The LHCal study

Particle identification

Oleksandr Ustinov, Yuriy Onishchuk

Kiev Taras Shevchenko National University

18-Sep-2016

General remarks

- **The LHCal as a part of the Forward Calorimeter**
- **Particle identification in the LHCal**
- **Particle type classification**
- **Machine learning to solve the classification problem**

Machine Learning

Main features:

- Machine learning (“**ML**”) explores the study and construction of algorithms that can learn from and make predictions on data
- Such algorithms operate by building a model from example inputs in order to make data-driven predictions or decisions,rather than following strictly static program instructions
- Used to devise complex models and algorithms that lend themselves to prediction

Types of ML:

- **Classification** – identifying to which category an object belongs to
- **Regression** – predicting a continuous-valued attribute associated with an object
- **Clustering** – automatic grouping of similar objects into sets
- **Model selection** – comparing, validating and choosing parameters and models
- **Dimension reduction** – reducing the number of random variables to consider
- **Preprocessing** – feature extraction and normalization

Machine Learning

Models for Classification in ML:

- Support Vector Machines
- Decision Trees
- Nearest Neighbors
- Naive Bayes

Use **Naive Bayes** algorithm from the Python scikit-learn package for Particle classification

Particle identification in LHCal

Used parameters to test ML possibilities in PID:

- Deposited Energy

$$E_{dep} = \Sigma_{i=0}^{n} E_i$$

- Number of cells n
- Mean Z-coordinate

$$\langle Z \rangle = \frac{\Sigma_{i=0}^{n} Z_i \cdot E_i}{\Sigma_{i=0}^{n} E_i} = \frac{\Sigma_{i=0}^{n} Z_i \cdot E_i}{E_{dep}}$$

- σ_L - longitudinal energy dissipation according to track direction:

$$\sigma_L^2 = \frac{\Sigma_{i=0}^{n} (L_i - L_{grav})^2 \cdot E_i}{E_{dep}}$$

  where L_{grav} – position of the track center of gravity
- σ_T - transversal energy dissipation according to track direction

$$\sigma_T^2 = \frac{\Sigma_{i=0}^{n} R_i^2 \cdot E_i}{E_{dep}}$$

  where R_i – distance to the track

Particle identification in LHCal

Particle types divided on **3 groupes**:

- **Muons** (μ) – exclusively ionisation energy losses
- **EMs** (electrons and γ) as EM shower produced particles
- **Hadrons** (π, K) as nuclearly and ionisationally interacted particles

Events with π and K like as μ – removed

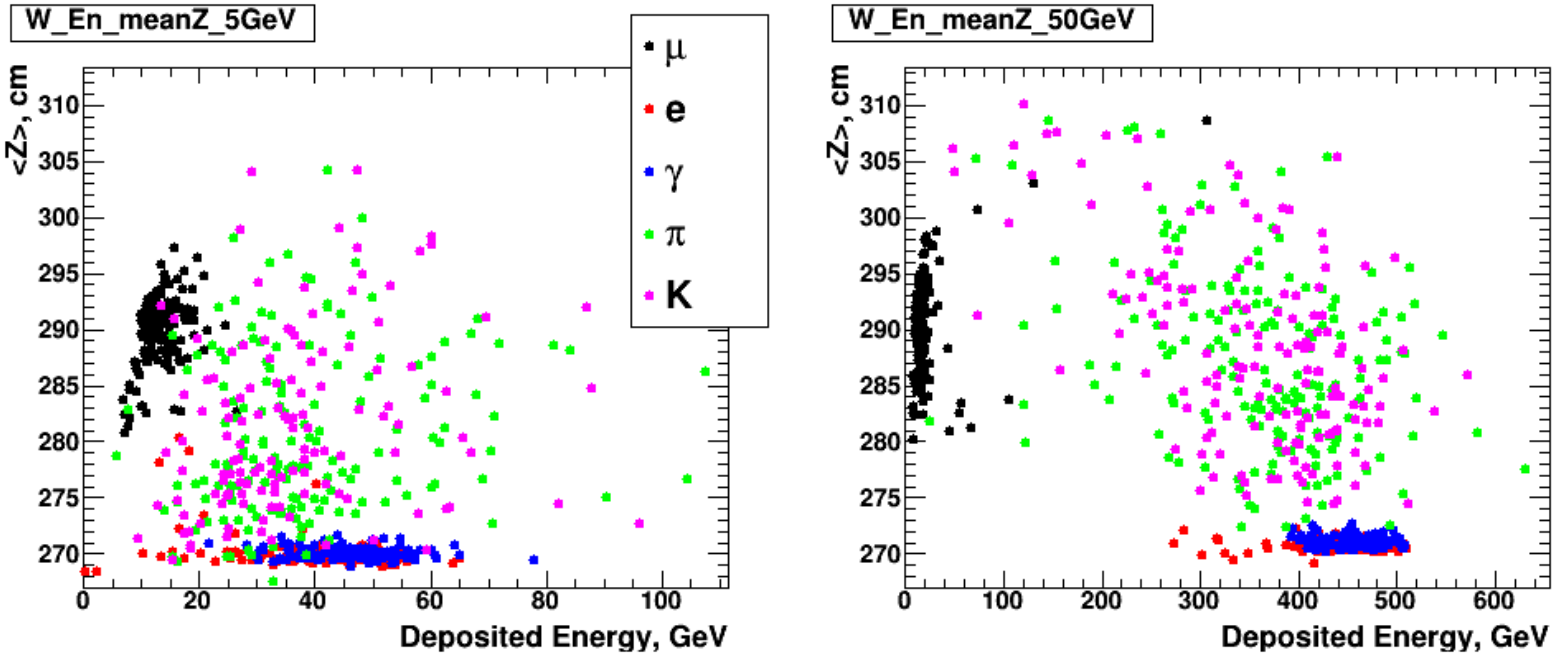
Tested combinations of parameters:

- E_{dep} vs $\langle Z \rangle$
- σ_L vs σ_T
- E_{dep} vs σ_L, σ_T
- $\langle Z \rangle$ vs σ_L, σ_T
- E_{dep} vs number of cells
- E_{dep} vs $\langle Z \rangle$ vs σ_L, σ_T
- E_{dep} vs $\langle Z \rangle$ vs number of cells vs σ_L, σ_T

Identification quality criterion:

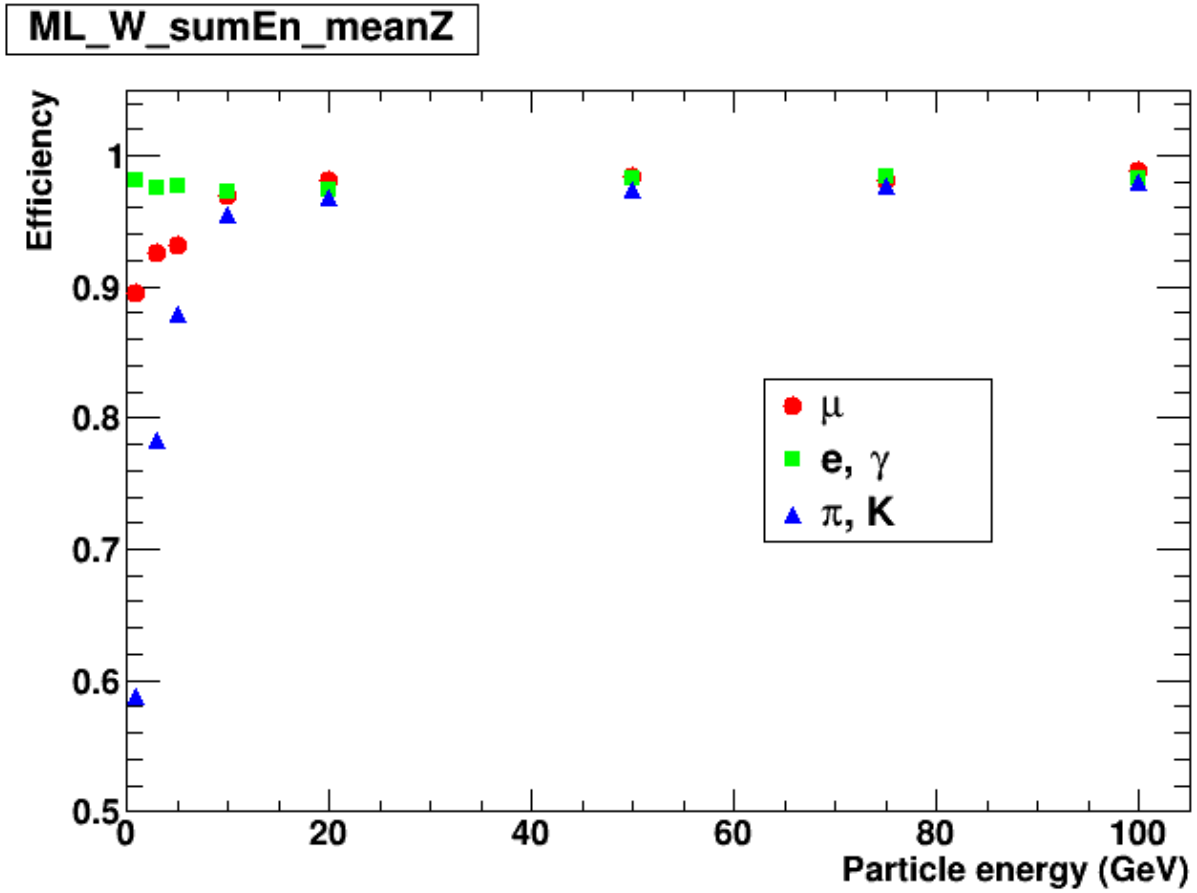
$$Efficiency = 100\% \cdot (Y^{true}_{prediction} / N_{samples})$$

E_{dep} vs $\langle Z \rangle$



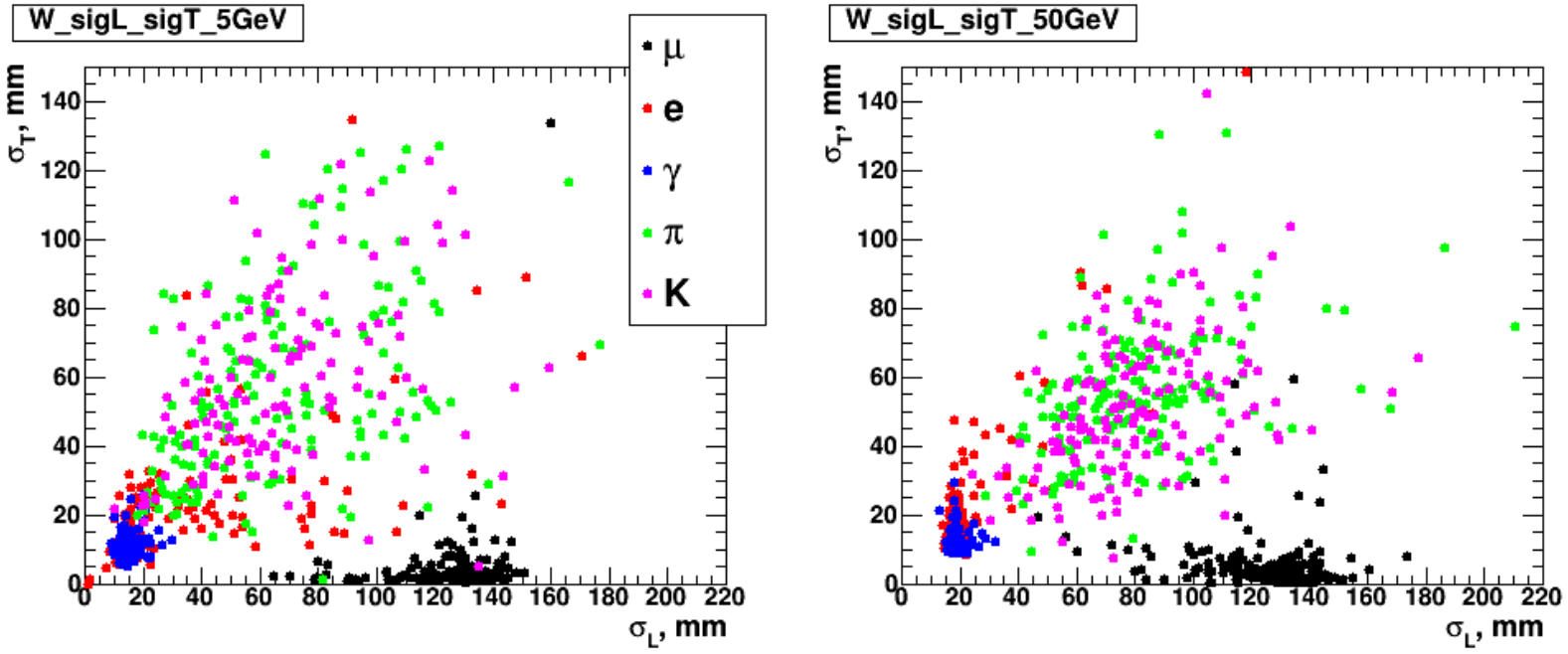
- Reliable localization of events caused by μ and EMs
- K and π populate an intermediate region
- Possible good separations for all particle types

E_{dep} vs $\langle Z \rangle$



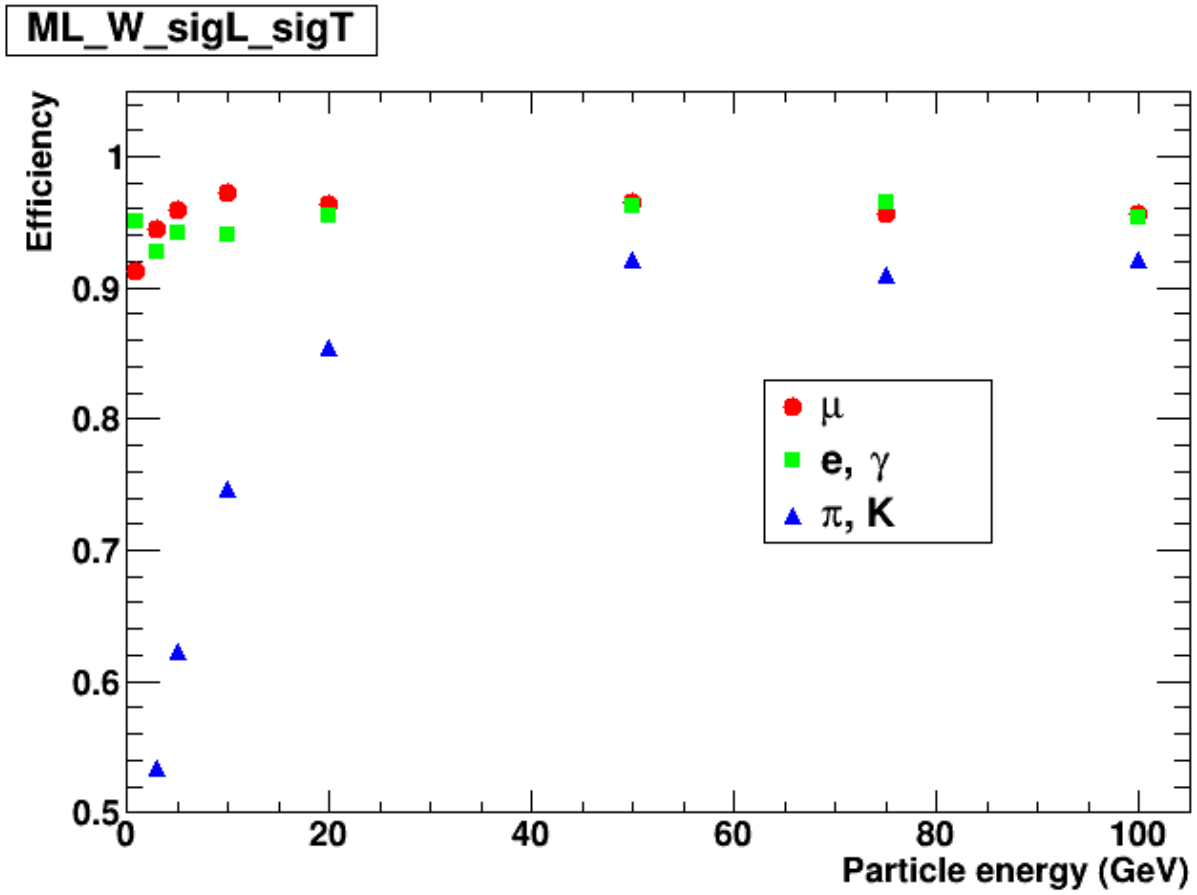
- Separation improves with energy enlargement
- Good Efficiency of the particle identification for EMs and μ within the whole energy range
- Good Efficiency for hadrons with Particle Energy > 10 GeV

σ_L vs σ_T



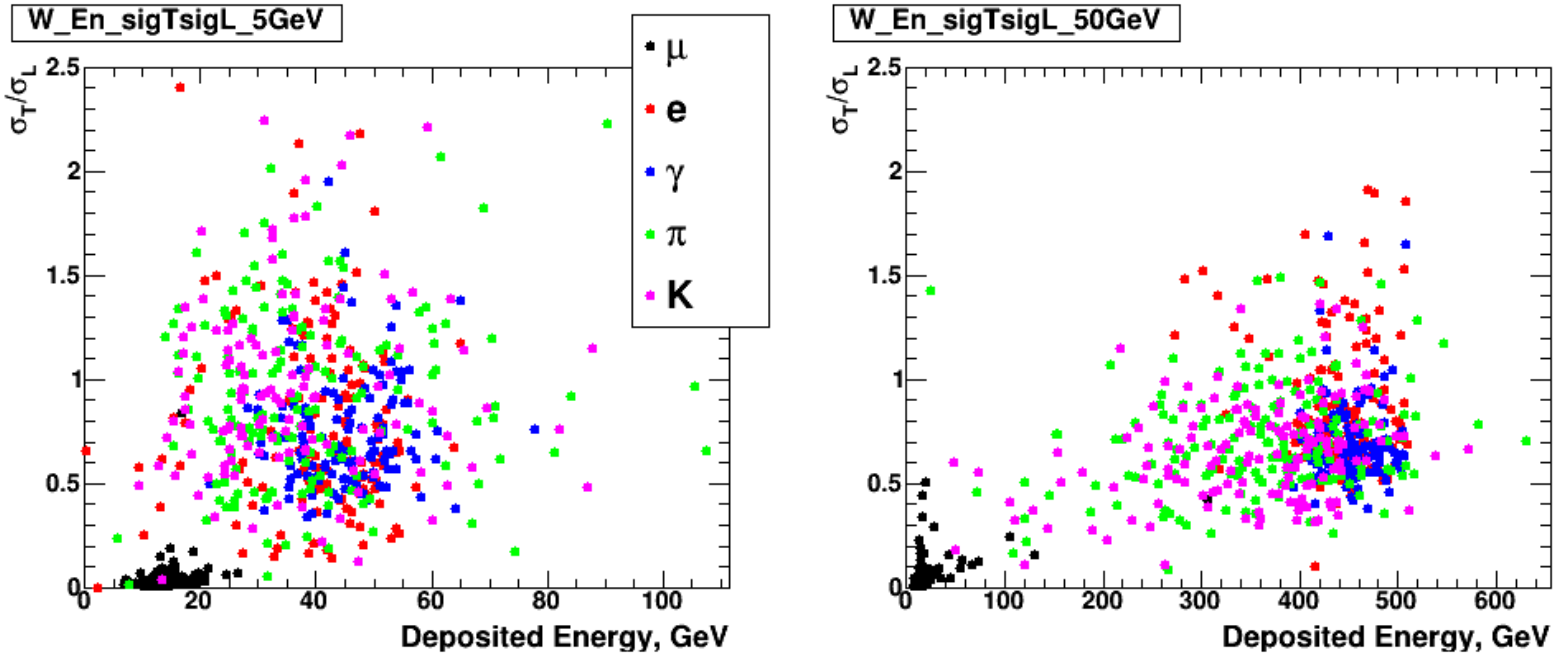
- For μ $\sigma_L \gg \sigma_T$ (long line tracks): Reliable localization of events
- EMs have $\sigma_L \simeq \sigma_T$ (good deposited energy localisation)
- Generally σ_L, σ_T for hadrons larger than for EMs (wide space dispersion of deposited energy)
- Low Particle Energy events overlapped with EMs (localisation in the Forward part of the LHCal)

σ_L vs σ_T



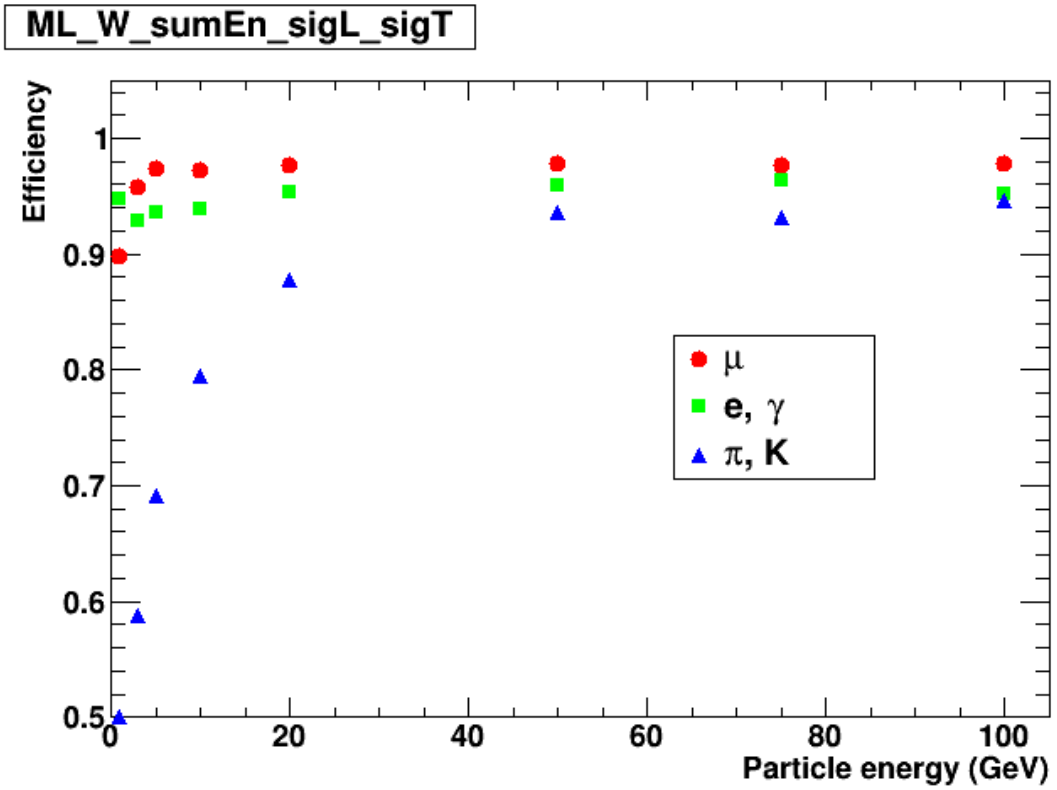
- Good Efficiency of the particle identification for EMs and μ within the whole energy range
- Poor Efficiency for hadrons with Particle Energy < 20 GeV

E_{dep} vs σ_L/σ_T



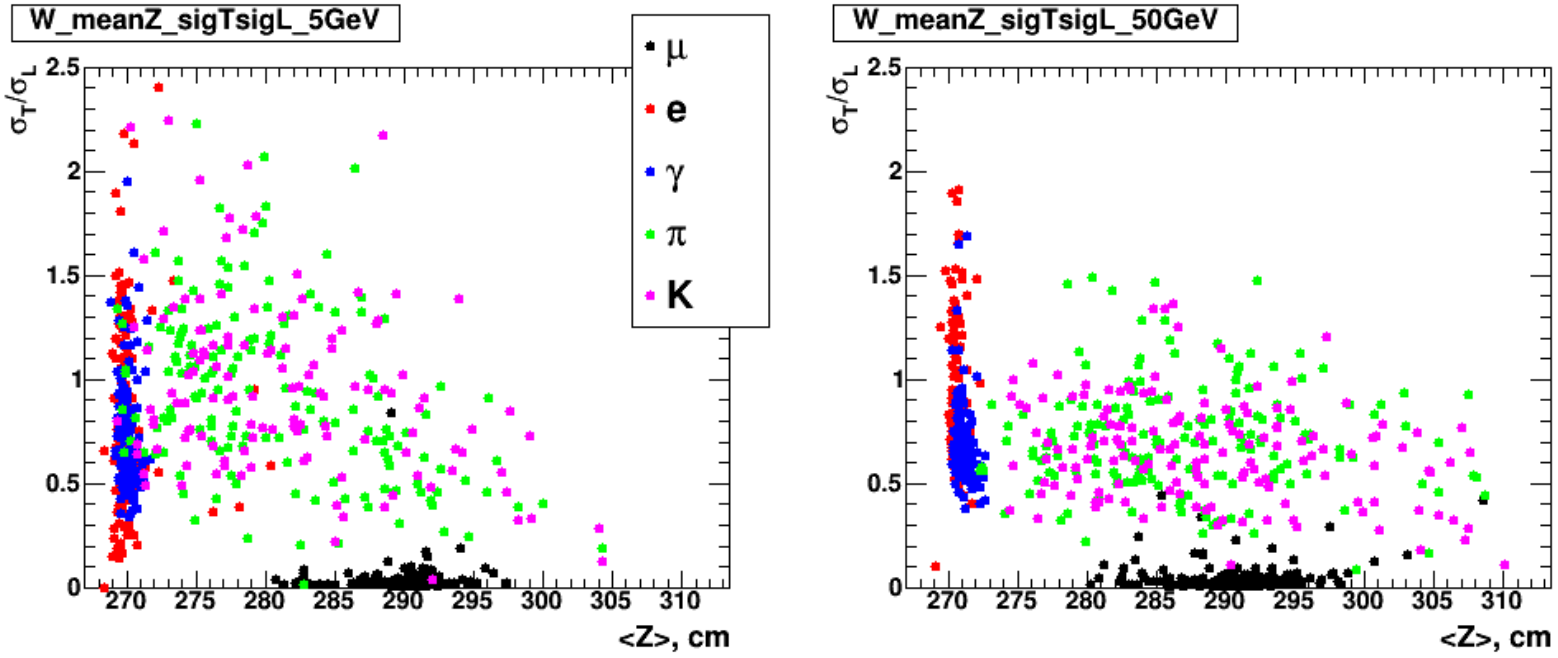
- Reliable localization of events caused by μ
- Sufficient overlapping of events for Hadrons and EMs
- Energy enlargement improves the separation of Hadrons and EMs

E_{dep} vs σ_L/σ_T



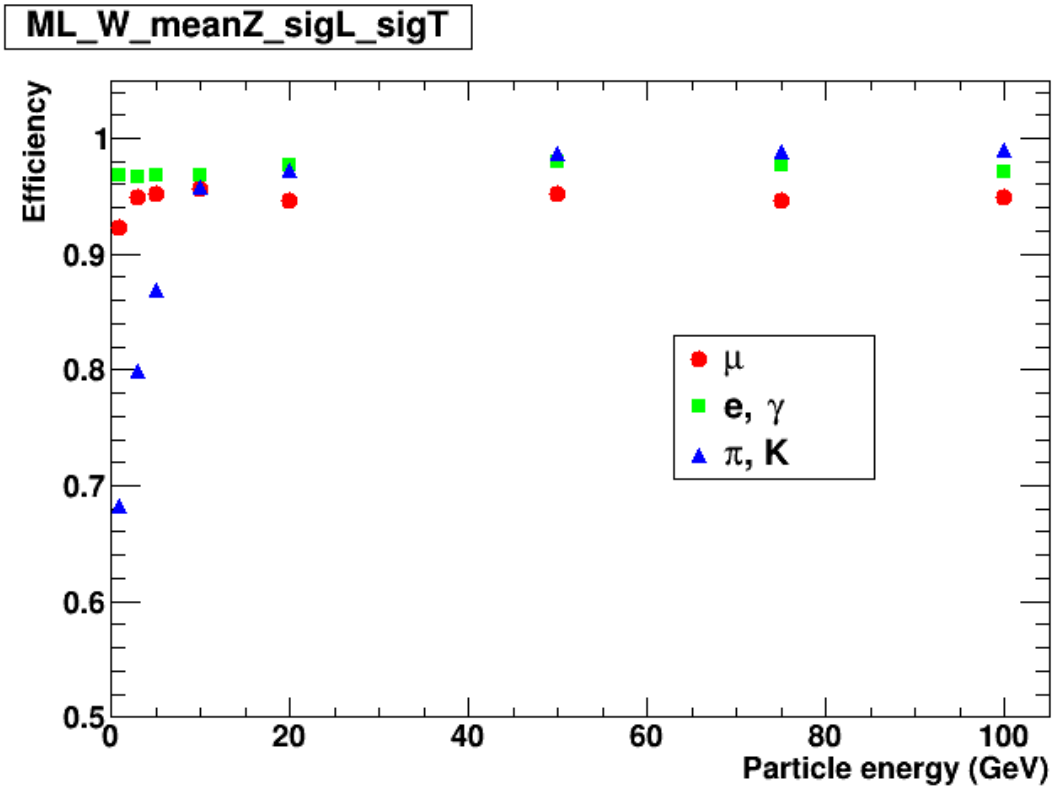
- Good Efficiency of the particle identification for EMs and μ within the whole energy range
- Poor Efficiency for hadrons with Particle Energy < 20 GeV

$\langle Z \rangle$ vs σ_L/σ_T



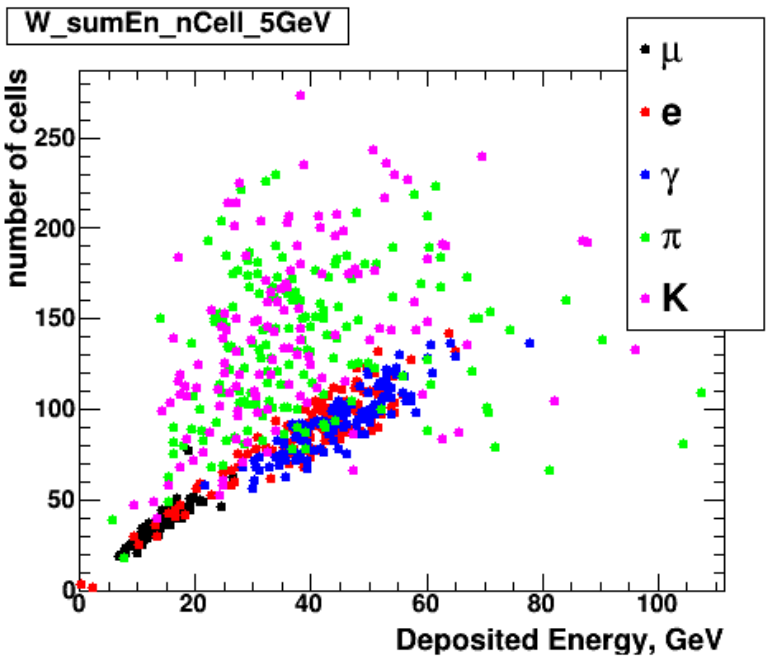
- Good separations for hadrons, μ, EM with Particle Energy > 10 GeV

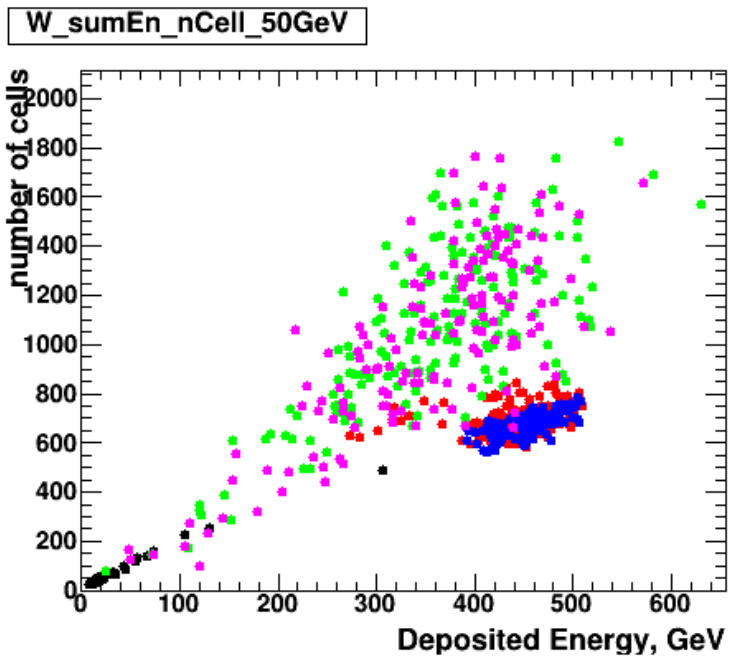
$\langle Z \rangle$ vs σ_L/σ_T



- Good Efficiency of the particle identification for EM and μ within the whole energy range
- Good Efficiency for hadrons with Particle Energy > 10 GeV

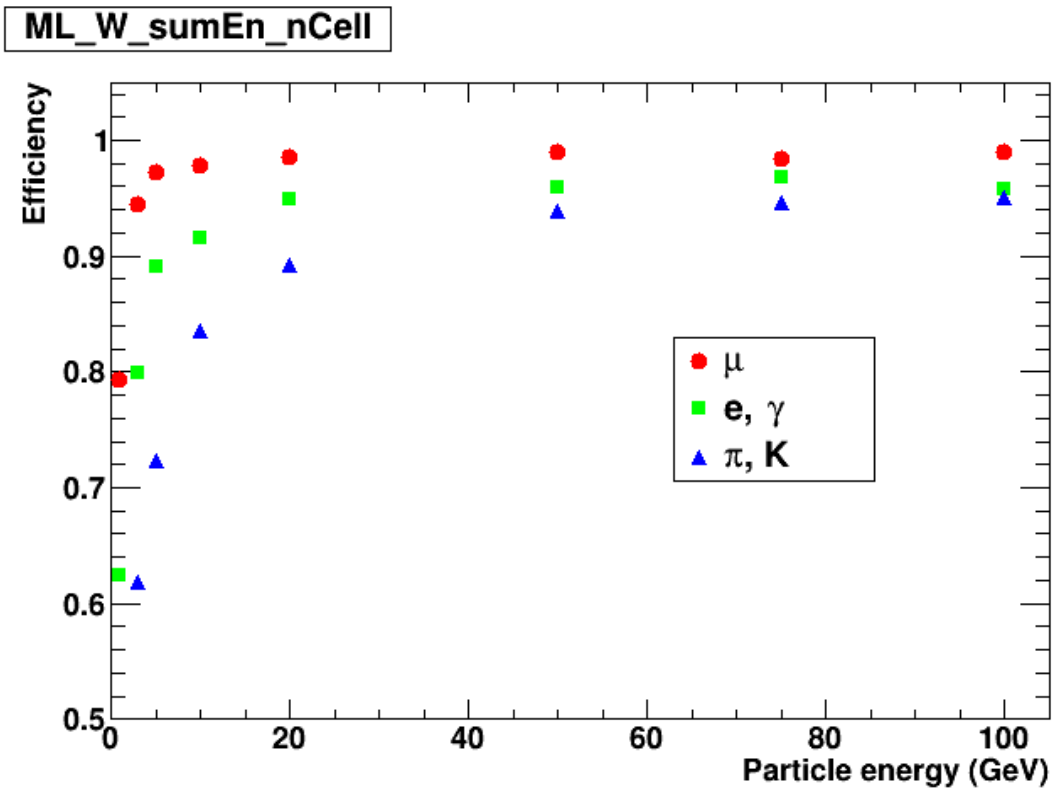
E_{dep} vs number of cells





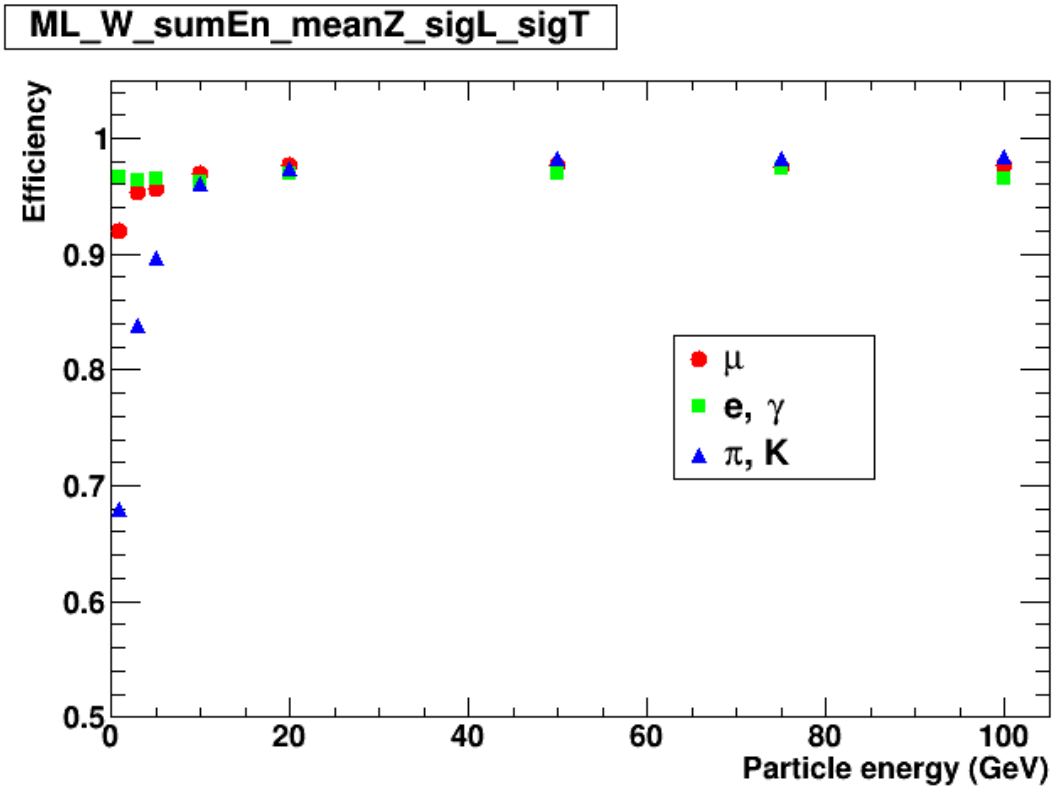
- μ and EMs events have linear dependencies vs Particle Energy
- Hadrons have more wider dispersion

E_{dep} vs number of cells



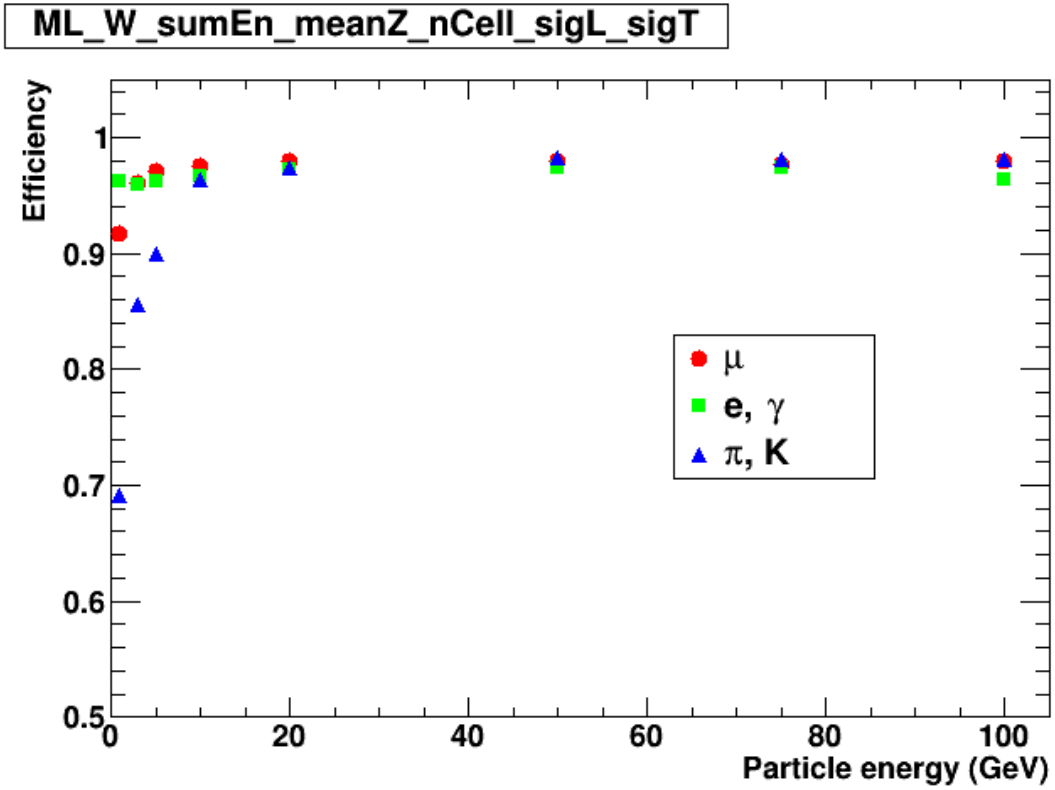
- Good Efficiency of the particle identification for EM and μ with Particle Energy > 10 GeV
- Poor Efficiency for hadrons with Particle Energy < 20 GeV

E_{dep} vs $\langle Z \rangle$ vs σ_L, σ_T



- The Efficiency dependence for 4 parameters
- Good Efficiency of the particle identification for EM and μ within the whole energy range
- Good Efficiency for hadrons with Particle Energy > 10 GeV

E_{dep} vs $\langle Z \rangle$ vs number of cells vs σ_L, σ_T



- The Efficiency dependence for 5 parameters
- The best results for the Efficiency
- Nice identification for all types of particles with energies larger than 10 GeV

Conclusions

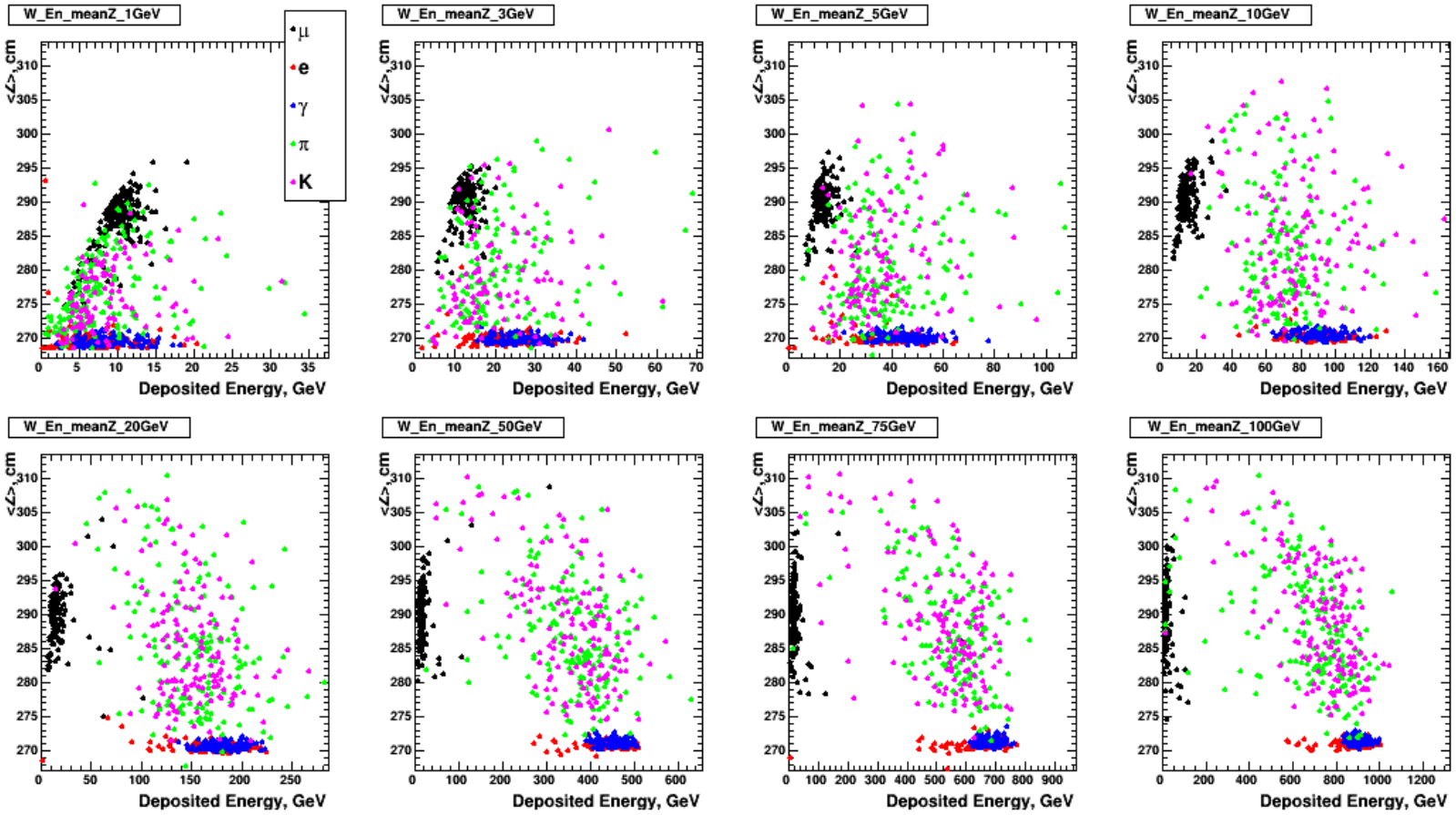
Conclusions

- Machine Learning possibilities in particle identification for the LHCal were studied
- Efficiency $\geq$ 95% for the particle identification have been obtained for the W-Si sandwich within the energy range of 1-100 GeV
- Nice identification based on ML for all types of particles with energies larger than 10 GeV using combination of 5 parameters

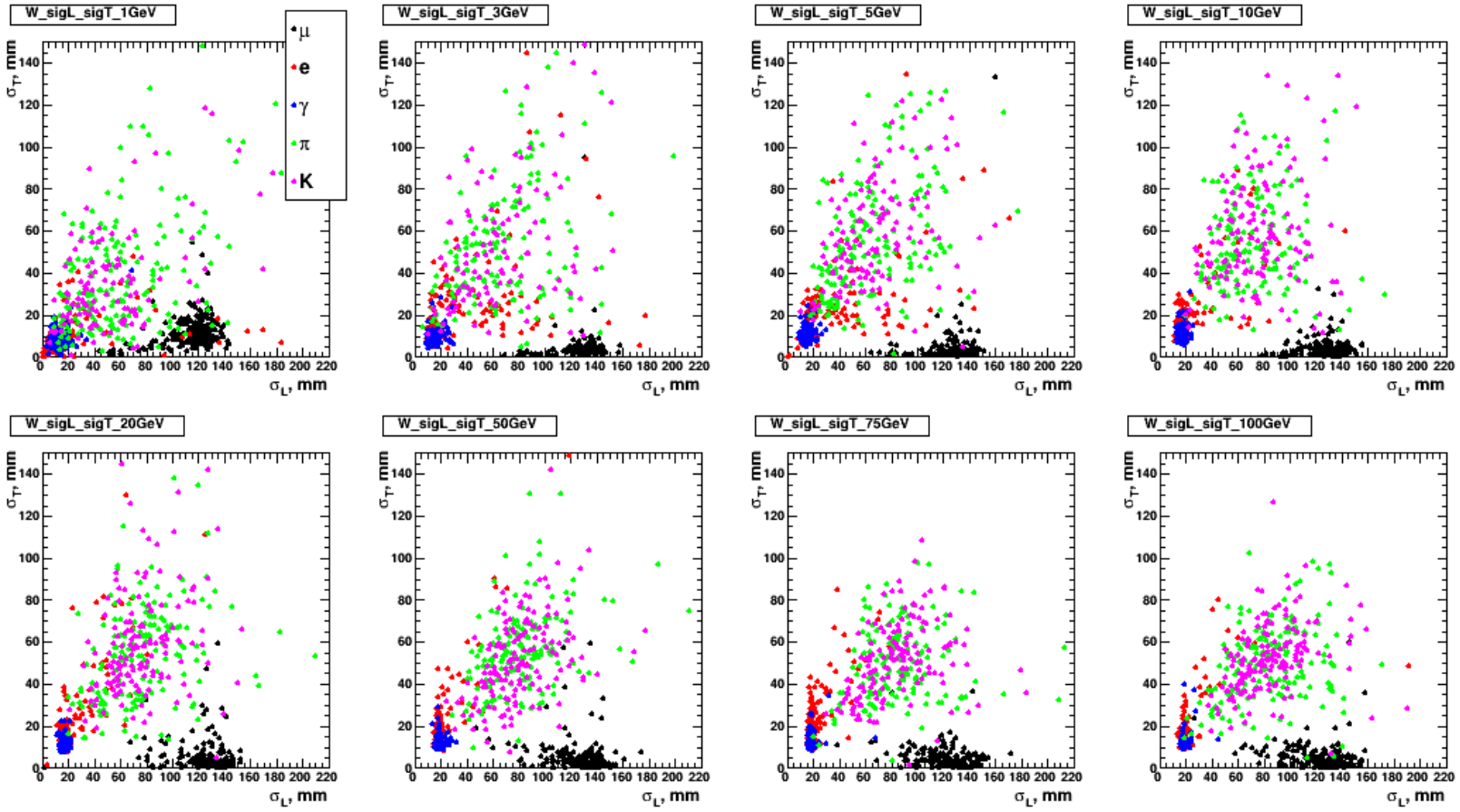
Thanks for attention

Upload slides

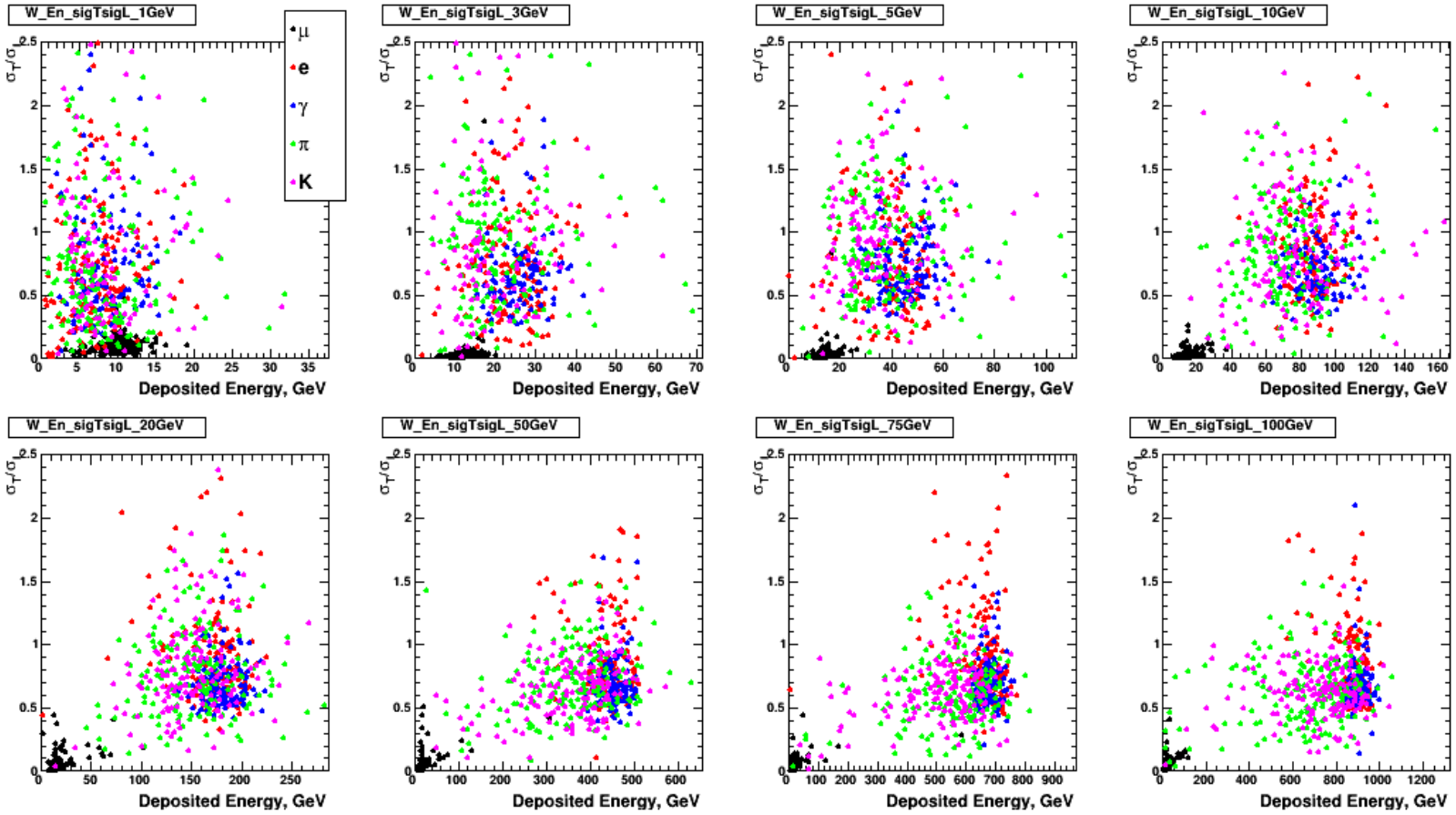
Deposited Energy vs $\langle Z \rangle$



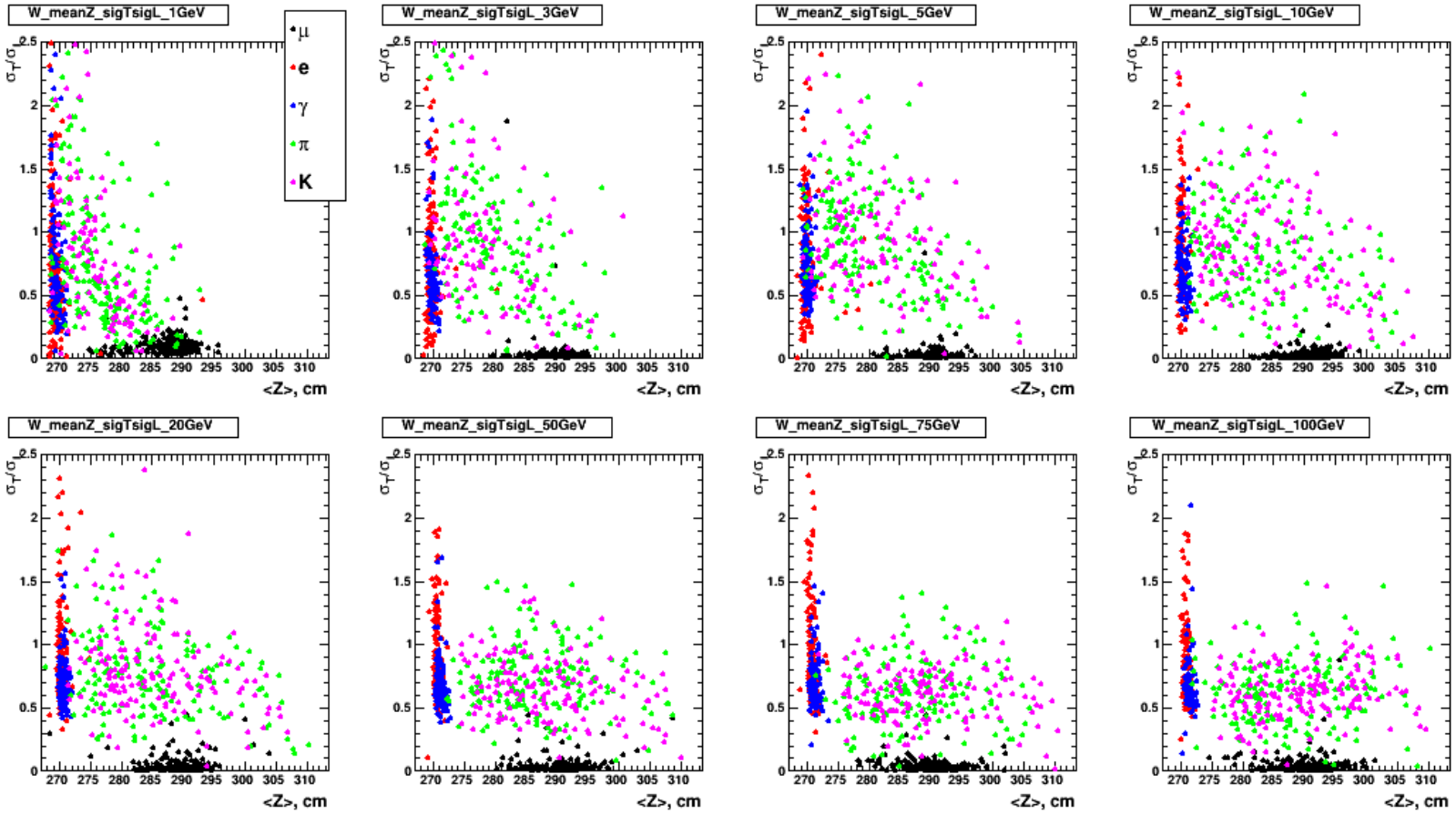
σ_L VS σ_T



Deposited Energy vs σ_L/σ_T



$\langle Z \rangle$ vs σ_L/σ_T



Deposited Energy vs number of cells