IMPROVING MATERIAL IDENTIFICATION IN MUON SCATTERING TOMOGRAPHY

Angel Bueno Rodriguez, Felix Sattler, Maximilian Perez Prada, Maurice Stephan, Sarah Barnes

4th MODE Workshop

Differentiable Programming for Experimental Design

Valencia (Spain)

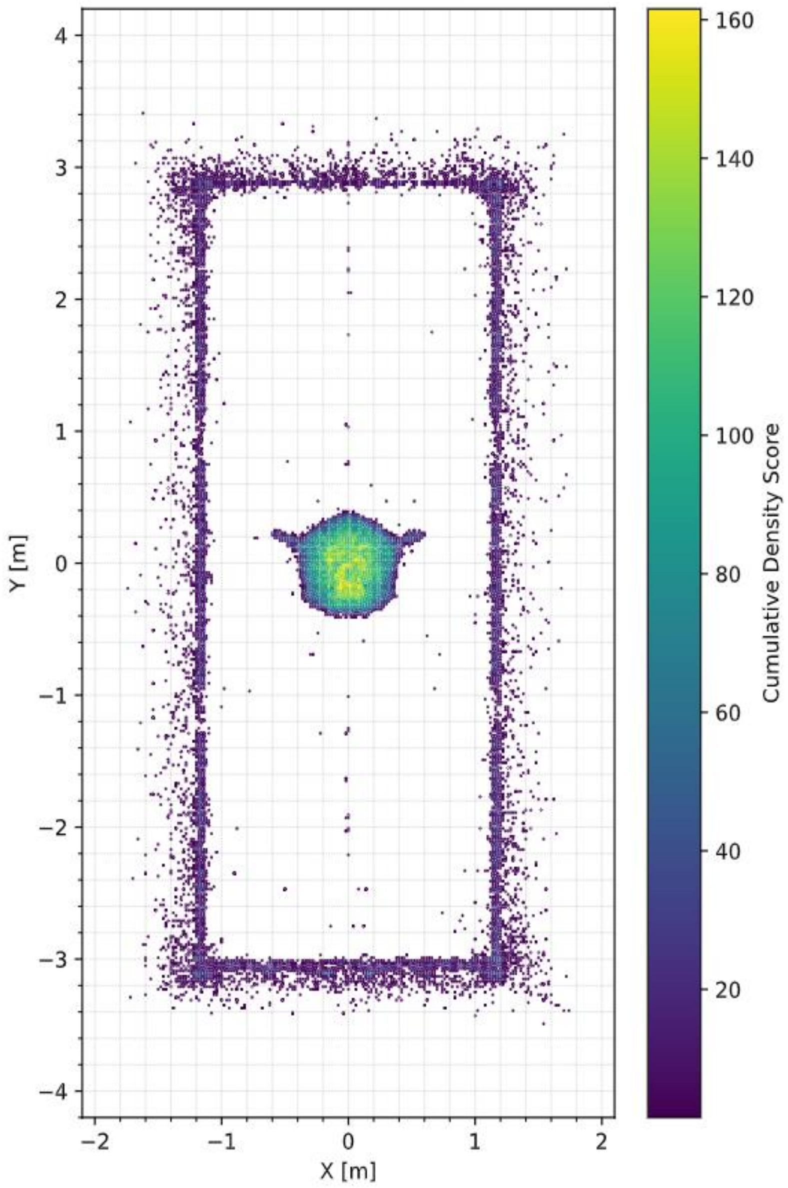
Machine learning in muon tomography

<u>Data is oil for the machine</u> **learning algorithms in muon tomography applications.** ***(Borozdin et al, 2023, Lagrange et al. 2024, PM Ruiz-del-Arbol et al. 2019, Schultz et al., 2007, …)***

Limited accessibility to experimental data constrains large-scale machine learning applications. Realistic simulations can alleviate this shortage by generating synthetic data, serving as a viable training alternative for machine learning models.

Existing solutions for creating **simulated data in particle physics is time-consuming** and limits the scalability of synthetic dataset creation.

We proposed a **novel workflow, B2G4, to accelerate the creation of detailed and scalable 3D scenes for particle physics simulations**:

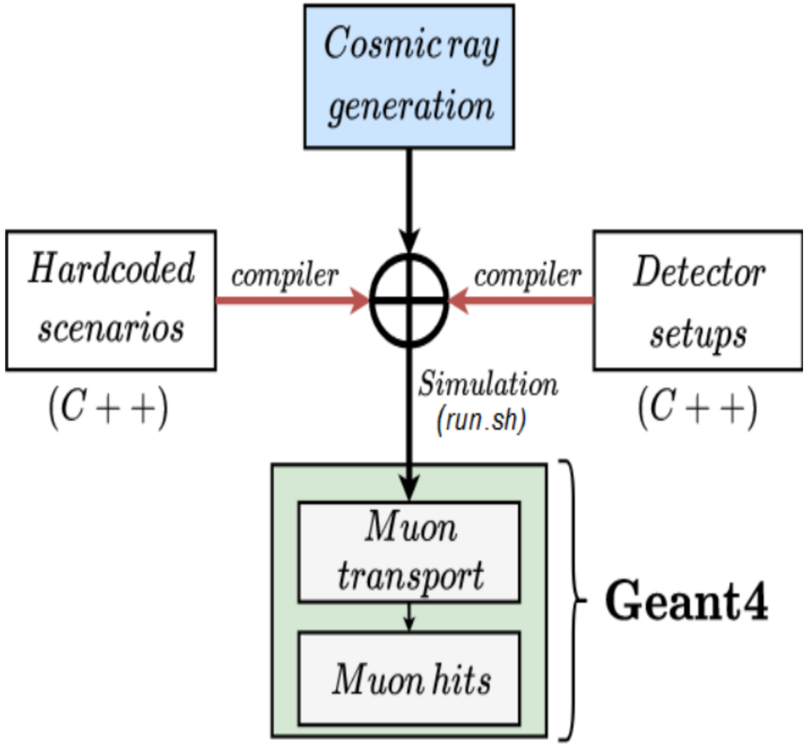


A. Bueno Rodriguez, F. Sattler, M. Perez Prada, M. Stephan, and S. Barnes, B2G4: A synthetic Data Pipeline for the Integration of Blender Models in Geant4 Simulation Toolkit", JAIS, vol. 2024, no. 1

Tackling the data scarcity problem: B2G4

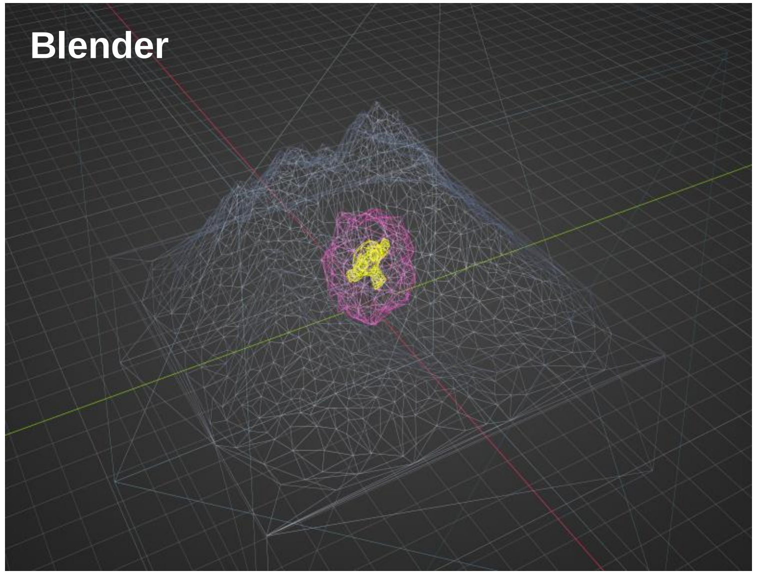
B2G4 is a novel framework that transplants highly detailed 3D scenes from Blender into Geant4 for a variety of physical applications, including muon tomography

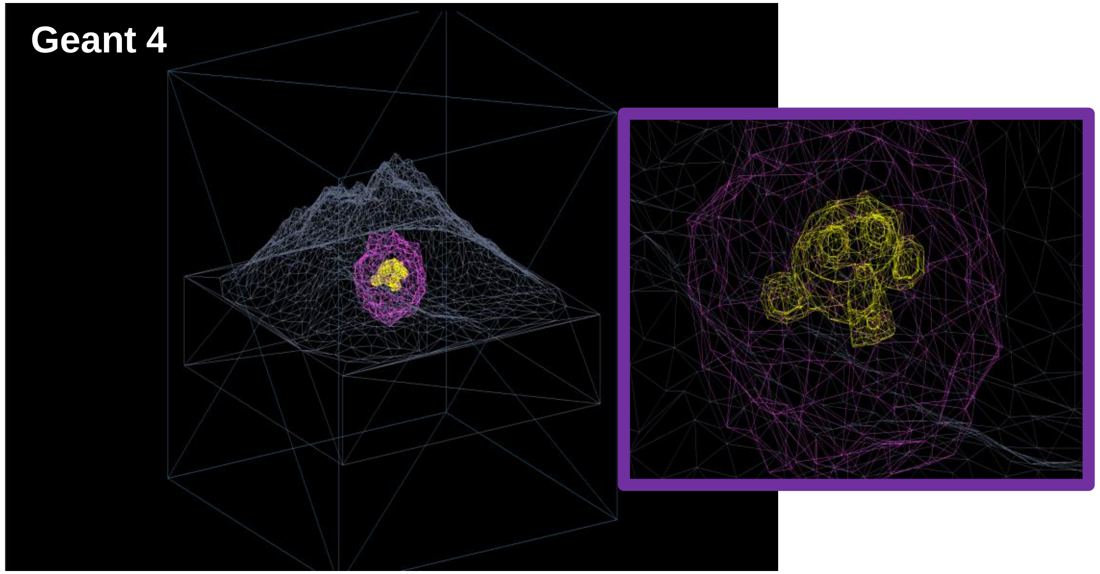
- **Geant4** supports the import of 3D models. Yet, Geant4 can be prone to parsing errors, with very limited visualization options
- **Simulation 3D scenes must be coded manually and compiled, even for minor changes.**
- **B2G4 provides a suite of Python scripts for Blender export and a C++ library for Geant4 imports:**
  - **3D scenes are simply created using Blender.**
  - B2G4 **translates the scene into a format readable by Geant4** for cosmic simulations
  - New model imports **do not require Geant4 code to be recompiled,** allowing for fast and automated dataset creation



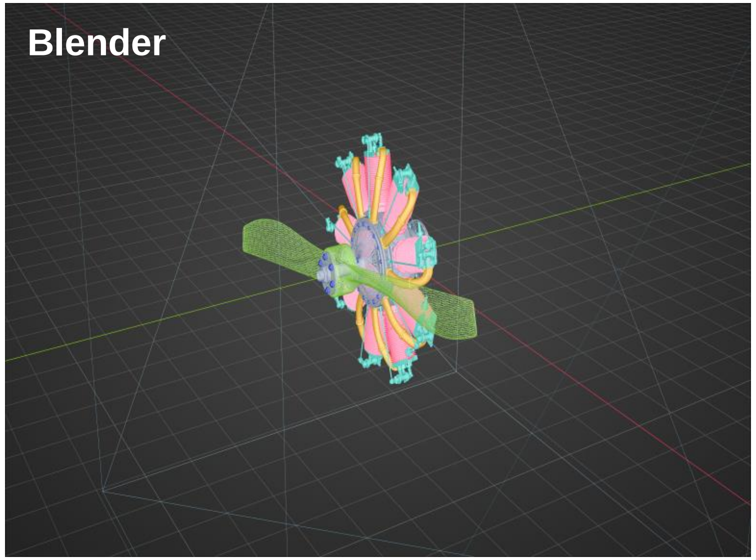
Standard setup for muon tomography simulations with Geant4

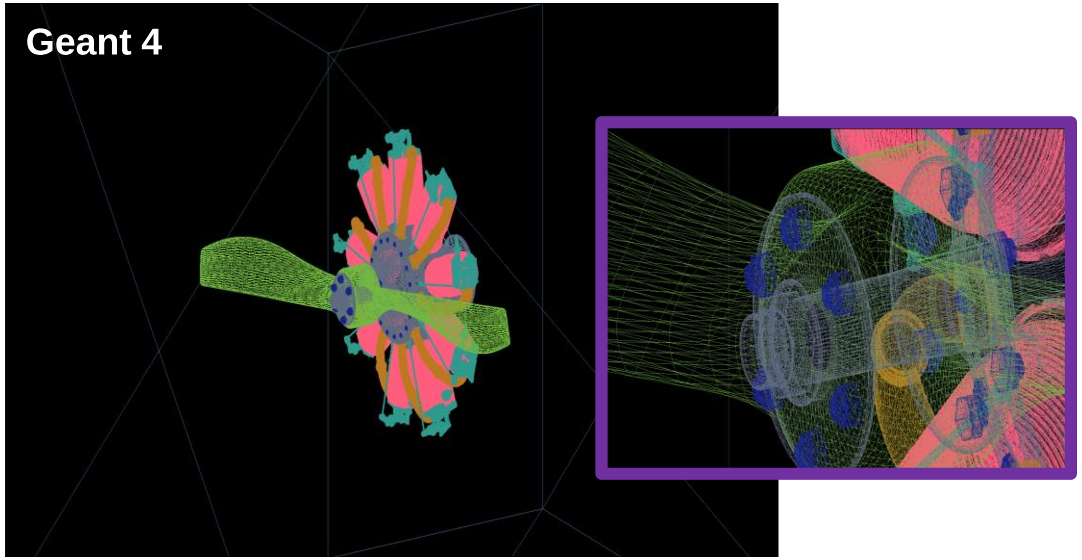
B2G4 Example scenes





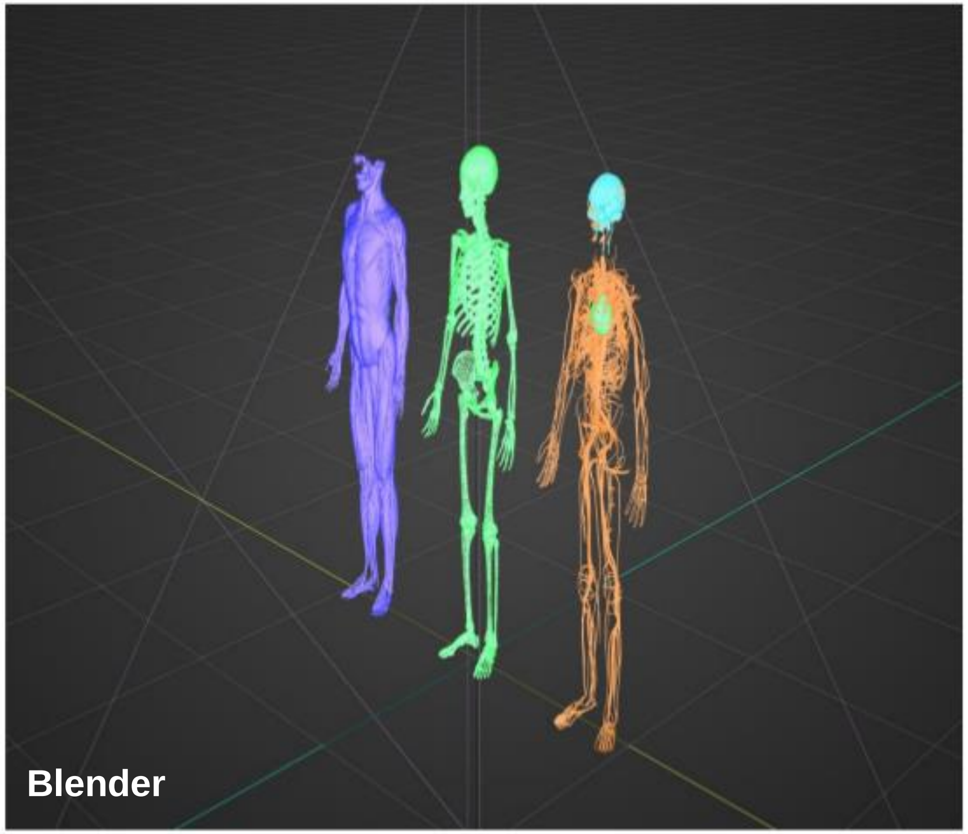
A monkey head made of gold inside an ancient structure buried in a mountain

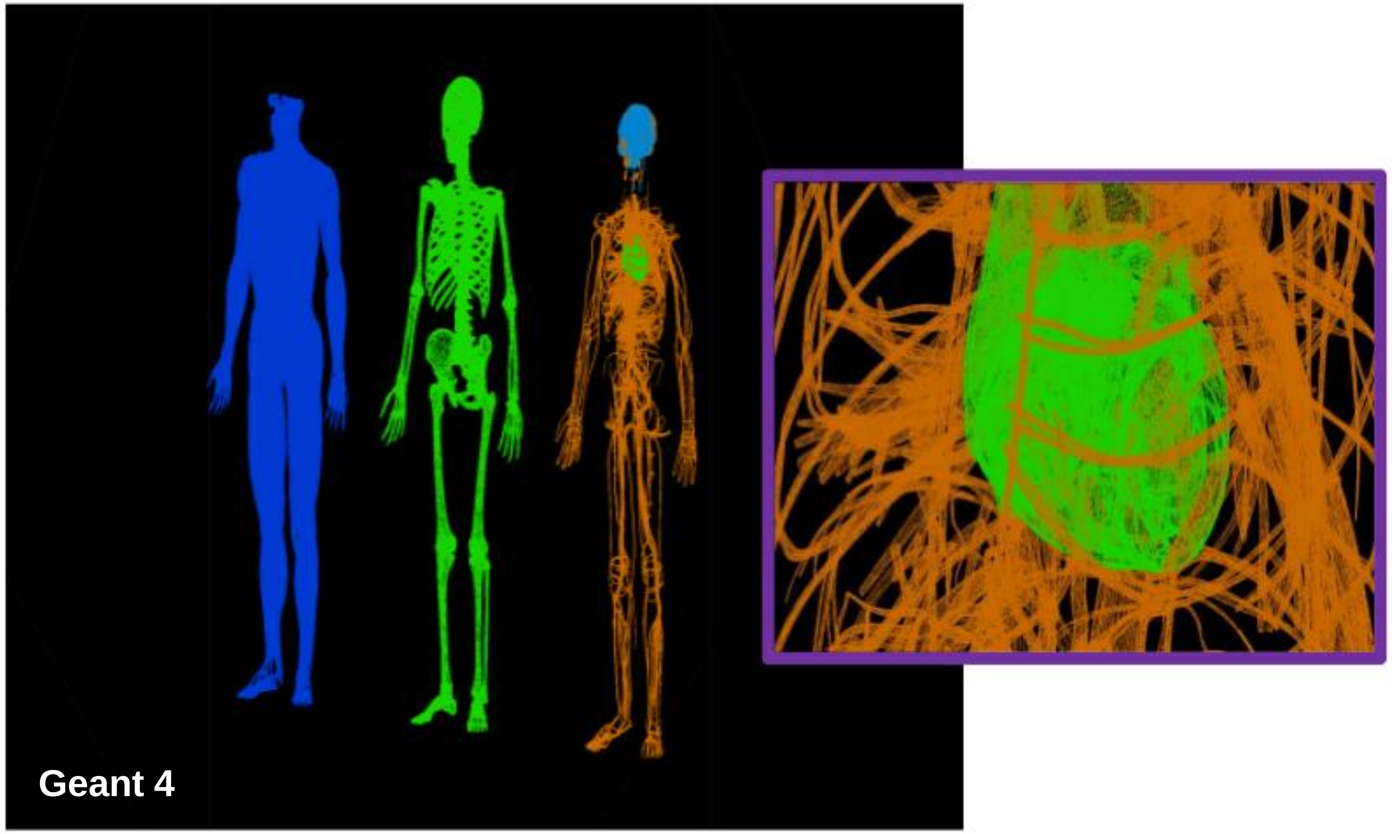




Aircraft engine. Metal bolts, propellers and cylinders from Blender are faithfully replicated in Geant4.

B2G4 Example scene (medicine)

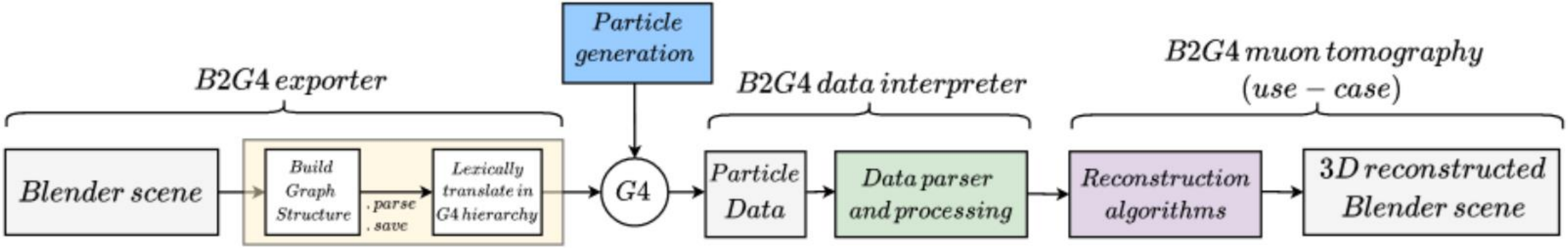




Three highly detailed human phantoms. Body parts can be easily modelled in Blender (top) and parsed into Geant4 (bottom), at full detail.

Triangles: 3.0 Million; exported in **30.6 seconds.**

B2G4 Workflow for Muon Tomography

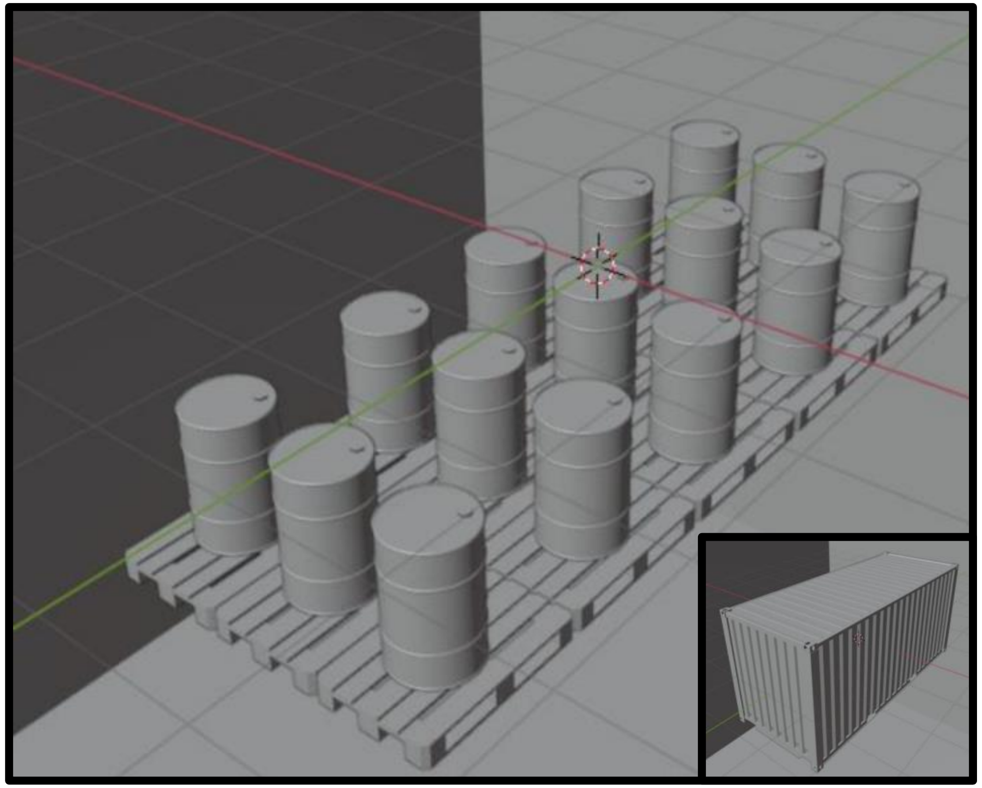


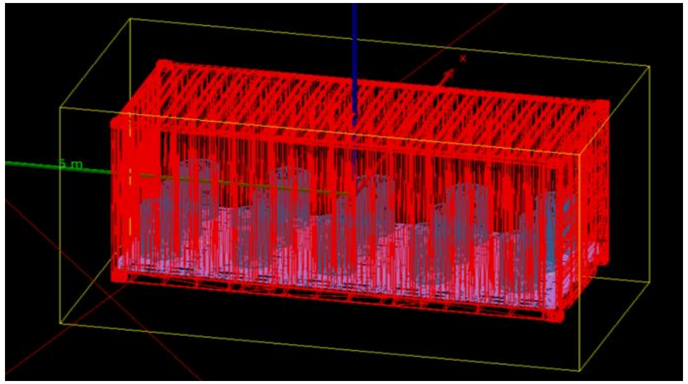
B2G4 end-to-end data pipeline.

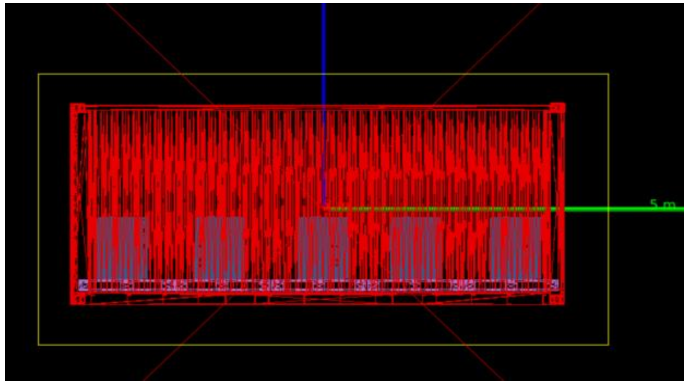
- **Scene is created in Blender:** B2G4 defines hierarchies in a way which mirrors the mother-daughter relationship model in Geant4 whilst employing instancing to avoid data duplication
- **Creation of graph data structure:** Physical and logical volume dictionaries representing the respective Geant4 objects are parsed as graphs and saved as a .JSON file. A dedicated parser generates Geant4 scene descriptions from the .JSON file.
- **Data interpreter:** Process and analyze simulation results tailored to the use case.

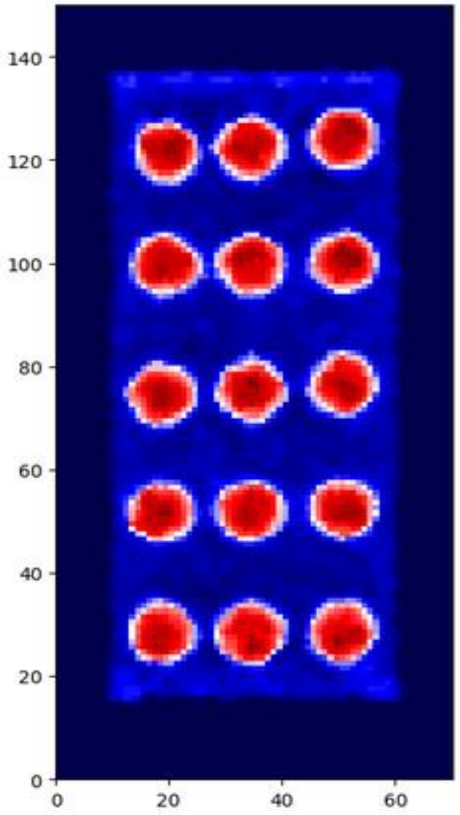
B2G4: Container simulations

B2G4 permits fast end-to-end simulation and reconstruction pipelines for physics applications, **including muon tomography for maritime containers**







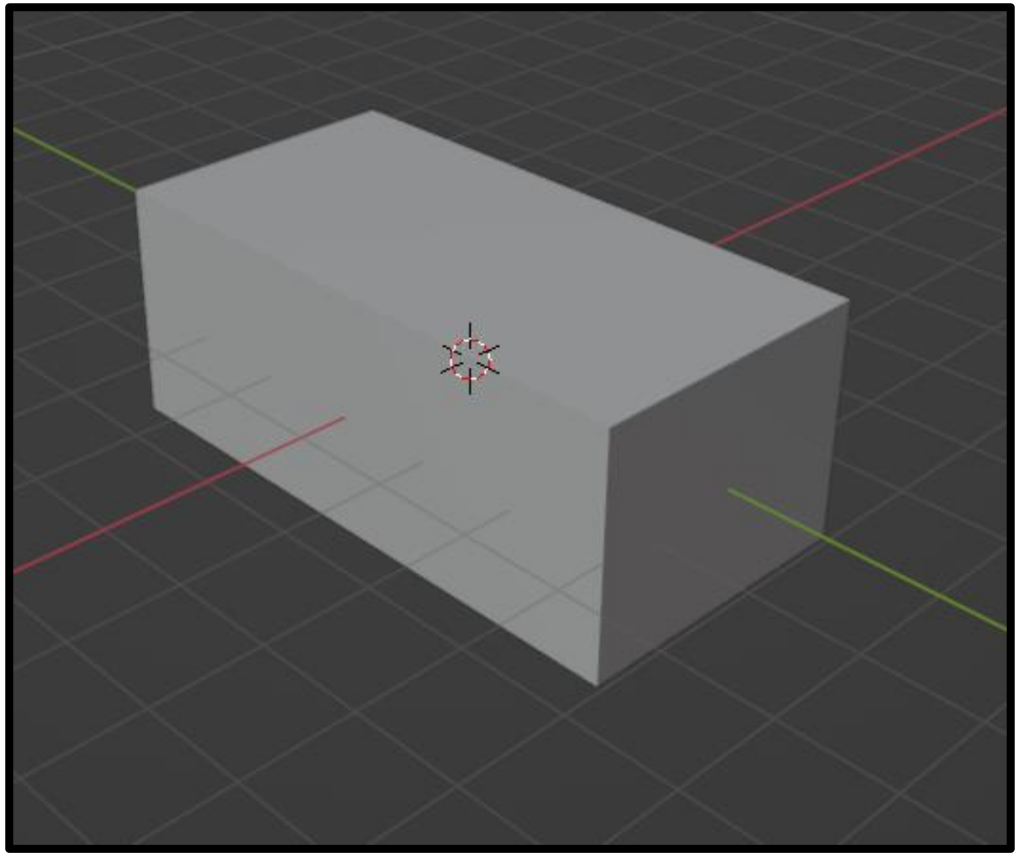


15 Barrels made of Iron (G4_Fe): (0.545, 0.797, 0.545) meters**, roughly centered** in the wooding palette
15 Wooden pallets. Each palette has a dimension of (0.76, 1.16, 0.14) meters. Random layout
1 container made of Stainless Steel (2.44, 6.07, 2,6) meters
10 Millions of muons simulated in Geant4.

HiRo: The Hidden Room dataset

- We take the classical tomography question and ask:

What do we have inside?

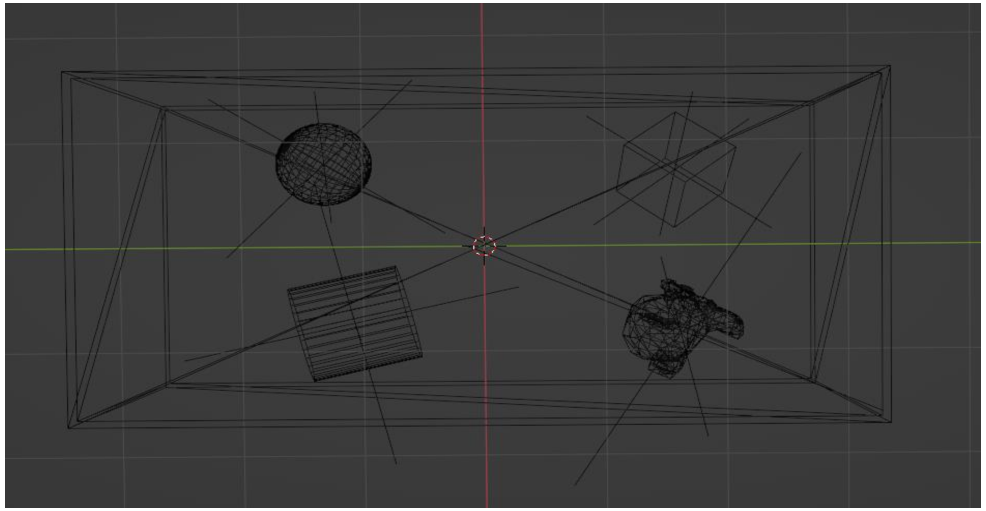


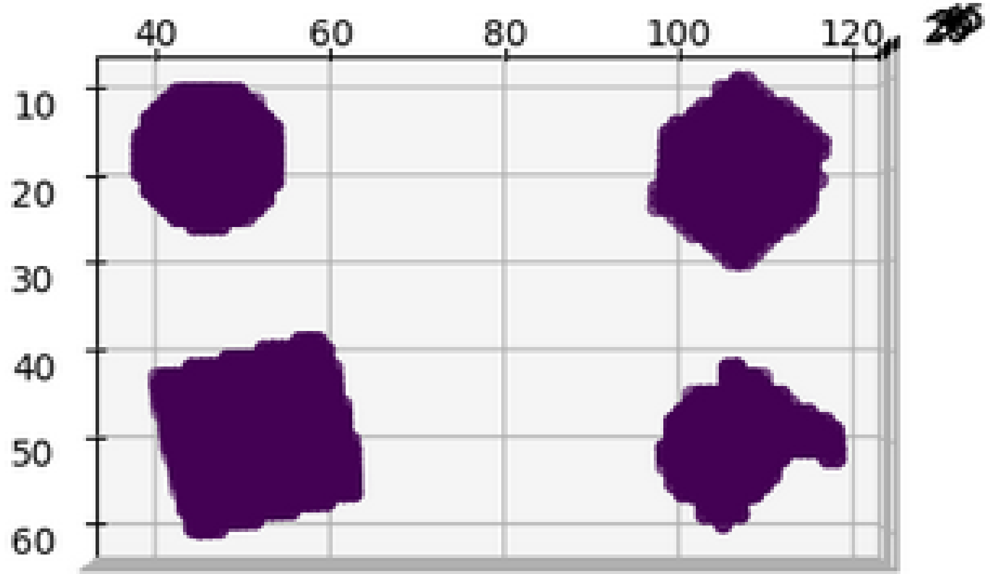
HiRo: The Hidden Room dataset

HiRo: A highly detailed 3D benchmark dataset for cosmic ray tomography

- **First iteration** contains >**500 scenes with primitive shapes** of different materials concealed in a hidden chamber of arbitrary composition
- **B2G4 improved to automate a probability-based randomization of objects**, positions, and materials, while keeping detailed information and object material properties
- **Ground truth generated at randomization time:** Materials ID and object specific metadata are generated as ground truth reference for machine learning applications

A. Bueno Rodriguez, F. Sattler, M. Perez Prada, M. Stephan, and S. Barnes. Scalable 3D data generation for machine learning applications in muon tomography (in preparation)





Using pillars to navigate latent spaces

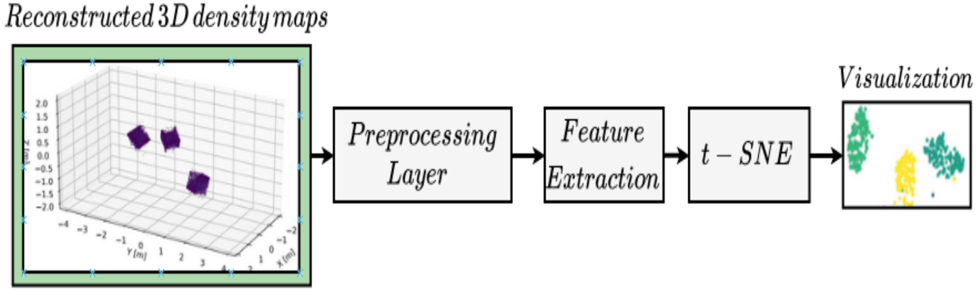
- **Objective:** To identify materials from reconstructed 3D structures through the use of embedding spaces.

    - To this end, we use ***t-Stochastic Neighbor Embedding (t-SNE)***: Dimensionality reduction technique that produces a low dimensional projection of the most important components of high dimensional data **(Hinton&Maaten, 2012)**

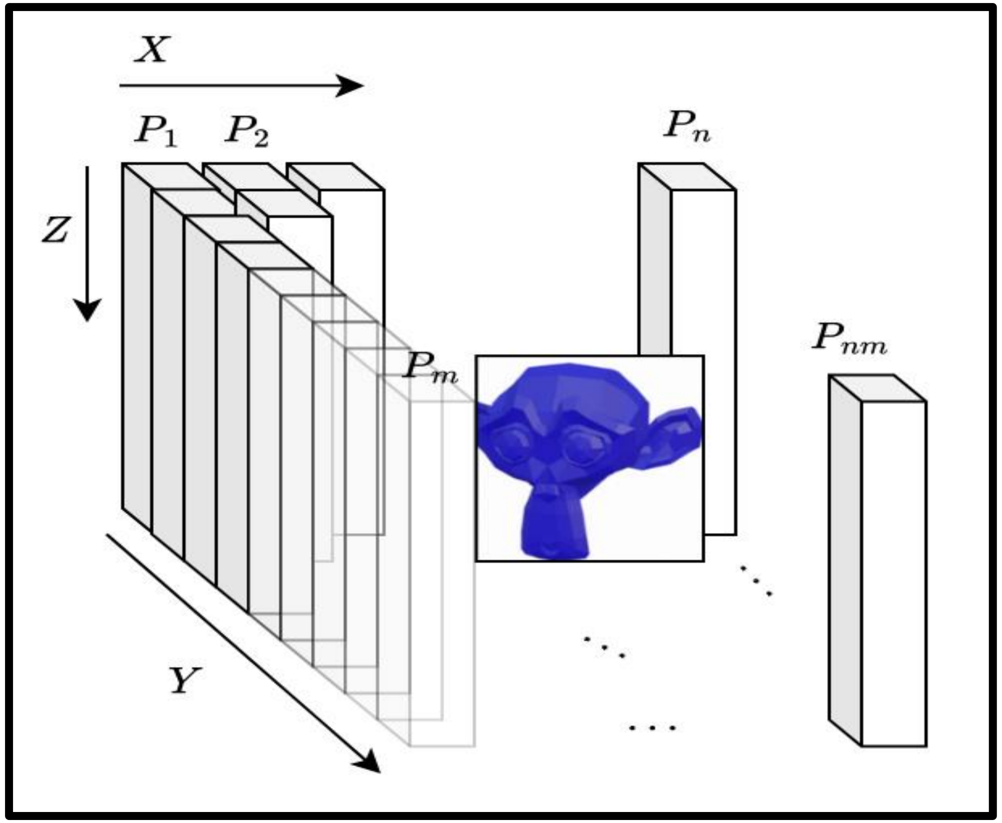
- **Methodology**: A feature-based approach relying on volumetric sampling, so-called **"pillars,"** which includes the calculation of 12 statistical measures, including averaging energy, mean, maximum, and, standard deviation.

A. Bueno Rodriguez, F. Sattler, M. Perez Prada, M. Stephan, and S. Barnes. Advanced Material Identification in Muon Tomography, MARESEC 2024. Muography track (submitted)

A

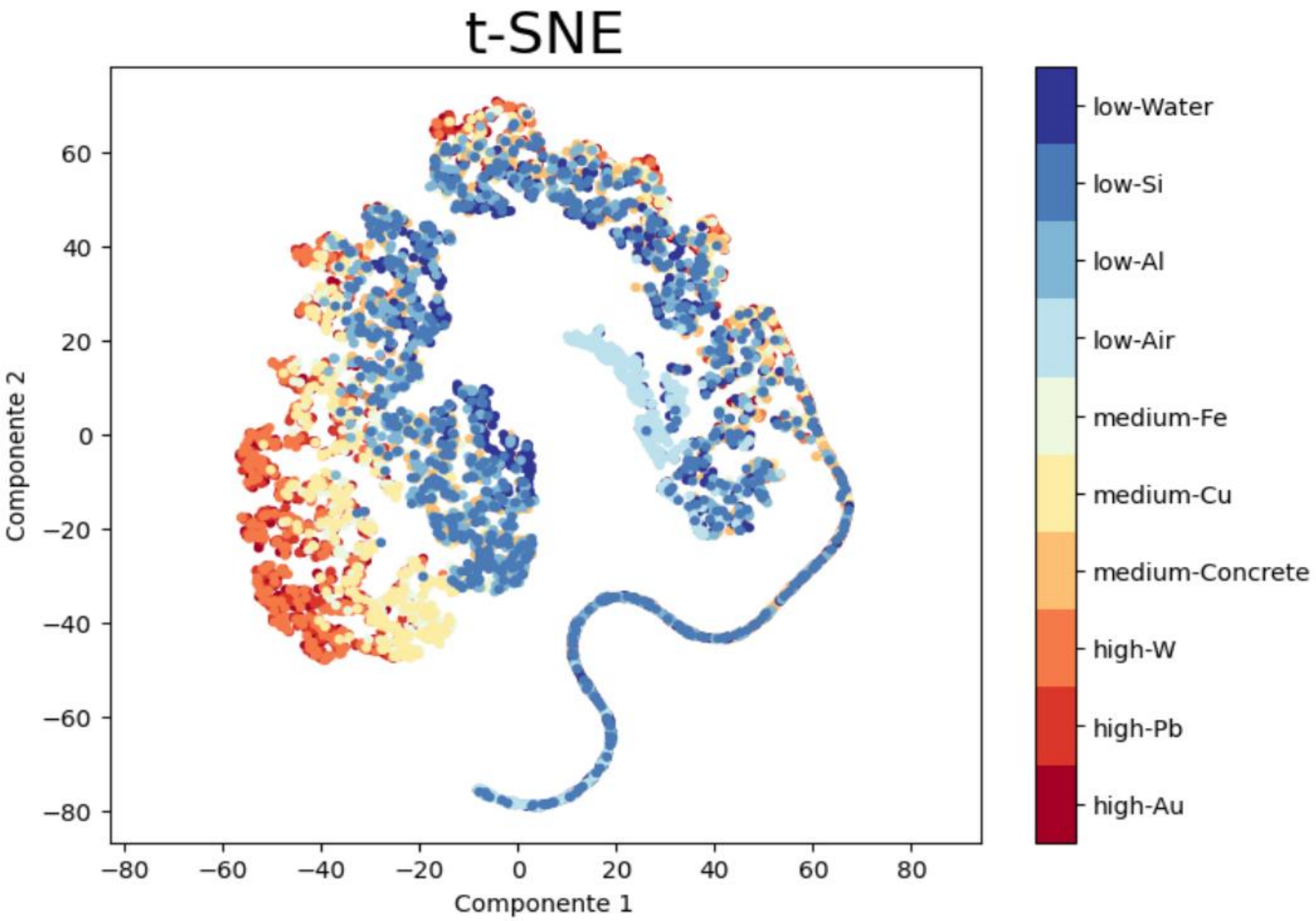


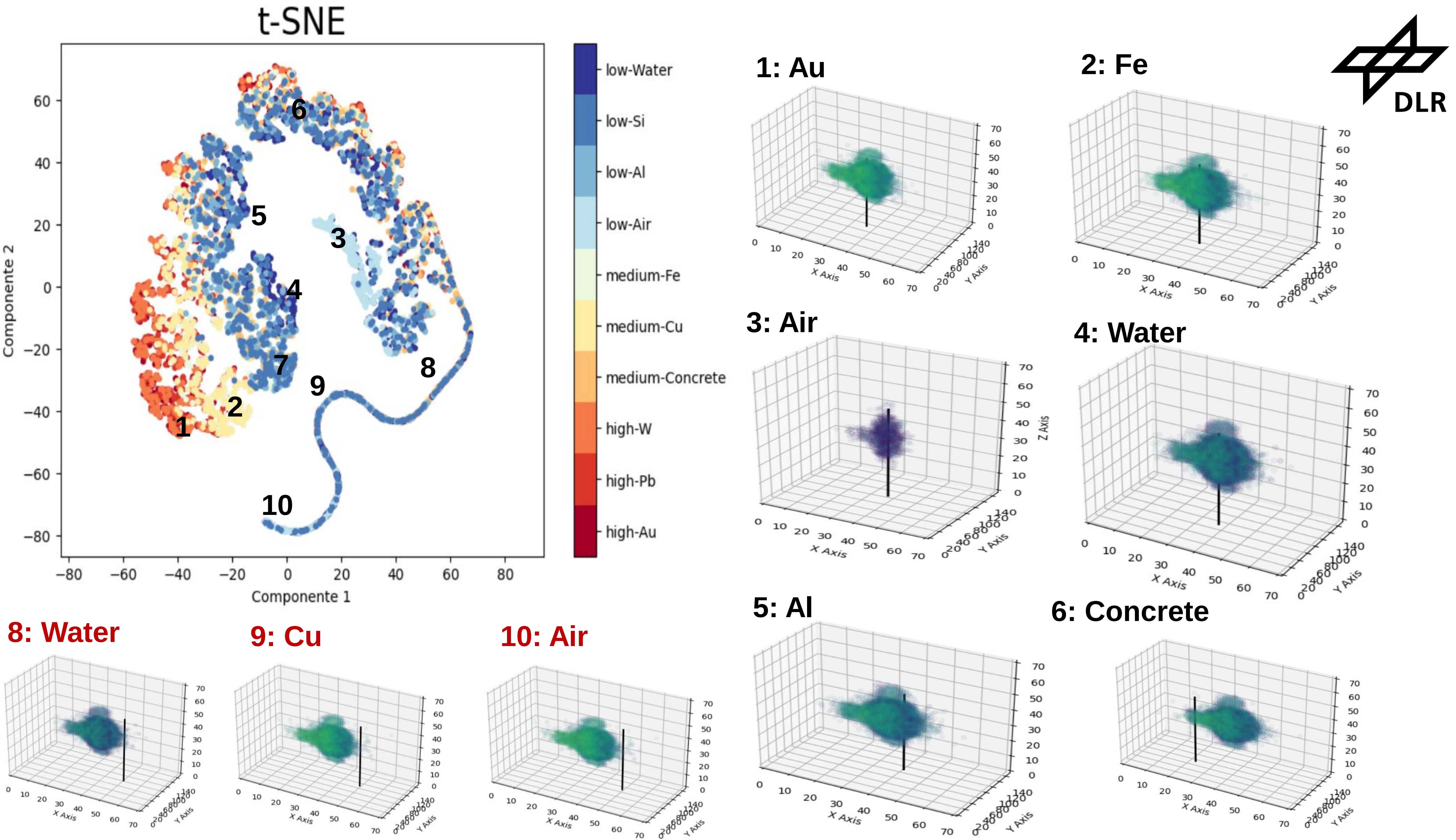
B



Experiment 1:

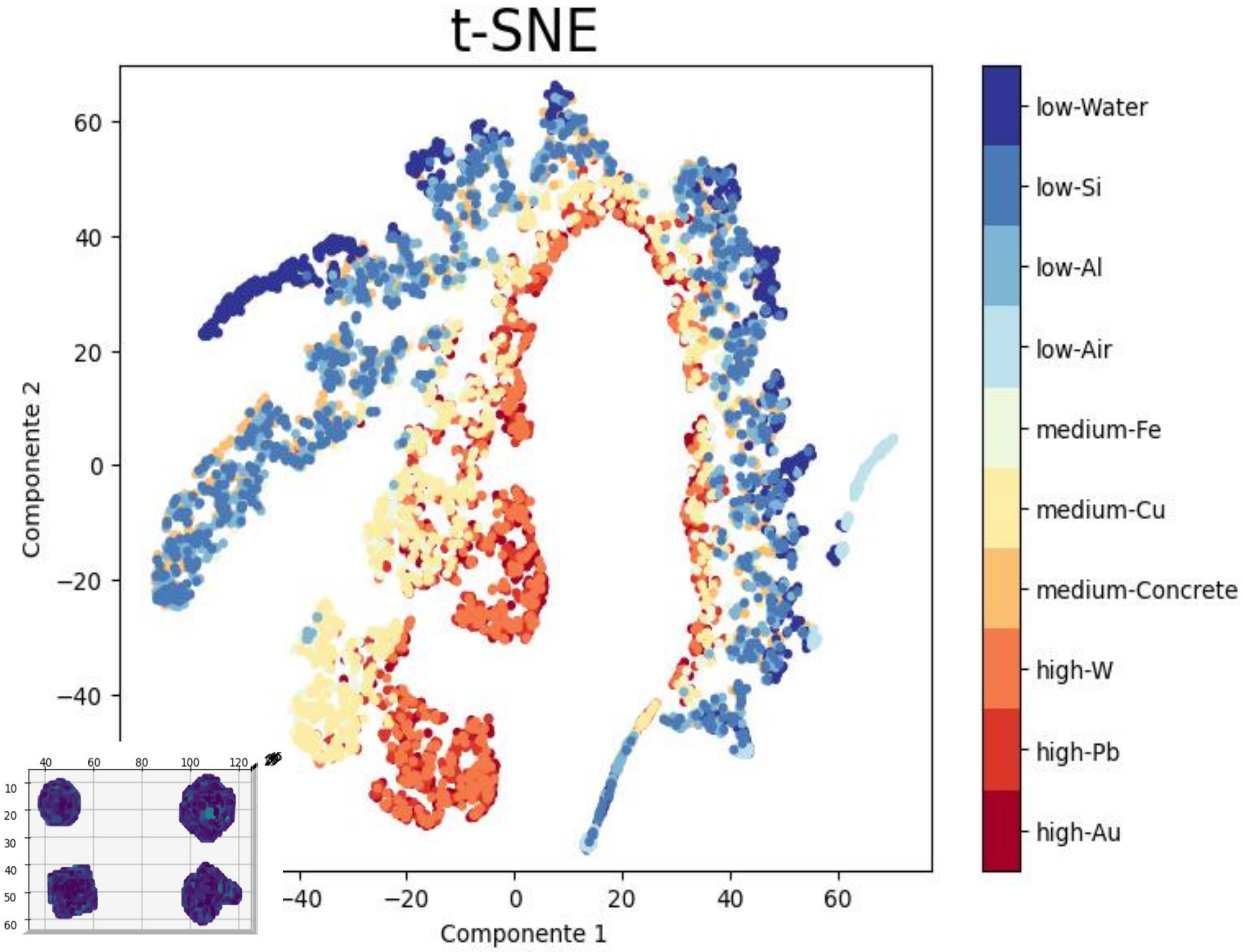
Monkey head, 10 materials, same location, hidden room, no preprocessing




t-SNE
Componente 1
Componente 2
low-Water
low-Si
low-Al
low-Air
medium-Fe
medium-Cu
medium-Concrete
high-W
high-Pb
high-Au
1: Au
2: Fe
3: Air
4: Water
5: Al
6: Concrete
8: Water
9: Cu
10: Air
X Axis
Y Axis
Z Axis
DLR

Experiment 2:

4 shapes scene (test scene), 10 materials. Pre-processed pillars with convolutional filters.



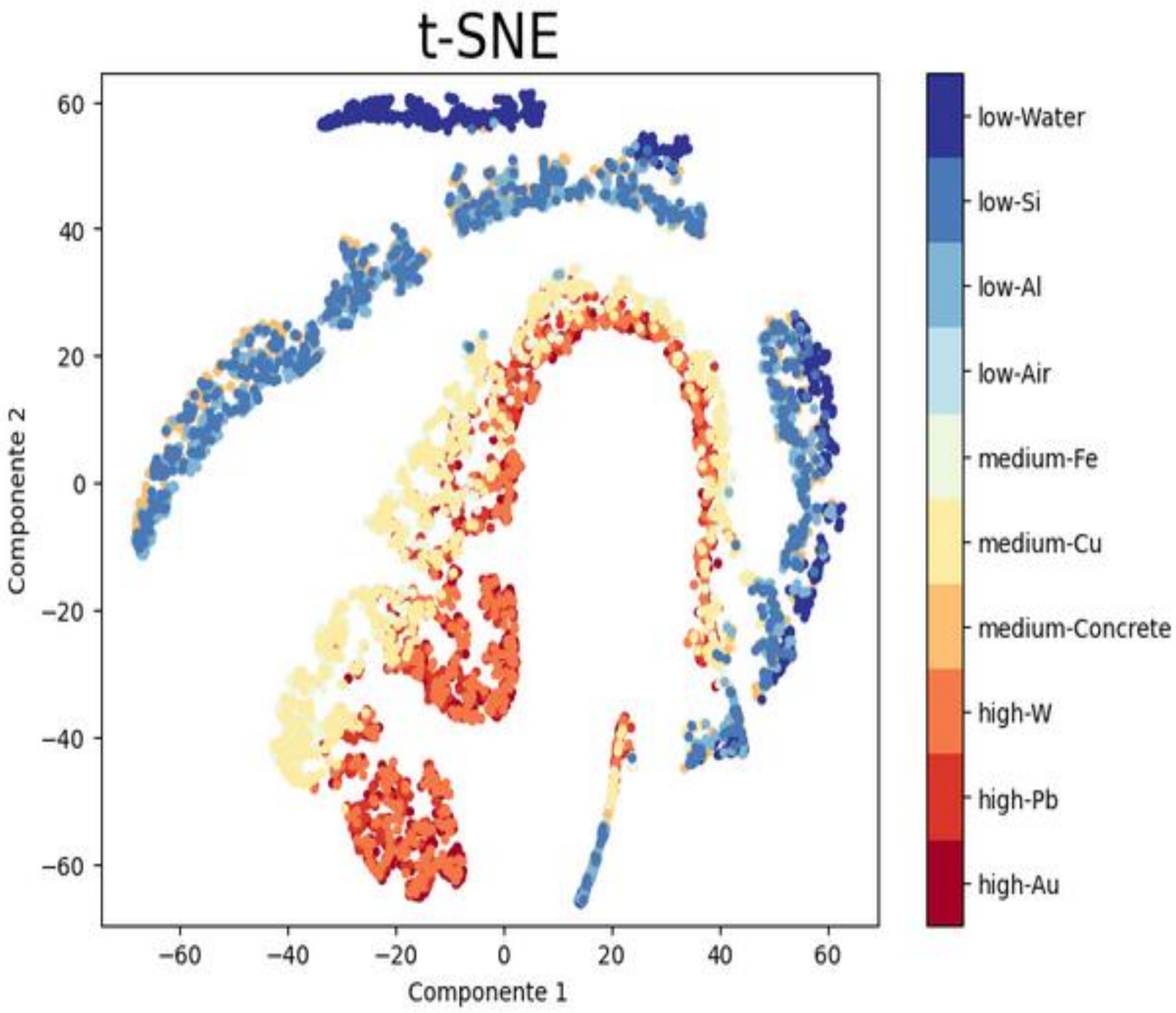
Sampling matters: embeddings are conditioned by the spatial distribution of scattering angles in different regions of interest.

Angles can be similar between different materials, but still show some separation in feature space… why?

Could this be related to ambiguities in the reconstruction?

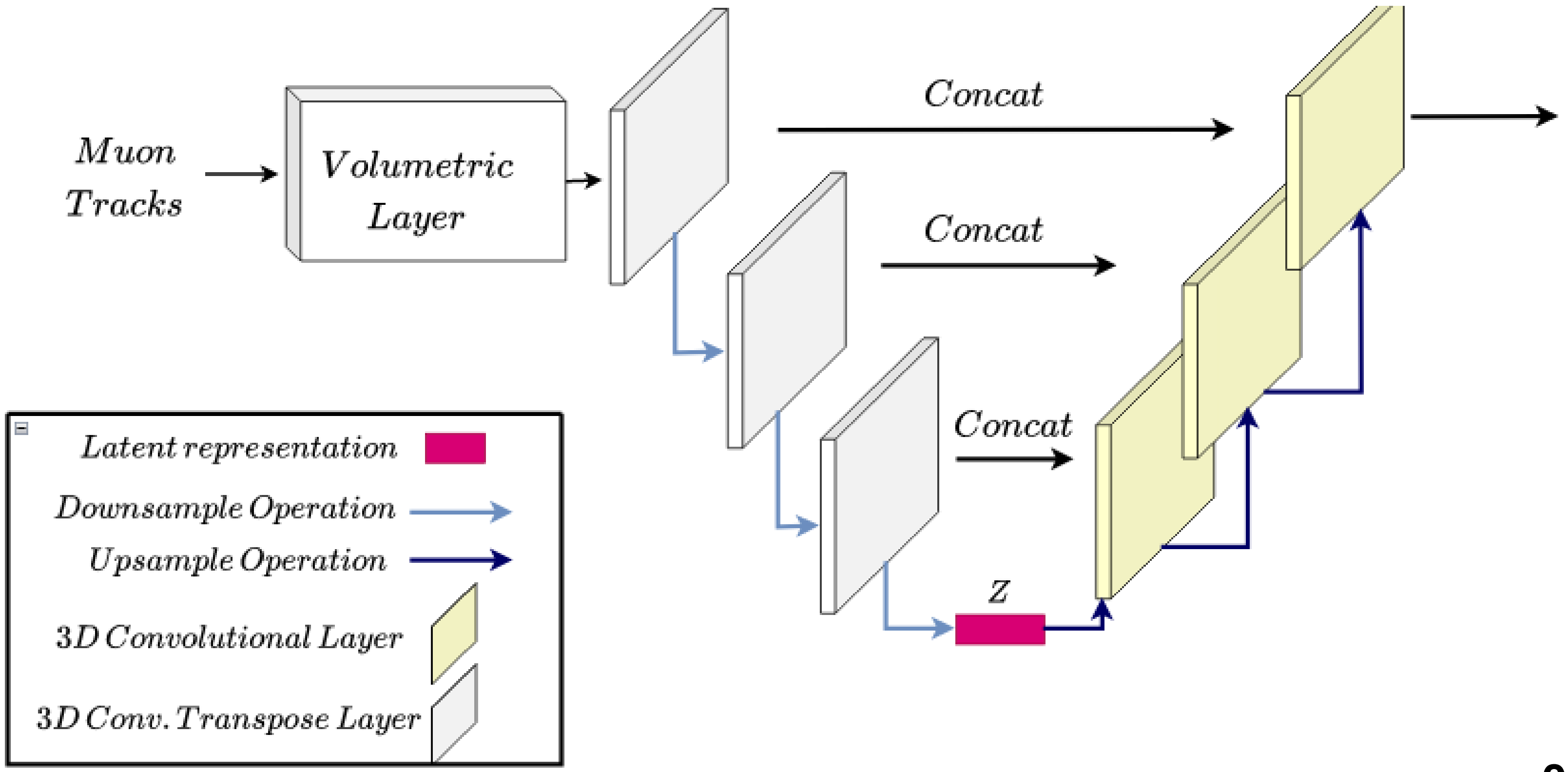
Experiment 2:

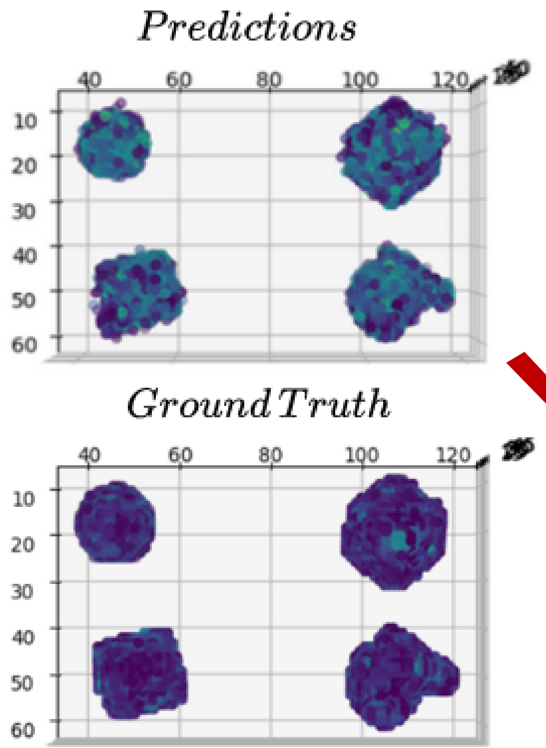
4 shapes (test scene), 10 materials. Pre-processed pillars with convolutional filters Considering energy from 2 to 10 GeV



- **Processing the scene is essential for ML**: a density gradient is clearly represented
- **When processed**, basic reconstruction methods (e.g: PoCA) capture relevant information and could be used as input for ML algorithms.
- **HOWEVER:**
  - **Shape and sampling strategy influences the representation**, but incorporating 3D features (convolutions) improves results
  - Material identification would benefit by incorporating **momentum and/or secondary particles information**

Experiment 3: Custom 3D U-NET using HiRo





WIP

0.70 seconds per prediction

Custom U-Net Architecture (Very shallow) Full coverage of the entire volumetric scene: the full hidden room at 5 cm voxel resolution, input size of (70, 150, 70) trained via MSE to predict denoised angles.

The designed U-net understands the spatial relationships and can reconstruct structural variations through the volume. Yet, **it is not perfect and not all shapes are faithfully predicted (see cylinder)**

Thanks !

Contact: angel.bueno@dlr.de