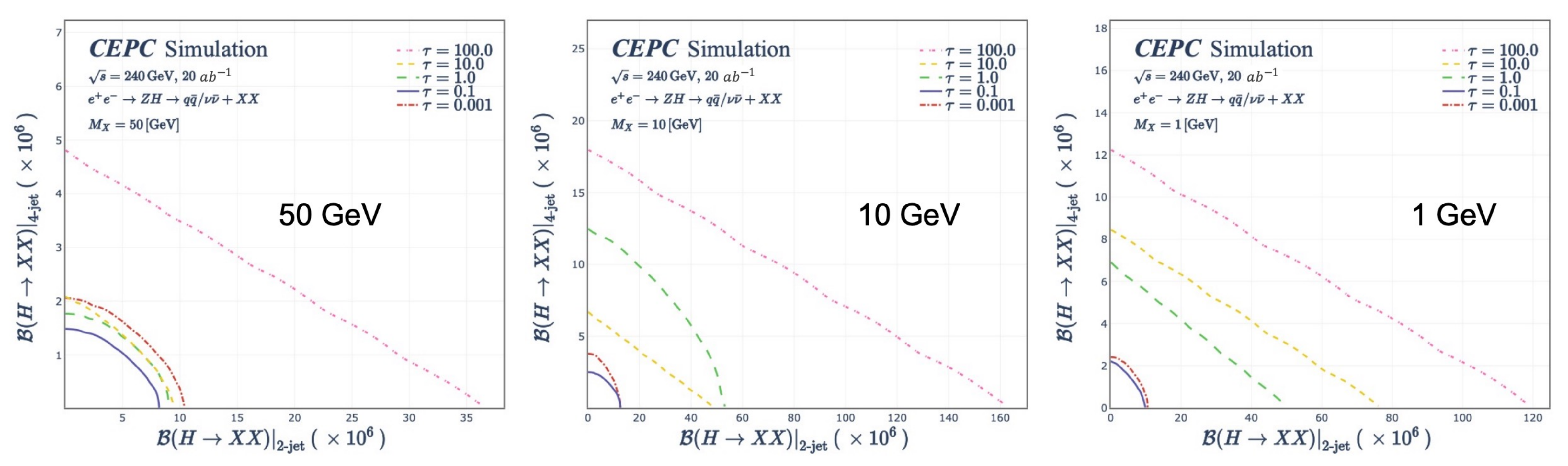
ILC Best limit: ~10^{-5}



LHC Best limit: ~10^{-3}

LLP 2D Sensitivity

- We also provide the 2D likelihood for 95% Confidence Level upper limit on BR(H → XX) with 2 jets and 4 jets final state
  - Keep ϵ_ν= BR(X→$\bar{\nu}\nu$)/BR(X→$\bar{q}$q) float during limit extraction
- Higher mass and shorter lifetime scenarios have better sensitivities



Local-contrastive-learning Machine (LCLM)

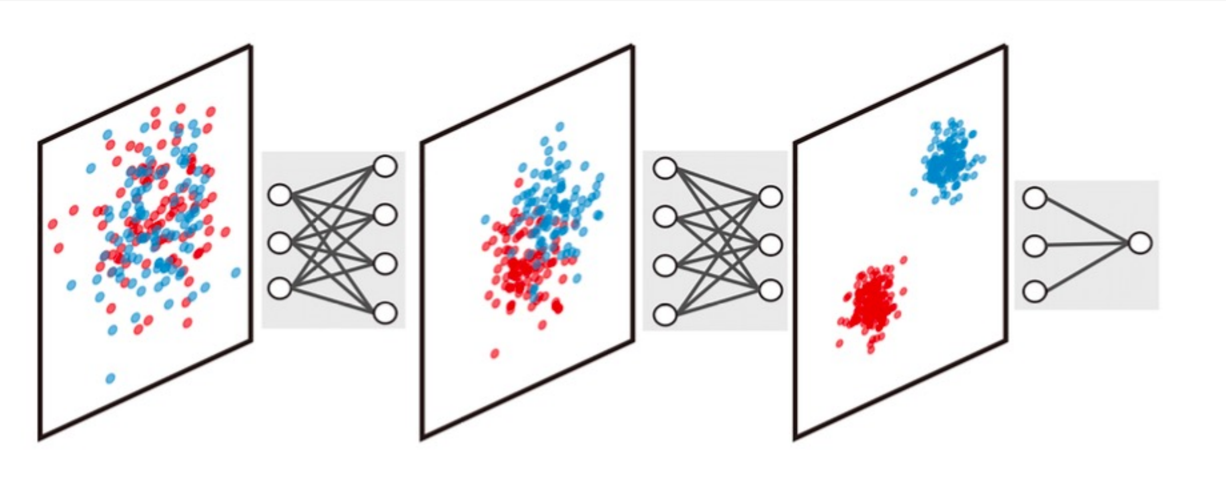
LCLM network utlizes the concept of the fermion and boson:

- Different pair of samples -> fermion
- Similar pair of samples -> boson

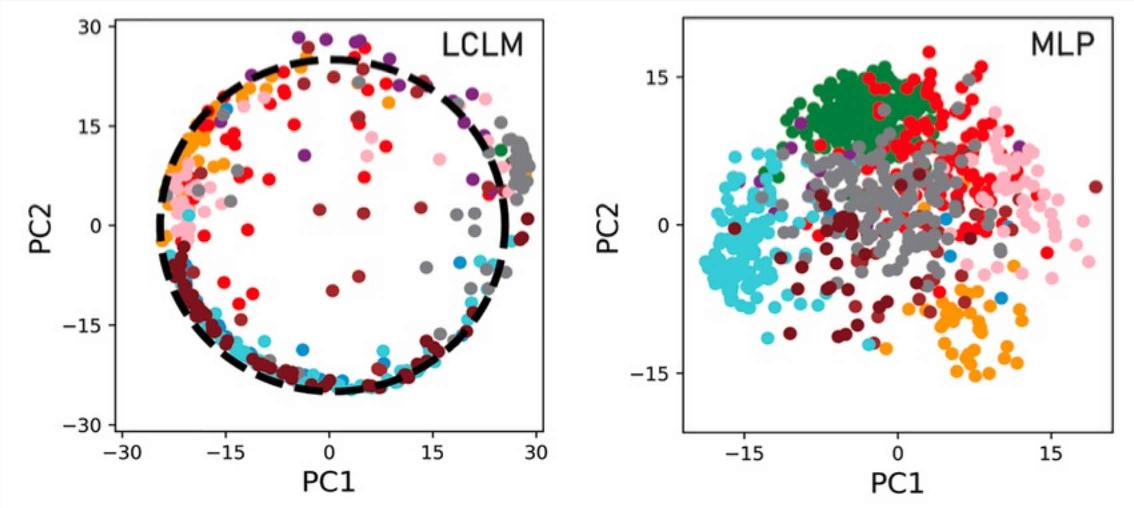
$$\mathcal{L} = \sum_{\mu=1}^{P} \frac{1}{2}\left[\sigma^{\mu} D_{\mu}^{2} + (1-\sigma^{\mu})\varphi\left(d_F - D_{\mu}^{2}\right)\right] + \frac{\lambda_w}{2}\|\mathbf{w}\|_2^2,$$

Fermi-boson network has good performance in both discrimination and robustness against class-preserving perturbations.

Fermi-boson network



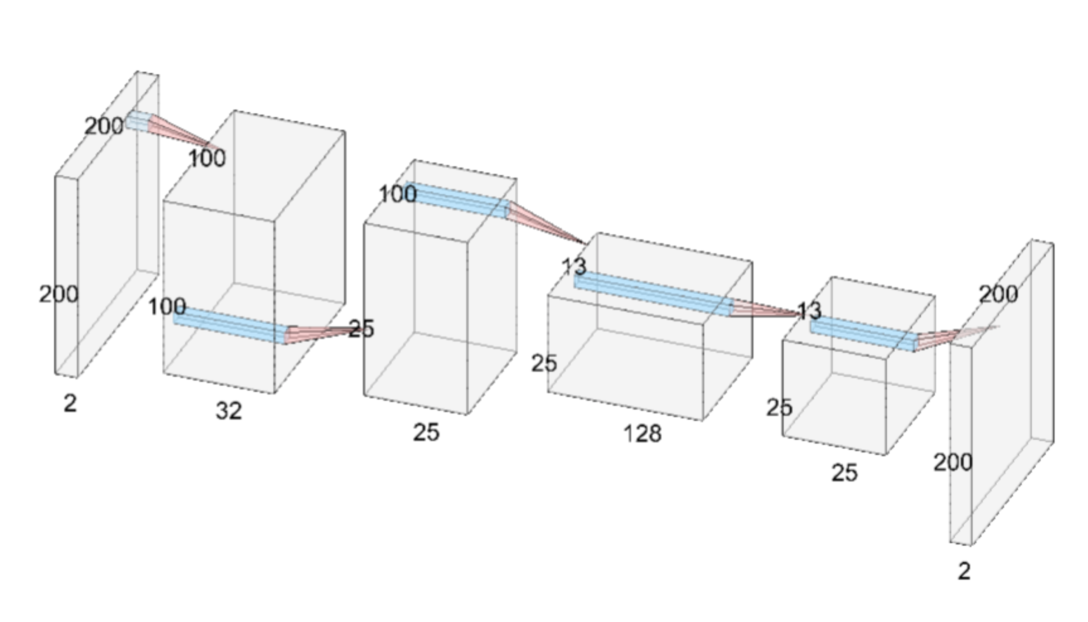
PCA outputs



[Sci. China Phys. Mech. Astron. 68, 210511 (2025)]

Fermi-Boson Network

- Training the CNN block with Fermi-boson loss: CNN block has 5 layers
- During calculating the Fermi-boson loss, a simple average pooling is applied
- Replacement: replacing the input by the LCLM output, [batch, 2, 100,100]



Training procedure

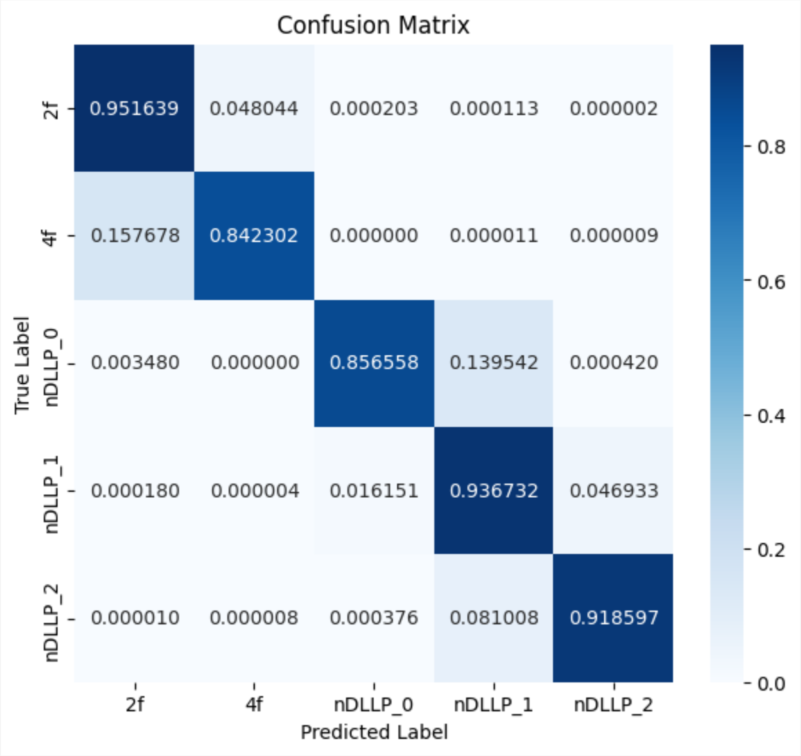
LLP dataset in shape of [200, 200] → LCLM *CNN block* *5 layers* → Dataset applied by the LCLM → ResNet 18 → Scores of 5 classes by the Softmax (**Confusion Matrix**)

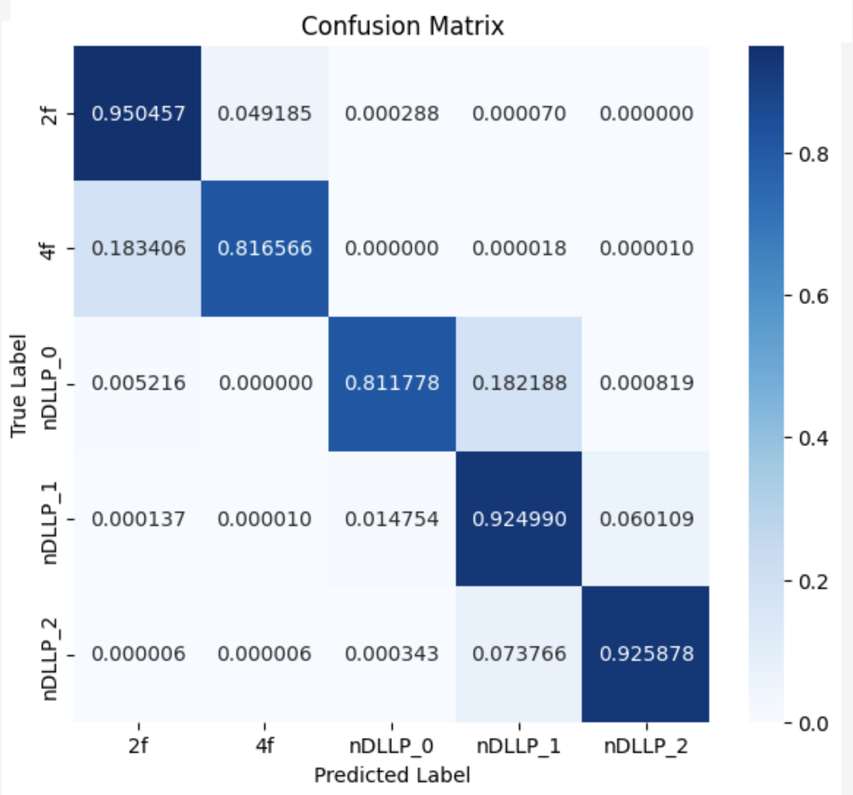
Trained for 10 epochs

Trained for 10 epochs

Fermi-Boson Network Performance

- The confusion matrix of the LCLM shows that the separation between the background and signal is larger for LCLM:
  - 5-class classification during training
    - SM Background: 2 fermions(2f) and 4 fermions (4f)
    - LLP Signal: nLLPs_0,1,2



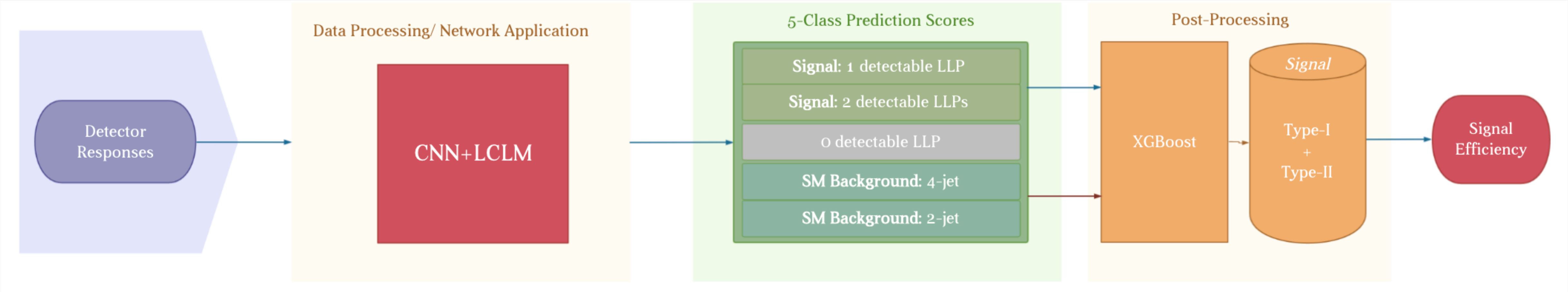


Default CNN, trace: 4.5
mass=50 GeV, lifetime=10 ns

LCLM trace: 4.4
mass=50 GeV, lifetime=10 ns

Fermi-Boson Network Performance

- LLP signal efficiency are derived after applying a score threshold to reject all backgrounds.
- XGBoost is used to for further optimizations when combing signal and background regions.
- Small difference seen in efficiency and uncertainty



category	LCLM+XGBoost (5×10^5 events)	CNN+XGBoost (5×10^5 events)
Signal efficiency	**0.99**	**0.98**
Training uncertainty	**0.30%**	**0.33%**

Summary and Outlook

LLPs Search with Deep Learning at Lepton Collider

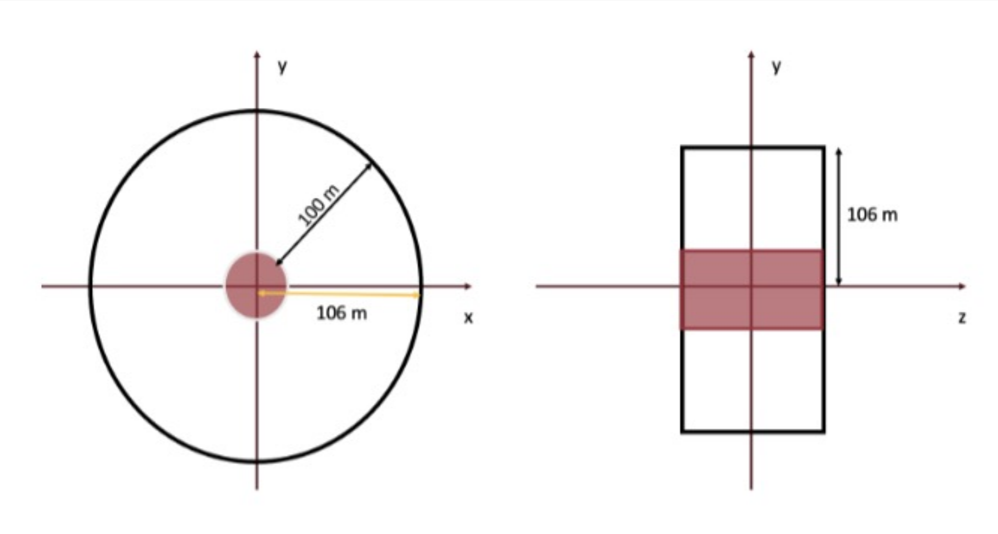
- Clean environment with distinct detector signature
- Best exclusion limit on BR(H→LLPs) @ 20 ab^{-1}: $1.2*10^{-6}$
- 1D and 2D sensitivity results

- Significant enhancement from deep learning techniques
  - Simplified analysis strategy compared to the traditional method
    - Low-level detector information without full reconstruction
  - Signal efficiency as high as 99%
  - biggest improvement in shorter lifetime region
- Application of LCLM algorithm to improve network stability and reduce training uncertainty
  - Study ongoing

Backups

Far Barrel Detector (FBD)

- The gain attributable to the FBD, located 100 meters outside the near detector, can be estimated by comparing the LLP signal yields (background free)



$$F_{gain} = \frac{\Delta\Omega}{4\pi}\left(\frac{1 - e^{-\frac{L+\Delta L}{d}}}{1 - e^{-\frac{L}{d}}} - 1\right) + 1$$

- L is the length of the muon to IP
- d is the exptected decay length of LLP
- $\frac{\Delta\Omega}{4\pi}$ is the angular acceptance
- ΔL is the gap between the FBD and the muon detector

- $\mathrm{p_{llp}} = 70\ \mathrm{GeV}, m_{llp} = 1\ GeV, \tau = 100ns, d = \gamma\beta c\tau = \frac{P}{m}c\tau \approx 70 * 3 * 10^{-8}\frac{m}{s} * 100 * 10^{-9}\mathrm{s} = 2100\mathrm{m}$
- The angular acceptance (ΔΩ /4π) is 0.72 (non-uniform distribution in theta, see backup slides)
- The gain at 100ns for 1GeV FBD is $F_{gain} = 1 + 0.72 \ * (\frac{1-e^{-\frac{106\,m}{2100\,m}}}{1-e^{-\frac{6\,m}{2100\,m}}} - 1) = 12.7$

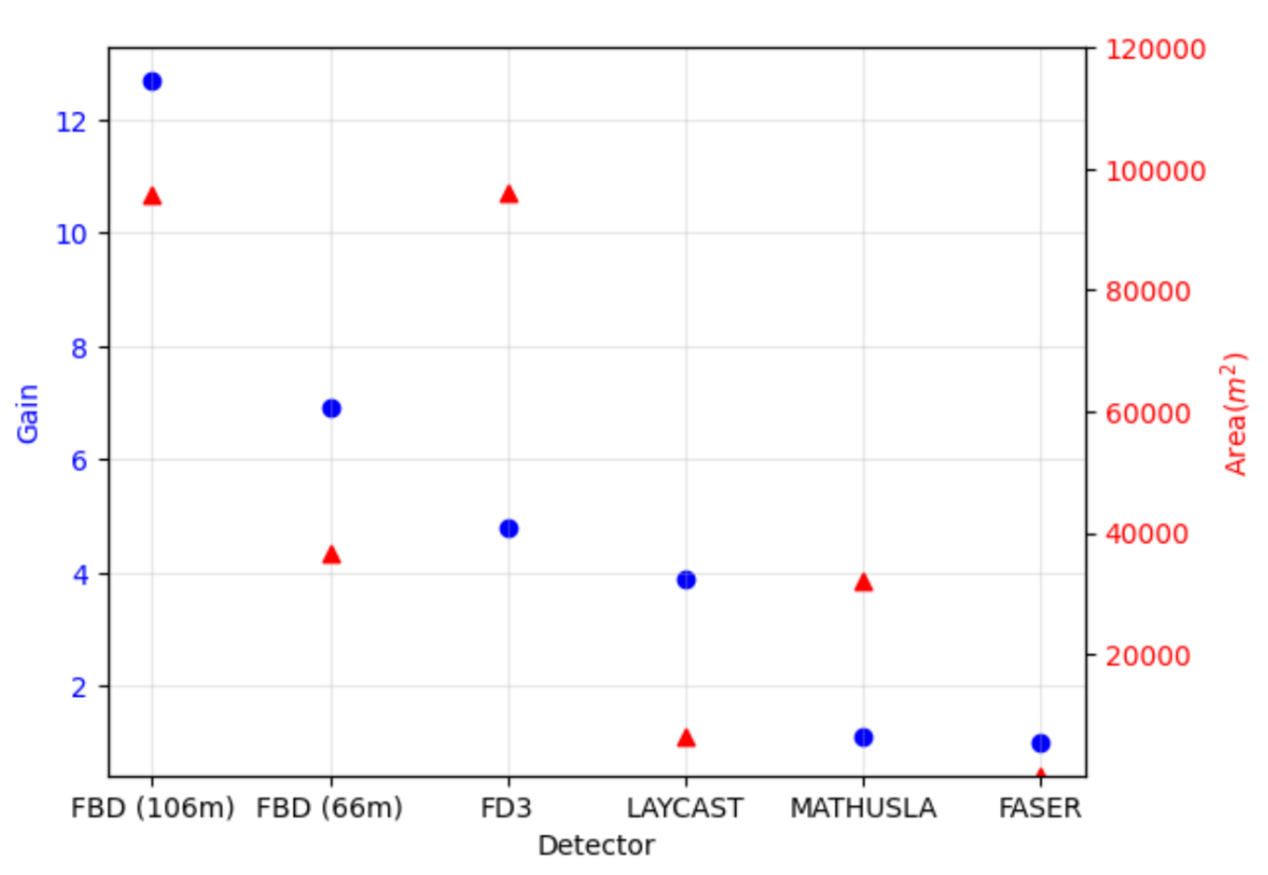
Far Barrel Detector (FBD)

- The gain attributable to the FBD, located 100 meters outside the near detector, can be estimated by comparing the LLP signal yields (background free)

$$F_{gain} = \frac{\Delta\Omega}{4\pi}\left(\frac{1-e^{-\frac{L+\Delta L}{d}}}{1-e^{-\frac{L}{d}}} - 1\right) + 1$$

- L is the length of the muon to IP
- d is the exptected decay length of LLP
- $\frac{\Delta\Omega}{4\pi}$ is the angular acceptance
- ΔL is the gap between the FBD and the muon detector

- Estimating the cost by the surface area and comparing the gain factor with different LLP detector scenarios
  - One advantage comes from good angular acceptance
  - Another big advantage comes from combined detection with near detector sharing time information



Cut-based Method

Selections	LLPs Signal with $Z \to j\bar{j}$	$ee \to q\bar{q}$	$ee \to ZH$
generated	1.0×10^6	2.5×10^8	0.99×10^7
decay in muon detector	134559	6516657	796596
$\lvert m_{q\bar{q}} - m_Z \rvert < 15GeV$	113723	4013875	39631
$\lvert m_{q\bar{q}} - m_H \rvert < 15GeV$	104942	229703	26862
$0.23 < y_{12} < 0.72$	93,517	129,546	20,041
$E_{2jets} > 30\text{GeV}$	69,468	72	16
$min(\Delta T_{j_1}, \Delta T_{j_2}) > 3\text{ns}$	68,368	50	11
Efficiency	**50.80%**	7.7×10^{-6}	1.4×10^{-5}