Refining CMS FastSim with ML-based regression

ML4Jets 2024-LPNHE, Paris, France

6 November, 2024

Sam Bein, Dorukhan Boncukcu, Acelya Güngördü, Kevin Pedro, Moritz Wolf, Mirac Vural
on behalf of the CMS Collaboration

samuel.bein@cern.ch

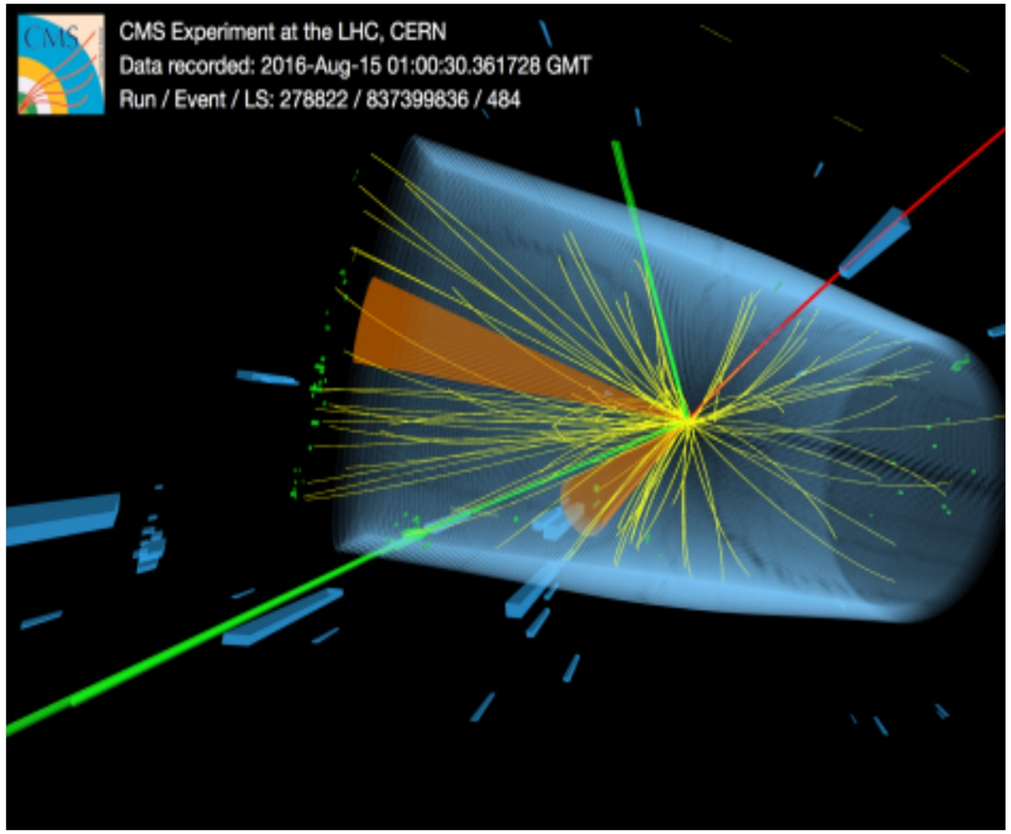
Two simulation engines

FullSim chain

- Detailed **Geant4**-based detector simulation customised for CMS, digitisation
- Full reconstruction as applied to data

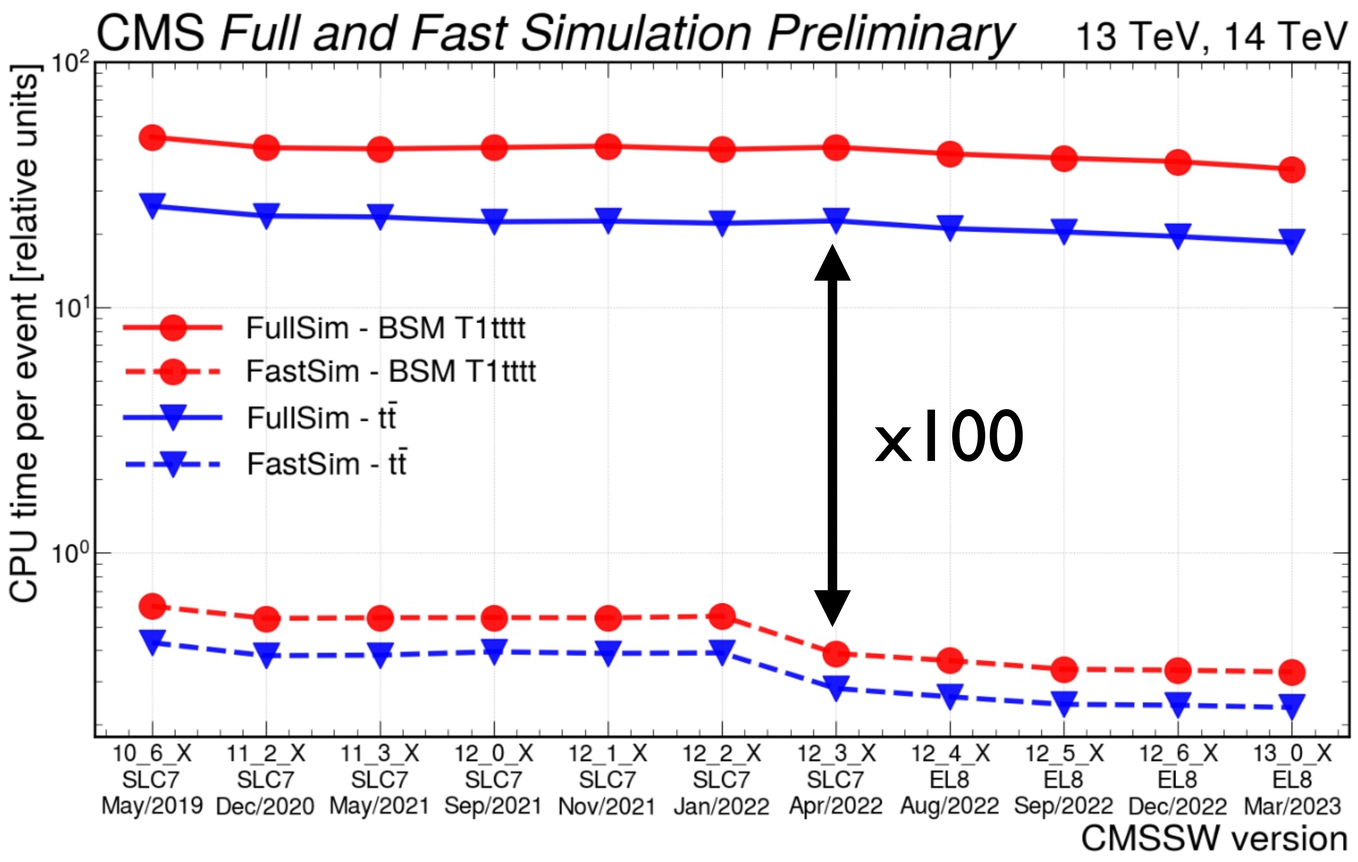
FastSim chain

- Fast **alternative** to FullSim
- Fast particle propagation, analytical interaction models, tracking

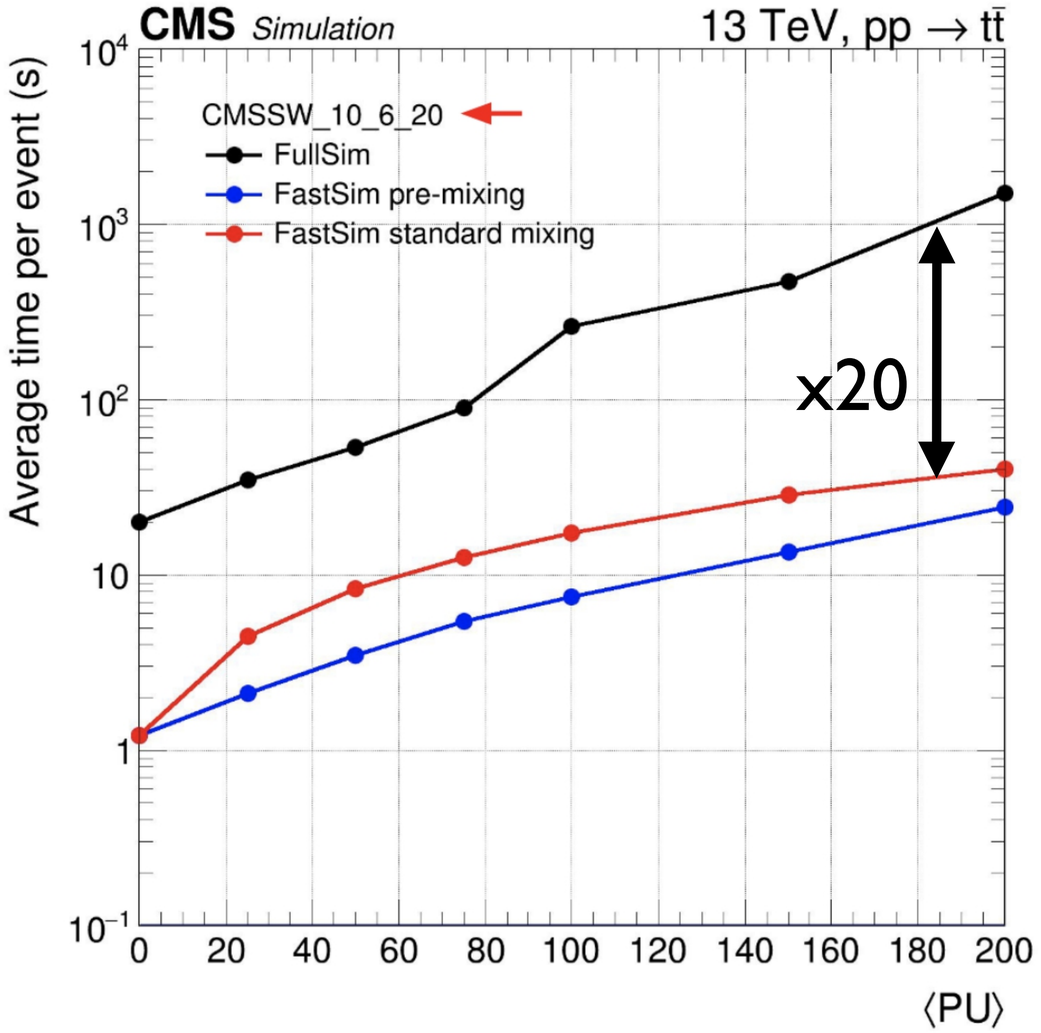


Events/second FastSim and FullSim

Detector step

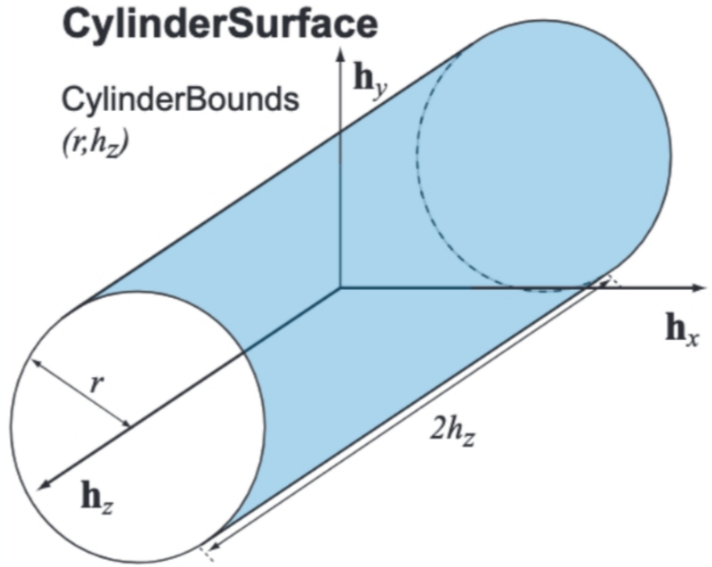


Complete chain



https://twiki.cern.ch/twiki/bin/view/CMSPublic/CMSOfflineComputingResults

FastSim simplified propagation

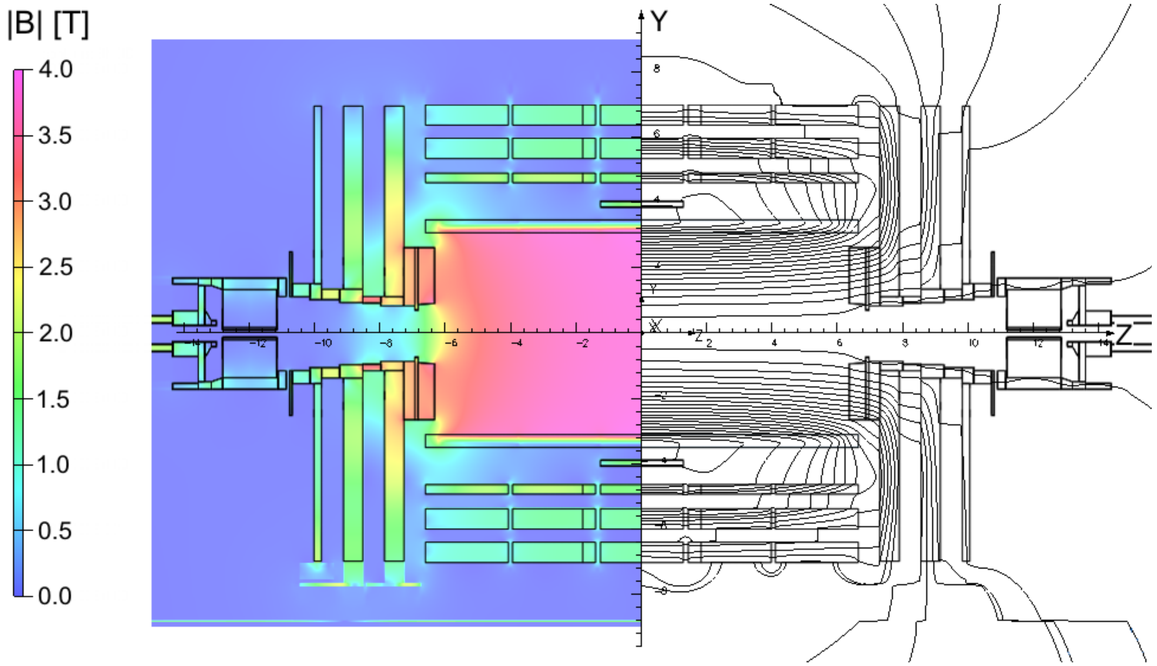


Geometry

- Infinitesimally thin layers
- Iterative tracking run over hit subsets

Transport

- x5 speed-up in reconstruction
- Underestimates fake-rate, biases b-tagging

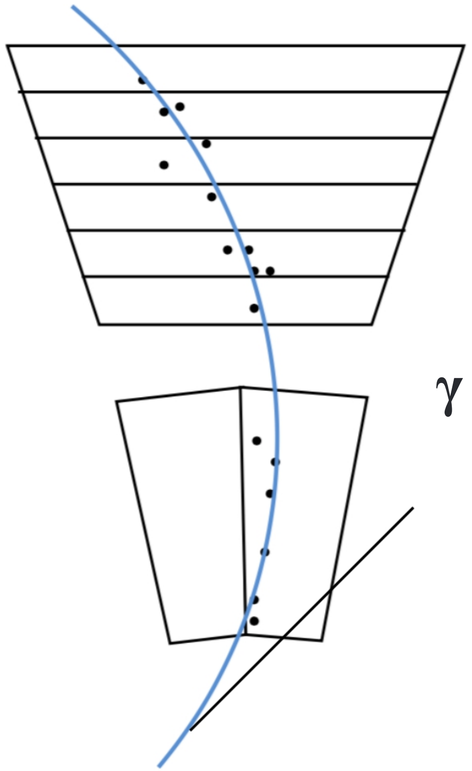


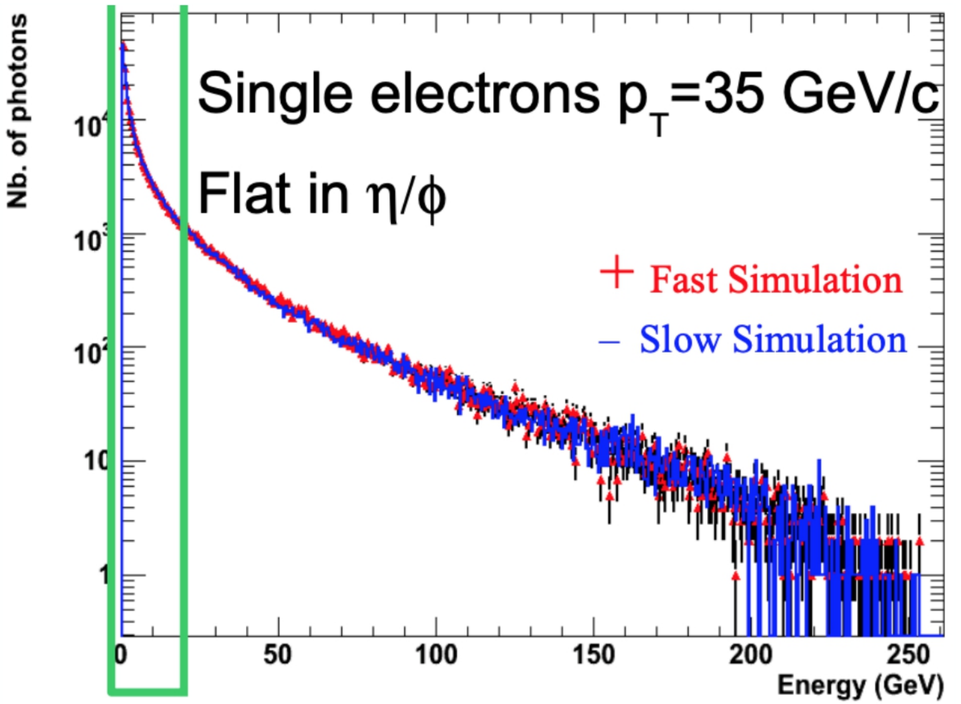
https://cds.cern.ch/record/1215500/plots

Analytical interaction models

Material interactions

- Energy loss by ionization
- Multiple coulomb scattering
- Bremsstrahlung
- $e^{\pm}$ conversion
- Elastic, inelastic nuclear interactions



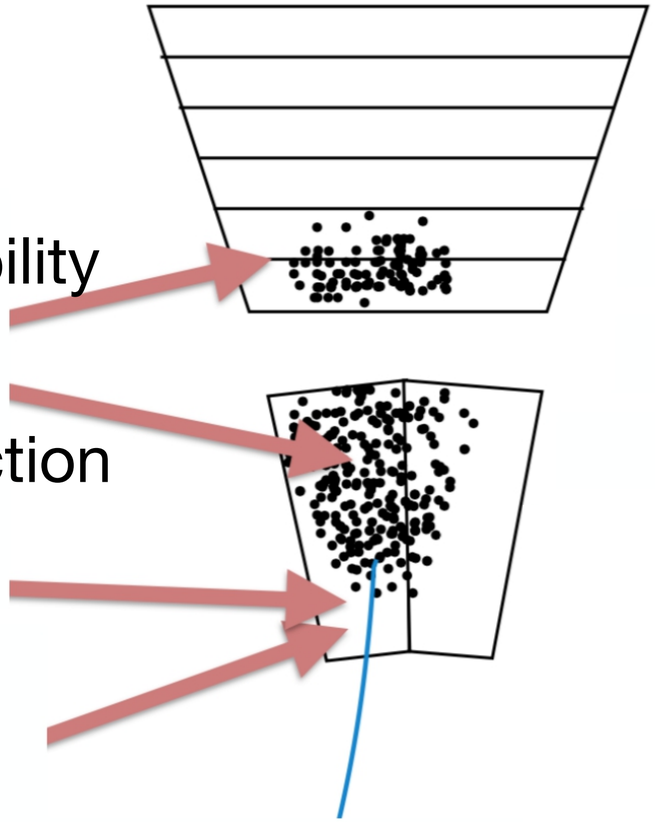


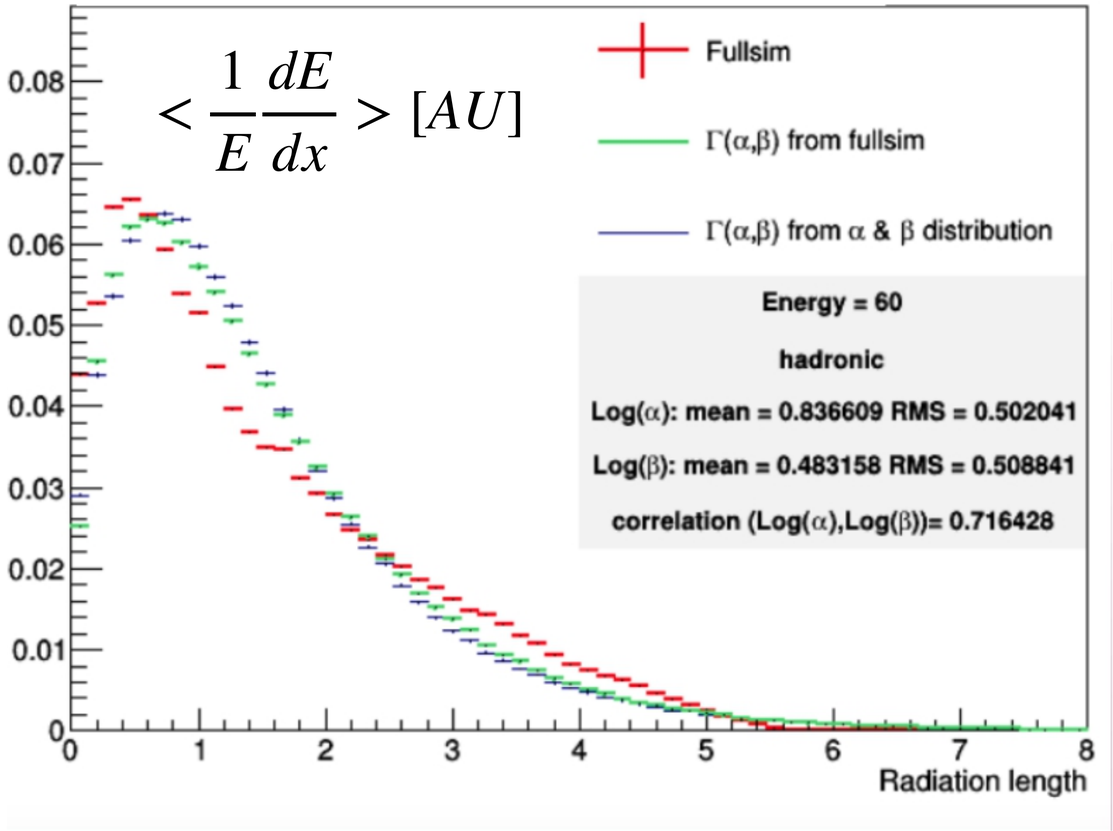
FastSim-CERN workshop

Calorimetry with GFLash

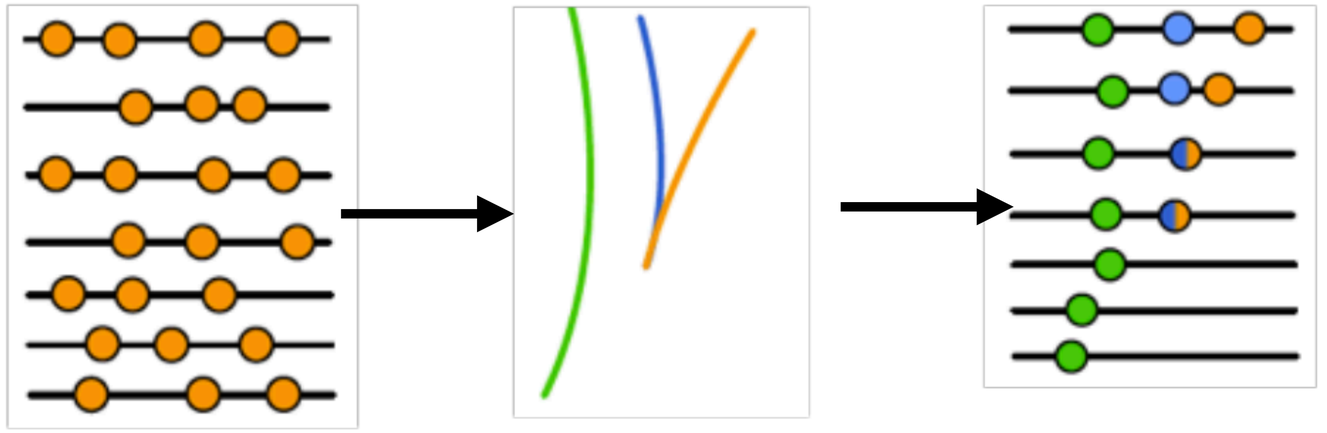
Particles reaching calorimeter

- Shower depositions modeled with probability distributions
- Start position as function of radiation lengths
  - GFlash - cite Grindhammer





Fast Reconstruction

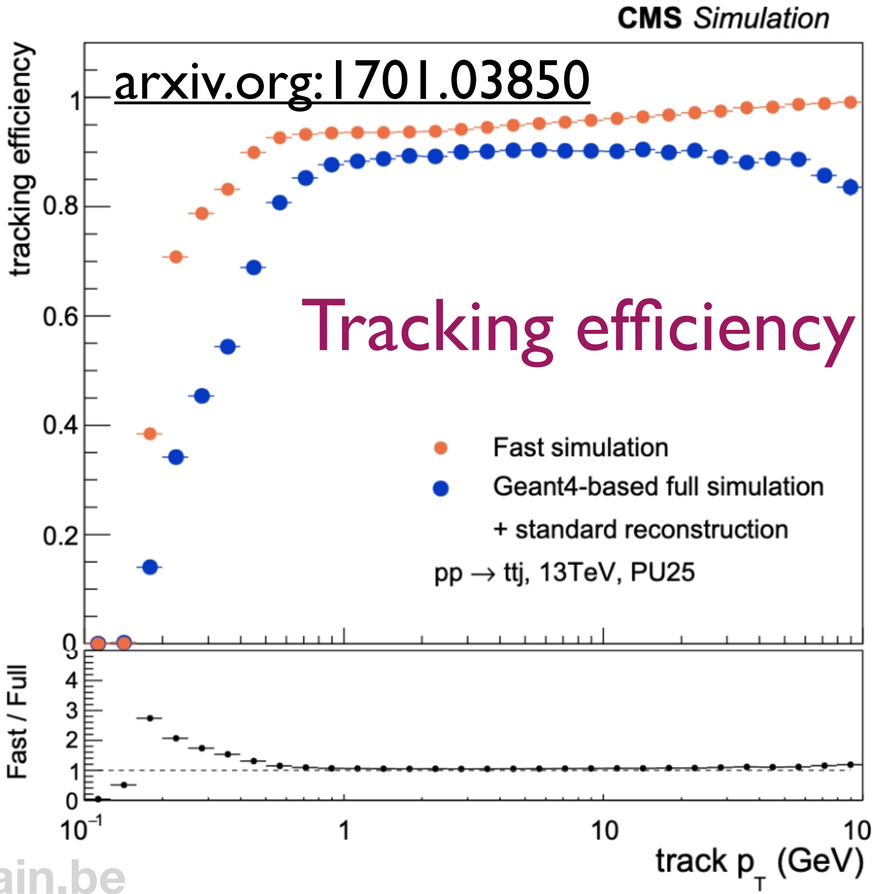


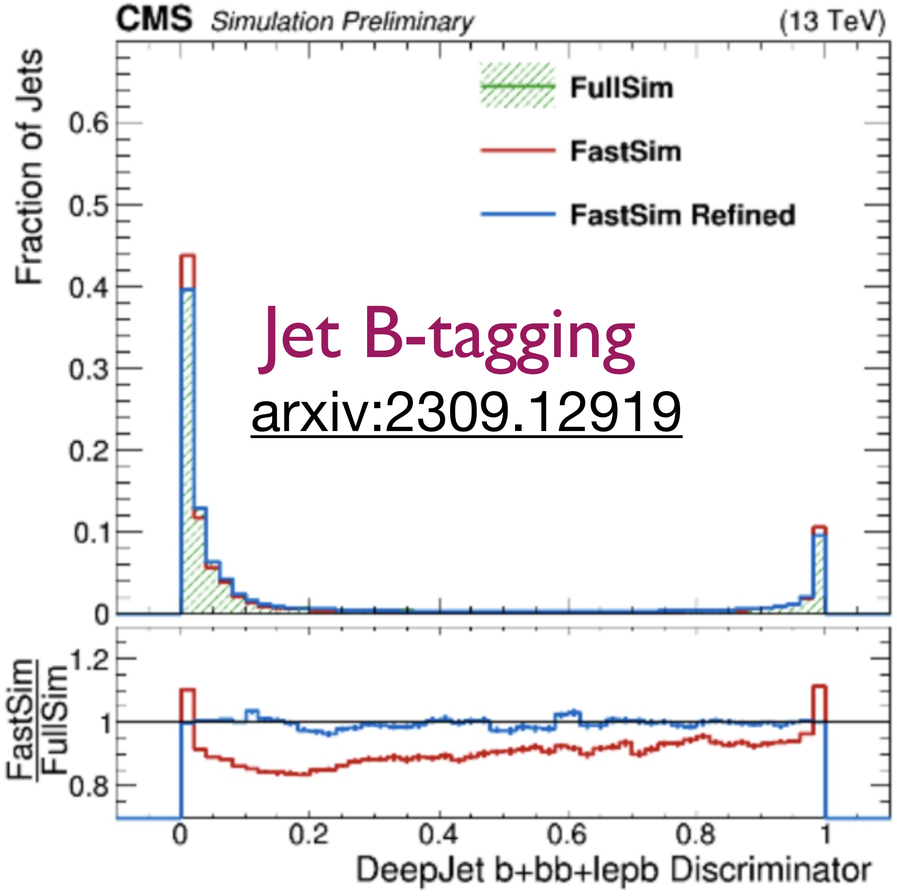
Fast tracking

- GEN-assisted hit/track association
- Iterative tracking run over hit subsets

Impact on performance

- x5 speed-up in reconstruction
- Underestimates fake-rate, biases b-tagging

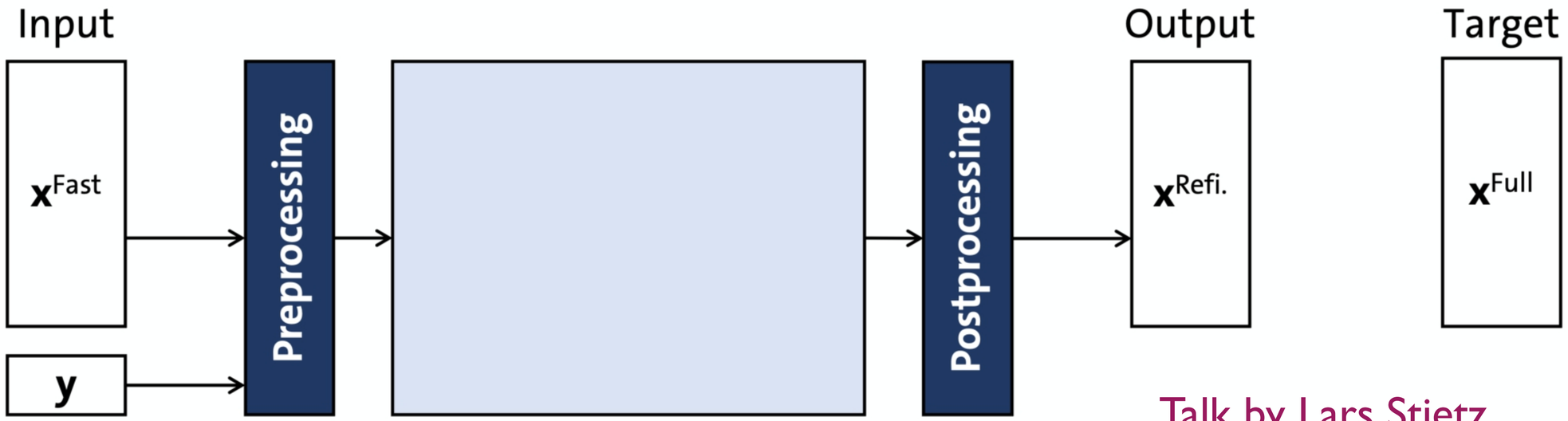




Regression based refinement

Ansatz

- Apply ML while leveraging maximum domain knowledge to yield accurate, detailed fast simulation
- Tweak the final output of the FastSim so that it better matches FullSim
  - *Fast Perfekt: regression-based refinement of FastSim* - arXiv:2410.15992 arxiv:2309.12919

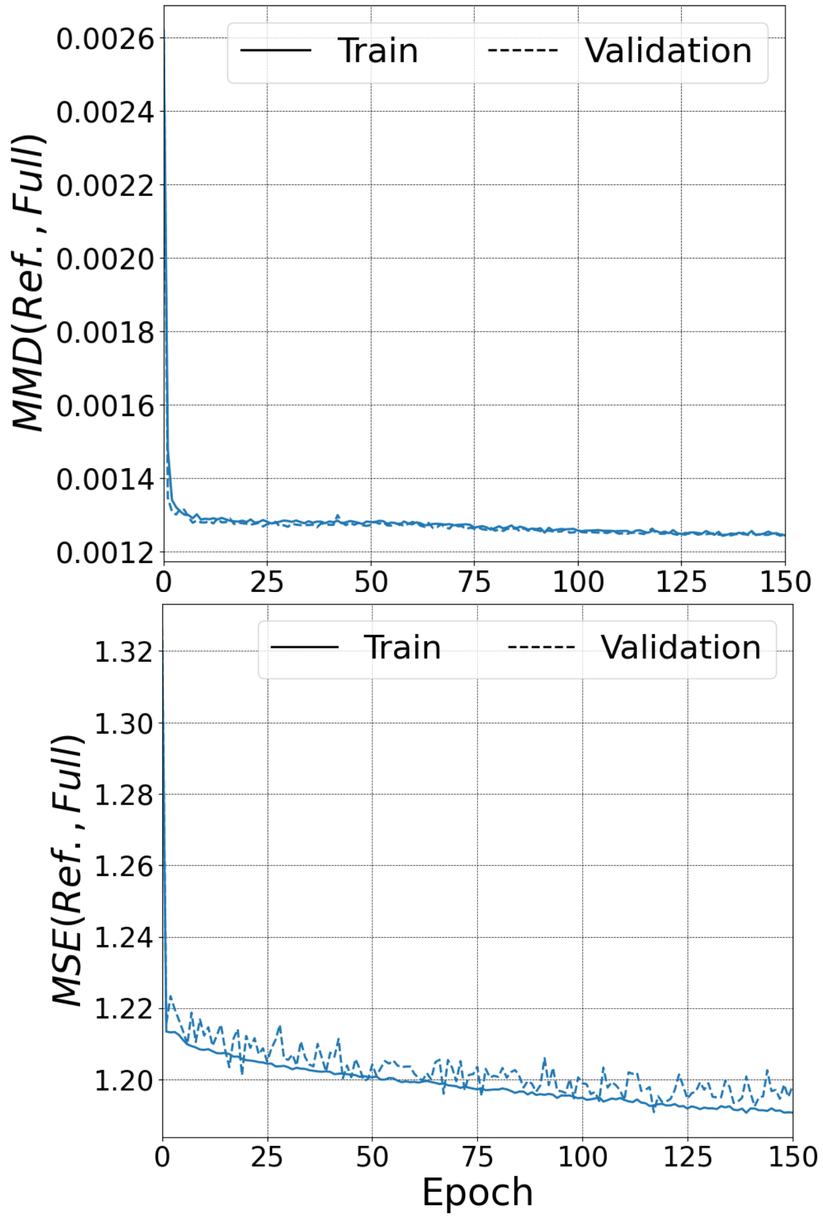


Talk by Lars Stietz

Regression based refinement

Apply Fast Perfekt

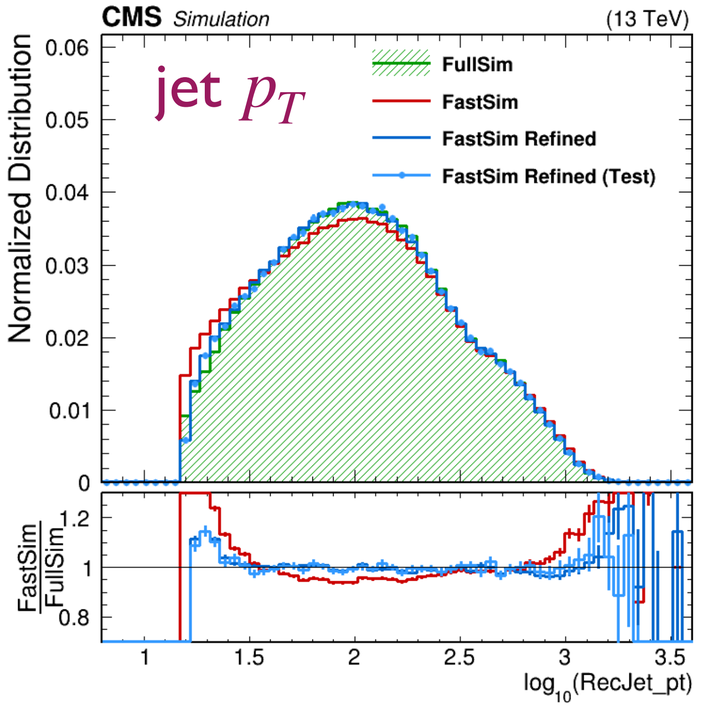
- Refine jets with regression NN
  - Kinematics, Flavour tagging variables
  - DeepJet, ParT Jet discriminators:
    - B, CvB, CvL, QG
  - Jet triplets: FastSim, FullSim, GEN
  - **Input**: FastSim jet+GEN jet p_T, η
  - **Output**: refined FastSim jet
- Single-stage MMD-based training

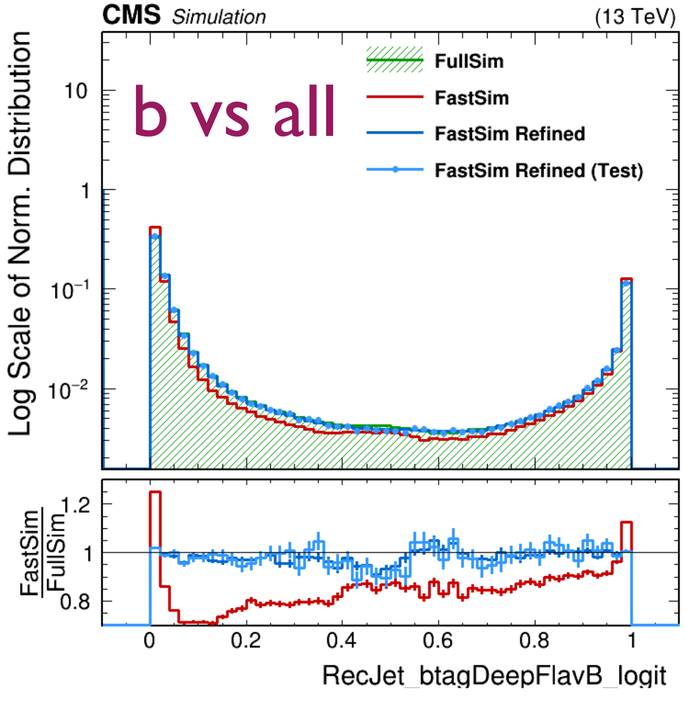


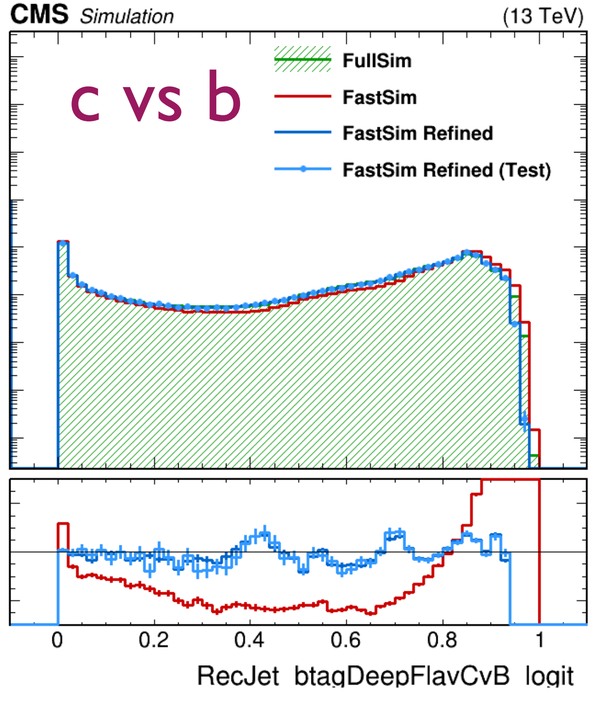
Regression based refinement

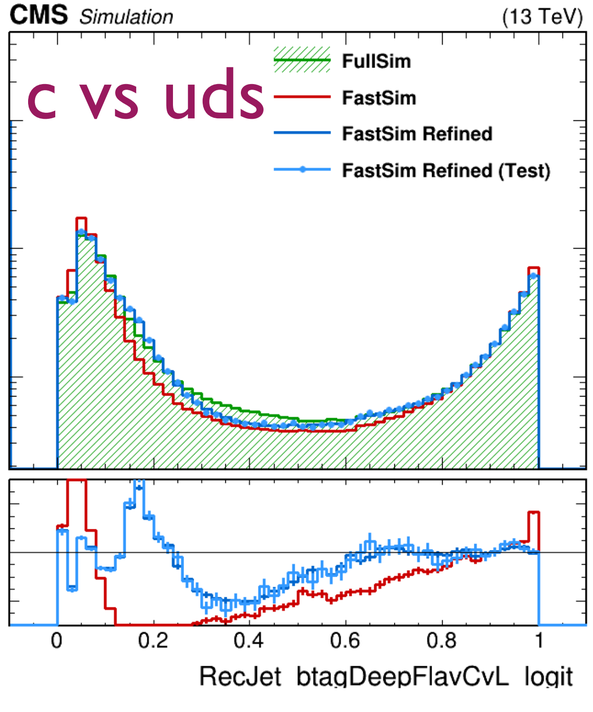
Comparing distributions

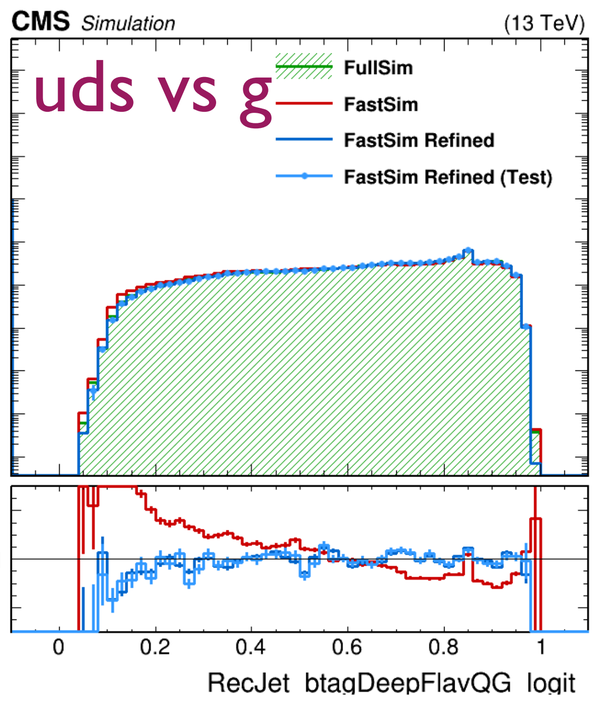
- Kinematics - pT
- Jet flavour taggers (DeepJet)







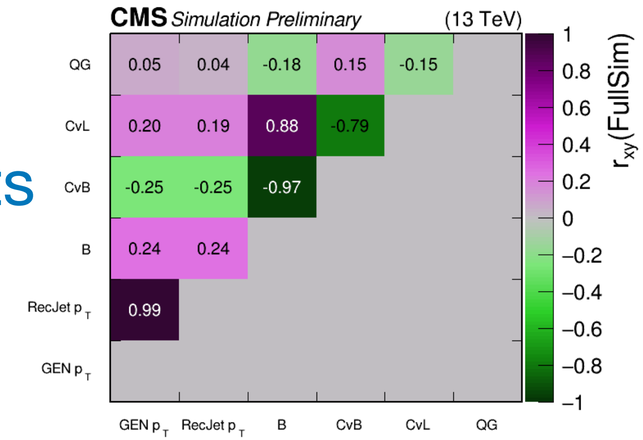




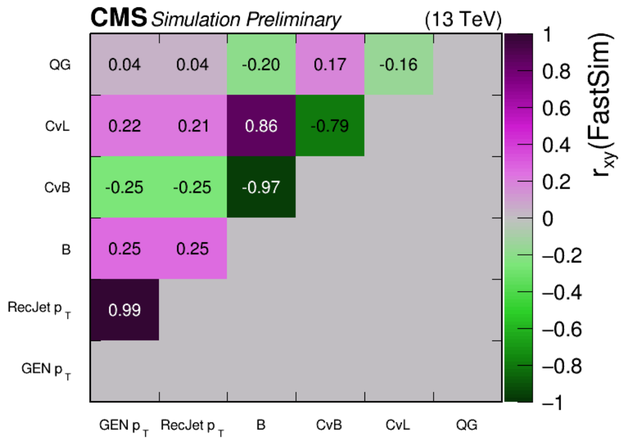
Refined correlations

Pearson coefficients

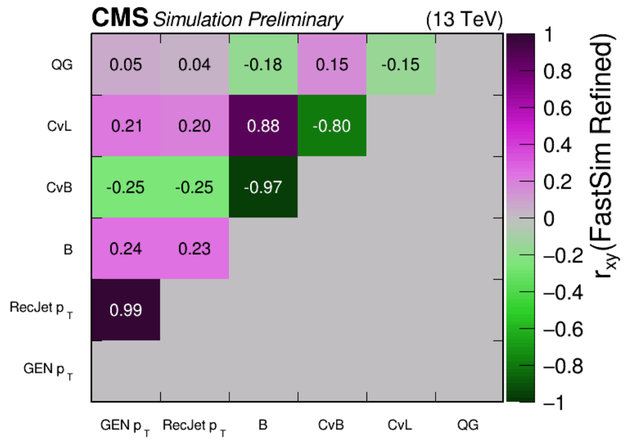
FullSim



FastSim

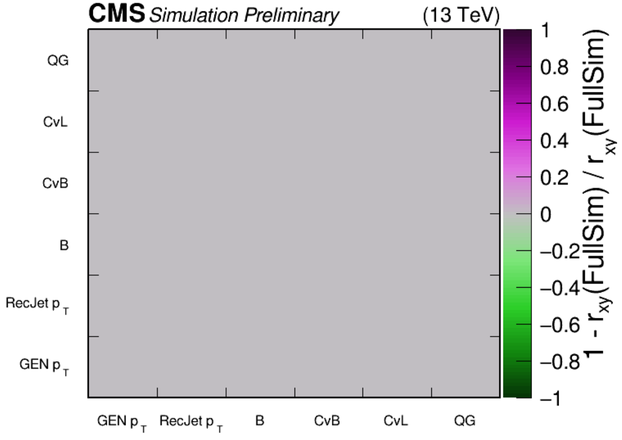


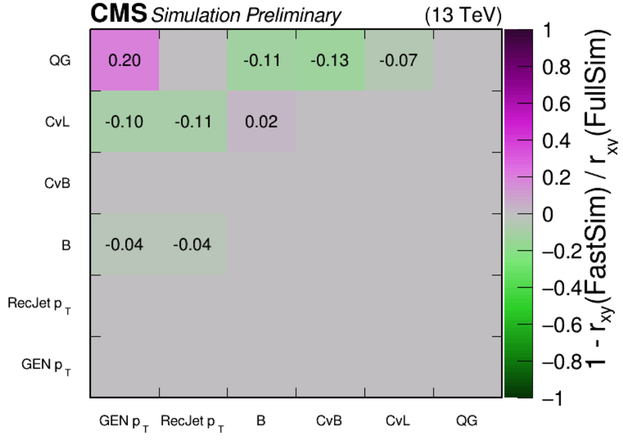
Refined FastSim

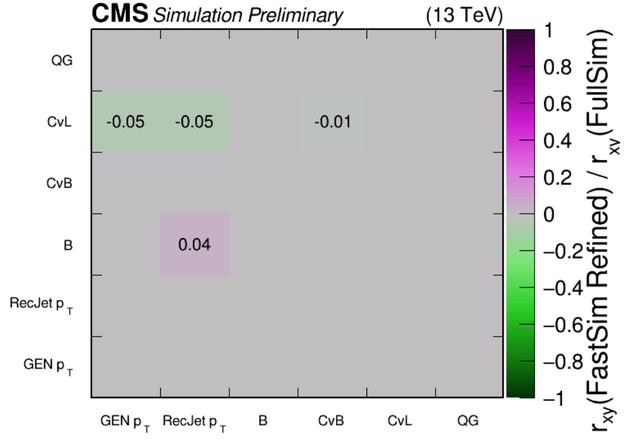


Residuals

x - FullSim

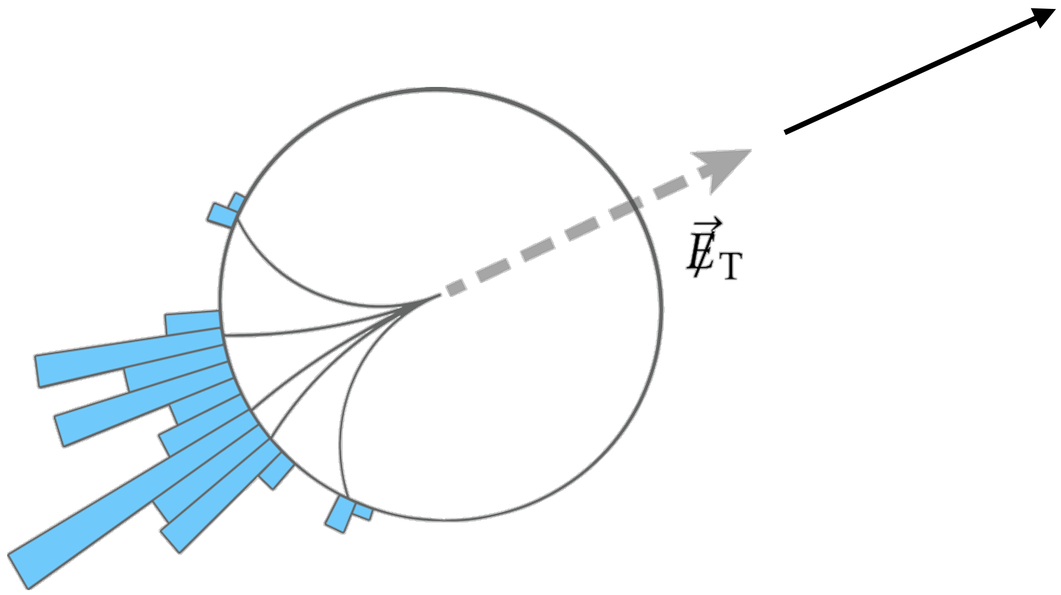


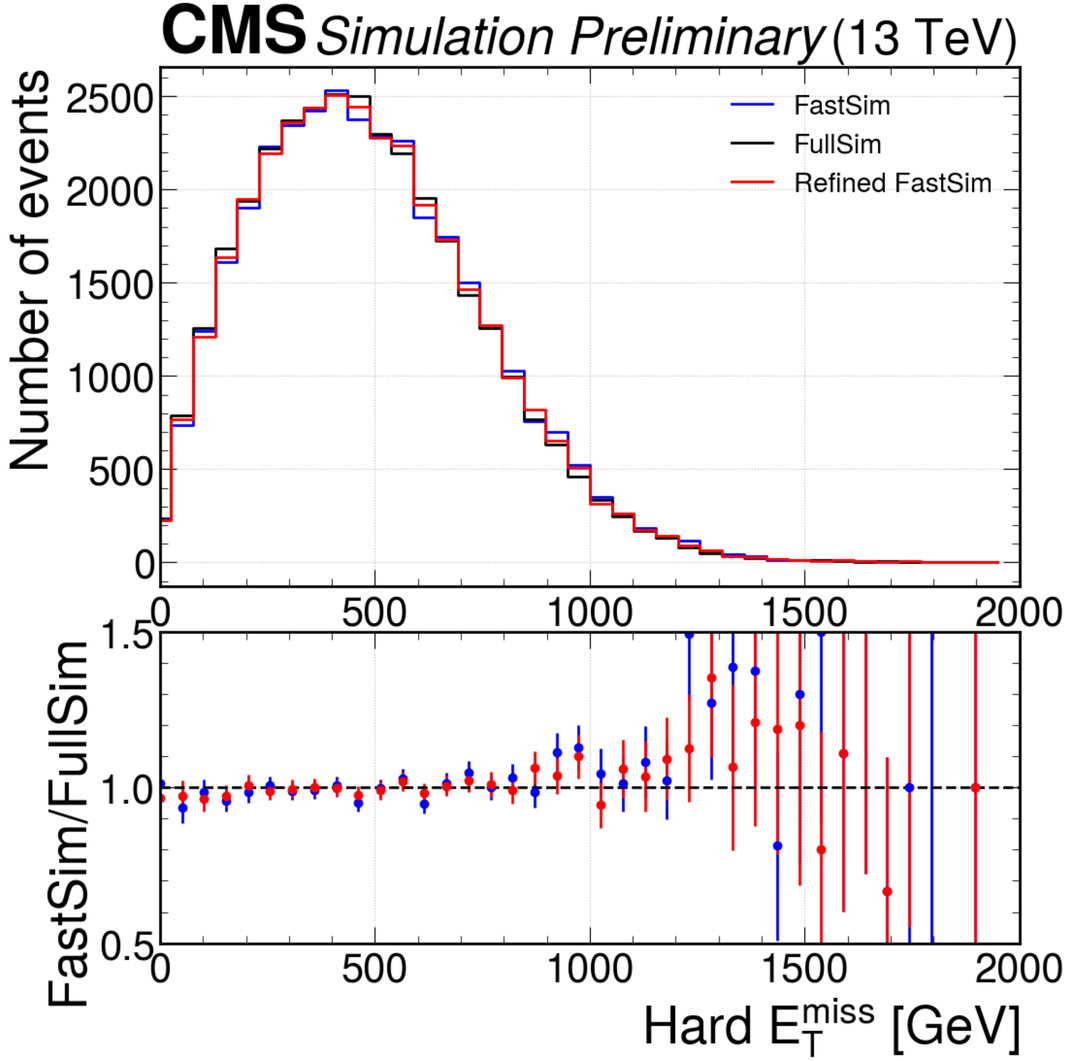




Propagation of refinement to MET

Type-1 MET correction

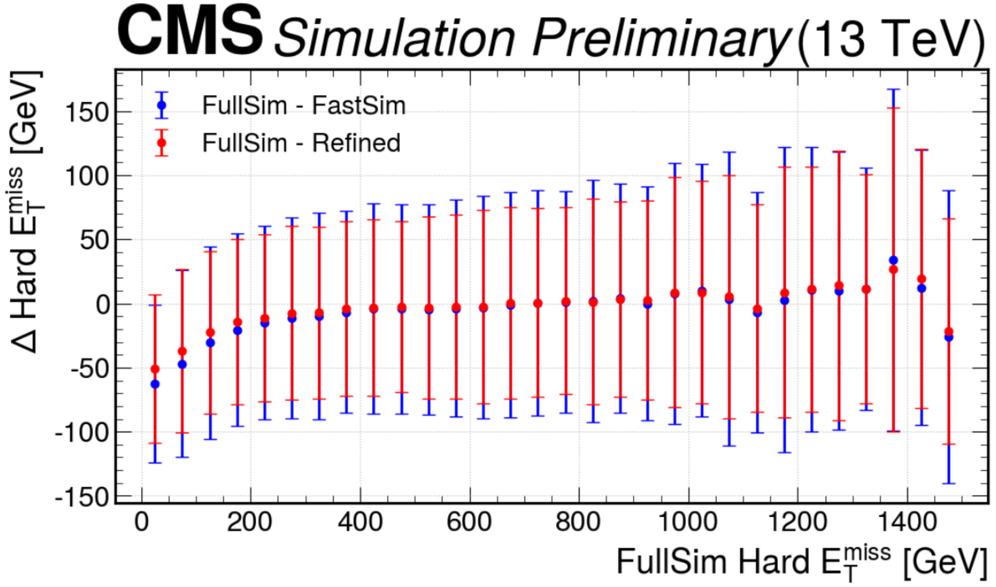


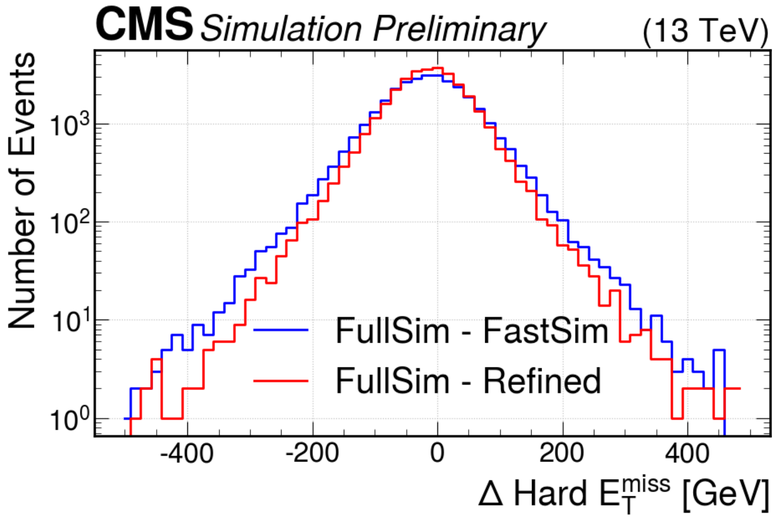


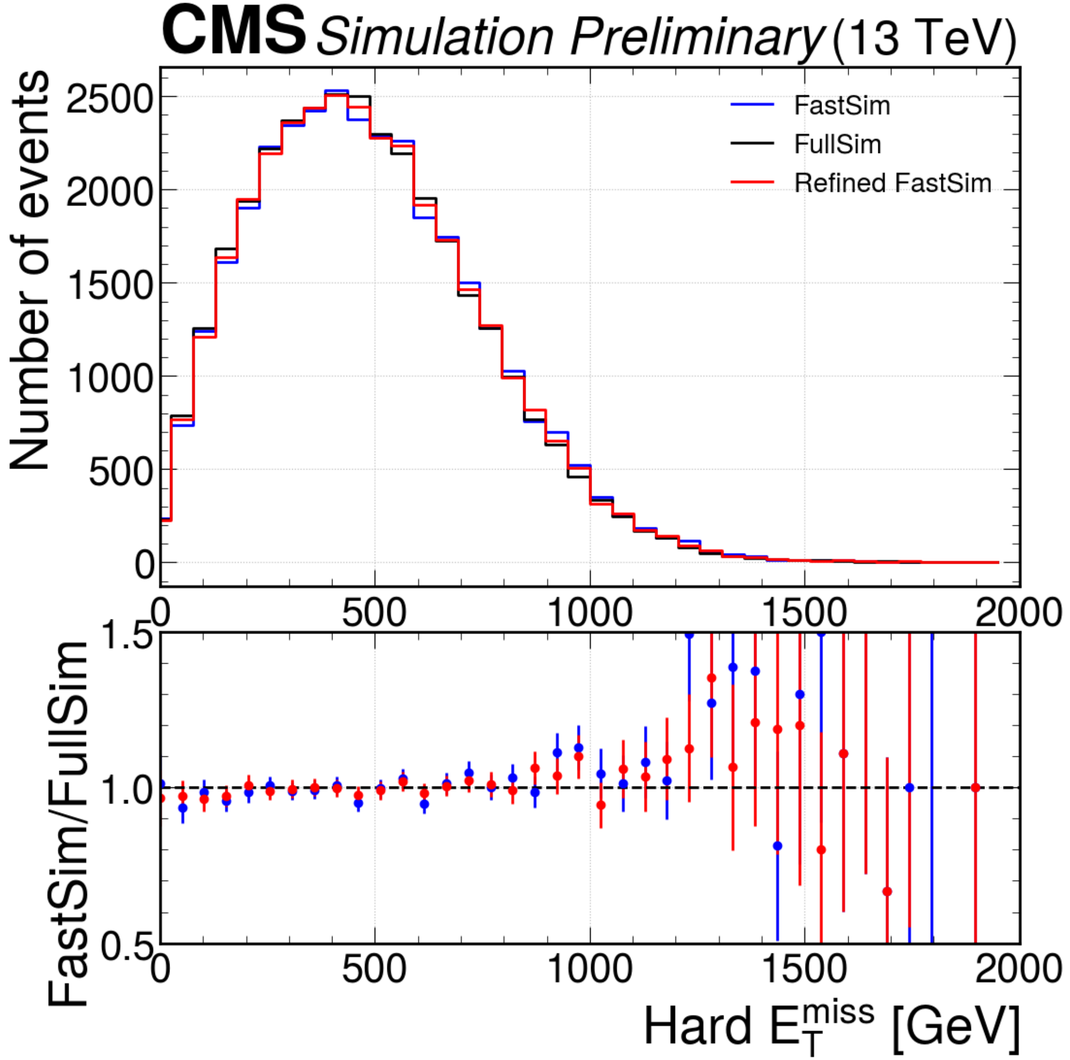
- Propagate refined jet p_T to MET

$$\text{refined}\ \vec{E}_T^{\text{miss}} = \text{Fast}\ \vec{E}_T^{\text{miss}} + \sum_{i \in \text{jets}} \vec{p}_{Ti}^{\,\text{Fast}} - \sum_{i \in \text{jets}} \vec{p}_{Ti}^{\,\text{refined}}$$

Propagation of refinement to MET







Summary and outlook

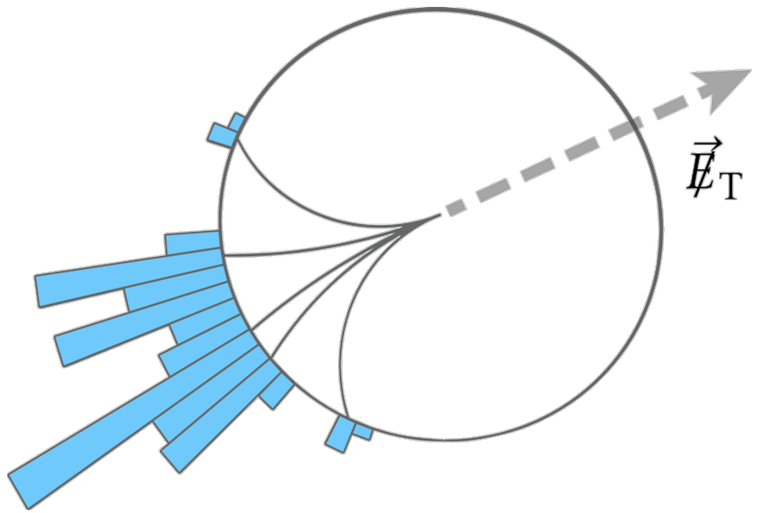
Fast Perfekt applied to CMS jets

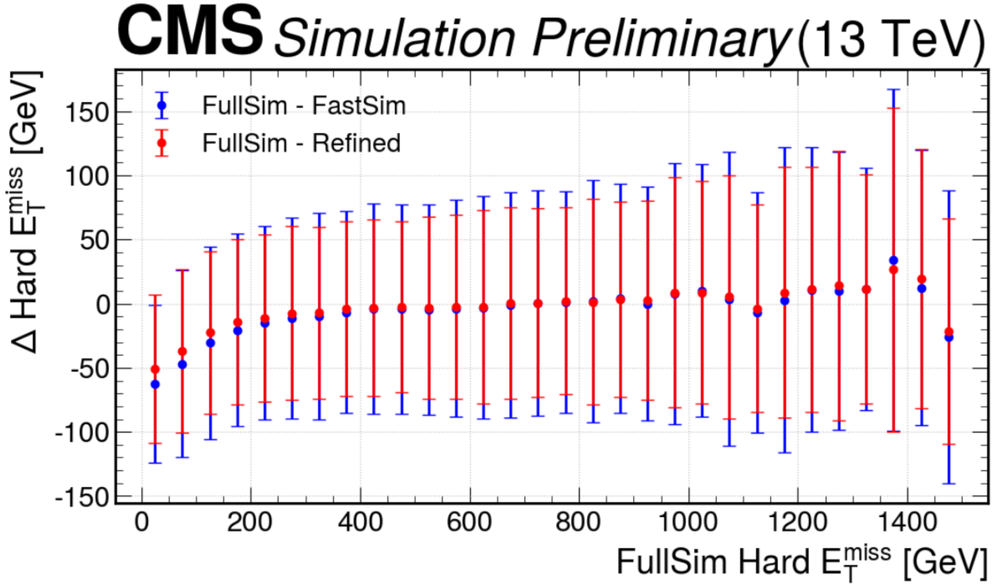
- Single-stage training based on unbiased MMD
  - Refinement with realistic production conditions, e.g., pile-up,
  - Flavor tagging observables, kinematics
  - Refinement propagated to event-level observables shows improvement
- Prototype in place for production, use with new data

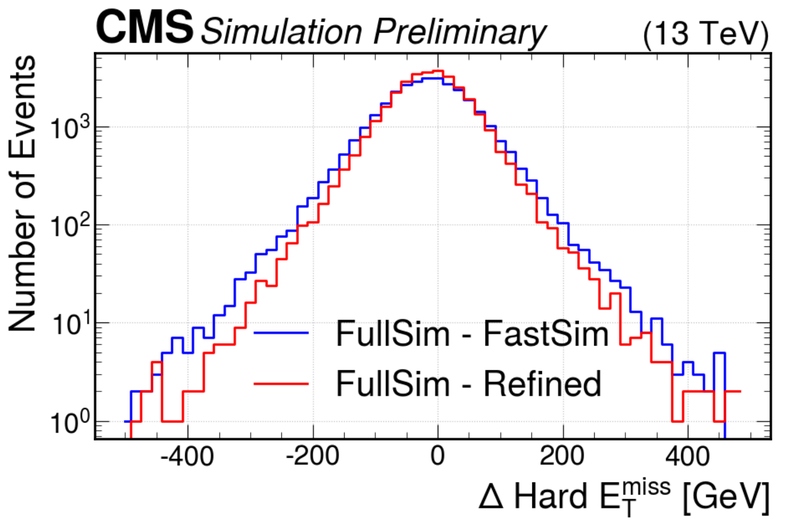
Thanks for your feedback!

Propagation of refinement to MET

Type-1 MET correction







- Propagate refined jet p_T to MET

$$\text{refined}\ \vec{E}_T^{\text{miss}} = \text{Fast}\ \vec{E}_T^{\text{miss}} + \sum_{i \in \text{jets}} \vec{p}_{Ti}^{\,\text{Fast}} - \sum_{i \in \text{jets}} \vec{p}_{Ti}^{\,\text{refined}}$$