interTwin

CERN – Fast Detector Simulation w/ GAN

CERN openlab summer student lecture, 7 August 2023

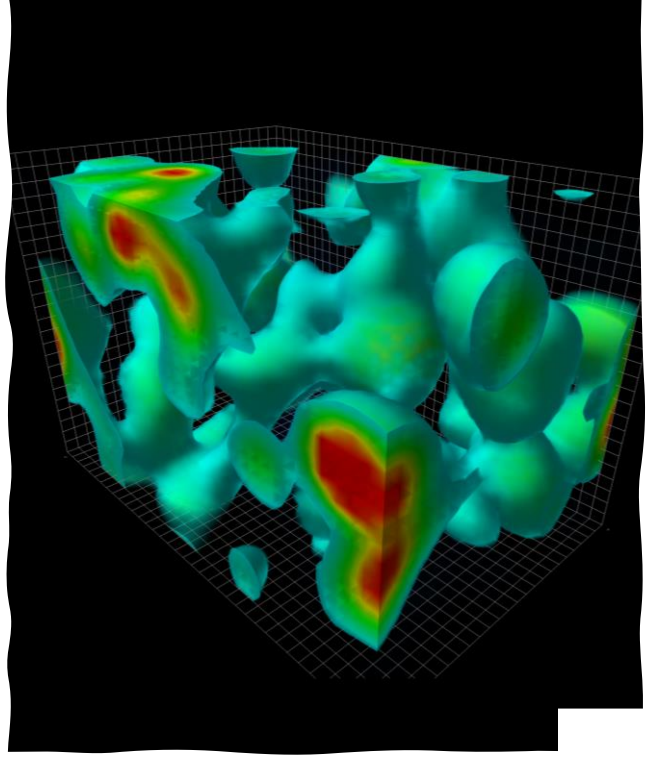
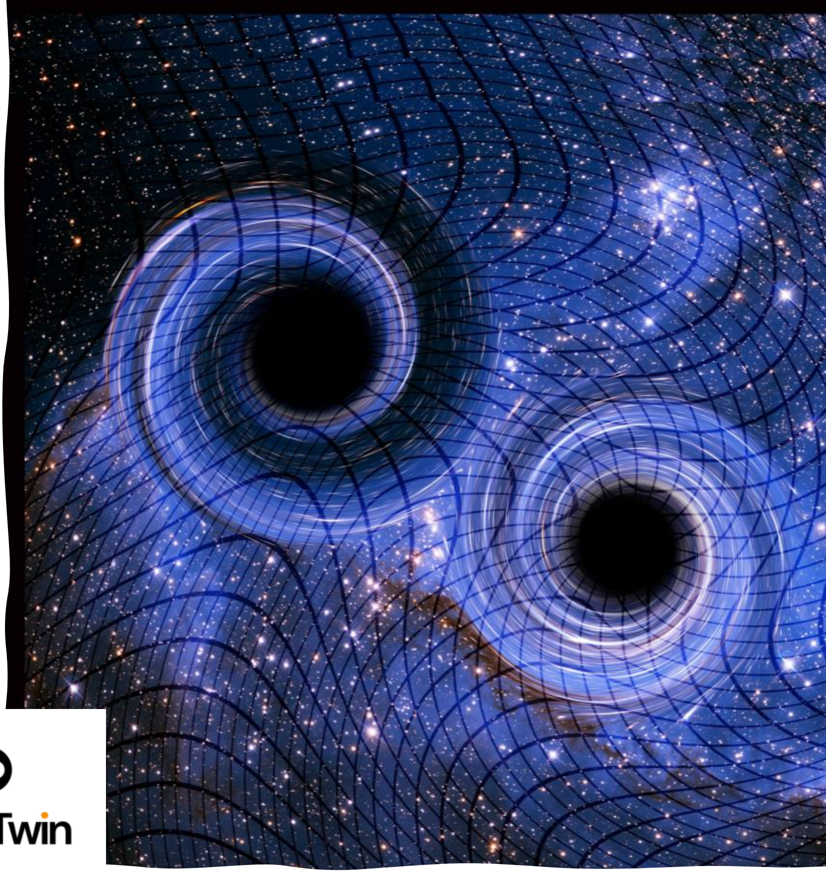
Kalliopi Tsolaki, Sofia Vallecorsa, CERN-IT
David Rousseau, CNRS/IN2P3-IJCLab

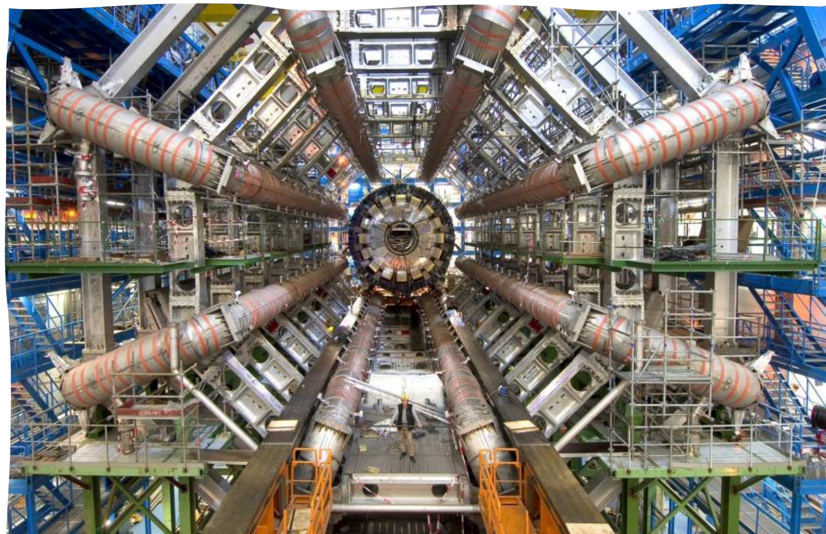
POWERED BY egi POWERED BY

Funded by the European Union
The interTwin project is funded by the European Union - Grant Agreement Number 101058386

interTwin Physics DT Applications

- Lattice QCD Simulations - HEP
- Particle Detector Simulation - HEP
- Noise Simulation for Radio Astronomy
- VIRGO Noise detector - Astrophysics



DT examples in HEP

electronics MDPI

Article

Practical Digital Twins Application to High Energy Systems: Thermal Protection for Multi-Detector

Andrzej Wojtulewicz [1], Paweł D. Domański [1], Maciej Czarnynoga [2], Monika Kutyła [2], Maciej Ławryńczuk [1,*], Robert Nebeluk [1], Sebastian Plamowski [1], Krystian Rosłon [2] and Krzysztof Zarzycki [1]

1 Institute of Control and Computation Engineering, Faculty of Electronics and Information Technology, Warsaw University of Technology, ul. Nowowiejska 15/19, 00-665 Warsaw, Poland; andrzej.wojtulewicz@pw.edu.pl (A.W.); pawel.domanski@pw.edu.pl (P.D.D.); robert.nebeluk@pw.edu.pl (R.N.); sebastian.plamowski@pw.edu.pl (S.P.); krzysztof.zarzycki@pw.edu.pl (K.Z.)
2 Faculty of Physics, Warsaw University of Technology, ul. Koszykowa 75, 00-662 Warsaw, Poland; maciej.czarnynoga@pw.edu.pl (M.C.); monika.kutyla@pw.edu.pl (M.K.); krystian.roslon@pw.edu.pl (K.R.)
* Correspondence: maciej.lawrynczuk@pw.edu.pl

sensors MDPI

Article

Digital Twins in the Practice of High-Energy Physics Experiments: A Gas System for the Multipurpose Detector

Patryk Chaber [1], Paweł D. Domański [1], Daniel Dąbrowski [2], Maciej Ławryńczuk [1,*], Robert Nebeluk [1], Sebastian Plamowski [1] and Krzysztof Zarzycki [1]

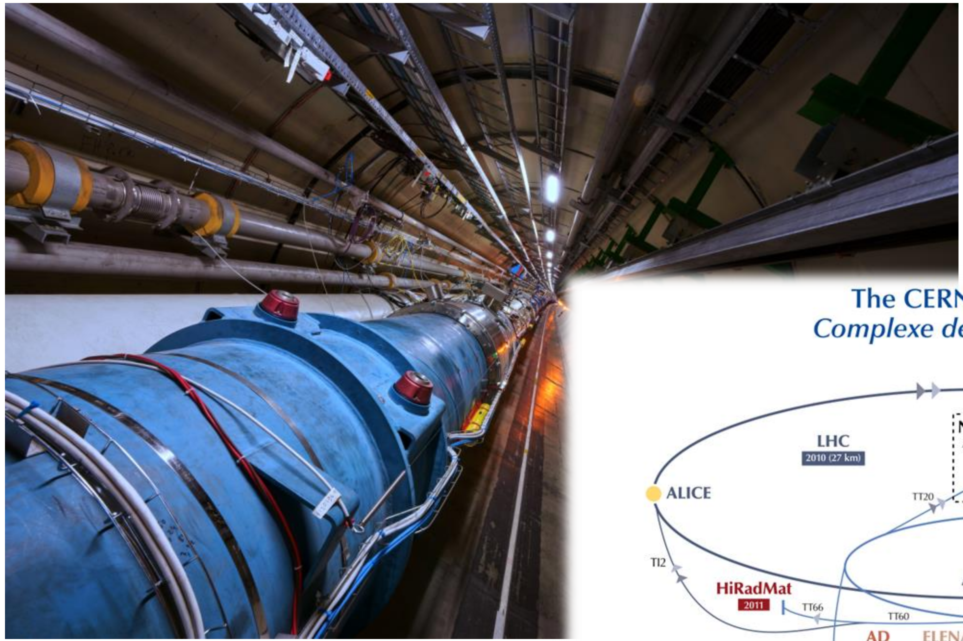
1 Institute of Control and Computation Engineering, Faculty of Electronics and Information Technology, Warsaw University of Technology, ul. Nowowiejska 15/19, 00-665 Warsaw, Poland; Patryk.Chaber@pw.edu.pl (P.C.); Pawel.Domanski@pw.edu.pl (P.D.D.); Robert.Nebeluk@pw.edu.pl (R.N.); Sebastian.Plamowski@pw.edu.pl (S.P.); Krzysztof.Zarzycki@pw.edu.pl (K.Z.)
2 The Faculty of Physics, Warsaw University of Technology, ul. Koszykowa 75, 00-662 Warsaw, Poland; Daniel.Dabrowski.dokt@pw.edu.pl
* Correspondence: Maciej.Lawrynczuk@pw.edu.pl

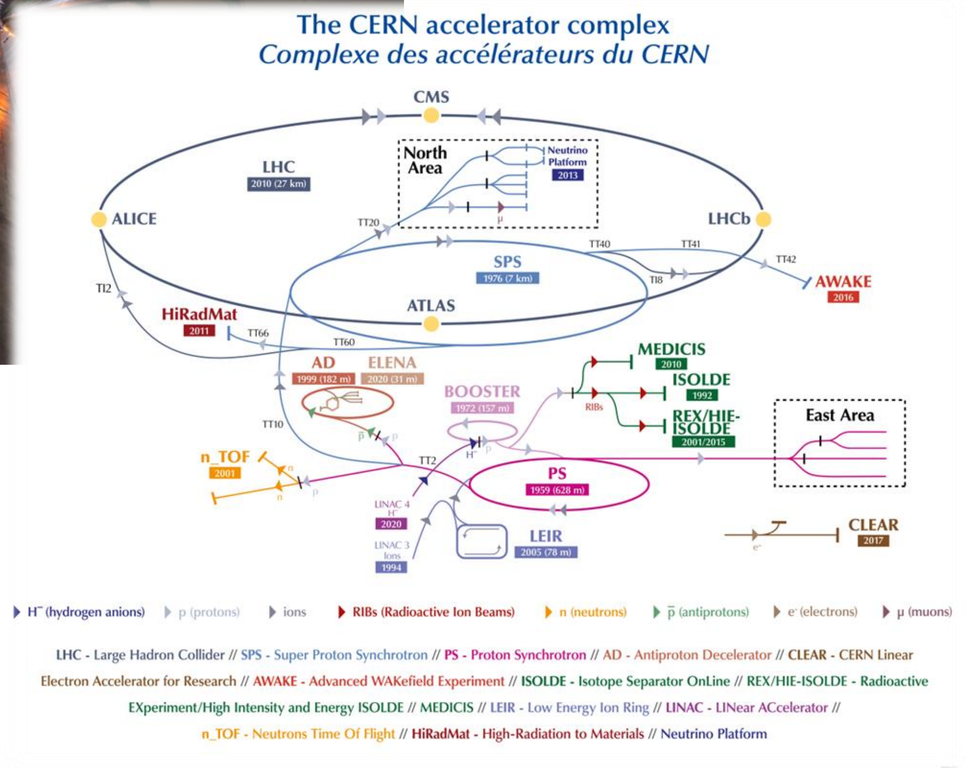
- Digital Twin developed for a thermal system [2]
  - maintaining proper thermal conditions for the operation of the Silicon Tracking Detectors
- DT approach to help in the design and the extension of the hardware and software objects

- Development of a control system using digital twins approach for a gas system delivering a specific mixture of gases to the multipurpose detector (MPD) [3]

LHC & Particle Detectors in a nutshell

- Large Hadron Collider (LHC) is the world's largest and most powerful particle accelerator
- 27 km ring of superconducting magnets
- 2 high-energy particle beams travel at close to the speed of light in opposite directions, before they are made to collide
- beams inside LHC are made to collide at four locations around the accelerator ring → positions of four experiments/particle detectors – ATLAS, CMS, ALICE and LHCb





Detector simulation – WLCG Overview

- Simulation of particle transport through matter is fundamental for interpreting the results of HEP experiments
- Particles undergo complex interactions while traversing the detector material with stochastic outcomes

- These processes are modelled with Monte Carlo (MC) techniques that rely on repeated random sampling
- MC simulation meets the theoretical predictions with a high degree of precision but is time and resource intensive

- Worldwide LHC Grid has currently more than 50% of its resources focused only to simulation [4]
  - WLCG provides global computing resources for the storage, distribution and analysis of the data generated by the LHC
- future High Luminosity LHC will require 100 times more simulated data [5]

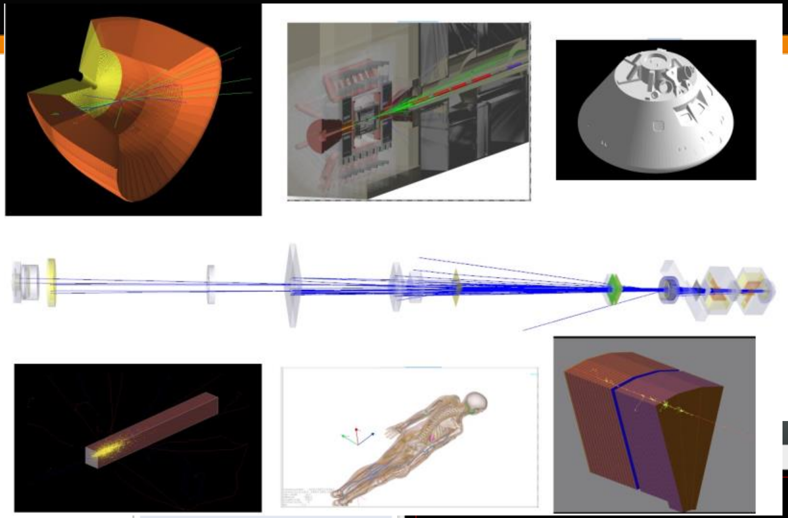
WLCG
Worldwide LHC Computing Grid

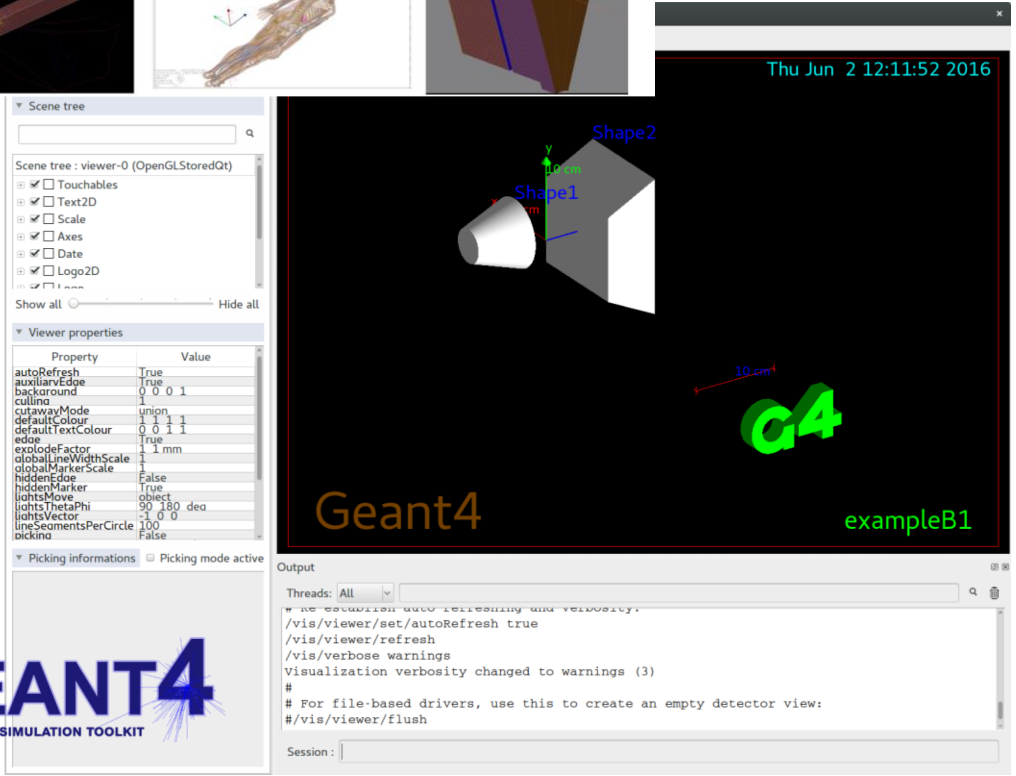
Detector simulation – Geant4 overview

- Geant4 is an **open-source toolkit** that enables **simulating particle passage** through matter
- written in C++
- **toolkit** means that there is no main program provided
- provides all the necessary **components** needed **to describe and to solve** particle transport **simulation** problems in the form of **user interfaces**
- provides necessary **tools** for users to write their simulation applications

- **problem definitions/description:** geometry, particles, physics, etc.
- **problem solution**: step-by-step particle transport computation based on MC methods

- each simulation problem requires different configuration that the user needs to define
- Visualization capabilities using OpenGL





Detector simulation – Motivation overview

Motivation

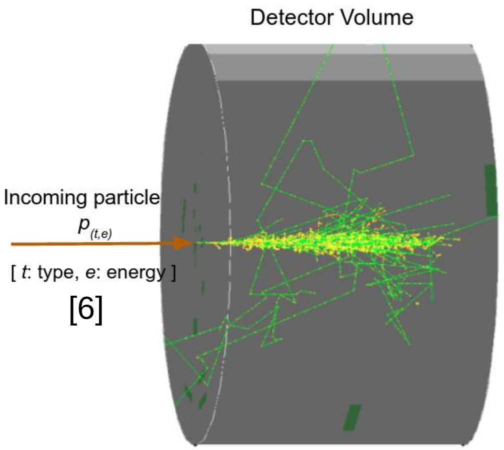
- The detailed (full) particle MC simulation is inherently slow, as well as a complex multi-dimensional problem
- Simulation takes substantial part of computing resources
- Calorimeters are the sub-detectors that are most time-consuming [6]

Solution

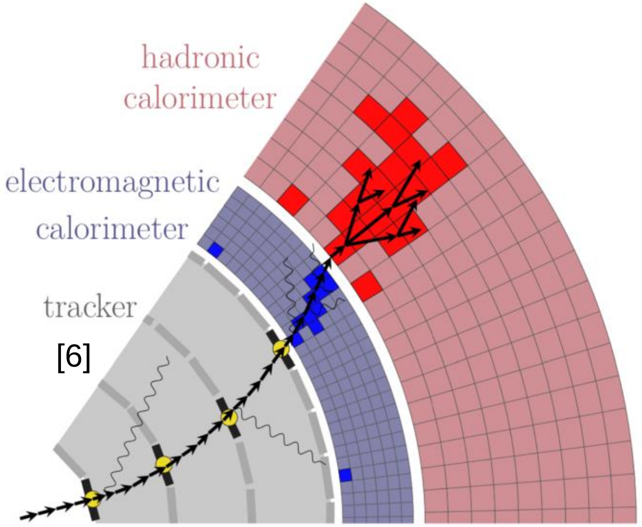
- HEP community is therefore highly motivated to explore fast alternatives, often trading some accuracy for speed
- Fast simulation is a set of established techniques that replaces parts of the detailed MC simulation with alternative approaches
- Here we are leveraging a generative adversarial network approach developed at our lab [1]

How?

- HEP detectors can be described as 3D cameras, taking pictures of particle collisions
- Calorimeters detect particles by measuring the energy deposited in interactions with matter
- Calorimeters consist of arrays of active sensor material and passive dense layers, which ensure that the incoming (primary) particle will deposit most of its energy inside their volume
- energy depositions in calorimeter cells can be compared to the monochromatic pixel intensities of a 3D image



[6]



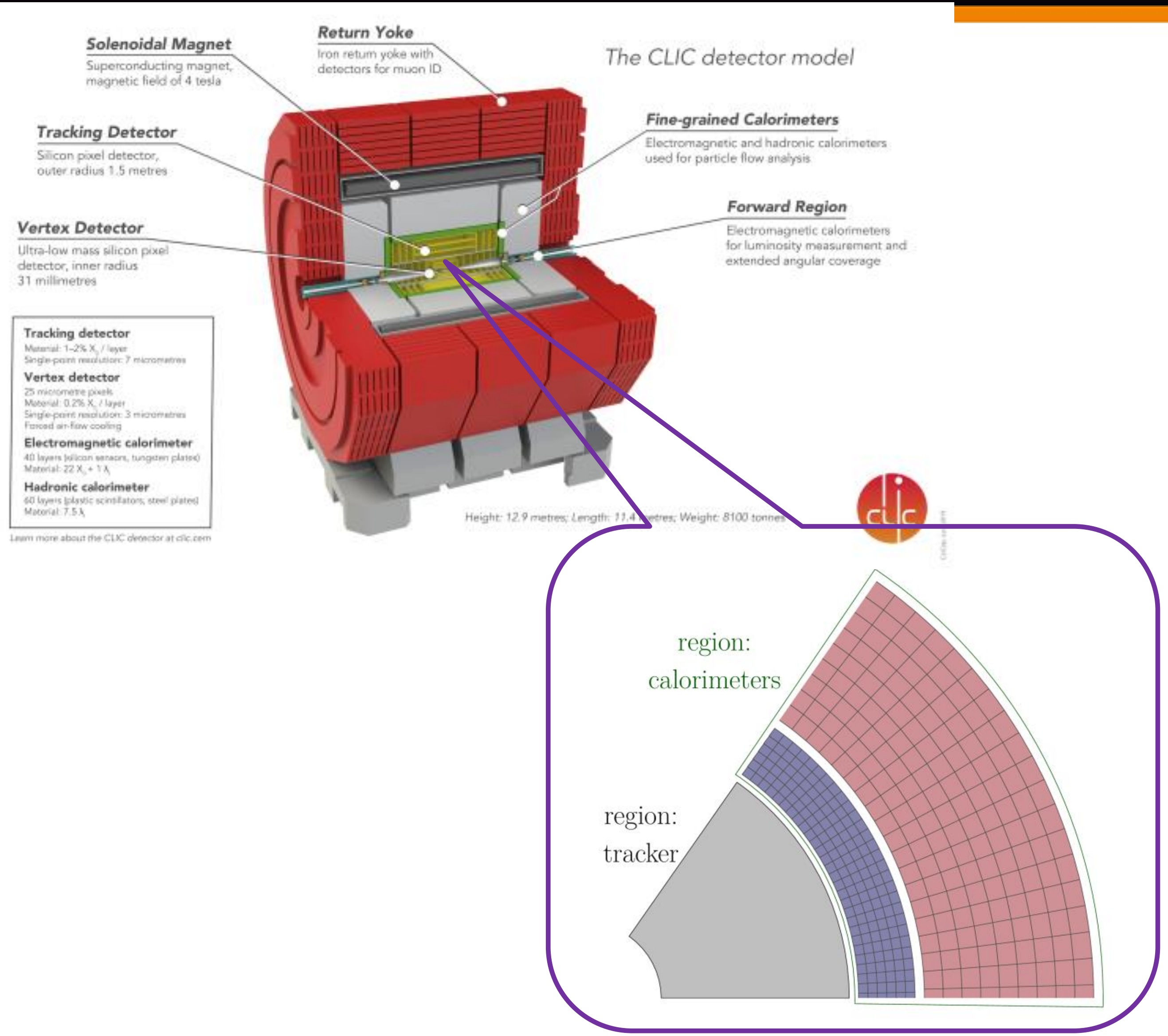
[6]

DT Particle Detector application - Scope

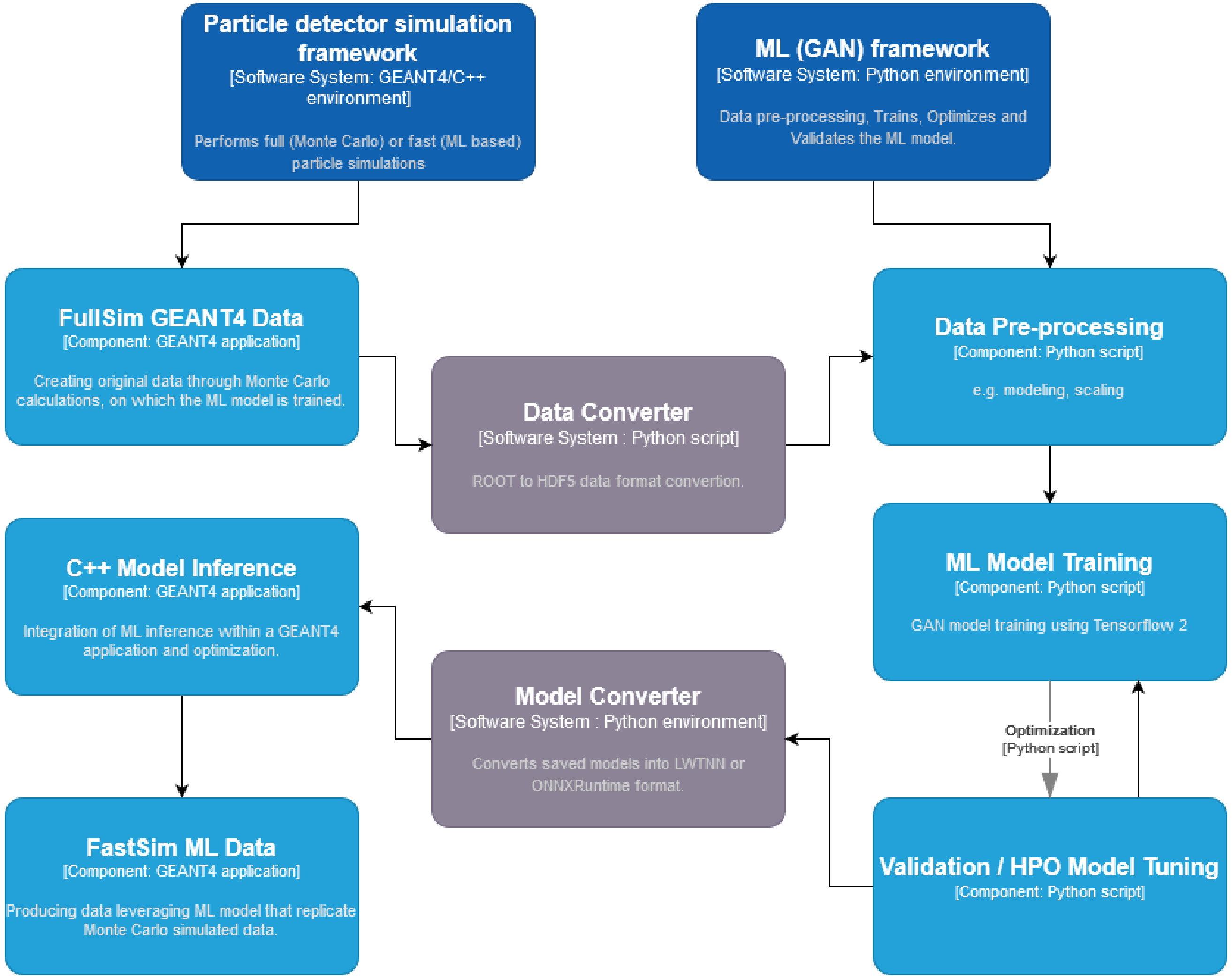
- **Detector Prototyping & Optimization**
  - Build **data-driven tool that simulates detector response** and integrates operation conditions from experimental setups (test-beams)

- **Online ML for Detectors**
  - adapt **real-time** detector and/or data acquisition configuration with respect to run conditions

- **Quality verification & Validation frameworks**
  - model **convergence and accuracy** of the generated data should be monitored
  - development of **sample-based validation framework** in collaboration with HEP community



Particle Detector DT Workflow composition



GOAL: Data-driven tool that simulates the output of a particle detector:
Train ML/DL models for generating data “similar” to training data, that exhibits the same physical properties

CERN’s application will bring data-driven prototypes to production level by:

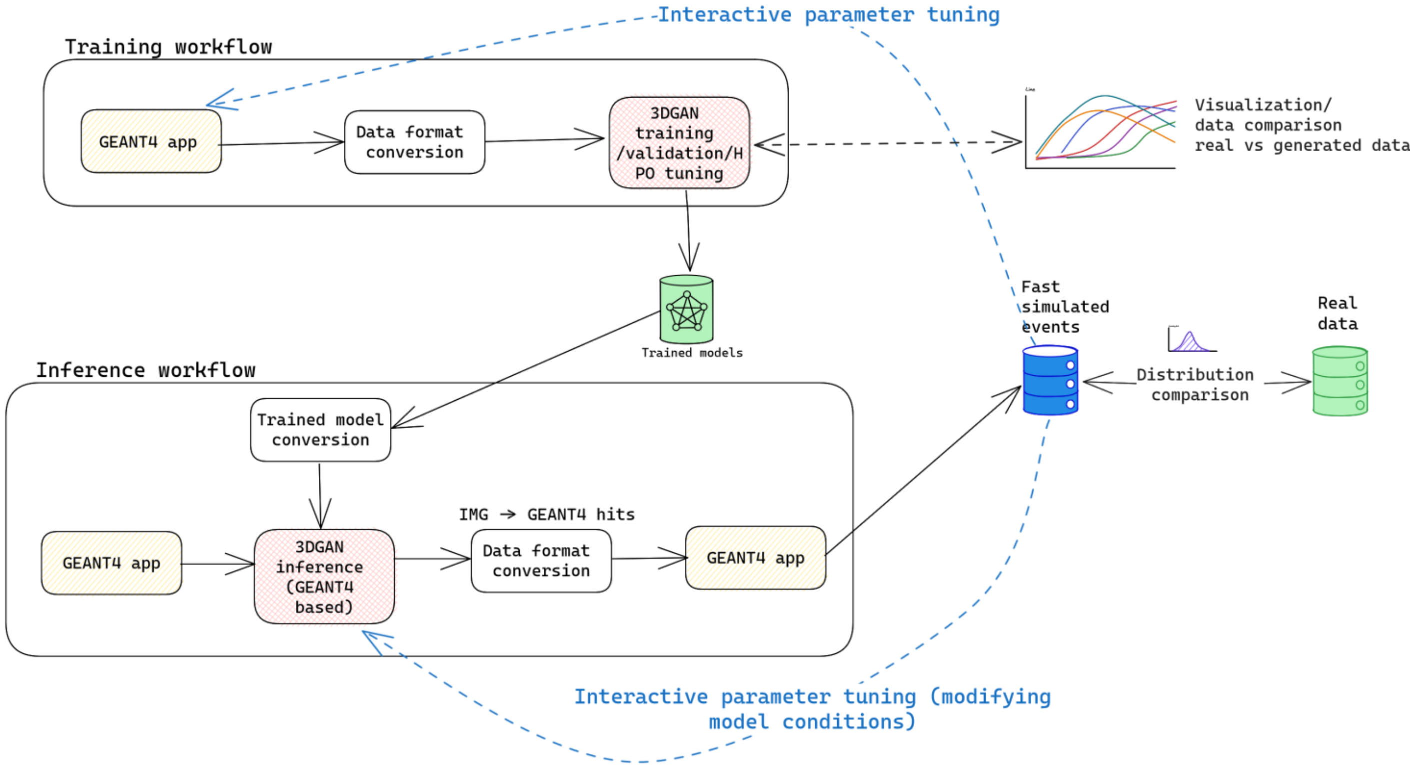
- integrating Monte Carlo based solutions
- establishing a flexible and detailed validation process

DT components:

- the simulation component that incorporates the Monte Carlo-based simulation framework (GEANT4)
- the deep learning (3D Generative Adversarial Network) component, which will produce deep learning models based on a specified particle detector set up.

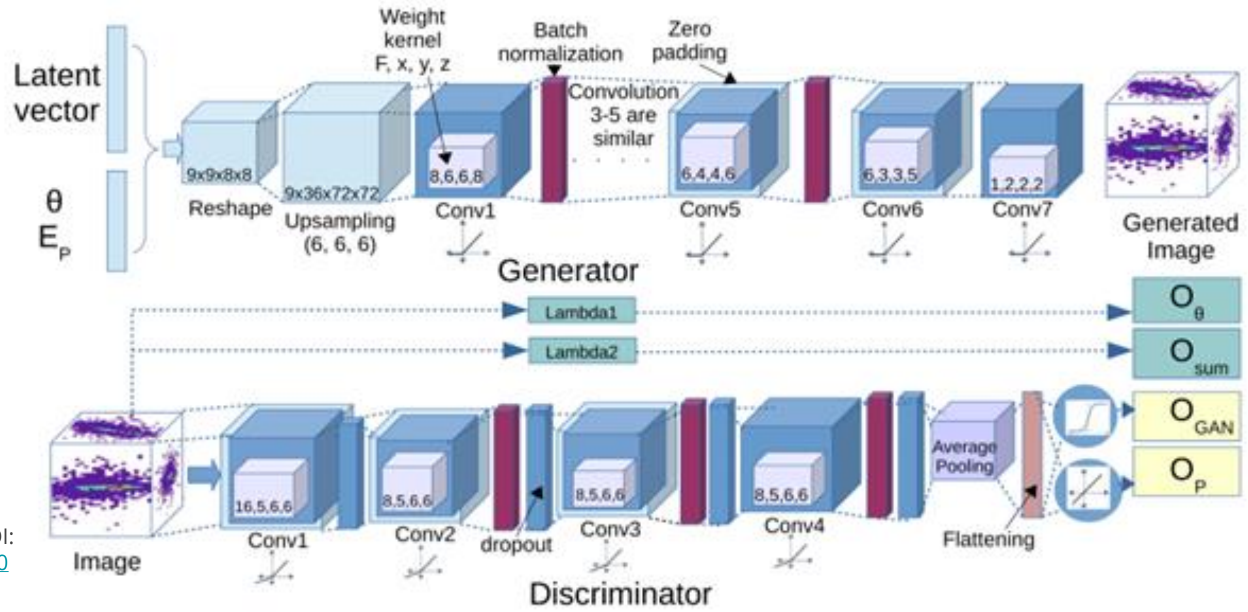
Fast detector simulation training and inference workflows

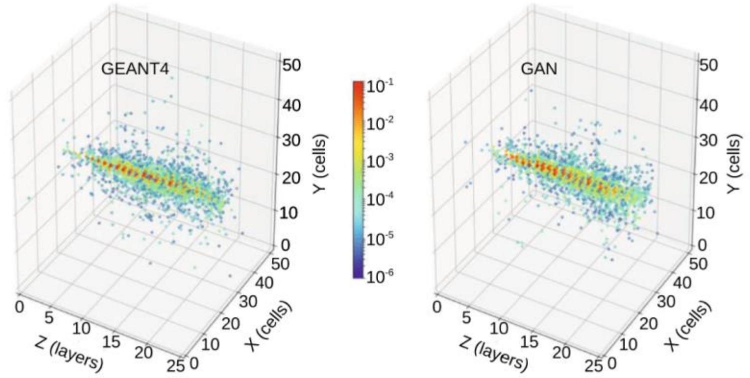
- **Generate input data for training**
  - Run Monte Carlo simulations locally using GEANT4 software → Output: ROOT files
- **Pre-process ROOT files before feeding the data into the ML model**
  - Input: ROOT files → Output: HDF5 files (decreased volume)
- **Store input and output data**
  - Object or local storage
- **Distributed training with multiple GPUs**
  - GANs, to be tested: Transformer-based models
- **Model inference**
- **Validation/Quality Check**
  - Comparing generated data with Monte Carlo data
  - Sample-based metrics
- **Continuous re-training**
  - Current state of use case (as described in proposal) is a static syn model of a detector.
  - Exploring of extending to an application capable of modelling in real time the detector's output in different operation conditions (beams and accelerator configurations) → continuous re-training on real data



CERN 3DGAN approach

- For the ML component of CERN’s thematic module, a Generative Adversarial Network based model is being leveraged
- 3DGAN: Simulation of a future high granularity calorimeter output as three dimensional images (51x51x25 pixels)*
- Three orders of magnitude speedup with respect to GEANT4
- Calorimeter cells are represented as monochromatic pixelated images with the cell energy depositions being the pixel intensities
- Our approach uses 3D convolution layers to represent the 3 spatial dimensions of the calorimeter images
- 3DGAN consists of 2 networks, the generator and discriminator





Example of a GEANT4 electron event (left) vs. an event generated by GAN (right) for the same initial conditions

***Fast Simulation of a High Granularity Calorimeter by Generative Adversarial Networks.** Gul Rukh Khattak et al. https://arxiv.org/abs/2109.07388 DOI: https://doi.org/10.48550/arXiv.2109.07388

DT capabilities and requirements

The designing process of our digital twin application and its capabilities led to identify application technical requirements, that were presented during past months and reported in D7.2

3DGAN model for fast detector calorimeter simulations

- The optimized 3DGAN model replicates the selected dataset (i.e. Monte Carlo simulation data).
- GPUs are required for the training phase.
- Tools for parallel training and hyper parameter optimization will be modified to accommodate the adversarial training process if needed.

Integration to MC-based framework and optimization of the data transformation pipelines

- Integration of the GAN models with the MC-based framework (GEANT4) will be implemented.
- The pre-trained model is stored and used for inference taking into account the data transformations between MC output and model input and vice versa.

Implementation of multi-modal probability learning capability

- The implementation of complex multivariate distributions based on a large range of input conditions.

Continuous retraining and quality verification

- Infrastructure to implement continuous training approach, updating the model configuration whenever new data becomes available.
- The convergence of the model should be monitored and the accuracy of the generated data.

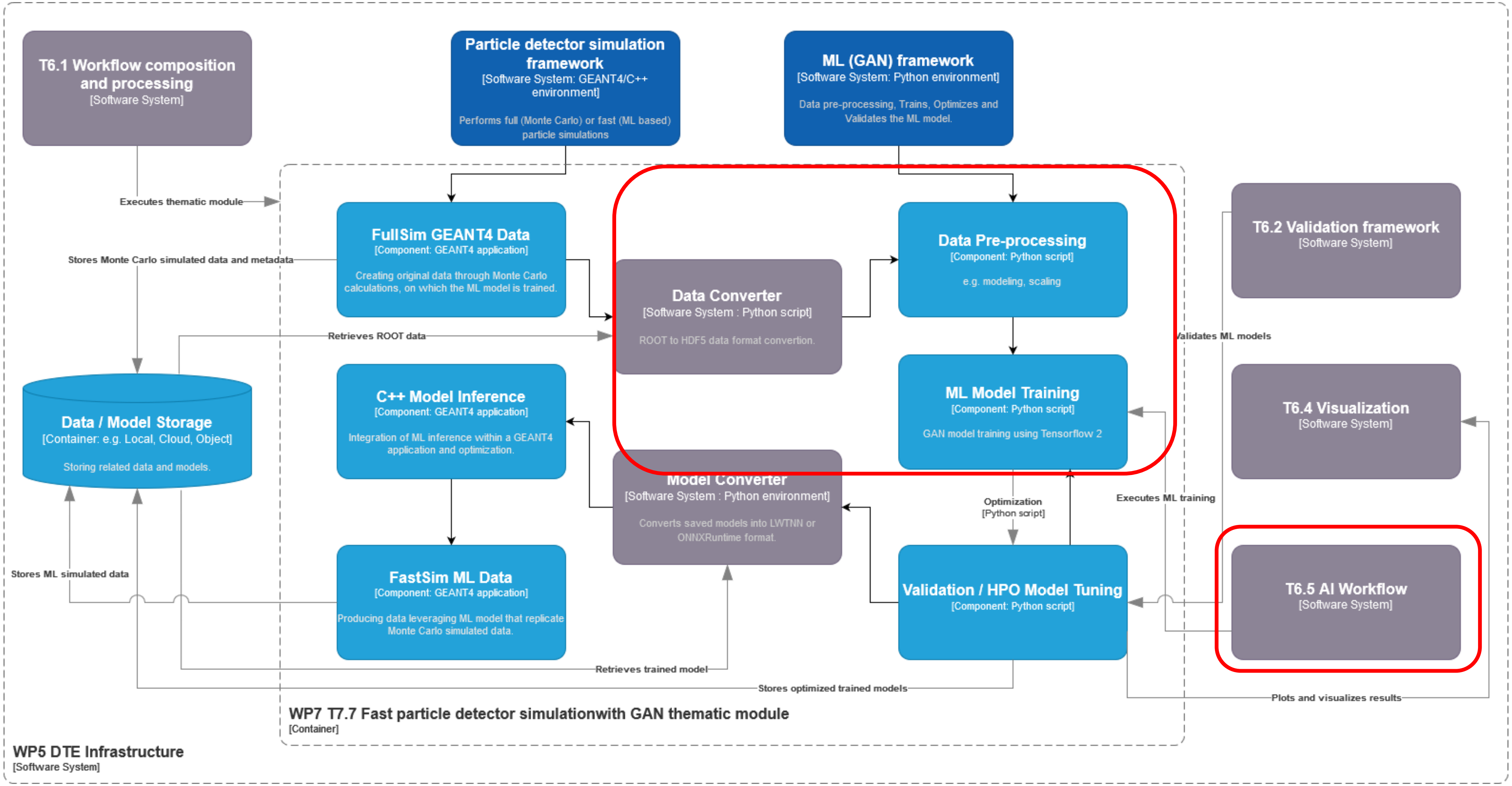
Validation

- Customizable validation framework will be developed in collaboration with HEP community experts

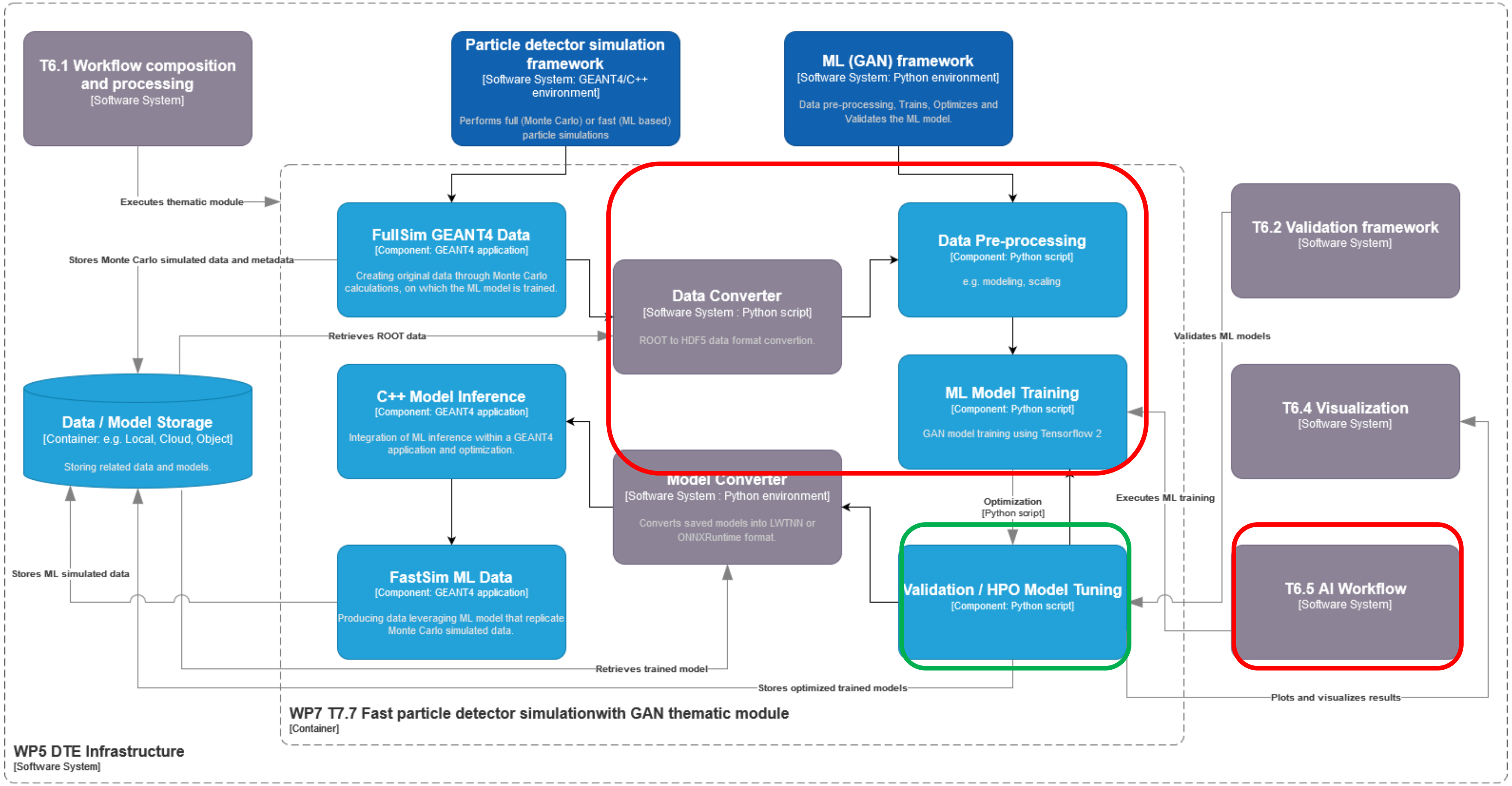
Fast simulation with GAN

- **Objectives**
  - Optimizing the Generative Adversarial Network (GAN)-based model developed for a selected set of detector geometries.
  - Integrating tools for distributed training and hyperparameter optimization.
  - Implementing validation techniques capable of assessing different performance aspects, such as accuracy and comparison to classical Monte Carlo, uncertainty estimation, coverage of the support space. This activity will be contributing to the development of an agreed validation standard among the HEP community.
- **Task background / overview**
  - GEANT4: simulation toolkit that performs particle physics simulations based on Monte Carlo methods (C++ framework)
    - Set of components including geometry and particle propagation descriptions, detector response modelling, event management, user interfaces and more...
  - Faster alternatives to Monte Carlo, including deep learning based techniques
  - Generative models (used in similar HEP applications) are able to combine deep learning with statistical inference and probabilistic modelling
    - deep learning fast simulation generates directly the detector output, without reproducing, step by step, each single particle that interacts with the detector material

Implemented components until now

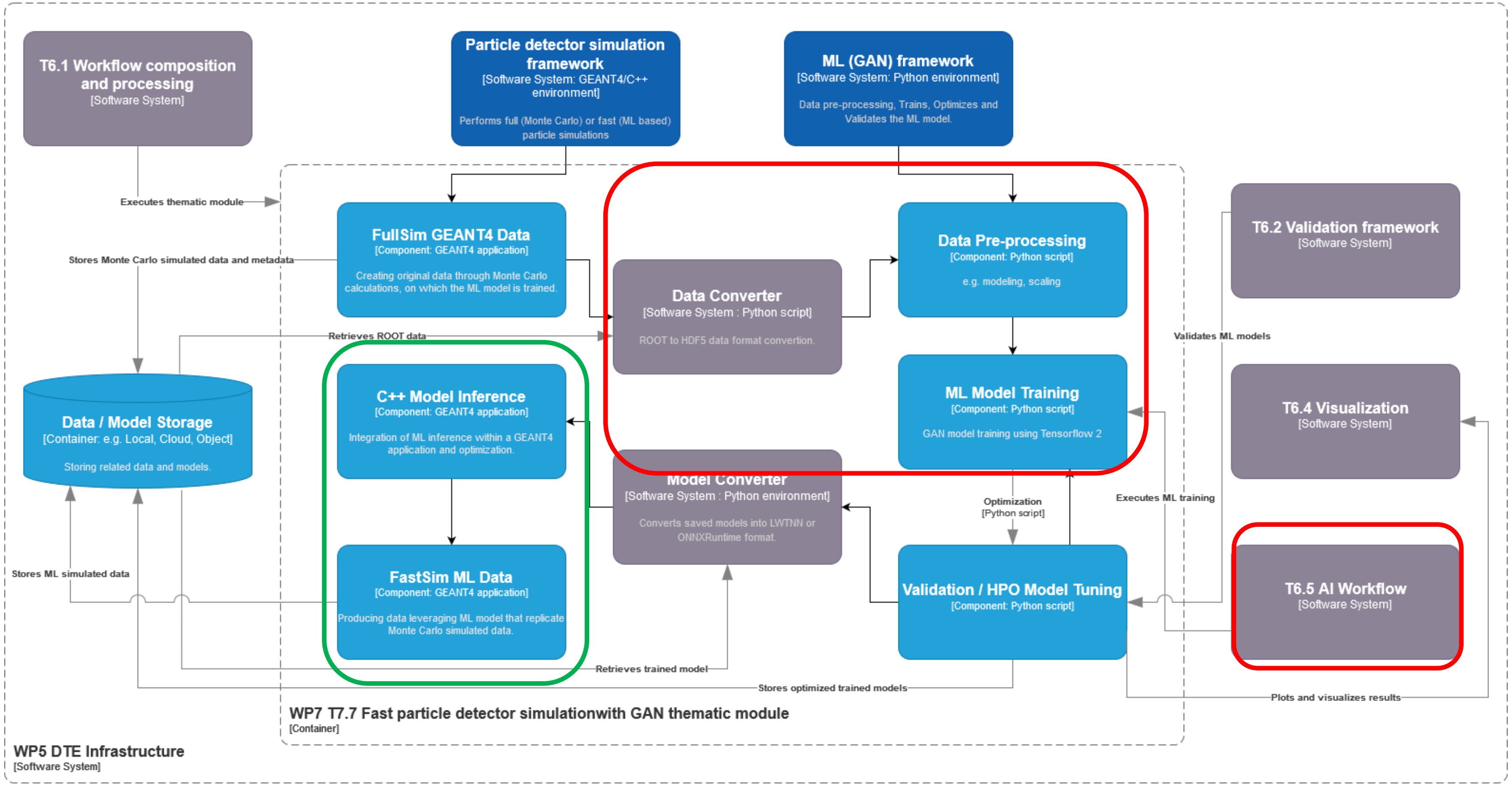


Next steps and challenges



- Implementation of validation techniques capable of assessing different aspects of performance, including accuracy and comparison to classical Monte Carlo
  - **Challenge**: the creation of an agreed validation standard at the community level

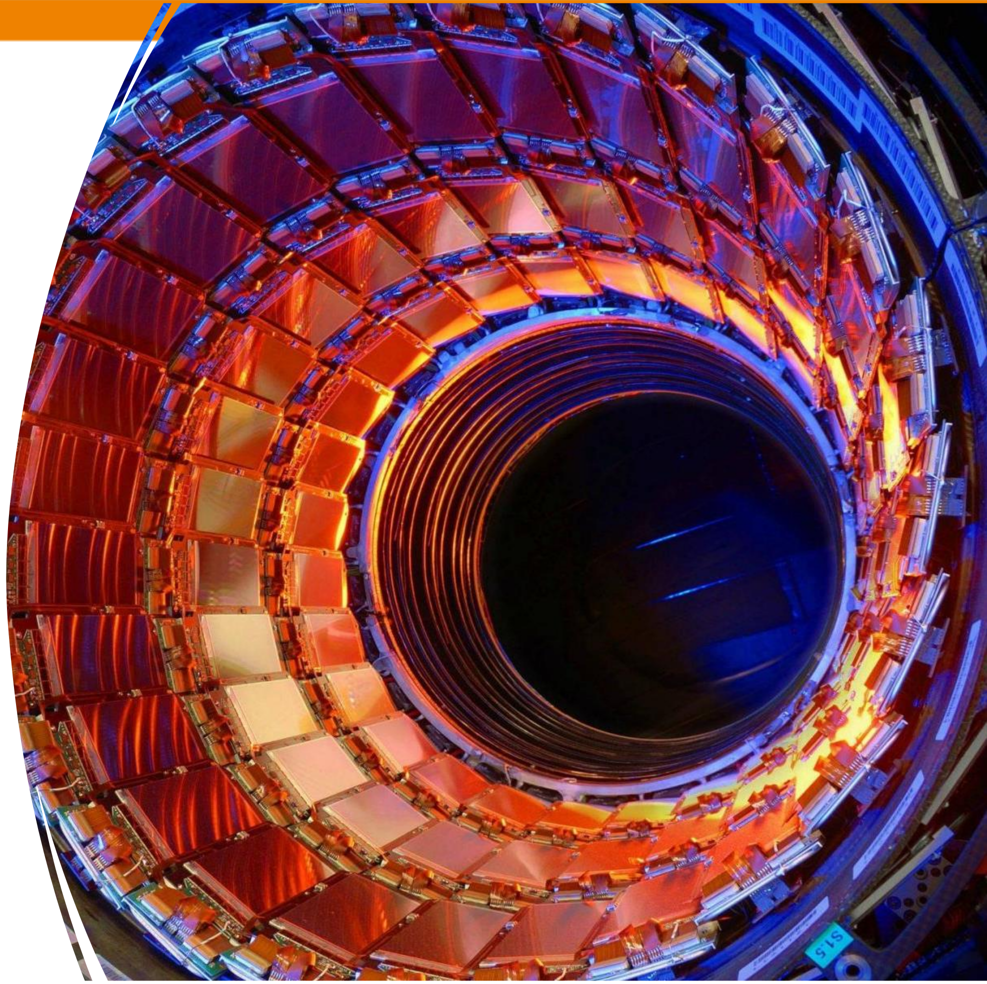
Next steps and challenges cont'd



- Integration of 3DGAN model to MC-based framework (GEANT4) and optimization of the data transformation pipelines
- The stored pre-trained model is used during inference inside GEANT4 → generation of fast simulation ML data
  - **Challenge**: attentive data transformations between MC output and model input and vice versa

THANK YOU!

QUESTIONS?



Contact details:

- Kalliopi Tsolaki: kalliopi.tsolaki@cern.ch
- Sofia Vallecorsa: sofia.vallecorsa@cern.ch
- David Rousseau: rousseau@ijclab.in2p3.fr

Annex

- interTwin: https://www.intertwin.eu/
- References:

[1] Khattak, G.R., Vallecorsa, S., Carminati, F. et al. Fast simulation of a high granularity calorimeter by generative adversarial networks. Eur. Phys. J. C 82, 386 (2022). https://doi.org/10.1140/epjc/s10052-022-10258-4

[2] Wojtulewicz, A.; Domański, P.D.; Czarnynoga, M.; Kutyła, M.; Ławryńczuk, M.; Nebeluk, R.; Plamowski, S.; Rosłon, K.; Zarzycki, K. Practical Digital Twins Application to High Energy Systems: Thermal Protection for Multi-Detector. Electronics 2022, 11, 2269. https://doi.org/10.3390/electronics11142269

[3] Chaber, P.; Domański, P.D.; Dąbrowski, D.; Ławryńczuk, M.; Nebeluk, R.; Plamowski, S.; Zarzycki, K. Digital Twins in the Practice of High-Energy Physics Experiments: A Gas System for the Multipurpose Detector. Sensors 2022, 22, 678. https://doi.org/10.3390/s22020678

[4] J. Albrecht et al., "A roadmap for HEP software and computing R&D for the 2020s," Comput. Softw. Big Sci., vol. 3, no. 1, p. 7, Mar. 2019

[5] G. Apollinari, I. A. BØjar, O. Brüning, P. Fessia, M. Lamont, L. Rossi, and L. Tavian, "High-luminosity large hadron collider (HL-LHC): Technical design report V. 0.1," CERN, Geneva, Switzerland, Tech. Rep. CERN-2017-007-M, 2017, pp. 1–516, vol. 4, DOI: https://doi.org/10.23731/CYRM-2017-004

[6] https://g4fastsim.web.cern.ch/