# MATTEO BARBETTI

Department of Physics and Astronomy, University of Florence Room 183, Via Sansone 1, 50019 Sesto Fiorentino (FI), Italy

matteo.barbetti@unifi.it

# **EDUCATION**

### University of Florence

Firenze, Italy

PH.D. IN SMART COMPUTING

● Nov 2020 – present

Ph.D. scholarship funded by INFN to carry out Machine Learning research for Physics applications Main activities carried out:

- Software development for Lamarr: the LHCb ultra-fast simulation framework
- Cloud applications development to coordinate Bayesian optimization studies
- Contribution to the LHCb distributed computing operations
- Application of Machine Learning techniques to Medical Physics

Advisory Board: L. Anderlini (INFN Firenze), D. Derkach (HSE University), M. Williams (MIT)

### University of Florence

Firenze, Italy

≥ M.Sc. IN PARTICLE PHYSICS

● Sep 2017 – Jun 2020

Thesis Title: "Techniques for parametric simulation with deep neural networks and implementation for the LHCb experiment at CERN and its future upgrades" (CERN-THESIS-2020-416)

Thesis Advisors: L. Anderlini, P. Lenzi

Thesis Award: "Giulia Vita Finzi" INFN award for the best thesis on computing and networks

Graduation Score: 110/110 cum laude

### University of Florence

Firenze, Italy

B.Sc. in Physics and Astrophysics

Sep 2013 − Sep 2017

Thesis Title: "Study of the charmonium resonances in  $B^+ \to p\bar{p}K^+$  and  $B^+ \to p\bar{p}\gamma K^+$  decays with the LHCb experiment at CERN" (CERN-THESIS-2017-496)

Thesis Advisors: L. Anderlini, G. Latino

Graduation Score: 110/110

### International Schools

4th International School on Open Science Cloud 🏶

Perugia, Italy

INFN, University of Perugia & University of Bologna

Nov 2022

Theme: Computing Models for Scientific Experiments

with diploma

8th Thematic CERN School of Computing

online

CERN School of Computing

Jun 2021

Theme: Scientific Software for Heterogeneous Architectures

with diploma

2nd Summer School in Physical Sensing and Processing

online

University of Bologna

Jul 2020

Theme: Sensing devices, DAQ systems, and data processing strategies

# RESEARCH ACTIVITIES

Since 2019, I am member of the LHCb experiment at CERN's Large Hadron Collider (LHC). In this context, I had the opportunity to focus on some of the most challenging computing developments for the next-generation High Energy Physics experiments, including the application of advanced *Deep Learning* techniques to large samples of proton-proton collision data and their training and deployment in a world-wide distributed environment.

In the same period, in collaboration with the Careggi University Hospital in Florence, I also had the opportunity to contribute to the research in Medical Physics by applying *Machine Learning* techniques to unpublished radiomic<sup>1</sup> datasets.

I am currently preparing my Ph.D. thesis in Smart Computing at the University of Florence with original contributions on LHCb Simulation and Radiomics.

My research activity is mainly devoted to the development of a novel LHCb ultra-fast simulation framework, called LAMARR [1] [2], that I have presented during the international workshop [ACAT22] and more recently at the international conference [CHEP23]. LAMARR consists of a pipeline of modular parameterizations designed to replace both the physics simulation and the reconstruction steps. Most of the parameterizations rely on Machine Learning algorithms [3], such as generative models, that were demonstrated to be able to well reproduce the distributions obtained from the detailed simulation [4]. Part of my Master Thesis was devoted to develop a prototype of such parameterizations: this preliminary study was honored with the national INFN award "Giulia Vita Finzi" for the best thesis on computing and networks.

I am an active developer within the LHCb Simulation team, with the responsibility for the parameterization of the LHCb Particle Identification (PID) system. Generative Deep Neural Networks succeed in reproducing the analysis-level classifiers defined by the PID reconstruction algorithms, showing promising results, which I presented at national [SIF21] and international [LTD] conferences. Parameterizing with high accuracy the PID multivariate distributions requires training such specialized models in GPU-intensive hyperparameter optimization campaigns.

To this end, to profit from multiple computing resources made opportunistically available to the LHCb Collaboration, I proposed a client-server approach with a centralized service defining the optimization procedure in multiple instances running in parallel on different providers. The approach was validated on resources made available to LHCb by CERN via the [CloudBank] initiative or by INFN and CINECA granting access to the Marconi 100 supercomputer. During the last few months, I have been converting the validated prototype into an INFN Cloud service, named Hopaas [5], featuring a web dashboard and a token-based user authentication procedure integrated to the INFN user identity management tools. I have presented Hopaas during the international workshop [ACAT22].

To enable the deployment of the optimized models in the LHCb software stack, running in multiple data centers around the world connected via the LHCb Computing Grid, I contributed to the development of scikinC [6], a tool able to translate Machine Learning models into C files that, once compiled, are distributed with the CernVM File System and dynamically linked to the LHCb applications.

My research is strongly connected to the data analysis effort within the LHCb group in Florence, exploiting advanced statistical treatment of the PID classifiers with Machine Learning techniques to shed light on the nature and the antiprotons observed in primary cosmic rays [7].

<sup>&</sup>lt;sup>1</sup>In the field of medicine, *Radiomics* is a new field of image analysis in which digital medical images are converted into quantitative high-dimensional data.

As part of the research activities for my Ph.D., I am exploring the application of Machine Learning techniques to Medical Physics to develop solutions suitable for the emerging "personalized medicine". In particular, I am working to build a Machine Learning model able to infer the histological type of mediastinal bulky lymphomas, one of the most common hematologic cancers, only using radiomic features drawn from PET images [8]. The aim of the study is to understand how PET radiomic features may predict lymphoma histology and in the future support its diagnosis, offering non-invasive solutions tailored to the single patient. I presented the promising preliminary results at national [SIF22] and international [ECMP22] conferences.

In 2020 I joined the *ML-INFN* project, a national initiative of INFN to coordinate the effort toward the widespread application of Machine Learning technologies to research activities. In 2022 I helped in organizing the Third ML-INFN Hackathon for which I gave two lectures, one about deep generative models and the other about Bayesian hyperparameter optimization [ML INFN22].

# INTERNSHIPS AND TUTORING

# University of Florence

Firenze, Italy

**➡** TUTOR SENIOR (LABORATORY TUTOR AND TEACHING ASSISTANT)

Jan 2021 – present

Scholarship funded by a local grant from the University of Florence for a total of 150 hours/year Main activities carried out:

- Introductory lessons to Python for third-year students of the B.Sc. in Physics
- Physics exercises discussed in class for the B.Sc. in Mathematics students
- Support to Physics Laboratory activities for the B.Sc. in Physics and Mathematics

**INFN** 

Firenze, Italy

**☐** INTERN

Curricular internship at the INFN Florence Division for a total of 150 hours

- Main activities carried out:
- Software development of deep generative models for the LHCb ultra-fast simulation
- Validation of the ultra-fast simulation approach

### **CERN**

Geneva, Switzerland

**≘** Research Intern

**U** Sep 2019 − Dec 2019

Internship funded by a national grant from INFN to spend three months at CERN Main activities carried out:

- Investigating Machine Learning techniques to parameterize the LHCb detector
- Software development for the LHCb ultra-fast simulation

### University of Florence

Firenze, Italy

**♣** Tutor (Laboratory Tutor and Teaching Assistant)

Oct 2018 – Dec 2020

Scholarship funded by a local grant from the University of Florence for a total of 200 hours/year Main activities carried out:

- Support to Physics Laboratory activities for the B.Sc. in Biological Sciences and Physics
- Physics exercises discussed in class for the B.Sc. in Biological Sciences students
- Introductory lessons to Python for third-year students of the B.Sc. in Physics

Matteo Barbetti Last updated: June, 2023

### University of Florence

Firenze, Italy

**韋** Tutor Junior (Student Assistant)

U Jun 2018 – Oct 2018

Scholarship funded by a local grant from the University of Florence for a total of 200 hours Main activities carried out:

- Communicator at "ScienzEstate", a dissemination event to promote scientific research at Florence
- Orientation service for students of the Science Faculty of the University of Florence
- Welcome service for students from the Erasmus Plus Program

CERN Geneva, Switzerland

**當** Intern

**Q** Jul 2017

Internship of two weeks at CERN funded by a local grant from the University of Florence Main activities carried out:

• Finalization of the bachelor thesis at CERN

### University of Florence

Firenze, Italy

**♣** Assistant Librarian

**U** Oct 2016 – Dec 2016

Scholarship funded by a local grant from the University of Florence for a total of 150 hours Main activities carried out:

- Support to students at the Science Library of the University of Florence
- Cataloguing of Geology historical theses

# CONTRIBUTIONS TO OPEN SOURCE SOFTWARE

### LHCb Fast Simulation

Set of tools and Python modules to prepare/process data, train Machine Learning models, and export them as C files to enable the integration within the LHCb software stack. Training scripts are designed to build parameterizations for various LHCb sub-detectors.

mbarbetti/calotron

**4** (

• mbarbetti/lb-pidsim-train

• landerlini/lb-trksim-train

• landerlini/scikinC

# Hyperparameter Optimization in the Cloud

Set of Python modules to deliver and access a cloud-based service for hyperparameter optimization through HTTP requests. Computing nodes contribute to optimization campaigns via token-based authentication, and a web dashboard allows to monitor the status of past and ongoing optimization studies.

• landerlini/hopaas

• landerlini/hopaas\_client

• mbarbetti/optunapi

**\$**()

#### Radiomics and Medical Physics

Set of notebooks and Python scripts for processing data, and for training Machine Learning models powered by radiomic features to infer the histological type of mediastinal bulky lymphoma.

• mbarbetti/mediastinal-lymphoma-classification



### Others

List of other personal contributions to open-source software. It includes a Jekyll theme for academic websites, a Python package to implement generative models in TensorFlow, and a repository containing notebooks and code for INFN education events.

• alshedivat/al-folio

**\*** ()

• mbarbetti/tf-gen-models

**&** (

• landerlini/mlinfn-advanced-hackathon

# 0

### LEADERSHIP & RESPONIBILITIES

### Italian National Institute for Nuclear Physics •

• Tutor at ML-INFN educational events

Jan 2020 – present

# LHCb Collaboration

• Data Quality, Computing and Simulation shifter

Mar 2021 - present

• Lamarr active developer for the Simulation Project

Nov 2020 – present

### Italian Association of Physics Students

• Deputy-President

Oct 2020 – Sep 2021

• National Secretary

**TEACHING** 

Oct 2019 - Sep 2021

• President of the Florence Local Committee

Nov 2018 – May 2019

Editorial Board Member of "Sistemi di Riferimento"Deputy-President of the Florence Local Committee

May 2018 – Sep 2021 Dec 2017 – Nov 2018

# INFN Educational Events, INFN

• Third ML-INFN Hackathon: Advanced Level – Lecturer and Hands-On Tutor

2022

### B.Sc. in Physics and Astrophysics, University of Florence

•	B015862:	Physics	Laboratory	III –	Laboratory	Tutor	and	Teaching	As sistant
---	----------	---------	------------	-------	------------	-------	-----	----------	------------

2020 - 2022

B015861: Physics Laboratory II – Laboratory Tutor
B015860: Physics Laboratory I – Laboratory Tutor

2023 2020 - 2021

#### B.Sc. in Mathematics, University of Florence

• B016237: Physics II with Laboratory – Labora	poratoru Tutor	Laboratory Tutor	v - Laborat	boratory	Lah	with	П	vsics	Ph <sup>·</sup>	B016237:	•
--	----------------	------------------	-------------	----------	-----	------	---	-------	-----------------	----------	---

2021 - 2023

• B016236: Physics I with Laboratory – Laboratory Tutor and Teaching Assistant

2022 - 2023

### B.Sc. in Biological Sciences, University of Florence

• B019238: Physics Laboratory for Biology - Laboratory Tutor and Teaching Assistant

2018 - 2020

• B019231: Physics – Teaching Assistant

2018 - 2019

### OUTREACH & DISSEMINATION

### Editorial board of "Sistemi di Riferimento"

online

AISF

May 2018 – Sep 2021

"SdR" is a dissemination project to promote Italian Physics Departments activities



### Science book "Invenzioni"

Sassi Junior & INFN

Jun 2021

Preparation of a paragraph dedicated to Artificial Intelligence



Last updated: June, 2023

Jury member for "Premio Galileo"

Padova, Italy

City of Padova & University of Padova

Oct 2018

"Premio Galileo" is a literary prize for science dissemination

Blog author online

Fisici Senza Palestra Apr 2016 - Oct 2017

Writing science dissemination articles for the "Fisici Senza Palestra" blog

# Organization of Outreach Events

### Live interview "Fisica del Clima" with Daniele Visioni

online

AISF & Cornell University

 $Mar\ 2021$ 

Organization of an interview about Climate Physics

# Live interview "Women in Science" with Anna Gregorio

online

AISF & University of Trieste

Feb 2021

Organization of an interview on the occasion of Women in Science International Day

# Live interview "COVID19" with Eugenio Valdano

online

AISF & INSERM

Apr 2020

Organization of an interview about statistical models for COVID-19 pandemic

### Outreach event "Tra clima e cocktail"

Firenze, Italy

AISF, Italian Climate Network, CNR & University of Florence

May 2019

Organization of an event aimed to raise awareness about climate change problem

**(1)** 

# Outreach event "Viaggio al Polo"

Firenze, Italy

AISF, Caffè-Scienza, INFN & University of Florence

May 2019

Organization of an event about intelligence according to various scientific domains

**(1)** 

#### Outreach event "Luminoscienza"

Firenze, Italy

AISF, LENS, University of Florence, INRIM & Caffè-Scienza

May 2018

Organization of three scientific evenings on the occasion of International Day of Light

0

### Seminar "The new particles of LHCb" by Lucio Anderlini

Firenze, Italy

AISF

Oct 2017

Organization of a seminar to discuss latest LHCb discoveries

**(1)** 

### Workshops & Conferences

CHEP 2023

ML-INFN & INFN Bari

Norfolk, Virginia (USA)

Jefferson Lab

May 2023

Oral: "The LHCb ultra-fast simulation option, Lamarr: design and validation"

### Third ML-INFN Hackathon: Advanced Level

Bari, Italy

Oral: "Introduction to generative models"

Nov 2022 Ø A ⊕

Oral: "Bayesian hyperparameter optimization"

Ø ⅓ **(** 

Note: Member of the organizing committee for this INFN educational event

Last updated: June, 2023

ACAT 2022 Bari, Italy Oct 2022 University of Bari, Polytechnic University of Bari & INFN Bari Ø A ⊕ Poster: "Lamarr: LHCb ultra-fast simulation based on machine learning models" Ø A ⊕ Poster: "Hyperparameter Optimization as a Service on INFN Cloud" 108° Congresso Nazionale della SIF 🏶 Milan, Italy Italian Physical Society (SIF) Sep 2022 Oral: "ML in the histological differentiation of mediastinal bulky lymphoma" 4th European Congress of Medical Physics 🏶 Dublin, Ireland European Federation of Organisations for Medical Physics Aug 2022 Oral: "ML in the histological differentiation of mediastinal bulky lymphoma" Learning To Discover 🏶 Orsay, France Institut Pascal Paris-Saclay Apr 2022 Oral: "Simulating the LHCb experiment with Generative Models" LPCC Fast Detector Simulation Workshop online Nov 2021 LHC Physics Centre at CERN Oral: "OptunAPI: API to distribute hyperparameters optimization through HTTP requests" 107° Congresso Nazionale della SIF online Italian Physical Society (SIF) Sep 2021 Oral: "Simulating the LHCb detector with GANs" Workshop della Commissione Calcolo e Reti dell'INFN 🏶 online May 2021 INFN Computing and Network Service Oral: "Simulating the LHCb detector with GANs" 1st CloudBank EU Workshop 🏶 online CERN IT & IPT Departments Apr 2021 Oral: "LHCb deployment in AWS" (restricted access)

### SELECTED PUBLICATIONS

- [1] L. Anderlini et al., Lamarr: the ultra-fast simulation option for the LHCb experiment, in 41st International Conference on High Energy Physics – PoS(ICHEP2022), 414 233, 2022
- [2] M. Barbetti, Lamarr: LHCb Ultra-Fast Simulation based on machine learning models deployed within Gauss, in 21st International Workshop on Advanced Computing and Analysis Techniques in Physics Research (ACAT 2022), arXiv:2303.11428
- [3] F. Ratnikov et al., A full detector description using neural network driven simulation, in 15th Pisa Meeting on Advanced Detectors, Nucl. Instrum. Meth. A 1046 (2023) 167591
- [4] L. Anderlini et al., Towards Reliable Neural Generative Modeling of Detectors, in 20th International Workshop on Advanced Computing and Analysis Techniques in Physics Research (ACAT 2021), J. Phys.: Conf. Ser. 2438 (2023) 012130, arXiv:2204.09947
- [5] M. Barbetti and L. Anderlini, Hyperparameter Optimization as a Service on INFN Cloud, in 21st International Workshop on Advanced Computing and Analysis Techniques in Physics Research (ACAT 2022), arXiv:2301.05522

[6] L. Anderlini and M. Barbetti, scikinC: a tool for deploying machine learning as binaries, in Computational Tools for High Energy Physics and Cosmology – PoS(CompTools2021), 409 034, 2022

- [7] LHCb Collaboration, Measurement of antiproton production from antihyperon decays in pHe collisions at  $\sqrt{s_{\mathrm{NN}}} = 110$  GeV, arXiv:2205.09009
- [8] E. M. Abenavoli et al., Characterization of mediastinal bulky lymphomas with FDG-PET-based radiomics and machine learning techniques, Cancers 15 (2023) 1931

# COMPUTER SKILLS

# LANGUAGES

$\mathbf{Git}\mathbf{Hub}$	https://github.com/mbarbetti	Italian	Native
Languages	Python, HTML, $C/C++$ , $TeX$	$\mathbf{English}$	Advanced
$\mathbf{OS}$	Mac OS, Windows, Linux	Spanish	Intermediate