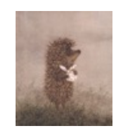
Machine Learning for Neutral Particles Identification in the LHCb Calorimeter

Viktoria Chekalina  Fedor Ratnikov

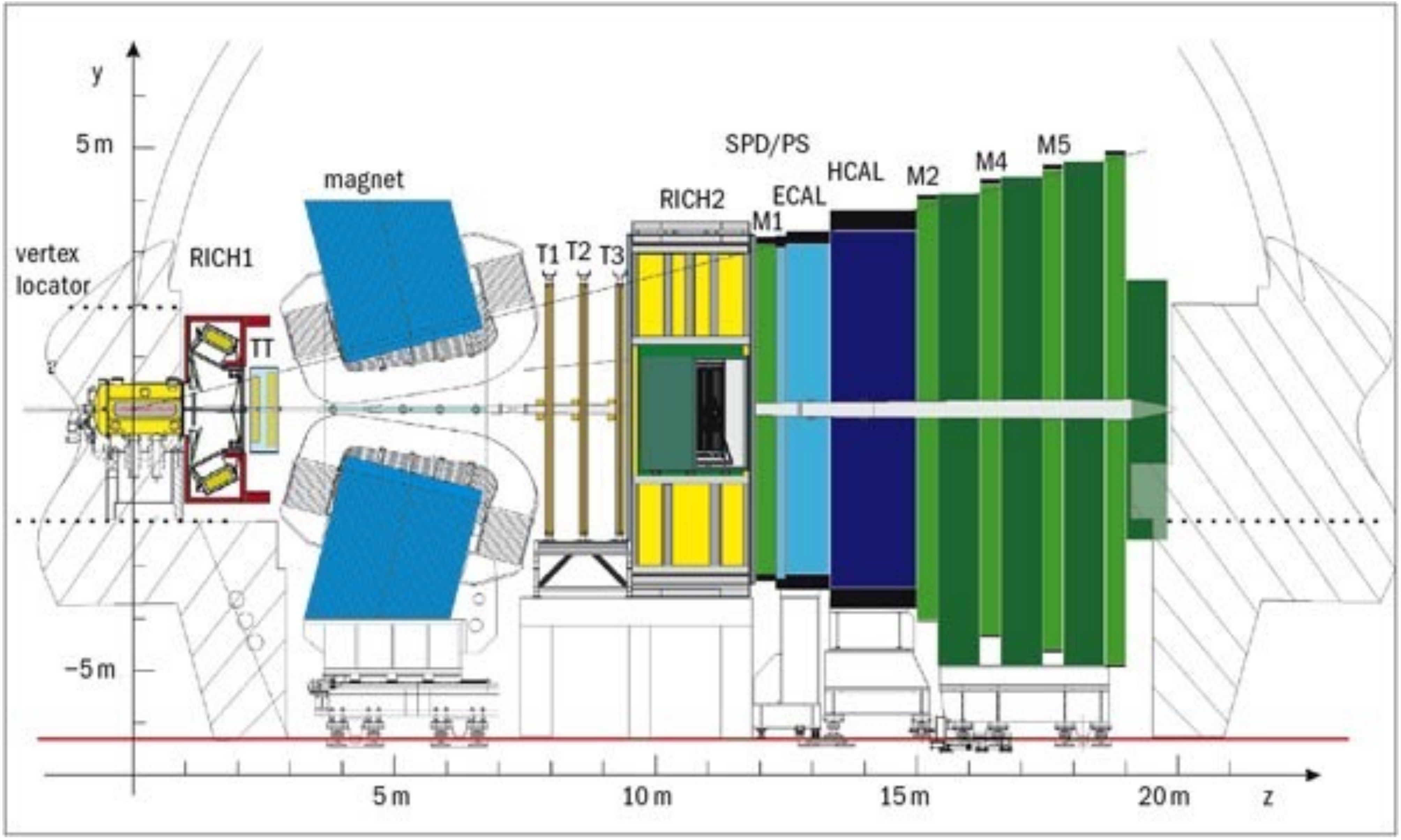
NRU Higher School of Economics,

Oct 2, 2017

Content

- LHCb ECAL and neutral particles identification
- Separation high energy photons from merged photons from π^0 decay as a ML problem
- Classifiers optimisation
- Comparing with baseline

LHCb Detector @LHC



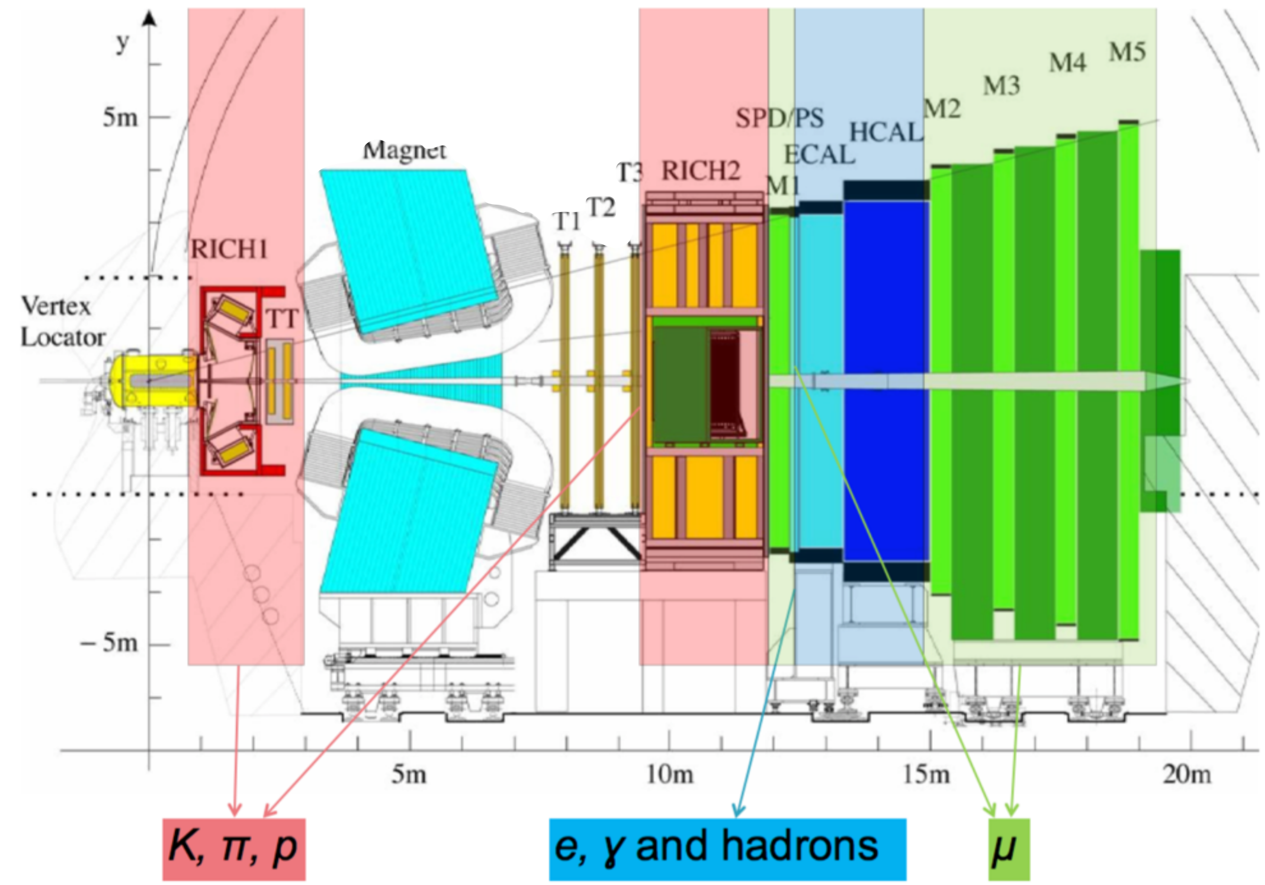
- One of four big LHC detectors

Int. J. Mod. Phys. A 30 (2015) 1530022

- Primary physics program: charm and beauty physics
  - need excellent particle ID

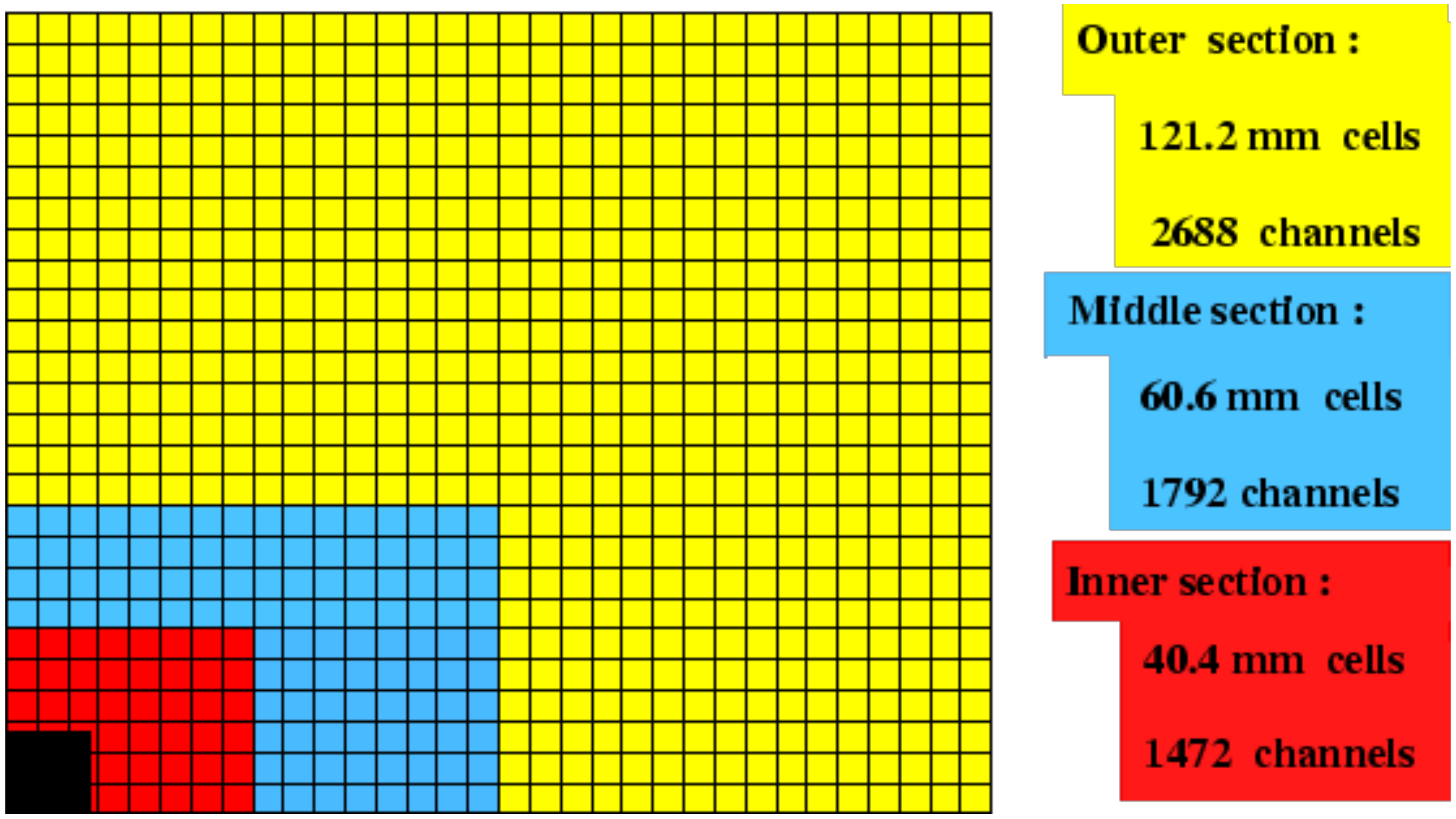
Particle Identification

- Charged particles
  - π, e, μ, K, p
    - see poster "Machine Learning based global particle identification algorithms at LHCb experiment" after this session
- Neutral particles
  - π^0, γ, K^0, n



- Radiative decays (γ) are very important for physics studies
- π^0 are massively produced in pp collisions at LHC
  - immediately decay: $\pi^0 \to \gamma\gamma$
    - photons are merging for high momentum π^0
  - huge background to photons from radiative decays
    - powerful separation is necessary

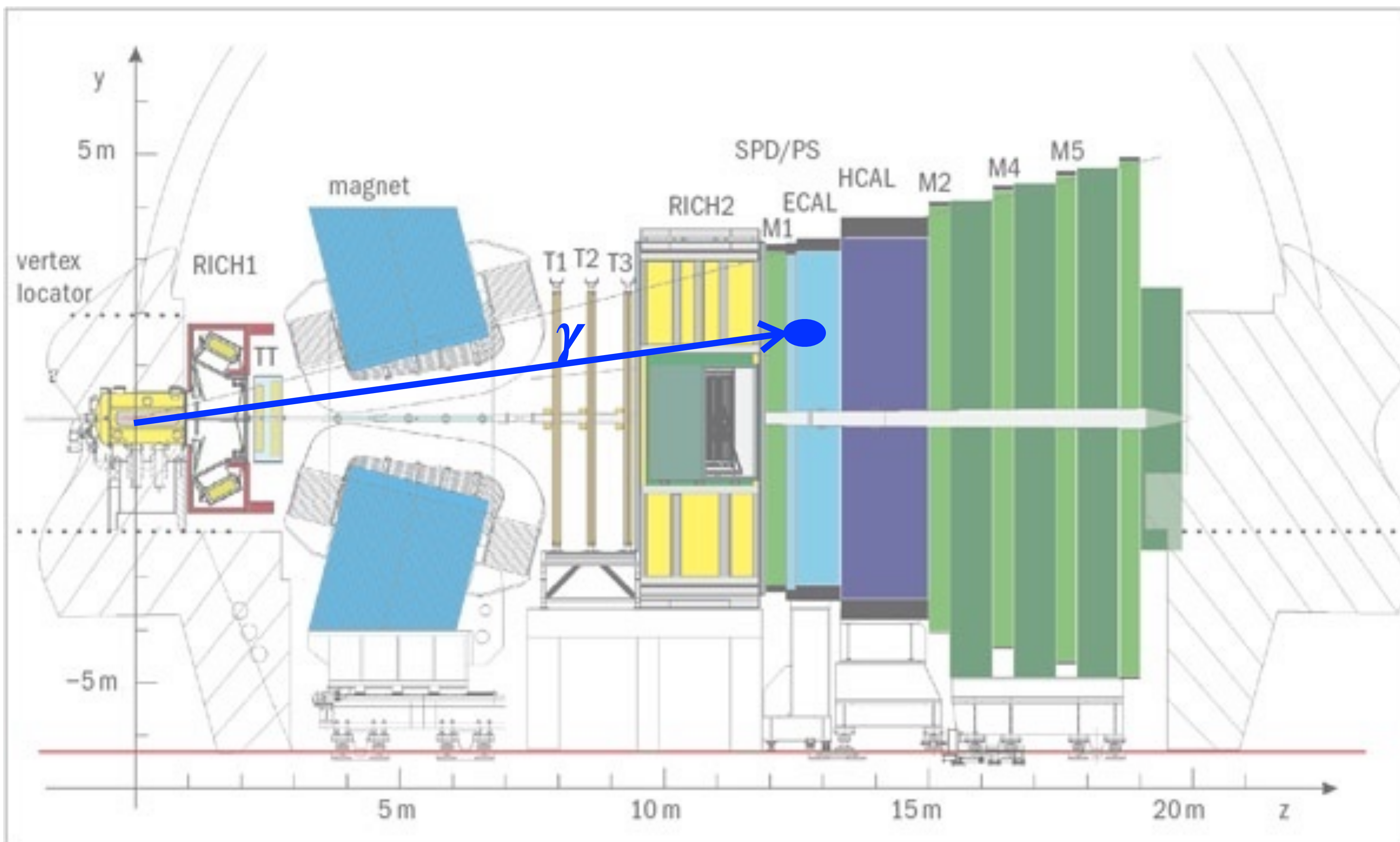
LHCb ECAL



- Shashlik technology
- 1×1, 2×2, 3×3 module granularities
  - pre-shower layer matching ECAL granularity

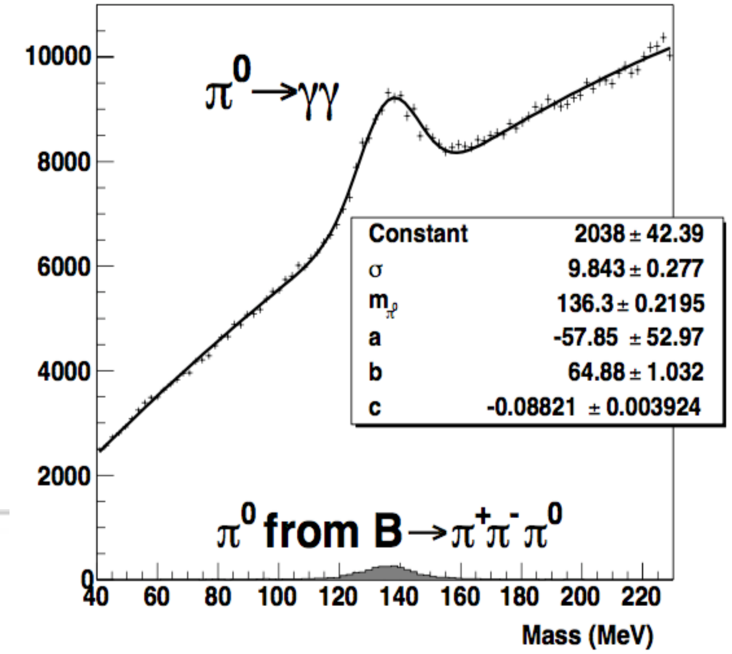
π^0 - γ Separation

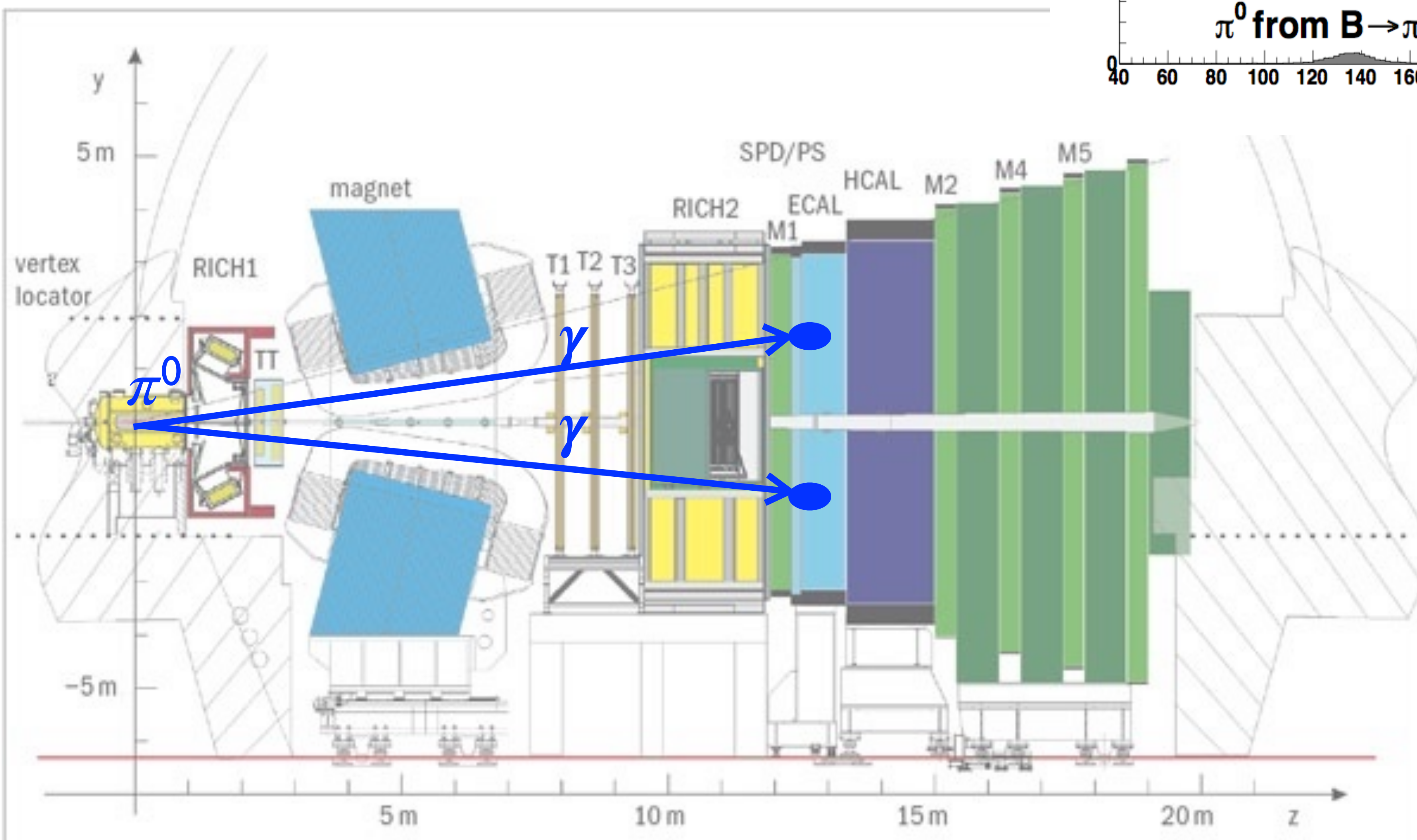
- Separated photon: stand alone cluster



π^0 - γ Separation

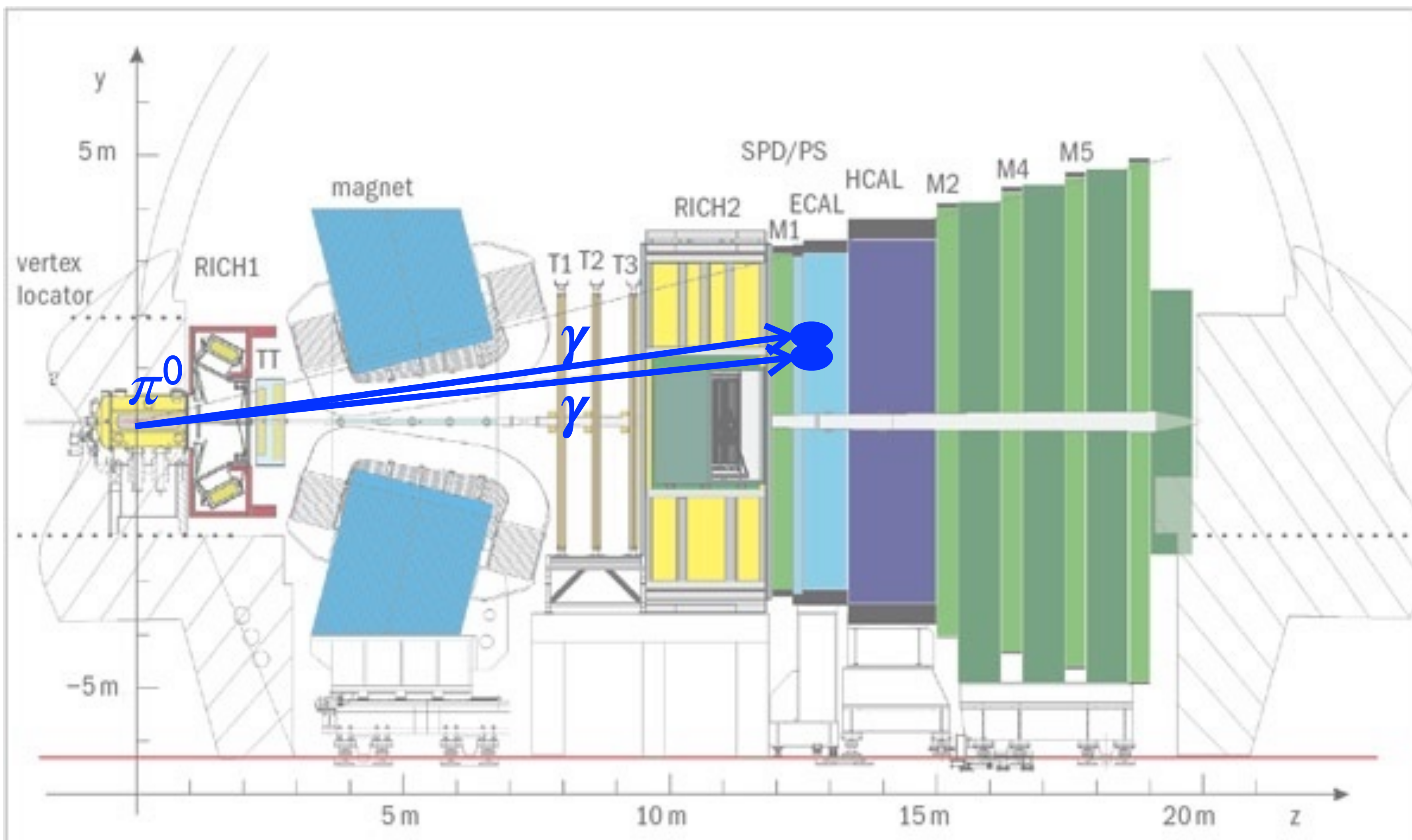
- π^0: two separate clusters





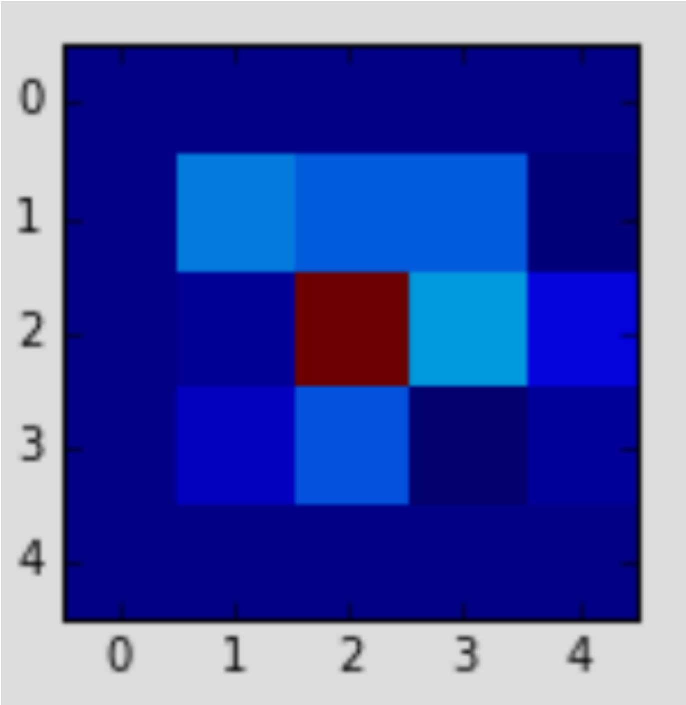
π^0 - γ Separation

◇ π^0: merged clusters fake stand alone photons

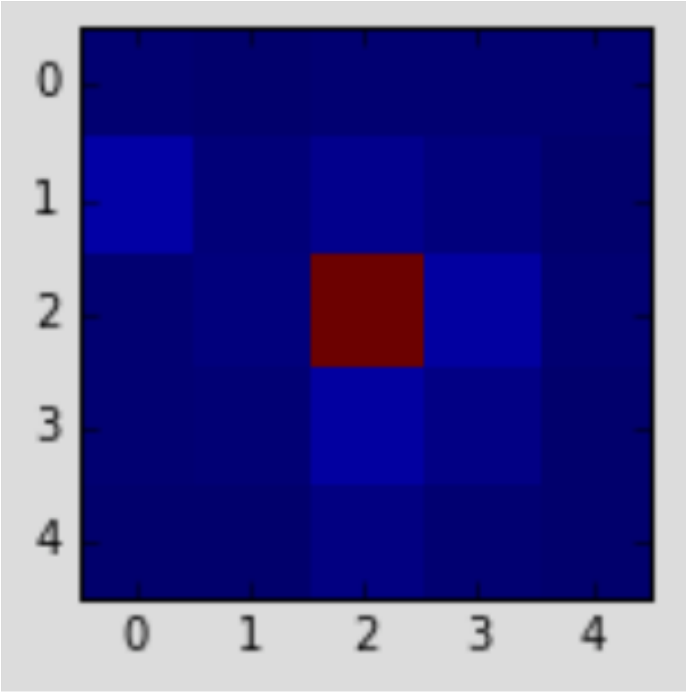


Separation Features

- π^0 - γ features



  - merged photons from π^0: elongated cluster
  - stand alone photon: round cluster
    - not straightforward at coarse granularity

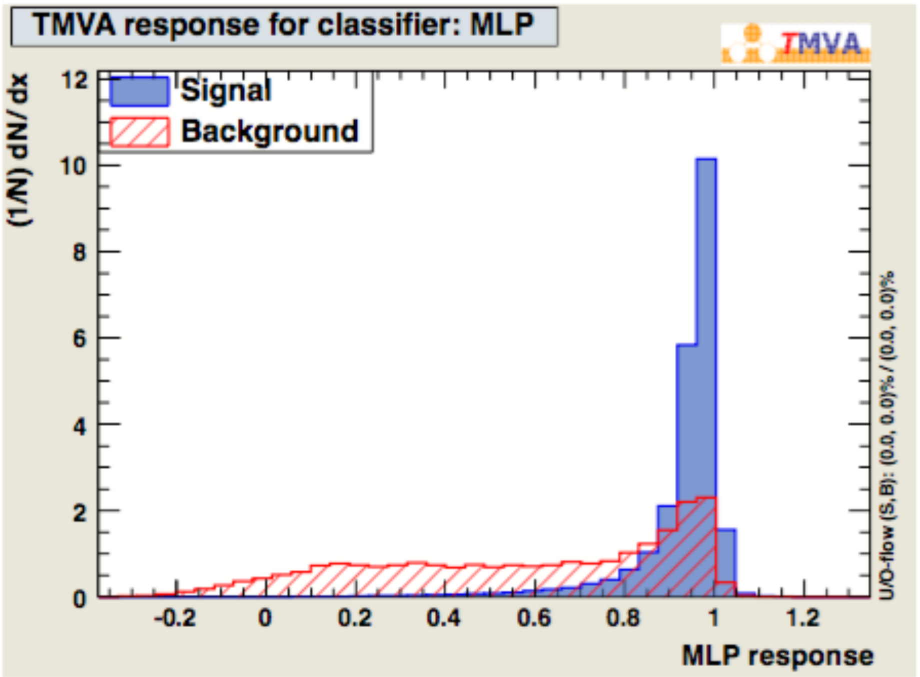


Baseline Approach

- Aggregate cluster constituents into "shape" and "asymmetry" features

LHCb-PUB-2015-016

- Learn behaviours using NN (2-layers MLP)



- Build classifier

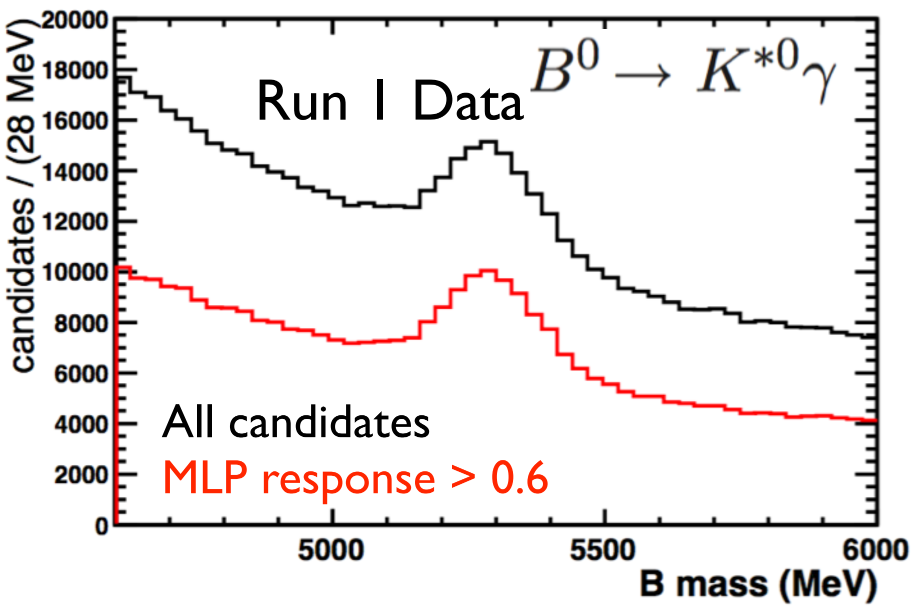
$$r2 = < r^2 > = S_{XX} + S_{YY}$$

$$\frac{E_{\text{seed}}}{E_{\text{cl}}} \qquad \frac{(E_{\text{seed}} + E_{2nd})}{E_{\text{cl}}}$$

$$r2r4 = 1 - \frac{< r^2 >^2}{< r^4 >}$$

$$\kappa = \sqrt{1 - 4\frac{S_{XX}S_{YY} - S_{XY}^2}{(S_{XX} + S_{YY})^2}} \qquad \text{asym} = \frac{S_{XY}}{\sqrt{S_{XX}S_{YY}}}$$

...



- Train
  - $B^0 \to K^{*0}\gamma$ vs mixture of $B^0 \to \pi^0 X$ decay modes

Plain Approach

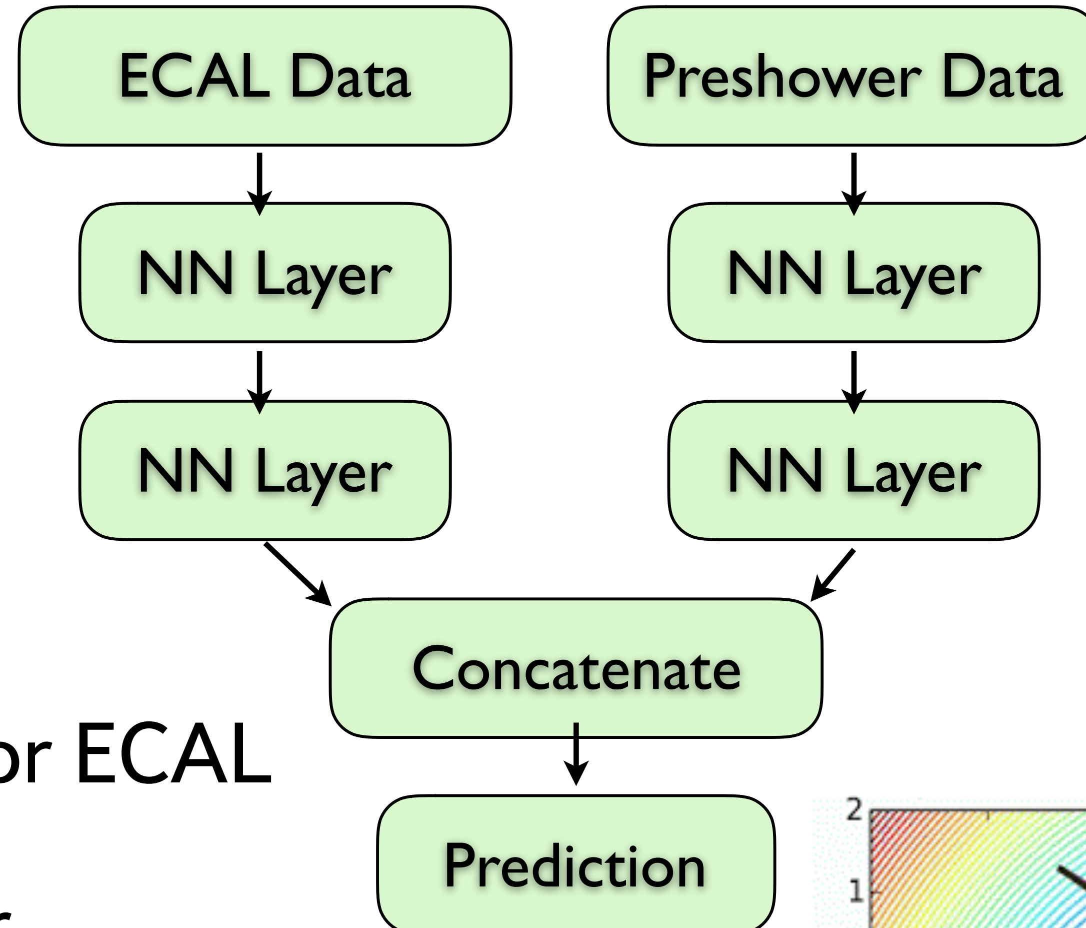
- Consider responses in 5×5 cells cluster around Rec/Calo/EcalClusters
  - for both ECAL and pre-shower detectors
- Build classifier on these 2×25 input features
  - i.e. from first principles
- Train
  - use kinematically similar π^0 and γ MC samples
    - B$\rightarrow K\pi\gamma$
    - B$\rightarrow K\pi\pi^0$
  - Consider potentially misleading candidates
    - $E_T > 2$ GeV
    - photons from π^0 are not more than 2 ECAL cells apart
  - 8 TeV MC data, 120K photons, 220K π^0
- Profit!

ML Technology

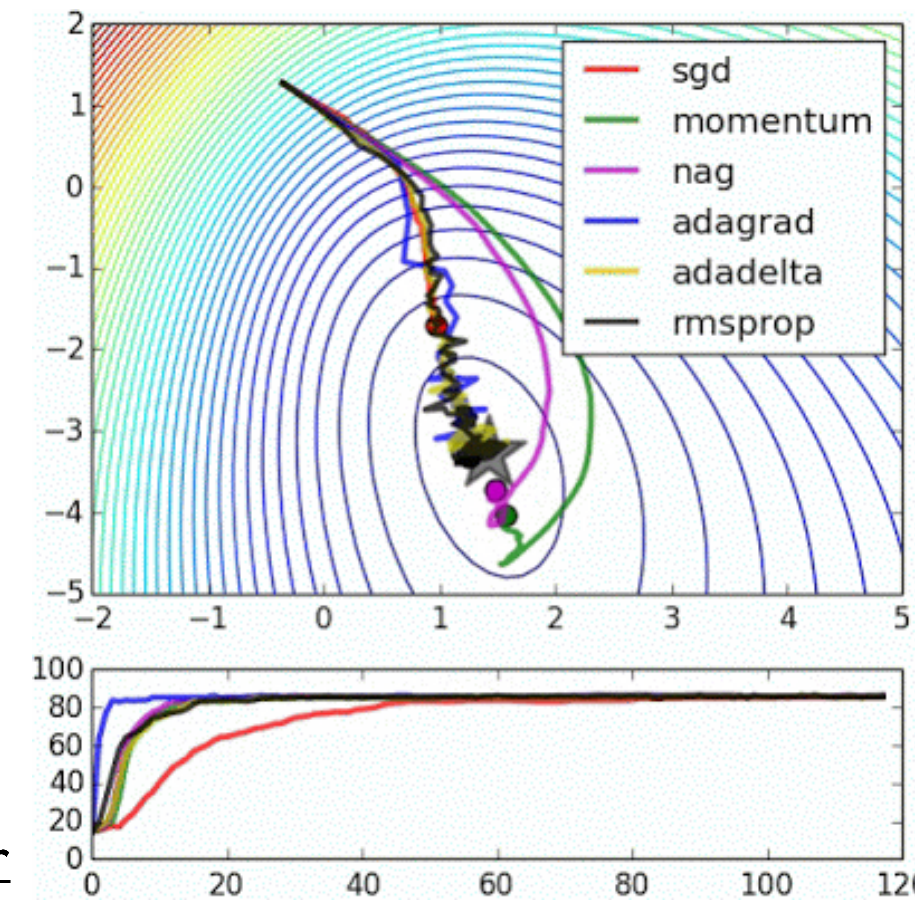
- Neural Networks vs BDT
  - try both approaches
  - optimise hyper-parameters of ML models
    - use inner area as a reference
  - pick the best approach

  - Out of the box
    - 1-layer NN, 80/50 ECAL/Preshower units: ROC AUC =0.86
    - XGBoost: ROC AUC =0.95

NN Optimisation

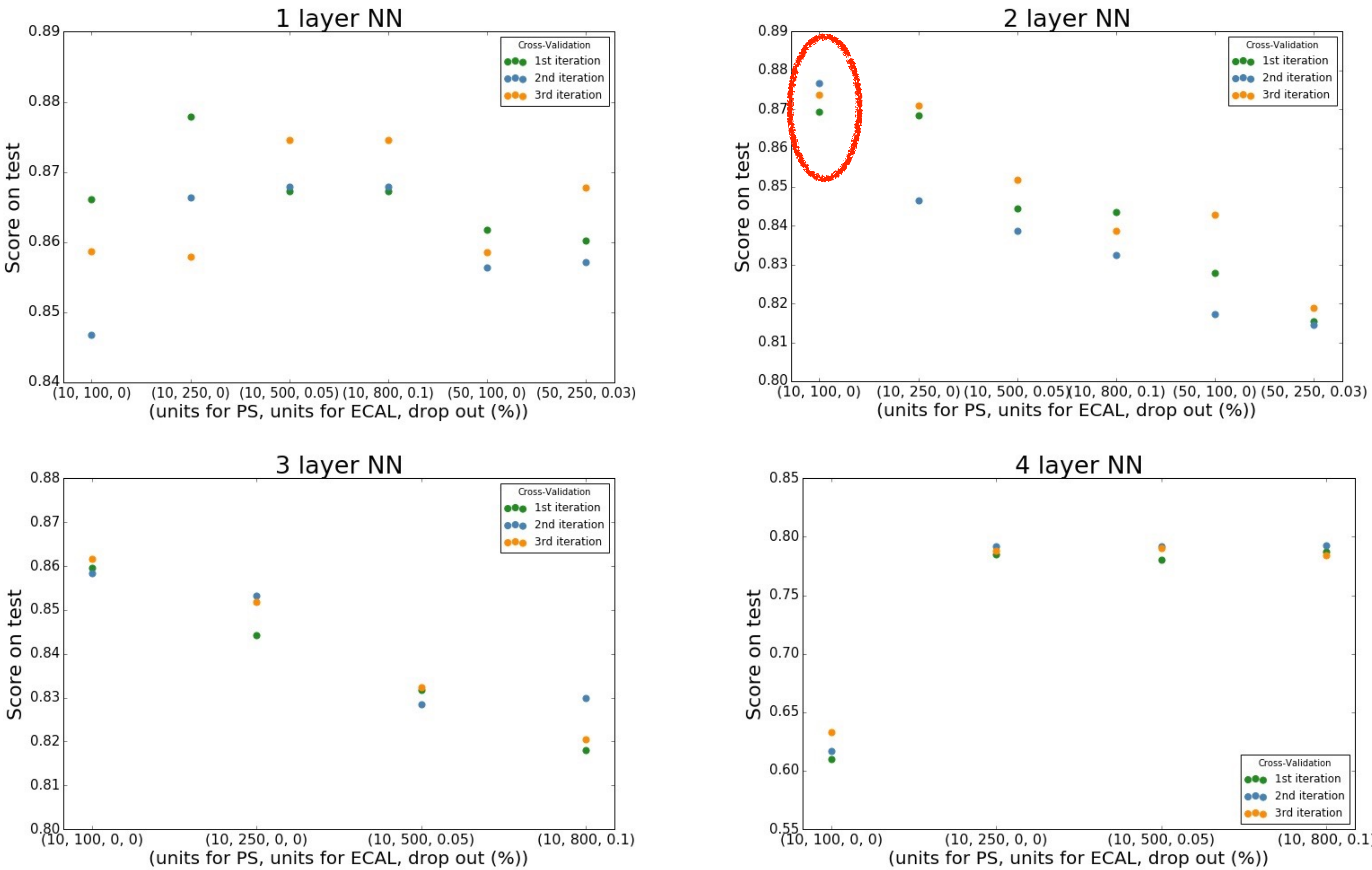


- NN depth
  - 1..4 hidden layers
- NN width
  - 100, 250, 500, 800 for ECAL
  - 10, 50 for Preshower
- Optimizer
  - Adamax, Adagrad, SGD

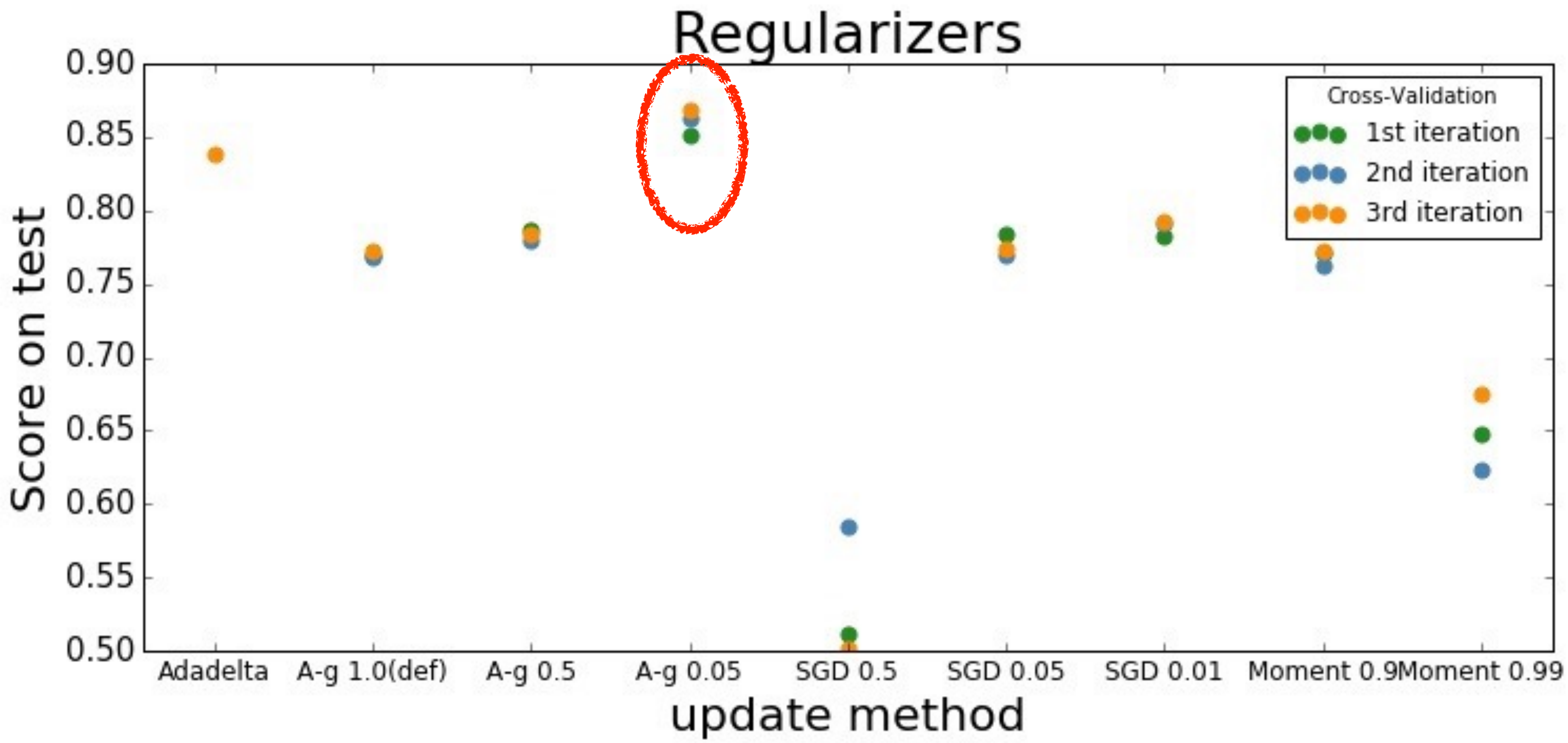


http://imgur.com/s25RsOr

NN Optimisation



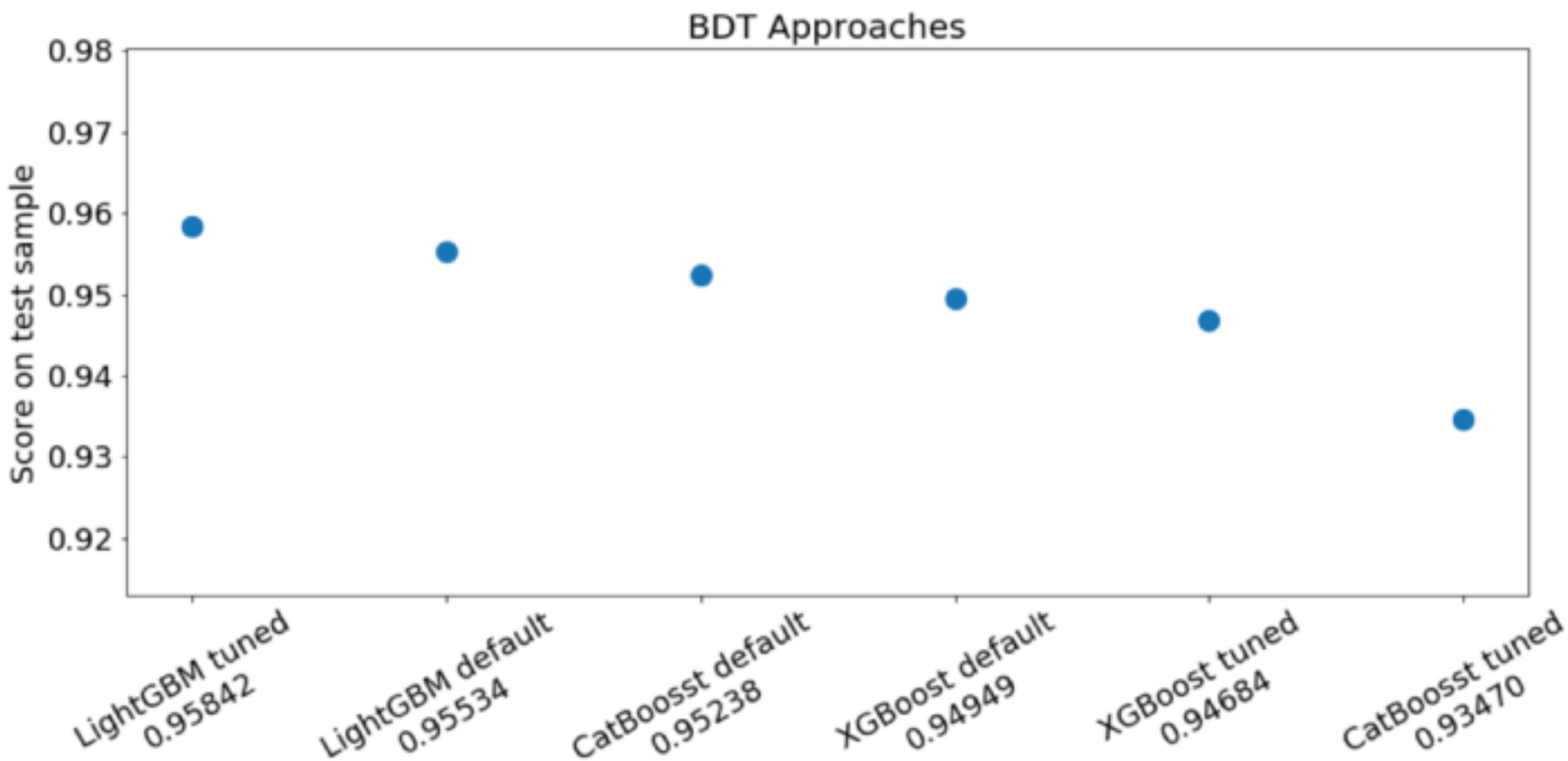
NN Optimisation



- Select minimizer
  - Adagrad(loss rate 0.05) win
- Final ROC AUC = 0.87

BDT

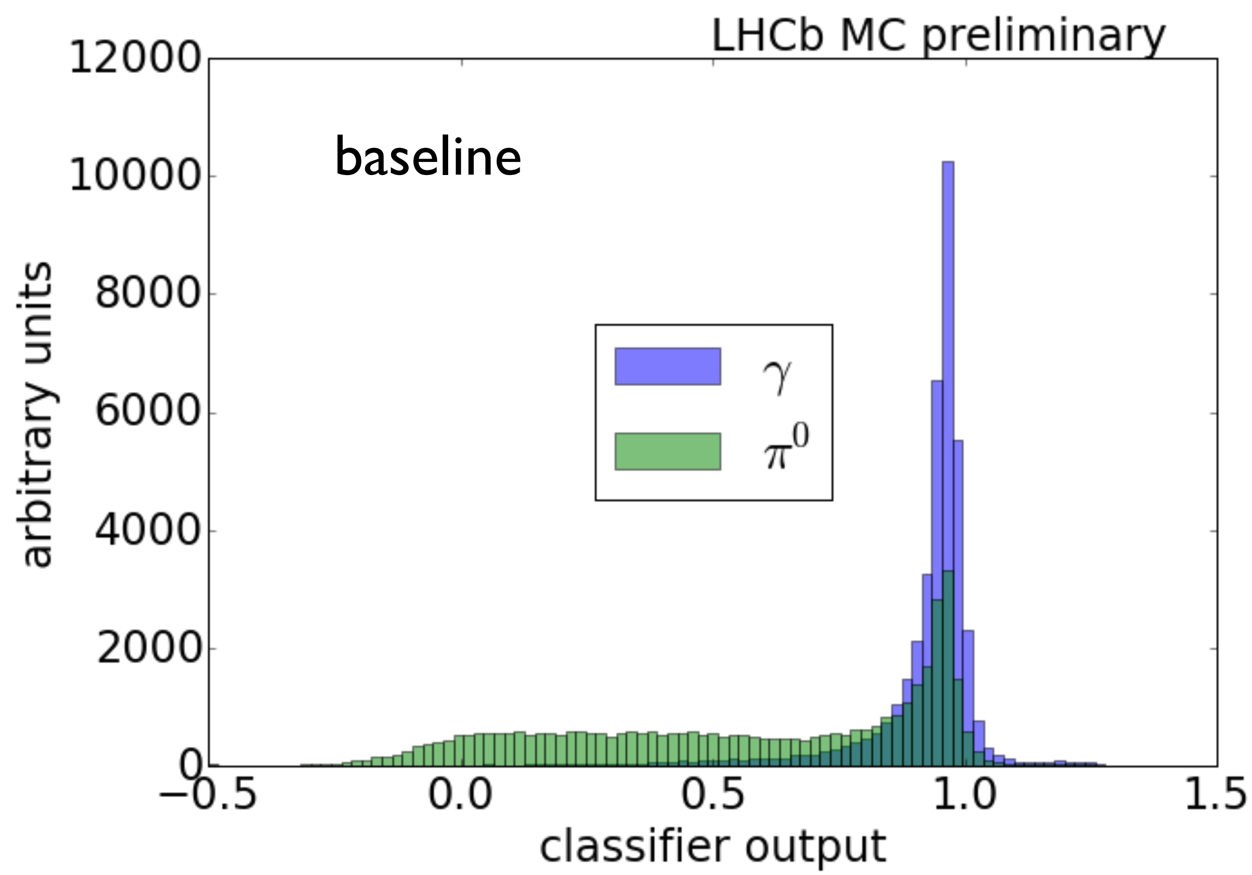
- Approaches
  - XGBoost
  - CatBoost
  - LightGBM

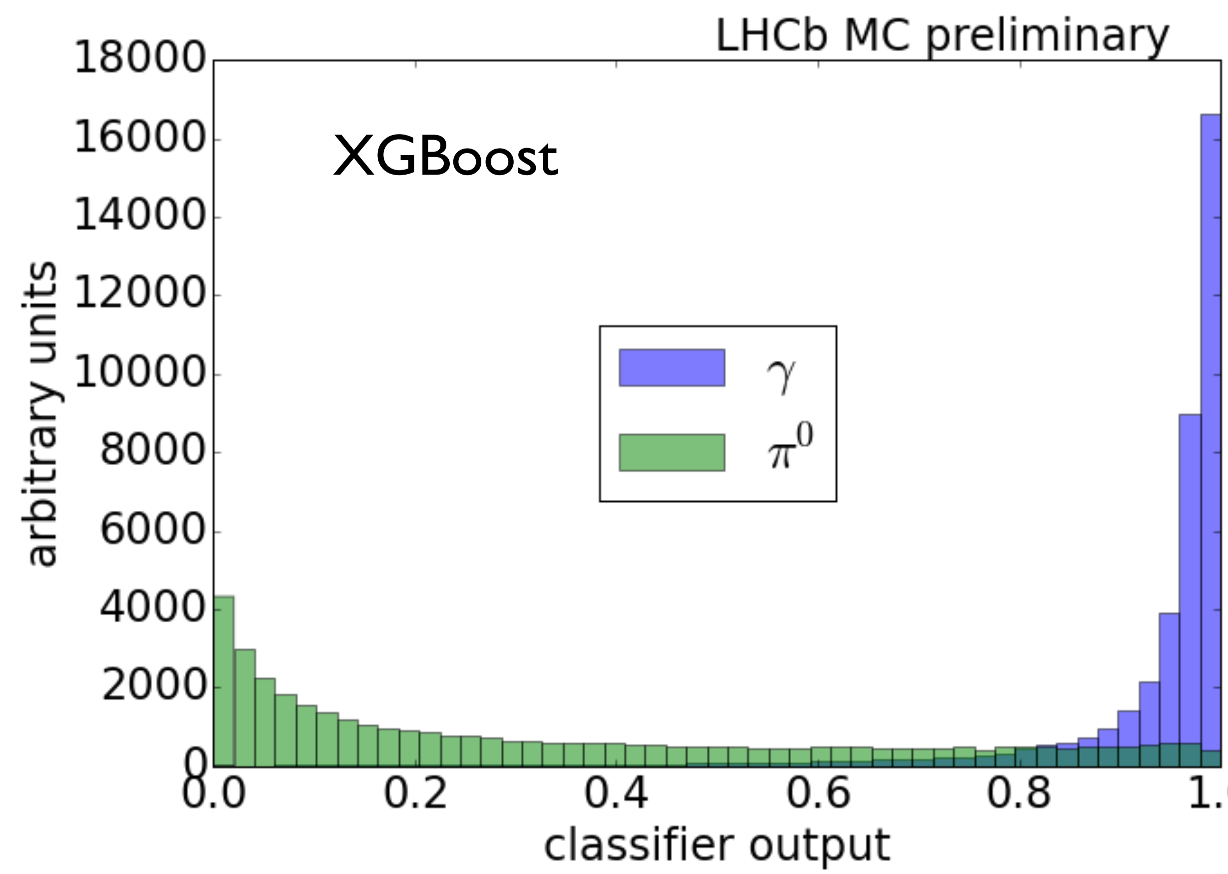


- Tune with
  - https://github.com/yandexdataschool/modelgym/

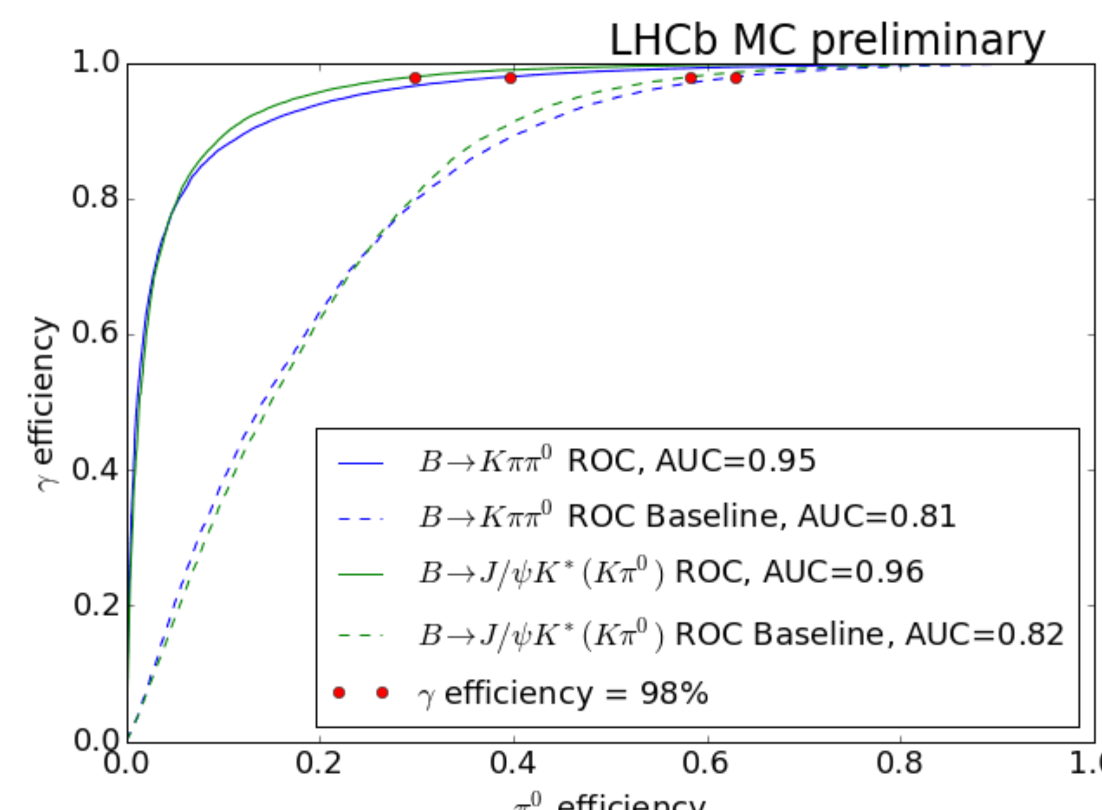
- Scores are about the same
  - continue using XGBoost

All Regions Together

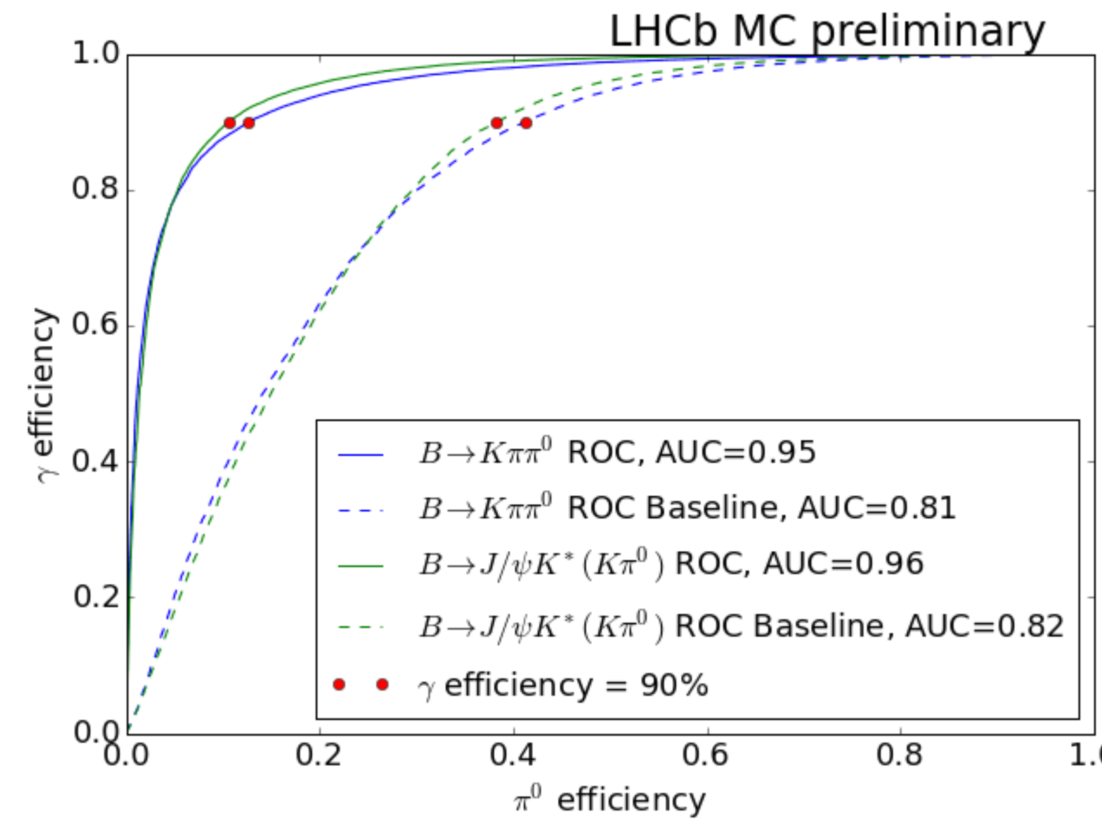


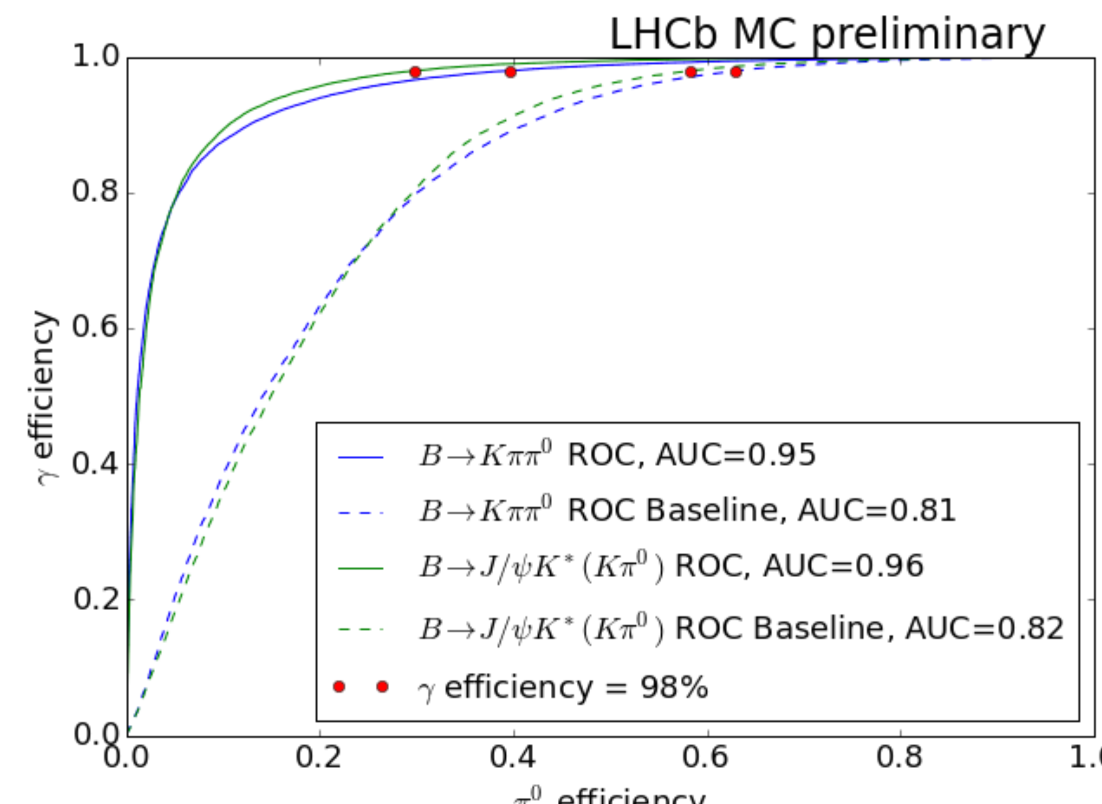


- Significant improvement of the quality
  - not that fantastic at higher efficiencies



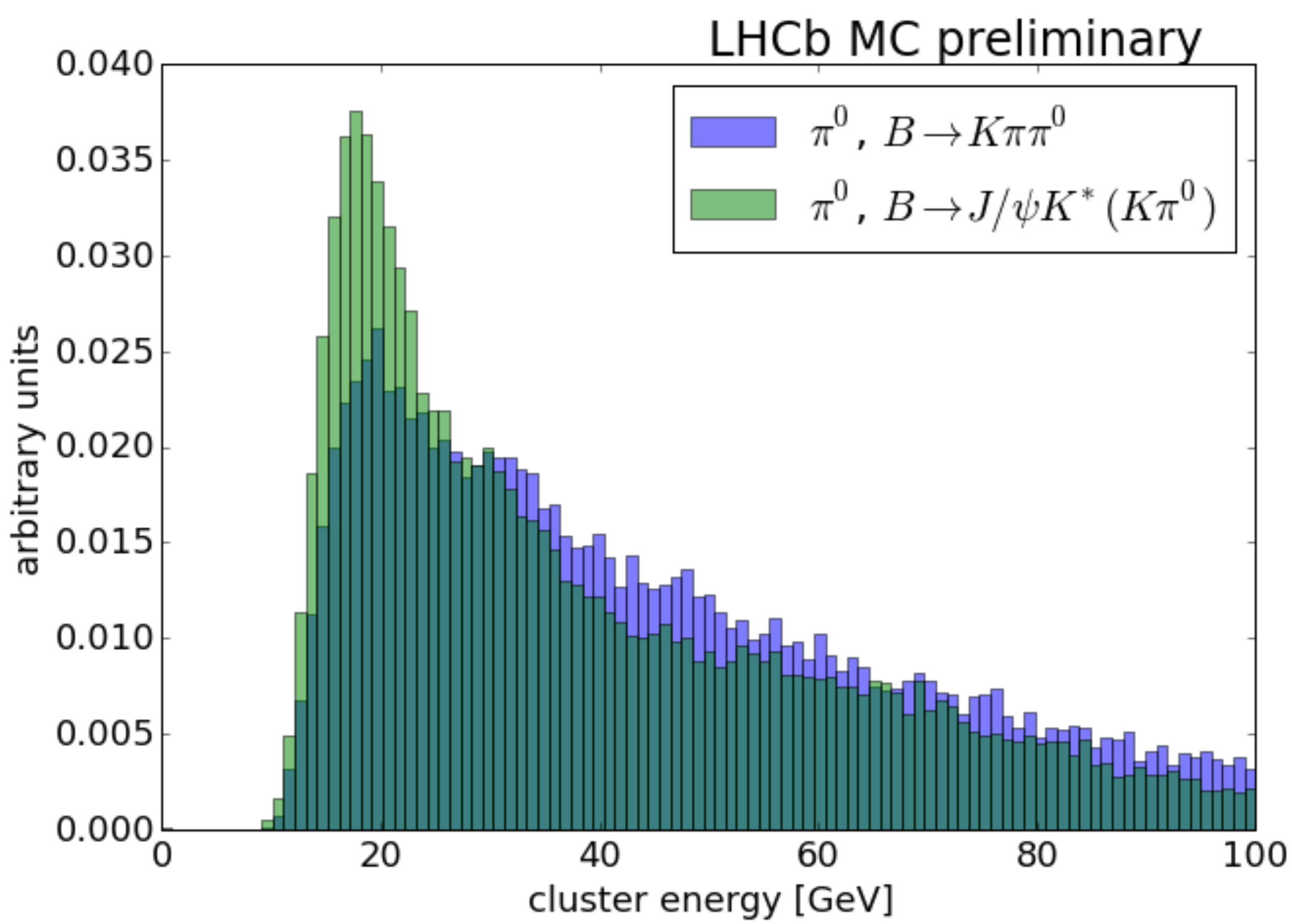
All Regions Together

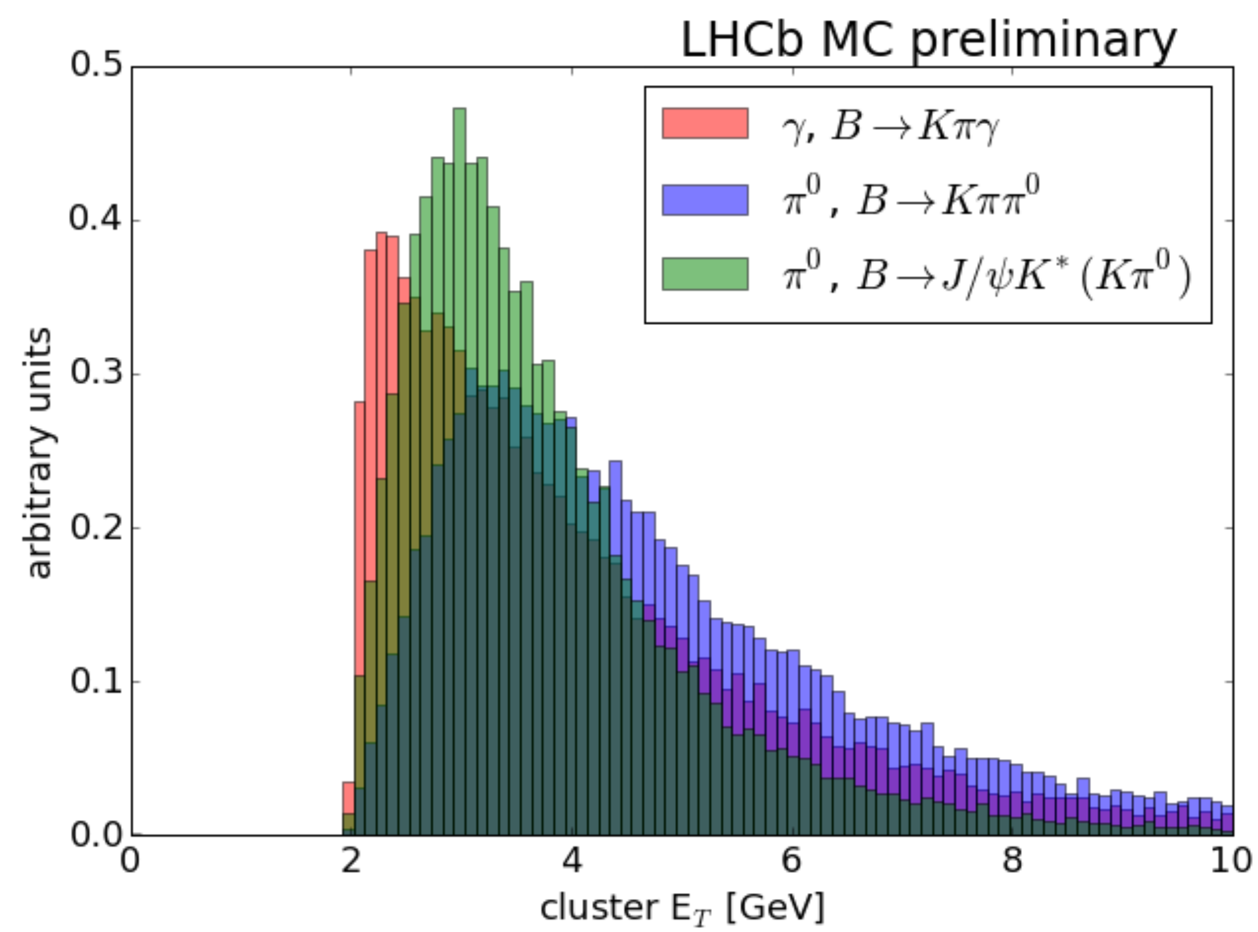




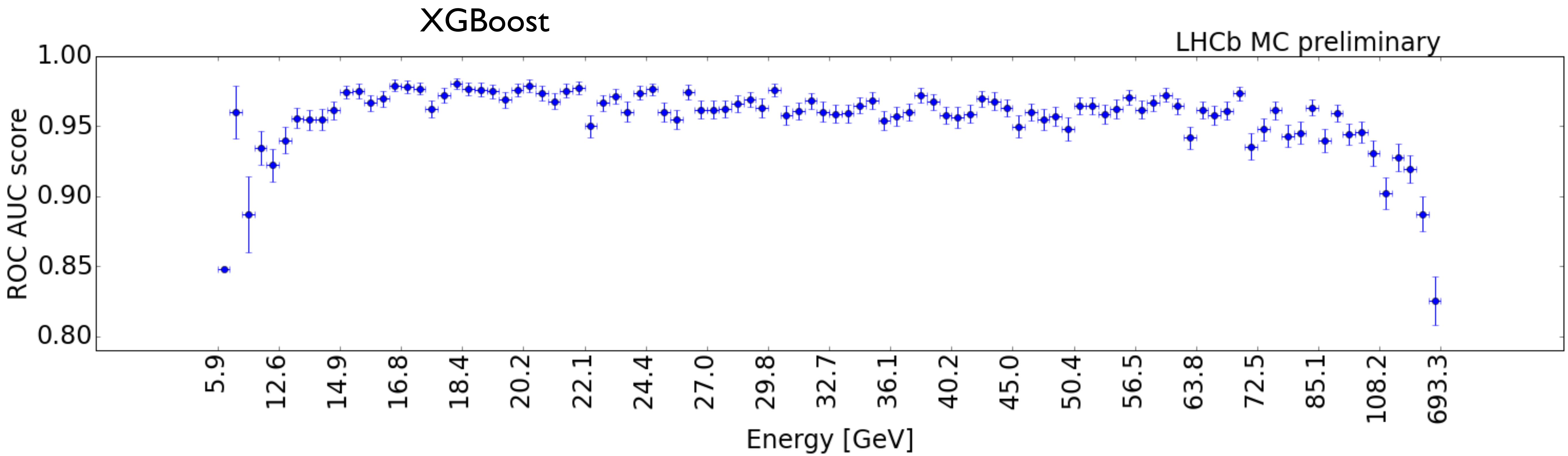
- Significant improvement of the quality
  - not that fantastic at higher efficiencies
  - though factor of 4 at moderate efficiencies

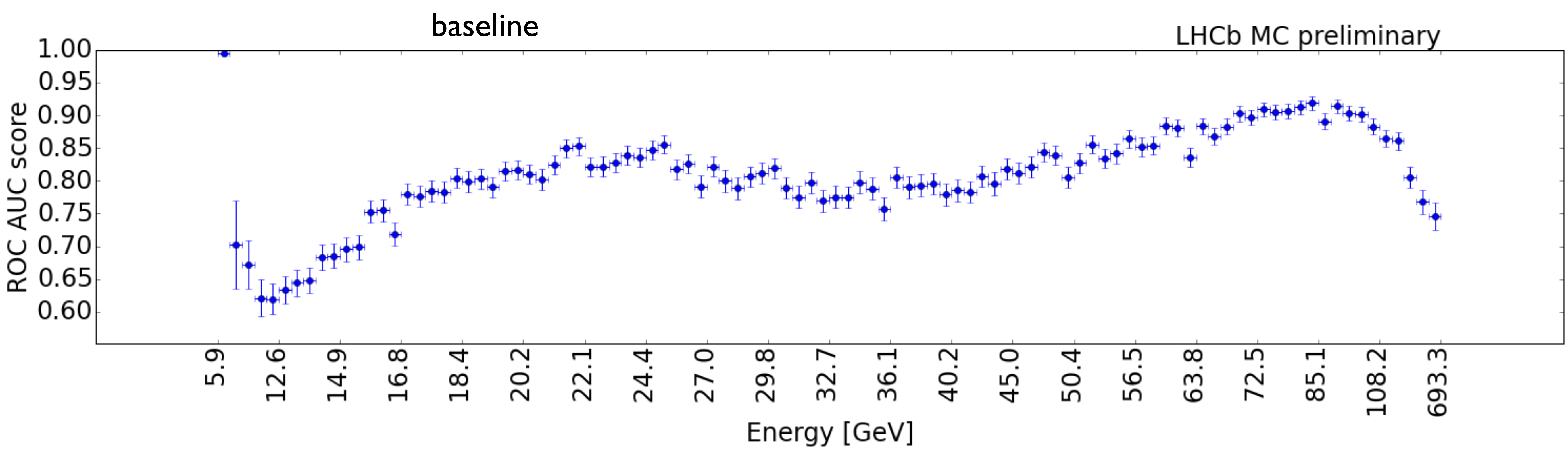
Samples Kinematics



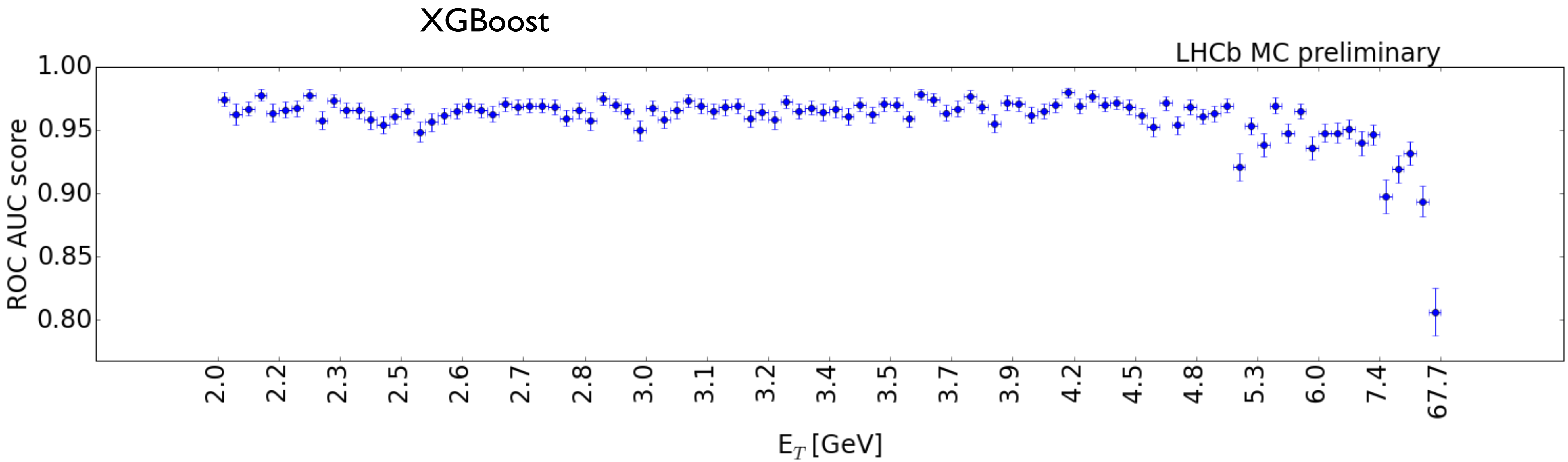


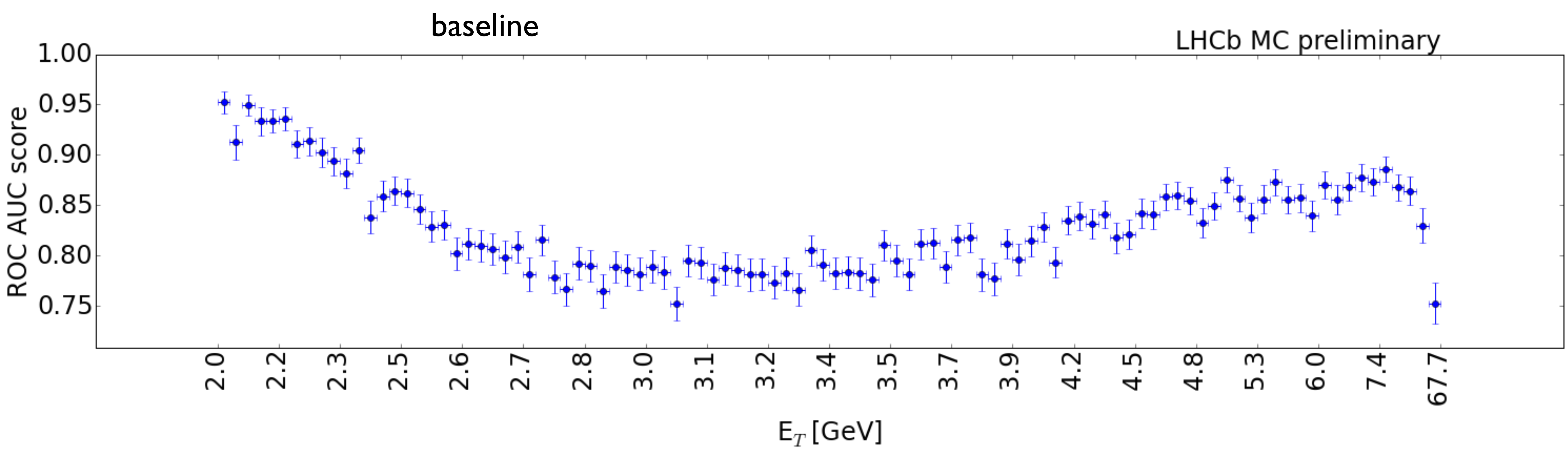
Momentum Dependencies



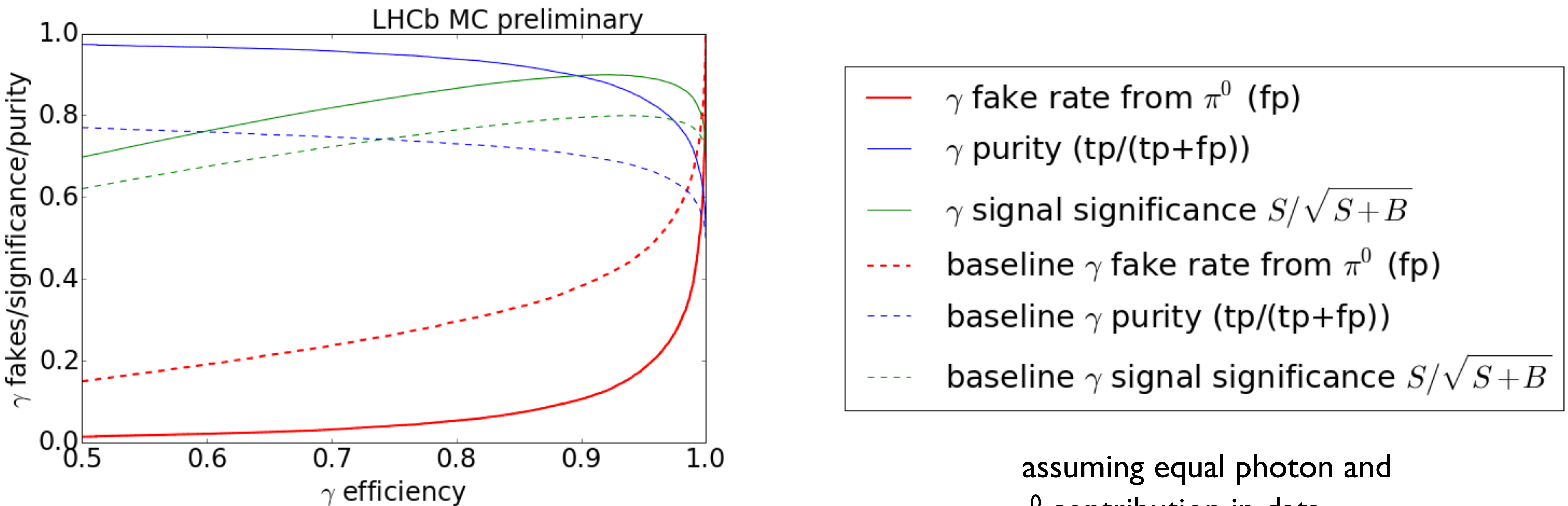


E_T Dependencies





Overall Performance



assuming equal photon and
π^0 contribution in data

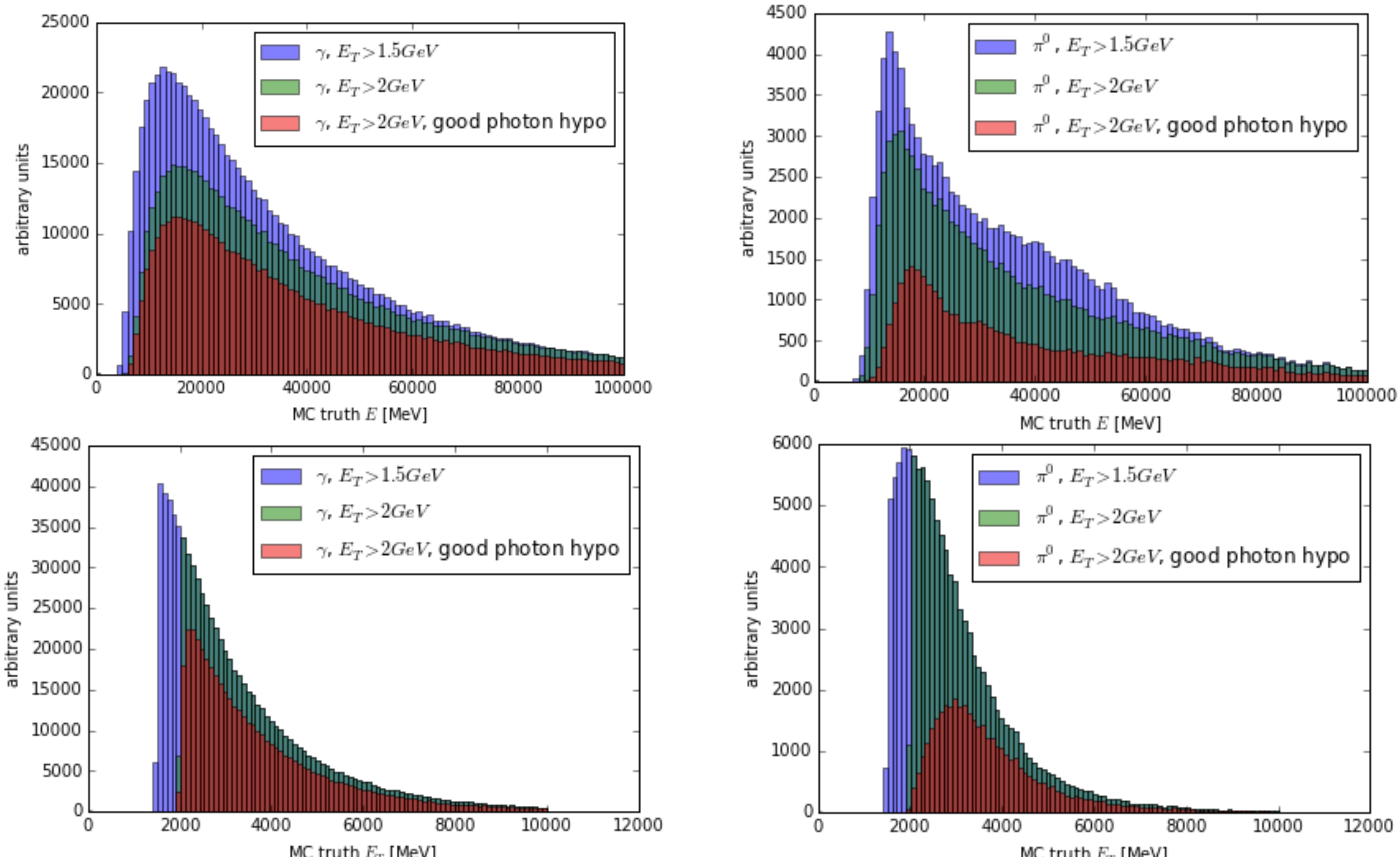
- Significant suppression of π^0 contribution is feasible for those analyses using photons

Conclusions

- We developed a new procedure to separate photons from merged π^0, based on first principles: direct use of energies deposited in calorimeter cells
- Diligent optimisation of both NN and BDT approaches was performed
- BDT shows better performance on this problem
- Promising possibility of aggressive background suppression is demonstrated on simulated data
- Performance tests on real data will follow

Backup

Kinematic Extension



◇ 30% efficiency losses due to hypothesis requirement

Train Samples Kinematics