HEAVY-FLAVOUR TAGGING FOR BOOSTED RESONANCES AND LARGE-CONE JETS IN CMS

ML4JETS 2018
FERMILAB
BATAVIA, IL, USA

NOVEMBER 15, 2018

Javier Duarte
Fermilab

OUTLINE

- Motivation: boosted Higgs and dark matter mediators
  - Double-b tagging
  - Background estimation methods
- Deep learning for large-cone jets
  - BEST
  - DeepAK8
  - Deep Double-b/c
- Summary and outlook

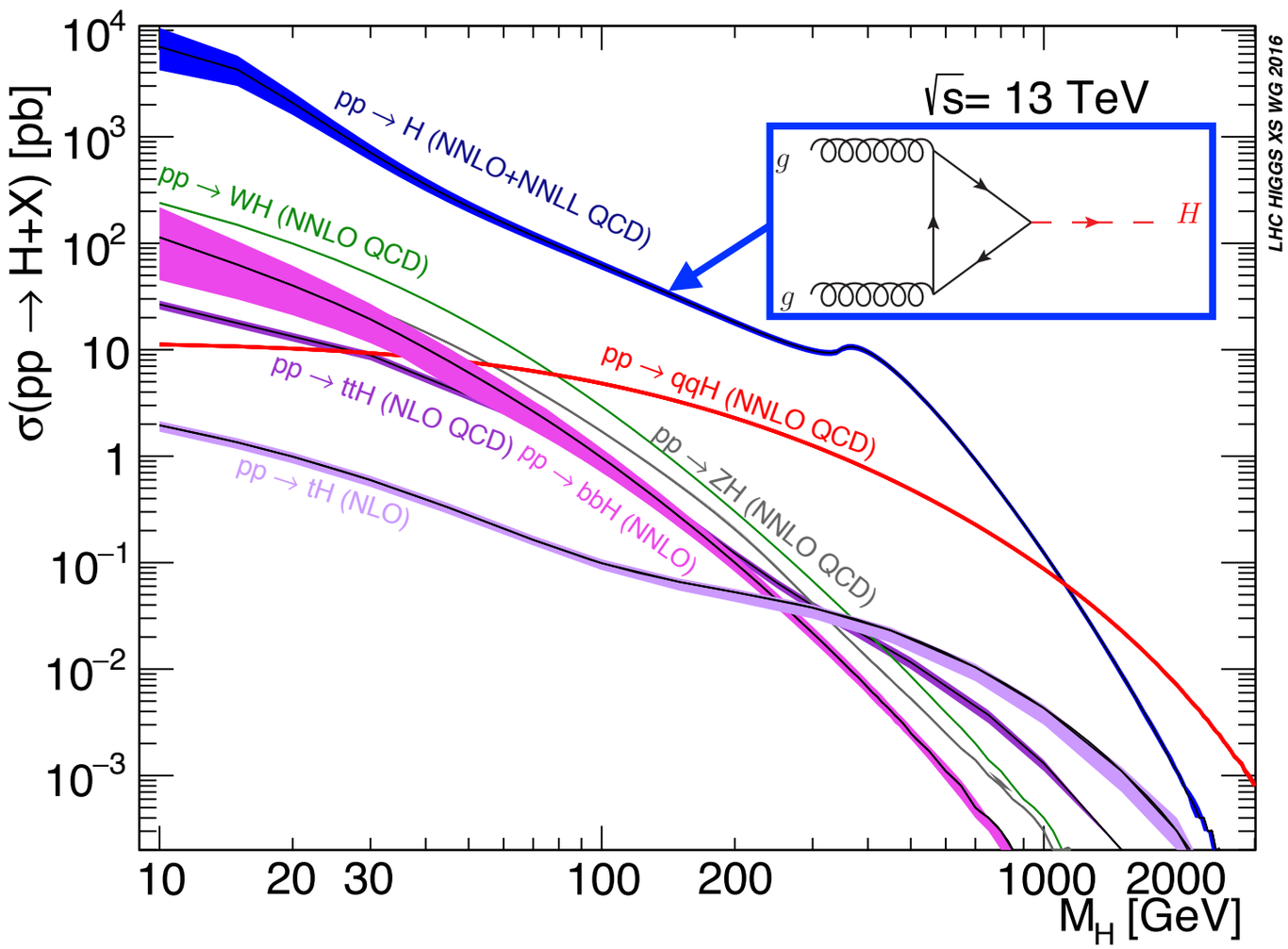
HEAVY FLAVOR TAGGING FOR BOOSTED RESONANCES

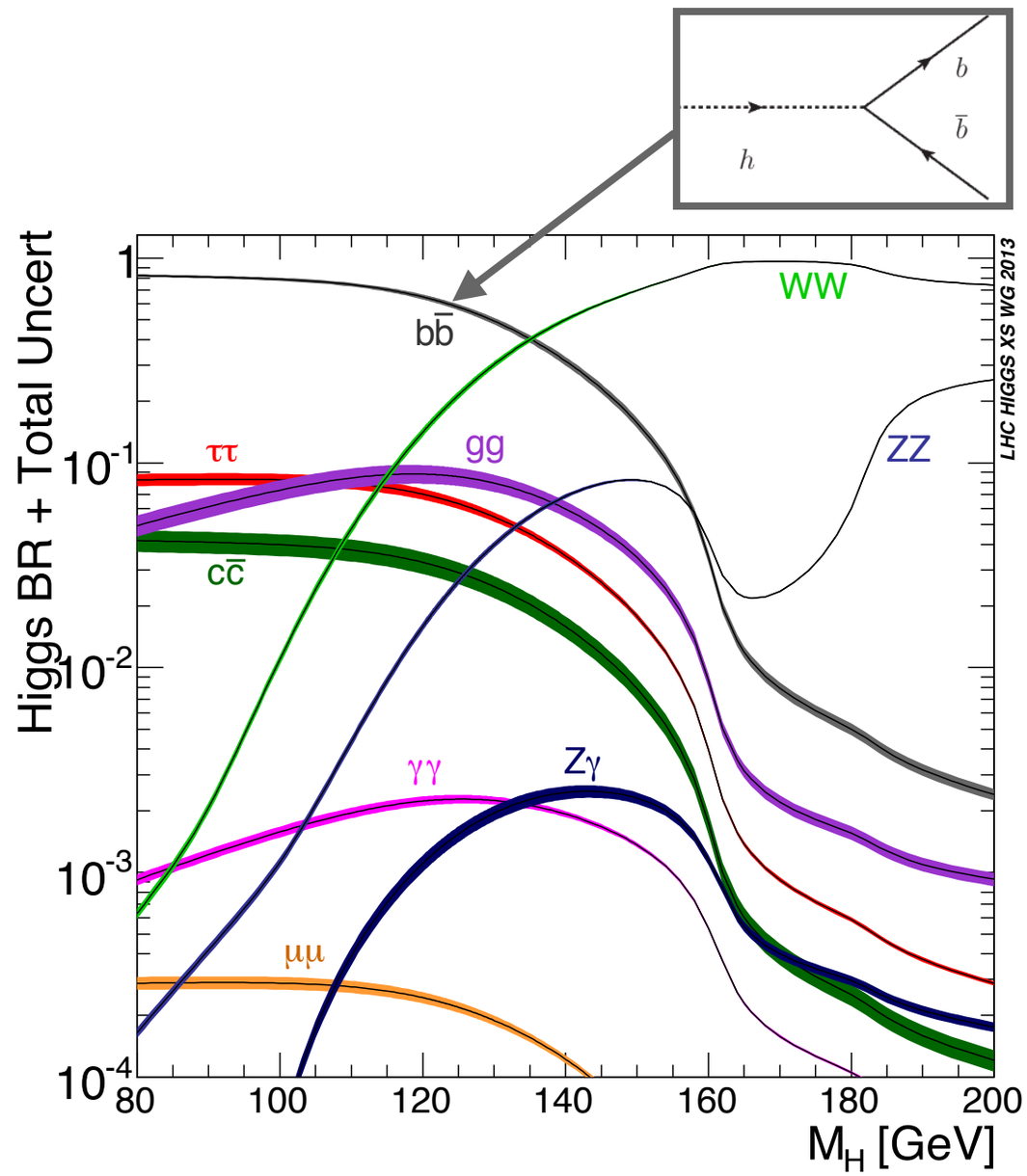
BOOSTED HIGGS AND SCALAR MEDIATORS

Javier Duarte
Fermilab

HIGGS AS A DIJET SEARCH?

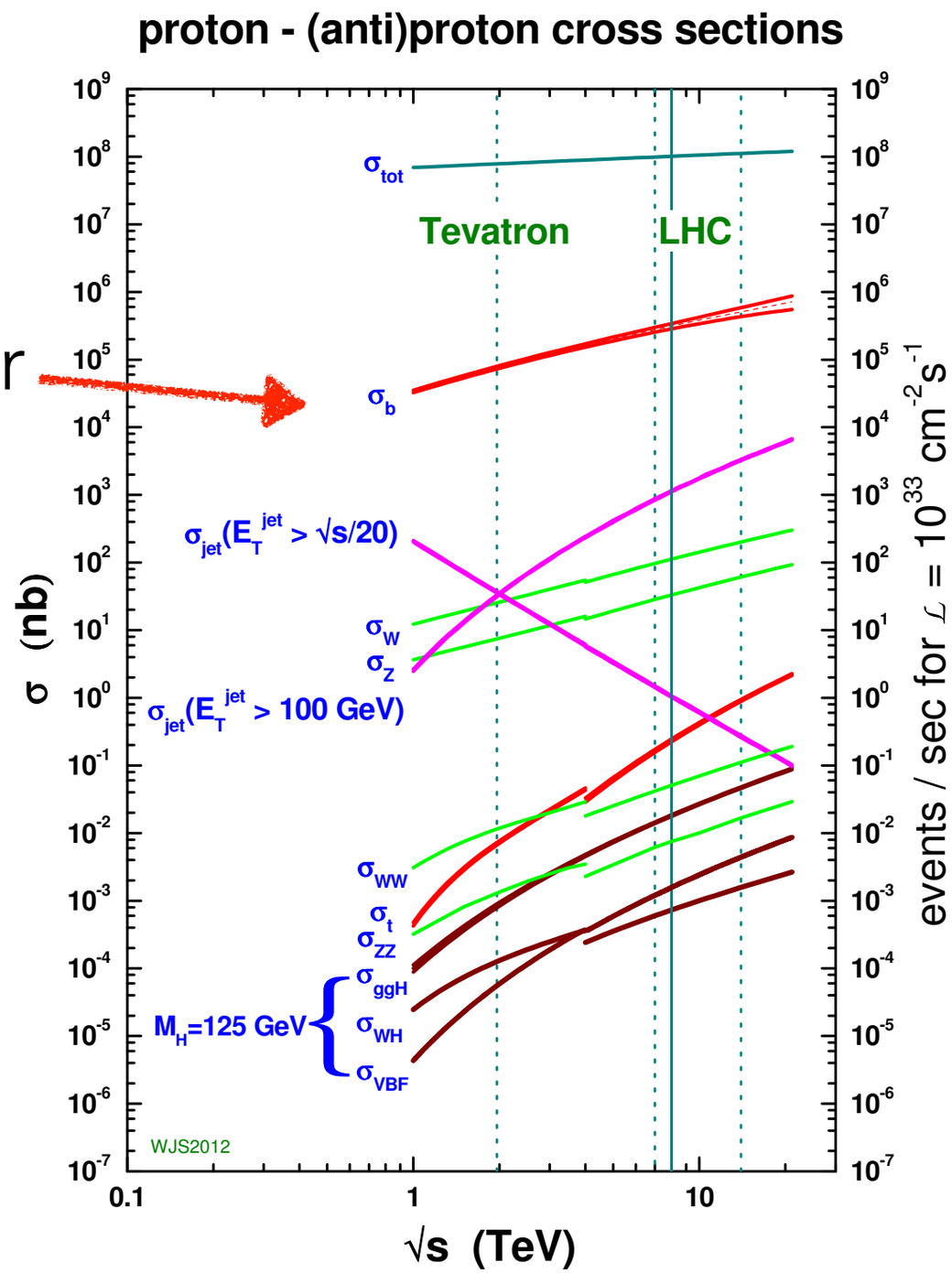
- Largest Higgs production/decay mode is gg → H → bb (>50%)





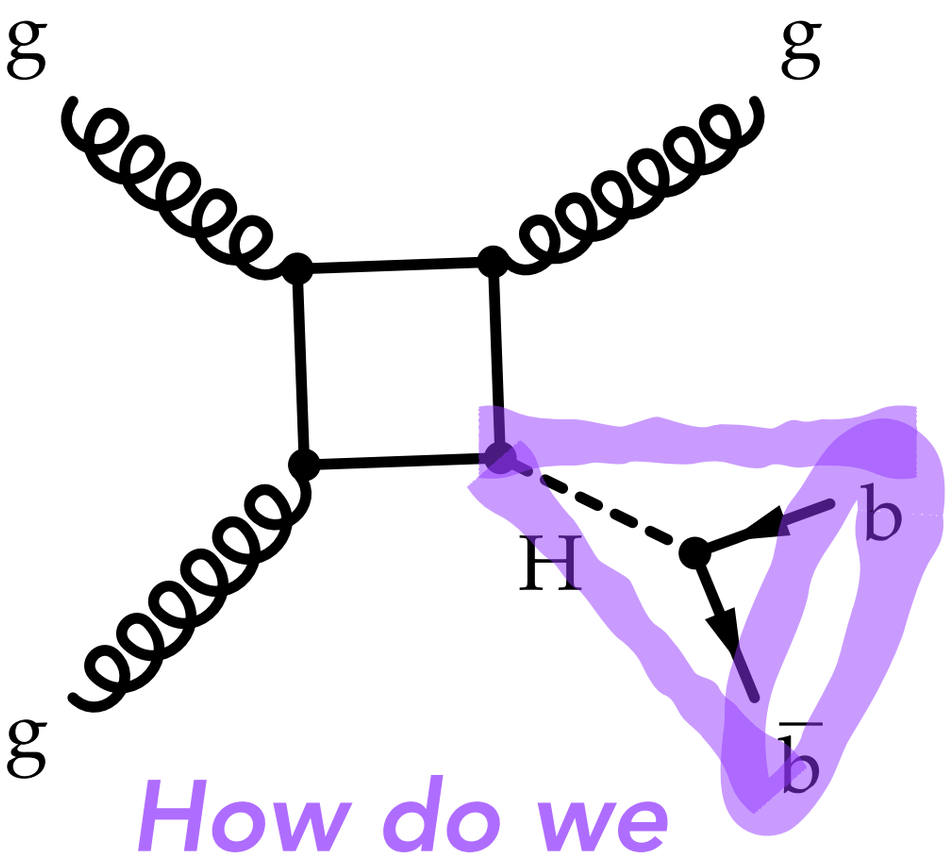
HIGGS AS A DIJET SEARCH?

- Largest Higgs production/decay mode is gg → H → bb (>50%)
- But, background is also immense
  - **QCD b production** is × 10^7 larger
- It seems hopeless…

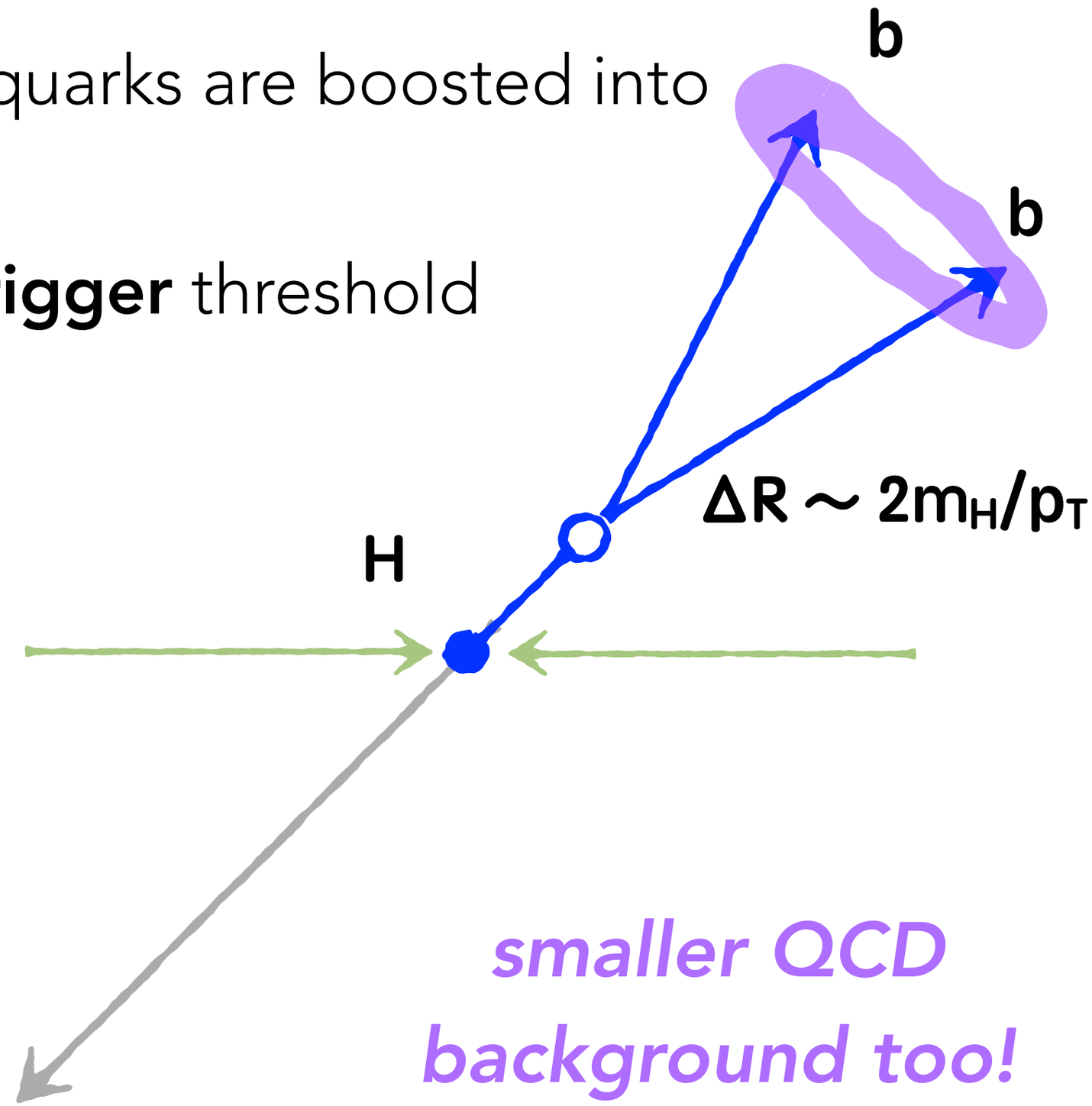


BOOSTED HIGGS TOPOLOGY

- At high p_T, the bottom quarks are boosted into a single large-radius jet
- ISR gets us above the **trigger** threshold

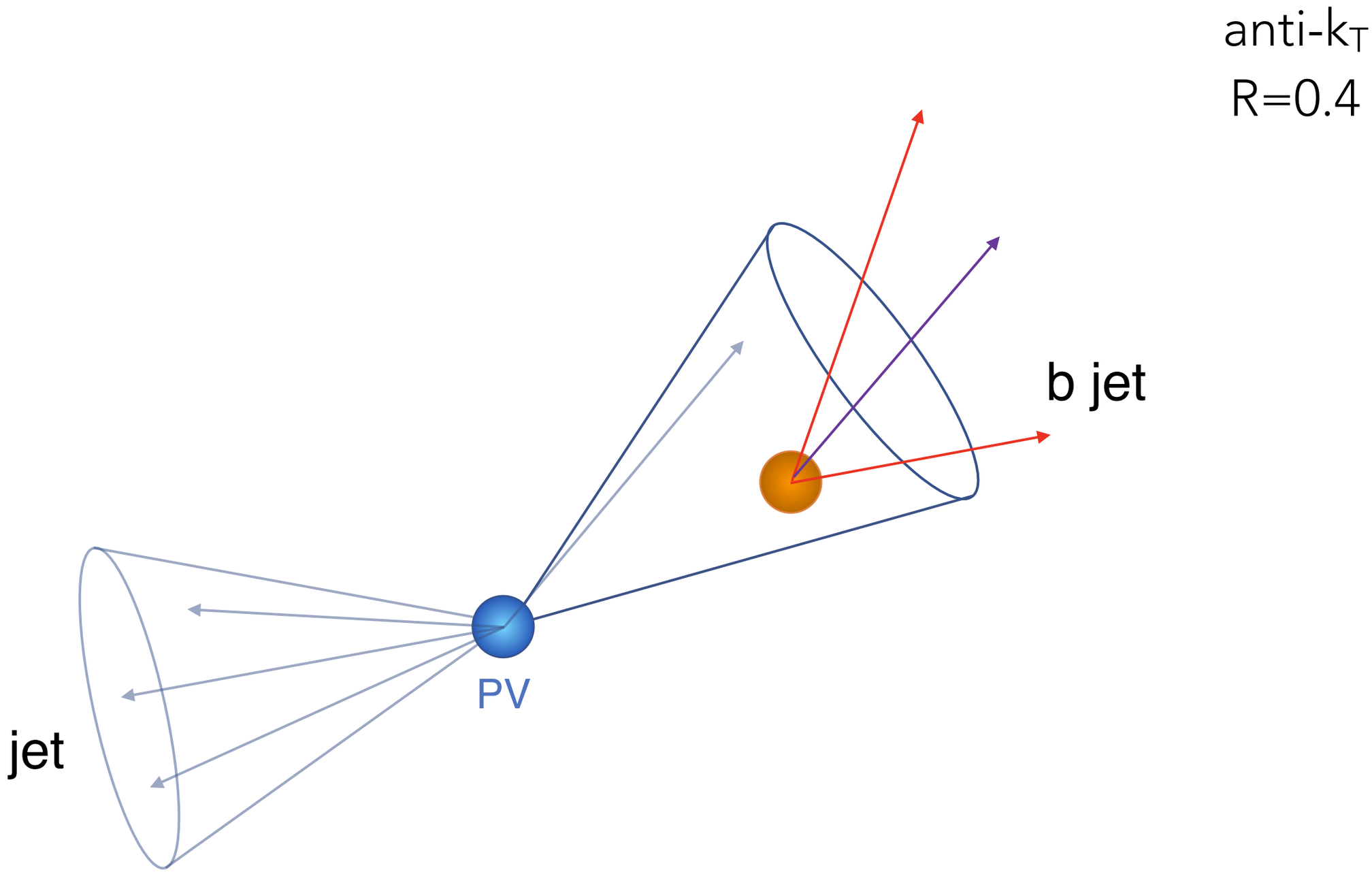


How do we tag these?



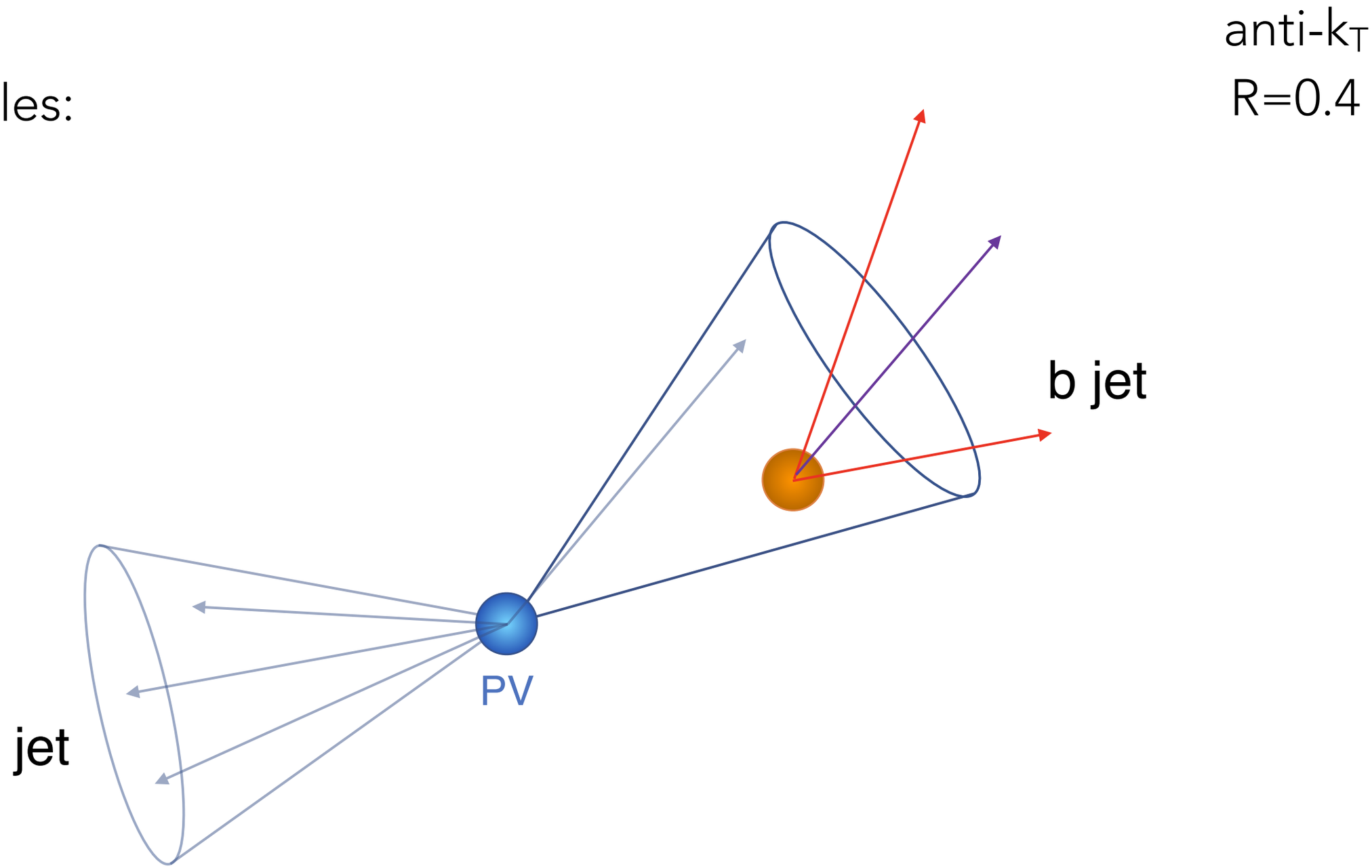
smaller QCD background too!

B TAGGING



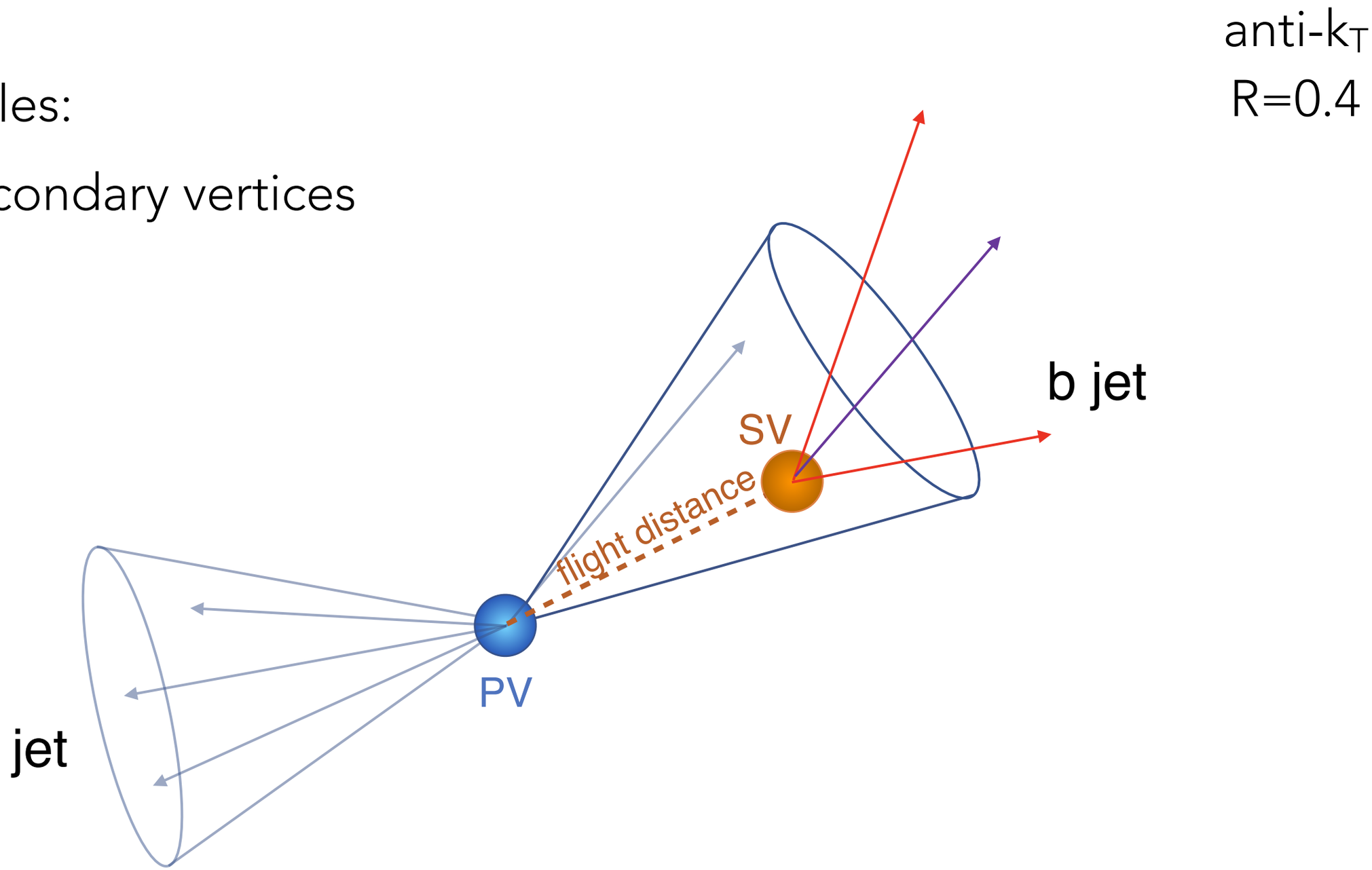
B TAGGING

- Handles:



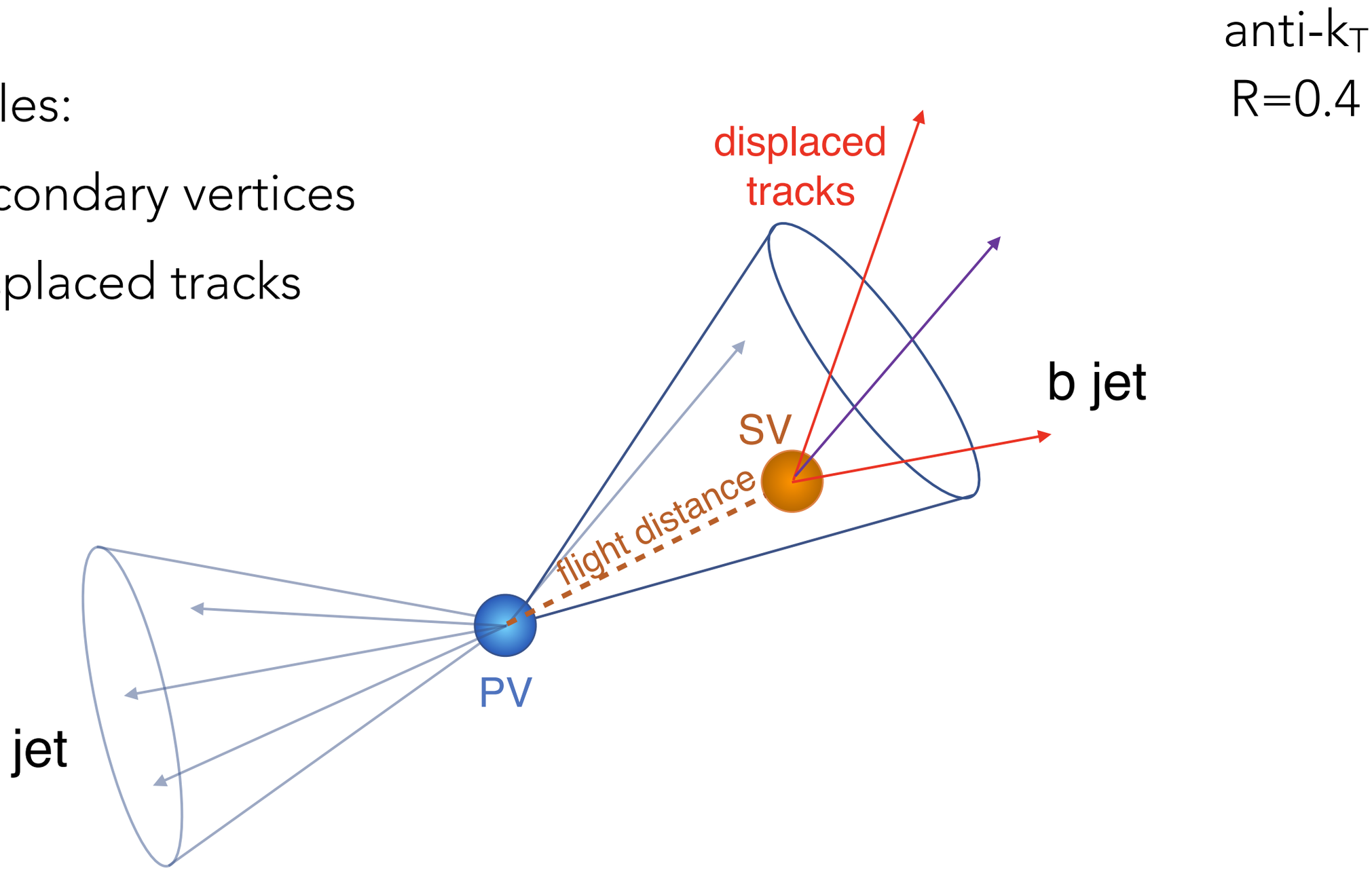
B TAGGING

- Handles:
  - secondary vertices



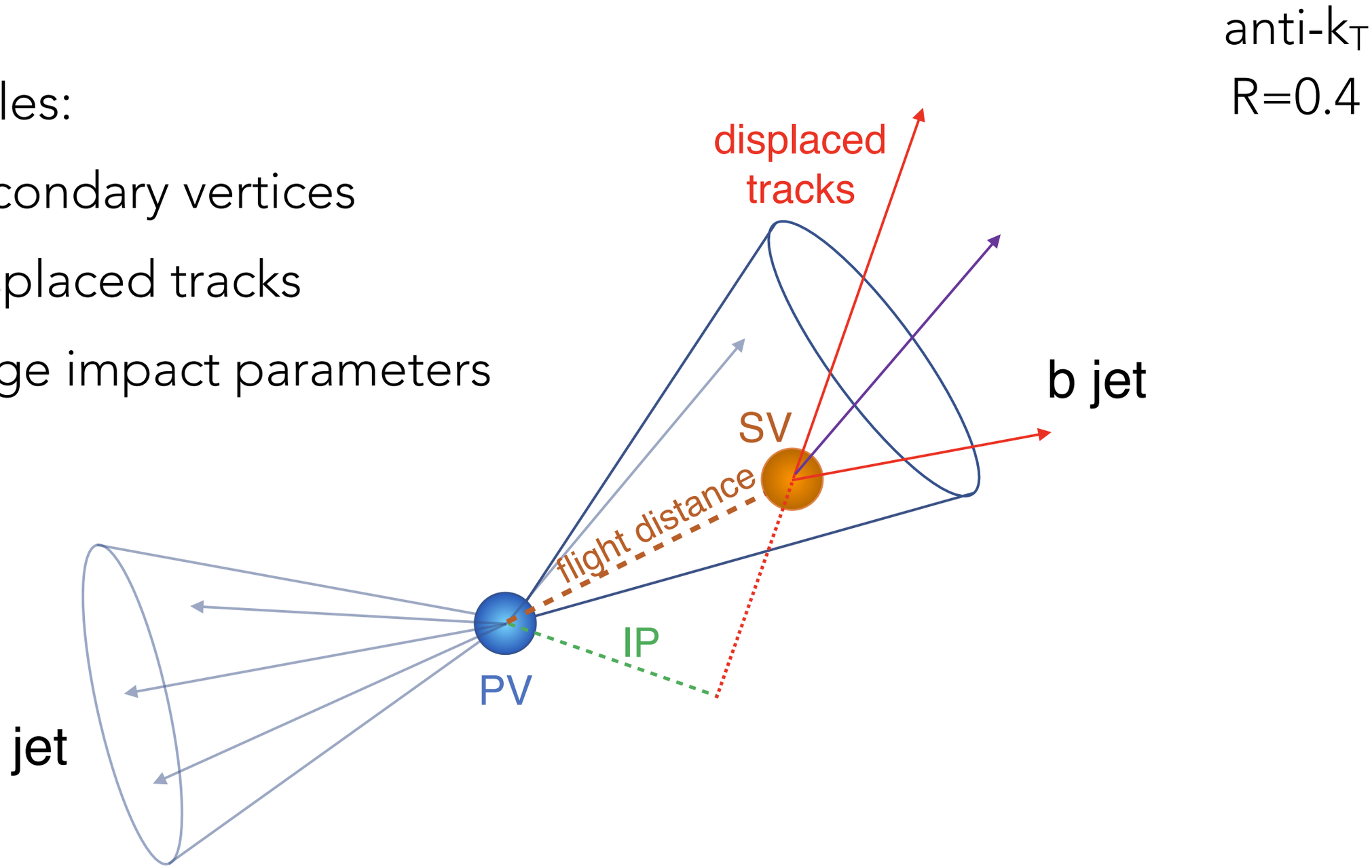
B TAGGING

- Handles:
  - secondary vertices
  - displaced tracks



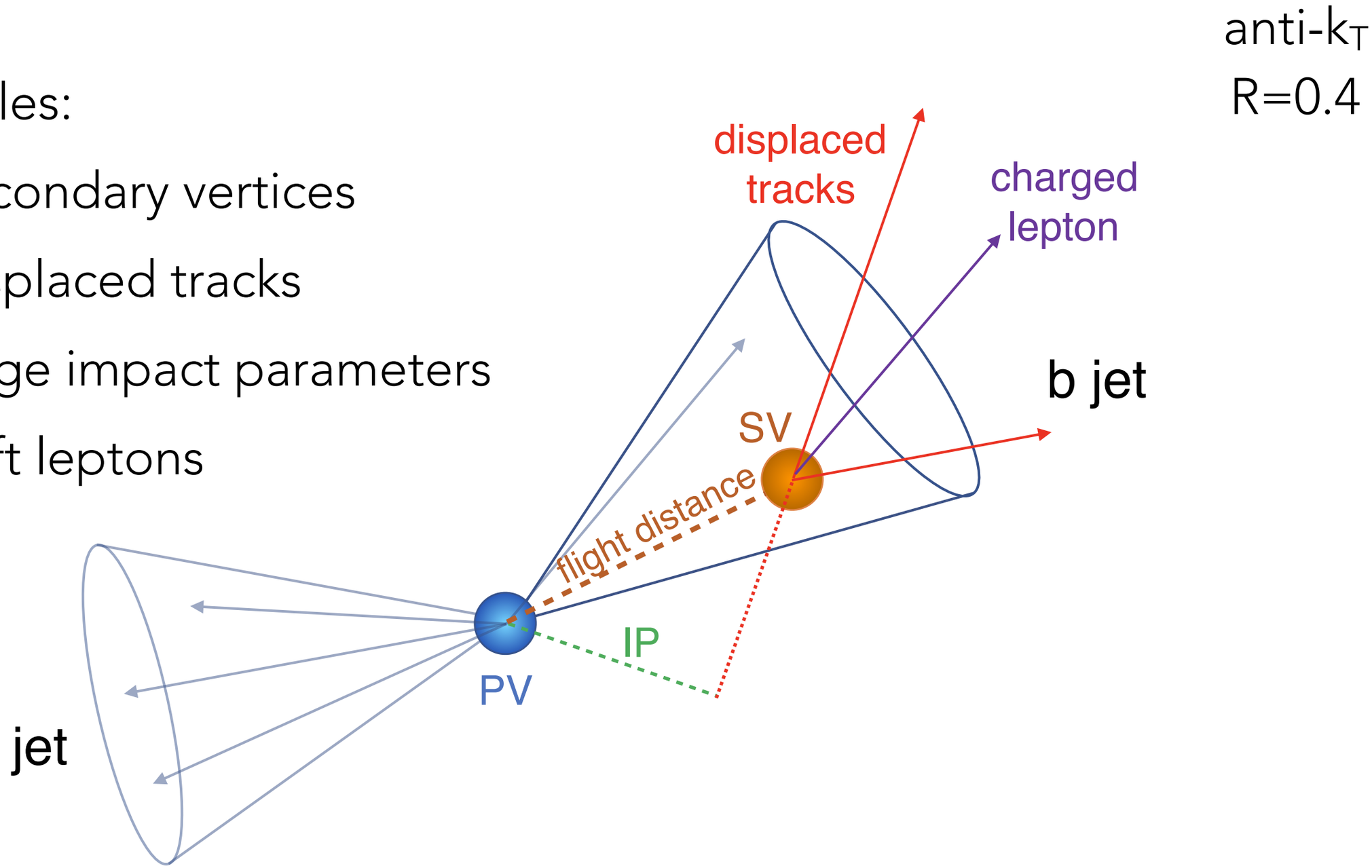
B TAGGING

- Handles:
  - secondary vertices
  - displaced tracks
  - large impact parameters



B TAGGING

- Handles:
  - secondary vertices
  - displaced tracks
  - large impact parameters
  - soft leptons

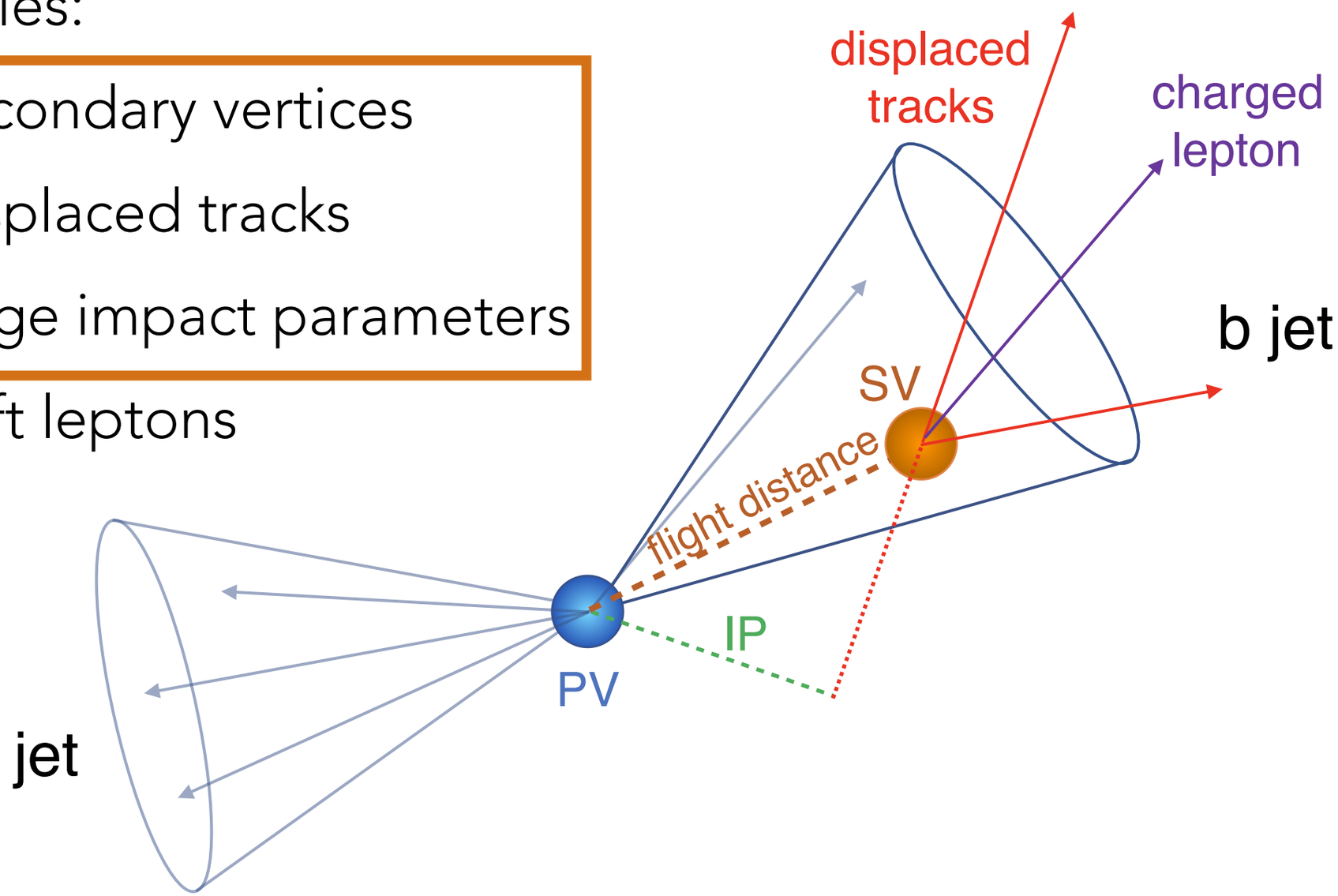


B TAGGING

anti-k_T
R=0.4

CSVv2

- Handles:
  - secondary vertices
  - displaced tracks
  - large impact parameters
  - soft leptons



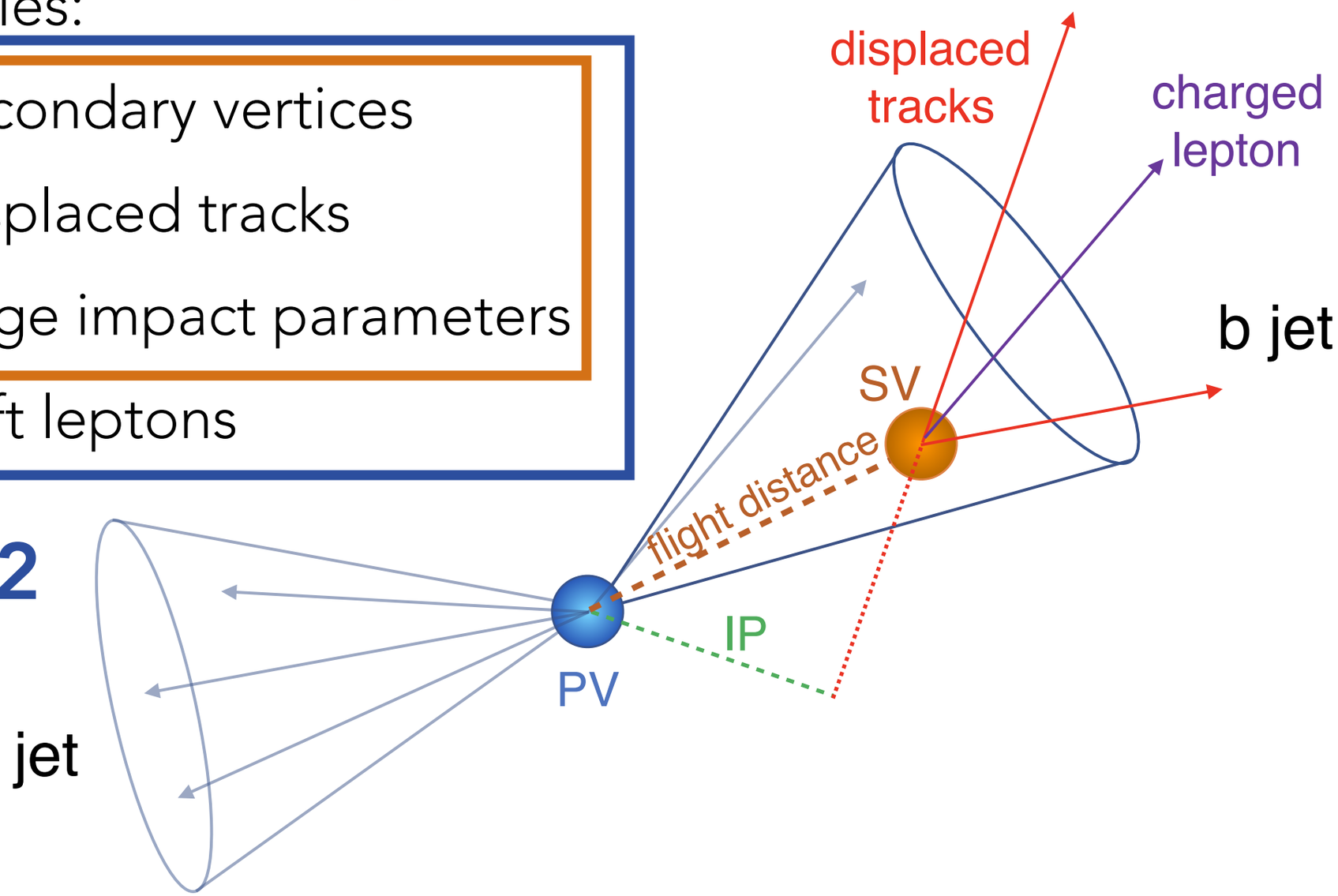
B TAGGING

anti-k_T
R=0.4

CSVv2

- Handles:
  - secondary vertices
  - displaced tracks
  - large impact parameters
  - soft leptons

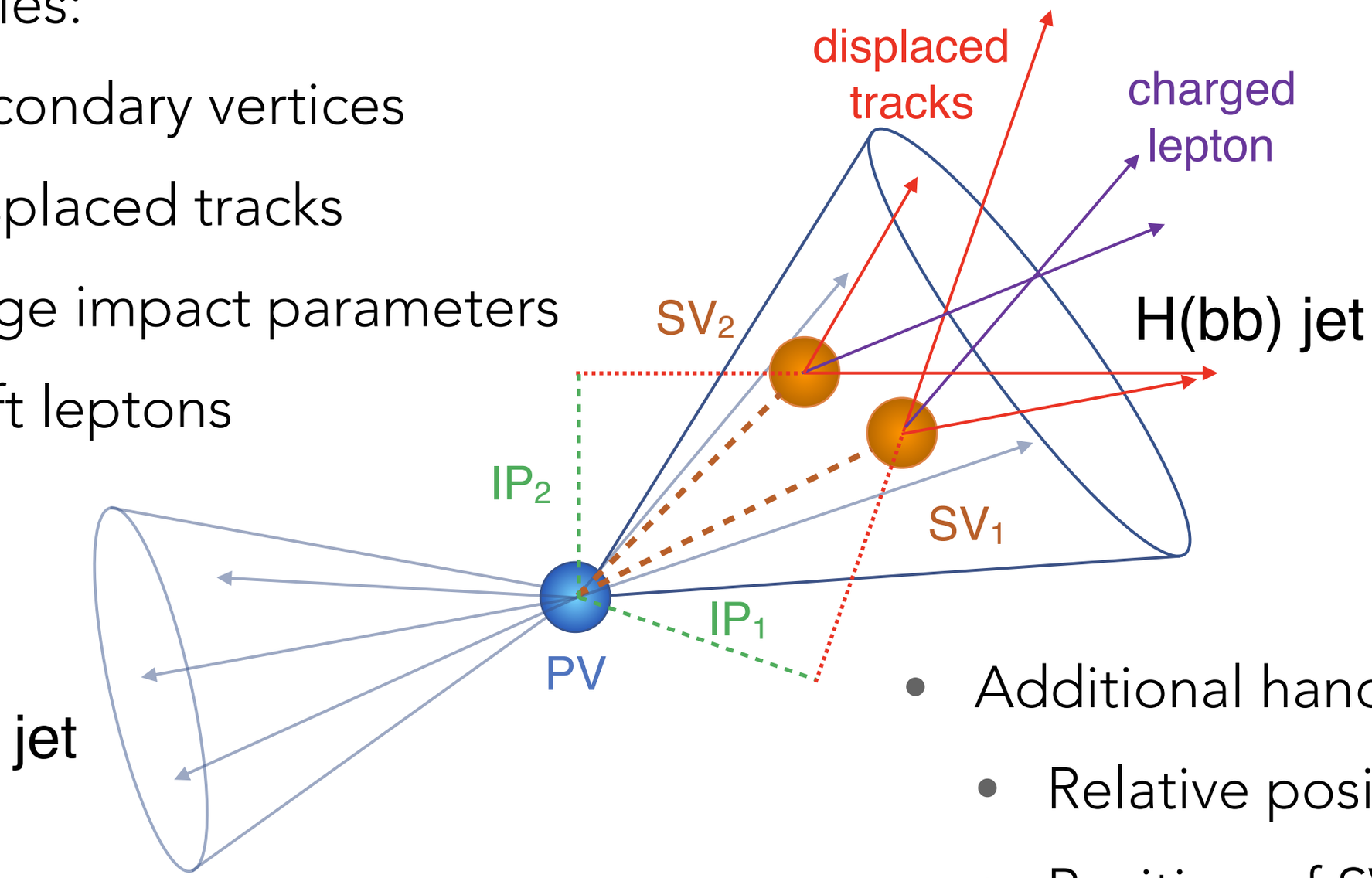
cMVAv2



DOUBLE-B TAGGING

anti-k_T
R=0.8

- Handles:
  - secondary vertices
  - displaced tracks
  - large impact parameters
  - soft leptons



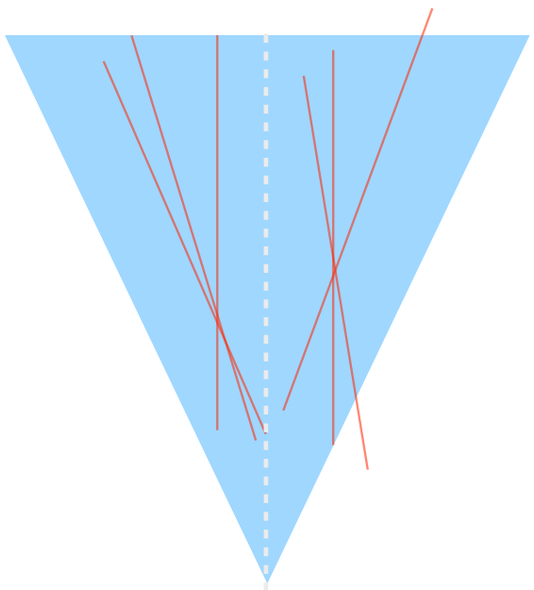
- Additional handles:
  - Relative position of SVs
  - Position of SVs relative to n-subjettiness axes

TAGGING APPROACHES

JINST 13 (2018) P05011

TAGGING APPROACHES

JINST 13 (2018) P05011

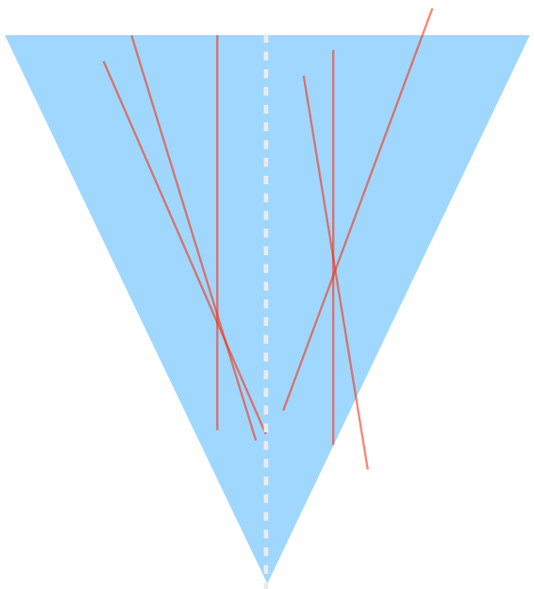


AK8 jet

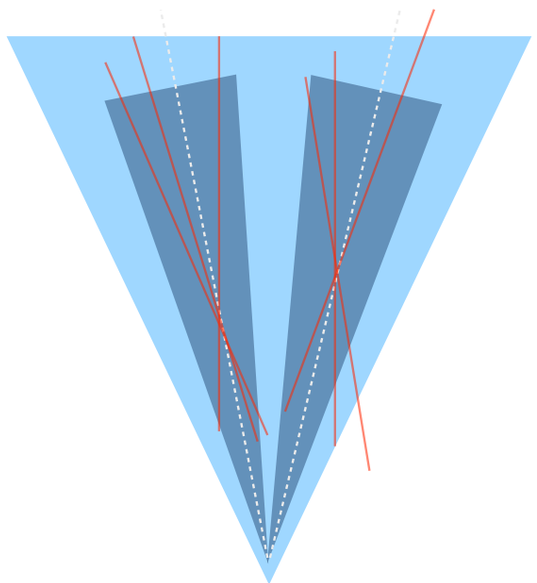
- Apply CSVv2 to AK8 jet (with looser requirements for the track-to-jet and vertex-to-jet association)

TAGGING APPROACHES

JINST 13 (2018) P05011



AK8 jet

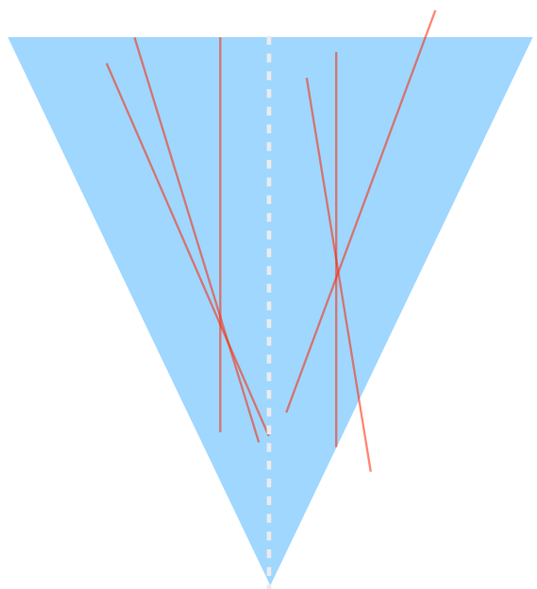


subjets

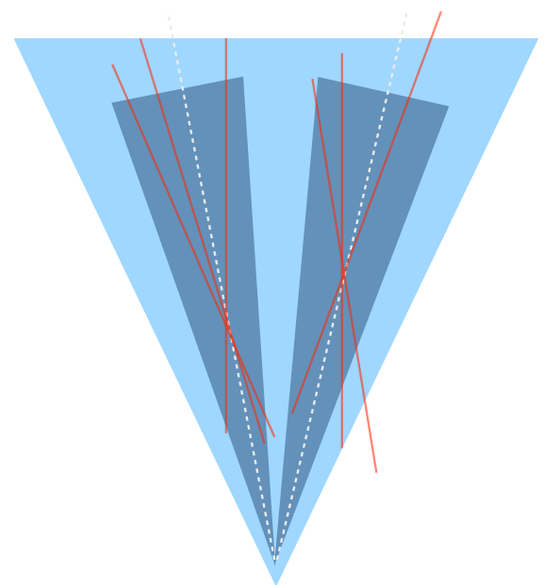
- Apply CSVv2 to AK8 jet (with looser requirements for the track-to-jet and vertex-to-jet association)
- Apply CSVv2 to AK4 subjets

TAGGING APPROACHES

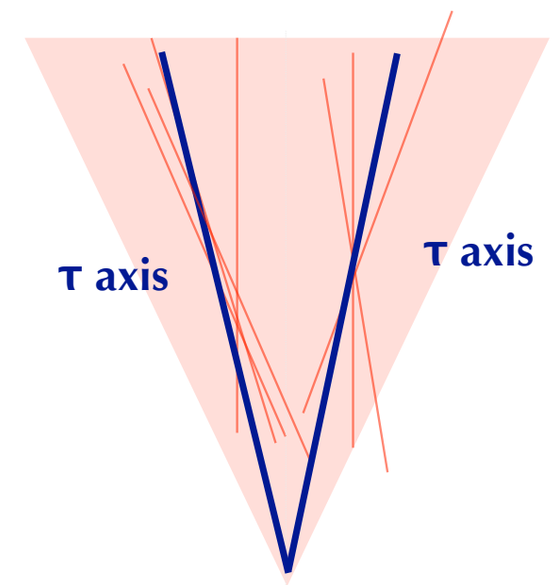
JINST 13 (2018) P05011



AK8 jet



subjets

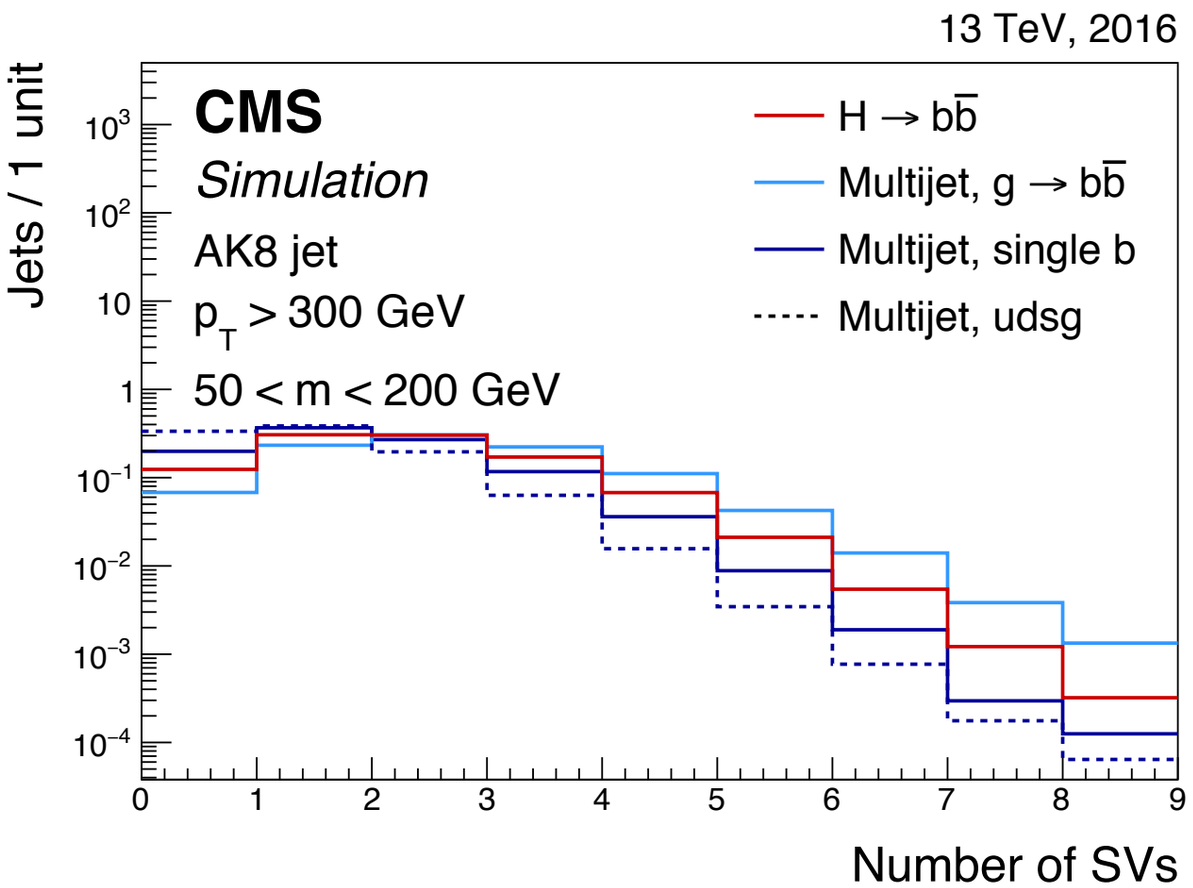


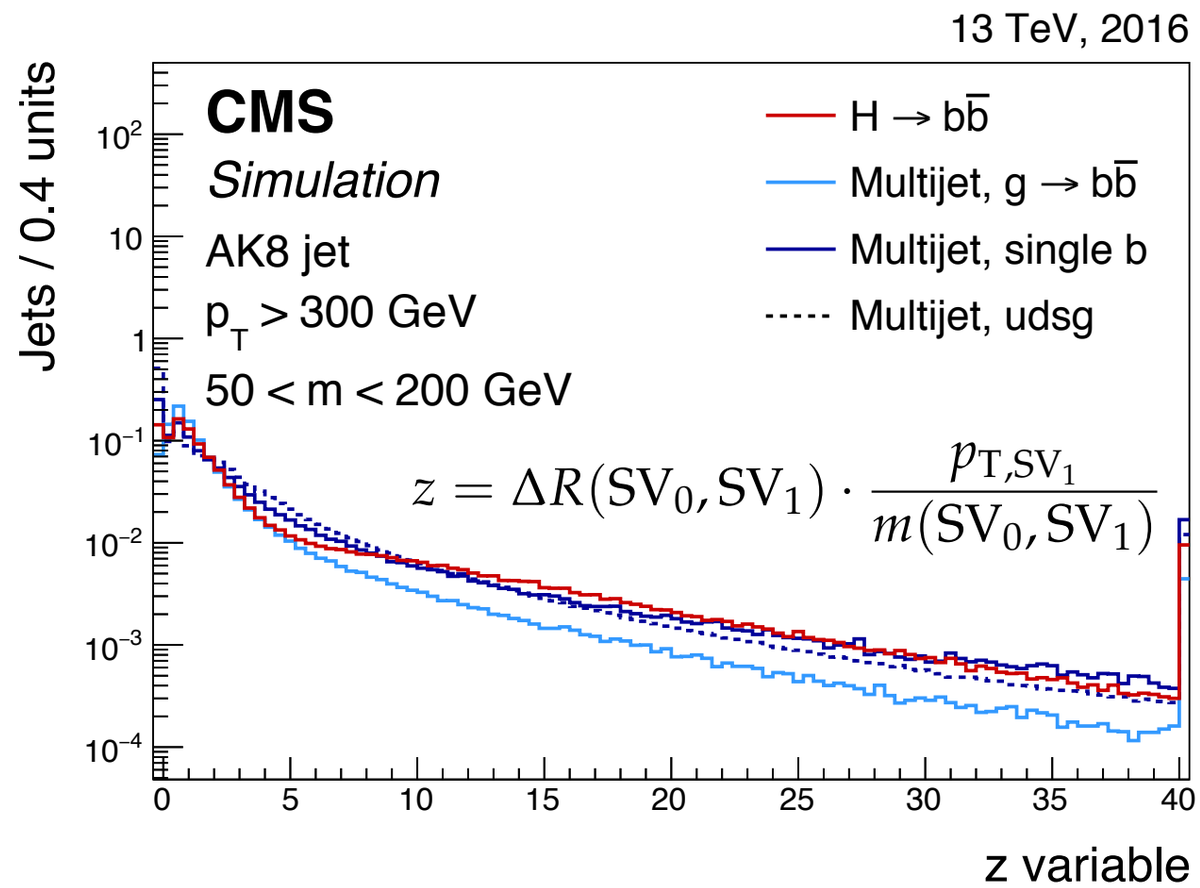
double-b

- Apply CSVv2 to AK8 jet (with looser requirements for the track-to-jet and vertex-to-jet association)
- Apply CSVv2 to AK4 subjets
- Train a new MVA with CSVv2 inputs + additional inputs (related to relative positions of SVs, displaced tracks, and n-subjettiness axes)

DOUBLE-B TAGGER

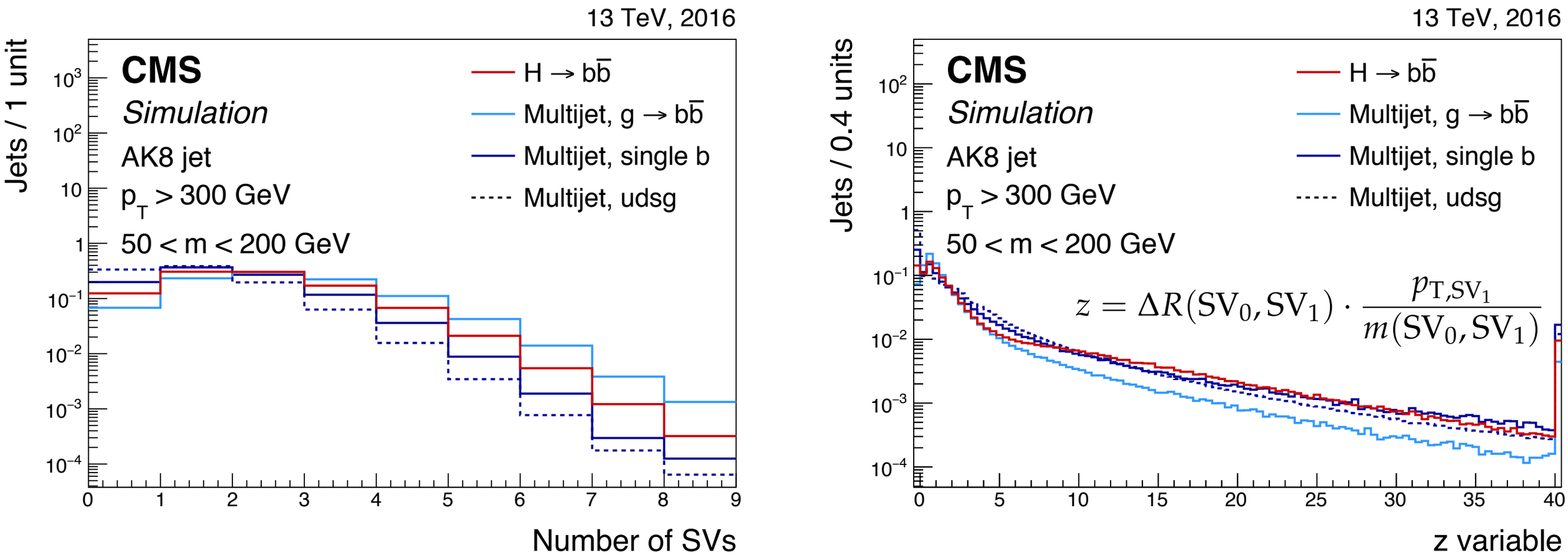
JINST 13 (2018) P05011





DOUBLE-B TAGGER

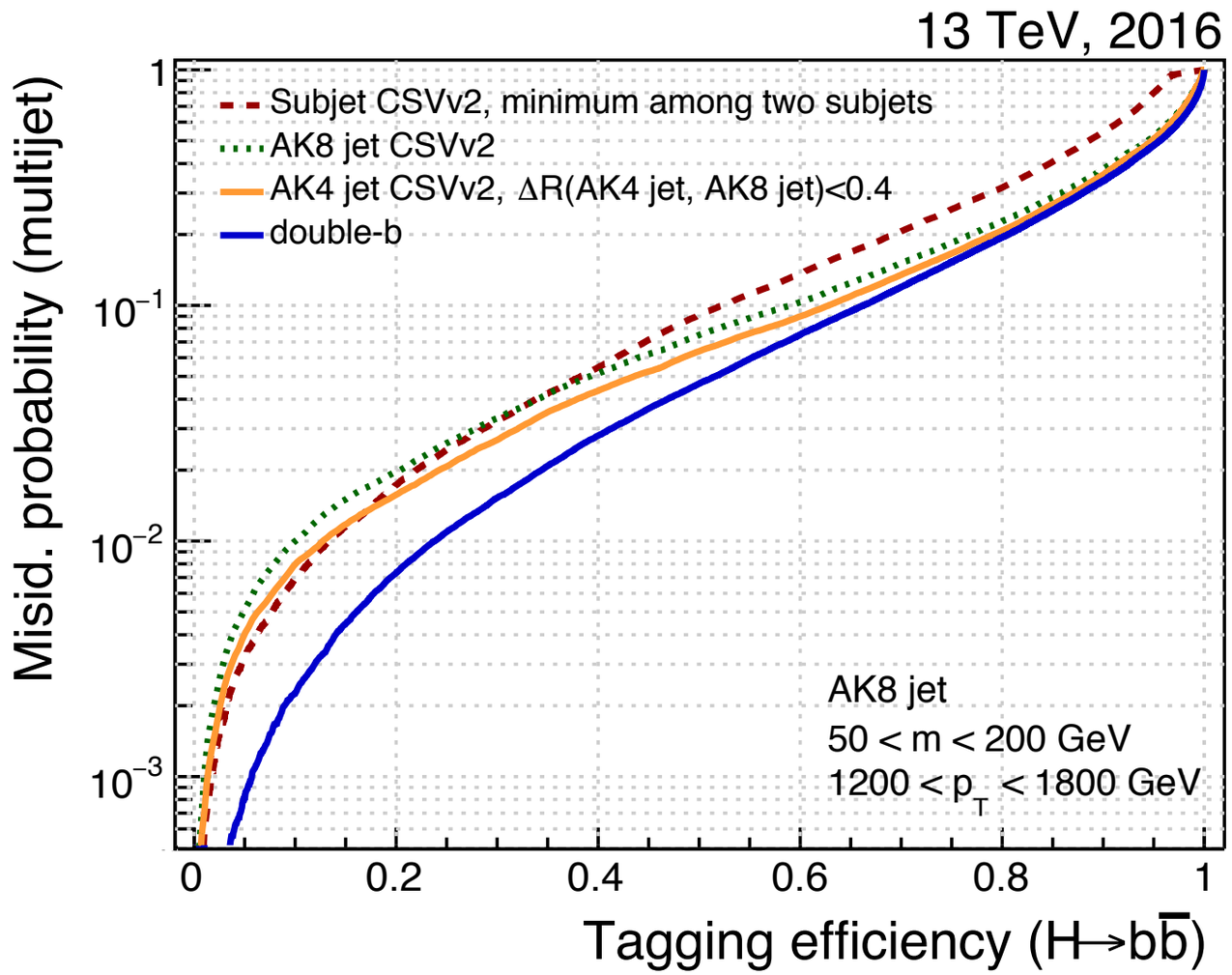
JINST 13 (2018) P05011

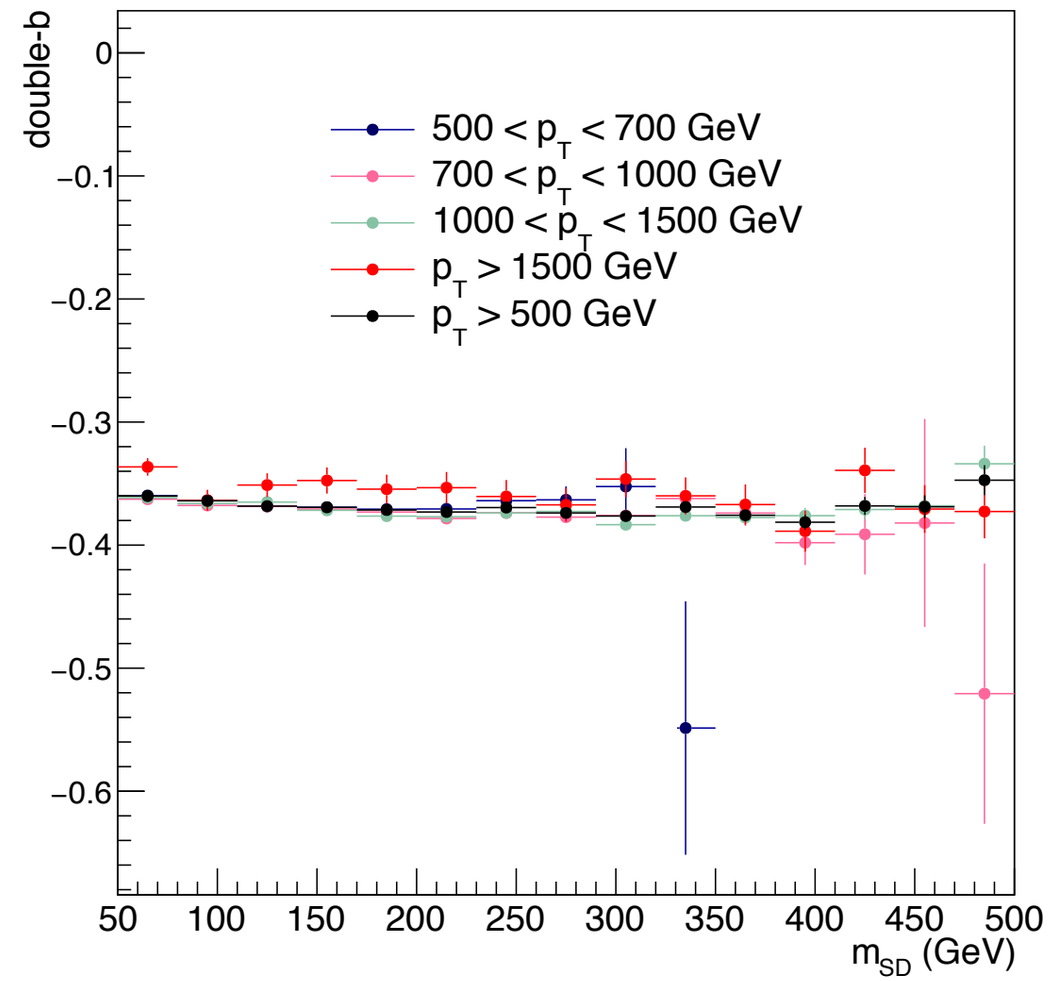


- Combines tracking and vertexing information in a boosted decision tree with 27 input observables

DOUBLE-B TAGGER

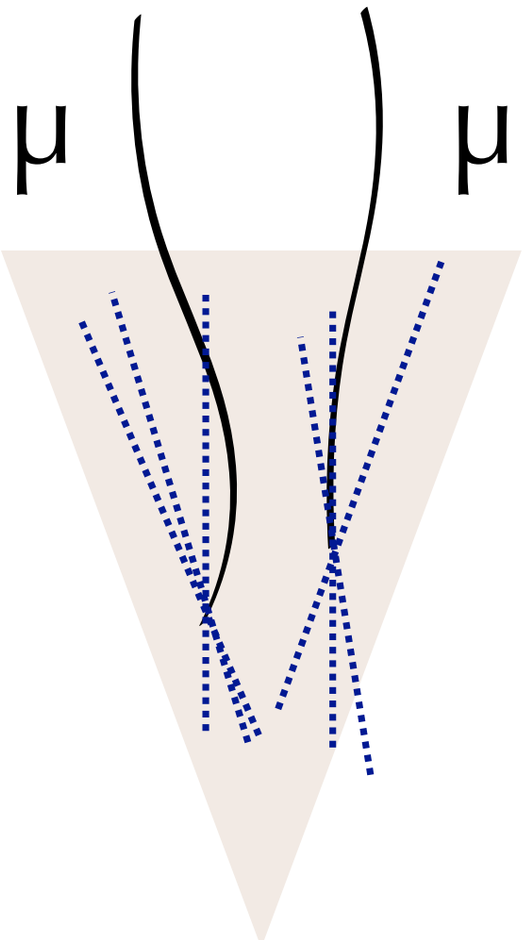
JINST 13 (2018) P05011

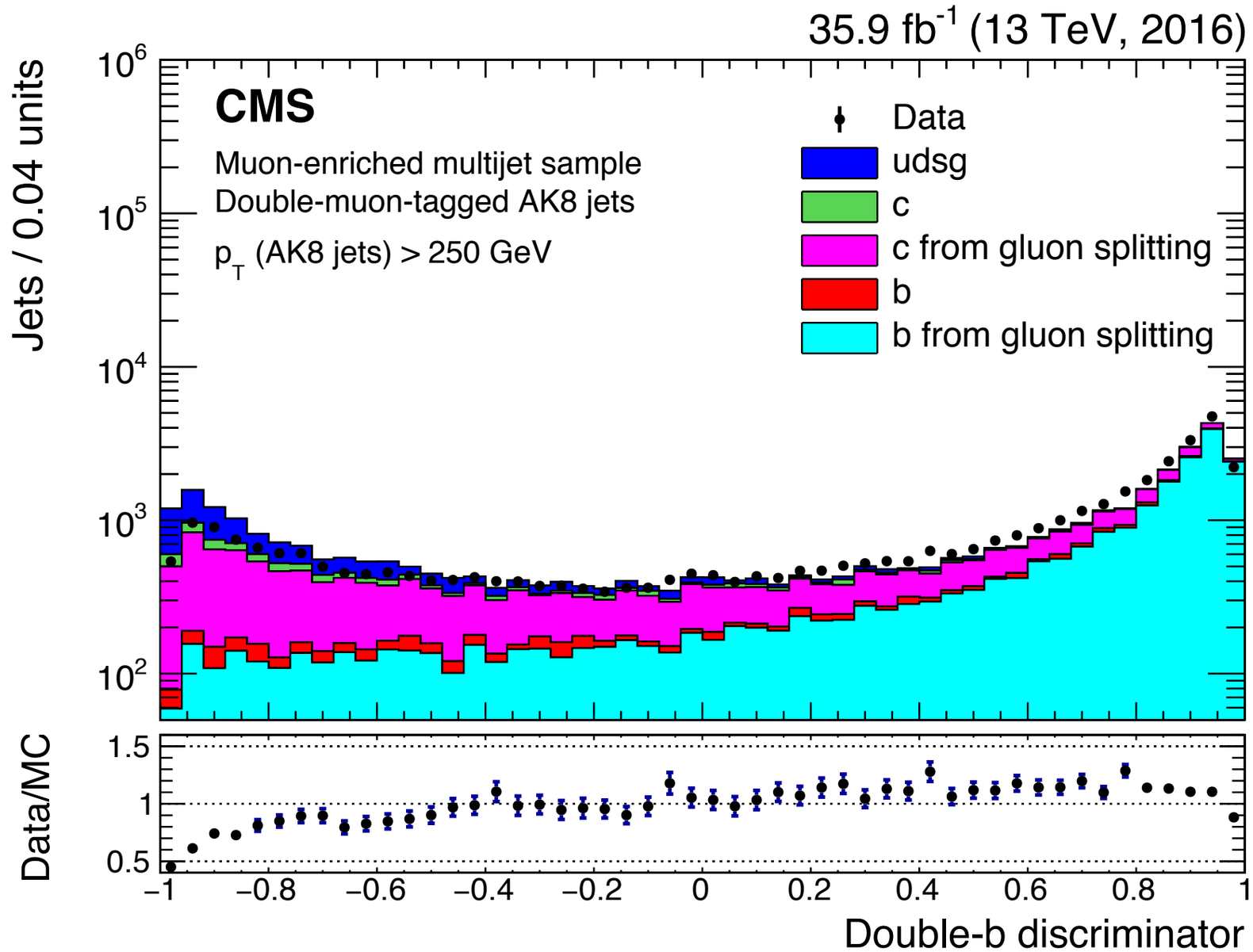




- Combines tracking and vertexing information in a boosted decision tree with 27 input observables
- No strong correlations in double-b tagger versus m_{SD} or p_T in QCD background

EFFICIENCY IN DATA

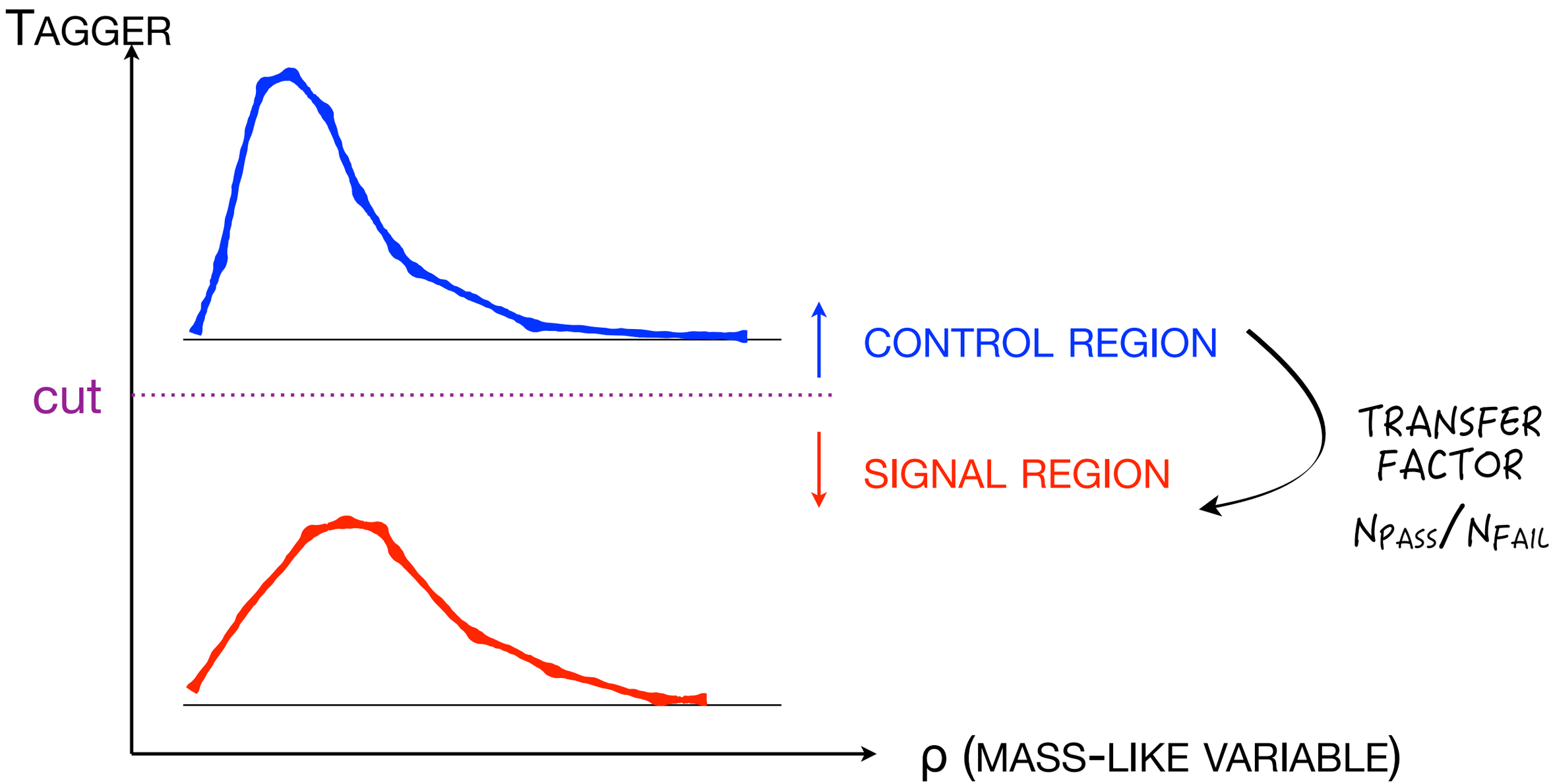




- Using g→bb jets as a proxy in double muon tagged jet sample
- Associated data/MC uncertainty 3-5%

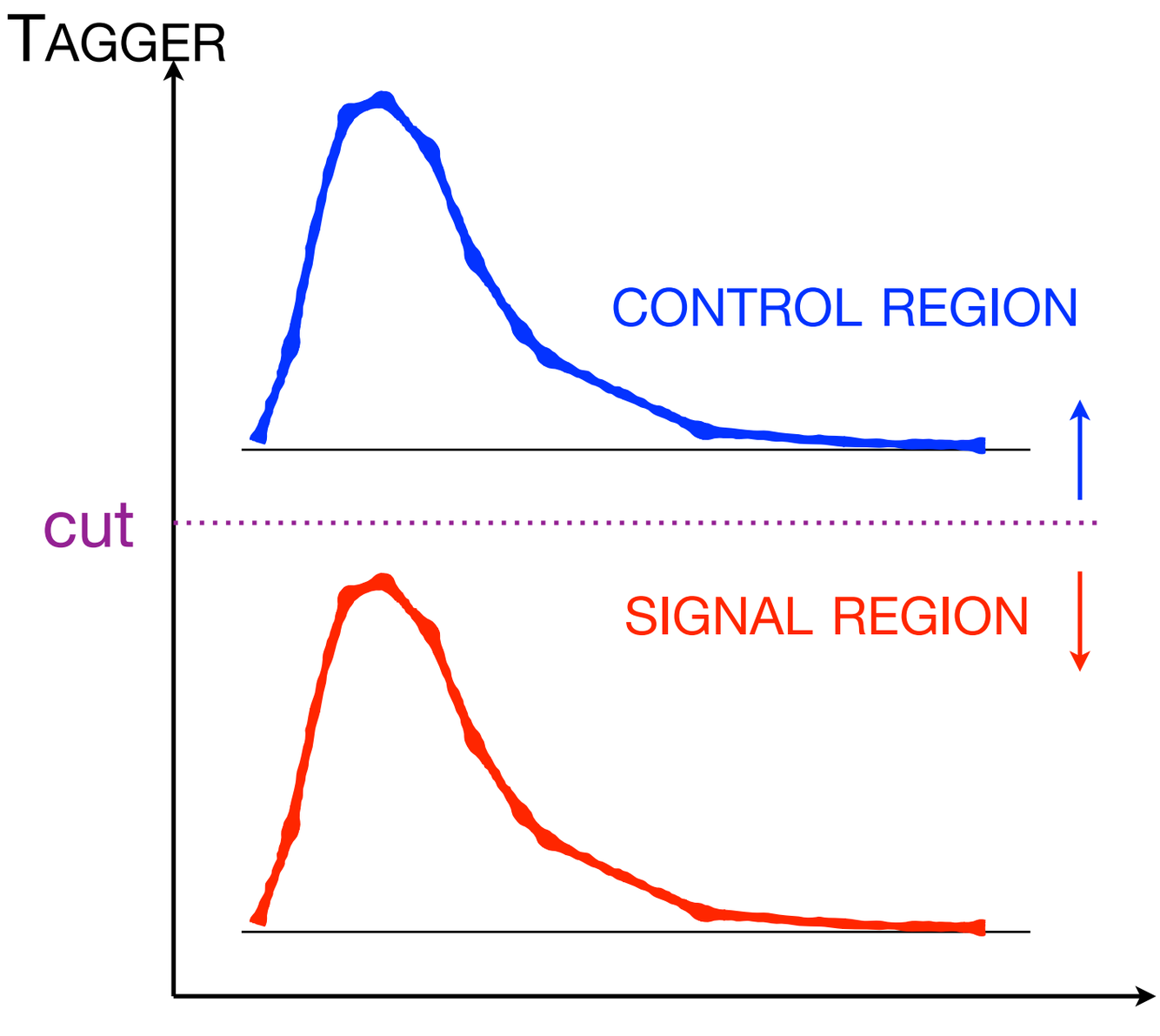
SIDEBAND QCD PREDICTION

- Core idea: predict QCD jet mass distribution from jets that fail your tagger
- Problem: what if cut on tagger sculpts jet mass distribution!?



SIDEBAND QCD PREDICTION

- Solution: ensure tagger is variable is decorrelated from jet mass and p_T
- Different ways to accomplish this…

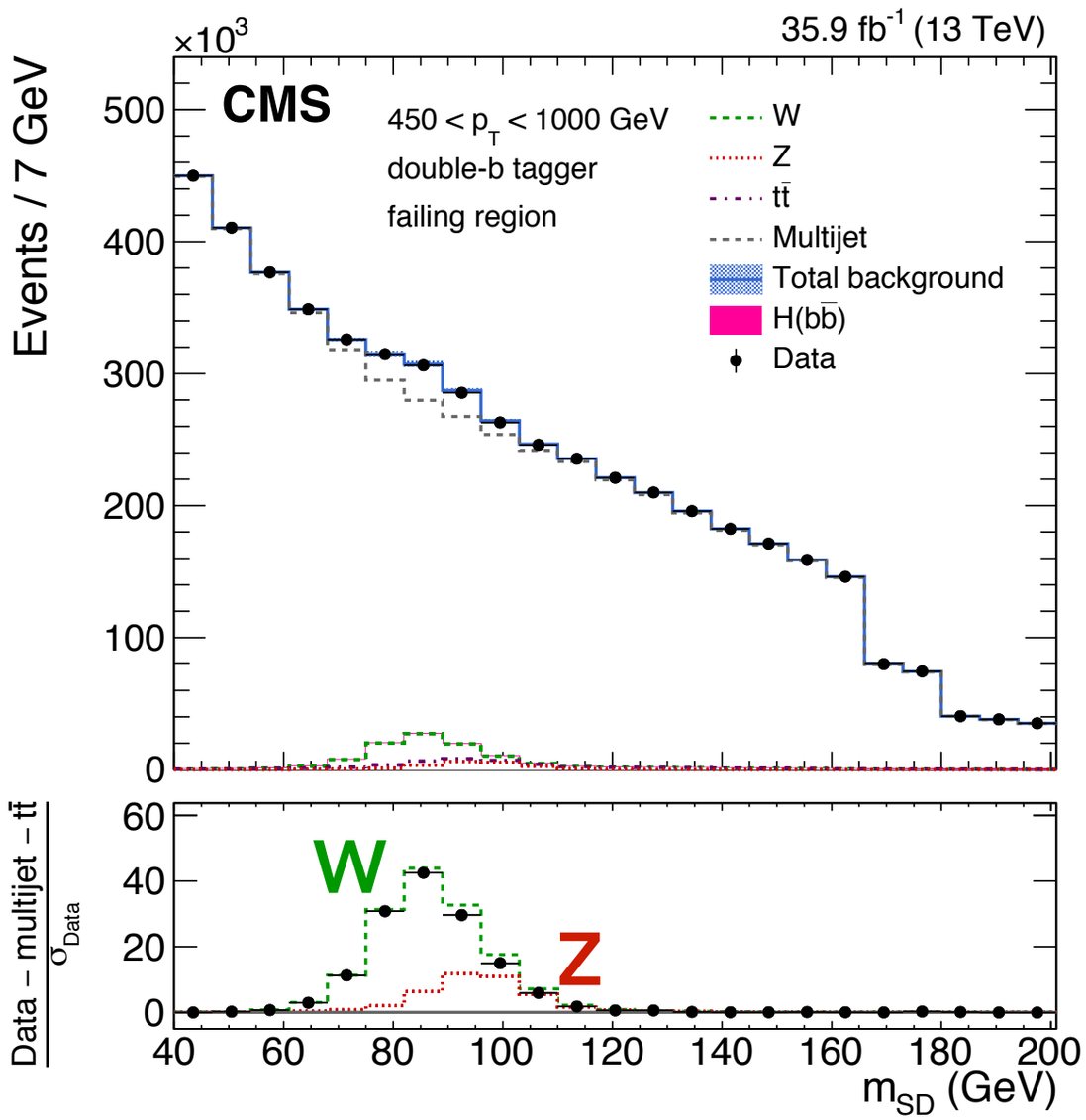


- Choose uncorrelated inputs
- "Divide out correlation" for correlated inputs, e.g. $p_{T\gamma\gamma}/m_{\gamma\gamma}$
- Decorrelate through loss function during training (domain adaptation)
- "Force" decorrelation through variable cuts (DDT)

PRL 120, 071802 (2018)

BOOSTED H(BB)+JET

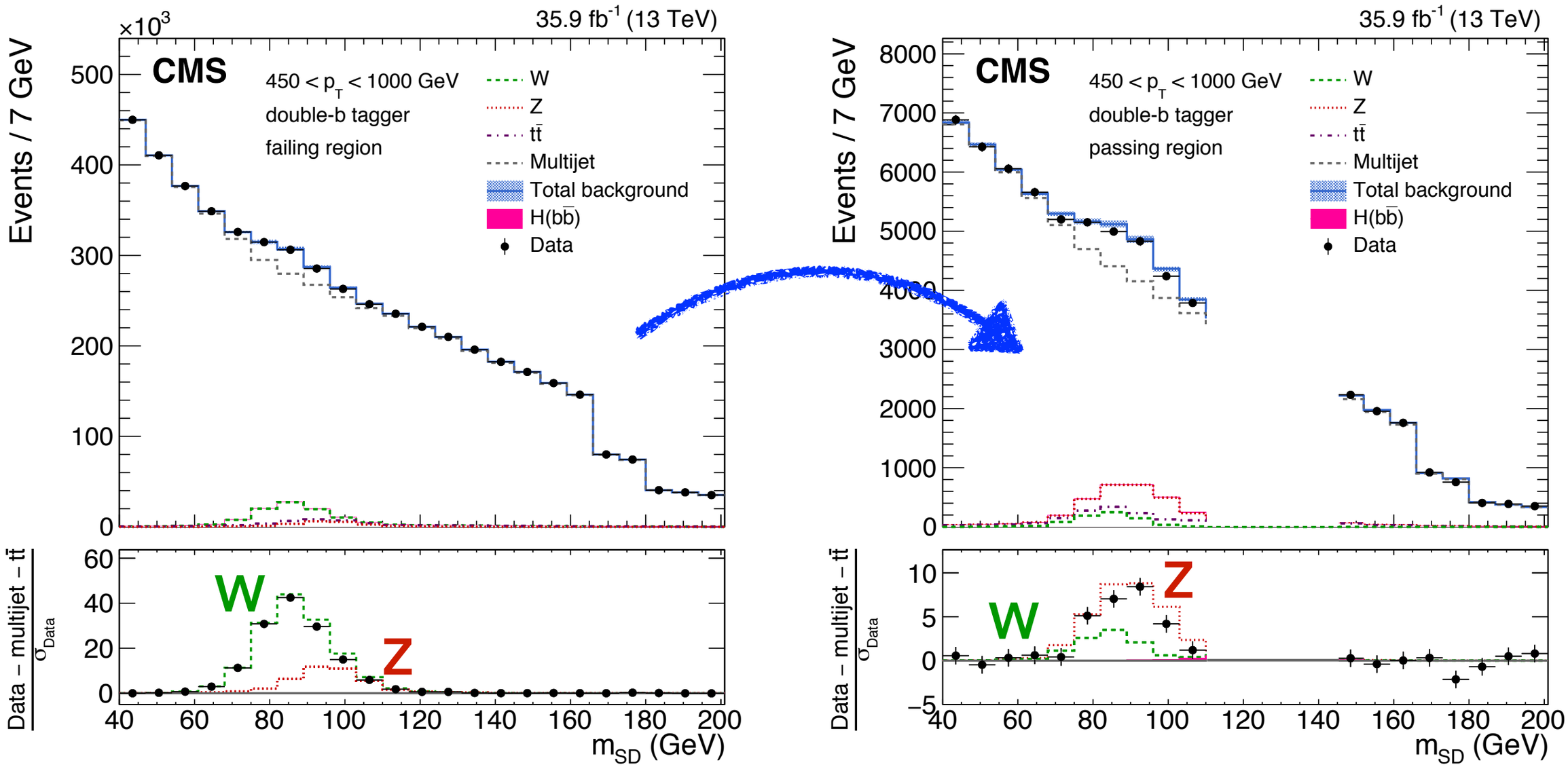
- Estimate QCD shape from events failing the double-b tagger



PRL 120, 071802 (2018)

BOOSTED H(BB)+JET

- Fit for Z(bb): constrains H(bb) systematic uncertainties



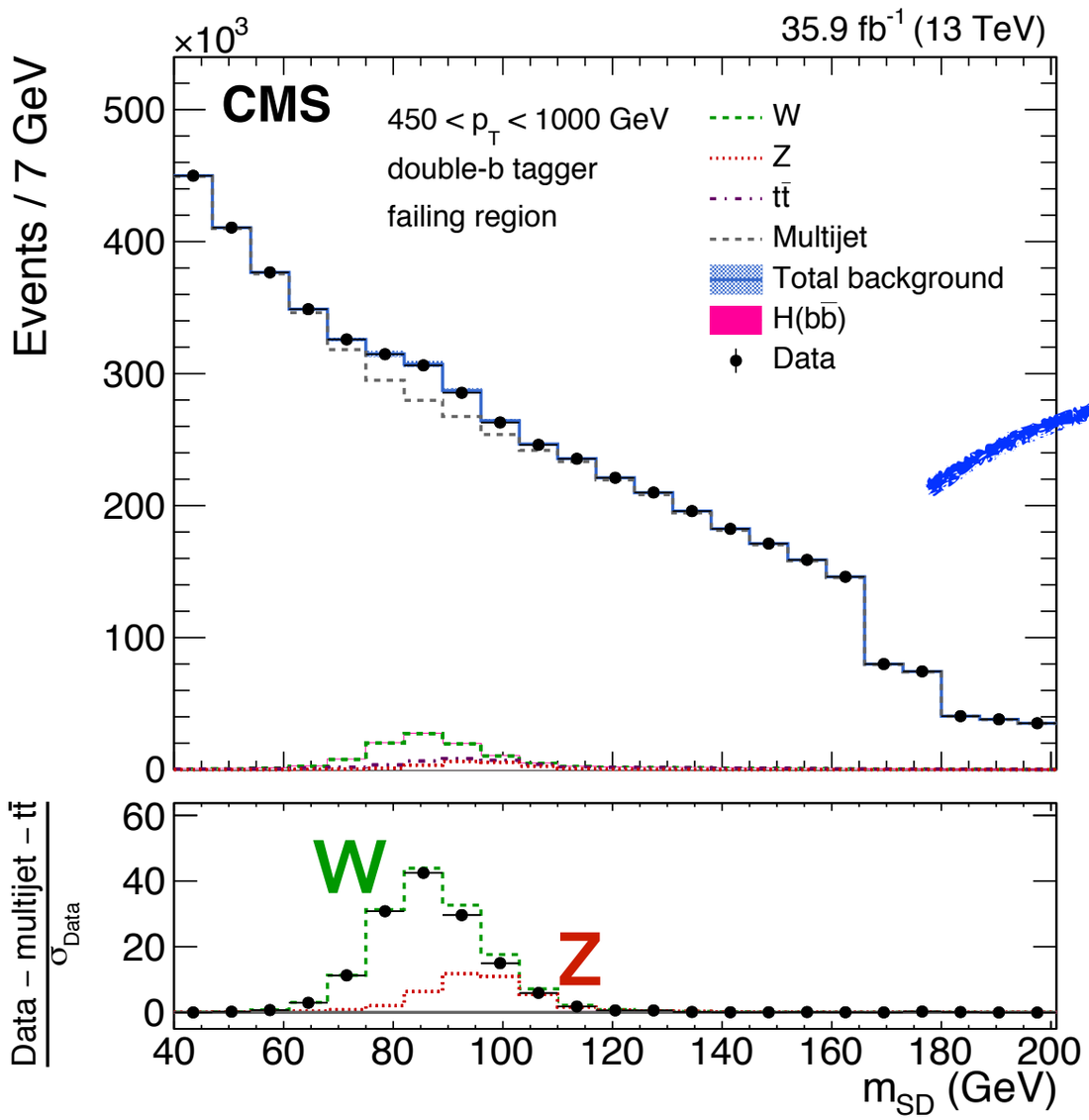
observed Z(bb) significance:

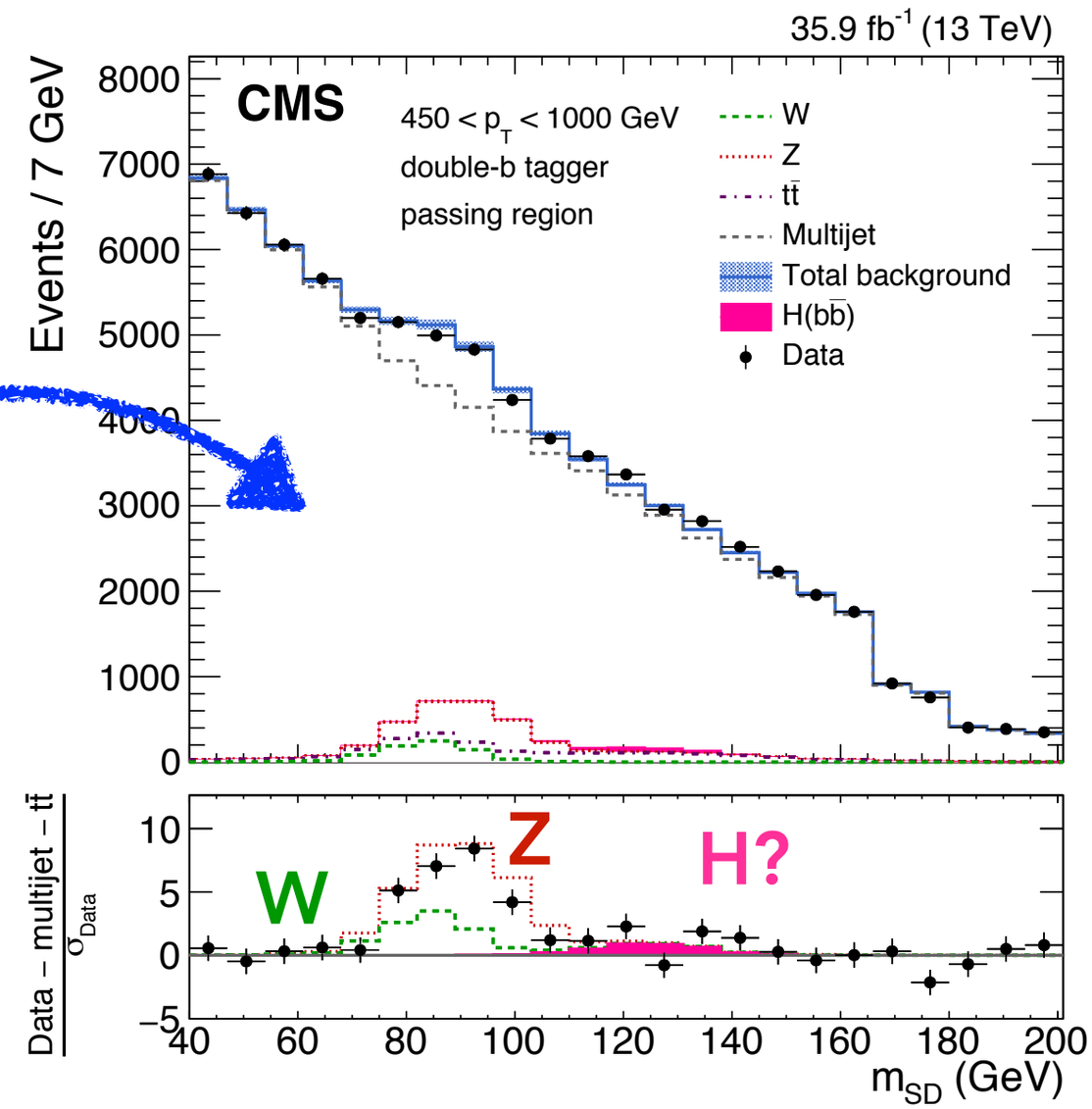
5.1σ, $\mu_Z = 0.78^{+0.23}_{-0.19}$

PRL 120, 071802 (2018)

BOOSTED H(BB)+JET

- Simultaneous fit for Z(bb) and H(bb)



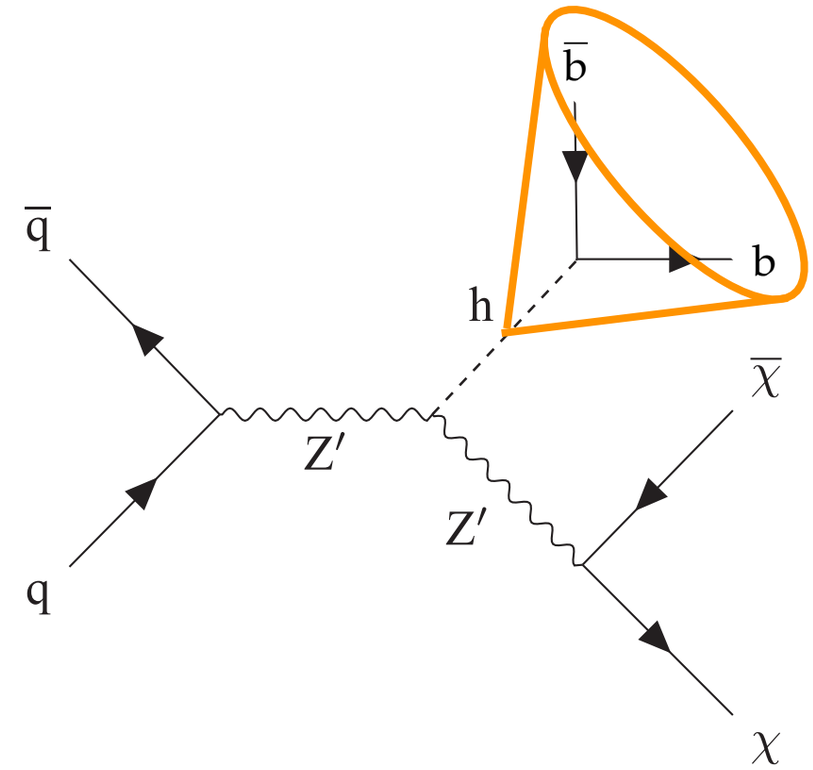


observed H(bb) significance: 1.5σ, $\mu_H = 2.3^{+1.8}_{-1.6}$

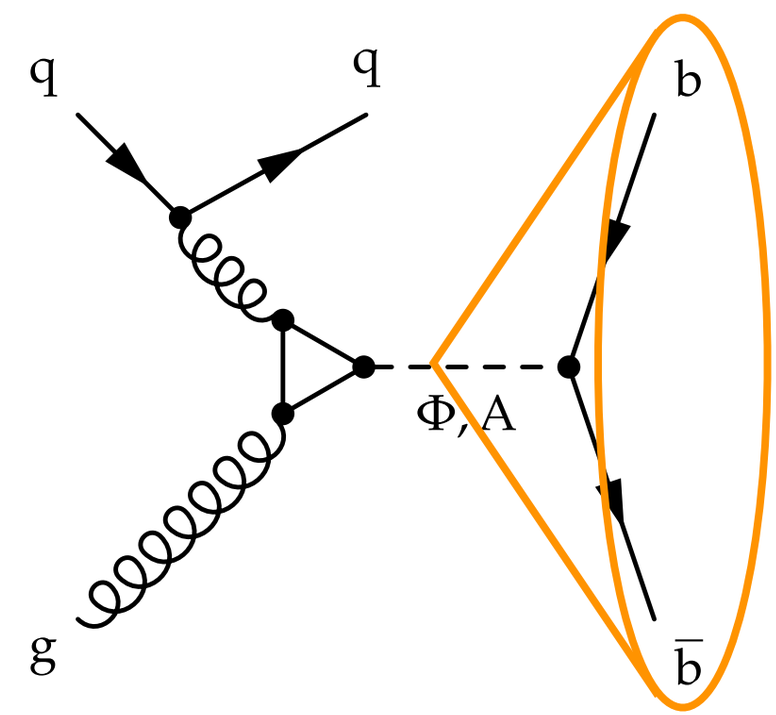
~3σ expected for 2016+2017+2018 data!

DARK MATTER MEDIATORS

- Search for spin-1 and spin-0 **dark matter mediators** that **interact** with **SM Higgs** (H(bb) + MET)



- Search for **Higgs-like** spin-0 **dark matter mediators** (Φ(bb) + jet)



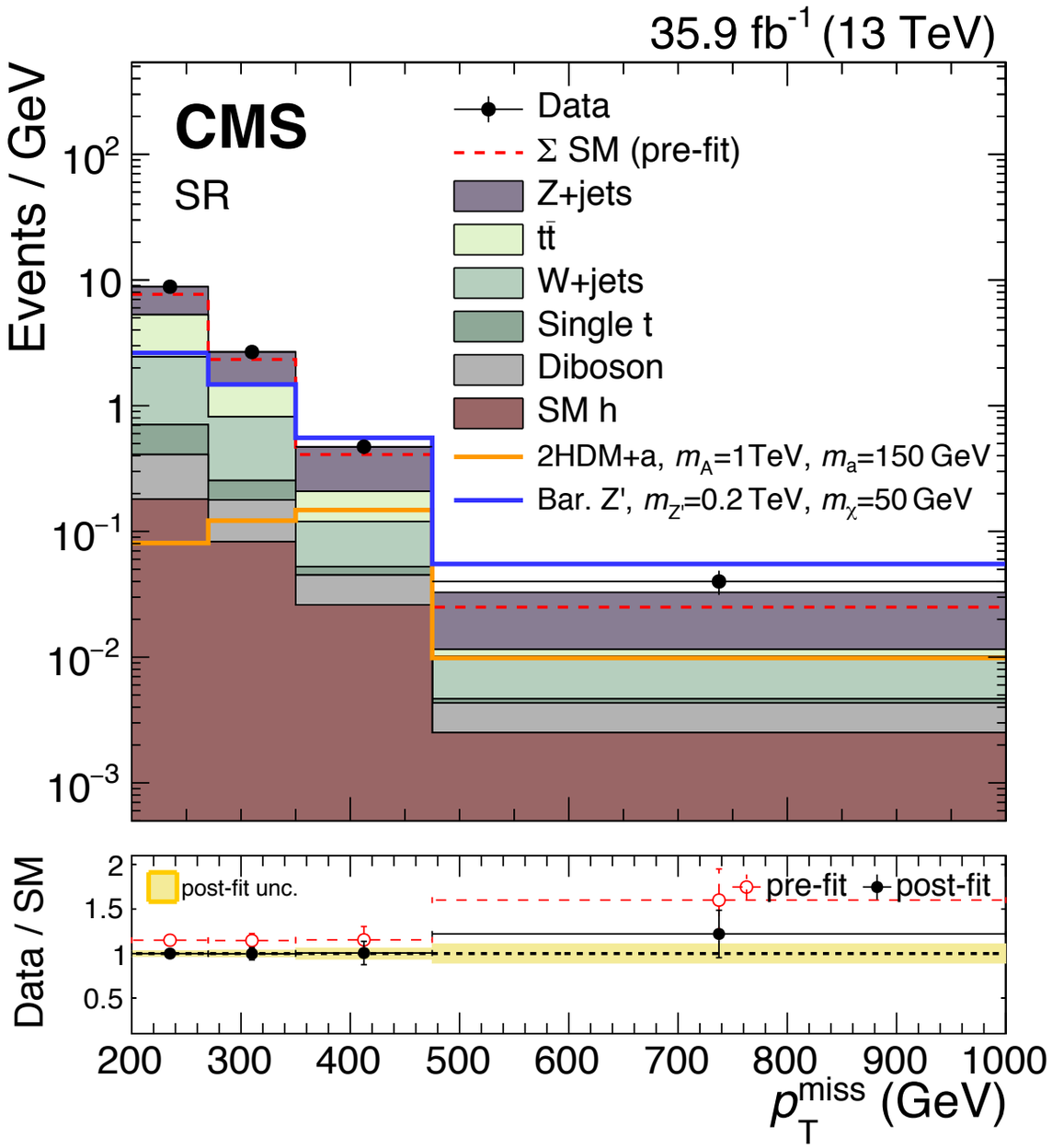
- **Retrained** R = 0.8 BDT for **R = 1.5** jets with one additional input: sub-leading subjet CSVv2 score

EXO-16-050

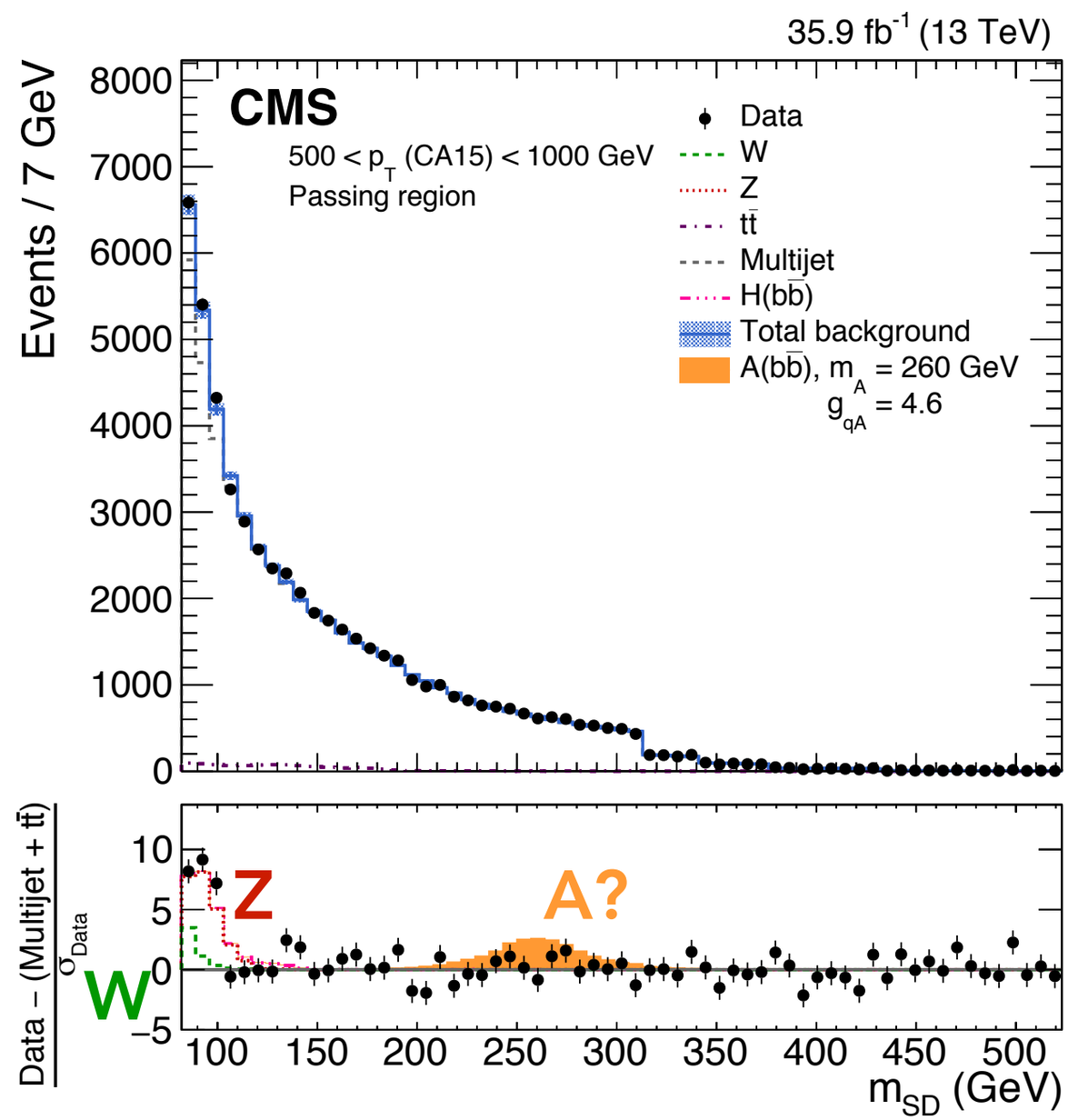
R = 1.5 SEARCHES

EXO-17-024

- Baryonic Z′ hypothesis at 200 GeV with m_χ = 50 GeV



- Pseudoscalar A(bb) hypothesis at 260 GeV



HEAVY FLAVOR TAGGING FOR BOOSTED RESONANCES

DEEP LEARNING FOR LARGE-CONE JETS

Javier Duarte
Fermilab

IMPROVING THE TAGGERS

- Dominant QCD background is a limiting feature of many searches
- Improve taggers with ***deep learning*** by leveraging:
  - Modern machine learning tools (Keras, TensorFlow, PyTorch, etc.)
  - Increased GPU availability (CMS, CERN, FNAL, AWS resources)
  - Shared data preprocessing and training framework:
    https://github.com/DL4Jets

Deep Learning Base AMI (Amazon Linux) Version 3.0 - ami-8cc478f4

Deep Learning base AMI with NVidia drivers like CUDA 8 and 9, CuDNN 6 and 7, CuBLAS 8 and 9, NCCL and more

Root Device Type: ebs Virtualization type: hvm

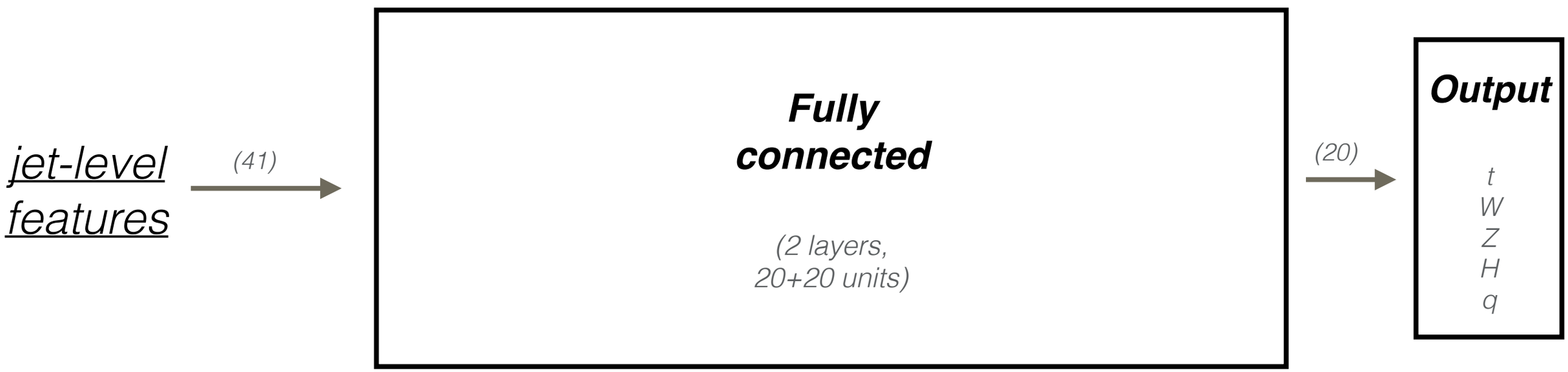
Instance Type Edit instance type

Instance Type	ECUs	vCPUs	Memory (GiB)	Instance Storage (GB)	EBS-Optimized Available	Network Performance
p3.16xlarge	188	64	488	EBS only	Yes	25 Gigabit

Cancel Previous Launch

DP-2017/027 BOOSTED EVENT SHAPE TAGGER

- Inputs are jet-level features evaluated in four different hypothesis rest frames (t, W, Z, H)

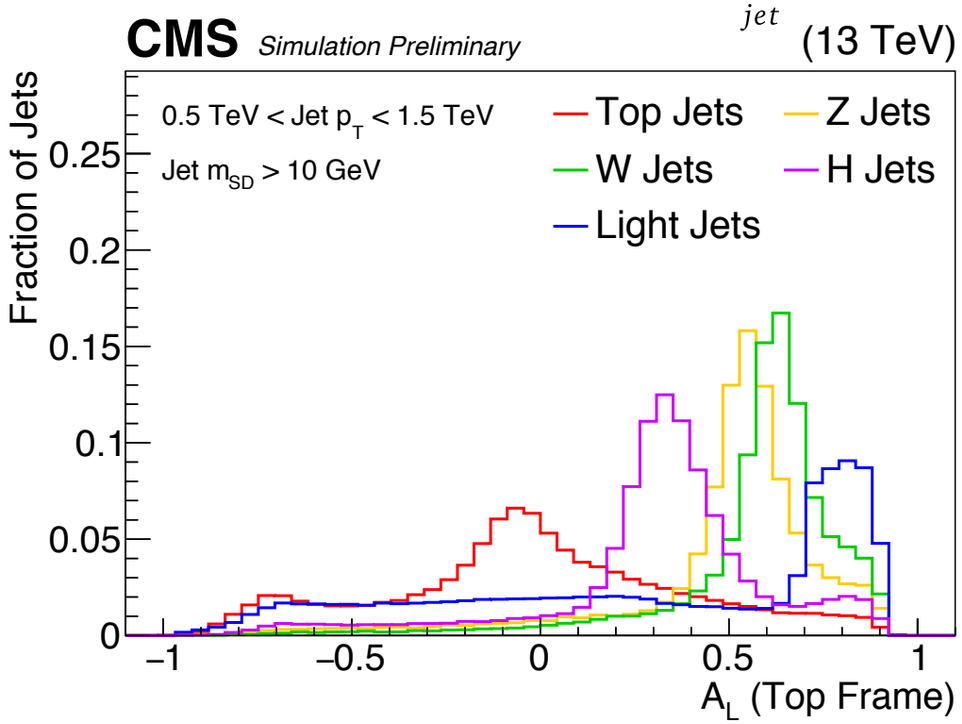


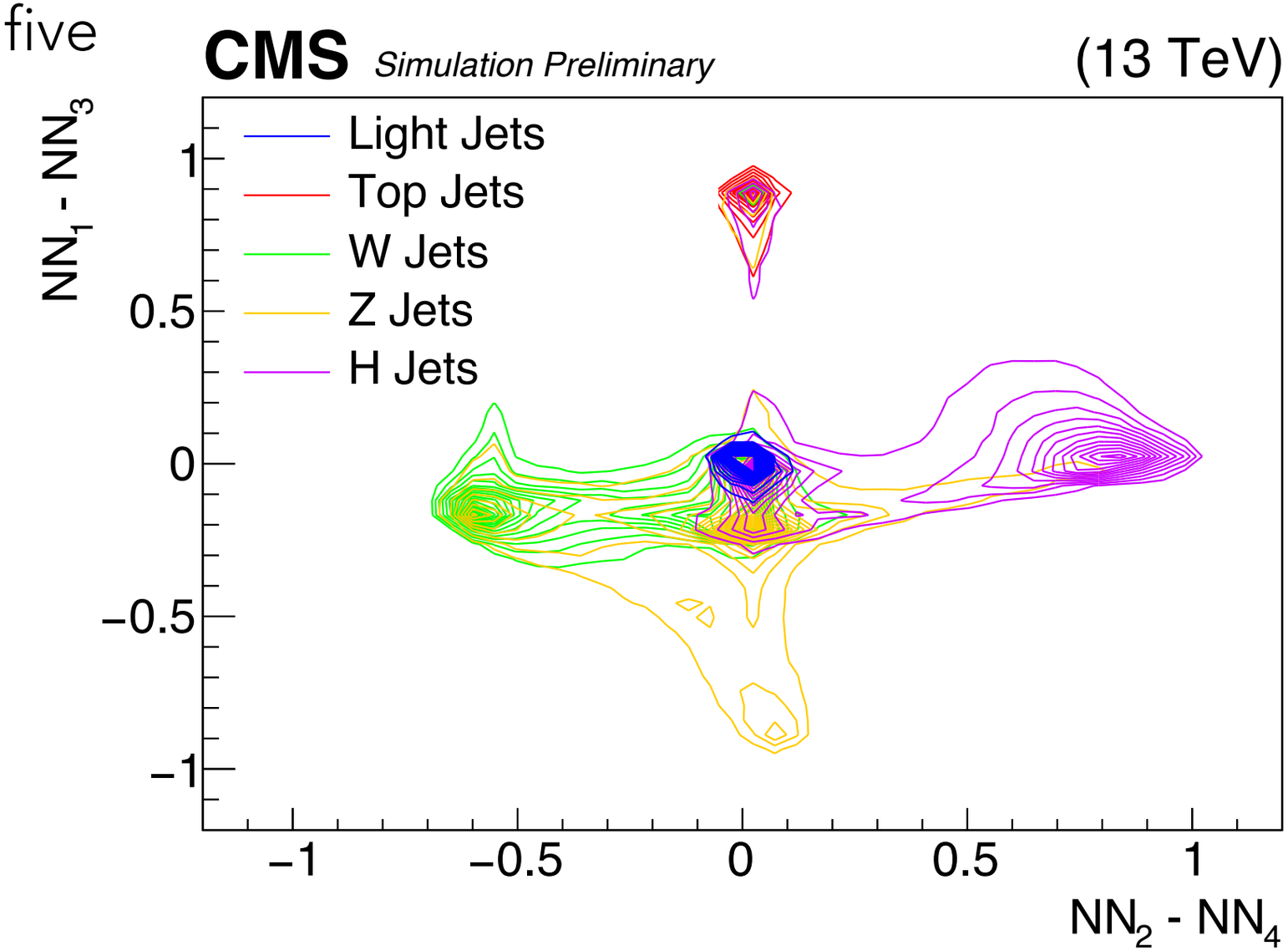
DP-2017/027 BOOSTED EVENT SHAPE TAGGER

- Boost jet into hypothesized rest frames and computes different variables
- When boosting to "correct" frame, jet constituents should be isotropic and show the N-prong structure
- Good separation between five particle classes

jet asymmetry in top rest frame

$$A_L = \frac{\sum_{jet} p_z^{jet}}{\sum_{jet} p^{jet}}$$



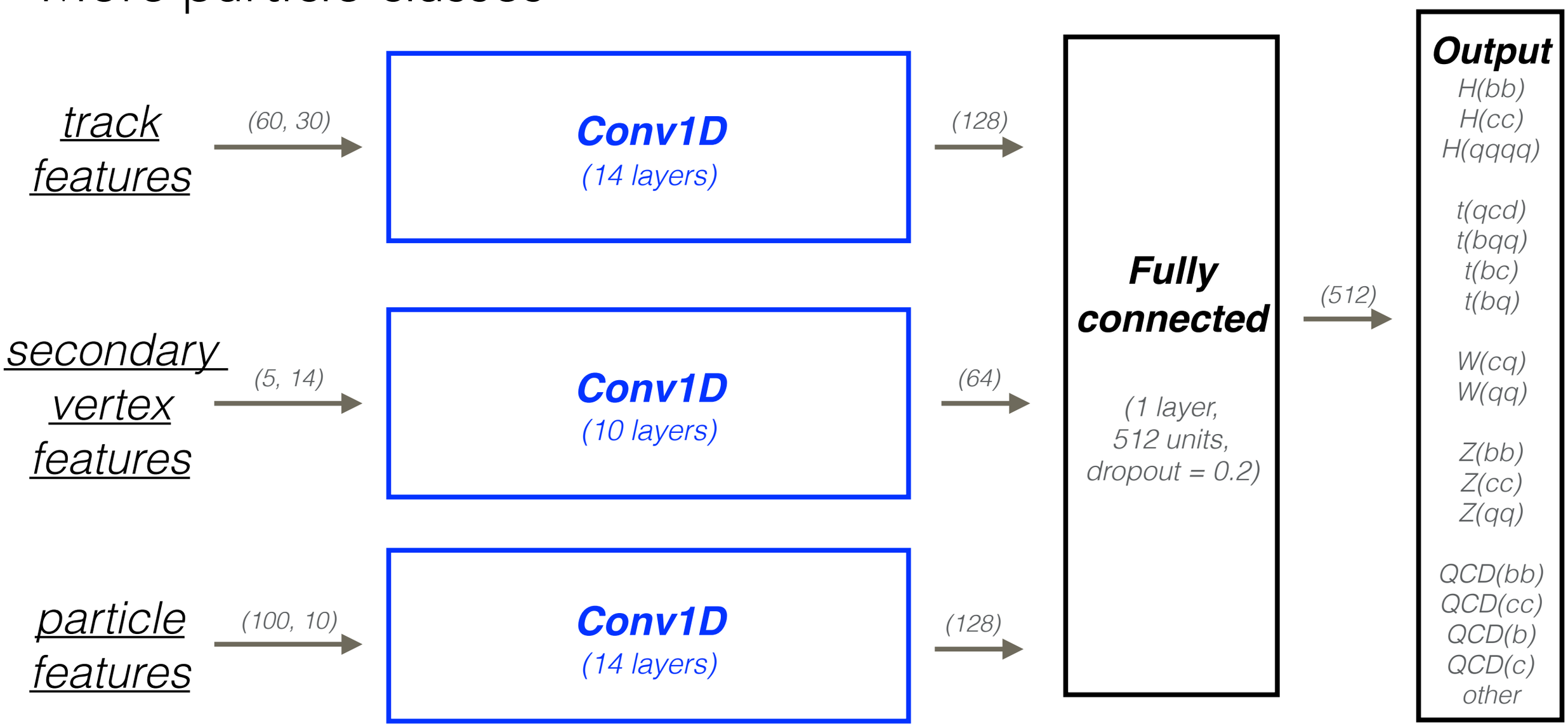


DP-2017/049

DEEP BOOSTED JET

- Designed for maximal performance
- More input variables (up to 2,870!)
- More particle classes

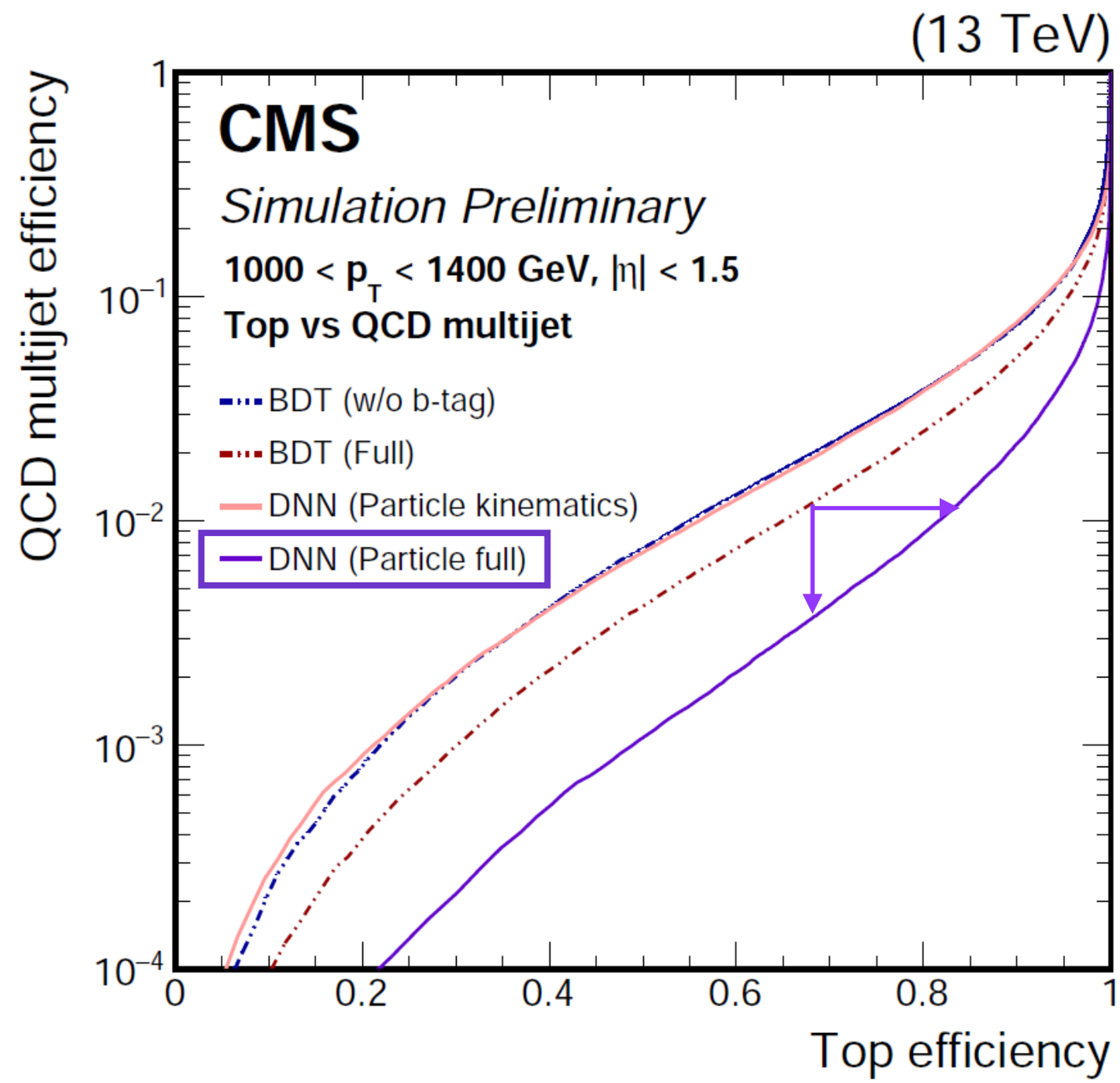
ADVERSARIAL APPROACH FOR MASS DECORRELATION



DP-2017/049

DEEP BOOSTED JET

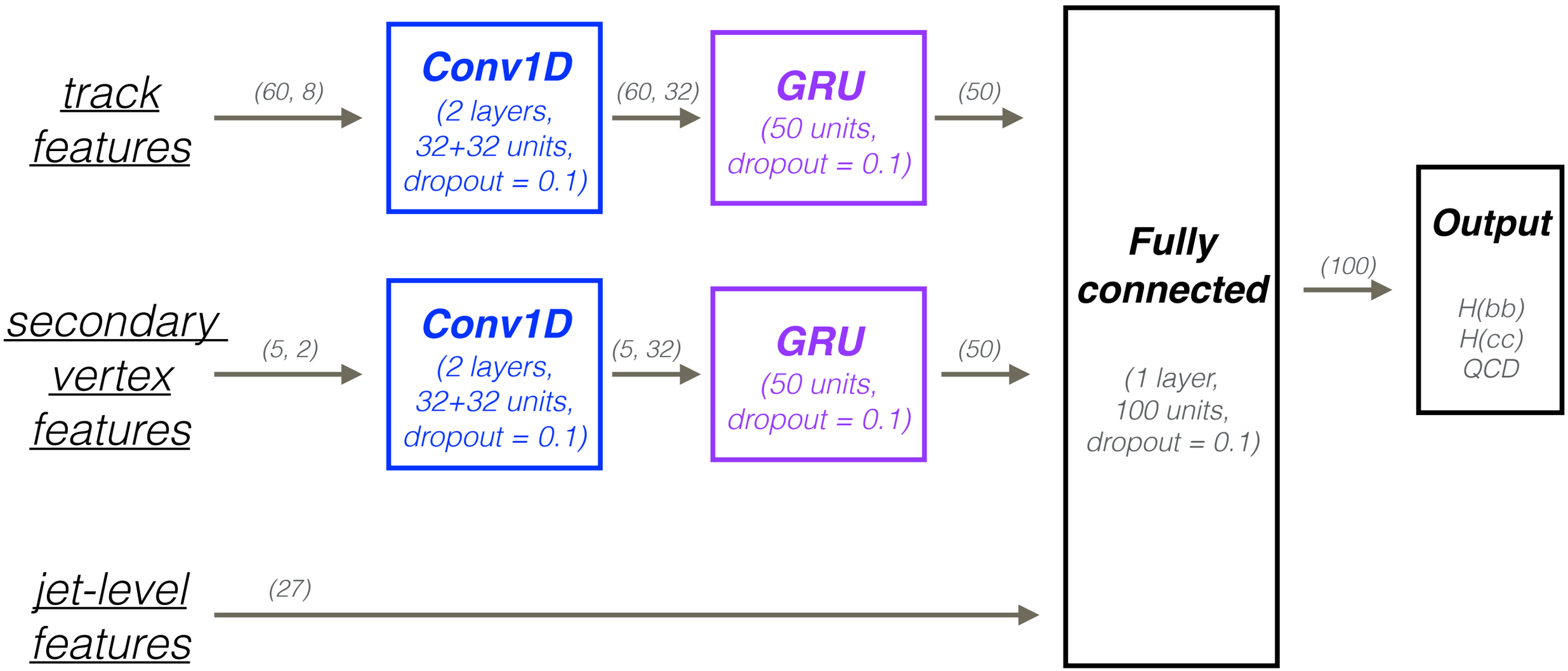
- Excellent performance vs. jet-level approaches
- Significant gain for particle-based DNN approach after including flavour
- Mass decorrelated version also trained with adversarial network: trade off between performance and decorrelation



DP-2018/033

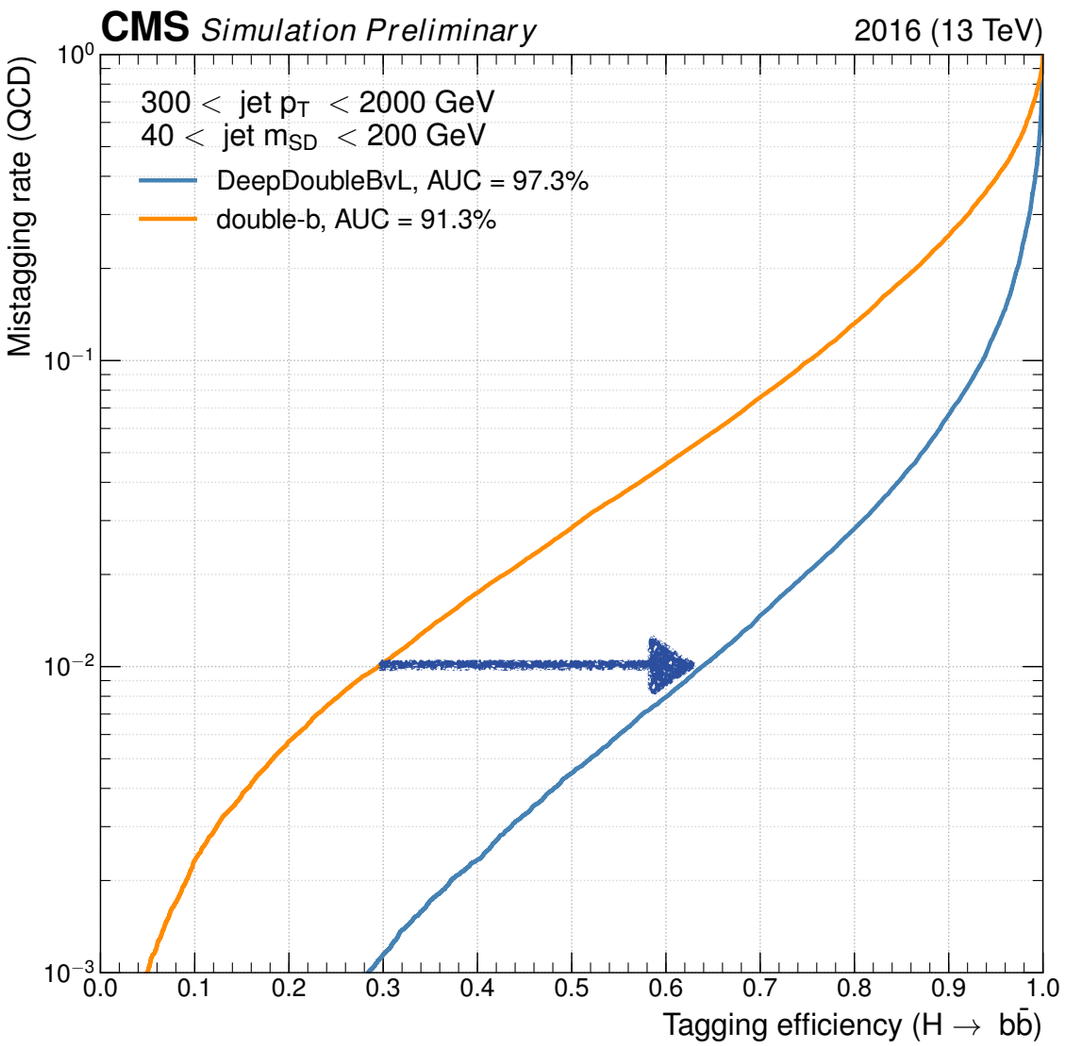
DEEP DOUBLE-B/C TAGGER

- **Convolutional** layers: used in image recognition, …
- **Recurrent** layers: used in language translation, …
- Reduced set of kinematic inputs to mitigate mass sculpting



DP-2018/033

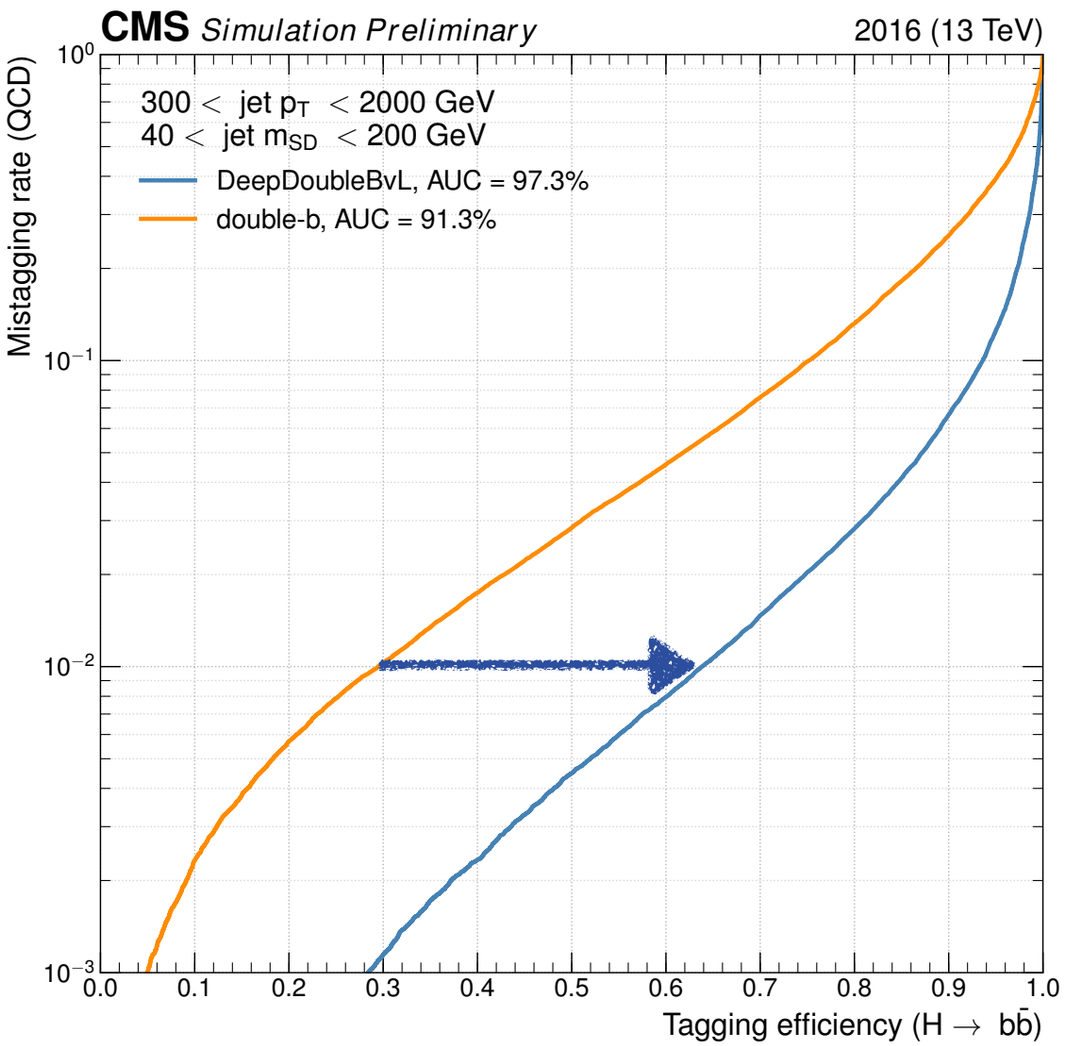
DEEP DOUBLE-B TAGGER



DP-2018/033

DEEP DOUBLE-B TAGGER

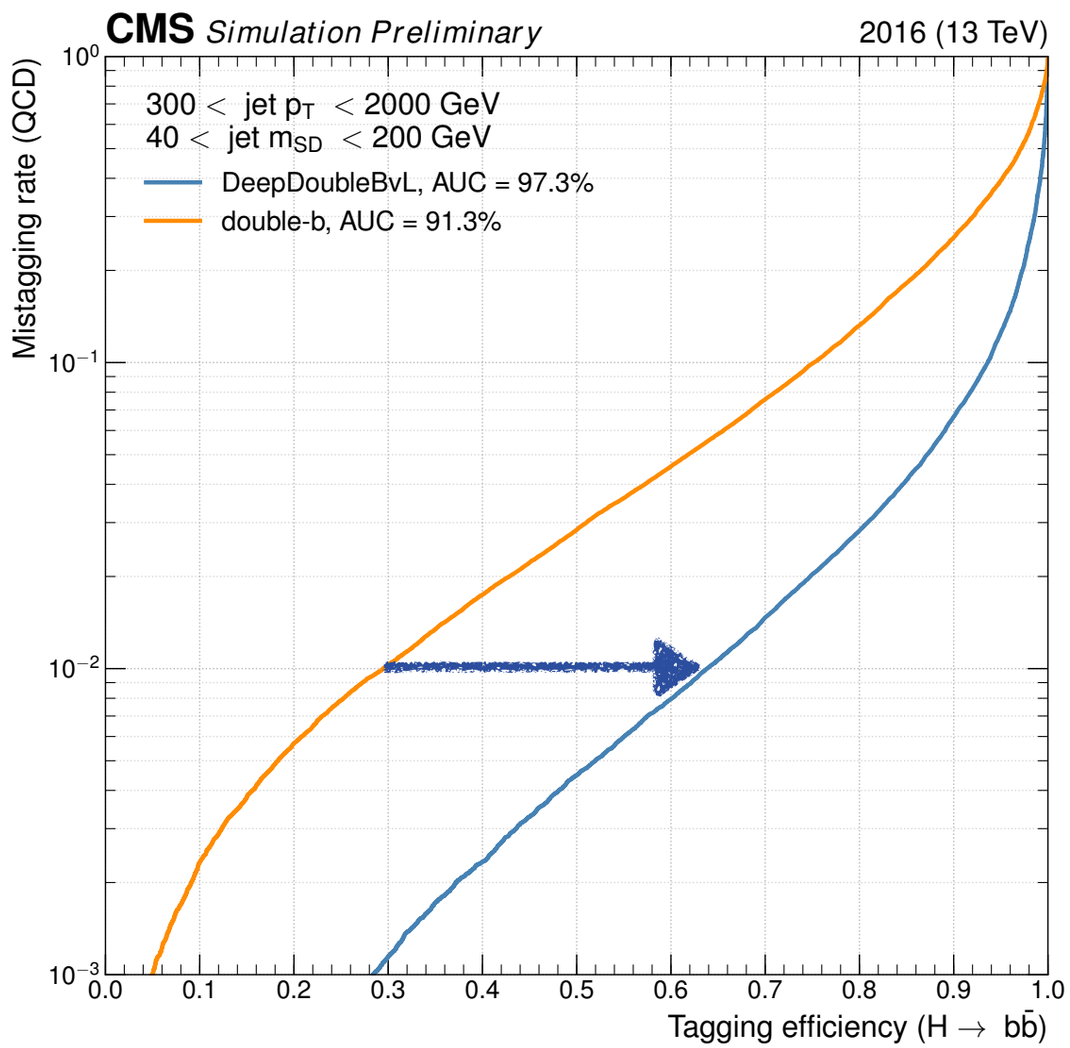
- Large performance gain over BDT

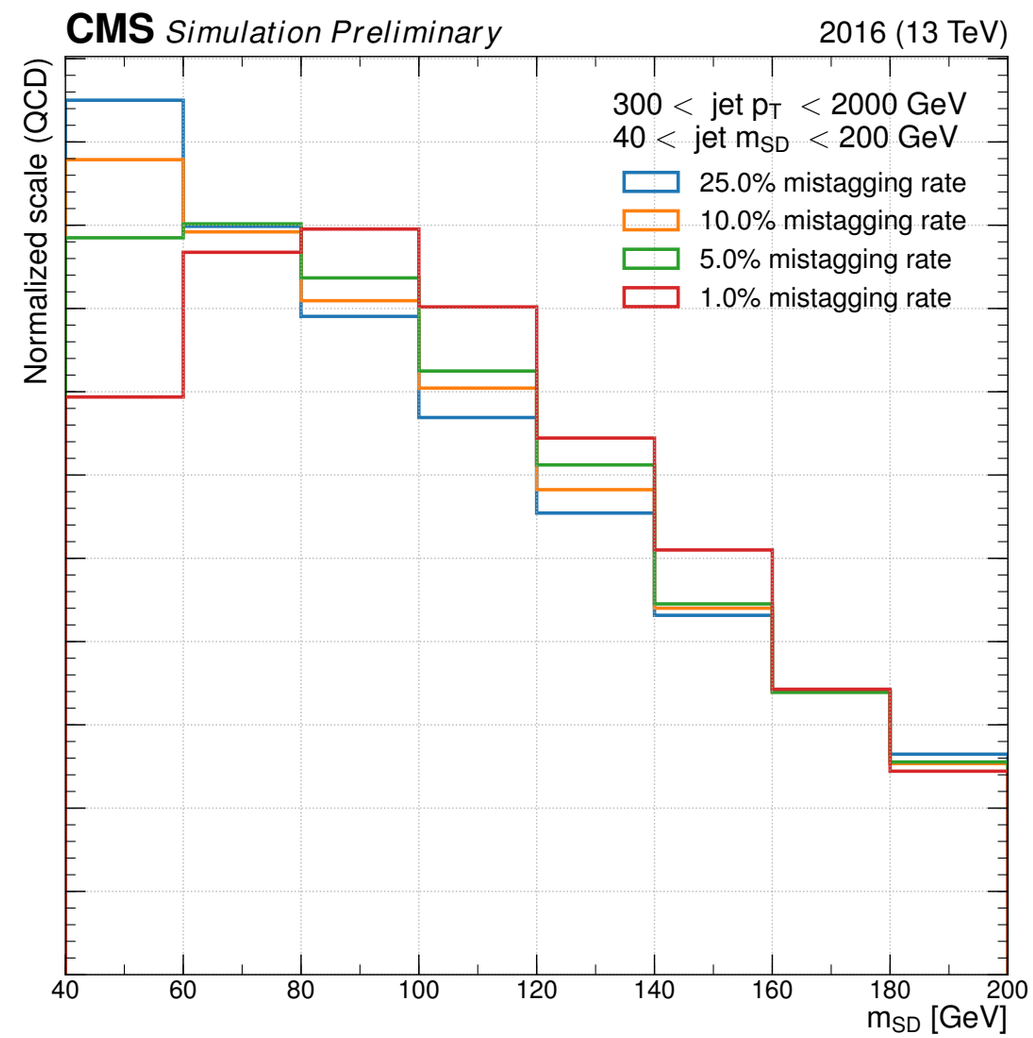


DP-2018/033

DEEP DOUBLE-B TAGGER

- Large performance gain over BDT
- Default algorithm still "learns" the mass ⇒ mass sculpting





HOW TO DECORRELATE?

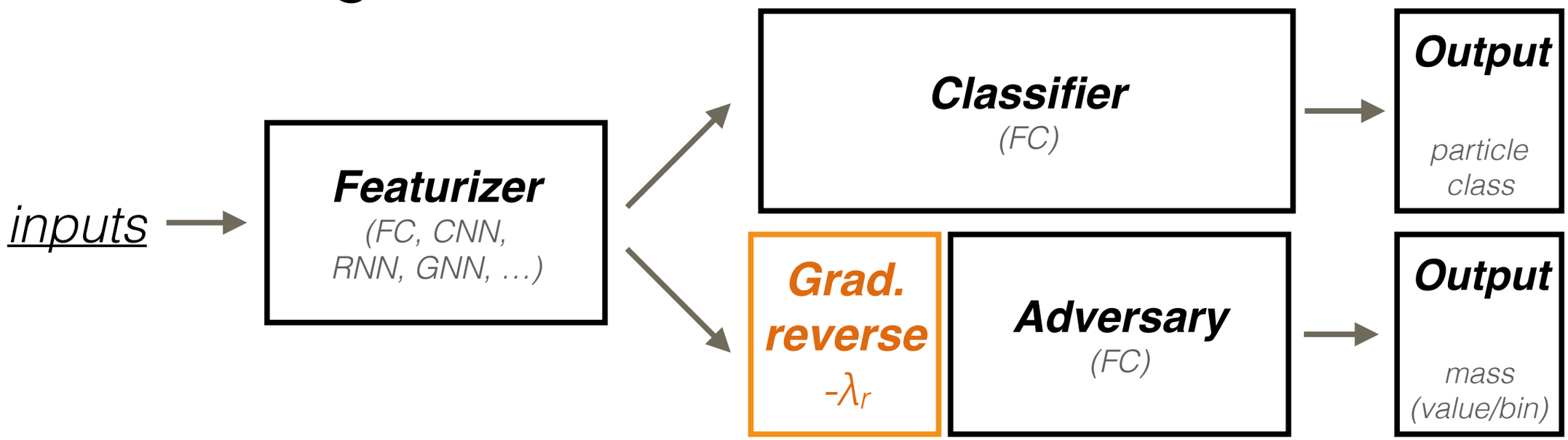
arXiv:1611.01046
arXiv:1409.7495

HOW TO DECORRELATE?

arXiv:1611.01046
arXiv:1409.7495

- Adversarial training

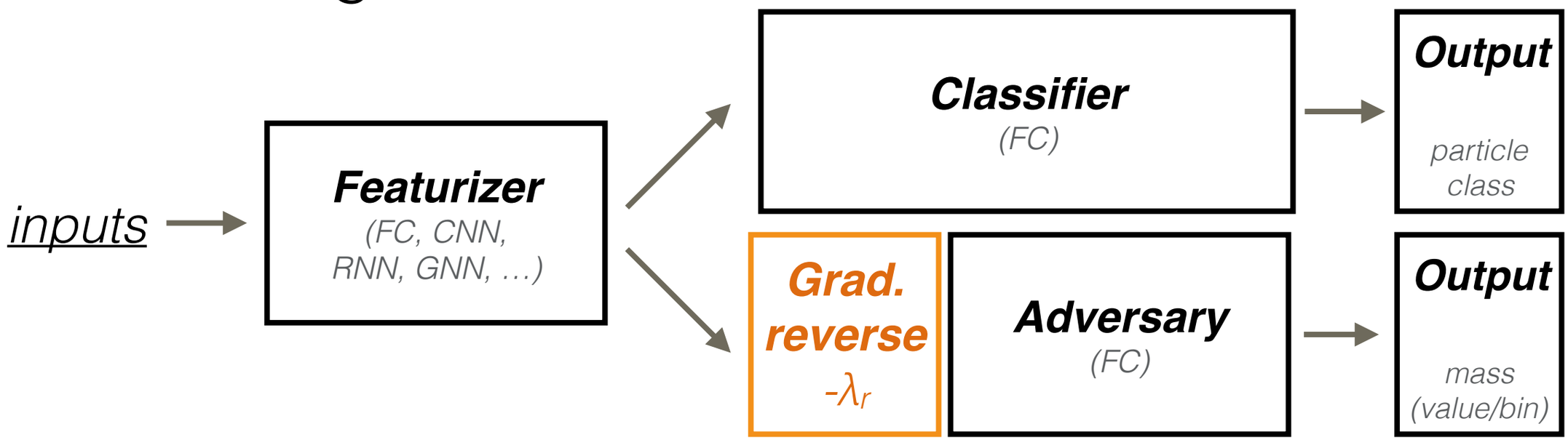
$$L = L_{\text{disc}} + \lambda L_{\text{adv}}$$



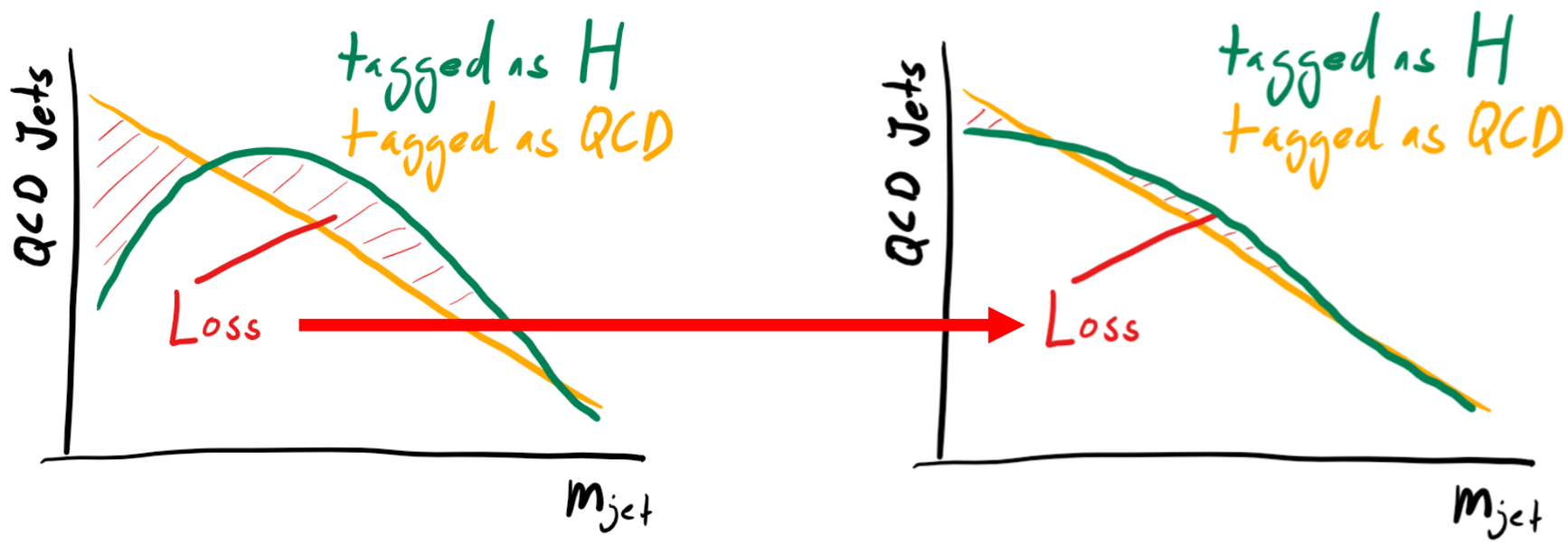
HOW TO DECORRELATE?

arXiv:1611.01046
arXiv:1409.7495

- Adversarial training $L = L_{\text{disc}} + \lambda L_{\text{adv}}$



- Dedicated "penalty term" $L = L_{\text{disc}} + \lambda D_{\text{KL}}$



PENALTY TERM

*$h_{b(s)}$: one-hot encoded vector of mass bin for background (signal)

PENALTY TERM

- Kullback-Liebler divergence is an information-theoretic measure of the distance between two distributions h_1 and h_2

$$D_{\mathrm{KL}}(h_1 \parallel h_2) = h_1 \log\left(\frac{h_1}{h_2}\right)$$

*$h_{b(s)}$: one-hot encoded vector of mass bin for background (signal)

PENALTY TERM

- Kullback-Liebler divergence is an information-theoretic measure of the distance between two distributions h_1 and h_2

$$D_{\mathrm{KL}}(h_1 \parallel h_2) = h_1 \log\left(\frac{h_1}{h_2}\right)$$

- Loss function composed of three terms:

*$h_{b(s)}$: one-hot encoded vector of mass bin for background (signal)

PENALTY TERM

- Kullback-Liebler divergence is an information-theoretic measure of the distance between two distributions h_1 and h_2

$$D_{\mathrm{KL}}(h_1 \parallel h_2) = h_1 \log\left(\frac{h_1}{h_2}\right)$$

- Loss function composed of three terms:

$$L = -y\log(p) - (1-y)\log(1-p)$$

categorical/binary cross entropy

*$h_{b(s)}$: one-hot encoded vector of mass bin for background (signal)

PENALTY TERM

- Kullback-Liebler divergence is an information-theoretic measure of the distance between two distributions h_1 and h_2

$$D_{\mathrm{KL}}(h_1 \parallel h_2) = h_1 \log\left(\frac{h_1}{h_2}\right)$$

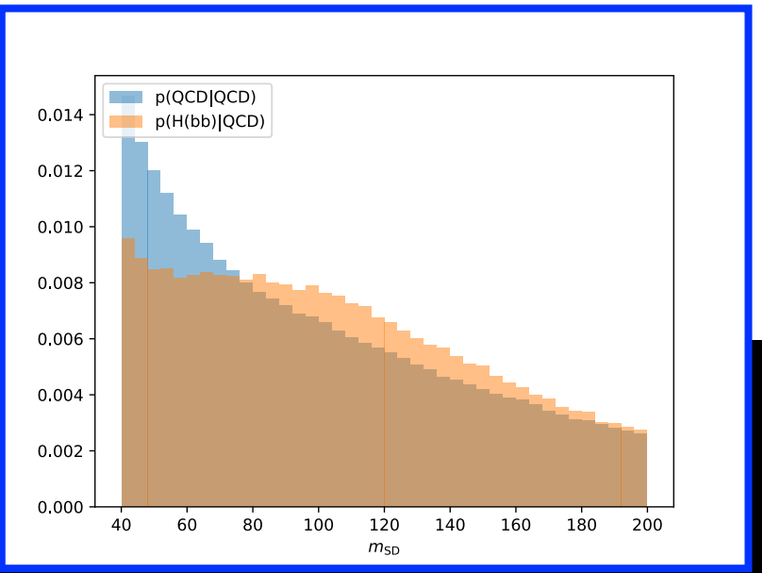
- Loss function composed of three terms:

$$L = -y\log(p) - (1-y)\log(1-p)$$

categorical/binary cross entropy

$$+\lambda D_{\mathrm{KL}}(p \cdot h_{\mathrm{b}} \parallel (1-p)\cdot h_{\mathrm{b}})$$

background sculpting penalty



*$h_{b(s)}$: one-hot encoded vector of mass bin for background (signal)

PENALTY TERM

- Kullback-Liebler divergence is an information-theoretic measure of the distance between two distributions h_1 and h_2

$$D_{\mathrm{KL}}(h_1 \parallel h_2) = h_1 \log\left(\frac{h_1}{h_2}\right)$$

- Loss function composed of three terms:

$$L = -y\log(p) - (1-y)\log(1-p)$$

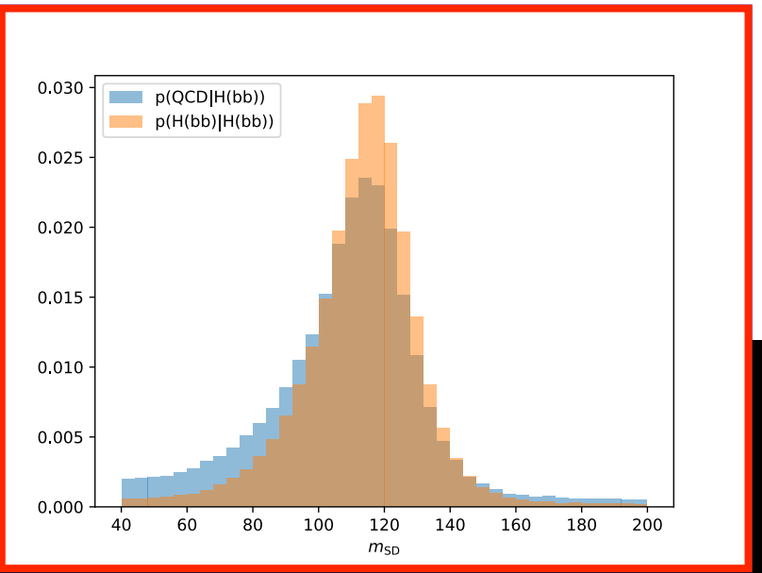
categorical/binary cross entropy

$$+\lambda D_{\mathrm{KL}}(p \cdot h_{\mathrm{b}} \parallel (1-p) \cdot h_{\mathrm{b}})$$

background sculpting penalty

$$+\lambda D_{\mathrm{KL}}(p \cdot h_{\mathrm{s}} \parallel (1-p) \cdot h_{\mathrm{s}})$$

signal sculpting penalty

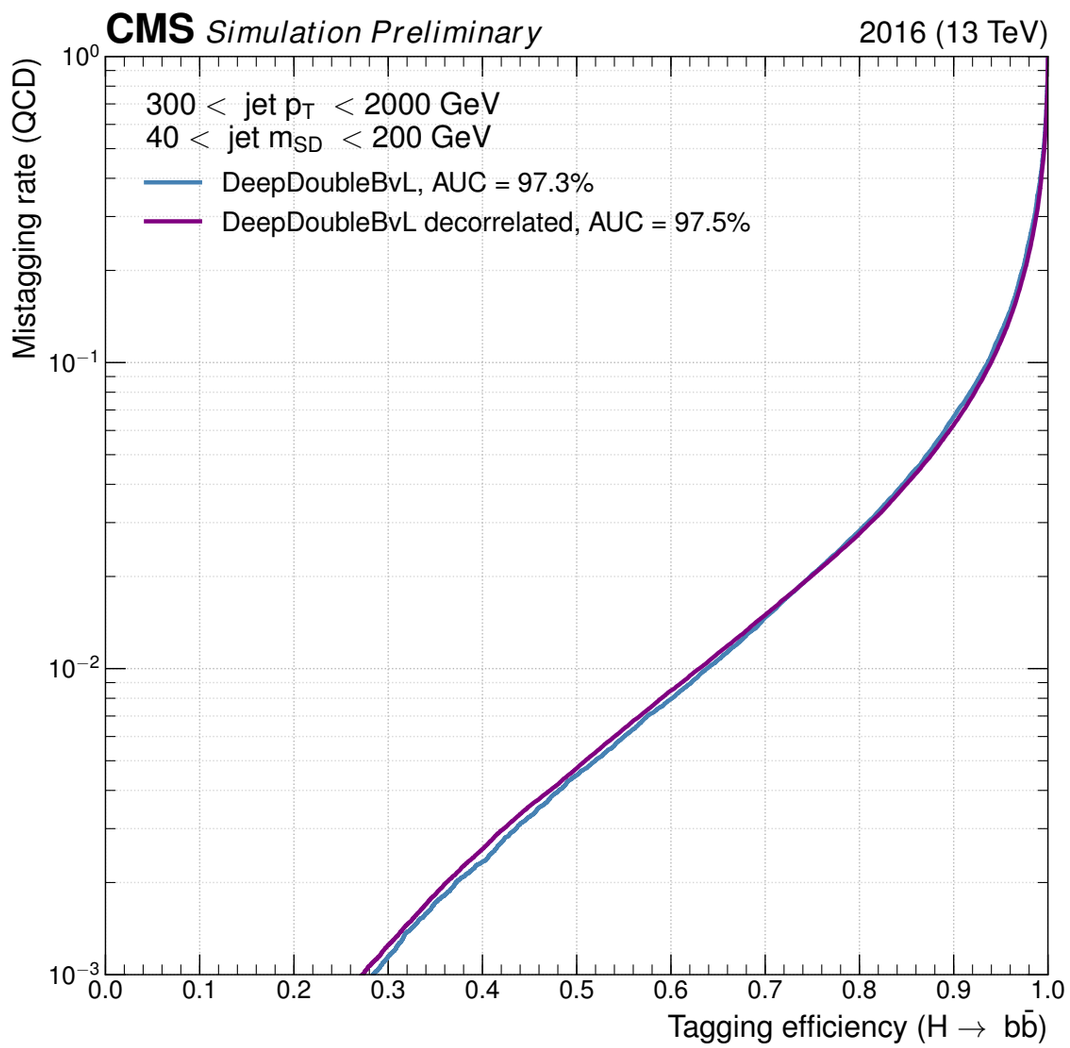


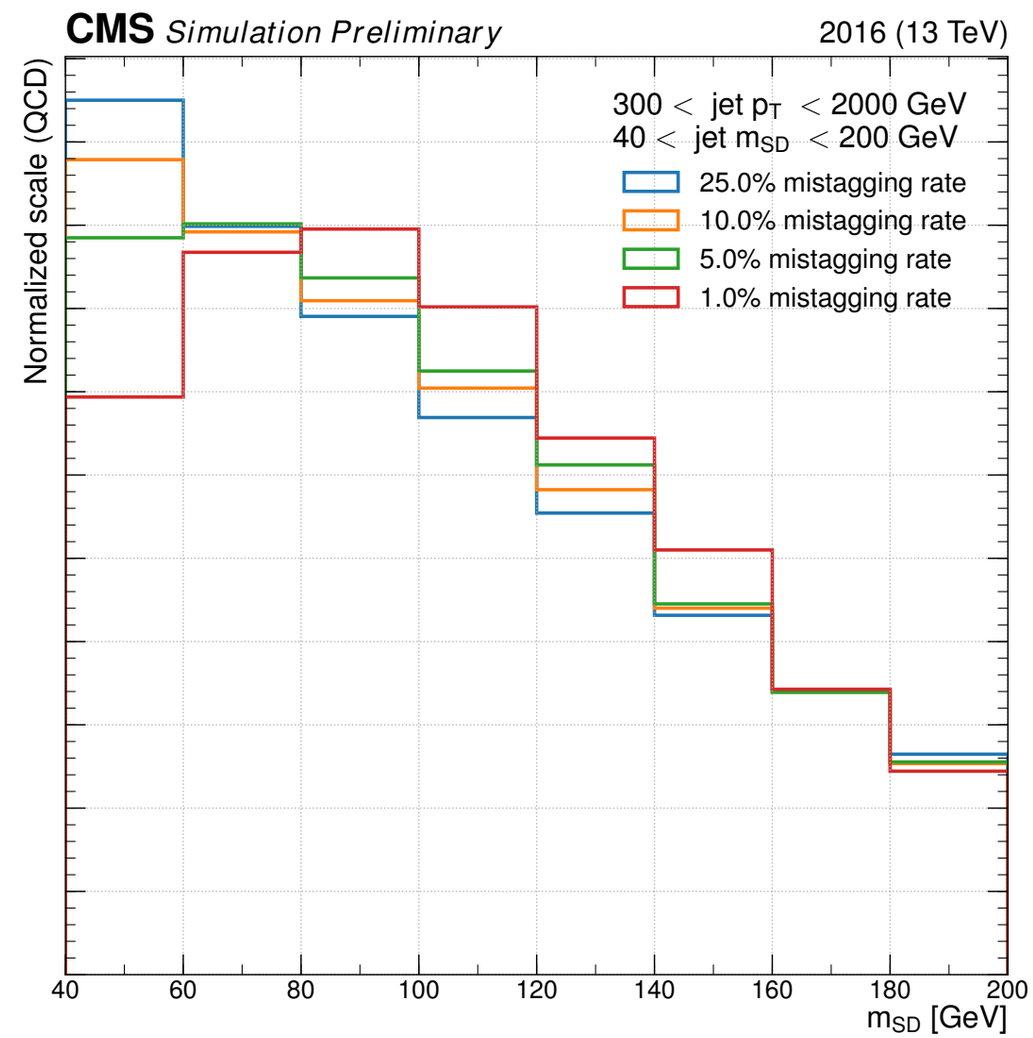
*$h_{b(s)}$: one-hot encoded vector of mass bin for background (signal)

DP-2018/033

DEEP DOUBLE-B TAGGER

- Dedicated "penalty term" based on Kullback-Leibler divergence mitigates mass sculpting

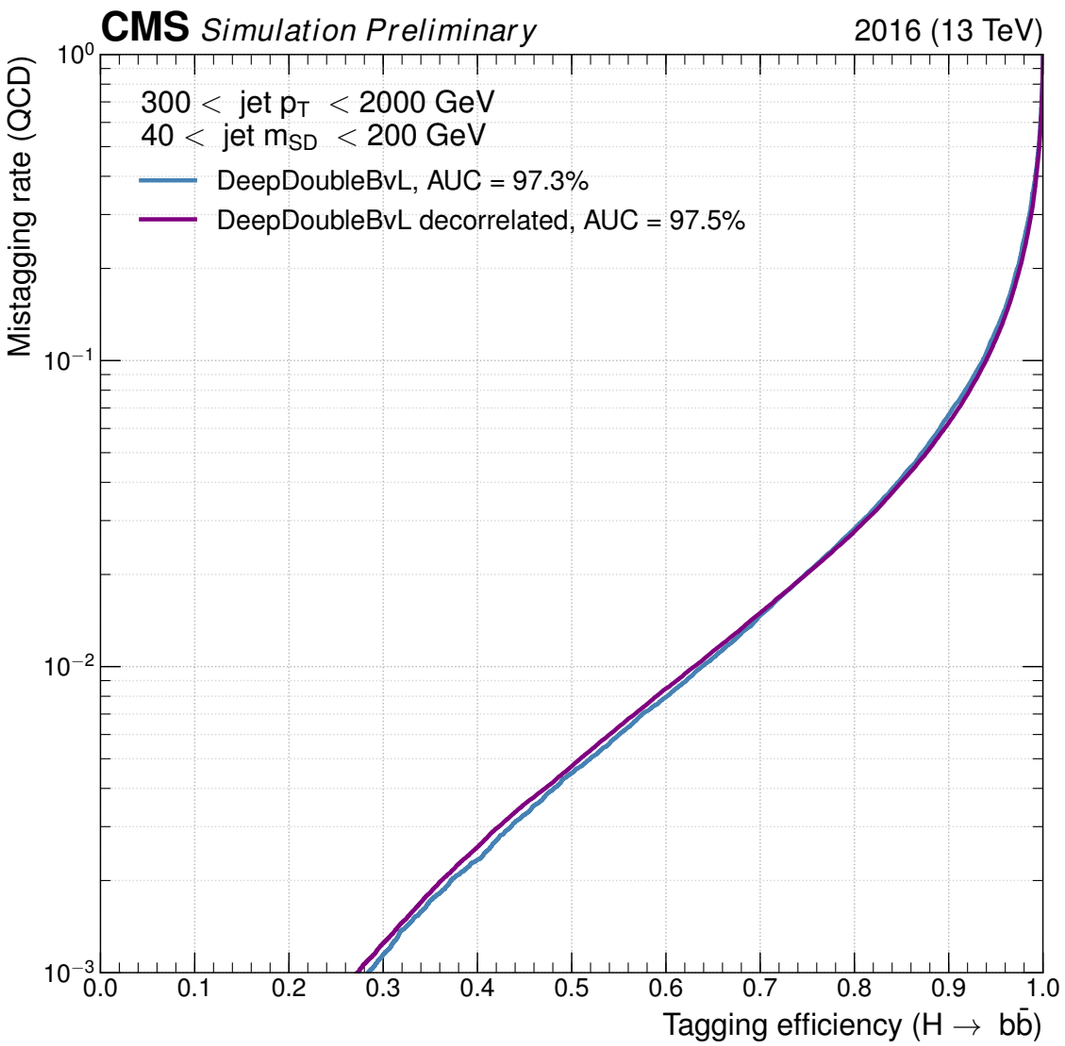


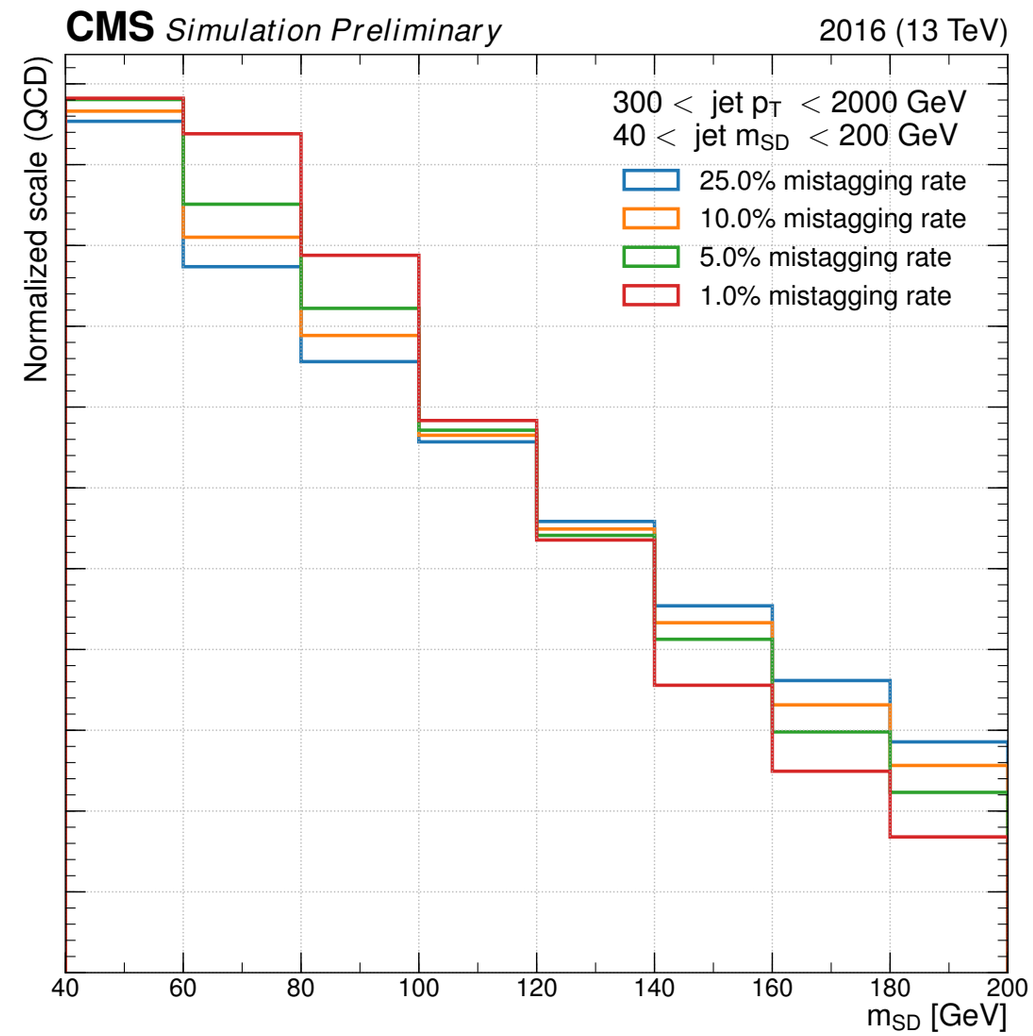


DP-2018/033

DEEP DOUBLE-B TAGGER

- Dedicated "penalty term" based on Kullback-Leibler divergence mitigates mass sculpting

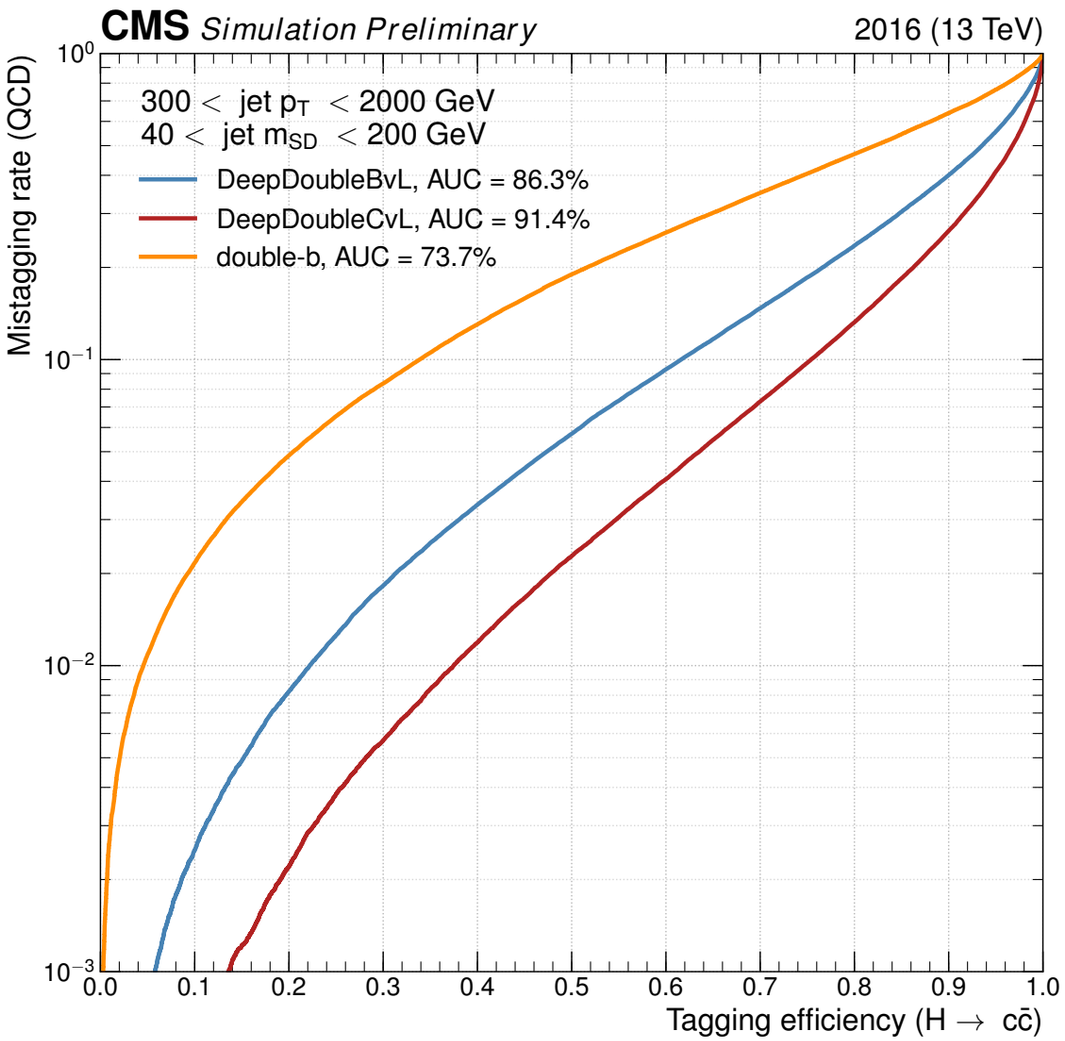


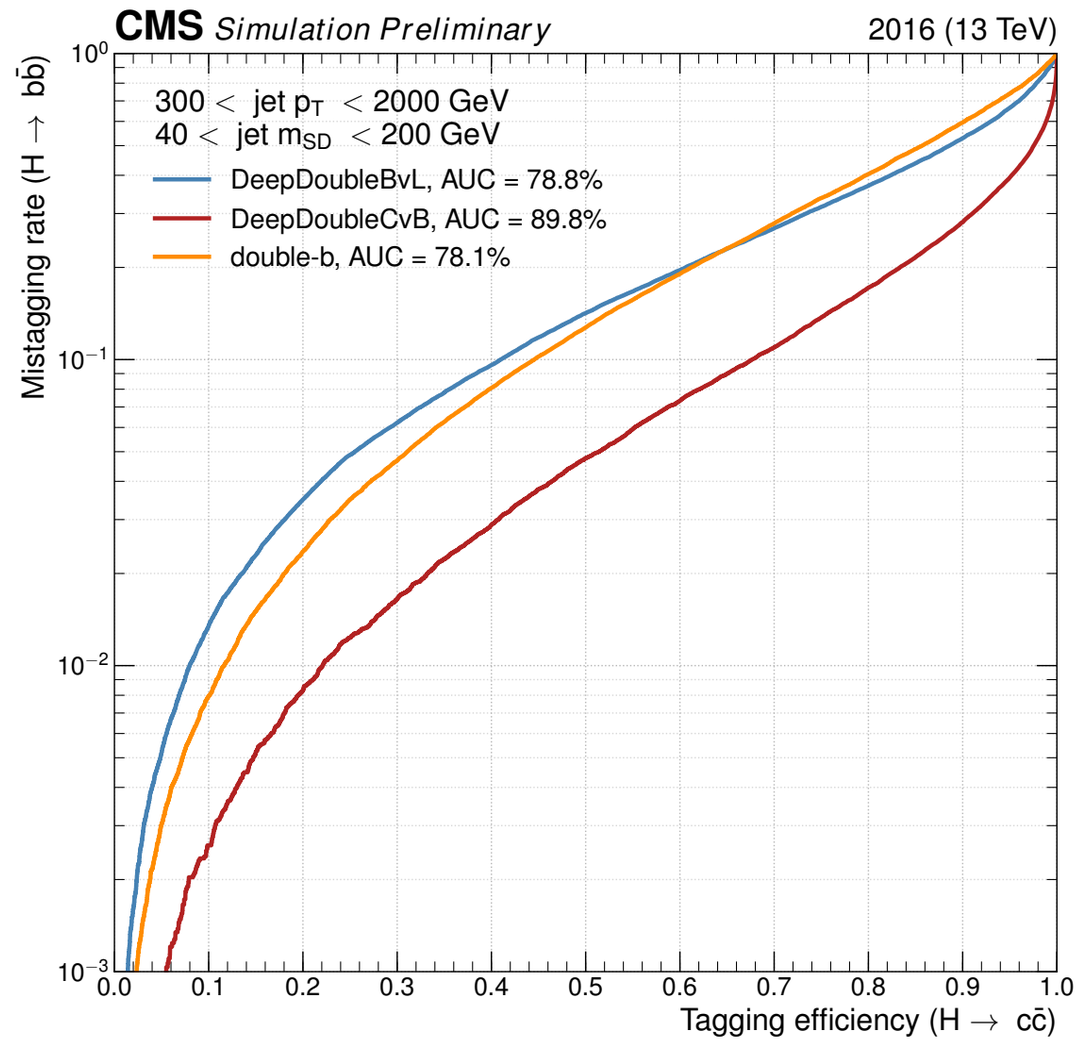


DP-2018/033

DEEP DOUBLE-C TAGGER

- Good performance for deep double-c for discriminating H(cc) from light jets and H(bb)!



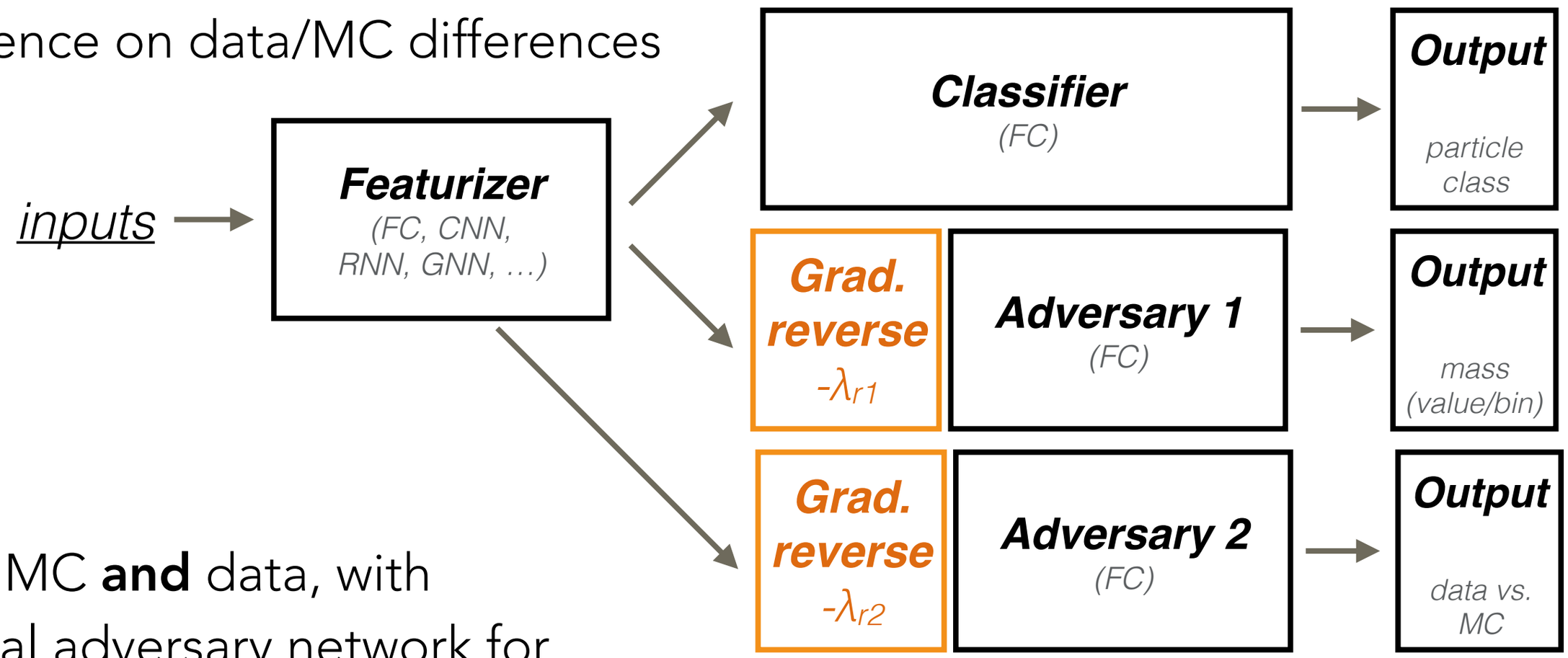


DATA/MC VALIDATION WITH ML

arXiv:1611.01046
arXiv:1409.7495

- Adversarial training to reduce dependence on data/MC differences

$$L = L_{\text{disc}} + \lambda_1 L_{\text{adv1}} + \lambda_2 L_{\text{adv2}}$$



- Train on MC **and** data, with additional adversary network for predicting whether it's data or MC

- Preliminary results **improve** Data/MC agreement

HEAVY FLAVOR TAGGING FOR BOOSTED RESONANCES

SUMMARY AND OUTLOOK

Javier Duarte
Fermilab

SUMMARY AND OUTLOOK

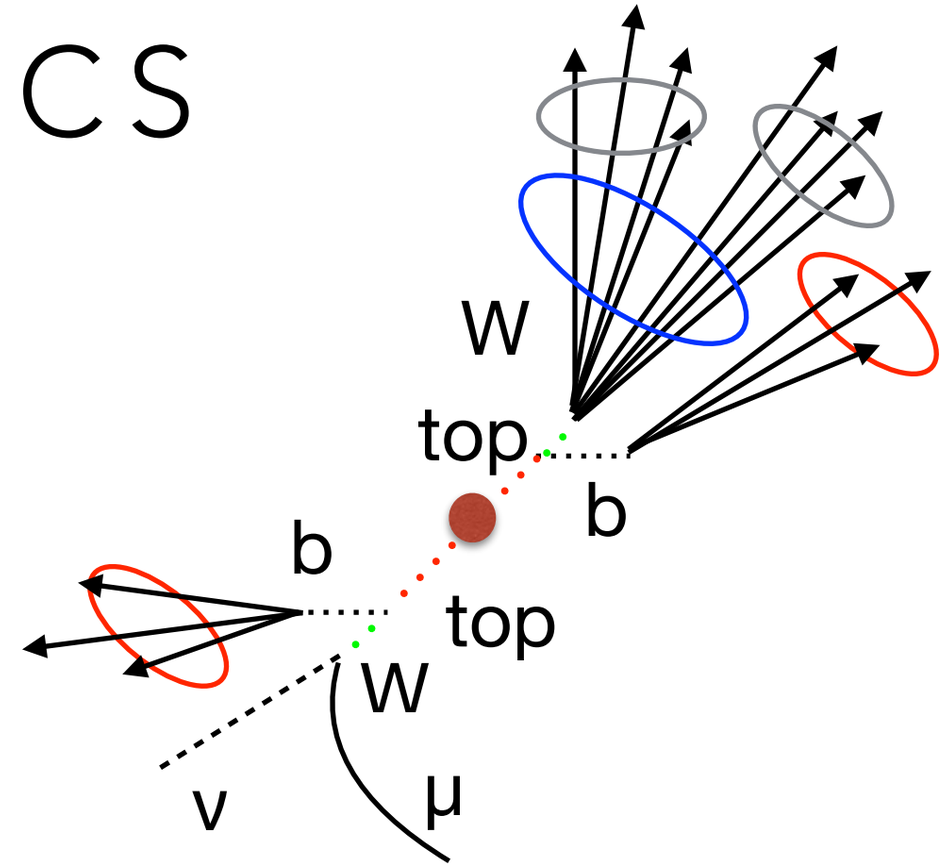
- Identifying boosted resonances is essential for the success of the LHC physics program (Higgs measurements, searches for new physics, etc…)
- Particle-level deep neural networks have improved tagger performance a great deal
  - Different approaches in CMS prioritize different goals; validation in data on-going
- Still room for improvement with ***physics-motivated*** architectures
- Many analyses will benefit from these developments!

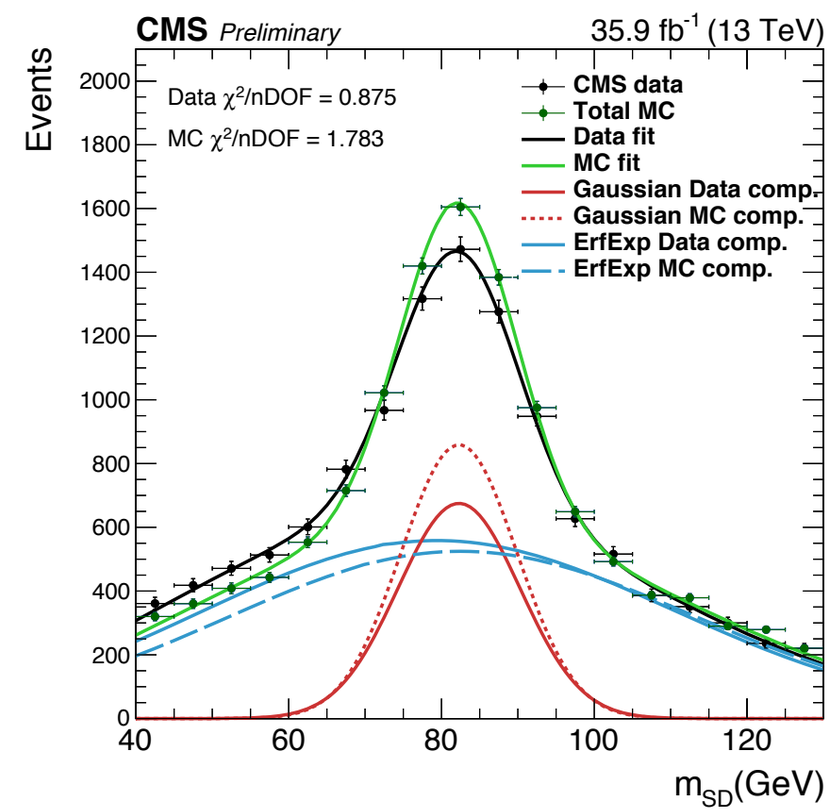
BACKUP

Javier Duarte
Fermilab

SYSTEMATICS

Systematic uncertainty source	Type (shape or normalization)	Relative size (or description)
QCD transfer factor	both	profile $a_{k\ell}$ and QCD normalization
Luminosity	normalization	2.5%
V-tag ($N_2^{1,DDT}$) efficiency	normalization	4.3%
Muon veto efficiency	normalization	0.5%
Electron veto efficiency	normalization	0.5%
Trigger efficiency	normalization	4%
Muon ID efficiency	shape	up to 0.2%
Muon isolation efficiency	shape	up to 0.1%
Muon trigger efficiency	shape	up to 8%
$t\bar{t}$ normalization SF	normalization	from 1μ CR: 8%
$t\bar{t}$ double-b mis-tag SF	normalization	from 1μ CR: 15%
W/Z NLO QCD corrections	normalization	10%
W/Z NLO EWK corrections	normalization	15% – 35%
W/Z NLO EWK ratio decorrelation	normalization	5% – 15%
double-b tagging efficiency	normalization	4%
Jet energy scale	normalization	up to 10%
Jet energy resolution	normalization	up to 15%
Jet mass scale	shape	shift m_{SD} peak by $\pm 0.4\%$
Jet mass resolution	shape	smear m_{SD} distribution by $\pm 9\%$
Jet mass scale p_T	normalization	0.4%/100 GeV (p_T)
Monte Carlo statistics	normalization	-
H p_T correction (gluon fusion)	both	30%

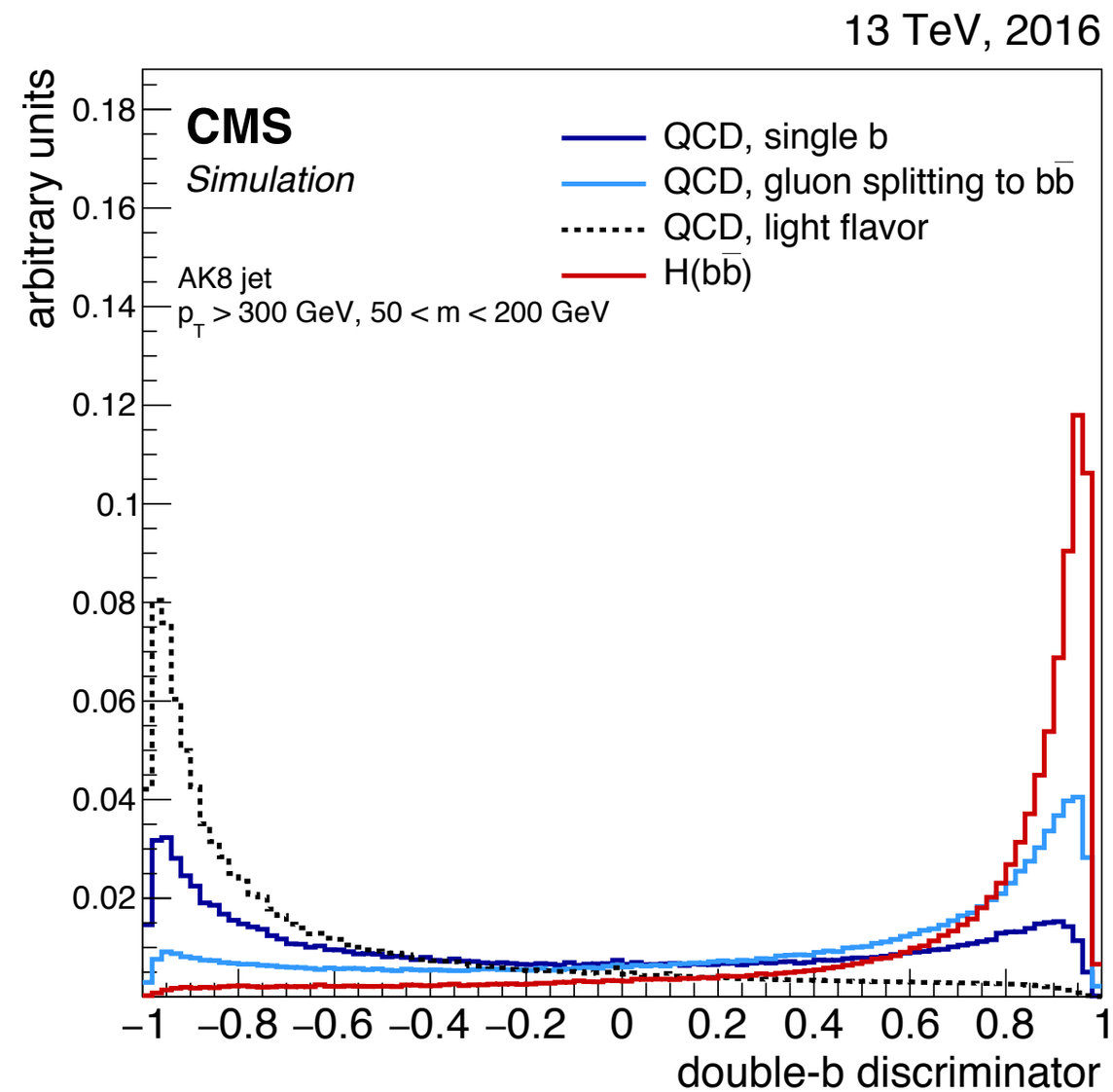




- Signal systematic uncertainties from merged W sample in semi-leptonic ttbar events (external constraint)
- SM candles: presence of W/Z(bb) in final jet mass distribution provides in-situ constraint
- Higgs p_T correction uncertainty of 30%

DOUBLE-B TAGGER

- Combines tracking and vertexing information in a multivariate classifier with 27 observables
- Targets the bb signal with additional aims:
  - jet mass and p_T independent
  - cover a very wide p_T range
  - inputs are chosen to avoid p_T correlation
    - e.g. no ΔR-like variables, no substructure info

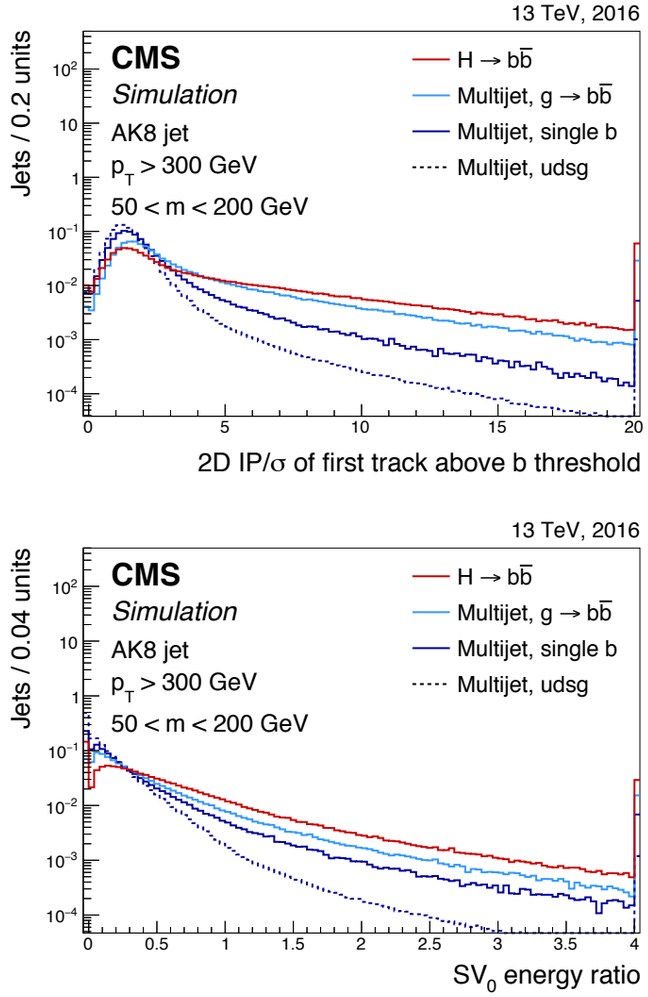


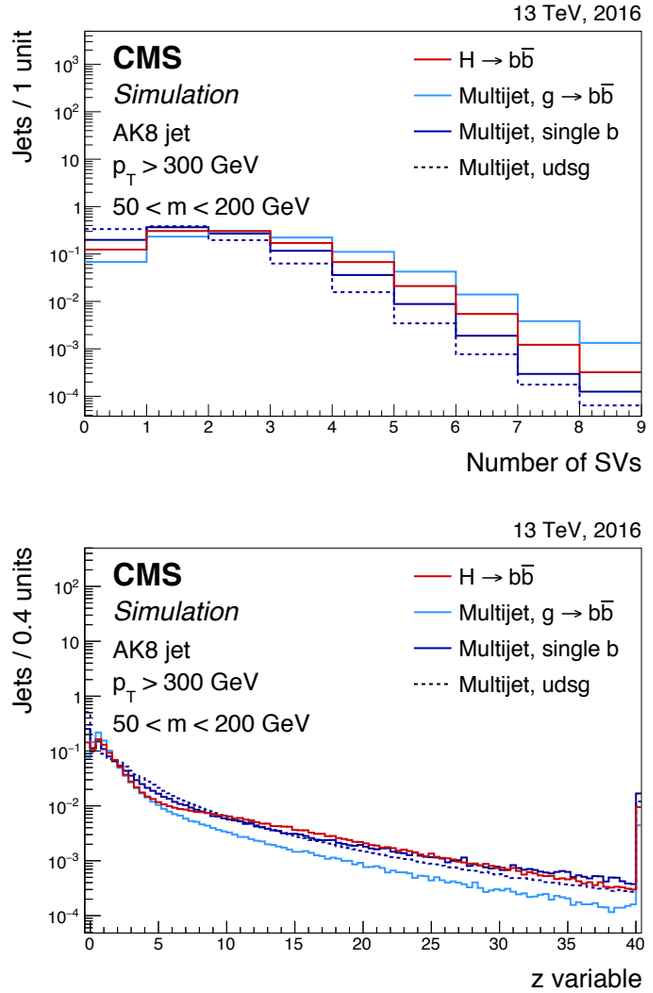
DOUBLE-B TAGGER INPUTS

- 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.





SAMPLE SIZES

- Train
  - H(bb) - 9 M
  - H(cc) - 10 M
  - QCD - 80 M
- Test
  - H(bb) - 600 k
  - H(cc) - 600 k
  - Z(qq) - 800 k
  - QCD - 1 M