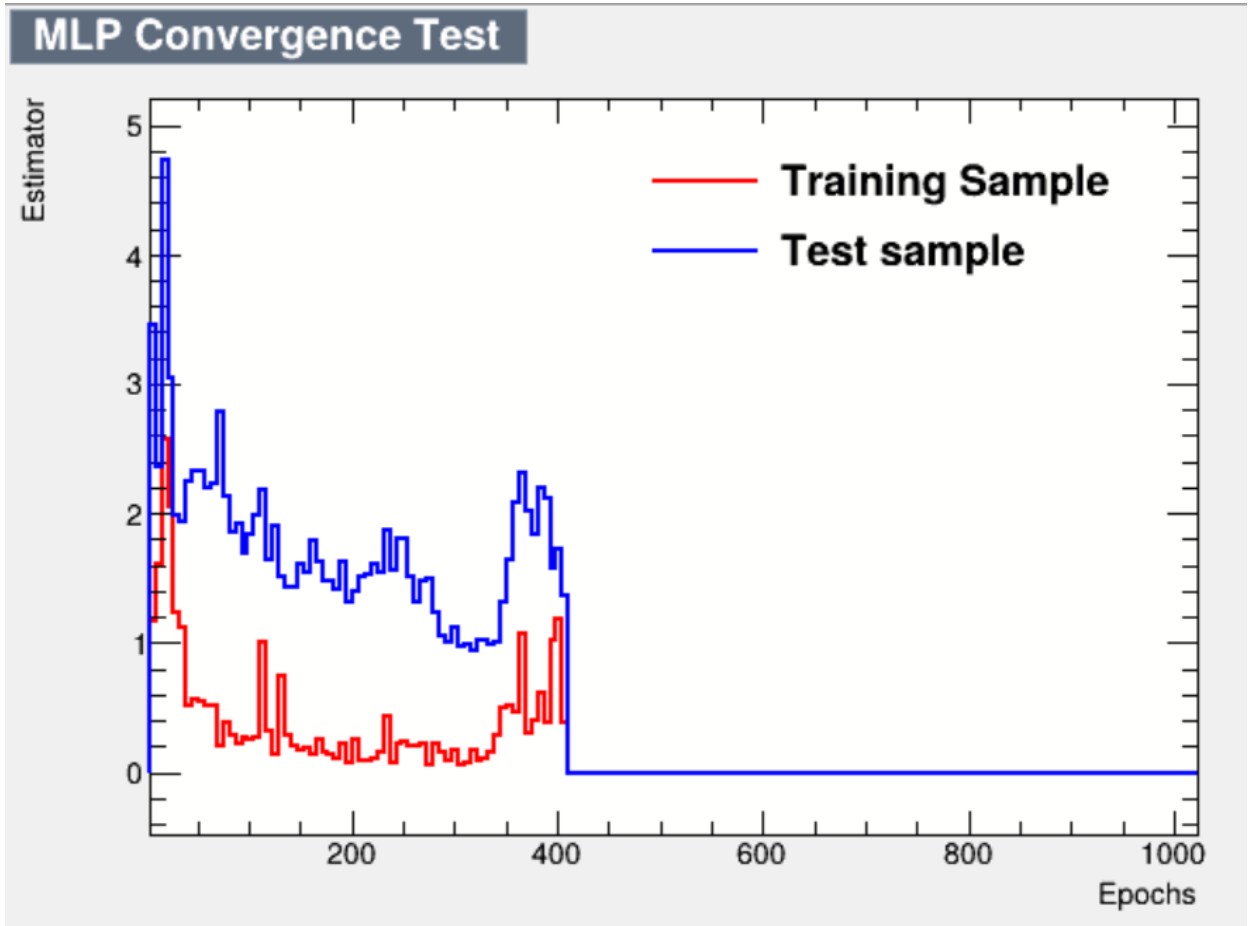
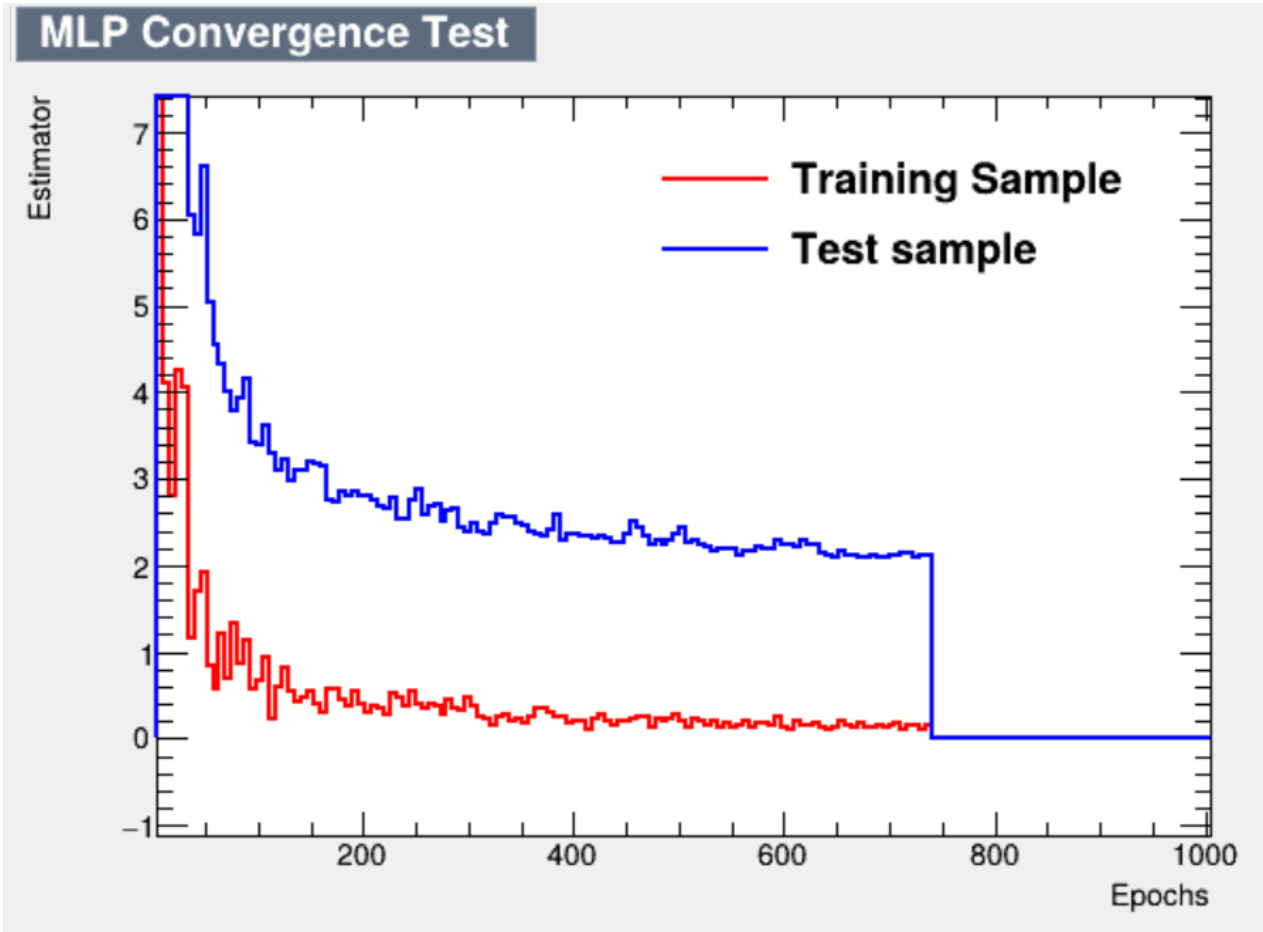
Refresher:
Lower loss/Estimator = better performance

Before Optimization

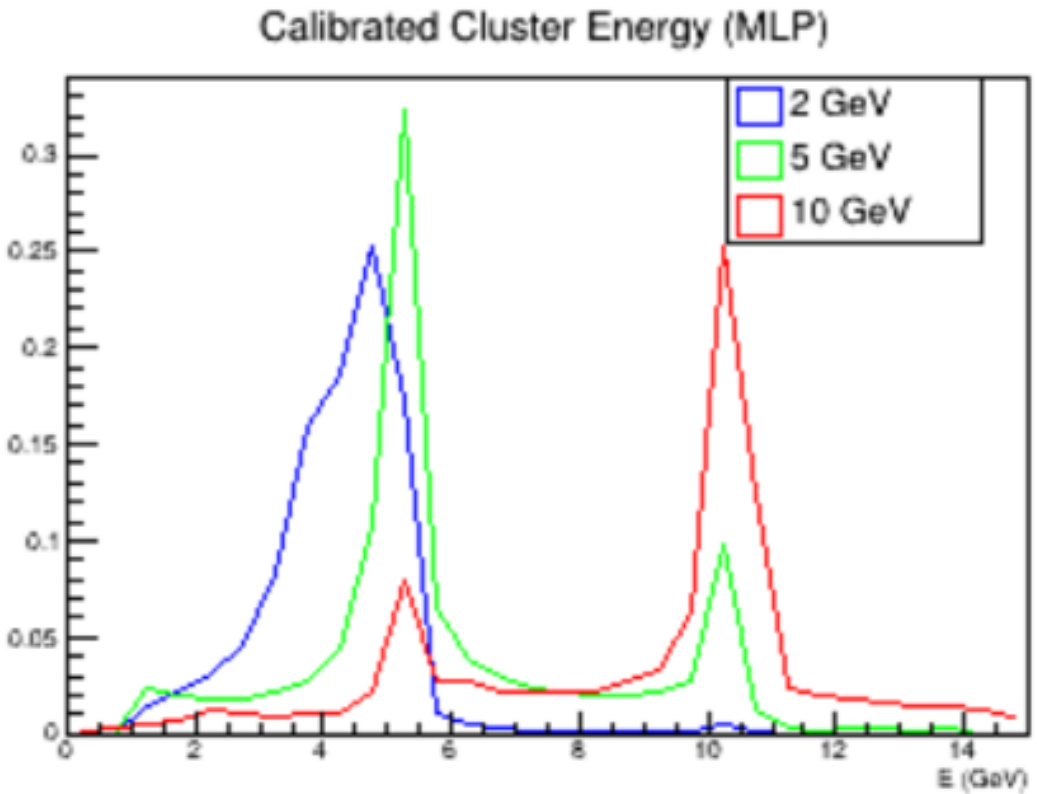


Initial Optimization

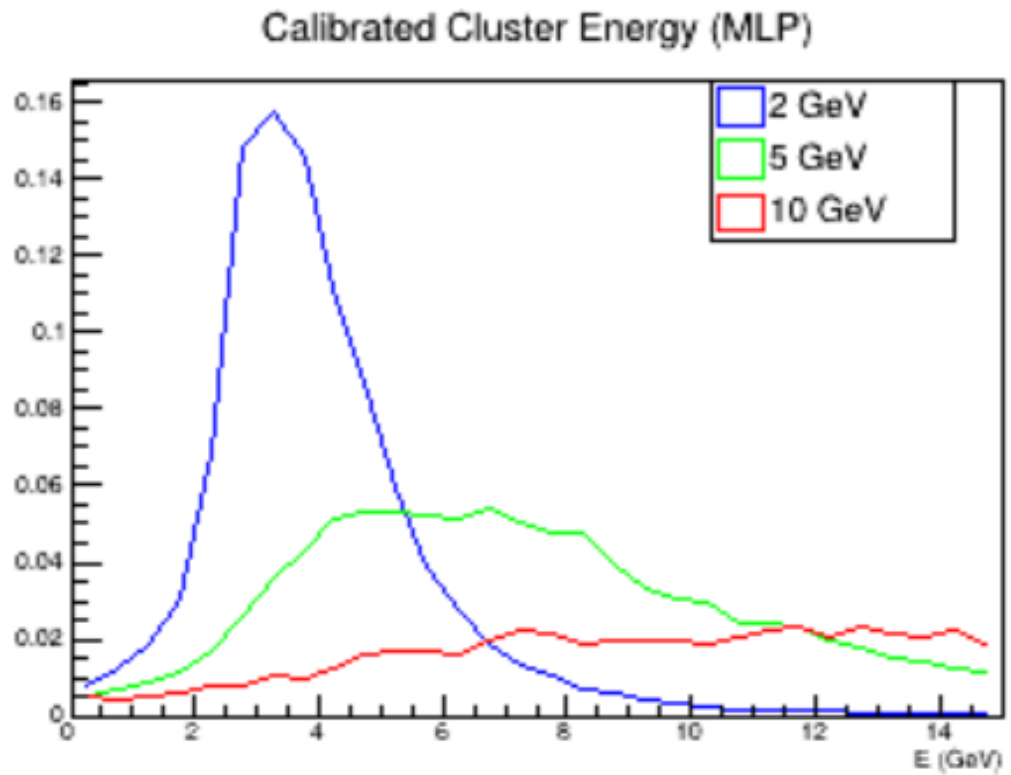


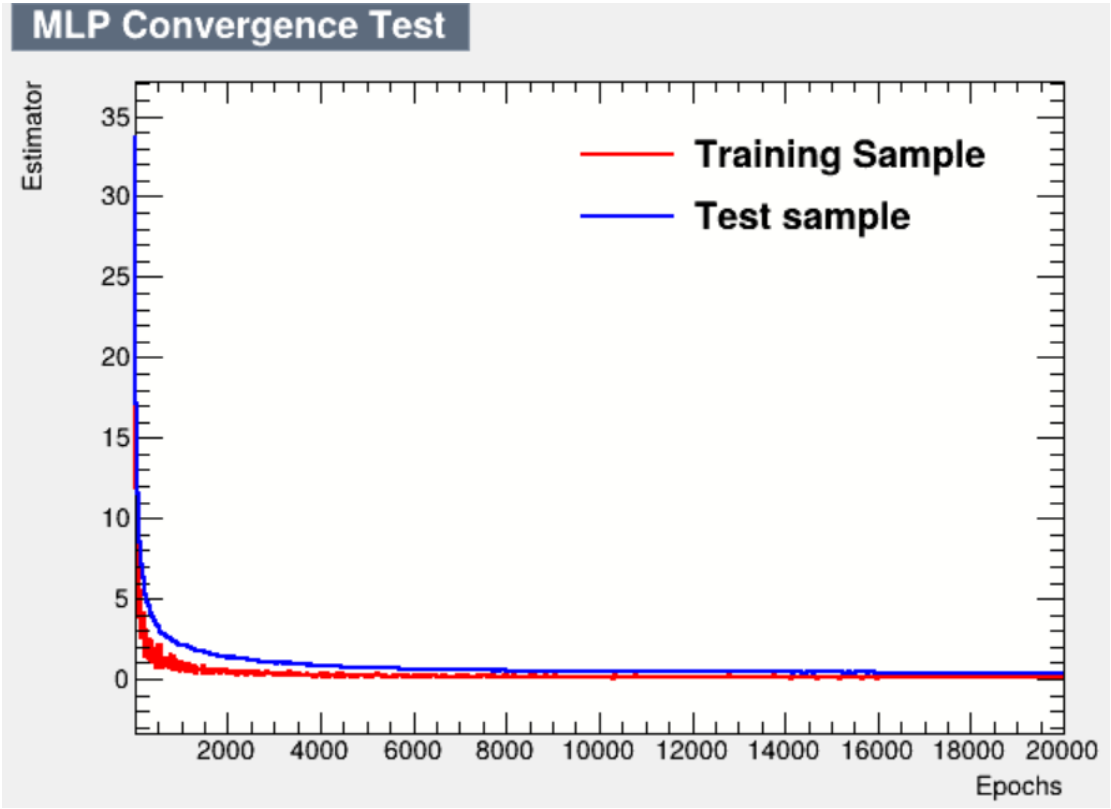
Old Calibrated Cluster Energies

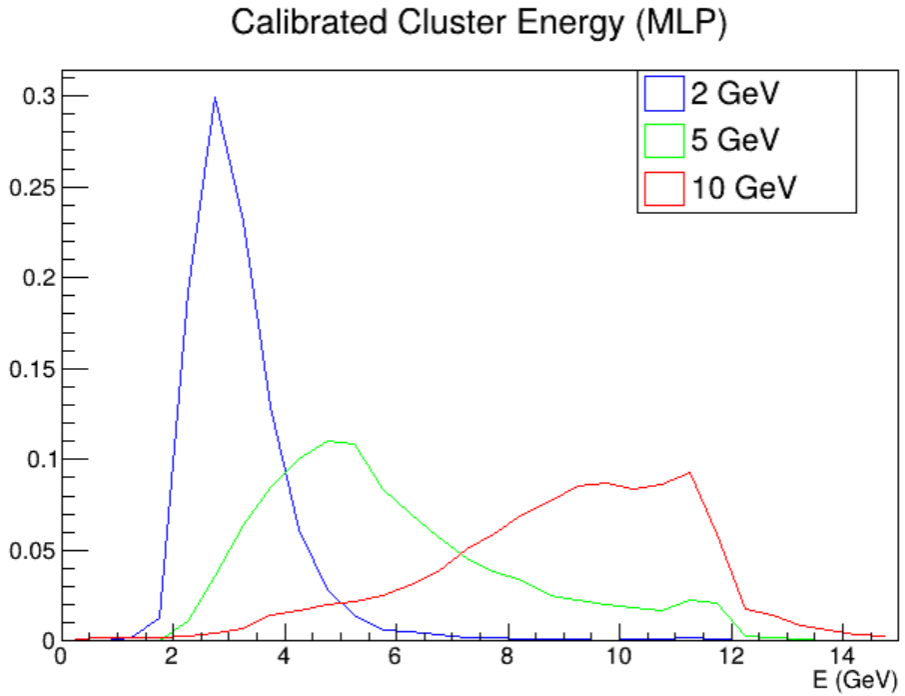
Before Optimization

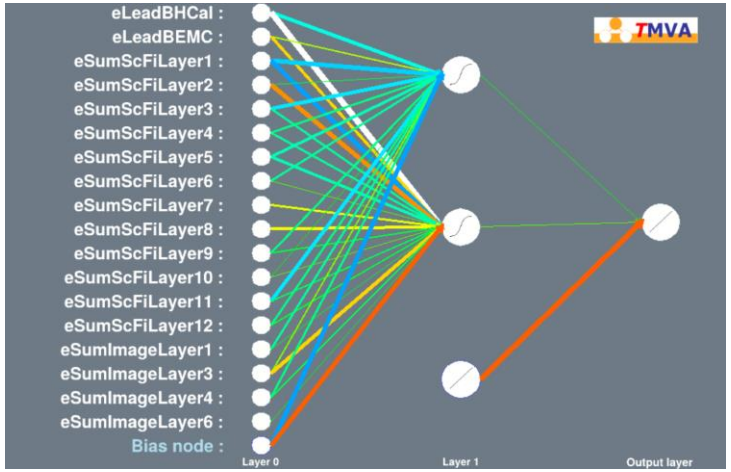


Initial Optimization







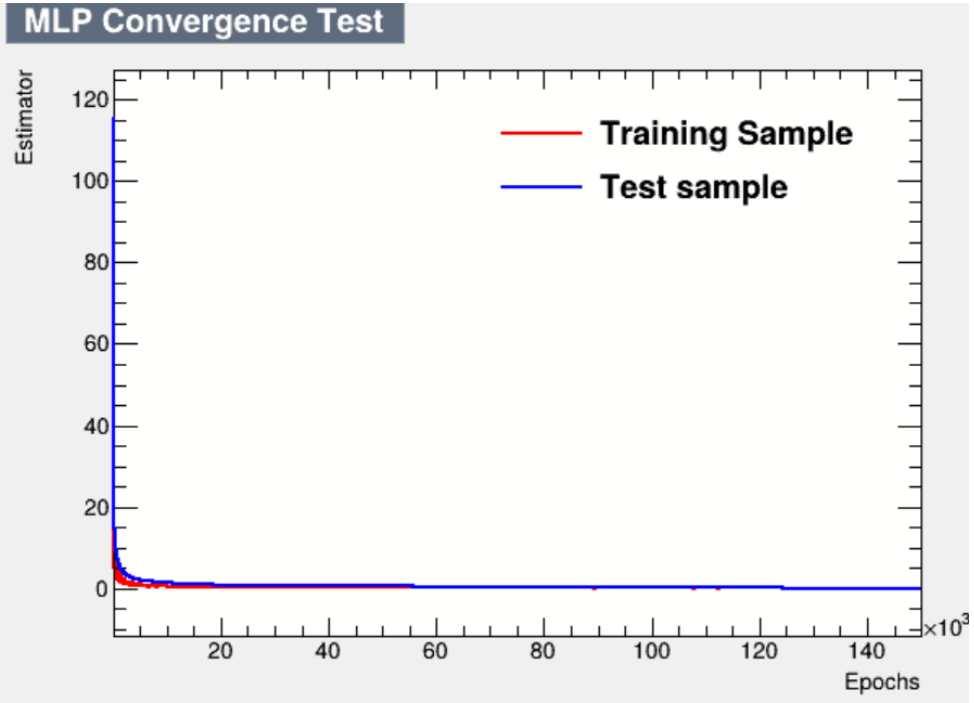


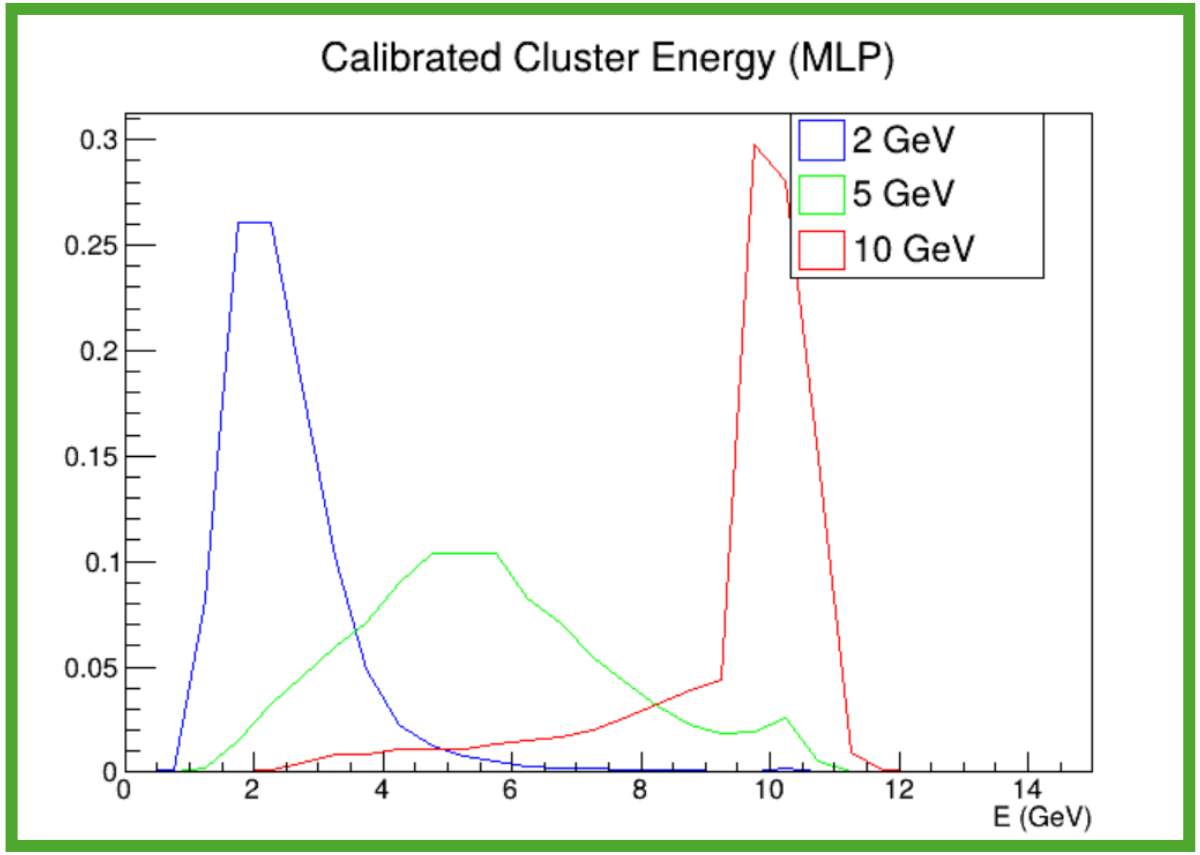
$$\eta_t = \frac{\eta_0}{1+\alpha \cdot t}$$

Carefully found a resonant η_0 (Learning Rate) and α (Decay Rate)
Essentially through (informed) trial and error

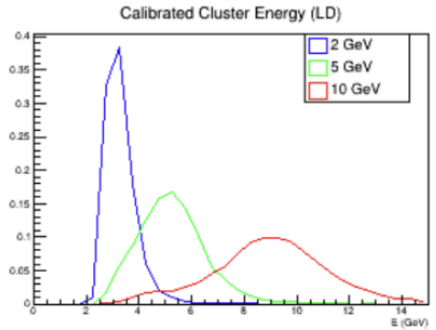
Need to find the minimum in the “loss landscape” – if initial LR is too big you pass it, if you decay too fast or start too slow you’ll run out of ‘gas’ before getting there, ideally you LR to decay as you train to get closer and closer to the minimum

We’re on a good trajectory, so let’s just run for more epochs:

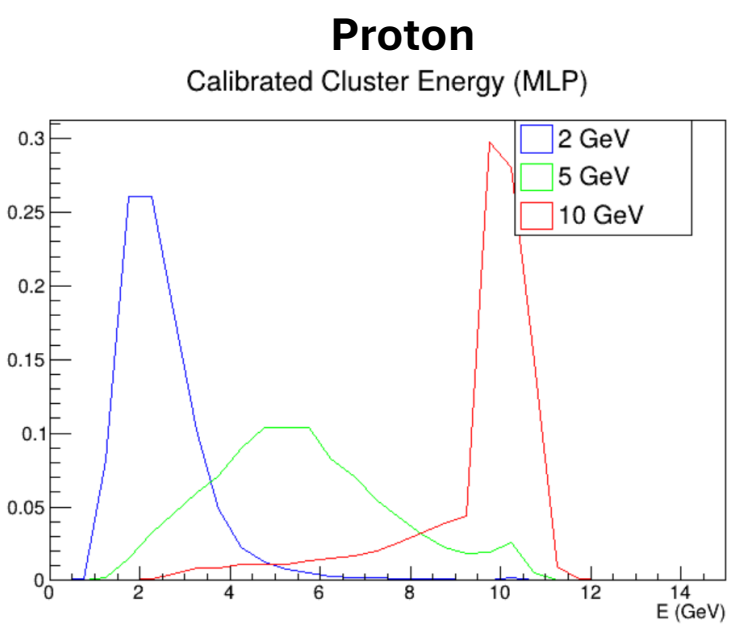




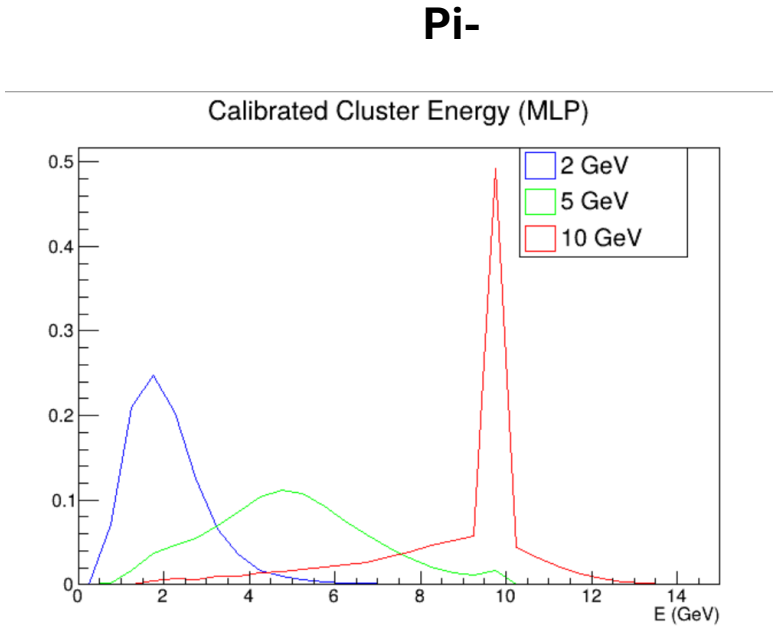
Finally starting to perform well – arguably better than LD:



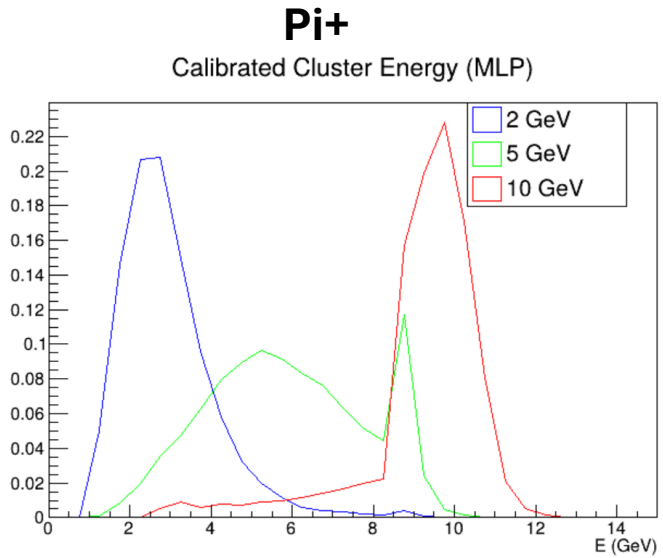
Proton



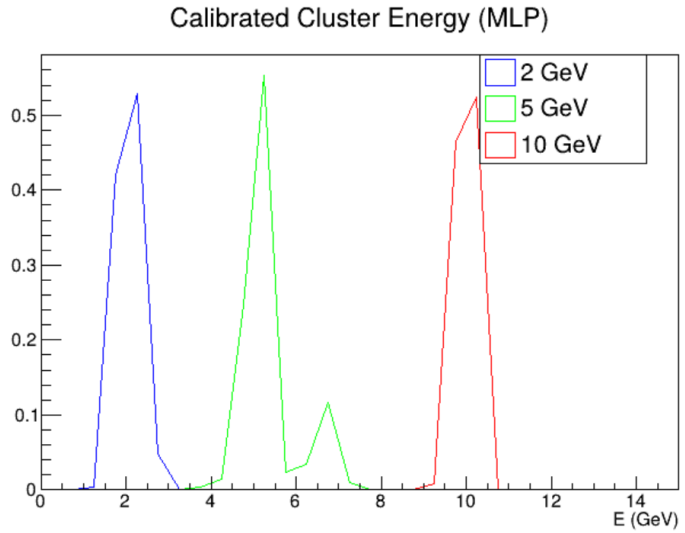
Pi-

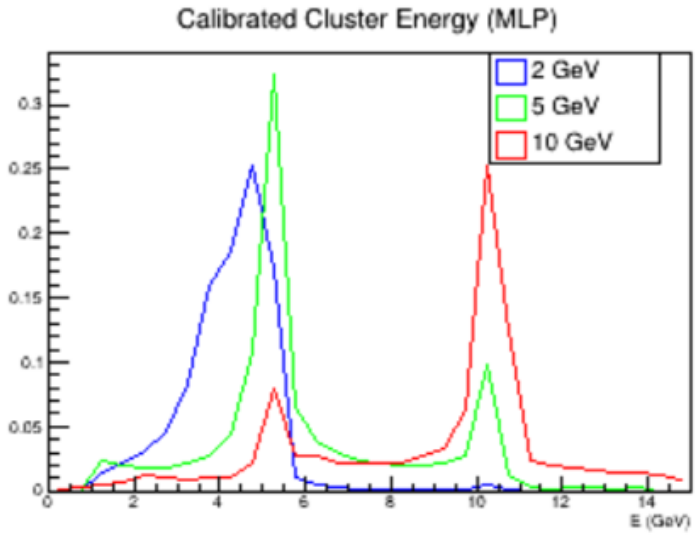


Pi+

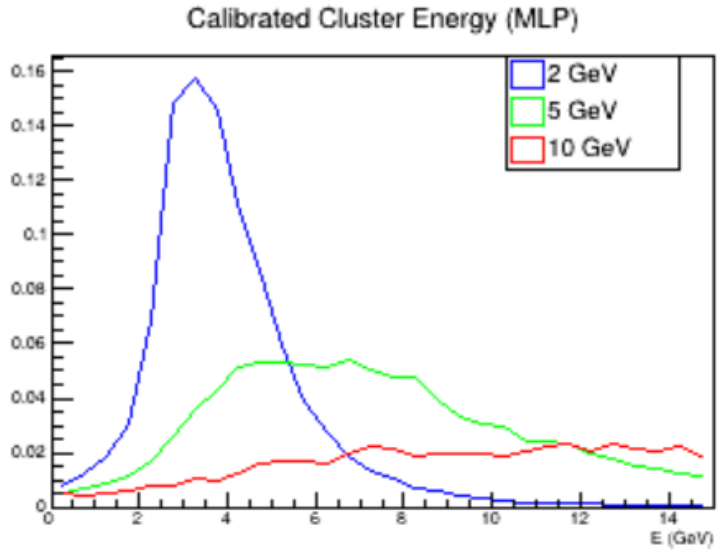


Pi0

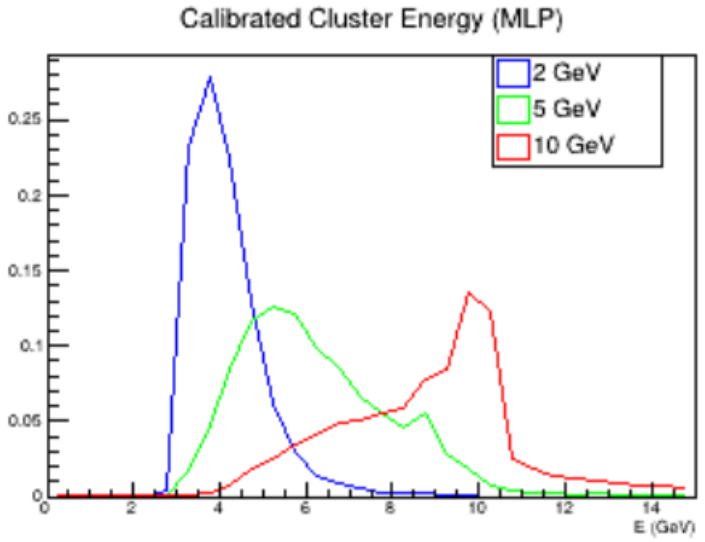




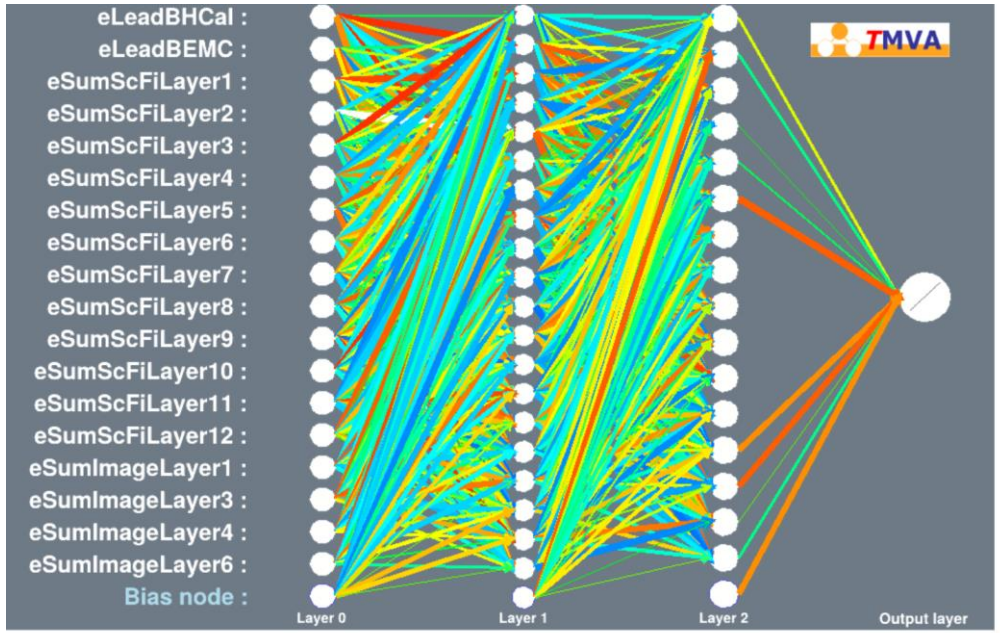
Initial

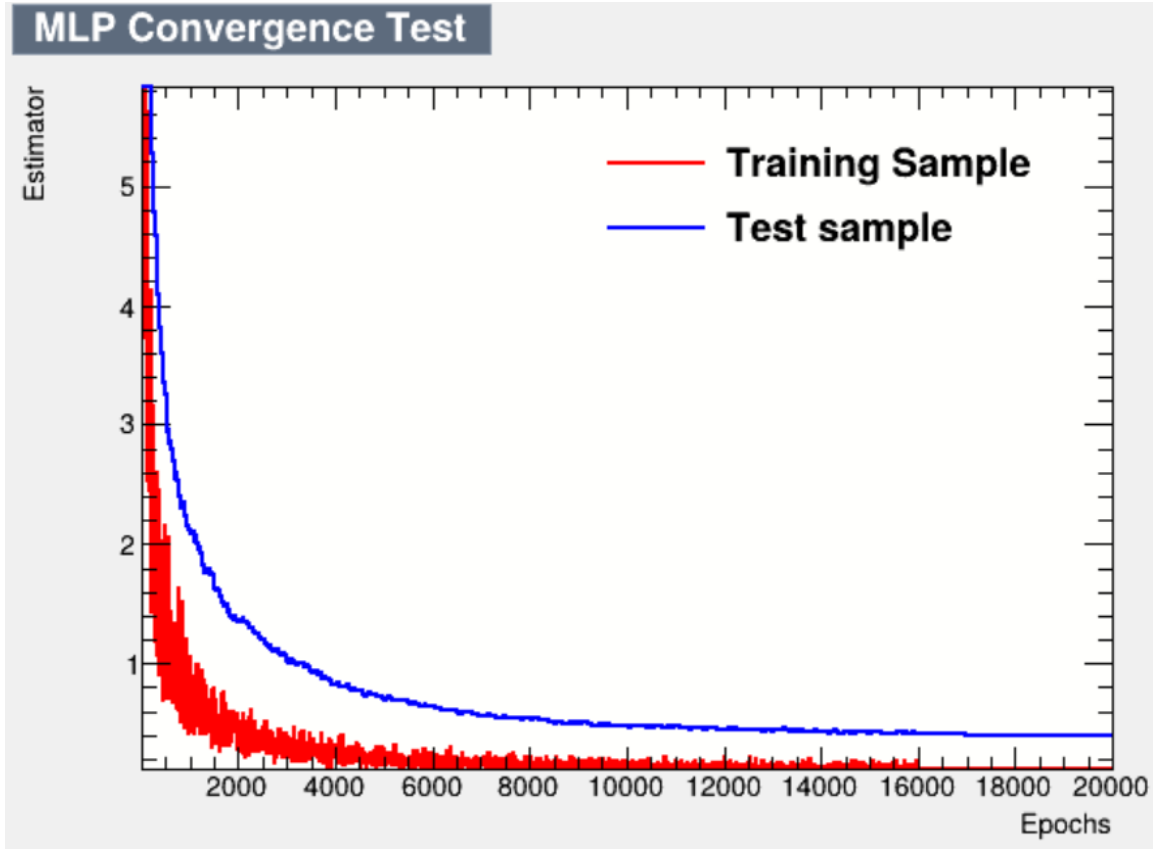


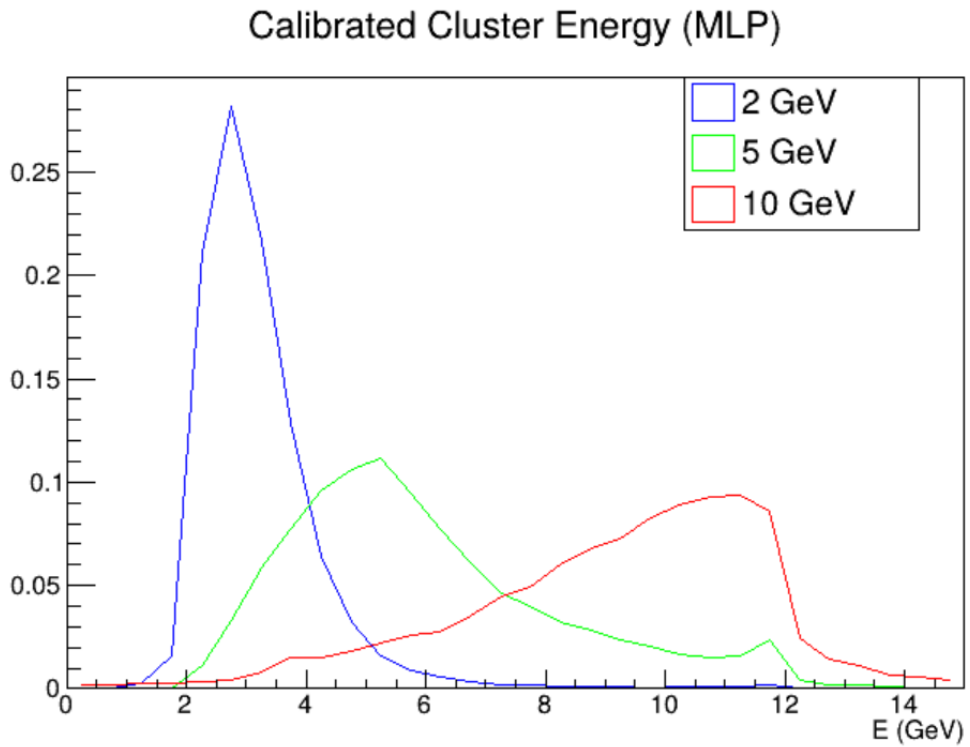
Tuning Round 1



Tuning Round 2




MLP Convergence Test
Estimator
Training Sample
Test sample
1
2
3
4
5
2000
4000
6000
8000
10000
12000
14000
16000
18000
20000
Epochs


Calibrated Cluster Energy (MLP)
2 GeV
5 GeV
10 GeV
0
0.05
0.1
0.15
0.2
0.25
0
2
4
6
8
10
12
14
E (GeV)