

Ultra-High-Resolution Detector Simulation with Intra-Event Aware GAN and Self- Supervised Relational Reasoning

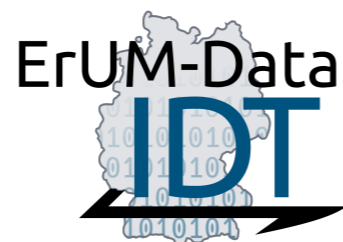
Hosein Hashemi ¹, Nikolai Hartmann ¹, Sahand Sharifzadeh ², James Kahn ³, Thomas Kuhr ¹

Ludwig-Maximilians-Universität München ^{1,2}

The ORIGINS Excellence Cluster ¹

Helmholtz AI ³

DeepMind ²



HELMHOLTZAI

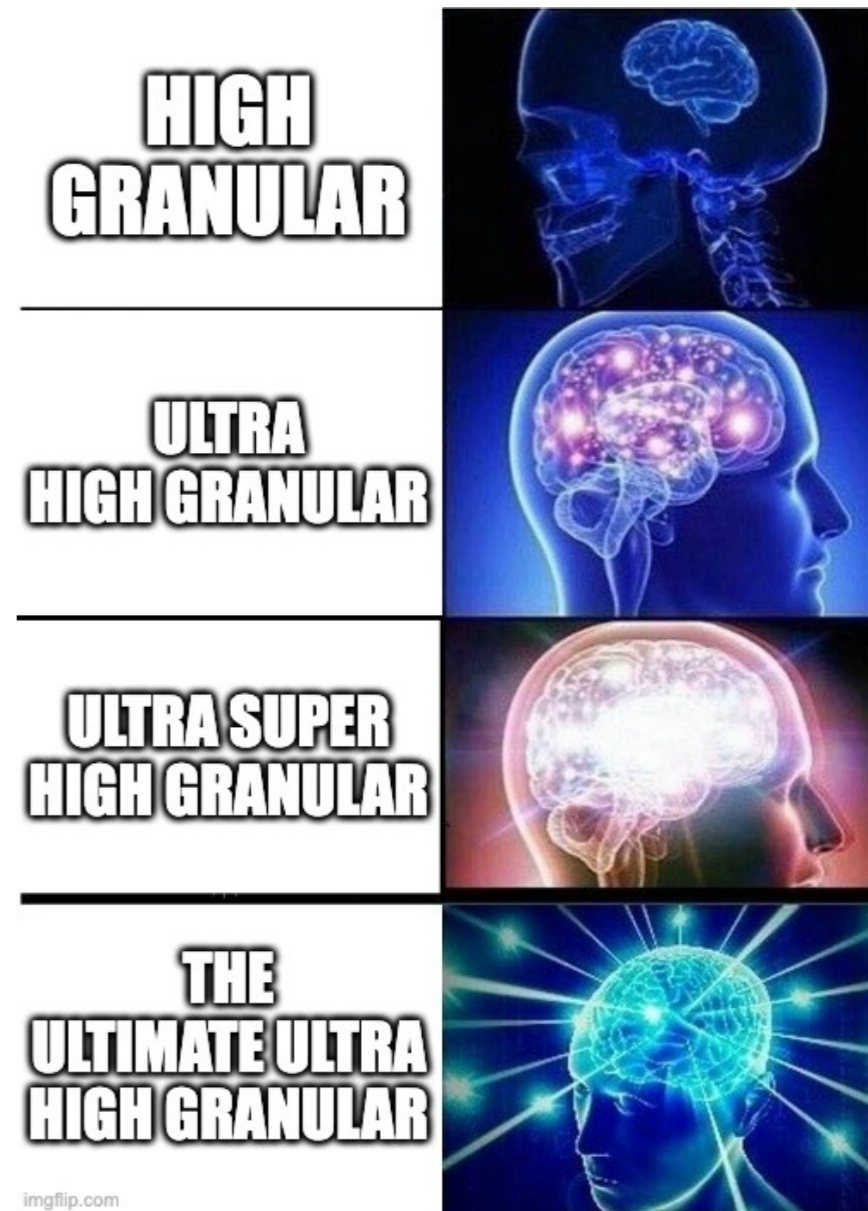
Introduction

► Let's get this straight once and for all:

- $O(100)$ → Low-granularity (Jets data)
- $O(1000)$ → Mid-granularity (simplified Calorimeter data)
- $O(10000)$ → High granularity (ILD Calo. prototype, D3 Calo. Challenge)
- $+O(10000000)$ → Ultra high granularity (HL-LHC, PXD)

► **The Pixel Vertex Detector (PXD):** is the innermost sub-detector for charged particles at Belle II.

► The PXD is assembled from **40 sensors**, where each sensor consists of 250×768 pixels → more than **7.5M information channels** per event → “**Ultra-High granularity**”



imgflip.com

Introduction

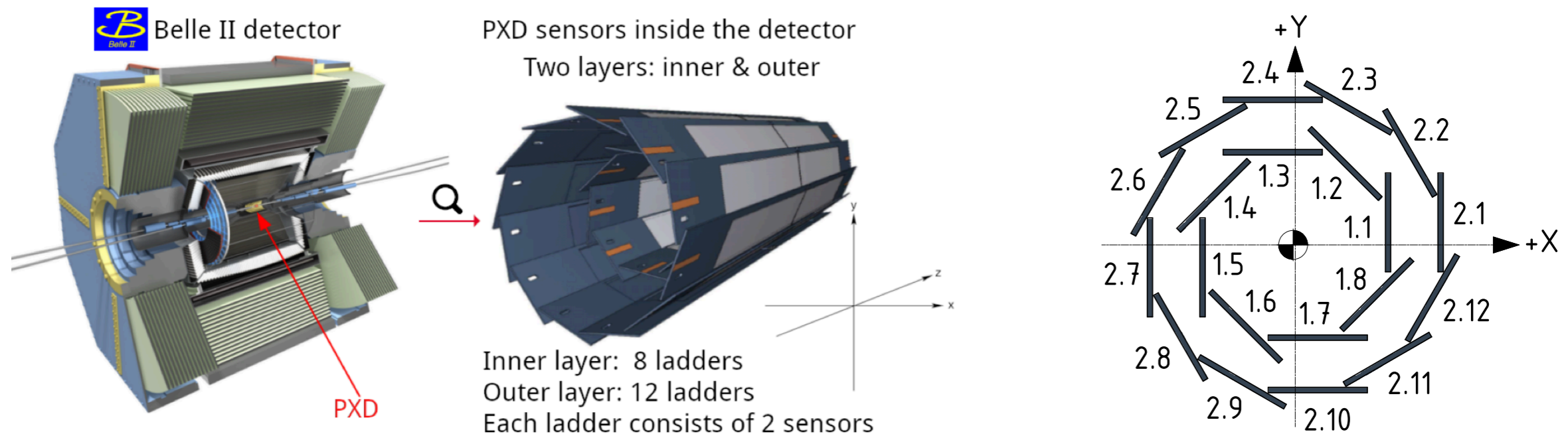
► Having an **annulus-like topology**:

- **The inner layer**: 16 sensors implemented into 8 ladders
- **The outer layer**: 24 sensors implemented into 12 ladders

► **Problem**:

- High time-complexity for simulating background events online
- High space-complexity for producing and storing PXD background data

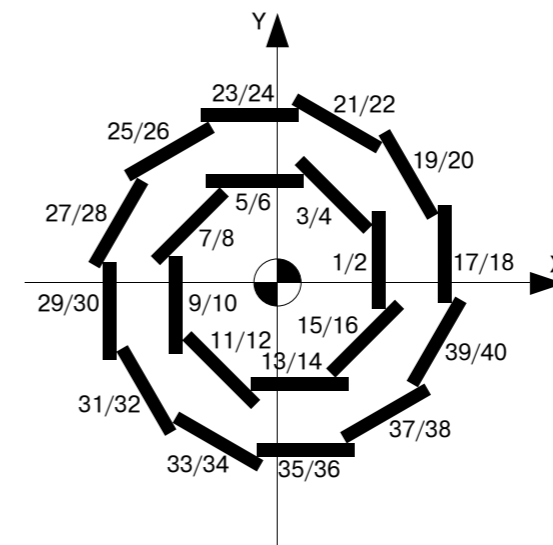
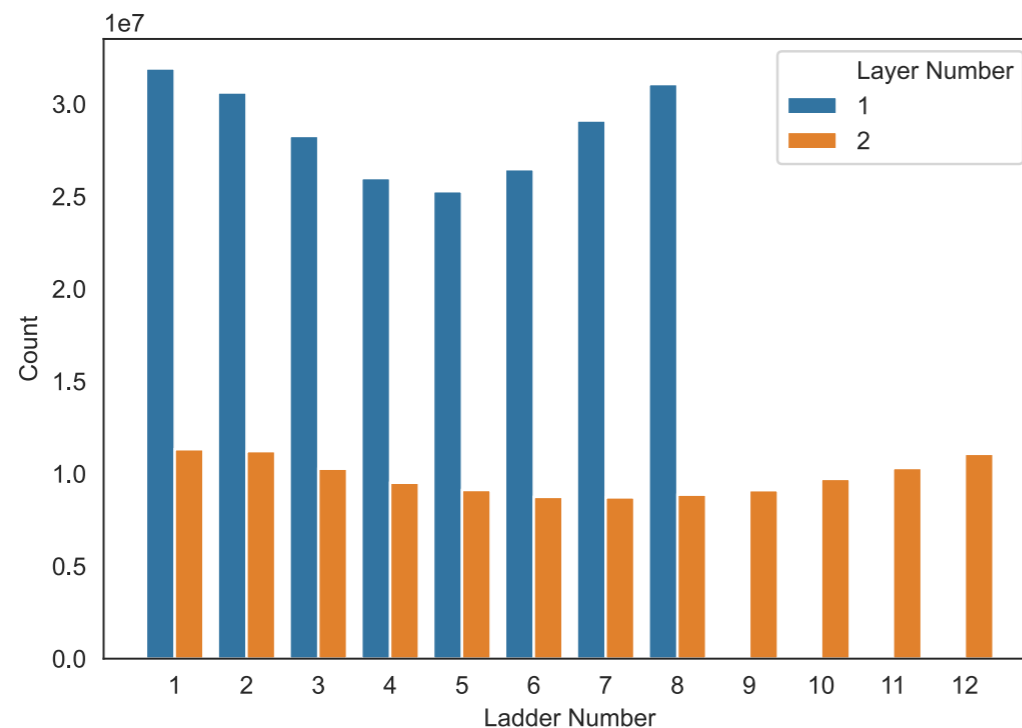
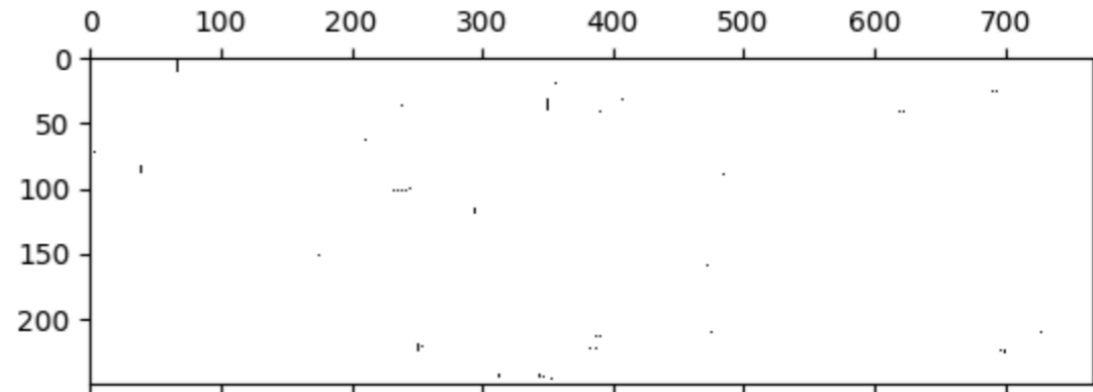
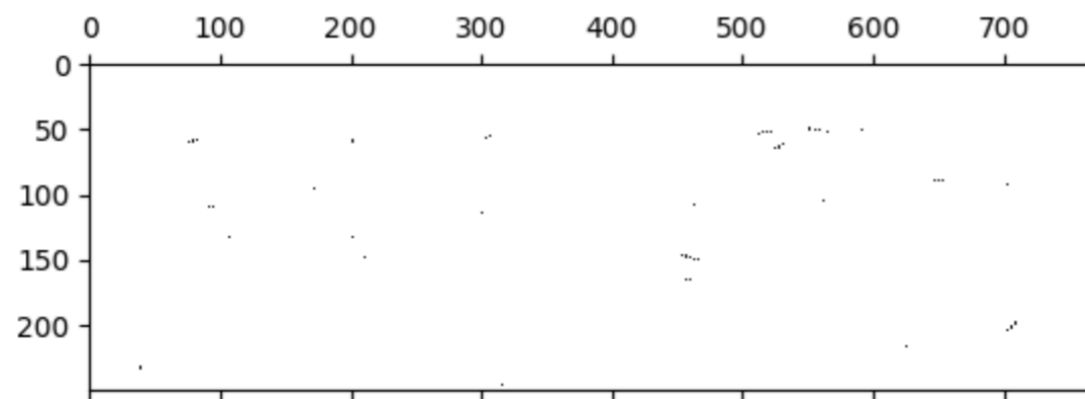
► **Solution**: **Amortised Simulation**: Generate the PXD background on the fly of analysis



Challenges

► PXD background generation challenges:

- **Ultra-High Resolution Data 40x250x768** → More than 7.5 M information channels
- **Non-Trivial (Annulus) Detector Topology** → [[1-40]] Sensor dependent information are not sequential
- **Extremely Sparse and Fine-Grained events** → Lack of continuity and connectivity of data manifold
- **Existence of Intra-Event sensor-by-sensor Correlation** → PXD hits are correlated within an event

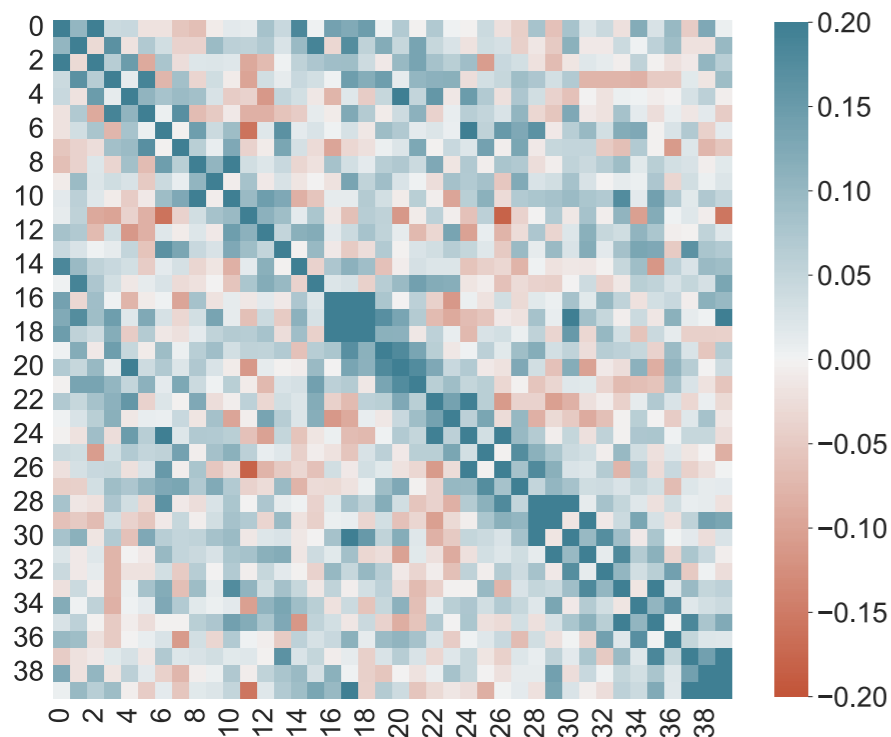


Challenges

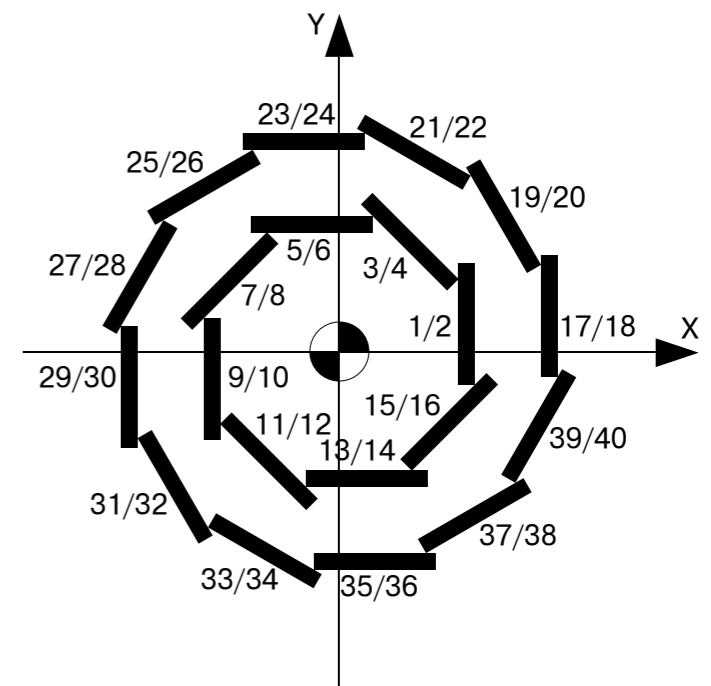
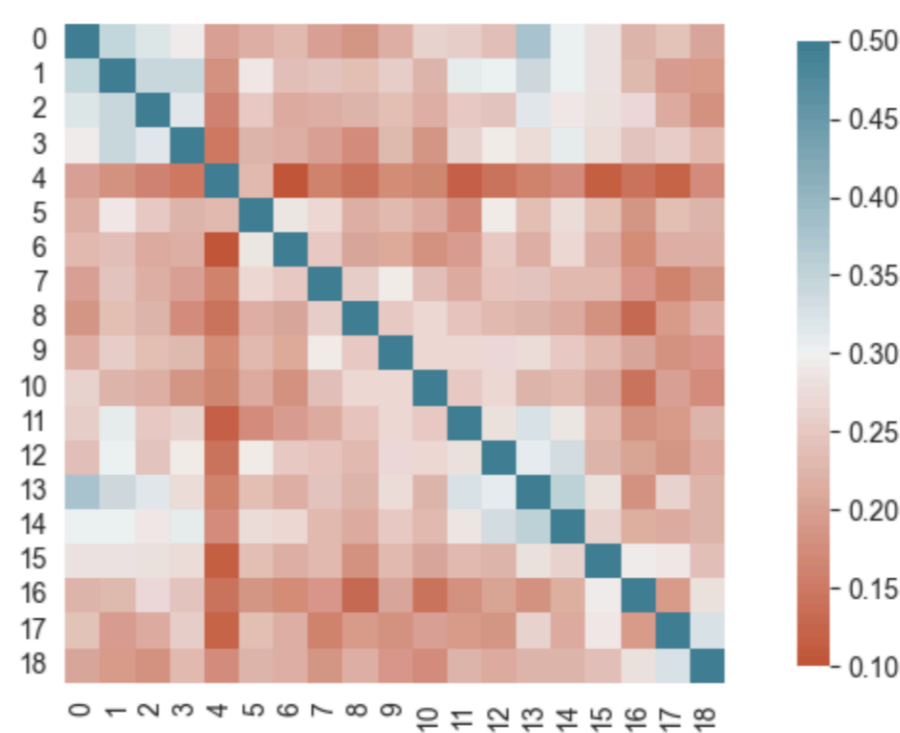
► PXD background generation challenges:

- **Ultra-High Resolution Data 40x250x768** → More than 7.5 M information channels
- **Non-Trivial (Annulus) Detector Topology** → [[1-40]] Sensor dependent information are not sequential
- **Extremely Sparse and Fine-Grained events** → Lack of continuity and connectedness of data manifold
- **Existence of Intra-Event Correlation** → PXD hits are correlated within an event

Geant4 Simulated Data



Real Data



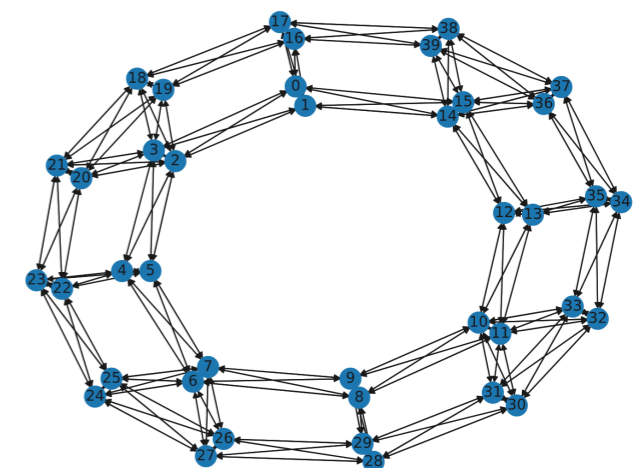
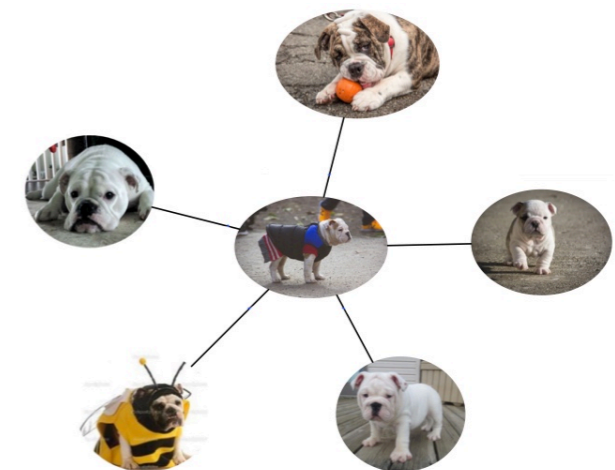
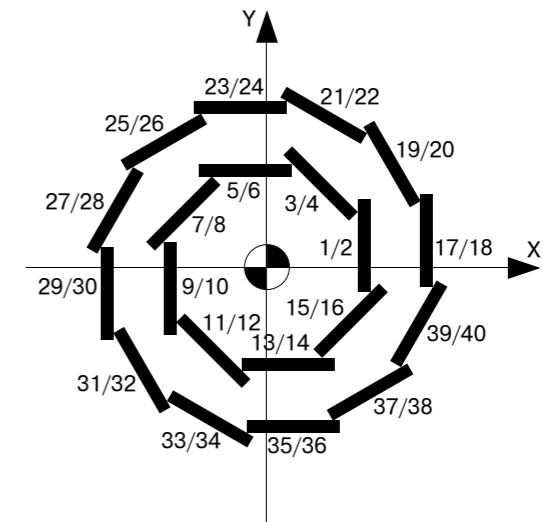
Task: Fine-Grained Generation

■ Conditional Deep Generative Model (cDGM):

✓ $[[1-40]]$ PXD sensors

■ Fine-grained conditional image generation:

- A. Different classes show both **statistical** and **semantic** similarity
- B. **Similar Natural datasets**: The Stanford Cars, iNaturalist
- C. The objective is to create objects from subordinate categories such as breeds of dogs or models of cars.
- D. The **small inter-class** and **large intra-class** variation inherent to fine-grained image analysis makes it a challenging problem.



Intra-Event Aware Reasoning: Theory



■ Traditional DGM:

- Treating the sensor/layer information the same as the hit/kinematics level information → almost works!
- It is like doing video generation while treating the temporal and spatial domain the same → Stationarity assumption!
- Convolutions introduce the bias of translation invariance → can be limiting when dealing with hits that contain patterns that change scale, rotate, or do other affine transformations through different sensors/ layers.

■ Paradigm shift in sampling: **Differentiate between Event features and sensor/hit features**

■ How to formulate it in a unified perspective?

☑ Theoretical Perspective:

A. Having an **Event** as a category, the objects of this category are the sensors/layers (finite sets of detector hits).

B. The map between the objects (morphisms) are the relations between the sensors/layers.

☑ How to use this relational inductive bias to approximate an Event?

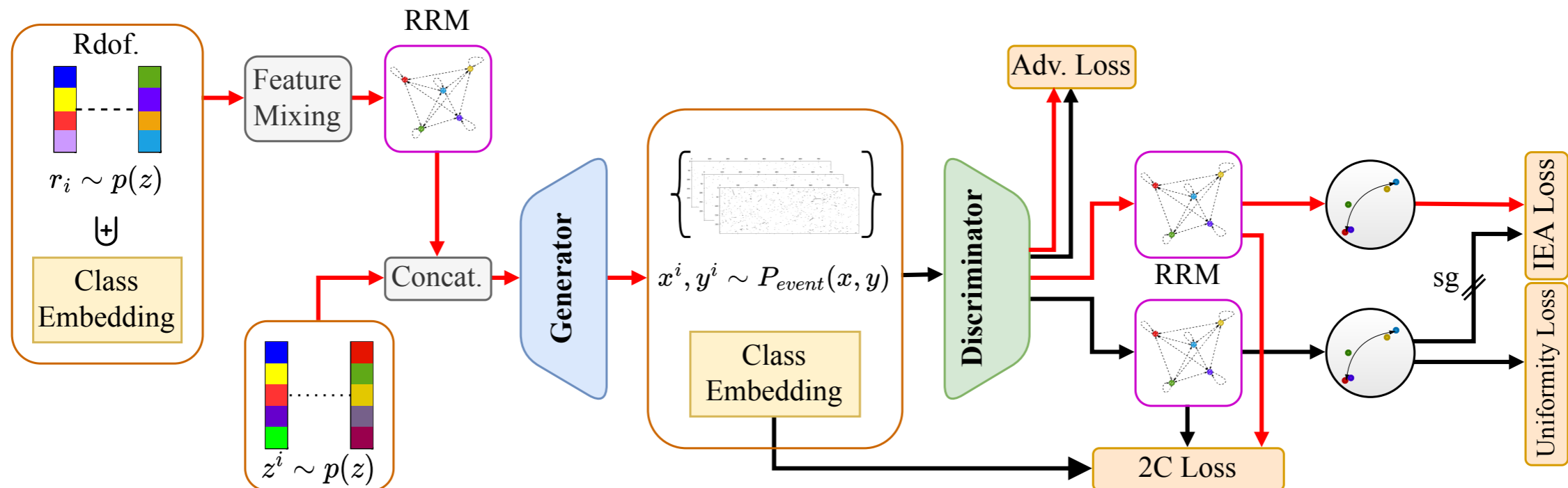
Bidirectional and one-shot, like Bert family: **IEA-GAN** ([arXiv:2303.08046](https://arxiv.org/abs/2303.08046))

IEA-GAN Model

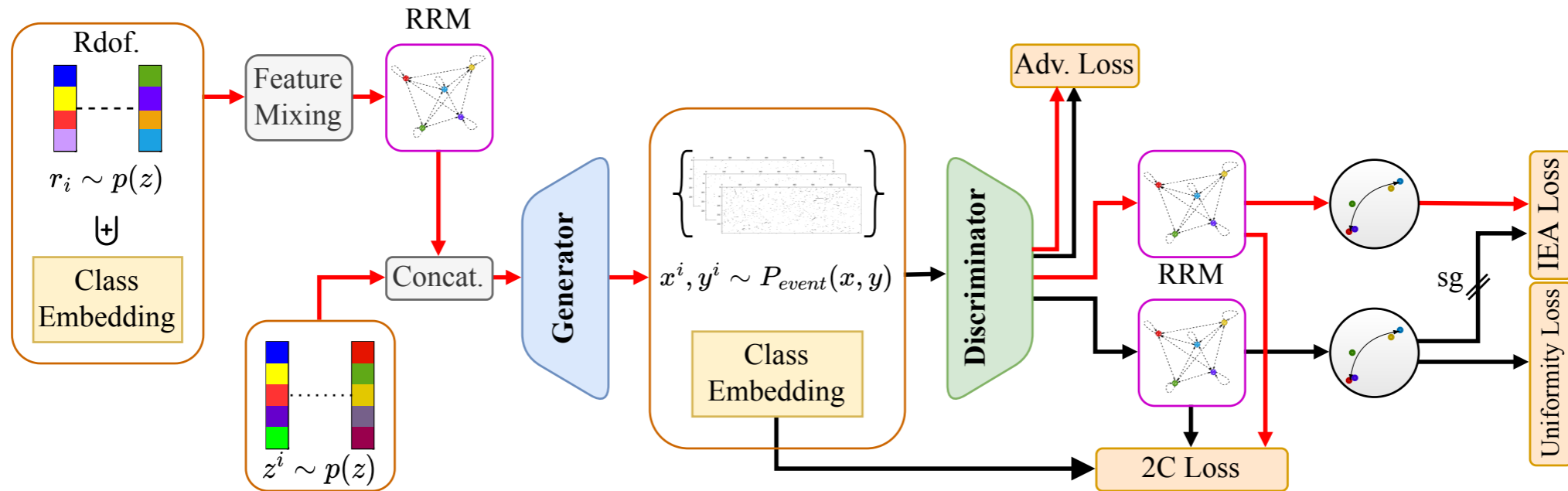


- Paradigm shift in sampling: **Intra-Event Aware Reasoning**
- How to approximate the concept of an “Event” while we simulate the detector response?
 - ✓ **Proper Sampling:** Defining an image per class sampler (**generating event by event**) and shuffling within each batch (event).
 - ✓ **Intra-event relational reasoning (Contextual):** Using a Relational Reasoning Module over an event to weight the importance of each sample with respect to each other. Thus, the model will understand the class-to-class relations in a single event.
 - ✓ **Maintaining the Discriminator's Generalisability (Information Entropy)**
 - ✓ **Transferring Discriminator's Intra-event contextual knowledge to the Generator**

Self-Supervised Learning + Knowledge Distillation + Relational Inductive Bias = IEA-GAN
(Deep Metric Learning)



IEA-GAN Model (Discriminator)



$$L_{dis} = L_{Adv} + \lambda_{2C} L_{2C} + \lambda_{uniform} L_{uniform}$$

$$2C \text{ Contrastive Loss: } L_{2C}(x_i, y_i; t) = -\frac{1}{N} \sum_{i=1}^N \log \left(\frac{\exp(h(x_i)^T e(y_i)/t)}{\exp(h(x_i)^T e(y_i)/t) + \sum_{k=1}^m \mathbf{1}_{k \neq i} \cdot \exp(h(x_i)^T h(x_k)/t)} \right)$$

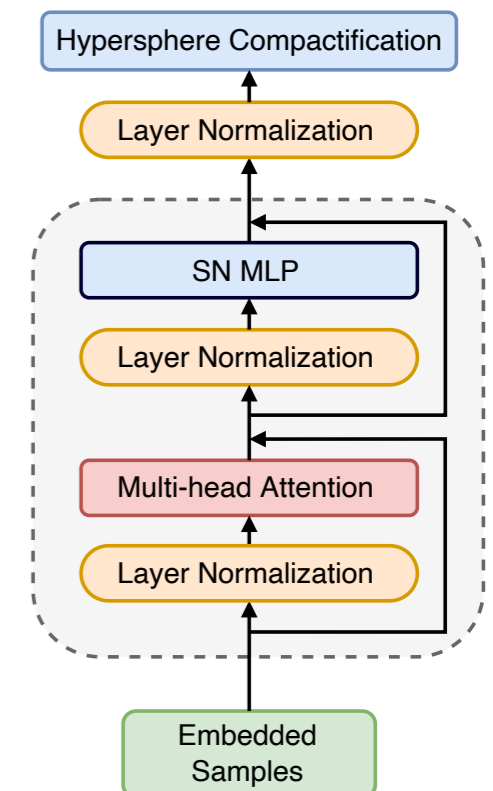
$$\text{Uniformity Loss: } L_{uniform}(h; t) = \log \mathbb{E}_{x_i, x_j \sim p_{data}} [\exp(-t \|h(x_i) - h(x_j)\|_2^2)]$$

► By imposing uniformity condition over the feature vectors on the unit hypersphere, they preserve as much information as possible since the uniform distribution carry high entropy.

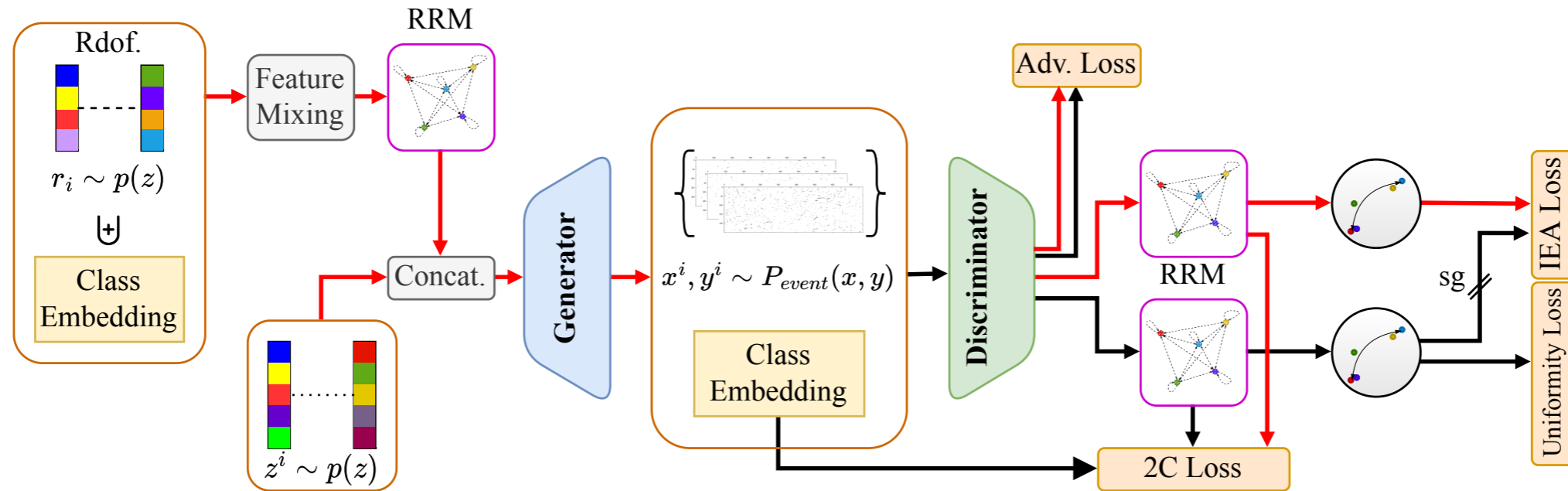
$h(\cdot)$: Relational embedding
 $e(\cdot)$: proxy (class embedding)

Hypersphere dimension: 1024
 SN-MLP dimension: 512
 Number of Heads: 4
 Number of Layers: 1

Relational Reasoning Module (RRM)



IEA-GAN Model (Generator)



$$L_{gen} = L_{Adv} + \lambda_{2C} L_{2C} + \lambda_{IEA} L_{IEA}$$

$$\text{2C Contrastive Loss: } L_{2C}(x_i, y_i; t) = -\frac{1}{N} \sum_{i=1}^N \log \left(\frac{\exp(h(x_i)^T e(y_i)/t)}{\exp(h(x_i)^T e(y_i)/t) + \sum_{k=1}^m \mathbf{1}_{k \neq i} \cdot \exp(h(x_i)^T h(x_k)/t)} \right)$$

$$\text{IEA Loss: } L_{IEA}(x_f, x_r) = D_{KL} \left(\sum_{i,j} \sigma(h(x_i^{(r)})h(x_j^{(r)})^T) \mid \sigma(h(x_i^{(f)})h(x_j^{(f)})^T) \right)$$

$h(.)$: Relational embedding

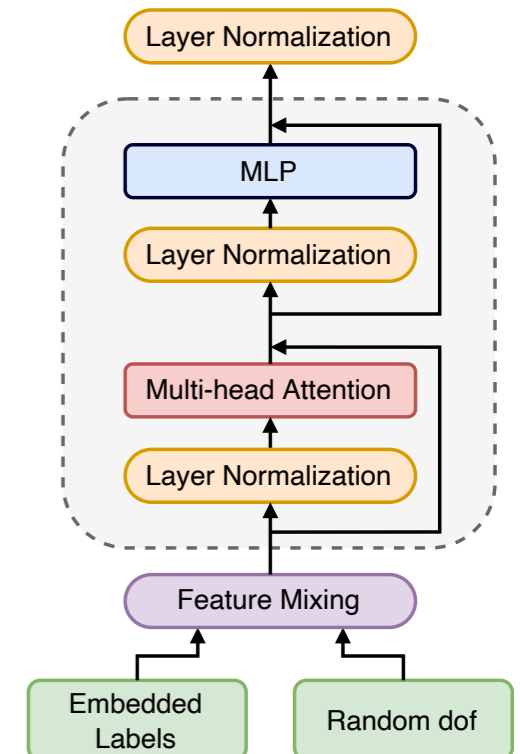
$e(.)$: proxy (class embedding)

$\sigma(.)$: Softmax function

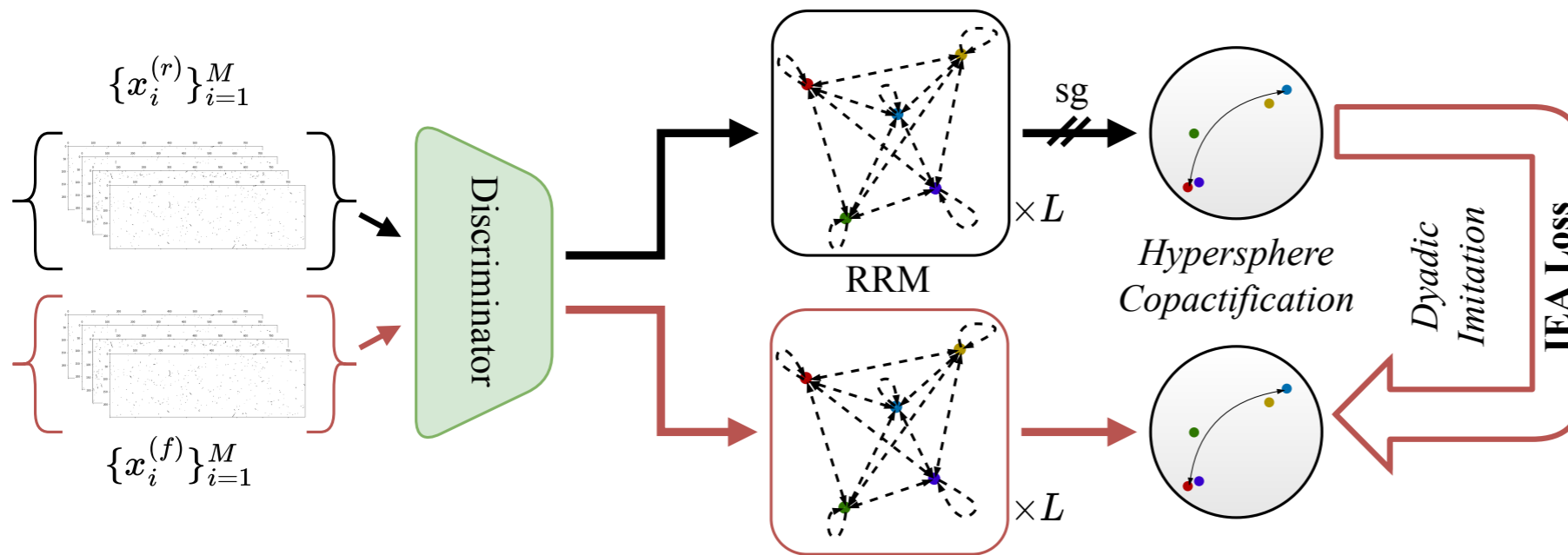
$x^{(f)}$: generated images

$x^{(r)}$: real images

Relational Reasoning Module (RRM)

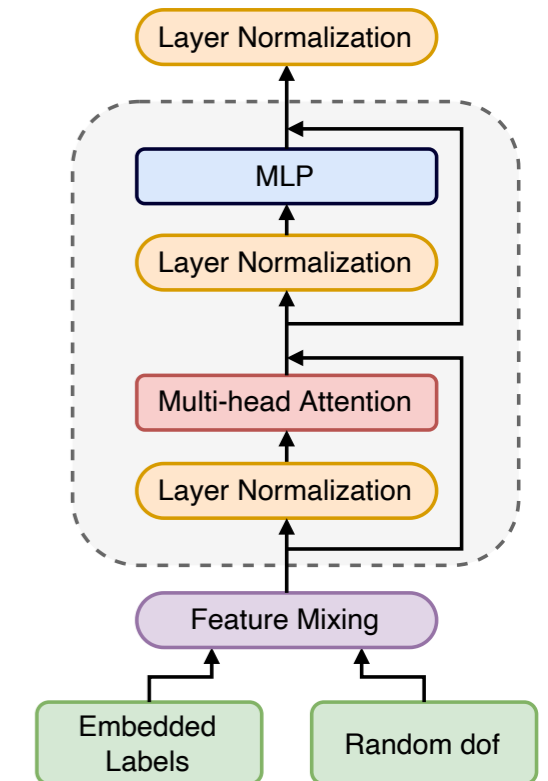


IEA-GAN Model (Generator)



IEA Loss:
$$L_{IEA}(x_f, x_r) = D_{KL}(\sum_{i,j} \sigma(h(x_i^{(r)})h(x_j^{(r)})^\top) | \sigma(h(x_i^{(f)})h(x_j^{(f)})^\top))$$

Relational Reasoning Module



Hypersphere dimension: 128

MLP dimension: 128

Number of Heads: 2

Number of Layers: 1

$h(.)$: Relational embedding

$e(.)$: proxy (class embedding)

$\sigma(.)$: Softmax function

$x^{(f)}$: generated images

$x^{(r)}$: real images

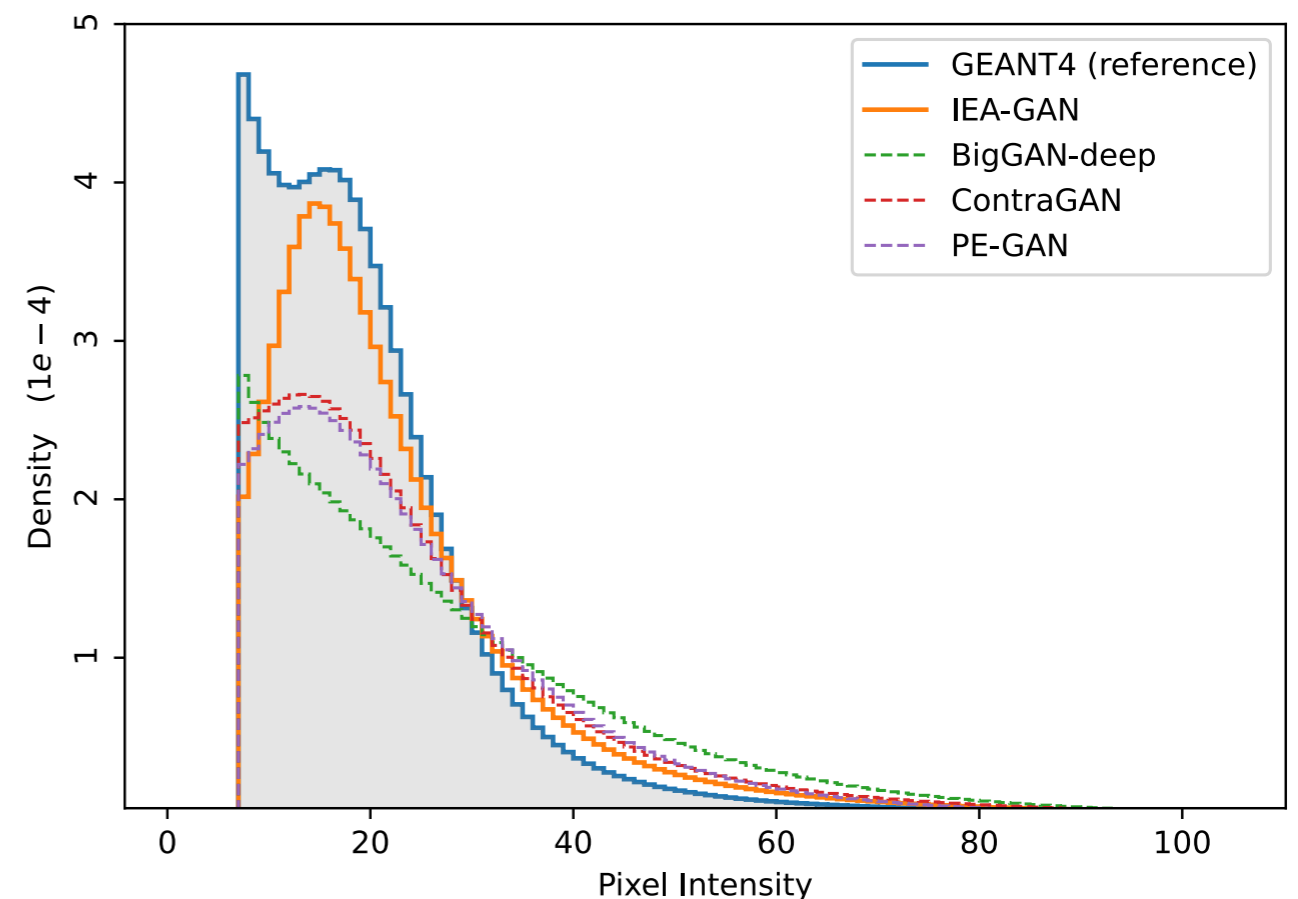
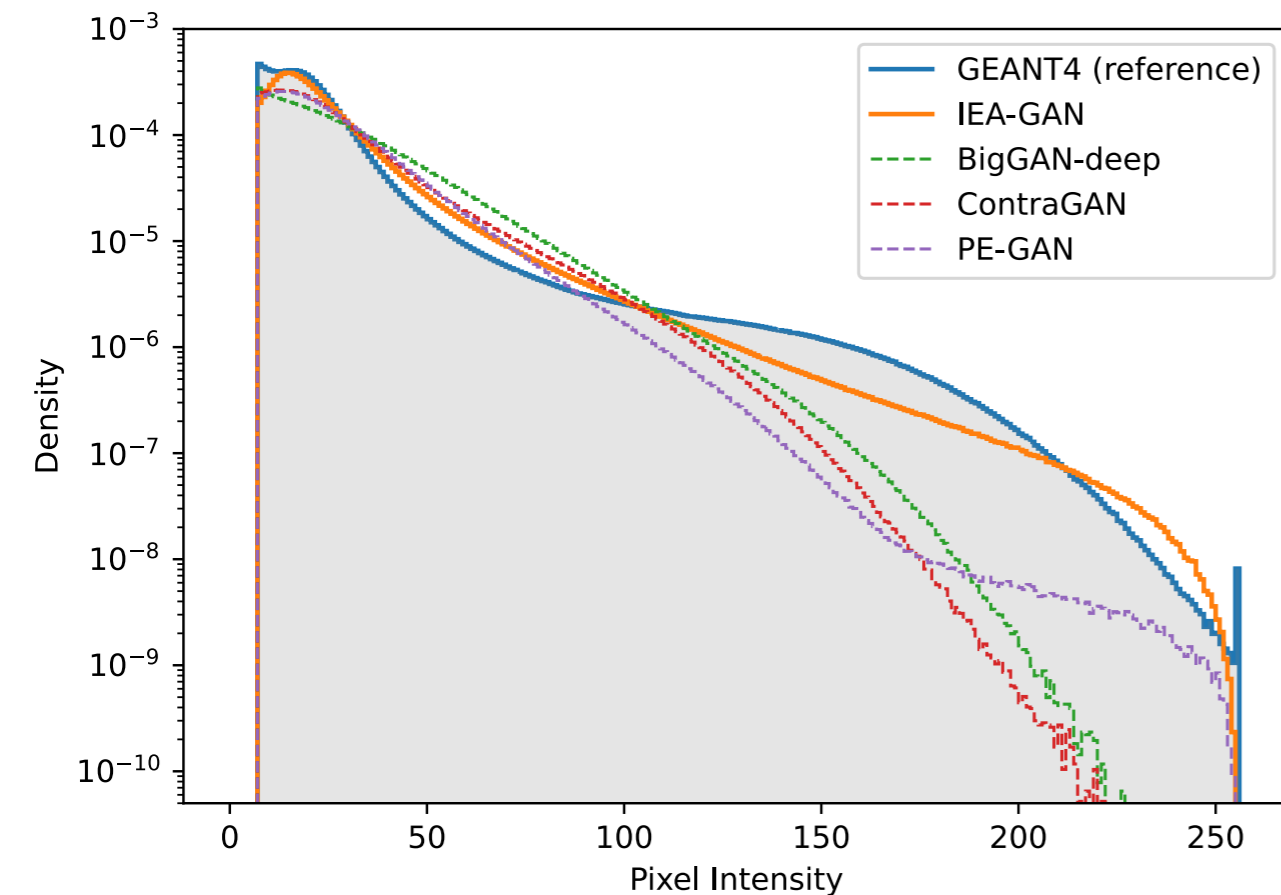
► Upon minimising it, we are putting a self-supervised penalising system over the intra-event awareness of the the generator by encouraging it to look for more detailed connections among the images.

► In the end we want to maximise the agreement of data points on two unit hyperspheres of real image and generated image embeddings.

Validation of generated PXD images

Validation Metrics over the test set in comparison to SOTA in High Resolution Image Generation:

Pixel Energy above the threshold:



BigGAN-deep, Brock, A., Donahue, J., Simonyan, K.: Large Scale GAN Training for High Fidelity Natural Image Synthesis, arXiv (2019). <https://doi.org/10.48550/arXiv.1809.11096>.

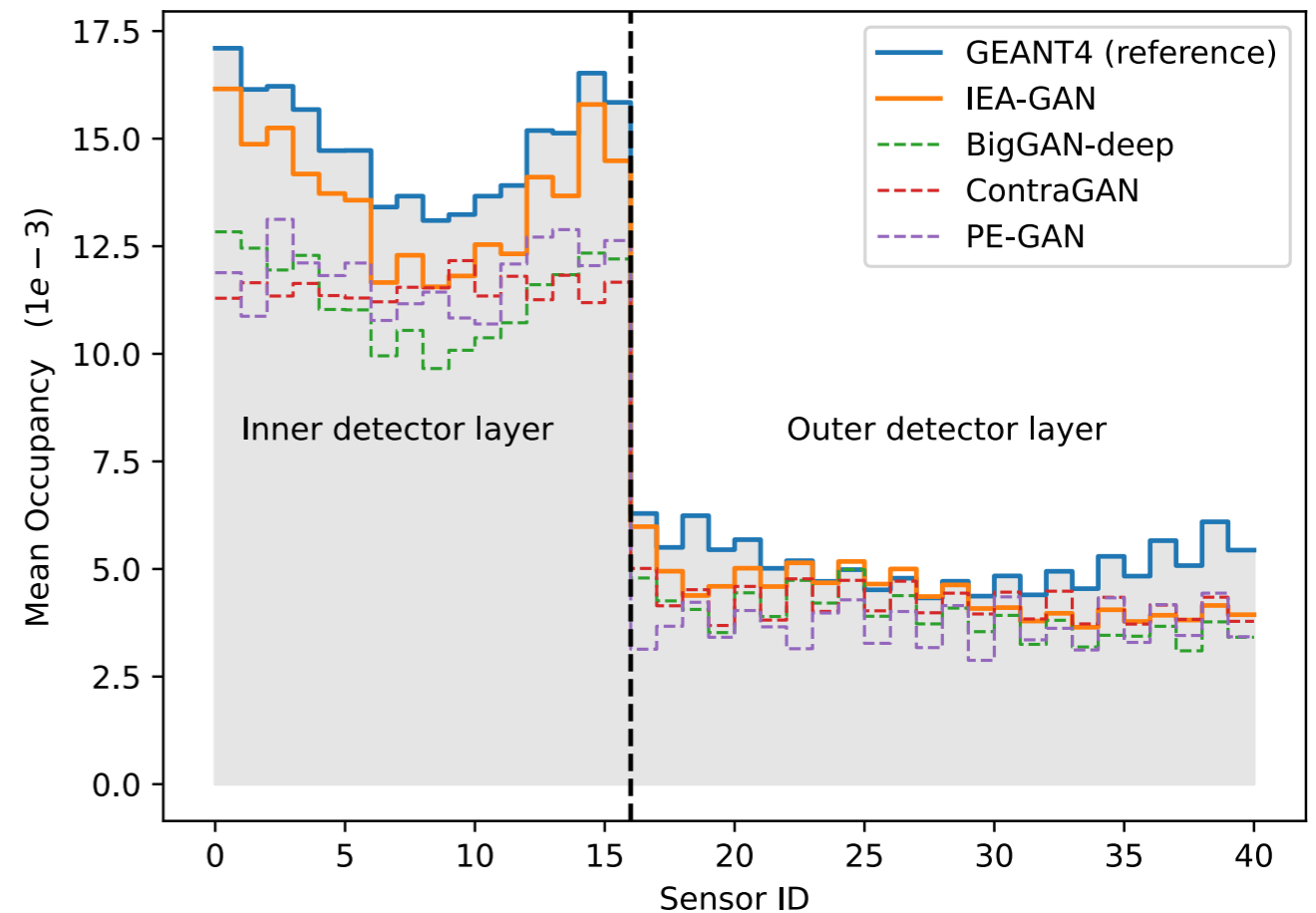
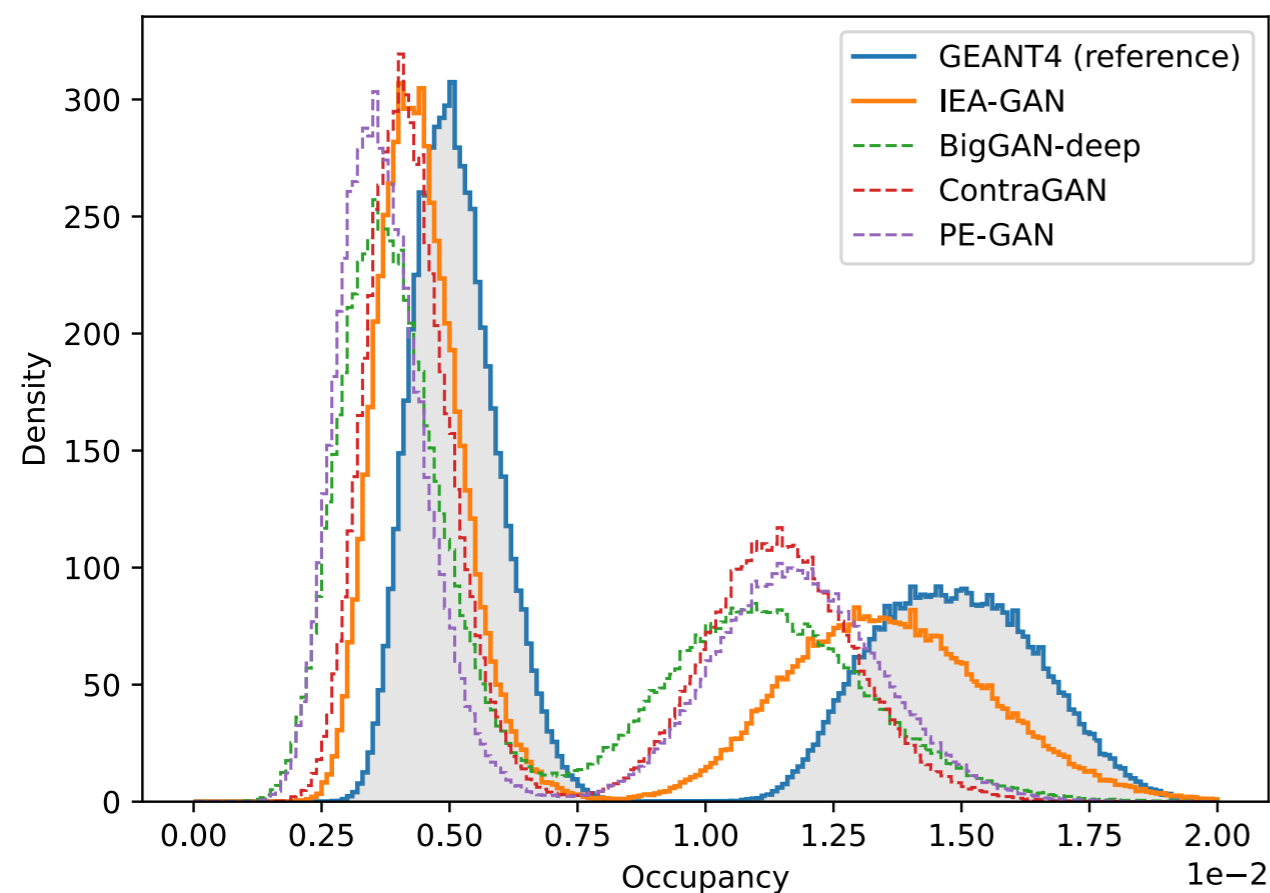
ContraGAN: Contrastive Learning for Conditional Image Generation. In: Advances in Neural Information Processing Systems, vol. 33, pp. 21357-21369.

PE-GAN: Hashemi et al.: Pixel Detector Background Generation using Generative Adversarial Networks at Belle II. vCHEP(2021). https://doi.org/10.1051/ep_jconf/202125103031

Validation of generated PXD images

Validation Metrics over the test set in comparison to SOTA in High Resolution Image Generation:

Occupancy Density and Mean Occupancy :



BigGAN-deep, Brock, A., Donahue, J., Simonyan, K.: Large Scale GAN Training for High Fidelity Natural Image Synthesis, arXiv (2019). <https://doi.org/10.48550/arXiv.1809.11096>.

ContraGAN: Contrastive Learning for Conditional Image Generation. In: Advances in Neural Information Processing Systems, vol. 33, pp. 21357-21369.

PE-GAN: Hashemi et al.: Pixel Detector Background Generation using Generative Adversarial Networks at Belle II. vCHEP(2021). https://doi.org/10.1051/ep_jconf/202125103031

Validation of generated PXD images



❖ Validation Metrics over the test set:

☑ FID and KID:

- The use of activations of the last layer from the **Inception-V3** model trained on the PXD images to summarise each image, gives the score. **The lower the FID/KID the better the image diversity and Fidelity.**

	WGAN-gp	BigGAN-deep	ContraGAN	PE-GAN	IEA-GAN
FID	12.09	4.40 ± 0.88	3.14 ± 0.74	2.61 ± 0.91	1.50 ± 0.16
KID	0.0096	0.0031 ± 0.0001	0.0015 ± 0.0002	0.0021 ± 0.0004	0.0010 ± 0.0002

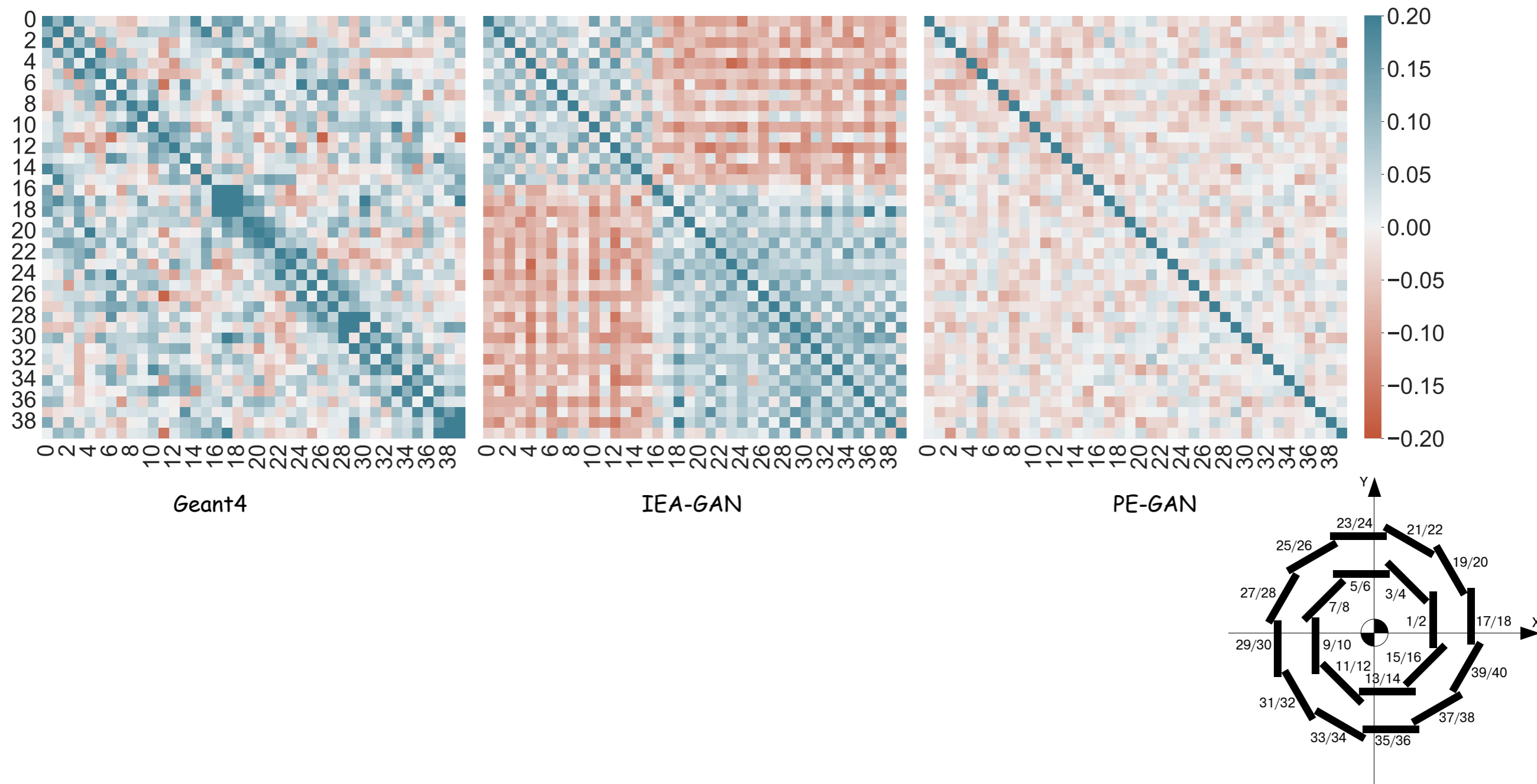
- Possible interpretation of FID at the pixel level:

Image Jitterings	FID
None	0
Random Masking (dead zones)	14.58
Random Noise	87.23
Random Rotation (30 degrees)	23.69
Random Rotation (10 degrees)	2.81
Random Translation (0.1, 0.1)	1.99
Random Shear (10, 10)	23.53
Random Zoom	9.06
High Intensity smearing	3.16
Low Intensity smearing	47.24

Validation of generated PXD images

Validation Metrics over the test set:

- ✓ Spearman's correlation between the occupancy of Geant4 simulated images (left), and generated images from IEA-GAN (center), generated images from PE-GAN (right).

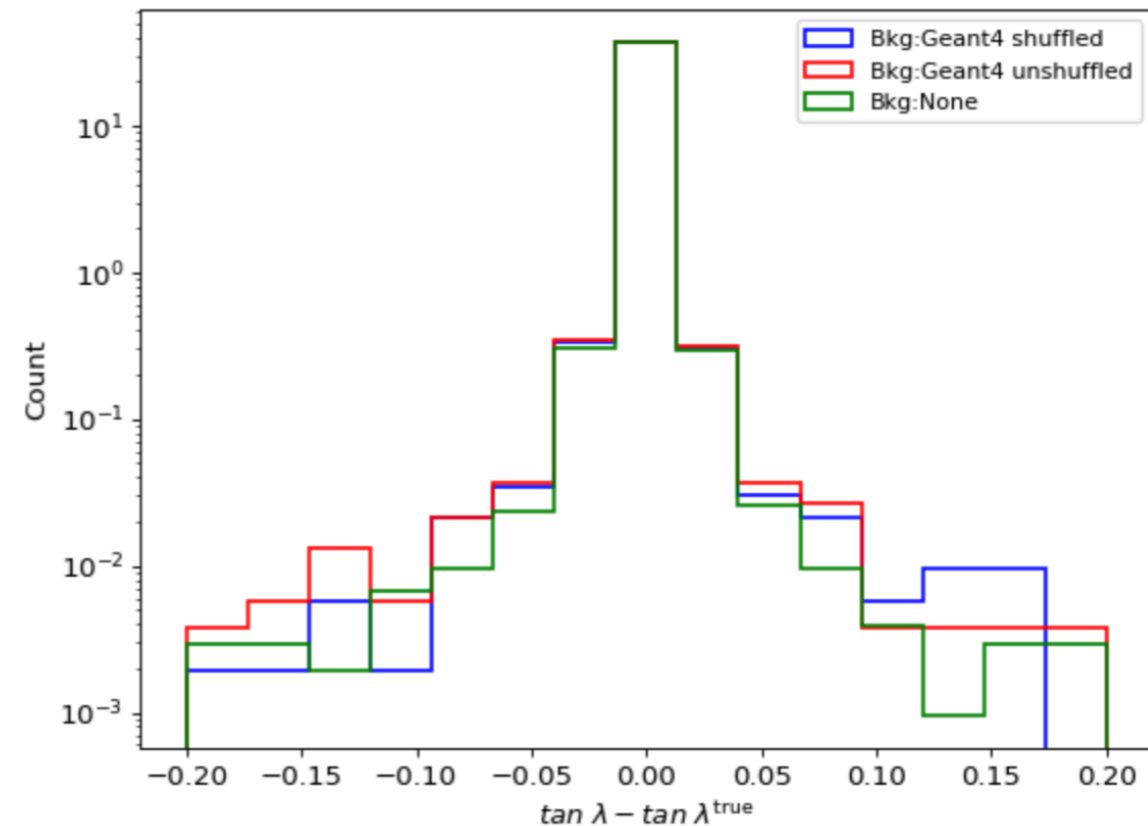
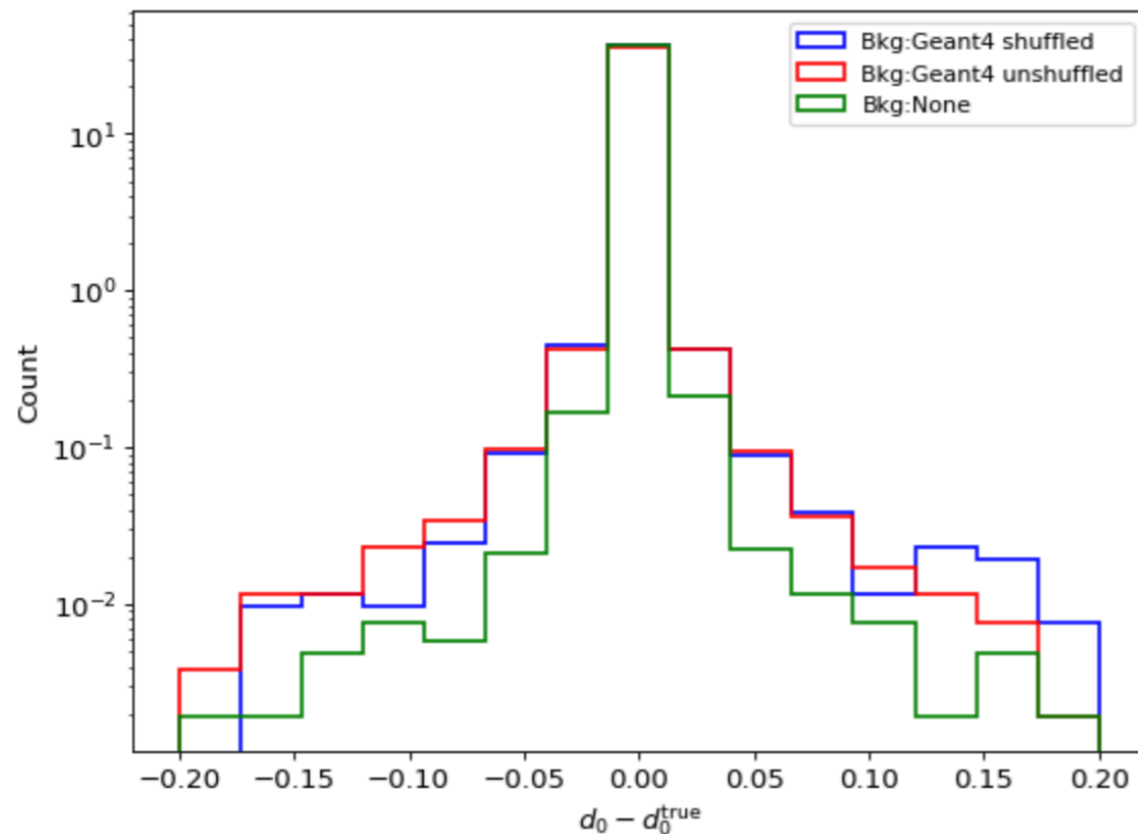


Validation of generated PXD images

❖ How important are these correlations?

- For high momentum regime $p_T > 0.4 \text{ GeV}$
- **Shuffling the events \rightarrow losing the correlation**

Parameter	Unbiased Resolution \pm error		KS statistic	p-value
	Shuffled Geant4	Unshuffled Geant4		
d_0	0.1343 ± 0.0007	0.0732 ± 0.0004	0.0067	0.7655
ϕ_0	0.2158 ± 0.0011	0.1859 ± 0.0009	0.0066	0.7899
z_0	5.0076 ± 0.0253	4.9341 ± 0.0249	0.0152	0.0211
ω	0.0010 ± 0.0000	0.0008 ± 0.0000	0.0138	0.0485
$\tan \lambda$	0.0388 ± 0.0002	0.0382 ± 0.0002	0.0167	0.0086



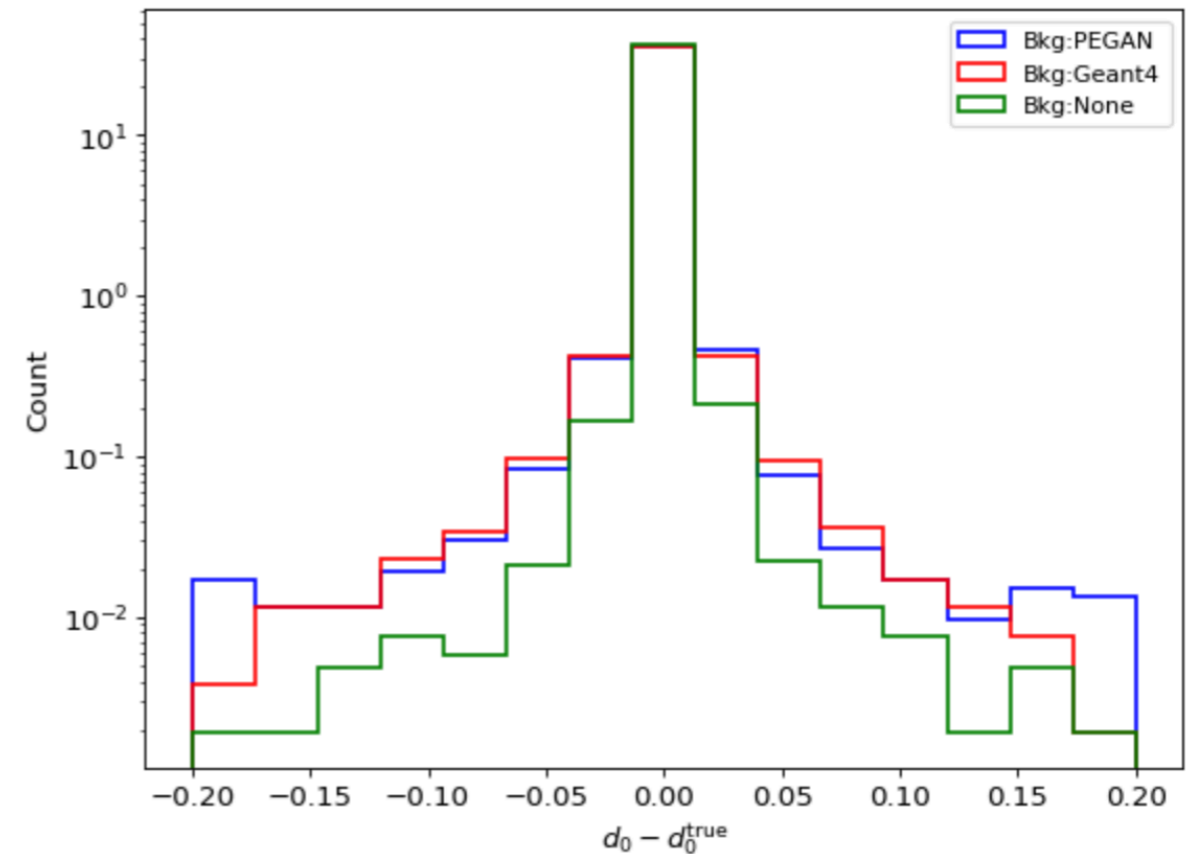
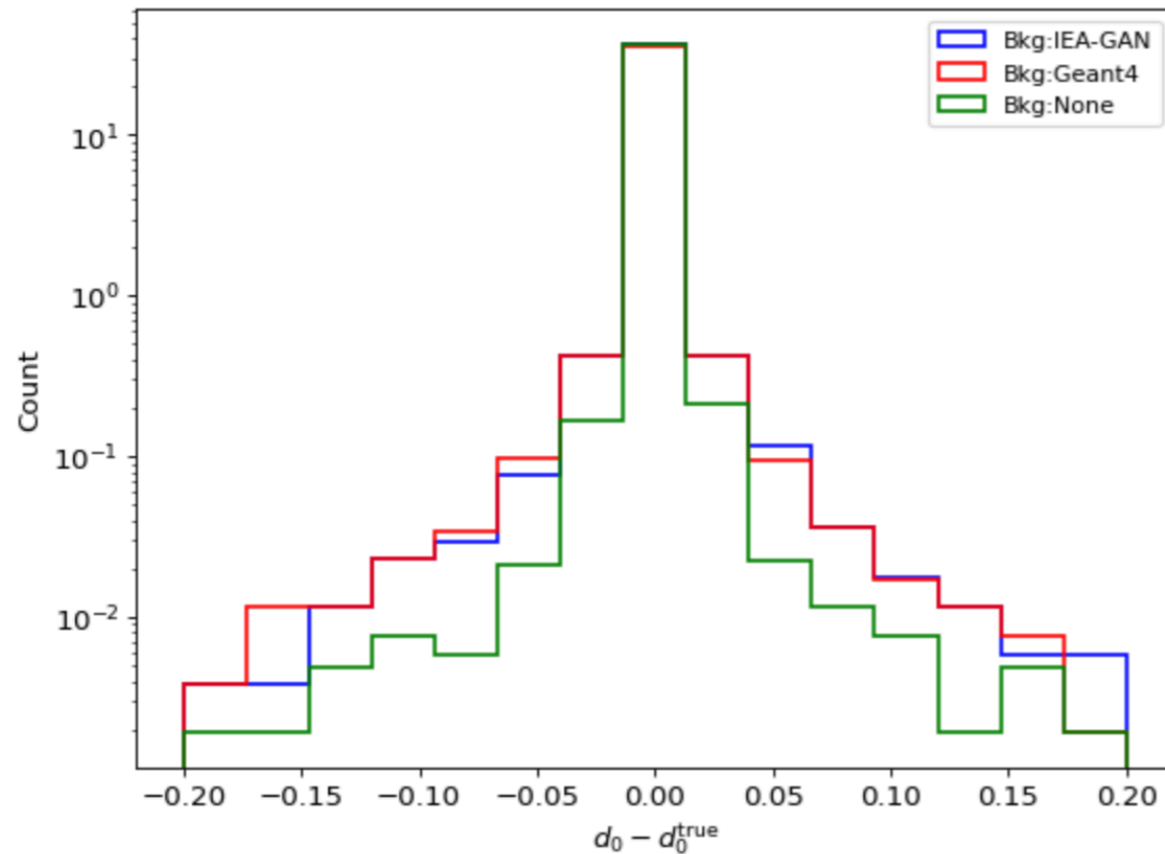
Validation of generated PXD images



❖ IEA-GAN vs PE-GAN:

- For high momentum regime $p_T > 0.4 \text{ GeV}$

Model	Parameter	Unbiased Resolution \pm error		KS statistic	p-value
		Model	Geant4		
PE-GAN	d_0	0.1375 ± 0.0007	0.0732 ± 0.0004	0.0156	0.0164
	ϕ_0	0.2207 ± 0.0011	0.1859 ± 0.0009	0.0120	0.1193
	z_0	6.9073 ± 0.0349	4.9341 ± 0.0249	0.0183	0.0029
	ω	0.0014 ± 0.0000	0.0008 ± 0.0000	0.0116	0.1425
	$\tan \lambda$	0.0579 ± 0.0003	0.0382 ± 0.0002	0.0179	0.0037
IEA-GAN	d_0	0.0762 ± 0.0004	0.0732 ± 0.0004	0.0104	0.2373
	ϕ_0	0.1905 ± 0.0010	0.1859 ± 0.0009	0.0109	0.1939
	z_0	5.1467 ± 0.0261	4.9341 ± 0.0249	0.0073	0.6814
	ω	0.0010 ± 0.0000	0.0008 ± 0.0000	0.0103	0.2537
	$\tan \lambda$	0.0412 ± 0.0002	0.0382 ± 0.0002	0.0068	0.7538



Summary and Outlook



☑ IEA-GAN:

- ▶ Successful simulation of **fine-grained**, **ultra-high resolution** (+7.5M), **correlated** PXD images based on the sensor positions.
- ▶ **Take-away messages:**
 - ☑ In general, if you wanna generate/approximate event based data, IEA mechanism is something to consider.
 - ☑ IEA-GAN being the first application of Self-Supervised Learning (SSL) in the detector simulation, SSL methods would create better opportunities to model particle physics fine-grained data.
- ▶ Accepted at ML4PS workshop at NeurIPS 2022: <https://ml4physicalsciences.github.io/2022/>
- ▶ Open-source code: <https://github.com/Hosein47/IEA-GAN>,
- ▶ Full Paper: ([arXiv:2303.08046](https://arxiv.org/abs/2303.08046))

Thank You
Let's Brainstorm Now



References

- * Kang, Minguk, and Jaesik Park. "Contragan: Contrastive learning for conditional image generation." *Advances in Neural Information Processing Systems* 33 (2020): 21357-21369
- * Hashemi, Hosein, et al. "Pixel Detector Background Generation using Generative Adversarial Networks at Belle II." *EPJ Web of Conferences*. Vol. 251. EDP Sciences, 2021.
- * Srebre, Matej, et al. "Generation of Belle II Pixel Detector Background Data with a GAN." *EPJ Web of Conferences*. Vol. 245. EDP Sciences, 2020.
- * Heusel, Martin, et al. "Gans trained by a two time-scale update rule converge to a local nash equilibrium." *Advances in neural information processing systems* 30 (2017).
- * Brock, Andrew, Jeff Donahue, and Karen Simonyan. "Large scale GAN training for high fidelity natural image synthesis." *arXiv preprint arXiv:1809.11096* (2018).



✓ The Base Model:

■ Technologies:

- ▶ Residual blocks
- ▶ Spectral Normalisation
- ▶ Orthogonal Weight init.
- ▶ Orthogonal regularisation
- ▶ Contrastive Learning
- ▶ Hinge Loss
- ▶ Consistency Regularisation
- ▶ Differentiable Augmentation
- ▶ IEA Loss
- ▶ 5×10^{-5} lr for both G and D

Algorithm 1 Intra-Event Aware GAN

Require: generator and discriminator parameters θ_G, θ_D , Intra-Event-aware coefficient λ_{IEA} , Uniformity coefficient $\lambda_{uniform}$ and hyperparameter s , Adam hyperparameters α, β_1, β_2 , event size M , number of discriminator iteration steps per generator iteration N_D

```

1: for number of training iterations do
2:   for  $t = 1, \dots, N_D$  do
3:     sample  $\{z^i\}_{i=1}^M \sim p(z)$ ,
4:      $\{x^i, y^i\}_{i=1}^M \sim p_{event}(x, y), \{r^i\}_{i=1}^M \sim p_{Rdof}(z)$       ▷ Event Sampling.
5:     for  $i = 1, \dots, M$  do
6:        $\ell_{D_{hinge}}^{(i)} \leftarrow \ell_{D_{hinge}}(x^{(i)}; G(z^i, y^i, r^i))$ 
7:     end for
8:      $\mathcal{L}_{D_{hinge}} \leftarrow \frac{1}{M} \sum_{i=1}^M \ell_{D_{hinge}}^{(i)}$ 
9:      $\mathcal{L}_{uniform} \leftarrow \mathcal{L}_{uniform}(x; s)$       ▷ The Uniformity Loss.
10:     $\mathcal{L}_{2C}^{real} \leftarrow \frac{1}{M} \sum_{i=1}^M \ell_{2C}(x^i, y^i)$ 
11:     $\theta_D \leftarrow Adam(\mathcal{L}_{D_{hinge}} + \lambda_{2C} \mathcal{L}_{2C}^{real} + \lambda_{uniform} \mathcal{L}_{uniform}, \alpha, \beta_1, \beta_2)$ 
12:  end for
13:  sample  $\{z^i\}_{i=1}^M \sim p(z)$ ,
14:  sample  $\{r^i\}_{i=1}^M \sim p_{Rdof}(z)$       ▷ Event Sampling.
15:  for  $i = 1, \dots, M$  do
16:     $\ell_{G_{hinge}}^{(i)} \leftarrow \ell_{G_{hinge}}(G(z^i, y^i, r^i))$ 
17:  end for
18:   $\mathcal{L}_{G_{hinge}} \leftarrow \frac{1}{M} \sum_{i=1}^M \ell_{G_{hinge}}^{(i)}$ 
19:   $\mathcal{L}_{IEA} \leftarrow \frac{1}{M} \sum_{i=1}^M \ell_{IEA}(G(z^i, y^i, r^i), x^i)$       ▷ The Intra-Event Aware Loss.
20:   $\mathcal{L}_{2C}^{fake} \leftarrow \frac{1}{M} \sum_{i=1}^M \ell_{2C}(G(z^i, y^i, r^i), y^i)$ 
21:   $\theta_G \leftarrow Adam(\mathcal{L}_{G_{hinge}} + \lambda_{2C} \mathcal{L}_{2C}^{fake} + \lambda_{IEA} \mathcal{L}_{IEA}, \alpha, \beta_1, \beta_2)$ 
22: end for

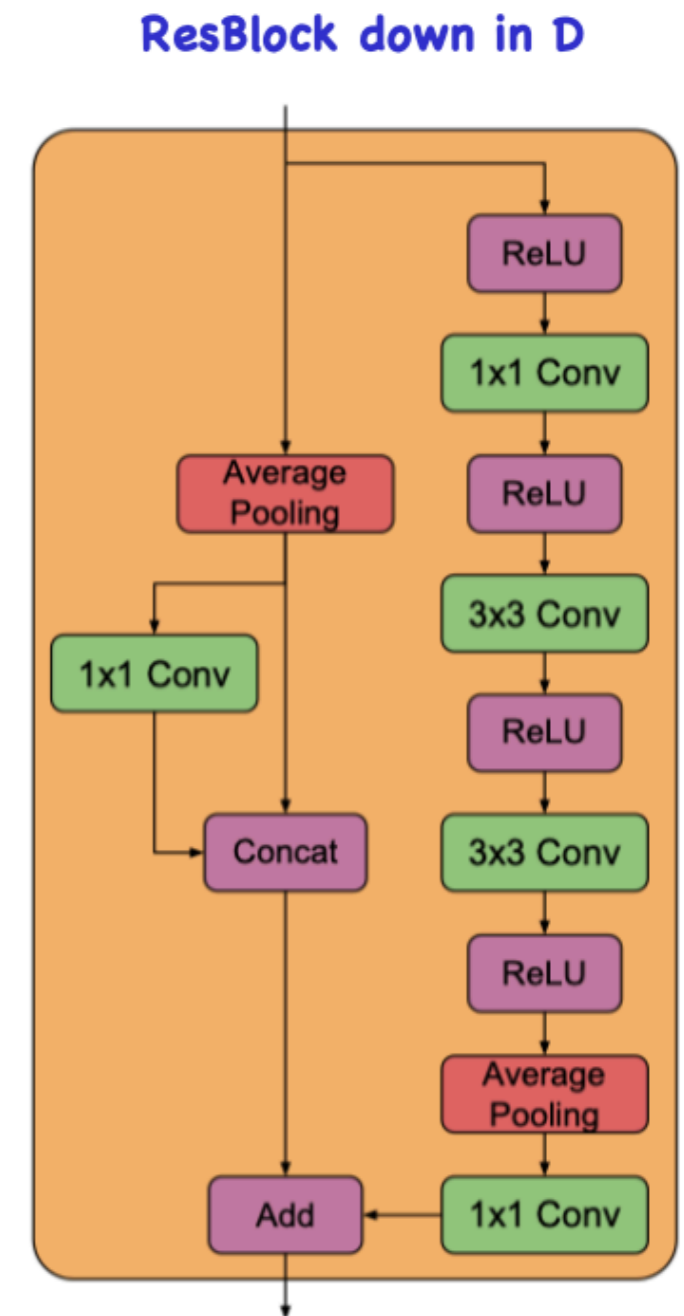
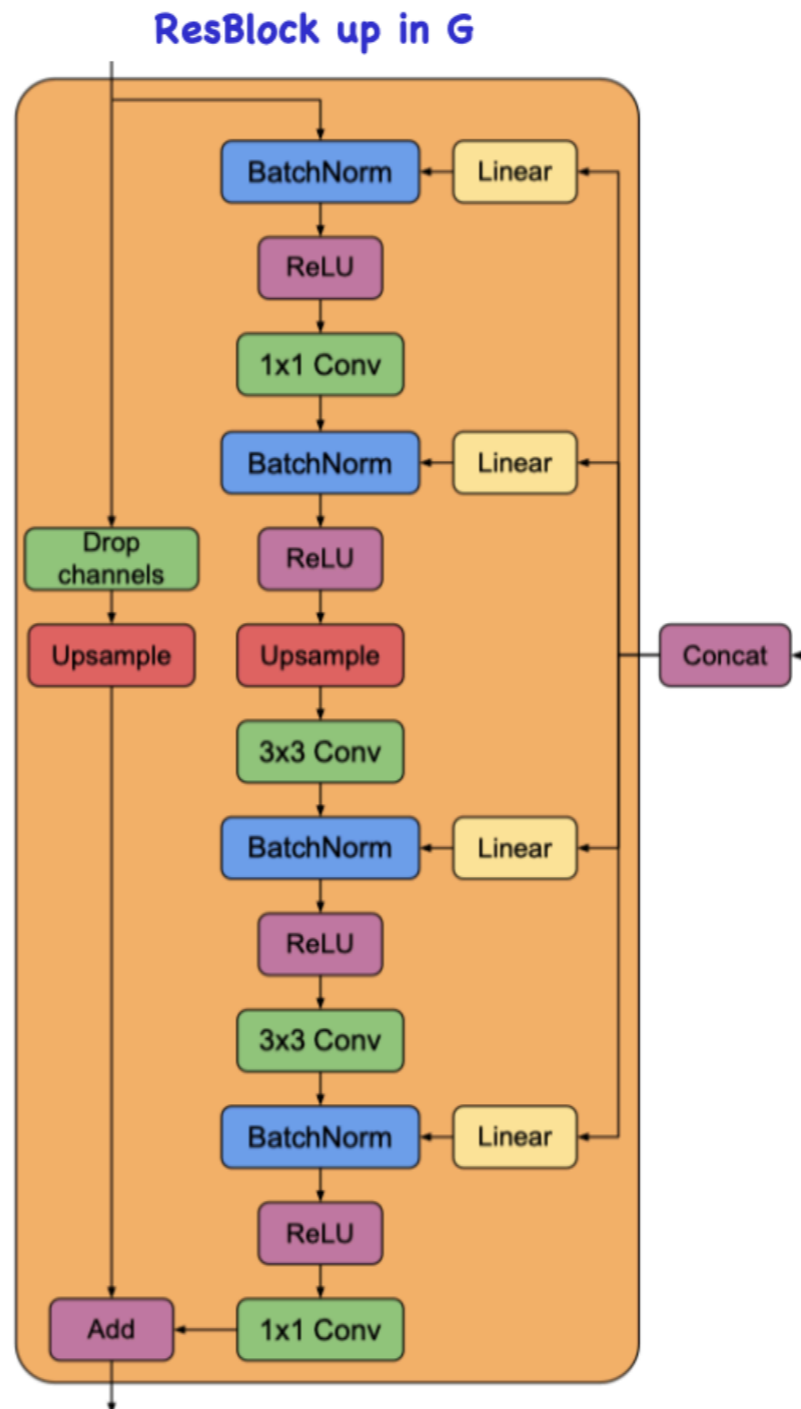
```

Back up Slides

✓ The Base Model:

■ Technologies:

- Residual blocks
- Spectral Normalisation
- Orthogonal Weight init.
- Orthogonal regularisation
- Contrastive Learning
- Hinge Loss
- Consistency Regularisation
- Differentiable Augmentation
- IEA Loss
- 5×10^{-5} lr for both G and D



Back up Slides

Table A1: FID comparison between IEA-GAN, IEA-GAN with RRM only, IEA-GAN with Uniformity loss only , and IEA-GAN with both IEA-loss, averaged across six random seeds.

	IEA-GAN	Only RRM	RRM with Uniformity	RRM with IEA-loss
FID	1.50 ± 0.16	2.74 ± 0.62	2.71 ± 0.14	3.42 ± 0.52

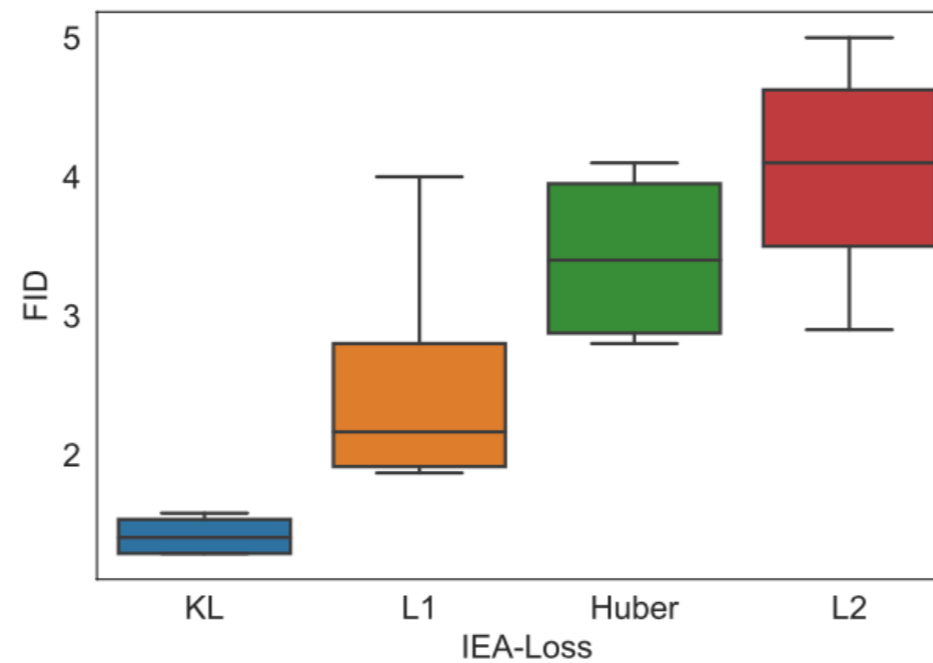


Fig. A1: Comparison of the FID between different IEA-losses

Back up Slides



Table B3: Computational performance of IEA-GAN and PE-GAN generators on a single core of an Intel Xeon Silver 4108 1.80GHz (CPU) and NVIDIA V100 with 32 GB of memory (GPU) compared to GEANT4. For the generative models, the mean and standard deviation obtained for sets of 1000 events. The time for GEANT4 refers to the theoretical time it would take to run the simulation of all background processes on-the-fly.

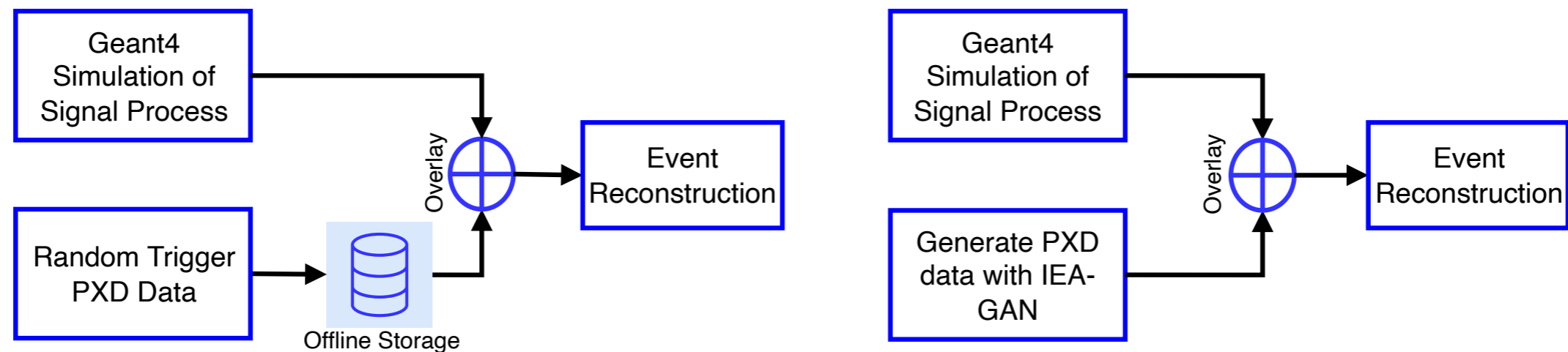
Hardware	Simulator	time/event[s]
CPU	Geant4	≈ 1500
	PE-GAN	11.781 ± 0.357
	IEA-GAN	10.159 ± 0.208
GPU	PE-GAN	0.090 ± 0.010
	IEA-GAN	0.070 ± 0.006

Overlay Problem

Realistic detector simulation has to take into account effects from background processes

- Simulation requires many PXD hitmaps with statistically independent background.
- Overlay hits from simulated background or random trigger data to hits from signal MC.
- PXD hits have the highest storage consumption.
- Requires distributing over all sites where MC is produced.

Solution: Generating the background data on the way of analysis with GANs instead of storing them.



Validation of generated PXD images



Validation Metrics over the test set:

Physics Analysis: Helix parameter resolutions

