

# Event Selection & Machine Learning for Detecting SUSY Events in CERN Collision Data

Khusanjon Bobokhojaev • Physics • California Lutheran University  
kbobokhojaev@callutheran.edu • Advisor: Dr. Sebastian Carron Montero

## Abstract

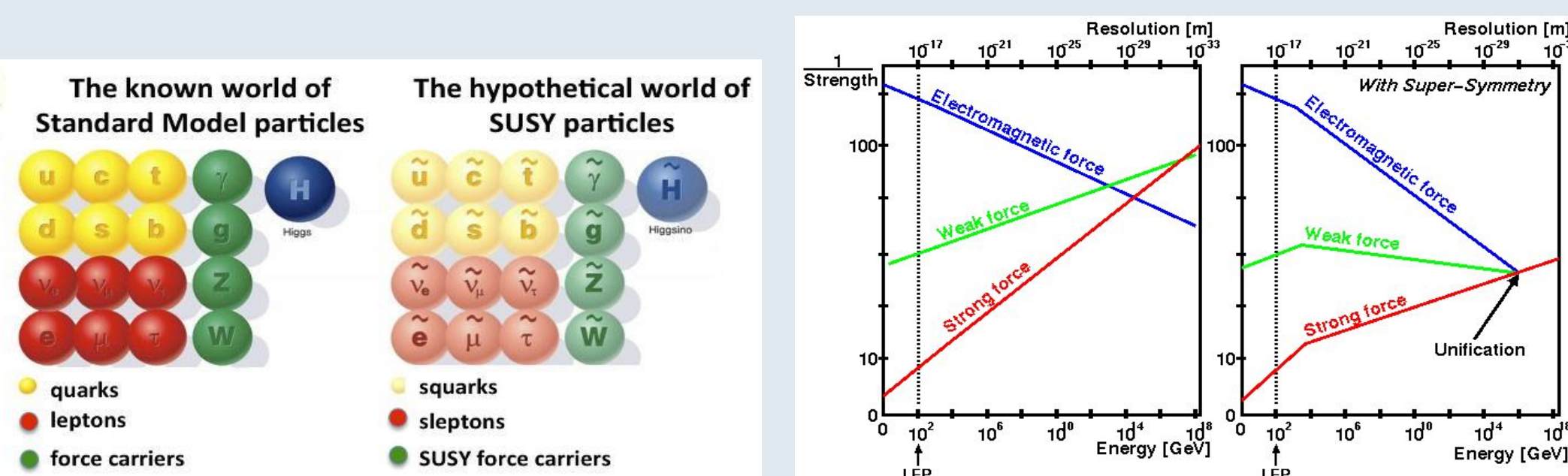
In the ongoing search for Supersymmetric (SUSY) particles, this project focuses on identifying potential signatures of the Supersymmetric Top Quark, or stop quark ( $t^*$ ), produced in proton-proton collisions at the Large Hadron Collider (LHC) at CERN. SUSY provides a theoretical framework extending the Standard Model, offering both a path toward unifying the fundamental forces and a compelling dark matter candidate—the lightest supersymmetric particle (LSP).

Earlier stages of this research developed a CMSSW-based analysis framework using FWLiteHistograms to process simulated CMS collision data through di-lepton decay channels—electron-electron ( $ee$ ), muon-muon ( $\mu\mu$ ), and electron-muon ( $e\mu$ ). Preselection cuts on kinematic variables such as missing transverse energy (MET), invariant mass, and lepton isolation were optimized to separate SUSY stop signals from dominant background processes including  $t\bar{t}$ , WZ, WW, and Drell-Yan events.

Building on that foundation, the current phase transitioned the analysis to a NanoAOD-based workflow utilizing the Coffea library and Awkward Arrays in a Jupyter environment, enabling faster and more scalable event processing. New b-tagging algorithms, jet energy scale (JES) corrections, and MET resolution refinements were implemented, alongside comprehensive systematic studies. The team is now awaiting updated simulation samples to validate the refined selection criteria and to train a boosted decision tree (BDT) classifier for improved SUSY signal discrimination. This work contributes to the broader effort to uncover evidence for supersymmetry—an essential step toward unifying the fundamental forces and identifying the nature of dark matter.

## Background

The **Standard Model** describes the elementary particles and their interactions through three fundamental forces: electromagnetic, weak, and strong. It is the most successful current theory to describe how particles make up all known matter. However, it fails to incorporate gravity and lacks a unified description of these forces. The theory of **Supersymmetry (SUSY)** elegantly appends the Standard Model by introducing partner particles for each fermion and boson [1-4].

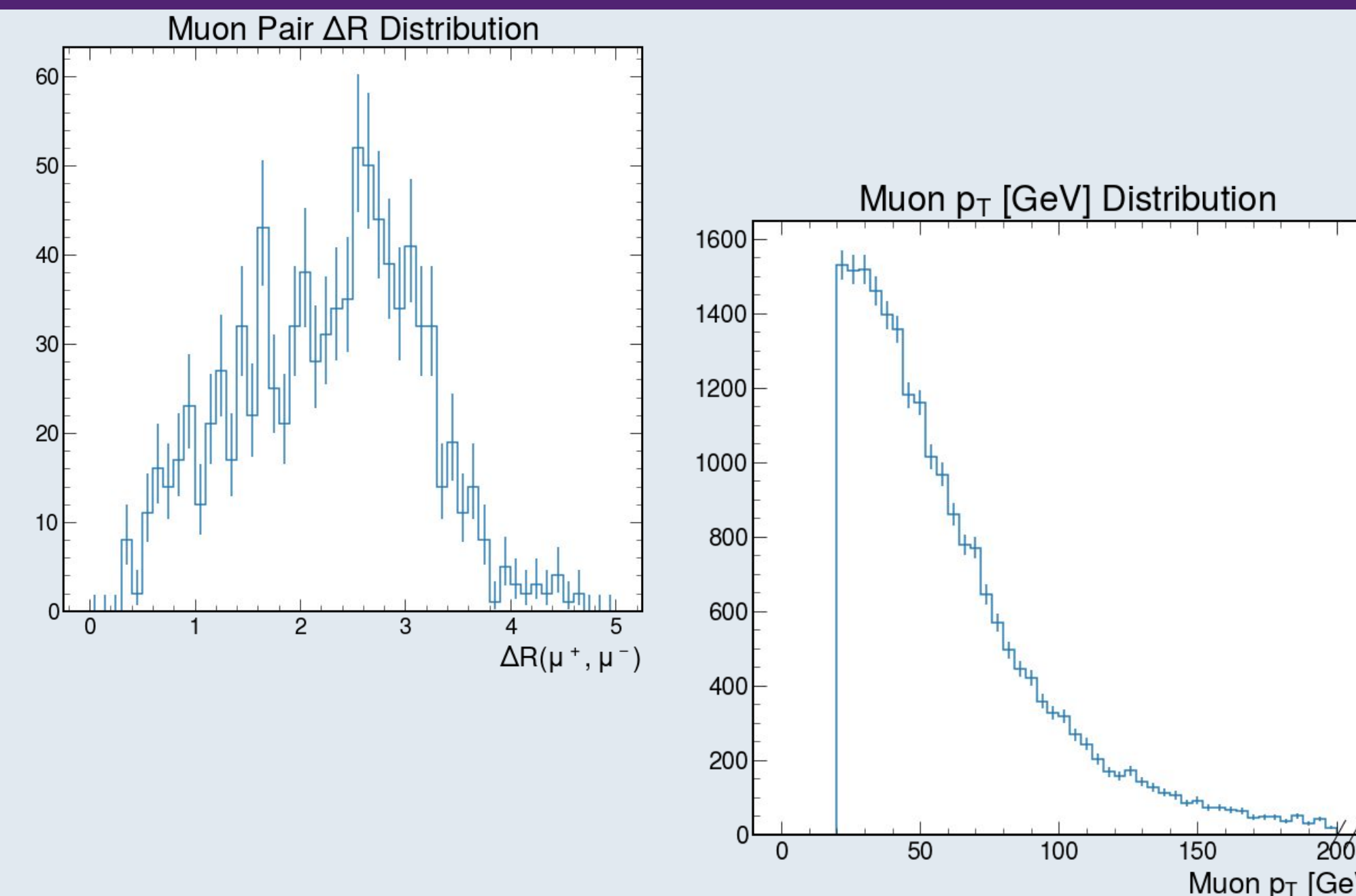


### Past Summer Work:

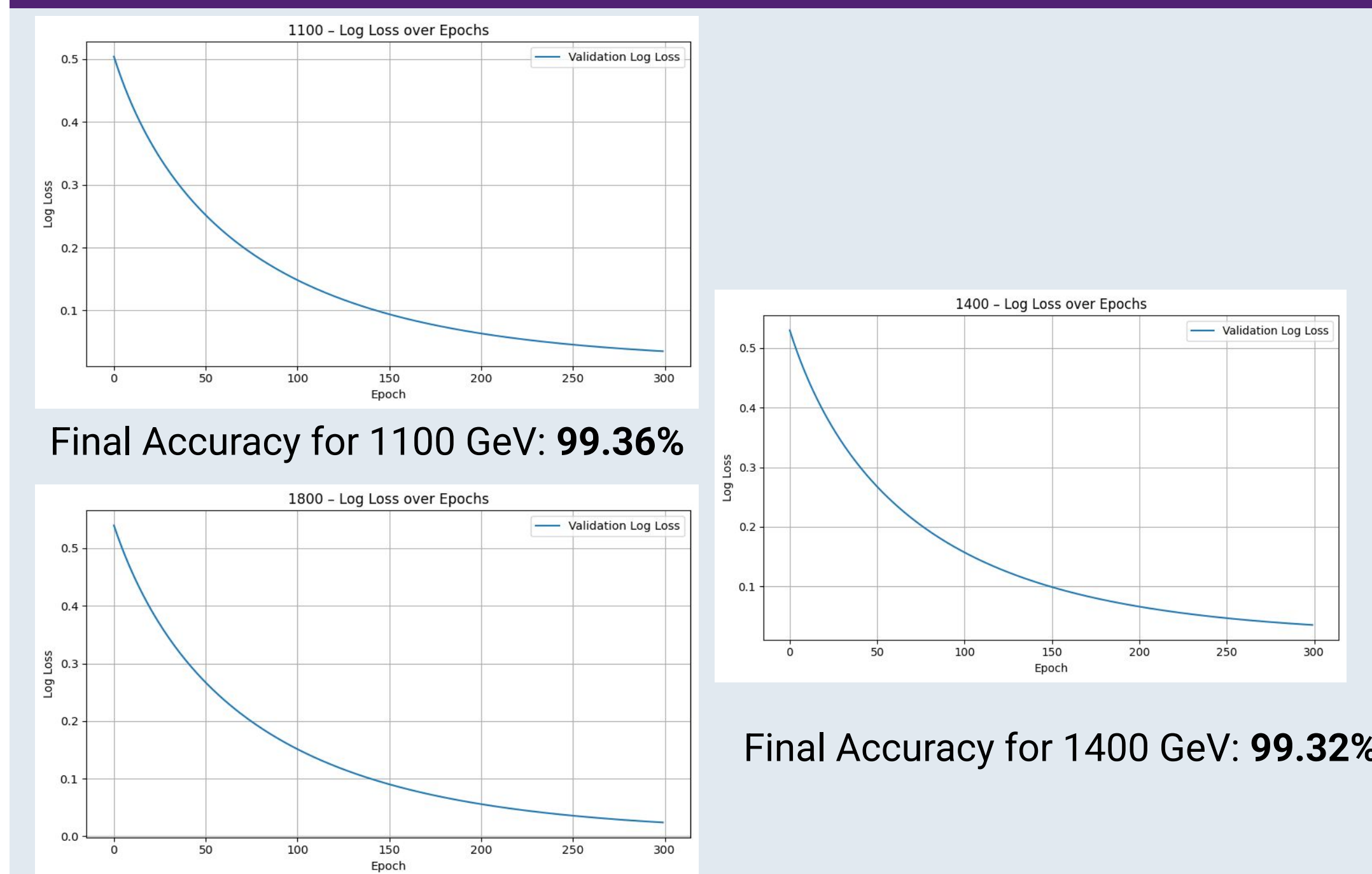
- Implementation of  $\mu\mu$  &  $e\mu$  Channels
  - The search for supersymmetric top quarks conducted was in the  $ee$ ,  $\mu\mu$ ,  $e\mu$  di-lepton decay channels.
    - Previous students had created the  $ee$  decay channel filtration.
    - Worked on the addition of the  $e\mu$  and  $\mu\mu$  channels.
      - Included studies for deltaR between muons, isolation of particles ("isolationR03" - cone radius of 0.3), dimuon transverse mass, collinearity of selected muons, matching event data muons to generator muons, and the corresponding for the  $e\mu$  channel.
    - After the slimming and filtering of the event data selections, the data was converted into "ntuple" files that could be fed into the ML algorithm.
- Machine Learning Algorithm
  - Neural Network Training:** Processed data was used to train a neural network with an 80:20 train-test split.
  - Training Variables:** Focused on distinguishing SUSY from Drell-Yan, WW, WZ, ZZ, and other backgrounds.
    - All 3 channels were trained with inputs as background and signal, and the data groups given were as follows:
      - Pt (Transverse Momentum), Eta, Phi, Vxy, NJets, missET
        - ML was trained on these parameters with each combination of SUSY and background processes.

## Optimization of Event Selection

- Monte Carlo Generator-Level Studies
  - Matched reconstructed muons to generator-level muons**
    - Counted how many muons came from W bosons in SUSY &  $T\bar{T}$ , and other background sources
    - Investigated "muons from muons" cases - found original particle mothers in decay process and filtered data
    - Extended decay categories to account for two muons in each event
- Variable Optimization
  - pT thresholds tested: 15, 20, 25, 30, 35 GeV
  - Number of quark jets (nj vs ak8jets): comparison and plotting
  - $\Delta R(\mu^+, \mu^-)$  cuts representing angle between muons



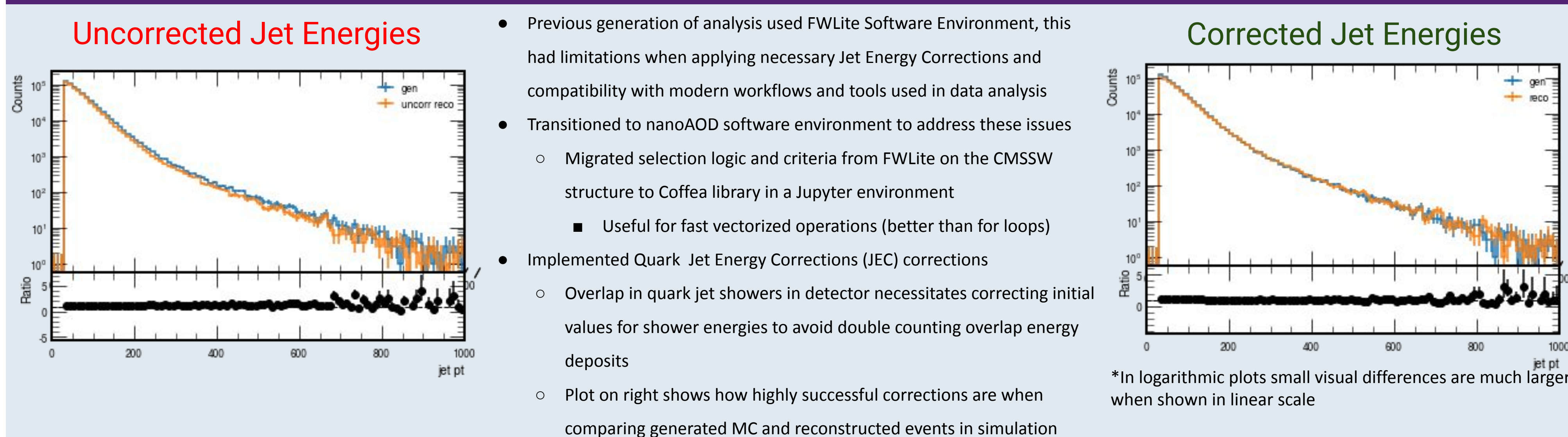
## Machine Learning: Boosted Random Forests



Final Accuracy for 1100 GeV: **99.36%**  
Final Accuracy for 1400 GeV: **99.32%**  
Final Accuracy for 1800 GeV: **99.81%**  
Testing Error as Function of ML Epoch, and Identification Accuracy for each assumed Stop Quark Mass

- Improving Particle Event Selection and Sorting
  - Modified dilepton selection in ntuple, so that we preferentially pick highest pT leptons
  - Optimized output parser program running over FWLite output files
- Machine Learning & Dimensionality Reduction
  - From originally used neural network model → **transitioned to Boosted Random Forest**
  - Principal Component Analysis applied** to analyze relative impact of model input variables
    - Removed lower impact variables to save computation time
  - GPU boosted **hyperparameter grid search to optimize model**
    - Achieved above 99.3% in final identification testing accuracy across all Stop Quark Assumed Masses**

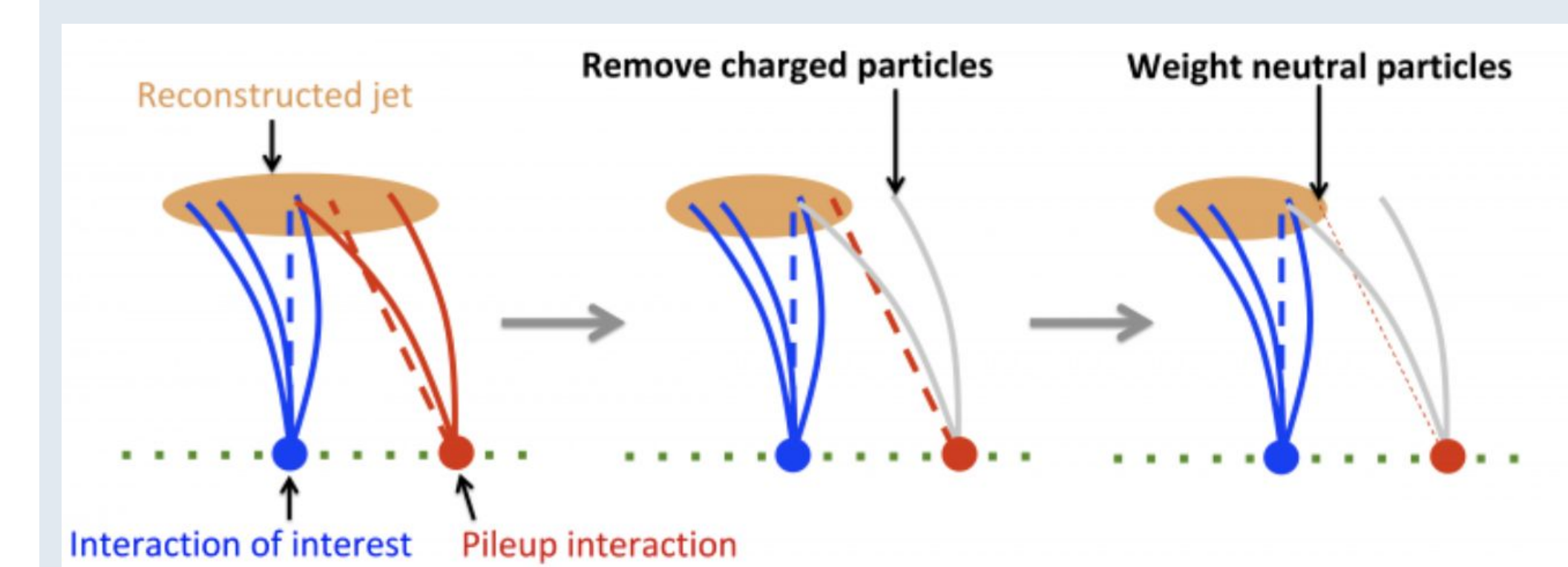
## Transition to nanoAOD



- Previous generation of analysis used FWLite Software Environment, this had limitations when applying necessary Jet Energy Corrections and compatibility with modern workflows and tools used in data analysis
- Transitioned to nanoAOD software environment to address these issues
  - Migrated selection logic and criteria from FWLite on the CMSSW structure to Coffea library in a Jupyter environment
    - Useful for fast vectorized operations (better than for loops)
- Implemented Quark Jet Energy Corrections (JEC) corrections
  - Overlap in quark jet showers in detector necessitates correcting initial values for shower energies to avoid double counting overlap energy deposits
  - Plot on right shows how highly successful corrections are when comparing generated MC and reconstructed events in simulation

## Current & Future Steps

- Current Steps:
  - Implementing b-quark identification algorithm to help remove Quantum Chromodynamic (QCD) background events
  - Further refine selection by studying lepton collinearity with jets within new nanoAOD data structure
  - Planning to implement corrections to Missing Transverse Energy (MET) that arises from recalculation of total energy after jet corrections
  - Apply selection criteria to updated Run 3 Monte Carlo simulations



- Future Steps
  - Most important studies are on expected sources of systematic error for final results, including:
    - Particle misidentification effects
    - Initial/Final quark state radiation correction effects
    - Collision Parton Distribution function effects
  - Present results at CERN

## Acknowledgements

This research was supported by the contributions of several individuals and organizations. Appreciation is extended to CERN and the SUSY group at CERN for providing access to data and resources. Thank you to Ryan Hoang, a former research student, for his work on the event selection code. Thank you to Carys Garvey who participated in team effort in the previous summer. Gratitude goes to Dr. Sebastian Carron Montero, advisor and professor, for his mentorship throughout this project. As well as the past funders - OURCS and VITAL foundations that funded this research project, and also the funder for this summer - ALLIES Program.

## References

- A. M. Serra, "The Standard Model," University of Zurich. Available: <https://www.physik.uzh.ch/groups/serra/StandardModel.html>
- MS Collaboration, "Seeking SUSY," CERN. Available: <https://cms.cern/physics/seeking-susy>
- CMS Collaboration, "Jets at CMS and the determination of their energy scale," CERN. Available: <https://cms.cern/physics/jets-determination-energy-scale>
- CMS Collaboration, "Jets and MET uncertainties," CERN. Available: <https://cms-opendata-guide.web.cern.ch/analysis/systematics/objectsuncertain/jetmetuncertain/>
- H. E. Haber and G. L. Kane, "The search for supersymmetry: probing physics beyond the standard model", Phys. Rept. 117 (1985) 75, doi:10.1016/0370-1573(85)90051-1.
- S. P. Martin, "A supersymmetry primer", in Perspectives on supersymmetry. Vol.2, G. L. Kane, ed., volume 21, p. 1. 2010. arXiv:hep-ph/9709356. doi:10.1142/9789812839657\_0001.