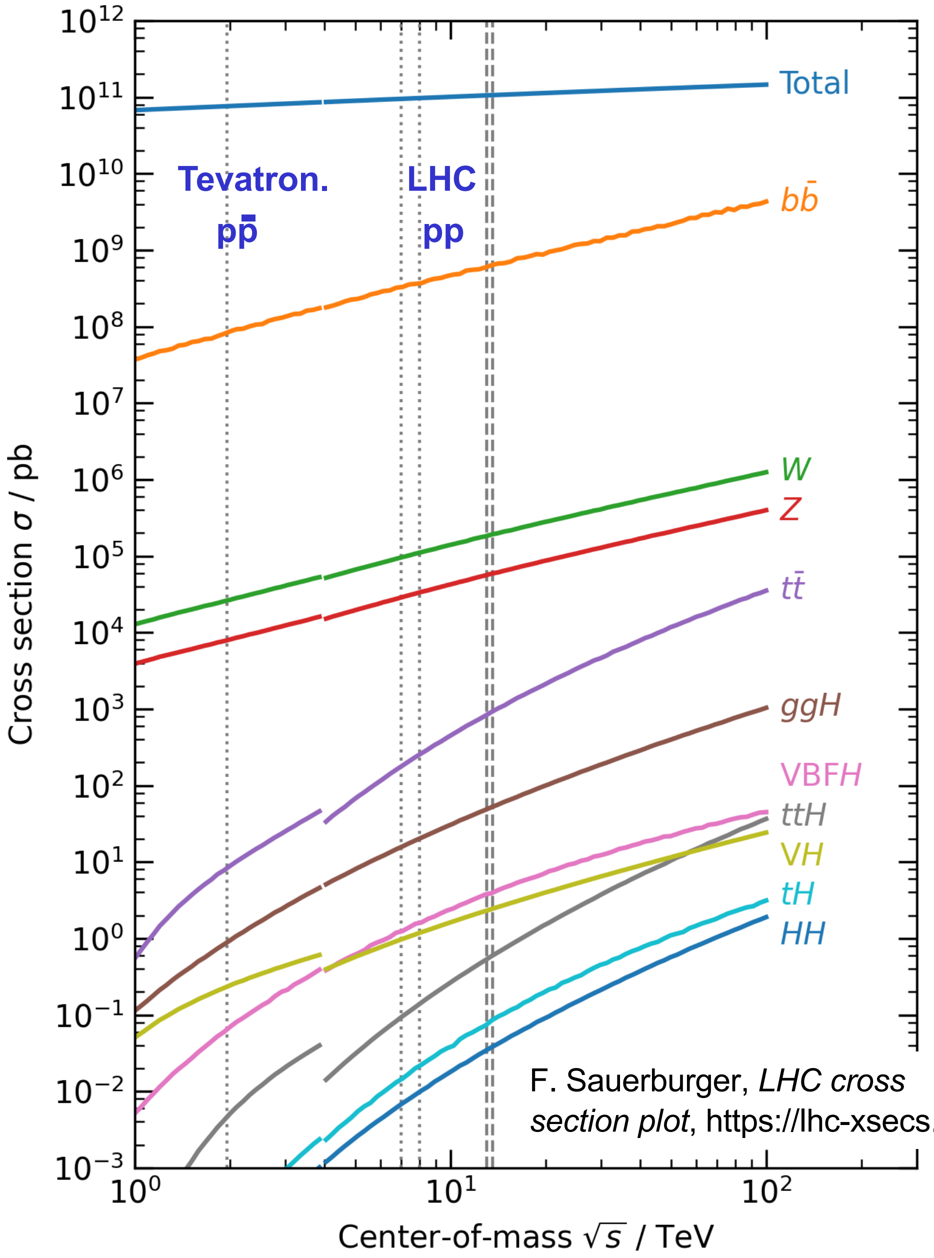
b-tagging: basis, concepts and beyond

Daniel Bloch, TOP LHC France, May 2023, IPHC Strasbourg

Introduction

why heavy flavour tagging ?



F. Sauerburger, *LHC cross section plot*, https://lhc-xsecs.org/

- jets from b and c quarks are involved in many Standard Model processes:
  - bb (cc) production is ~1% (10%) of total inelastic cross section at LHC
  - BR(W → cs) = 33%, BR(Z → bb) = 15%
  - BR(t → Wb) ~ 100%
  - BR(H → bb) = 58%, BR(H → cc) = 3%
- the large bb and tt cross sections allow to **calibrate** the b-tag algorithms
- tagging b jets are necessary to both remove the large SM backgrounds and to perform **SM precision** measurements: t+X, tt, H, HH, …
- and allow **new physics searches** in many many channels:
  - heavy vectors : W' → tb, Z' → bb
  - new scalars : h → bb, H^+ → bc
  - SUSY, etc…

short history

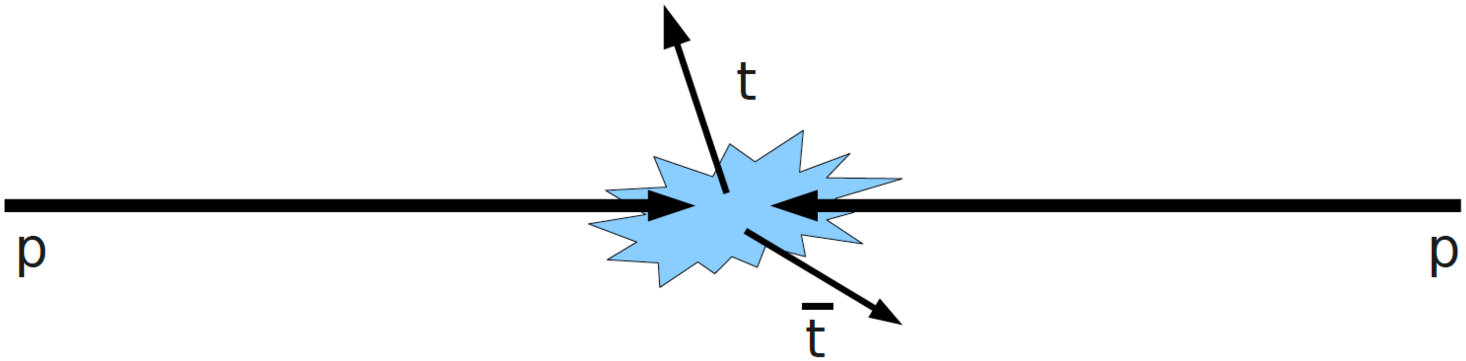
- Theory:
  - 1964, uds quark model: Gell-Mann, Zweig
  - 1961-1968, EW standard model: Glashow, Salam, Weinberg
  - 1964, H scalar field: Englert, Brout, Higgs
  - 1970, charm prediction: Glashow, Iliopoulos, Maiani
  - 1973, bottom and top prediction: Kobayashi, Maskawa

- Experimental discoveries:
  - 1974, charm: Ting at BNL, Richter at SLAC
  - 1977, bottom: Lederman at FNAL
  - 1983, W and Z: Rubbia, UA1, UA2 at CERN
  - 1995, top: CDF, D0 at FNAL
  - 2012, H boson: ATLAS, CMS at CERN

- Silicon vertex detectors at colliders: first used at
  - Mark II (1989), then SLD(1991-1998) at SLAC SLC (e^+e^- 91 GeV)
  - LEP experiments (1989-2000) at CERN (e^+e^- 88-208 GeV)
  - CDF (1992-2011) and D0 (1995-2011) at Tevatron FNAL ($p\bar{p}$ 1.8 TeV)
  - and everywhere: Hera (1991-2007), BaBar (1998-2008), Belle (1999-2010), LHC (since 2009), Belle II (since 2018), etc…

a simplified event

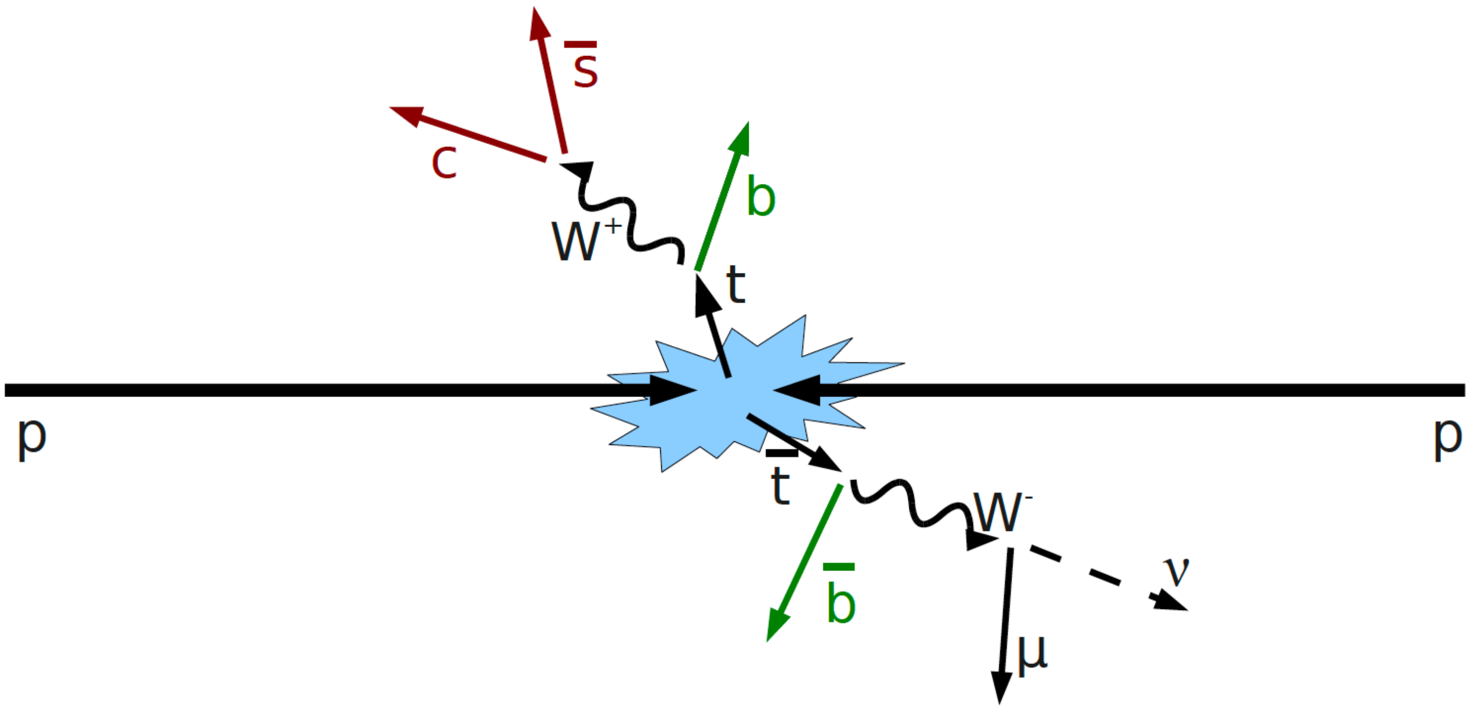
from Christophe Saout, DESY 2009



collision → production of heavy resonance (here top quark pair)

a simplified event

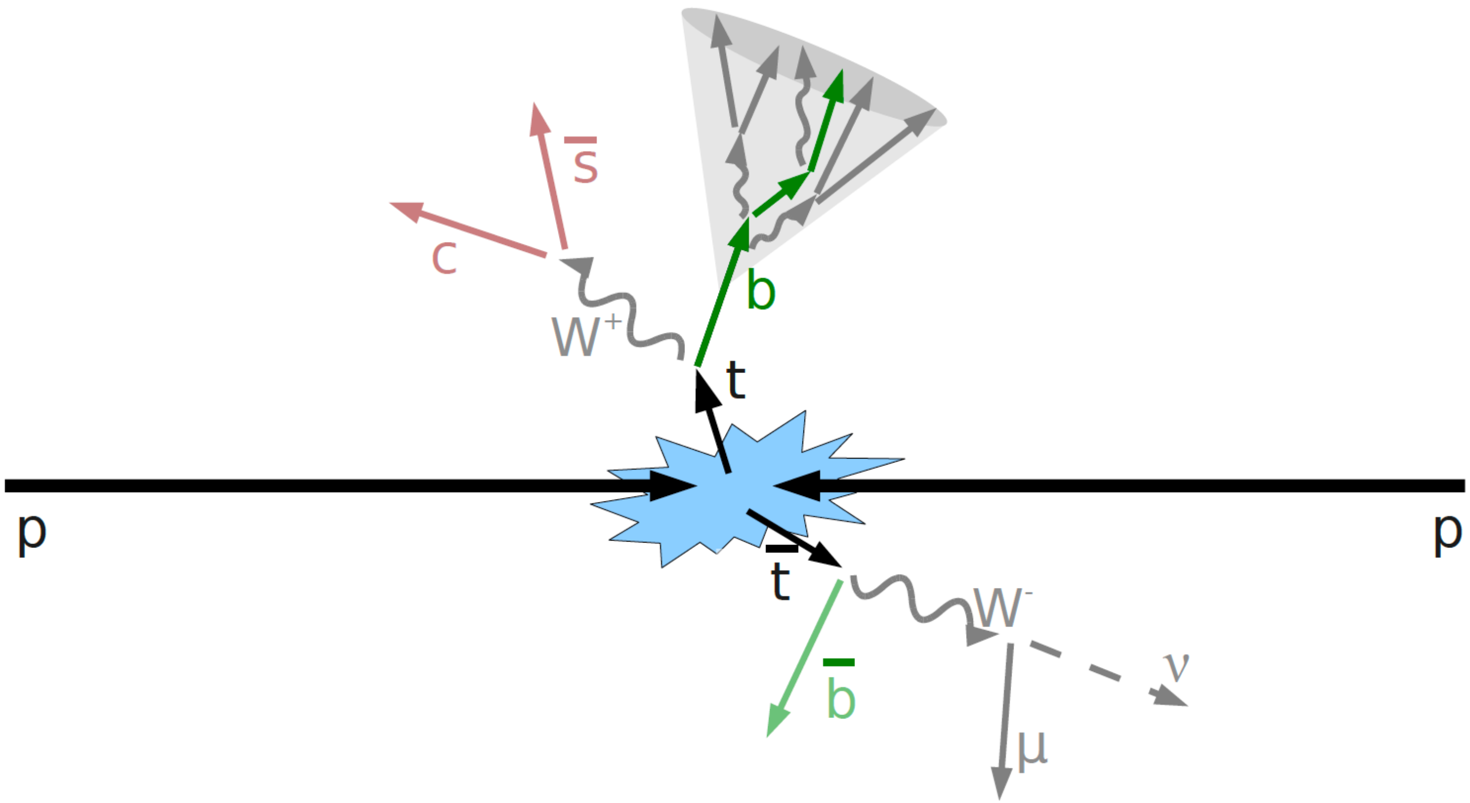
from Christophe Saout, DESY 2009



heavy resonance decay → b quark (and more) in the final state

a simplified event

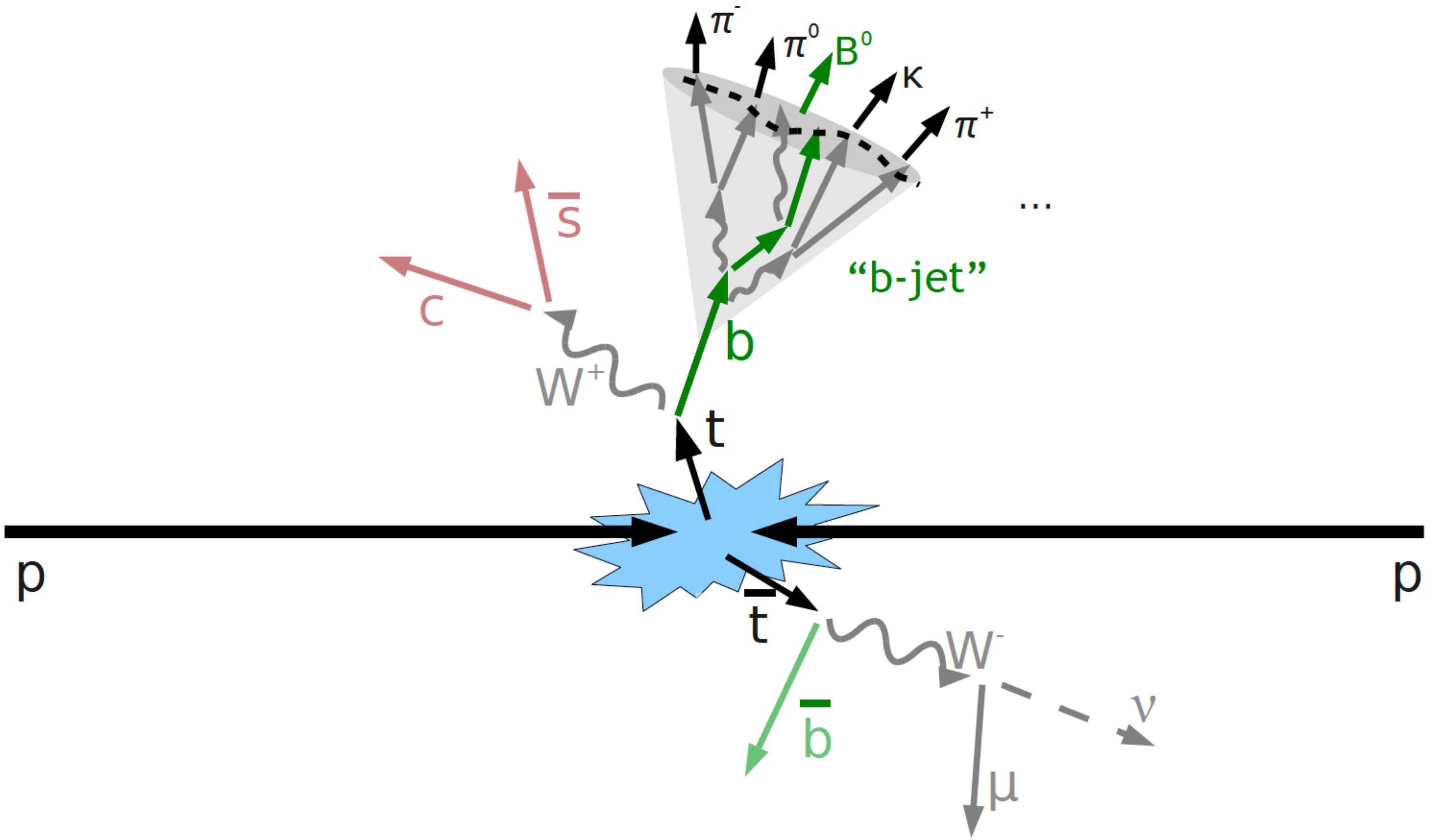
from Christophe Saout, DESY 2009



parton shower, gluon radiation, gluon splitting, hadronization of quarks

a simplified event

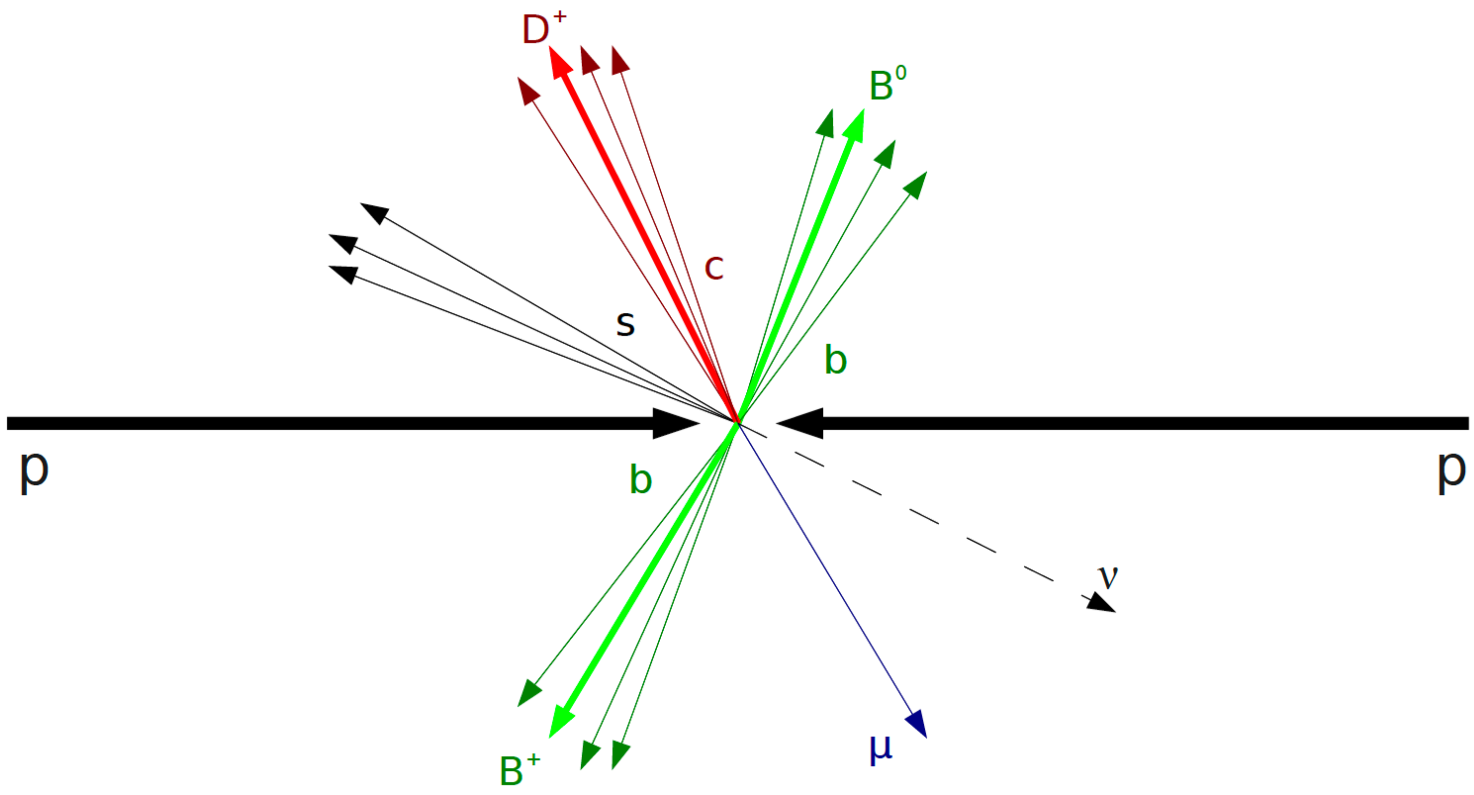
from Christophe Saout, DESY 2009



b-hadron and c-hadron decays, jets of particles

a simplified event

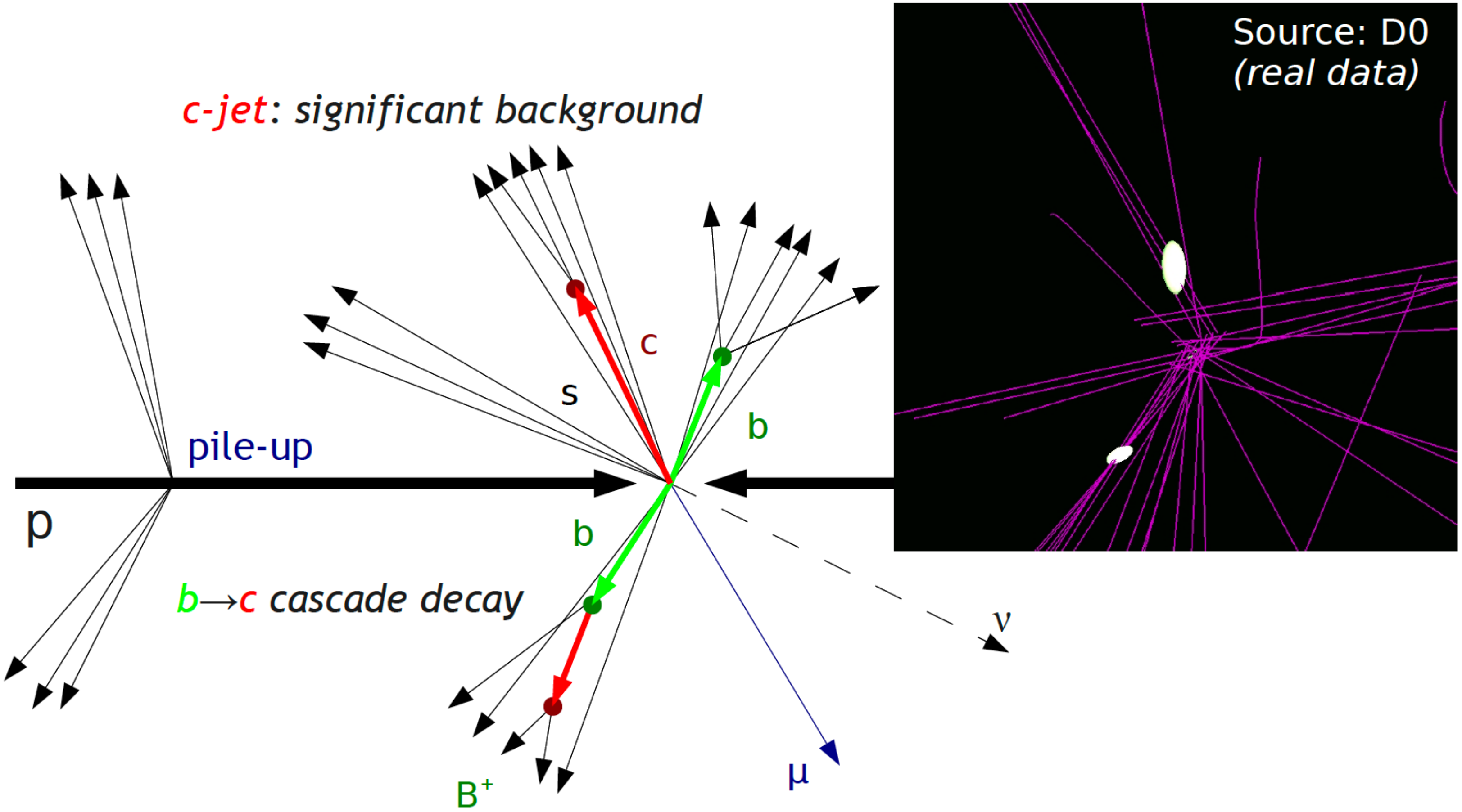
from Christophe Saout, DESY 2009



what we see in the detector: jets of particles, leptons, missing E_T

a simplified event

from Christophe Saout, DESY 2009



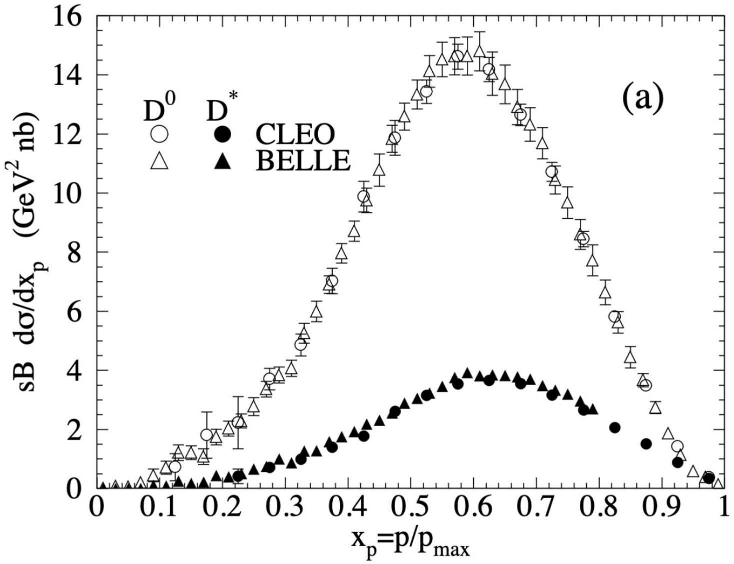
precise track reconstruction allows us to see secondary vertices from b or c decays: this is how we can perform b-tagging

Chap. I
charm and beauty properties

from PDG 2022

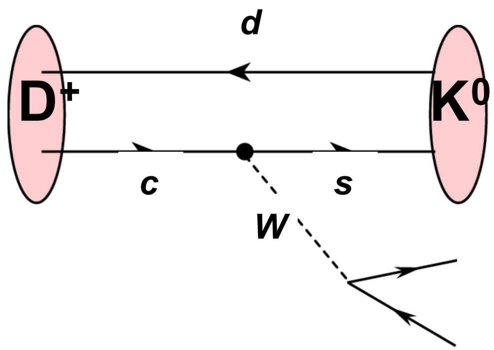
charm production & decay

- charm fragmentation
  - c quark can hadronize into excited D**, D* or to c-hadrons, with strong or e.m decays of D** and D*
  - <E(c-hadron) / E(c quark)> ~ 0.5-0.6
  - gluon splitting to $c\bar{c}$ = 3.0+_0.5%



- charm decay
  - m(D^0) = 1.865, m(D^+) = 1.870, m(D_s) = 1.968, m(Λ_c) = 2.286 GeV
  - $\tau(D^0)$ ~ 0.41 ps, $\tau(D^\pm)$ ~ 1.03 ps, τ(Ds) ~ 0.50 ps, $\tau(\Lambda_c)$ ~ 0.20 ps
  - ~94% c → s decay, ~11% c → e+X, ~12% c → μ+X, 0.1% c → $\tau\nu$
  - on average: ~2 charged particles from c-hadron decay

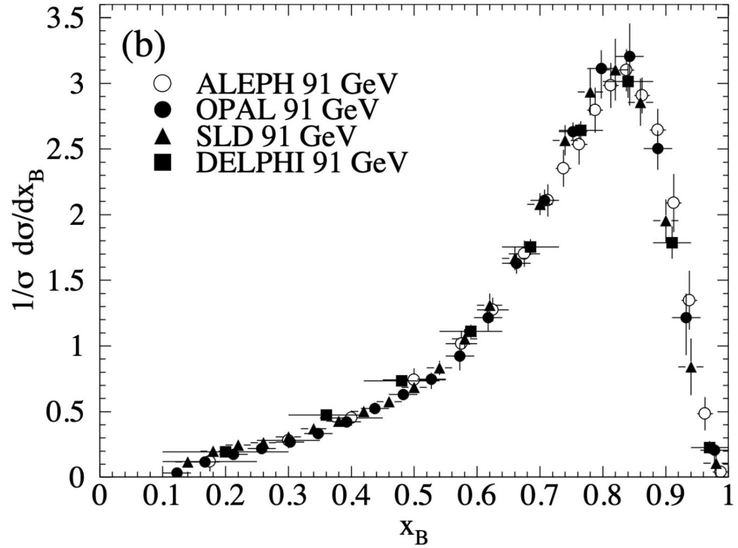
$$\| V_{CKM} \| = \begin{pmatrix} V_{ud} = 0.974 & V_{us} = 0.225 & V_{ub} = 0.004 \\ V_{cd} = 0.225 & V_{cs} = 0.973 & V_{cb} = 0.042 \\ V_{td} = 0.009 & V_{tb} = 0.041 & V_{tb} = 0.999 \end{pmatrix}$$



from PDG 2022

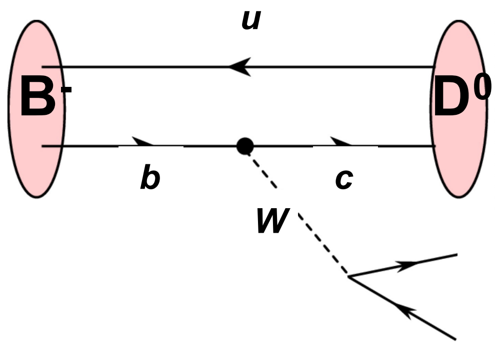
beauty production & decay

- b fragmentation: (from Z → $b\bar{b}$ meas)
  - b quark hadronize to excited B** or B* (~ 87% of the time) which decays (strongly or e.m.) to b-hadrons, giving:
  - 41% B^0, 41% $B^\pm$, 10% B_s, 8% b-baryon
  - <E(b-hadron) / E(b quark)> ~ 0.7
  - gluon splitting to $b\bar{b}$ = 0.28+_0.07%



- b decay
  - $m(B^0) = 5.280$, $m(B^+) = 5.279$, $m(B_s) = 5.367$, $m(\Lambda_b) = 5.620$ GeV
  - $\tau(B^0) \sim 1.52$ ps, $\tau(B^\pm) \sim 1.64$ ps, $\tau(B_s) \sim 1.52$ ps, $\tau(\Lambda_b) \sim 1.47$ ps
  - 98% b → c decay, 11% b → c eν, 11% b → c μν, 2% b → c τν decays
  - 20+_6% decay into 2 charm hadrons
  - on average: ~5 charged particles from b-hadron decay

$$\| V_{CKM} \| = \begin{pmatrix} V_{ud} = 0.974 & V_{us} = 0.225 & V_{ub} = 0.004 \\ V_{cd} = 0.225 & V_{cs} = 0.973 & V_{cb} = 0.042 \\ V_{td} = 0.009 & V_{tb} = 0.041 & V_{tb} = 0.999 \end{pmatrix}$$



b properties

- long lifetime ~1.57 ps and hard fragmentation

  => $<\beta\gamma c\tau> = E_B / m_B\, c\tau_B = 0.3 * 1.57 * 0.7\, E_b[\mathrm{GeV}] / 5.3 =$
  - 4 mm at E(b quark) = 60 GeV
  - 4 cm at 600 GeV (beyond beam pipe and first pixel layer)
- high track multiplicity: allows 2^{ndary} vertex reconstruction
- 22% semi-muonic decay (including muon from c-hadron)
  - interesting for low-level b-taggers
  - very useful for b-tagging efficiency measurement from data
- b backgrounds
  - multijet QCD events (gluon split): ~0.3% $b\bar{b}$ and ~3% $c\bar{c}$
  - charm has τ_c ~ 0.4-1 ps and some rather hard fragmentation

    => c production is a significant background for b-tagging
- c-properties
  - intermediate between b and light properties
  - c-tagging is more difficult, but feasible too

Chap. II
what do we need before tagging ?

what do you need before tagging ?

- Excellent detector performance, especially for tracks with pixel detectors
- Tracking in high pileup environment
- Primary vertex reconstruction, pileup discrimination
- Jet reconstruction
- Secondary vertex reconstruction

- Heavy flavour tagging is at the very end of a long chain of object reconstruction
- One needs also a very good detector simulation and a constant survey of data/simulation comparisons at all level in the event reconstruction, in order to improve the detector response in the simulation

example: the CMS tracker

Position information from finely segmented silicon sensors:

- Record the path of charged particles
- Measure momentum from bending radius in the magnetic field
- Reconstruct primary and secondary vertices

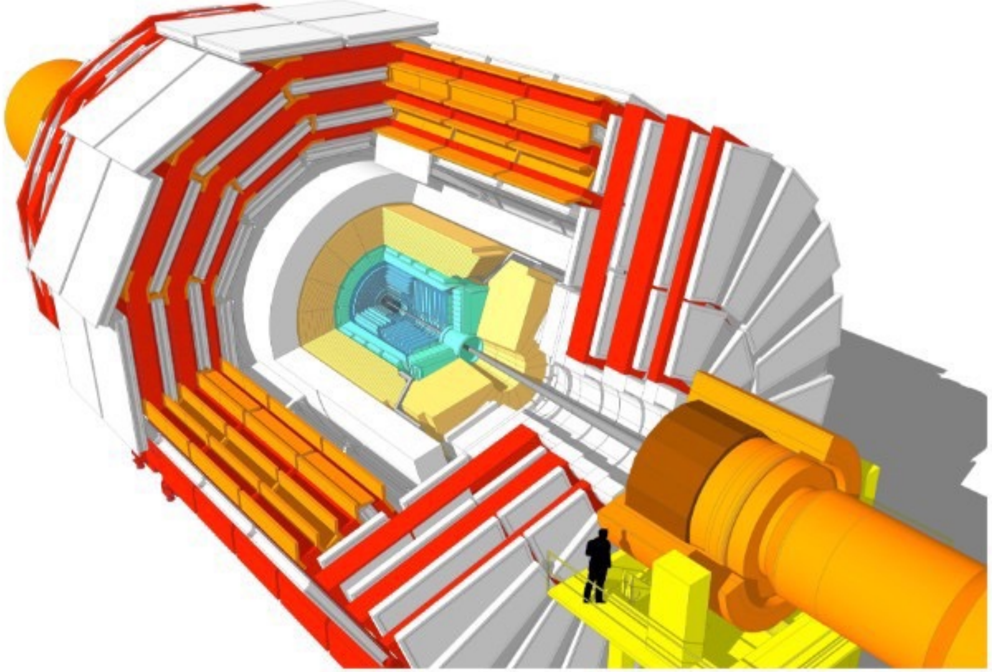
performance:
[typically ~15 hits per track]
$\sigma(p_T)/p_T$ ~ 1-2% @ 100 GeV/c
$\sigma(IP)$ ~ 10-20 μm @ 10-100 GeV/c

Requirements:

- **High resolution** & **low occupancy**:
  resolve and isolate individual tracks, reconstruct vertices

  Finer granularity is needed closer to the IP [High particle density, small tracking volume]

- **High rate capability**:
  fast charge collection time and read-out electronics to keep up with the expected event rates
- **Low material budget**:
  minimize multiple scattering
- **Radiation hardness**:
  innermost subdetectors
  ⇒ receive highest particle fluence



immersed in a 3.8 T magnetic field

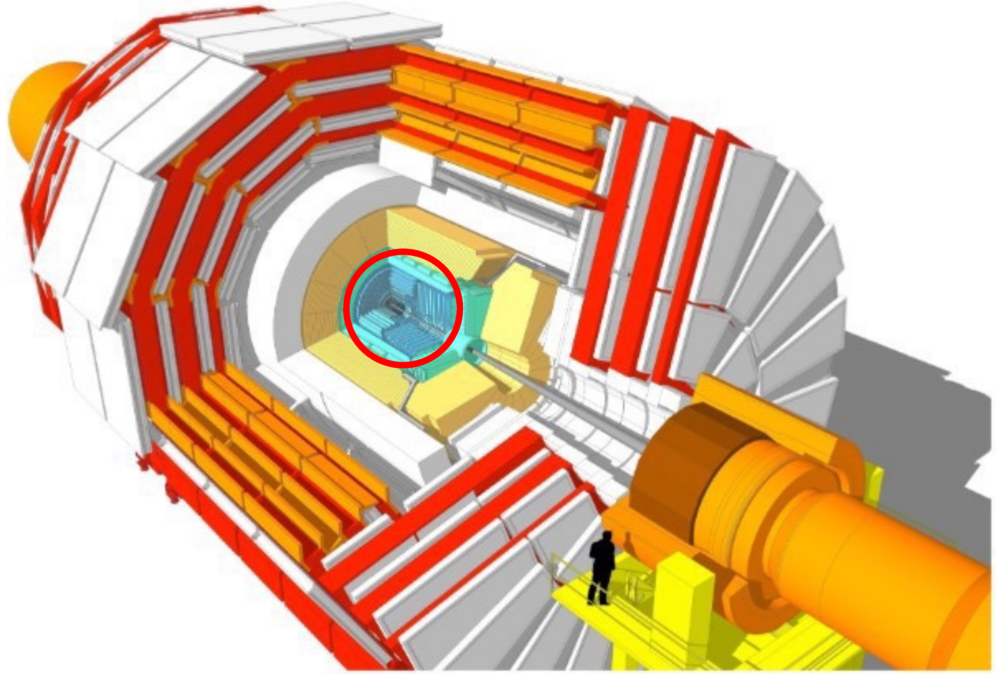
example: the CMS tracker

Position information from finely segmented silicon sensors:
- Record the path of charged particles
- Measure momentum from bending radius in the magnetic field
- Reconstruct primary and secondary vertices

performance:
[typically ~15 hits per track]
$\sigma(p_T)/p_T$ ~ 1-2% @ 100 GeV/c
σ(IP) ~ 10-20 μm @ 10-100 GeV/c

Requirements:
- **High resolution** & **low occupancy**:
  resolve and isolate individual tracks, reconstruct vertices

  Finer granularity is needed closer to the IP [High particle density, small tracking volume]
- **High rate capability**:
  fast charge collection time and read-out electronics to keep up with the expected event rates
- **Low material budget**:
  minimize multiple scattering
- **Radiation hardness**:
  innermost subdetectors
  ⇒ receive highest particle fluence



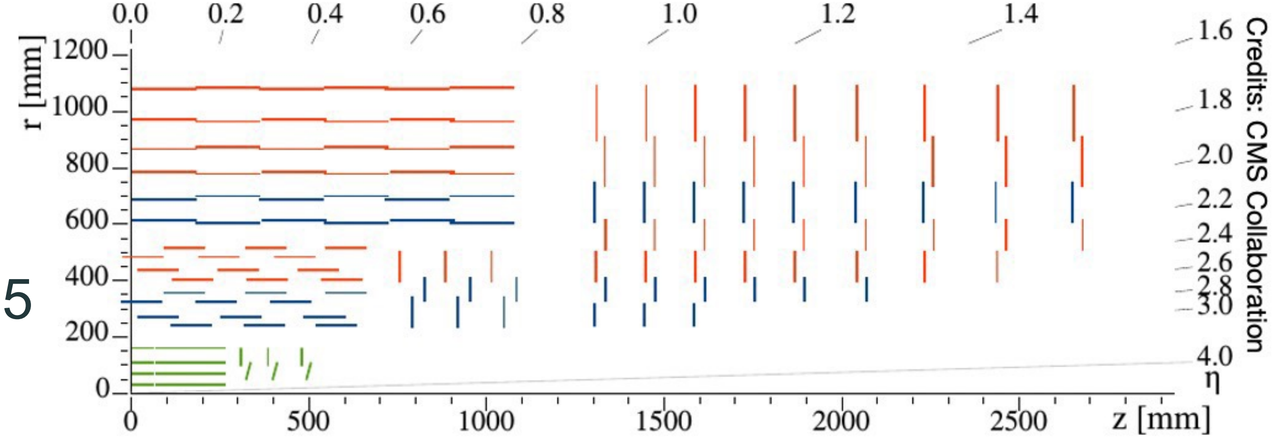
immersed in a 3.8 T magnetic field

CMS tracker

double sided strips
single sided strips
pixels

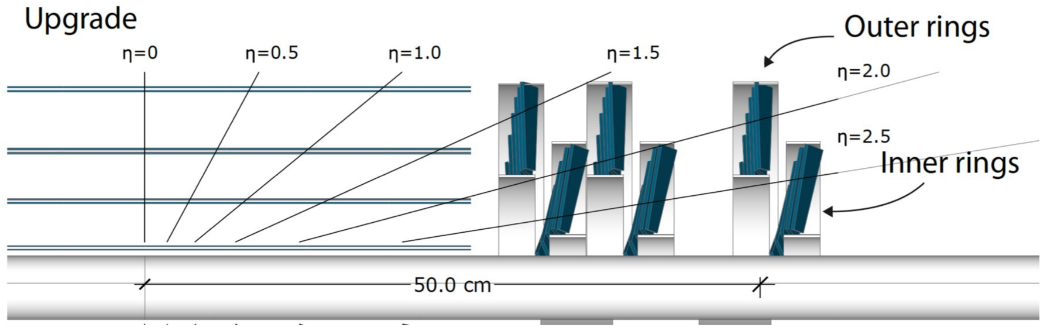
strips:

- O(10 million strips)
- O(200 m^2 sensors)
- coverage up to $|\eta|<2.5$



pixels: crucial for tagging of b and c quarks

- 127 million pixels
- 100x150 μm^2 size
- hit resolution 10x(20-40) μm^2
- coverage up to $|\eta|<2.5$
- 1rst layer as close to 2.9 cm from beam crossing point
- 4 barrel layers, 3 endcap disks (x2)

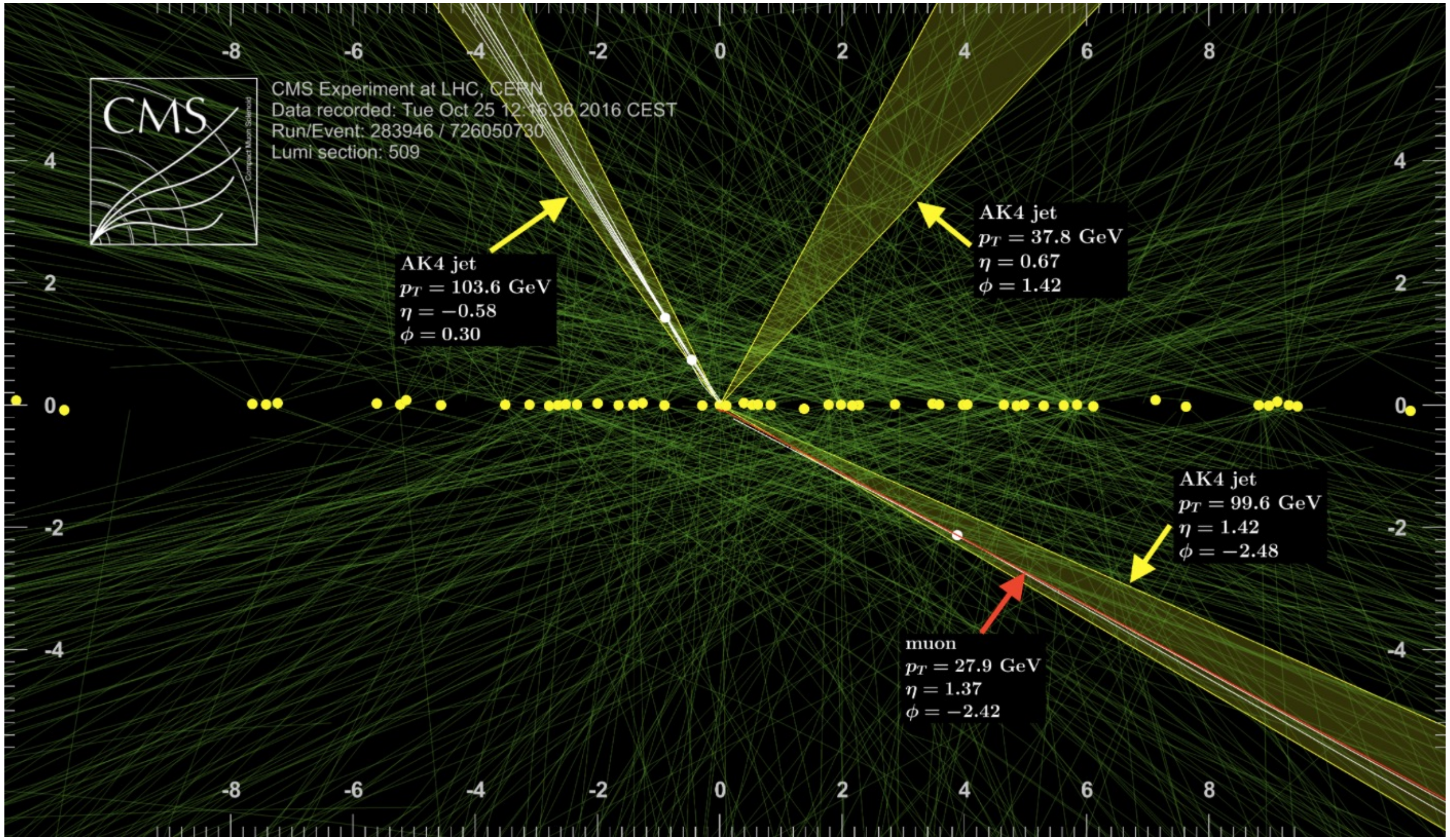


tracking

- The reconstruction of charged particle tracks and vertices is fundamental for the reconstruction of every type of physics event:
  - directly used in the reconstruction of charged hadrons, electrons, and muons
  - needed to distinguish charged and neutral hadrons and discriminating electrons from photons
  - crucial ingredient for higher level objects like b-tagged jets, c-tagged jets or taus
  - association of tracks to vertices needed to distinguish particles from the hard interaction from pileup vertices
  - secondary vertices crucial to track the decay chains of particles

tracking: challenge

CMS DP Note 2017-032



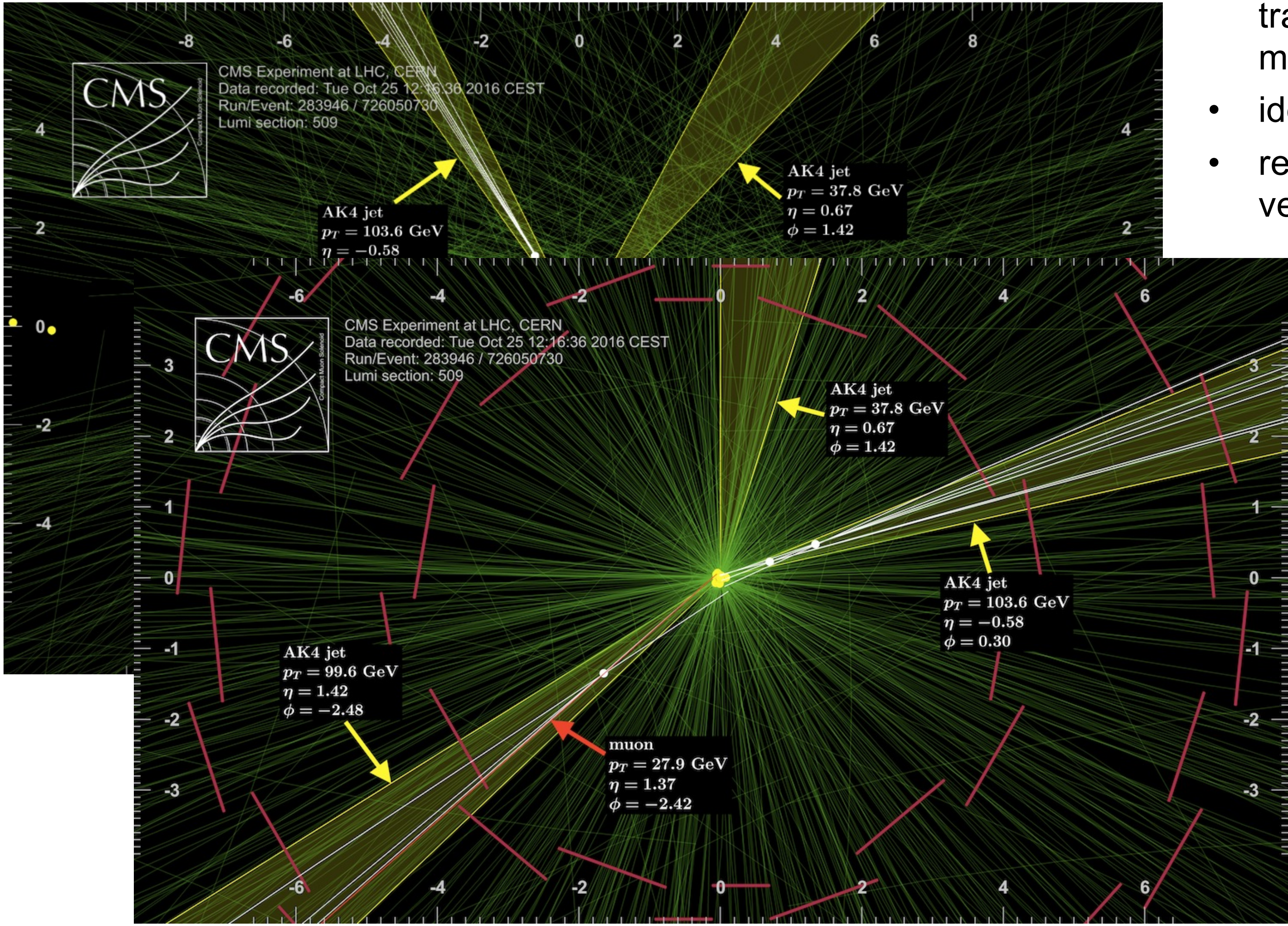
a typical
3 jets event

53 primary
vertices

typically 30 charged particles within the tracker acceptance per proton-proton collision and ~ 25-60 collisions per event: ~O(1000) charged particles per event need to be reconstructed.

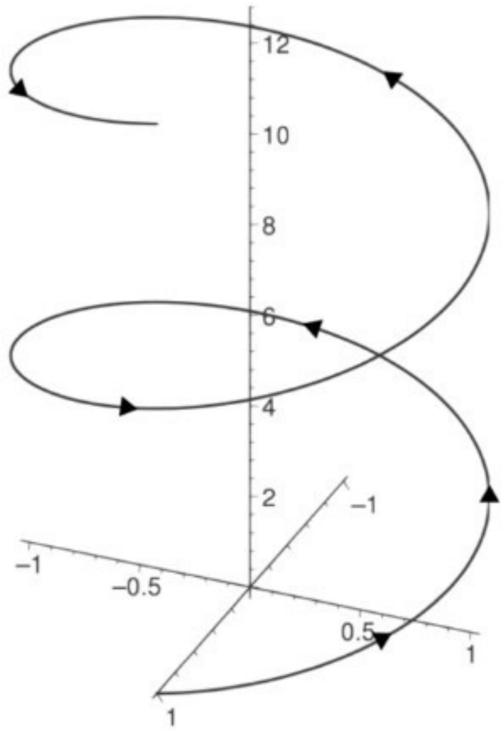
tracking: challenge

CMS DP Note 2017-032



- measure hits
- reconstruct tracks and measure p_T
- identify e, μ
- reconstruct vertices

trajectory parametrization



A helical trajectory can be expressed by 5 parameters,
but the parameterization is not unique.

Given one parameterization, we can always re-express the same trajectory in another parameterization.

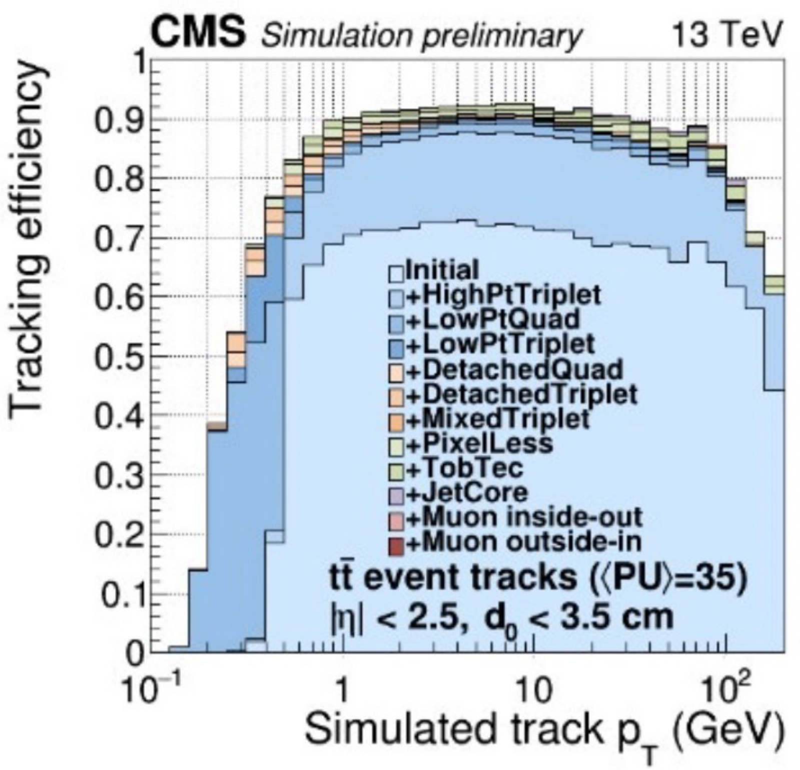
In general terms, the five parameters are:

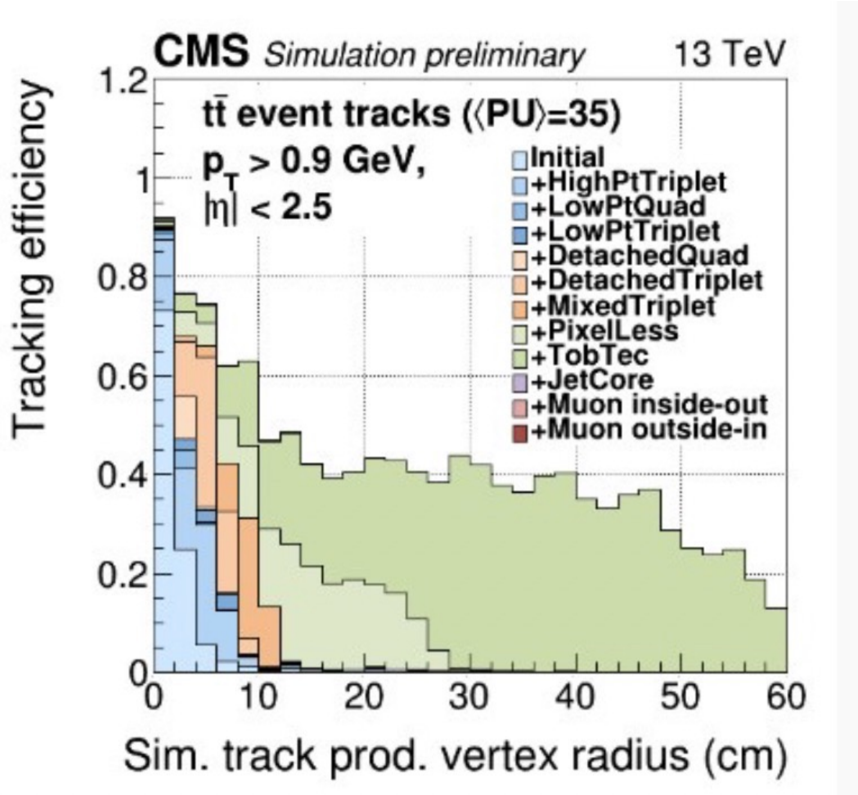
- **signed radius of curvature** (units of cm),
  which is proportional to particle charge
  divided by the **transverse momentum**, $\mathbf{p_T}$, (units of GeV);
- **angle of the trajectory** at a given point on the helix,
  in the plane **transverse to the beamline** (usually called $\boldsymbol{\varphi}$);
- **angle of the trajectory** at a given point on the helix
  with respect to the beamline (θ, or equivalently $\lambda = \pi/2 - \theta$),
  which is usually expressed
  in terms of pseudorapidity $\boldsymbol{\eta} = -\ln(\tan(\theta/2))$);
- offset or "**impact parameter**" relative to some reference point
  (usually the beamspot or a selected primary vertex),
  in the plane **transverse to the beamline** (usually called $\mathbf{d_{xy}}$);
- **impact parameter** relative to a reference point
  (beamspot or a selected primary vertex),
  along the beamline (usually called $\mathbf{d_z}$)

iterative tracking

In CMS, the track reconstruction is an iterative procedure:

- high quality tracks are reconstructed first
- their hits are removed
- and other tracks are reconstructed from the remaining hits

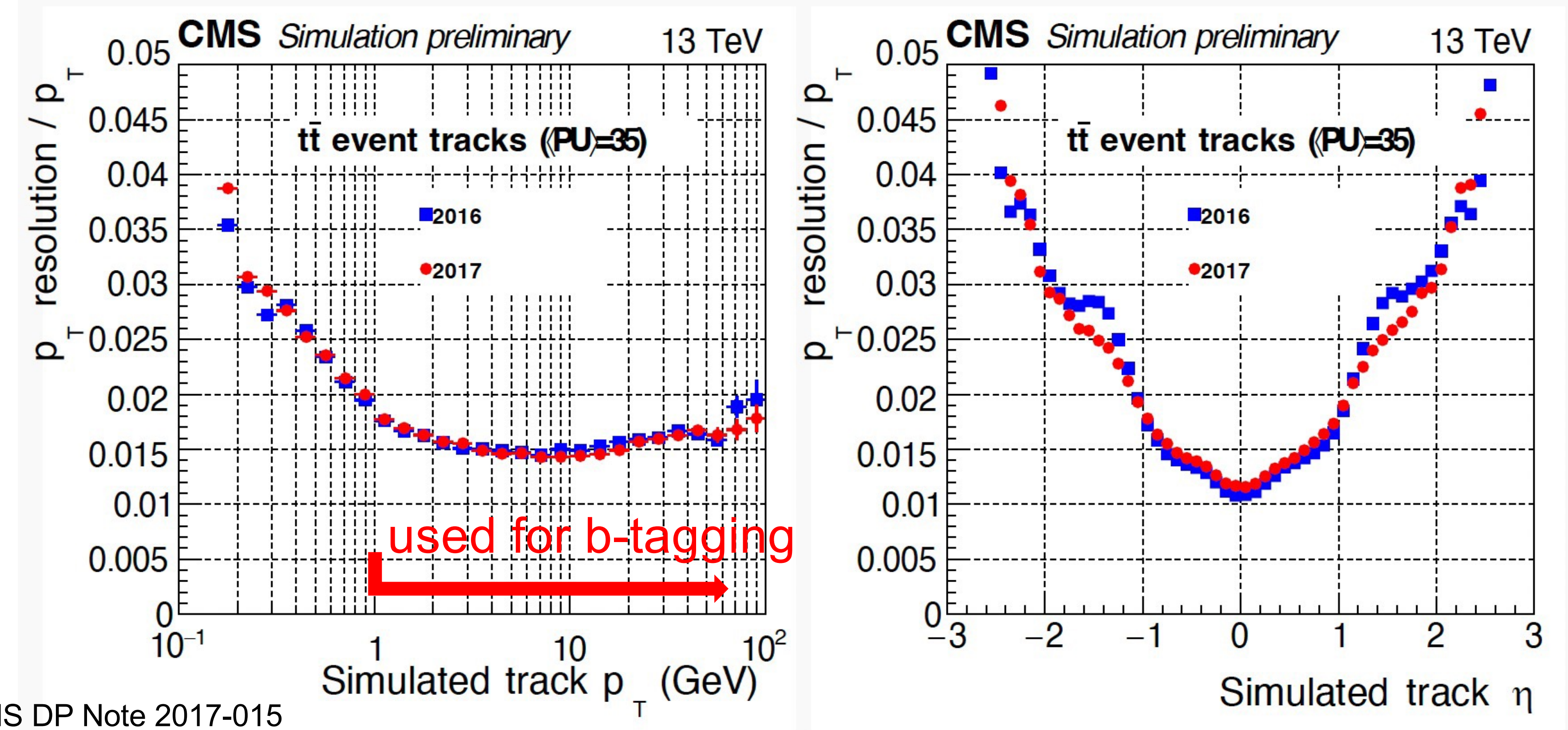




CMS DP Note 2017-015

tracking : momentum resolution

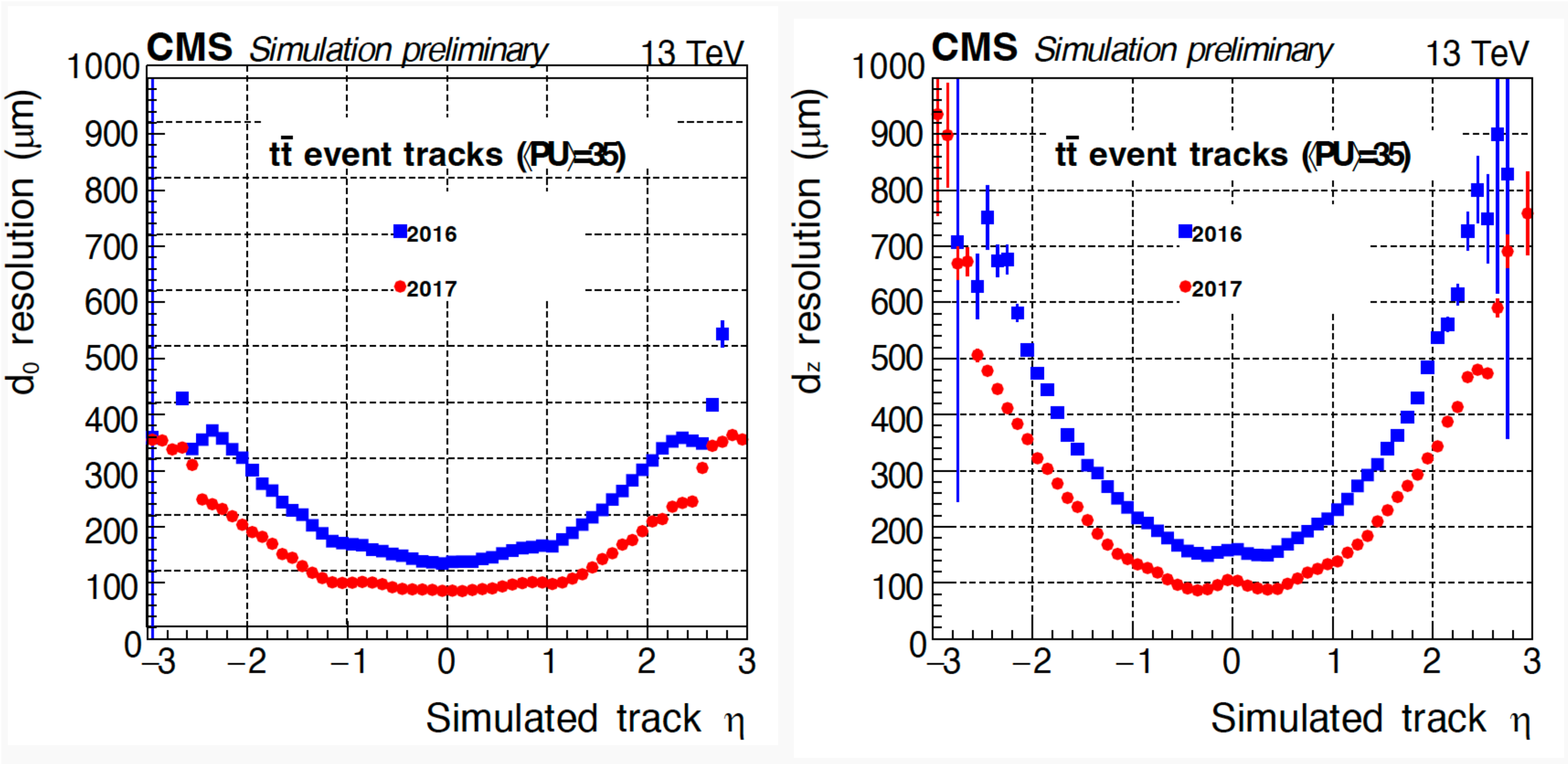
- p_T resolution is 3-4% for low p_T tracks (< 1 GeV) due to multiple scattering
- reaches ~1.5% at p_T ~10 GeV
- then degrades at high p_T due to less bending in magnetic field
- p_T resolution is best for central tracks



CMS DP Note 2017-015

tracking: impact parameter resolution

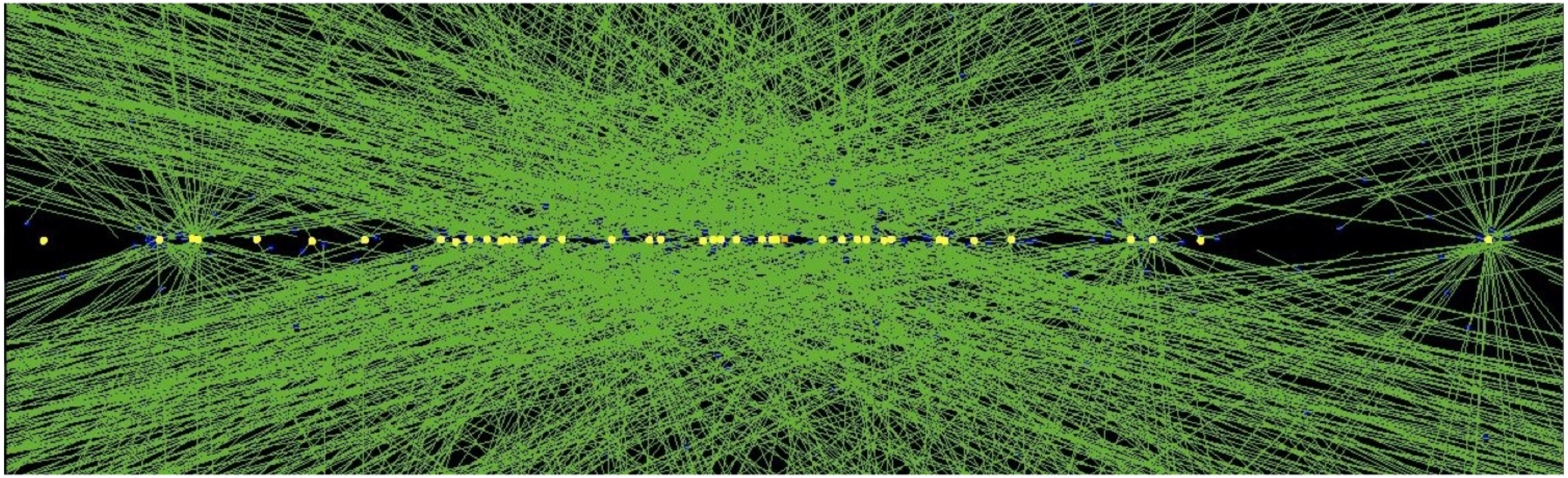
- impact parameter resolution ~80 (100) μm for central tracks in transverse plane (along the beam axis)
- but degrades in forward direction up to ~200 (500) μm



CMS DP Note 2017-015

primary vertex (PV)

- The reconstruction and correct identification of the vertex of the hard interaction in an event is of critical importance to correctly select the final state objects
- We also need to reconstruct as many of the PU vertices as possible to allow an efficient PU suppression in jets



CMS event with 40 pileup at 7 TeV

primary vertex (PV)

- The vertexing algorithm selects good tracks originating from the interaction region around the beam spot and clusters them according to the z coordinate of their point of closest approach to the center of the beam spot

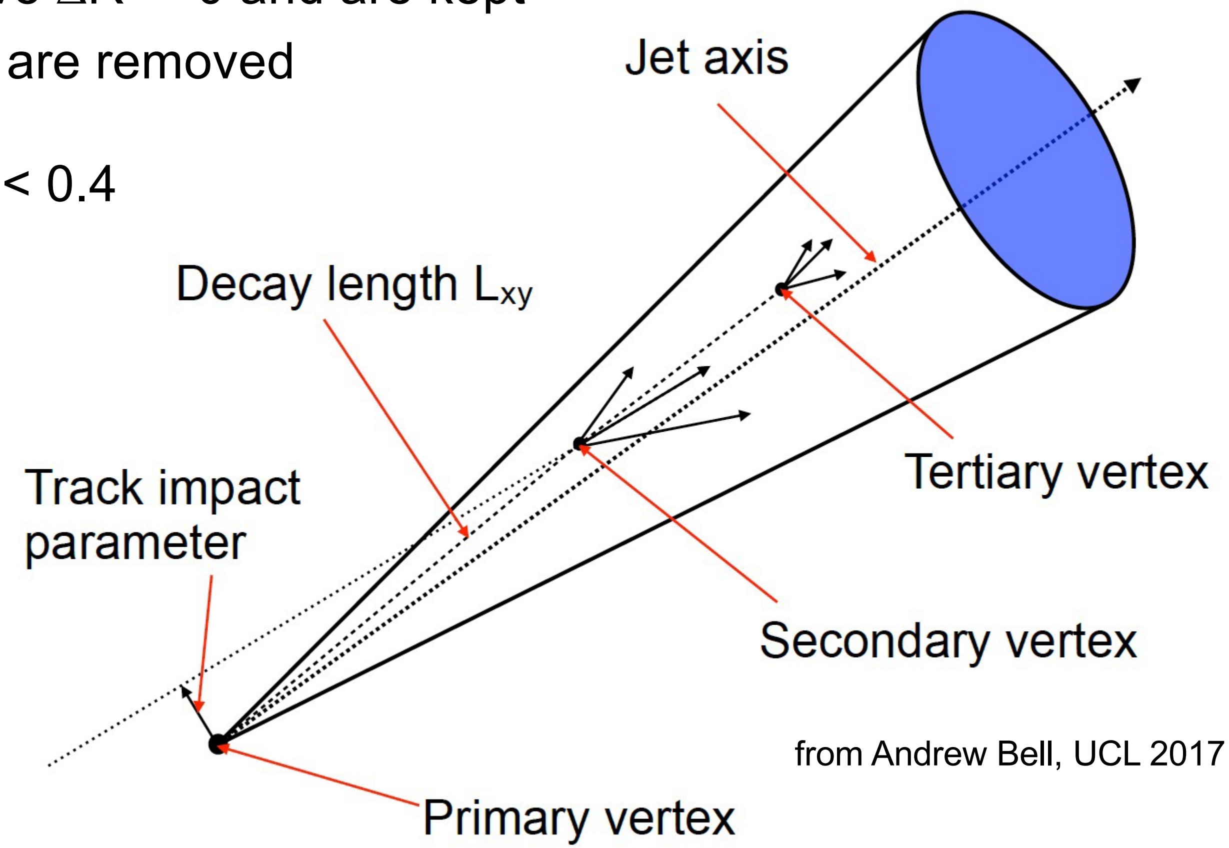
- When we cluster tracks into vertices, at the same time we want to resolve nearby vertices to separate the primary interaction from PU vertices and prevent vertex merging, but we also do not want to split genuine vertices into two, so the PV reconstruction proceeds as:
  - Track selection: based on deterministic annealing, inspired from theromodynamic, to group tracks with common « temperature »
  - Vertex fit: based on adaptive vertex fitter, weighting each track associated to a given vertex, downweighting the wrong tracks (outliers)
  - Merging of vertices starts for distance < 300 μm
  - Vertices closer than 100 μm can't be separated

- The typical PV reconstruction efficiency is ~99% for the hard interaction in ttbar events, and ~80% for PU vertices (up to a nb. of PU < 70)

b jet

- using the particle flow method (in CMS) particles are reconstructed with the full event information (tracks, calorimeter hits, muons hits)
- cluster particles into jets : if they are close in distance $\Delta R \min(p_T^1, p_T^2)^\alpha$ with cone size $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ and $\alpha = -1$: Anti-kT algorithm
- collinear particles have $\Delta R \rightarrow 0$ and are kept
- soft particles (low p_T) are removed

- narrow jet (AK4): $\Delta R < 0.4$

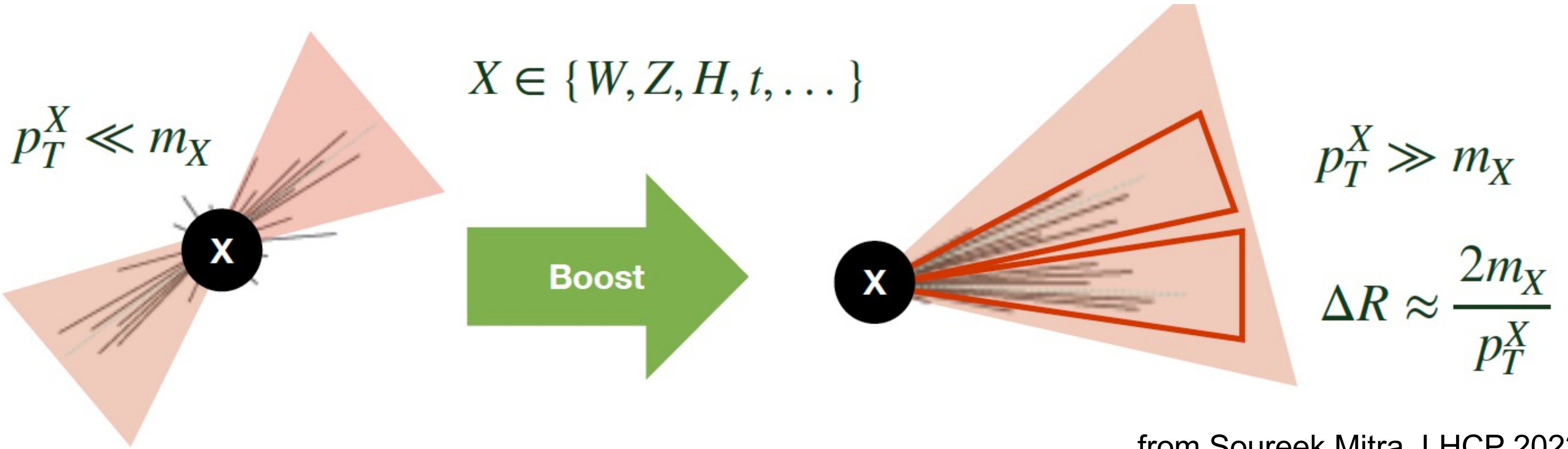


from Andrew Bell, UCL 2017

b jet

- using the particle flow method (in CMS) particles are reconstructed with the full event information (tracks, calorimeter hits, muons hits)
- cluster particles into jets : if they are close in distance $\Delta R \min(p_T^1, p_T^2)^\alpha$ with cone size $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ and $\alpha = -1$: Anti-kT algorithm
- collinear particles have $\Delta R \to 0$ and are kept
- soft particles (low p_T) are removed

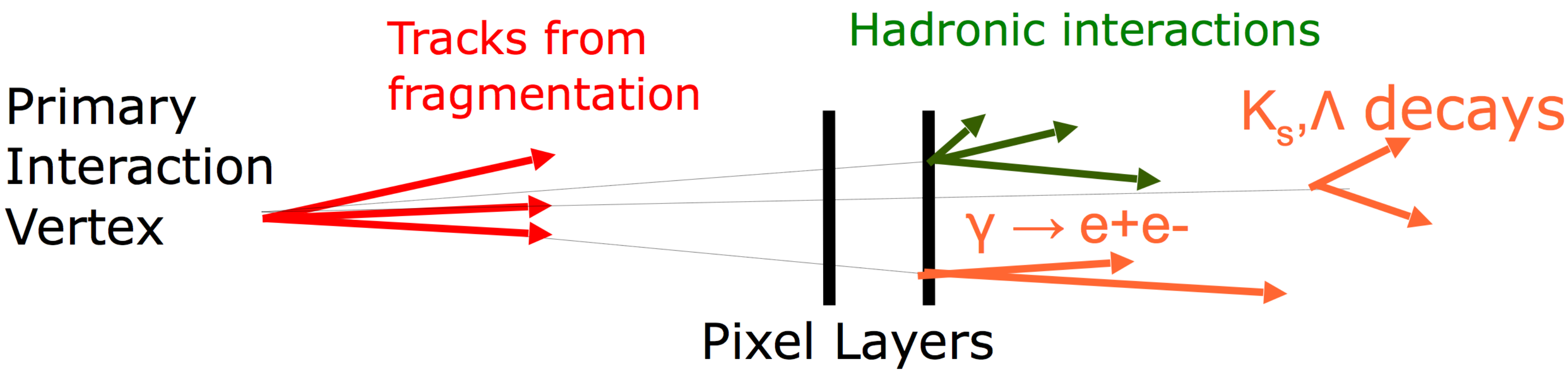
- wide jet (AK8): $\Delta R < 0.8$ for boosted objects as for high p_T $H \to bb$, or W' or Z' searches (needs $\Delta R = 0.8$ for $p_T(H) > 300$ GeV)



from Soureek Mitra, LHCP 2022

versus light jet

- in a light-flavour jet (from u, d, s quark or gluon), most tracks come directly from quark fragmentation
- but sometime can result in a displaced vertex and look like a b-jet:
  - interactions with the detector material
  - photon conversions
  - long-lived $K^0_s \rightarrow \pi^+\pi^-$ or $\Lambda \rightarrow p\pi^-$
  - badly measured tracks (with poor resolution, or fakes)
  - track from pileup, originating from another primary vertex

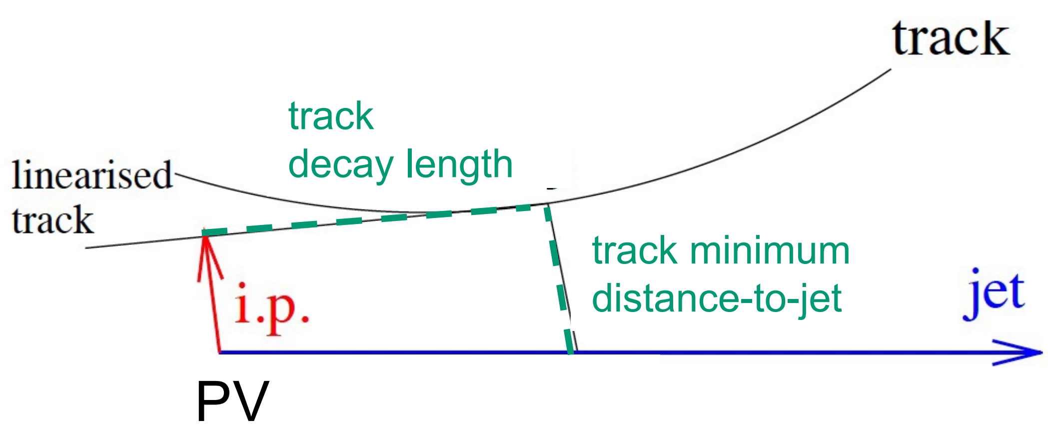


from Andrew Bell, UCL 2017

track selection

some track selection can help to discriminate between those coming from b or c decay and from ordinary jet fragmentation

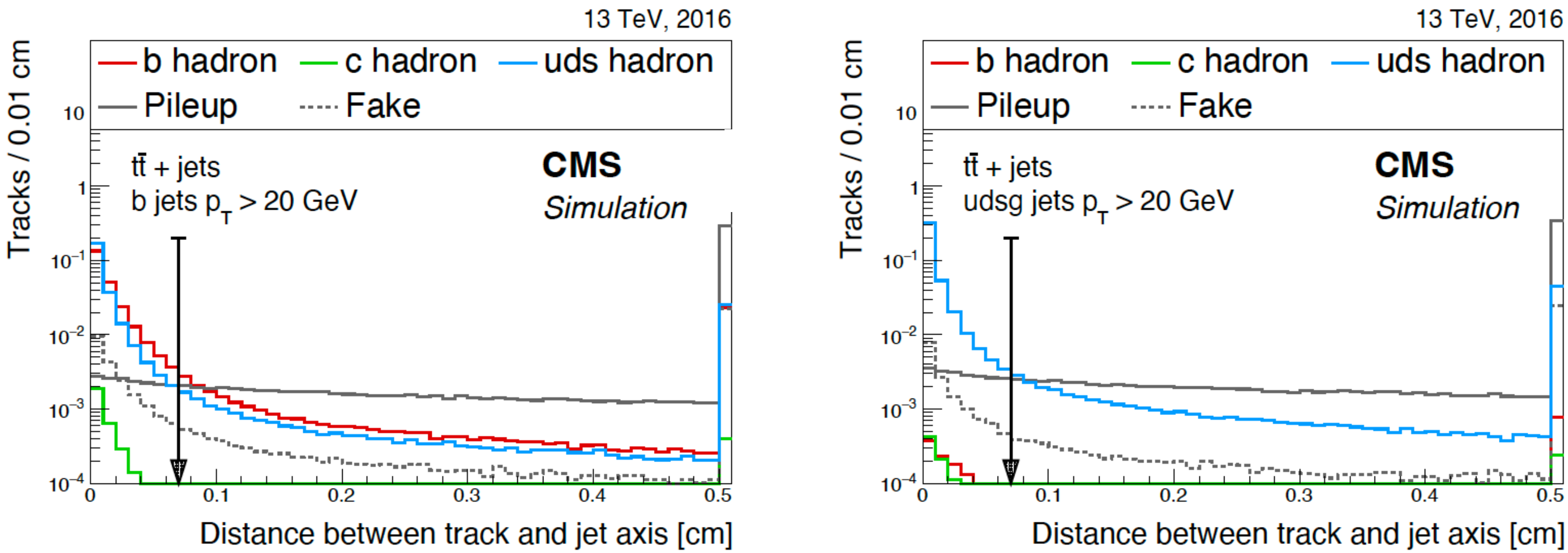
- quality cuts: p_T > 1 GeV, χ^2/dof < 5, ≥ 1 pixel hits
- against pileup:
  - $|d_{xy}| < 2$ mm
  - track distance-to-jet axis: closest 3D distance from track to jet < 0.7 mm
  - track decay length: corresponding distance to the primary vertex < 5 cm



track selection

some track selection can help to discriminate between those coming from b or c decay and from ordinary jet fragmentation

- quality cuts: p_T > 1 GeV, χ^2/dof < 5, ≥ 1 pixel hits
- against pileup:
  - $|d_{xy}|$ < 2 mm
  - track distance-to-jet axis: closest 3D distance from track to jet < 0.7 mm
  - track decay length: corresponding distance to the primary vertex < 5 cm

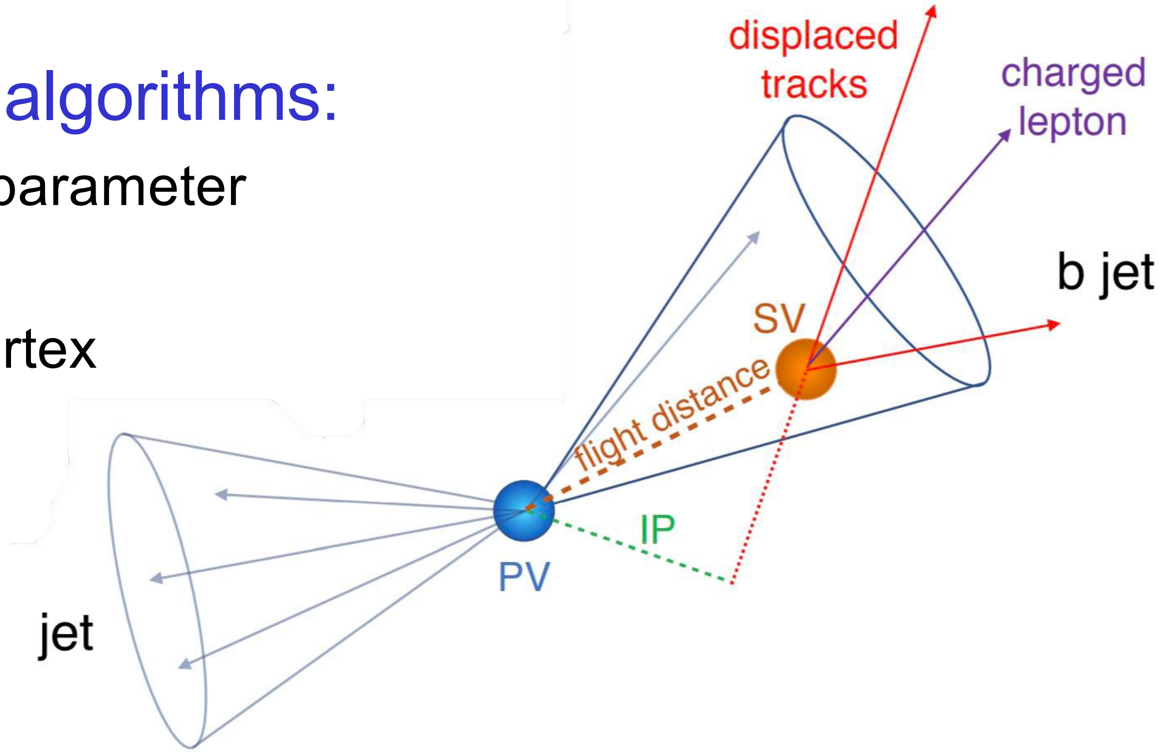


CMS, JINST 13 (2018) P05011

Chap. III
how to tag ?

tagging algorithms

- Basic tagging algorithms:
  - track impact parameter
  - soft lepton
  - secondary vertex

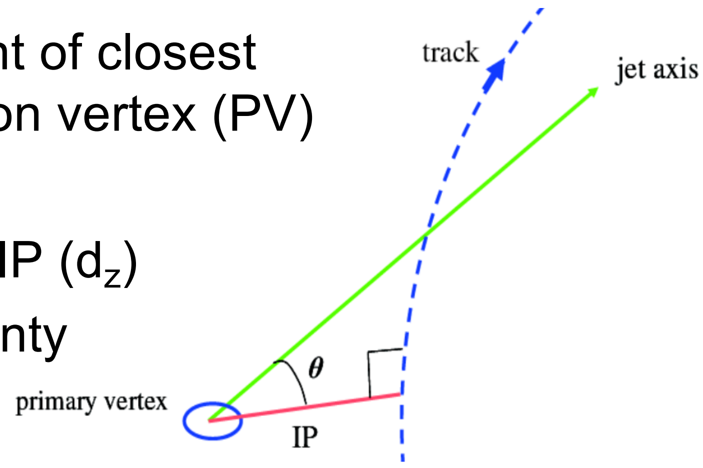


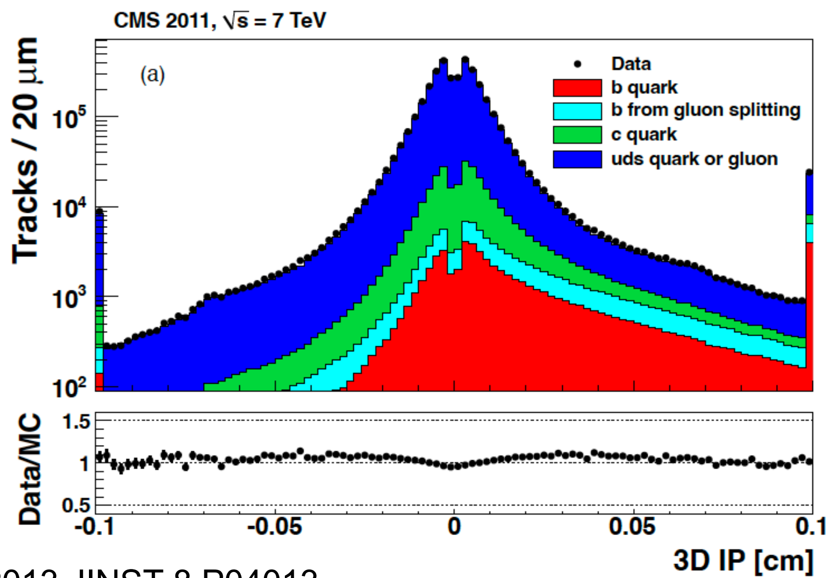
- Multivariate algorithms
  - Boosted Decision Trees
  - Neural Networks
  - Deep Learning

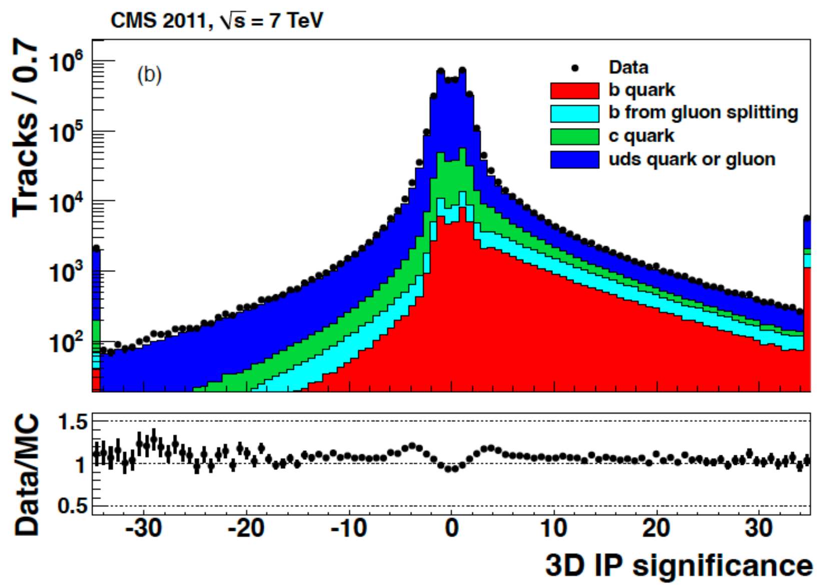
  …

impact parameter

- impact parameters (IP) defined as signed point of closest approach with respect to the primary interaction vertex (PV)
  - in transverse plane (d_{xy} or d_0)
  - or in 3D, from which infer the longitudinal IP (d_z)
- IP significance is the IP divided by its uncertainty
- signed IP: positive if it crosses the jet axis downstream from the PV, otherwise negative
- infer templates (from simulation) for b, c and light jets



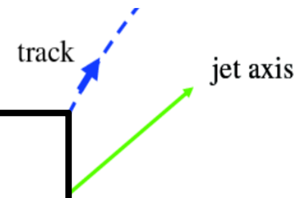




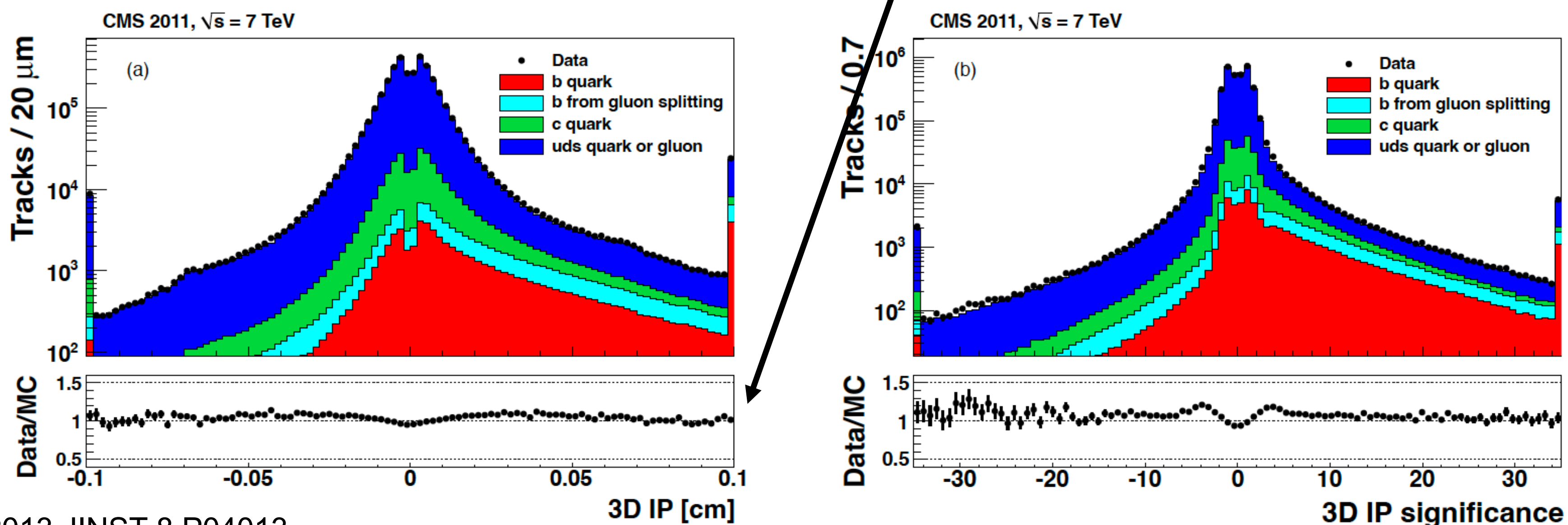
CMS, 2013 JINST 8 P04013

impact parameter

- impact parameters (IP) defined as signed point of closest approa
  - in t
  - or i
- IP sign
- signed downst
- infer te



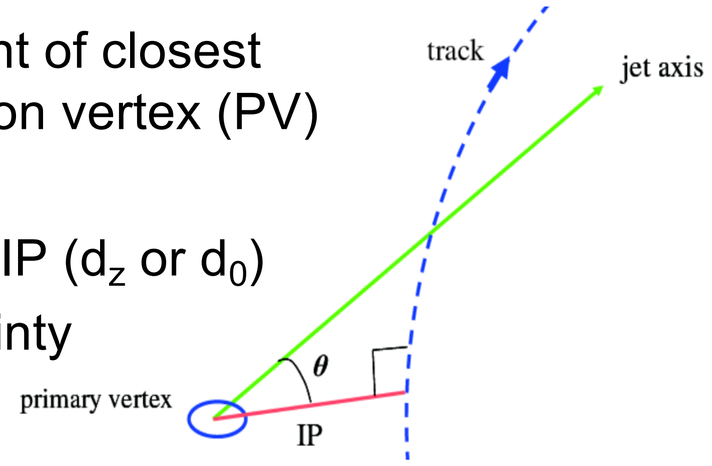
note the rather good data – MC agreement which is necessary to use further such basic criteria in higher level taggers !

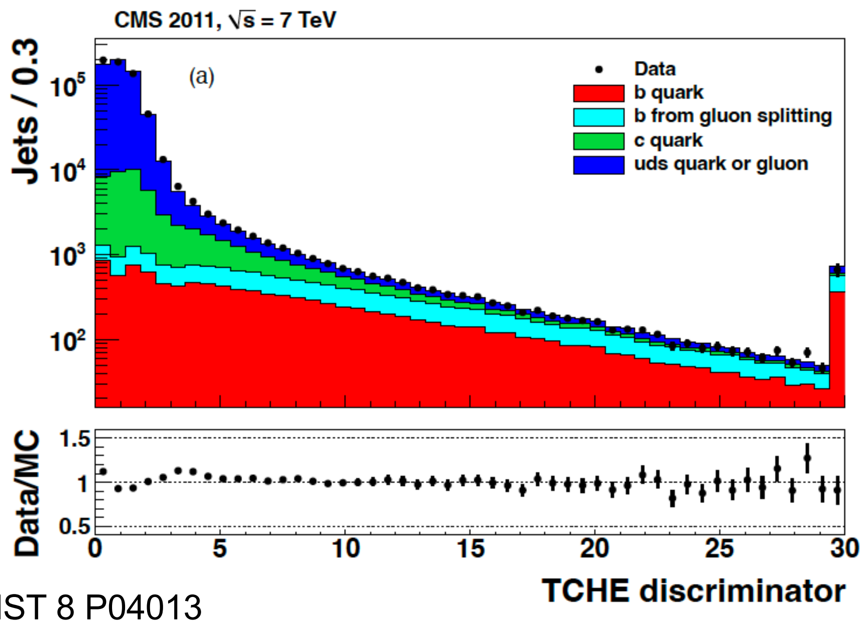


CMS, 2013 JINST 8 P04013

impact parameter

- impact parameters (IP) defined as signed point of closest approach with respect to the primary interaction vertex (PV)
  - in transverse plane (d_{xy} or d_0)
  - or in 3D, from which infer the longitudinal IP (d_z or d_0)
- IP significance is the IP divided by its uncertainty
- signed IP: positive if it crosses the jet axis downstream from the PV, otherwise negative
- infer templates (from simulation) for b, c and light jets





CMS, 2013 JINST 8 P04013

can order tracks in a jet by decreasing IP significance value and consider the 2nd or 3rd highest value (as in CMS)

(here the 2nd highest IP significance

more IPs: likelihood ratio

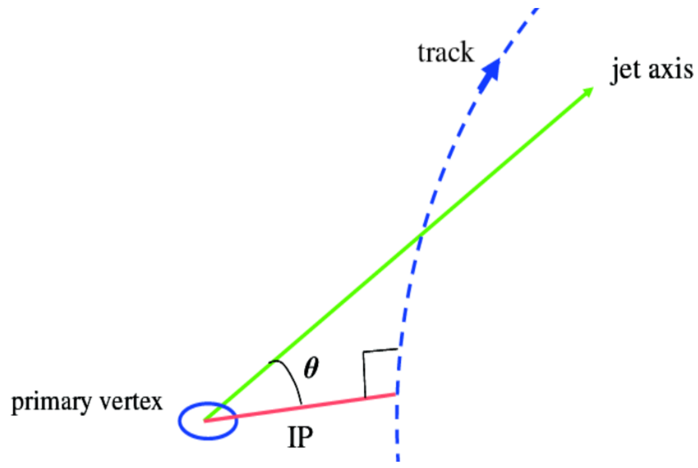
from Andrew Bell, UCL 2017

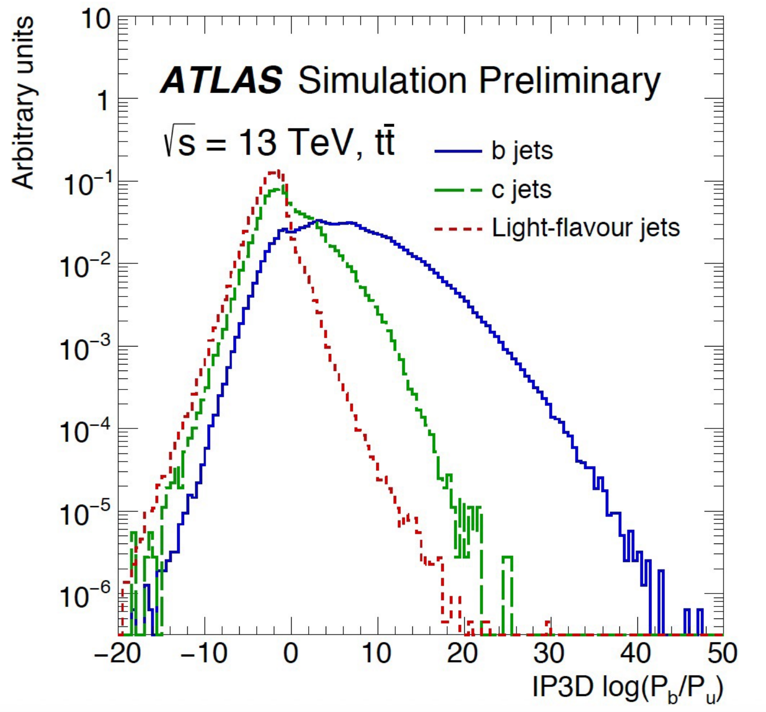
$$\log(P_i/P_j) = \log\left(\frac{\prod_{m=1}^{N} PDF_i(IP_m)}{\prod_{m=1}^{N} PDF_j(IP_m)}\right)$$

Product over all associated tracks

Probability for jet to have flavour j

PDF for track of flavour j, with IP_m





combine all IP significances into a likelihood ratio (as in ATLAS)

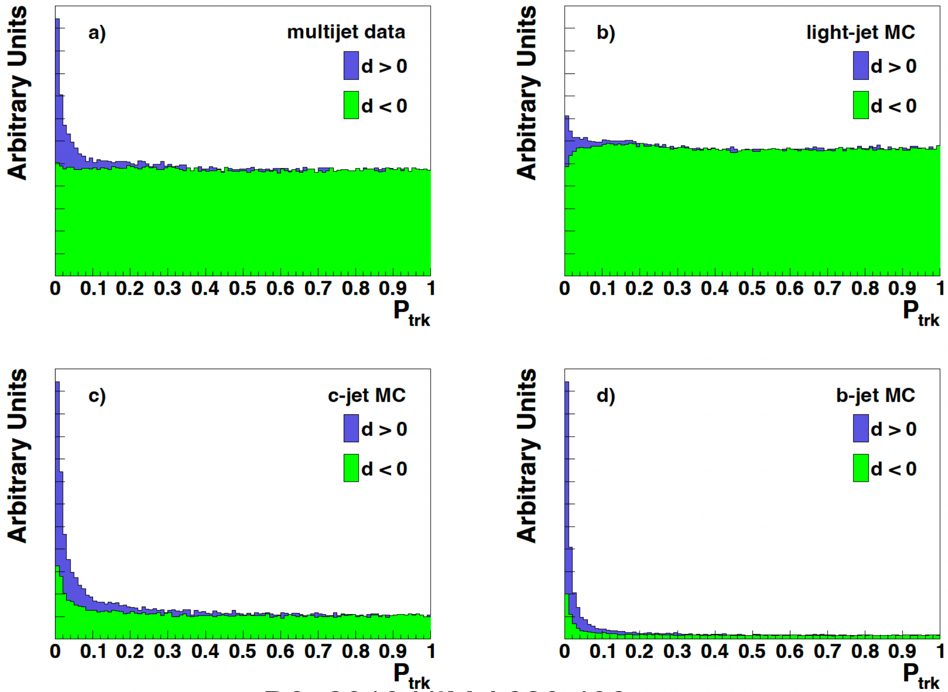
more IPs: ``probability''

- originally developped at LEP (ALEPH, DELPHI)
- for each track, compute the probability P_{tr} that it comes from the PV, using the IP significance
- use tracks with negative IP to calibrate it:

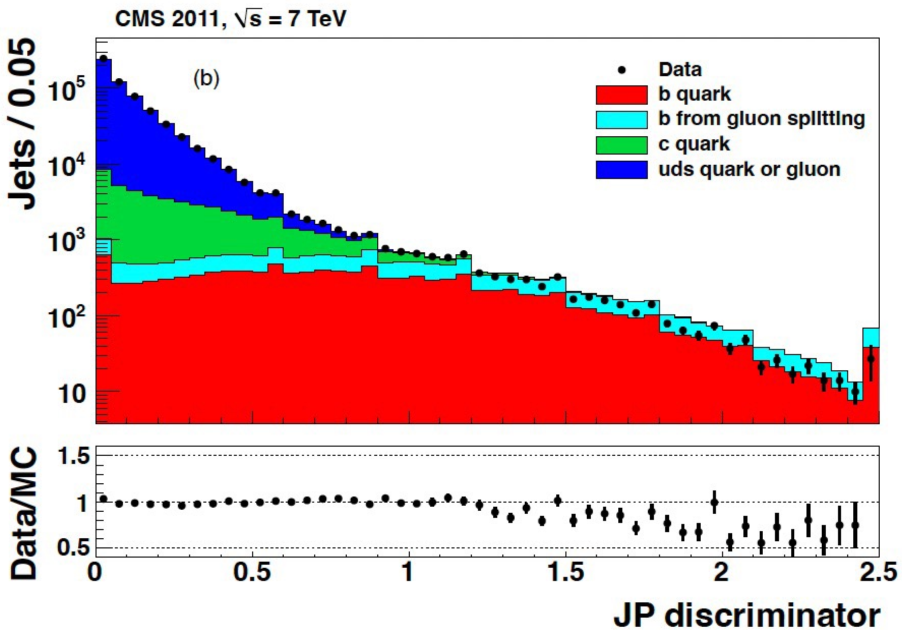
$$P_{\mathrm{tr}} = \int_{-\infty}^{-|\mathrm{IP}|/\sigma} \mathcal{R}(s)\,\mathrm{d}s.$$

- use tracks with positive IP to estimate a Jet ``Probability'' P_{j}, using the product Π of all P_{tr} in the jet:

$$P_{\mathrm{j}} = \Pi \sum_{\mathrm{tr}=0}^{N-1} \frac{(-\ln \Pi)^{\mathrm{tr}}}{\mathrm{tr}!},$$



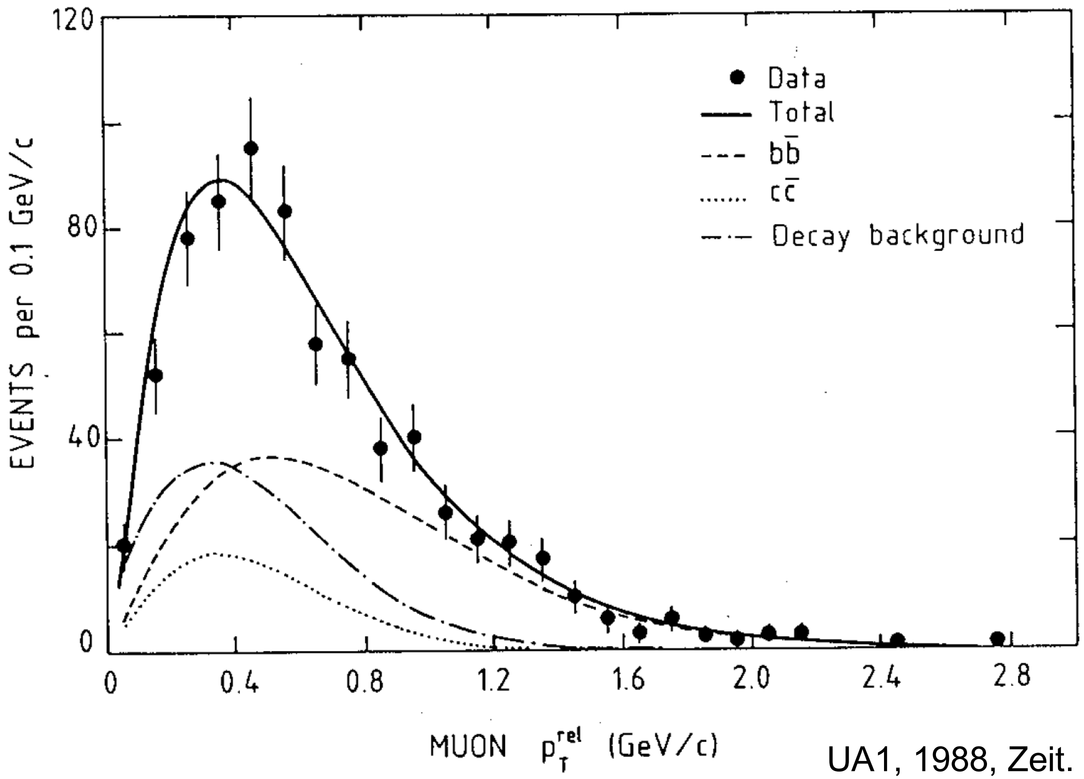
D0, 2010 NIM A620 490



CMS, 2013 JINST 8 P04013

soft lepton tag

- 23% b quark decays with a muon (23% with an electron)
- thanks to the high b mass, the p_T of the lepton with respect to its jet axis (p_T^{rel}) is larger in b decay than in c decay, or in light jets:
- this was the primary b jet identifier, before the arrival of pixel detectors



UA1:
first b jet
measument
at collider

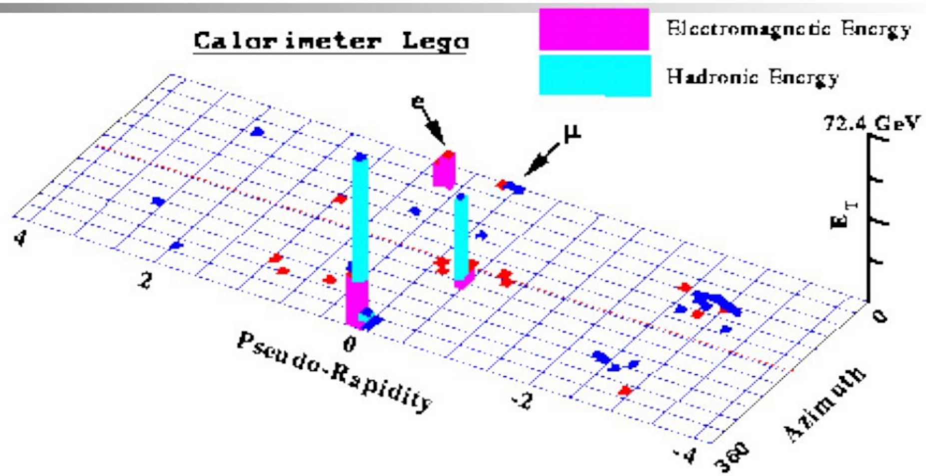
UA1, 1988, Zeit. Phys. C37, 489

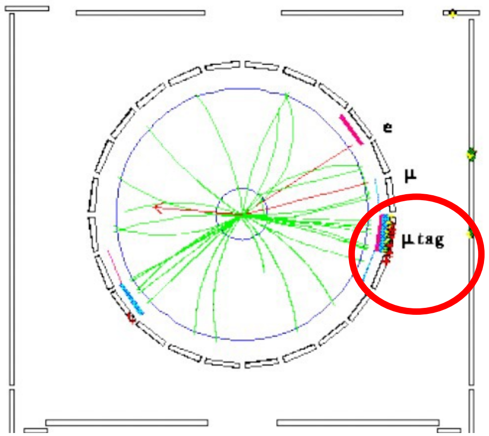
soft lepton tag

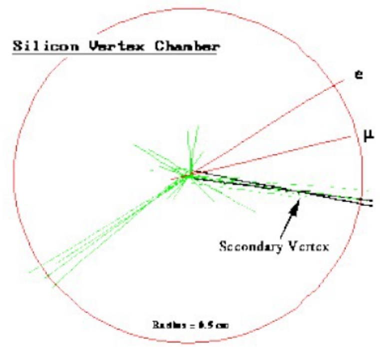
- was a key for the **top quark discovery** in CDF and D0

The Golden Event (CDF)

- "DPF event"
  - Oct. 22, 1992
  - That year ALL candidate events were "named"
  - eμ + 2 jet event
    - 1 jet tagged by both SLT and SVX
  - Decide not to declare discovery on 1 event
    - D0 similar experience



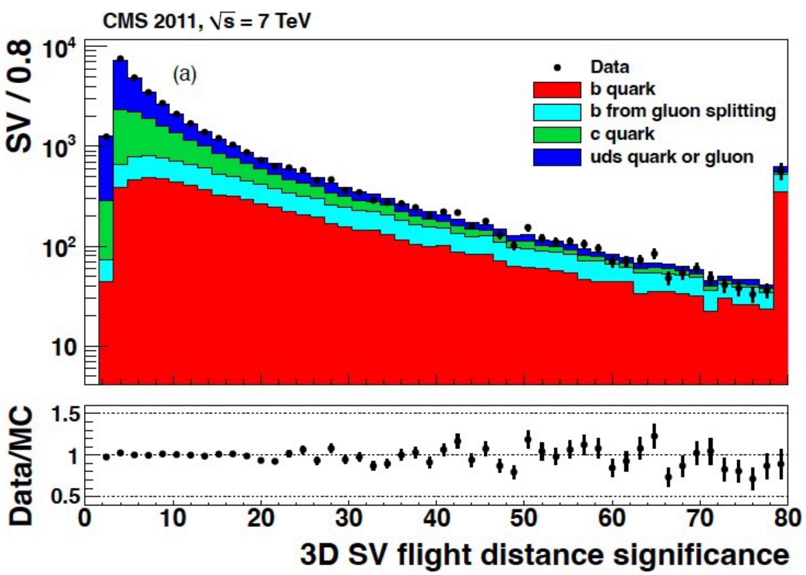


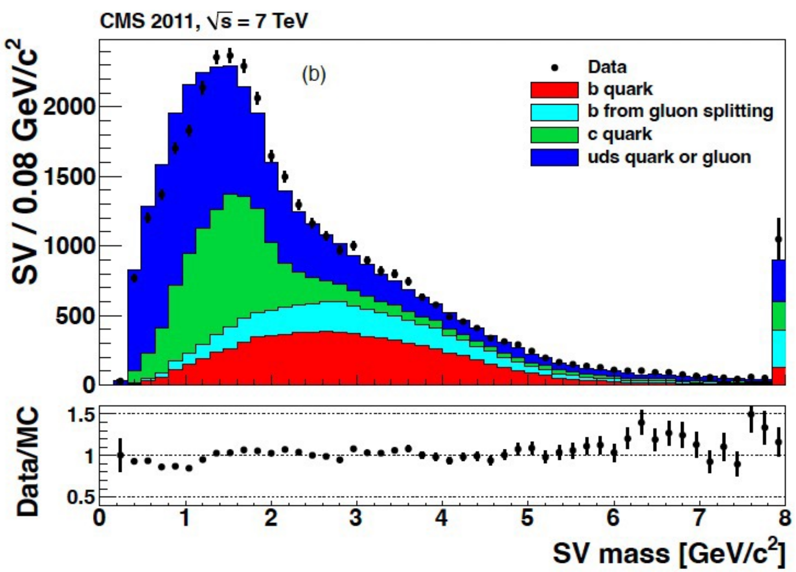


from Robert Roser, CTEQ school 2012

secondary vertex (SV)

- several reconstruction methods can be used: here the inclusive vertex finding is summarized (see CMS, JINST 13 (2018) P05011). It was originally introduced to measure the correlation between close b decays:
  - use all tracks in the event (not only those in a jet), cluster those which are close in distance and far from the PV
  - then apply an adaptive vertex fitter, drop any SV if it is too close from another SV (and less well measured)
  - remove tracks more compatible with the PV than with a SV
  - then refit the SVs and apply a last cleaning

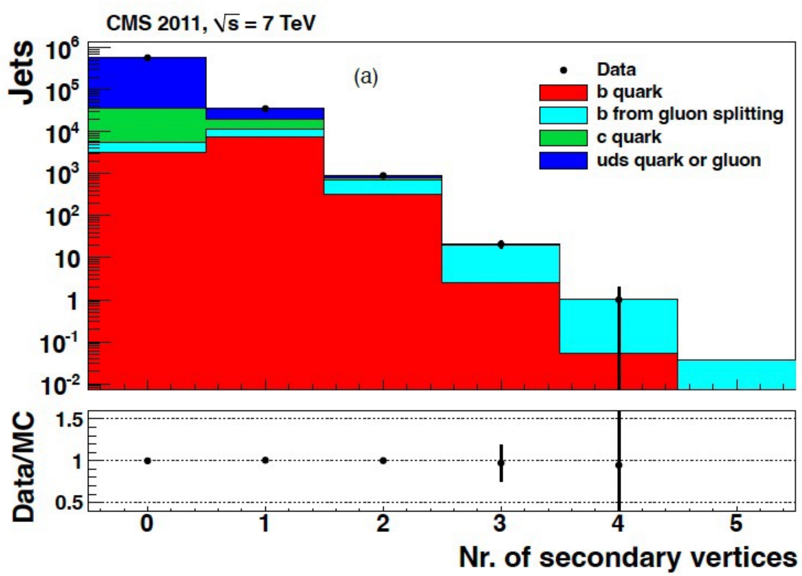


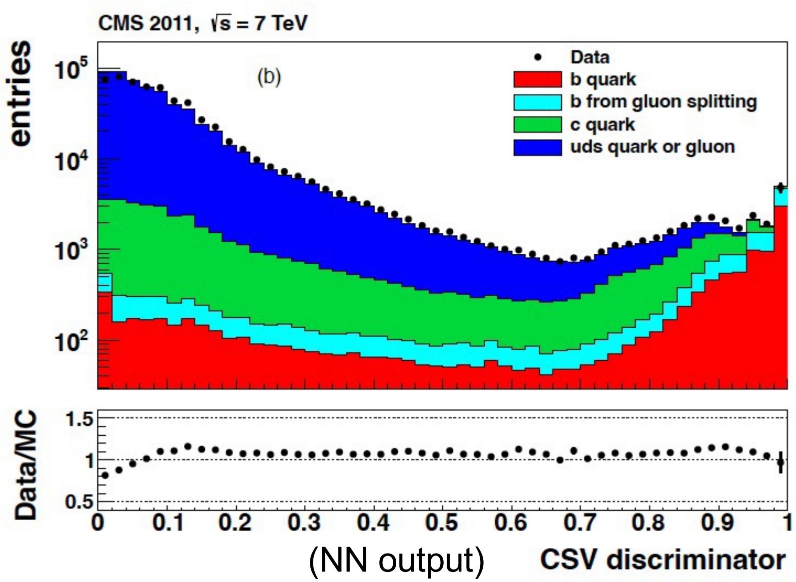


CMS, 2013 JINST 8 P04013

SV: from a simple tagger to a Neural Network

- variable used:
  - nb. of SVs and track multiplicity at SV
  - decay length significance: PV-SV distance divided by its error, if several SVs, can consider 2nd or 3rd highest value -> simple SV tag
  - SV mass: invariant mass of tracks attached to a SV
  - SV energy ratio: energy sum of tracks at SV / that of all tracks in jet
  - use also informations from single tracks at high IP in the jet
  - then can combine everything into a Neural Network





CMS, 2013 JINST 8 P04013

jet flavour and efficiency

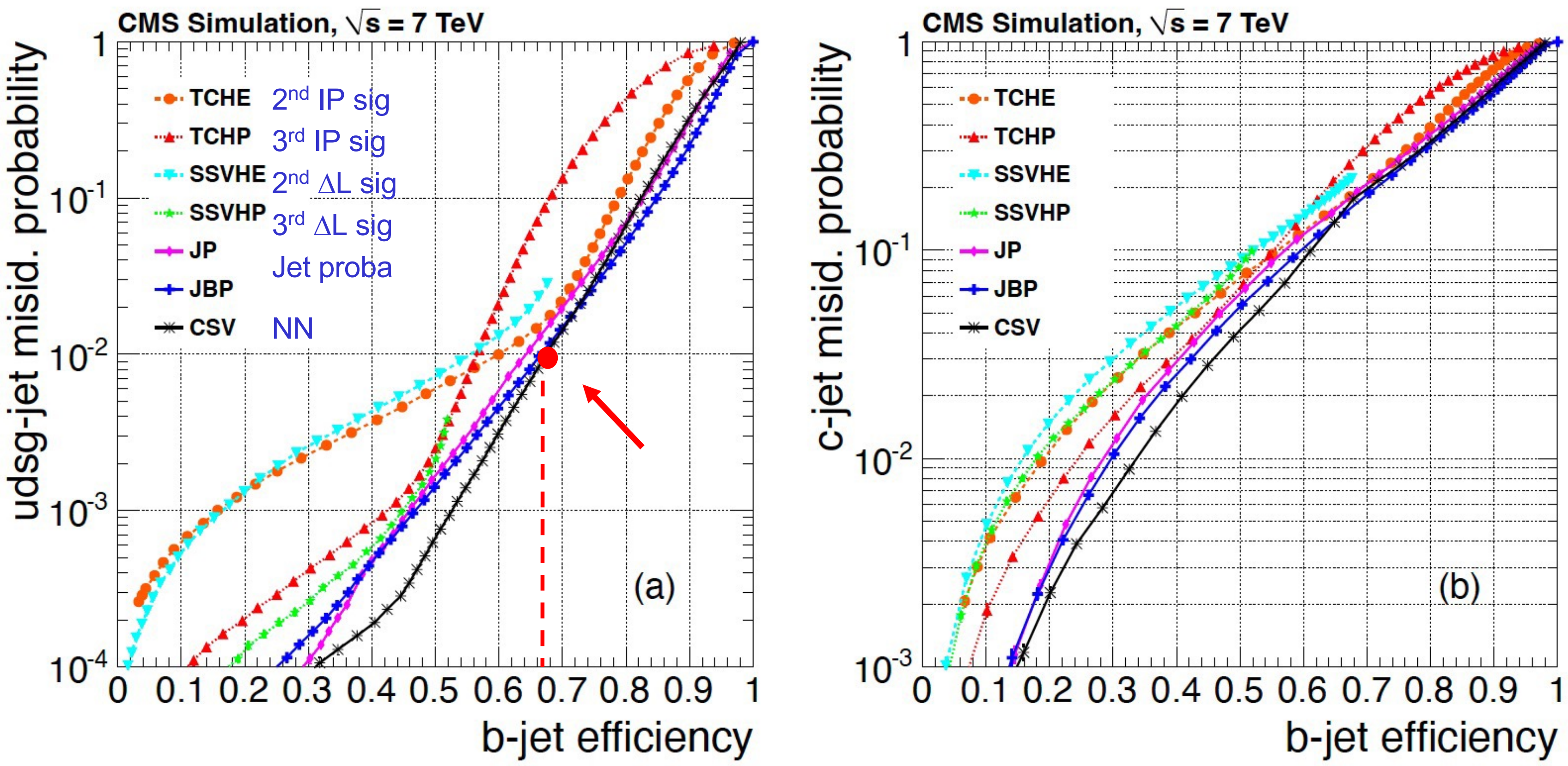
- in the simulation, a b jet is defined as containing a final state b-hadron (within $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ < 0.4 for AK4 jet),
- if there are no b-hadron, a c jet is looked for
- if there are no b- or c-hadron, generated jets are considered (using final state generated particles except neutrinos, with the same AK clustering): if a generated jet is at $\Delta R < 0.25$ from the reconstructed jet and with $p_T > 8$ GeV, it is defined as a light jet
- remaining jets are undefined (can be from pileup, or from a lepton, or fake)

- the efficiency to tag a b jet is the number of tagged b jets divided by the number of reconstructed b jets (same definition for c jets)
- the mistag rate, or misidentification probability, is also an efficiency but often applied to light jets: nb. of tagged light jets / nb. of reconstructed light jets

- the performance curve (or ROC curve) is a 2D representation of the tagging efficiencies for two different flavours

performance of early taggers

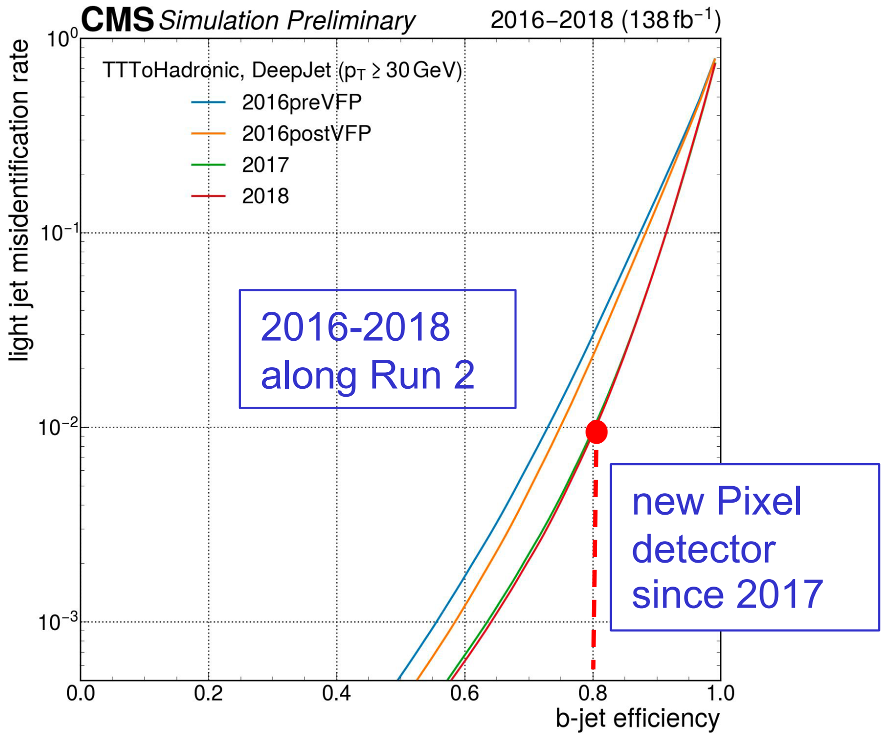
- simple taggers could reach ~50% b-tag efficiency for 1% light mistag
- NN reached 66% b-tag efficiency for 1% light and 15% charm mistag
- the user can rely for each tagger on dedicated operating points, corresponding to 10% (loose), 1% (medium), 0.1% tight light mistag



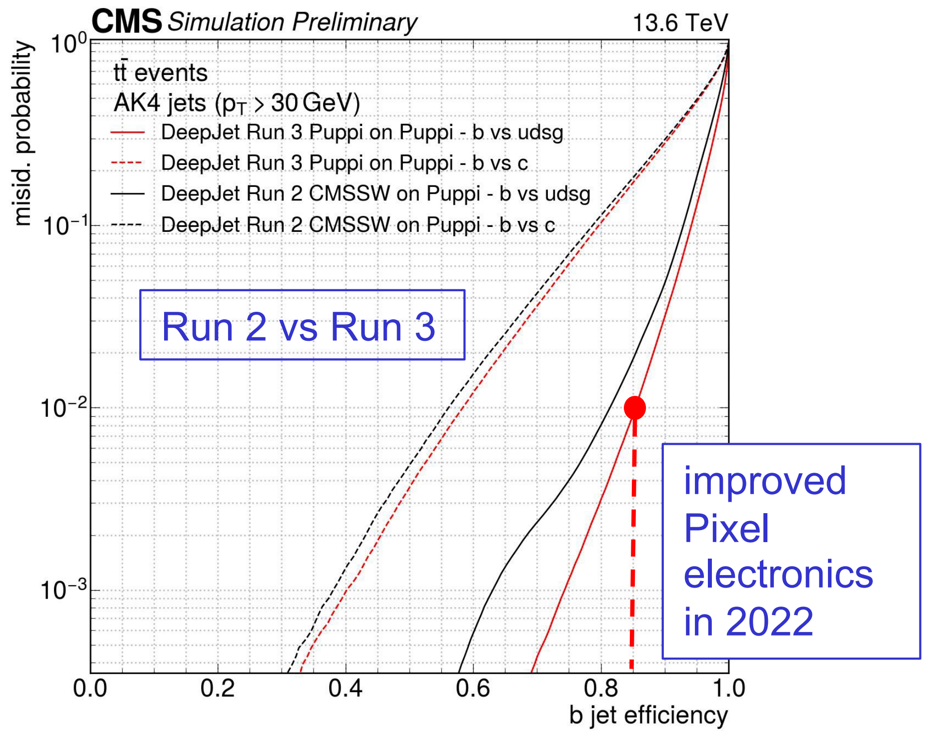
CMS, 2013 JINST 8 P04013

multi-variate taggers

- fastly growing evolution of the technology
  - Boosted Decision Trees
  - Neural Networks
  - Deep Neural Networks, with increased SV and IP informations and larger number of involved tracks
  - Graph Neural Networks, …



CMS DP Note 2023-005



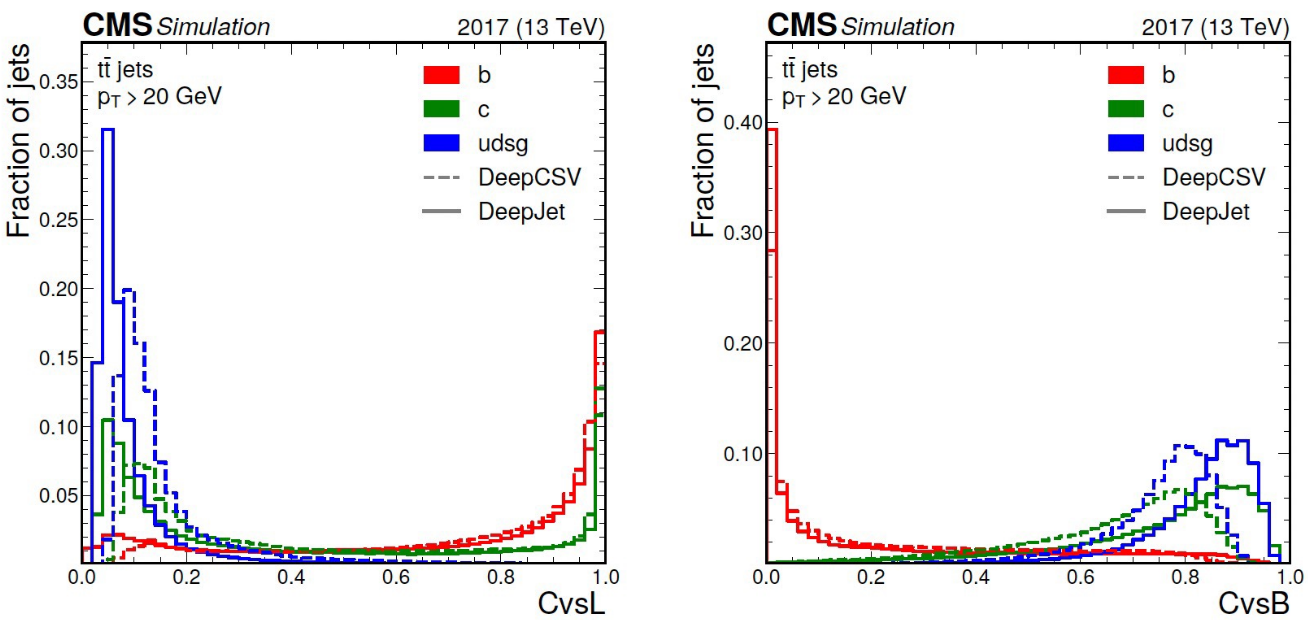
CMS DP Note 2023-012

charm tagging

charm properties are in between b and light flavour properties

- so 2 discriminants have to be used, one against light (CvsL), the other against b (CvsB)
- using Deep NN taggers, one get probabilities for a jet to contain either b, bb, c, uds and g which can be combined for c-tagging

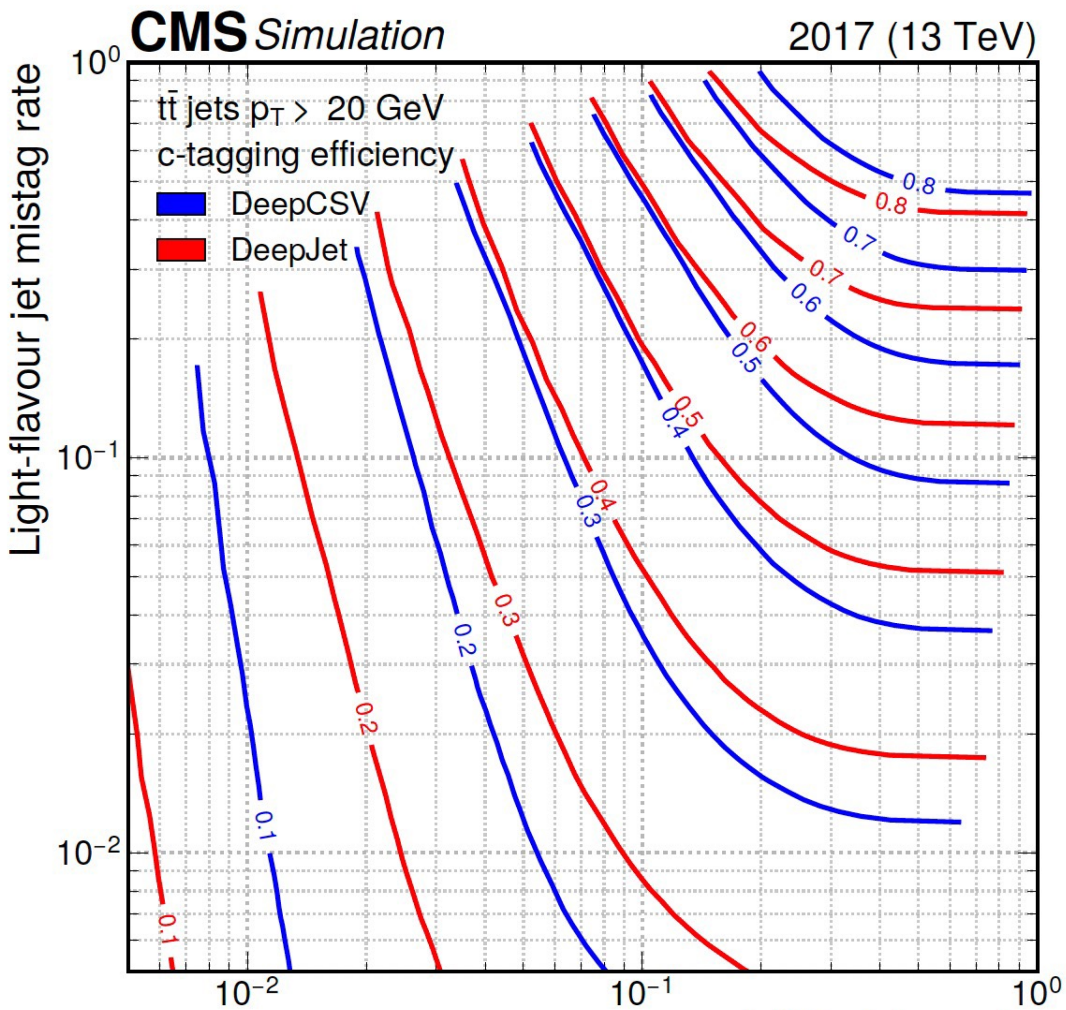
Tagger	BvsC/L	CvsB	CvsL
DeepCSV	$P(\mathrm{b})+P(\mathrm{bb})$	$\frac{P(\mathrm{c})}{P(\mathrm{c})+P(\mathrm{b})+P(\mathrm{bb})}$	$\frac{P(\mathrm{c})}{P(\mathrm{c})+P(\mathrm{udsg})}$
DeepJet	$P(\mathrm{b})+P(\mathrm{bb})+P(\mathrm{b_{lep}})$	$\frac{P(\mathrm{c})}{P(\mathrm{c})+P(\mathrm{b})+P(\mathrm{bb})+P(\mathrm{b_{lep}})}$	$\frac{P(\mathrm{c})}{P(\mathrm{c})+P(\mathrm{uds})+P(\mathrm{g})}$



CMS, 2022 JINST 17 P03014

charm tagging

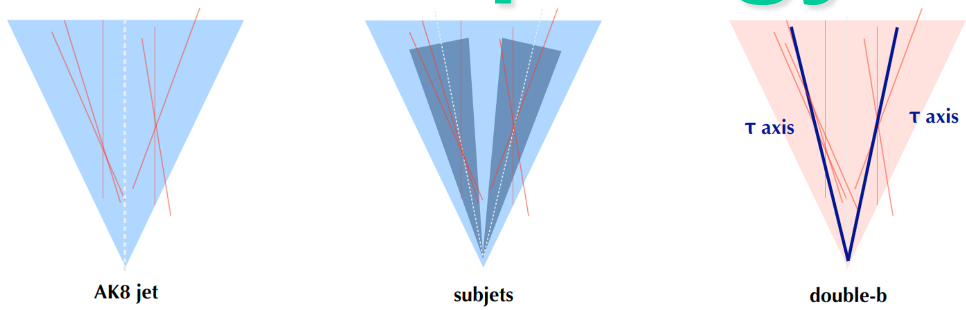
- by choosing appropriate selections (operating points) on CvsB and CvsL, one can get different c-tagging efficiencies



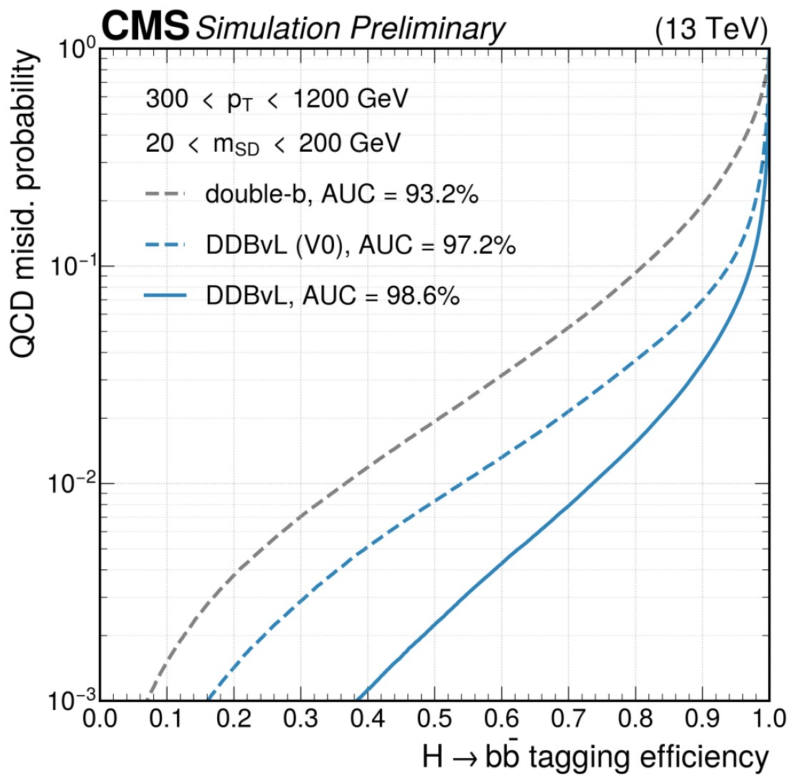
CMS, 2022 JINST 17 P03014

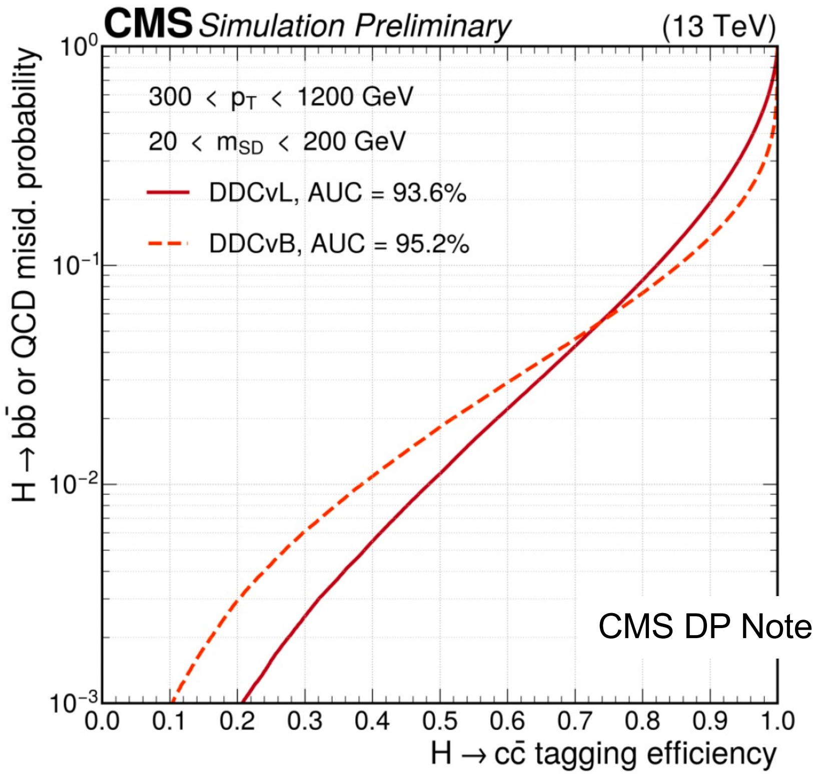
b (or c) jets in boosted topology

- developed at LHC,
  jet substructure technics can be applied for fat (AK8) jets at high p_T containing 2 close b quarks.



- the first b-taggers used the existing AK4 taggers, but applied to the subjet components. Dedicated double b-taggers were proposed to exploit more information from all the tracks, SVs and subjets (BDT, and now deep NN), while being less sensitive to to the subjet pair invariant mass

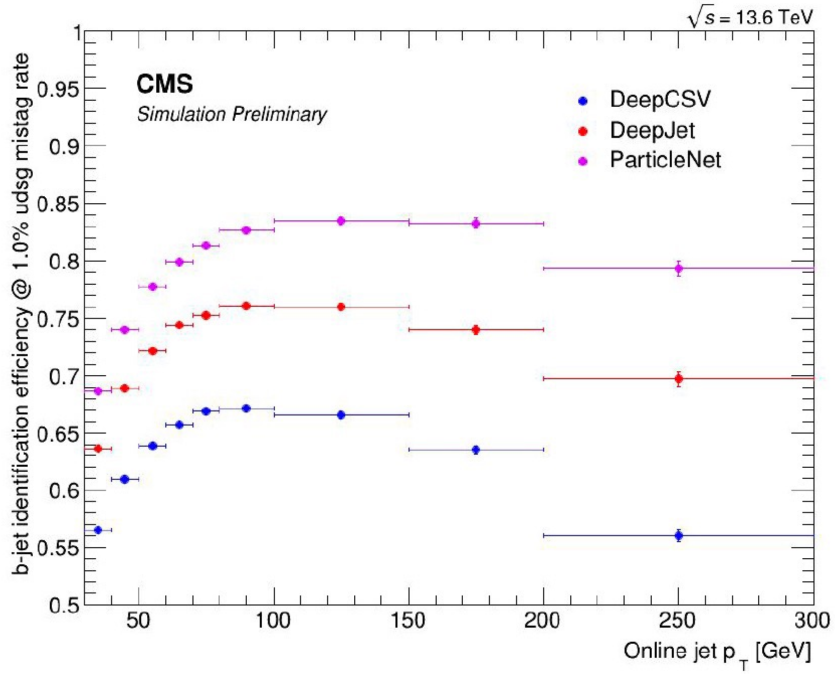


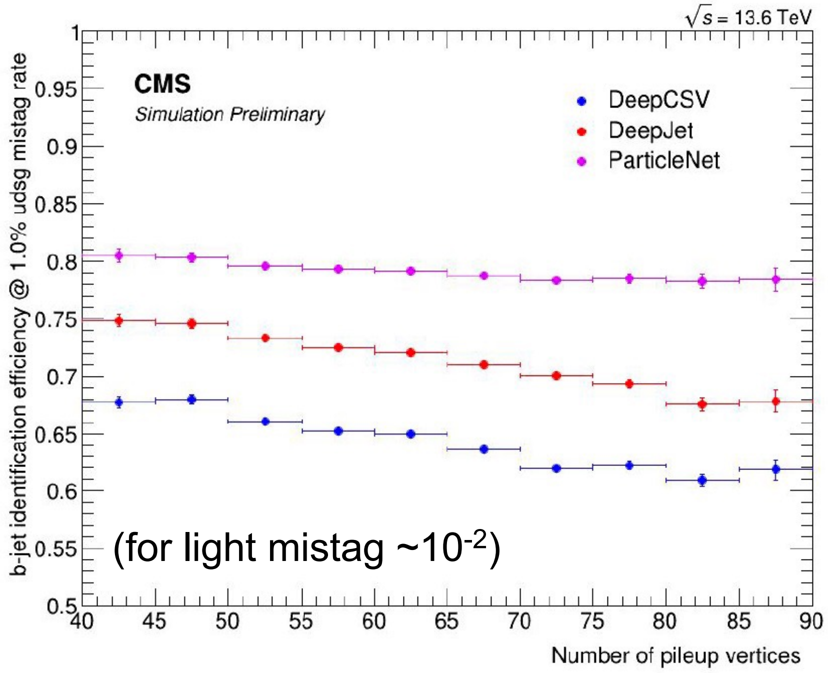


CMS DP Note 2022-041

b-tagging in trigger

- triggering on b jets was pioneered by CDF at Tevatron
- at LHC, a 1[rst] level (hardware) trigger is based on calorimeter and muon information and filters the interesting events with an output rate of 100 kHz
- then a fast computer farms can process online the (almost) full event reconstruction with an output rate ~2 kHz (High Level Trigger)
- a fast primary vertex reconstruction is applied in CMS, giving the PV z position by projecting the hits from the pixel layers, then a regional track reconstruction is applied using first the hits in pixels, then in strips





CMS DP Note 2023-021

Chap. IV
how to calibrate ?

tagging efficiency measurement

- in order to tag b (or c) jets physics analyses can use either
  - operating points, providing different tagging efficiencies and mistag rates
  - or the full shape of the b- (or c-) tagger output

  then provide a scale factor SF = data/MC ratio of efficiencies

- the efficiency measurement has to rely mostly on data, with minimal MC usage. Good to have several methods in order to validate the measurements !

- for b-tagging, one can benefit from the large amount of b jets in multijet (QCD) events, from the semi-leptonic b decays to get muon-jet b-enriched data, and from the two b jets present in each ttbar event

- for c-tagging, one can rely both on ttbar events (using $W \rightarrow cs$ decays) and on the $pp \rightarrow Wc$ process

- for double-b jets, one can use multijet events with an energetic gluon splitting $g \rightarrow bb$, both b giving a semi-muonic decay

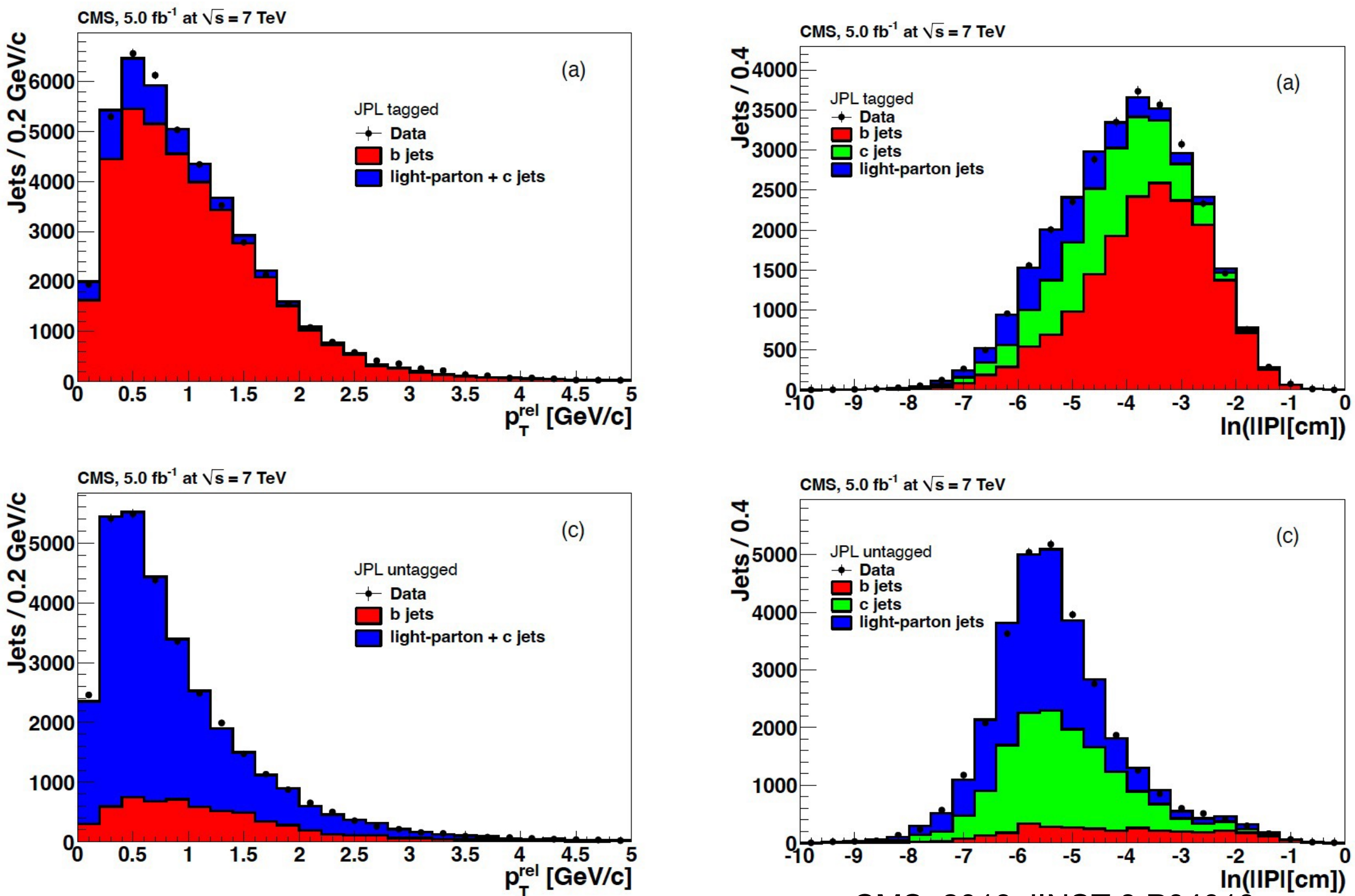
- for light jets, evaluate mistagging by using only negative IP tracks or negative SV decay lengths (as inputs to the tagger algo)

b-tag eff. meas.: multijets

three methods have been proposed (as developed in D0/CDF)

muon pTrel and IP:

- use the muon p_T relative to its jet axis, build templates for b from MC and for light+c from data multijet events; for high jet p_T, can use instead the muon IP



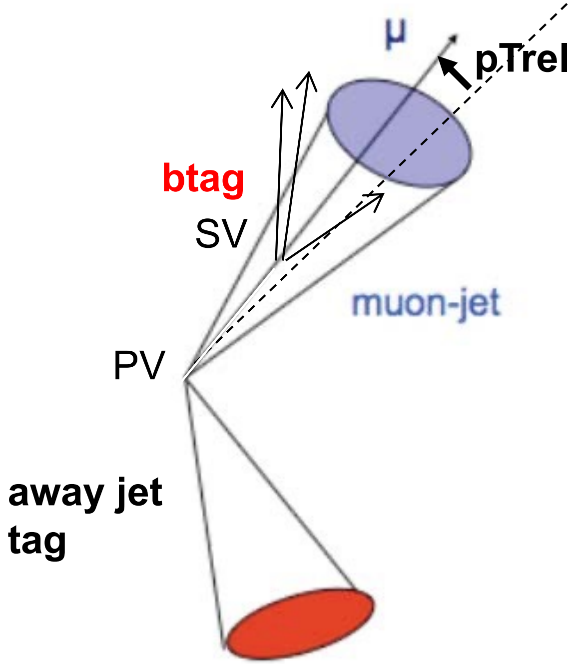
CMS, 2013 JINST 8 P04013

b-tag eff. meas.: multijets

System8: invented at IPHC for D0 (Benoit Clément 2003, PhD 2006)

- consider to tag (*WP*) both on the muon-jet and on an away jet (with another tagger) in the event, using also the ptRel (*Ref*) in the muon-jet (so 3 b-tagging criteria), this gives a system of 8 (non-linear) equations with 8 unknowns to be solved, including the b-tagging efficiency of interest
- advantage: no template shape to handle
- disadvantage: needs to know correlation coefficients, equations not easy to solve

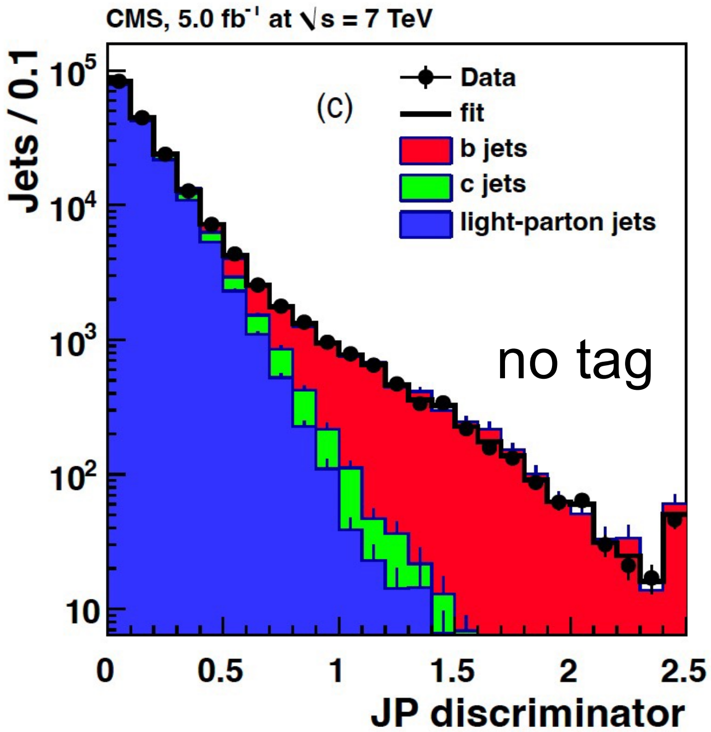
Tags:	
no tag	$n = n_b + n_{c\ell}$
away tag	$p = p_b + p_{c\ell}$
btag	$n^{tag} = \varepsilon_b^{tag} n_b + \varepsilon_{c\ell}^{tag} n_{c\ell}$
btag+away	$p^{tag} = \beta_{12}\, \varepsilon_b^{tag} p_b + \alpha_{12}\, \varepsilon_{c\ell}^{tag} p_{c\ell}$
pTrel	$n^{p_{Trel}} = \varepsilon_b^{p_{Trel}} n_b + \varepsilon_{c\ell}^{p_{Trel}} n_{c\ell}$
away+pTrel	$p^{p_{Trel}} = \beta_{23}\, \varepsilon_b^{p_{Trel}} p_b + \alpha_{23}\, \varepsilon_{c\ell}^{p_{Trel}} p_{c\ell}$
btag+pTrel	$n^{tag,p_{Trel}} = \beta_{13}\, \varepsilon_b^{tag} \varepsilon_b^{p_{Trel}} n_b + \alpha_{13}\, \varepsilon_{c\ell}^{tag} \varepsilon_{c\ell}^{p_{Trel}} n_{c\ell}$
all tags	$p^{tag,p_{Trel}} = \beta_{123}\, \varepsilon_b^{tag} \varepsilon_b^{p_{Trel}} p_b + \alpha_{123}\, \varepsilon_{c\ell}^{tag} \varepsilon_{c\ell}^{p_{Trel}} p_{c\ell}\,.$

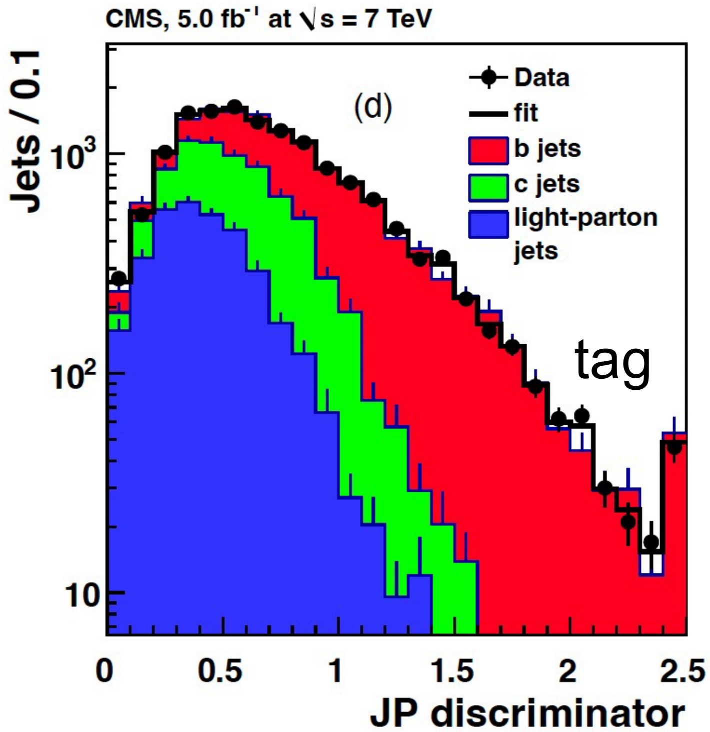


b-tag eff. meas.: multijets

3. LifeTime method: developed first at IPHC in CMS

- similarly to pTrel, template shape fits on SV mass (when available) or on Jet Probability distributions
- advantage: can be used on muon-jet events, or on multijet events (no muon request) up to very high p_T values
- disadvantage: rely on shapes from MC (for SV), and on JP calibration



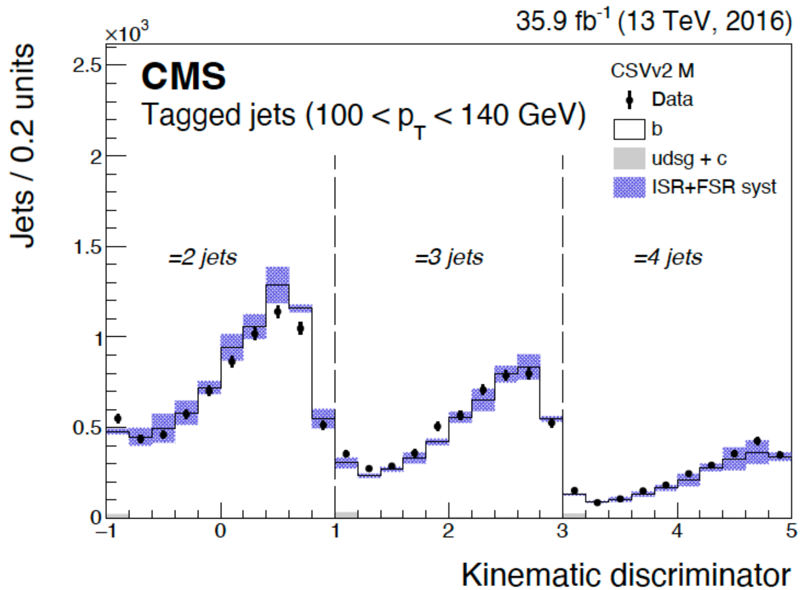


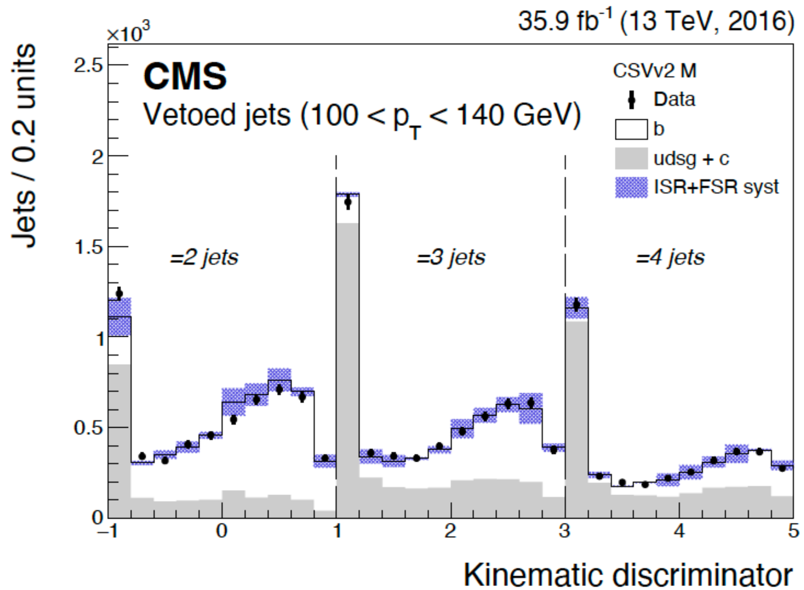
CMS, 2013 JINST 8 P04013

b-tag eff. meas.: ttbar

can benefit from both dilepton and lepton+jets ttbar final states,
those methods were initiated at Tevatron, but much more developed at LHC:
several methods proposed, here just some examples

- dilepton: select the b jet candidate using kinematic variables based on a BDT (for instance), then perform a likelihood fit to derive the nb. of tagged jets and the data/MC scale factor of the b-tag efficiency ε_b



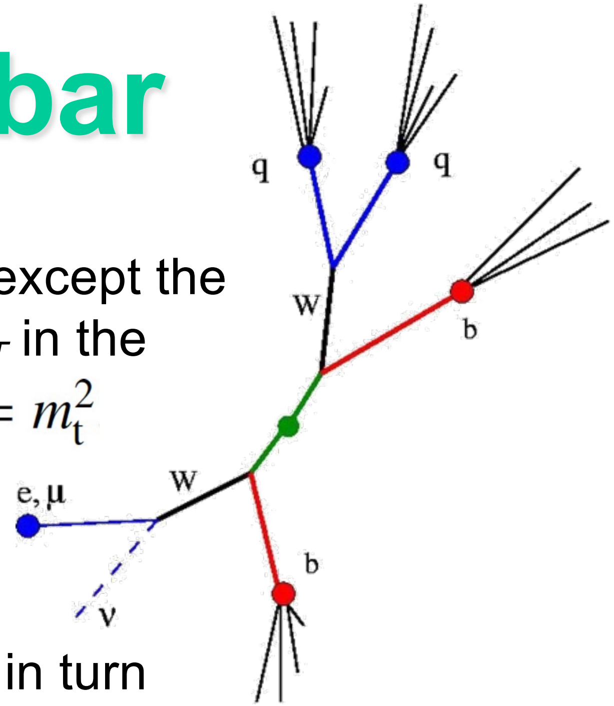


is cross-checked with a simple counting method (proposed at IPHC):

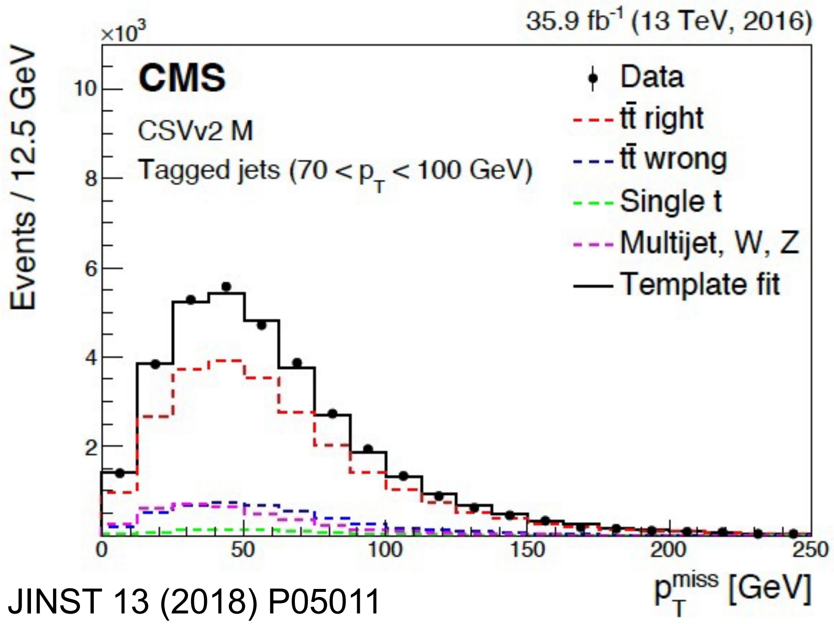
$$N_{2\ \text{b-tagged}} - N_{2\ \text{b-tagged}}^{\text{non-b jet}} = \varepsilon_b^2 n_{2\ \text{b jets}},$$

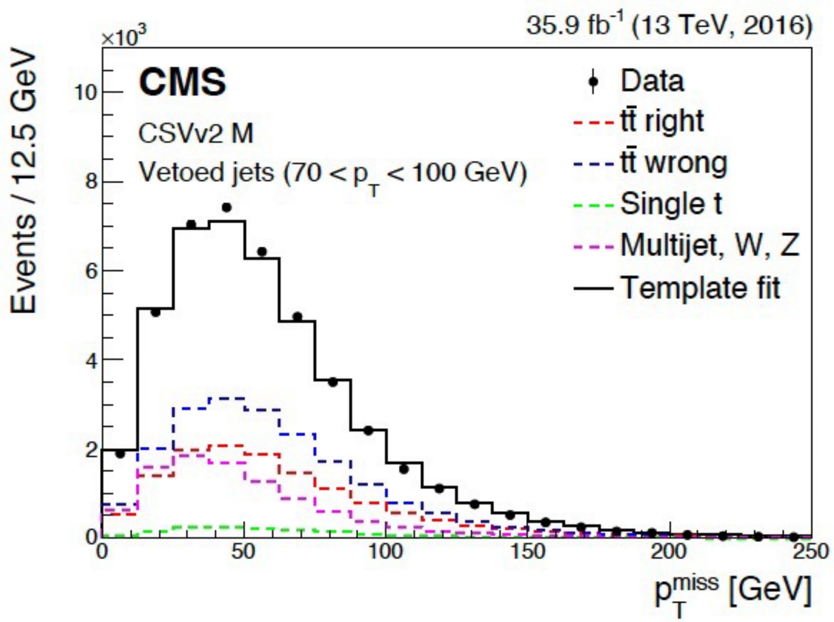
CMS, JINST 13 (2018) P05011

b-tag eff. meas.: ttbar



- lepton+jets: all objects can be fully reconstructed, except the neutrino which is approximated using the missing p_T in the event to satisfy: $(p_\nu + p_\ell)^2 = m_{\mathrm{W}}^2 \quad (p_\nu + p_\ell + p_{\mathrm{b},\ell})^2 = m_{\mathrm{t}}^2$
  - then a likelihood λ is built, which maximizes the association of all jets to either the hadronic or to the leptonic side
  - both leptonic and hadronic sides can be tagged in turn
  - the tagging efficiency is estimated on the other side by considering template fits of the log λ or p_T^{miss} distribution, with or without tagging

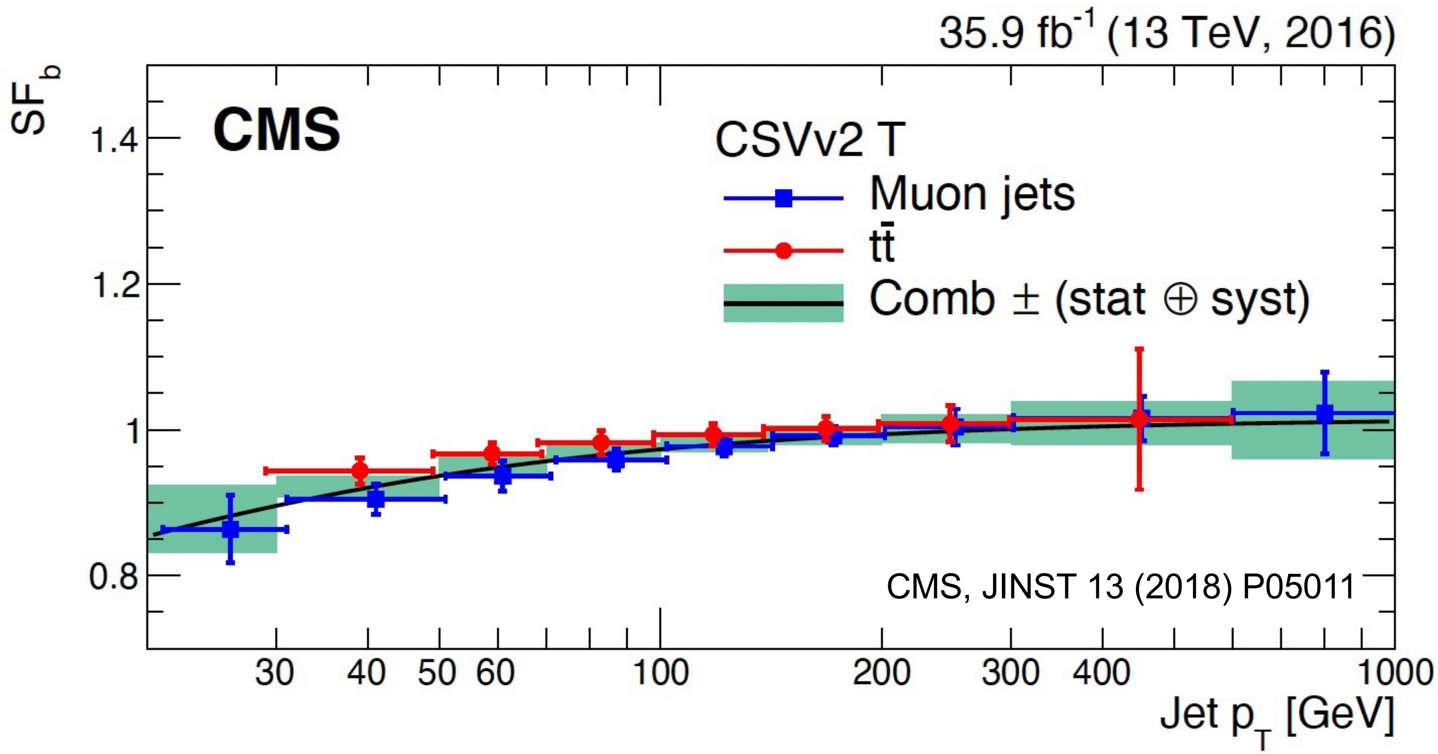




CMS, JINST 13 (2018) P05011

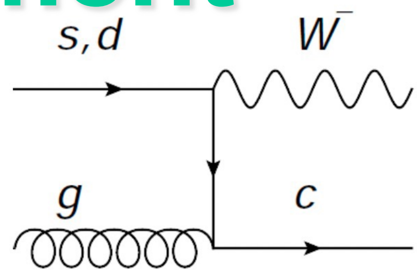
data/MC b-tag efficiency ratio

- taking into account all systematics (which are numerous and need to very carefully checked, the evil hides in the details…) one can combine the muon-jet and multijet results and compare them to the combined ttbar results (→ agreement ~1%)



- the full procedure is repeated for each data taking year and each data reprocessing, in order to deliver those correction factors, vs jet p_T or vs tagger shape, for all physics analyses

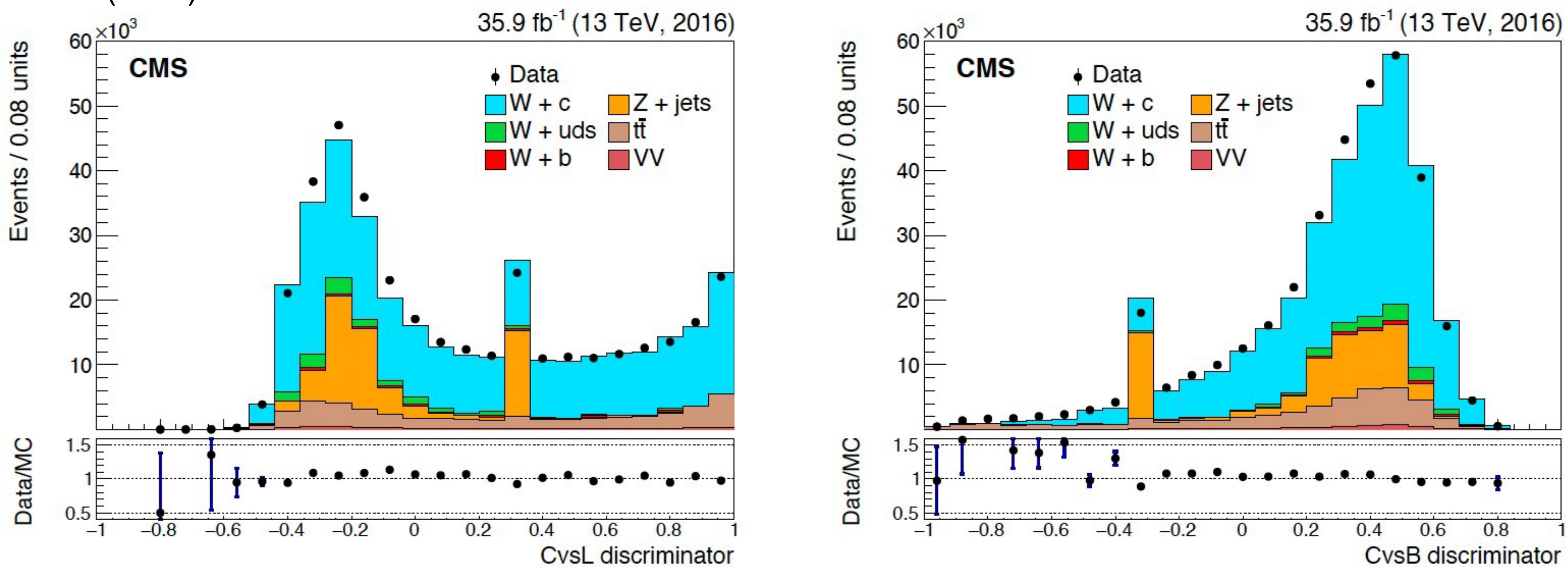
c-tag efficiency measurement



- Wc process with W → eν or μν decay:
  the c jet contains a soft muon (from charm semi-leptonic decay) such that the lepton from W and muon from c have opposite charge, while the background has no charge preference,
- thus the same-sign background can be subtracted and background ``free'' distributions can be fitted with or w/o c jet tagging applied
- one can then infer the c jet tagging efficiency from data

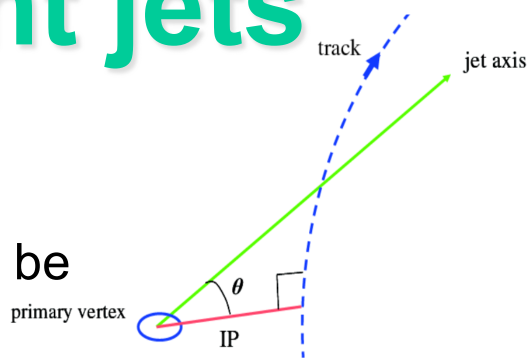
$$\varepsilon_c = \frac{N(\mathrm{W+c})^{\mathrm{OS\text{-}SS}}_{\mathrm{tagged}}}{N(\mathrm{W+c})^{\mathrm{OS\text{-}SS}}},$$

CMS, JINST 13 (2018) P05011



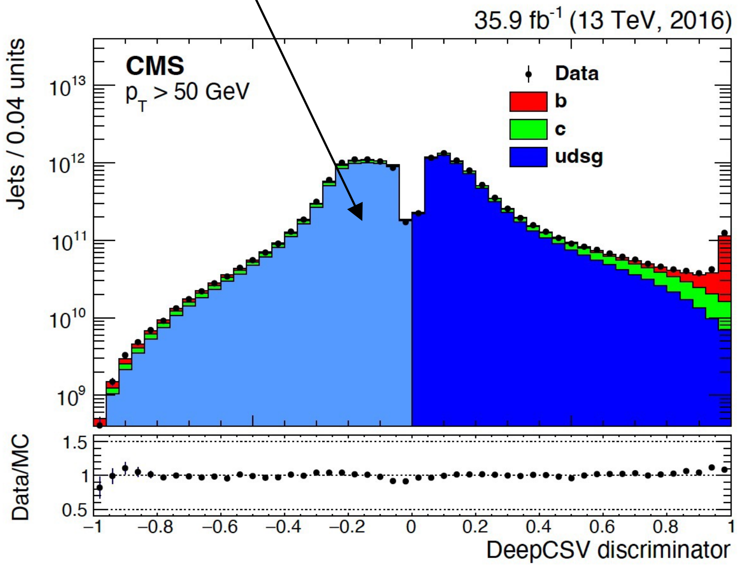
- ttbar events can also be used in the l+jets channel with a W → cs or ud pair which can be selected and is enriched in c jet candidates

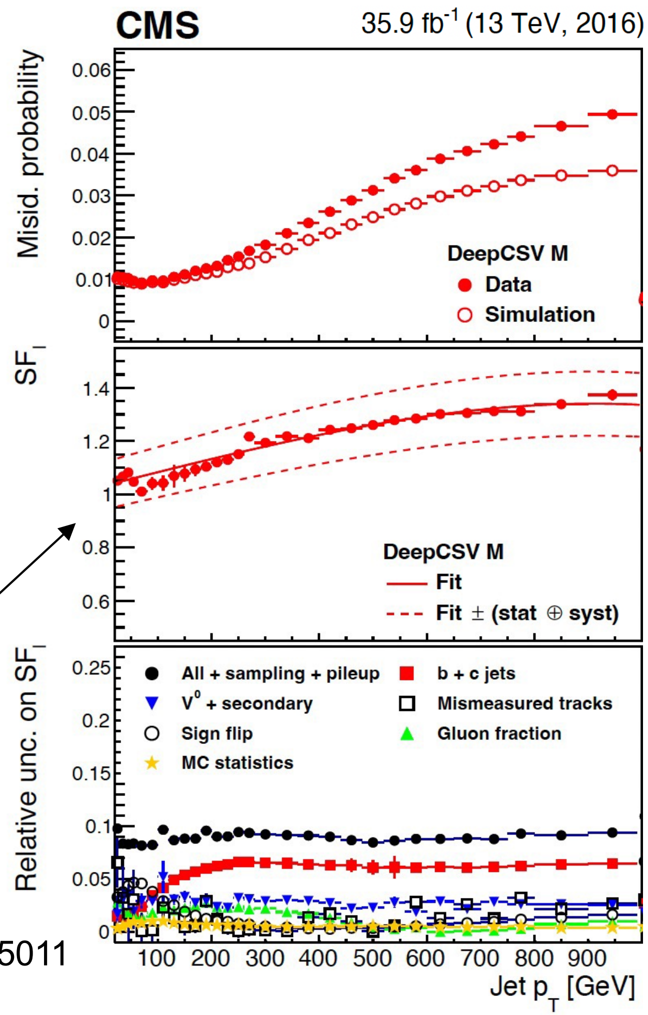
mistag efficiency of light jets



negative tag method: invented in CDF and D0

- mistagging from the huge amount of uds and g jets can be evaluated by using the same tagger as for b-tagging, but restricting its input quantities to tracks with negative IP and to SVs with negative decay length





- using multijet (QCD) events, data/MC mistag efficiency ratio is derived with uncertainty ~10% (including various systematic sources)

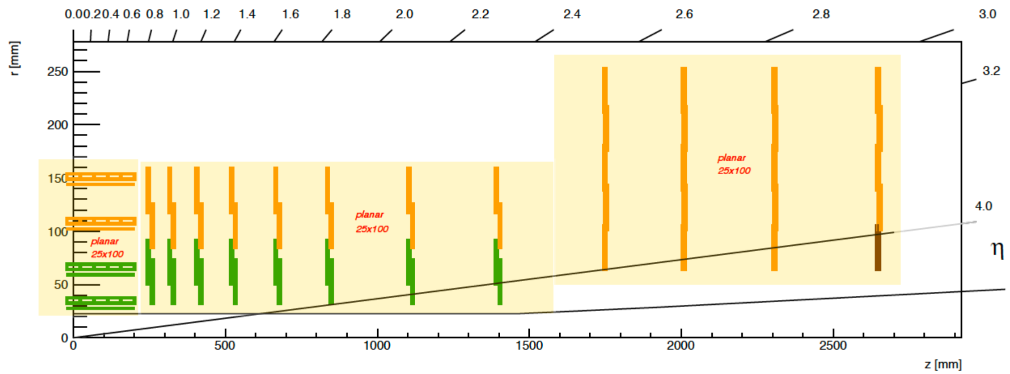
CMS, JINST 13 (2018) P05011

Conclusion and prospect

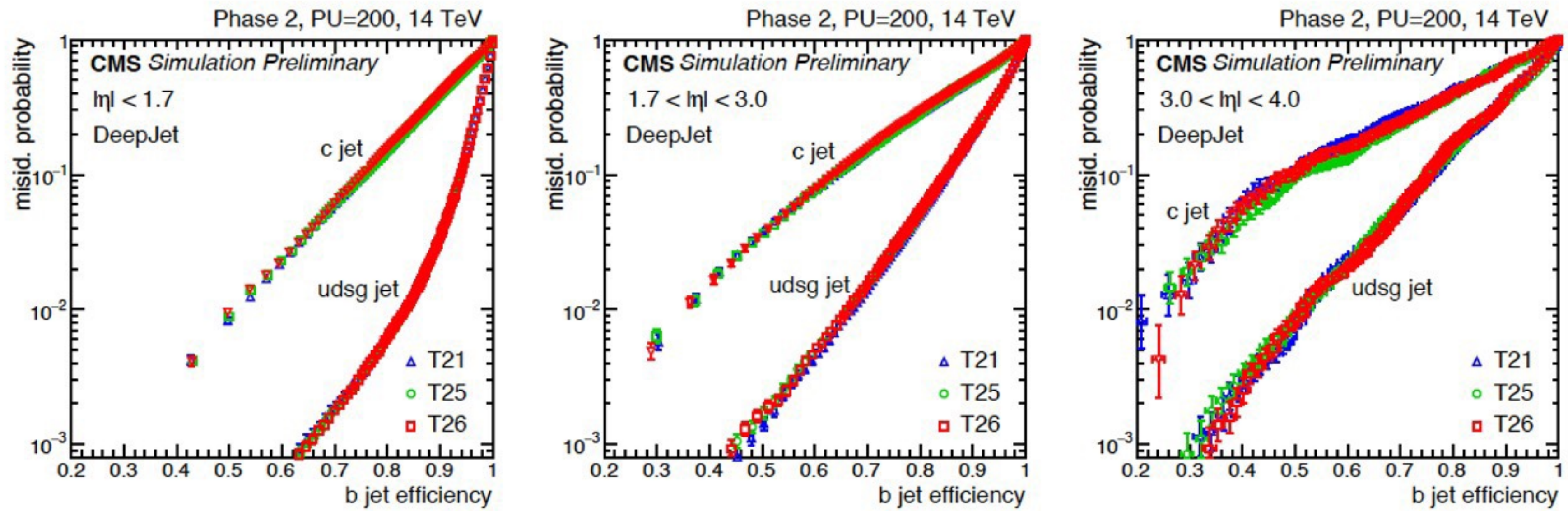
the future: tagging at HL-LHC

- at HL-LHC, the pixel detector will need to cope with higher instantaneous luminosity (x 2-4), higher pileup (~140-200) and higher radiation damage

- the next pixel detector will have a smaller granularity: 25x100 μm^2 (instead of 100x150 μm^2 at present), and will extend up to $|\eta| < 4$ (instead of < 2.5 at present)



- the expected performance are very good, despite the harsh conditions



CMS DP Note 2022-071