

# Kinematic Reconstruction of Top Quarks Using CMS Open Data: Development and Validation of Kinematic Reconstruction Methods for Top-Quark Pairs in the Dilepton Decay Channel

**Applicant:** Spitsyn Mykyta  
**Project duration:** 12 Weeks

**Mentor:** Oleksandr Zenaiev, Hamburg University  
**Proposed start date:** 1 July 2026

## Introduction and Project Description

**Motivation.** The Compact Muon Solenoid (CMS) is a general-purpose detector at the CERN Large Hadron Collider (LHC) [1]. The top quark is one of the most important particles in the Standard Model. Due to its exceptionally large mass, it plays a unique role in electroweak symmetry breaking and provides a sensitive probe of potential physics beyond the SM. Furthermore, precise knowledge of the  $t\bar{t}$  invariant mass distribution,  $M(t\bar{t})$ , is important both for top-quark pole mass measurements and for searches for heavy resonances beyond the SM. In the dilepton decay channel,  $t\bar{t} \rightarrow W^+bW^- \bar{b} \rightarrow \ell^+\nu b \ell^-\bar{\nu}\bar{b}$ , in addition to the two-lepton signature, two undetected neutrinos are produced, making measurements of the  $t\bar{t}$  kinematic distribution especially challenging. Accurate reconstruction methods are therefore essential for precision measurements and searches for new phenomena in top-quark production.

**Kinematic Reconstruction Algorithms.** The reconstruction procedure aims to achieve both high resolution of the observable of interest and high reconstruction efficiency for  $t\bar{t}$  events. Several reconstruction techniques have been developed and employed by the CMS experiment.

*Full Kinematic Reconstruction (FKR)* [3] reconstructs the individual top-quark decays by imposing the on-shell constraints on the masses of  $W$  bosons,  $t$  and  $\bar{t}$  quarks, leading to a system of fourth-order polynomial equations for the neutrino longitudinal momenta. Although computationally demanding and limited by ambiguities arising from multiple or absent solutions, FKR remains attractive due to its physically interpretable reconstruction of the individual  $t$  and  $\bar{t}$  quarks and its potential for future ML-assisted extensions.

*Loose Kinematic Reconstruction (LKR)* [4] instead reconstructs the full  $t\bar{t}$  system as a single object, approximating the longitudinal momentum and energy of the neutrino pair using the measured lepton system. Recently, this method was improved to *LKRv3* [5] by rescaling the system energy and longitudinal momentum while preserving rapidity. This substantially improves the reconstruction accuracy of  $M(t\bar{t})$ , especially in the threshold production region.

*ML-based reconstruction methods* have also been explored. Such methods can exploit multidimensional correlations between reconstructed objects and generator-level information, potentially providing significant improvements in resolution and stability. However, their use in top-quark analyses has so far remained rather limited [4].

**Proposed Work.** In the proposed project, LKRv3 will serve as the baseline for the development of a new generic ML-based reconstruction framework for dileptonic  $t\bar{t}$  events. For this project, the Run 1 dataset collected at a center-of-mass energy of 7 TeV will be used, which is publicly available through the CERN Open Data Portal [2], enabling community-driven studies. This dataset is available together with the corresponding Monte Carlo (MC) simulations of  $t\bar{t}$  production and various background processes.

The project will include the following studies:

1. Development and training of a Multi-Layer Perceptron (MLP) architecture using various input configurations: only raw kinematic variables; raw variables combined with the LKRv3 output; and raw variables combined with both LKRv3 and FKR outputs.
2. Extension of the best-performing MLP architecture to probabilistic and graph-based approaches, including a Normalizing Flow head and, as a further extension, a Graph Neural Network (GNN). The performance of all models will be compared with the FKR, LKR, and LKRv3 baselines.
3. Generation of additional Monte Carlo samples to study model dependence, systematic effects, and possible biases induced by physics beyond the SM.
4. Measurement of the  $M(t\bar{t})$  differential cross sections using the improved reconstruction method, followed by comparisons with published CMS measurements.

The ultimate goal of the project is the development of a robust and generic ML-based reconstruction approach for dileptonic  $t\bar{t}$  events that could be adopted in future analyses by the LHC experiments and potentially form the basis of a dedicated publication.

## Preliminary Timeline

**Week 1.** Literature review on differential  $t\bar{t}$  cross-section measurements and existing kinematic reconstruction techniques, and setup of the analysis framework.

**Week 2.** Validation and comparison of the existing  $t\bar{t}$  kinematic reconstruction methods using CMS Open Data and MC simulations.

**Weeks 3–4.** Development and implementation of a baseline MLP architecture. Training and evaluation of the model using different input feature configurations.

**Weeks 5–8.** Extension of the reconstruction framework to architectures based on Normalizing Flows and exploration of additional ML approaches, including GNN. Quantification of reconstruction performance relative to the MLP baseline and conventional methods.

**Weeks 9–10.** Measurement of differential  $t\bar{t}$  cross sections using the best-performing reconstruction model and cross-validation against published CMS results.

**Weeks 11–12.** Finalization of the analysis framework, preparation of the final report and documentation, cleanup and publication of the analysis code repository.

## References

- [1] CMS Collaboration, *The CMS experiment at the CERN LHC*, JINST **3** (2008) S08004.
- [2] CMS Collaboration, *CERN Open Data Portal*, <https://opendata.cern.ch>
- [3] L. Sonnenschein, *Analytical solution of  $t\bar{t}$  dilepton equations*, Phys. Rev. D **73** (2006) 054015 [arXiv:hep-ph/0603011].
- [4] CMS Collaboration, *Review of top quark mass measurements in CMS*, Phys.Rept. **1115** (2025) 116–218 [arXiv:2403.01313].
- [5] M. Spitsyn, V. Aushev, O. Zenaiev, *Improvement of the loose kinematic reconstruction method of  $t\bar{t}$  using CMS open data*, KINR Annual Workshop HEP-TEC-2026, Kyiv, Ukraine (January 2026).