Quark/gluon jet discrimination using Machine Learning

Adam Tomczak, Dzmitry Budkouski

Supervisors: dr hab. Marcin Chrząszcz, dr Andrzej Siódmok

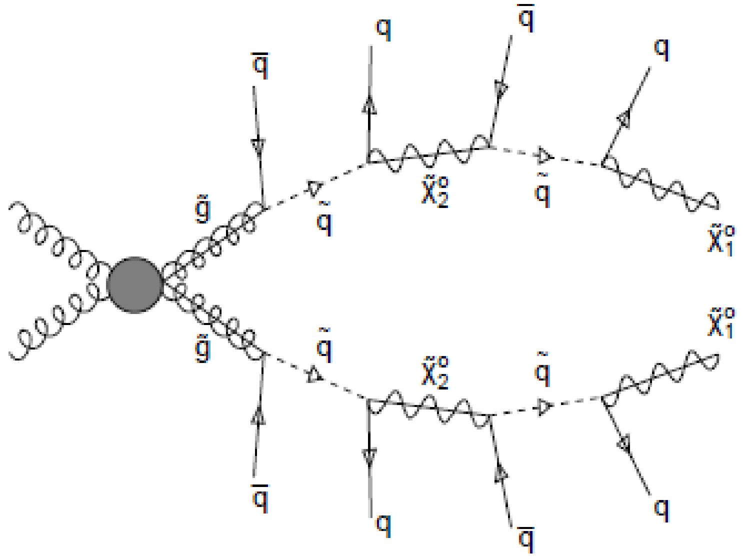
Institute of Nuclear Physics
Polish Academy of Sciences
Particle Physics Summer Student Programme

IFJ PAN, Krakow
2nd August, 2019

Motivation

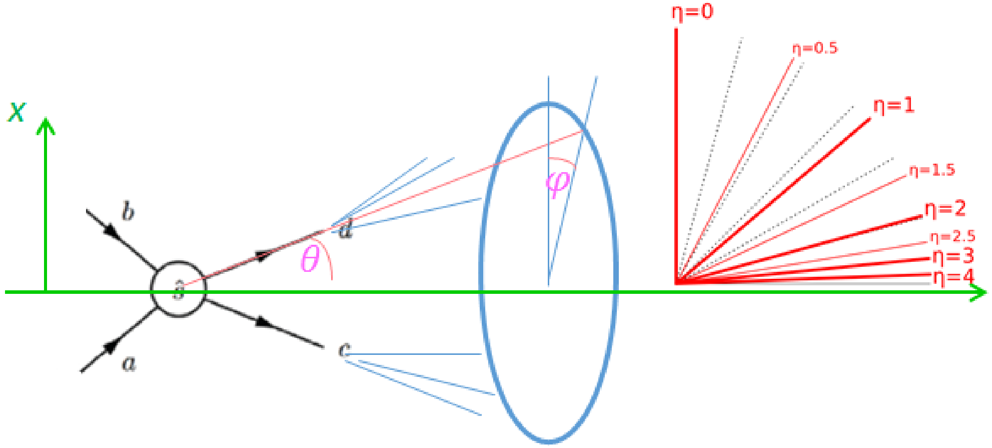
- Abundance of jets in HEP experiments
- The need for quick and accurate Event Selection
- Origins of jets
  Gluons -> Standard Model QCD Background
  Quark -> a signature of Beyond SM processes
- Jet spectrum discrimination as a tool

- Our goal: computationally efficient model for discrimination of quark and gluon jet abundance in a sample
- Method: neural networks with a constrained number of inputs and simplifying cuts in data analysis

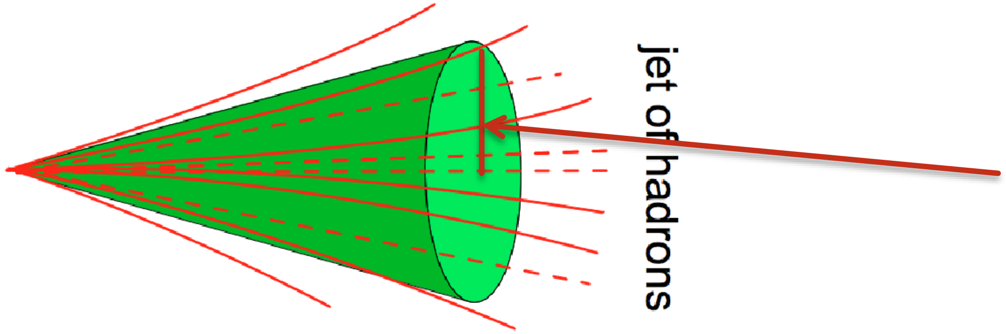


Hadronic Jets

- Hadronic Jet Definition
  A narrow cone of hadrons and other particles produced by the hadronization of a color particles (quark/gluon)
- Kinematic variables and variables for jet description



$$\eta = \frac{1}{2}\, ln \frac{|p| + p_L}{|p| - p_L} = -\ln[tan\frac{\theta}{2}]$$



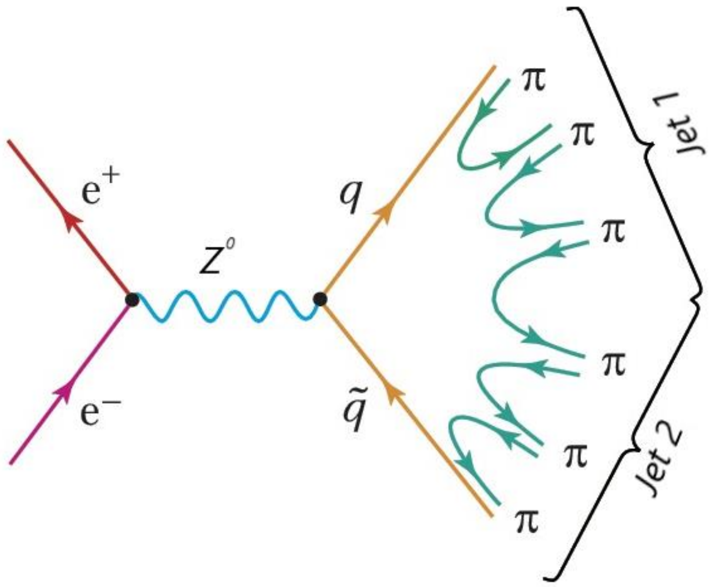
Angular distance in (η, φ) - coordinates:
$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\varphi)^2}$

Data Generation

- Herwig 7.1.5 GPMC
  (cluster hadronization model)
- Energy of the processes -> 91,2 GeV
- Clean quark and gluon samples generated from LEP simulations

$$ee \rightarrow H \rightarrow gg \qquad ee \rightarrow Z \rightarrow qq$$

- Quark vs gluon jet substructure differences



Analysis: packages and variables

Packages used for analysis:

- The Rivet toolkit with an *Anti-*k_T algorithm
- YODA
- ROOT
- Matplotlib

General Variables:

- Energy of jet
- P_t of jet
- Invariant mass of jet
- Charge particle multiplicity
- Width of jet (ΔR)
- Charge of jet
- Phi, pseudorapidity (η)

Substructure Variables:

- Energy Correlation Fuction

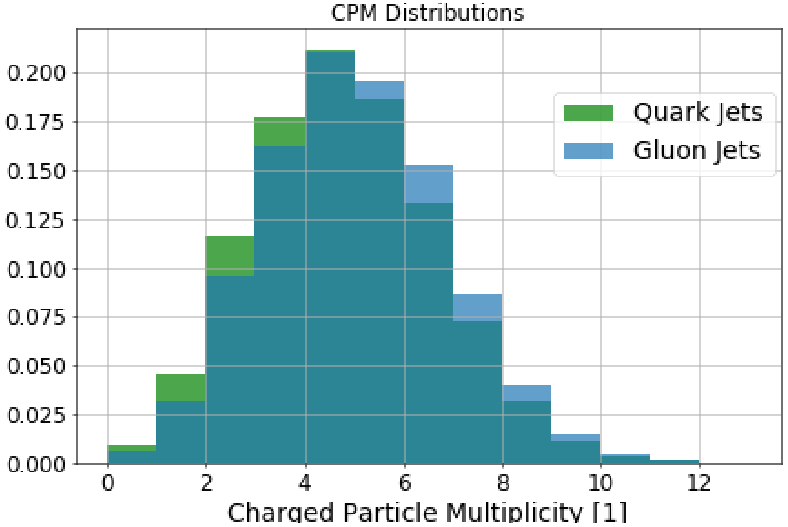
  $e_2^{0.2} = \sum_{i<j\in jet} z_i z_j \; \Delta R_{ij}^{0,2} \qquad z_i = \frac{P_{t,i}}{\sum_{j\in jet} P_{t,j}}$

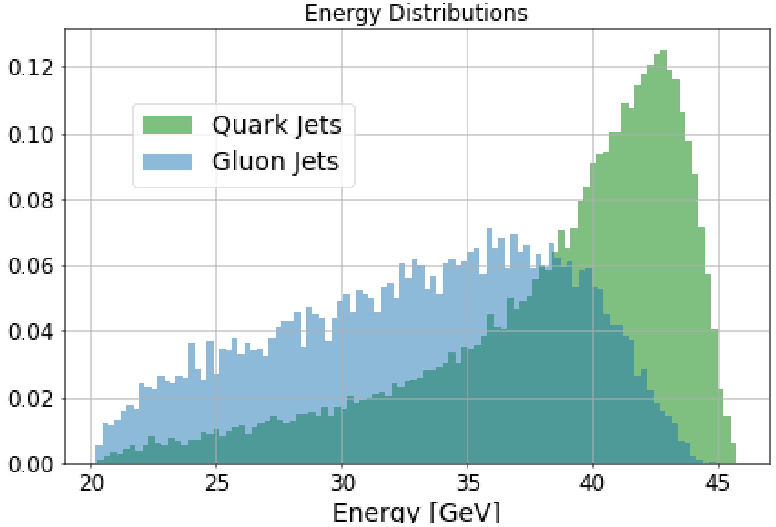
- Jet Angularity

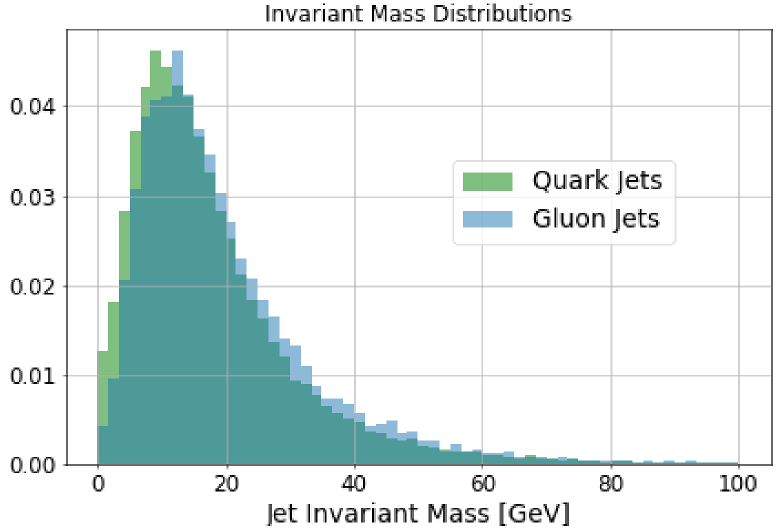
  $\lambda_2^1 = \sum_{i\in Jet} z_i \, \theta_i^2 \qquad \theta_i \equiv \frac{R_{i,jet}}{\Delta R}$

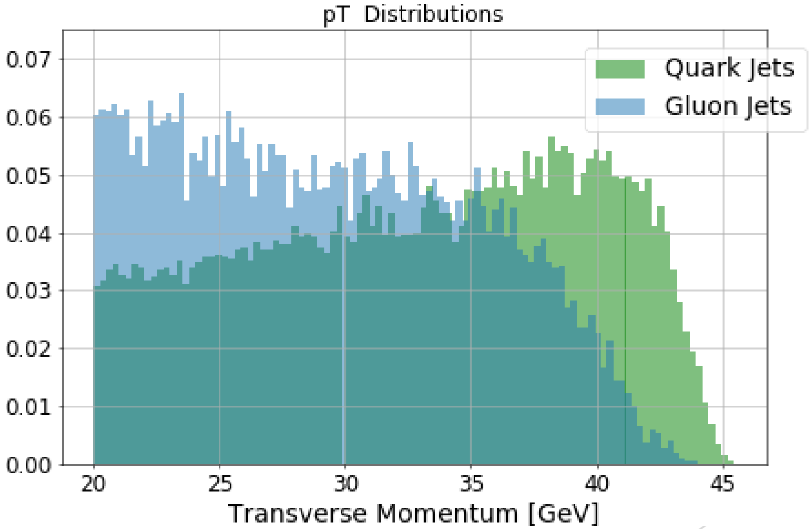
- Pull (along phi)
- Generalized moments

Distributions of variables



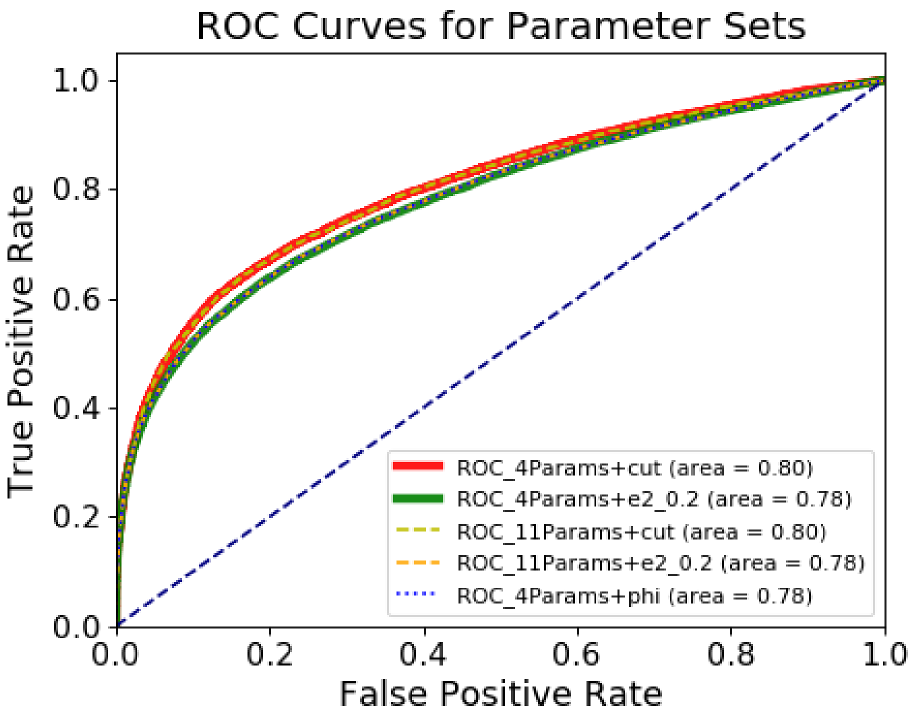






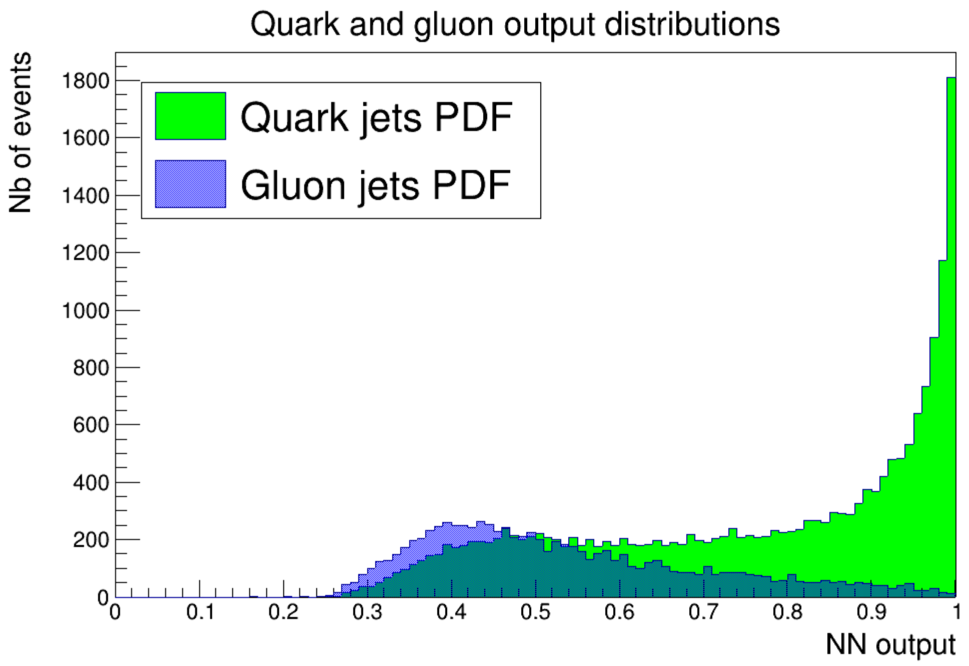
Neural Network

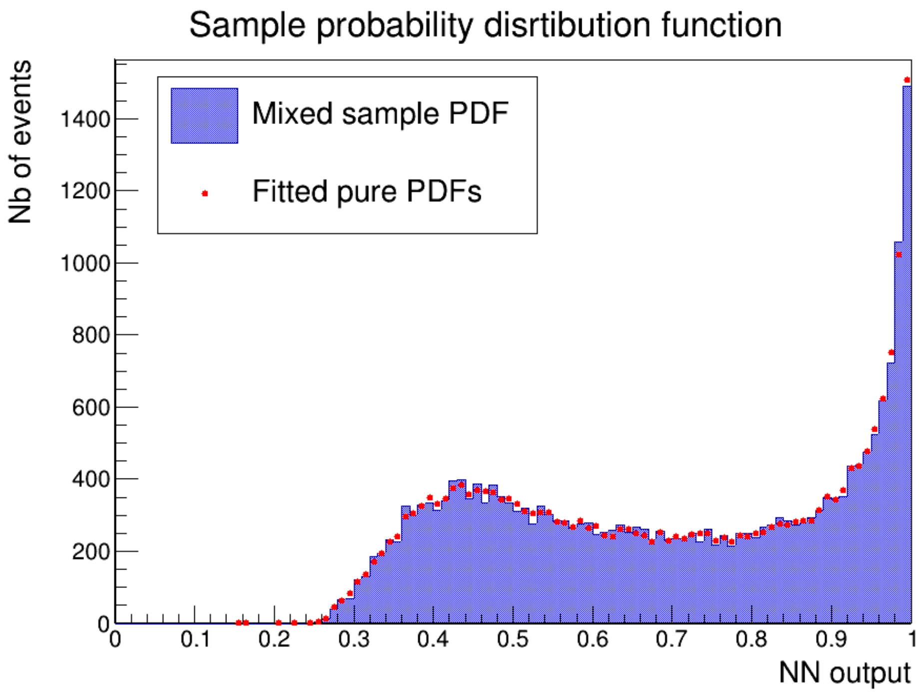
- Hep_ml python package
- Multilayer Perceptron Classifier (feedforward NN)
- Variety of parameter combination and cuts tried
- For the final NN, the employed setup was
  - ➢ two *hidden* layers with 6 and 3 neurons in each
  - ➢ loss function – logarithmic
  - ➢ optimization method – irprop



- Final set of variables and cuts:
  - ➢ Energy of jet
  - ➢ P_t of jet
  - ➢ Invariant mass of jet
  - ➢ Charge particle multiplicity
  - ➢ $e_2^{0.2} < 2.7$ cut

Results





Sample 1

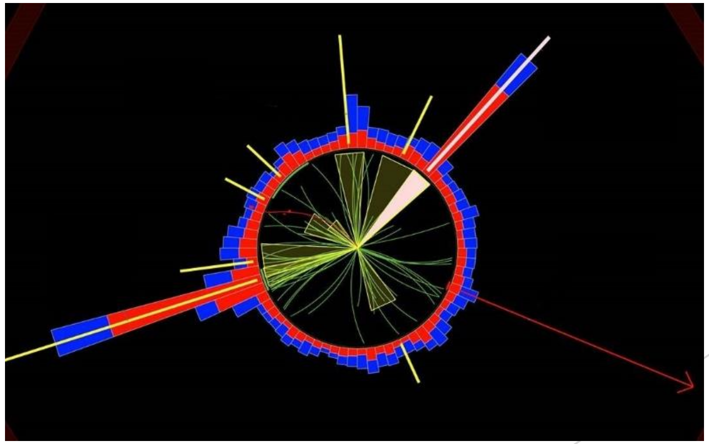
Gluon True	$30{,}3\%$
Quark True	$69{,}7\%$
Gluon Fit	$(29{,}7 \pm 0.9)\%$
Quark Fit	$(70{,}2 \pm 1.1)\%$

Sample 2

Gluon True	$50{,}1\%$
Quark True	$49{,}9\%$
Gluon Fit	$(52{,}8 \pm 1{,}5)\%$
Quark Fit	$(47{,}2 \pm 1{,}2)\%$

Conclusion and Outlook

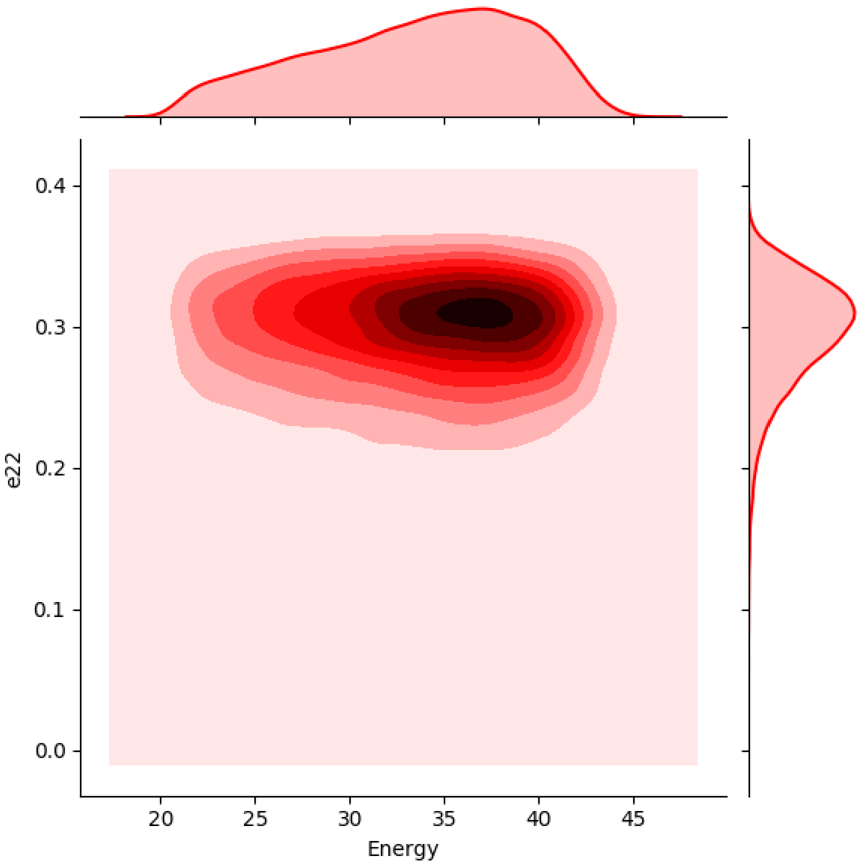
- Two pure samples generated
- Parameter distributions and correlations investigated
- Parameter sets and cuts tested by ROC curves
- Neural network trained and used to determine output PDFs for quarks and gluons
- Mixed samples fitted by PDFs with precision $(1 - 1.5)\%$
- The area under the ROC curve suggests single jet classification accuracy of about 80% is achieved



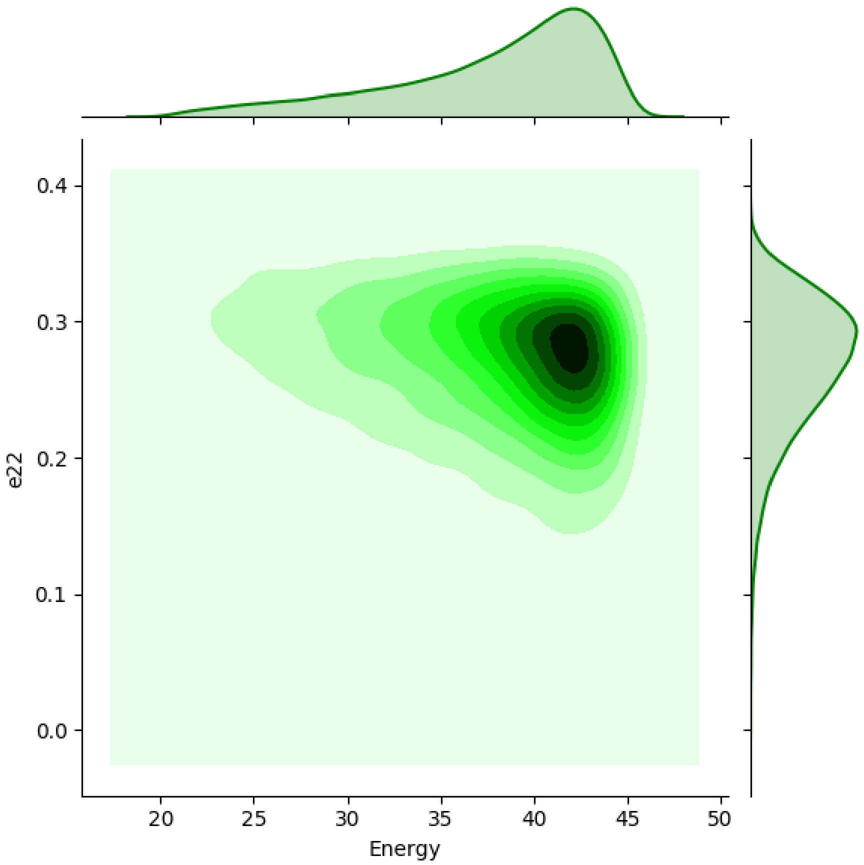
- Possible extensions of the project
  - ➢ Verification on experimental data
  - ➢ Extension to more complex collisions
  - ➢ Comparison of different GPMC's
  - ➢ Application of complex neural networks
  - ➢ Increase in the energy range

Thank you for your attention!

Back Up – e2_0.2 cut

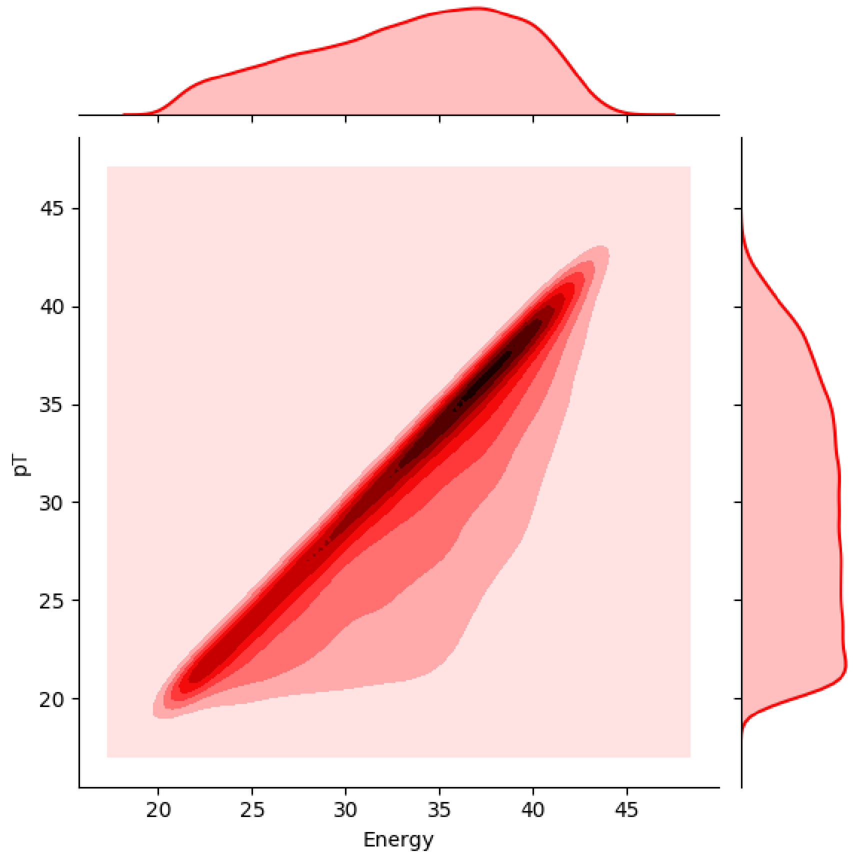


Gluons

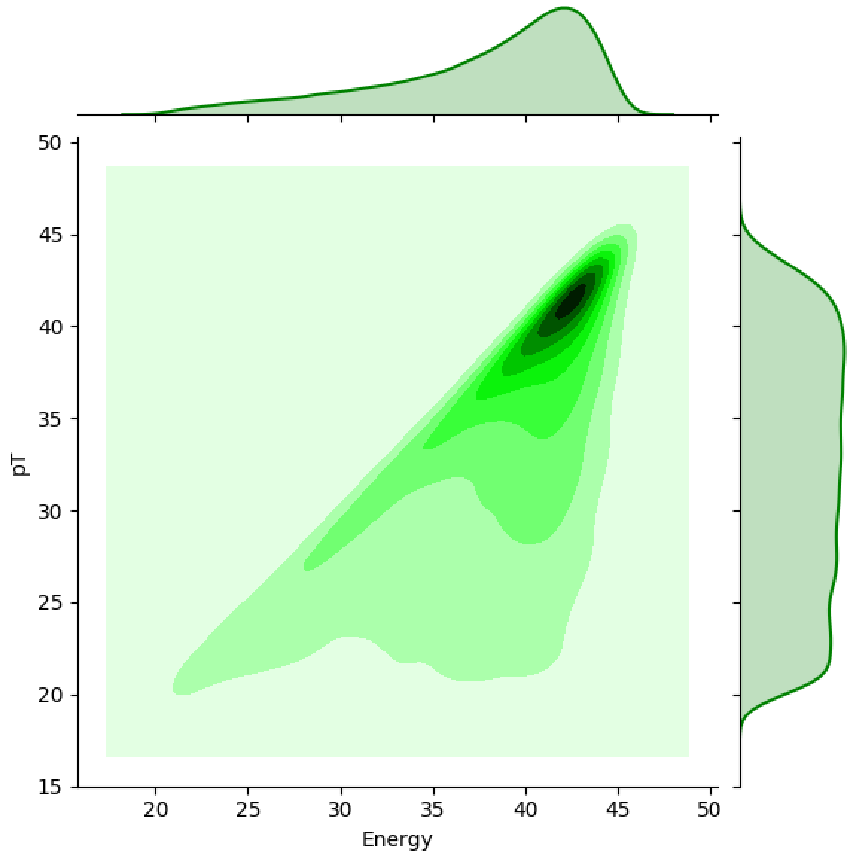


Quarks

Back Up – pT-E correlation

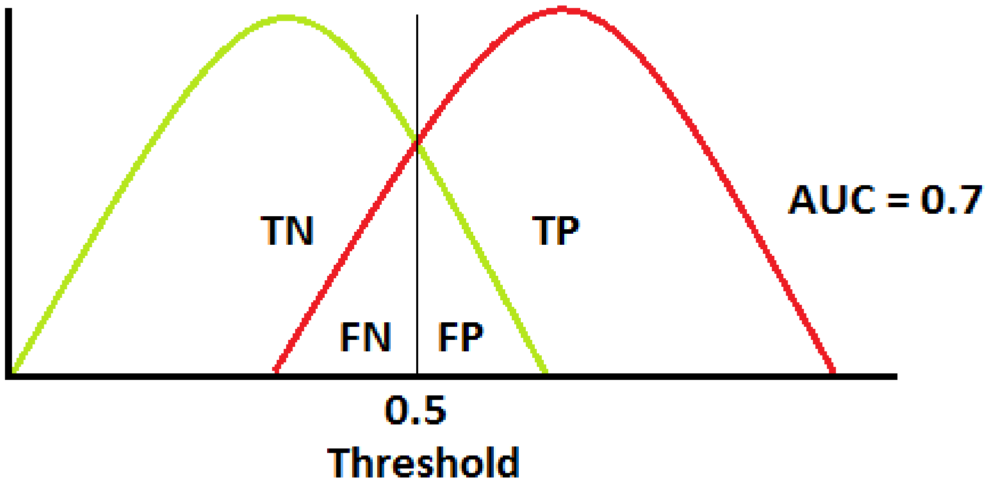


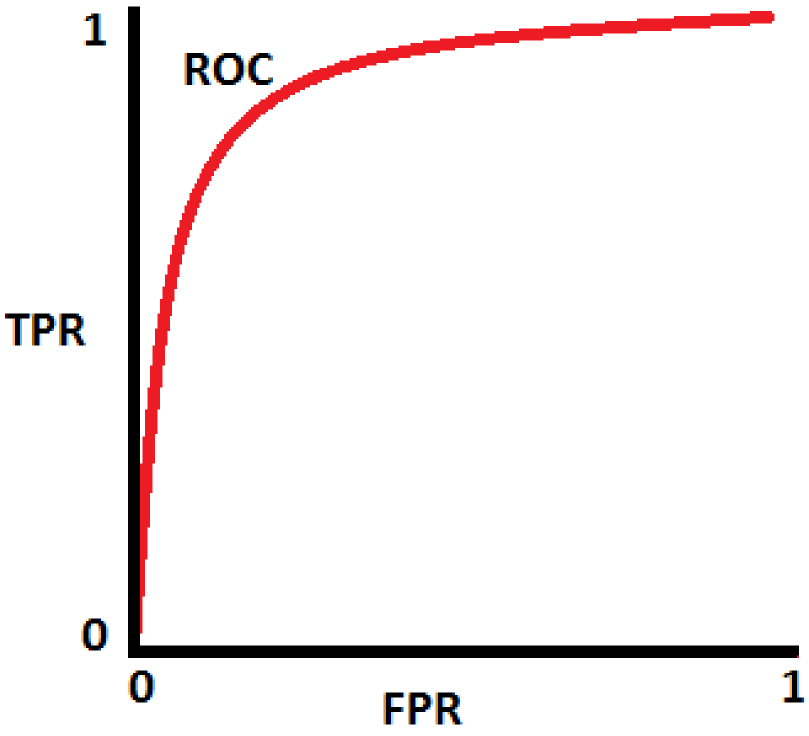
Gluons



Quarks

Back up – ROC curves





Back up up – ROC curves 2