ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

Super-Resolution Surrogate Model for Accelerated Geant4 Simulations

Gallo G.[*,1], Cirrone G.A.P.[2,4], Fattori S.[2], Ientile V.[3], Sciuto A.[2], Tricomi A.[1,3,4]

(1) Department of Physics and Astronomy "E. Majorana", University of Catania, Catania, Italy
(2) National Institute for Nuclear Physics, Southern National Lab, Catania, Italy
(3) National Institute for Nuclear Physics, Catania Division, Catania, Italy
(4) Sicilian Center of Nuclear Physics and Physics of Matter (CSFNSM), Catania, Italy
(*) Speaker

ICSC and Spoke2 – Where Are We Now?,
Catania December 10-12, 2024

ICSC

Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

SPOKE 2 - FUNDAMENTAL RESEARCH & SPACE ECONOMY

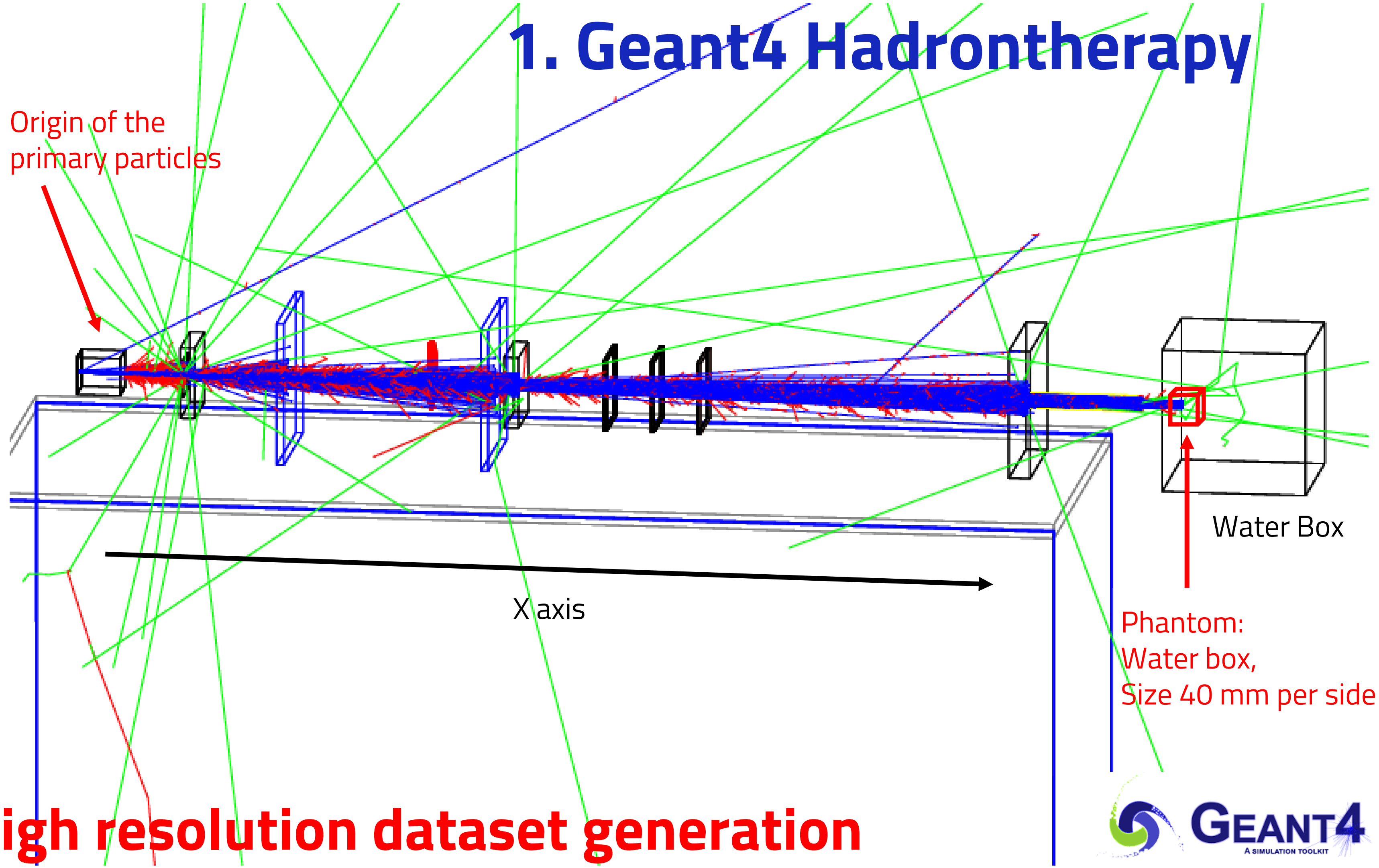
- **WP 6**
- **UC 2.6.2 – ML_GEANT4_WP6**

Enhancing Geant4 Monte Carlo Simulations through Machine Learning Integration

1. Geant4 Hadrontherapy

- Primary particles are protons of 63.5 MeV at the origin (62 MeV at phantom entrance)
- The phantom is subdivided in N slabs or "voxels" along the X axis
- The result of the simulation are:

- [ ] Dose deposit in the phantom
- [ ] Particle fluence
- [ ] LET of primary particle and secondary ions



>>> High quality, high resolution dataset generation

GEANT4 A SIMULATION TOOLKIT

https://geant4.web.cern.ch/docs/advanced_examples_doc/example_hadrontherapy

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

2. Super-Resolution in Machine Learning

Goal:

- Predicting results that are **comparable** to high-resolution simulations

- **Challenge**: reduce computation cost by working with a lower voxel density

Approach:

- **Input**: low-resolution data (lower density voxel)

- **Output**: improved simulations at higher resolution, predicted by the model.

LR

SR (x4)

3. Datasets production for ML training

Generating **high-density** and low-density datasets for studying and validating linear energy transfer (LET) calculation.

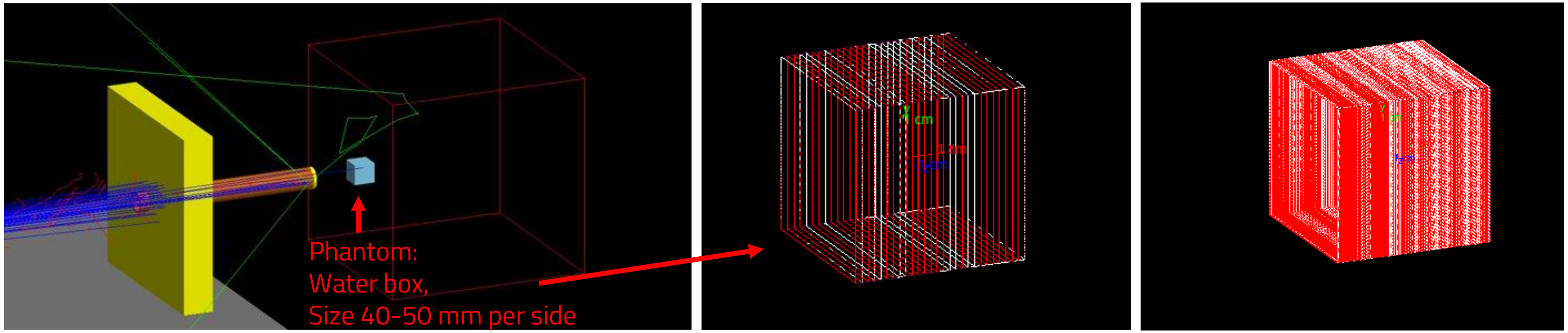
- How deep can we go with voxel sizes in Geant4?
- Up to which voxel size does the Geant4 code generate a physically correct simulation?
- How many primary particles do we need to run to get a reliable data set without blowing up the computational cost?

Range Cut (~ threshold on particle production)

- ❑ 1mm
- ❑ 100 µm
- ❑ 50 µm
- ❑ 10 µm
- ❑ 1 µm
- ❑ 0.5 µm

Voxel Size:

- ➢ 1 mm
- ➢ 100 µm
- ➢ 10 µm
- ➢ 1 µm



4. Execution time

Cut (u)	Voxelsize (u)	Exec Time	CPU (%)	MEM (%)
1000	100	00h.00m.54s	78.96	0.11
100	100	00h.01m.08s	80.97	0.10
50	100	00h.01m.37s	84.47	0.12
10	100	00h.04m.25s	88.46	0.18
1	100	00h.33m.15s	92.03	0.20
0.5	100	00h.33m.20s	92.32	0.20
1000	10	00h.07m.27s	89.84	0.18
100	10	00h.08m.02s	89.81	0.19
50	10	00h.09m.08s	88.72	0.19
10	10	00h.18m.13s	90.50	0.20
1	10	00h.55m.52s	93.28	0.30
0.5	10	00h.56m.01s	92.78	0.30
1000	1	01h.17m.07s	90.07	0.20
100	1	01h.19m.27s	89.92	0.20
50	1	01h.24m.57s	91.63	0.22
10	1	01h.48m.07s	92.10	0.30
1	1	03h.14m.48s	93.31	0.50
0.5	1	03h.15m.00s	93.55	0.50

N particles = 1 x 10^6

- AMD EPYC 7552 96 Core
- 512 GB RAM

N=96 Threads

INFN CATANIA
Istituto Nazionale di Fisica Nucleare
Sezione di Catania

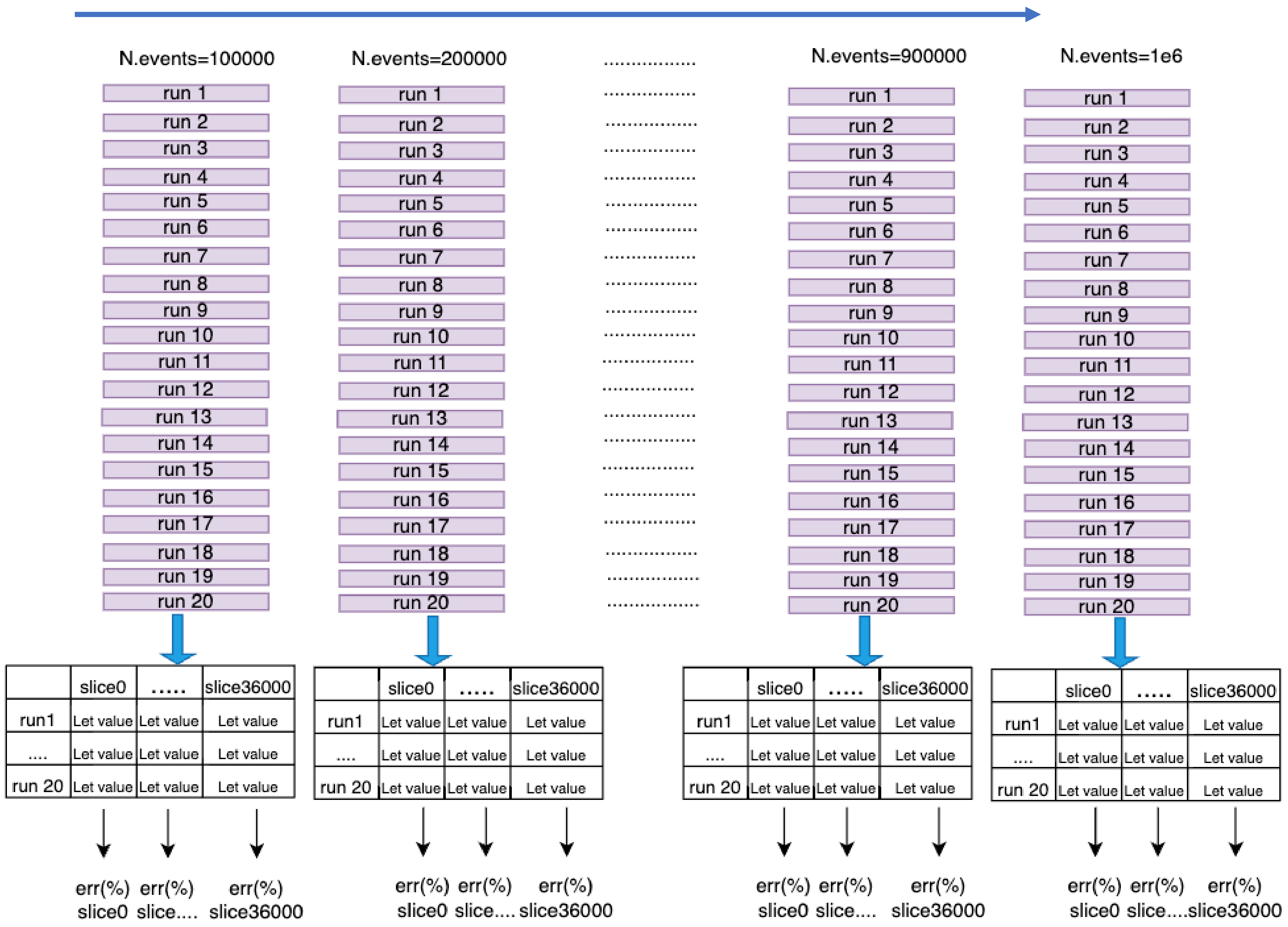
Range Cut (~ threshold on particle production)

- ❑ 1mm
- ❑ 100 µm
- ❑ 50 µm
- ❑ 10 µm
- ❑ 1 µm
- ❑ 0.5 µm

Voxel Size:

- ➢ 100 µm
- ➢ 10 µm
- ➢ 1 µm

5. Datasets validation



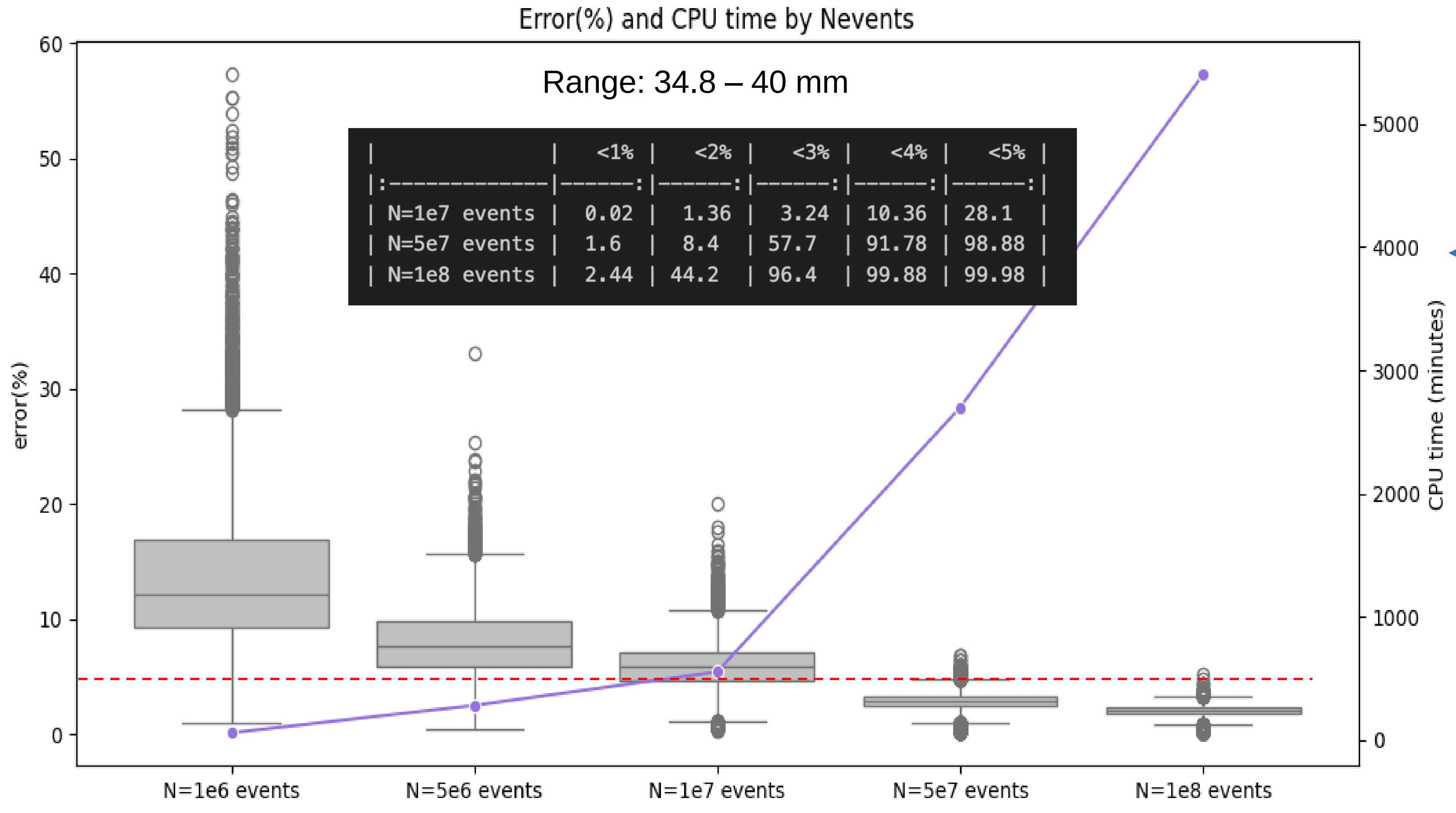
How many particles for a reliable dataset?

- Method definition: N events *vs* err(%) per slice

INFN CATANIA
Istituto Nazionale di Fisica Nucleare
Sezione di Catania

5. Datasets validation

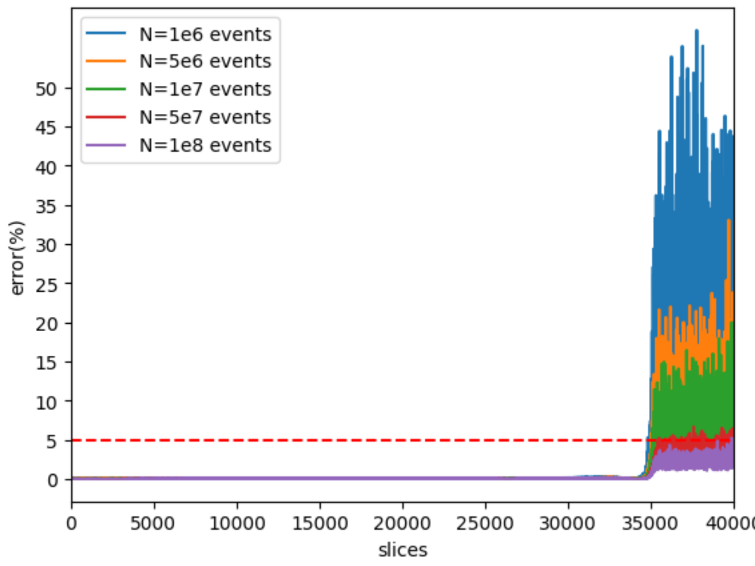
- AMD EPYC 7552 96 Core
- 512 GB RAM
- N=96 Threads



	<1%	<2%	<3%	<4%	<5%
N=1e7 events	0.02	1.36	3.24	10.36	28.1
N=5e7 events	1.6	8.4	57.7	91.78	98.88
N=1e8 events	2.44	44.2	96.4	99.88	99.98

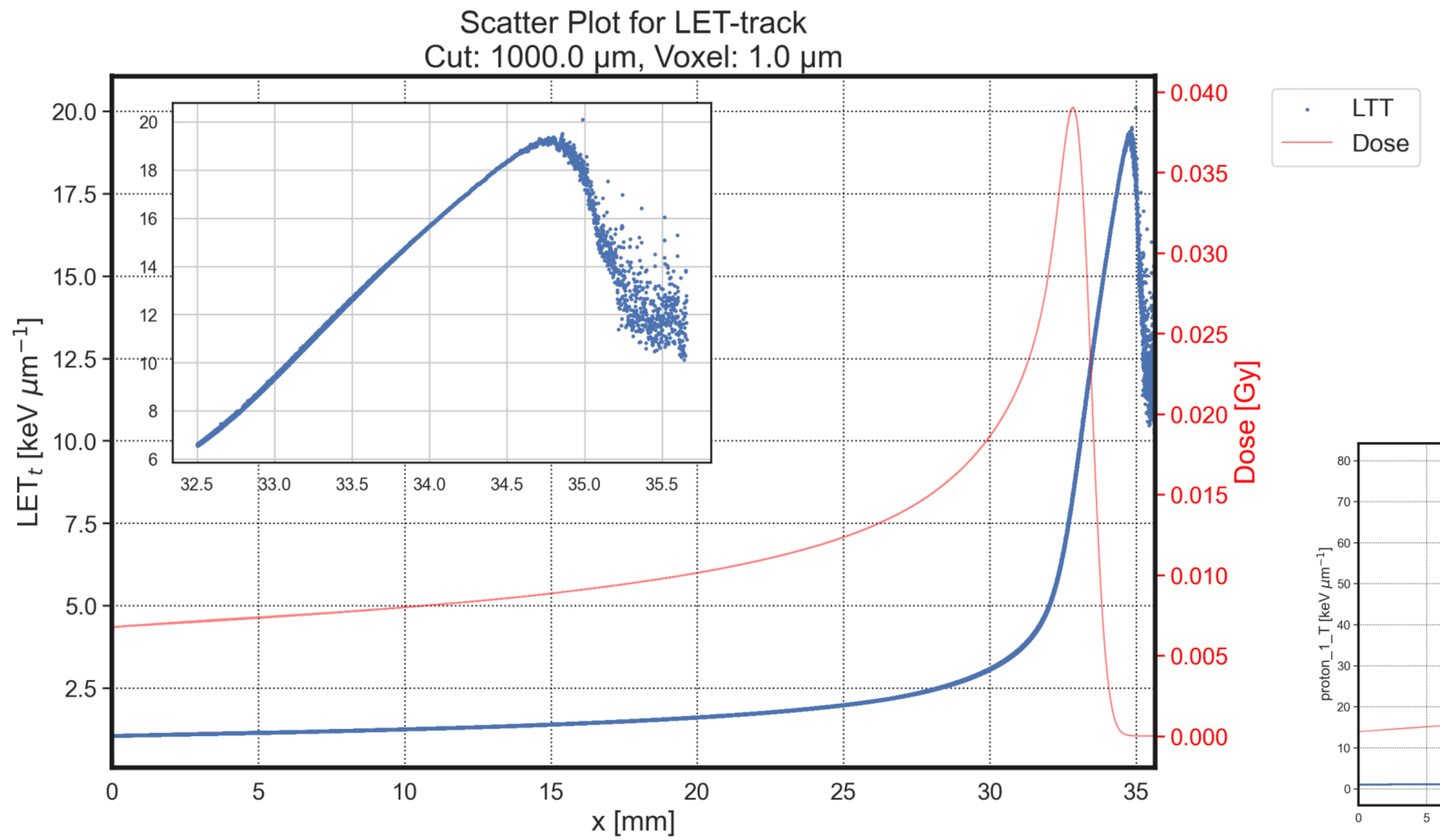
- The data considered for the box plots are err% (calculated on the 20 runs where the number of particles was fixed) for each voxel in the range indicated.

Voxel size = 1um, cut = 1 mm.



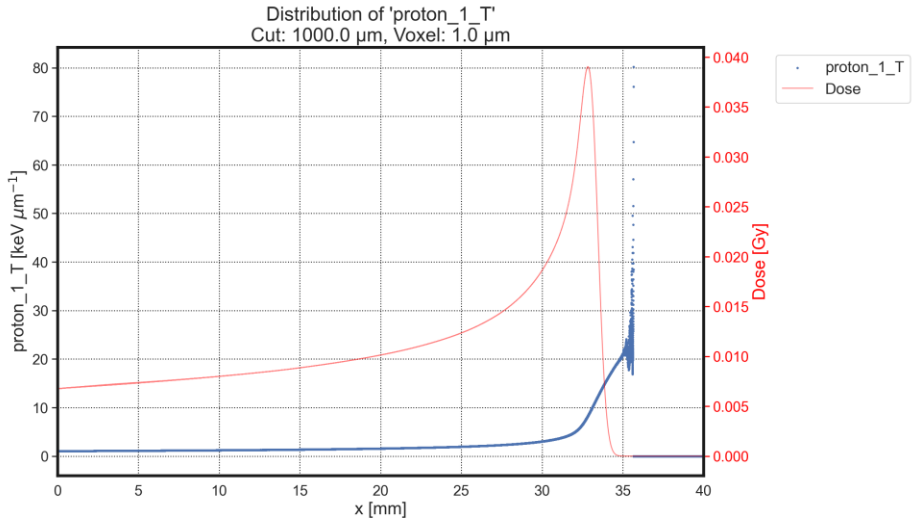
5. Dataset validation

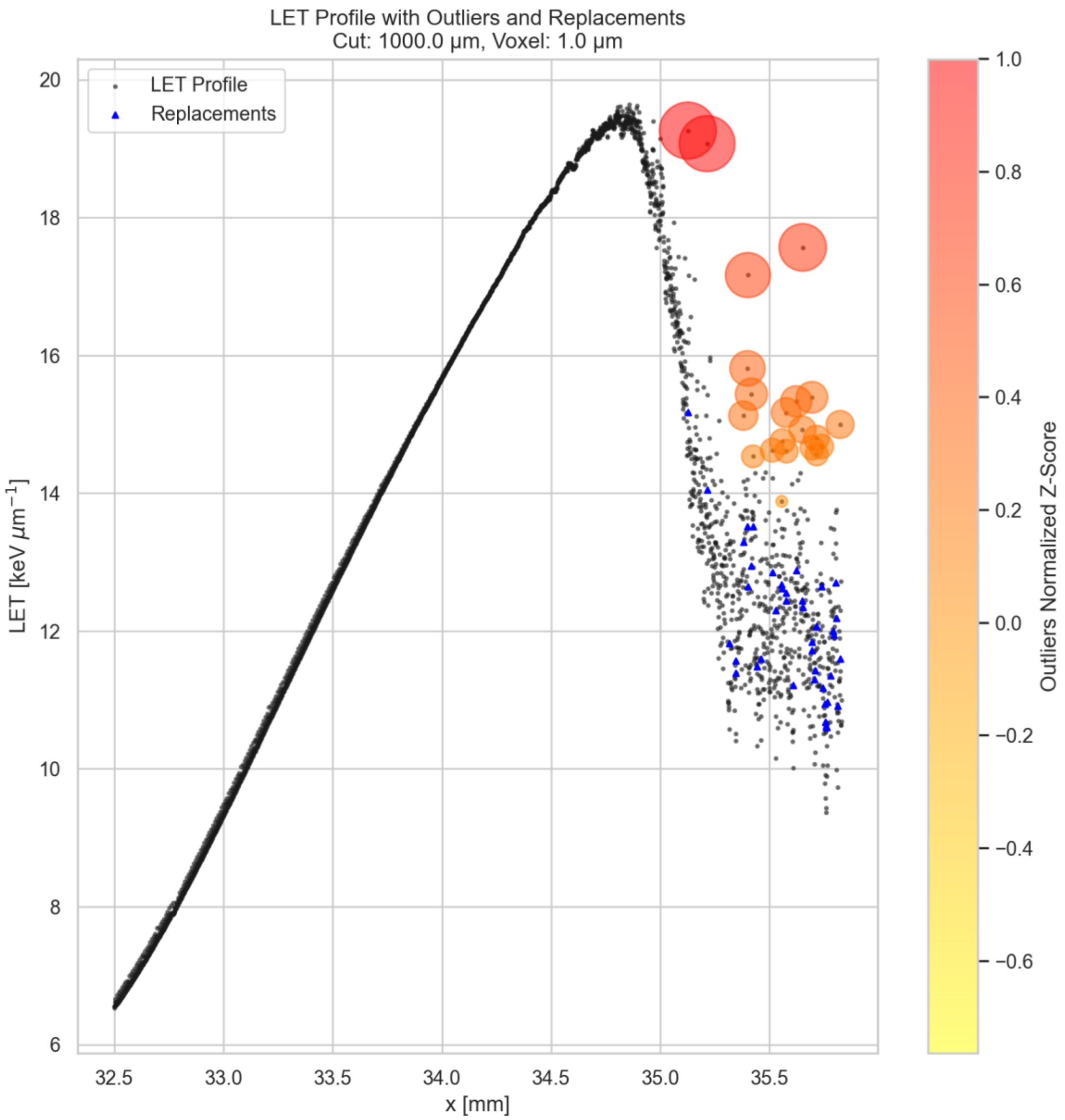
1e8 primary particles;
Ep = 62 MeV at phantom entrance



Dataset early stop:

- To limit the quality degradation of the LET profile, the data set is truncated corresponding to where the primary particle beam stops (~35.66 mm in water).





6. Data Exploration

Dataset

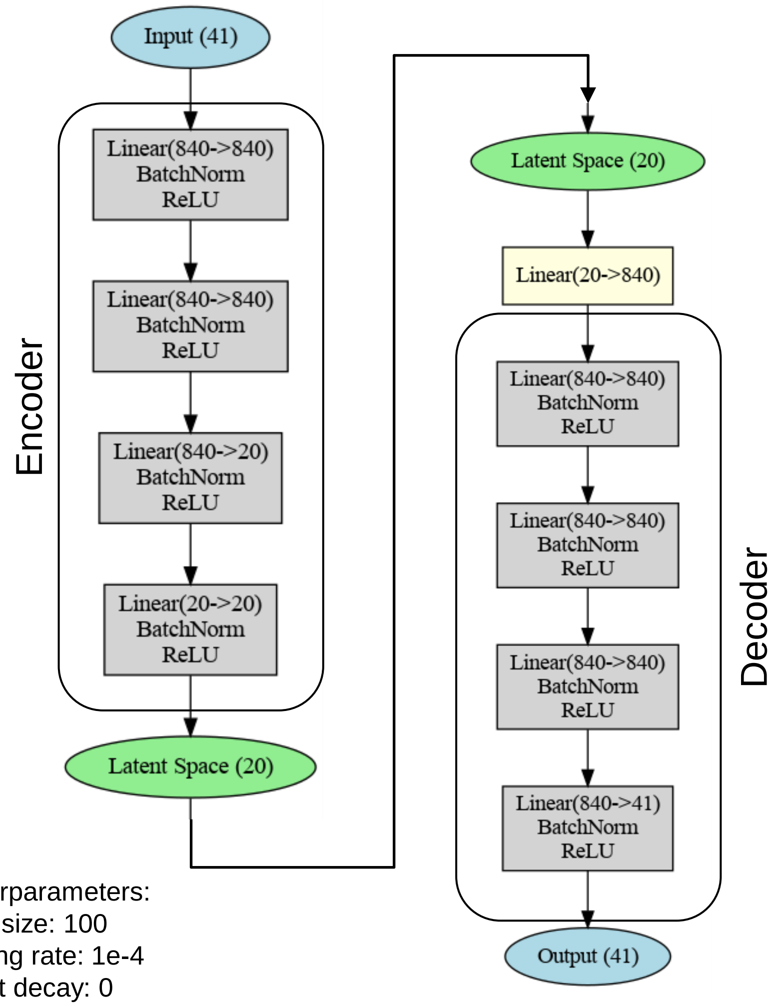
- runbeam 1 x 10^8,
- cut 1000 um,
- voxel 1 um

Dataset cleanup:

Several combinations of first outlier detection and value replacement algorithms have been tested, including a denoising autoencoder.
The best results were obtained with:

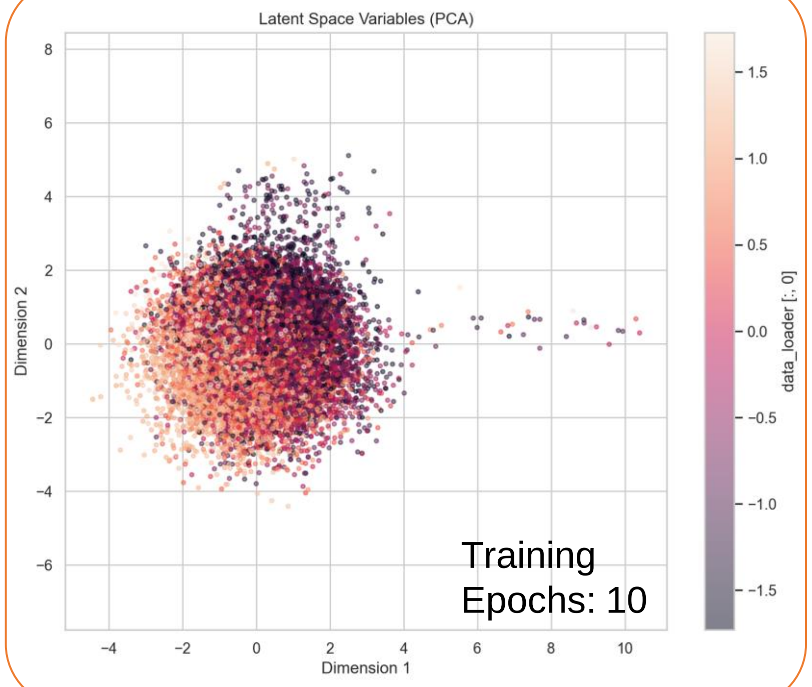
- outliers_method = 'DBSCAN',
- replace_method = 'kneighbors_regressor'

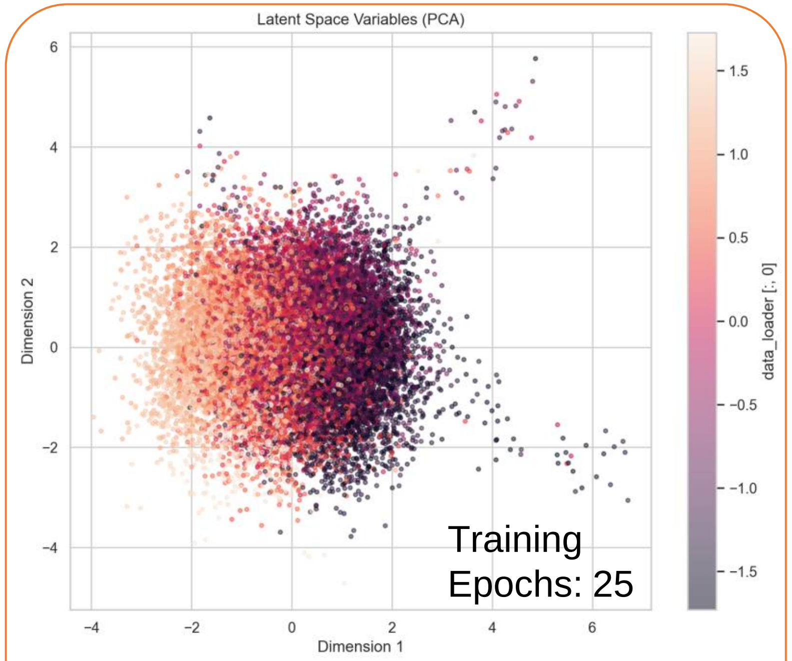
7. Super-Resolution model = Variational Autoencoder

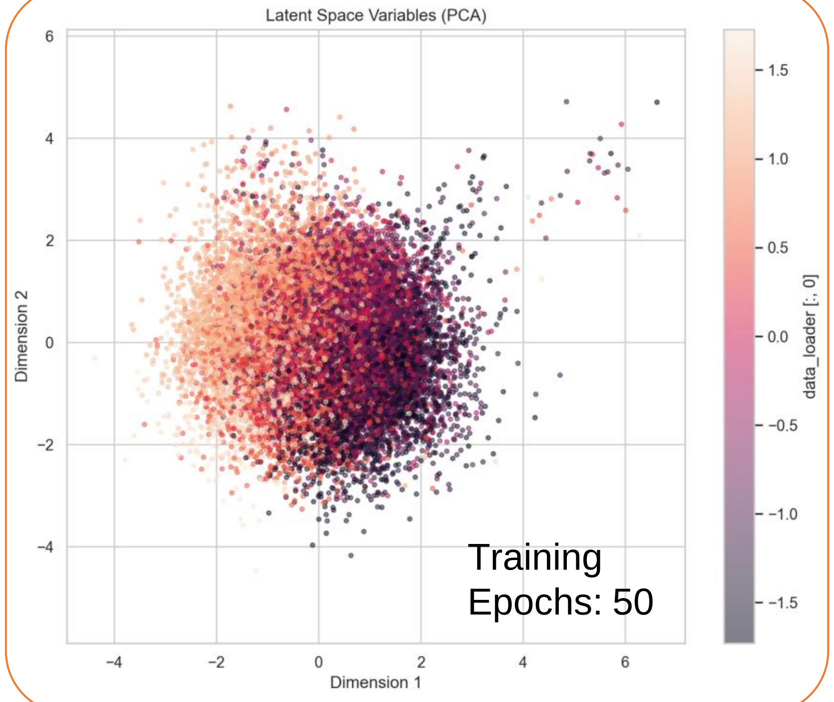


Other hyperparameters:
- batch size: 100
- learning rate: 1e-4
- weight decay: 0

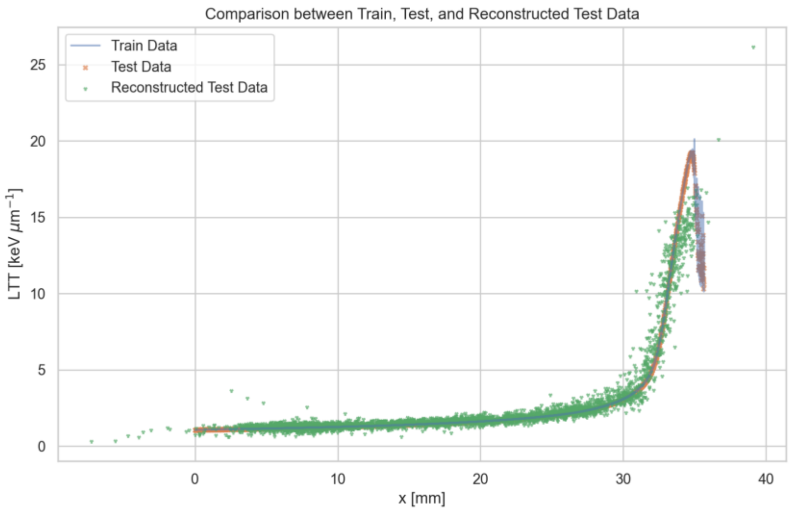
7. Super-Resolution model: Latent Space Evolution



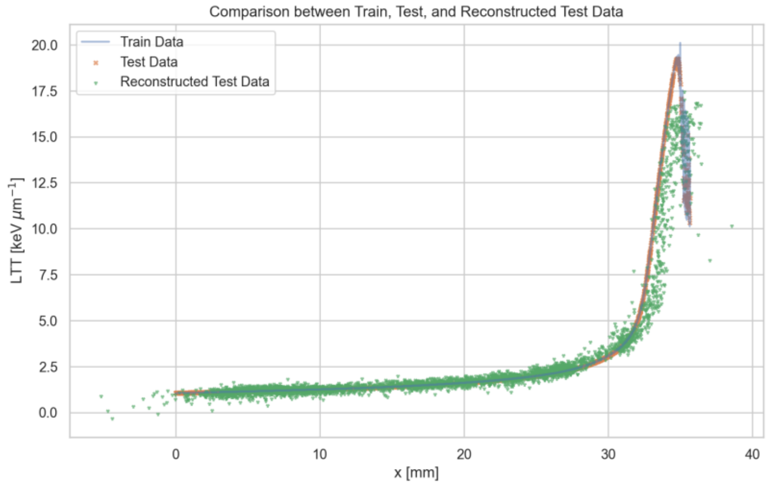




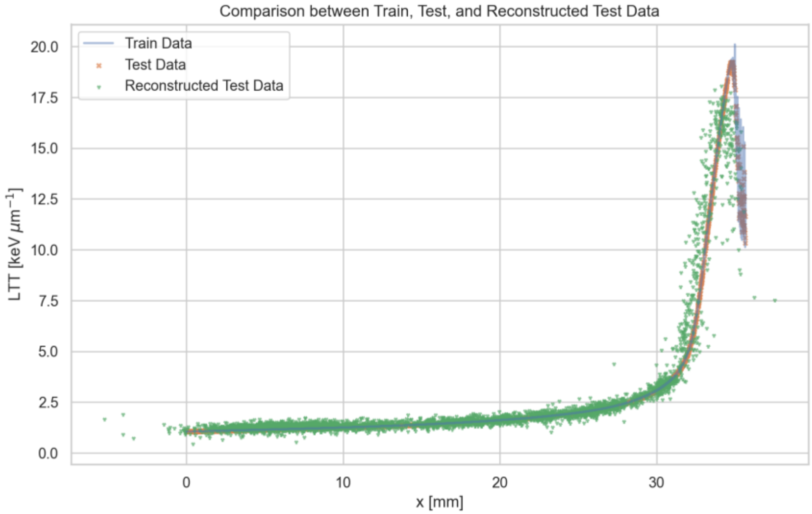
7. Super-Resolution model: Test vs Reconstructed Data



Training Epochs: 10



Training Epochs: 25



Training Epochs: 50

8. Where are we now?

KPI ID	Description	Acceptance threshold
KPI2.6.1.1	Publications	1
KPI2.6.1.2	Presentation at conferences	1 ☑
KPI2.6.1.3	Publicly available Code repository	1 ☑
KPI2.6.1.4	Use case Test Datasets defined	1 ☑
KPI2.6.1.5	Geant4 Algorithms to be used as targets for a ML optimization	1 ☑
KPI2.6.1.6	Efficiency Gain on the same hardware: The improved simulation, when run on the same hardware as the standard simulation, should achieve at least a 20% reduction in the time taken to generate predictions.	20% in time reduction, with acceptable physics performance

Achievements:

☑ Talk delivered at Congress of the Italian Physical Society (Bologna, September 9-13, 2024)

☑ 1st release of a public accessible code repository

☑ Testing and validation on chosen dataset

☑ LET calculation algorithm is the primary target for optimization by ML model for super-resolution task

8. Where are we now?

KPI ID	Description	Acceptance threshold
KPI2.6.1.1	Publications	1
KPI2.6.1.2	Presentation at conferences	1 ☑
KPI2.6.1.3	Publicly available Code repository	1 ☑
KPI2.6.1.4	Use case Test Datasets defined	1 ☑
KPI2.6.1.5	Geant4 Algorithms to be used as targets for a ML optimization	1 ☑
KPI2.6.1.6	Efficiency Gain on the same hardware: The improved simulation, when run on the same hardware as the standard simulation, should achieve at least a 20% reduction in the time taken to generate predictions.	20% in time reduction, with acceptable physics performance

Next Steps:

- ❑ Publish a paper in a peer-reviewed journal as soon as the ML model produces satisfactory results
- ❑ Create a pipeline to directly interface the ML model with the MC simulation and estimate the efficiency gain
- ❑ Test the ML model on the entire dataset, including the tail after the Bragg Peak
- ❑ Test of additional model strategies (stacked or chained AE, GAN…)

ICSC
Centro Nazionale di Ricerca in HPC,
Big Data and Quantum Computing

Università di Catania

INFN CATANIA
Istituto Nazionale di Fisica Nucleare
Sezione di Catania

www.supercomputing-icsc.it

Acknowledgements

"This work is supported by ICSC – Centro Nazionale di Ricerca in High Performance Computing, Big Data and Quantum Computing, funded by European Union – NextGenerationEU".

Thank you for the attention!

Resources Required

LEONARDO BOOSTER-GPU	
Minimum Number of GPU hours	3000
Optimal Number of GPU hours	6000
Maximum number of usable GPU	2 (via tensorflow/ pytorch)
Total RAM	512 GB

CLOUD		
Data storage	5 TB	
vCPU	Number	24
	Time (Hours)	1600
	Optimal number of core	90
	Total RAM	512 GB

4. Execution time

Cut(um)	Voxelsize(um)	Exec Time	CPU(%)	MEM(%)
1000	100	00h.00m.11s	16.72	0.08
100	100	00h.00m.12s	18.39	0.08
50	100	00h.00m.15s	20.48	0.09
10	100	00h.00m.26s	26.96	0.09
1	100	00h.02m.33s	31.86	0.1
0.5	100	00h.02m.36s	31.83	0.1
1000	10	00h.00m.43s	28.99	0.1
100	10	00h.00m.45s	29.27	0.1
50	10	00h.00m.51s	29.42	0.1
10	10	00h.01m.28s	30.98	0.1
1	10	00h.04m.12s	32.23	0.1
0.5	10	00h.04m.10s	32.26	0.1
1000	1	00h.06m.14s	31.72	0.1
100	1	00h.06m.25s	31.7	0.1
50	1	00h.07m.10s	31.62	0.1
10	1	00h.08m.47s	31.85	0.1
1	1	00h.15m.04s	32.3	0.2
0.5	1	00h.15m.00s	32.25	0.2

N particles = 5 x 10^5

- AMD EPYC 7552 96 Core
- 512 GB RAM

N=32 Threads

INFN CATANIA
Istituto Nazionale di Fisica Nucleare
Sezione di Catania

Range Cut (~ threshold on particle production)

- ❑ 1mm
- ❑ 100 µm
- ❑ 50 µm
- ❑ 10 µm
- ❑ 1 µm
- ❑ 0.5 µm

Voxel Size:

- ➢ 100 µm
- ➢ 10 µm
- ➢ 1 µm