国立大学法人
岩手大学
IWATE UNIVERSITY

Reconstructing Sparticle masses using Generative Machine Learning

Based on Eur.Phys.J.ST 235 (2026) 7, 1903-1914, with R. Barman & A. Choudhury

The International Workshop on Future Linear Colliders 2026 (LCWS2026)

Subhadeep Sarkar

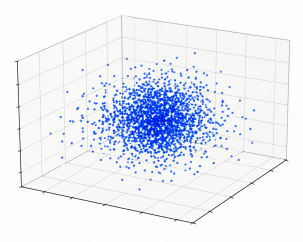
Email: sarkar@iwate-u.ac.jp

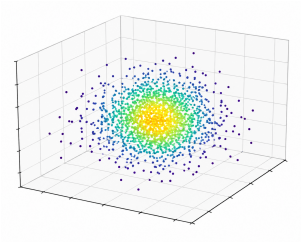
Need for ML

- More data $\rightarrow$ hard to deal with
- High-dimensional data $\&$ complex relations
- Testing several BSM scenarios

> ☞ Generative models can unfold high-dimensional data
> ☞ Fast likelihood-inference method

[1907.10621], [2205.07912]



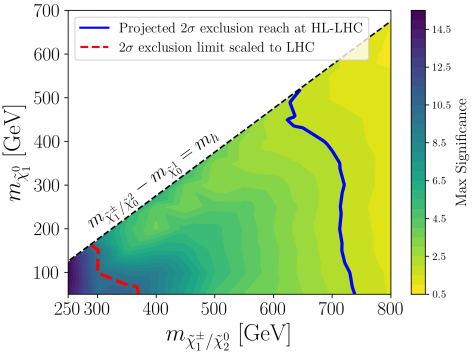


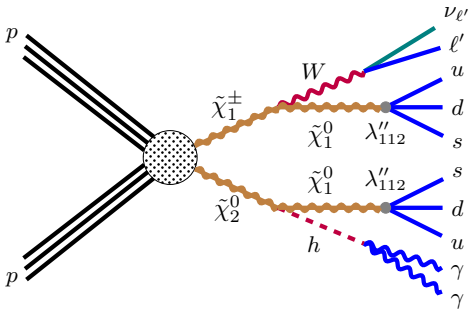
Generative models can estimate the probability distribution of data

If we observe an excess over the SM $\rightarrow$ How accurately can we reconstruct the underlying new particle masses across a wide BSM parameter space?

If we observe an excess over the SM → How accurately can we reconstruct the underlying new particle masses across a wide BSM parameter space?

- A test Scenario:
  $pp \to \tilde{\chi}_1^\pm \tilde{\chi}_2^0 \to 1\ell + 2\gamma + \text{jets}$ at HL-LHC
- Reconstruct $\tilde{\chi}_1^0$ & $\tilde{\chi}_2^0$ masses





We use a hybrid Transformer-based Encoder + Diffusion Network

Model Overview



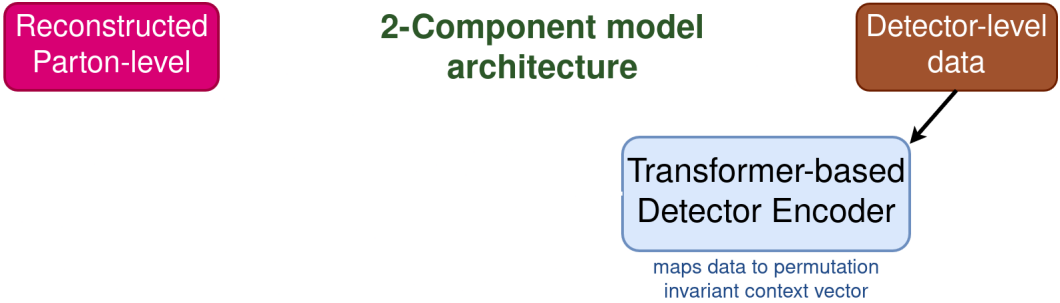
Model Overview

Reconstructed Parton-level

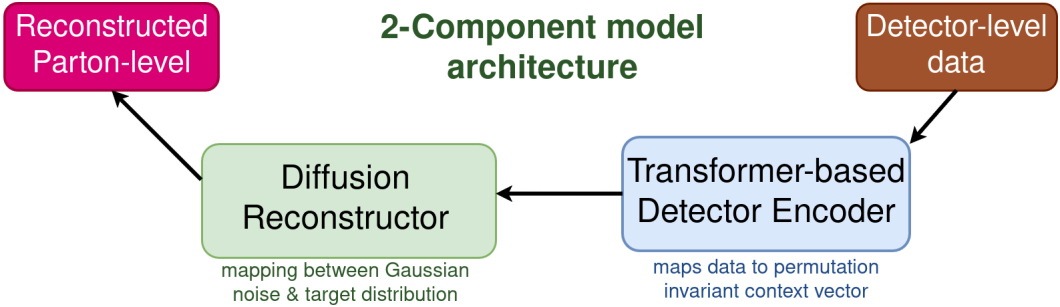
2-Component model architecture

Detector-level data

Model Overview



Model Overview



Transformer encoder

Detector-level final state: $\ell + 2\gamma + \text{jets} + \not{E}_T$

Inputs

$\{p_x, p_y, p_z, E\}_i : i = l, \gamma_1, \gamma_2, j_n \ (2 \leq n \leq 4)$
$\Delta R_{min}(l, j_n), \Delta R_{max}(l, j_n), \Delta R_{min}(l, \gamma)$
$\Delta R_{max}(l, \gamma), \Delta R_{min}(\gamma, j_n), \Delta R_{max}(\gamma, j_n)$
$\sum_{i=1}^{n} p_T(j_i), \not{E}_T$

Self-Attention

- Each token attends all tokens in same sequence
- Assigns **Query (Q)**, **Key (K)** and **Value (V)**
- Outputs weighted ($A(Q, K, V)$) tokens

Vaswini et.al. "Attention Is All You Need"

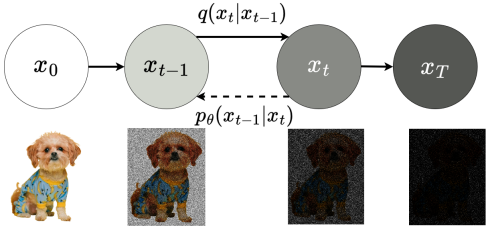
$\ell(p_x, p_y, p_z, E) \Rightarrow d = 256$ latent rep.
$\gamma_a(p_x, p_y, p_z, E) \Rightarrow d = 256$ rep. (a=1,2)
$j_b(p_x, p_y, p_z, E) \Rightarrow d = 256$ rep. (b=1,2,3,4)
Others $\Rightarrow d = 256$ rep.

256-dim context vector C is constructed by averaging over the 8 tokens.

Diffusion Model

$$\ell + 2\gamma + \text{jets} + \not{E}_T \quad \rightarrow\rightarrow \quad \tilde{\chi}_2^0\{p_x, p_y, p_z, E\}, \tilde{\chi}_1^0\{p_x, p_y, p_z, E\}$$

- ☆ Iteratively add noise to data (Forward process)
- ☆ Learn to restore data back to original (Reverse process)



Training

- ✎ C and $(pe)_{d_t}$ passed through an MLP
- ✎ Noised target (dim 8) is concatenated $\rightarrow$ fed into an MLP to predict noise $\eta(C, (pe)_{d_t}, x_p^t)$
- ✎ Loss $\mathcal{L} = \mathbb{E}_{x_p^0, t, \epsilon} ||\epsilon - \eta(C, (pe)_{d_t}, x_p^t)||^2$

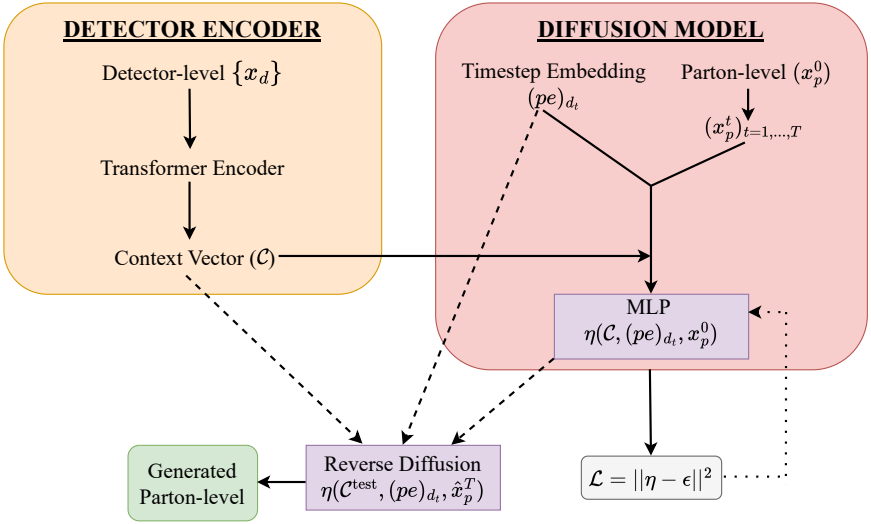
Testing

Go to previous step $\hat{x}_p^{T-1} = \mu_T + \sqrt{\beta_T}\, z$

$$\mu_T = \frac{1}{\sqrt{\alpha_T}} \left(\hat{x}_p^T - \frac{1-\alpha_T}{\sqrt{1-\bar{\alpha}_T}} \eta \left(C, (pe)_{d_t}, \hat{x}_p^T \right) \right)$$

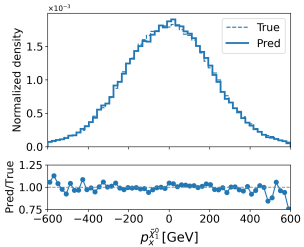
J. Ho et.al. (2006.11239)

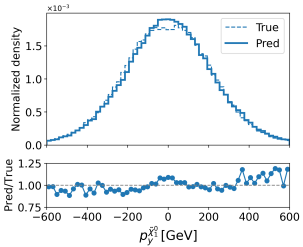
Model Schematics



Parton-level distribution for p_x & p_y of $\tilde{\chi}_1^0$: True Vs. Prediction

- **Fixed Benchmark:**
  $m_{\tilde{\chi}_2^0} = 600$ **GeV,**
  $m_{\tilde{\chi}_1^0} = 200$ **GeV**

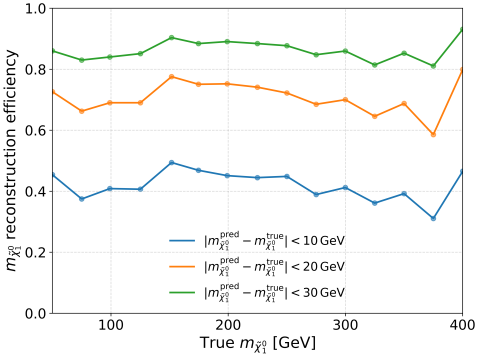




Fixed $m_{\tilde{\chi}_2^0}$, varying $m_{\tilde{\chi}_1^0}$

- $m_{\tilde{\chi}_2^0} = 600$ GeV
- Trained with $m_{\tilde{\chi}_1^0} \in [50, 400]$ GeV with 50 GeV step size
- Tested with 25 GeV step size

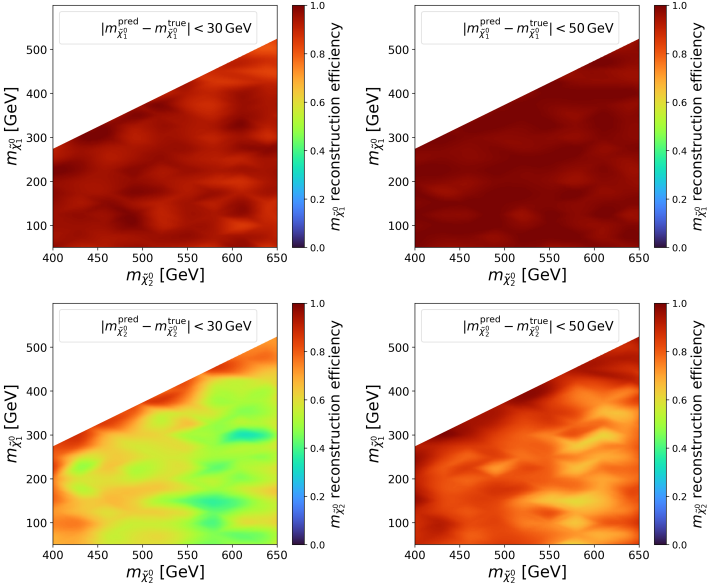
$\epsilon_{m_{\tilde{\chi}_1^0}} \gtrsim 80\%$ for $\Delta m < 30$ GeV



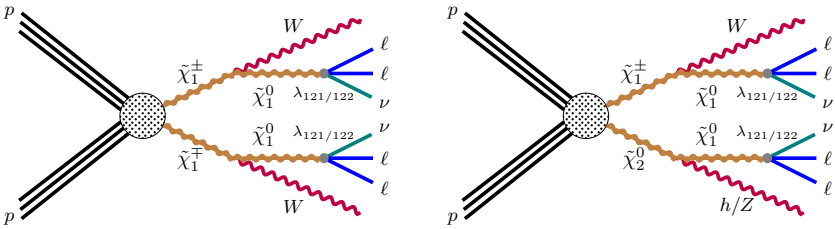
Reconstruction Efficiency

- **Varying both $m_{\tilde{\chi}_1^0}, m_{\tilde{\chi}_2^0}$**
- For training
  $m_{\tilde{\chi}_2^0} \in [400, 600]$ GeV
  $m_{\tilde{\chi}_1^0} \in [50, m_{\tilde{\chi}_2^0} - m_h]$

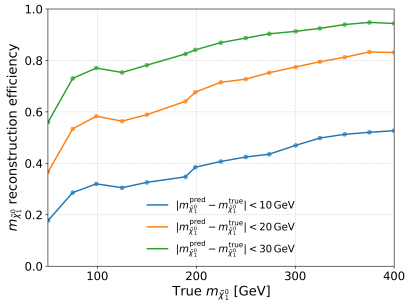
✎ $\epsilon_{m_{\tilde{\chi}_1^0}} \gtrsim 70\%$ for $\Delta m = 30$ GeV
✎ Less for $m_{\tilde{\chi}_2^0}$

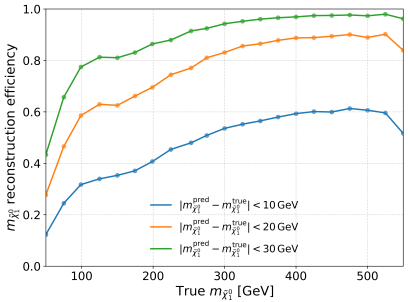


Test Scenario 2: The LLE Benchmark Scenario



Reconstruction Efficiencies





Summary & outlook

- We introduce a 2-component network: A transformer-based detector encoder and a Diffusion network
- For a test scenario, we select a *UDD*-type RPV SUSY channel, and we reach $\gtrsim 70\%$ reconstruct efficiency for $\tilde{\chi}_1^0$ mass with a tolerance of $\Delta m = 30$ GeV
- For *LLE*-type scenario, we reach $\gtrsim 80\%$ reconstruction efficiency for the $\tilde{\chi}_1^0$ mass for Δm=30 GeV
- Our architecture can be easily adapted to other BSM scenarios and other collider datasets too.