Machine Learning Flavour Tagging for Future Higgs Factories

Mareike Meyer

Material to be shown at ECFA workshop in Paestum

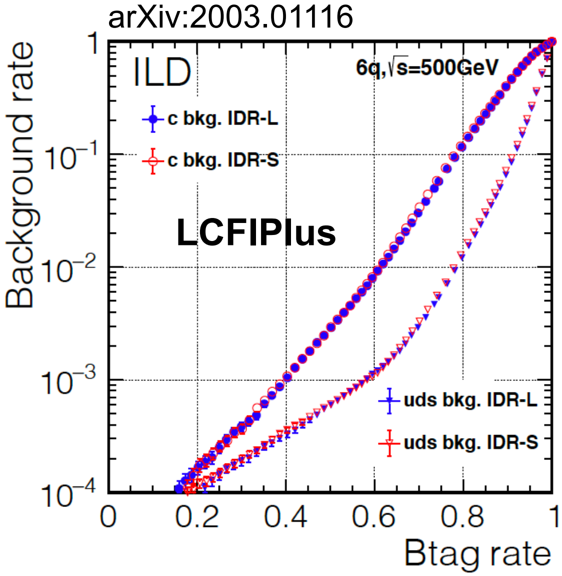
ILD Software and Analysis meeting, 05/10/2023

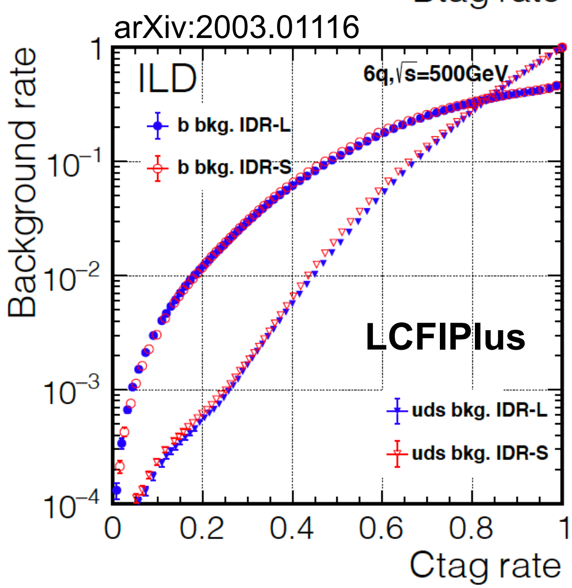
HELMHOLTZ RESEARCH FOR GRAND CHALLENGES

Introduction

- current standard: **LCFIPlus**
  - arXiv:1506.08371, https://github.com/lcfiplus/LCFIPlus
- based on TMVA (BDTs)

➡ Can the **heavy flavour tagging** be **improved** by replacing the BDTs used in LCFIPlus **with (deep) NNs**?





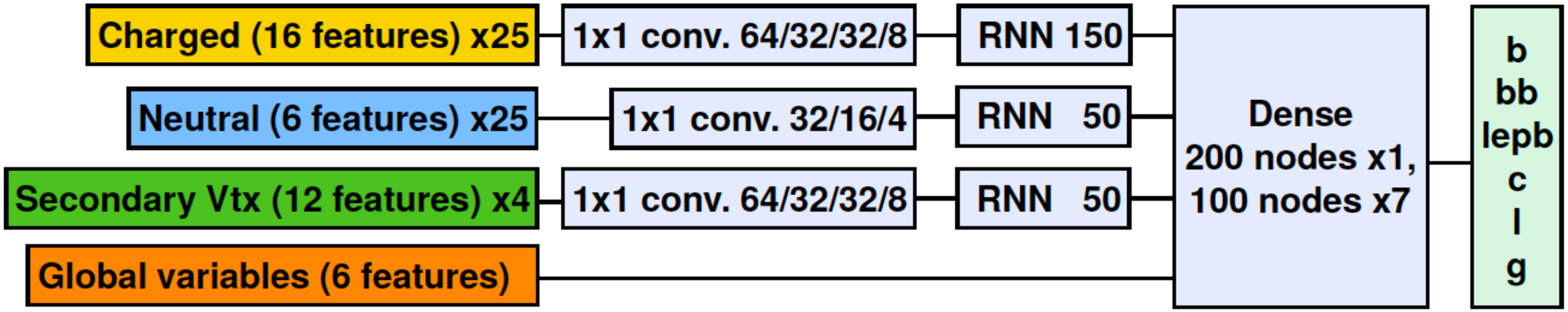
this work:

Apply the following taggers to ILD:

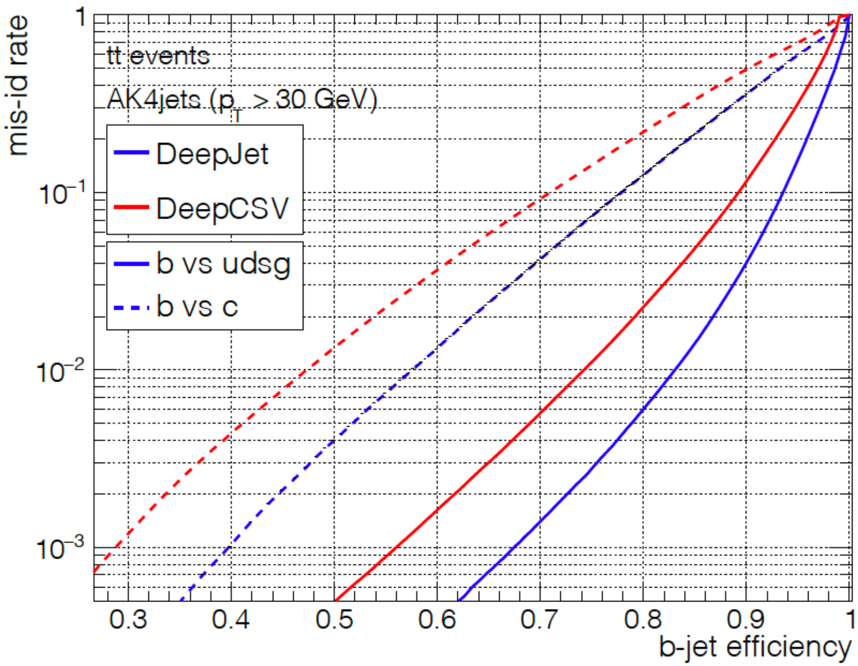
- **CMS DeepJet**
  - „Jet Flavour Classification Using DeepJet“, arXiv:2008.10519
  - „Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV“, arXiv:1712.07158
- **ParticleNet**
  - „Jet Tagging via Particle Clouds“, arXiv:1902.08570
  - „Pushing the Limit of Jet Tagging With Graph Neural Networks“, Huilin Qu, talk at ML4Jets2021, July 7, 2021

CMS DeepJet

arXiv:2008.10519, arXiv:1712.07158



- successfully applied in many recent CMS analyses
- allows for usage of low-level features from many jet constituents
- able to deal with variable length sequence of inputs
- allows for ordering of particles according to their assumed importance
- large gain in performance compared e.g. to FCNN (DeepCSV)



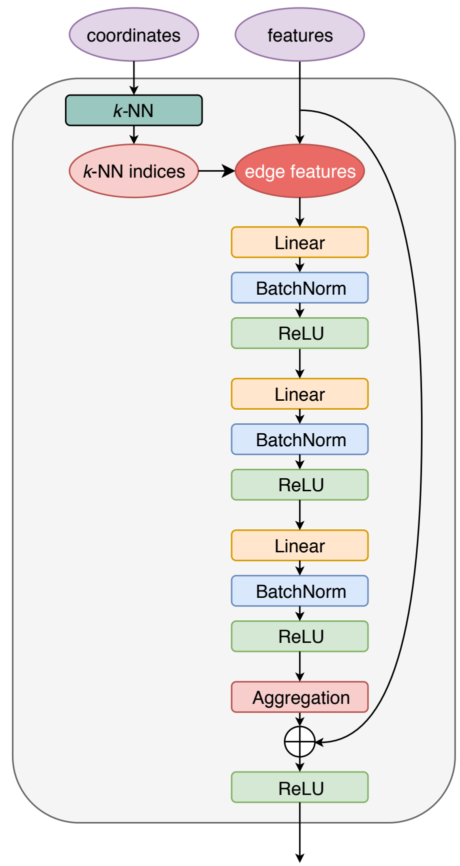
Particle Net

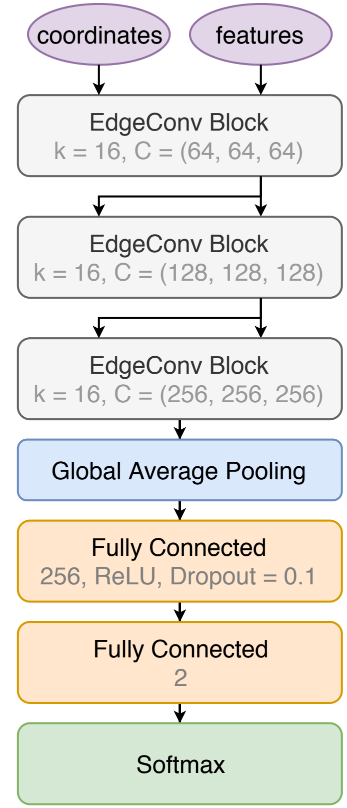
arXiv:1902.08570, „Pushing the Limit of Jet Tagging With Graph Neural Networks“, Huilin Qu, talk at ML4Jets2021, July 7, 2021

- based on **Dynamic Graph CNN** (Y. Wang et al., arXiv:1801.07829)
- treat jet as „particle cloud“, input: **jet constituents**
- key building block: **EdgeConv**
  - treat particle cloud as a graph, each point is a vertex, edges are constructed as connections between each points and k nearest neighboring points
  - learn an „edge feature“ for each pair:

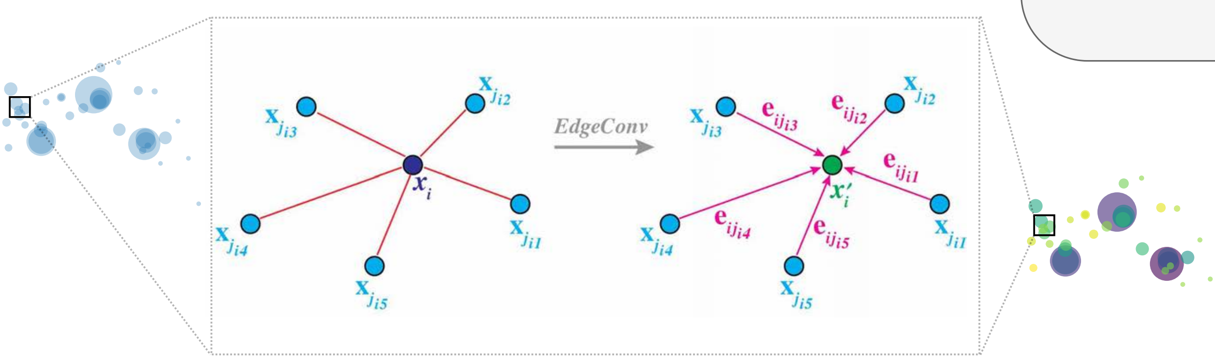
$$e_{ij} = \mathrm{MLP}(x_i, x_j)$$

  - MLP: parameters shared among all edges
  - aggregation of edge features: $x_i' = \mathrm{mean}_j\, e_{ij}$





(a) ParticleNet



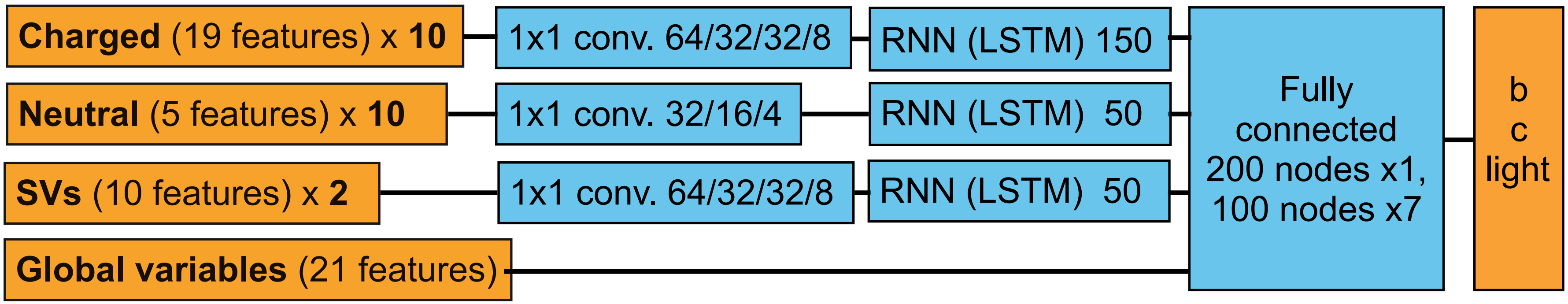
Training data

- train on **events with 6 jets** (b,c,u,d,s)
  - /pnfs/desy.de/ilc/prod/ilc/mc-opt-3/ild/dst-merged/500-TDR_ws/flavortag/ILD_l5_o1_v02/v02-00-01/
- run PV & SV finder, jet clustering and vertex refinement of LCFIPlus

- split sample into training, validation and test (75% / 12.5% / 12.5%)
- number of jets in **training data**:
  - b jets: 434116
  - c jets: 484034
  - light jets: 1449546
  - ➡ over-sampling of b and c jets performed to get same number of b,c & light jets
  - ➡ **total number of jets in training data**: 3 * 1449546 = 4348638
- number of jets in **validation data**:
  - b jets: 72443
  - c jets: 80890
  - light jets: 241283

DeepJet

DeepJet: architecture & data pre-processing

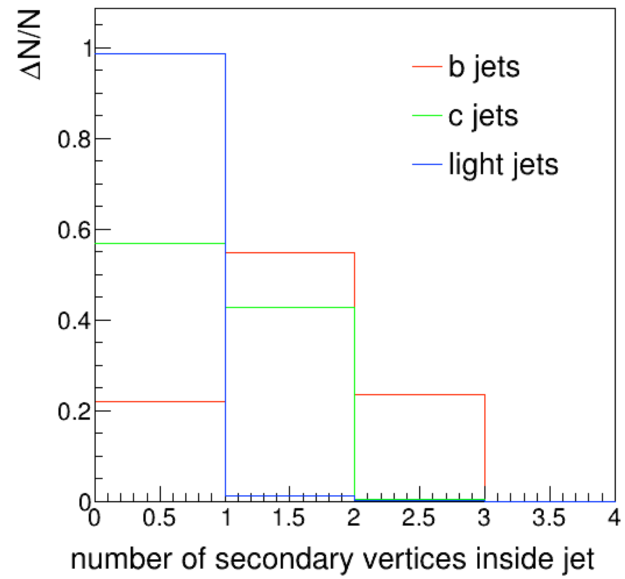


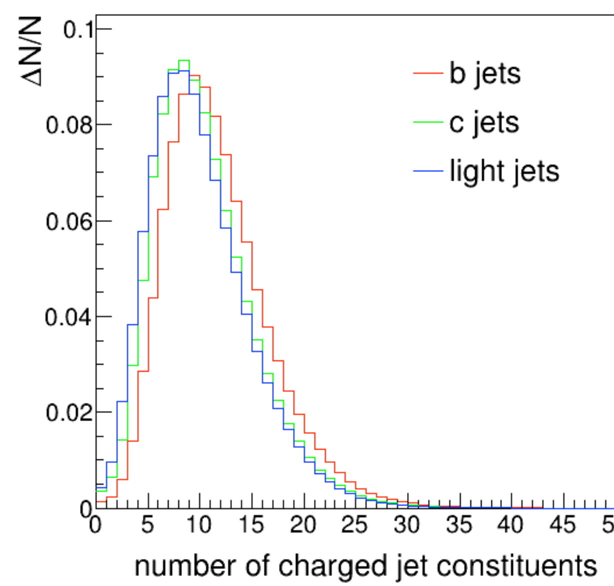
- classify jets into **three classes**: b jets, c jets & light jets
- **ordering of input particles** by (as applied in CMS)
  - impact parameter significance for charged jet constituents
  - shortest angular distance to a secondary vertex (by momentum if there is no secondary vertex) for neutral jet constituents
  - flight distance significance for secondary vertices
- if a value of a features is not available, the value is set to -10
- **normalize input features** to mean 0, std 1

DeepJet: input features - global variables

21 input features

- jet momentum
- jet transverse momentum
- number of charged jet constituents
- number of neutral jet constituents
- number of secondary vertices
- additional variables from LCFIPlus:
  - mass of all tracks with d0/z0 significance > 5σ
  - product of b/c/light-quark probabilities of d0/z0 values of all tracks, using b/c/light-quark d0/z0 distributions
  - joint probability in the r-φ plane / in the z projection using all tracks (with IP significance > 5σ)
  - vertex probability taking into account all tracks associated to vertex
  - distance and its significance between the first and second vertex in the jet
  - mass of the vertex (pT - corrected)
  - vertex probability of all vertices

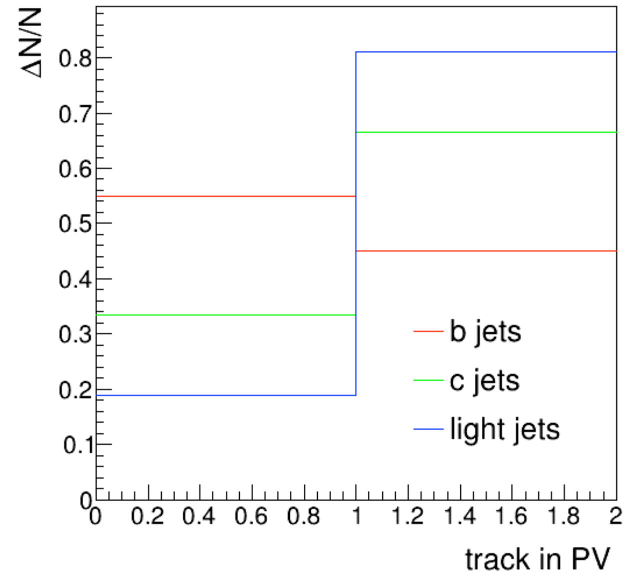


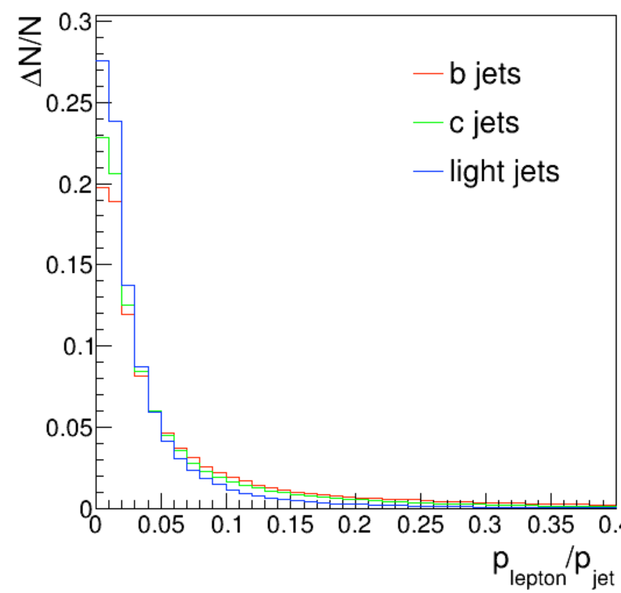


DeepJet: input features - charged jet constituents

- track momentum / jet momentum
- transverse track momentum relative to jet
- dot product of jet and track momentum w.r.t. jet momentum
- ΔR(track, jet),
- d0, d0 significance
- Z0, Z0 significance
- 3D impact parameter, 3D impact parameter significance
- track reconstructed in PV?
- is electron?, is muon?, lepton momentum relative to jet, lepton transverse momentum relative to the jet, lepton momentum / jet momentum
- kaon-ness of charged particles, track momentum fraction weighted with kaon-ness
- χ2/ndf

19 input features

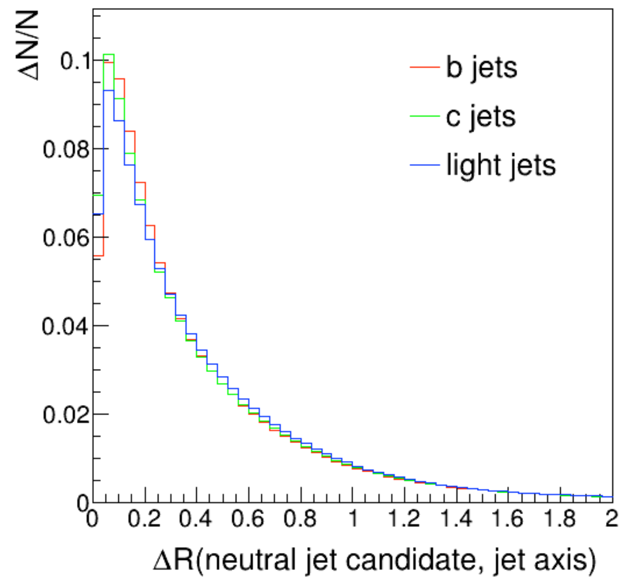


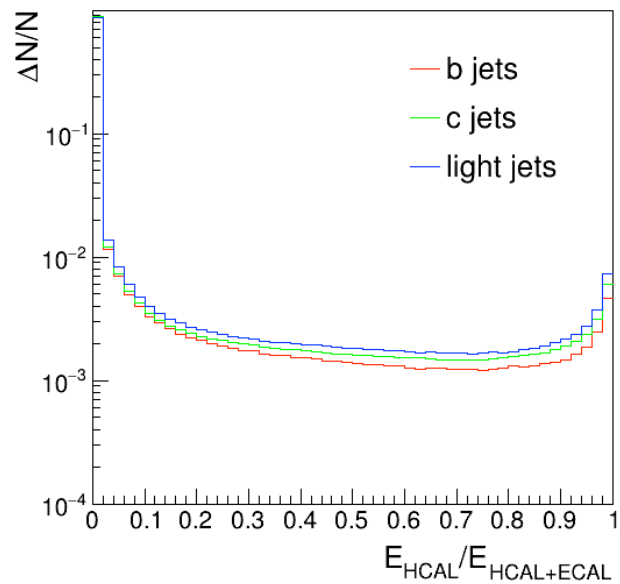


DeepJet: input features - neutral jet constituents

- momentum of neutral jet constituent
- fraction of the jet momentum carried by neutral jet constituent
- ΔR(jet axis, neutral candidate),
- is photon?
- fraction of neutral candidate energy deposited in the hadronic calorimeter

5 input features

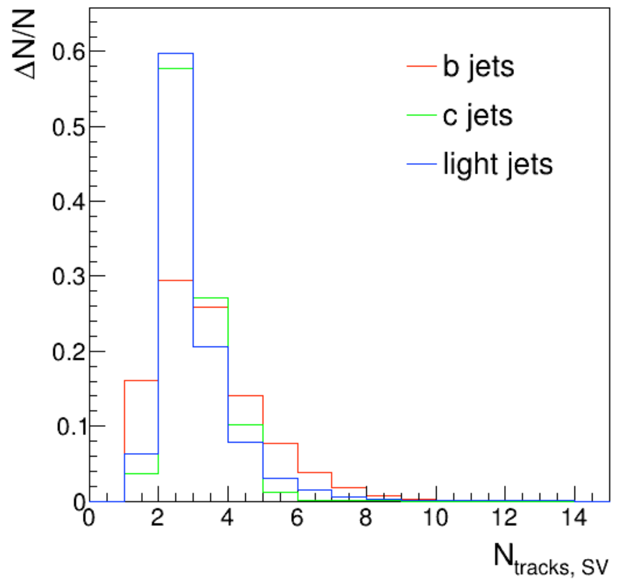


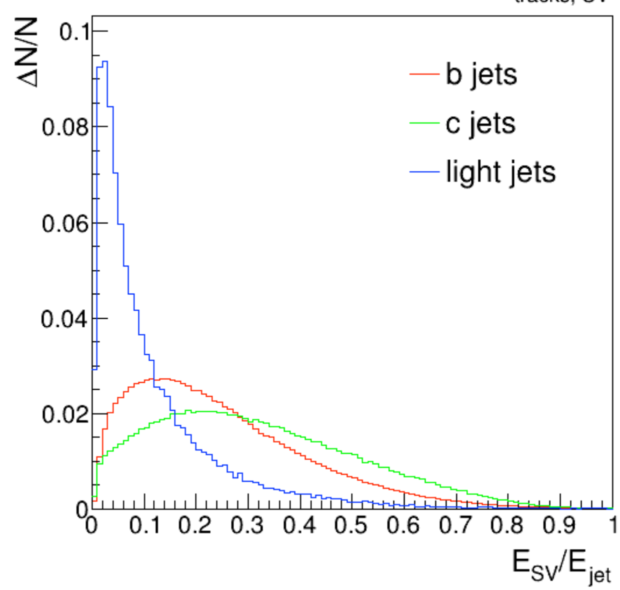


DeepJet: input features - secondary vertices

- SV mass
- number of tracks in SV
- ΔR(SV, jet)
- SV energy / jet energy
- SV energy
- cosine of the angle between the secondary vertex flight direction and the direction of the secodary vertex momentum
- 3D impact parameter, 3D impact parameter significance
- χ2, ndf

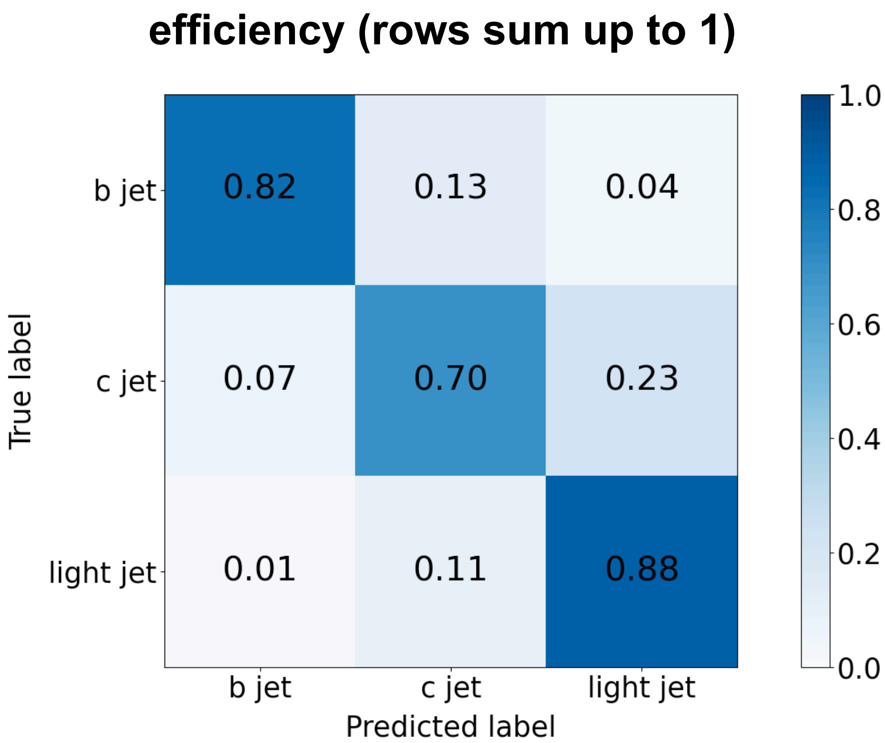
10 input features

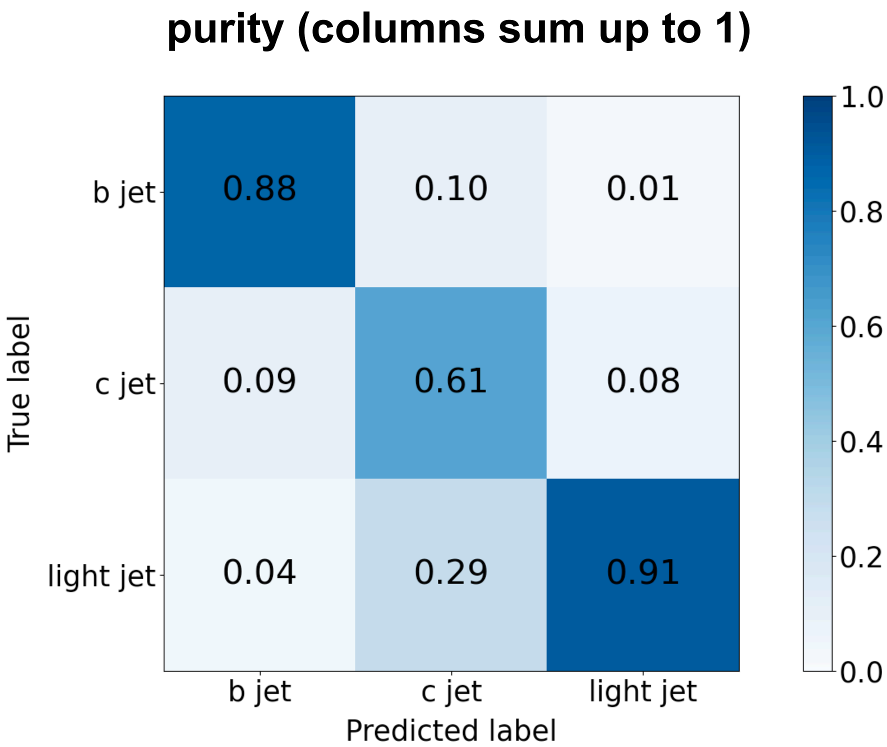




DeepJet: confusion matrices

validation data

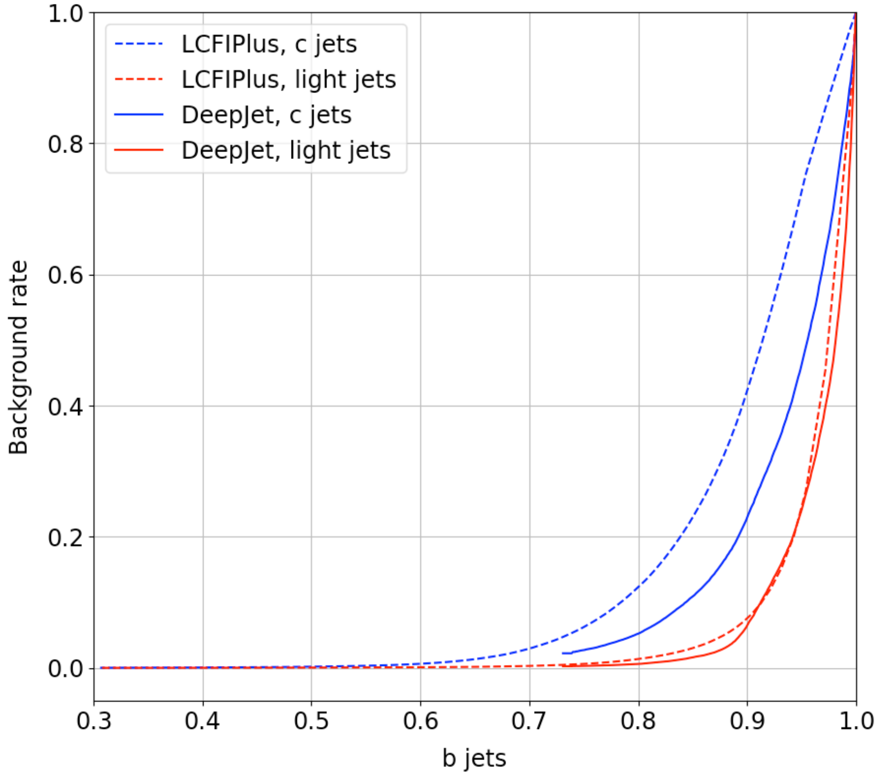


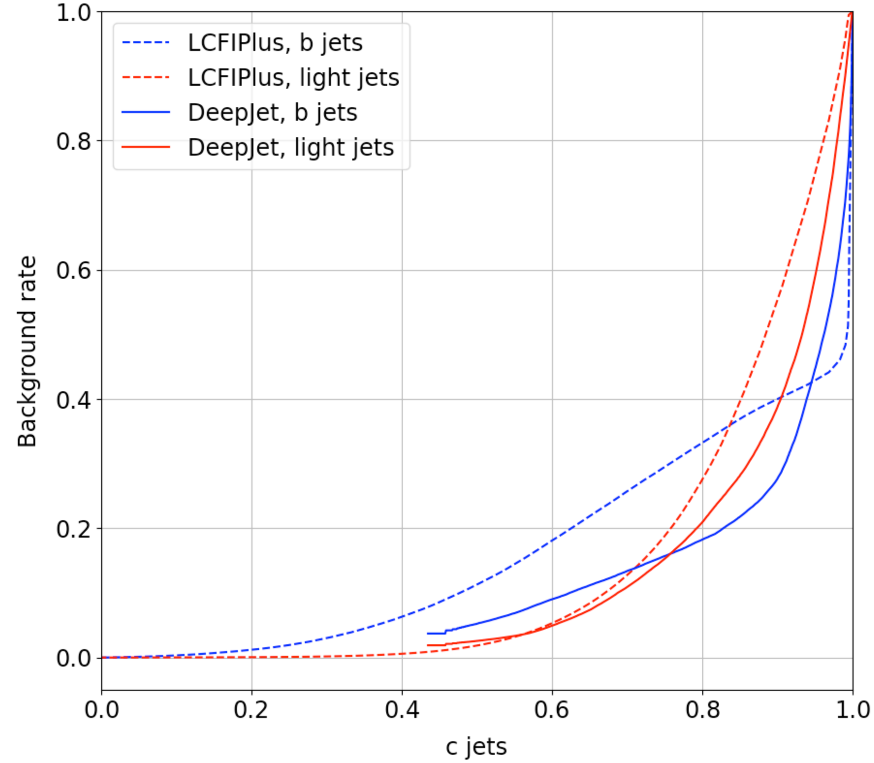


- identification efficiencies of over 80% for b jets & light jets
- c jet identification efficiency lower (70%)
- especially separation between c jets and light jets should be improved

DeepJet: ROC curves - comparison to LCFIPlus

validation data





- better performance of DeepJet training over large parts of the b and c tagging efficiencies

ParticleNet

ParticleNet: data pre-processing

- classify jets into **three classes**: b jets, c jets & light jets
- **no ordering of input particles**
- if a value of a features is not available, the value is set to -10
- **normalize input features** to mean 0, std 1

ParticleNet: input features

Variable	Definition
$\Delta\eta$	difference in pseudorapidity between the particle and the jet axis
$\Delta\phi$	difference in azimuthal angle between the particle and the jet axis
$\log p_T$	logarithm of the particle's p_T
$\log E$	logarithm of the particle's energy
$\log \frac{p_T}{p_T(\text{jet})}$	logarithm of the particle's p_T relative to the jet p_T
$\log \frac{E}{E(\text{jet})}$	logarithm of the particle's energy relative to the jet energy
ΔR	angular separation between the particle and the jet axis ($\sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$)
q	electric charge of the particle
`isElectron`	if the particle is an electron
`isMuon`	if the particle is a muon
`isChargedHadron`	if the particle is a charged hadron
`isNeutralHadron`	if the particle is a neutral hadron
`isPhoton`	if the particle is a photon

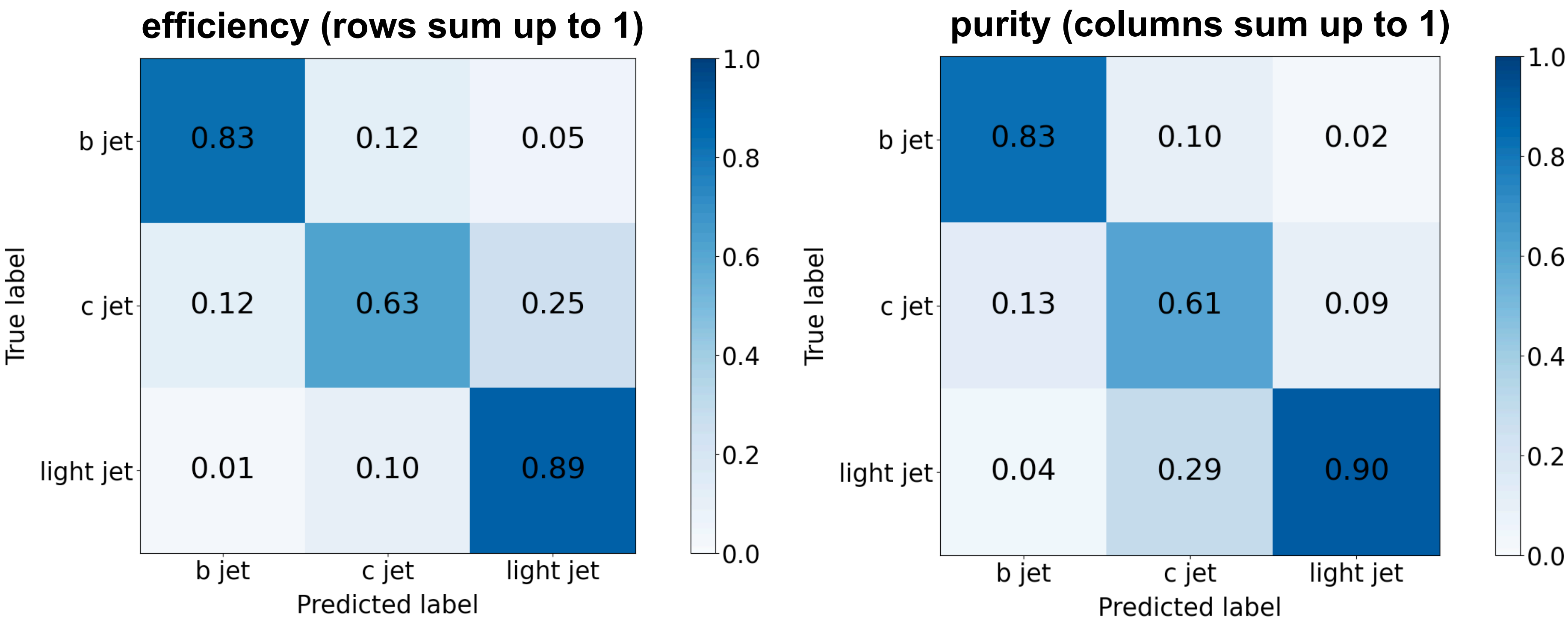
2 SVs,
all jet constituents

- d0, Z0, 3D IP + significances
- Track used in PV?, lepton momentum relative to jet, lepton momentum fraction, kaoness, weighted kaoness momentum fraction, dot product jet and track (norm), HCAL fraction, Chi2/NDF

- **secondary vertices**:
  - coordinates: Δη, ΔΦ
  - features: log(pT), mass, number of tracks, χ2/ndf, 2D & 3D IP and their significances, cos(mom, pos), energy/jet energy, energy, rapidity

ParticleNet: confusion matrices

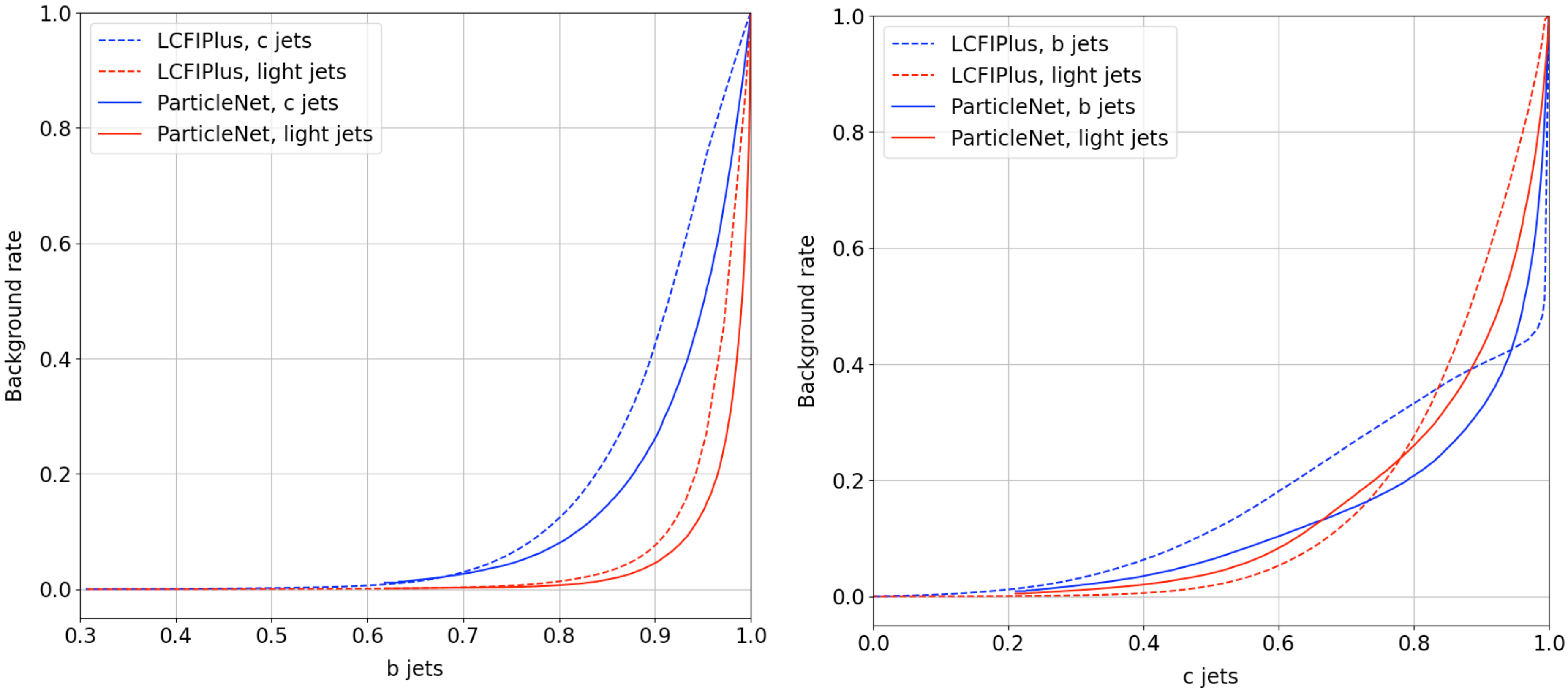
validation data



- identification efficiencies of over 83% for b jets & light jets
- c jet identification quite low (63%)
- especially separation between c jets and light jets should be improved, larger confusion of c jets with b jets than with DeepJet training

ParticleNet: ROC curves - comparison to LCFIPlus

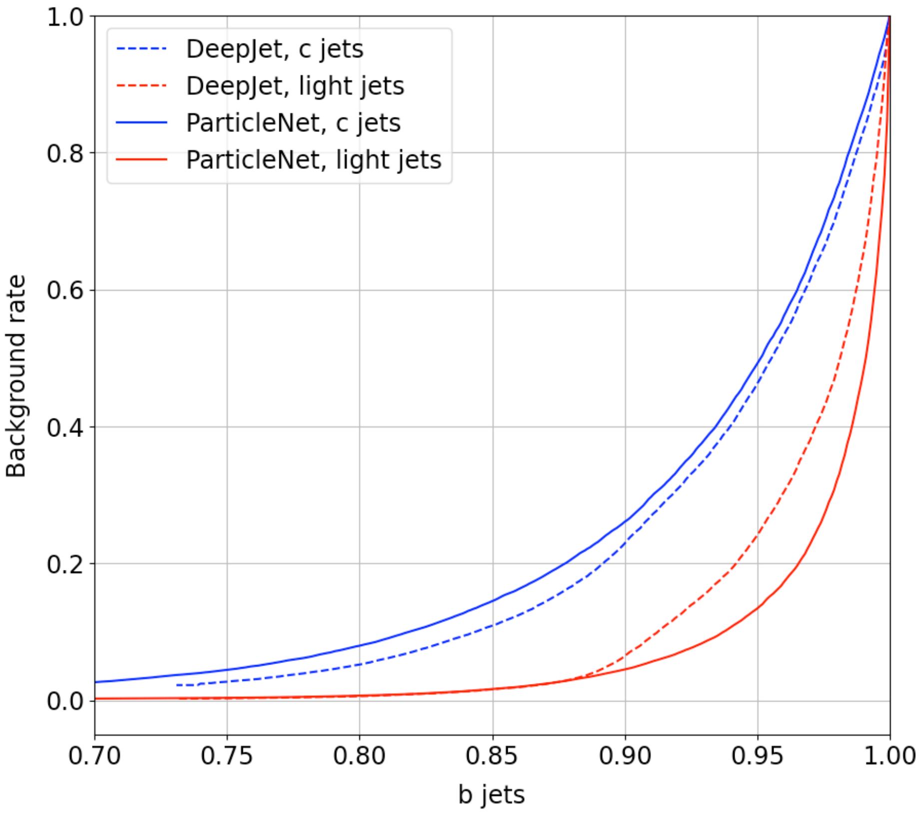
validation data

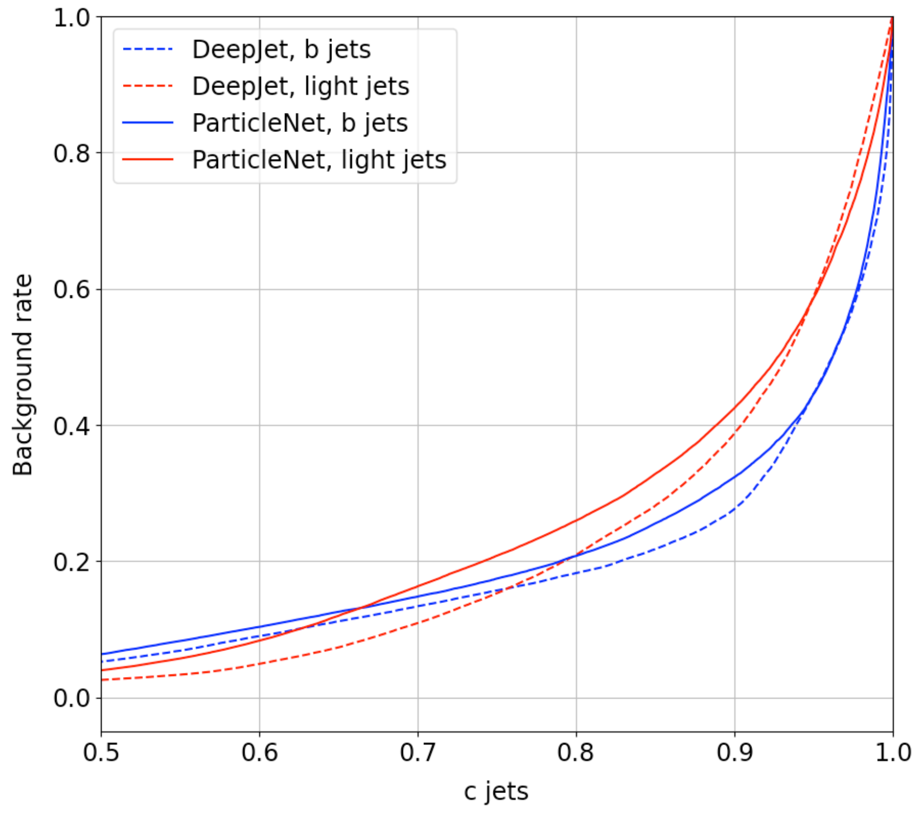


- better performance than LCFIPlus over large parts of the b and c tagging efficiencies
- one of the first trainings with this architecture, a lot of possibilities for optimization (architecture, hyperparameters, features, over-training in c-jet category…)

ParticleNet: ROC curves - comparison to DeepJet

validation data





- better performance of DeepJet training for b vs. c identification and for c vs. b & light jet identification
- better performance of ParticleNet for b jet vs. light jet indentification

Summary & outlook

- application of CMS DeepJet tagger and ParticleNet to ILD
- (large) improvements in b and c jet identification vs. c/b and light jet background w.r.t. default LCFIPlus used in ILD
- ParticleNet model not yet optimized

    ➡ a lot of possibilities to further improve performance

Outlook:

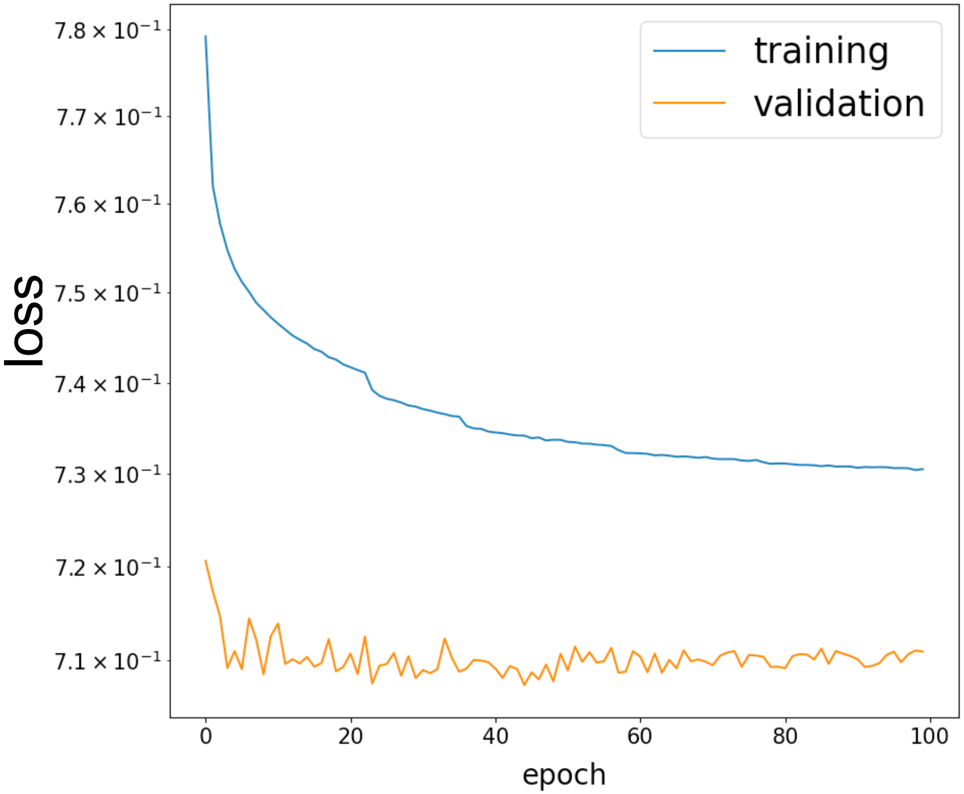
- further optimization of ParticleNet model
- study s-tagging efficiency
- integrate into iLCSoft/Key4hep to make the taggers usable for others

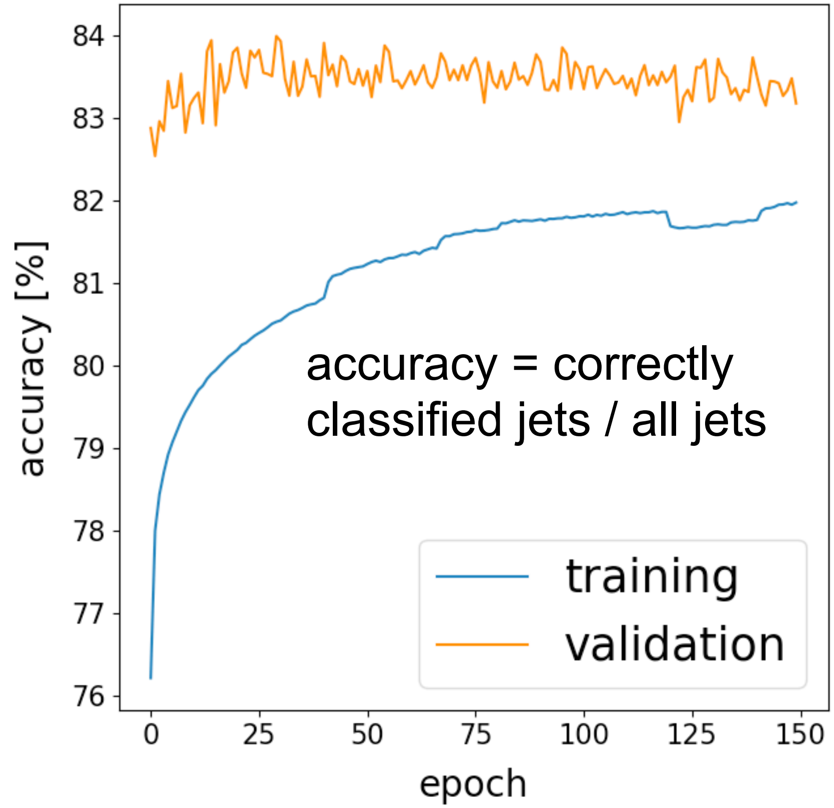
Backup

Training DeepJet

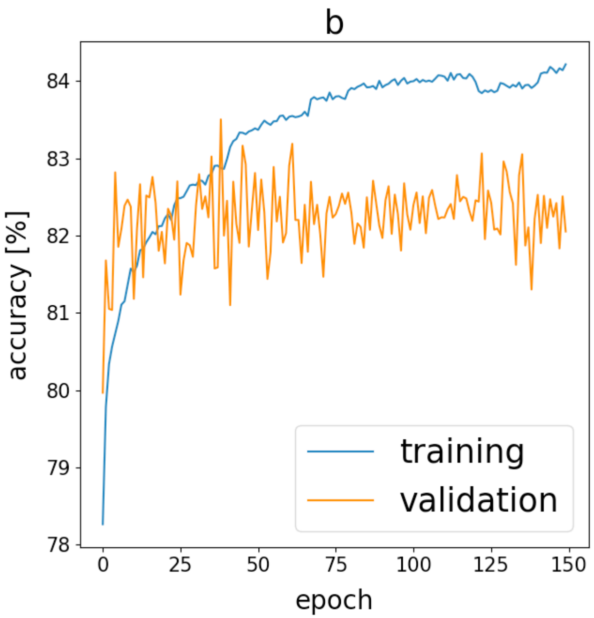
- activation functions: relu / softmax (last layer)
- cross entropy loss
- optimizer: Adam
- regularization: batch normalization, dropout (0.1)
- batch size: 200
- learning rate: 0.0003
- number of epochs: 100
- Xavier weight initialization

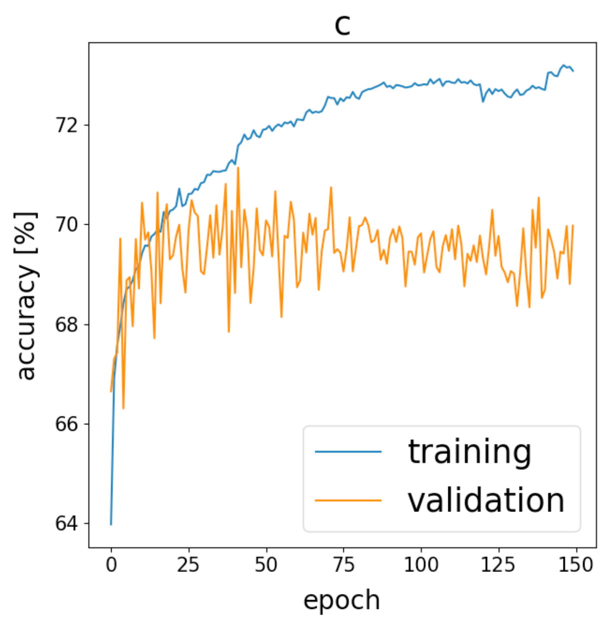
DeepJet: loss & accuracy

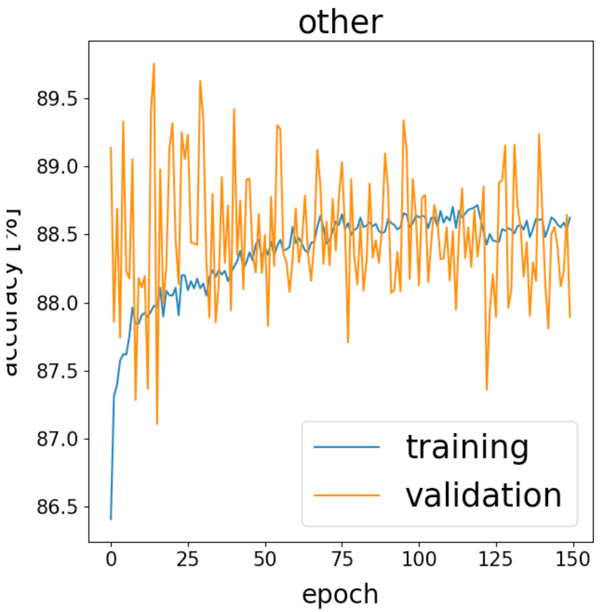




DeepJet: loss & accuracy

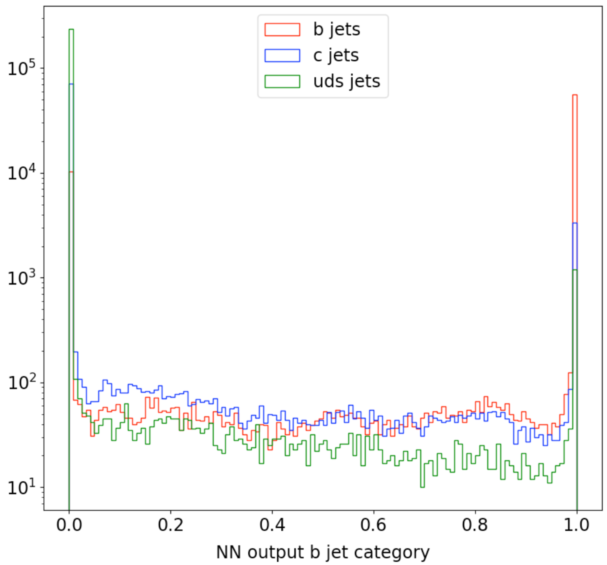


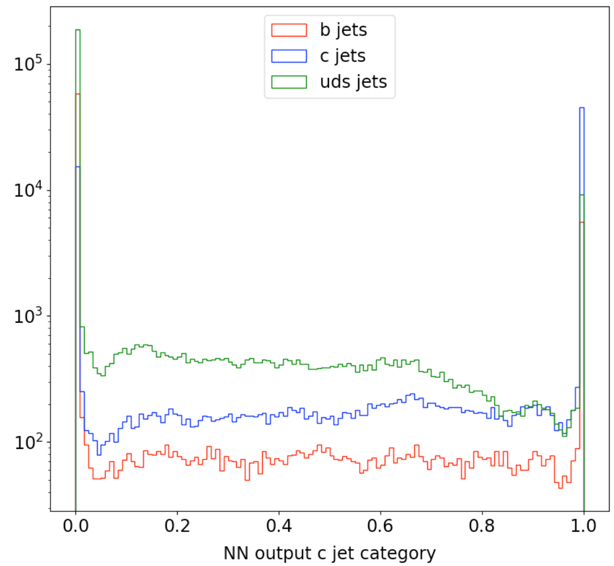


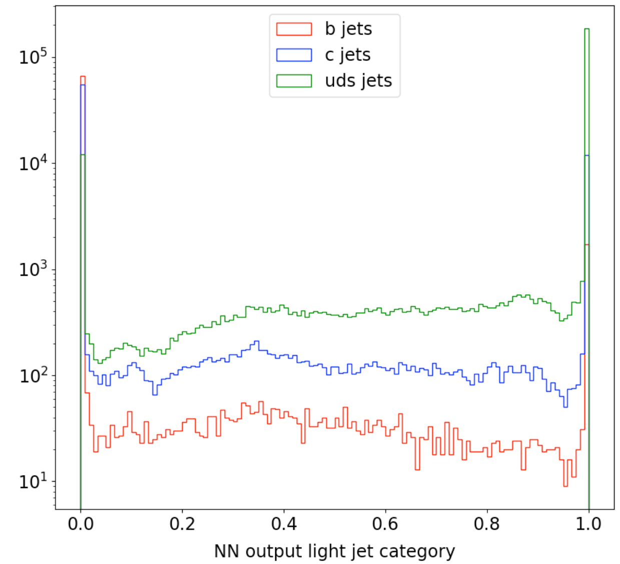


accuracy = correctly classified jets / all jets

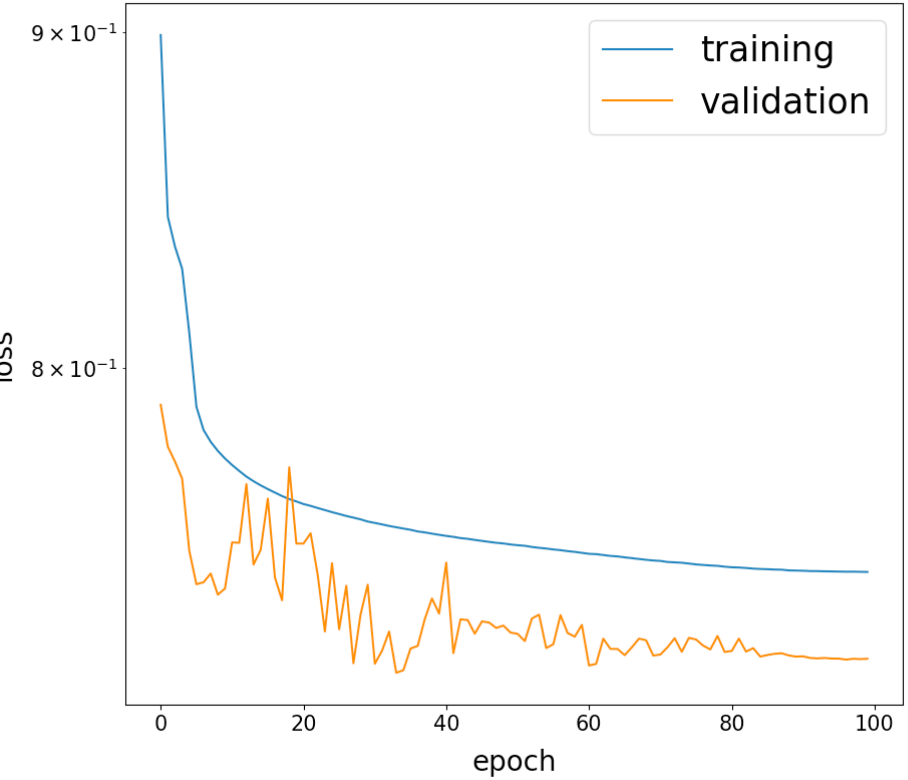
DeepJet: NN output

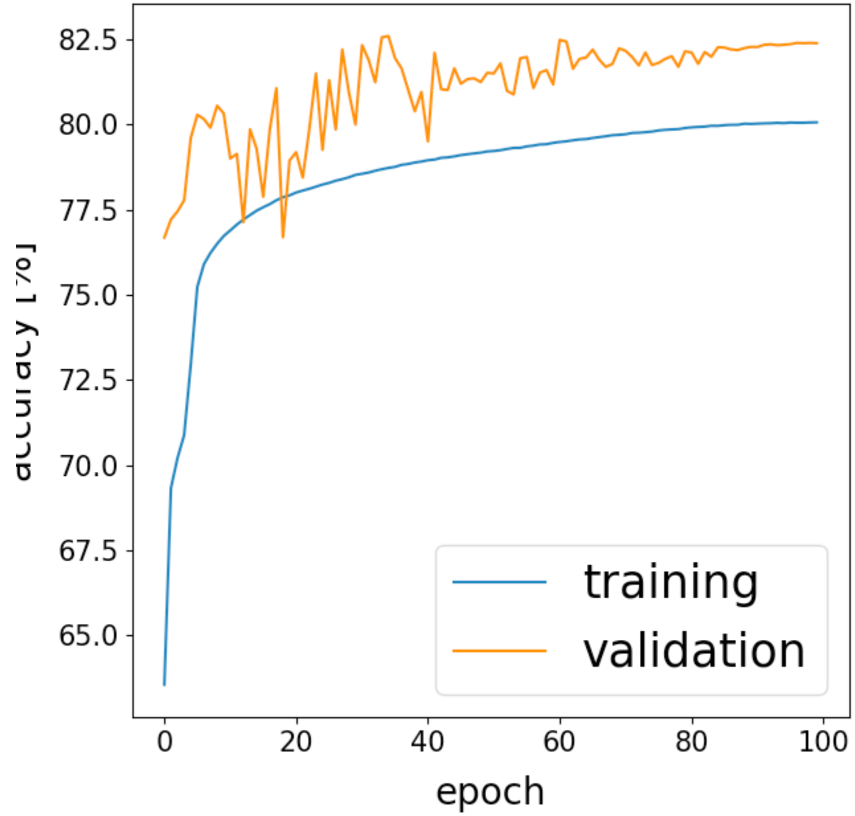




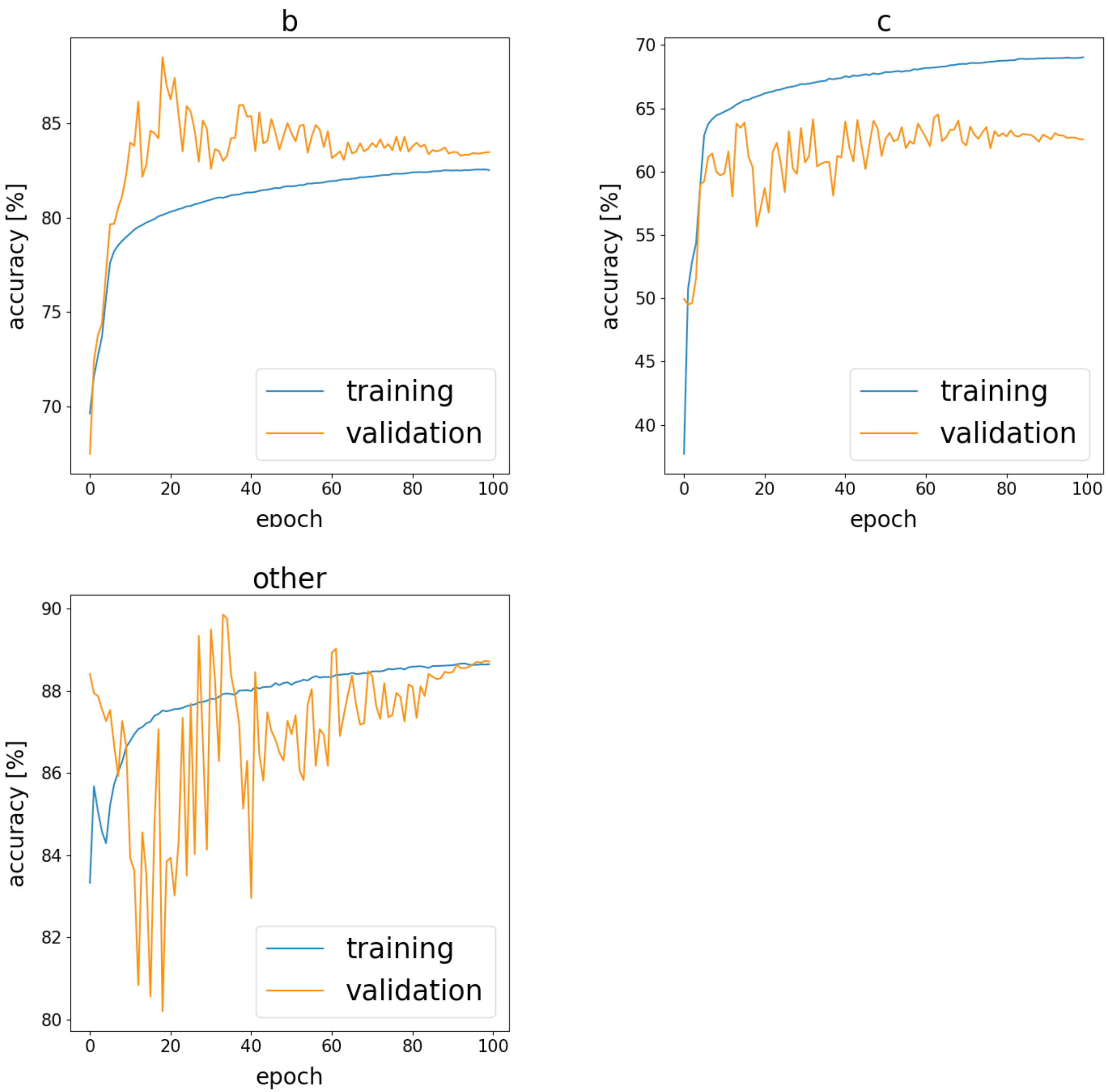


ParticleNet: loss & accuracy

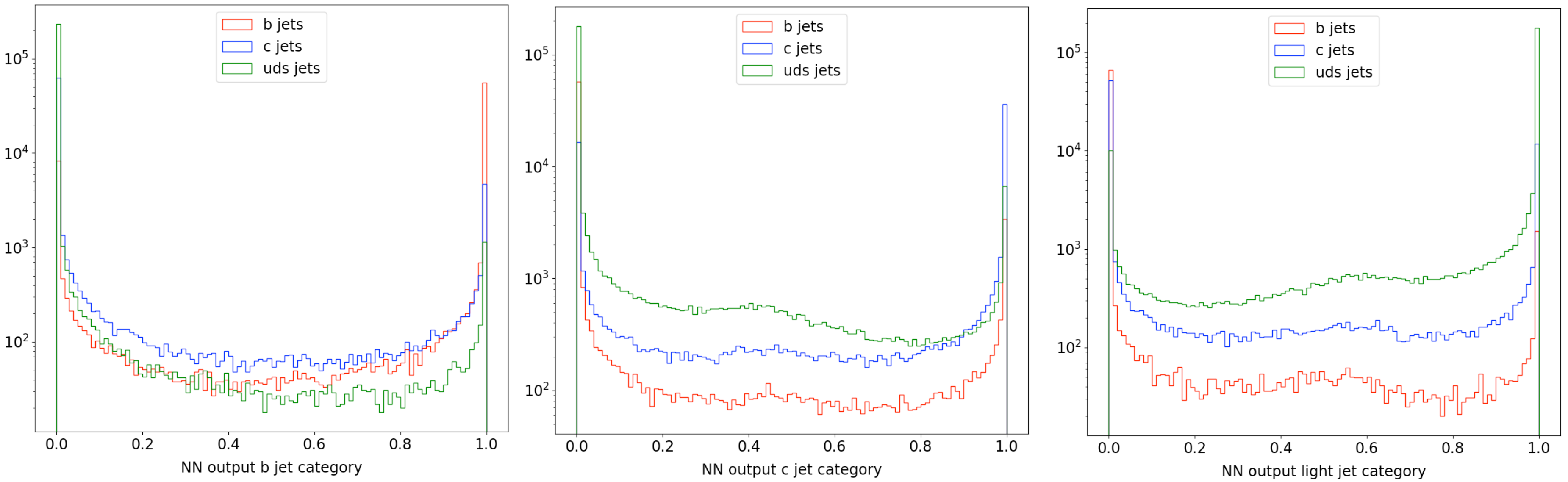




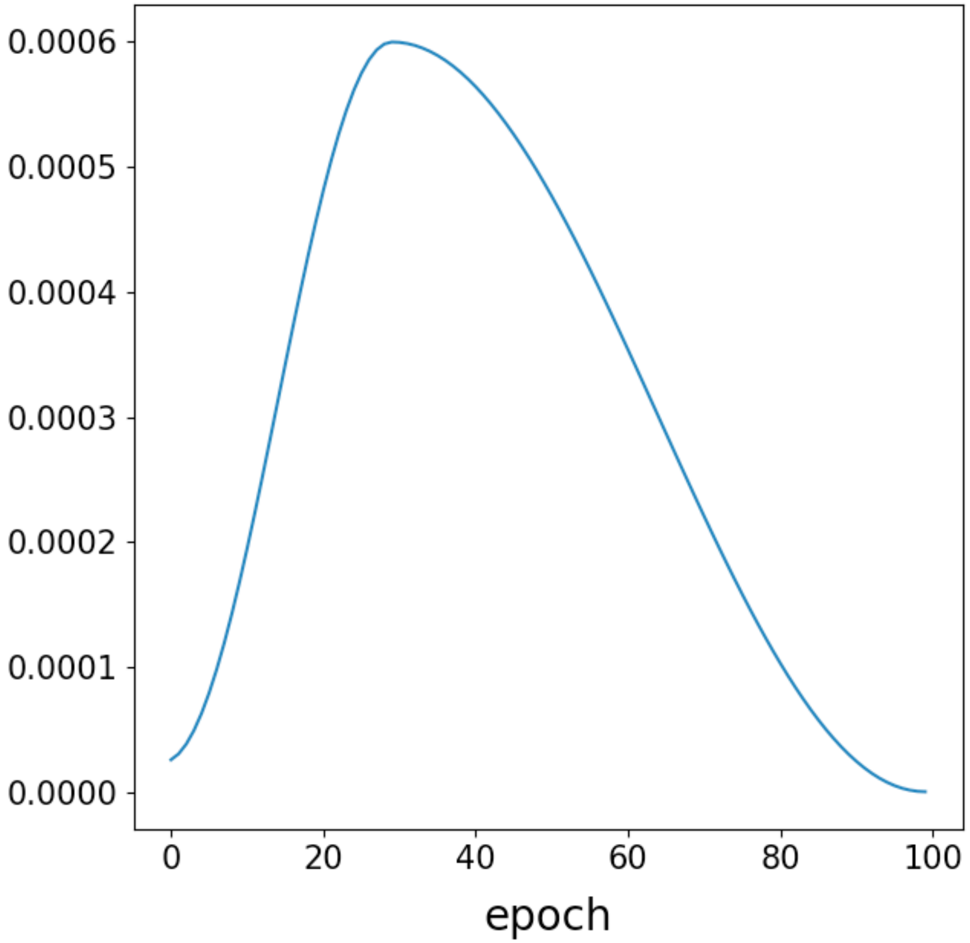
ParticleNet: accuracies



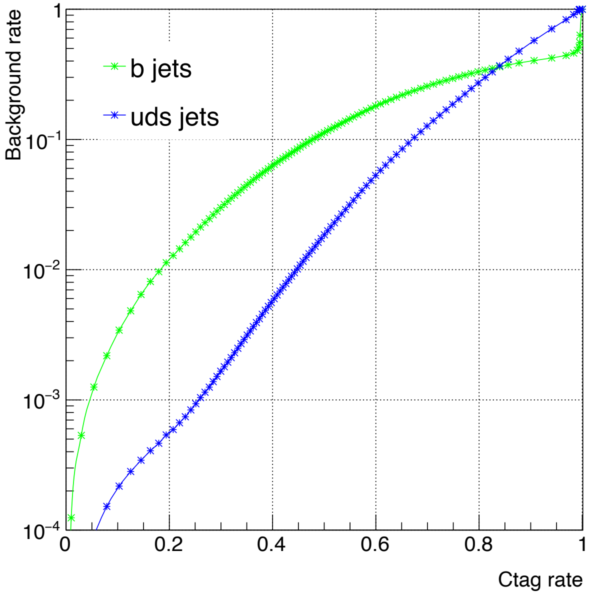
ParticleNet: NN output

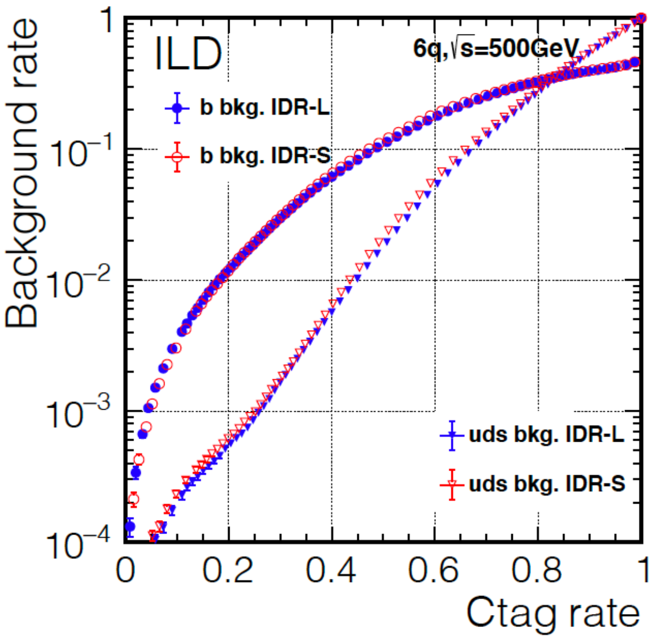


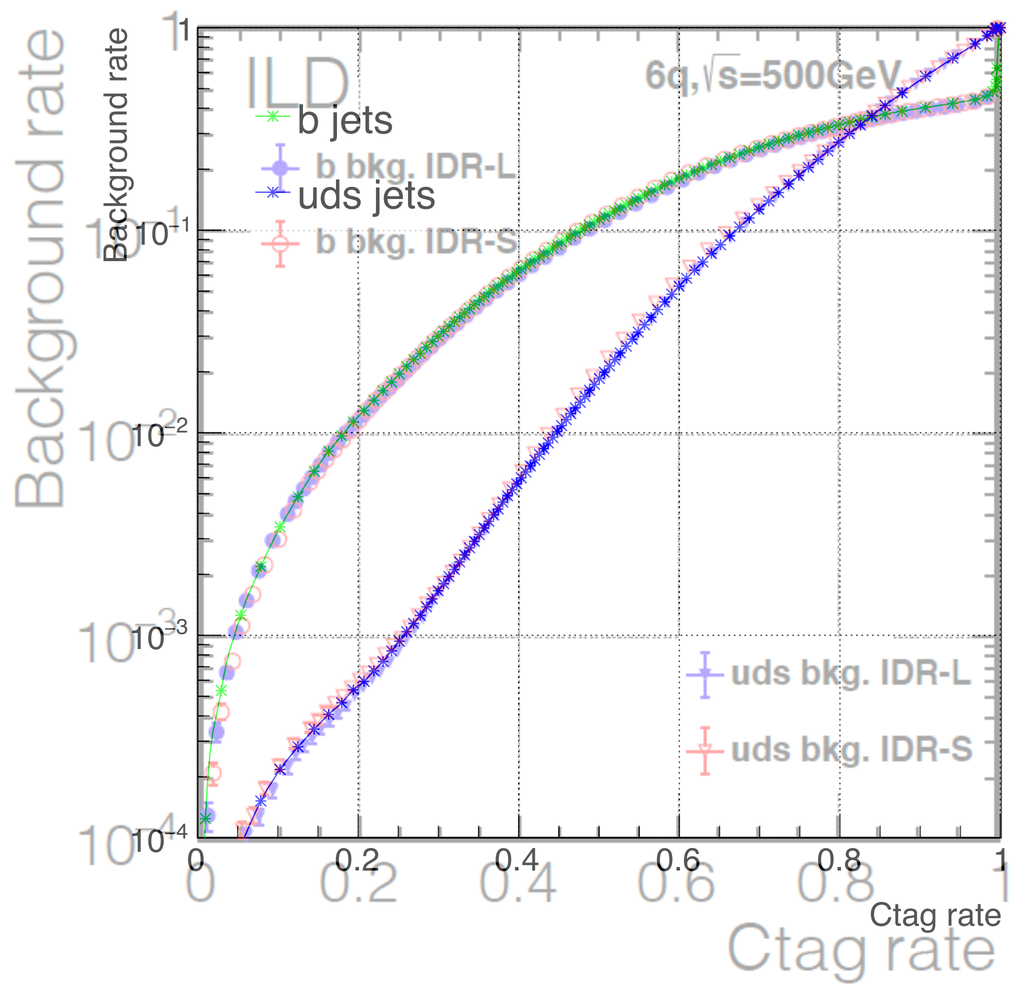
ParticleNet: learning rate



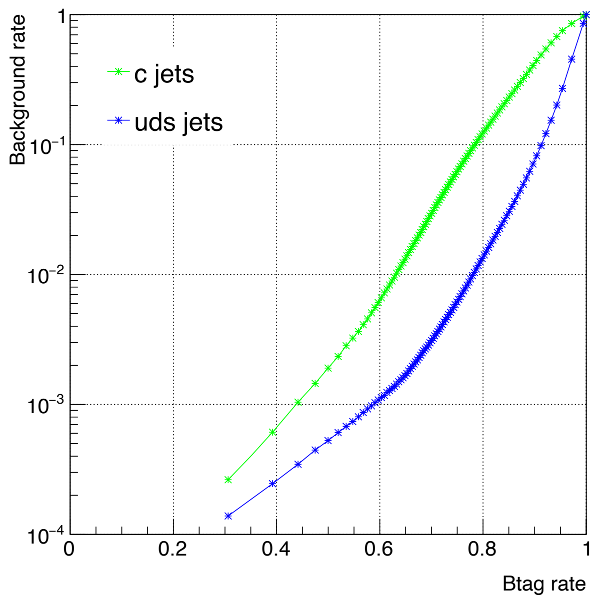
Performance LCFIPlus

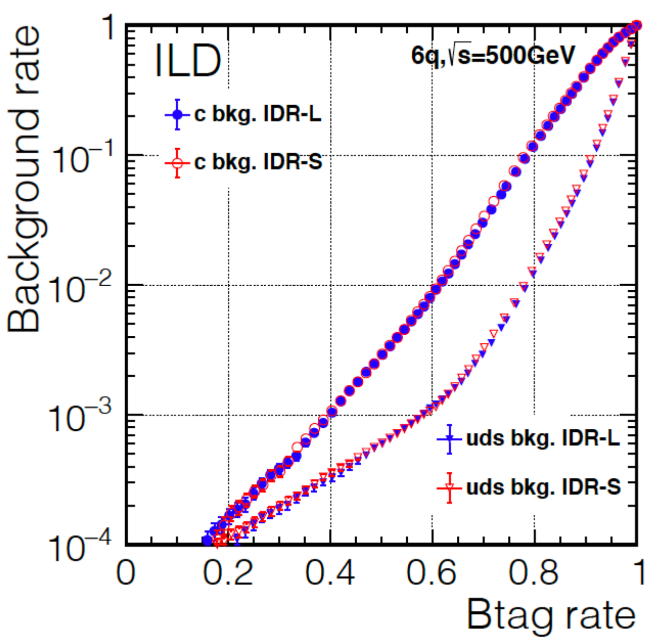


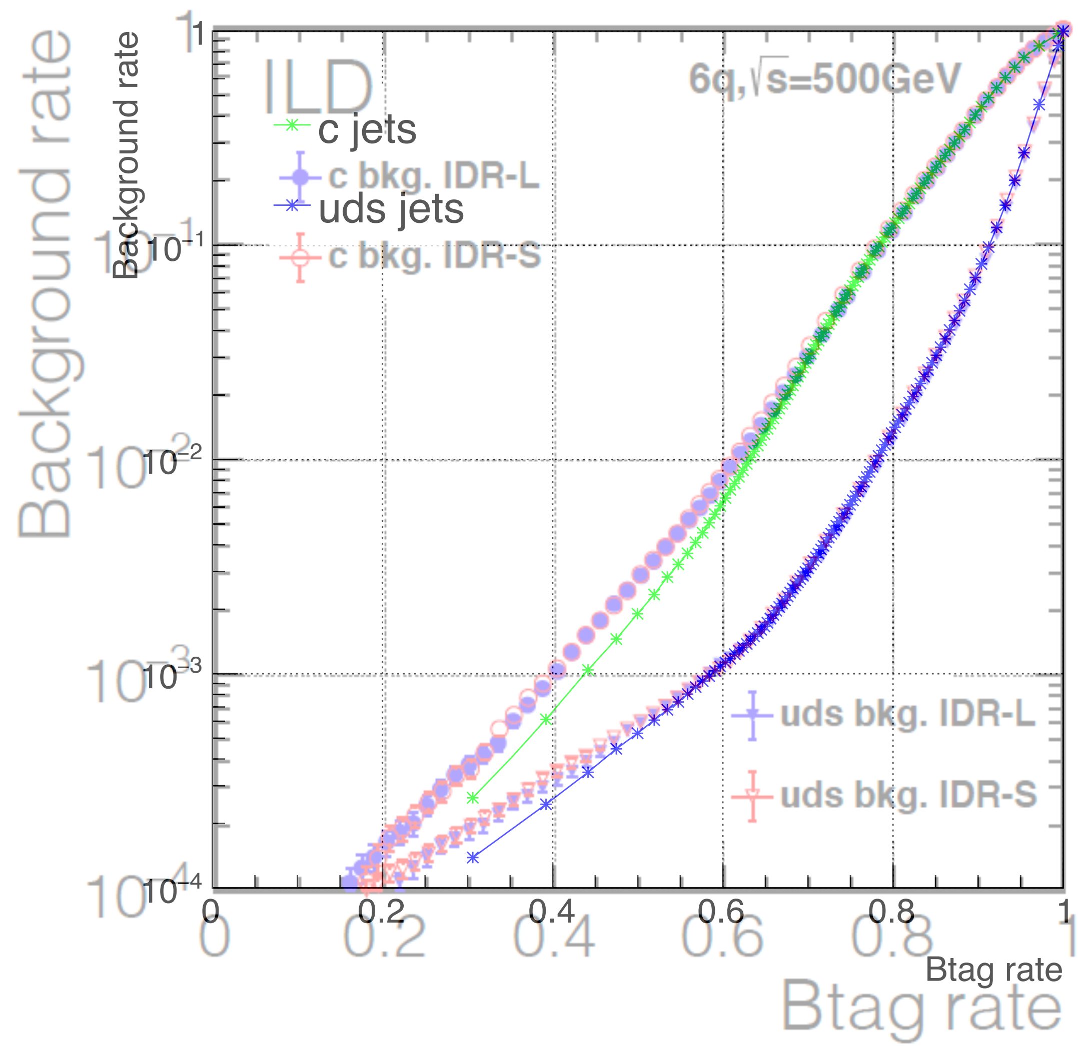




Performance LCFIPlus





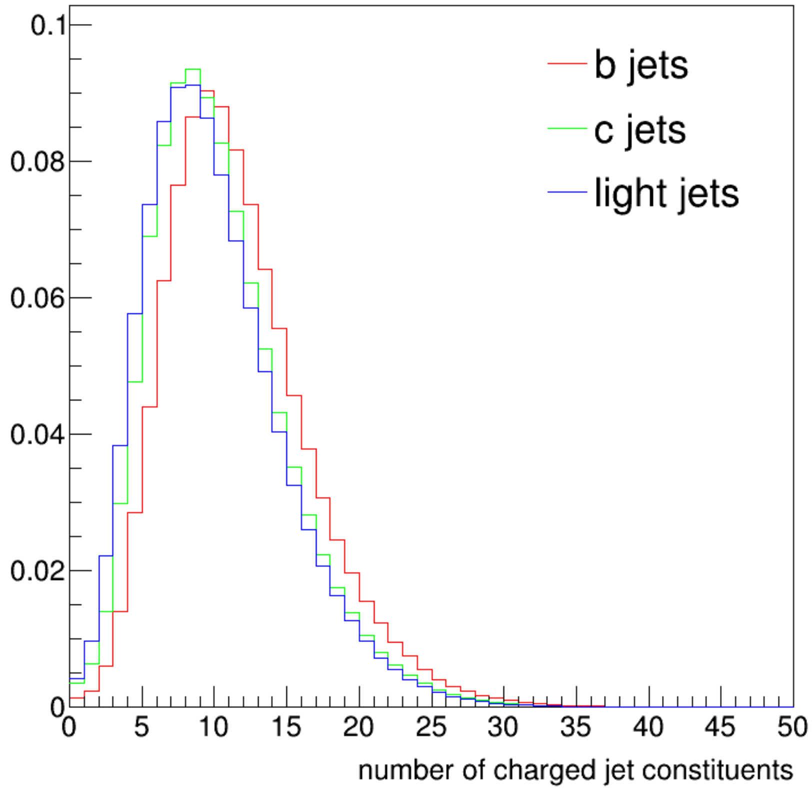


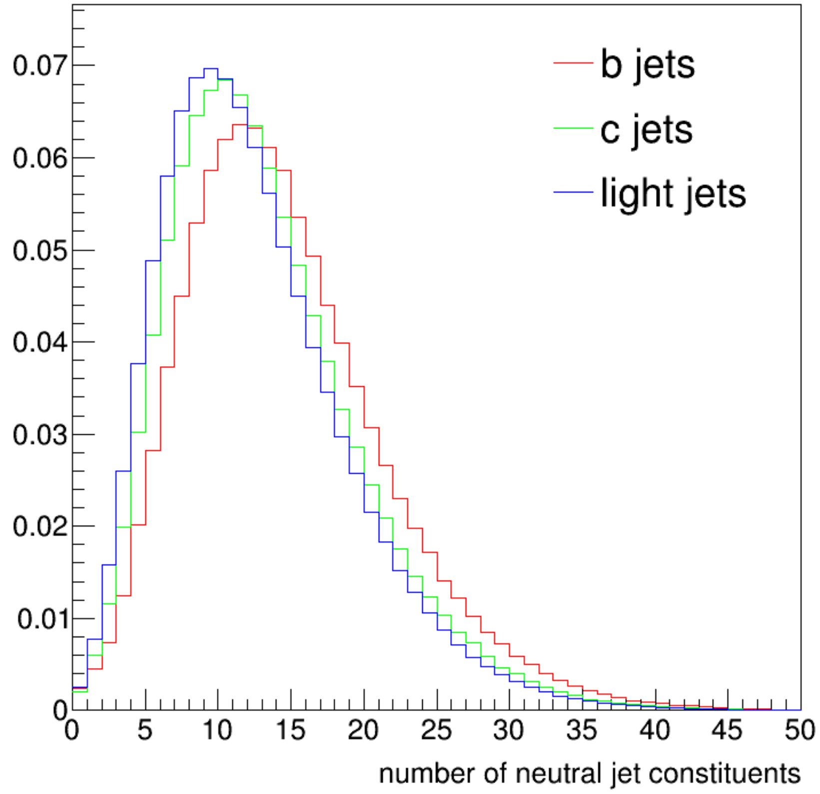
Variables used by LCFIPlus

Name	Description	Normalization factor	Used by category
trk1d0sig	d0 significance of track with highest d0 significance	1	A, B, C, D
trk2d0sig	d0 significance of track with second highest d0 significance	1	A, B, C, D
trk1z0sig	z0 significance of track with highest d0 significance	1	A, B, C, D
trk2z0sig	z0 significance of track with second highest d0 significance	1	A, B, C, D
trk1pt	transverse momentum of track with highest d0 significance	$1/E_{\mathrm{jet}}$	A, B, C, D
trk2pt	transverse momentum of track with second highest d0 significance	$1/E_{\mathrm{jet}}$	A, B, C, D
jprobr	joint probability in the r-phi plane using all tracks	1	A, B, C, D
jprobr5sigma	joint probability in the r-phi plane using all tracks having impact parameter significance exceeding 5 sigma	1	A, B, C, D
jprobz	joint probability in the z projection using all tracks	1	A, B, C, D
jprobz5sigma	joint probability in the z projection using all tracks having impact parameter significance exceeding 5 sigma	1	A, B, C, D
d0bprob	product of b-quark probabilities of d0 values for all tracks, using b/c/q d0 distributions	1	A, B, C, D
d0cprob	product of c-quark probabilities of d0 values for all tracks, using b/c/q d0 distributions	1	A, B, C, D
d0qprob	product of q-quark probabilities of d0 values for all tracks, using b/c/q d0 distributions	1	A, B, C, D
z0bprob	product of b-quark probabilities of z0 values for all tracks, using b/c/q z0 distributions	1	A, B, C, D
z0cprob	product of c-quark probabilities of z0 values for all tracks, using b/c/q z0 distributions	1	A, B, C, D
z0qprob	product of q-quark probabilities of z0 values for all tracks, using b/c/q z0 distributions	1	A, B, C, D
nmuon	number of identified muons	1	A, B, C, D
nelectron	number of identified electrons	1	A, B, C, D
trkmass	mass of all tracks exceeding 5 sigma significance in d0/z0 values	1	A, B, C, D

Variables used by LCFIPlus

Name	Description	Normalization factor	Used by category
1vtxprob	vertex probability with all tracks associated in vertices combined	1	B, C, D
vtxlen1	decay length of the first vertex in the jet (zero if no vertex is found)	$1/E_{\mathrm{jet}}$	B, C, D
vtxlen2	decay length of the second vertex in the jet (zero if number of vertex is less than two)	$1/E_{\mathrm{jet}}$	D
vtxlen12	distance between the first and second vertex (zero if number of vertex is less than two)	$1/E_{\mathrm{jet}}$	D
vtxsig1	decay length significance of the first vertex in the jet (zero if no vertex is found)	$1/E_{\mathrm{jet}}$	B, C, D
vtxsig2	decay length significance of the second vertex in the jet (zero if number of vertex is less than two)	$1/E_{\mathrm{jet}}$	D
vtxsig12	vtxlen12 divided by its error as computed from the sum of the covariance matrix of the first and second vertices, projected along the line connecting the two vertices	$1/E_{\mathrm{jet}}$	D
vtxdirang1	the angle between the momentum (computed as a vector sum of track momenta) and the displacement of the first vertex	E_{jet}	B, C, D
vtxdirang2	the angle between the momentum (computed as a vector sum of track momenta) and the displacement of the second vertex	E_{jet}	D
vtxmult1	number of tracks included in the first vertex (zero if no vertex is found)	1	B, C, D
vtxmult2	number of tracks included in the second vertex (zero if number of vertex is less than two)	1	D
vtxmult	number of tracks which are used to form secondary vertices (summed for all vertices)	1	D
vtxmom1	magnitude of the vector sum of the momenta of all tracks combined into the first vertex	$1/E_{\mathrm{jet}}$	B, C, D
vtxmom2	magnitude of the vector sum of the momenta of all tracks combined into the second vertex	$1/E_{\mathrm{jet}}$	D
vtxmass1	mass of the first vertex computed from the sum of track four-momenta	1	B, C, D
vtxmass2	mass of the second vertex computed from the sum of track four-momenta	1	D
vtxmass	vertex mass as computed from the sum of four momenta of all tracks forming secondary vertices	1	B, C, D
vtxmasspc	mass of the vertex with minimum pt correction allowed by the error matrices of the primary and secondary vertices	1	B, C, D
vtxprob	vertex probability; for multiple vertices, the probability P is computed as 1-P = (1-P1)(1-P2)...(1-PN)	1	B, C, D


b jets
c jets
light jets
0.1
0.08
0.06
0.04
0.02
0
0
5
10
15
20
25
30
35
40
45
50
number of charged jet constituents


b jets
c jets
light jets
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
0
5
10
15
20
25
30
35
40
45
50
number of neutral jet constituents