







Differentiable Modeling for Calorimeter Simulation using Diffusion Models

Xuan Tung Nguyen^{*}, Long Chen, Tommaso Dorigo, Nicolas R. Gauger, Federico Nardi, Julien Donini

Fifth MODE Workshop on Differentiable Programming for Experiment Design Kolymbari, Crete, Greece June 12th, 2025

Objectives

Surrogates for Calorimeter Optimization?

- Goal: Gradient-based optimization of calorimeter designs.
- **Challenge**: GEANT4 is not natively differentiable \rightarrow ongoing efforts
- **Proposal**: Develop a high-fidelity, differentiable **surrogate model**.
- Candidate: State-of-the-art generative model

Diffusion Models (DMs)

Strong conditioning capabilities

Fully differentiable via backpropagation



T.Dorigo et al., "Toward the End-to-End Optimization of Particle Physics Instruments with Differentiable Programming", arXiv 2203.13818.

Diffusion model for the surrogate



Diffusion model for the surrogate



Challenge with Surrogate for Design Optimization

Challenge: Surrogate model training in design optimization ideally requires exponentially many samples as the design-space increases.

- Need to model entire design space accurately (Inputs are multi-dimensional → DM must learn over this space).
- Full simulation for whole design space is **intractable** (Can't cover all configurations with data alone).



Pre-training + Post-training

• **Pre-train** a foundation model \rightarrow learns global landscape approximately.



Pre-training + Post-training

- **Pre-train** a foundation model \rightarrow learns global landscape approximately.
- **Post-train** on current design point \rightarrow learns local landscape accurately.



Pre-training + Post-training

- **Pre-train** a foundation model \rightarrow learns global landscape approximately.
- **Post-train** on current design point \rightarrow learns local landscape accurately.
- Backpropagation of diffusion model to its conditioning gives us gradient surrogate.
- Enables gradient-based optimization.



Diffusion Model

Denoising Diffusion Probabilistic Models



- Forward process: gradually add noise.
- Reverse process: gradually remove noise.

Forward diffusion

$$d\boldsymbol{x}_t = f(t)\boldsymbol{x}_t dt + g(t)d\boldsymbol{w}_t$$

Backward diffusion

$$d\boldsymbol{x}_t = \left[f(t)\boldsymbol{x}_t - \boldsymbol{g}(t)^2 \,\nabla_{\boldsymbol{x}_t} \log p_t(\boldsymbol{x}_t)\right] dt + g(t) d\boldsymbol{w}_t$$

Diffusion Model

Denoising Diffusion Probabilistic Models

• General generative model is built based on **probability density function**:

P.D.F (probability density function):

$$p_{\theta}(x) = rac{e^{-f_{\theta}(x)}}{Z(\theta)} \iff Z(\theta) = \int e^{-f_{\theta}(x)}$$

• DDPMs use score function instead → **score-based model**:

$$\nabla_x log(p_x)$$

• The sampling process of DM is iterative:

$$x_i \leftarrow x_i + \epsilon \nabla_x log(p_x) + \sqrt{2\epsilon} z_i, i = 0, 1, ..., K$$



Song, Yang. "Generative Modeling by Estimating Gradients of the Data Distribution." *yang-song net*, 5 May 2021, https://yang-song.netblog/2021/score/.



We can add an extra parameter y to the score-function to guide the model based on our design \rightarrow toward **conditional** denoising diffusion probabilistic models.

Diffusion Model

Conditional Denoising Diffusion Probabilistic Models





- Conditioning = guide the generation process.
- Learn the conditioned data distribution.





Sampling

Efficient Sampling



Sampling

Efficient Sampling



DDIM (deterministic sampling):

$$x_{1-1} = \sqrt{\bar{\alpha}_{t-1}}x_0 + \sqrt{1 - \bar{\alpha}_{t-1}}\epsilon_{\theta}(x_t, t))$$

- Efficient sampling.
- Requires fewer steps.
- Deterministic: can be differentiated.

Sampling

Efficient Sampling



DDIM (deterministic sampling):

$$x_{1-1} = \sqrt{\bar{\alpha}_{t-1}} x_0 + \sqrt{1 - \bar{\alpha}_{t-1}} \epsilon_{\theta}(x_t, t))$$

- Efficient sampling.
- Requires fewer steps.
- Deterministic: can be differentiated.

Post-train model: Adapter



Post-train model: Low-rank Adaptation (LoRA)



Post-train model: Low-rank Adaptation (LoRA)



Output_{pretrain} = W * X $W_{LoRA} = W + r$ Where r is low-rank update $r = W_A * W_B$

Differentiable Modeling for Calorimeter Simulation using Diffusion Models

Post-train model: Low-rank update (LoRA)

 $Output_{pretrain} = W * X$ $W_{LoRA} = W + r$ Where r is low-rank update $r = W_A * W_B$



We can adapt our model with different configuration (eg. CRILIN, PANDA, etc.) using LoRA post-training, require only small amount of data and training time

A Case Study

Muon Collider

- CRILIN project: Optimization of FbF₂ calorimeter (ECAL).
- Pre-train on 5x5x5 Geant4 datasets with diverse cell sizes, materials, energies.
- Post-train on 1x1x4cm³ cells (CRILIN baseline).
- Used in loop: update design \rightarrow post-train \rightarrow compute gradient \rightarrow next iteration.



Training Setup

X 🔺

Training Data:

- Simulation: Generated using GEANT4.
- **Dataset Size**: 100,000 calorimeter shower images.
- Energy levels: 1, 10, 50, 100, 200 GeV.
- Granularity: 5 x 5 x 5 cells.
- Material: Lead fluoride (PbF₂), Lead Tungstate (PbWO₄).
- Addition labels: 5 different cell size configurations:
 - 2 x 2 x 6 cm³
 - 3 x 3 x 8 cm³
 - 4 x 4 x 10 cm³
 - 5 x 5 x 15 cm³
 - 6 x 6 x 20 cm³



Comparison of GEANT4 and DDPM-generated showers:

- DDPMs accurately reproduces GEANT4-generated showers.
- No significant visual differences across energy levels.
- Shows that our diffusion model effectively learns that shower distributions.



GEANT4 Ground-truth

CDDPMs Generated (at 300 epochs)



Energy profile at early epoch





Material: PbF2, Cell size: xy = 4 cm, z = 10 cm

- Two distributions closely matched across all energy levels. ٠
- Improved peak alignment and energy deposition spread.

- Huge mismatch between generated and ground-truth profiles.
- Energy depositions are underestimated, and peak positions are deviate.

Our model significantly enhances fidelity and successfully learns the shower calorimeter showers characteristics over time.

0.4 -

Material: PbF2, Cell size: xy = 4 cm, z = 10 cm 1 GeV 10 GeV 50 GeV - Ground truth Generated 0.8 2.5 0.6

1.0 -융 0.2 0.5 -0.5 -Ś 10 15 20 25 15 20 25 10 15 25 Ground truth 1.0 -Generated 8.0 a 0.6 1.5 0.4 10 0.0 -15 20 25 15 20 25 30 15 20 25 20 25 5 10 30 10 10 10 15 Coordinate Coordinate Coordinate Coordinate

15

1.0

100 GeV

2.5 -

2.0

1.5



- Quantify the physical fidelity of the generated shower, we use **physics-motivated metrics**.
- Total Energy (E):

$$E = \sum_{x,y} I(x,y)$$

Sum of intensities (energy deposits) over the full grid.

• Energy-Weighted Radius (R_e):

$$R_{
m E} = rac{\sum_{x,y} I(x,y) \left[(x-ar{x})^2 + (y-ar{y})^2
ight]}{\sum_{x,y} I(x,y)}$$

Describes spread of the shower around its center.

• Shower Dispersion (σ_{γ}) :

$$\sigma_y = \sqrt{\frac{\sum_y (y - \bar{y})^2 I_y y}{\sum_y I_y(y)}}$$

Measures vertical spread using energy profile $I_y(y)$.

Mean Squared Error of each metric across training epochs





These results demonstrate that the model not only learns to generate visually plausible showers, but it generate highly accurate **HEP simulation data** \rightarrow it can serve as **high fidelity surrogate**.

Model performance on unseen cell size configurations

The conditional DDPMs was tested on cell size configuration that were not included in the training set to evaluate its ability to generalize.

- Model captures the overall distribution but exhibits some deviate in shape.
- The transverse profiles are well-aligned with the ground-truth.



Pre-train predicted Material: PbF2, Cell size: xy = 3 cm, z = 14 cm

Results on Unseen Data post-train

Pre-trained Energy profile

LoRA post-trained Energy profile

LoRA post-train Material: PbF2, Cell size: xy = 3 cm, z = 14 cm





Pre-train predicted Material: PbF2, Cell size: xy = 3 cm, z = 14 cm

- Generated energy deposition tends to overshoot the ground truth. ٠
- LoRA fine-tuning align closely with the ground truth across all energy levels.
- Post-training reduces discrepancies with little training time.

Conclusion of working progress

Advantages of the Pre-train + Post-train Approach

- Achieves high fidelity across a wide design space.
- Requires only **limited data** for local adaptation.

Speed Benefit

- For simple designs (e.g., homogeneous calorimeters), the speed gain is modest.
- For complex or hybrid geometries, surrogate-based sampling will be significantly faster.

Next Steps

- Evaluate the **quality of gradients** from the surrogate.
- Apply gradient-based optimization to real design tasks.





- A diffusion model surrogate can generate high-fidelity calorimeter showers, conditioned on design.
- Pre-train + post-train to account for the large design space.
- DDIM sampling makes inference fast and differentiable.
- Gradients can be obtained through backpropagated through diffusion model to the conditioning parameters.
- Optimization with surrogate gradients is ongoing.









R

Thank you for your attention!