

Foundation Models

Autoregressive Generation

SPADE

SPLit And DELay Embeddings for Autoregressive High-Granularity Calorimeter Simulation

EUCAIFCON 2026 · 26 AUGUST 2026 · KIP, HEIDELBERG

AUTHORS

Joschka Birk, Frank Gaede, Anna Hallin, Gregor Kasieczka, Martina Mozzanica, Henning Rose

✉ henning.rose@uni-hamburg.de | [arXiv:2606.11304](https://arxiv.org/abs/2606.11304) | github.com/uhh-pd-ml/SPADE



CLUSTER OF EXCELLENCE
QUANTUM UNIVERSE



From Text to Tensors

We all use LLMs. This is how they work.



"The quick fox"

Text

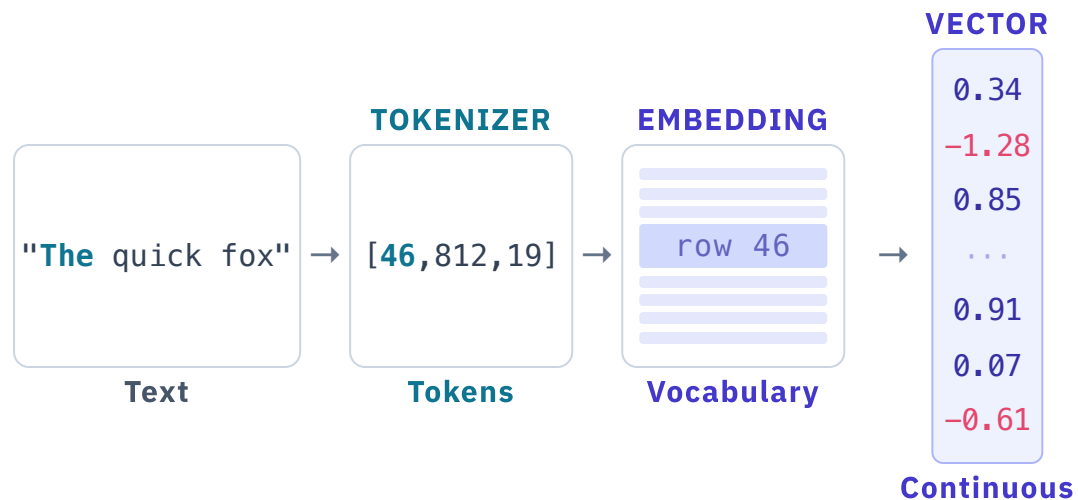
From Text to Tensors

We all use LLMs. This is how they work.



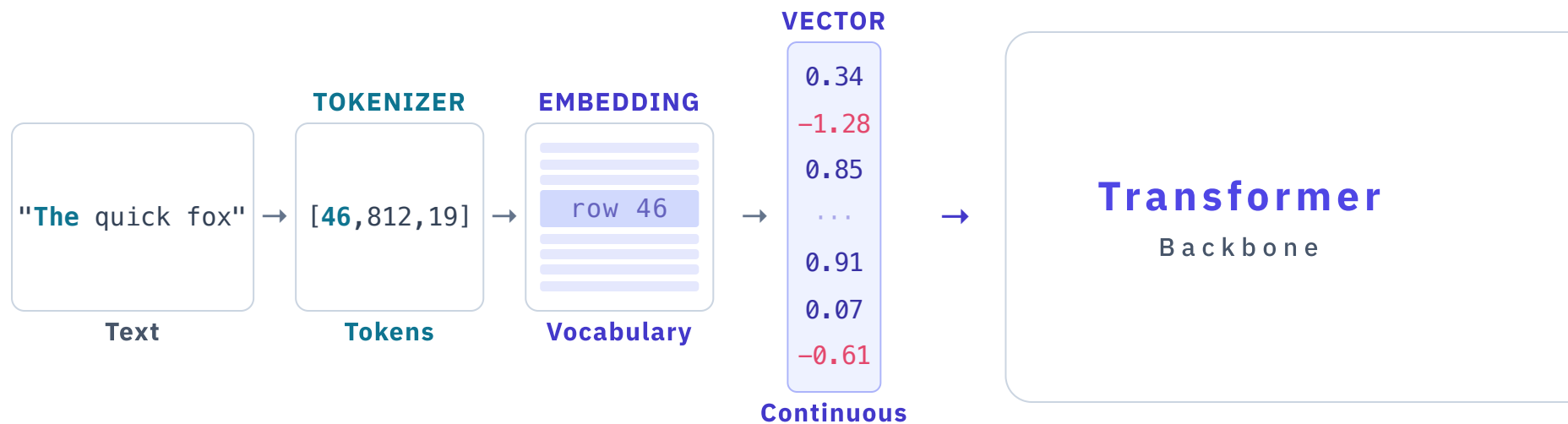
From Text to Tensors

We all use LLMs. This is how they work.



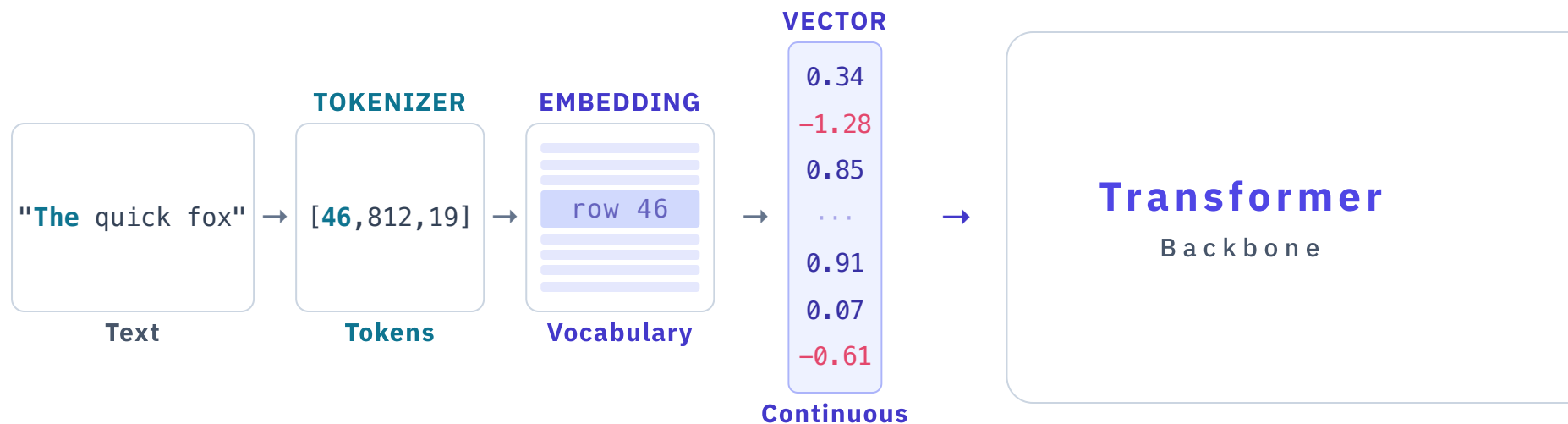
From Text to Tensors

We all use LLMs. This is how they work.



From Text to Tensors

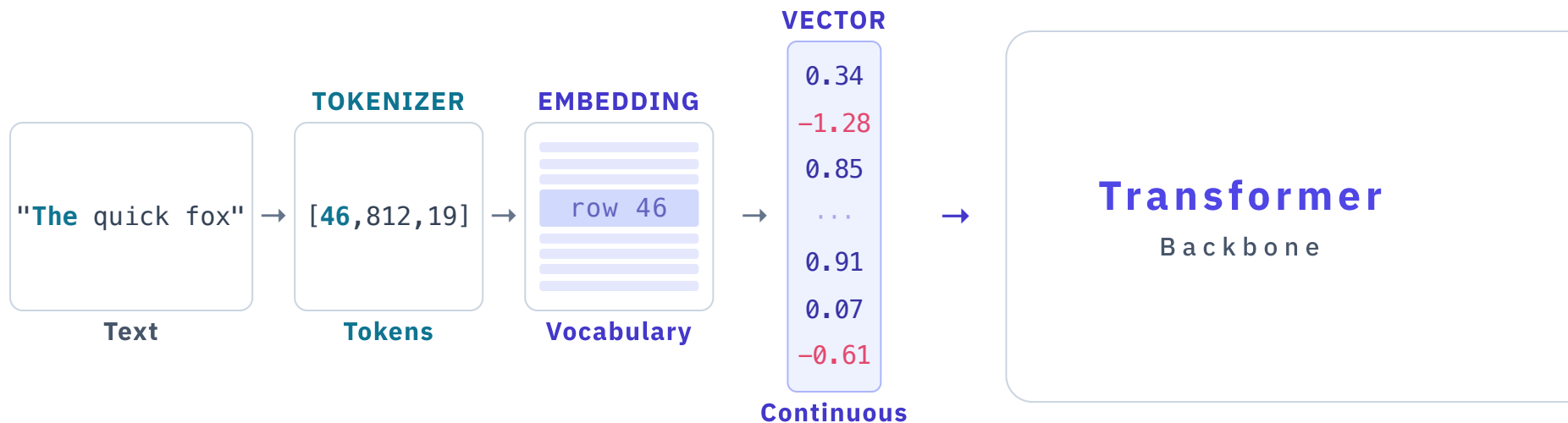
We all use LLMs. This is how they work.



Text: a finite vocabulary. Every token is one lookup.

From Text to Tensors

We all use LLMs. This is how they work.



Text: a finite vocabulary. Every token is one lookup.

Physics: continuous data, no vocabulary. Nothing to look up.

Dataset & Geometry

γ -showers ^[3]

High-granularity ILD ECAL ^[4]
30×30×30 sensor grid

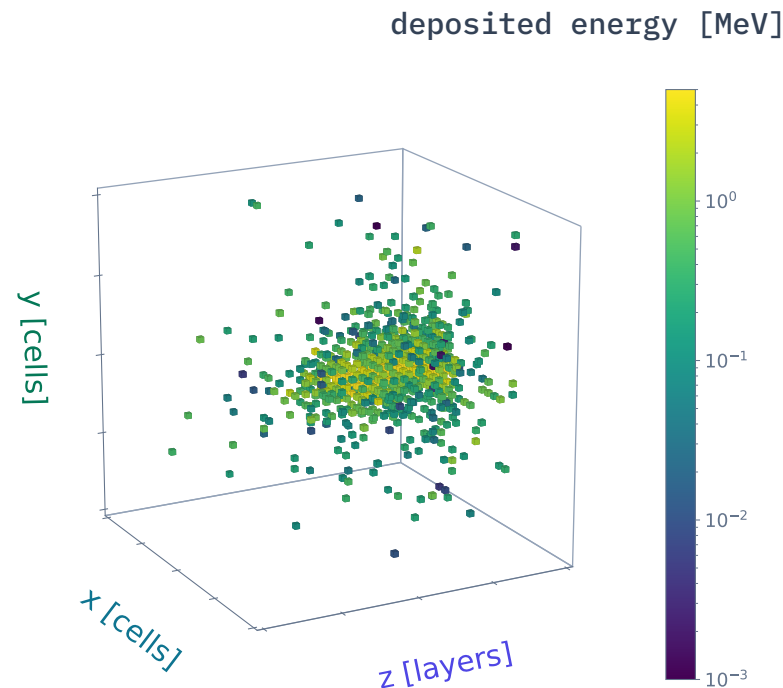
Incident energy: 10–100 GeV

Hits per shower: 100–1700

Structure: 4D, sparse

^[3] [Buhmann et al. 2005.05334 \(2020\)](#)

^[4] [ILD Collaboration 2003.01116 \(2020\)](#)



Dataset & Geometry

γ -showers ^[3]

High-granularity ILD ECAL ^[4]

30×30×30 sensor grid

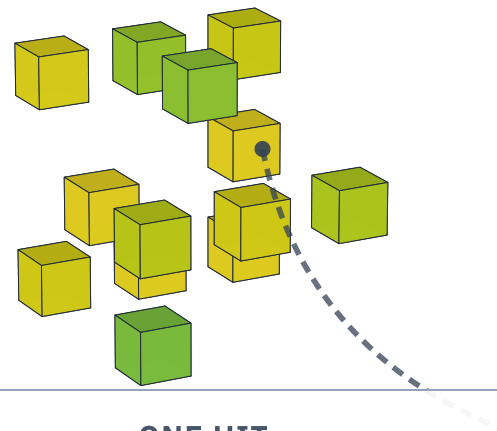
Incident energy: 10–100 GeV

Hits per shower: 100–1700

Structure: 4D, sparse

[3] Buhmann et al. 2005.05334 (2020)

[4] ILD Collaboration 2003.01116 (2020)



ONE HIT

[z: 5, x: 14, y: 22, E: 2.1 MeV]

Dataset & Geometry

△ Granularity Bottleneck

One token per voxel. The vocabulary multiplies:

$$\text{Vocabulary} = N_x \times N_y \times N_z$$

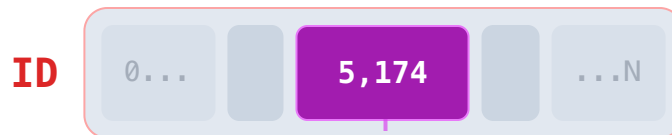
30×30×30 grid 27,000

Real ILD ECAL^[4] ~30 M

[4] [ILD Collaboration 2003.01116 \(2020\)](#)

SINGLE VOCABULARY

1 voxel = 1 token ID



ONE HIT

[z: 5, x: 14, y: 22, E: 2.1 MeV]

Dataset & Geometry

⚡ Factorized Vocabularies

One vocabulary per coordinate. The vocabulary adds:

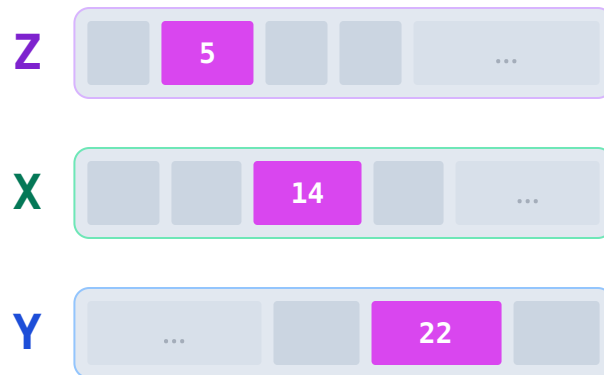
$$\text{Vocabulary} = N_x + N_y + N_z$$

30×30×30 grid **90** (vs 27,000)

Real ILD ECAL^[4] **~hundreds**

[4] [ILD Collaboration 2003.01116 \(2020\)](#)

INDEPENDENT VOCABULARIES



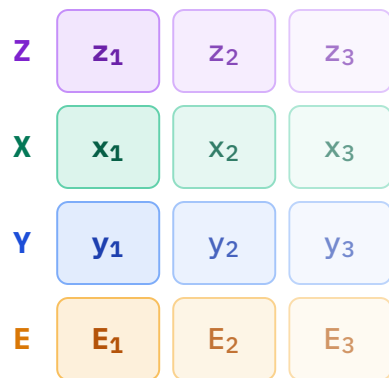
ONE HIT

[z: 5, x: 14, y: 22, E: 2.1 MeV]

The Delay Mechanic



NO DELAY: FOUR PARALLEL STREAMS



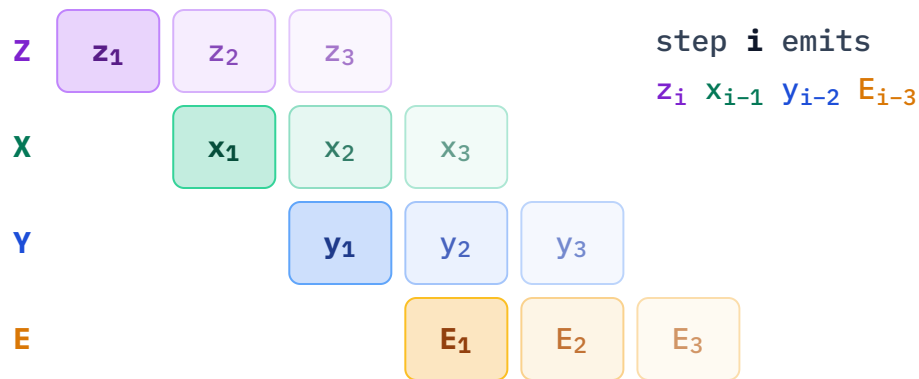
Parallel: $P(z) \cdot P(x) \cdot P(y) \cdot P(E)$

Coordinates guessed independently. The geometry falls apart.

The Delay Mechanic



SPLIT-AND-DELAY: EACH STREAM SHIFTED BY ONE STEP



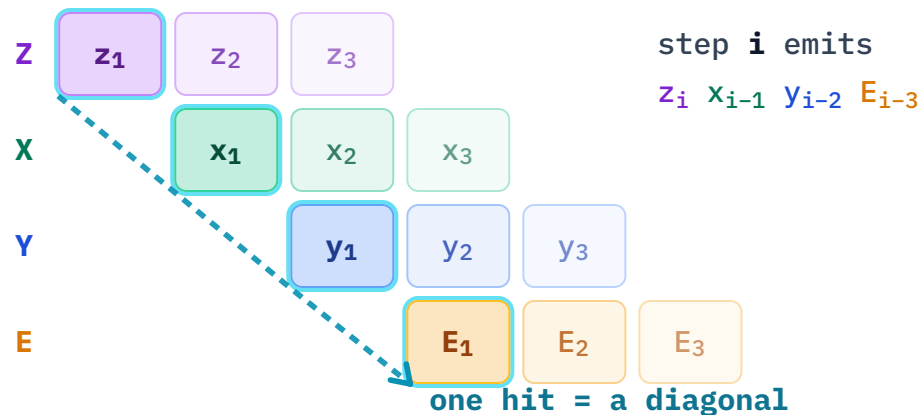
Split-and-Delay: $P(z) \cdot P(x|z) \cdot P(y|z,x) \cdot P(E|z,x,y)$

Exact within-hit chain rule. No 4× longer sequence.

The Delay Mechanic



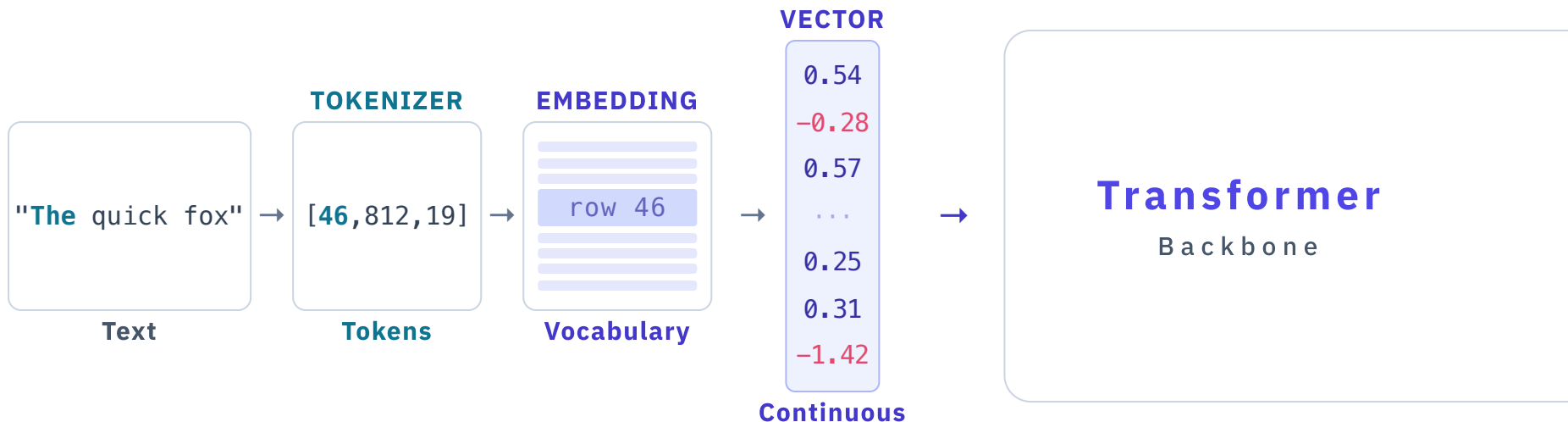
SPLIT-AND-DELAY: EACH STREAM SHIFTED BY ONE STEP



Split-and-Delay: $P(z) \cdot P(x|z) \cdot P(y|z,x) \cdot P(E|z,x,y)$

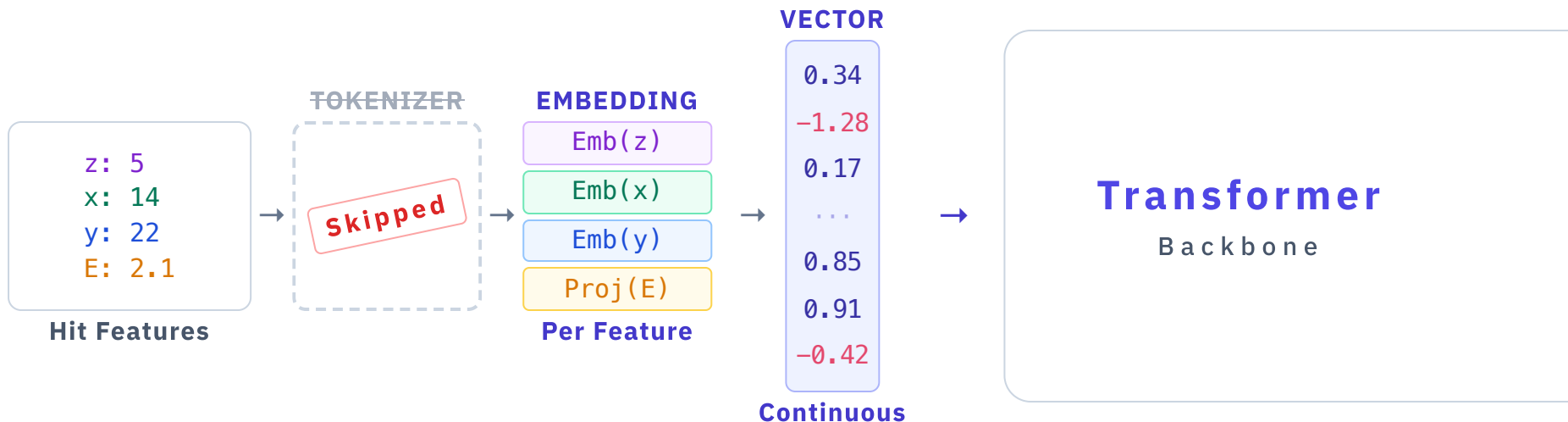
Exact within-hit chain rule. No 4× longer sequence.

Embedding



Text: a finite vocabulary. One lookup per token.

Embedding



Physics: one embedding per feature. No tokenizer.

Output Heads & Sampling



Transformer

BACKBONE

Output Heads & Sampling



Transformer

BACKBONE



Vector

-0.27

-0.54

0.17

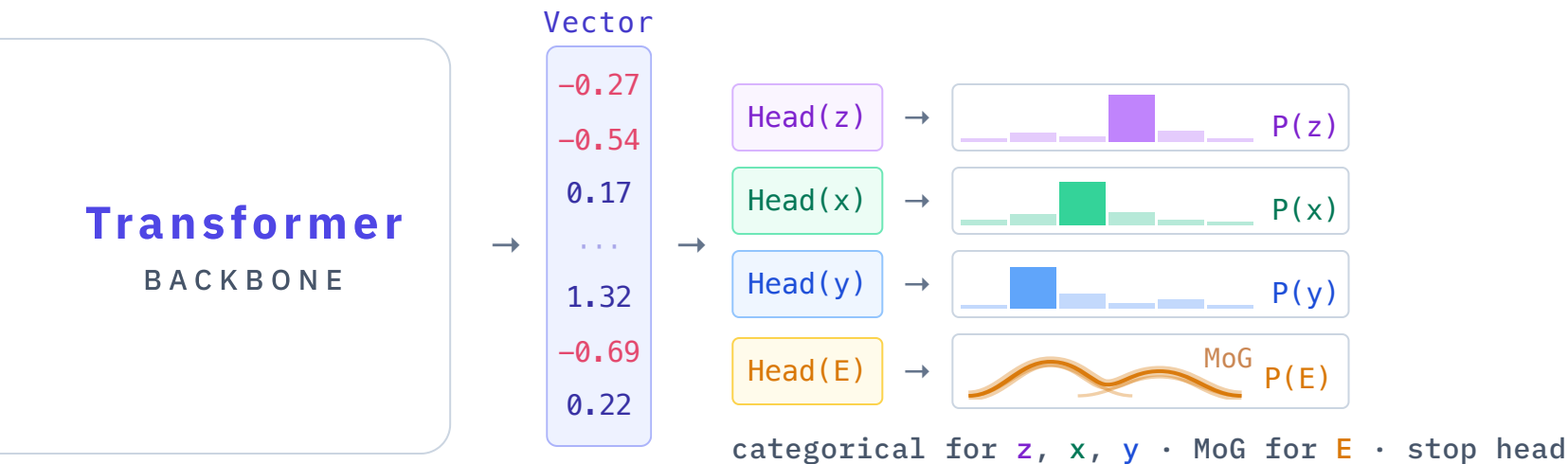
...

1.32

-0.69

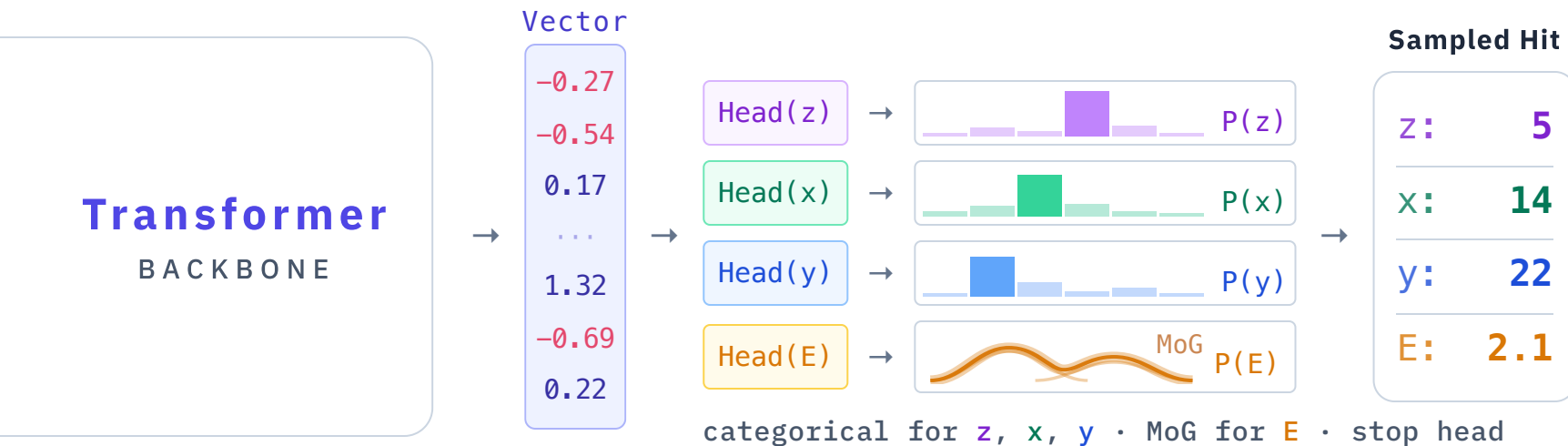
0.22

Output Heads & Sampling

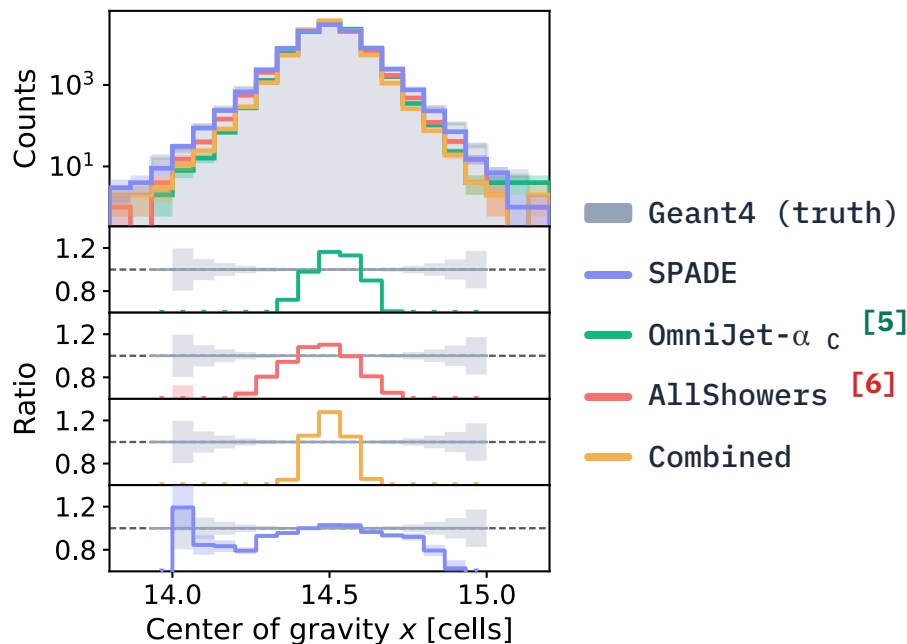


Output Heads & Sampling

S P ♠ D E
SPLIT & DELAY



Spatial Fidelity: Center of Gravity

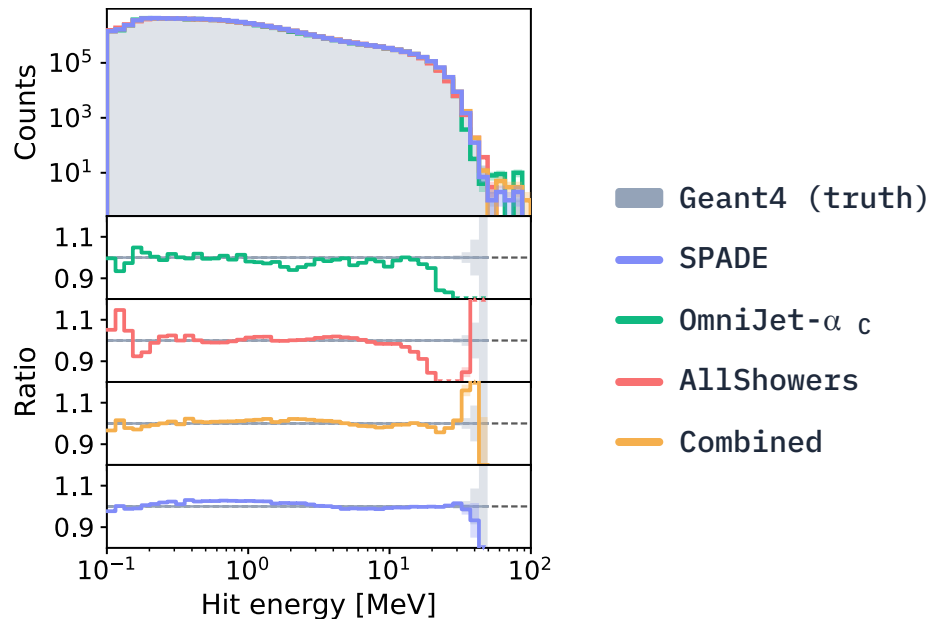


- **SPADE** models the center of gravity in x, core and tails
- The delay recovers the x and E correlation inside a hit
- Outperforms **OmniJet- α_c** and **AllShowers**

[5] [Birk et al. 2501.05534](#)

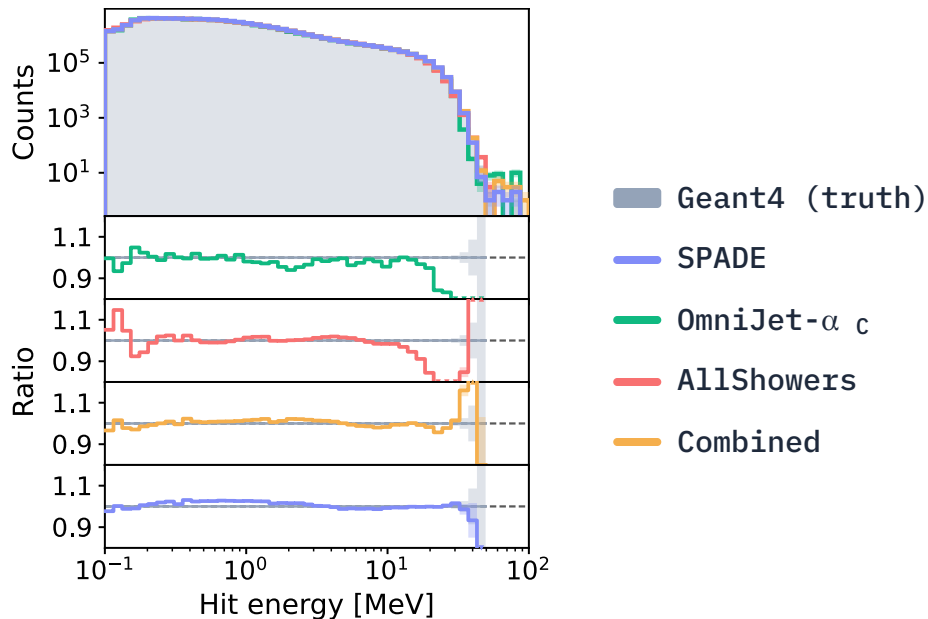
[6] [Buss et al. 2601.11716](#)

Energy Fidelity: Hit Spectrum



- **SPADE** models the energy directly. Accurate over 3 orders of magnitude
- Continuous head, conditioned on the coordinates
- No codebook limits in the tails, unlike **OmniJet- α_c**

Conclusion



SPADE needs no tokenizer. It trains natively on the detector data.

Nothing here is calorimeter-specific. One method for all high dimensional physics data.

arXiv:2606.11304

github.com/uhh-pd-ml/SPADE

henning.rose@uni-hamburg.de