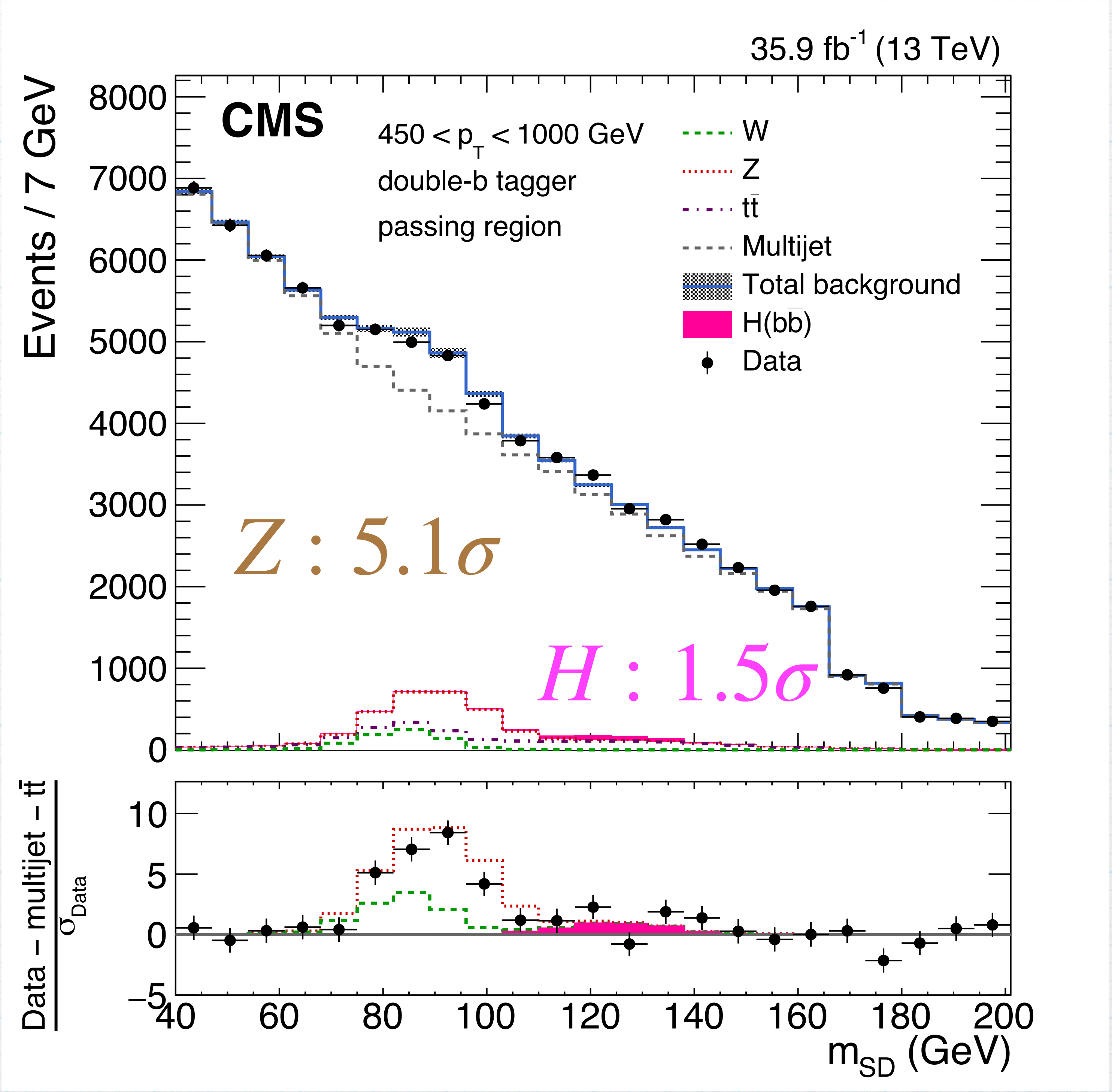
Jets and their structure

Simone Marzani
Università di Genova &
INFN Sezione di Genova

CTEQ 2024
summer school

Lecture 2: jet substructure

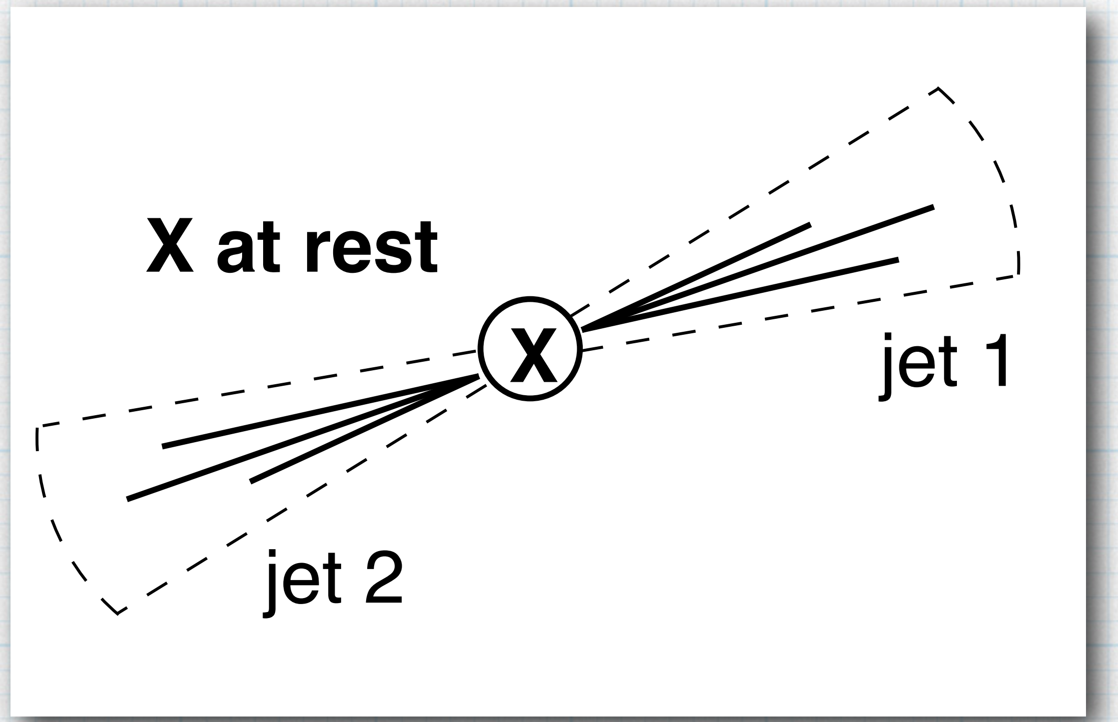
- boosted-objects physics
- grooming and tagging
- calculations for jet substructure

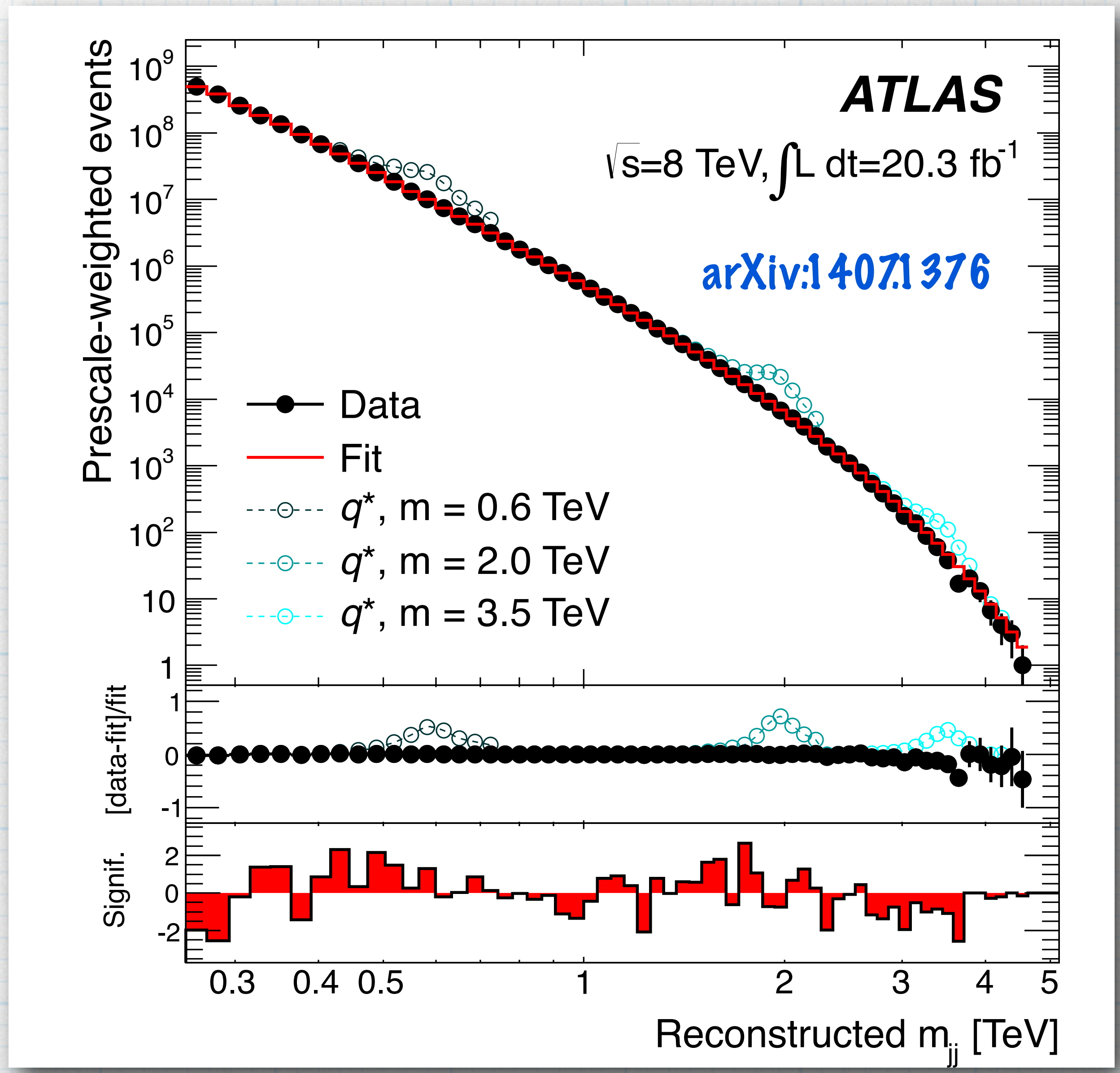


the (ambitious) target of this lecture is to understand this plot

searching for new particles (I)

* Standard analysis: the heavy particle X decays into two partons, reconstructed as two jets



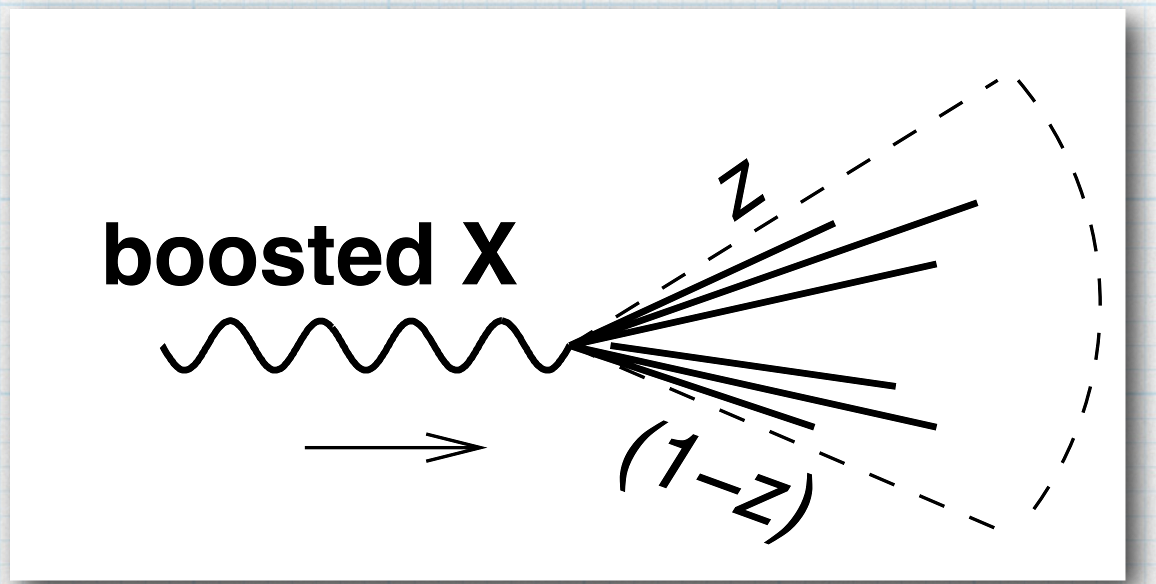


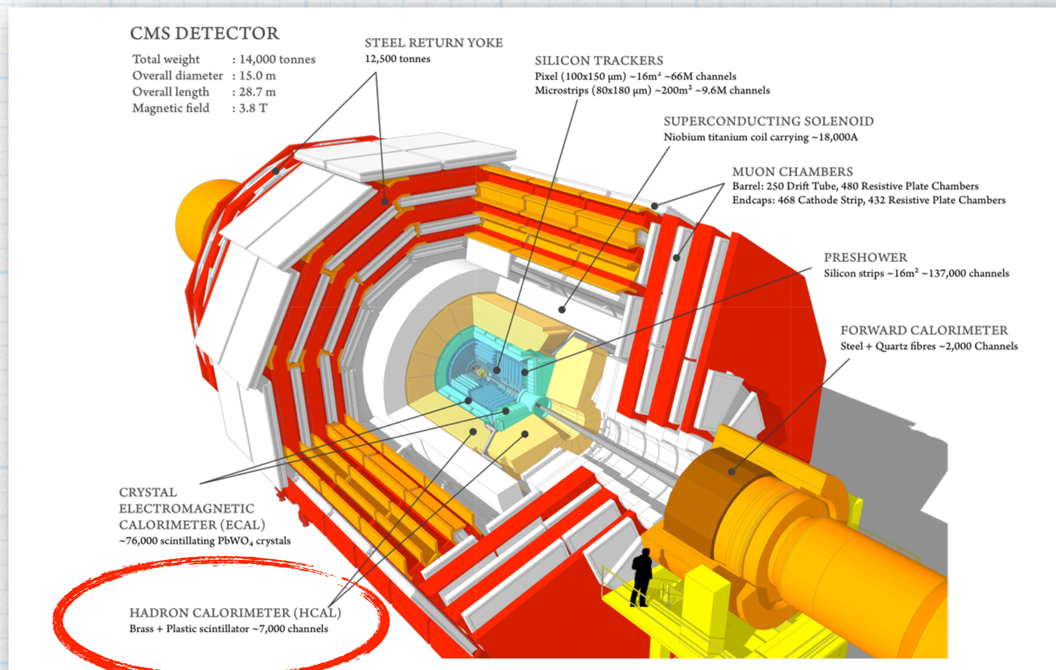
* Look for bumps in the dijet invariant mass distribution

* What about EW-scale particles at the LHC ?

searching for new particles (II)

* LHC energy (10^4 GeV) $\gg$ electro-weak scale (10^2 GeV)
* EW-scale particles (new physics, Z/W/H/top) are abundantly produced with a large boost



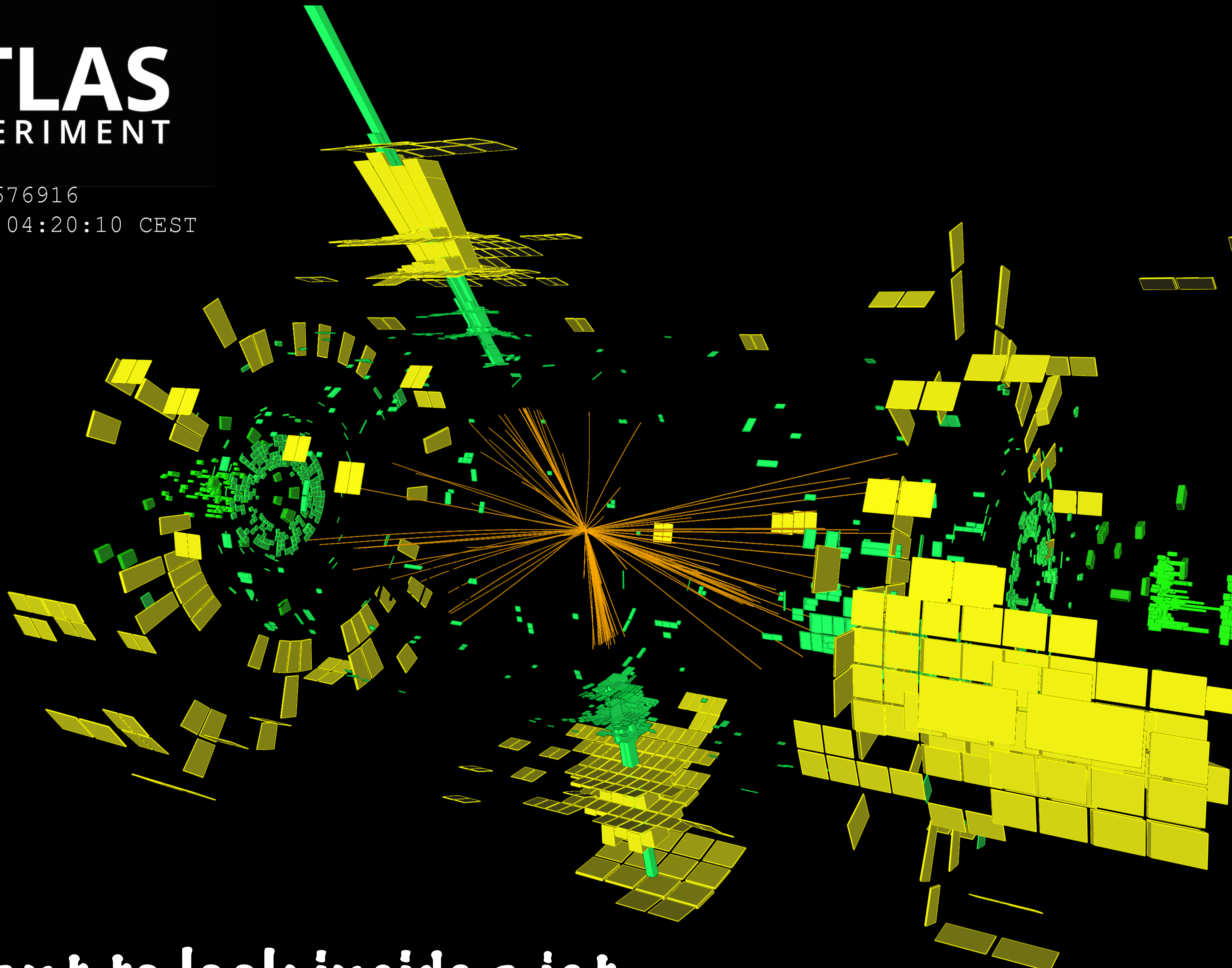


source CMS

* their decay-products are then collimated
* if they decay into hadrons, we end up with localised deposition of energy in the hadronic calorimeter: a jet

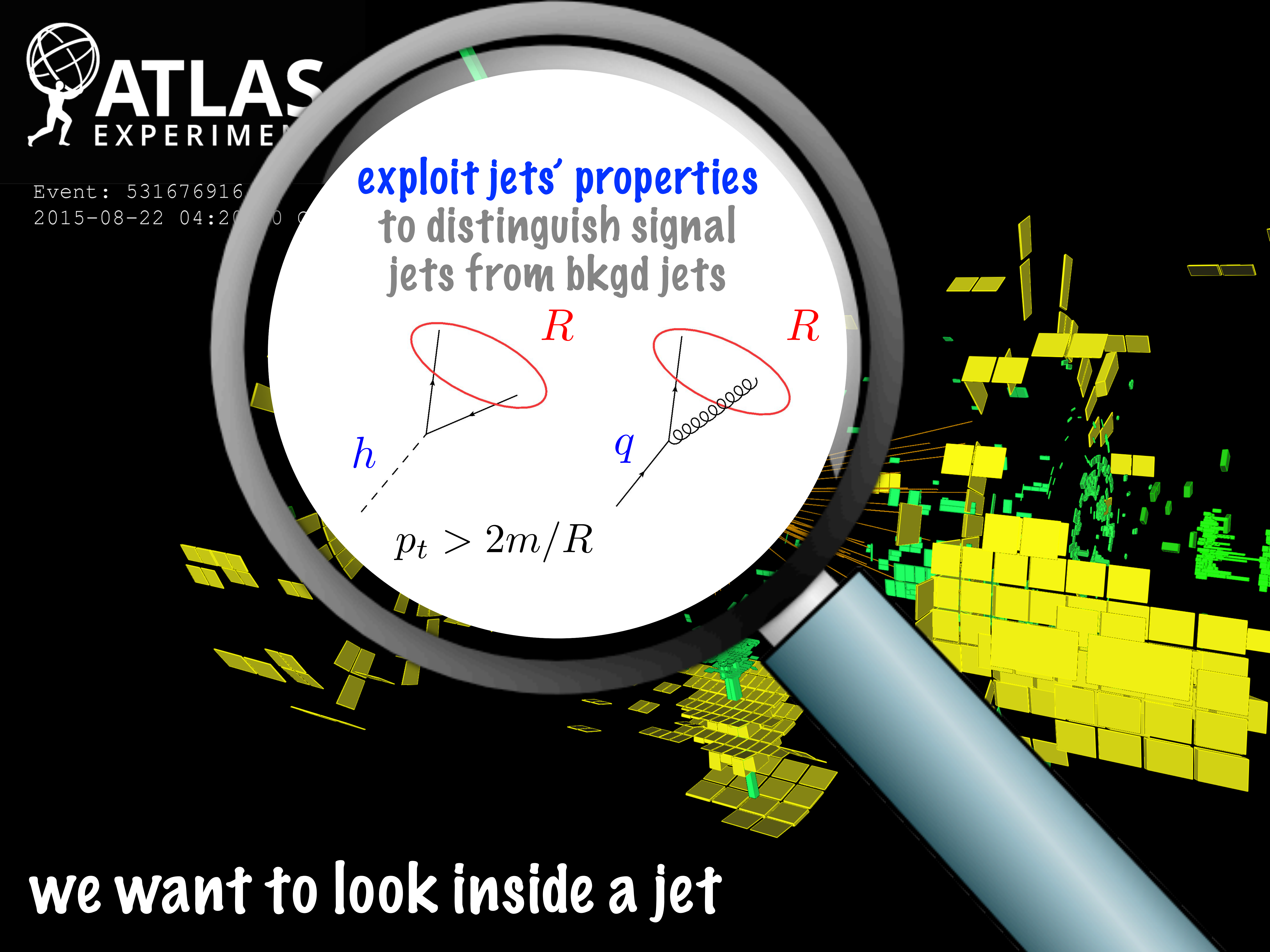
ATLAS EXPERIMENT

Event: 531676916
2015-08-22 04:20:10 CEST



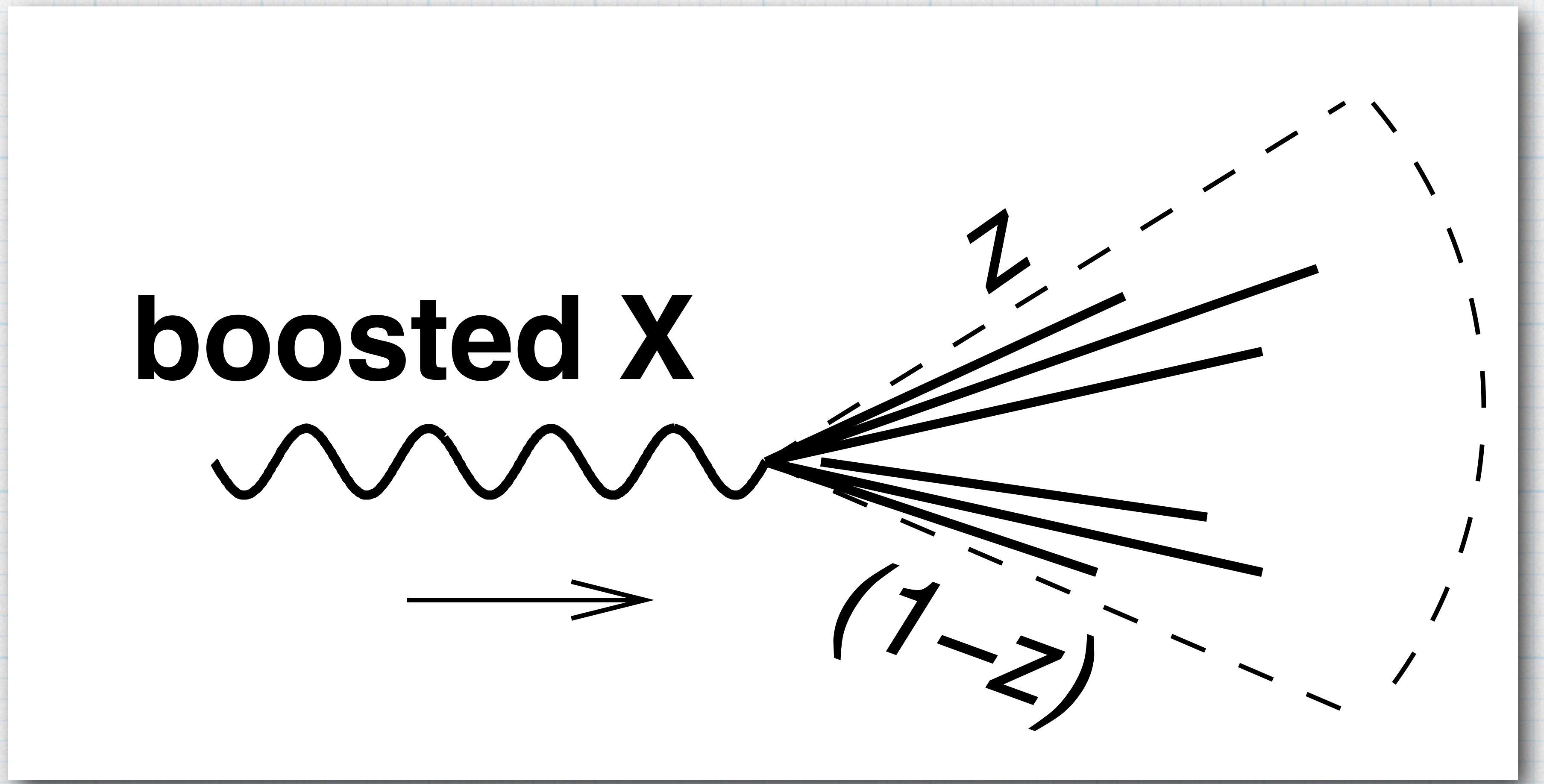
we want to look inside a jet

we want to look inside a jet


ATLAS
Event: 531676916
2015-08-22 04:2
exploit jets' properties
to distinguish signal
jets from bkgd jets
R
R
h
q
$p_t > 2m/R$
we want to look inside a jet

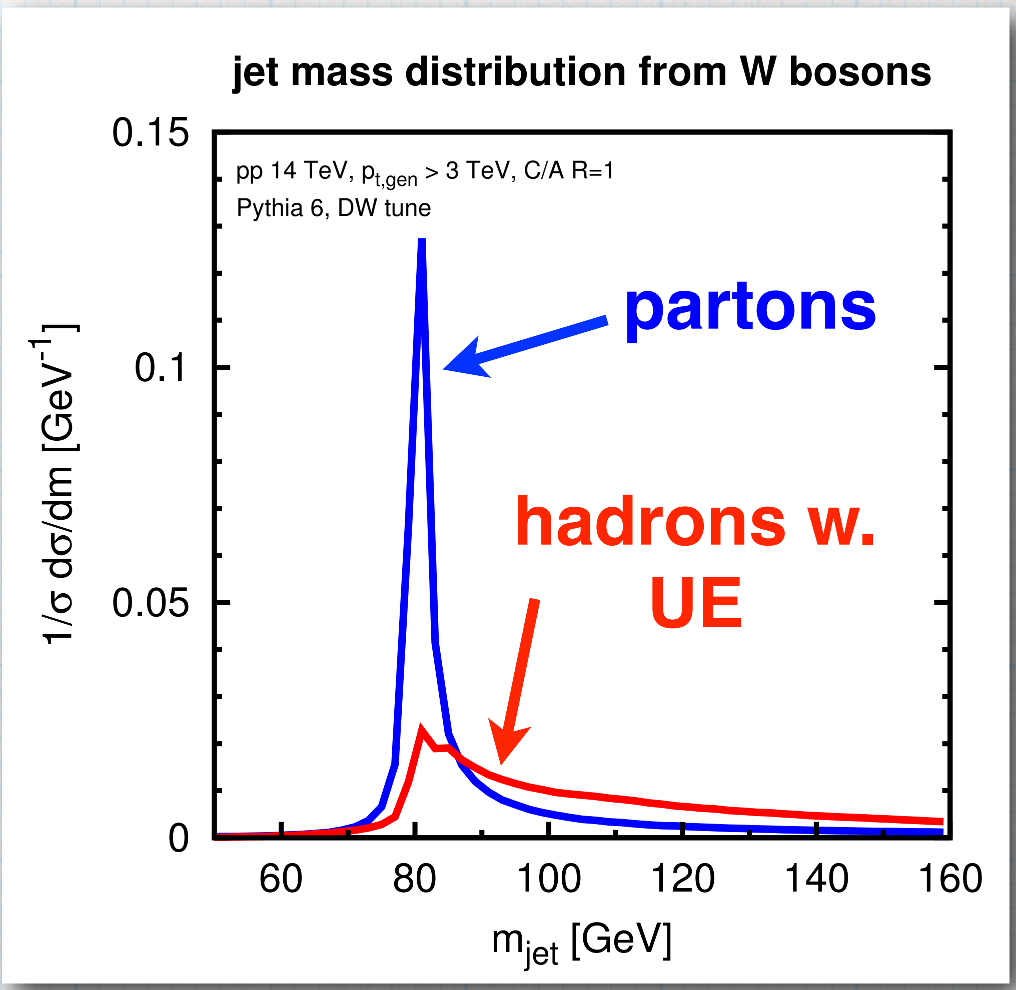
signal-jet mass

* first jet-observable that comes to mind
* signal jets should have a mass distribution peaked near the resonance



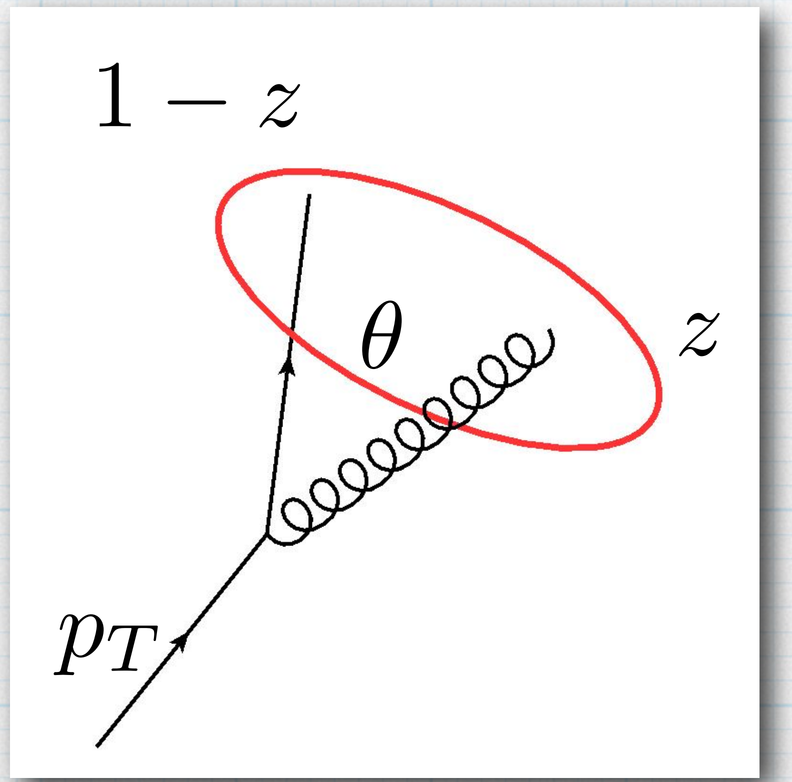
signal-jet mass

* first jet-observable that comes to mind
* signal jets should have a mass distribution peaked near the resonance
* however, that's a simple partonic picture
* perturbative and non-pert. emissions from the qqb pair broadens and shift the peak
* underlying event and pile-up typically enhance the jet mass



QCD-jet mass

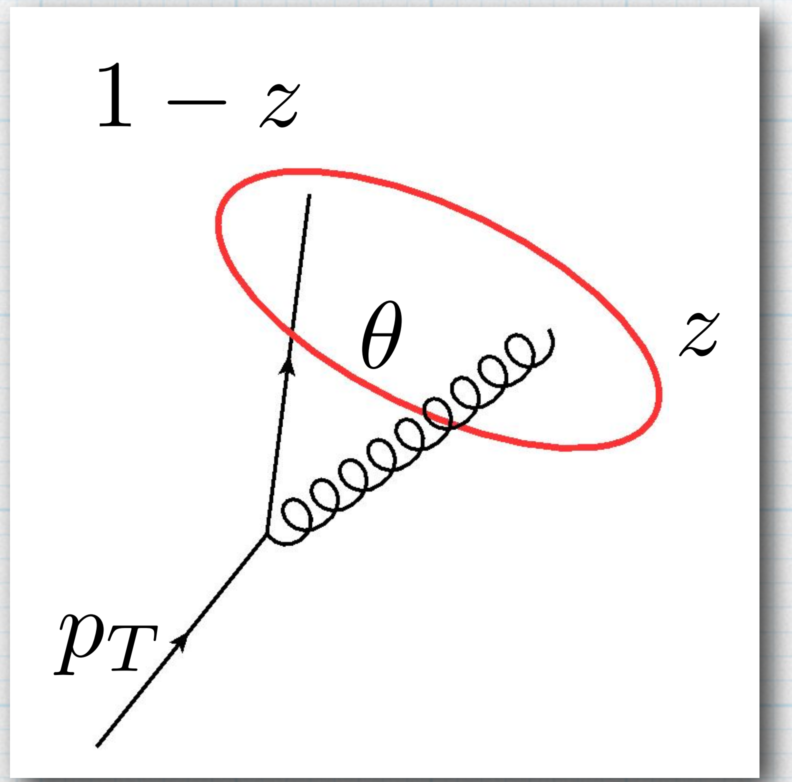
* first jet-observable that comes to mind
* background (QCD) jets acquire mass through showering



$$m^2 = 2p_q \cdot p_g \simeq z(1-z)\theta^2 p_T^2$$

QCD-jet mass

* first jet-observable that comes to mind
* background (QCD) jets acquire mass through showering

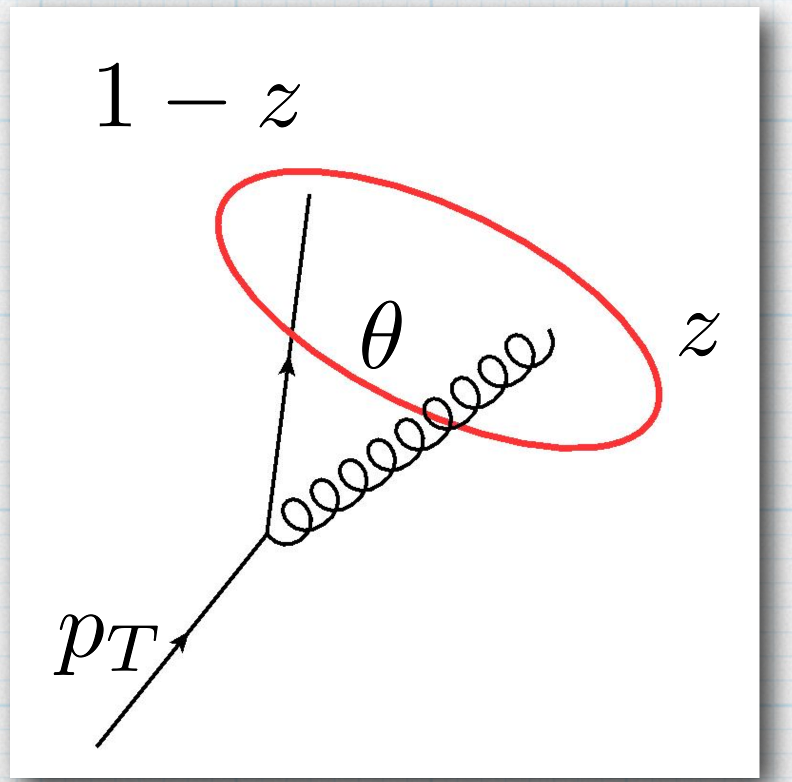


$$m^2 = 2p_q \cdot p_g \simeq z(1-z)\theta^2 p_T^2$$

$$\langle m^2 \rangle \simeq \frac{\alpha_s}{2\pi} p_T^2 \int_0^{R^2} \frac{d\theta^2}{\theta^2} \int_0^1 dz z(1-z)\theta^2 P_{gq}(z)$$

QCD-jet mass

* first jet-observable that comes to mind
* background (QCD) jets acquire mass through showering



$$m^2 = 2p_q \cdot p_g \simeq z(1-z)\theta^2 p_T^2$$

$$\langle m^2 \rangle \simeq \frac{\alpha_s}{2\pi} p_T^2 \int_0^{R^2} \frac{d\theta^2}{\theta^2} \int_0^1 dz z(1-z)\theta^2 P_{gq}(z)$$

$$= \frac{\alpha_s C_F}{\pi} p_T^2 R^2 \int_0^1 dz z(1-z) \frac{2-2z+z^2}{2z}$$

mass grows with p_T

$$= 3/8$$

jet properties

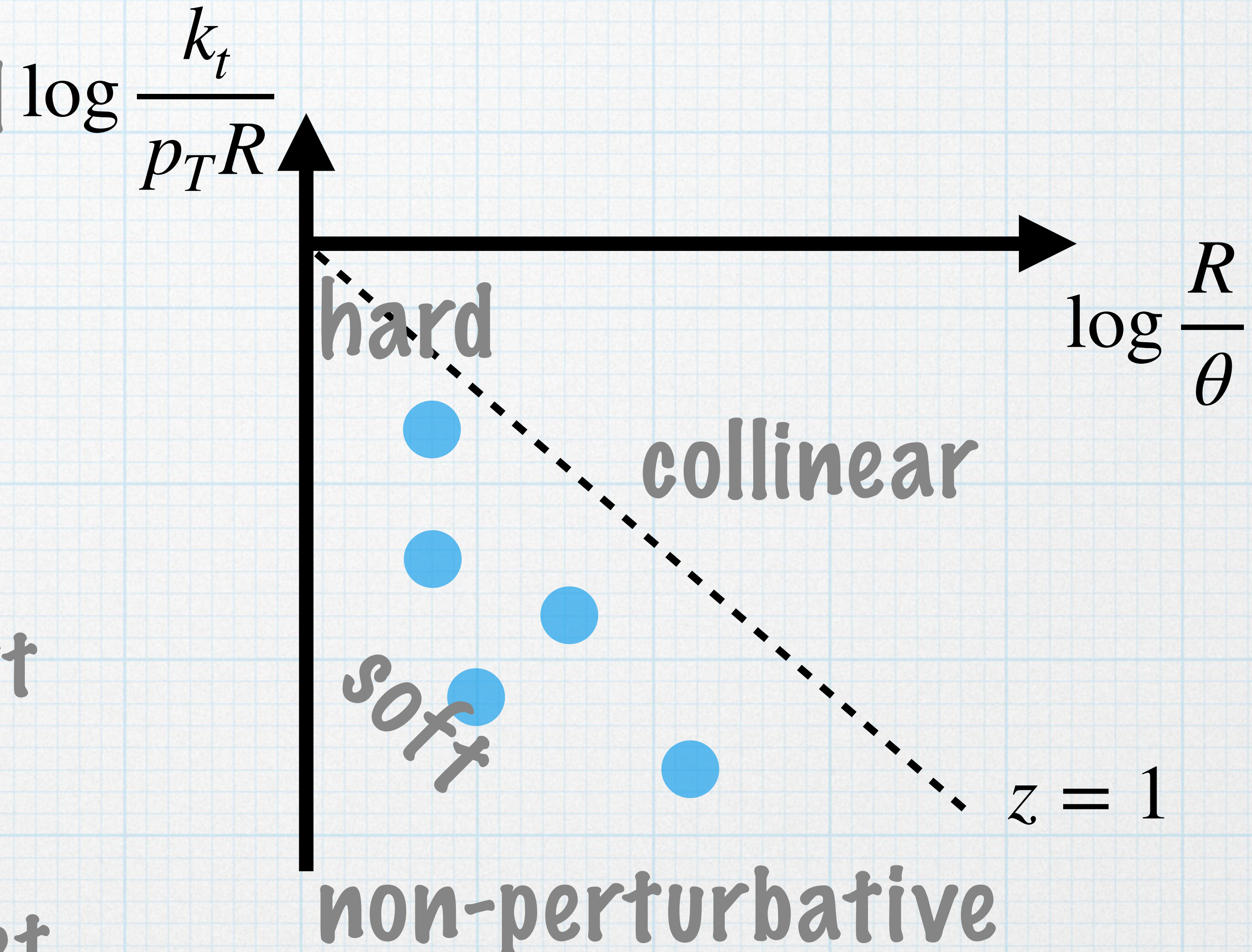
* we want to studies the properties of jets
* hence, we resolve a (high p_t) jet down to a smaller scale, e.g. its mass
* large logarithms appear invalidating the fixed-order expansion
* we need to reorganise the calculation so that we can consider any number of soft/collinear partons: resummation
* vast field with many approaches: dQCD, SCET, etc.

aside: the Lund plane

* the Lund plane is a powerful representation of soft-collinear emissions kinematics

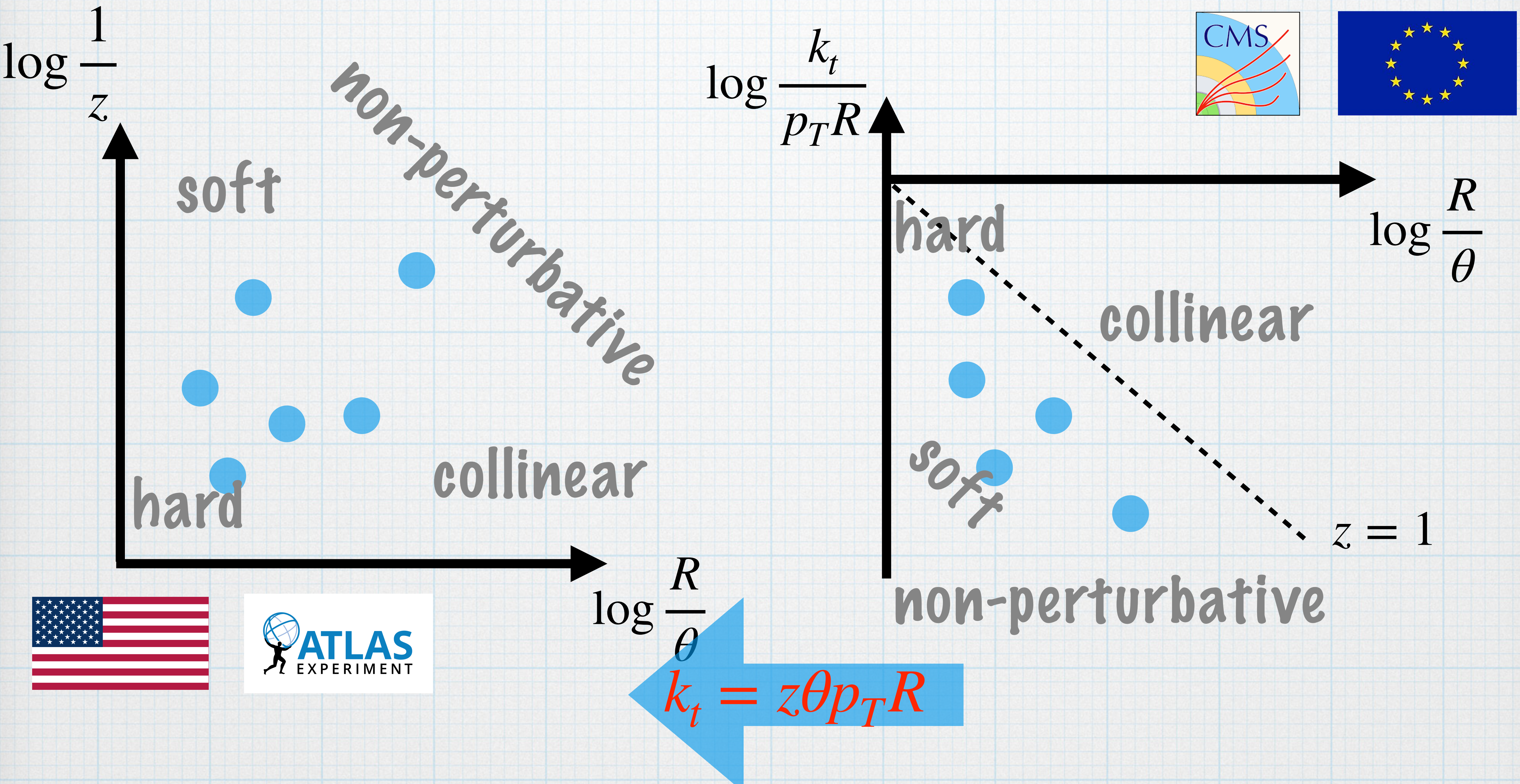
* as the name suggests it was first developed in the context of Monte Carlo studies

* useful representation of a jet (also for ML!)



* soft-collinear emissions populate the Lund plane uniformly: equal area = equal probability
* now also a powerful observables (measured at the LHC)

aside: the Lund plane



- soft-collinear emissions populate the Lund plane uniformly: equal area = equal probability
- now also a powerful observables (measured at the LHC)

how do we model it ?

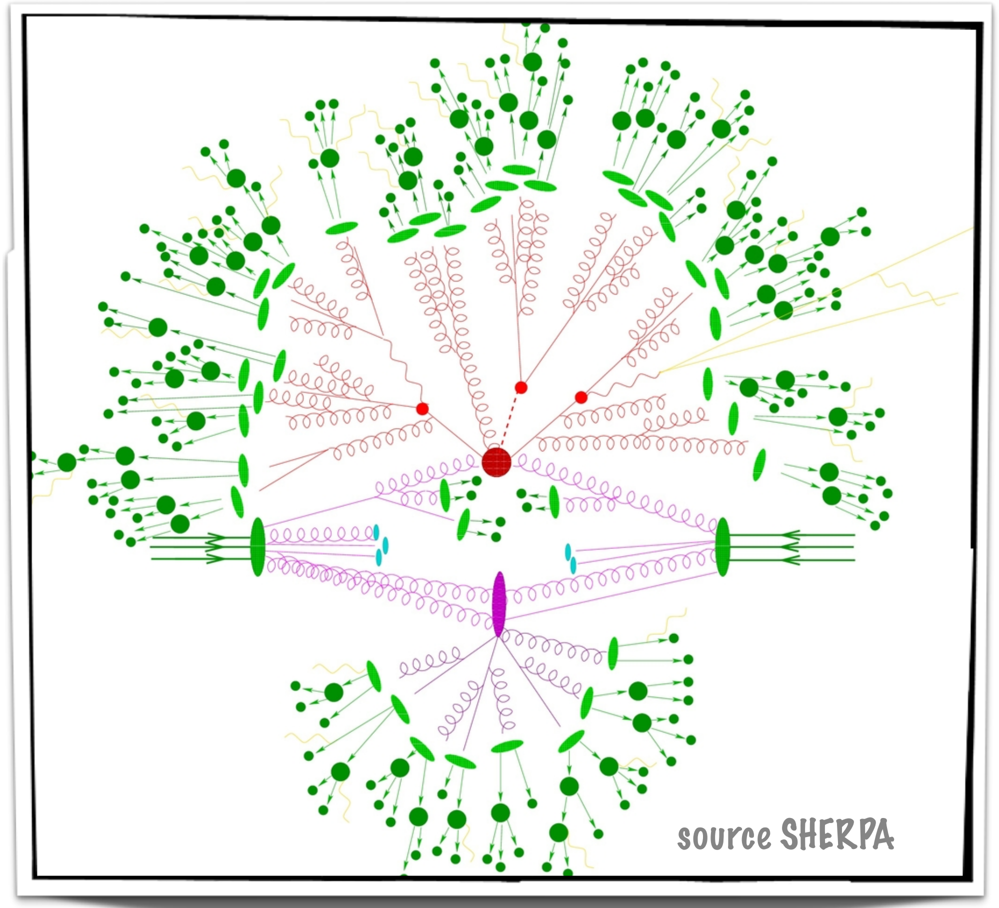
* jet properties: we want to compute x-sections and distributions with many particles in the final state
* fixed-order perturbation theory seems inadequate
* interesting physics happens at small angular separation and small energies
* all-order (resummed) calculations are possible and necessary !

Monte Carlo Parton Showers

emissions at small angles factorize

$$d\sigma_{n+1} \simeq d\sigma_n \frac{\alpha_s}{2\pi} \frac{d\theta^2}{\theta^2} dz d\phi P(z,\phi)$$

we can write a computer program that simulates these classical branchings



how do we model it ?

- jet properties: we want to compute x-sections and distributions with many particles in the final state
- fixed-order perturbation theory seems inadequate
- interesting physics happens at small angular separation and small energies
- all-order (resummed) calculations are possible and necessary !

Analytic Resummation

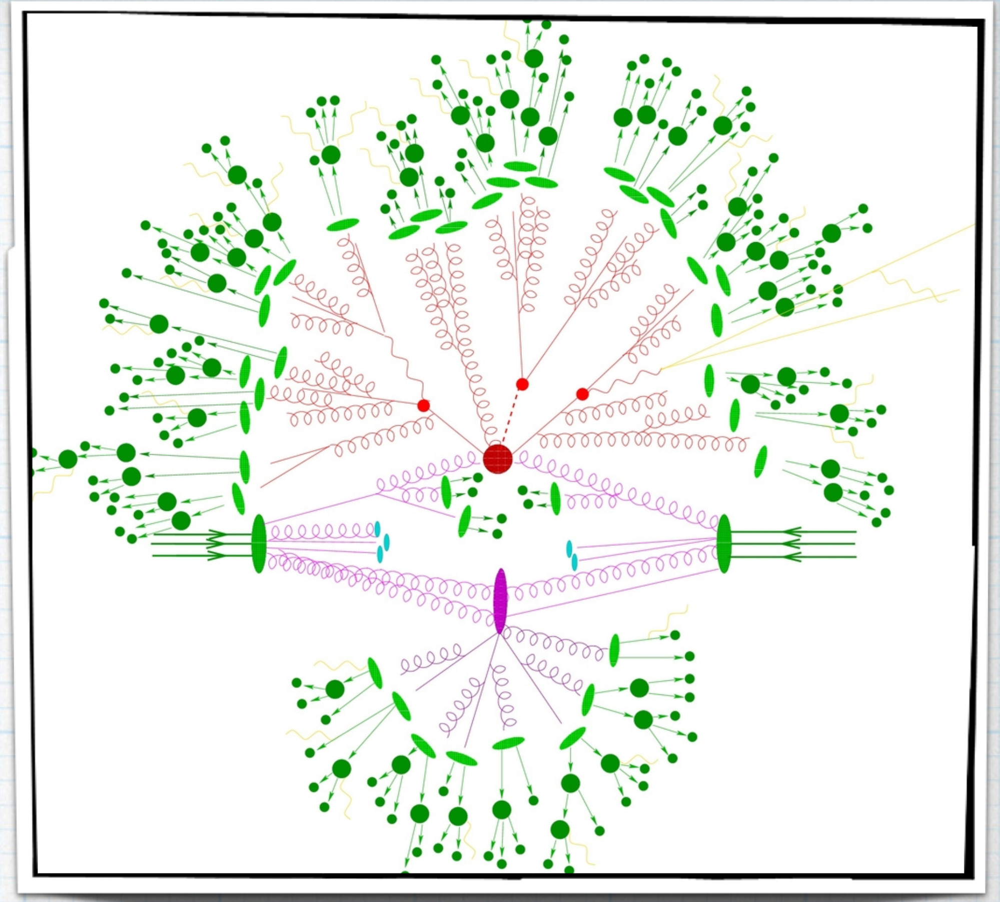
emissions at small angles factorize

$$d\sigma_{n+1} \simeq d\sigma_n \frac{\alpha_s}{2\pi} \frac{d\theta^2}{\theta^2} dz d\phi P(z,\phi)$$

soft emissions factorise in a subtle way

$$d\sigma_{n+1} \simeq d\sigma_n \frac{\alpha_s}{2\pi} d\theta dz d\phi \sum_{i<j} C_{ij} D_{ij}(z,\theta,\phi)$$

$$\sigma_{res} = g_0 \exp[g_1(\alpha_s L)/\alpha_s + g_2(\alpha_s L) + \alpha_s g_3(\alpha_s L) + \ldots]$$



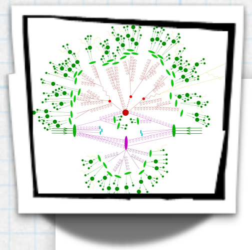
* powerful general-purpose tools
* provide fully differential events on which any observable can be measured
* interfaced with non-perturbative models to give a realistic description
* theoretical accuracy difficult to assess (often low)

VS

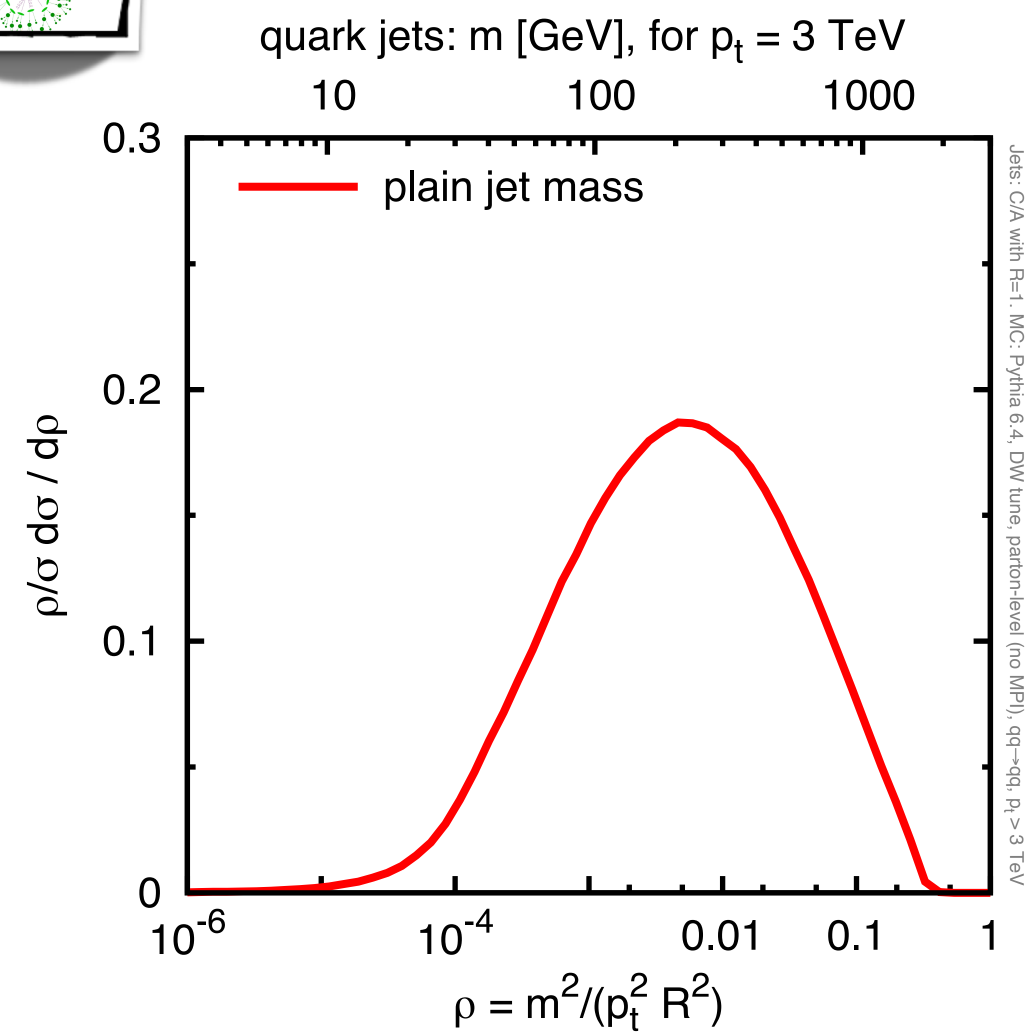
$$\sigma_{res} = g_0 \exp[g_1(\alpha_s L)/\alpha_s + g_2(\alpha_s L) + \alpha_s g_3(\alpha_s L) + \ldots]$$

* feasible for a limited number of observables
* well defined and improvable accuracy
* state-of-the art (resummation + fixed order)
* provide insights and understanding

the jet mass



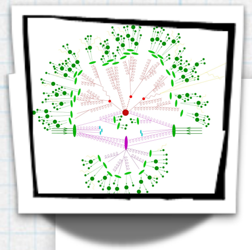
$\sigma_{res} = g_0 \exp[g_1(\alpha_s L)/\alpha_s + g_2(\alpha_s L) + \alpha_s g_3(\alpha_s L) + \ldots]$

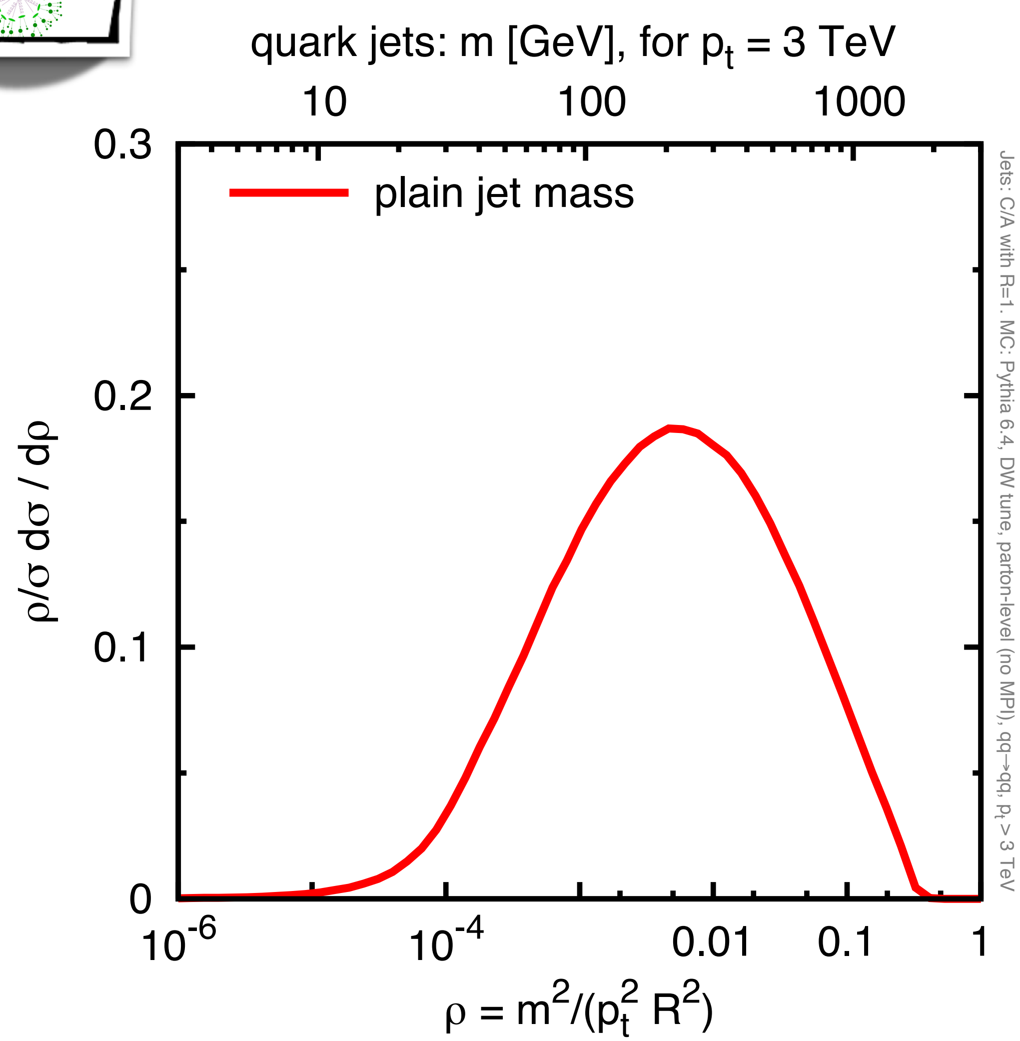


* plain jet mass: Sudakov peak, where does it come from ?
* let's do an easy calculation: one gluon emission in the collinear limit

the jet mass

$\sigma_{res} = g_0 \exp[g_1(\alpha_s L)/\alpha_s + g_2(\alpha_s L) + \alpha_s g_3(\alpha_s L) + \ldots]$



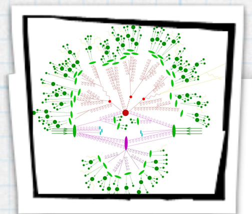


$$\frac{1}{\sigma}\frac{d\sigma^{\mathrm{LO}}}{dm^2} = \frac{\alpha_s}{2\pi}\int_0^{R^2}\frac{d\theta^2}{\theta^2}\int_0^1 dz P_{gq}(z) \times \delta\left(m^2 - z(1-z)\theta^2 p_T^2\right)$$

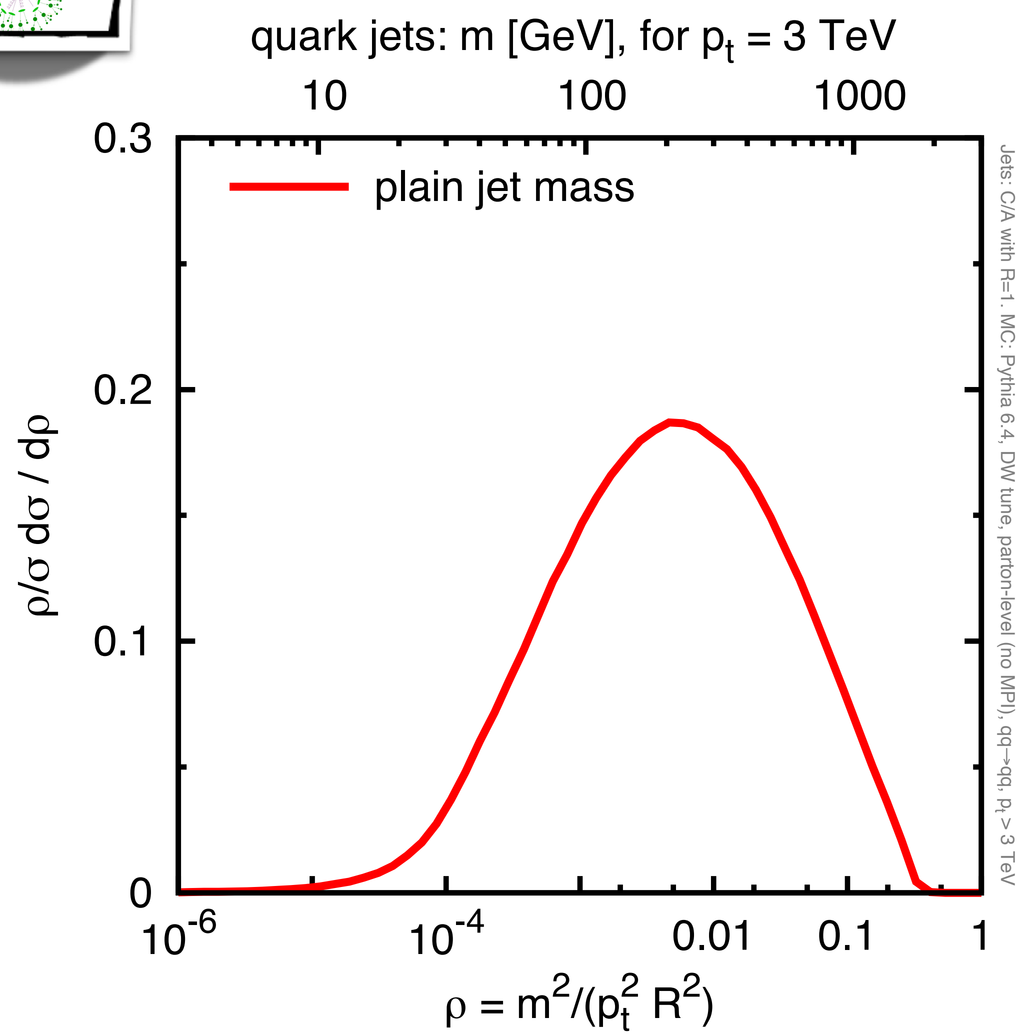
we demand a mass m

* plain jet mass: Sudakov peak, where does it come from ?
* let's do an easy calculation: one gluon emission in the collinear limit

the jet mass



$\sigma_{res} = g_0 \exp[g_1(\alpha_s L)/\alpha_s + g_2(\alpha_s L) + \alpha_s g_3(\alpha_s L) + \ldots]$



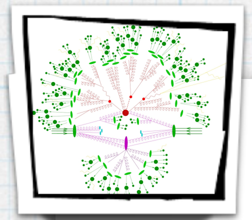
$$\frac{1}{\sigma}\frac{d\sigma^{\rm LO}}{dm^2} = \frac{\alpha_s}{2\pi}\int_0^{R^2}\frac{d\theta^2}{\theta^2}\int_0^1 dz P_{gq}(z) \times \delta\left(m^2 - z(1-z)\theta^2 p_T^2\right)$$

$$= \frac{\alpha_s C_F}{2\,\pi} m^{-2}\int_{m^2/(p_T^2 R^2)}^1 dz \quad P_{gq}(z)$$

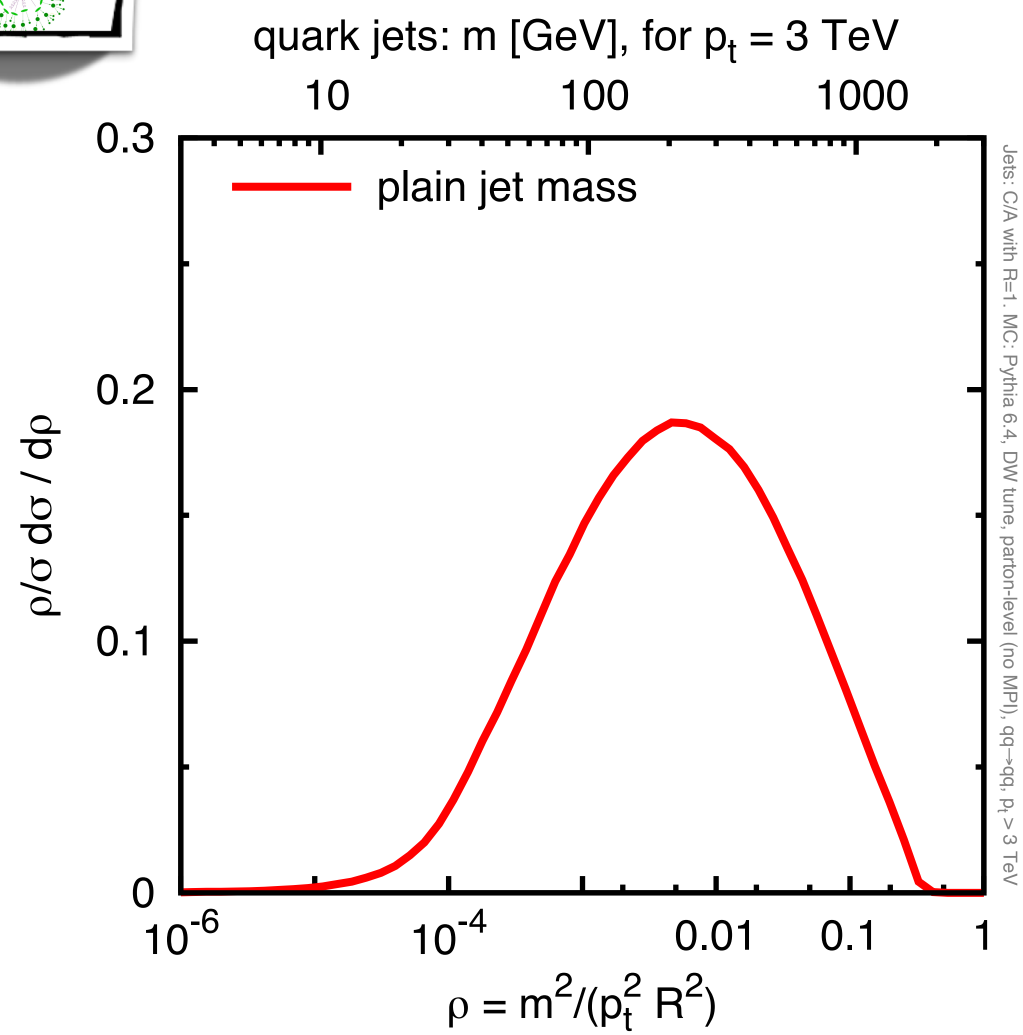
we do the angular integral with the delta function

* plain jet mass: Sudakov peak, where does it come from ?
* let's do an easy calculation: one gluon emission in the collinear limit

the jet mass



$\sigma_{res} = g_0 \exp[g_1(\alpha_s L)/\alpha_s + g_2(\alpha_s L) + \alpha_s g_3(\alpha_s L) + \ldots]$



$$\frac{1}{\sigma}\frac{d\sigma^{\rm LO}}{dm^2}$$

$$= \frac{\alpha_s}{2\pi}\int_0^{R^2}\frac{d\theta^2}{\theta^2}\int_0^1 dz P_{gq}(z)$$

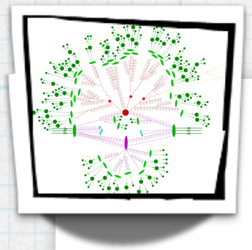
$$\times\,\delta\left(m^2 - z(1-z)\theta^2 p_T^2\right)$$

$$= \frac{\alpha_s C_F}{2\,\pi} m^{-2}\int_{m^2/(p_T^2 R^2)}^{1} dz \quad P_{gq}(z)$$

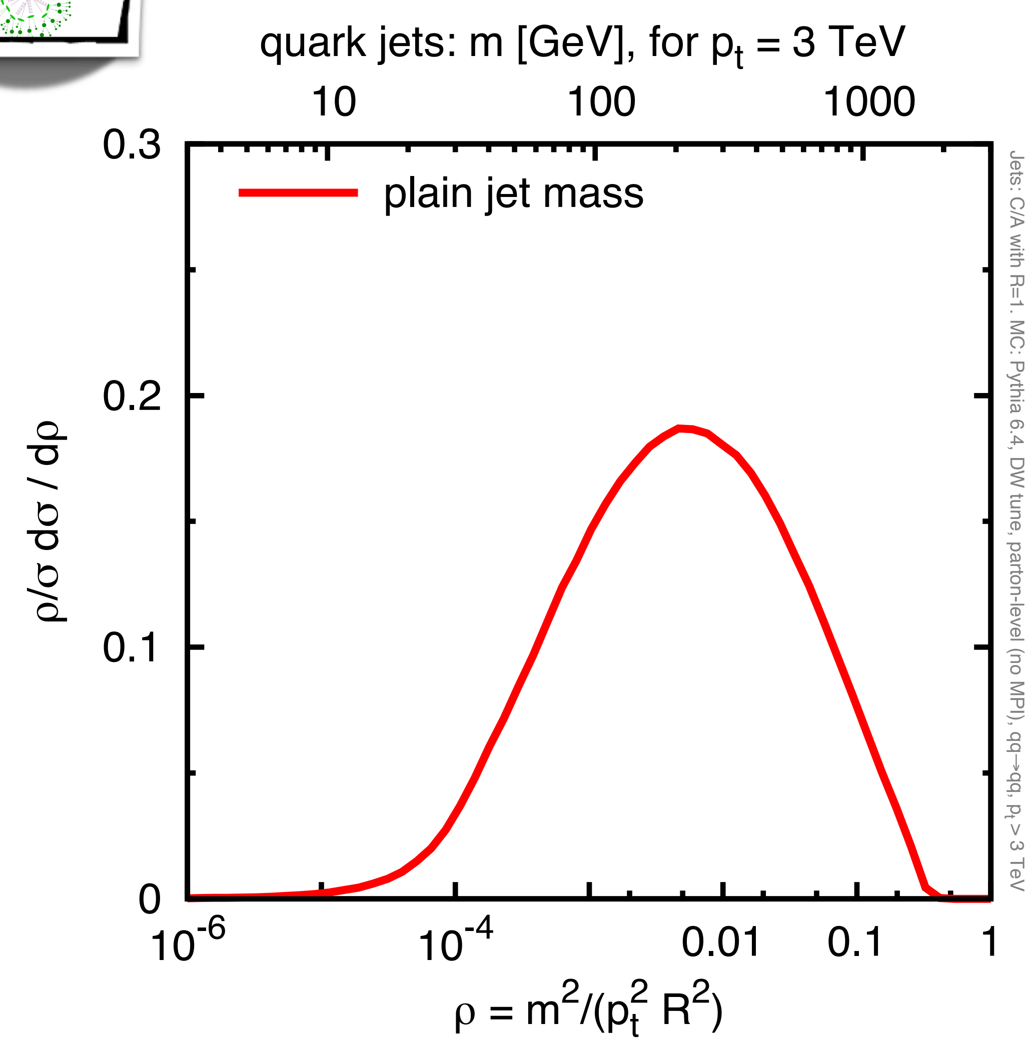
$$\simeq \frac{\alpha_s C_F}{\pi} m^{-2}\left[\ln\frac{p_T^2 R^2}{m^2} - \frac{3}{4}\right].$$

- plain jet mass: Sudakov peak, where does it come from ?
- let's do an easy calculation: one gluon emission in the collinear limit

the jet mass



$\sigma_{res} = g_0 \exp[g_1(\alpha_s L)/\alpha_s + g_2(\alpha_s L) + \alpha_s g_3(\alpha_s L) + \ldots]$



$$\frac{1}{\sigma}\frac{d\sigma^{\rm LO}}{dm^2}$$

$$= \frac{\alpha_s}{2\pi}\int_0^{R^2}\frac{d\theta^2}{\theta^2}\int_0^1 dz P_{gq}(z)$$

$$\times\, \delta\left(m^2 - z(1-z)\theta^2 p_T^2\right)$$

$$= \frac{\alpha_s C_F}{2\,\pi} m^{-2}\int_{m^2/(p_T^2 R^2)}^1 dz \qquad P_{gq}(z)$$

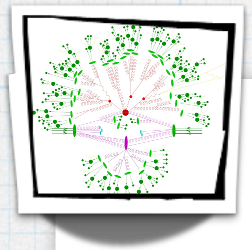
$$\simeq \frac{\alpha_s C_F}{\pi} m^{-2}\left[\ln\frac{p_T^2 R^2}{m^2} - \frac{3}{4}\right]$$

double log: soft & coll.

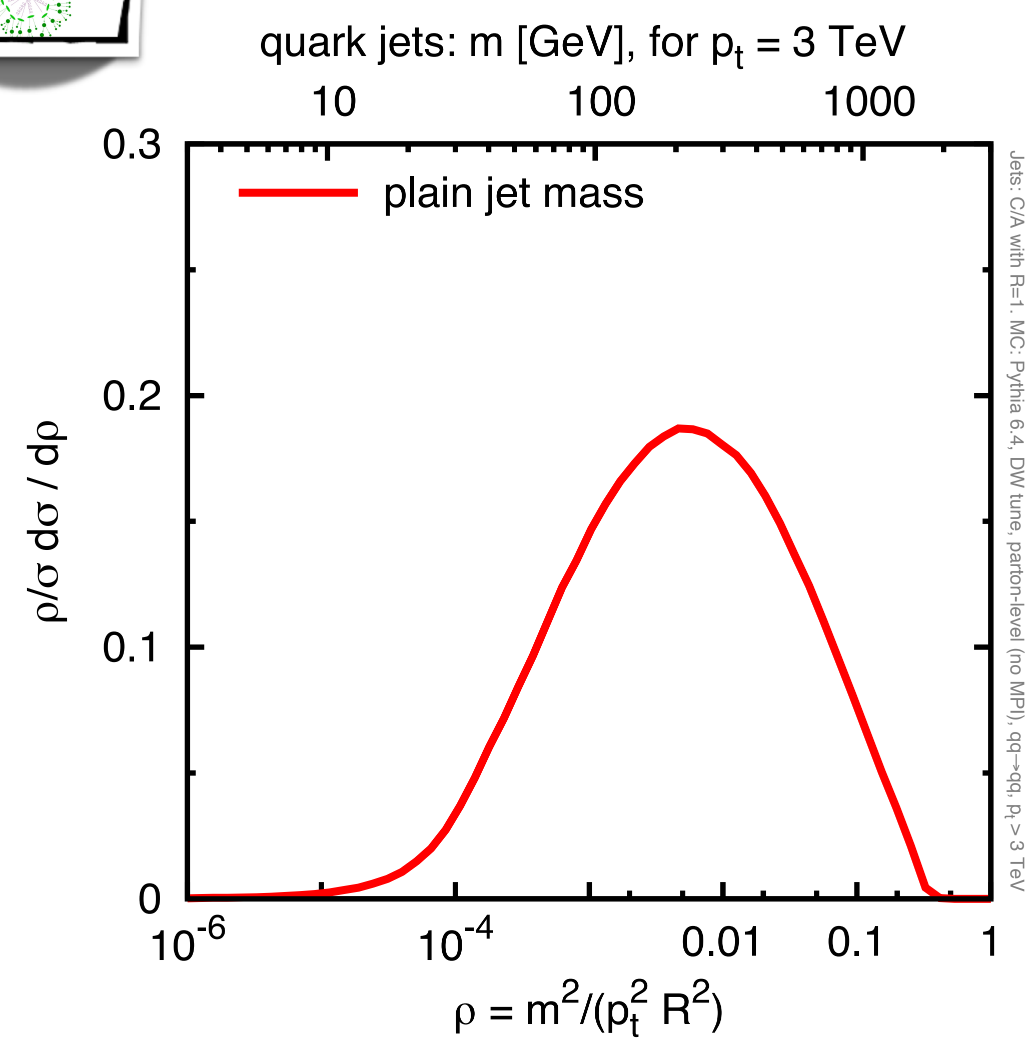
single log: hard coll.

- plain jet mass: Sudakov peak, where does it come from ?
- let's do an easy calculation: one gluon emission in the collinear limit

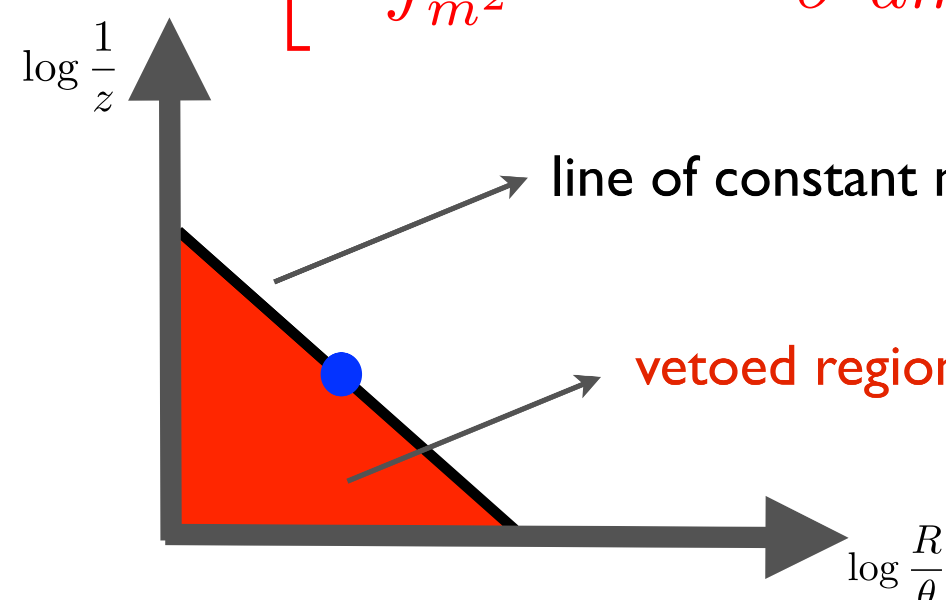
the jet mass



$\sigma_{res} = g_0 \exp[g_1(\alpha_s L)/\alpha_s + g_2(\alpha_s L) + \alpha_s g_3(\alpha_s L) + \dots]$

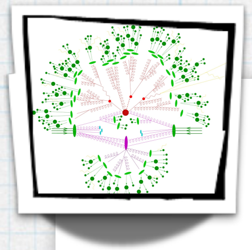


$$\frac{1}{\sigma}\frac{d\sigma^{LL}}{dm^2} = \frac{1}{\sigma}\frac{d\sigma^{LO}}{dm^2} \times \exp\left[-\int_{m^2}^{p_T^2 R^2} dm'^2 \frac{1}{\sigma}\frac{d\sigma^{LO}}{dm'^2}\right]$$

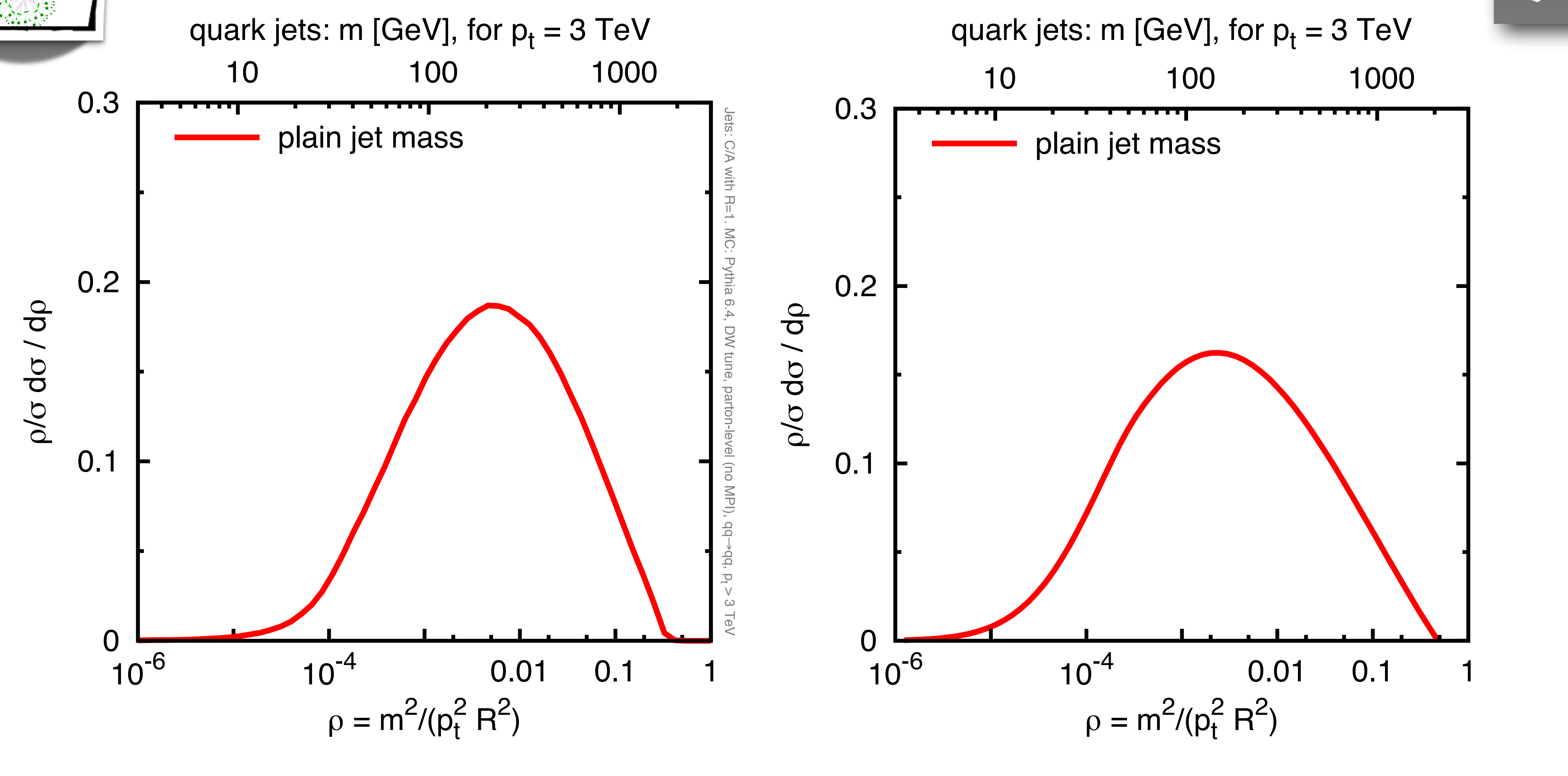


- all-order leading logs: veto emissions which would give too big a mass
- exponential that gives the no-emission probability

the jet mass



$\sigma_{res} = g_0 \exp[g_1(\alpha_s L)/\alpha_s + g_2(\alpha_s L) + \alpha_s g_3(\alpha_s L) + \dots]$

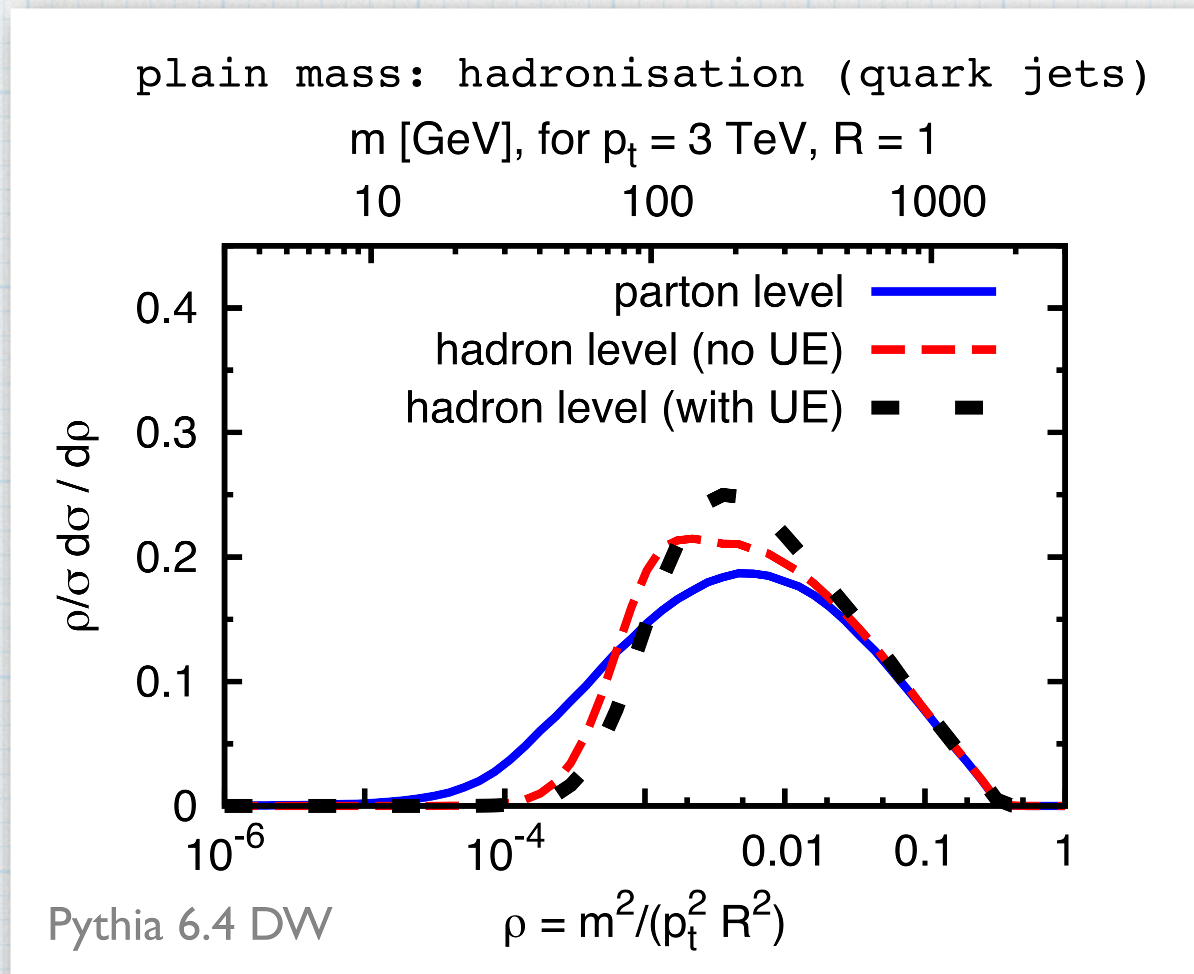


* all-order leading logs: veto emissions which would give too big a mass
* exponential that gives the no-emission probability

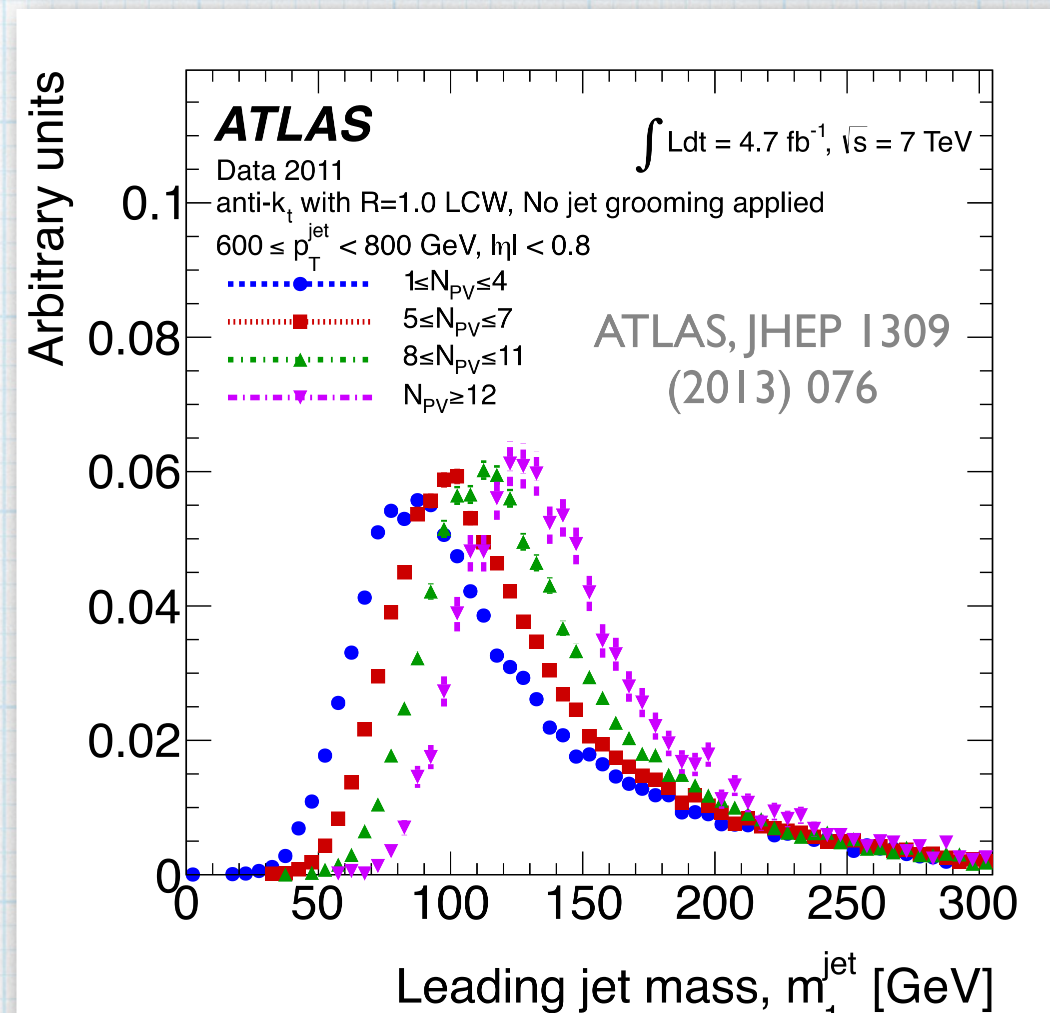
QCD-jet mass: NP effects

* first jet-observable that comes to mind

* background (QCD) jets receive important non-pert contributions



hadronisation and UE



pile-up (data!)

beyond the mass: substructure

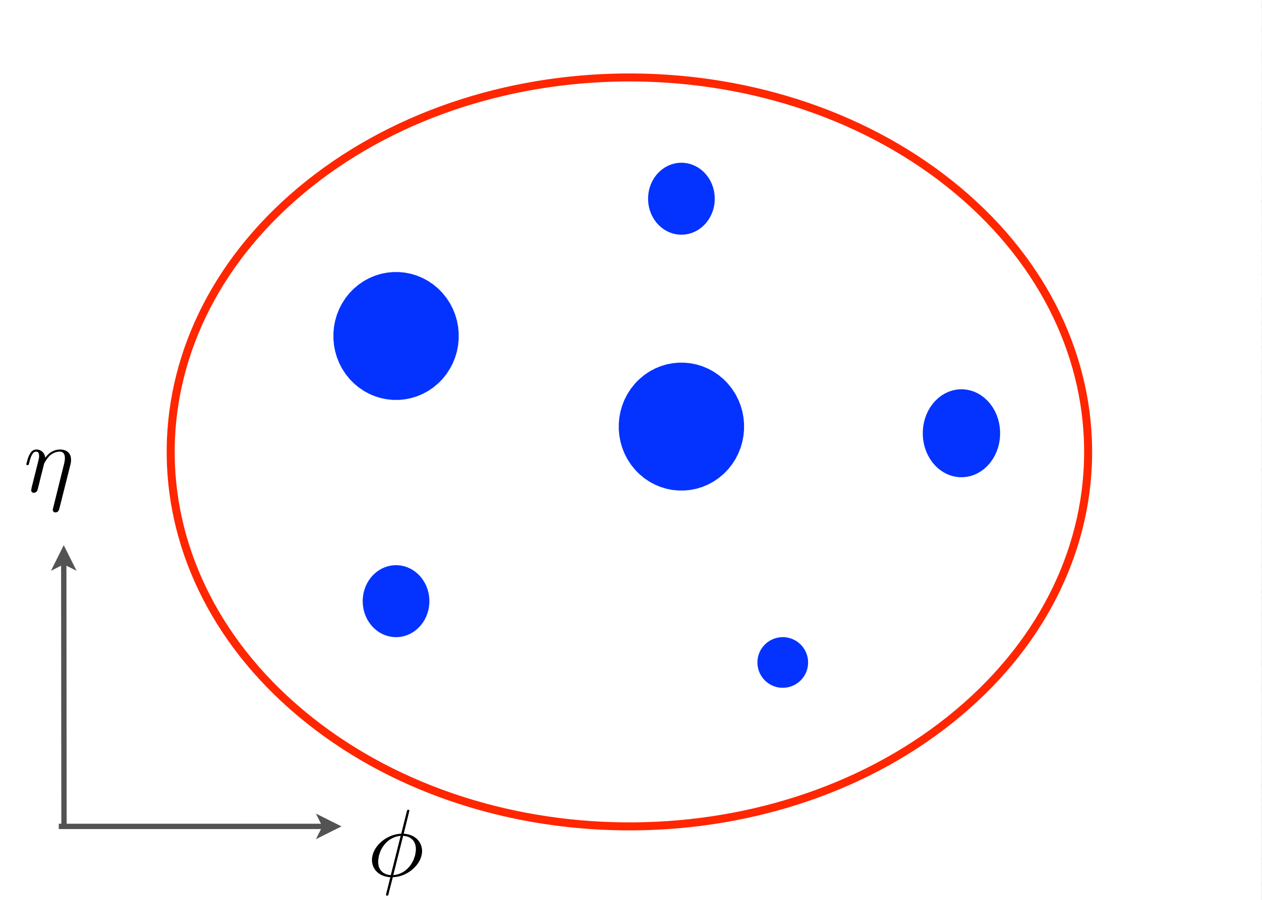
* need to go beyond the mass and exploit jet substructure : grooming and tagging:
    * clean the jets up by removing soft radiation
    * identify the features of hard decays and cut on them

beyond the mass: substructure

* need to go beyond the mass and exploit jet substructure : grooming and tagging:

  * clean the jets up by removing soft radiation

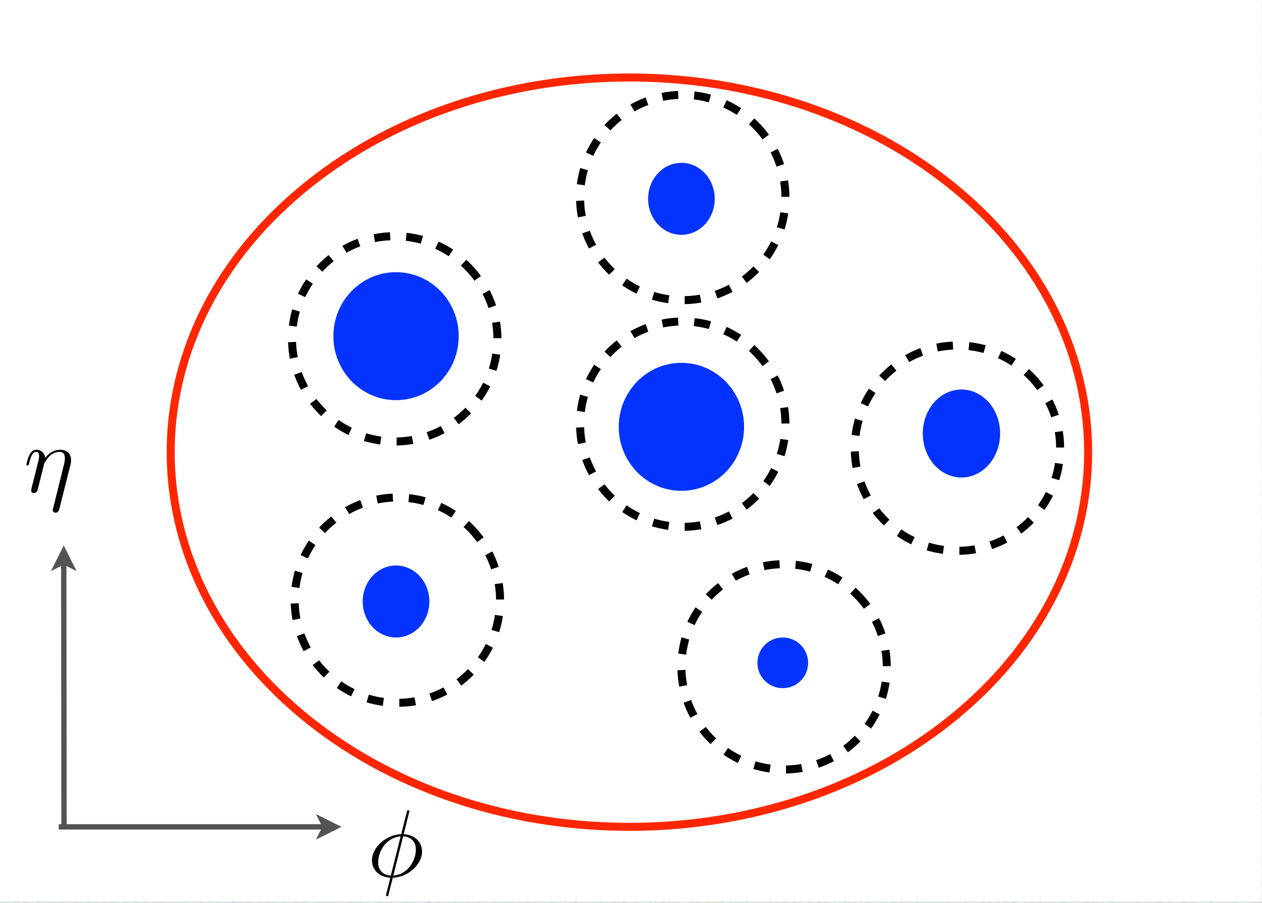
  * identify the features of hard decays and cut on them



core-idea for grooming:

beyond the mass: substructure

* need to go beyond the mass and exploit jet substructure : **grooming** and **tagging**:
  * clean the jets up by removing soft radiation
  * identify the features of hard decays and cut on them

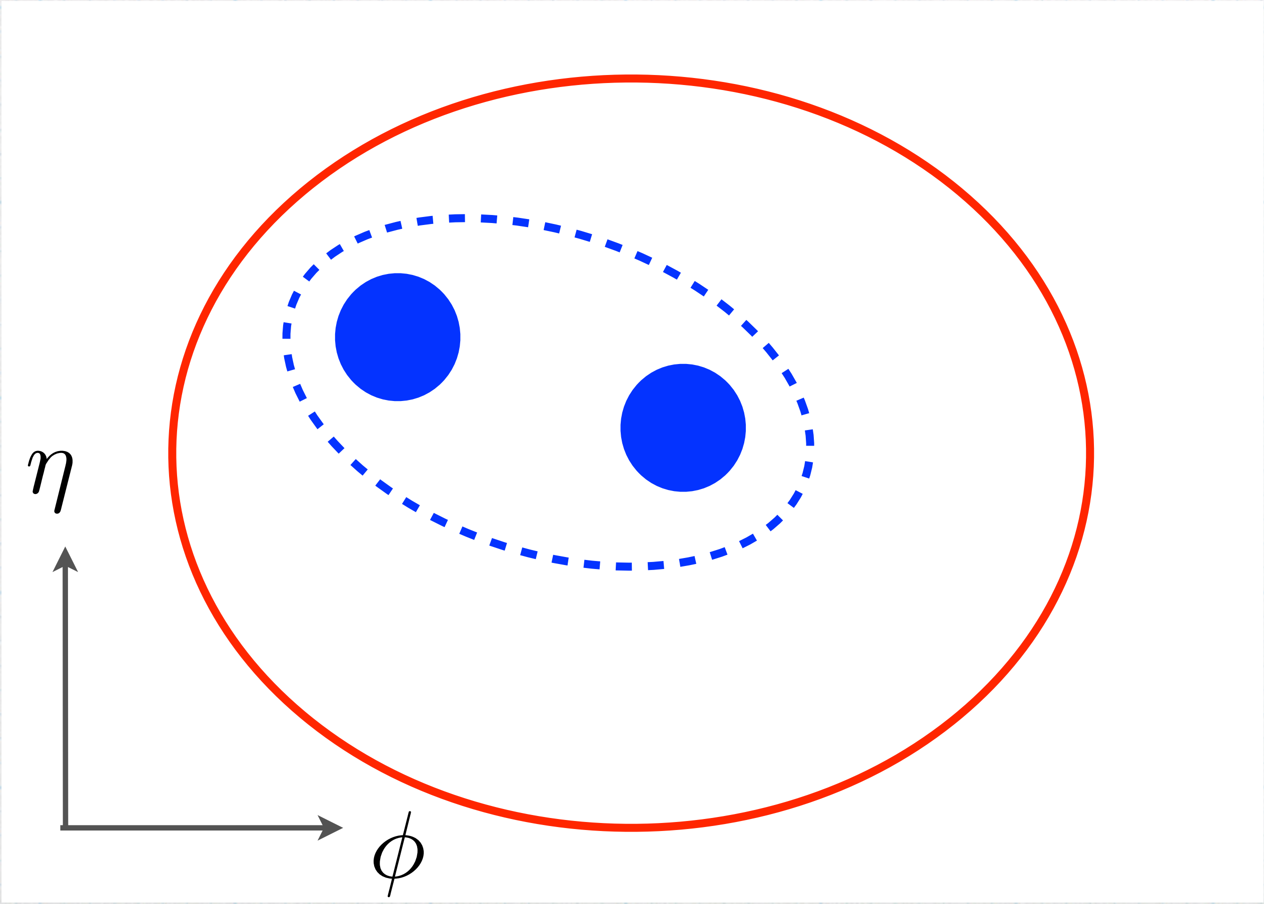


core-idea for grooming:

* identify the "right" angular scale

beyond the mass: substructure

* need to go beyond the mass and exploit jet substructure : grooming and tagging:
    * clean the jets up by removing soft radiation
    * identify the features of hard decays and cut on them



core-idea for grooming:

* identify the "right" angular scale
* throw away what is soft & large angle
* left with a groomed jet

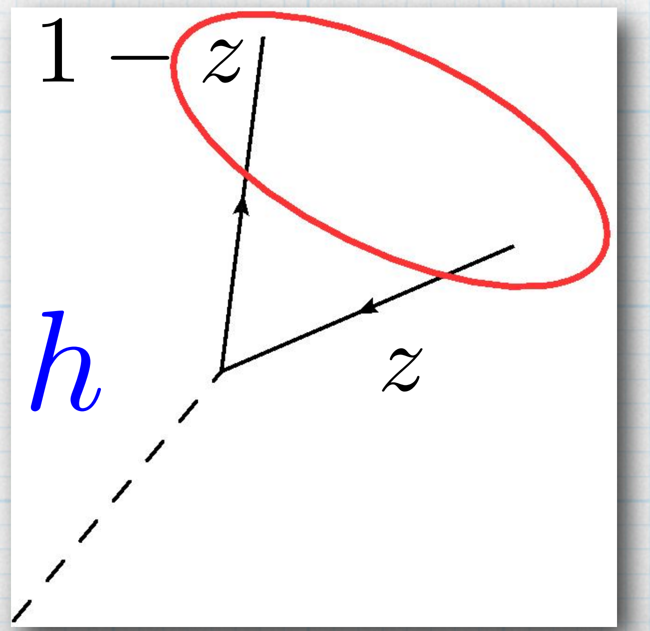
beyond the mass: substructure

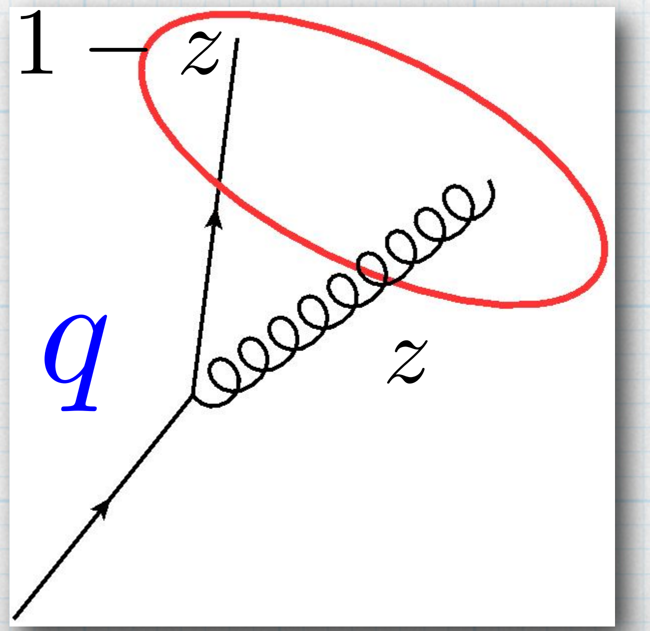
* need to go beyond the mass and exploit jet substructure : grooming and tagging:

  * clean the jets up by removing soft radiation

  * identify the features of hard decays and cut on them

core-idea for 2-body tagging:





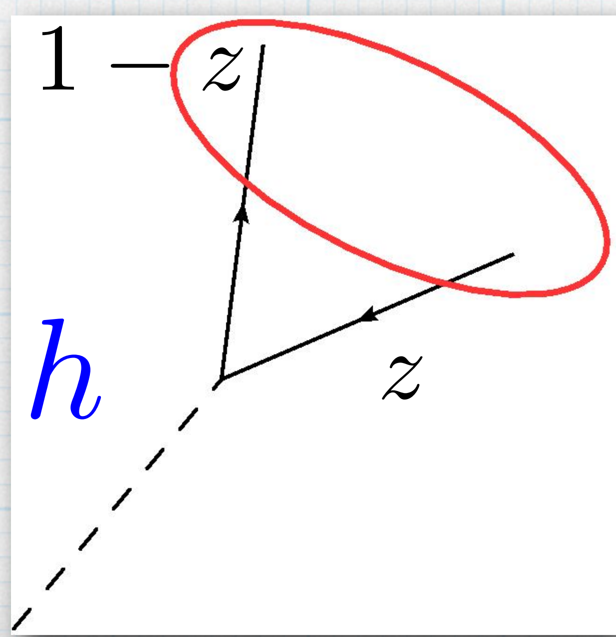
beyond the mass: substructure

* need to go beyond the mass and exploit jet substructure : **grooming** and **tagging**:
  * clean the jets up by removing soft radiation
  * identify the features of hard decays and cut on them

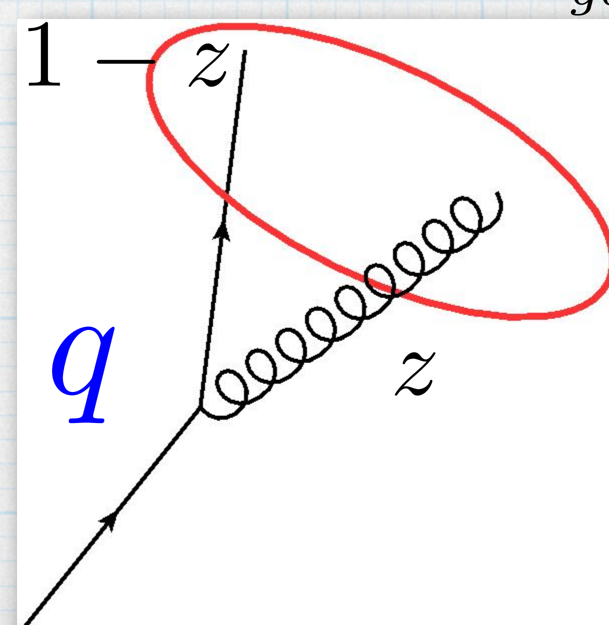
core-idea for 2-body tagging: $\min(z, 1-z) > z_{\rm cut}$

$P_{h\to q\bar{q}} = 1$

symmetric sharing of the energy



$$P_{gq} = C_F \frac{1+(1-z)^2}{z}$$

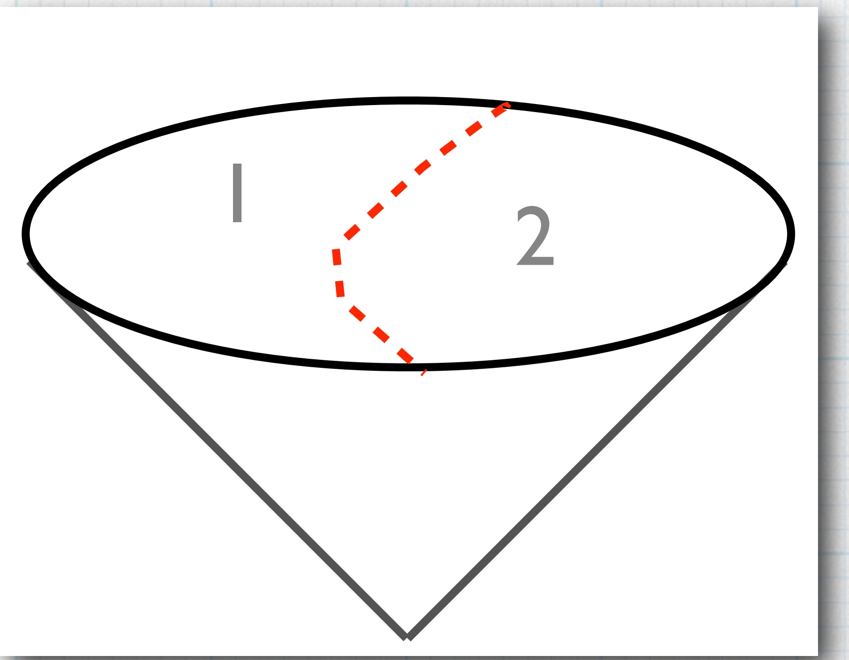


asymmetric sharing of the energy

analytic understanding at work: soft drop

Larkoski, SM, Soyez and Thaler (2014)

1. Undo the last stage of the C/A clustering. Label the two subjets j_1 and j_2.



2. If
$$\frac{\min(p_{T1}, p_{T2})}{p_{T1} + p_{T2}} > z_{\rm cut}\left(\frac{\Delta R_{12}}{R_0}\right)^{\beta}$$
then deem j to be the soft-drop jet.

3. Otherwise redefine j to be the harder subjet and iterate.

1-prong jets can be either kept (grooming mode) or discarded (tagging mode)

- generalisation of the (modified) Mass Drop procedure
- no mass drop condition (not so important)
- mMDT recovered for β=0
- some inspiration from semi-classical jets

Butterworth, Davison, Rubin and Salam (2008)
Dasgupta, Fregoso, SM and Salam (2013)

Tseng and Evans (2013)

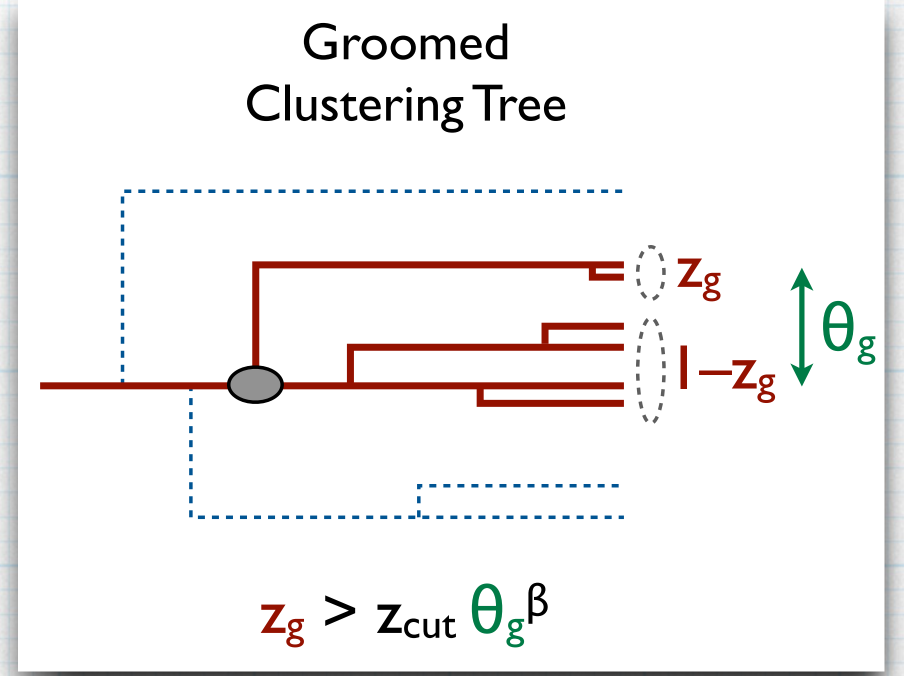
analytic understanding at work: soft drop

Larkoski, SM, Soyez and Thaler (2014)

1. Undo the last stage of the C/A clustering. Label the two subjets j_1 and j_2.

2. If $$\frac{\min(p_{T1}, p_{T2})}{p_{T1} + p_{T2}} > z_{\rm cut} \left(\frac{\Delta R_{12}}{R_0}\right)^{\beta}$$

then deem j to be the soft-drop jet.



3. Otherwise redefine j to be the harder subjet and iterate.

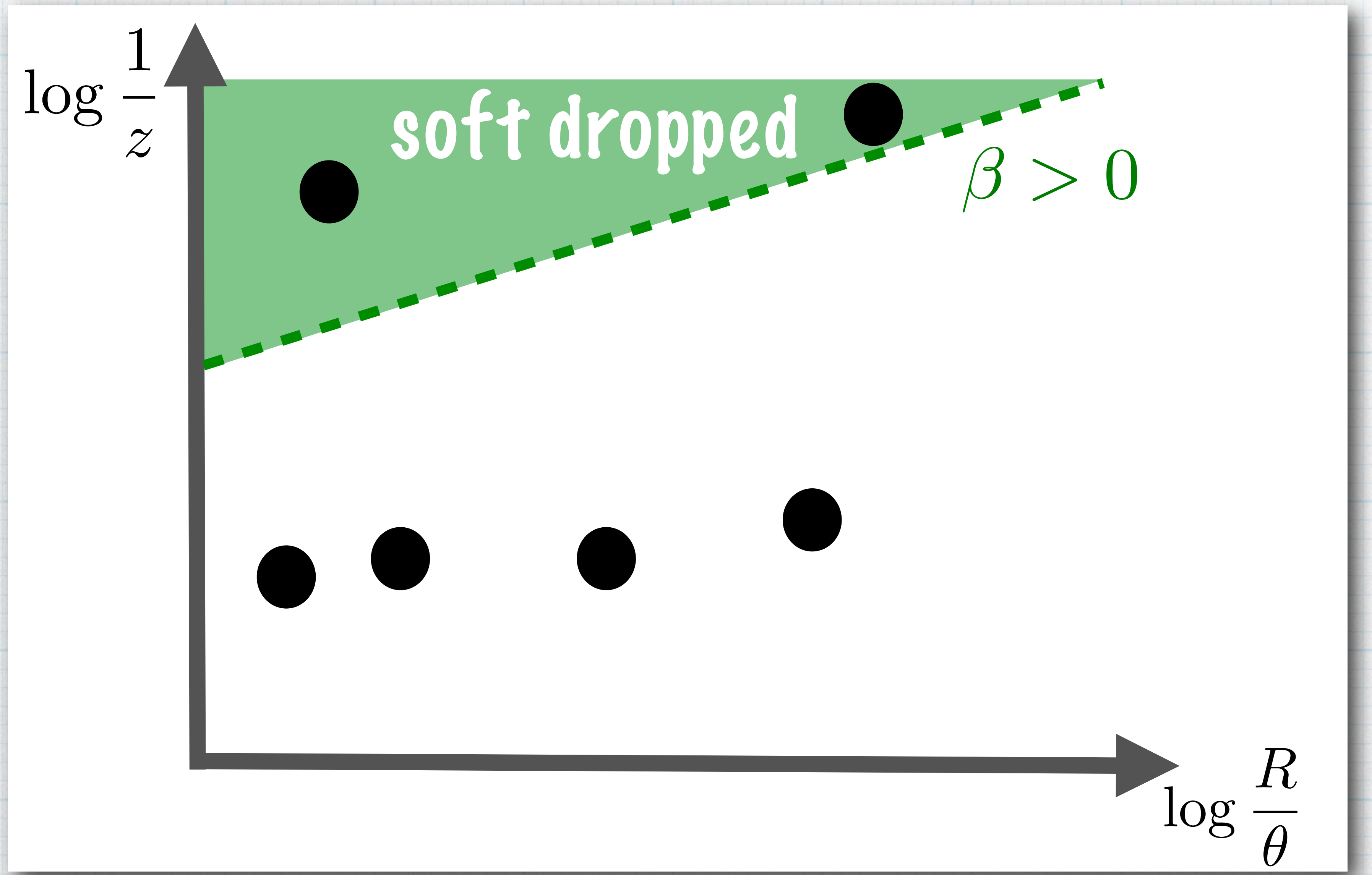
1-prong jets can be either kept (grooming mode) or discarded (tagging mode)

* generalisation of the (modified) Mass Drop procedure
* no mass drop condition (not so important)
* mMDT recovered for β=0
* some inspiration from semi-classical jets

Butterworth, Davison, Rubin and Salam (2008)
Dasgupta, Fregoso, SM and Salam (2013)

Tseng and Evans (2013)

soft drop as a groomer

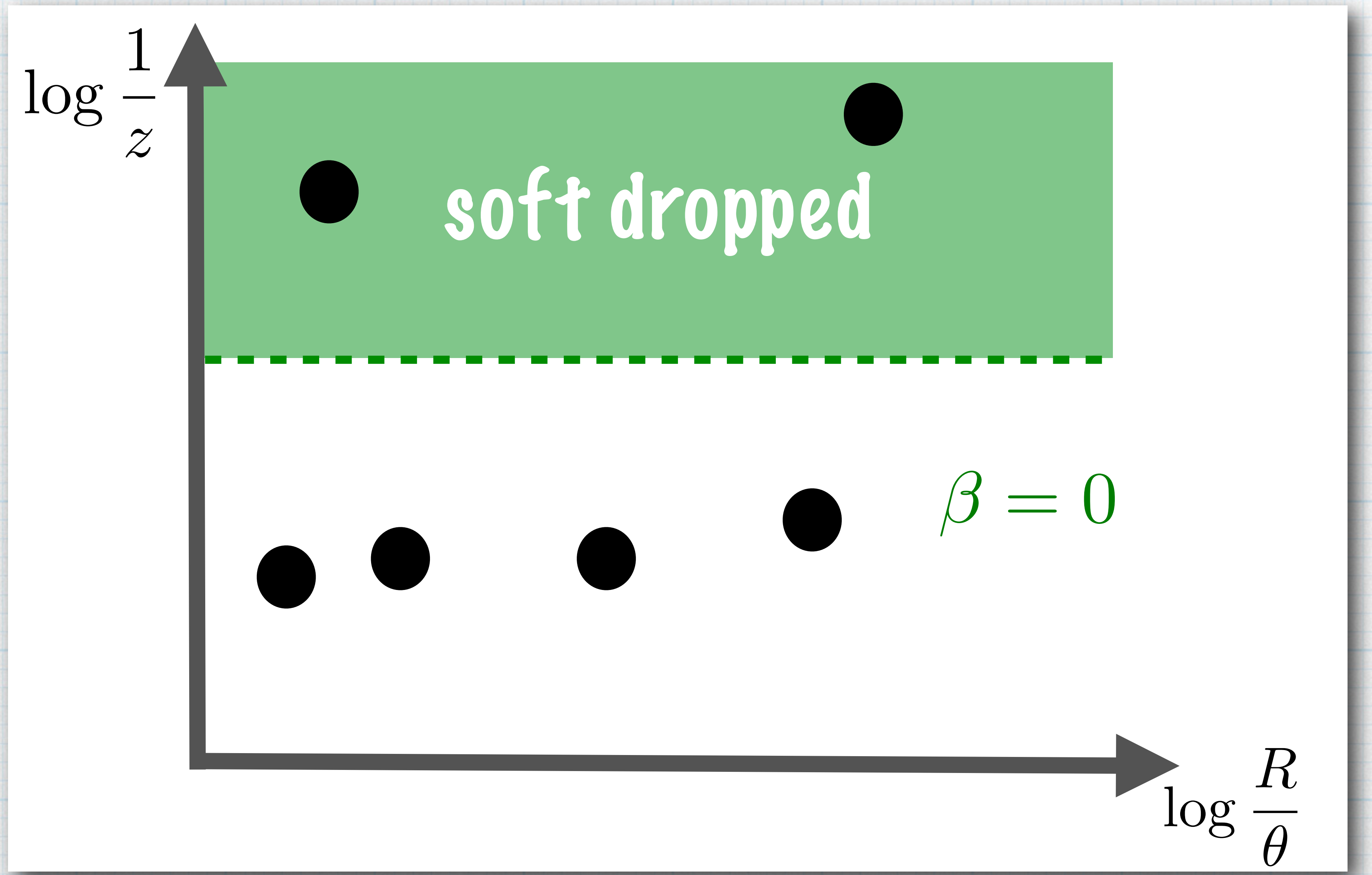


* useful to consider the soft-gluon phase space
* soft-drop condition becomes

$$z > z_{\text{cut}} \left(\frac{\theta}{R} \right)^{\beta}$$

* soft drop always removes soft radiation entirely (hence the name)
* for β>0 soft-collinear is partially removed

soft drop and mMDT

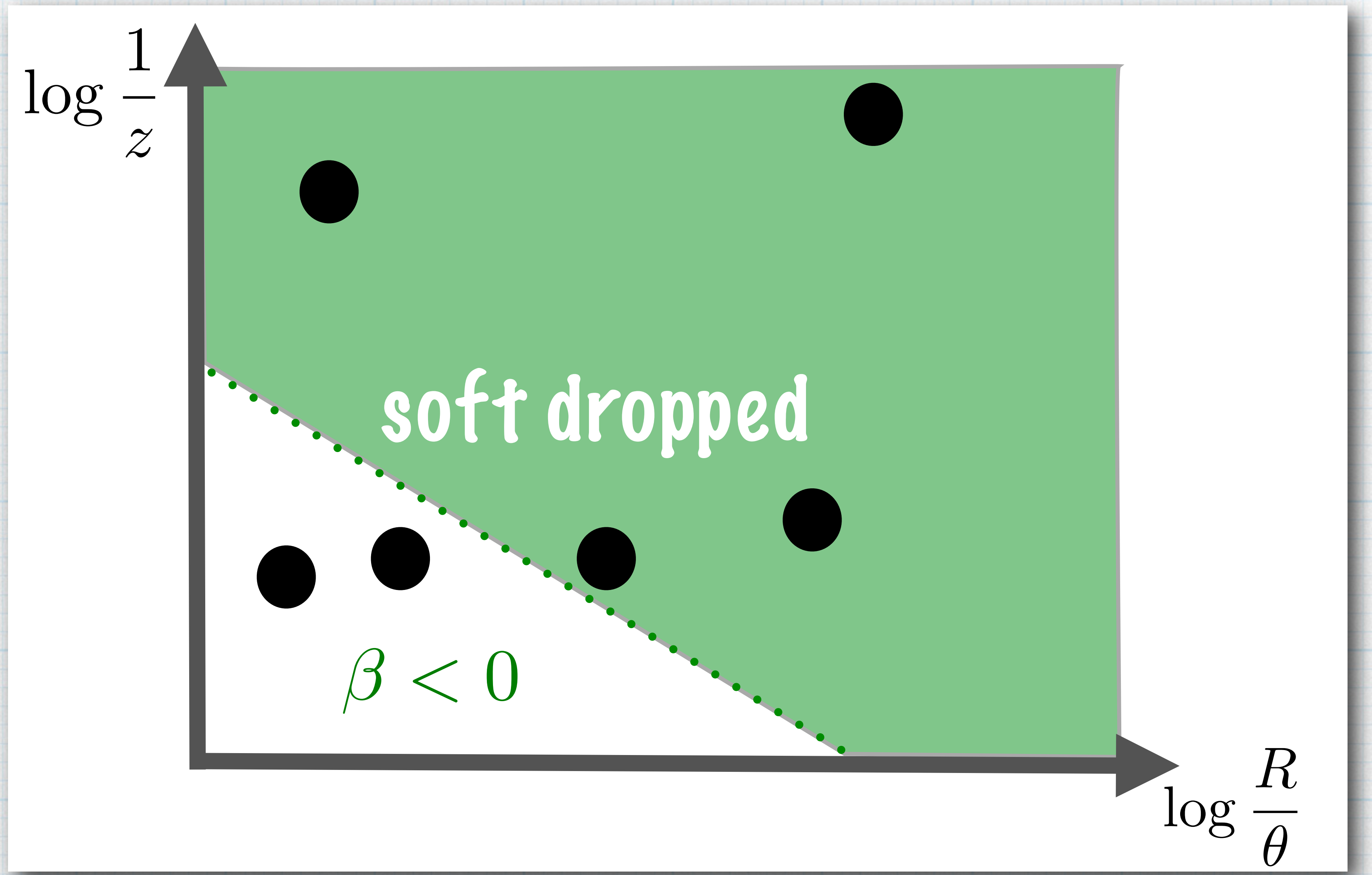


* soft-drop condition becomes

$$z > z_{\mathrm{cut}} \left(\frac{\theta}{R} \right)^{\beta}$$

* for β=0 soft-collinear is also entirely removed (mMDT limit)

soft drop as a tagger

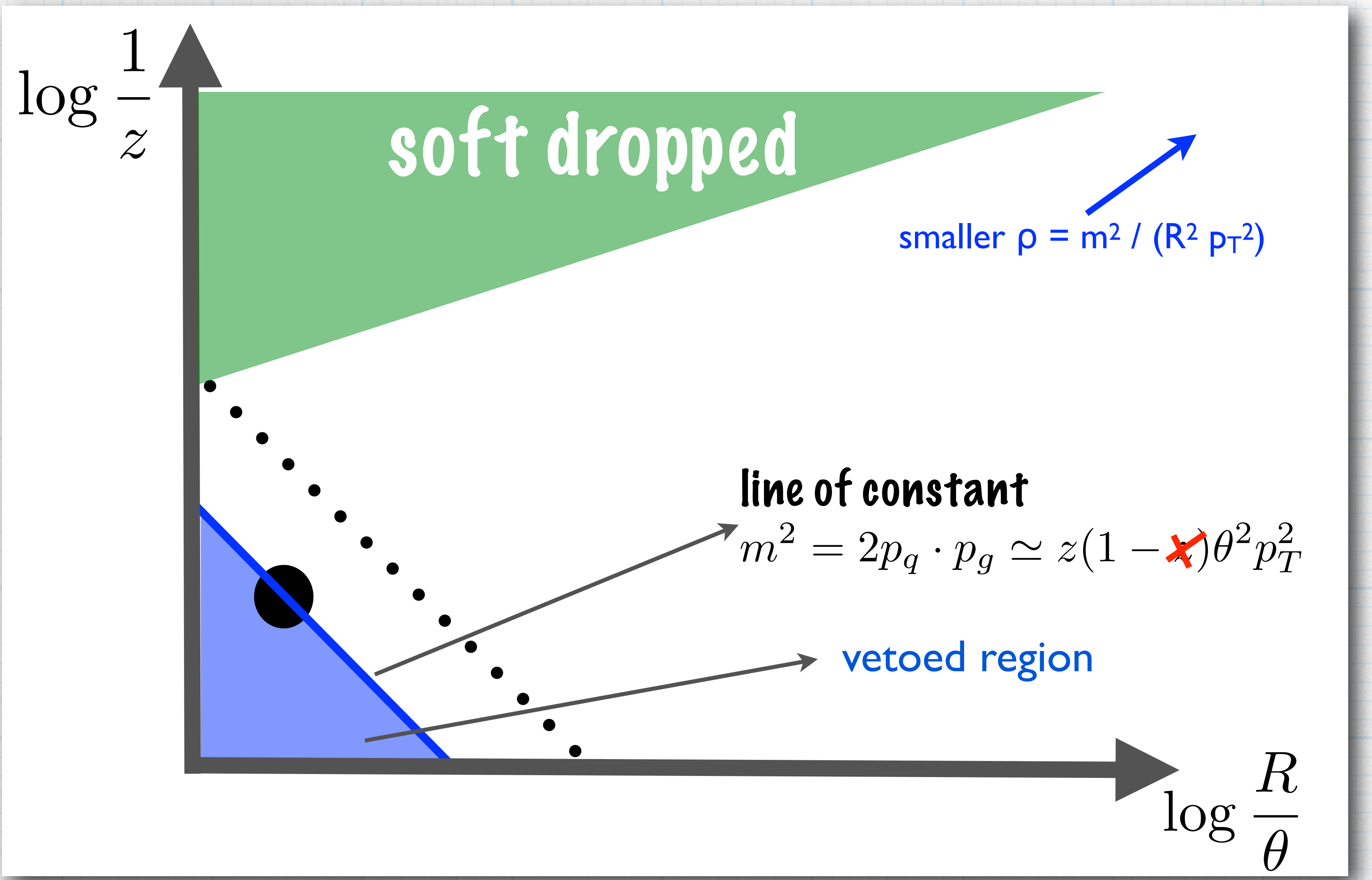


* soft-drop condition becomes

$$z > z_{\text{cut}} \left(\frac{\theta}{R} \right)^{\beta}$$

* for β<0 some hard-collinear is also partially removed

soft-drop mass at LL

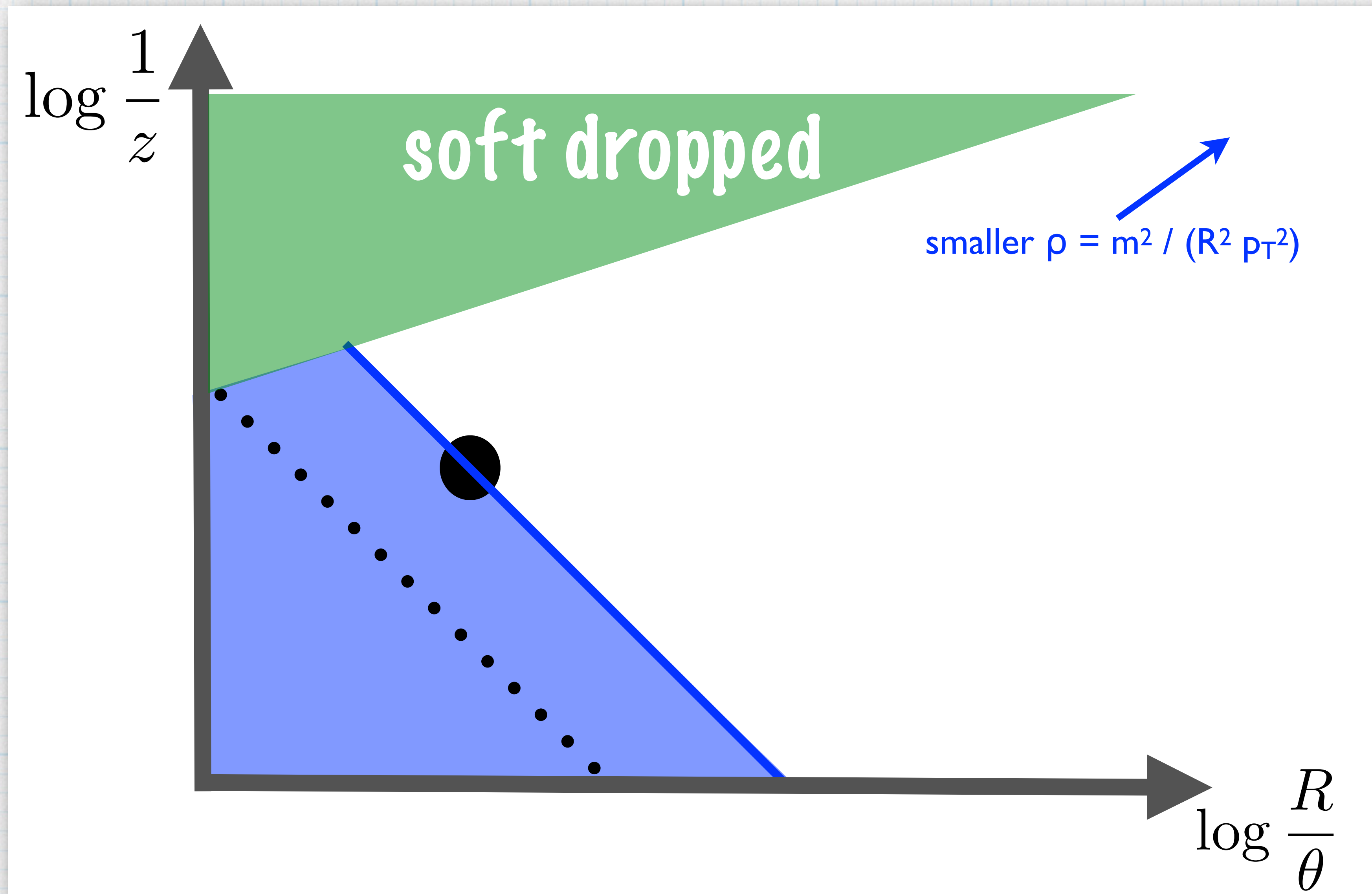


- one emission sets a **mass m**
- veto emissions that would give too big a mass
- **soft drop here has no effect**

$$\Sigma(\rho) \equiv \int^{\rho} d\rho' \frac{1}{\sigma}\frac{d\sigma}{d\rho'}$$

$$\Sigma^{(\text{s.d.})} = \exp\left[-\frac{\alpha_s C_F}{2\pi}\left[-\frac{3}{2}\ln\frac{1}{\rho} + \Theta(\rho > z_{\text{cut}})\ln^2\frac{1}{\rho} \right.\right.$$
$$\left.\left. + \Theta(z_{\text{cut}} > \rho)\left(\frac{\beta}{2+\beta}\ln^2\frac{1}{\rho} + \frac{2}{2+\beta}\left(\ln^2\frac{1}{z_{\text{cut}}} + 2\ln\frac{z_{\text{cut}}}{\rho}\ln\frac{1}{z_{\text{cut}}} \right)\right)\right]\right]$$

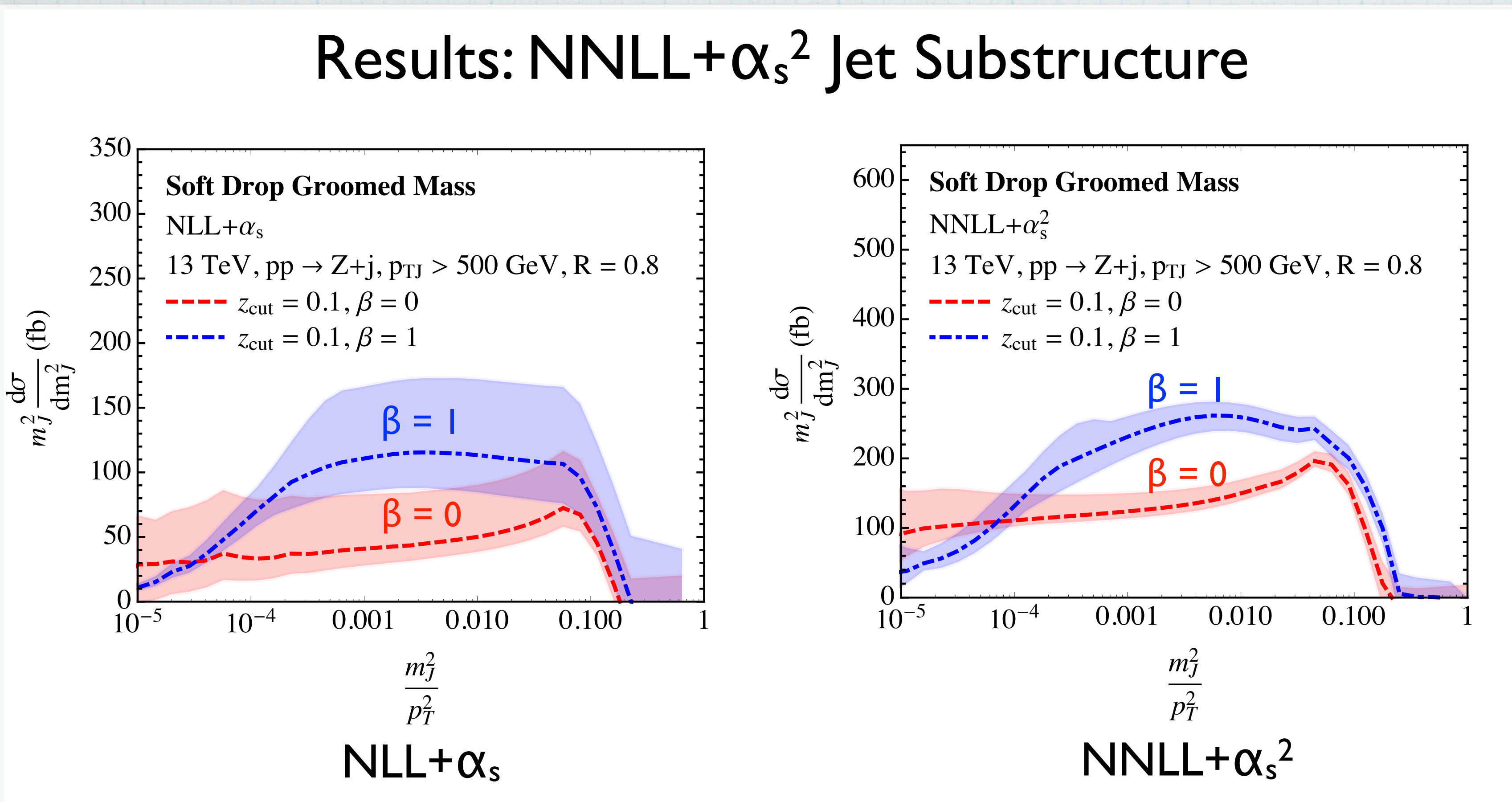
soft-drop mass at LL



* only one transition point at $\rho = z_{cut}$
* soft & soft collinear radiation is partially removed
* only single logs for β=0!

$$\Sigma^{(\mathrm{s.d.})} = \exp\left[-\frac{\alpha_s C_F}{2\pi} \left[-\frac{3}{2}\ln\frac{1}{\rho} + \Theta(\rho > z_{\mathrm{cut}}) \ln^2\frac{1}{\rho} + \Theta(z_{\mathrm{cut}} > \rho) \left(\frac{\beta}{2+\beta}\ln^2\frac{1}{\rho} + \frac{2}{2+\beta}\left(\ln^2\frac{1}{z_{\mathrm{cut}}} + 2\ln\frac{z_{\mathrm{cut}}}{\rho}\ln\frac{1}{z_{\mathrm{cut}}} \right) \right) \right] \right]$$

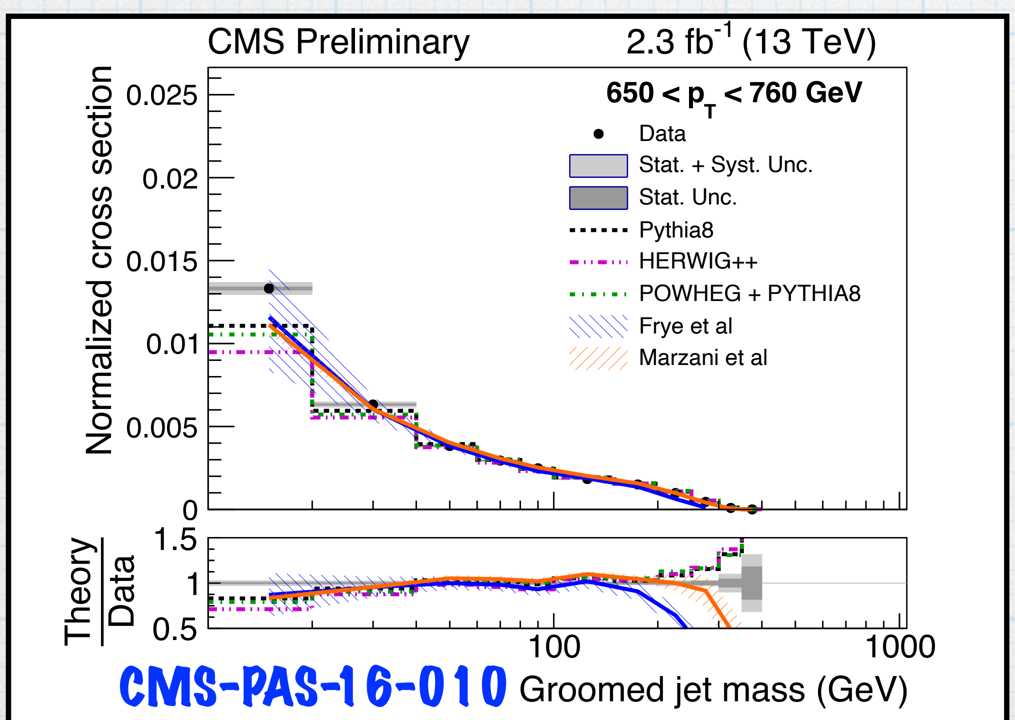
precision jet substructure



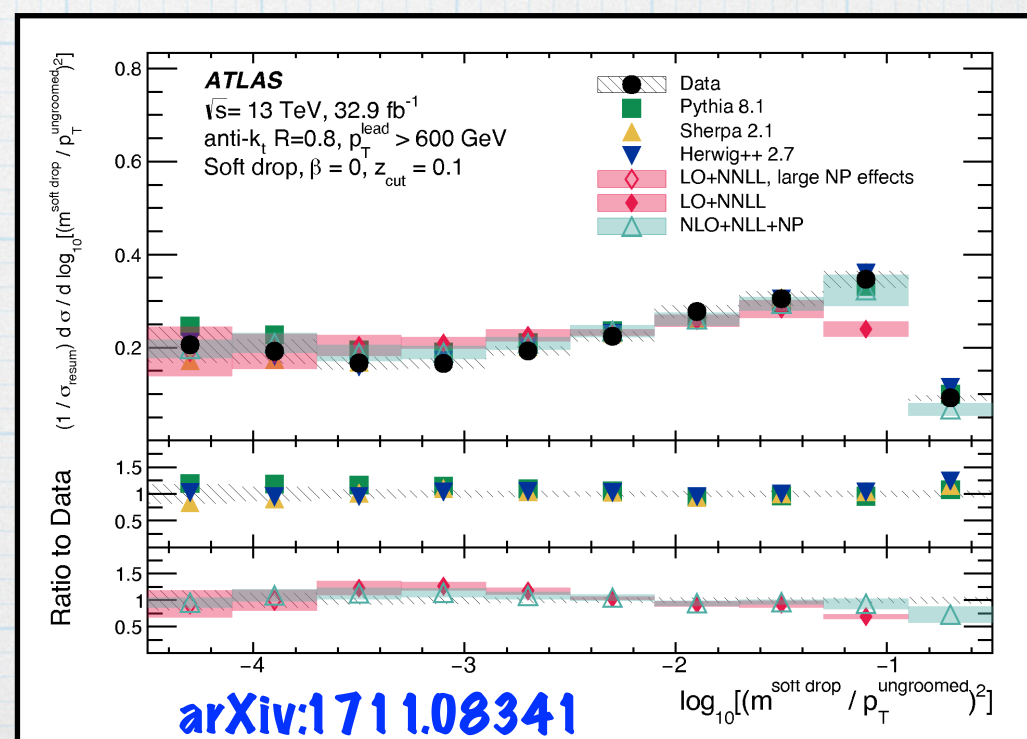
- using SCET, precision pushed to NNLL
- no non-global logs
- no colour correlations

Frye, Larkoski, Schwartz, Yan (2016)

and data!

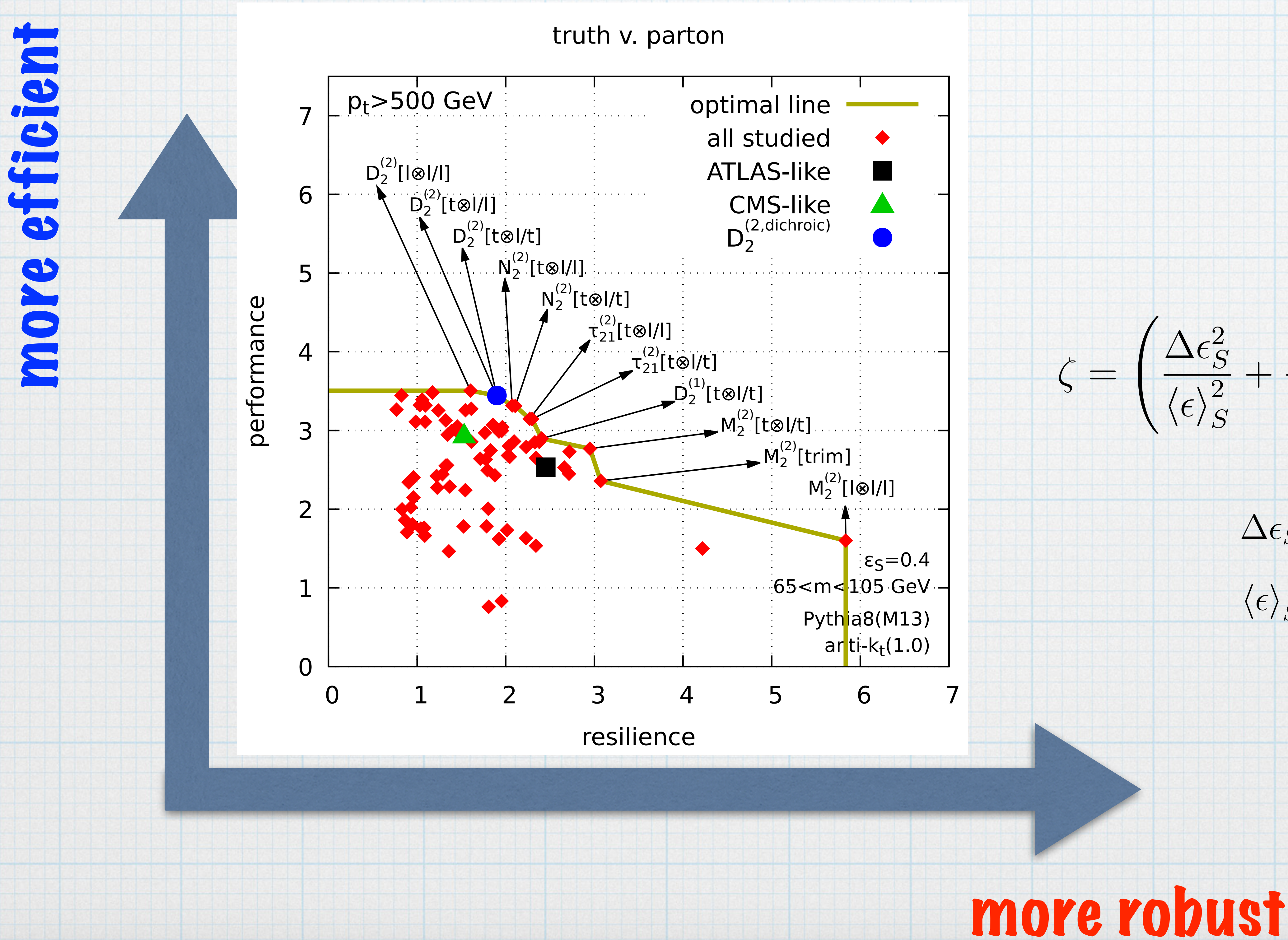


CMS-PAS-16-010



arXiv:1711.08341

performance & resilience



$$\zeta = \left(\frac{\Delta\epsilon_S^2}{\langle\epsilon\rangle_S^2} + \frac{\Delta\epsilon_B^2}{\langle\epsilon\rangle_B^2} \right)^{-1/2}$$

$$\Delta\epsilon_{S,B} = \epsilon_{S,B} - \epsilon'_{S,B},$$

$$\langle\epsilon\rangle_{S,B} = \frac{1}{2}\left(\epsilon_{S,B} + \epsilon'_{S,B}\right)$$

summary of lecture 2

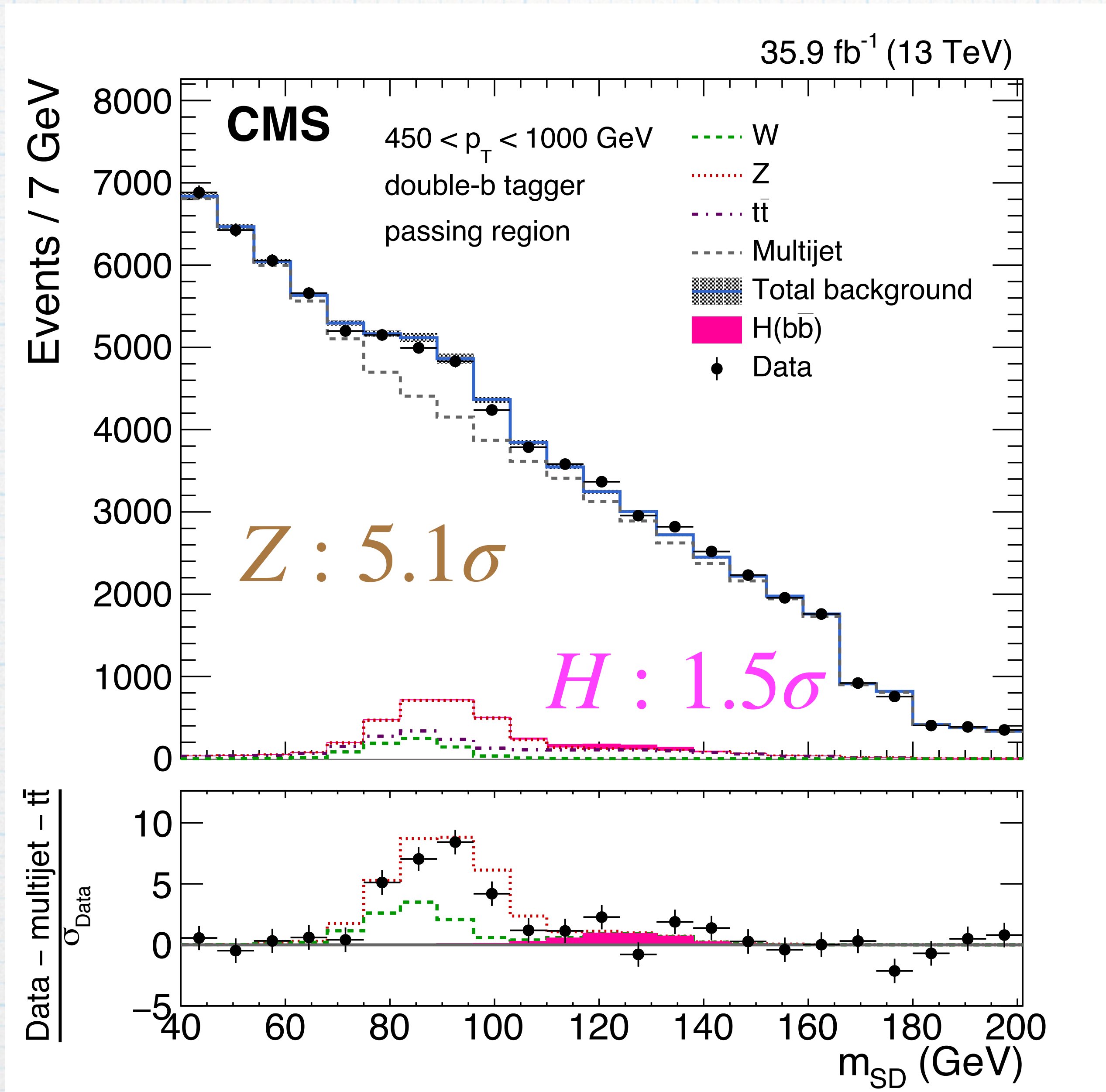
more efficient

* boosted-object physics
* grooming and tagging
* precision substructure physics with soft-drop

more robust

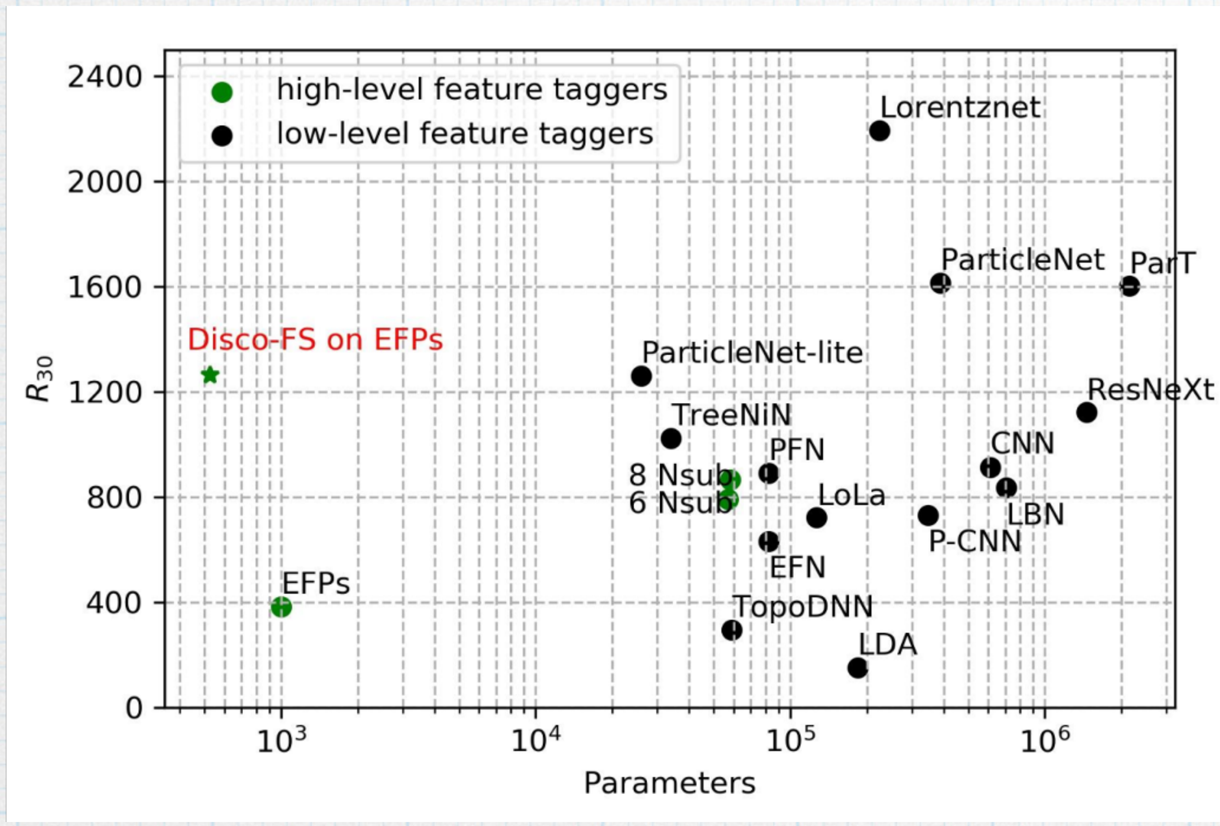
summary of lecture 2

more efficient



more robust

looking at ML



adapted from Ranit Das talk at BOOST 2022

more efficient

more robust ???

* challenge: what is the best metric to measure robustness for ML algorithms?

homework 3

* Gluon splitting into bottom quarks g→bb is important for H→bb studies. What's its average mass? (take m_b=0)