Full-Geometry Fast Calorimeter Simulation and Steps towards a Back Transformation

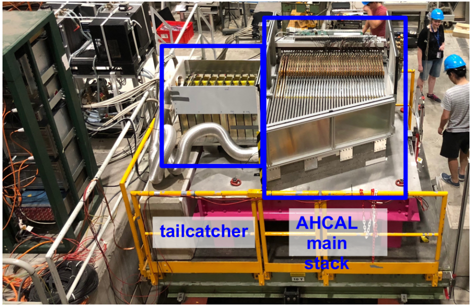
Julian Utehs
In collaboration with Stan Lai and André Wilhahn

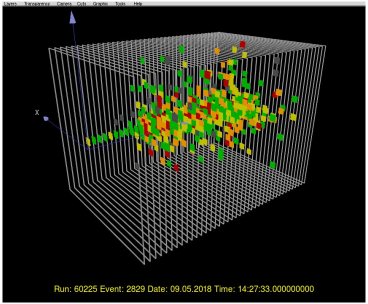
II. Physikalisches Institut, Georg-August-Universität Göttingen
CALICE Collaboration Meeting

September 27 - 29, 2023

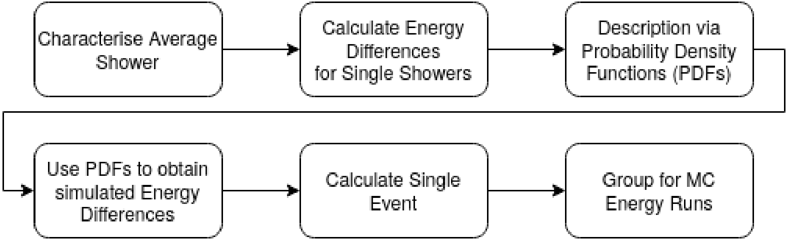
The AHCAL Prototype

- Consists of 38 active layers of 24×24 scintillator tiles ($3 \times 3\,\text{cm}^2$) embedded in stainless-steel absorber structure with analogue SiPM read-out electronics
- Full simulation on particle level is not data-driven and might need many adjustments
- Investigation of **data-driven fast calorimeter simulation** based on pion showers
  $\Rightarrow$ Aim: Better performance and higher accuracy



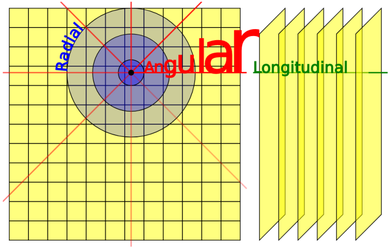


Characterisation of Shower Events



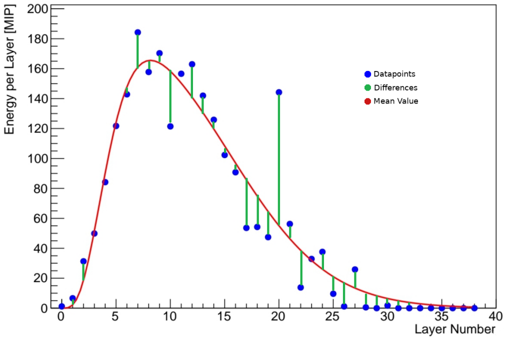
The **pion shower** disperses in a cone-shaped manner. Hence, it is divided into three parts for a parameterisation:

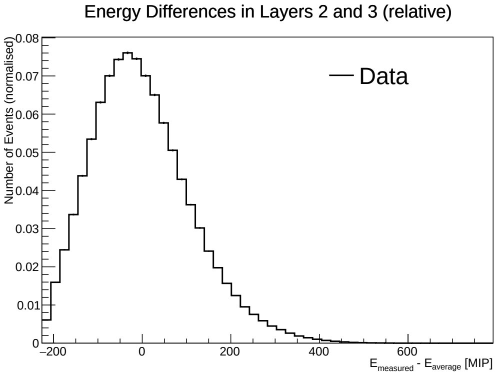
- Longitudinal
- Radial
- Angular



Data Preparation

- Consider differences in energy deposition between single and average shower instead of absolute energies
- Investigate energy deposition relative to shower start layer and center of gravity





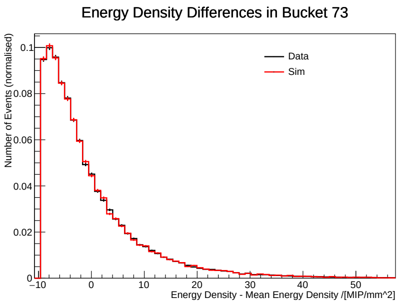
Kernel Density Estimators

- Want to find PDF of dataset x_1, x_2, ..., x_n
- Define Kernel Density Estimator (KDE) with bandwidth h as:

$$f(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right)$$

with

$$K(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right)$$



- PDF = sum of all Gaussian kernels
- Choice of bandwidth determines smoothness of PDF
- Apply KDE of energy differences simultaneously on layer groups

Cuts and Configurations

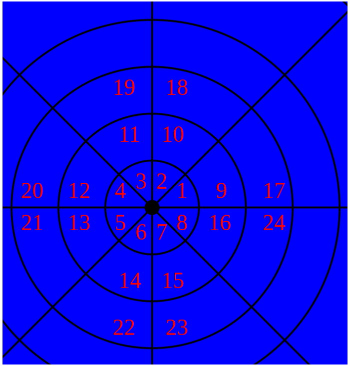
Cuts

- Applied particle identification using BDT-techniques to remove beam contamination
- Exclude first physical AHCAL layer in order to minimise uncertainty in shower start finding algorithm
- Apply low-energy cut to remove muons from dataset
- Exclude events starting in layer 11 or later to minimise leakage

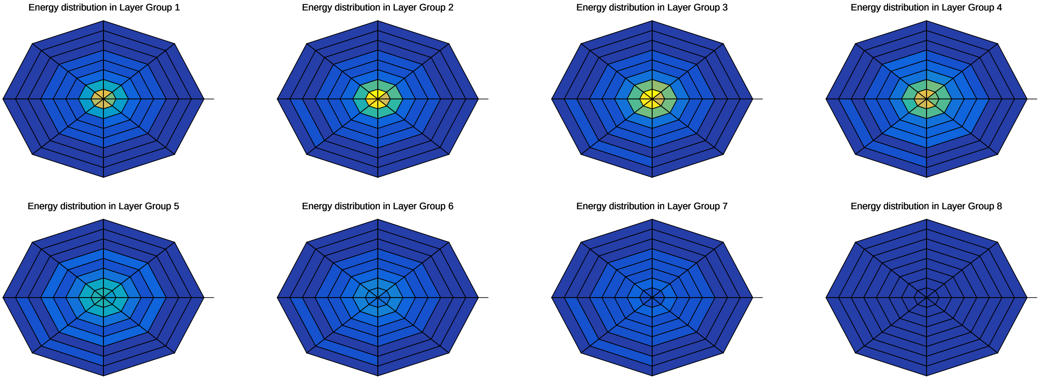
Cuts and Configurations

Configurations

- Layer Groups {1-2, 3-4, 5-6, 7-8, 9-12, 13-16, 17-24, 25-40}
- Angular Groups $\{\frac{\pi}{4}, \frac{2\pi}{4}, \frac{3\pi}{4}, \frac{4\pi}{4}, \frac{5\pi}{4}, \frac{6\pi}{4}, \frac{7\pi}{4}, \frac{8\pi}{4}\}$
- Radial Groups {25, 50, 75, 100, 150, 200, 300, 400} (in mm from Center of Gravity)

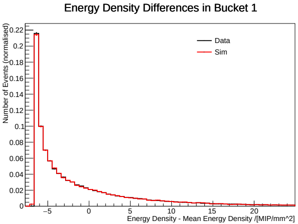


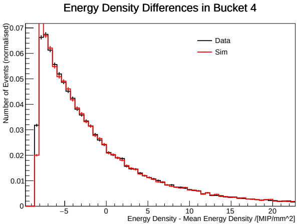
Mean Energy per Bucket

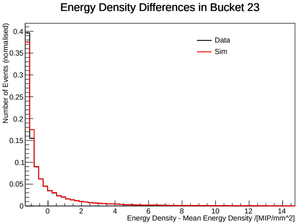


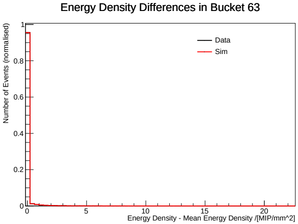
Energy Differences per Bucket in One-step Simulation of 3 Spatial Dimensions

- One-step Simulation of 3 Spatial Dimensions shows good agreement









Pros and Cons

Pro

- One-step Simulation of 3 Spatial Dimensions possible
- Directly usable values

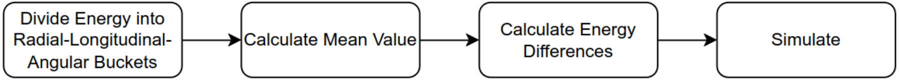
Con

- Analysis takes quite a time ($\sim$ 100 seconds per 100 events $\Rightarrow$ 3.5 seconds per 100 events)
- Low number of hits in the buckets $\Rightarrow$ Low statistics!

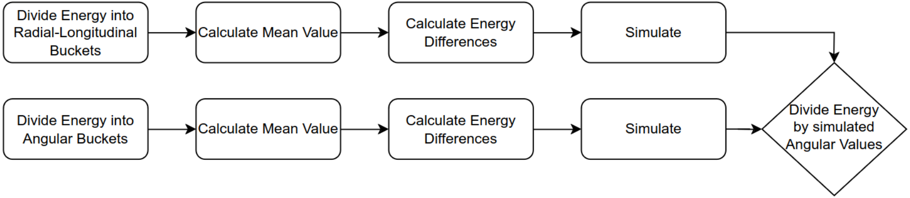
Idea:
Why not use a Two-Step Simulation: First calculate the Radial-Longitudinal part and allocate the angular part later?

Two-step Simulation

One-step Simulation:

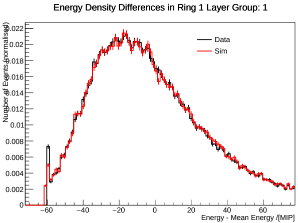


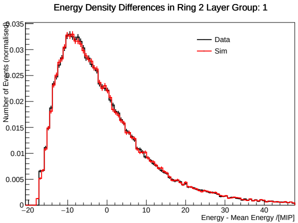
Two-step Simulation:

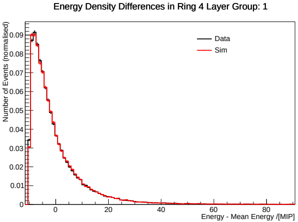


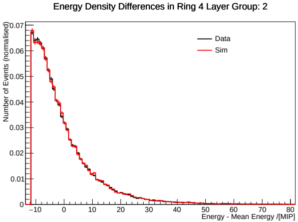
Energy Differences per bucket in Radial-Longitudinal Analysis

- Radial-Longitudinal Simulation shows good agreement



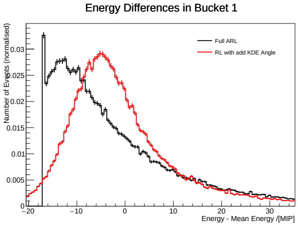


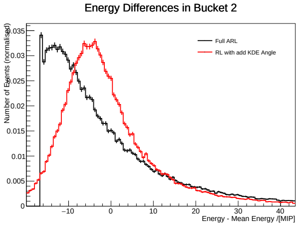


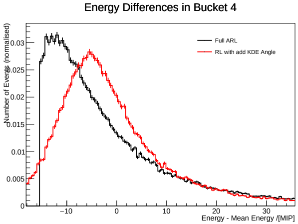


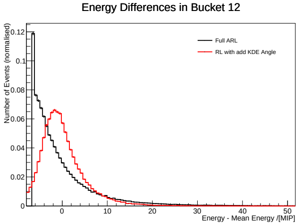
Energy Differences - One-step Sim. compared to Two-step Sim.

- One-Step Simulation and Two-step Simulation are not in agreement



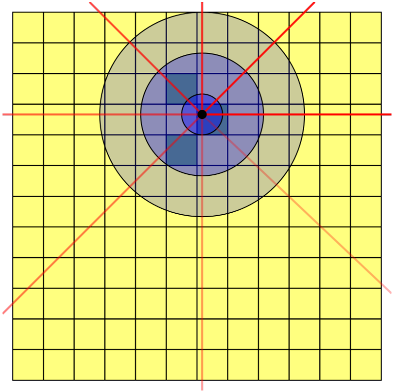


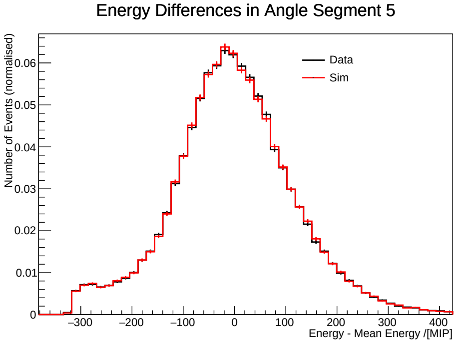




Problems

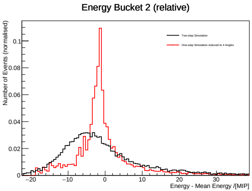
- Angular Part corresponds to a gaussian curve
- Angular Part in combination with small radii can easily be zero

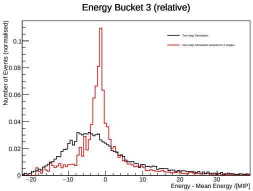


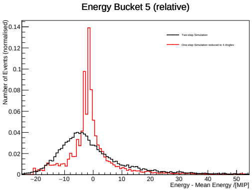


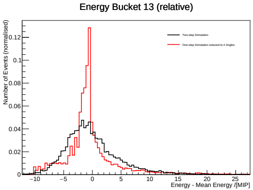
Problem Verification - Energy Differences per bucket Two-step Simulation in comparison to One-step Simulation reduced to 4 Angles

- Comparison of artifically reduced OSS is similar to TSS



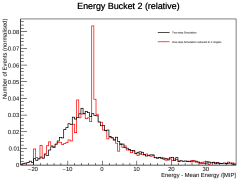


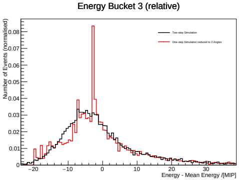


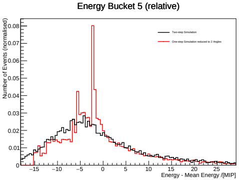


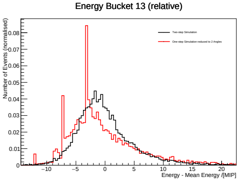
Problem Verification - Energy Differences per bucket Two-step Simulation in comparison to One-Step Simulation reduced to 2 Angles

- OSS artificially reduced to 2 Angles is even more similar
  $\Rightarrow$ A lost Correlation is not reversed easily









Pros and Cons

Pro

- Radial-Longitudinal Simulation with added Angular Part was implemented
- Faster

Con

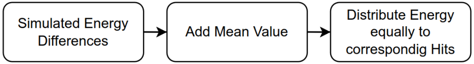
- Results in Energies below zero (!)
- It does not seem to fit very well (maybe we have to include the gap rejection for the event selection again)

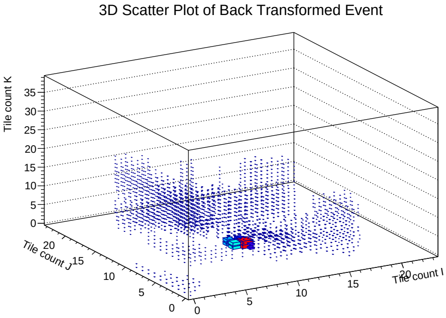
Idea:

Why not decrease executing time for full ARL by using a Lookup table?

First Back Transformation

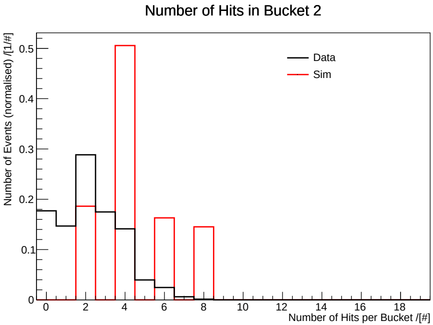
- One-step Simulation Data can be used to obtain slcio Files again (details to this process in the talk from André)
- slcio Files can be used to obtain Root Files again
- First approach equally distributes the energy in the corresponding hits

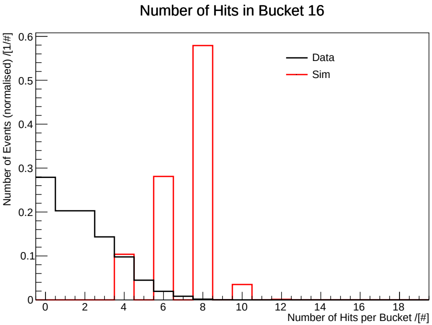




Problems with the Back Transformation

- Back Transformation does visually not show good agreement with Data
- Energy is equally split into the number of hits of every single Bucket
- Number of Hits does therefore not correlate with the input





Summary of Simulation

Pros:

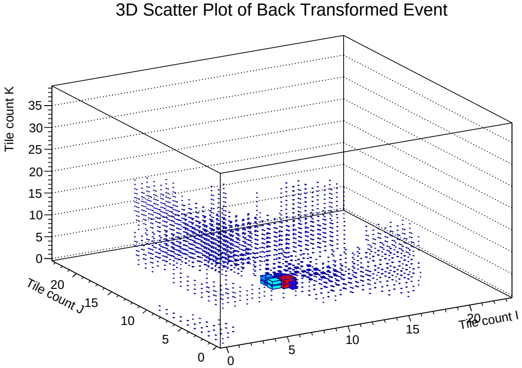
- One-step Simulation does fulfill the requirements to be used as an implementation
- Back transformation is working in principle
- One-step Simulation might has several additional improvements, so that a much faster simulation can be implemented

Cons:

- One-step Simulation is still not a Fast Calorimeter Simulation
- Two-step Simulation is neither possible nor easy to handle
- Back Transformation needs adjustments in order to split the energy correctly into the hits
  $\Rightarrow$ Whether this is possible, is under investigation

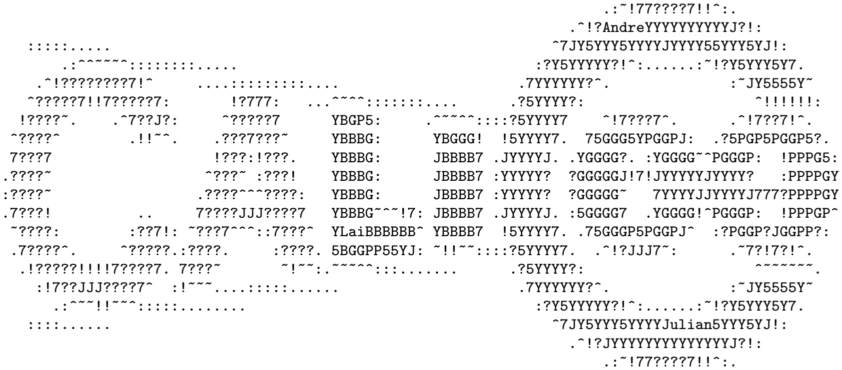
Outlook

- Improve the calculation time of the One-step Simulation further
- Implementing a working example for a Back Transformation
- Comparison to already established Monte-Carlo Simulation



Questions?

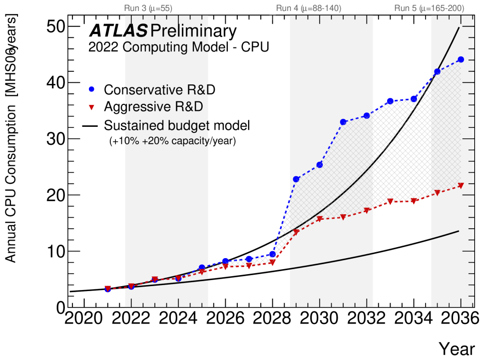
Questions?

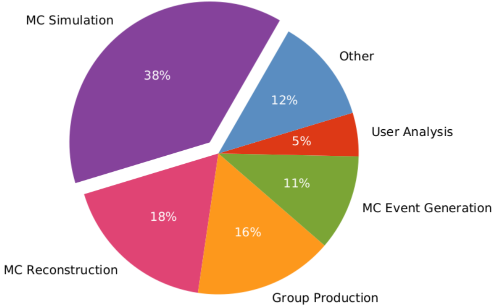


Backup Slides

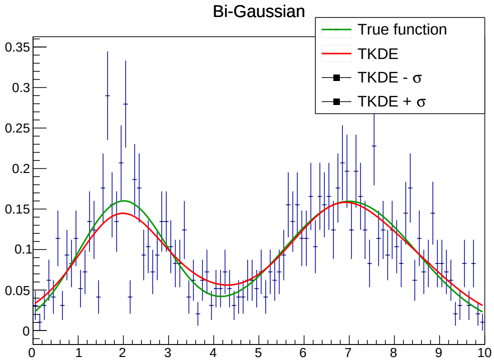
Motivaton for Fast Simulation

- CPU consumption of MC simulations increases with occupancy/granularity
- Up to 90 % of calculation time is needed for the calorimeter
- Saving of computational resources will become necessary sooner or later





Kernel Density Estimators



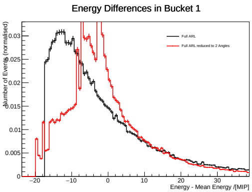
- Generalise to d dimensions:

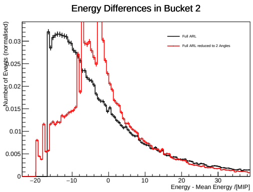
$$f(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^{n} |\mathbf{H}|^{-1/2} K\left(\mathbf{H}^{-1/2}(\mathbf{x} - \mathbf{x}_i)\right)$$

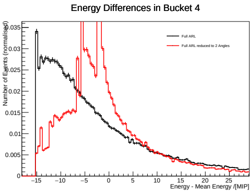
- $\mathbf{x}$: d-dimensional data vector; $\mathbf{H}$: $d \times d$ bandwidth matrix
- Have chosen $\mathbf{H} = h^2\mathbf{C}$ where $\mathbf{C}$ is the covariance matrix of the dataset

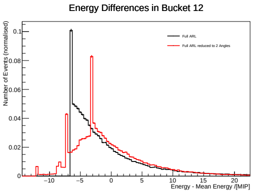
Problem Verification - Energy Differences per bucket Full ARL in comparison to Full ARL Reduced to 2 Angles

- Full ARL artificially reduced to 2 Angles (Mean value of 4 angles)



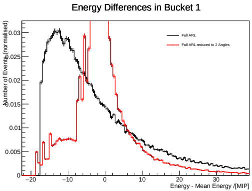


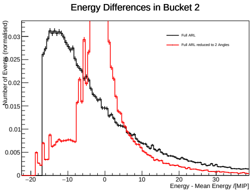


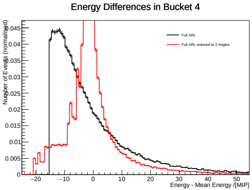


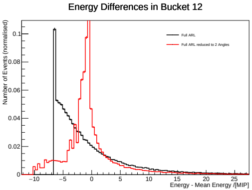
Problem Verification - Energy Differences per bucket Full ARL in comparison to Full ARL Reduced to 4 Angles

- Full ARL artificially reduced to 4 Angles (Mean value of 2 angles)









Problem Verification - Energy Differences per bucket 'RL + KDE Angular' in comparison to 'Full ARL with calculated 2 Angles'

- Full ARL calculated only with 2 Angles is also not in agreement