



UNIVERSITÀ  
DEGLI STUDI  
FIRENZE

Scuola di  
Scienze Matematiche  
Fisiche e Naturali  
Corso di Laurea Magistrale  
in Scienze Fisiche e Astrofisiche

**Relatore:**

Dott. Lucio Anderlini  
anderlinil@fi.infn.it

**Correlatore:**

Prof. Piergiulio Lenzi  
piergiulio.lenzi@unifi.it

**Candidato:**

Matteo Barbetti  
matteo.barbetti@stud.unifi.it

**Tecniche di simulazione parametrica con reti neurali profonde e loro implementazione per l'esperimento LHCb al CERN e le sue future evoluzioni**

L'esperimento LHCb è uno dei quattro grandi rivelatori lungo l'acceleratore di adroni, *Large Hadron Collider*, al CERN, ed è dedicato allo studio dei quark pesanti prodotti in collisioni protone-protone. L'obiettivo principale di LHCb è la ricerca di effetti indiretti di fisica oltre il Modello Standard in processi che violano la simmetria  $CP$ , oppure in decadimenti rari dei quark  $b$  e  $c$  rari. Al fine di migliorare la precisione statistica, a partire dal Run 3, LHCb opererà con un *trigger* puramente software che consentirà di accrescere di un ordine di grandezza la mole di dati raccolti. Ciò permetterà misure con precisione mai raggiunta prima, purché la Collaborazione riesca a procurare campioni simulati altrettanto grandi. Con tutta probabilità, la produzione di questi campioni dominerà le richieste di risorse computazionali nelle future evoluzioni dell'esperimento LHCb. Riprodurre accuratamente tutti i processi fisici dalla collisione protone-protone all'interazione radiazione-materia all'interno dei rivelatori (approccio denominato simulazione *completa*) è già oggi insostenibile, così che l'evoluzione verso tecniche di simulazione veloce è indispensabile per trarre vantaggio dagli upgrade del rivelatore. La simulazione *ultra-veloce* richiede un ridotto quantitativo di risorse computazionali, rinunciando a simulare l'interazione radiazione-materia e parametrizzando direttamente la risposta ad alto-livello del rivelatore. I sotto-sistemi di LHCb sono basati su processi fisici molto diversi tra loro che rendono la costruzione di una parametrizzazione un'operazione non semplice: per esempio, il processo di identificazione delle particelle (*Particle Identification*, PID) combina l'informazione di rivelatori a immagine Cherenkov, calorimetri ed un rivelatore per muoni. Le parametrizzazioni possono essere costruite in modo efficace con reti neurali generative addestrate con lo schema a reti antagoniste (GAN). Questi algoritmi consentono infatti di riprodurre distribuzioni di probabilità, anche molto varie, addestrando un modello generativo sui dati.

Ampia parte di questo lavoro di tesi è stata dedicata allo sviluppo e all'implementazione di algoritmi GAN moderni per modellizzare la risposta ad alto livello dei rivelatori per l'identificazione delle particelle a LHCb. Questi algoritmi sono stati addestrati direttamente sui dati di calibrazione raccolti nel 2016 con una procedura di selezione appositamente studiata per evitare distorsioni nelle distribuzioni delle risposte dei rivelatori. In particolare, ho modificato la procedura di apprendimento delle reti per sottrarre statisticamente la componente di fondo irriducibile presente nei campioni di calibrazione e ho sviluppato un algoritmo indipendente in grado di misurare la qualità dei campioni generati. Con lo stesso approccio ho parametrizzato anche le variabili ottenute combinando le risposte dei diversi rivelatori. Il sistema di reti neurali completo permette, date informazioni di natura generica, quali la quantità di moto e l'inclinazione con cui le particelle penetrano nel rivelatore, di sintetizzare tutte le variabili utili a definire criteri di selezione relativi all'identificazione delle particelle cariche. Infine, ho contribuito alla nascita di **mambah**, un pacchetto Python che fornisce strutture di dati utili a rappresentare gli eventi di decadimento in Fisica delle Alte Energie. Tutte le strutture definite in **mambah** sono progettate per processare *batch* di eventi in parallelo, traendo profitto dall'utilizzo di acceleratori hardware quali GPU ed FPGA. Nel contesto del progetto di **mamba** ho contribuito all'implementazione delle funzioni di gestione dei database relazionali e al progetto del modulo di **mambah** dedicato alla simulazione, chiamato **mambah.sim**. Il modulo **mambah.sim** consente di simulare selettivamente alcune particelle prodotte nelle collisioni protone-protone per poi propagarle attraverso il rivelatore attraverso parametrizzazioni delle funzioni di efficienza e risoluzione. Ho integrato in **mambah.sim** le parametrizzazioni da me costruite per i rivelatori di identificazione delle particelle a LHCb, dimostrandone le eccellenti proprietà di generalizzazione a decadimenti diversi da quelli usati durante la fase di addestramento delle reti. Infine, ho mostrato come i campioni prodotti con **mambah.sim** siano competitivi con i campioni ottenuti con Simulazione Completa, pur assicurando una riduzione significativa dei costi in termini di risorse computazionali.