Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani
ICSC

Advanced Machine Learning. Flash Simulation and bleeding edge applications

FlashSim: *February status report*

Lucio Anderlini
Istituto Nazionale di Fisica Nucleare, Sezione di Firenze

External Partner

Who we are

Staff members:

- Alessandro Bombini [j], INFN
- Giuseppe Piparo[l], INFN
- Maurizio Martinelli [a], Università Milano Bicocca
- Simone Capelli [a], Università Milano Bicocca
- Federica Maria Simone [i], Politecnico di Bari
- Nicola De Filippis [i], Politecnico di Bari
- Vieri Candelise [h], Università di Trieste
- Giuseppe Della Ricca [h], Università di Trieste
- Valentina Zaccolo [k], Università di Trieste
- Mattia Faggin [k], Università di Trieste
- Lorenzo Rinaldi [e], Università di Bologna
- Piergiulio Lenzi [g], Università di Firenze
- Vitaliano Ciulli [g], Università di Firenze
- Sharam Rahatlou[h], Università Roma 1
- Daniele del Re [h], Università Roma 1
- Lorenzo Capriotti [f], Università di Ferrara
- Francesco Conventi [e], Università di Napoli
- Francesco Cirotto [e], Università di Napoli

PhD students:

- Francesco Vaselli [c], Scuola Normale Superiore di Pisa
- Matteo Barbetti [b], Università di Firenze
- Muhammad Numan Anwar [j], Politecnico di Bari
- Benedetta Camaiani [g], Università di Firenze
- Alkis Papanastassiou [g], Università di Firenze
- Antonio D'Avanzo [e], Università di Napoli

External collaborators:

- Andrea Rizzi [c], Università di Pisa

Ongoing activities

- ALICE - Timing-based vertex-reconstruction validation for the triggerless DAQ
- ATLAS - Fast simulation
- ATLAS - 4D reconstruction algorithms
- ATLAS - Anomaly detection for full-hadronic final states
- CMS - Flashsim
- CMS - Theory-independent classifiers for CMS using domain adaptation
- CMS - Data anomaly detection for data validation and certification
- LHCb - Lamarr (Flashsim) - PID and Calorimetry
- LHCb - Lamarr (Flashsim) - Tracking
- LHCb - Parametrization of Cherenkov detector path inRICH detectors
- LHCb - Flash simulation of resistive solid-state detectors
- LHCf - Reconstruction of multiple calorimetric clusters

Simulation
Reconstruction
Monitoring
Data analysis

Hear-beat survey: *responses*

- LHCf - Reconstruction of multiple calorimetric clusters: ***up and running***
- LHCb - Flash simulation of resistive solid-state detectors: ***up and running***
- LHCb - Tracking parametrizations for Flash Sim: ***up and running***
  - *Repo: https://github.com/s-capelli/lb-trksim-train/tree/notebooks-bicocca*
  - *Abstract at ACAT 2024 accepted as poster*
- LHCb - Tracking parametrizations for Particle ID and Calorimetry: ***up and running***
- CMS - Data anomaly detection for data validation and certification: ***up and running***
  - *Abstract to AISSAI2024 accepted!*
  - *Repo: https://github.com/fsimone91/MuonAE*
- ALICE - Reconstruction of Omega decays with ML techniques: ***restarting***

The current setup made available by AI_INFN is a too scattered into many R&D clusters were different R&D are being performed. We need to converge quickly to a more stable setup.
Can we have a formal request for resources to INFN CSN5?

We receive many complaints on the stability and performance of MinIO (DataCloud will phase it out).
Can we kindly push for a quick convergence on the identification of a cloud storage solution.

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani
ICSC

Development of algorithms for flash simulation

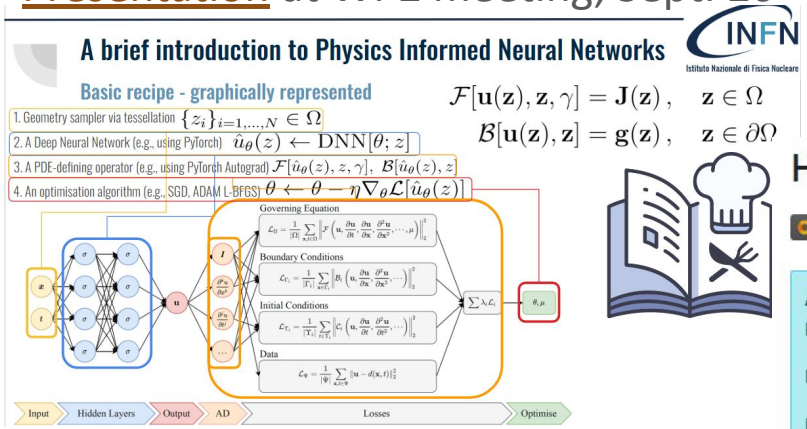
Effort focused on the development of algorithms for the flash simulation.

Generative Adversarial Networks

PIDGAN [LHCb]

Algorithms*	Avail	Test	Lipschitzianity**	Design inspired by	Tutorial
GAN	✅	✅	❌	1, 8, 9	Open in Colab
BceGAN	✅	✅	❌	2, 8, 9	Open in Colab
LSGAN	✅	✅	❌	3, 8, 9	Open in Colab
WGAN	✅	✅	✅	4, 9	Open in Colab
WGAN_GP	✅	✅	✅	5, 9	Open in Colab
CramerGAN	✅	✅	✅	6, 9	Open in Colab
WGAN_ALP	✅	✅	✅	7, 9	Open in Colab
BceGAN_GP	✅	✅	✅	2, 5, 9	Open in Colab
BceGAN_ALP	✅	✅	✅	2, 7, 9	Open in Colab

Presentation at WP2 meeting, Sept. 26th 2023



Hands-on Deep Generative Models

Tutorial [CMS]

Open in Colab | Open in Kaggle | Open in GitHub

Authors: M. Barbetti (INFN-CNAF), S. Capelli (INFN-MiB), F. Vaselli (INFN-Pisa)
Date created: 30/10/2023

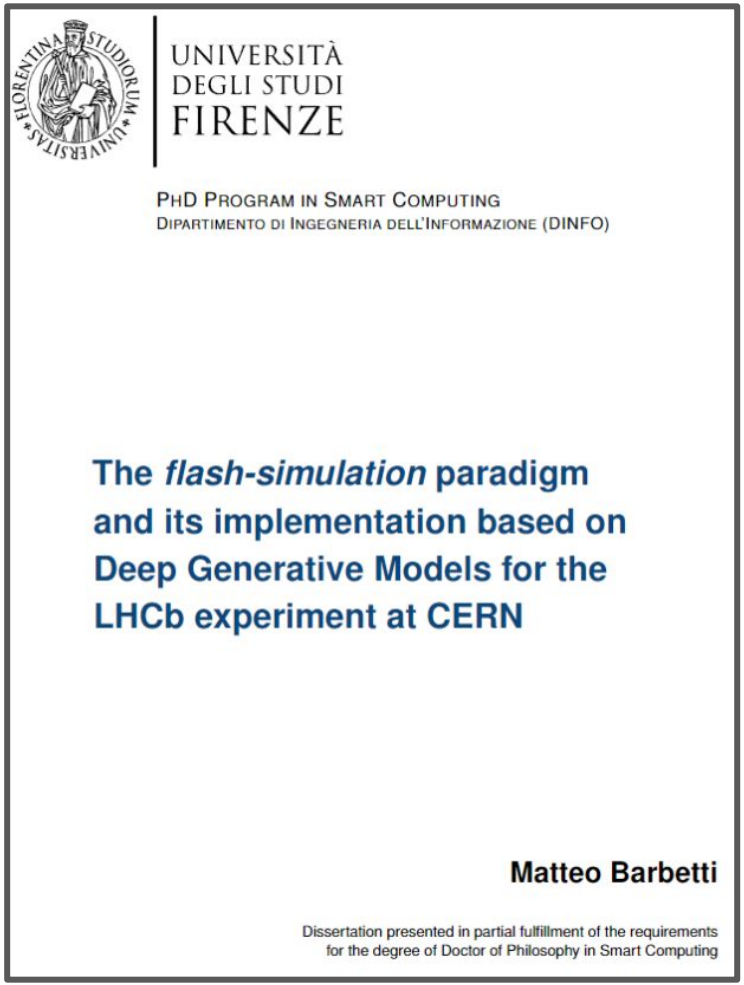
Last modified: 12/11/2023

Description: This hands-on demonstrates how deep generative models succeed in reproducing the high-level response of a generic HEP experiment, offering a viable solution to reduce the pressure on the computing budget for simulation production. In particular, in this notebook we will test the performance of Generative Adversarial Networks and Normalizing Flows to parameterize the errors introduced during the detection and reconstruction of high-energy particle jets using the CMS detector.

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani
ICSC

Documentation

- Ph.D. Thesis Matteo Barbetti
- Advanced Hackathon of ML-INFN [link]
- Notebooks on Lamarr Tracking [link]
- Doxygen on SQLamarr [link]
- PIDGan tutorials [link]



UNIVERSITÀ DEGLI STUDI FIRENZE

PHD PROGRAM IN SMART COMPUTING
DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE (DINFO)

The *flash-simulation* paradigm and its implementation based on Deep Generative Models for the LHCb experiment at CERN

Matteo Barbetti

Dissertation presented in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Smart Computing

Infrastructure for training

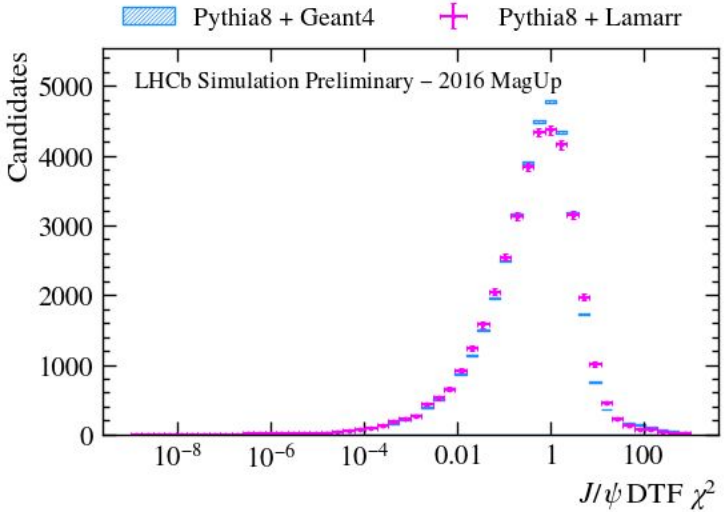
- Uploaded to GitHub the Helm chart of the Kubernetes platform to access GPU resources in cloud: https://github.com/landerlini/ai-infn-platform

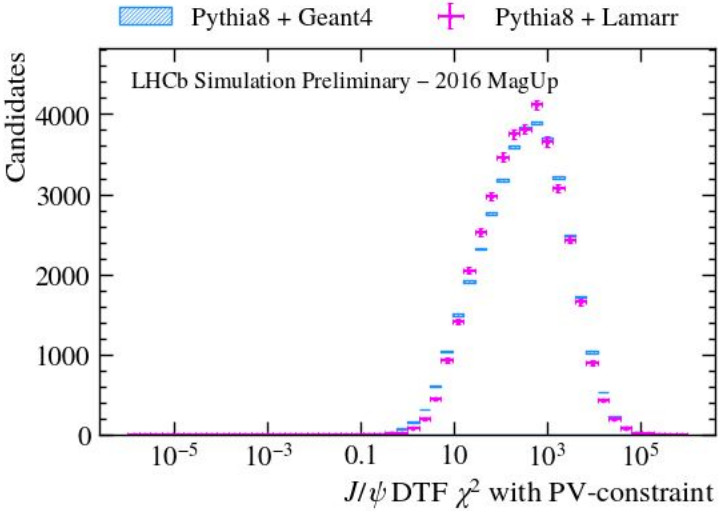
  Features:
  - JupyterHub with IAM token authentication
  - POSIX interface INFN Cloud MinIO
  - Multi-node distributed filesystem based on NFS
  - Rudimental batch system, tested for Geant4-based simulations
- Missing feature for the needs of the Flagship: **cvmfs**.
  - Checked with WP5 the feature does not break the offloading model based on InterLink
  - I would like and endorsement of CSN5 before requesting support for ICSC usecases on these resources
- DNS entry registered, deployment still not ready to production: hub.ai.cloud.infn.it
- 4 flagship use-cases tested on development clusters (see below).

Infrastructure for validation

The validation workflow for Lamarr (LHCb flashsim) has been ported to Snakemake, but it has not been containerized yet.

It has been tested locally for the decay B $\rightarrow \chi_c$ K showing promising results for charged particles and photons, but more tests are needed to validate the full pipeline.





KPIs

KPI ID	Description	Acceptance threshold	2024-02-13
KPI2.2.1.1	N_{MC} billion events obtained from ML-based simulation, as demonstrated by official links in experiments' simulation databases	N_{MC} >= 1	1 M events (completed: 0.1%)
KPI2.2.1.2	N_{EXP} experiments have tested a machine-learning based simulation	N_{EXP} >= 2	0 experiment (completed: 0%)
KPI2.2.1.3	Machine-learning use-cases tested in the context of the CN were presented at N_{CONF} international and national events	N_{CONF} >= 3	3 use-cases (since Sept. '23) (completed: 100%)
KPI2.2.1.4	N_{UC} different machine-learning use-cases were tested in the context of the CN and made available in git repositories	N_{UC} >= 5	4 use-cases (completed: 80%)

List of conferences for KPI2.2.1.3

- L.A., Generative models at the LHC, ALPACA workshop 2023, Trento
- B. Camaiani, Example of adaptation domain in High Energy Physics, XAI 2023, Milano
- A. Papanastassiou, “Anomaly detection with autoencoders for data quality monitoring in HEP”, XAI 2023, Milano

Abstract accepted (not counted in KPIs, yet. Right?)

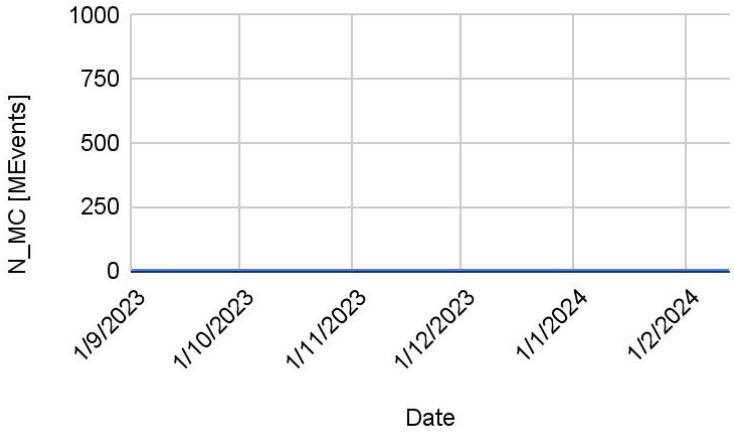
- *Lamarr: implementing the flash-simulation paradigm at LHCb*, ACAT 2024
- F. Simone, *Anomaly detection for data quality monitoring of the CMS detector*, AISSAI 2024

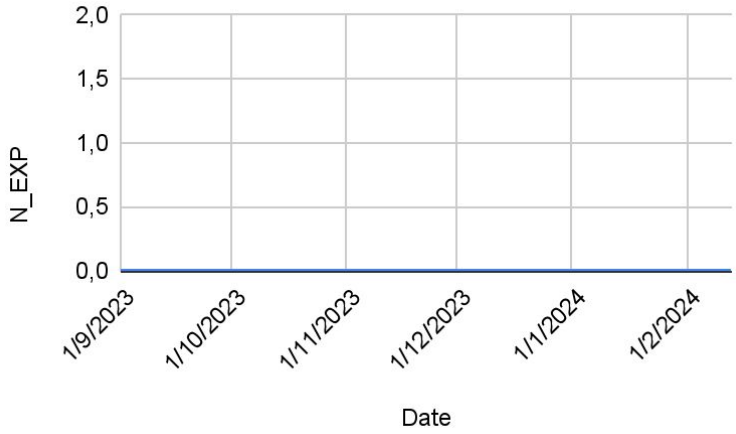
Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani
ICSC

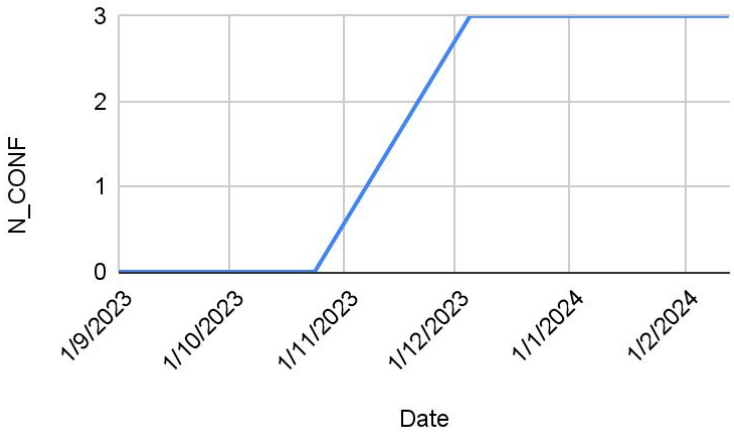
List of use-cases tested on the platform (4/5)

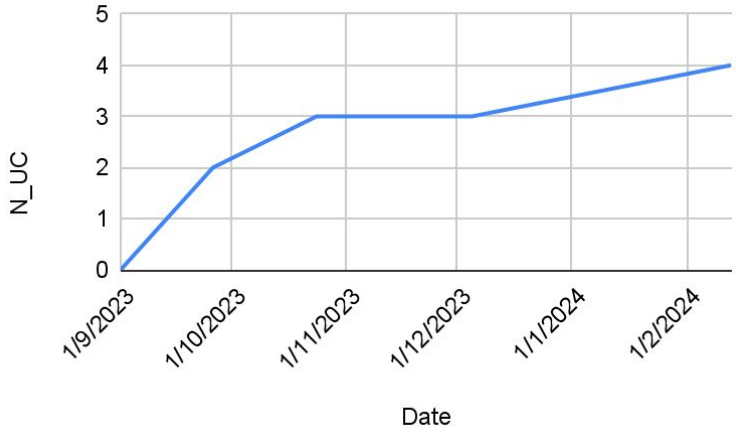
- Lamarr, the ultra-fast simulation option for the LHCb experiment (tracking parametrizations)
- Lamarr, the ultra-fast simulation option for the LHCb experiment (particle identification and neutral reconstruction parametrizations)
- Theory-independent classifiers for the data analysis with the CMS experiment
- Machine-learning-based simulation of the response of resistive solid-state detector to the charge generated by a traversing minimum-ionizing particles

KPIs


N_MC [MEvents]
1000
750
500
250
0
1/9/2023
1/10/2023
1/11/2023
1/12/2023
1/1/2024
1/2/2024
Date


N_EXP
2,0
1,5
1,0
0,5
0,0
1/9/2023
1/10/2023
1/11/2023
1/12/2023
1/1/2024
1/2/2024
Date


N_CONF
3
2
1
0
1/9/2023
1/10/2023
1/11/2023
1/12/2023
1/1/2024
1/2/2024
Date


N_UC
5
4
3
2
1
0
1/9/2023
1/10/2023
1/11/2023
1/12/2023
1/1/2024
1/2/2024
Date

On resources

The HPC bubbles requested by this flagship should be commissioned by summer 2024.

In the meanwhile:

- **Compute**:
  - ML_INFN is providing GPUs for **interactive development** (no batch)
  - some of us is starting playing with ***Leonardo Booster*** for practicing **ML with batch jobs** (spoke, please clarify rules and procedures)
  - ML_INFN plans for experimental support for batch jobs since early 2024, (but very tight schedule)
- **Storage:**
  - cache and ephemeral storage is available through INFN Cloud (few TB)
  - “permanent” storage for data relevant for audit under discussion, for the time being, we are encouraged using minio.cloud.infn.it

Get in touch! The machinering is slowly starting to spin, since getting accustomed to the environment may require time, if you plan to use ICSC resources at some point, we suggest you start practicing with the temporary, resource-limited environment we have now.