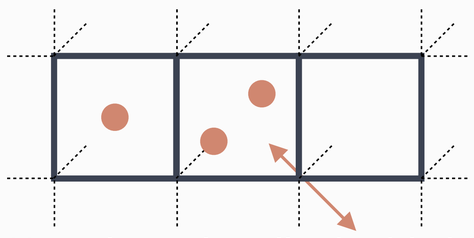
The Calorimeter Pyramid

Rethinking the design of generative calorimeter shower models

Simon Schnake[1,2], Dirk Krücker[1], Benno Käch[1,4], Moritz Scham[1,2,3], Kerstin Borras[1,2]

We aim to generate calorimeter showers as **point clouds**. As a result, we must remap the continuously generated points $(x, y, z, E)_i$ to the respective calorimeter cells.



How do we handle double hits?

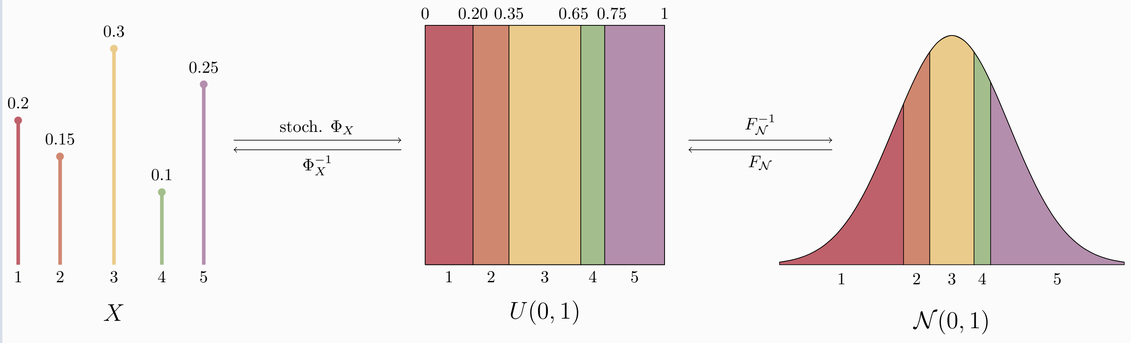
In our approach, we split the task into two parts:

- Generate the hit cells $(x, y, z)_i$ first
- Generate the energies E_i later

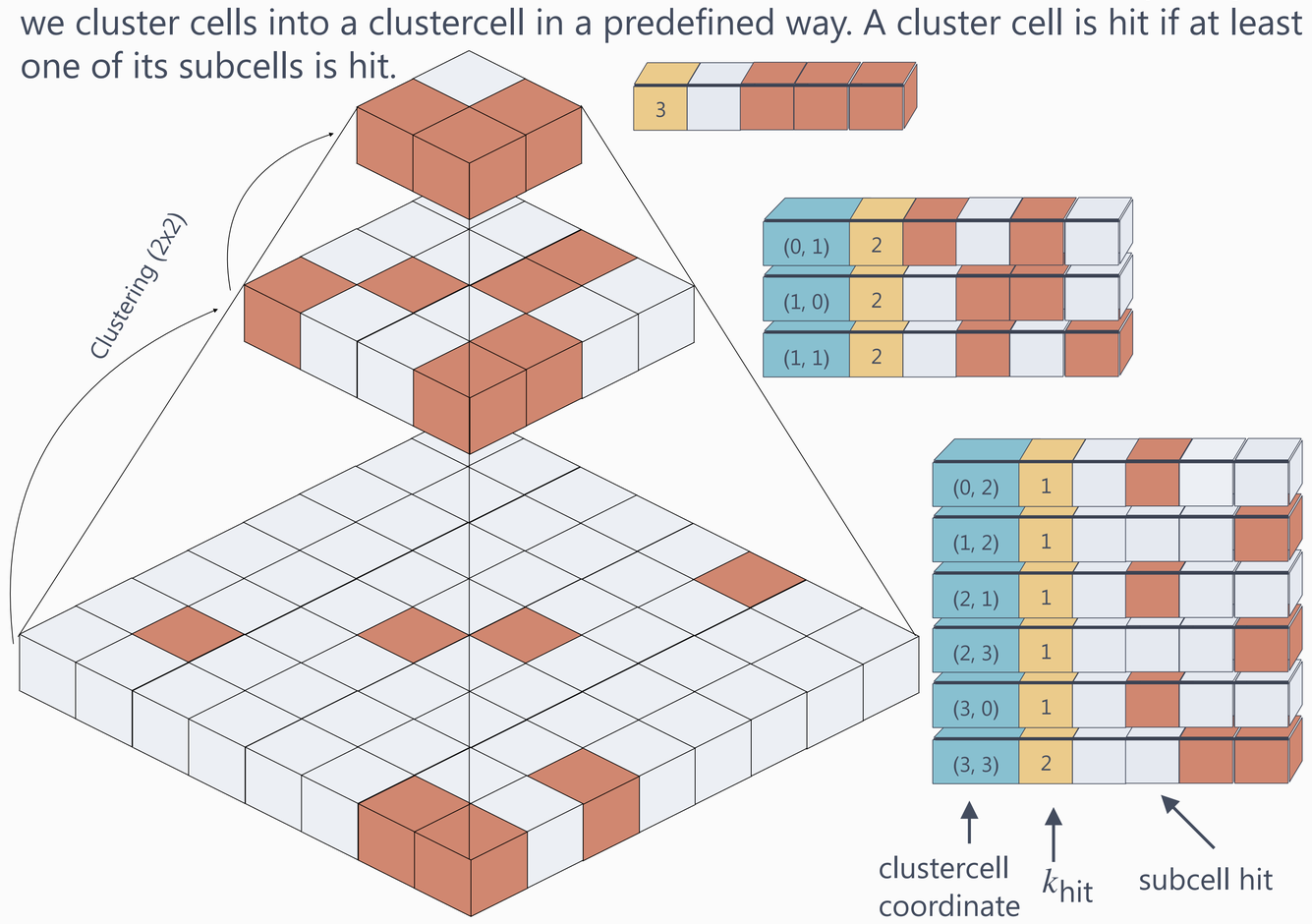
Here we present the first part: How to generate hit cells.

On each stage the models task is to model which subcells are hit of all hit cells. We have to model the subcells for a variable length cell list. We can employ a set based generative model here. For each cell we want to know for all subcells, whether they are hit or not. We want to employ the *Gumbel top-k trick*[Kool19], therefore we have to determine the number of hit subcells k_{hits} and than sample the hit subcells without replacement according to their hit probabilities.

For the generation of k_{hit}s we dequantize the discrete values into continuous ones using the **CDFDequantization**:



For each cell, we want to know whether it is a hit or not: To utilize the sparsity, we cluster cells into a clustercell in a predefined way. A cluster cell is hit if at least one of its subcells is hit.

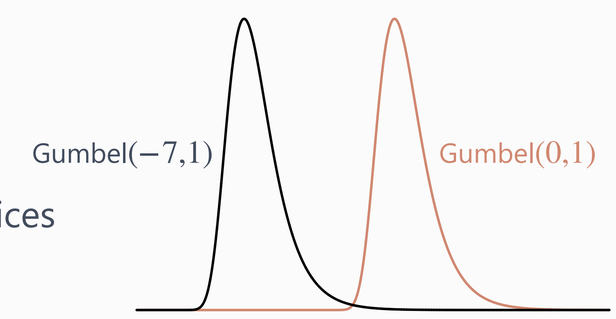


In the generation process, we have one model for each stage. Each stage generates the hit subcells for all hit cluster cells conditioned on the stage above.

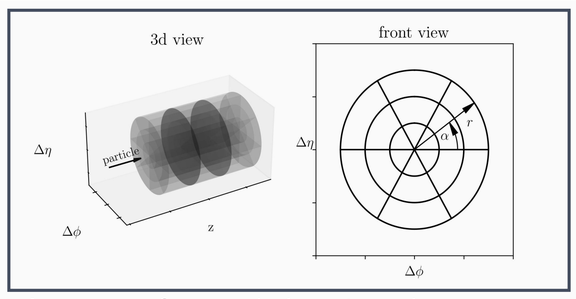
For each subcell, a probability of being hit is $p = 1$ for hit subcells and 0 for non-hit subcells. By allowing a small probability ($\approx 10^{-3}$) for non-hit cells to be hit, the hits can be dequantized to X.

$$X = \begin{cases} x \sim \text{Gumbel}(0,1), & \text{if } p = 1 \\ x \sim \text{Gumbel}(-7,1), & \text{if } p = 0 \end{cases}$$

The quantization is done by taking the indices of the k_{hit} largest of X



CaloChallenge Dataset 2:
DOI 10.5281/
calochallenge.github.io/homepage/



- 100k GEANT4-simulated electrons showers for training/testing
- Energies with log-uniform distribution [1 GeV,1 TeV]
- Concentric cylinder detector geometry
- 45 layers (z) x 9 radial segments (r) x 16 angular segments (α) =6480 voxels

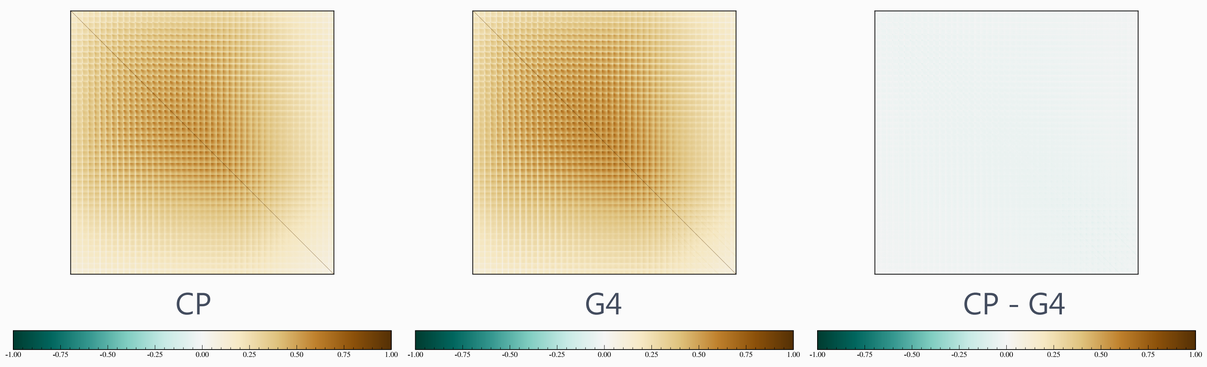
Clustering Prescription:

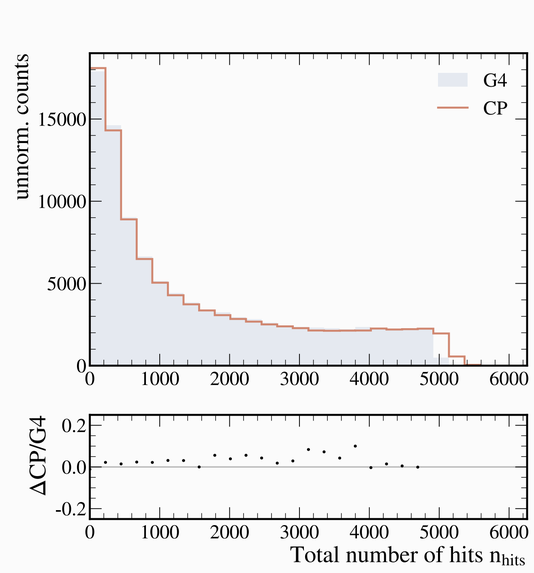
- 3 Stages
- Cluster all cells in α
- Cluster all cells in r

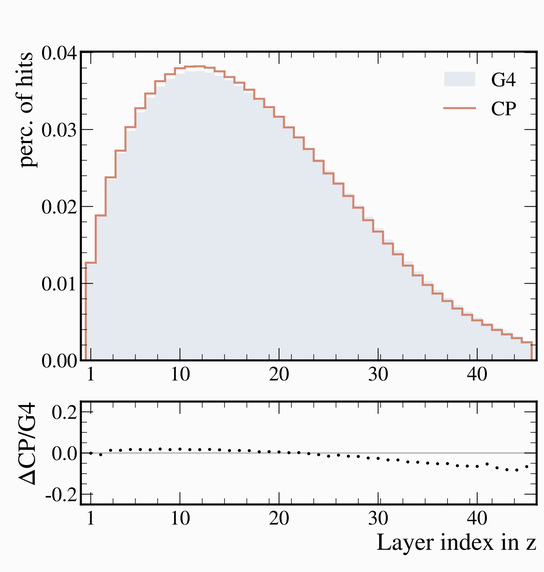
Models:

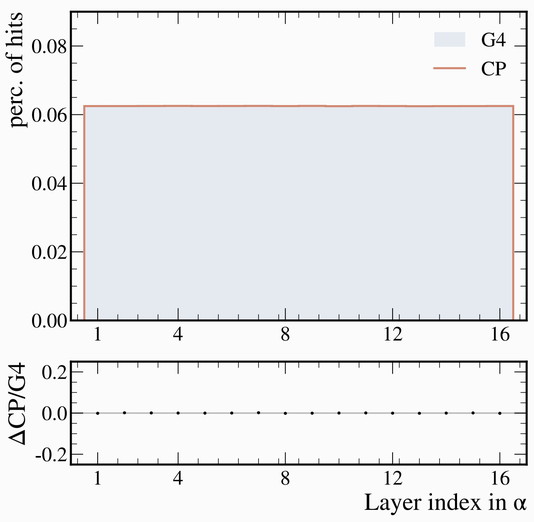
- Stage 1: RQS Coupling Flow
- Stage 2: RQS Perm. Inv. Coupling Flow
- Stage 3: RQS Perm. Inv. Coupling Flow

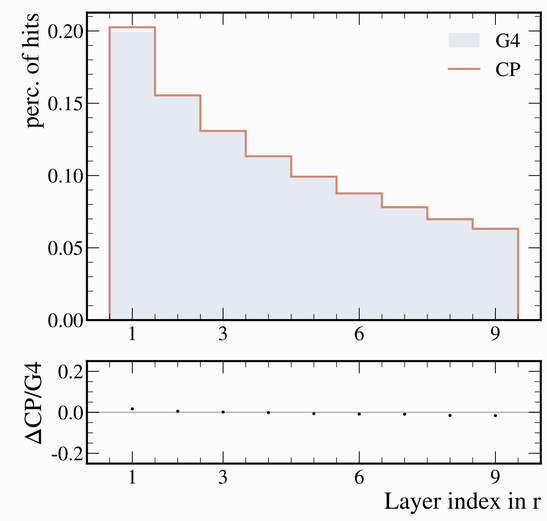
Pearson correlation coefficients between the hits of all cells











(1) DESY, Hamburg (2) RWTH University of Aachen (3) FZJ and Jülich Supercomputer Center (4) Hamburg University