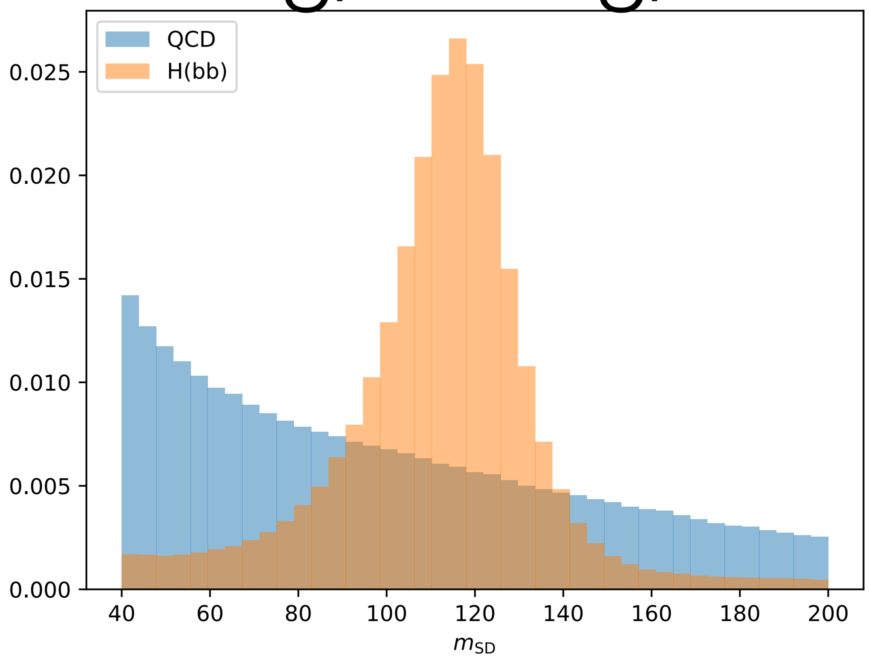
deep double-b tagger

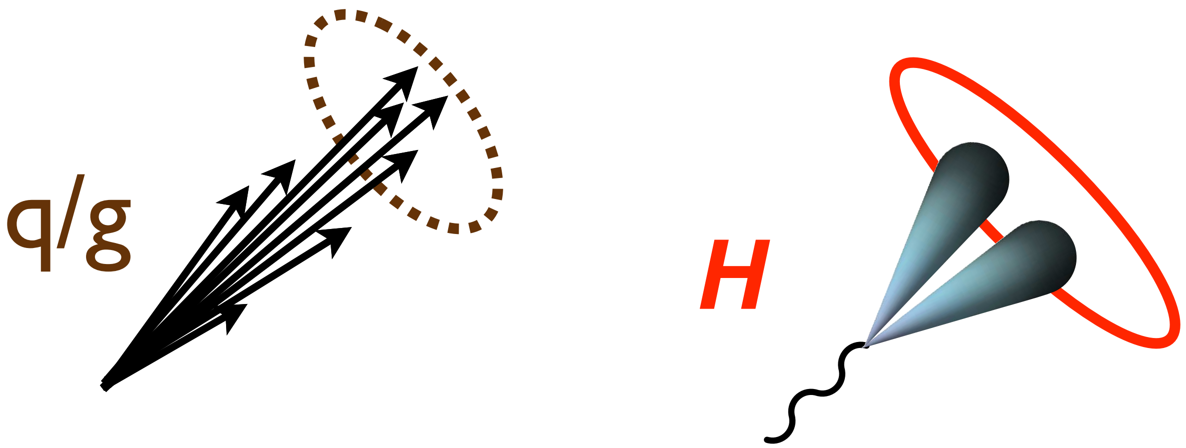
Javier Duarte, Eric Scotti, Caterina Vernieri

11/29/2017

Overview

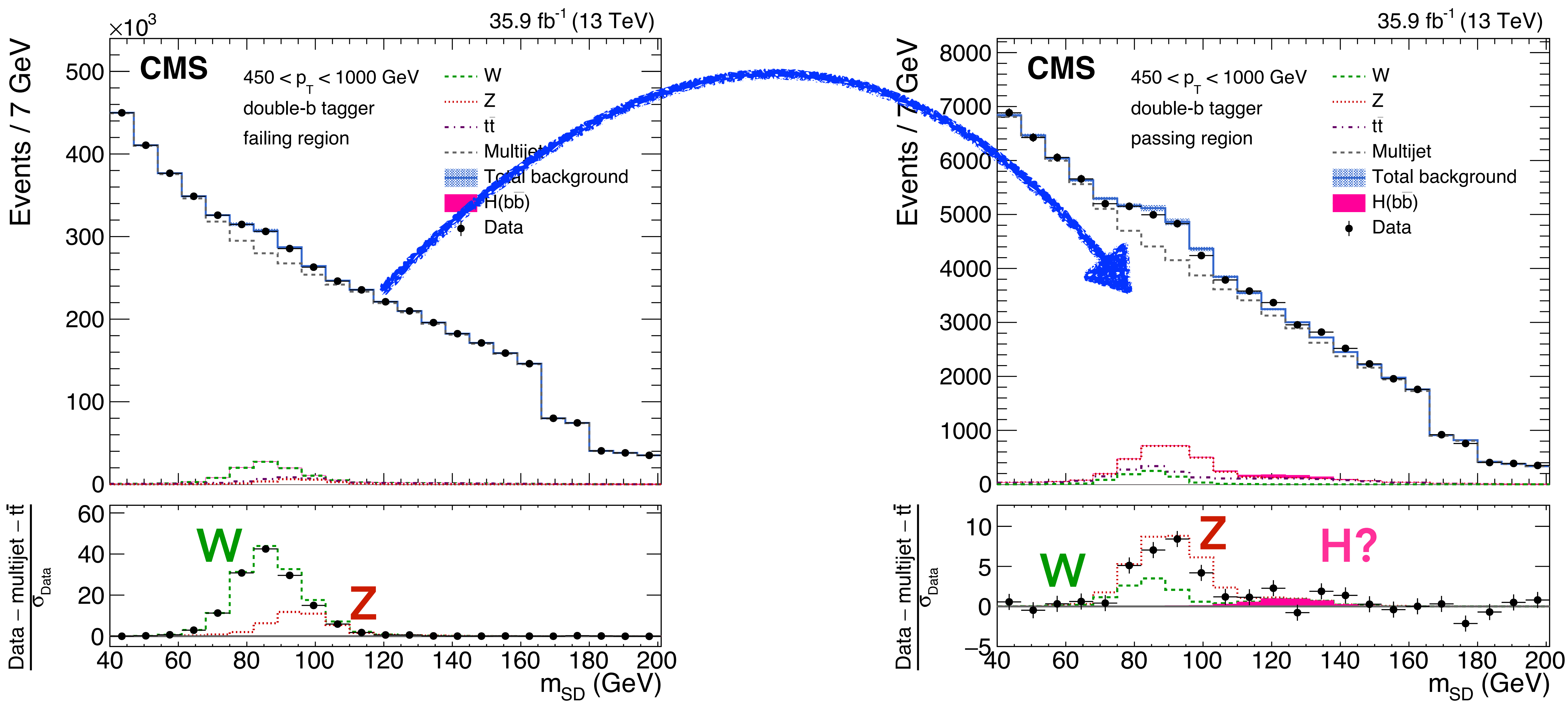
- Input data: CMS simulation of Higgs jets and quark/gluon initiated jets (“QCD”) in [40, 200 GeV] mass range and [300, 2500 GeV] p_T range
- QCD samples: 2.4M, Higgs samples: 2.0M training/testing/validation split: 0.6/0.2/0.2





- Goal: good H(bb) identification without inducing significant mass sculpting (important for data-driven background estimation methods)

Example



- Background prediction in HIG-17-010 proceeds by using taking the m_{SD} shape for the QCD prediction from the anti-double-b tagged region modulated by a smooth transfer factor

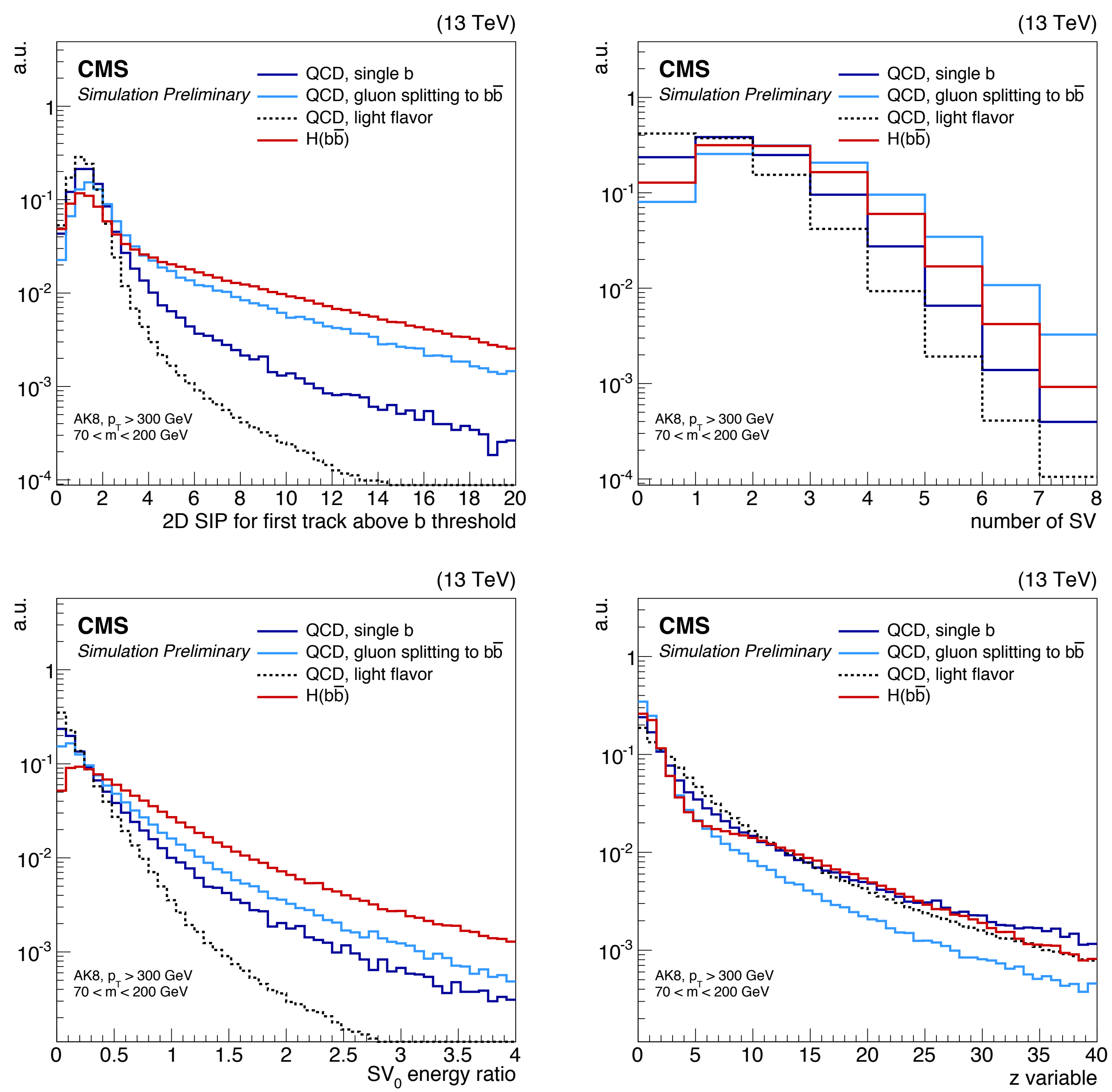
BDT double-b Inputs

double-b features (27)

- The first four SIP values for selected tracks ordered in decreasing SIP;
- For each τ-axis we consider the first two SIP values for their respective associated tracks ordered in decreasing SIP, to further discriminate against single b quark and light flavor jets from QCD when one or both SV are not reconstructed due to IVF inefficiencies;
- The measured IP significance in the plane transverse to the beam axis, 2D SIP, of the first two tracks (first track) that raises the SV invariant mass above the bottom (charm) threshold of 5.2 (1.5) GeV;
- The number of SV associated to the jet;
- The significance of the 2D distance between the primary vertex and the secondary vertex, flight distance, for the SV with the smallest 3D flight distance uncertainty, for each of the two τ-axes;
- The ΔR between the SVs with the smallest 3D flight distance uncertainty and its τ-axis, for each of the two τ-axes;
- The relative pseudorapidity, η_{rel}, of the tracks from all SVs with respect to their τ-axis for the three leading tracks ordered in increasing η_{rel}, for each of the two τ-axes;
- The total SV mass, defined as the total mass of all SVs associated to a given τ-axis, for each of the two τ-axes;
- The ratio of the total SV energy, defined as the total energy of all SVs associated to a given τ-axis, and the total energy of all the tracks associated to the fat jet that are consistent with the primary vertex, for each of the two τ-axes;
- The information related to the two-SV system, the z variable, defined as:

$$z = \Delta R(\mathrm{SV}_0, \mathrm{SV}_1) \cdot \frac{p_{\mathrm{T,SV_1}}}{m(\mathrm{SV}_0, \mathrm{SV}_1)} \tag{2}$$

where SV_0 and SV_1 are SVs with the smallest 3D flight distance uncertainty. The z variable helps rejecting the $\mathrm{b\bar{b}}$ background from gluon splitting relying on the different kinematic properties compared to the $\mathrm{b\bar{b}}$ pair from the decay of a massive resonance.



deep AK8 Inputs

note: “track” features includes also leptons and it could be a challenge for validation

track features (60×30)

Table 10: Full list of charged PF candidate features used as input to the DeepAK8 network

feature	comment
trackEtaRel	BTV
trackPtRatio	BTV
trackPParRatio	BTV
trackSip2dVal	BTV
trackSip2dSig	BTV
trackSip3dVal	BTV
trackSip3dSig	BTV
trackJetDistVal	BTV
$p_T(cPF)/p_T(j)$	
$E_{rel}(cPF)$	
$\Delta\phi(cPF, j)$	
$\Delta\eta(cPF, j)$	
$\Delta R(cPF, j)$	
$\Delta R_m(cPF, SV)$	
$\Delta R(cPF, \text{subjet } 1)$	
$\Delta R(cPF, \text{subjet } 2)$	
χ^2_n	
quality	
d_z	
S_z	
d_{xy}	
S_{xy}	
track_dptdpt	track covariance
track_detadeta	track covariance
track_dphidphi	track covariance
track_dxydxy	track covariance
track_dzdz	track covariance
track_dxydz	track covariance
track_dphidxy	track covariance
track_dlambdadz	track covariance

PF cand. features (100×10)

Table 11: Full list of inclusive PF candidate features used as input to the DeepAK8 network

feature
$p_T(PF)/p_T(j)$
$E_{rel}(PF)$
$\Delta\phi(PF, j)$
$\Delta\eta(PF, j)$
$\Delta R(PF, j)$
$\Delta R_m(PF, SV)$
$\Delta R(PF, \text{subjet } 1)$
$\Delta R(PF, \text{subjet } 2)$
$w_p(PF)$
f_{HCAL}

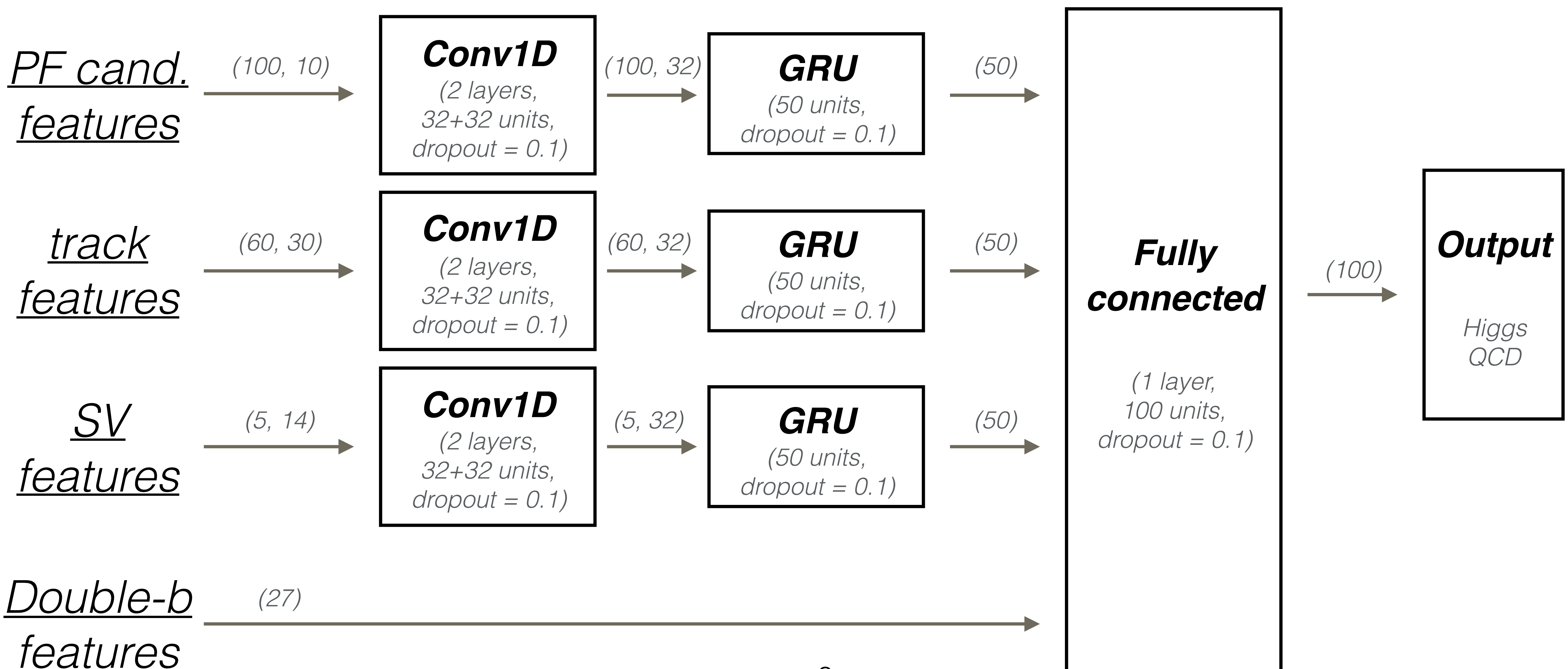
SV features (5×14)

Table 12: Full list of secondary vertex features used as input to the DeepAK8 network

feature
$p_T(SV)/p_T(j)$
$E_{rel}(SV)$
$\Delta\phi(SV, j)$
$\Delta\eta(SV, j)$
$\Delta R(SV, j)$
$p_T(SV)$
m_{SV}
$N_{\text{tracks}}(SV)$
$\chi^2_n(SV)$
$d_{xy}(SV)$
$S_{xy}(SV)$
$d_{3D}(SV)$
$S_{3D}(SV)$
$\cos\theta(SV)$

Conv1D + GRU network topology

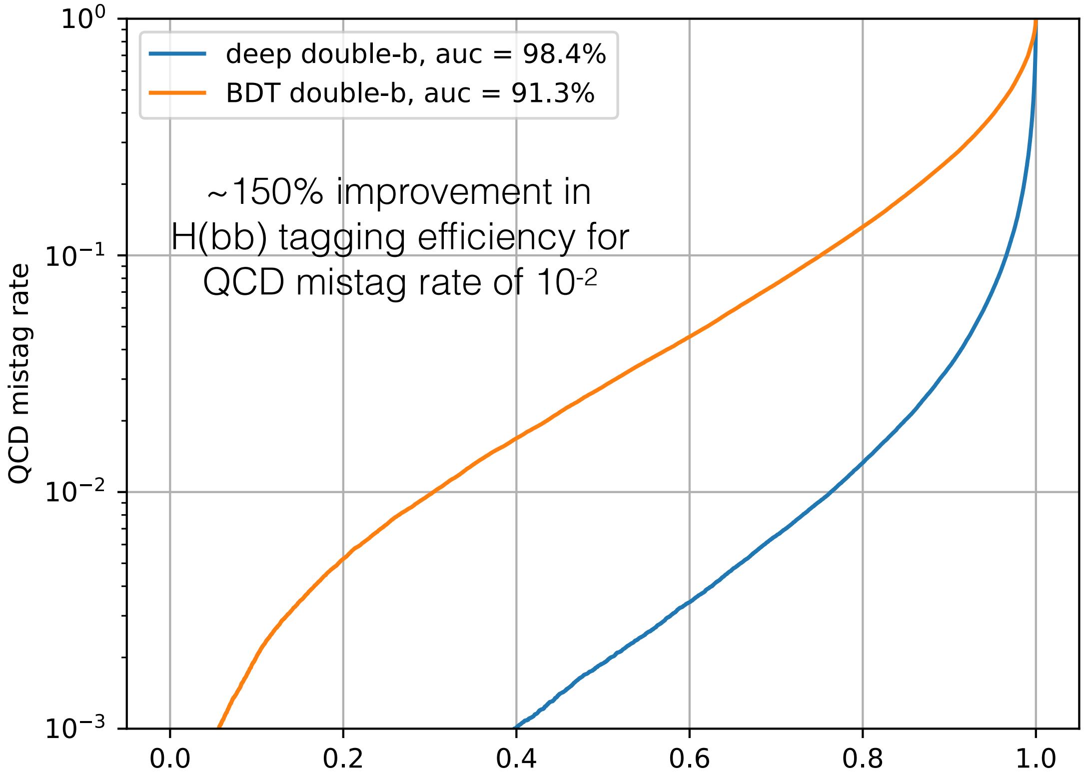
- 27 high-level (double-b) features + 100×10 PF candidate features + 60×30 track features + 5×14 secondary vertex features per Higgs-candidate jet
- Conv1D with kernel size 1 = Time-distributed dense = apply same dense network to each PF candidate / track / SV
- GRU = Gated Recurrent Unit = Recurrent network to reduce dimensionality of output from Conv1D layers (100×32, 60×32, 5×32) → (50, 50, 50)

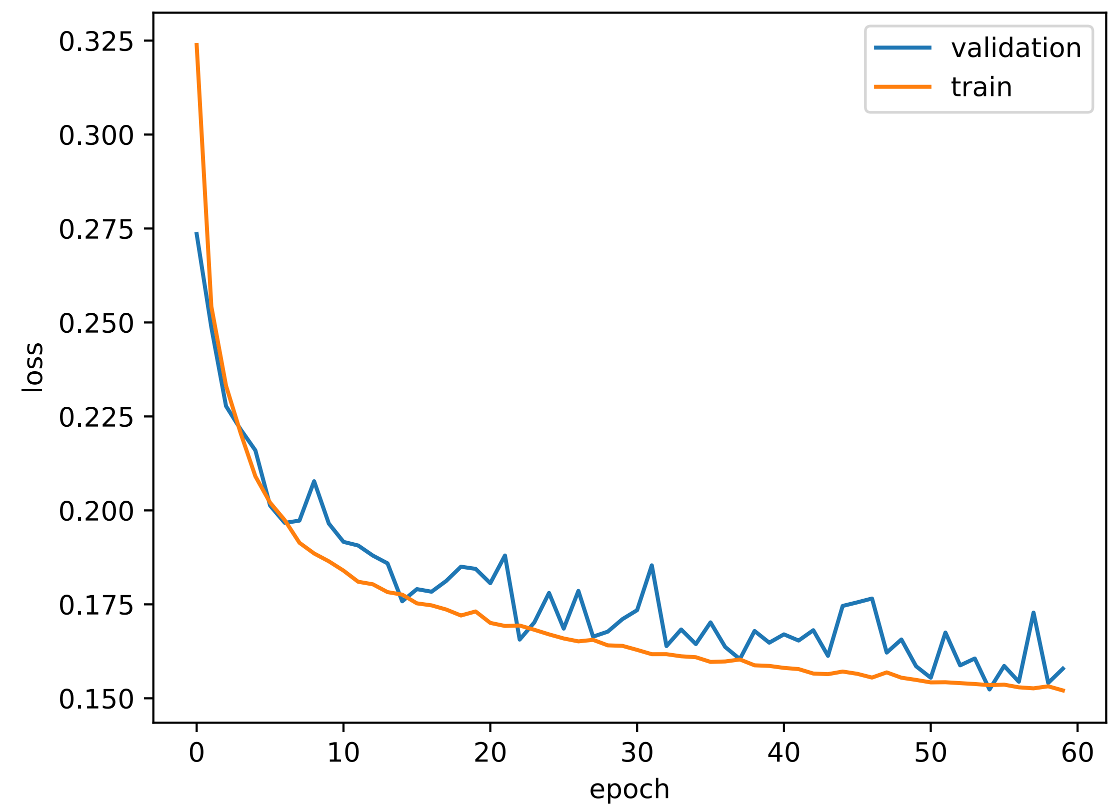


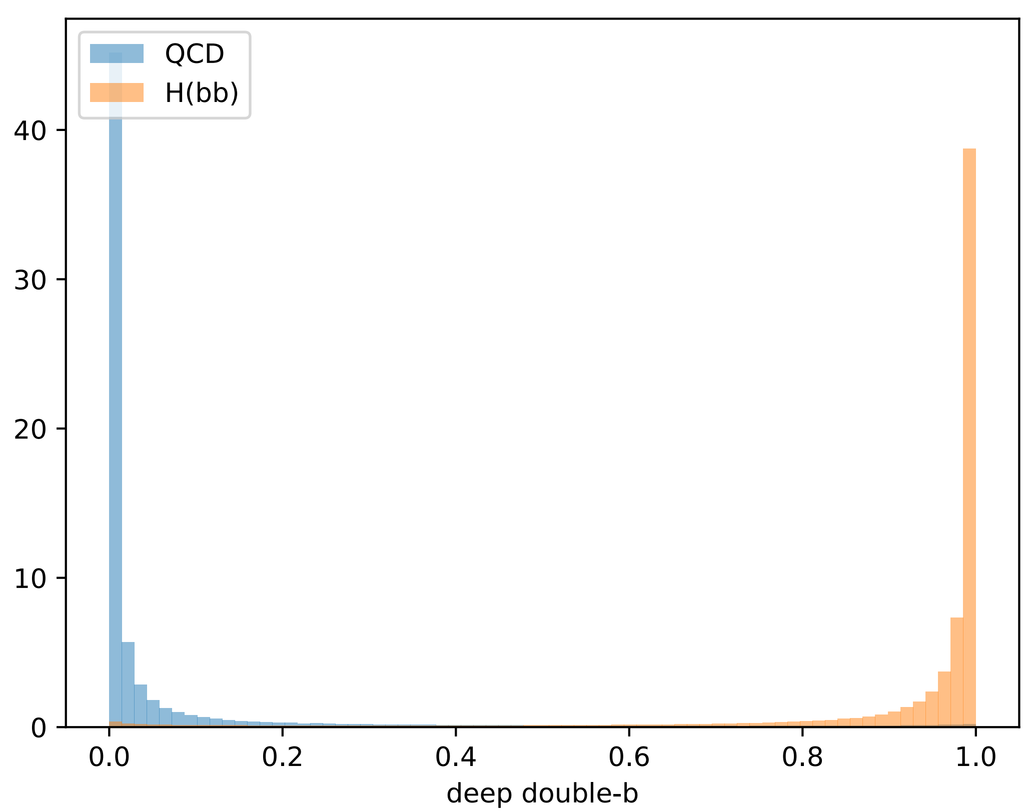
Conv1D + GRU network topology

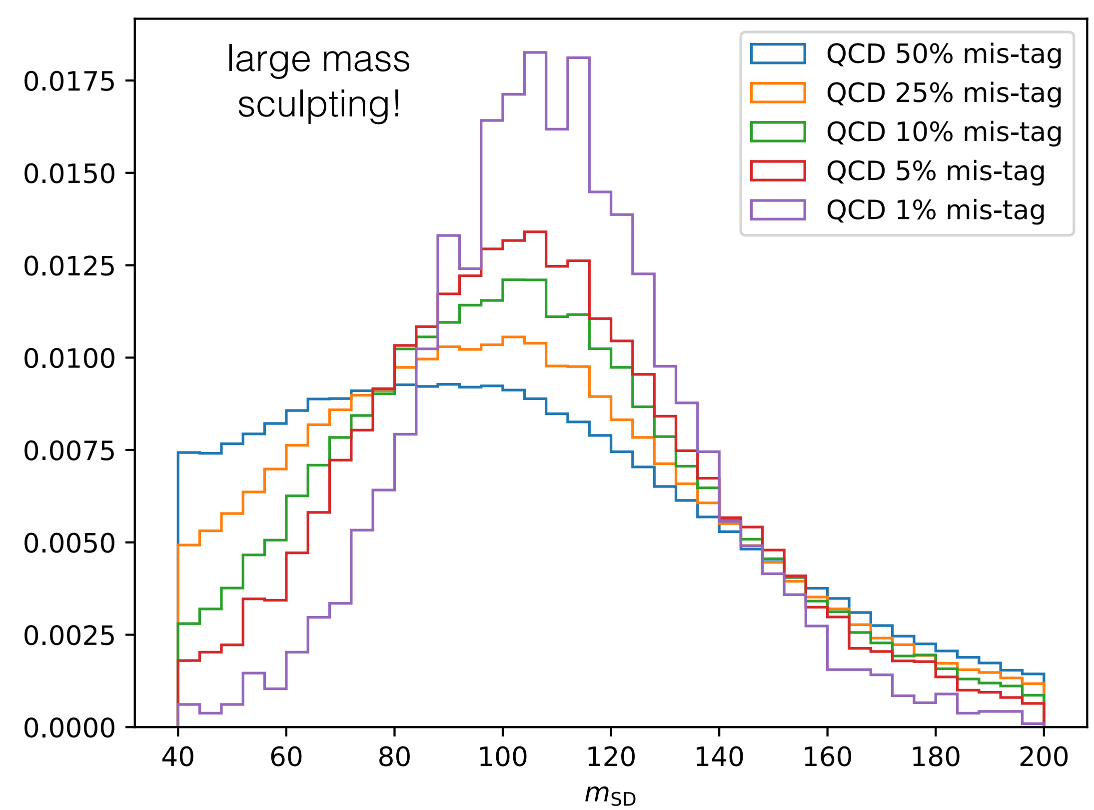
```
Layer (type)                     Output Shape          Param #     Connected to
====================================================================================================
input_2 (InputLayer)             (None, 100, 10)       0
____________________________________________________________________________________________________
input_3 (InputLayer)             (None, 60, 30)        0
____________________________________________________________________________________________________
input_4 (InputLayer)             (None, 5, 14)         0
____________________________________________________________________________________________________
pf_conv1 (Conv1D)                (None, 100, 32)       320
____________________________________________________________________________________________________
cpf_conv1 (Conv1D)               (None, 60, 32)        960
____________________________________________________________________________________________________
sv_conv1 (Conv1D)                (None, 5, 32)         448
____________________________________________________________________________________________________
spatial_dropout1d_1 (SpatialDrop (None, 100, 32)       0
____________________________________________________________________________________________________
spatial_dropout1d_3 (SpatialDrop (None, 60, 32)        0
____________________________________________________________________________________________________
spatial_dropout1d_5 (SpatialDrop (None, 5, 32)         0
____________________________________________________________________________________________________
pf_conv2 (Conv1D)                (None, 100, 32)       1024
____________________________________________________________________________________________________
cpf_conv2 (Conv1D)               (None, 60, 32)        1024
____________________________________________________________________________________________________
sv_conv2 (Conv1D)                (None, 5, 32)         1024
____________________________________________________________________________________________________
spatial_dropout1d_2 (SpatialDrop (None, 100, 32)       0
____________________________________________________________________________________________________
spatial_dropout1d_4 (SpatialDrop (None, 60, 32)        0
____________________________________________________________________________________________________
spatial_dropout1d_6 (SpatialDrop (None, 5, 32)         0
____________________________________________________________________________________________________
input_1 (InputLayer)             (None, 1, 27)         0
____________________________________________________________________________________________________
gru_1 (GRU)                      (None, 50)            12450
____________________________________________________________________________________________________
gru_2 (GRU)                      (None, 50)            12450
____________________________________________________________________________________________________
gru_3 (GRU)                      (None, 50)            12450
____________________________________________________________________________________________________
flatten_1 (Flatten)              (None, 27)            0
____________________________________________________________________________________________________
dropout_1 (Dropout)              (None, 50)            0
____________________________________________________________________________________________________
dropout_2 (Dropout)              (None, 50)            0
____________________________________________________________________________________________________
dropout_3 (Dropout)              (None, 50)            0
____________________________________________________________________________________________________
concat (Concatenate)             (None, 177)           0
____________________________________________________________________________________________________
fc1_relu (Dense)                 (None, 100)           17800
____________________________________________________________________________________________________
fc1_dropout (Dropout)            (None, 100)           0
____________________________________________________________________________________________________
softmax (Dense)                  (None, 2)             202
====================================================================================================
Total params: 60,152.0
```

Performance
(double-b + PF + track + SV)

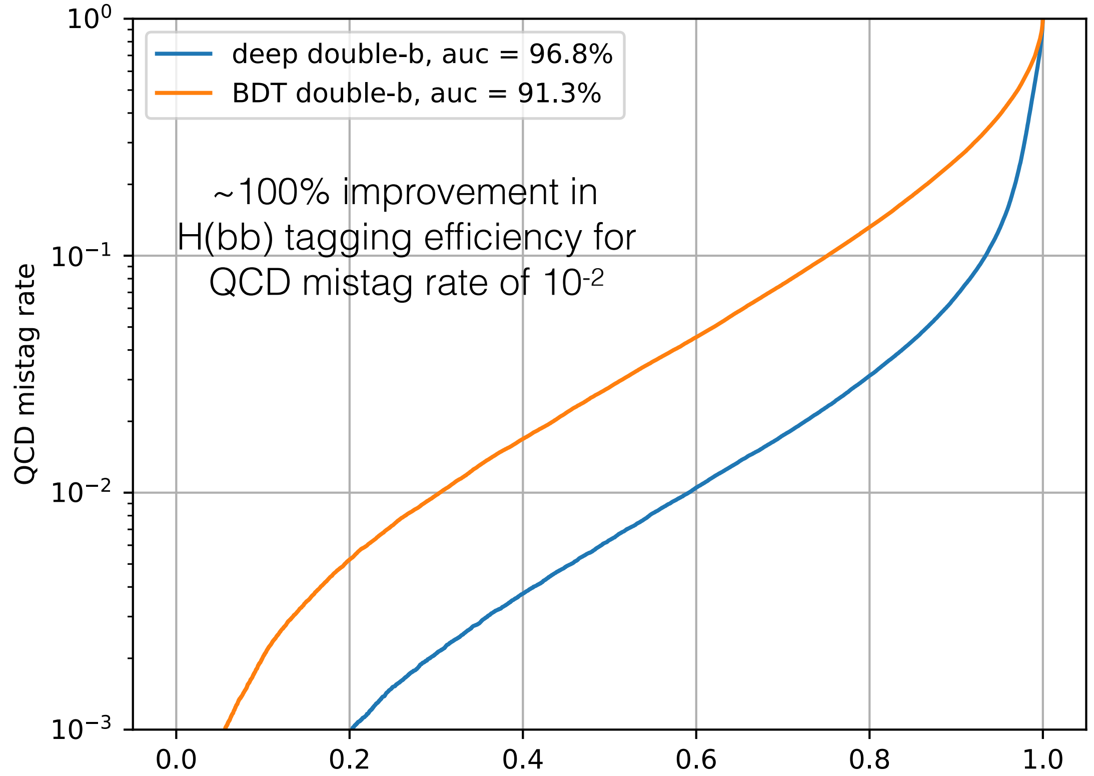


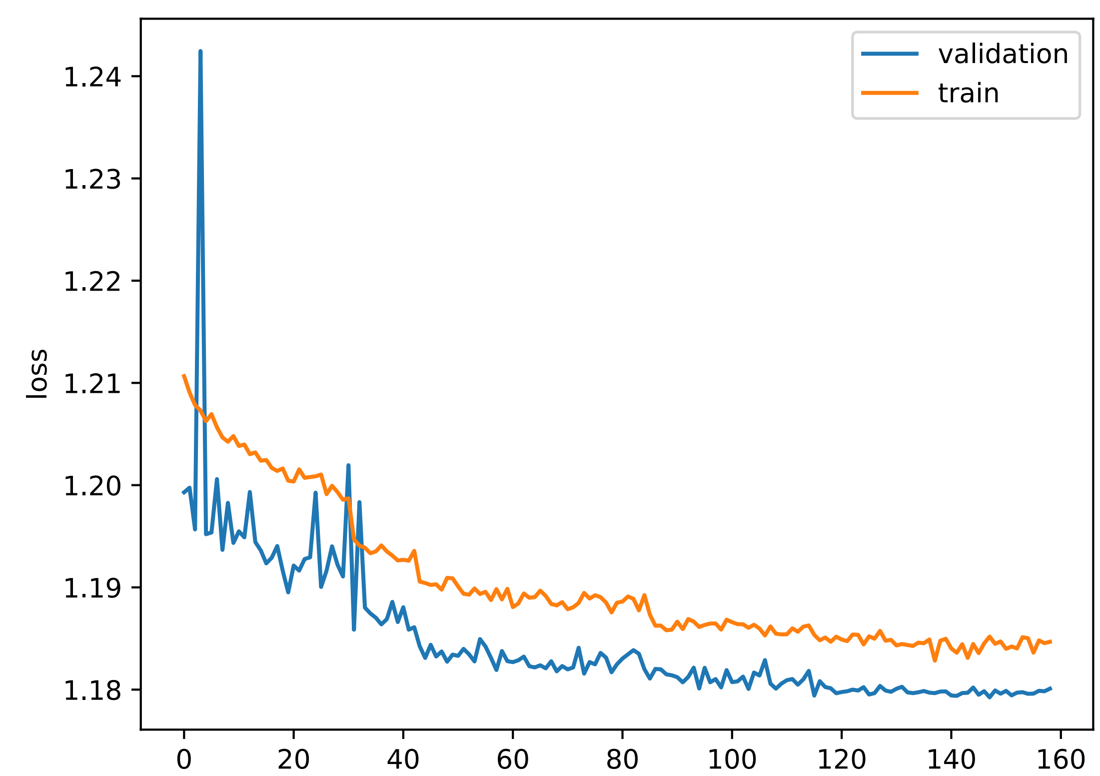


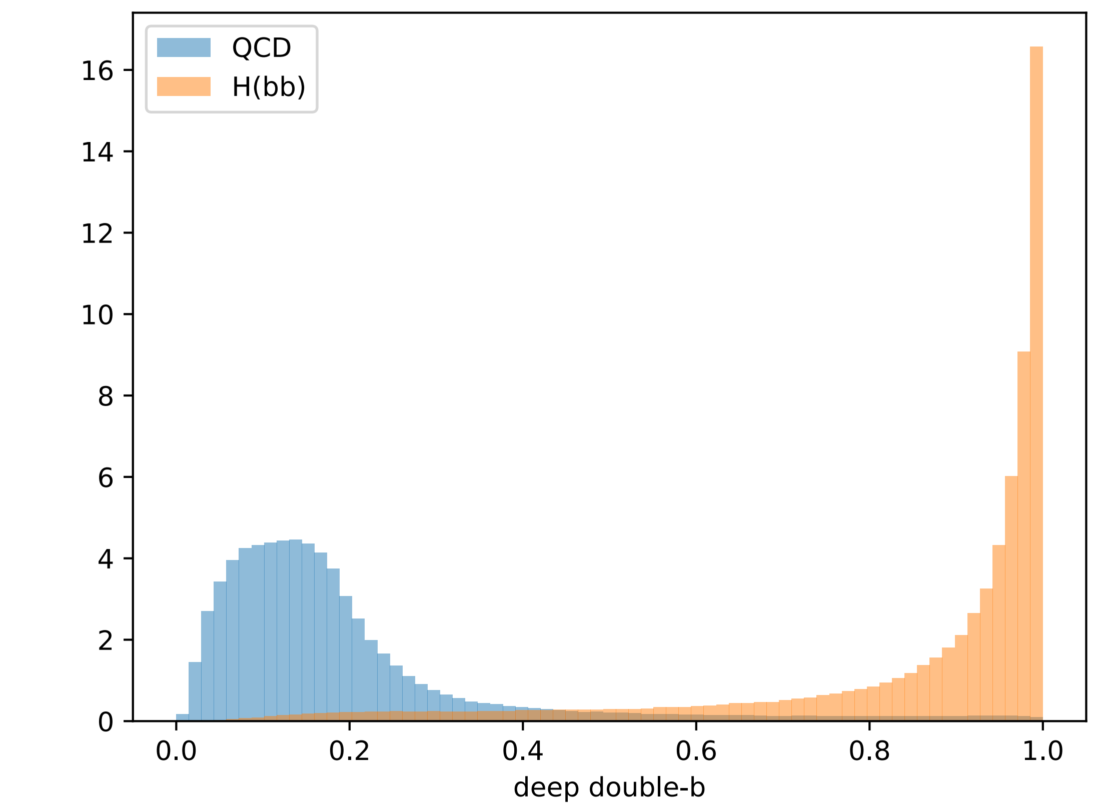


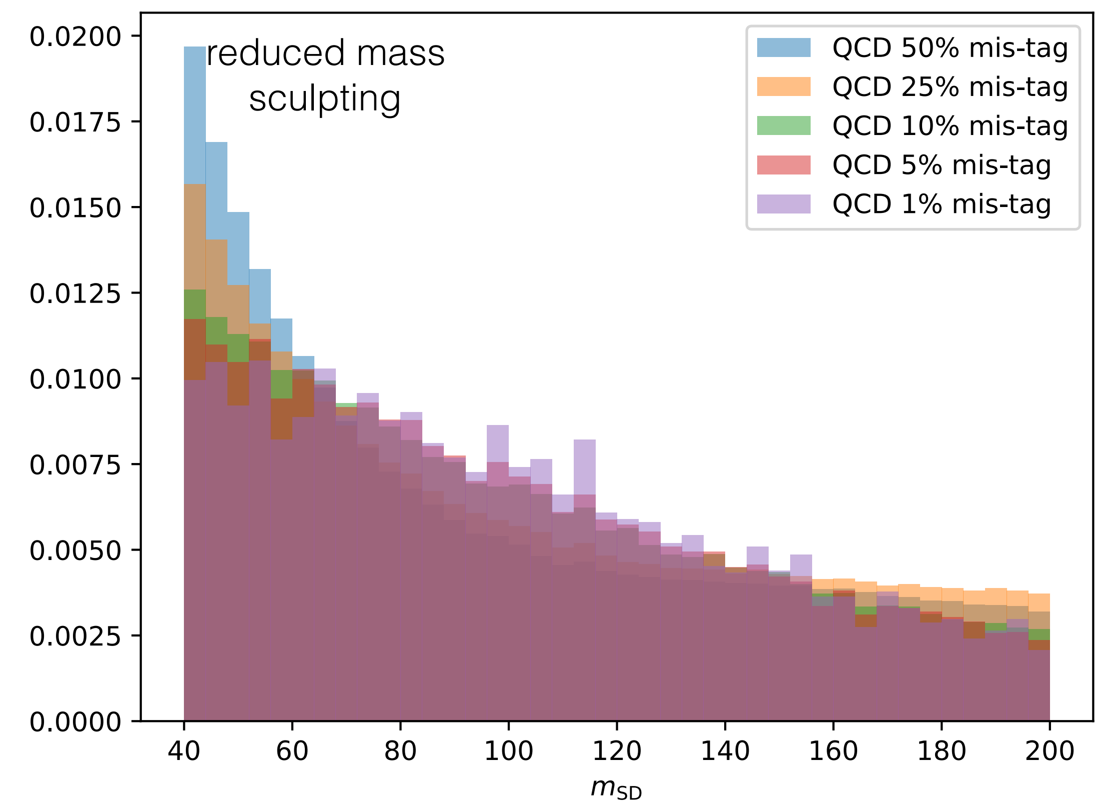


Performance with mass sculpting penalty (double-b + track + SV)



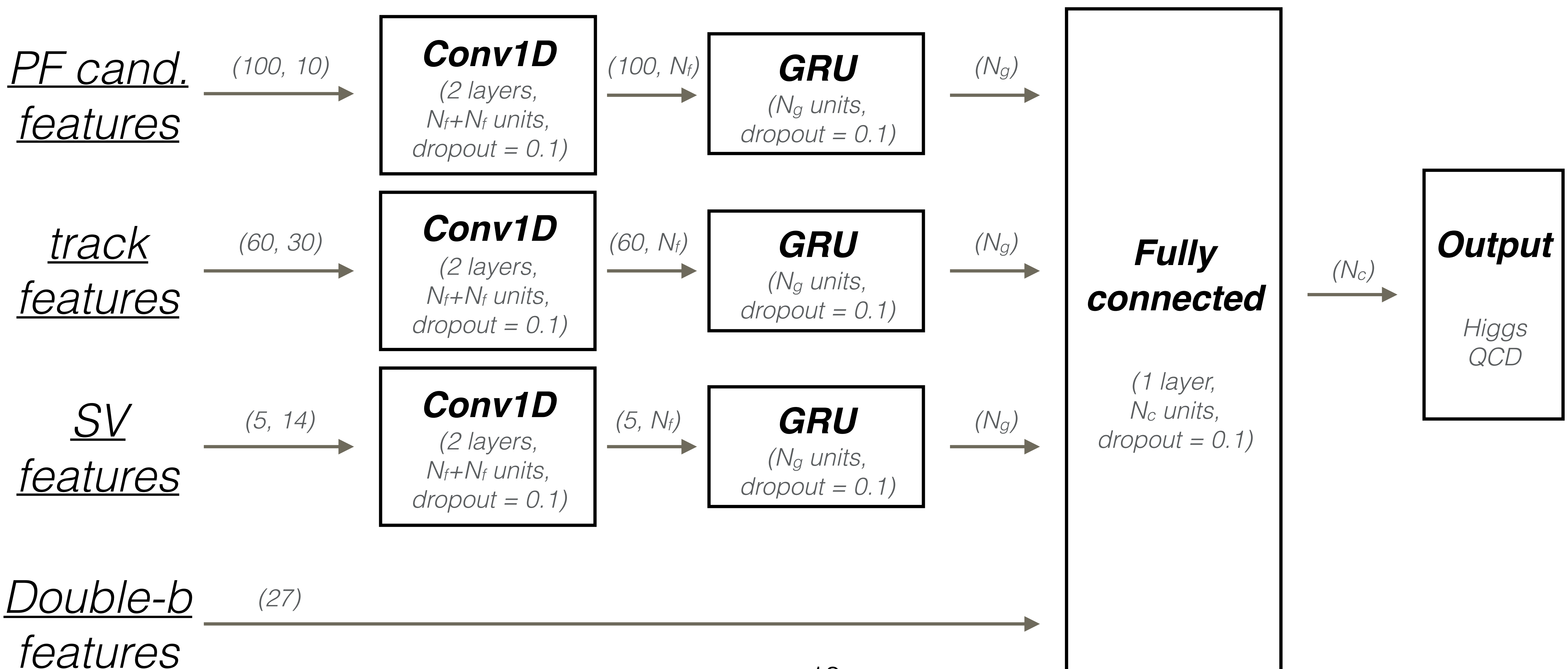






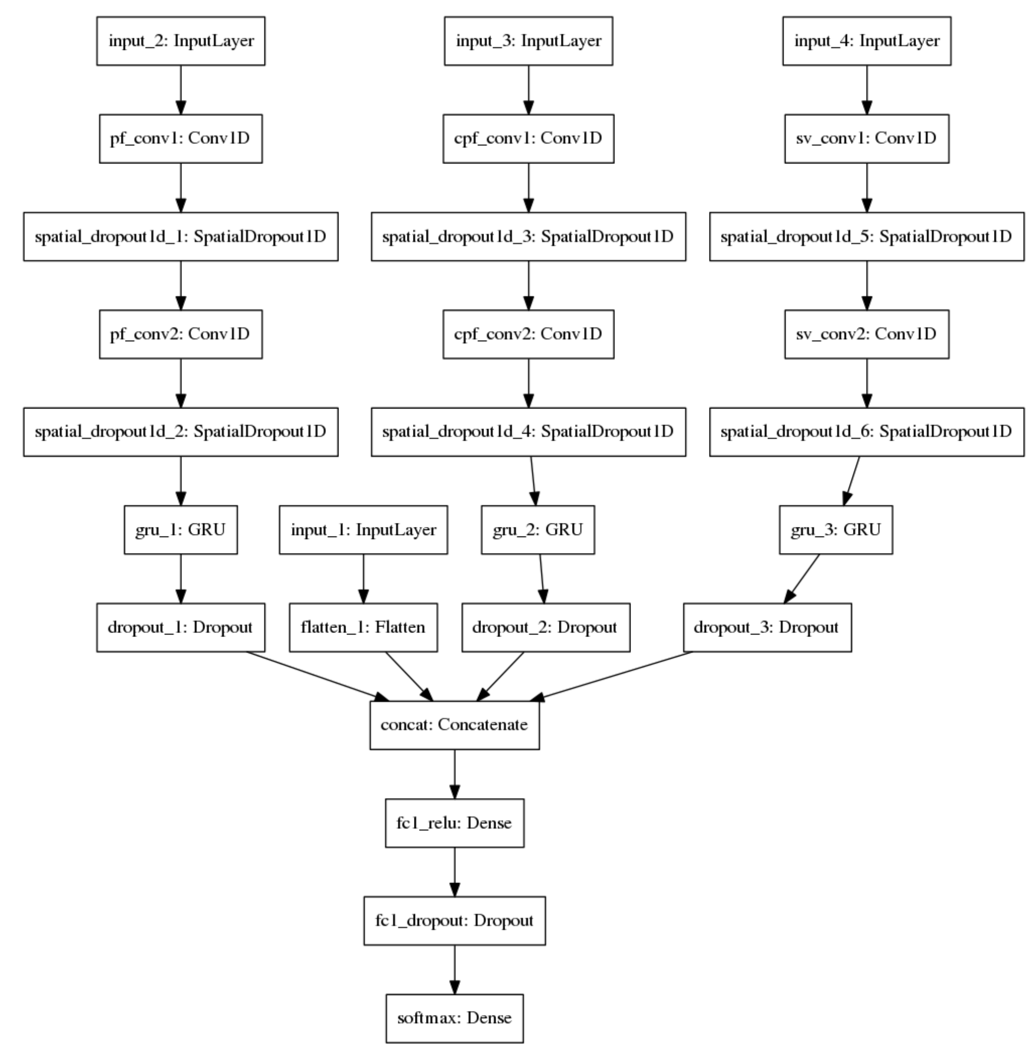
Conv1D + GRU network topology

- 27 high-level (double-b) features + 100×10 PF candidate features + 60×30 track features + 5×14 secondary vertex features per Higgs-candidate jet
- Conv1D with kernel size 1 = Time-distributed dense = apply same dense network to each PF candidate / track / SV
- GRU = Gated Recurrent Unit = Recurrent network to reduce dimensionality of output from Conv1D layers ($100 \times N_f$, $60 \times N_f$, $5 \times N_f$) → (N_g, N_g, N_g)



Reminder: p2 vs p3

- We only saw 20-40% improvement in training time when running on p3 vs p2
- Could be because our “deep” model is still too simple to stress system…
- Considering extending the model (more layers, different architecture, etc.)



Backup