Deep Generative Models for Calorimeter Simulations

— Vienna Workshop on Simulations (VIEWS) , Vienna, AT —

Claudius Krause

Institute of High Energy Physics (HEPHY), Austrian Academy of Sciences (OeAW)

April 27, 2024

Deep Generative Models …

… are machine-learning models that “generate” new samples of a (complicated) $p(x)$.

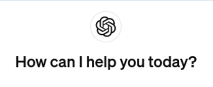
… can be understood as fancy random number generators, with the numbers being:

- pixels of an image

“Calorimeter Simulation” via midjourney.com

$\Rightarrow$ image generators like `MidJourney`, `DALL·E`

- translated to words

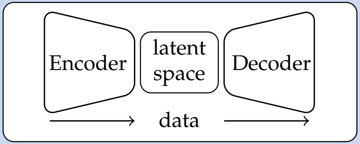


$\Rightarrow$ chatbots like `ChatGPT`, `GitHub CoPilot`

The Landscape of Generative Models.

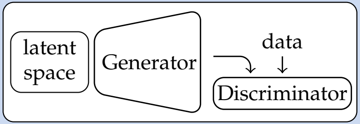
Variational Autioencoder (VAE)

⇒ Compressing data through a bottleneck.



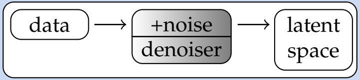
Generative Adversarial Network (GAN)

⇒ Generator and Discriminator play a game against each other.



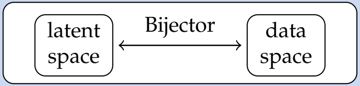
Diffusion Models

⇒ Gradually add noise and revert.



Normalizing Flows

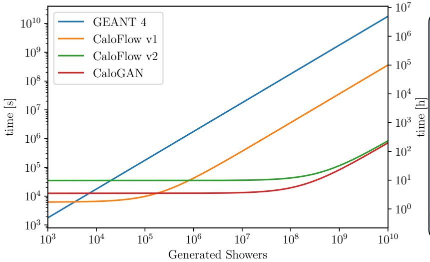
⇒ Bijective map to a known distribution.



Generative Models are Fast Surrogates!

	CaloFlow*		CaloGAN*	Geant4†
	teacher [2106.05285]	student [2110.11377]	[1712.10321]	
training	22+82 min	+ 480 min	210 min	0 min
generation time per shower	36.2 ms	**0.08 ms**	**0.07 ms**	1772 ms

*: on Rutgers Titan V GPU, †: on the CPU of CaloGAN: Paganini, de Oliveira, Nachman [1712.10321, PRD]



Statistical Amplification:

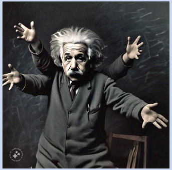
- implicit bias of the DGM lets us interpolate $\Rightarrow$ extract more info
- effect diminishes with increasing sample size
- strongly depends on use case (combinatorics!)

Matchev et al.[2002.06307], Butter et al.[2008.06545], Bieringer et al.[2202.07352]

How can we evaluate generative models?

In text / image / video generation: "by eye".
⇒ Our brains are incredible good at this task, but it doesn't scale.

imagined with Meta AI.



In (high-energy) physics: depends on the application
⇒ We want to correctly cover $p(x)$ of the entire phase space.

- in importance sampling or MCMC, a bad DGM just means an inefficient setup.
- in simulation (like end-to-end or calorimeter), we need to be more careful!

⇒ Histograms are always just a 1-dim projection!

⇒ Better: the classifier test

A Classifier provides the "ultimate metric".

According to the Neyman-Pearson Lemma we have:

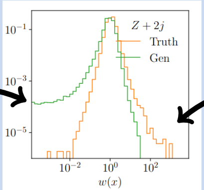
- The likelihood ratio is the most powerful test statistic to distinguish two samples.
- A powerful classifier trained to distinguish the samples should therefore learn (something monotonically related to) $w = \frac{p_{\text{data}}}{p_{\text{model}}}$.
- If this classifier is confused, we conclude $\Rightarrow \; p_{\text{data}}(x) = p_{\text{model}}(x)$

$\Rightarrow$ This captures the full phase space incl. correlations.

CK/D. Shih [2106.05285, PRD]

Failure modes of the model can now be seen in the w histogram:

Data manifold over-populated by model: $\Rightarrow$ missmodeled feature



Data manifold not populated by model: $\Rightarrow$ missed feature

R. Das, CK, et al. [2305.16774, SciPost]

Now let's apply all of this to Calorimeter Showers.

Deep Generative Models can work with different data representations:

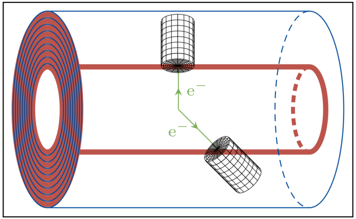
- GEANT4 Hits $\Rightarrow$ point clouds
- Detector Cells $\Rightarrow$ (irregular?) grid
- Shower-specific coordinate system $\Rightarrow$ voxels

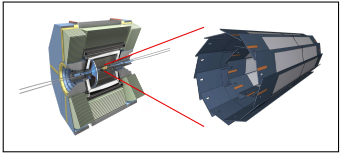
Crucial Differences to (most of) the other talks at VIEWS

- Considered energy: Calorimeter showers at colliders are usually $\mathcal{O}(\text{GeV})$
- Physics accuracy: assume GEANT4 is ground truth and try to emulate it.
- Experimental setup: at LHC usually single showers

Deep Generative Models for Shower Simulations

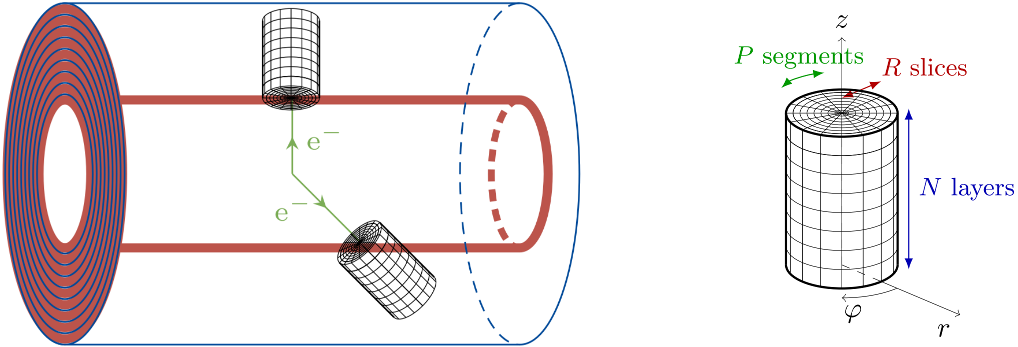
Part I: Single Particle Showers





Part II: Event Background at Belle-II

Digitized showers are like images.



Application to full detector: center coordinate system at shower.

https://g4fastsim.web.cern.ch/docs/ml_workflow/#dataset-description

Towards deployment in FastSimulation.

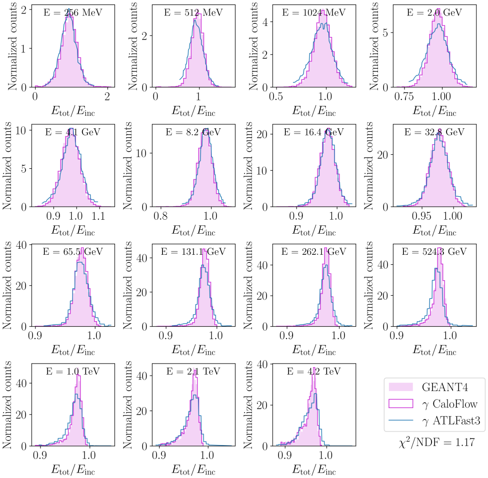
Have a rapidly evolving field: need a survey of current approaches on a common dataset!

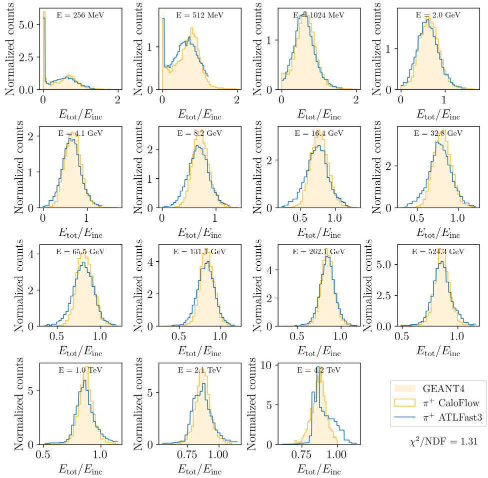
⇒ Fast Calorimeter Challenge 2022 https://calochallenge.github.io/homepage/

Michele Faucci Giannelli, Gregor Kasieczka, **CK**, Ben Nachman, Dalila Salamani, David Shih, and Anna Zaborowska

- Dataset 1: AtlFast3 trainig data (γ: 368, π: 533 voxels)
  [2109.02551, Comput.Softw.Big Sci.] $E_{\text{inc}} \in [256\text{ MeV}, 4.2\text{ TeV}]$
- Dataset 2: simulated detector (e^-: 6480 voxels) $E_{\text{inc}} \in [1\text{ GeV}, 1\text{ TeV}]$
- Dataset 3: simulated detector (e^-: 40500 voxels) $E_{\text{inc}} \in [1\text{ GeV}, 1\text{ TeV}]$

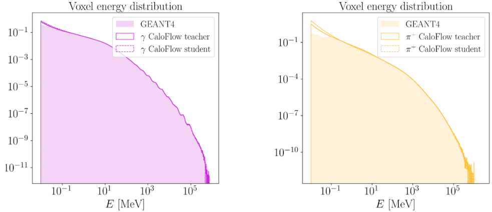
One example: CaloFlow works well on dataset 1.



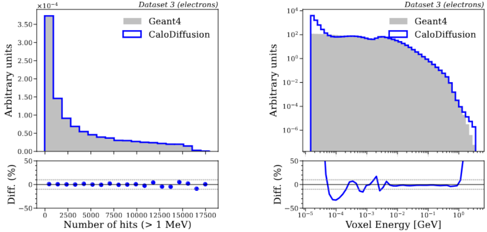


Comparing CK/Pang/Shih [2210.14245] to AtlFast3 [2109.02551, Comput.Softw.Big Sci.]

More Examples for Normalizing Flows and Diffusion Models

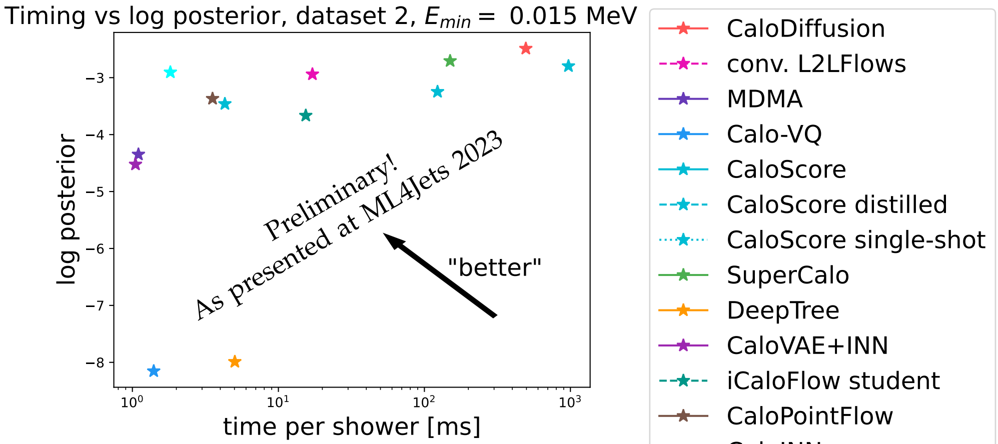


CK/Pang/Shih [2210.14245]



Amram/Pedro [2308.03876]

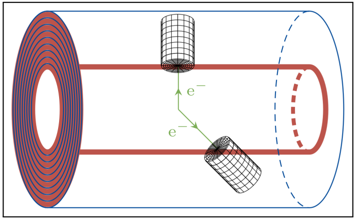
Preliminary Evaluation of ds2

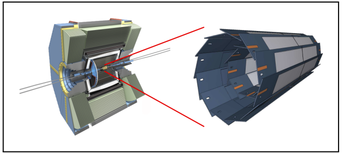


GEANT4: $\mathcal{O}(10^5)$ ms

Deep Generative Models for Shower Simulations

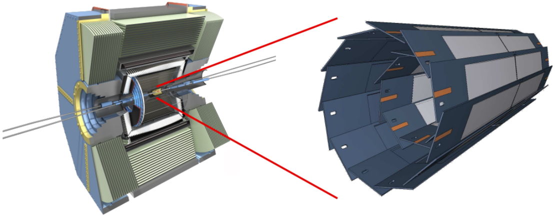
Part I: Single Particle Showers

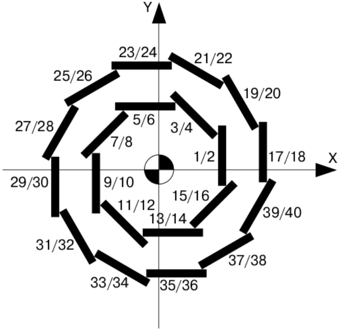




Part II: Event Background at Belle-II

Going from single showers to events: Belle-II



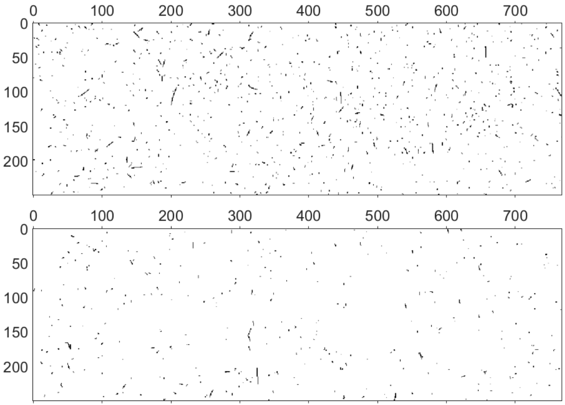


Hashemi et al. [2303.08046/acc. in Nature Comm.]

The Belle-II experiment at KEK in Japan

- Precision e^+e^- collider at the $\Upsilon(4S)$ resonance.
- The Pixel Vertex Detector (PXD) is the innermost detector component.

The PXD Background is complex and high-dimensional.



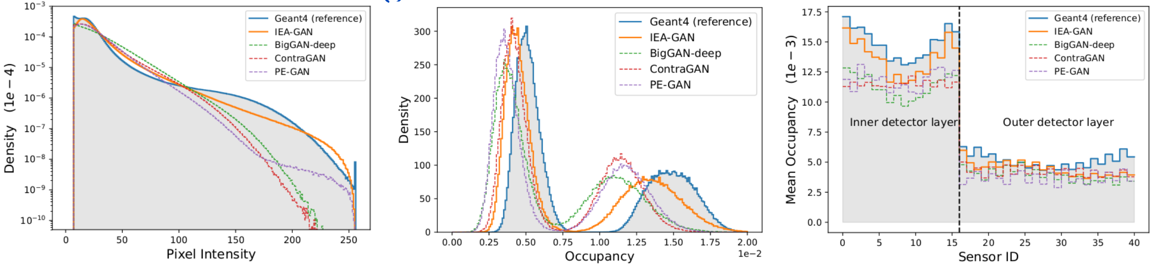
Hashemi et al. [2303.08046/acc. in Nature Comm.]

The PXD has 16+24 sensors with 250$\times$768 pixel each $\Rightarrow$ 7.7M channels.

Main backgrounds in the PXD are

- synchrotron radiation
- residual gas collisions
- Bhabha scattering
- …

This is the highest-dimensional dataset in HEP.



Hardware	Simulator	time/event [s]	Storage [Mb]	Speed-up
CPU	Geant4	≈ 1500	≈ 2000	1
	PE-GAN	11.781 ± 0.357	≈ 47	$\approx \times 127$
	IEA-GAN	10.159 ± 0.208	≈ 47	$\approx \times \mathbf{147}$
GPU	PE-GAN	0.090 ± 0.010	≈ 47	$\approx \times 16667$
	IEA-GAN	0.070 ± 0.006	≈ 47	$\approx \times 21429$

Hashemi et al. [2303.08046/acc. in Nature Comm.]

Deep Generative Models for Shower Simulations

- Deep Generative Models are promising candidates for fast surrogates.
- The field is moving fast, driven by progress in computer vision / image generation.

Recent Review: [2312.09597]

- DGMs are $\mathcal{O}(10^4)\times$ faster than GEANT4.
- For $\mathcal{O}(10^2)$ channels, samples are indistinguishable.
- For $\mathcal{O}(10^3)$ channels, samples are almost indistinguishable.
- For $\mathcal{O}(10^4 - 10^6)$ channels, we are reaching the current limits.

via midjourney: "Albert Einstein smiling while having fun coding"