The Strong Interaction and LHC phenomenology

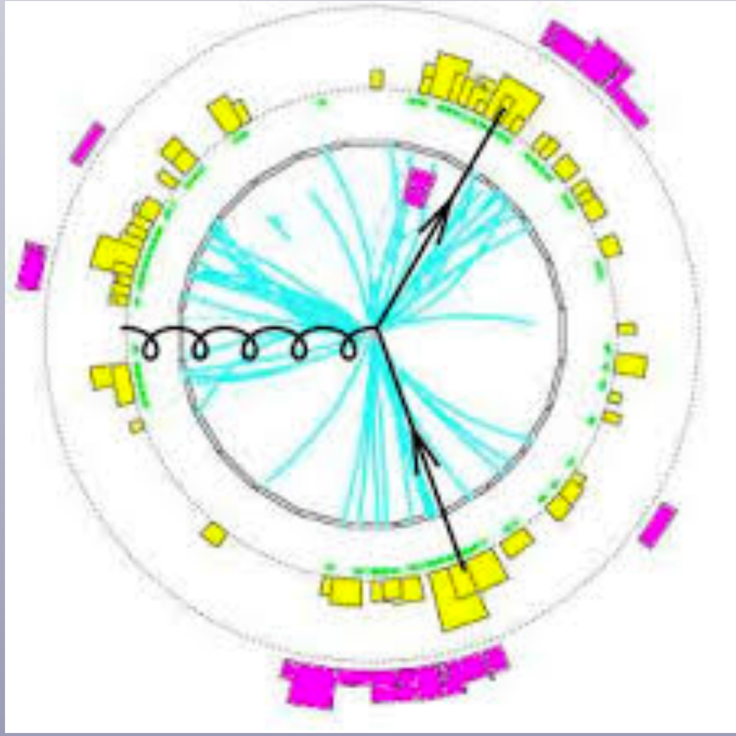
Juan Rojo
STFC Rutherford Fellow
University of Oxford

Theoretical Physics Graduate School course

Lecture 8:
Jet Reconstruction and Jet Substructure

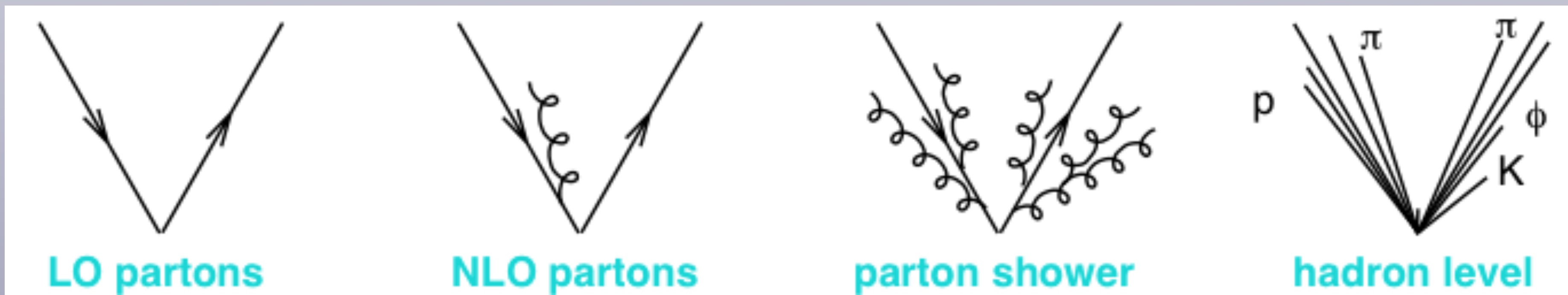
Jet Reconstruction

Main Reference:
G. Salam, “Towards Jetography”, arxiv:0906.1833



Jets in hadron collisions

- QCD calculations are provided in terms of **quarks and gluons in the final state**
- After hard scattering, quarks and gluons follow a **branching process** and then **hadronize**, leading to a collimated bunch of hadrons as characteristic signal n the detector: **a QCD jet**
- A quantitative robust **mapping between final state hadrons**, observed in the detector, and **partons from** the hard-scattering, is required to compare theory with data: a **jet algorithm**
- We already saw a first example of jet algorithm: the **Sterman-Weinberg jets** in electron-positron annihilation (which however in **collinear unsafe at O(α_S^2)**)
- Here we explore **modern jet reconstruction algorithms** with enfasis on LHC phenomenology



A jet is a fundamentally ambiguous concept
A jet algorithm is required to provide a jet definition
Good jet algorithms can be applied to partons, hadrons and calorimeter cells

Jet algorithms

- The mapping between quarks and **gluons / hadrons / calorimeter towers** and **jets** is provided by the so called **jet algorithms**

- In addition, we need to specify which four-momenta is assigned to the combination of two particles when forming the jet: this is known as the **recombination scheme**

- The **jet algorithm** with its corresponding parameters, like jet radius, together with the **recombination scheme,** fully determine the **jet definition**

- In 1990, the **Snowmass accord** was set out to specify the **desirable properties** of good jet algorithms

Several important properties that should be met by a jet definition are [3]:

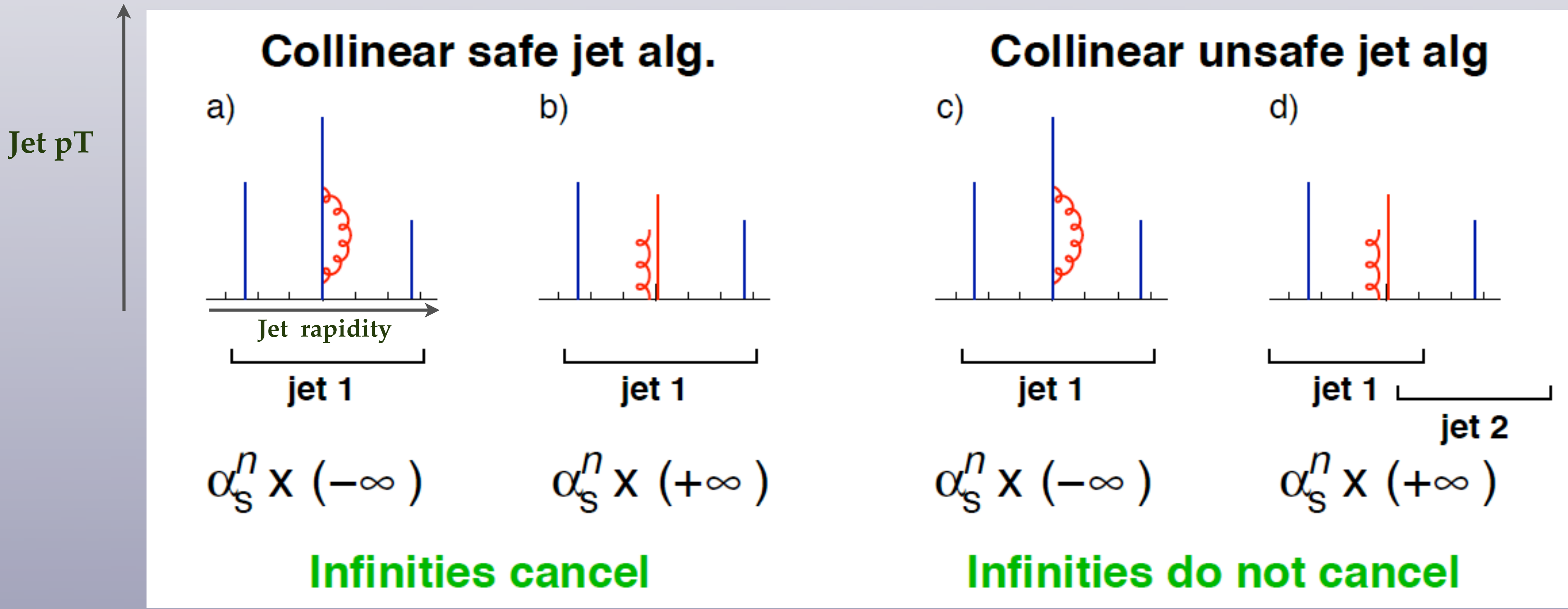
1. Simple to implement in an experimental analysis;
2. Simple to implement in the theoretical calculation;
3. Defined at any order of perturbation theory;
4. Yields finite cross sections at any order of perturbation theory;
5. Yields a cross section that is relatively insensitive to hadronisation.

- A crucial property of a jet algorithm is that it should be **infrared safe**, to avoid depending on the unknown, long-distance properties of QCD where perturbation theory breaks downs

- Many of the jet algorithms in the last 20 years were **infrared unsafe.** Fortunately, this is not the case at the LHC anymore

Jet algorithms

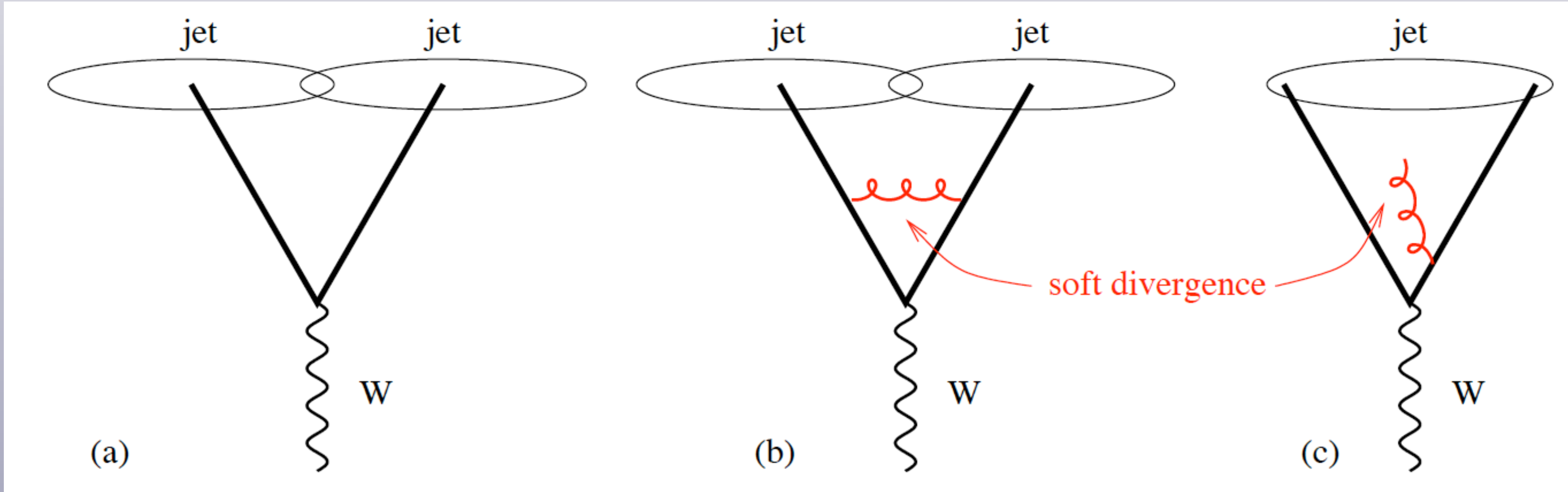
- In the particular case of jet algorithms, **infrared safety** can be formulated as the requirement that if the **final state particles** are modified by a **soft emission** or a **collinear splitting** then the set of hard jets found should be unchanged
- Failing this criterion, a jet definition will produce **infinite results** at some point in the perturbative expansion because of the **lack of cancellation** of infrared divergences



In the **IRC unsafe algorithm**, a **collinear splitting** leads to a **different set of final state jets** and thus to the lack of cancellation of soft and collinear divergences (KLN theorem)

Jet algorithms

- In the particular case of jet algorithms, **infrared safety** can be formulated as the requirement that if the **final state particles** are modified by a **soft emission** or a **collinear splitting** then the set of hard jets found should be unchanged
- Failing this criterion, a jet definition will produce **infinite results** at some point in the perturbative expansion because of the **lack of cancellation** of infrared divergences



In the **IRC unsafe algorithm**, a **soft emission** leads to a **different set of final state jets** and thus to the lack of cancellation of soft and collinear divergences (KLN theorem)

IRC unsafety

From the practical point of view, using a **IRC unsafe algorithm** severely degrades the value of any jet measurement because it becomes impossible to compare with many **theoretical calculations**

Last meaningful perturbative order for various IRC unsafe algorithms

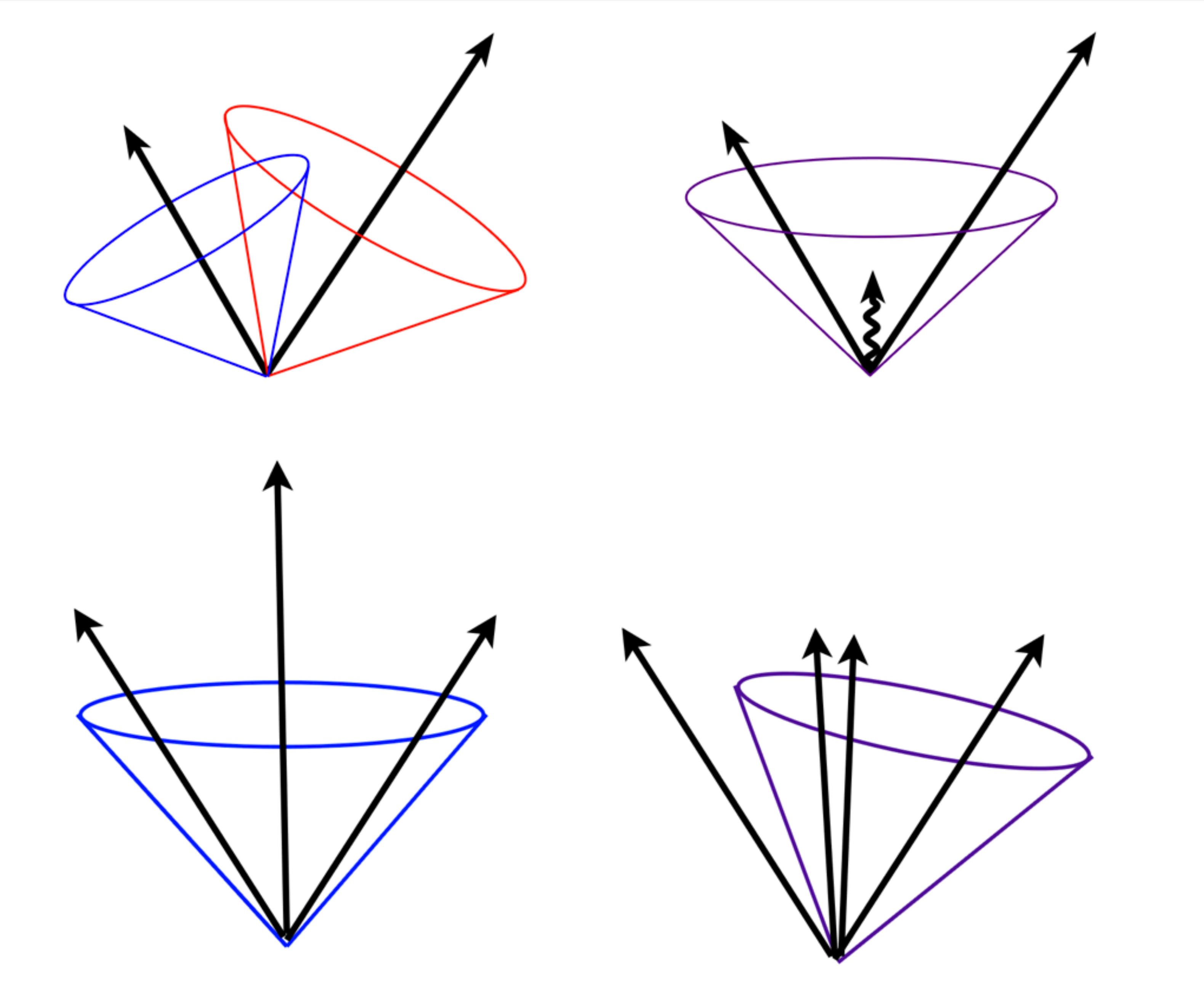
Observable	IR_{2+1}	IR_{3+1}, Coll_{3+1}
Inclusive jet cross section	LO	NLO
$W/Z/H$ + 1-jet cross section	LO	NLO
3-jet cross section	none	LO
$W/Z/H$ + 2-jet cross section	none	LO
jet masses in 3-jet and $W/Z/H$ + 2-jet events	none	none

Until 2007, all the **cone jet algorithms used** (JetClu, MidPoint, ...) were IRC unsafe

It is possible to construct a cone jet algorithm which is IRC safe: **SISCone**

In the following we concentrate on **sequential recombination algorithms** which are the ones mostly used by LHC analysis

Jet cone algorithms



Defining jets in terms of cones, it is easy to find **infrared unsafy**

A crucial problem was the choice of **seeds** for the search of stable cones

Cone algorithms were popular in the experiments because of their apparently regular shape, which simplified treatment of jet corrections

Sequential recombination jet algorithms

These jet algorithms are based on defining some measure of **how likely two partons are to have arised from a QCD splitting,** and proceed sequentially to construct the jet by reconstructing the partons which are closer in this measure

The **inclusive kt algorithm** for hadron collisions can be formulated as follows. First define

$$
\begin{aligned}
d_{ij} &\equiv \min\left(p_{T,i}^2, p_{T,j}^2\right) \frac{\Delta R_{ij}^2}{R^2} \qquad \Delta R_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2 \\
d_{iB} &\equiv p_{T,i}^2
\end{aligned}
$$

were **R** is the **jet radius**

The basic idea is that this distance reproduces the **infrared singularity structure of QCD branchings**

$$
d\sigma_{n+1} = d\sigma_n \frac{dt}{t} dz \frac{\alpha_s}{2\pi} \hat{P}_{ab}(z)
$$

Collinear singularity

Soft Singularities

The kt algorithm distance is **smallest** when the **QCD matrix element is dominant**

The algorithm also accounts for **initial state soft and collinear splittings** introducing the distance to the hadrom beam

$$
t \simeq M_{ij} \simeq z(1-z)\, p_{T,ij}^2\, \theta^2 \qquad p_{T,i} \sim z\, p_{T,ij} \qquad p_{T,j} \sim (1-z)\, p_{T,ij}
$$

Sequential recombination jet algorithms

These jet algorithms are based on defining some measure of **how likely two partons are to have arised from a QCD splitting,** and proceed sequentially to construct the jet by reconstructing the partons which are closer in this measure

The **inclusive kt algorithm** for hadron collisions can be formulated as follows. First define

$$d_{ij} \equiv \min\left(p_{T,i}^2, p_{T,j}^2\right) \frac{\Delta R_{ij}^2}{R^2} \qquad \Delta R_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$$

$$d_{iB} \equiv p_{T,i}^2$$

- From the list of all final state particles, determine all the **distances d_{ij} and d_{iB}**
- Find the minimum distance
- If it is a d_{ij}, recombine particles **i** and **j** and go back to the first step
- Otherwise, declare **particle i** to be a **jet** and **remove** from the list of particles. Back to step 1
- Stop the algorithm when no particles remain

It is important to emphasize that **sequential recombination algorithms** can be applied exactly in the same way to **QCD partons**, **hadrons** or at the **detector level**

Allows straightforward comparison between **data and theory**

Sequential recombination jet algorithms

It is possible to **generalize the kt algorithm** by introducing a modified distance as follows

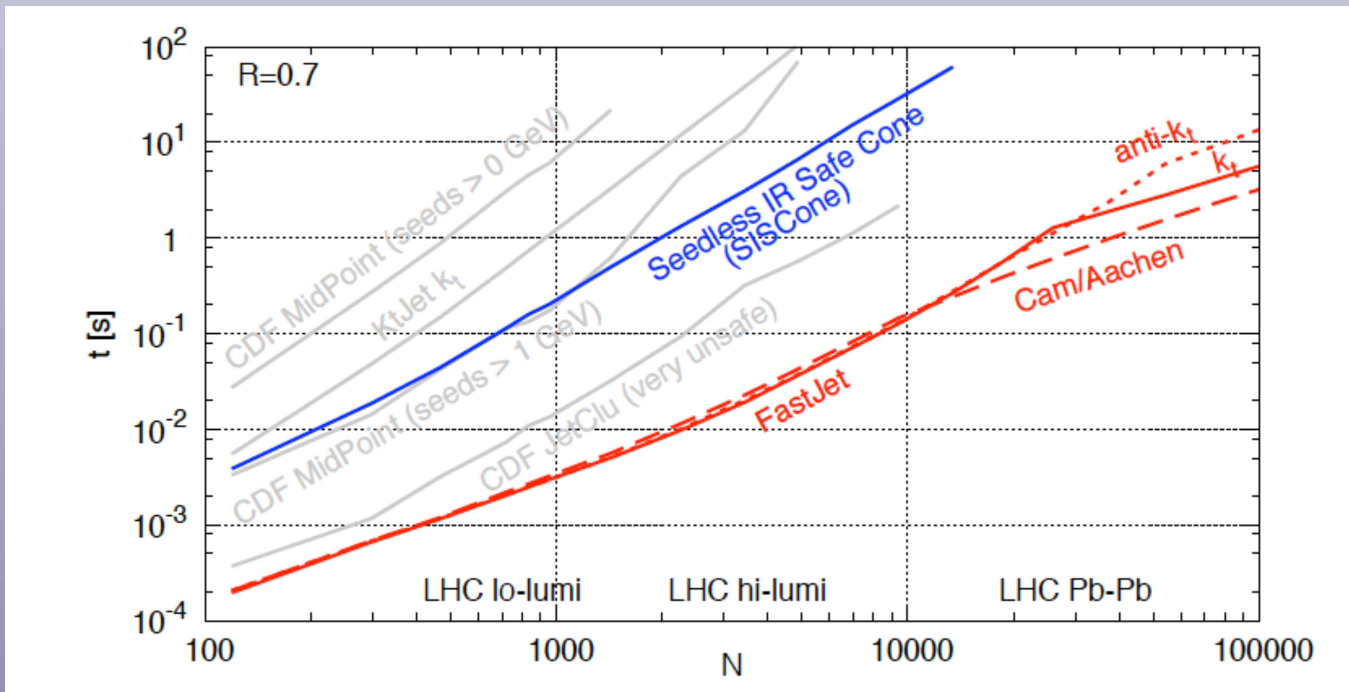
$$d_{ij} = \min(p_{ti}^{2p}, p_{tj}^{2p}) \frac{\Delta R_{ij}^2}{R^2}, \qquad \Delta R_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2,$$

$$d_{iB} = p_{ti}^{2p},$$

- **p = 1 ->** kt algorithm: follows QCD branching structure in pt and in angle
- **p = 0 -> Cambridge/Aachen:** follows QCD branching structure **only in angle**
- **p = -1 -> Anti-kT algorithm:** unrelated to QCD branching structure, with clustering measure favouring recombination of high-pT particles

By construction, these sequential recombination algorithms are **infrared safe**

At the LHC, the default jet algorithm is the **Anti-KT** algorithm, for reasons that we discuss now

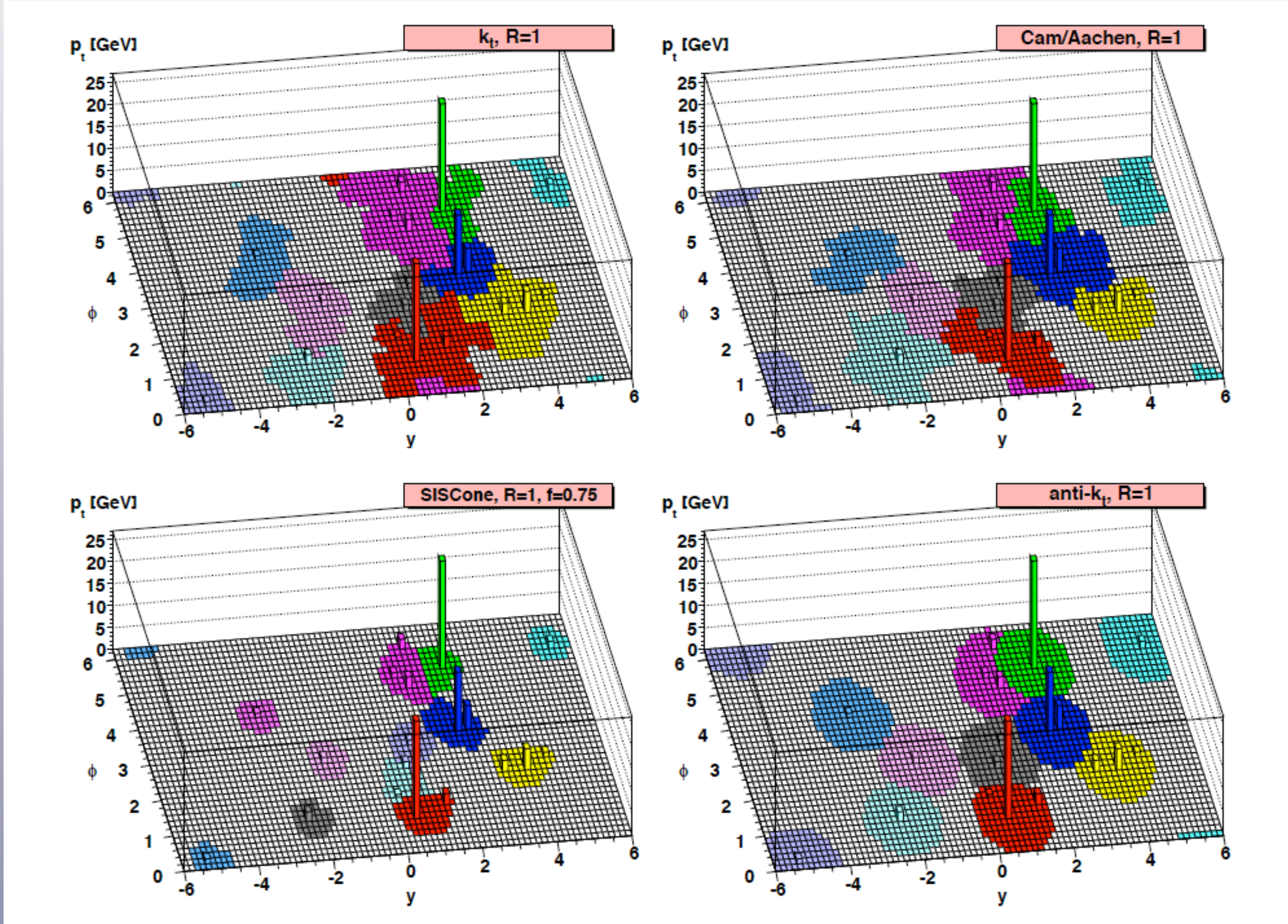


Original implementations of kt algorithm very **slow, T=O(N³)**, making it unpractical for high-multiplicity hadron collisions

Modern implementations (**FastJet**) much more efficient using computational geometry, amd achieve **T=O(N log N)**

Jet areas

The **catchment area of jets** is defined by adding a very large number of extremely soft particles to the event, called **ghost particles,** and measuring how many of then end up in a given jet



Jet areas

The **catchment area of jets** is defined by adding a very large number of extremely soft particles to the event, called **ghost particles,** and measuring how many of then end up in a given jet

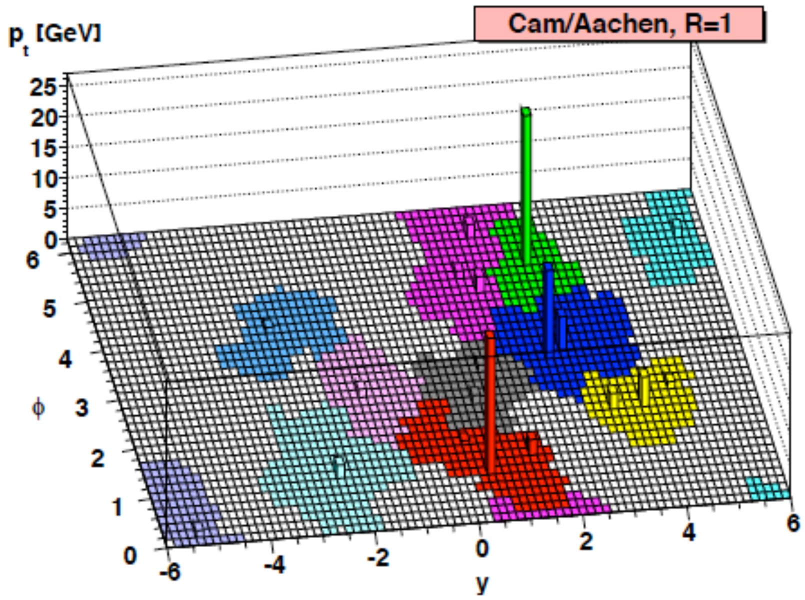
Jet areas in the **anti-kt algo** are much more regular than in **kt** and **C/A** algos

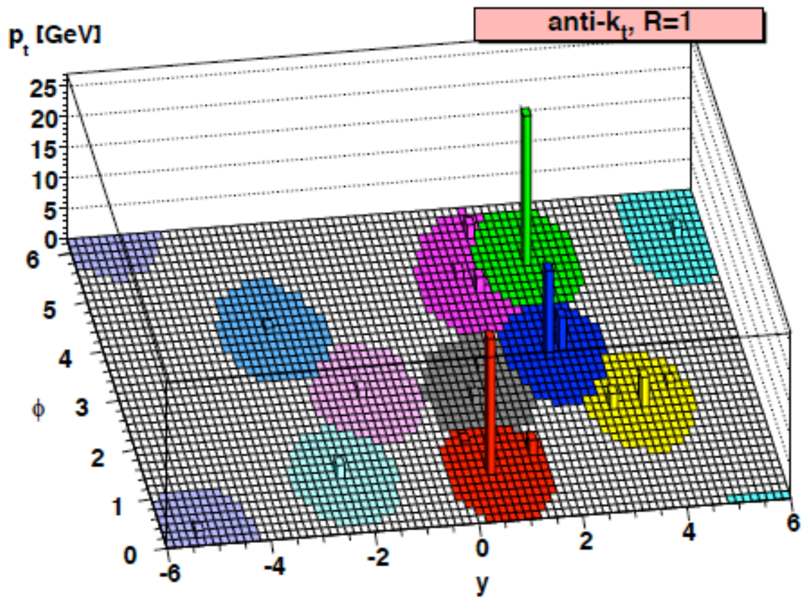
This is of course an essential advantage from the experimental point of view, both for **jet energy calibration** and subtraction of **underlying event** and **pileup**

This important property arises because **kt** and **C/A** first cluster soft and collinear branchings, and only at the end the high-pT particles, leading to the irregular shape

On the other hand, **anti-kt** clusters soft particles only at the end, leading to the regular, cone-like, geometry

In this sense, **anti-kt** is the perfect **cone jet algorithm**





Using jets

An essential parameter in all sequential recombination algorithms is the value of the **jet radius R**

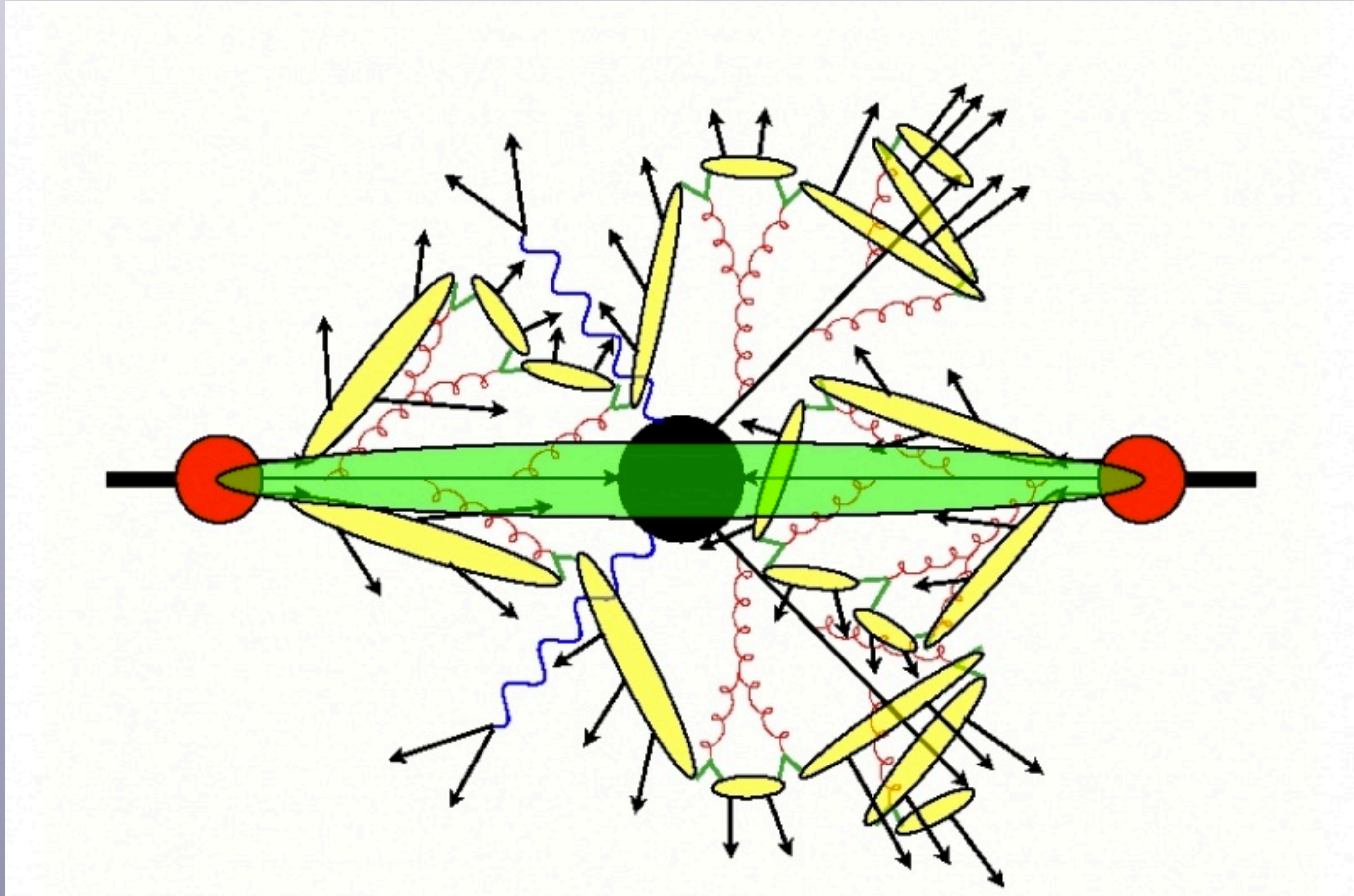
Its optimization depends on details of analysis, and needs to account for competition of different effects

1) **Underlying event** (multiple semi-hard interactions in the same pp collision) and **pile-up** (multiple semi-hard interactions from different pp collisions in the same **bunch crossing**)

UE and pileup are a source of **roughly constant background noise** that adds to the **jet pT**

$$\langle \delta p_t \rangle_{\mathrm{UE}} \simeq \Lambda_{\mathrm{UE}} R J_1(R) = \Lambda_{\mathrm{UE}} \left(\frac{R^2}{2} - \frac{R^4}{8} + \dots \right) .$$

So a large value of R
will pick up too much noise



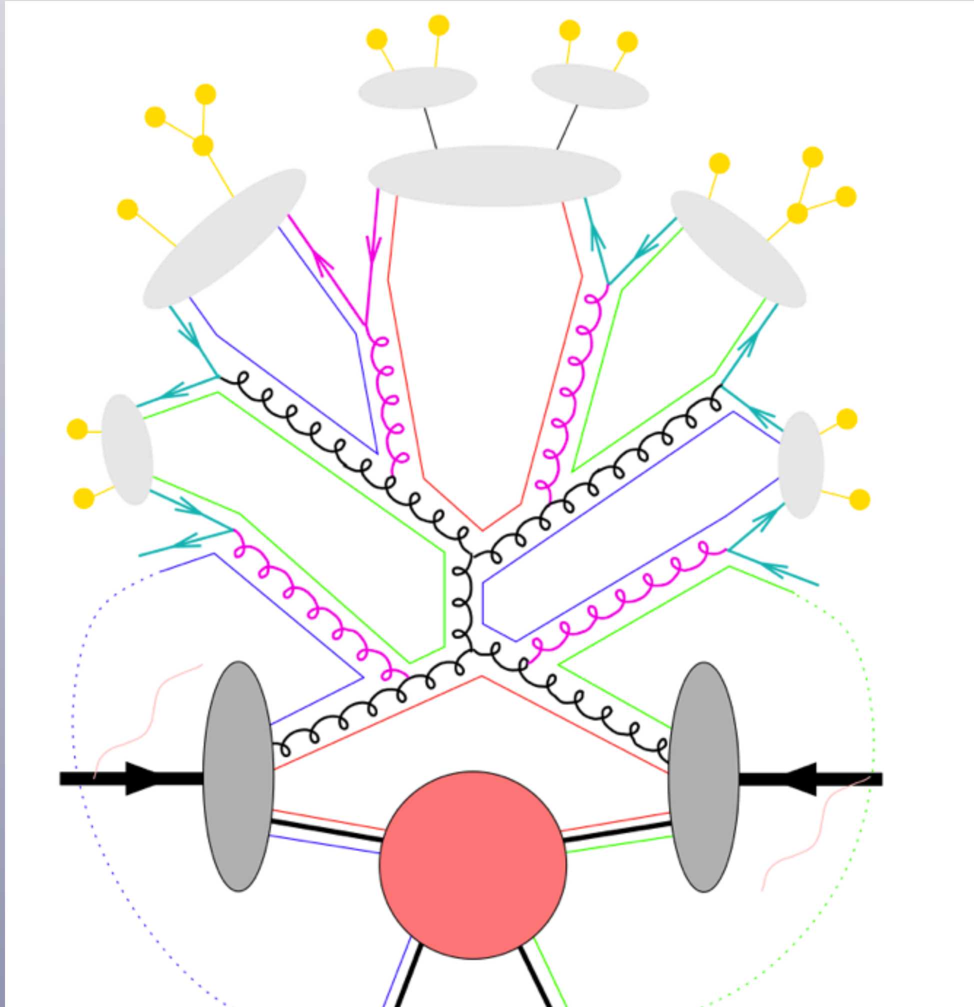
Using jets

An essential parameter in all sequential recombination algorithms is the value of the **jet radius R**

Its optimization depends on details of analysis, and needs to account for competition of different effects

2) **Hadronization:** the formation of **color-singlet hadrons** from colored partons is a **local effect,** and if **R** is too small then we find **large non-perturbative corrections**

$$\langle \delta p_t \rangle_{\mathrm{NP}}^{qq' \to qq'} = \frac{\Lambda}{\pi} \left[-\frac{2}{R} C_F + \frac{1}{8} R \left(5 C_F - \frac{9}{N_c} \right) + \mathcal{O}\left(R^2\right) \right]$$



Using jets

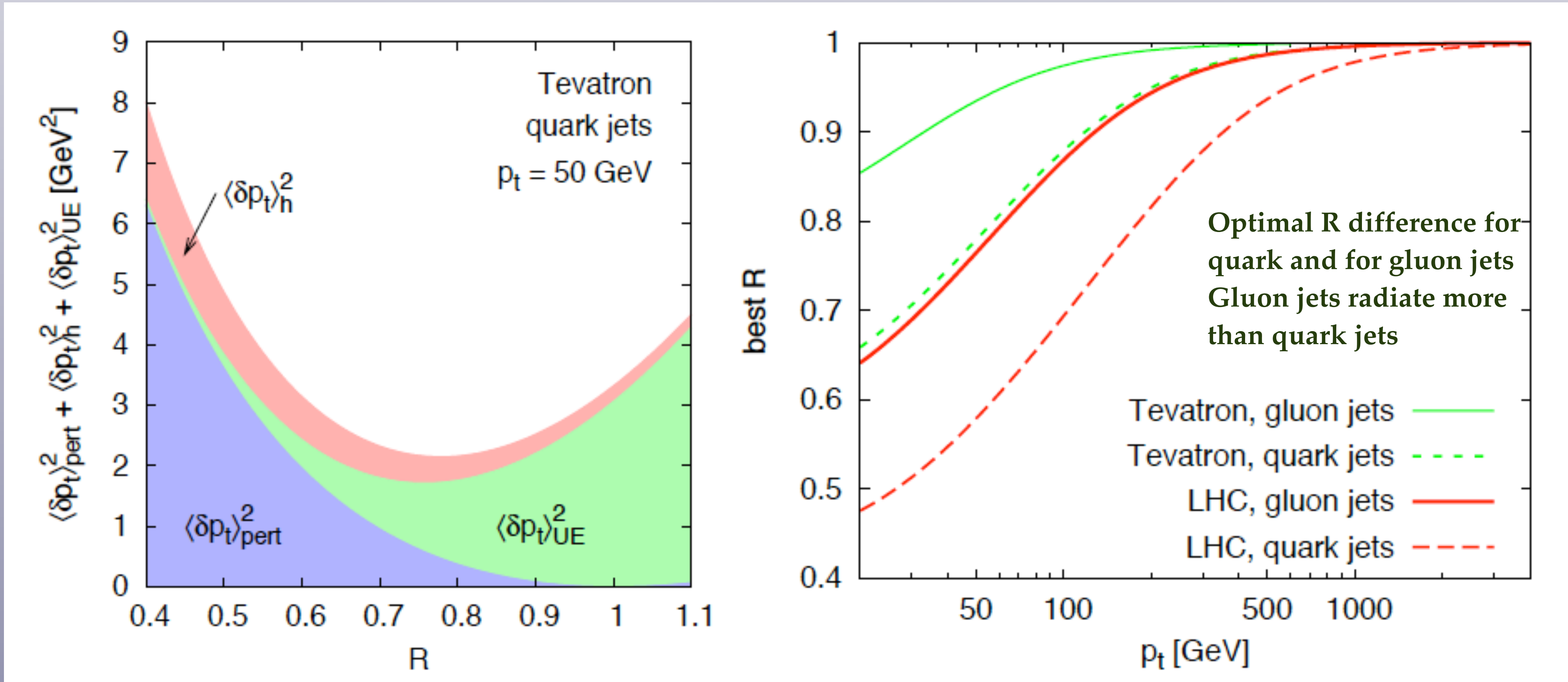
An essential parameter in all sequential recombination algorithms is the value of the **jet radius R**

Its optimization depends on details of analysis, and needs to account for competition of different effects

3) **Out-of-cone perturbative radiation:** if the jet radius is too small, most of the **pt of the original parton** will be left out of the **jet pt because of QCD radiation,** leading to an undesired mismatch between partons and hadrons

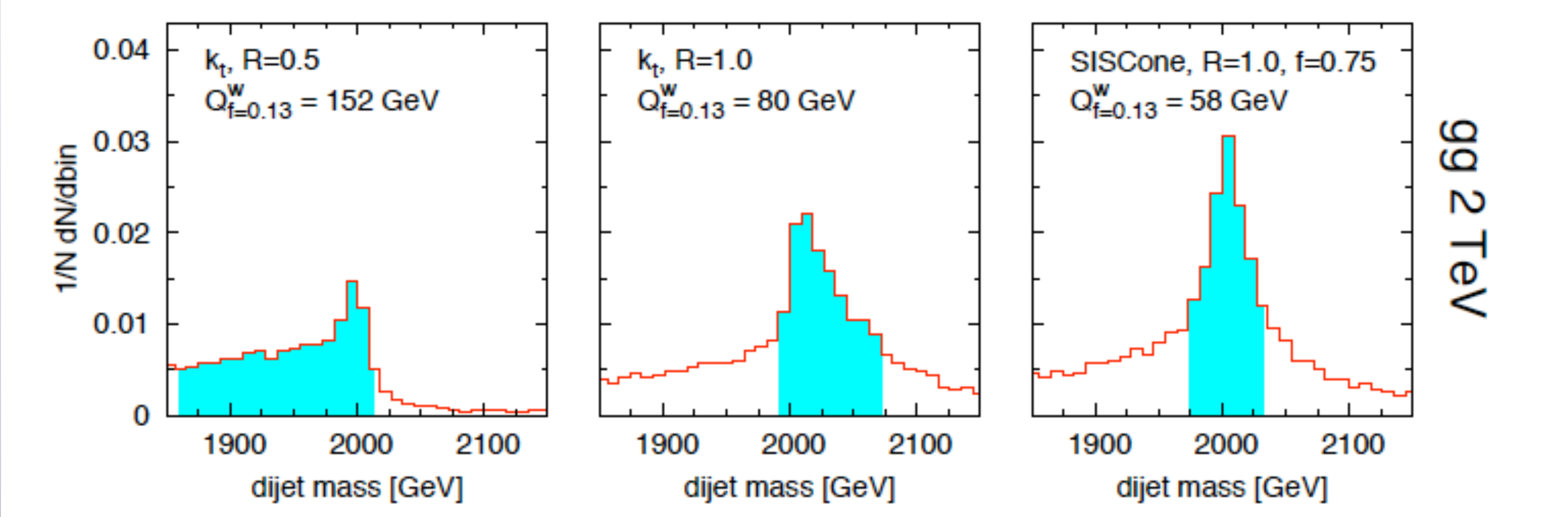
$$\frac{\langle \delta p_t \rangle_{\text{pert}}}{p_t} = \frac{\alpha_s}{\pi} L_i \ln R + \mathcal{O}(\alpha_s),$$

By simultaneously minimizing the three contributions, we can determine **optimal value of R**



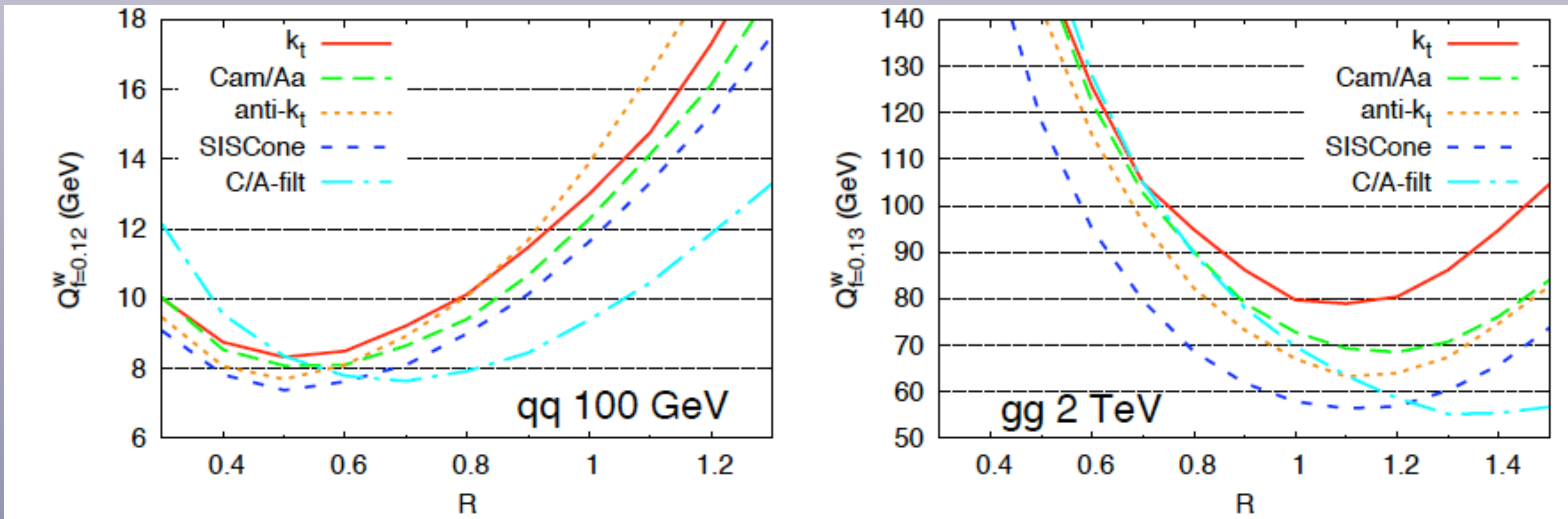
Jet quality

It is also possible to quantify the performance of **different jet algorithms** with **different jet radius** for specific analysis, for example for the kinematic reconstructions of the mass of a heavy resonance



Anti-kt and SIScone outperform kt and Cambridge/Aachen in most cases

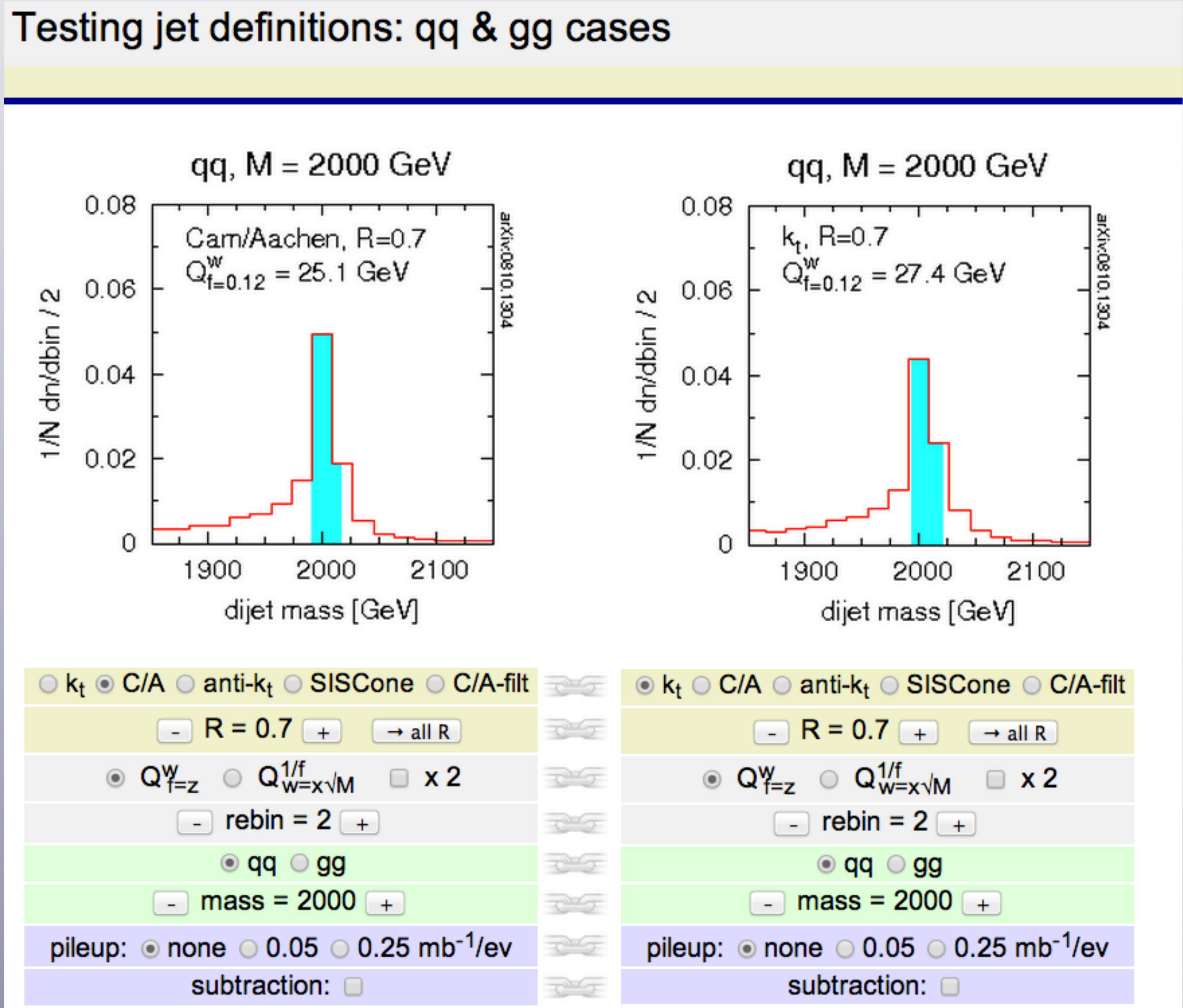
At the LHC, for **TeV-scale particles** decaying into gluons, **larger R** would be beneficial



Jet quality

Let's try live to find the **optimal jet definition for kinematic reconstructions at LHC**

http://www.lpthe.jussieu.fr/~salam/jet-quality/



Test also impact of **UE/ pileup subtraction**

Study how **different masses** require **different jet radius**

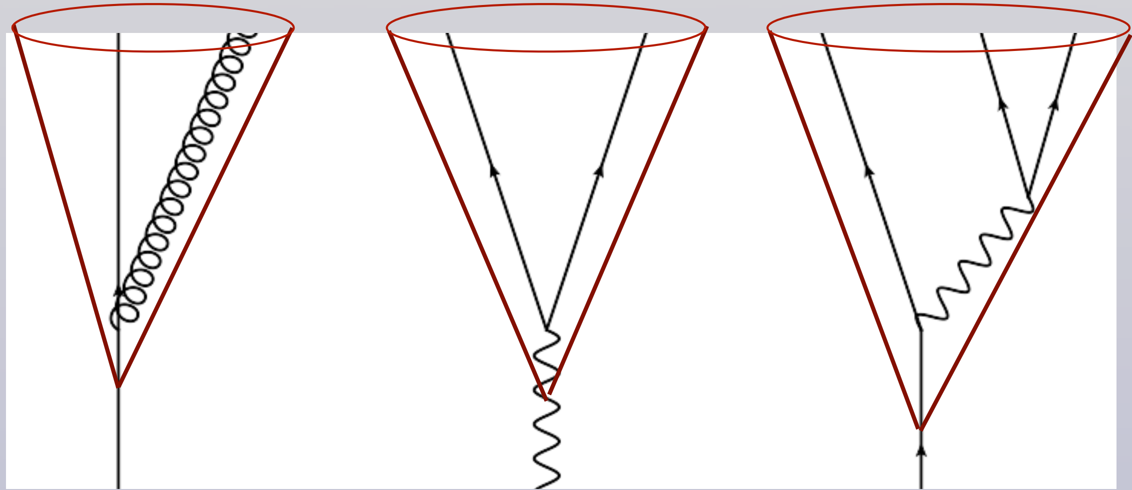
Quantify the differences between **quark jets** and **gluon jets**

Jet Substructure

Basic reference:
"Jet substructure as a new Higgs search channel at the LHC", arxiv:0802.2470

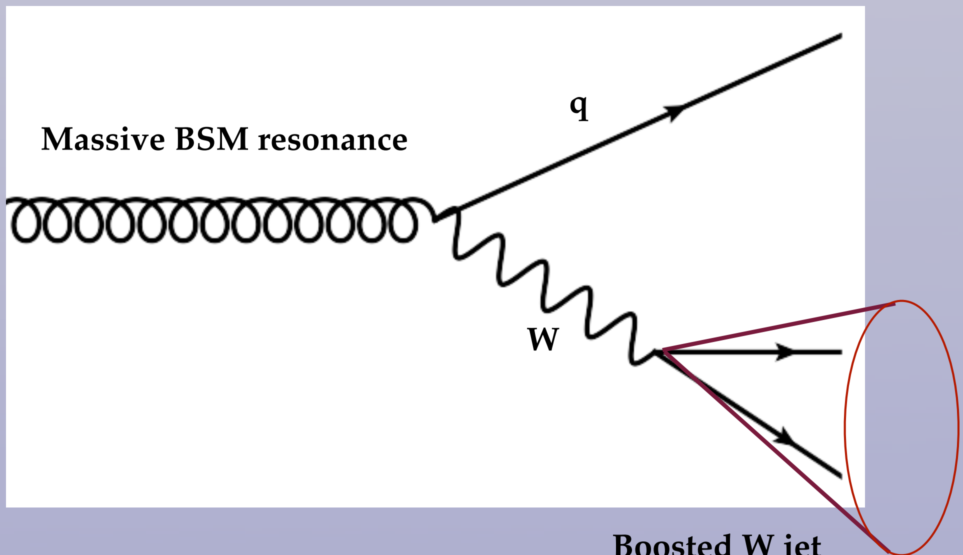
Jet substructure

- In the decays of a **massive enough resonances, boosted prongs** can often be collimated into a **single jet**
- Different **jet substructure** in these jets and QCD jets provide strong background suppression in **BSM searches**



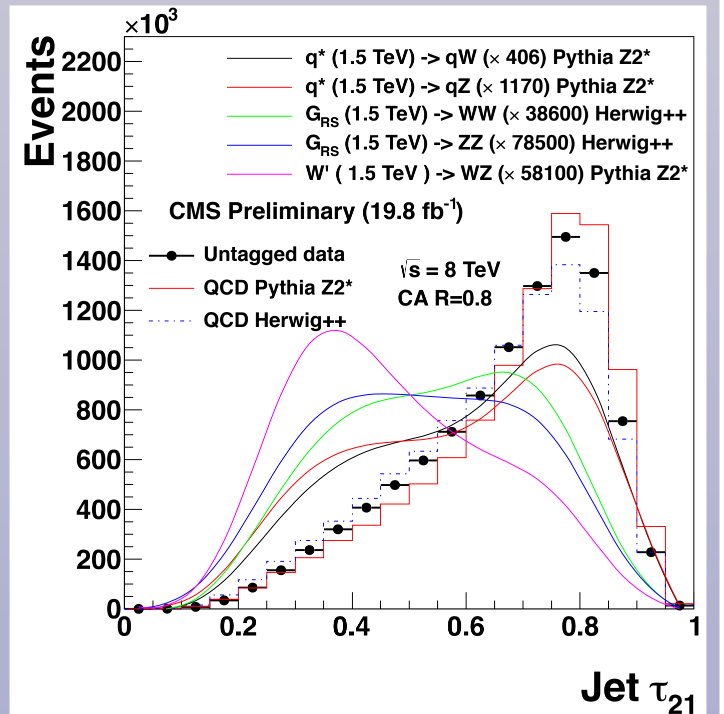
QCD jet **Boosted W jet** **Boosted top quark jet**

- LHC analysis are using more and more jet substructure techniques (also for Higgs)
- As illustration, recent CMS search for **q* -> qV in the tagged dijet final state**
- Discriminating variable: **different jet shape/mass in signal and in QCD background**



Boosted W jet

CMS PAS-EXO-12-024



Jet substructure

- To understand under which **kinematics** jet substructure is needed, let's consider the decay of a **Higgs boson** into a **b-quark pair**
- For simplicity, we assume that the **Higgs** has been produced **centrally**. Then, taking into account the general parametrization of four-momenta in hadronic collisions, we have

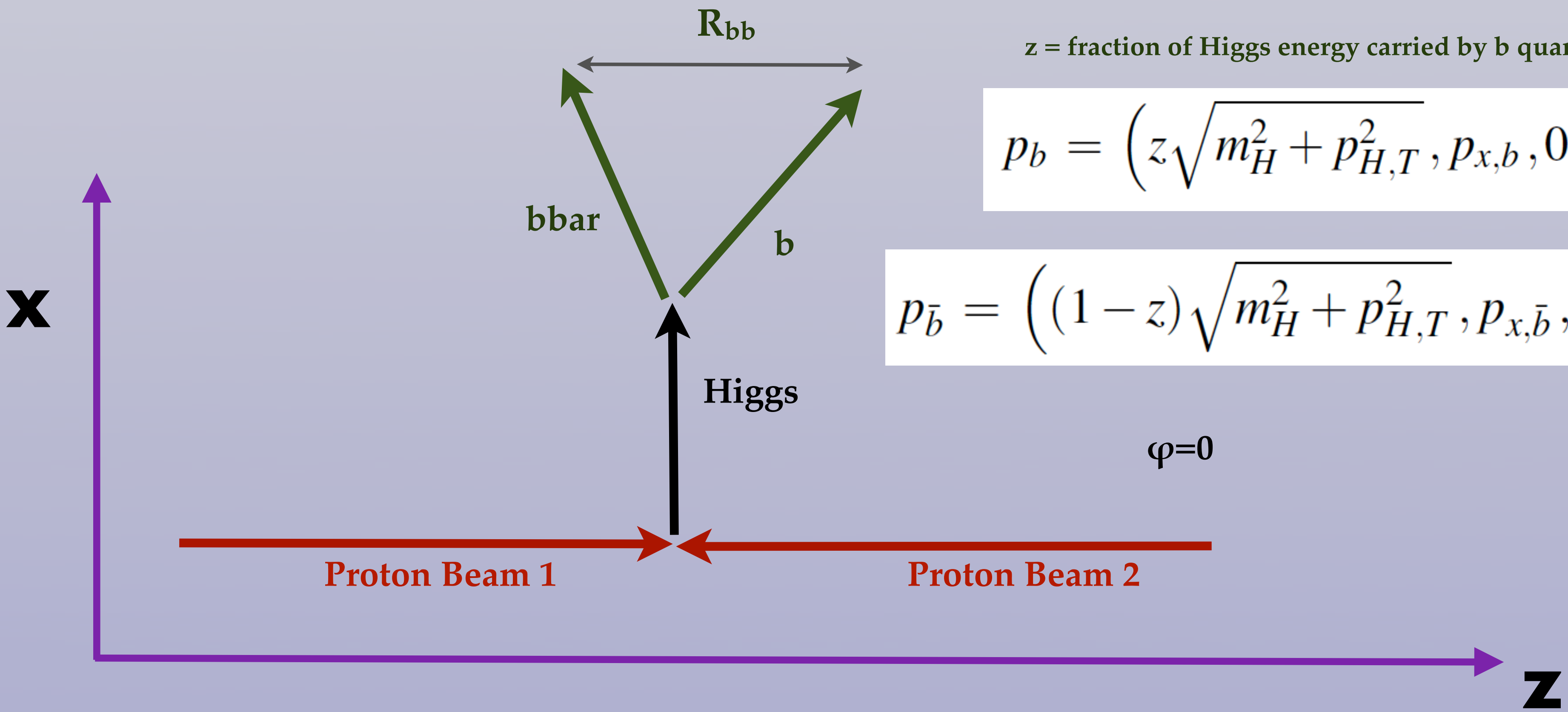
$$p = (E, p_x, p_y, p_z) = (m_T \cosh y, |p_T| \cos\phi, |p_T| \sin\phi, m_T \sinh y)$$

$$p_H = \left(\sqrt{m_H^2 + p_{H,T}^2}\, , p_{T,H}\, , 0 , 0\right)$$

z = fraction of Higgs energy carried by b quark

$$p_b = \left(z\sqrt{m_H^2 + p_{H,T}^2}\, , p_{x,b}\, , 0 , p_{z,b}\right)$$

$$p_{\bar{b}} = \left((1-z)\sqrt{m_H^2 + p_{H,T}^2}\, , p_{x,\bar{b}}\, , 0 , -p_{z,b}\right)$$



Jet substructure

- Now, using **four-momentum conservation** and neglecting the masses of the bottom quarks, we find the following expressions for the bottom quarks momenta

$$
\begin{aligned}
p_{x,b} &= \frac{p_T}{2} + (2z-1)\,\frac{p_{T,H}^2 + m_H^2}{2\,p_T} \\
p_{x,\bar{b}} &= \frac{p_T}{2} - (2z-1)\,\frac{p_{T,H}^2 + m_H^2}{2\,p_T} \\
p_{z,b} &= \left[(p_{T,H}^2 + m_H^2)\left(z^2 - z + \frac{1}{2}\right) - \frac{p_T^2}{4} - (2z-1)^2\,\frac{(p_{T,H}^2 + m_H^2)^2}{4\,p_T^2} \right]^{1/2}
\end{aligned}
$$

- We want now to consider the **boosted limit,** defined as the regime in which the **transverse momentum** of the heavy particle that is decaying is **much larger than its mass.** In this limit the above formulae **simplify** to

$$
p_{T,H} \gg m_H \quad \longrightarrow \quad
\begin{aligned}
p_{x,b} &= z\,p_{T,H} \\
p_{x,\bar{b}} &= (1-z)\,p_{T,H} \\
p_{z,b} &= \sqrt{z(1-z)}\,m_H
\end{aligned}
$$

Jet substructure

- So in the **boosted limit**, the expression for the **angular separation** between the two b quarks is

$$\begin{aligned} p_{x,b} &= z\, p_{T,H} \\ p_{x,\bar{b}} &= (1-z)\, p_{T,H} \\ p_{z,b} &= \sqrt{z(1-z)}\, m_H \end{aligned} \longrightarrow R_{b\bar{b}} = \frac{p_{z,b}}{p_{x,b}} - \frac{p_{z,\bar{b}}}{p_{x,\bar{b}}} = \frac{1}{\sqrt{z(1-z)}} \frac{m_H}{p_{T,H}}$$

Therefore, if $\mathbf{R_{bb} < R}$, **the jet radius**, the two bottom quarks will be part of a **single hadronic jet,** and therefore, the **traditional identification** of the Higgs boson from the **sum of two b-tagget jets** becomes impossible

For $\mathbf{R = 0.5}$, this relation implies that events with $\mathbf{p_{T,H} > 500}$ **GeV** will be **lost** in traditional analysis

This is not very good, since high $\mathbf{p_{T,H}}$ events are also those where the Standard Model backgrounds are smaller

- To be able to sort out this problem, a number of **jet substructure techniques** have been developed in order to **efficiently distinguish jets with non-trivial substructure** from the usual QCD jets
- Basic ideas are common
  - QCD jets are likely to have generated from **soft/collinear splittings**, while in the decay of heavy resonances all prongs share similar fractions of energy
  - For QCD jets, the **jet mass** does not show any structure, while for heavy resonance decays it should be peaked around the resonance mass
  - Removing **soft radiation** should leave signal events unchanged while decreasing the contamination of QCD events

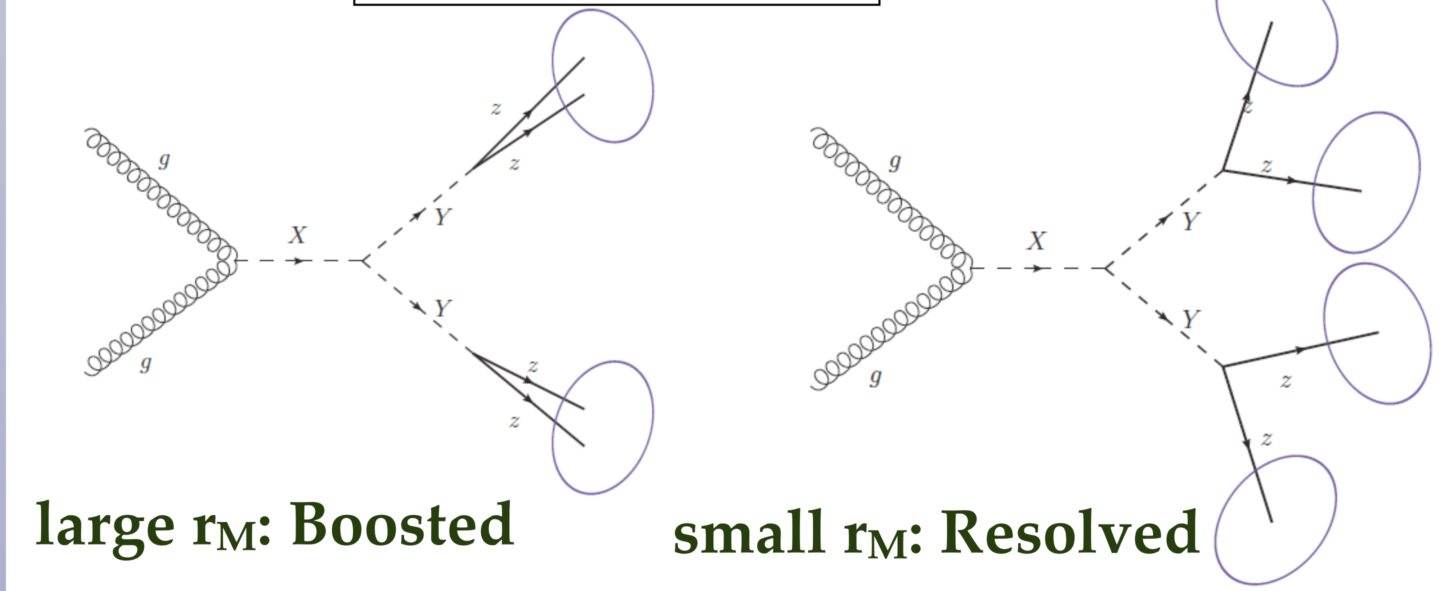
Jet substructure

- Many BSM scenarios involve **resonant pair production** of heavy (SM and BSM) particles
- In the spirit of **Simplified Models,** we assume that the underlying process is

$$\boxed{pp \to X \to 2Y \to 4z}$$

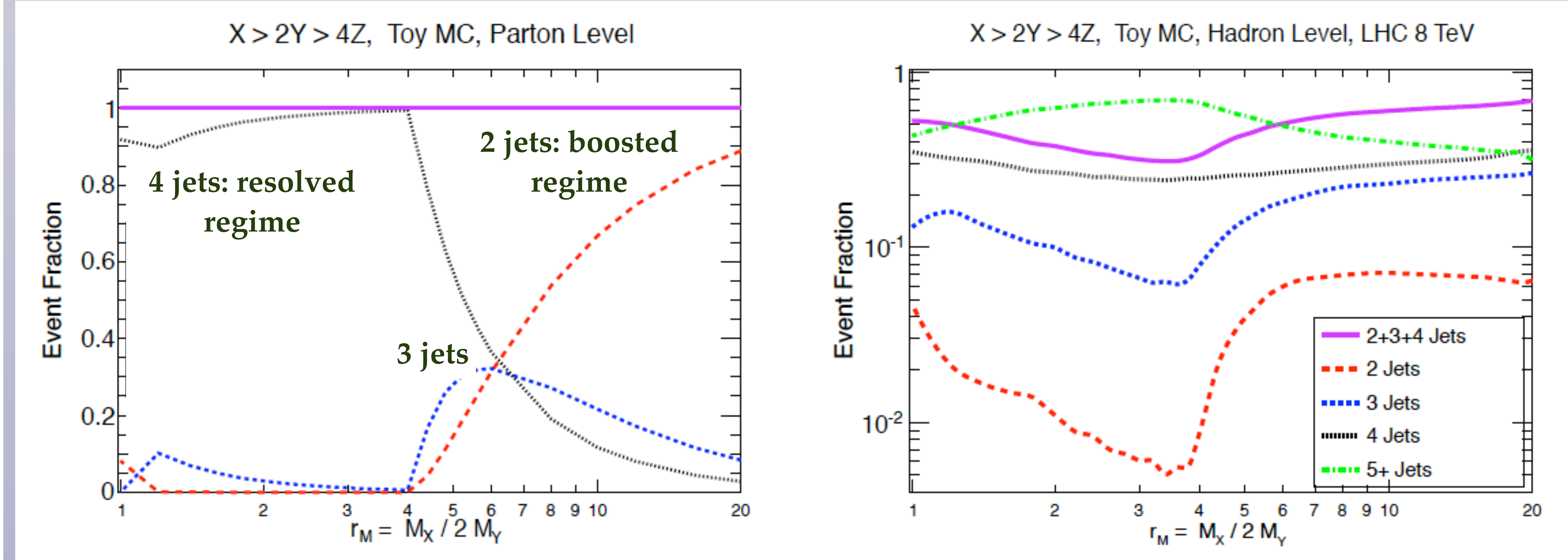
- Implemented in a **toy parton-level Monte Carlo** event generator
- Depending on the value of the mass ratio **$r_M = M_X/2M_Y$** different final state topologies
- For large **r_M** the intermediate heavy particles **Y** will be **highly boosted,** and thus their decay products **z** will be close in the detector
- For small **r_M** the **Y** particles are produced close to rest, and the four decay particles **z** are well separated in the detector (**resolved regime**)

$$\boxed{pp \to X \to 2Y \to 4z}$$



Jet substructure

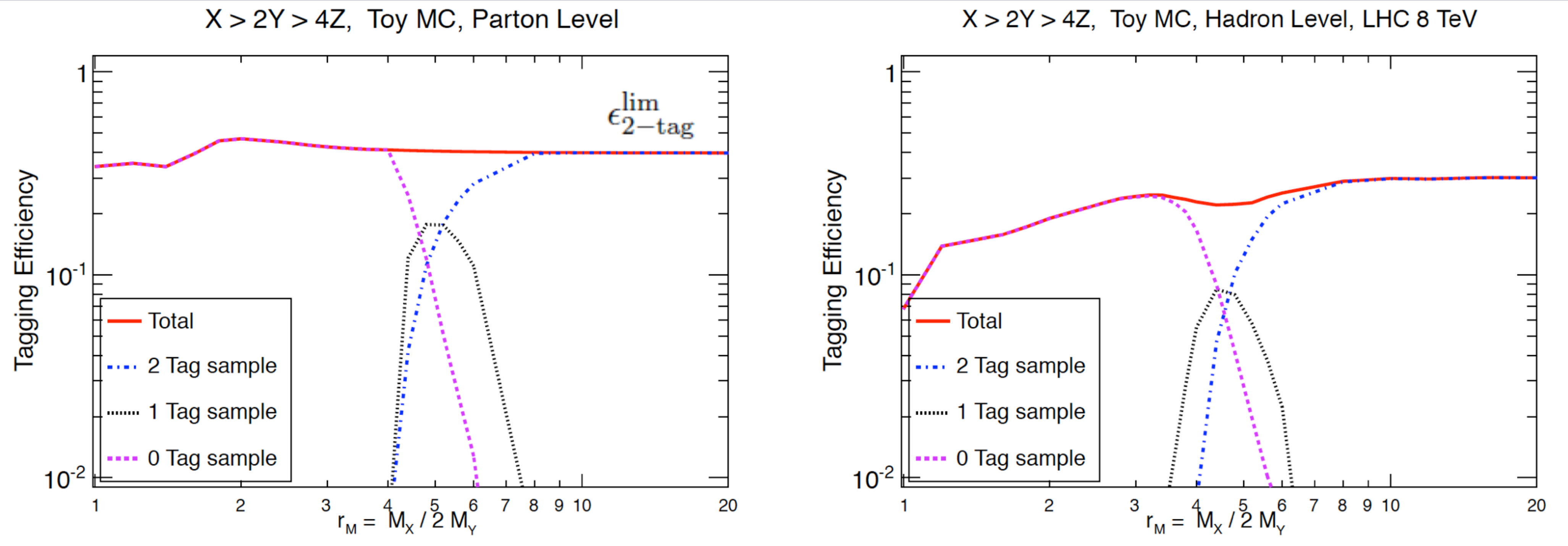
- At **parton level**, without cuts, the classification of the event topology (boosted, resolved or intermediate) is trivial **based on the number of jets**
- But at **hadron level** with **realistic cuts** such naive classification is not feasible
- A more robust event classification achieved based on the number of **jet substructure tags**



Juan Rojo

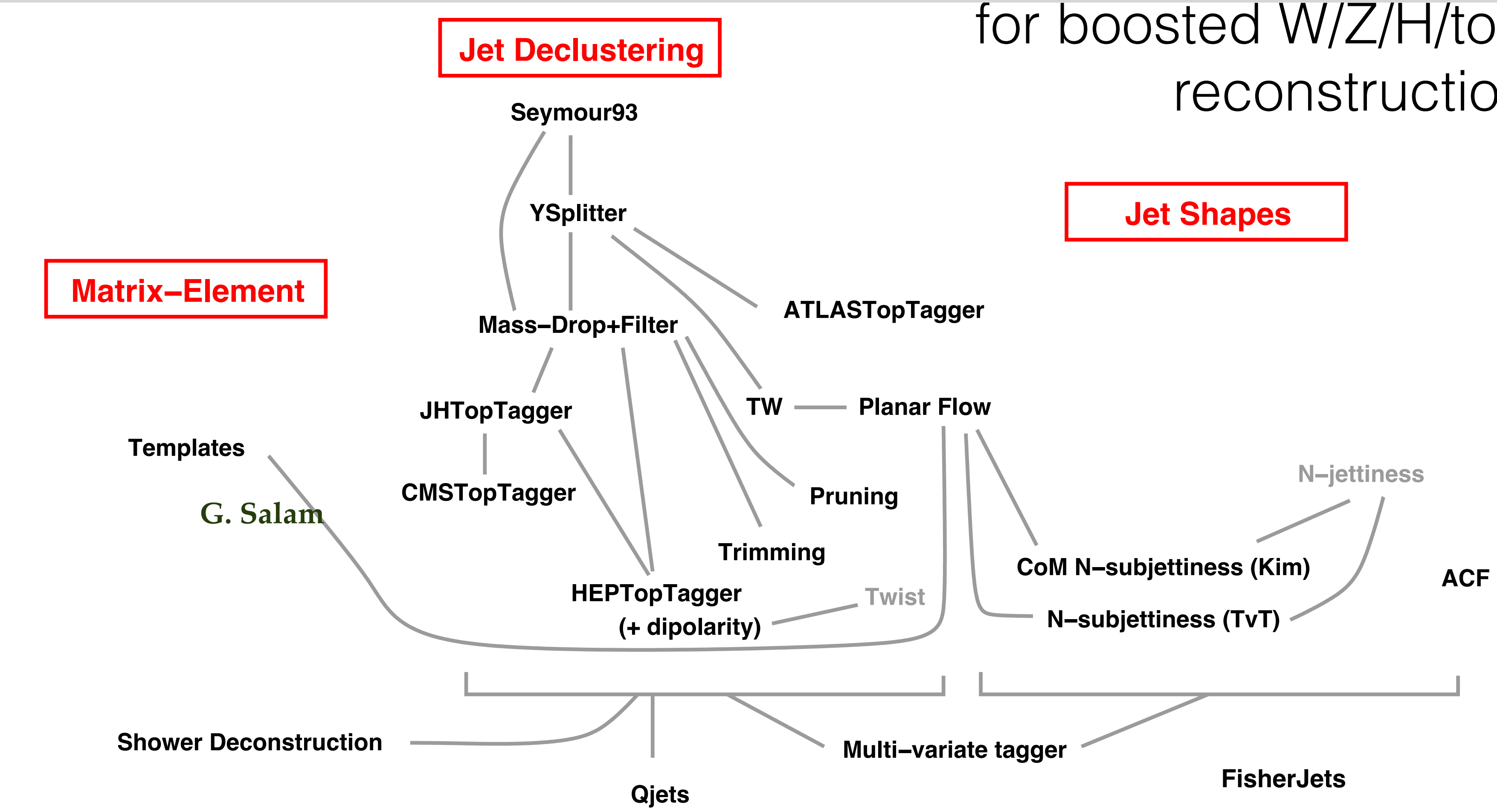
Jet substructure

- Tagging efficiency **independent of the value of the mass ratio** (except hadron level small $\mathbf{r_M}$)
- **Smooth interpolation** between the boosted and resolved regimes is possible



- A search strategy based only on the **traditional, resolved jet configuration** would fail dramatically for heavy resonances
- However, note also that a purely boosted search also fails to explore the low and medium mass region of the BSM resonance

Jet substructure techniques



A **plethora of different methods** have been developed, and some of them are implemented in the **ATLAS and CMS** analysis software

Impossible to cover all of them, here concentrate in one of the most widely used: the **mass-drop tagger**

The BDRS mass-drop tagger

This substructure tagger is based on the **Cambridge-Aachen algorithm** with radius **R**

- First of all, calculate the **angular distance** between all objets that enter jet clustering

$$\Delta R_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$$

- **Recombine** the two objects with the **smallest angular distance**
- Recompute the angular distances, and iterate until $\mathbf{\Delta R_{ij} > R}$: the remaining objects are classified as **jets**

Once we have the **hard jets,** we can take each of these, **j**, and explore the possibility that it comes from the decay of some heavy particle:

- Break the jets into two **subjets, $\mathbf{j_1}$** and **$\mathbf{j_2}$,** reverting the last stage of the clustering, with $\mathbf{m_{j1} > m_{j2}}$
- We first require a **substantial mass drop**: the subjet masses are much smaller than the original jet mass

$$m_{j1} \le \mu\, m_j$$

- Then we require that the splitting is not too asymmetric

$$y_{\rm mdt} \equiv \frac{\min(p_{T,j1}, p_{T,j2})}{m_j^2} \Delta R_{j1,j2j}^2 \ge y_{\rm cut}$$

- If the two conditions are satisfied, the jet **j** is classified as composed by the **decay products of some heavy resonance**, else define **$\mathbf{j = j_1}$** and continue until there are no more subjets

The BDRS mass-drop tagger

Let us try to understand the **physical motivation** for the BDRS mass-drop tagger, using what we know on the **structure of QCD parton branchings**

BDRS requires a **substantial mass drop**

$$m_{j1} \le \mu \, m_j$$

Now, QCD jets, generated by the parton branching process, are characterized that the fact that with high likelihood branchings would have been either **soft or collinear,** due to structure of QCD matrix element

$$d\sigma_{n+1} = d\sigma_n \frac{dt}{t} dz \frac{\alpha_s}{2\pi} \hat{P}_{ab}(z)$$

On the other hand, the **matrix element for the decay** of heavy resonances, say the Higgs, does not contain such infrared divergences

Applying the formulae of the QCD parton branching to the case of QCD jets we can write

$$m_j^2 = m_{j1}^2 + m_{j2}^2 + 2z(1-z)E_j^2\,\theta_{j1,j2}^2$$

which now, neglecting the mass of the softest subjet, we can write

$$m_{j1}^2 \simeq m_j^2 \left(1 - \frac{2E_j^2}{m_j^2}\theta_{j1,j2}^2\right) \le \mu \, m_j^2$$

Therefore, the **BDRS cut in μ disfavors collinear splittings**, characteristic of QCD branching not absent in heavy resonance decays

The BDRS mass-drop tagger

Let us try to understand the **physical motivation** for the BDRS mass-drop tagger, using what we know on the **structure of QCD parton branchings**

BDRS also requires that the **splitting is not too symmetric**

$$y_{\rm mdt} \equiv \frac{\min(p_{T,j1}, p_{T,j2})}{m_j^2} \Delta R^2_{j1,j2j} \geq y_{\rm cut}$$

Using the **formulae for the boosted regime** that we have derived previously, it is easy to see that

$$y_{\rm mdt} \equiv \frac{\min(p_{T,j1}, p_{T,j2})}{m_j^2} \Delta R^2_{j1,j2j} \simeq \frac{\min\left(z_{j1}^2, z_{j2}^2\right)}{z_{j1}, z_{j2}} = \frac{\min\left(z_{j1}, z_{j2}\right)}{\max\left(z_{j1}, z_{j2}\right)} \geq y_{\rm cut}$$

Therefore, this cut penalizes kinematical configurations where the **branching is too asymmetric**.

This is the case of QCD, where due to the **soft singularities in the splitting function** typically lead to

$$z_{j1} \gg z_{j2} \quad \text{or} \quad z_{j1} \ll z_{j2}$$

As opposed to the **decays of a heavy resonance** into two prongs of similar mass, where we have

$$z_{j1} \sim z_{j2}$$

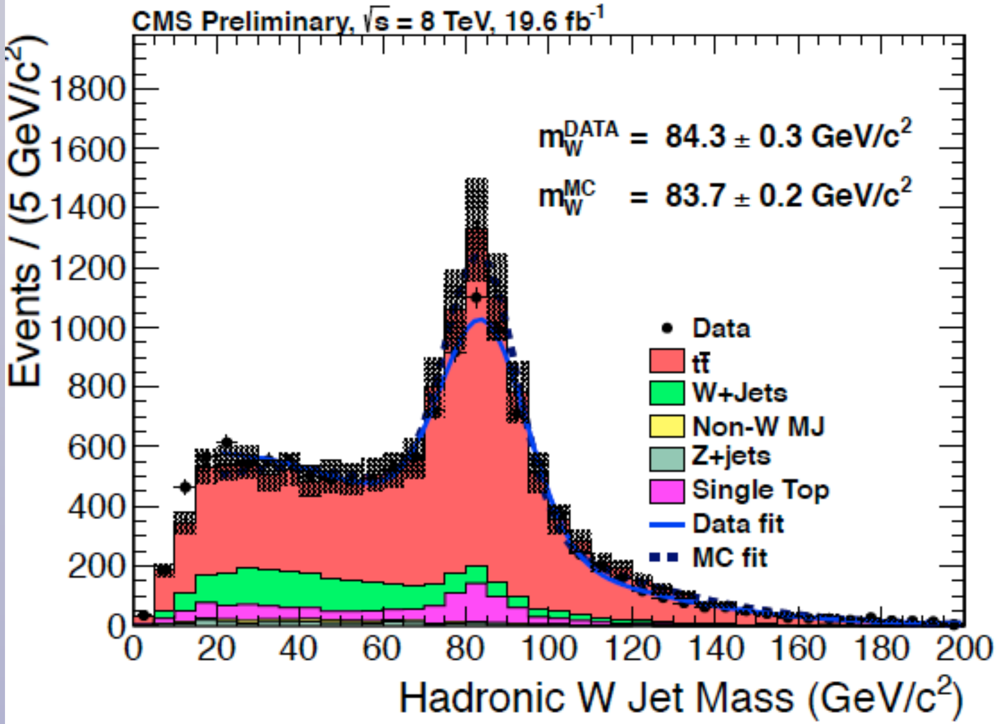
So **asymmetry cut** in BDRS tagger is efficient to **suppress QCD splittings** as compared to heavy resonance decays

Jet substructure at LHC

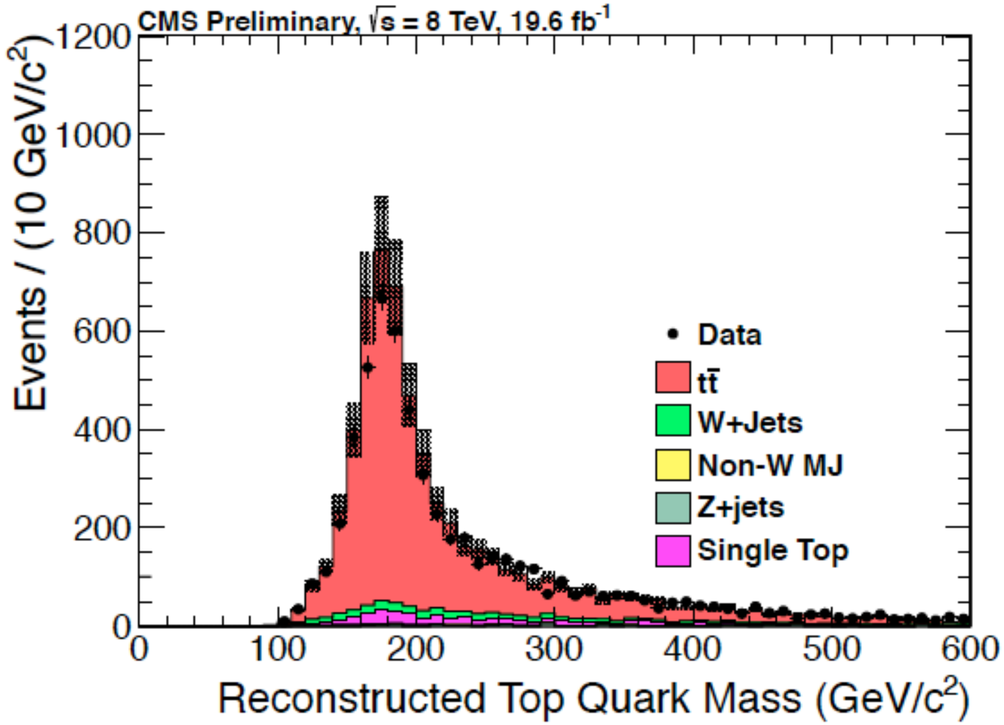
Some LHC analysis have already implemented **jet substructure tools**, and even more will do so in the Run II where the increased energies will enhance the importance of the **boosted topologies**

Hadronic boosted top decays, t -> W(-> qq')b, with decay products of W in same hadronic jet

W-tagged jet mass:



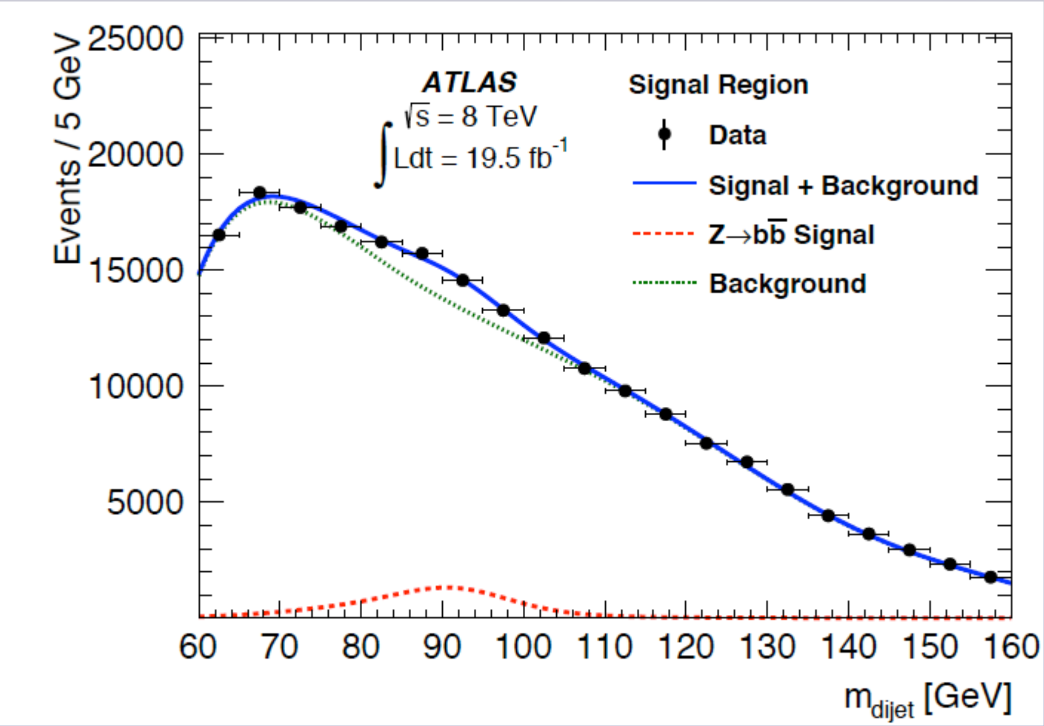
W-jet combined with a b-jet:

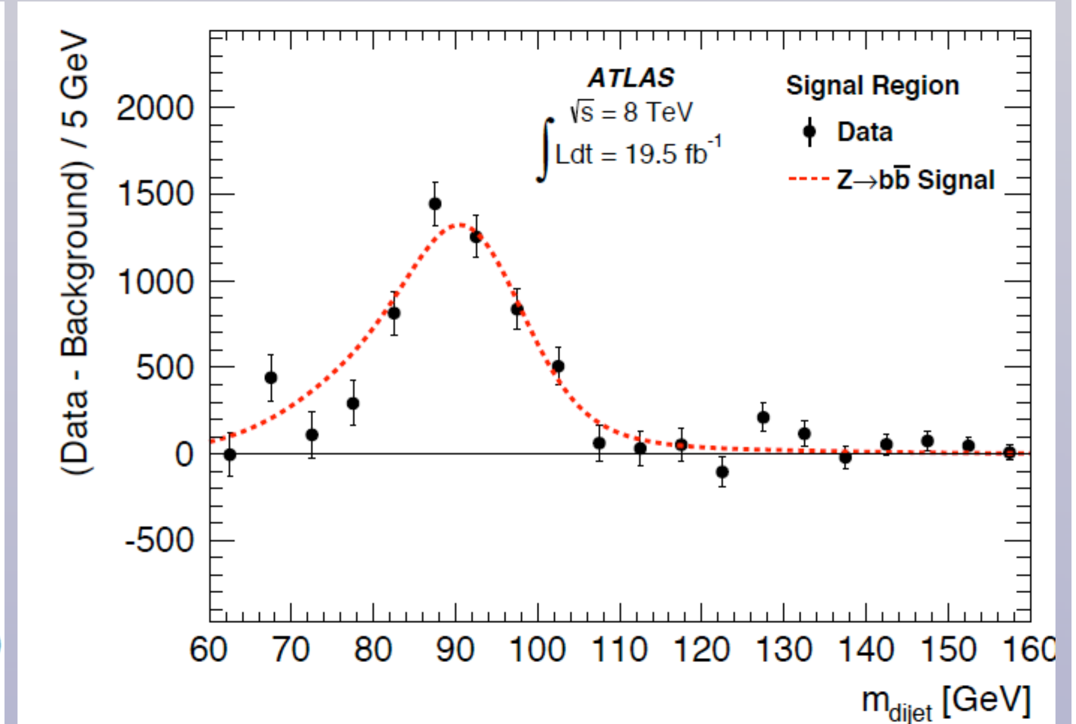


Jet substructure at LHC

Some LHC analysis have already implemented **jet substructure tools**, and even more will do so in the Run II where the increased energies will enhance the importance of the **boosted topologies**

Observation of the Z -> bb decays in high transverse momentum events



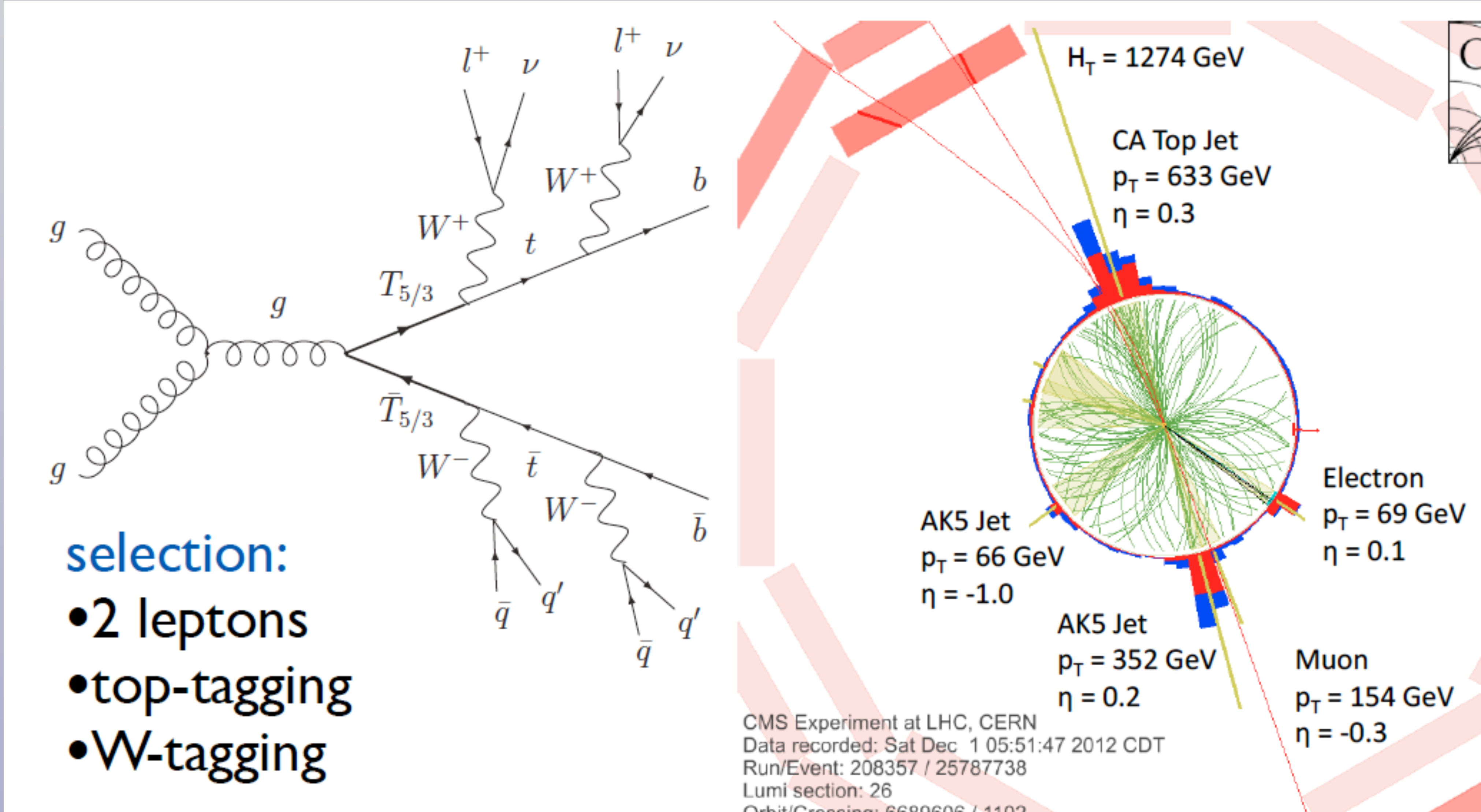


Jet substructure techniques essential to **increase S/B** in boosted topologies

Jet substructure at LHC

Some LHC analysis have already implemented **jet substructure tools**, and even more will do so in the Run II where the increased energies will enhance the importance of the **boosted topologies**

Search for fermionic top partners in boosted topologies

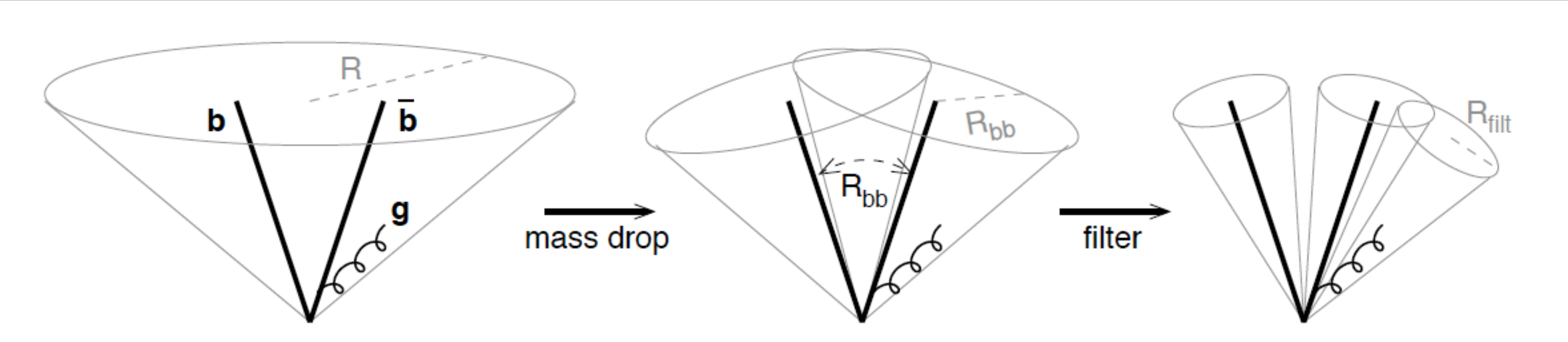


Jet substructure and pile-up removal

Jet substructure is also very useful in the context of **soft background** removal in hard jets

The basic idea is, when **undoing the jet clustering,** remove the **too soft subjets** that are likely to arise from the Underlying Event / Pile-Up rather than from the hard event

A well-known example of these techniques is **jet filtering**



Once a mass-drop has been identified, keep only the three hardest subjets of the jet with **radius Rfilt,** and discard all the rest of the event

Jet reconstruction and substructure

In this lecture we studied the definition of **QCD jets at hadron colliders,** and the role played by **jet substructure techniques** in boosted topologies:

☑ Modern jet reconstruction techniques are **infrared safe** to all orders, and fast efficient numerical implementations are available

☑ The **anti-kt algorithm** is the default choice of LHC experiments: combines the nice properties of sequential recombination algorithms with a regular, cone-like geometry

☑ Various effects contaminate the i**nterpretation of the transverse momentum of the jet** with that or the original parton: hadronization, underlying event, pile-up, out-of-cone radiation

☑ Jet substructure is a set of techniques to distinguish **hadronic decays of heavy resonances** (SM or BSM) as compared to QCD backgrounds

☑ **Boosted techniques** allow to substantially enhance S/B in high-mass regions which are most sensitive to BSM searches

Next we move back from the final state to the initial state, and review the procedure with which we can determine the **Parton Distribution Functions** of the proton