Heavy flavour tagging

CMS POS 2017

Why?

QCD:

“All jets are created equal…”

Why?

QCD:

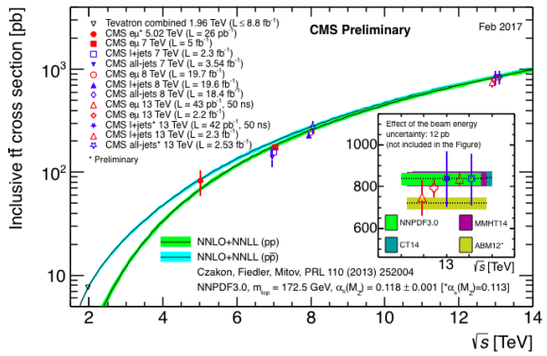
“All jets are created equal…”

EWK:

“…but some jets are more equal than others.”

Why? - from our spokesman LHCP highlights

Top Pair Cross Sections



CMS: 835 ± 33 pb
Theory: 816 ± 42 pb

Top pair rate is > 10 Hz, enabling us to address much more precise questions

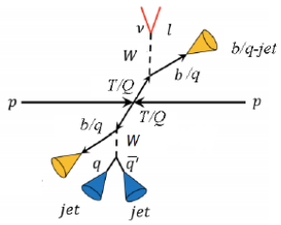
- Single and double differential cross sections
- Rare (FCNC) decays
- CP violation (a beginning)
- Width and more complex methods for measuring the mass

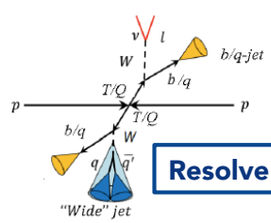
Top pair production at 13 TeV CM energy is mainly (80%) produced by gluons, providing important information on the gluon distribution at relatively high x, up to ~0.25

Factory	Quark	Cross Section (nb)	Luminosity ($cm^{-2}s^{-1}$)
B (KEKb)	Bottom	1.15 (Y(4S))	2.11×10^{34}
LHC	Top	0.82 (incl t-t)	1.51×10^{34}

Boosted Objects, e.g., Vector Like Quarks

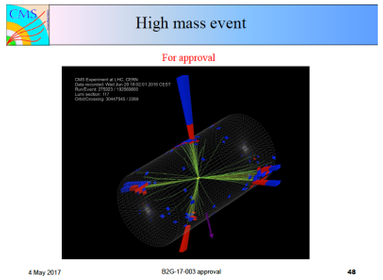
$$T\bar{T} \rightarrow bW^{+}\bar{b}W^{-} \rightarrow b\ell\nu\bar{b}q\bar{q}'$$





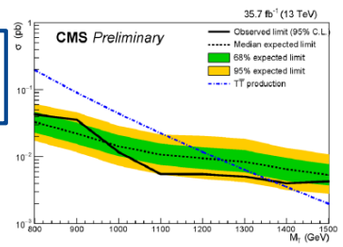
B2G-17-003

Resolve into two subjets



AK8 jets

Limit:
Expected: 1245 GeV
Observed: 1365 GeV

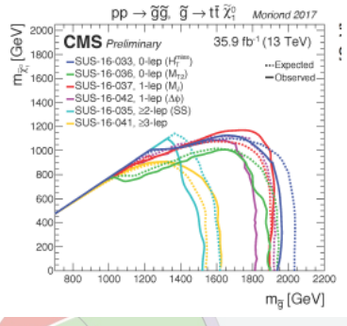


SUSY Searches

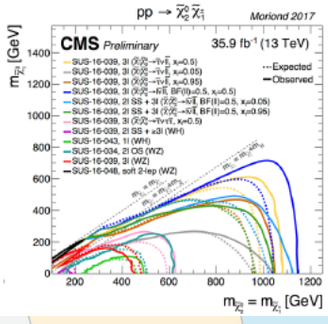
Broad program: 19 searches completed with full 2016 CMS dataset, with several already submitted to journals

- Probing different models (inclusive production, strong and electroweak production, and 3rd generation sparticles (stops)
- Different final states(with leptons, photons, jets) and analysis techniques

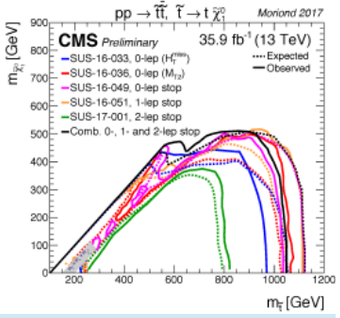
gluinos (inclusive)



electroweakinos



stops (3rd gen)



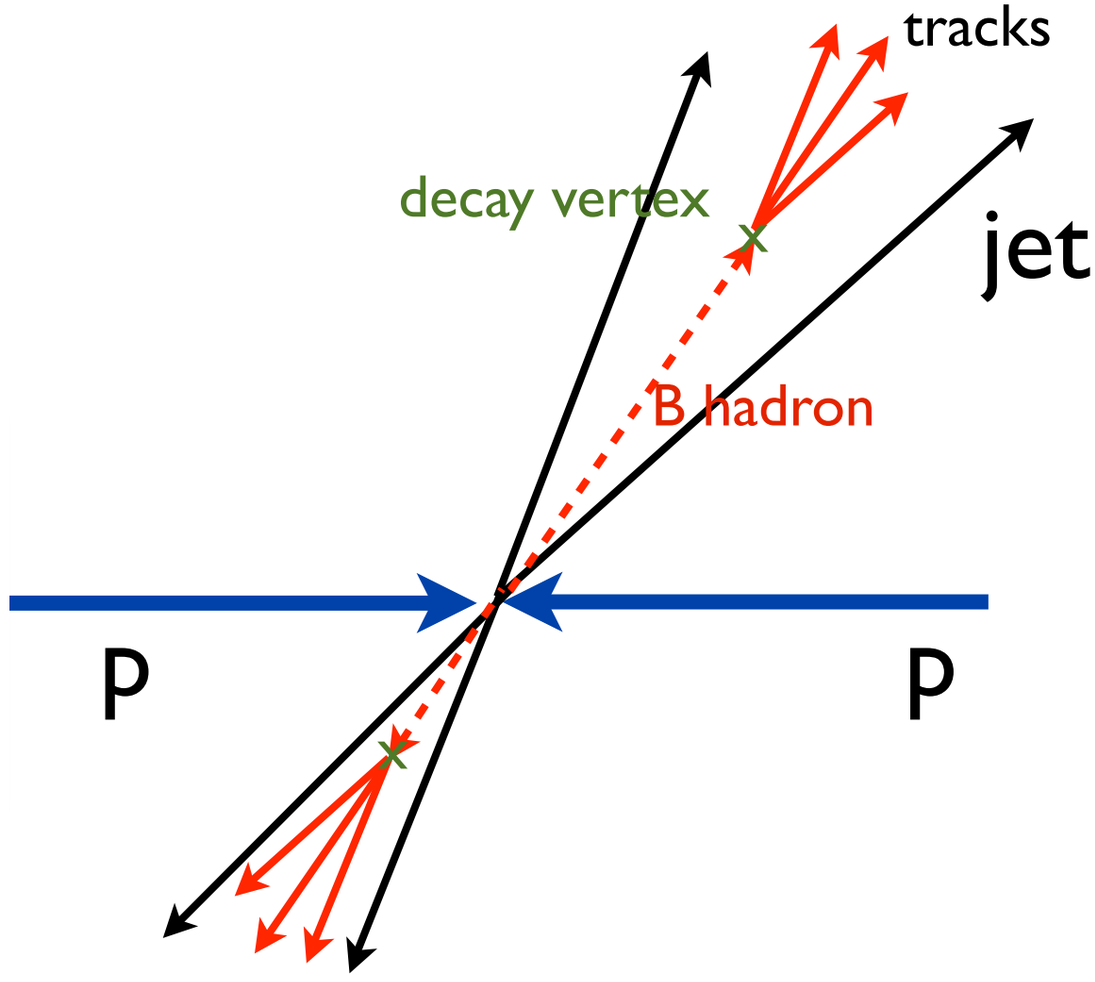
… and many more

How?

At LHC energies b-quarks hadronise and produce **jets** containing **B-hadrons** (B^+, B^-, B^0, B_s, Λ_B)

Their lifetime (~1.5ps) and Lorentz boost lead to **displaced decay vertices**

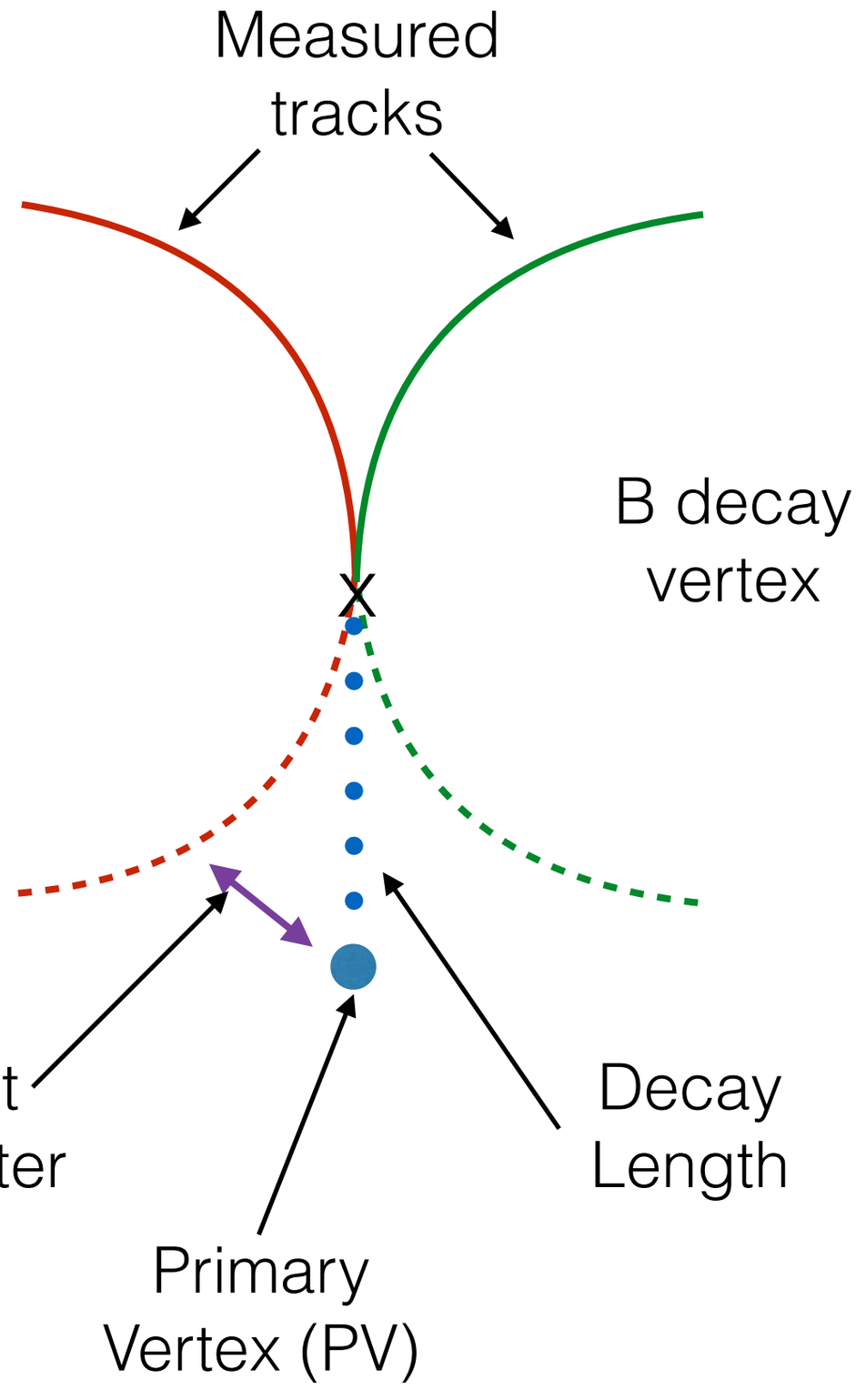
Look for displaced tracks and secondary vertices within the jet



How? - Impact Parameter

Defined as the distance between as the track and the **P**rimary **V**ertex at their point of closest approach

In the relativistic limit IP is ~ Lorentz invariant.
Decay length is not



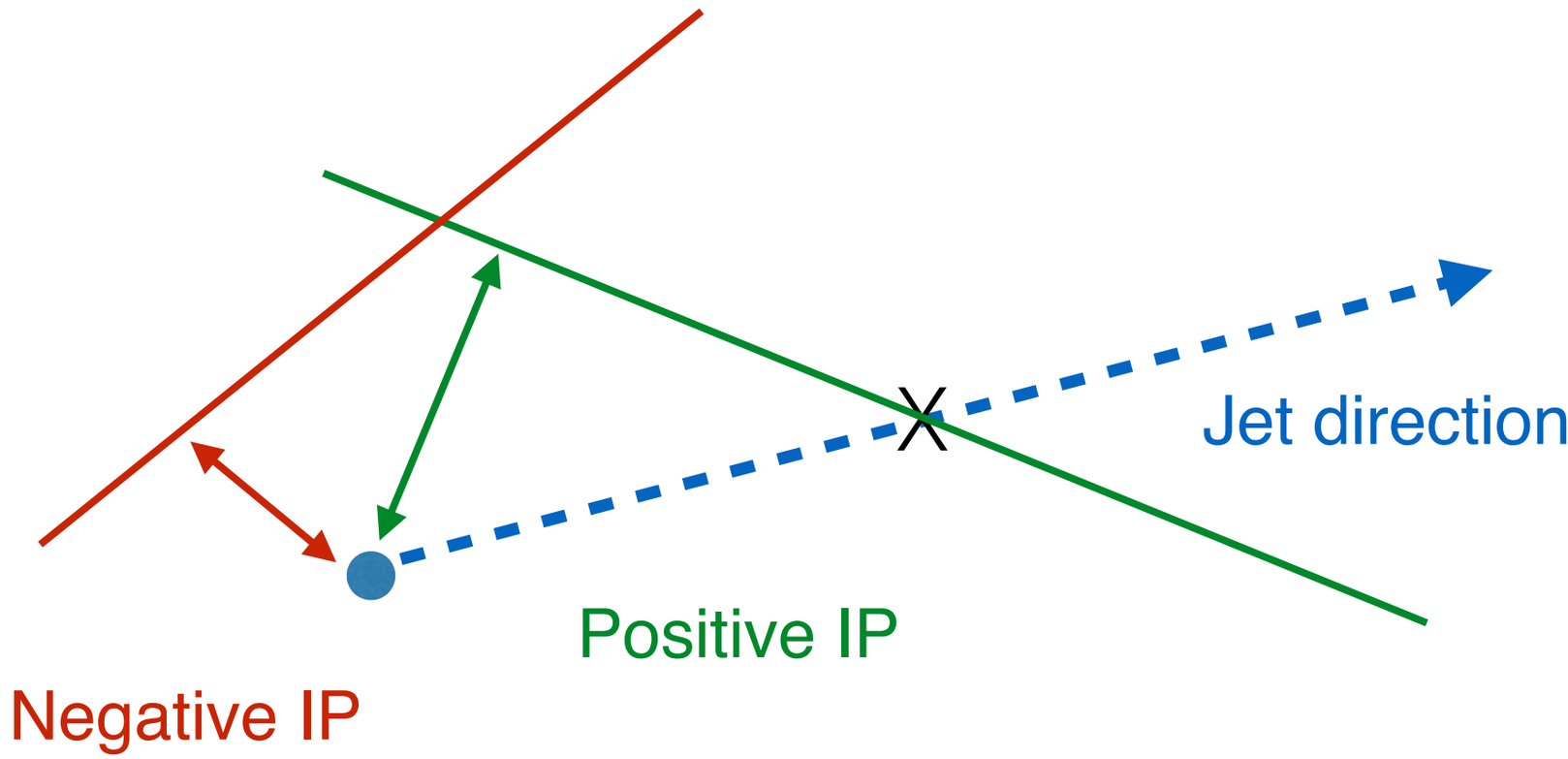
How? - Impact Parameter

The scalar product of the IP segment and the jet direction determines the sign

In a perfect world:

- the IP distribution of *light-flavour* jets would be perfectly symmetric around 0 (and perfectly gaussian, because of various effects entering)
- the distribution of *b-jets* would be mostly positive

In reality: light jets are slightly asymmetric and b-jets have negative IPs as well



How? - Impact Parameter

The scalar product of the IP segment and the jet direction determines the sign

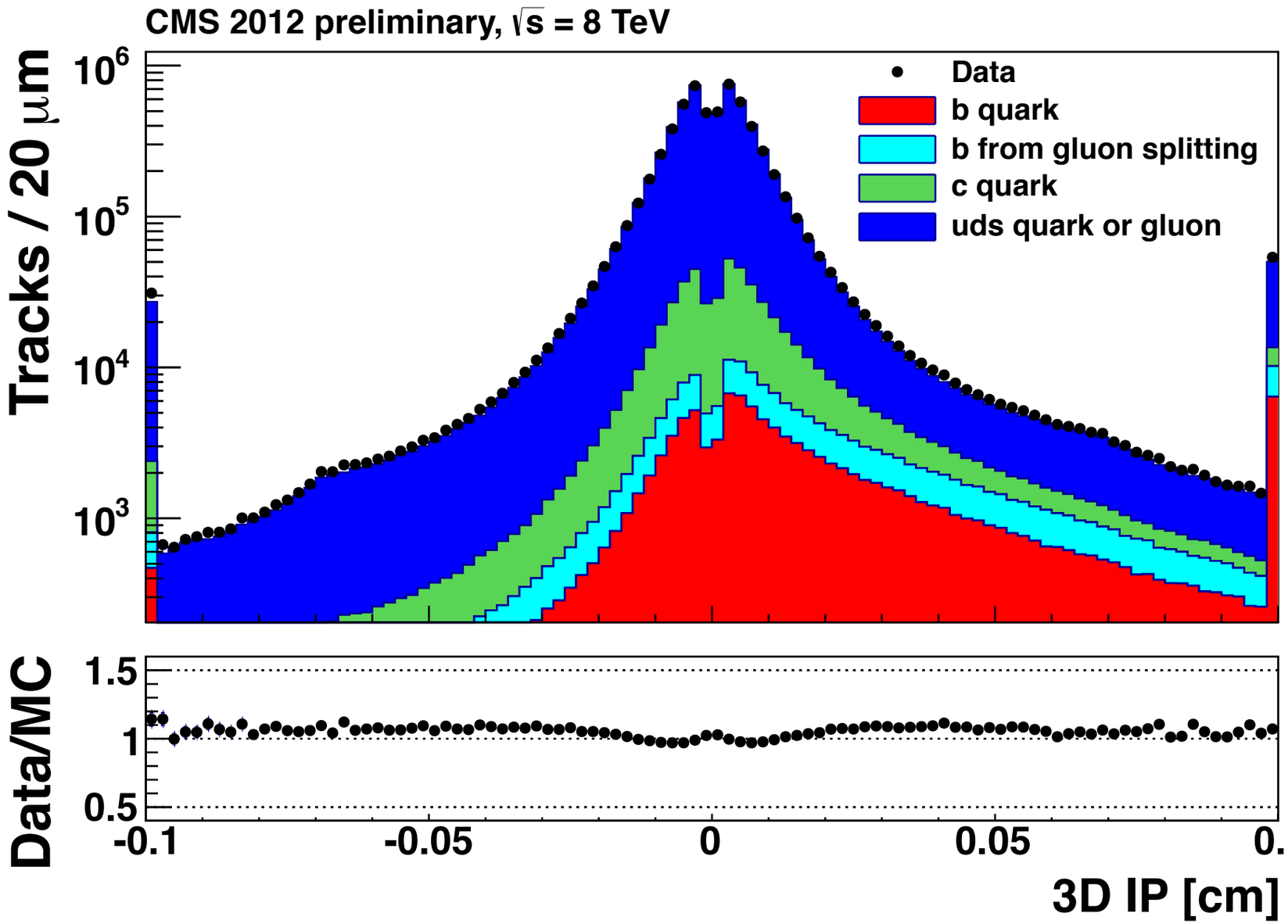
In a perfe

- the IP c
  around
  enterin
- the dist

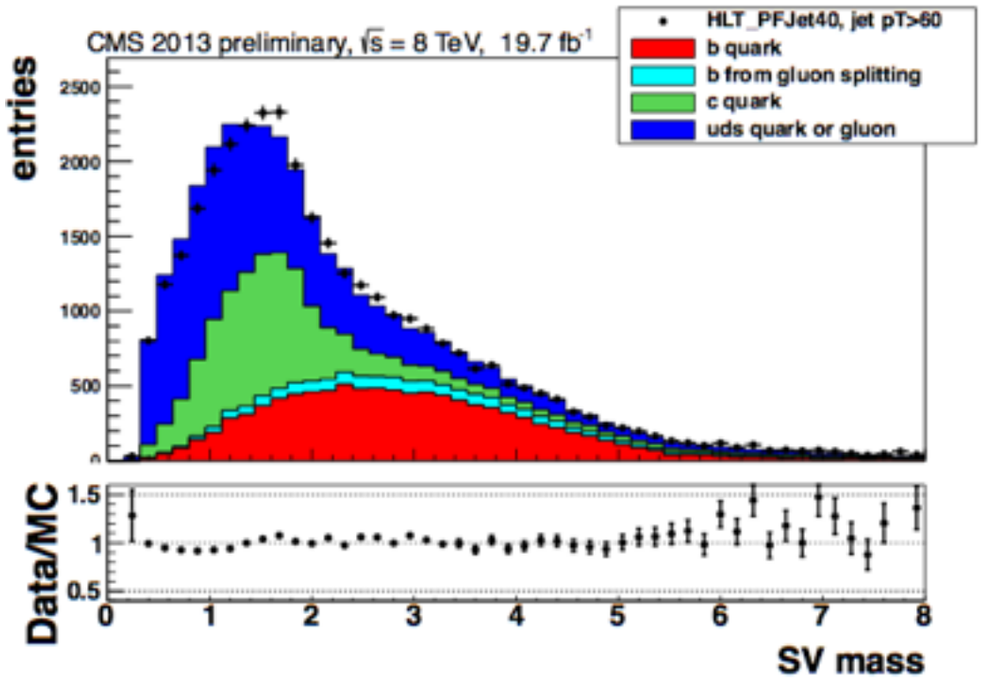
In reality:
as well

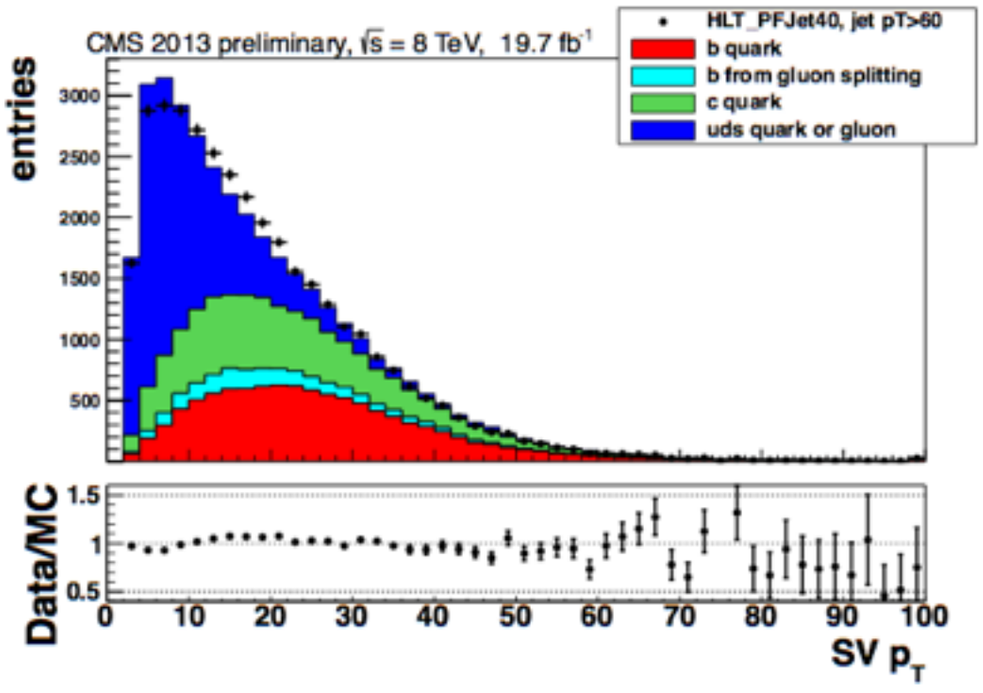
netric

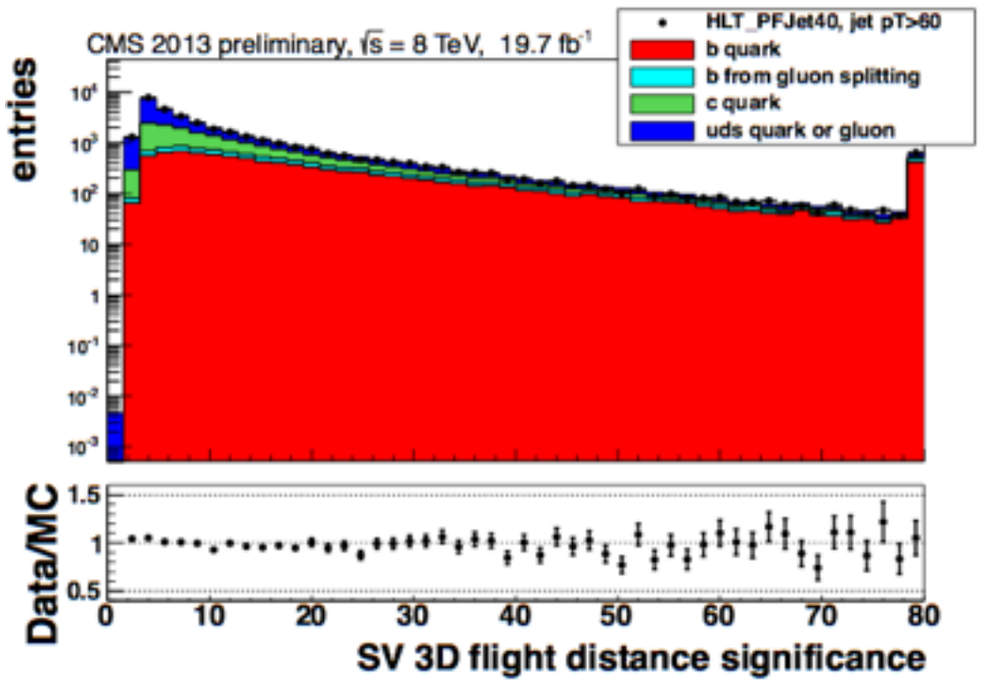
ative IPs



How? - Secondary Vertices



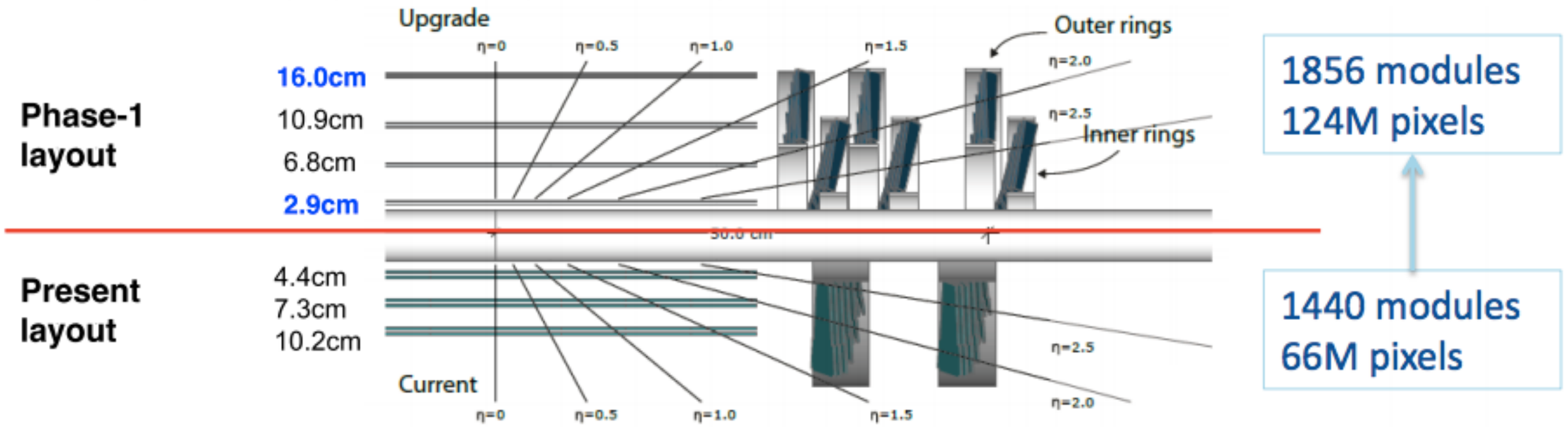




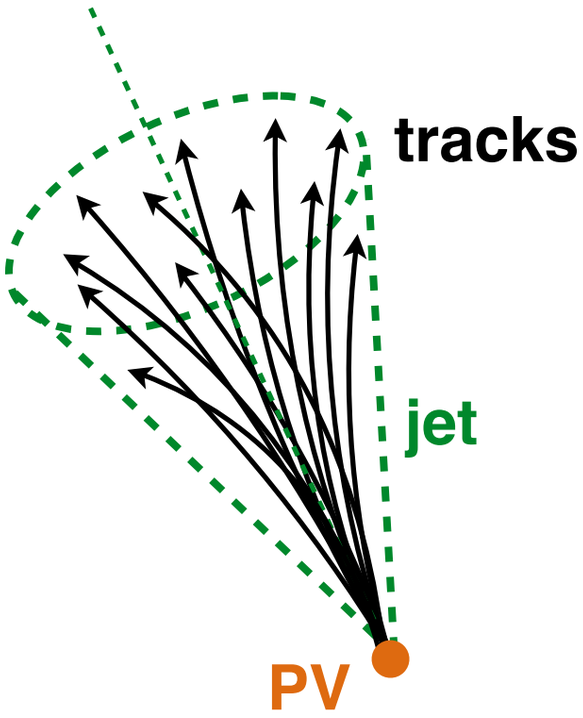
… and many more

How? - Tracker is very important to us!

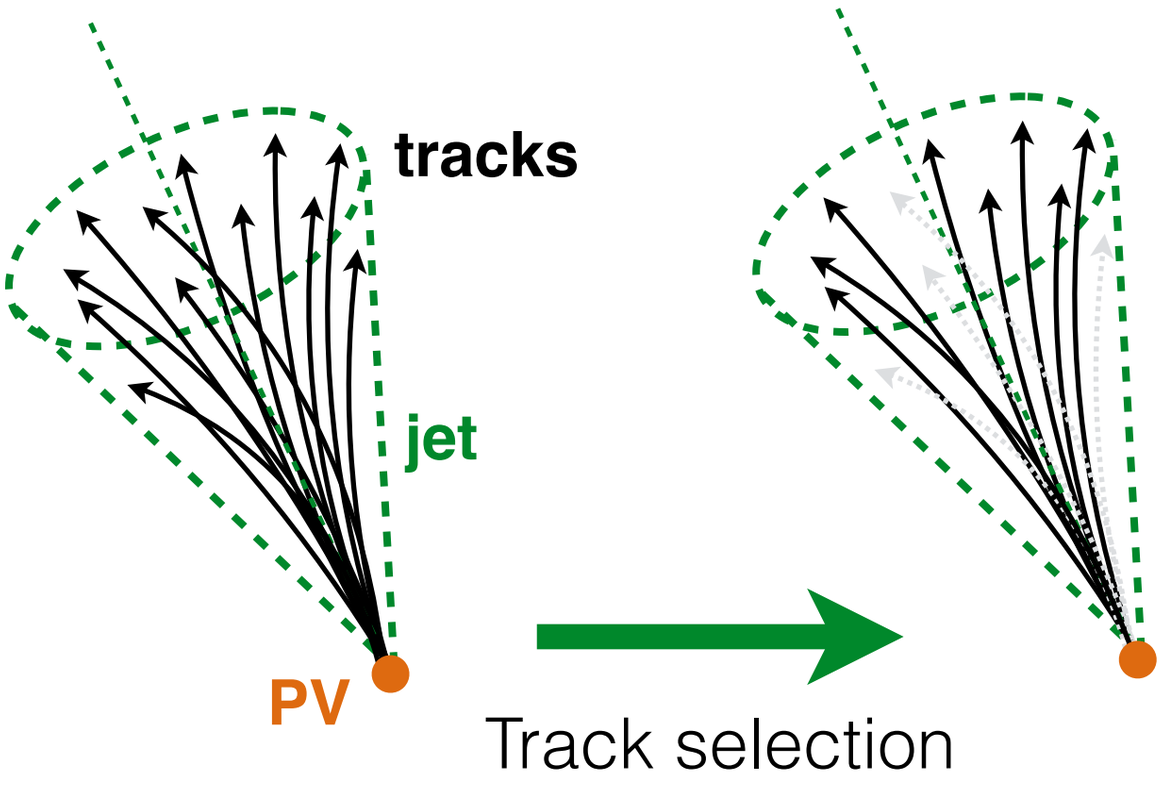
- Move from 3- to 4-hit coverage to increase redundancy and track finding efficiency
- Sensor technology, pixel size and module concept very similar to fit into existing infrastructure
- Move from analog to digital readout chip (ROC)
  - reduced buffer overflow and inefficiency
- **Barrel:** 1184 modules, 48M ⇛ 79M pixels
- **Forward:** 672 modules, 18M ⇛ 45M pixels



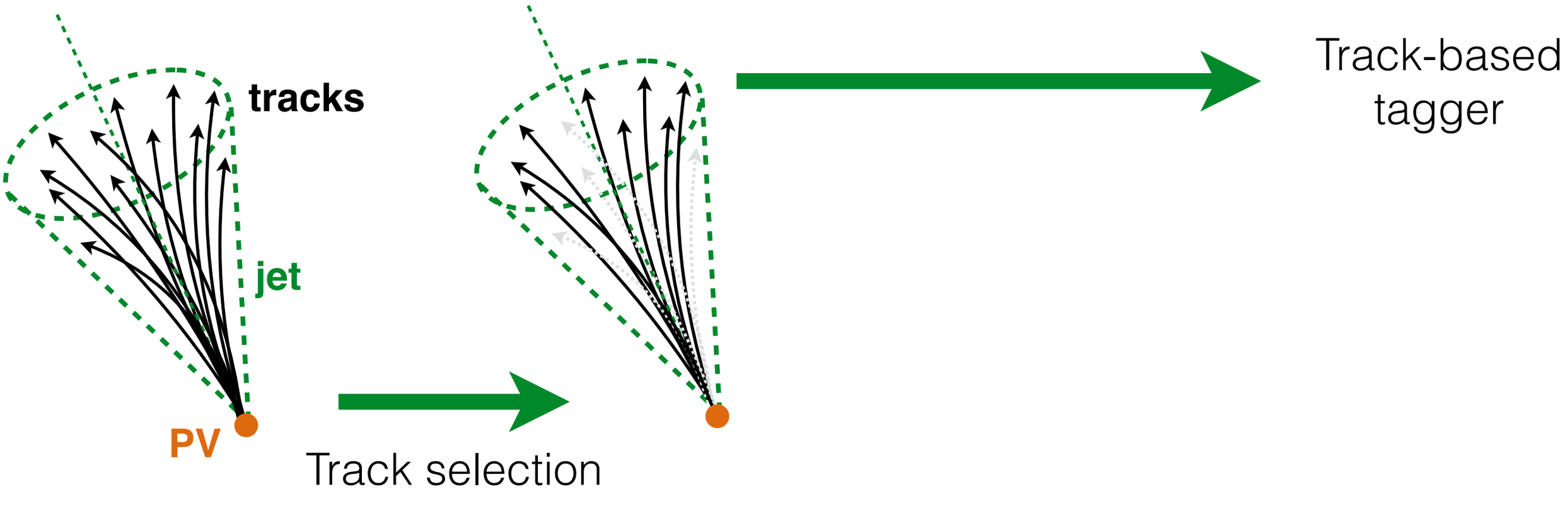
The b-tagging workflow



The b-tagging workflow

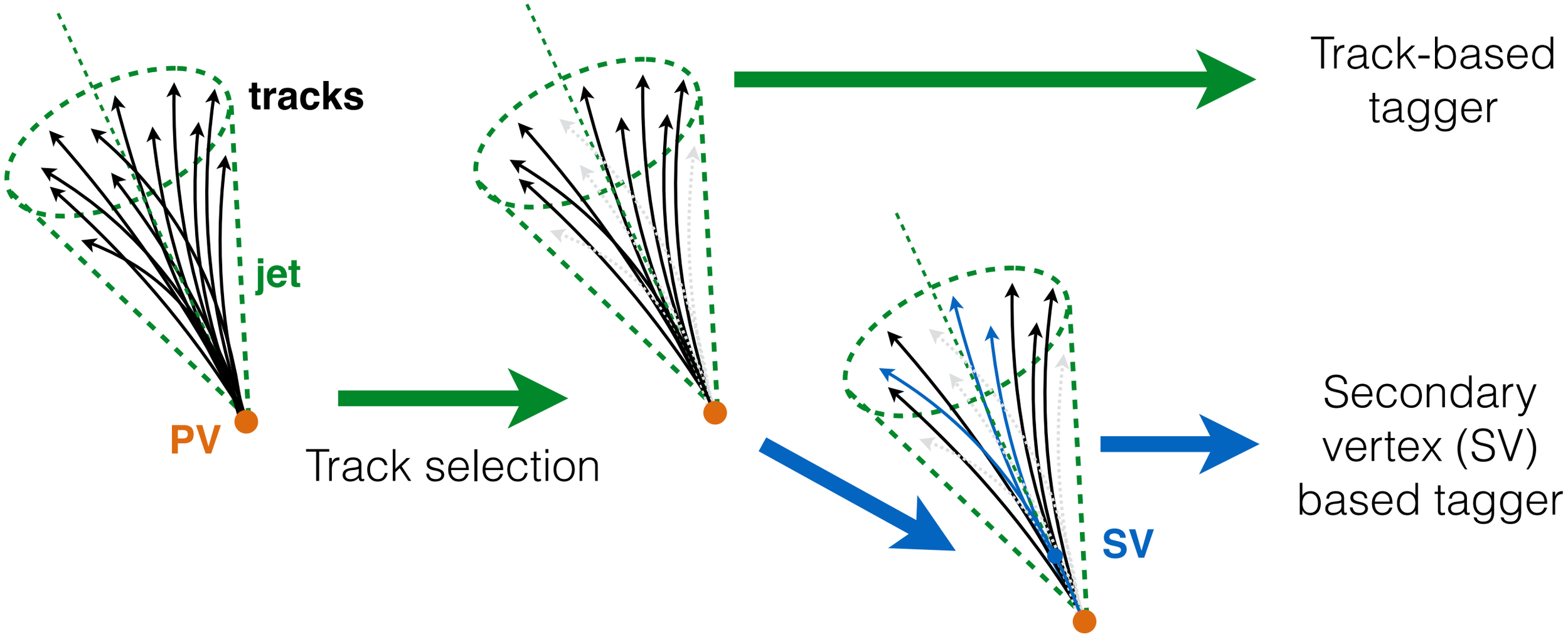


The b-tagging workflow



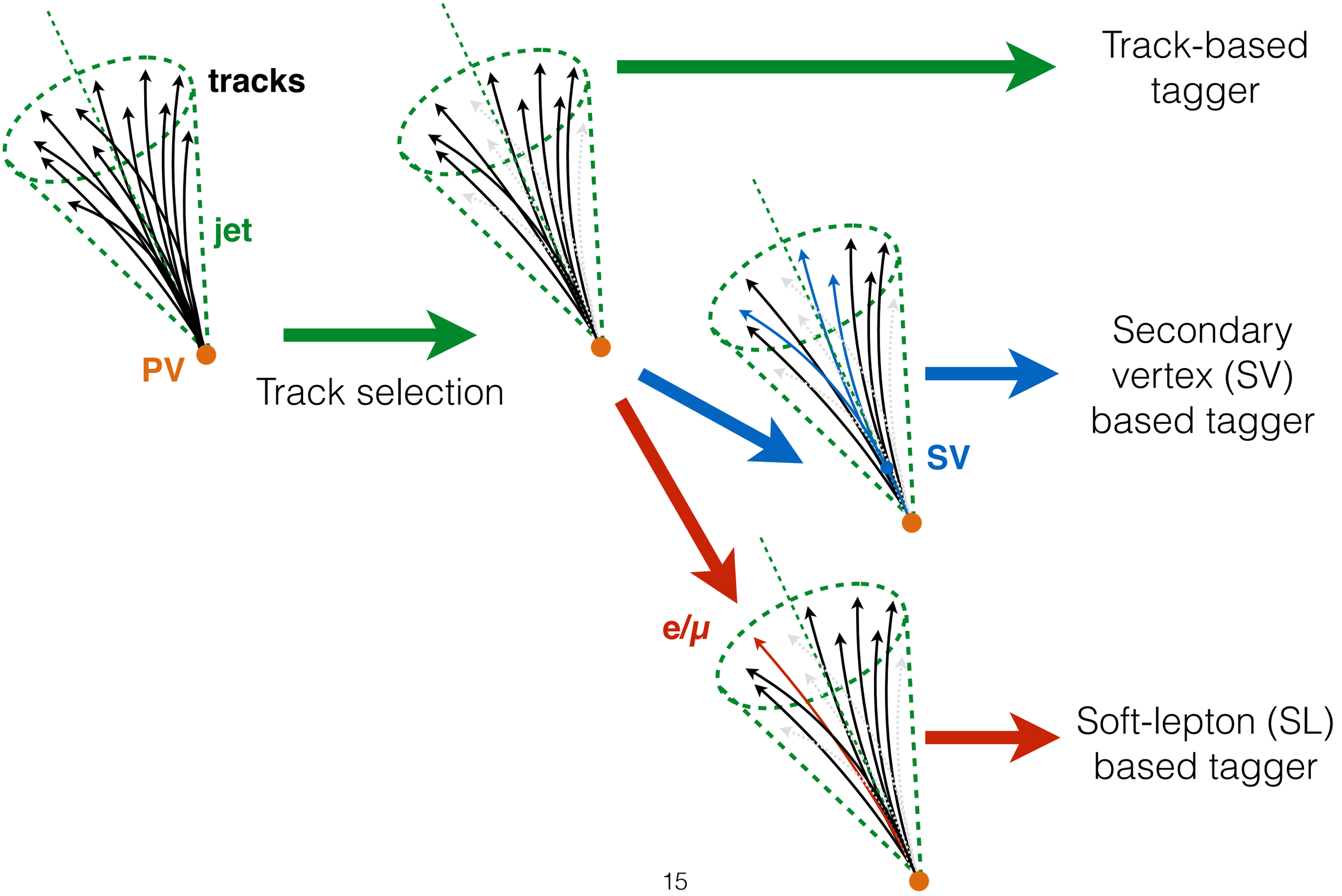
e.g.: **J**et **P**robability (**JP**)

The b-tagging workflow

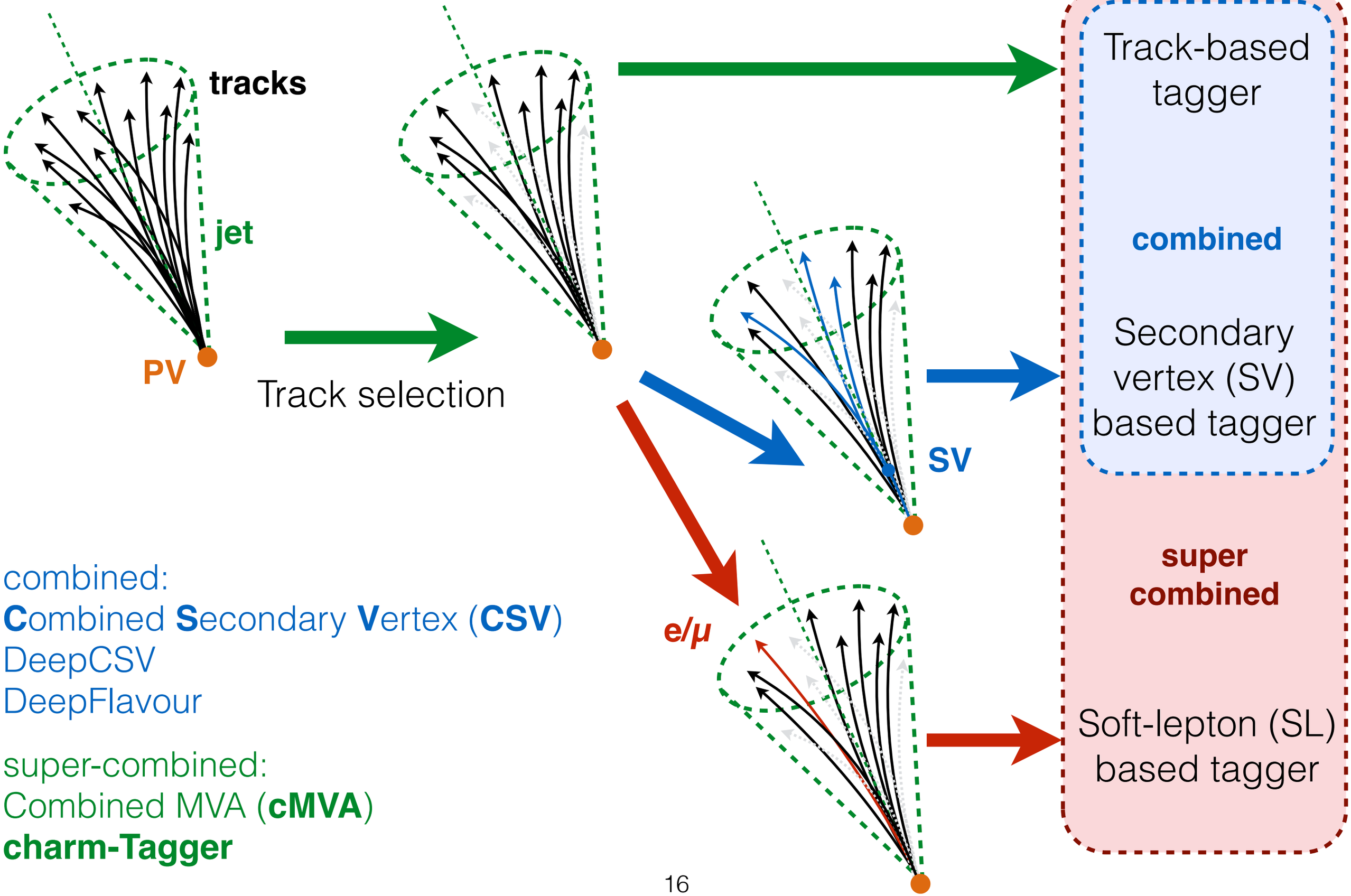


- **A**daptive **V**ertex **R**econstruction **(AVR)**: applied on tracks associated to the jet
- **I**nclusive **V**ertex **F**itter **(IVF)**: on the full set of tracks recorded in the event (SV ΔR-matched to jet).
  Current reconstruction default

The b-tagging workflow



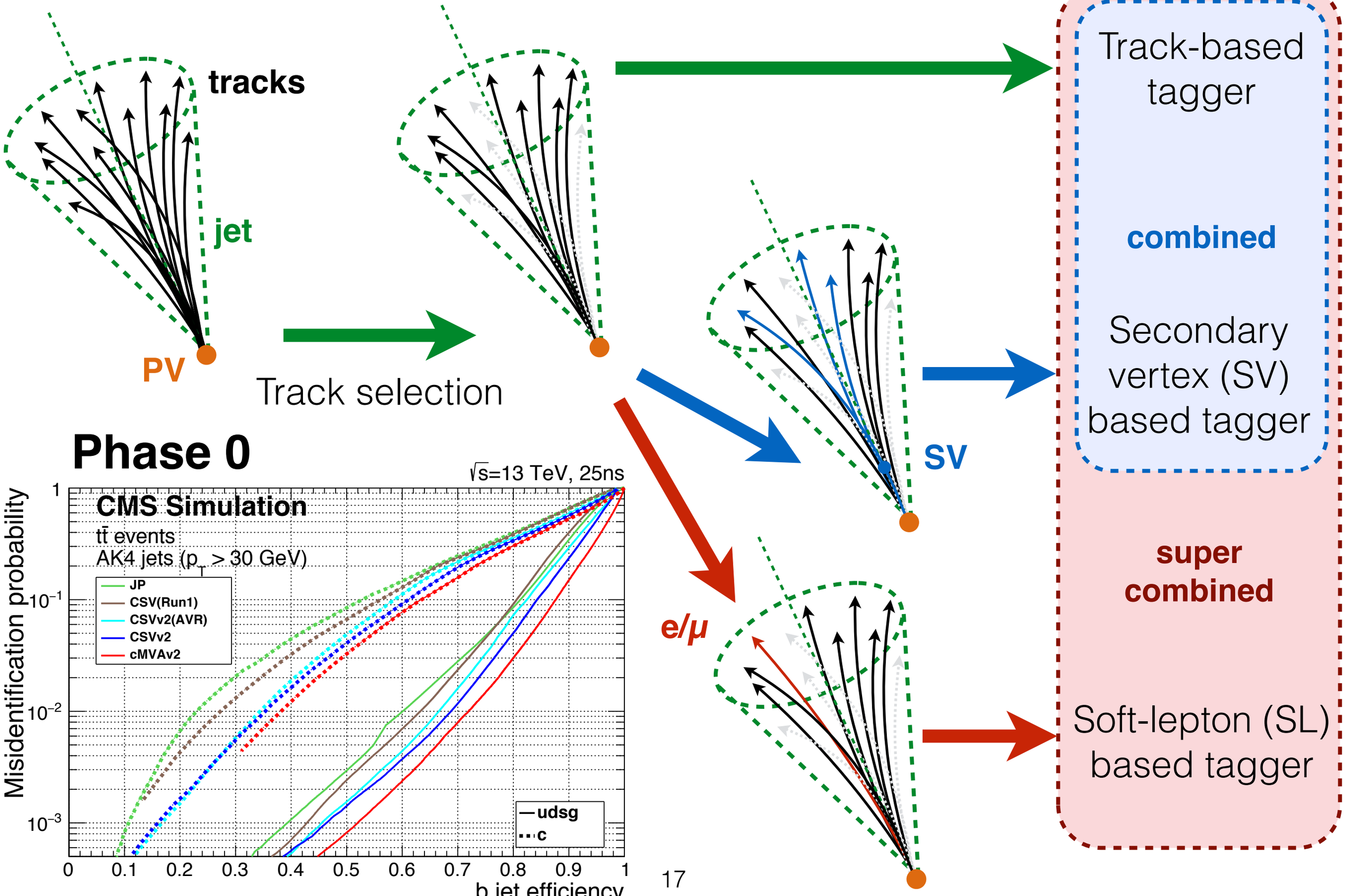
The b-tagging workflow



combined:
Combined **S**econdary **V**ertex (**CSV**)
DeepCSV
DeepFlavour

super-combined:
Combined MVA (**cMVA**)
charm-Tagger

The b-tagging workflow



Phase 0

The b-tagging workflow

tracks

jet

PV

Track selection

SV

e/μ

Track-based tagger

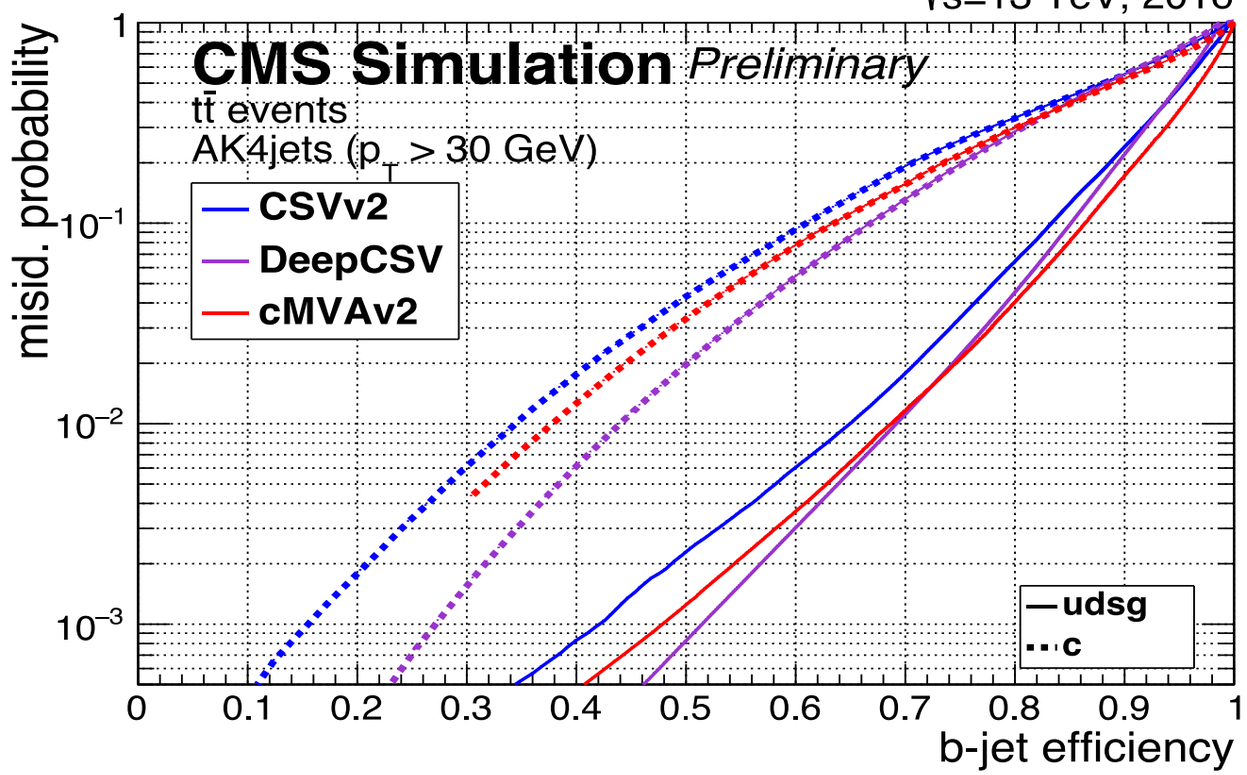
combined

Secondary vertex (SV) based tagger

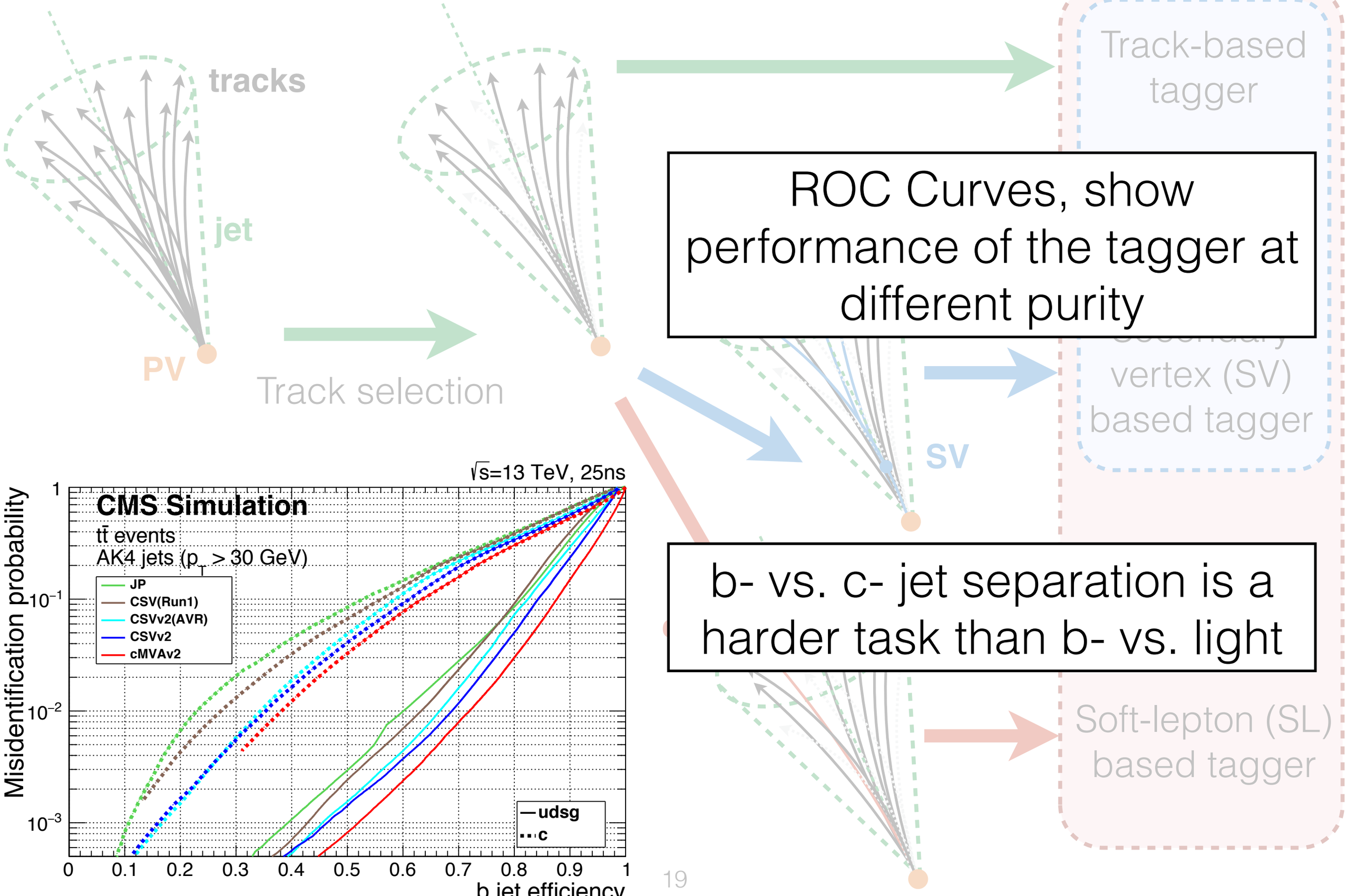
super combined

Soft-lepton (SL) based tagger

Phase I



The b-tagging workflow



How? - How many input variables?

```
(['Cpfcan_BtagPf_trackEtaRel',
  'Cpfcan_BtagPf_trackPtRel',
  'Cpfcan_BtagPf_trackPPar',
  'Cpfcan_BtagPf_trackDeltaR',
  'Cpfcan_BtagPf_trackPParRatio',
  'Cpfcan_BtagPf_trackSip2dVal',
  'Cpfcan_BtagPf_trackSip2dSig',
  'Cpfcan_BtagPf_trackSip3dVal',
  'Cpfcan_BtagPf_trackSip3dSig',
  'Cpfcan_BtagPf_trackJetDistVal',
  'Cpfcan_BtagPf_trackJetDistSig',

  'Cpfcan_ptrel',
  'Cpfcan_drminsv',
  'Cpfcan_fromPV',
  'Cpfcan_VTX_ass',
  'Cpfcan_puppiw',
  'Cpfcan_chi2',
  'Cpfcan_quality'
   ],
  25)
```

```
(['Npfcan_ptrel',
  'Npfcan_deltaR',
  'Npfcan_isGamma',
  'Npfcan_HadFrac',
  'Npfcan_drminsv',
  'Npfcan_puppiw'
 ],
25)
```

```
(['sv_pt',
  'sv_deltaR',
  'sv_mass',
  'sv_ntracks',
  'sv_chi2',
  'sv_normchi2',
  'sv_dxy',
  'sv_dxysig',
  'sv_d3d',
  'sv_d3dsig',
  'sv_costhetasvpv',
  'sv_enratio',
],
4)
```

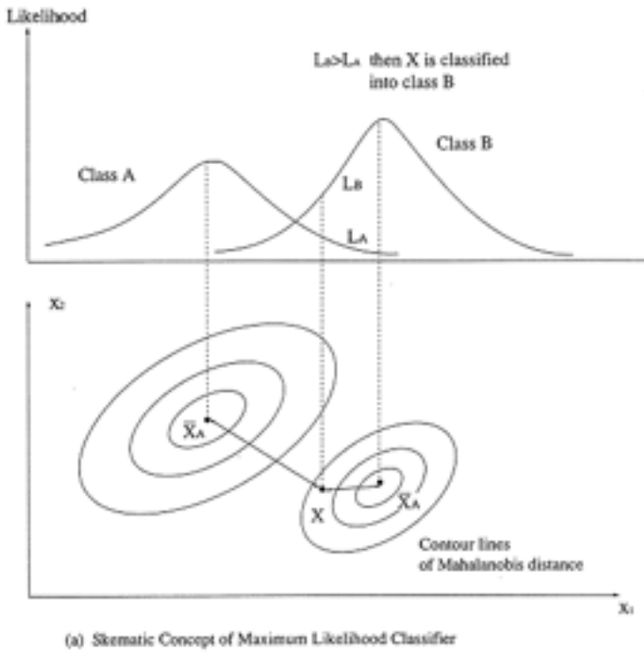
```
(['jet_pt', 'jet_eta',
  'nCpfcand','nNpfcand',
  'nsv','npv',
  'TagVarCSV_trackSumJetEtRatio',
  'TagVarCSV_trackSumJetDeltaR',
  'TagVarCSV_vertexCategory',
  'TagVarCSV_trackSip2dValAboveCharm',
  'TagVarCSV_trackSip2dSigAboveCharm',
  'TagVarCSV_trackSip3dValAboveCharm',
  'TagVarCSV_trackSip3dSigAboveCharm',
  'TagVarCSV_jetNSelectedTracks',
  'TagVarCSV_jetNTracksEtaRel'])
```

Potentially, a **LOT**!

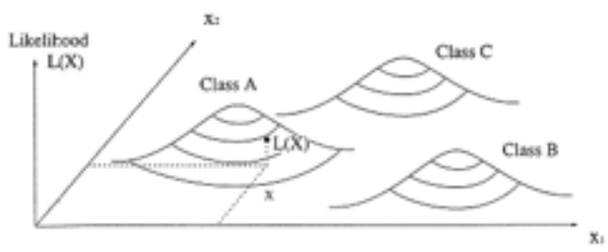
How? - Machine learning

Likelihood classifier

e.g. JP



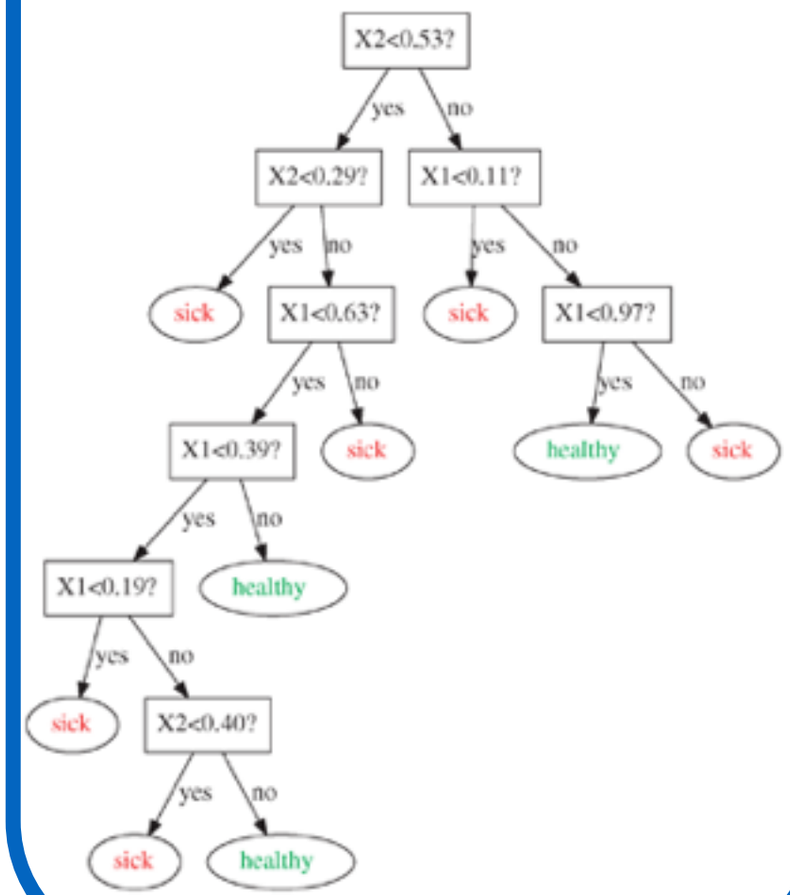
(a) Skematic Concept of Maximum Likelihood Classifier



(b) Likelihood Functions in 3D Expression

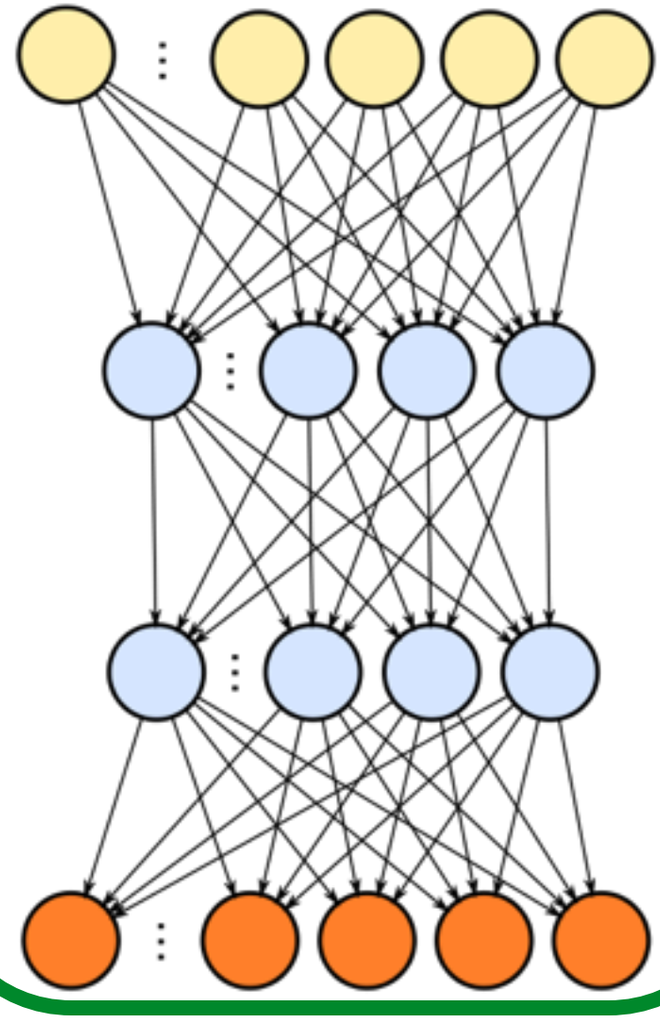
BDT

e.g. cMVA

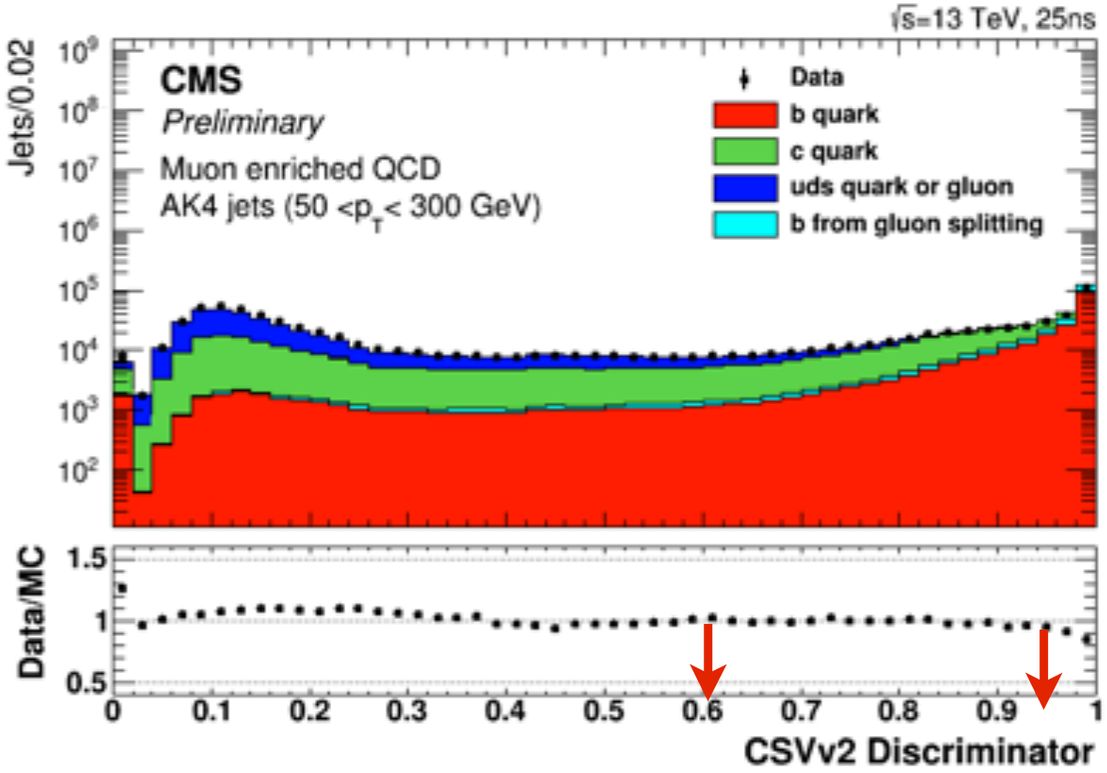


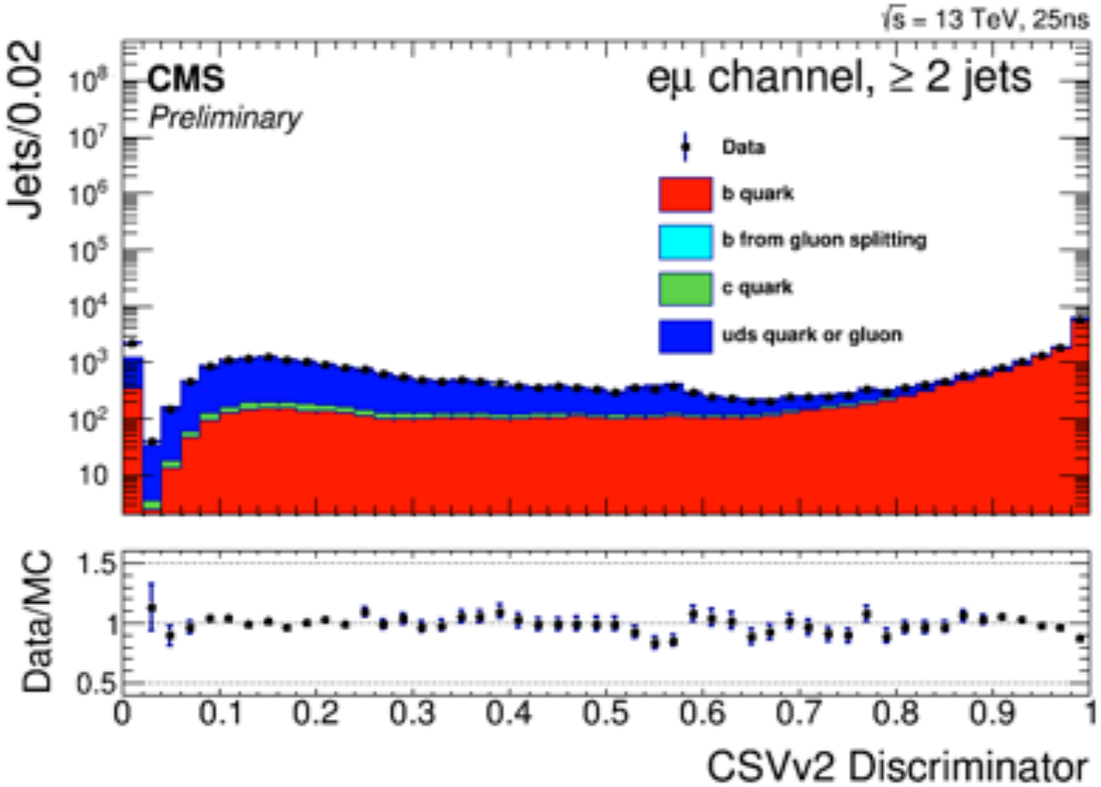
(Deep) Neural Networks

e.g. CSV, DeepCSV



Discriminators





define three operating points based on mistag probabilities (0.1%, 1% and 10%)

Tagger name in CMSSW	Tagger name	WP name	WP Discr cut
pfCombinedInclusiveSecondaryVertexV2BJetTags	CombinedSecondaryVertex v2	CSVv2L	0.5426
	CombinedSecondaryVertex v2	CSVv2M	0.8484
	CombinedSecondaryVertex v2	CSVv2T	0.9535

Dependence on kinematics

low-momentum:

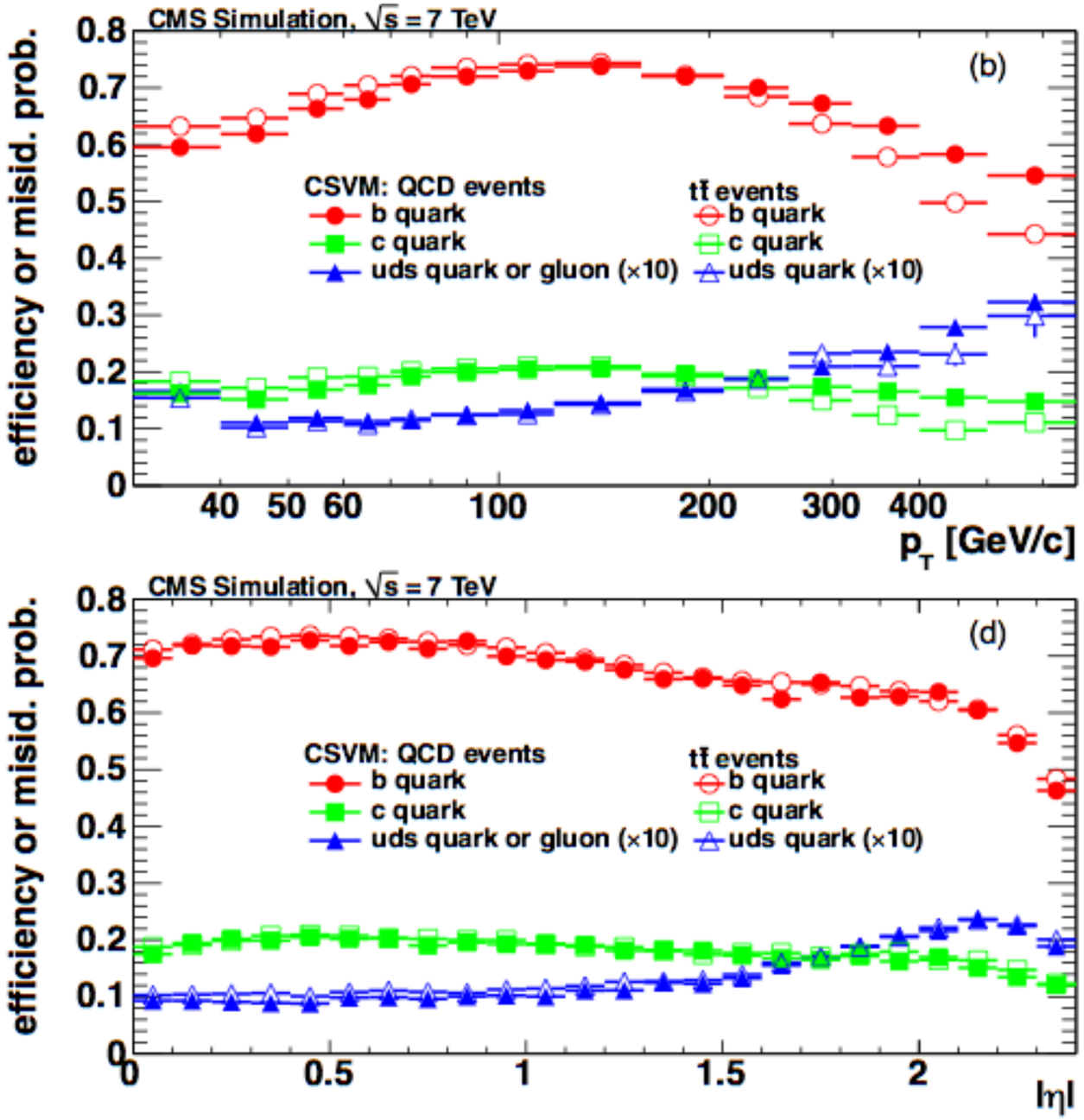
- more multiple coulomb scattering
- smaller Lorentz factor → B hadron decays closer to primary vertex (more difficult to resolve)

high-momentum:

- more collimated jets
- dense environment difficult for tracking
- smaller track curvature
- close-by-tracks create overlapping hits in the detector

forward region:

- edge effects
- detector layers further away from PV



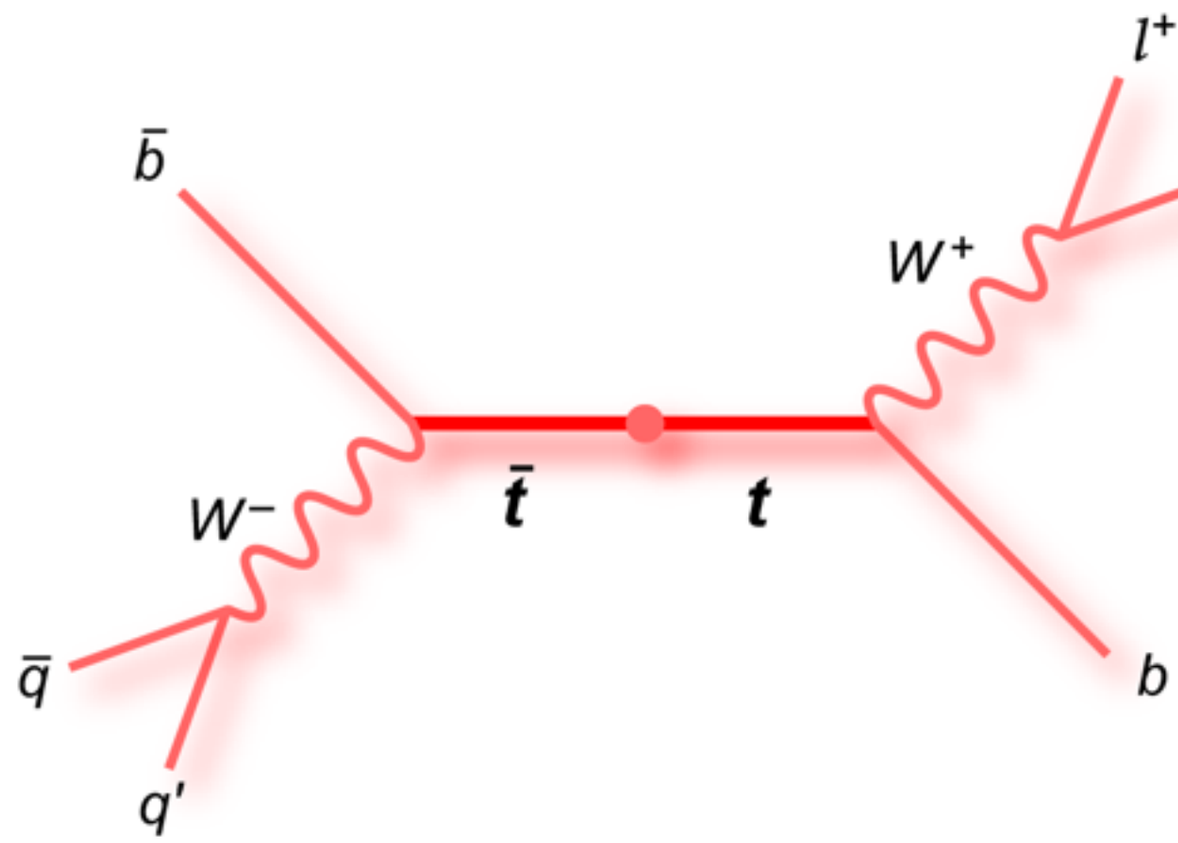
Performance measurements

events with muons in jets:

- use characteristic properties of the muon from B decays to measure (fit) the b-content
- relative momentum to the jet axis (ptrel) is large because of the B hadron mass
- possible to use IP of muon as well (although correlated with b-taggers)

events with top quarks:

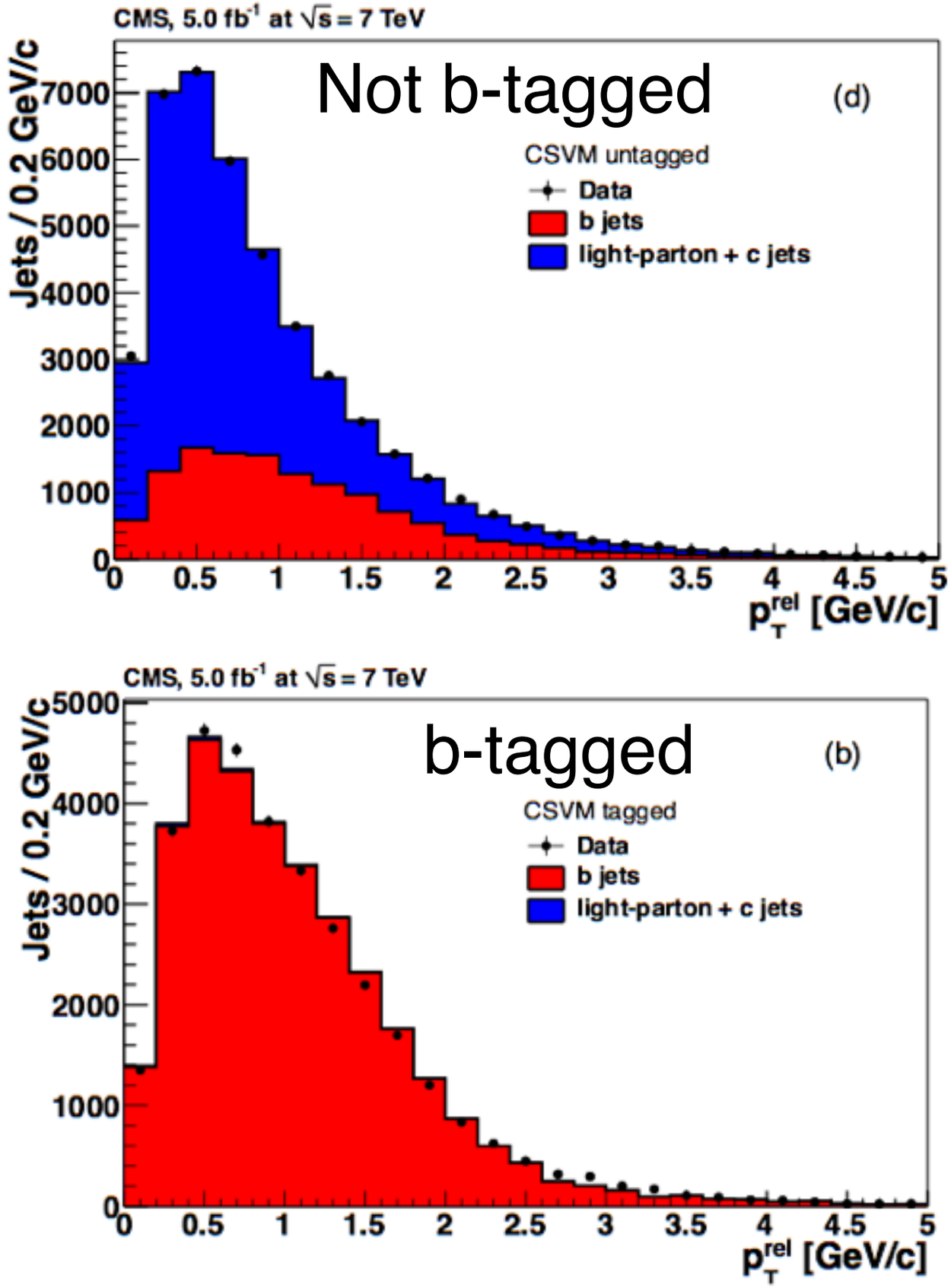
- select sample with a high fraction of top quark events
- the branching fractions of top quarks are well known
  - the flavour composition of the sample is known as well



Example, the p_T^{rel} method

- split sample with jets containing muons into two sets: b-tagged and not b-tagged
- fit the blue (light jets) and red (b-jet) shapes to match the data, extract the fractions f_b^{tag} and f_b^{untag}

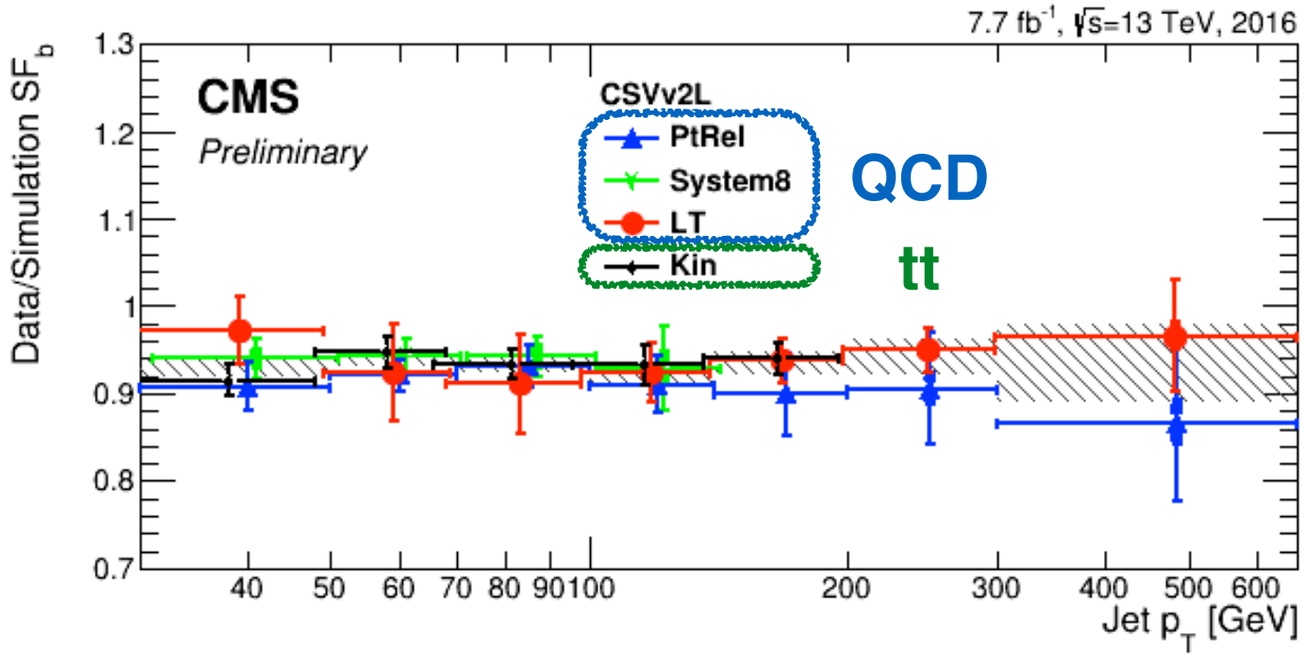
$$\varepsilon_{\mathrm{b}}^{\mathrm{tag}} = \frac{f_{\mathrm{b}}^{\mathrm{tag}} \cdot N_{\mathrm{data}}^{\mathrm{tag}}}{f_{\mathrm{b}}^{\mathrm{tag}} \cdot N_{\mathrm{data}}^{\mathrm{tag}} + f_{\mathrm{b}}^{\mathrm{untag}} \cdot N_{\mathrm{data}}^{\mathrm{untag}}}$$

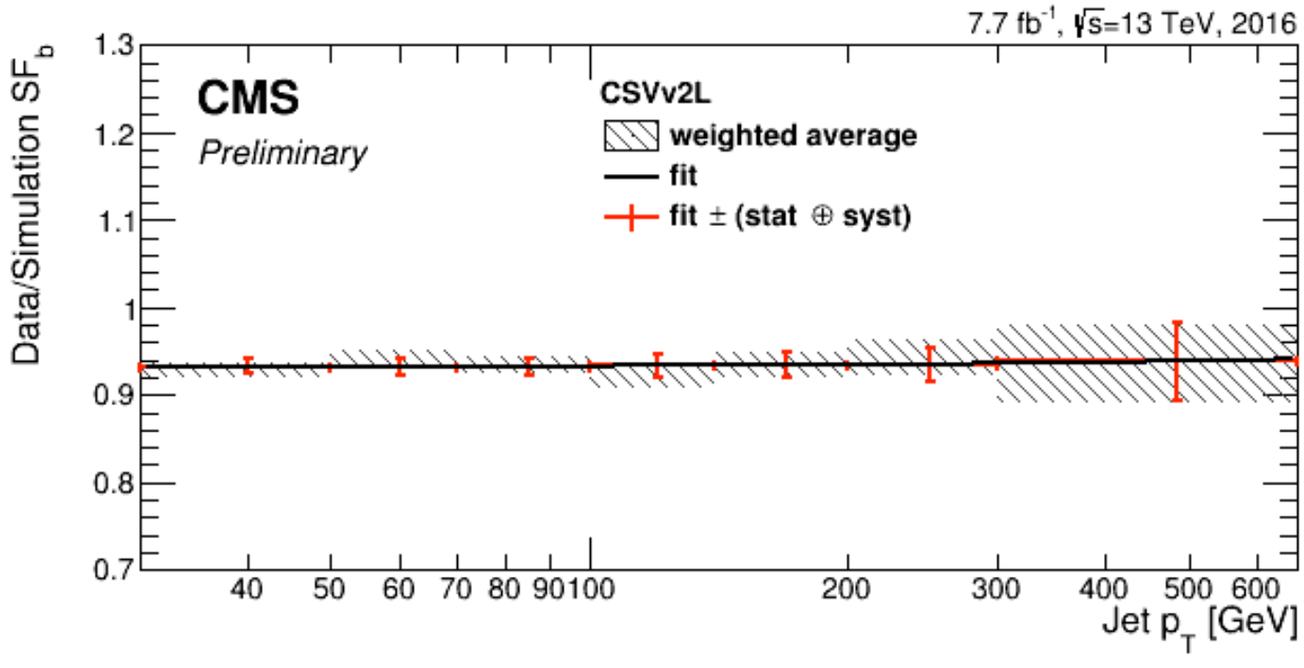


Performance measurements

example for CSV algorithm:

- agreement between data and simulation at 5% level
- associated uncertainties at 5% level as well
- result expressed as “scale factor” to be applied to the simulation
- what are reasons the scale factor is not exactly one?





Data/MC Scale Factors

- scale factors derived on jet-by-jet basis
- to be applied in simulated events
- various methods exist
- all methods have advantages and disadvantages
- choice for an SF method depends on analysis

jet-by-jet updating

flip b-tag status of individual jets based on drawing random numbers

example: switch 5% of b-tagged b-jets to untagged (for SF=0.95)

event reweighting

calculate event weight based on flavour combination

example: apply weight of 0.95 * 0.95 for events with two b-tagged b-jets (and nothing else!)

… and other options. If in doubt contact the hypernews

Tips and tricks to survive software writing

- Program defensively: when making an assumption (e.g. this pointer is valid, this number is > 0) **check it!** If not learn to raise/throw appropriate exceptions
- git is your friend, especially when trying to understand what changed
  - … and commit often, after every change
- learn to read the compiler errors, most of the time are quite descriptive
- learn how to use a debugger, pdb is straightforward, gdb much less
- command line parsers are a great tool to avoid having lots of files doing just slightly different things
- TEST YOUR CODE BEFORE SUBMITTING!

“the difference between a top grader and a bottom grader is not the amount of mistakes per page, but how quickly he/she can find them”

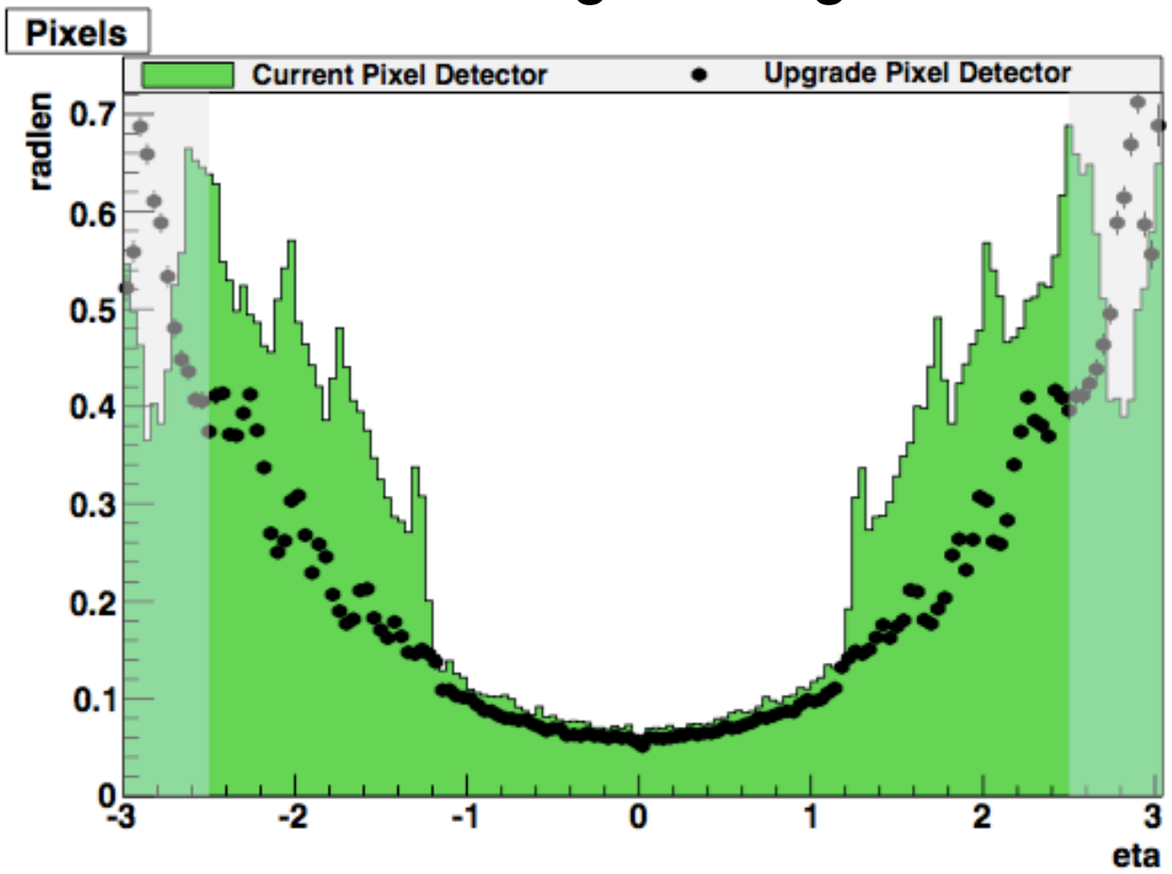
my math professor

Back-up

Phase I improved performance

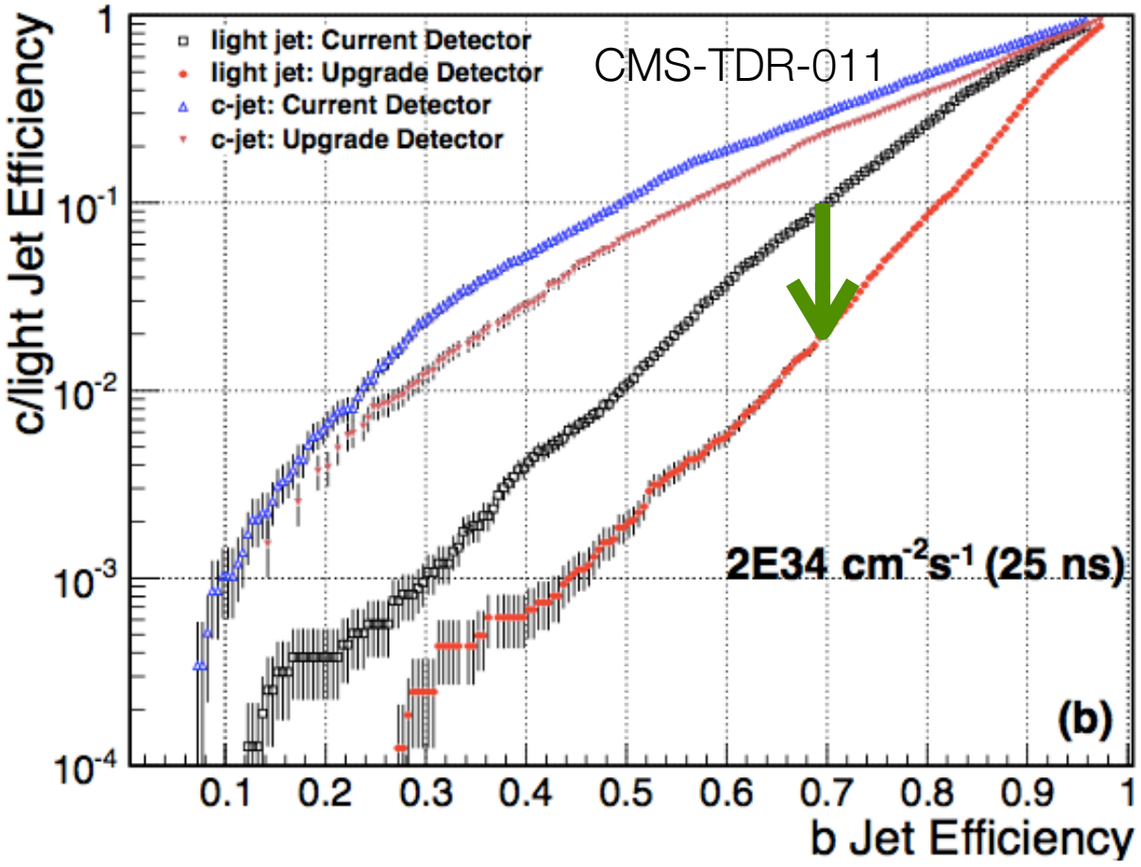
- Reduced mass by a factor two: from one-phase C_6F_{16} to two-phase CO_2 cooling system
- Improved **redundancy** ⇛ impacts **tracking efficiency** and purity
- Move **closer to beam** ⇛ improves **vertexing** and b-tagging

services moved outside the tracking coverage



but one more layer

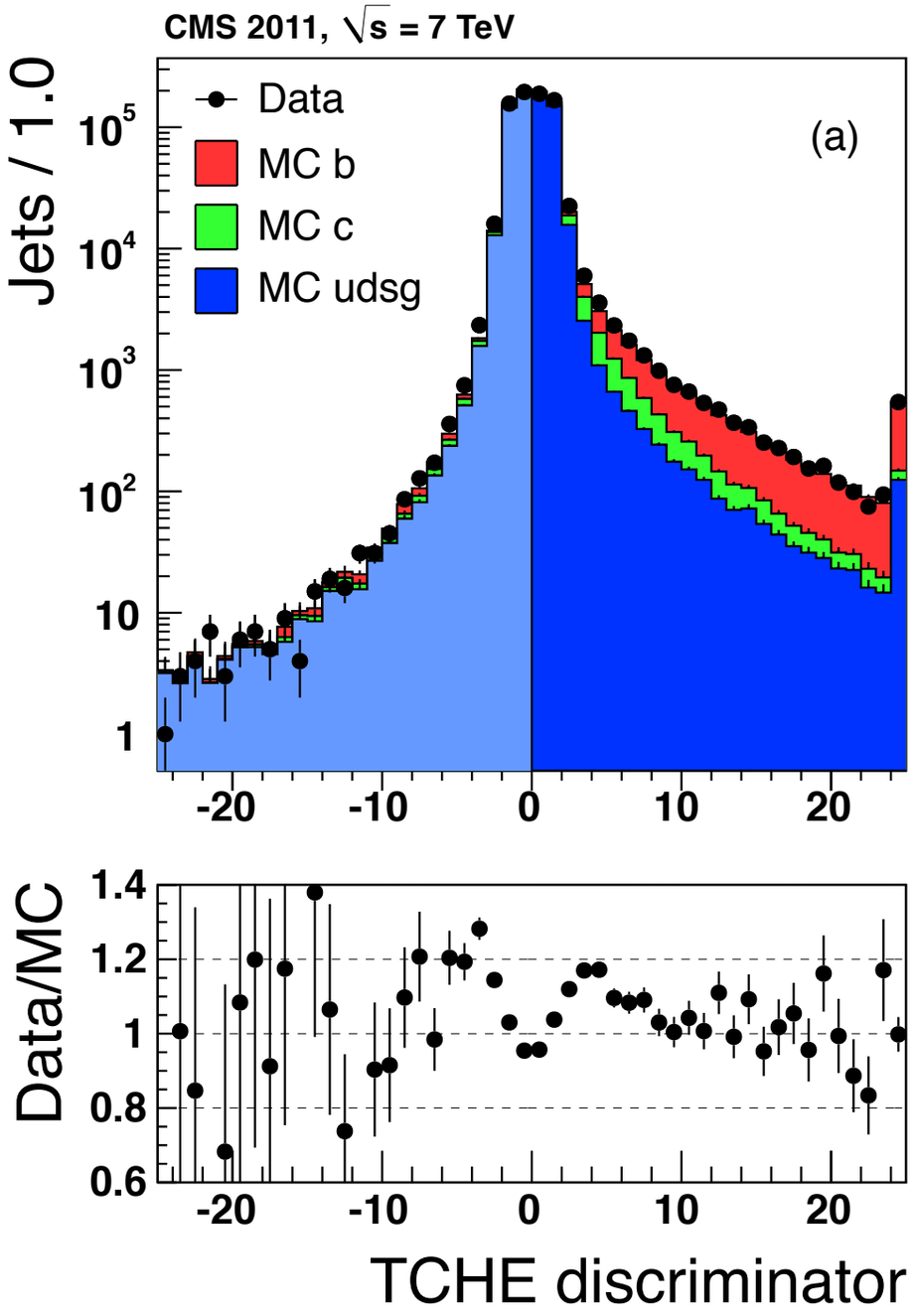
factor 5 in reduction of mistag @70% b efficiency

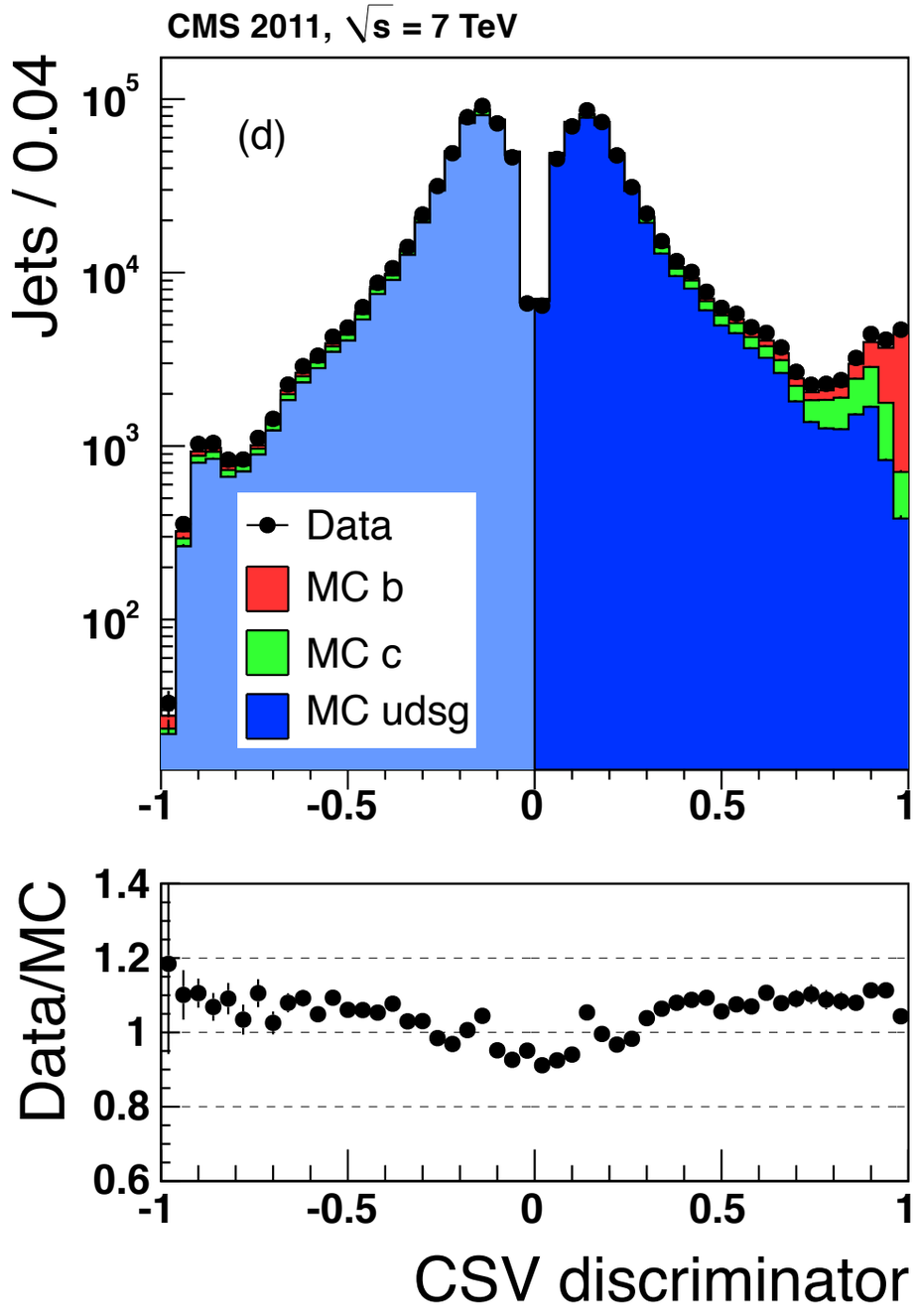


measurement of mistag rate

sign creates asymmetric discriminator distributions

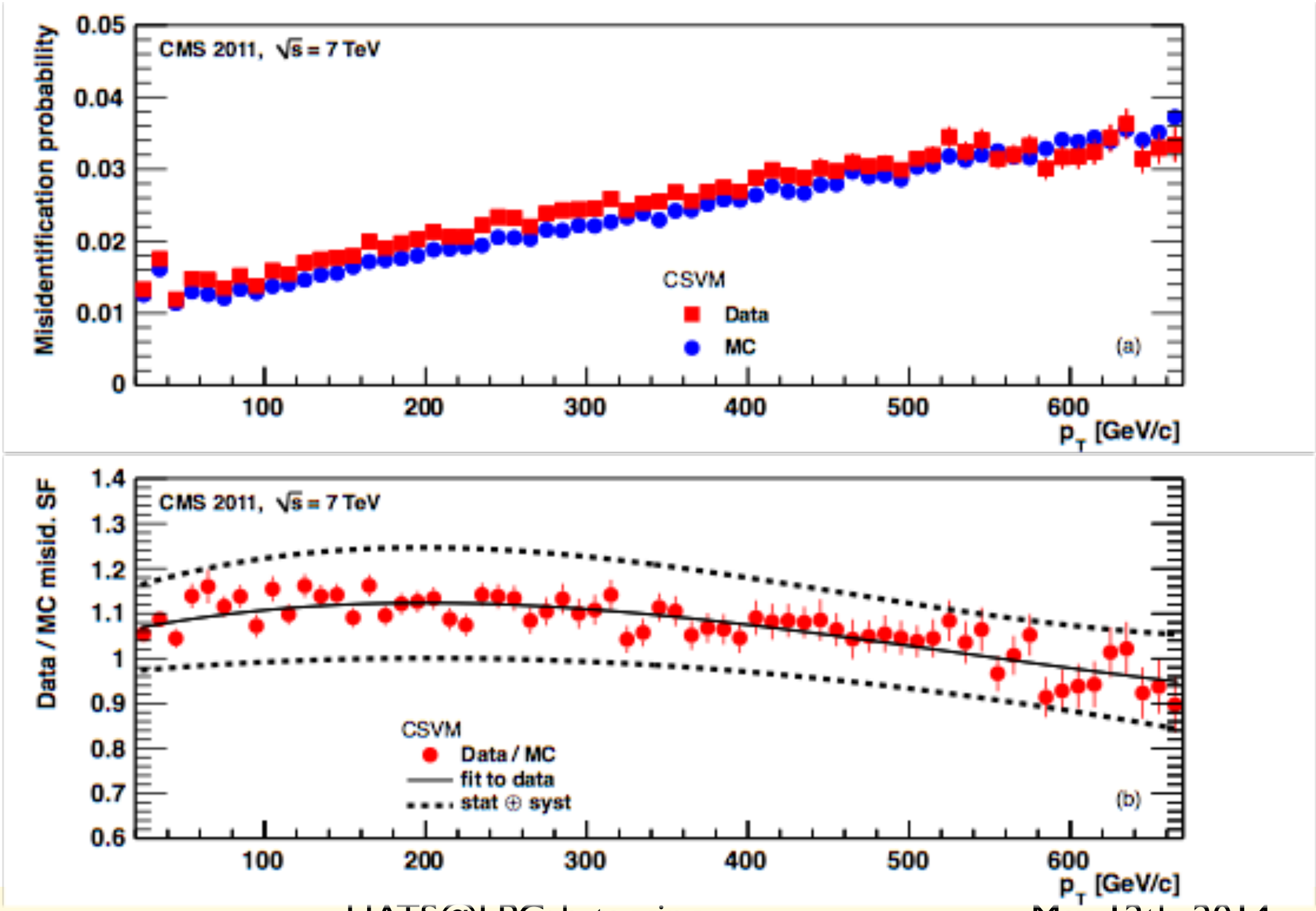
➡ negative side can be used to predict positive side for light flavour jets





measurement of mistag rate

projection from negative to positive side needs "correction" factor (taken e.g. from simulation): $\varepsilon_{\text{data}}^{\text{misid}} = \varepsilon_{\text{data}}^{-} \cdot R_{\text{light}}$



event reweighting

- an event with a given configuration b-jets and light flavour jets has the probabilities:

$$P(\mathrm{MC}) = \prod_{i=\mathrm{tagged}} \varepsilon_i \prod_{j=\mathrm{not\ tagged}} (1 - \varepsilon_j)$$

$$P(\mathrm{DATA}) = \prod_{i=\mathrm{tagged}} \mathrm{SF}_i \varepsilon_i \prod_{j=\mathrm{not\ tagged}} (1 - \mathrm{SF}_j \varepsilon_j)$$

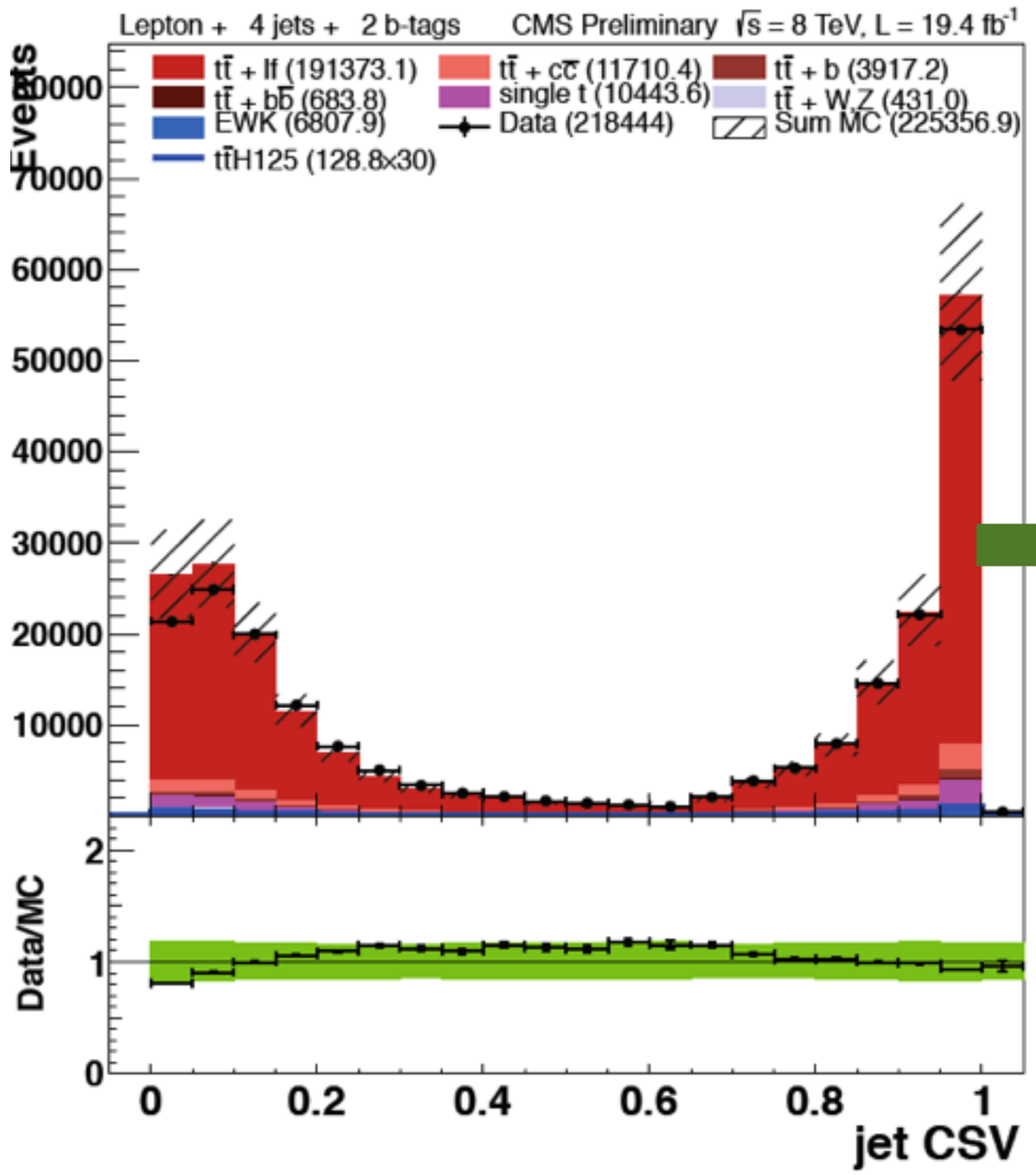
- then the weight is simply: $w = \frac{P(\mathrm{DATA})}{P(\mathrm{MC})}$

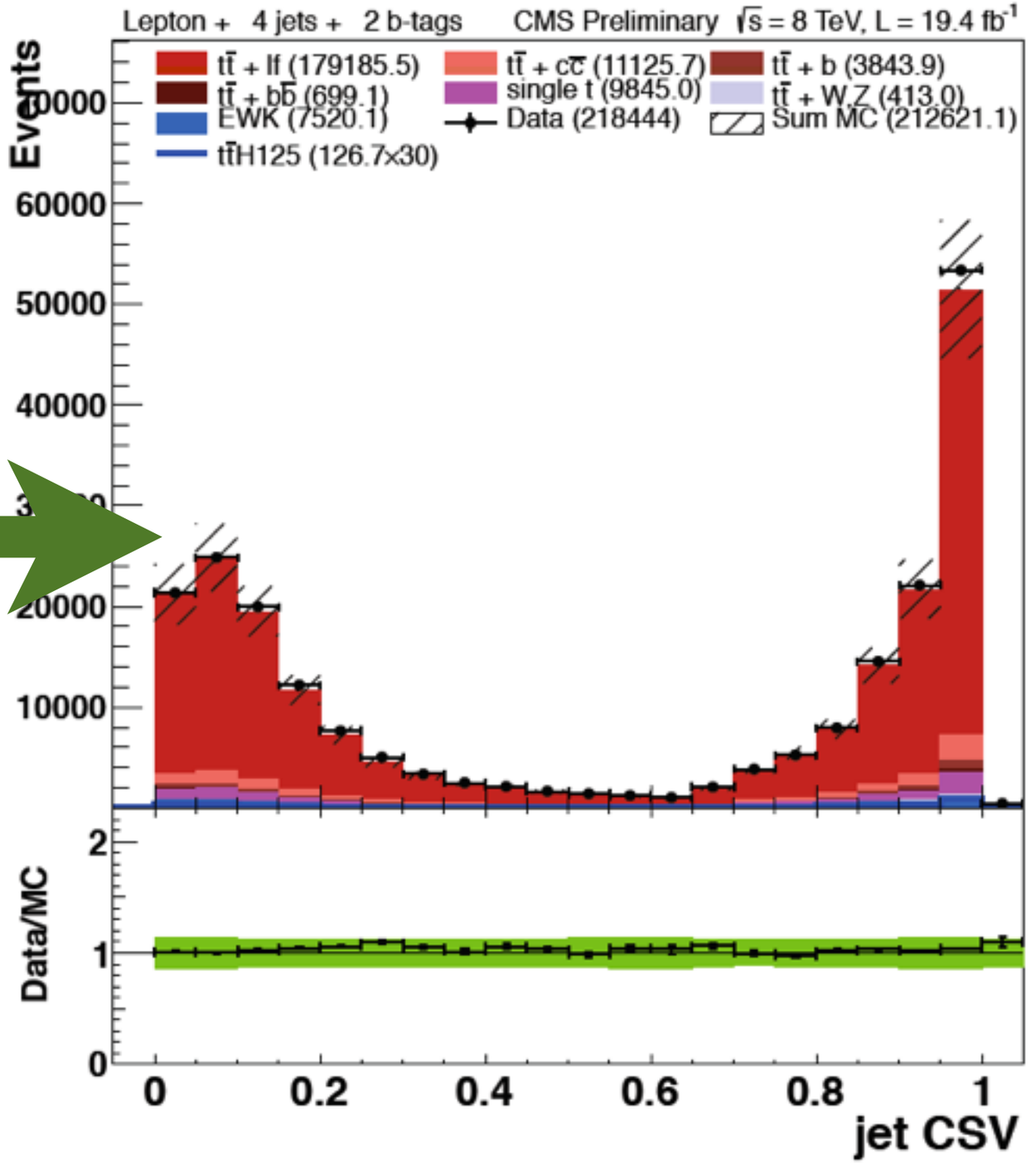
- relatively simple and straightforward method
- absolute efficiencies and mistag rates are required
- other methods exist which ONLY require the SF
- see https://twiki.cern.ch/twiki/bin/view/CMS/BTagSFMethods for a complete overview of methods

discriminator reshaping

- correcting efficiencies at fixed operating points does not improve agreeement between distributions (e.g. discriminator distribution)

- a relatively new method is to reweight the discriminator itself
- the discriminator can then be used later in the analysis (e.g. as input to MVA methods)

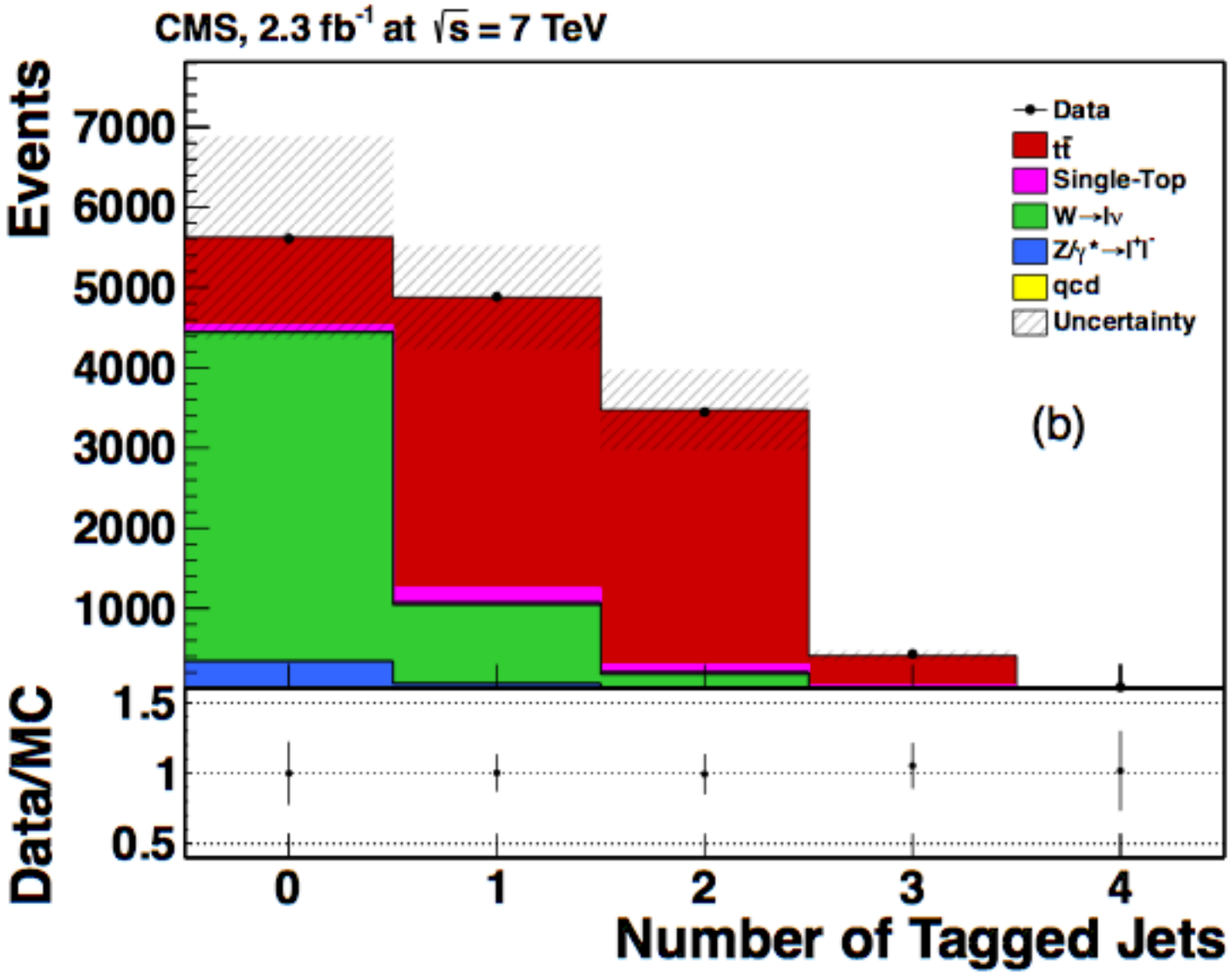




efficiency measurements

method with top-quarks:

- count number of b-tagged jets in a sample of top-quarks



- the background needs to be known (simulation)

efficiency measurements

- the number of events with n b-tagged jets can be written as

$$\langle N_n \rangle = L \cdot \sigma_{t\bar{t}} \cdot \varepsilon \cdot \sum_{i,j,k} F_{ijk} \sum_{i'+j'+k'=n}^{i' \le i, j' \le j, k' \le k} [C_i^{i'} \varepsilon_b^{i'} (1-\varepsilon_b)^{(i-i')} C_j^{j'} \varepsilon_c^{j'} (1-\varepsilon_c)^{(j-j')} C_k^{k'} \varepsilon_l^{k'} (1-\varepsilon_l)^{(k-k')}]$$

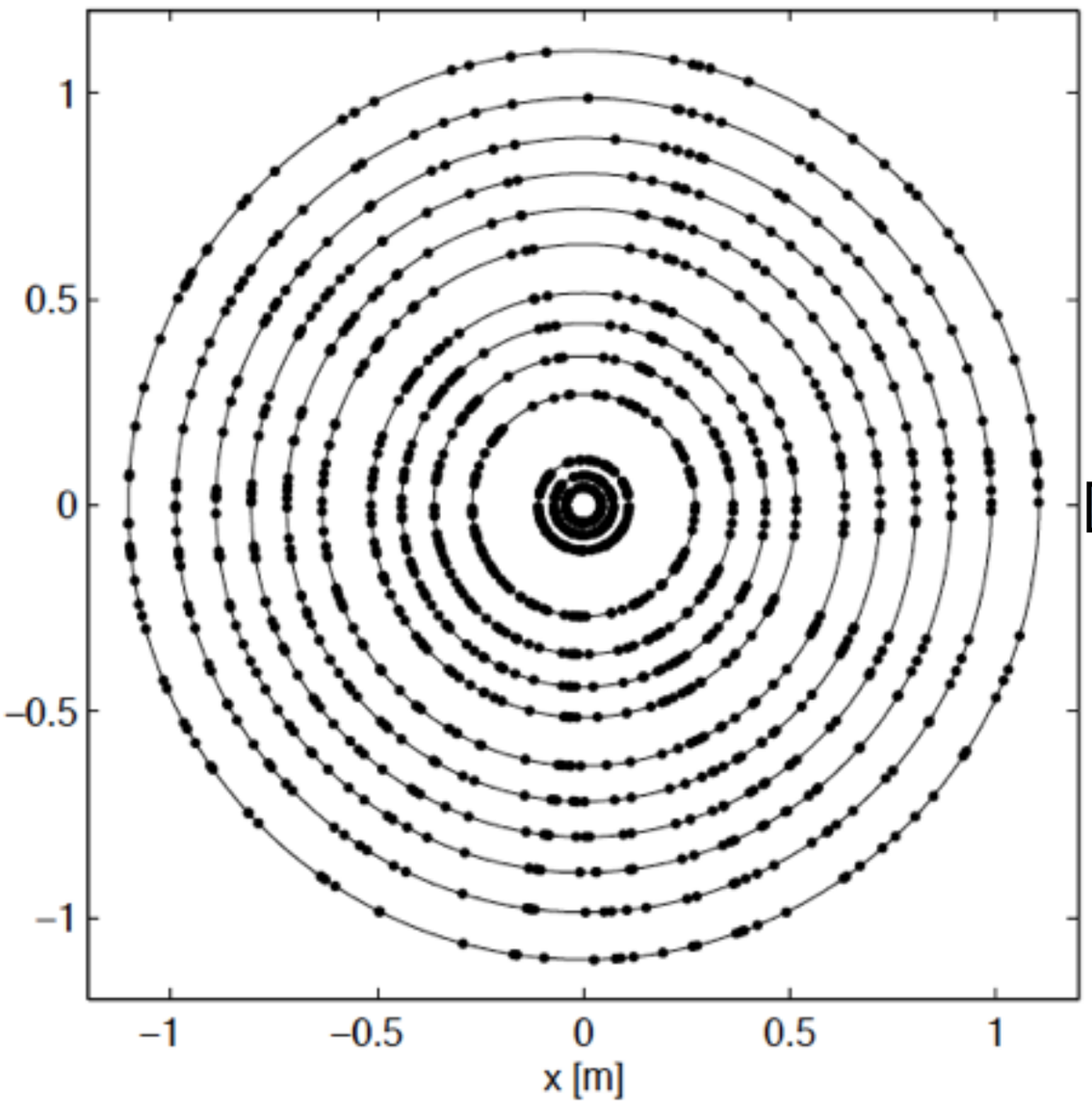
- where $C_i^{i'}$ is the binomial coefficient and ε are the efficiencies
- F_{ijk} is the fraction of events with i b-jets, j c-jets and k light jets
- example: the F_{112} term contributes to the bin with one b-tag as

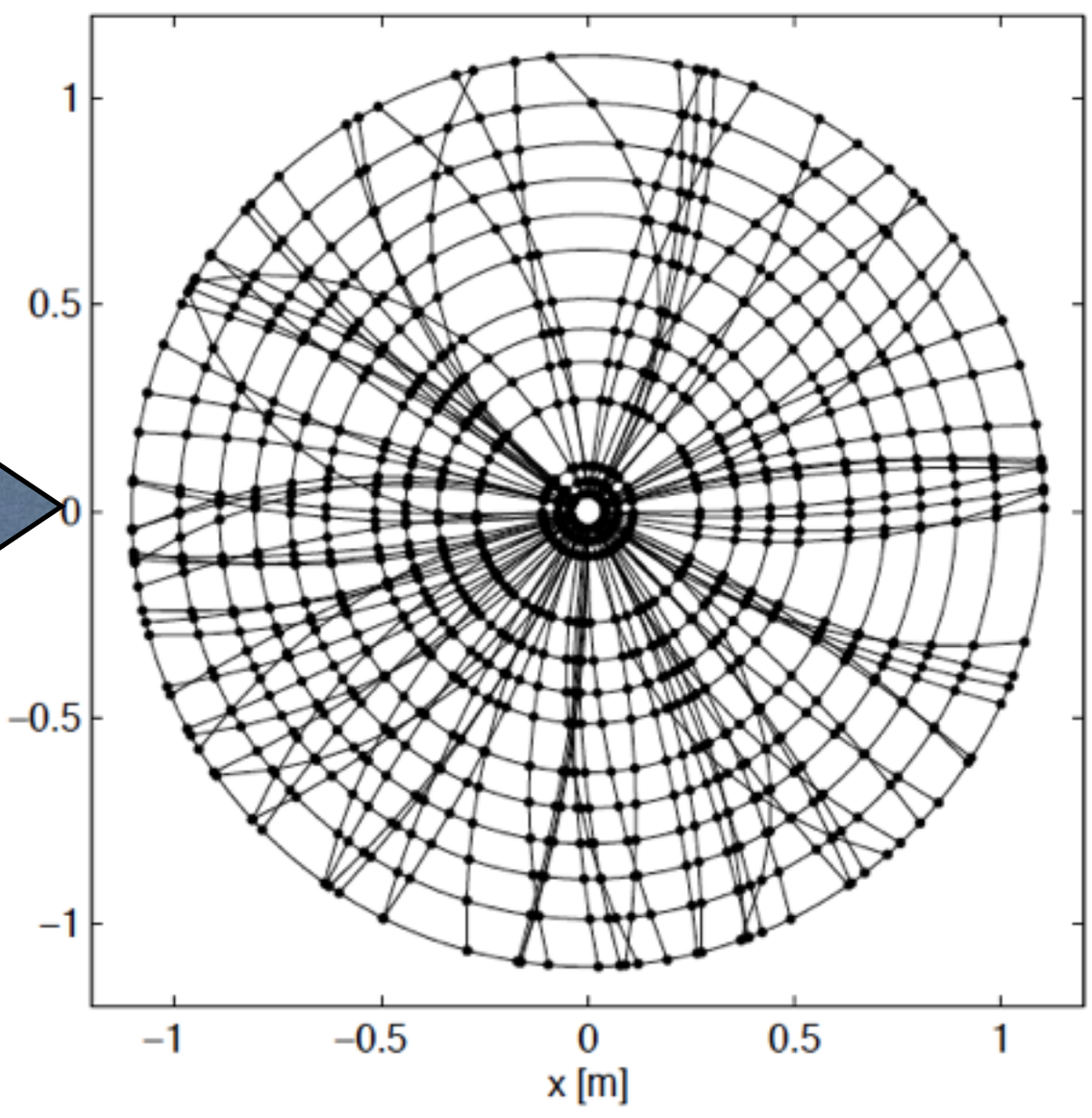
$$\langle N_1 \rangle \propto F_{112} \times \left(\underbrace{1 \cdot \varepsilon_b (1-\varepsilon_c)(1-\varepsilon_l)^2}_{\text{the b jet}} + \underbrace{1 \cdot (1-\varepsilon_b)\varepsilon_c (1-\varepsilon_l)^2}_{\text{the c jet}} + \underbrace{2 \cdot (1-\varepsilon_b)(1-\varepsilon_c)\varepsilon_l (1-\varepsilon_l)}_{\text{the light-parton jet}} \right)$$

- the efficiencies are parameters in a likelihood fit, in which the function is minimized:

$$\mathcal{L} = -2 \log \prod_n \mathcal{P}oisson(N_n, \langle N_n \rangle)$$

track reconstruction





- classification or pattern recognition problem
- b-tagging stands and falls with the quality of tracks

new workflow in more detail