Jets and their structure

Simone Marzani
Università di Genova &
INFN Sezione di Genova

INFN

QCD Masterclass 2019

Lecture plan

* lecture 1: jets and jet algorithms
* lecture 2: calculating jet properties
* lecture 3: jet substructure
* lecture 4: more advanced topics & curiosities

resources

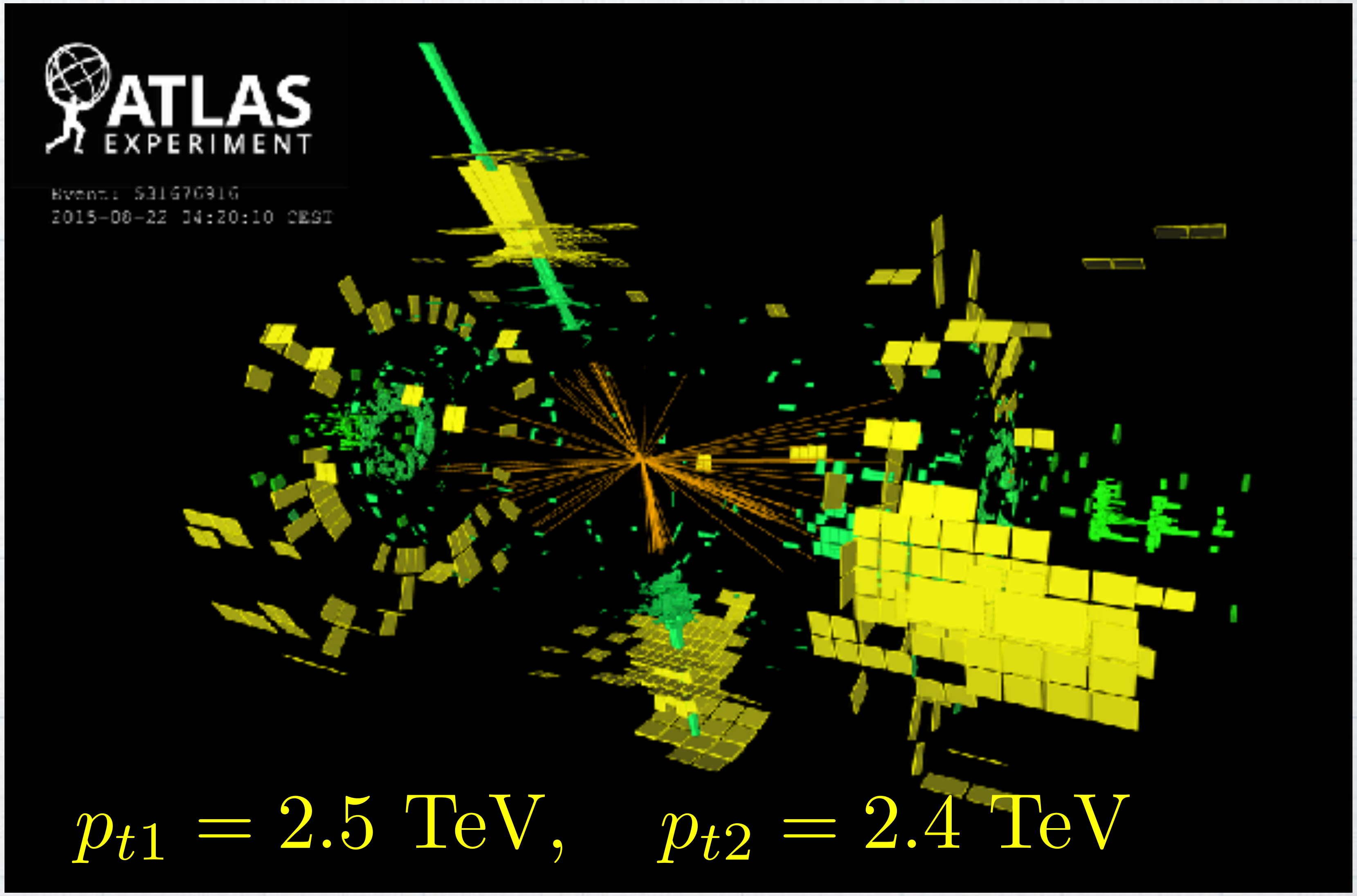
* SM, M. Spannowsky, G. Soyez, "Looking inside jets: an introduction to jet substructure and boosted-object phenomenology"
* the BOOST report series
* Les Houches reports 2015 & 2017
* G. Salam: "Towards jetography"
* G. Soyez: "Pileup mitigation at the LHC: a theorist's view"
* Gras et al. "Systematics of quark/gluon tagging"

Lecture plan

* lecture 1: jets and jet algorithms
* lecture 2: calculating jet properties
* lecture 3: jet substructure
* lecture 4: more advanced topics & curiosities

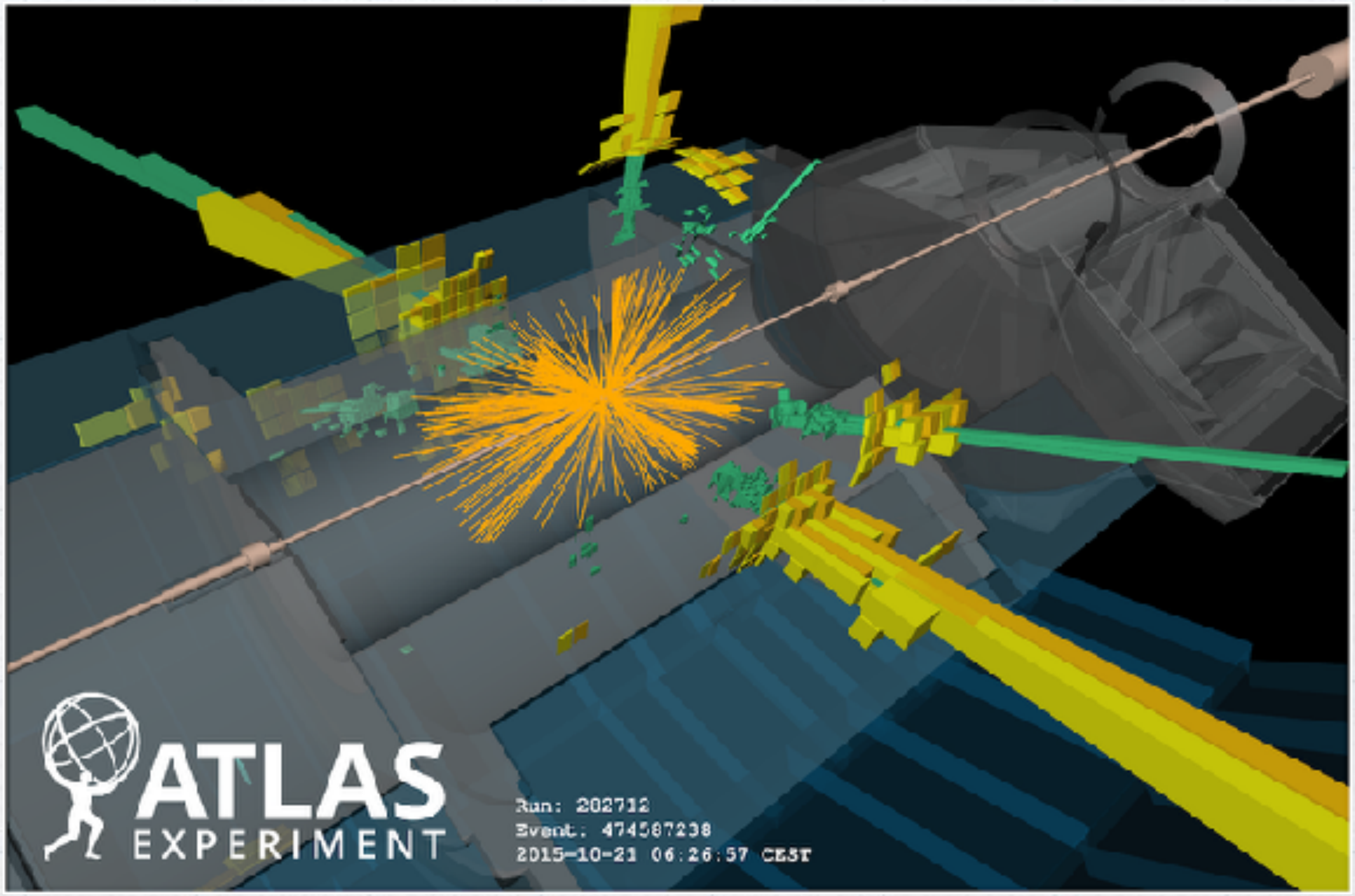
Lecture 1: Jets and jet algorithms

* jet definition(s)
* IRC safety
* sequential recombination
* pile-up



jets for experimentalists

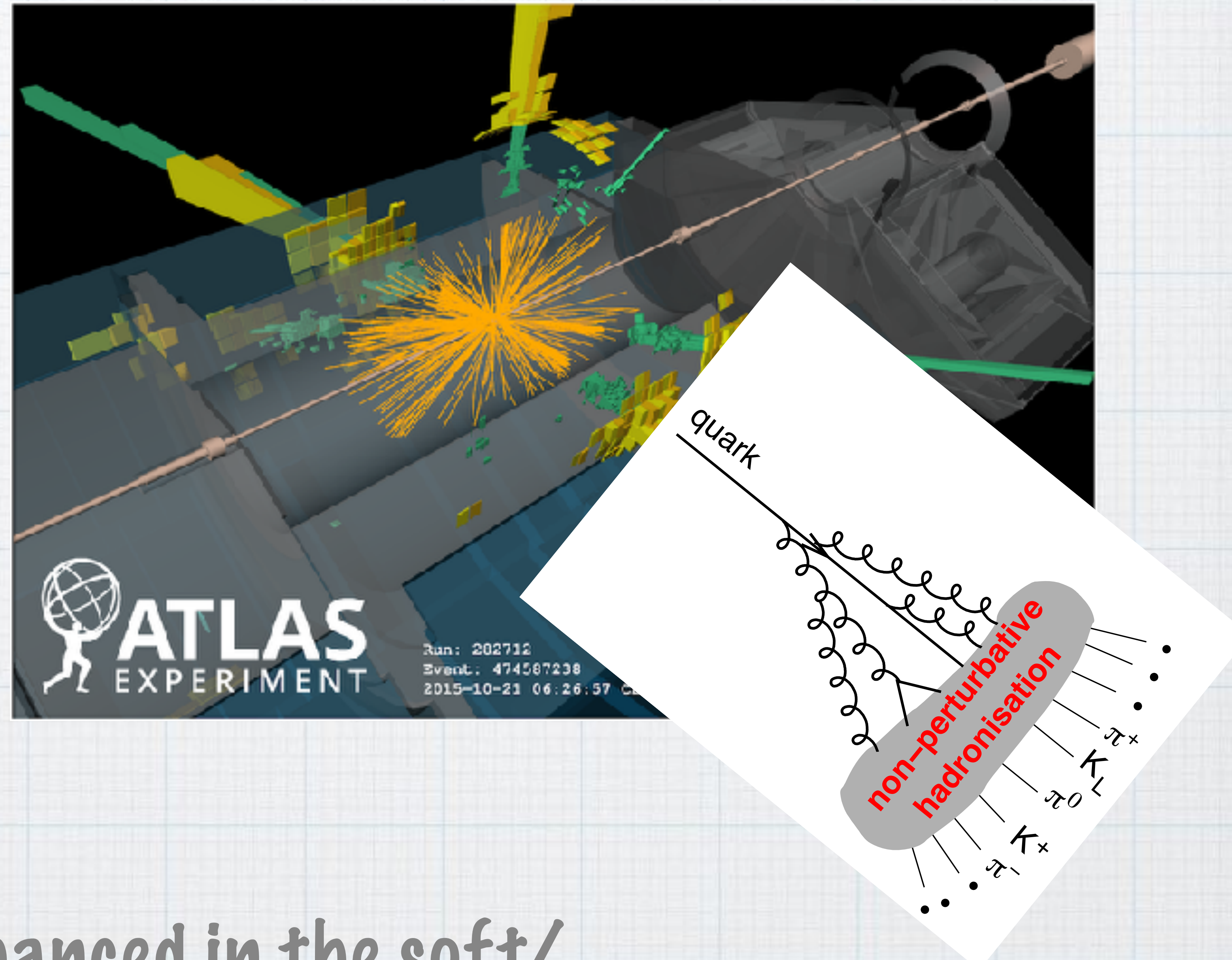
* high-energy collisions ofter results into collimated sprays of particles



* why?

jets for experimentalists

* high-energy collisions offer results into collimated sprays of particles



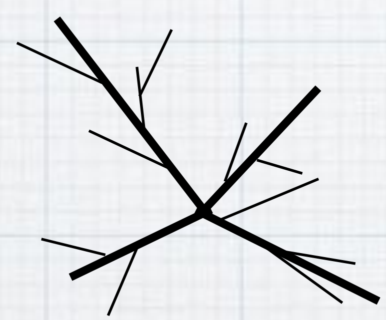
* why?

gluon emission enhanced in the soft/collinear limit

$$\int \frac{dE}{E}\frac{d\theta}{\theta}\alpha_s \gg 1$$

jets for theorists

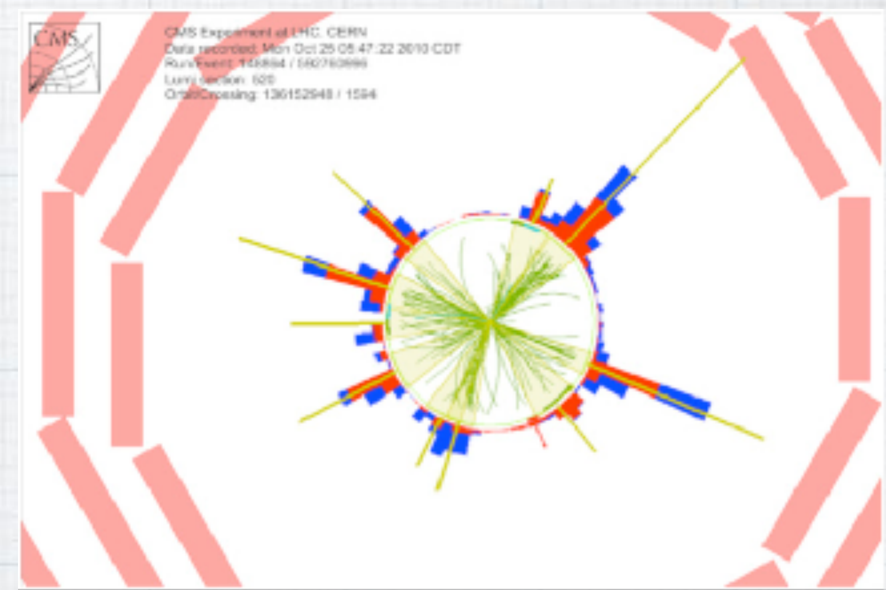
* jets are extremely useful for theorists
* powerful way of turning calculations into predictions



theory-land:
(quarks & gluons)

??

jets

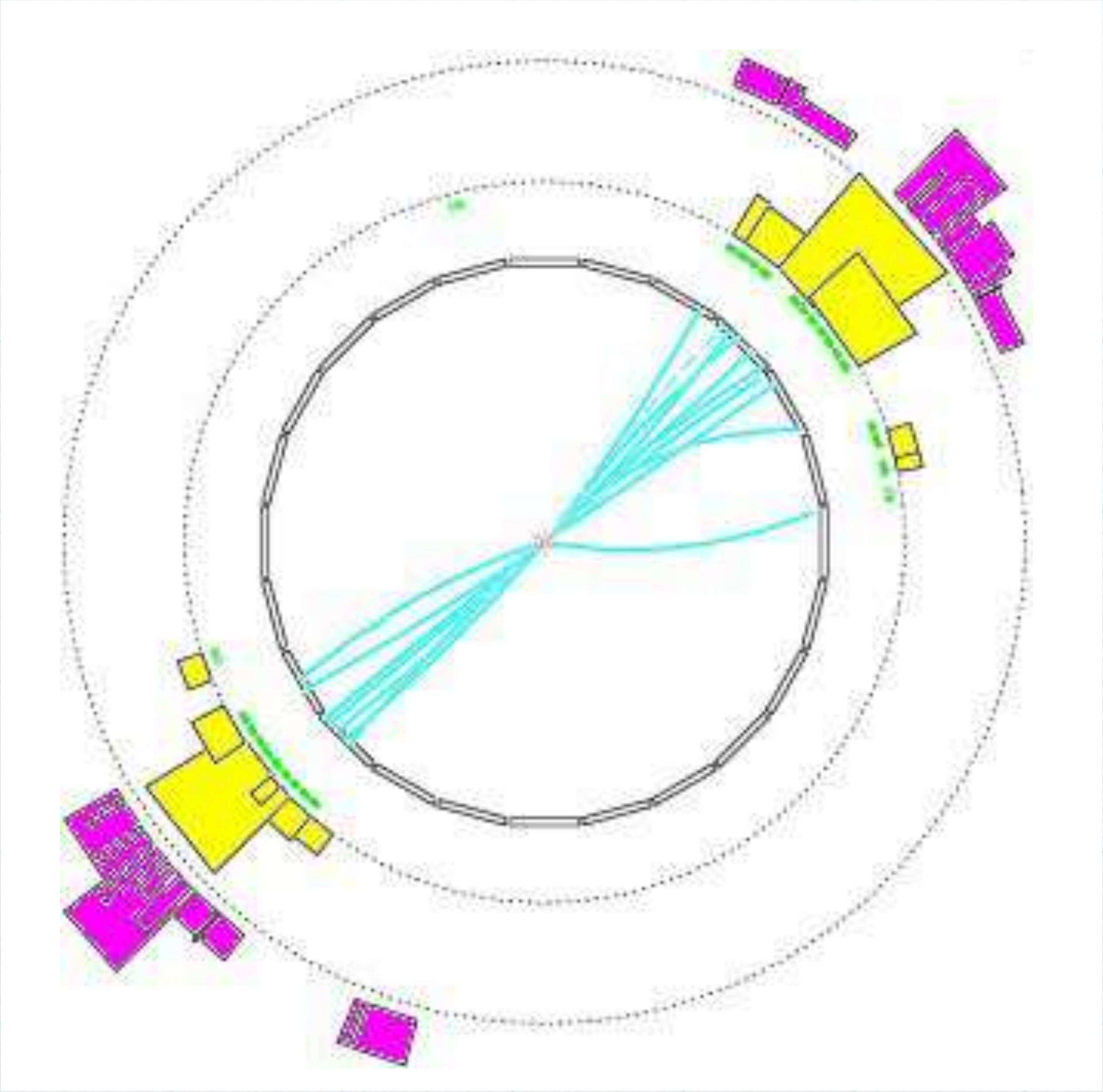


real data
(π,K,p,e,μ)

thanks to jets we can
reduce the complexity of the final state, simplifying many hadrons to simpler objects that one can hope to calculate

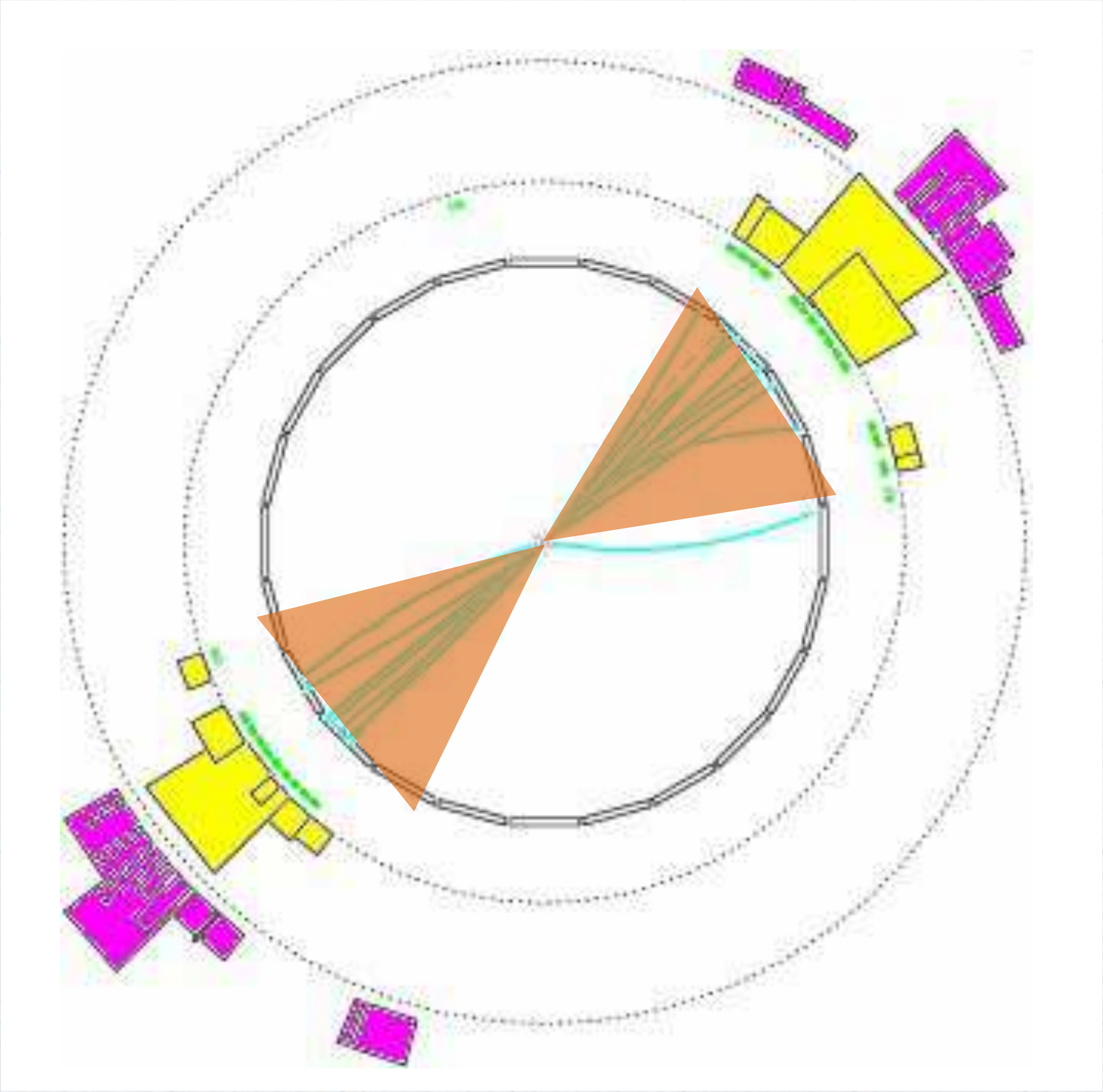
what is a jet?

* how many jets do you see?



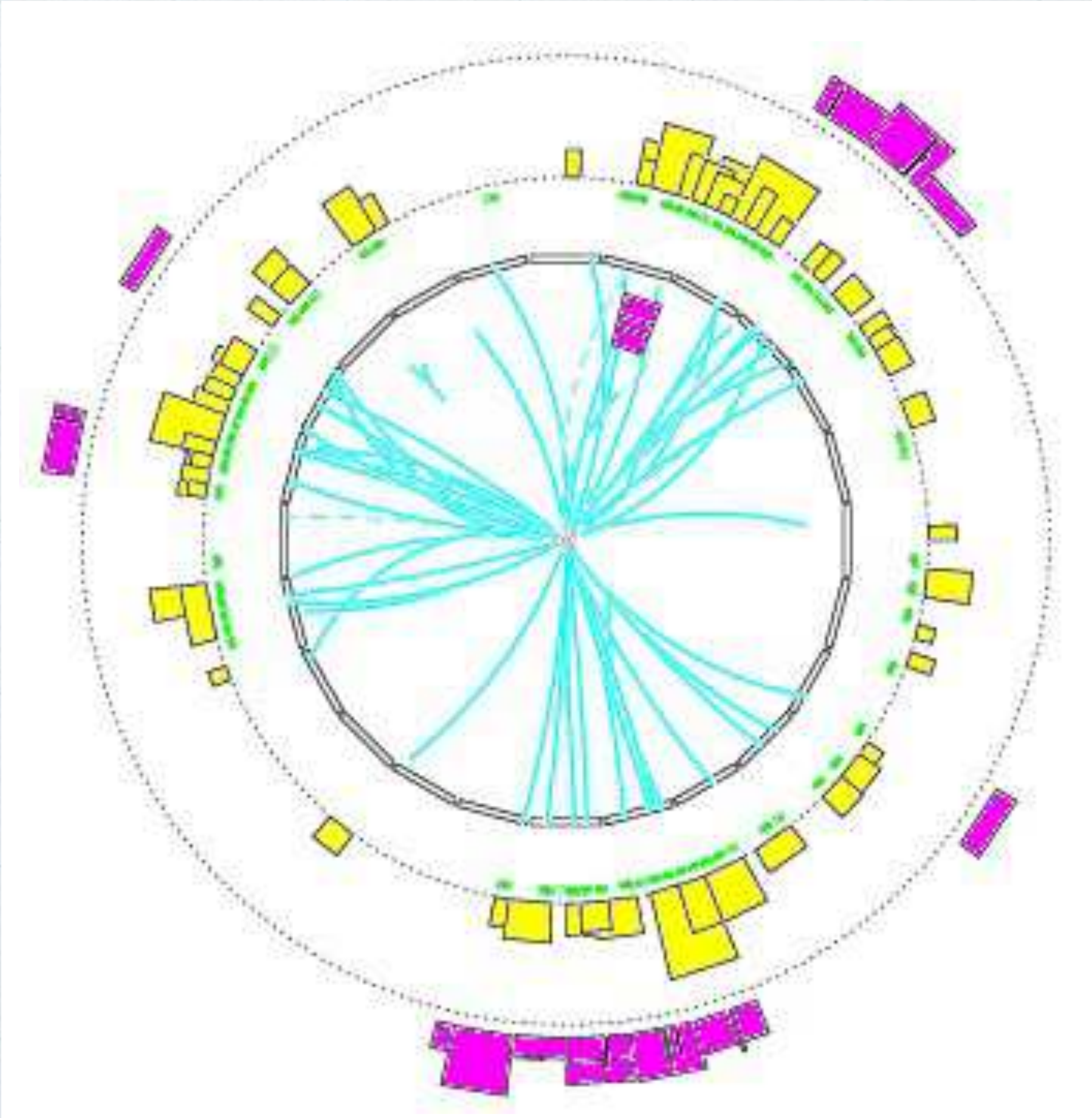
what is a jet?

* how many jets do you see?
* two is probably a good guess
* eyeballing not good enough!



what is a jet?

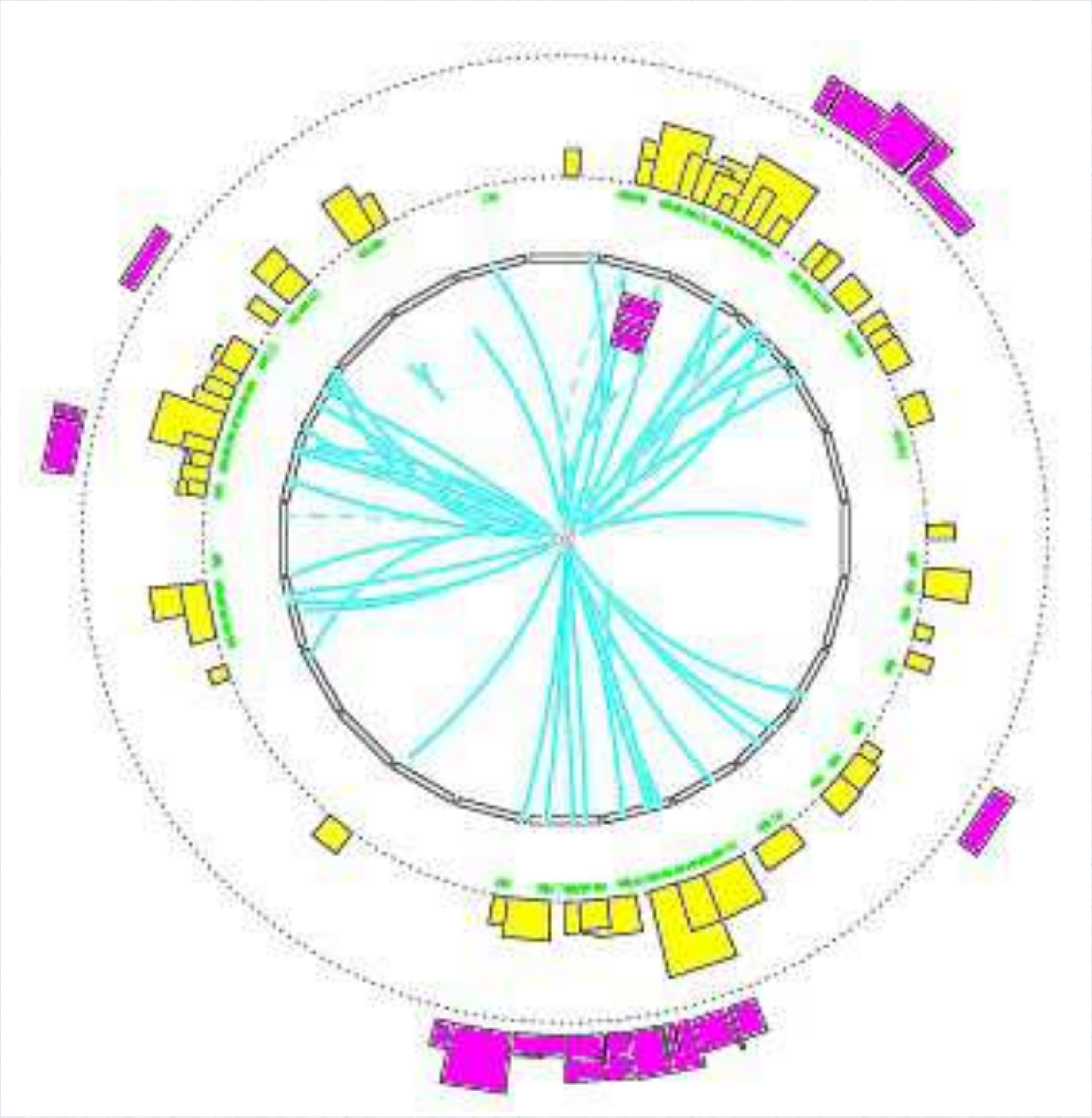
* what about now ?



what is a jet?

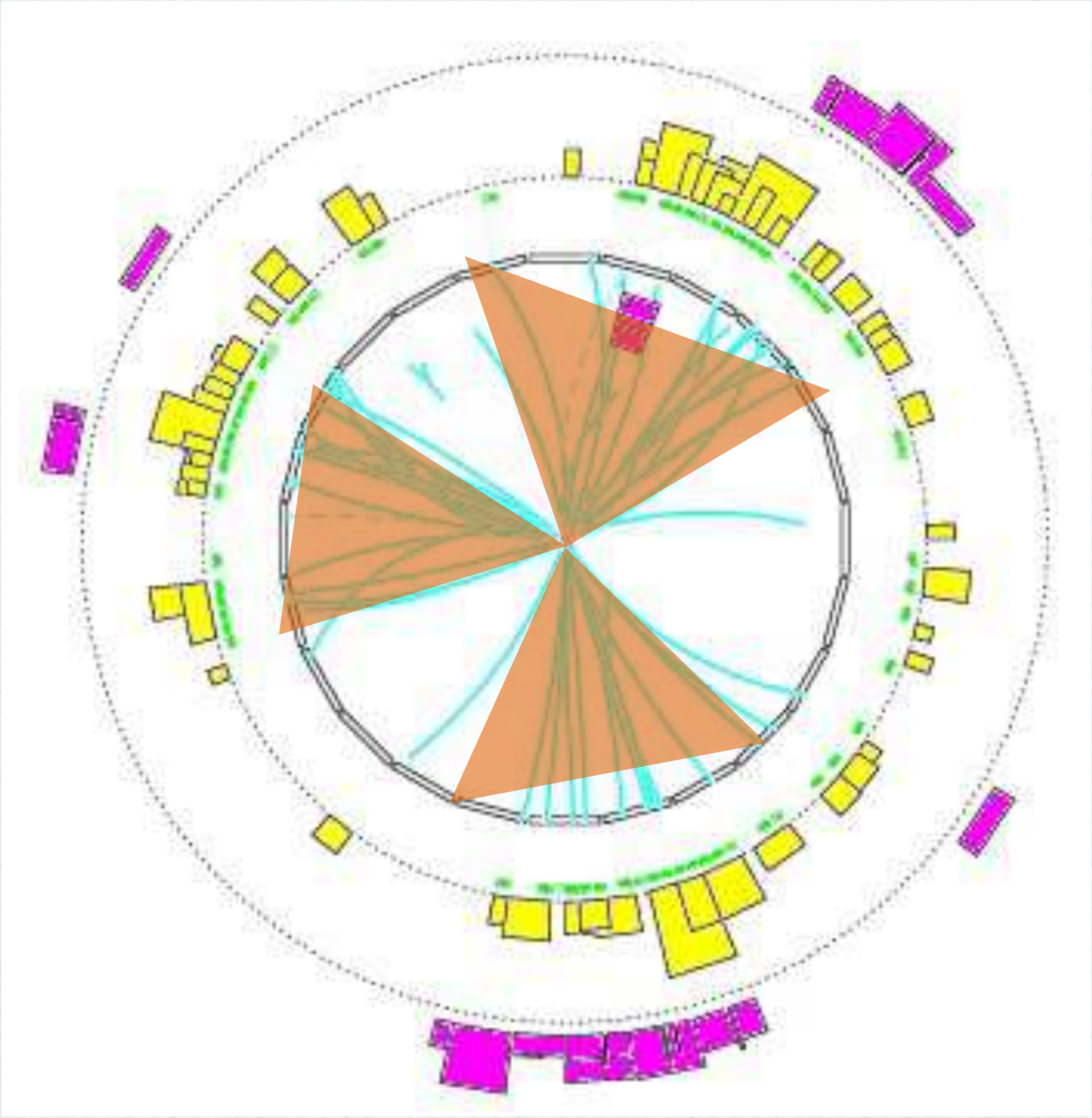
* what about now ?

* messy events are more ambiguous



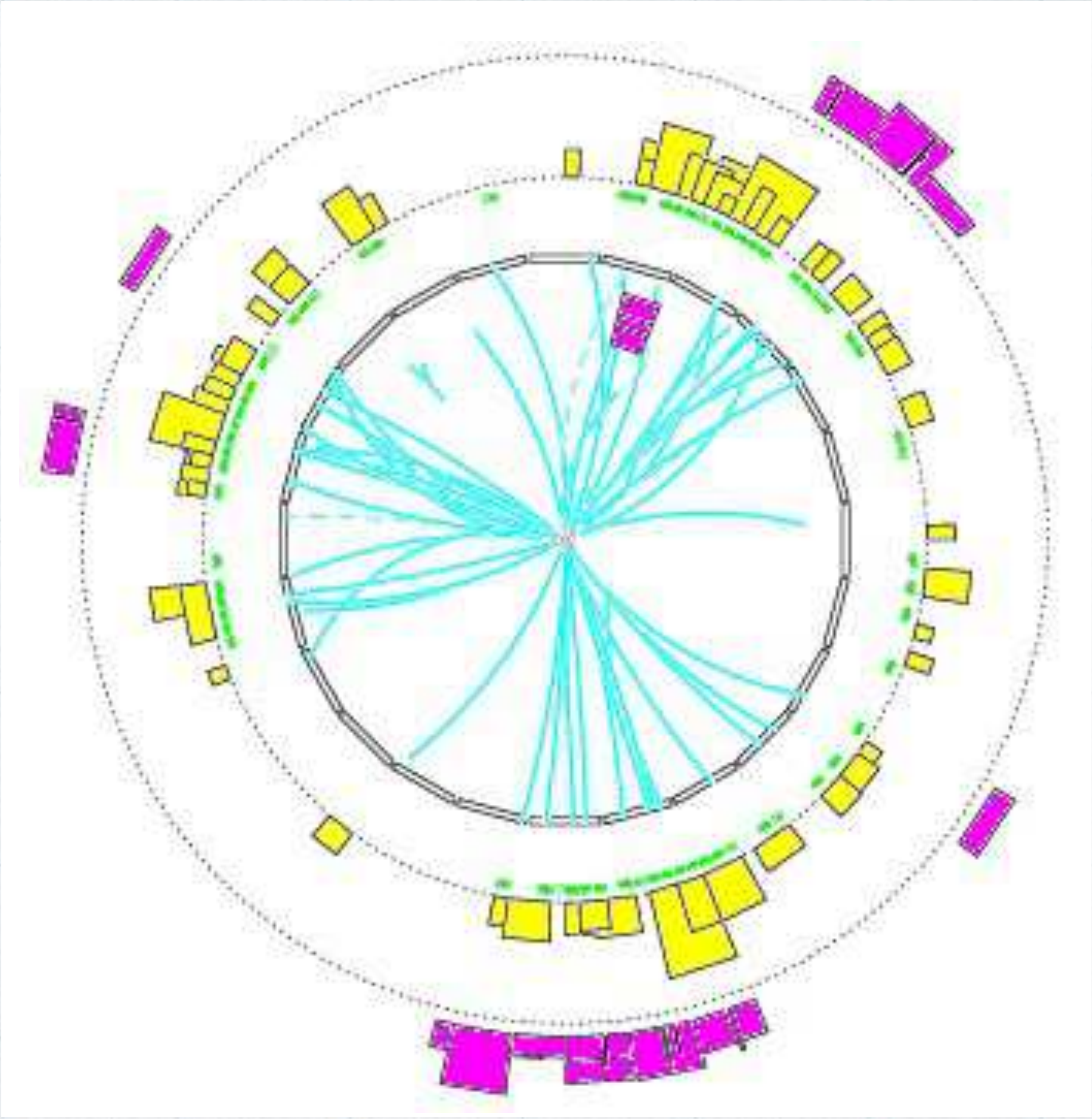
what is a jet?

- what about now ?
- messy events are more ambiguous
- 3 jet event?



what is a jet?

* what about now ?

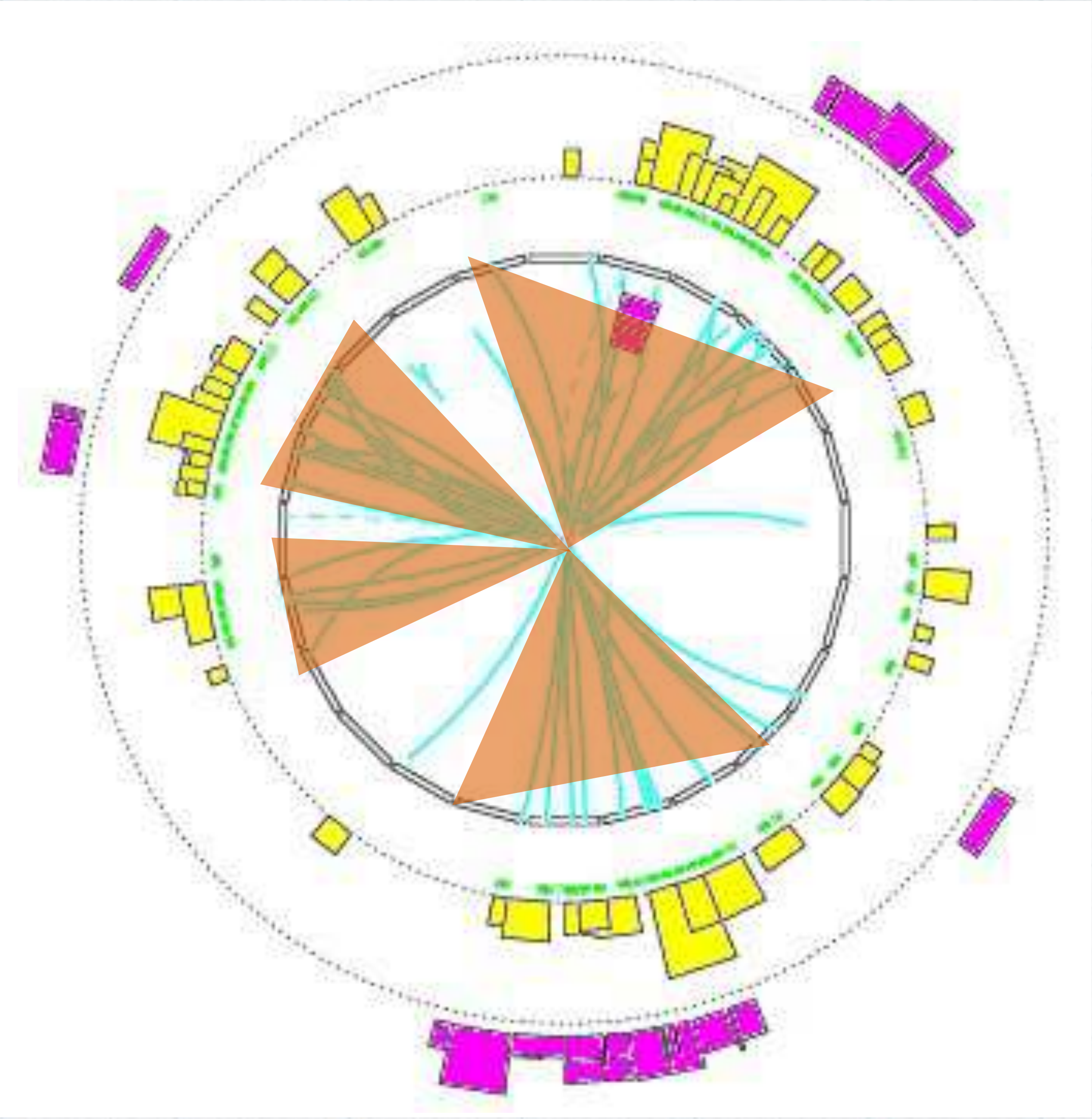


what is a jet?

* what about now ?

* messy events are more ambiguous

* or 4 jet event?



* we need a way to define jets in a given event

jet definition

a jet algorithm
+
its parameters (e.g. R)
+
a recombination scheme
=
a jet definition

* examples of recombination schemes:
  * E-scheme: sum all the four momenta
  * winner-take-all

jet clustering algorithm

* an algorithm that maps the momenta of the final state particles into the momenta of a certain number of jets

$$\{p_i\} \xrightarrow{\text{jet algorithm}} \{j_k\}$$

particles,
4-momenta,
calorimeter towers, ...

often comes with resolution parameters e.g. R

jets

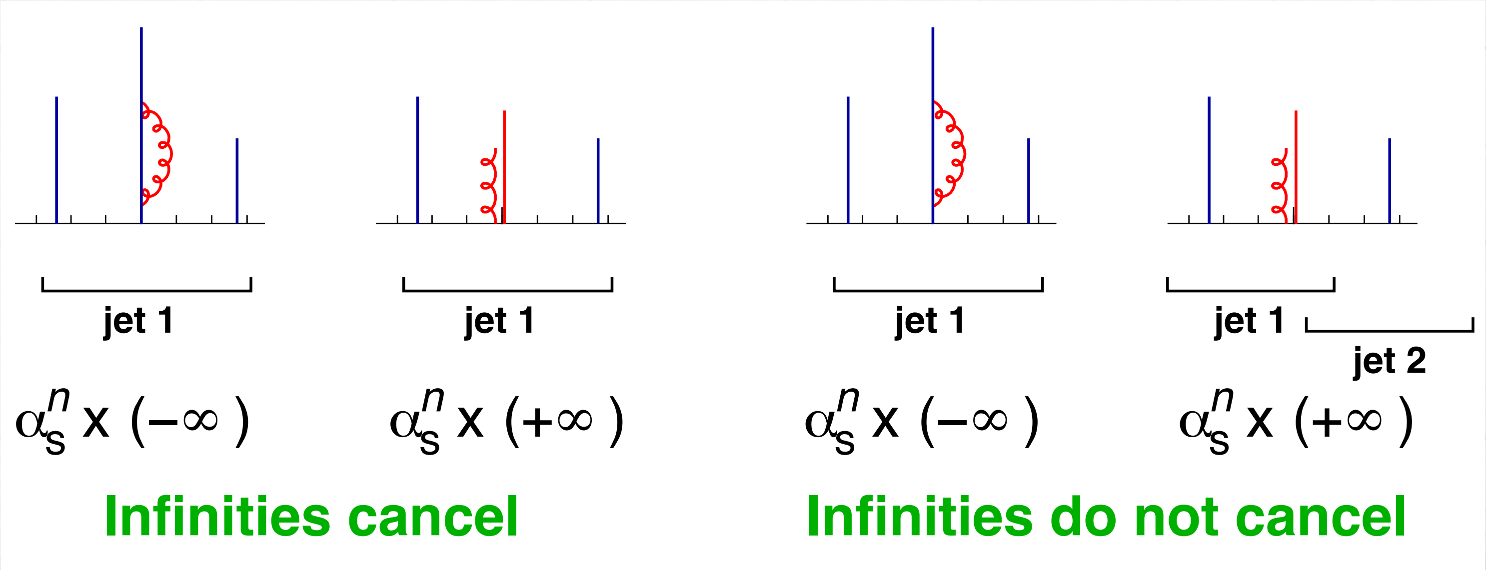
* jet definitions must make sense for both theorists and experimentalists!

what do theorists want?

- Infra-Red and Collinear Safety!
- An observable is **IRC safe** if, in the limit of a collinear splitting, or the emission of an infinitely soft particle, the observable remains **unchanged:**

$$O(X; p_1, \ldots, p_n, p_{n+1} \to 0) \to O(X; p_1, \ldots, p_n)$$

$$O(X; p_1, \ldots, p_n \parallel p_{n+1}) \to O(X; p_1, \ldots, p_n + p_{n+1})$$



we need IRC safety if we want to compute things beyond LO!

homework 1

* which of the following observables are IRC safe (assuming the jet has been selected in an IRC safe fashion)?

  * the jet invariant mass

  * the invariant mass of tracks in a jet

  * generalised angularities (assume κ, $\beta>0$)

$$\lambda_{\kappa,\beta} = \sum_{i \in jet} \left(\frac{p_{Ti}}{p_T} \right)^{\kappa} \theta_i^{\beta}$$

what do experimentalists want?

* jet algorithms must be usable on real events
* fast and easy to calibrate
* a thousand particles in each event
* CMS high-level trigger output rate 50kHz

types of algorithms

* sequential recombination algorithms
  * bottom-up approach: combine particles starting from closest ones
  * how? Choose a distance measure, iterate recombination until few objects left, call them jets
  * usually trivially made IRC safe, but their algorithmically complex (unless you're clever)
  * Examples: Jade, k_t, Cambridge/ Aachen, anti-k_t ...

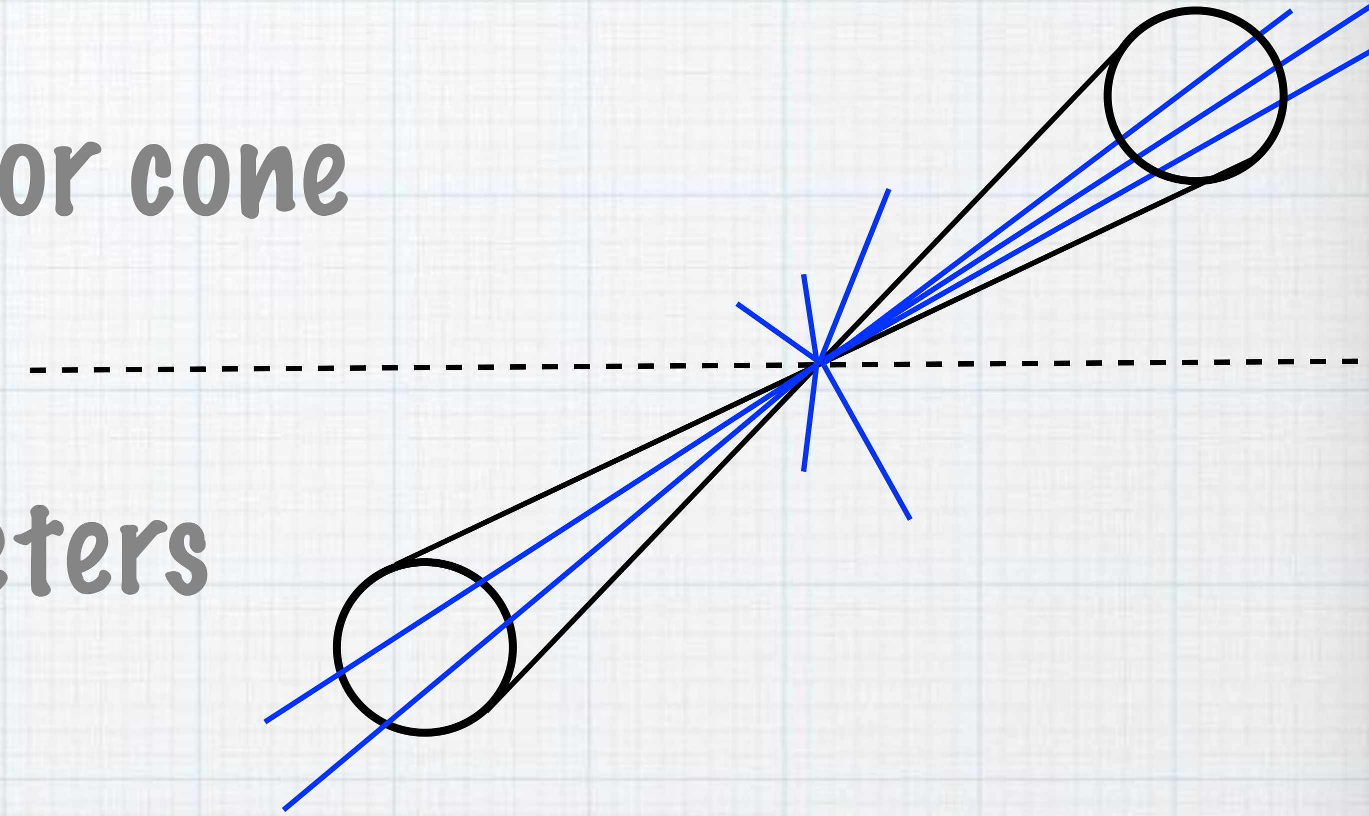
* cone algorithms
  * top-down approach: find coarse regions of energy flow.
  * how? Find stable cones (i.e. their axis coincides with sum of momenta of particles in it)
  * can be programmed to be fairly fast, at the price of being complex and IRC unsafe
  * Examples: JetClu, MidPoint, ATLAS cone, CMS cone, SISCone ...

for a complete review see G. Salam, Towards jetography (2009)

a bit of history

* first calculation done for cone algorithm

* two resolution parameters



To study jets, we consider the partial cross section $\sigma(E,\theta,\Omega,\epsilon,\delta)$ for e^+e^- hadron production events, in which all but a fraction $\epsilon \ll 1$ of the total e^+e^- energy E is emitted within some pair of oppositely directed cones of half-angle $\delta \ll 1$, lying within two fixed cones of solid angle Ω (with $\pi\delta^2 \ll \Omega \ll 1$) at an angle θ to the e^+e^- beam line. We expect this to be measur-

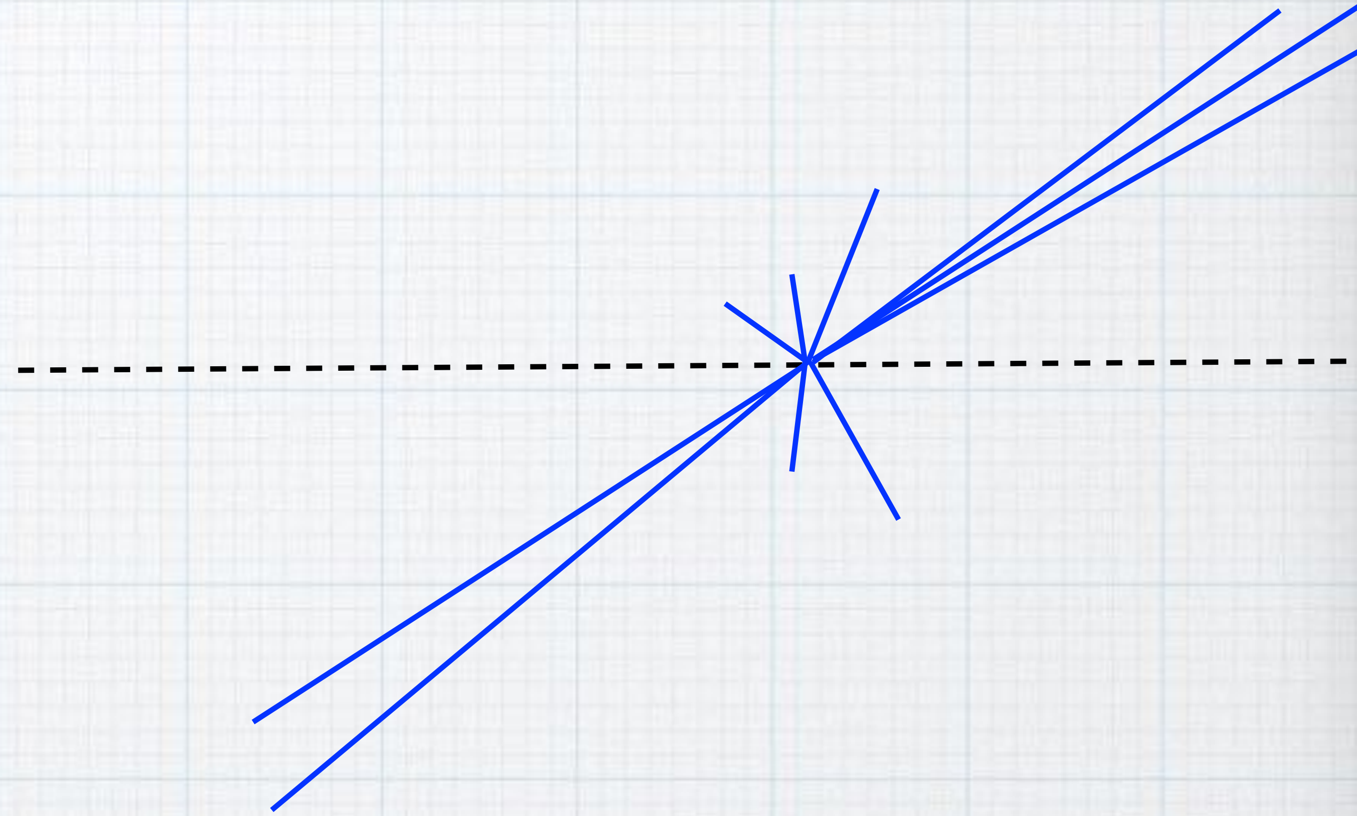
Sterman and Weinberg,
Phys. Rev. Lett. 39, 1436 (1977):

sequential recombination

* start with a list of particles,

* compute all distances d_{ij} and d_{iB}

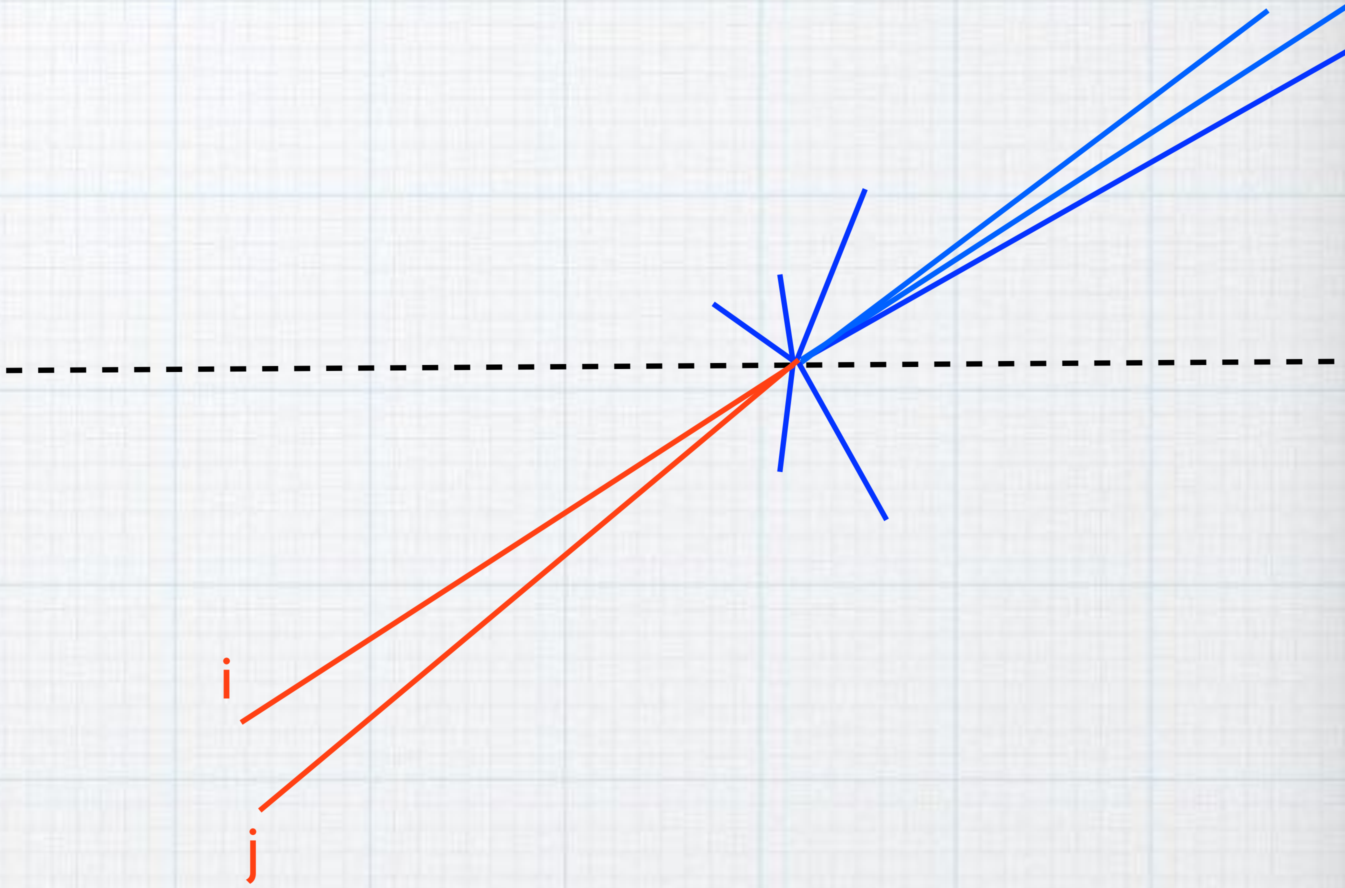
* find the minimum of all d_{ij} and d_{iB}



d_{ij} (weighted) distance between i j
d_{iB} external parameter or distance from the beam ...

sequential recombination

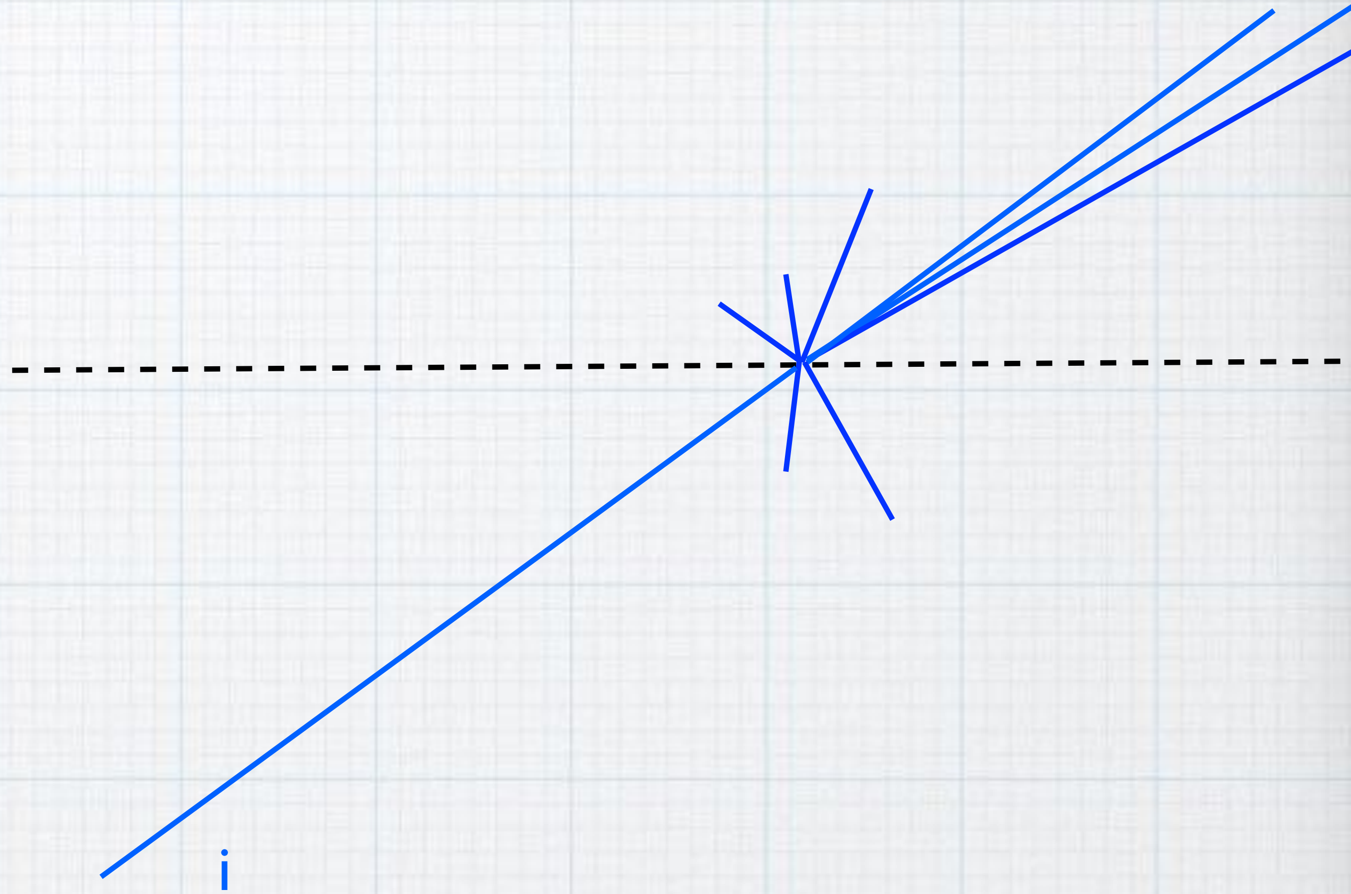
* start with a list of particles,
* compute all distances d_{ij} and d_{iB}
* find the minimum of all d_{ij} and d_{iB}
* if the minimum is a d_{ij}, recombine i and j and iterate



d_{ij} (weighted) distance between i j
d_{iB} external parameter or distance from the beam ...

sequential recombination

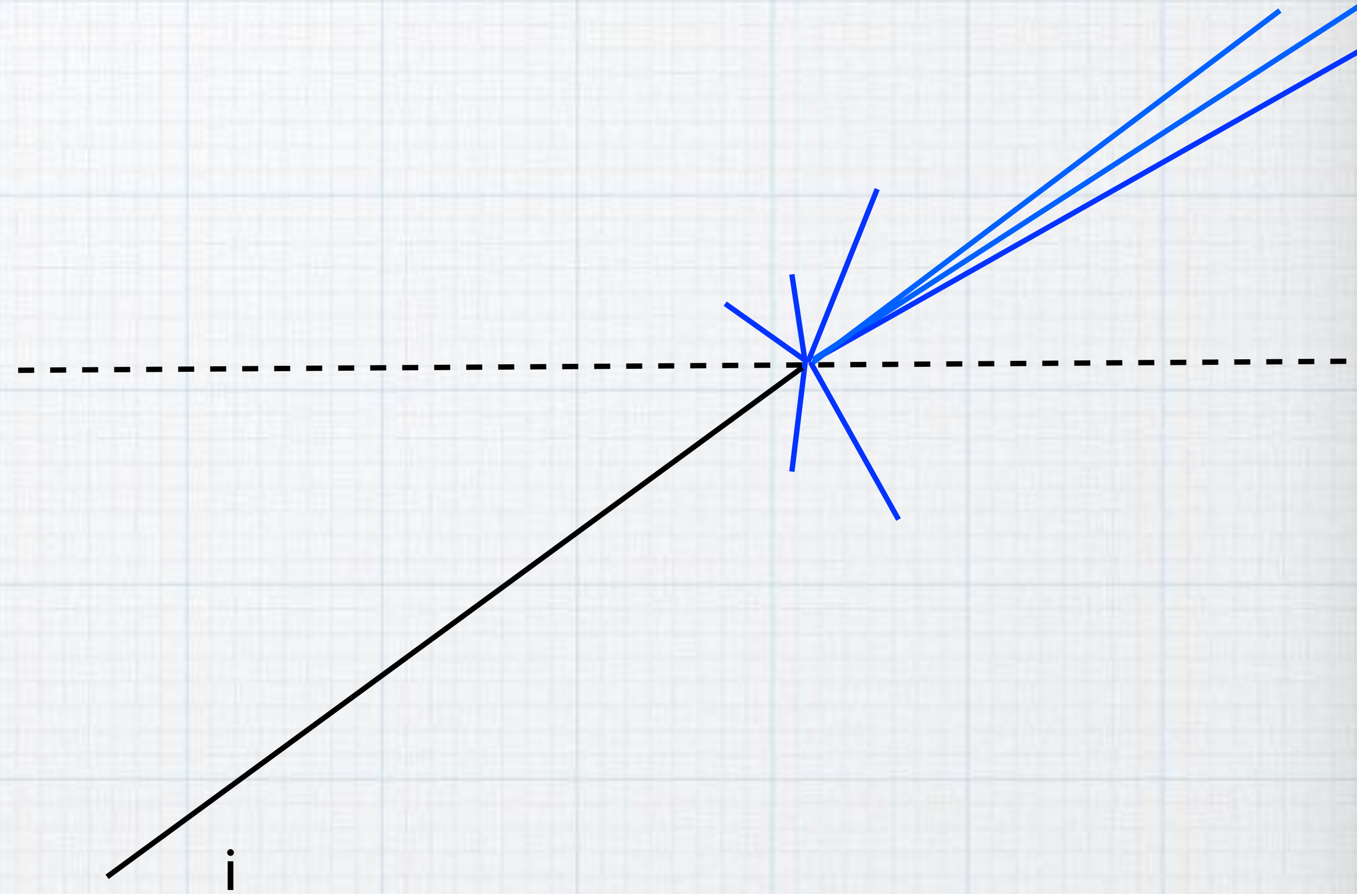
- start with a list of particles,
- compute all distances d_{ij} and d_{iB}
- find the minimum of all d_{ij} and d_{iB}
- if the minimum is a d_{ij}, recombine i and j and iterate



d_{ij} (weighted) distance between i j
d_{iB} external parameter or distance from the beam ...

sequential recombination

* start with a list of particles,
* compute all distances d_{ij} and d_{iB}
* find the minimum of all d_{ij} and d_{iB}
* if the minimum is a d_{ij}, recombine i and j and iterate
* otherwise call i a final-state jet, remove it from the list and iterate



d_{ij} (weighted) distance between i j
d_{iB} external parameter or distance from the beam ...

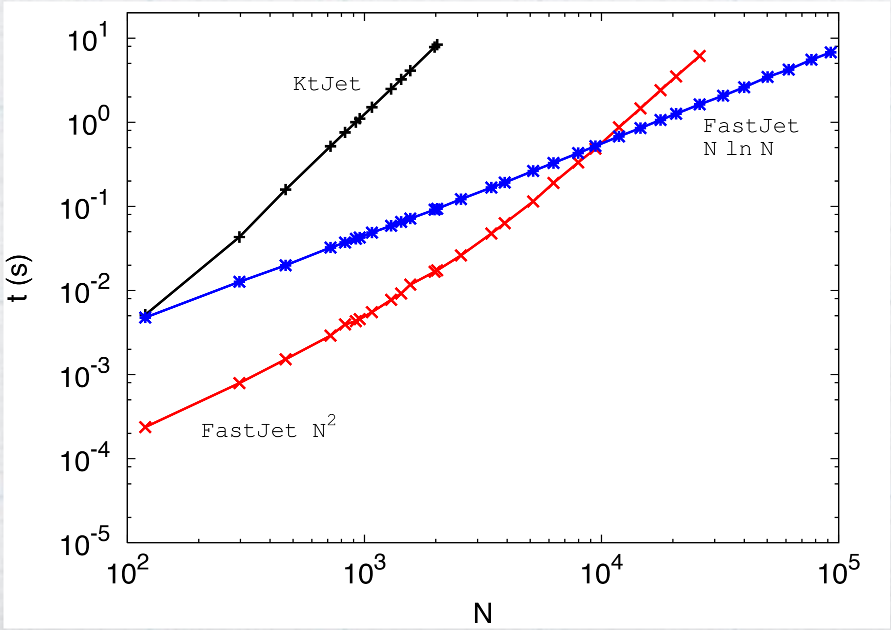
speeding-up the algorithms

* from combinatorics sequential recombination should scale like N^3
* an approach based on geometry (Voronoi diagrams) leads to notable improvements
* Sequential recombination algorithms could be implemented with $O(N^2)$ or even $O(N \ln N)$ complexity rather than $O(N^3)$

Cacciari, Salam, 2006

* Cone algorithms could be implemented exactly (and therefore made IRC safe) with $O(N^2 \ln N)$ rather than $O(N2^N)$ complexity

Salam, Soyez, 2007



method implemented in FastJet

the generalised k_t family

* actual choice of d_{ij} determines the algorithm

$$d_{ij} = \min(p_{ti}^{2p}, p_{tj}^{2p}) \frac{\Delta y^2 + \Delta\phi^2}{R^2}$$

IRC behaviour

p = 1 k_t algorithm

S. Catani, Y. Dokshitzer, M. Seymour and B. Webber, Nucl. Phys. B406 (1993) 187
S.D. Ellis and D.E. Soper, Phys. Rev. D48 (1993) 3160

new soft particle (pt $\to$0) means that d $\to$ 0 $\Rightarrow$ clustered first, no effect on jets

new collinear particle ($\Delta y^2+\Delta\Phi^2$ $\to$ 0) means that d $\to$ 0 $\Rightarrow$ clustered first, no effect on jets

p = 0 Cambridge/Aachen algorithm

Y. Dokshitzer, G. Leder, S.Moretti and B. Webber, JHEP 08 (1997) 001
M.Wobisch and T.Wengler, hep-ph/9907280

new soft particle (p_t $\to$0) can be new jet of zero momentum $\Rightarrow$ no effect on hard jets

new collinear particle ($\Delta y^2+\Delta\Phi^2$ $\to$ 0) means that d $\to$ 0 $\Rightarrow$ clustered first, no effect on jets

p = -1 anti-k_t algorithm

M. Cacciari, G. Salam and G. Soyez, arXiv:0802.1189

new soft particle (p_t $\to$0) means d $\to\infty$ $\Rightarrow$ clustered last or new zero-jet, no effect on hard jets

new collinear particle ($\Delta y^2+\Delta\Phi^2$ $\to$ 0) means that d $\to$ 0 $\Rightarrow$ clustered first, no effect on jets

the k_t algorithm

* the k_t distance is the inverse of the QCD splitting probability

$$\frac{dP_{k\to ij}}{dE_i d\theta_{ij}} \sim \frac{\alpha_s}{\min(E_i, E_j)\theta_{ij}}$$

* the algorithm roughly inverts the QCD shower, bringing us back to the hard scattering
* the clustering history has physical meaning
* jets grow around soft particles, which is a problem in a noisy environment as the LHC

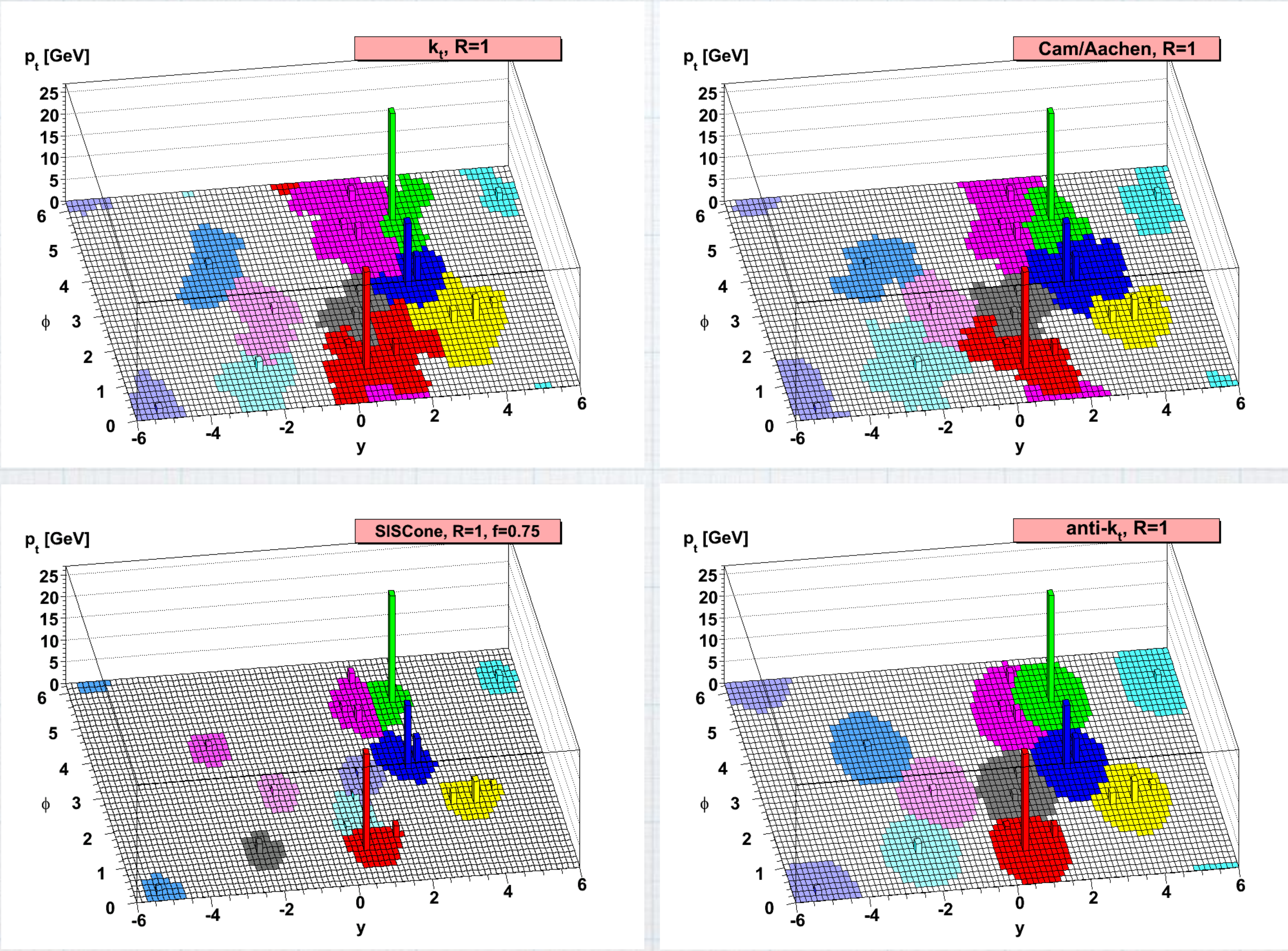
the anti-k_t algorithm

* with this measure soft particles are always far away
* jets grow around hard cores
* if no other hard particles are around the algorithm provides (ironically) perfect cones
* however, the clustering history carries little physics (re-clustering)

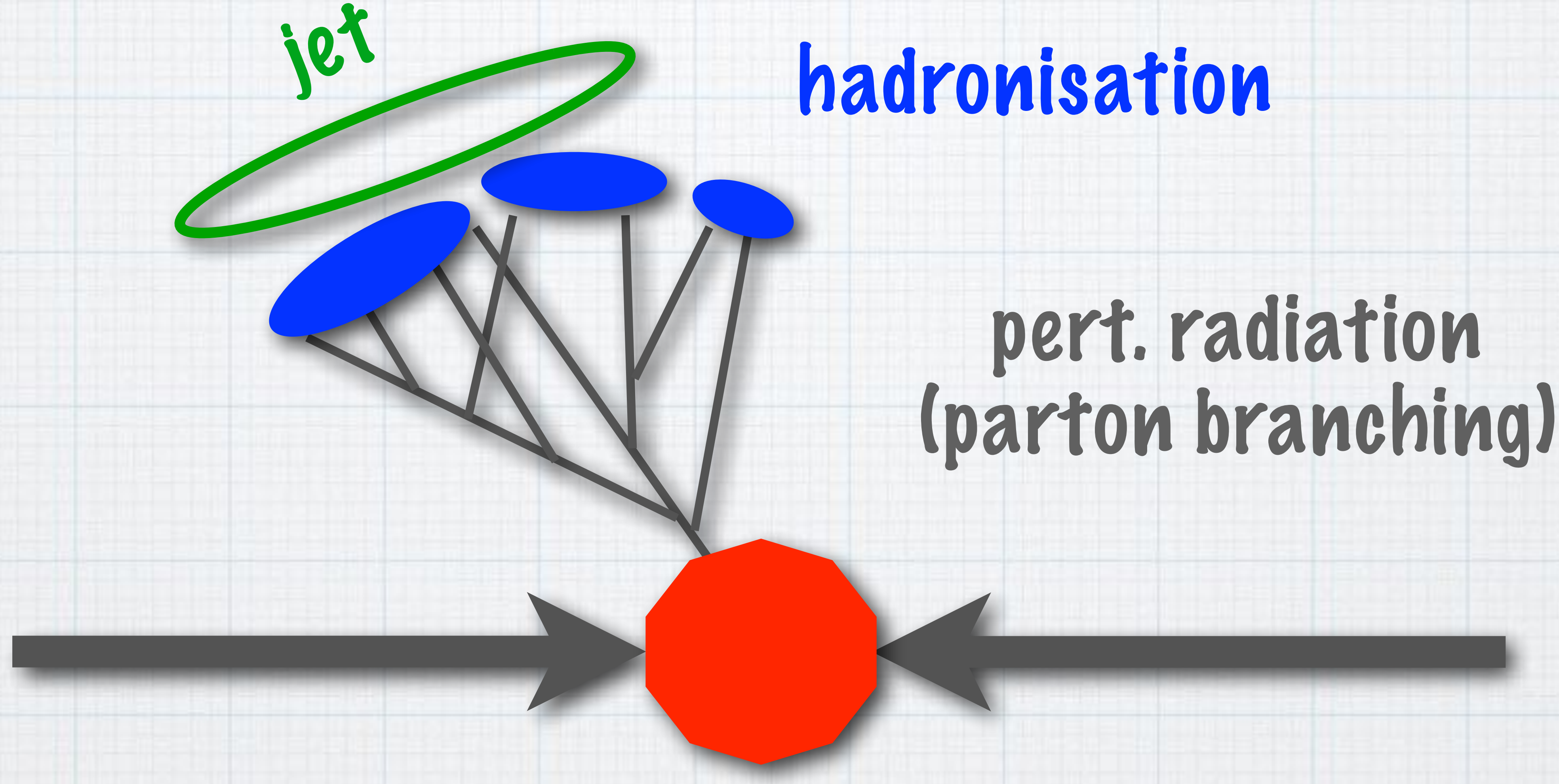
homework 2

* show that for an event made up of two particles all gen. k_t algorithms recombine them is their azimuth-rapidity distance is less than R
* things dramatically changes with many particles!

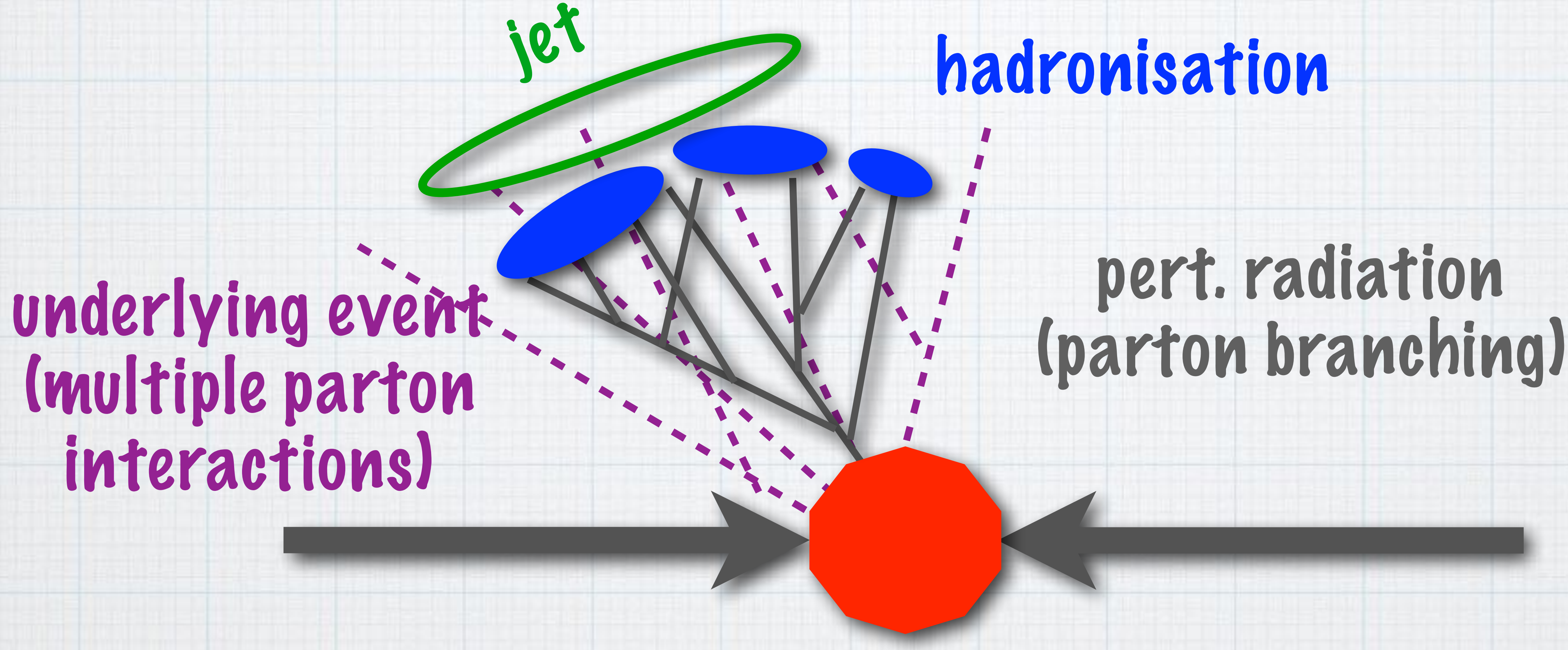
comparing them all



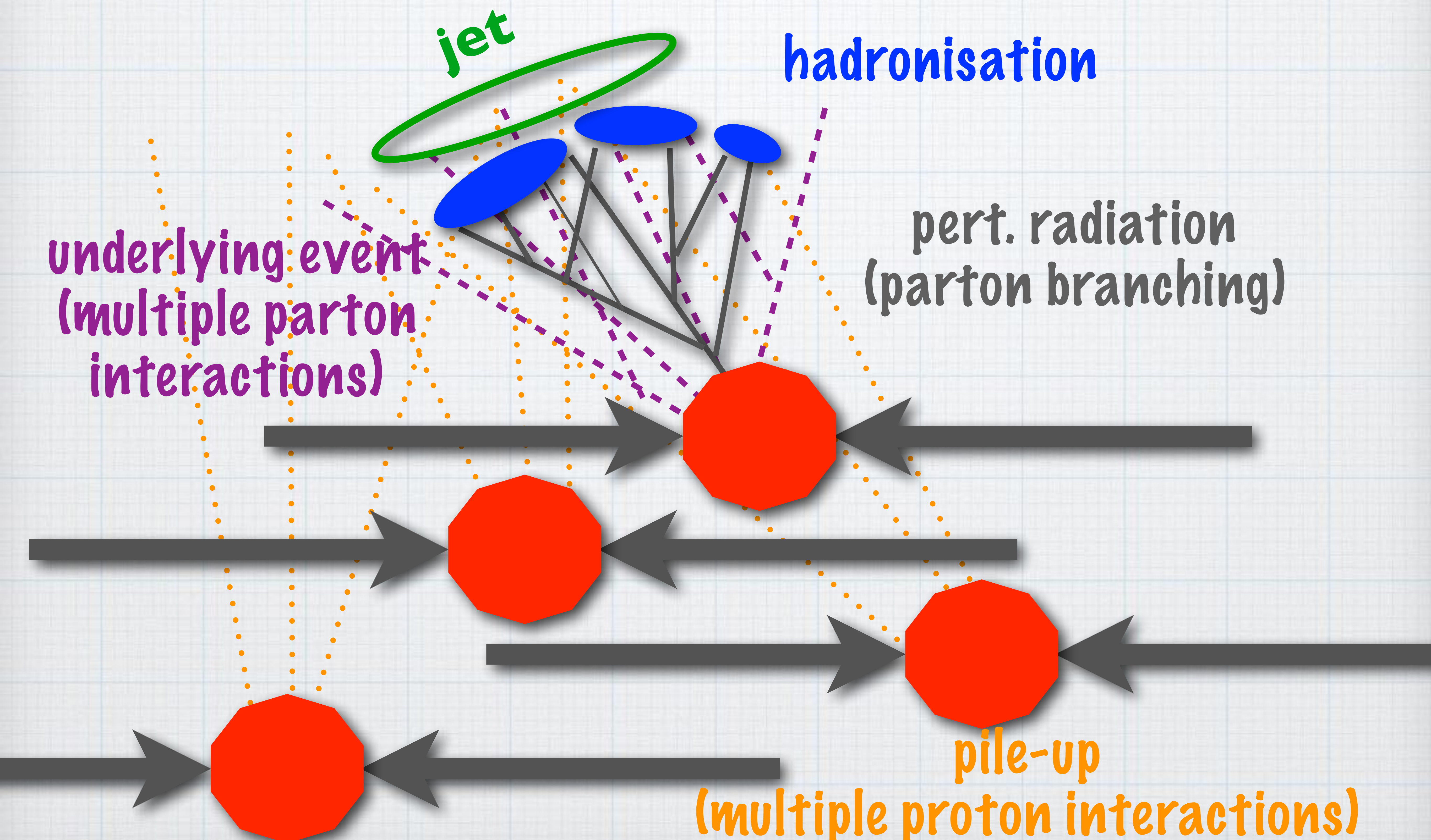
a useful cartoon



a useful cartoon



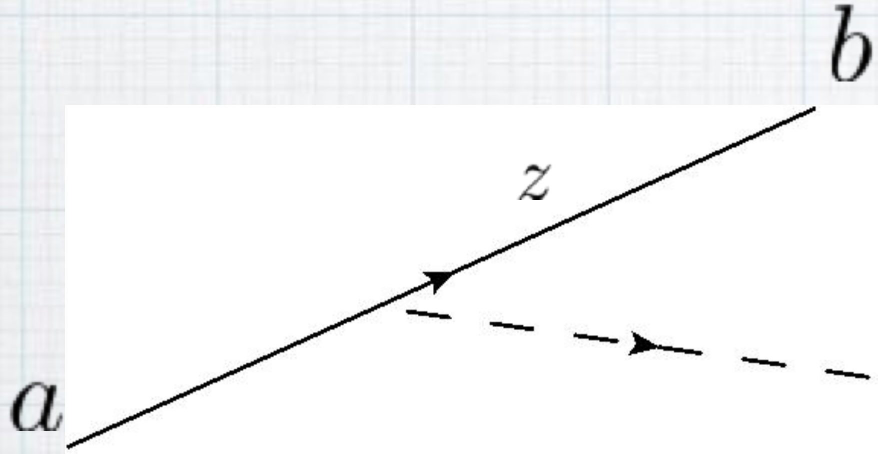
a useful cartoon



preview:
QCD in the soft/collinear limit

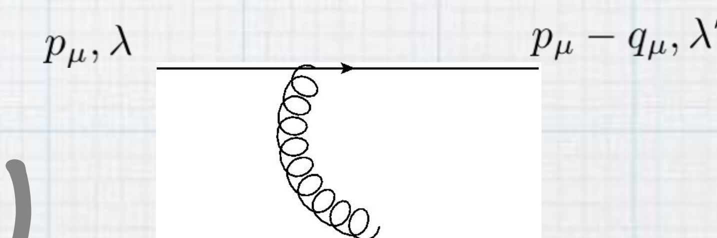
* factorisation properties in both limits

* collinear limit (semi-classical)

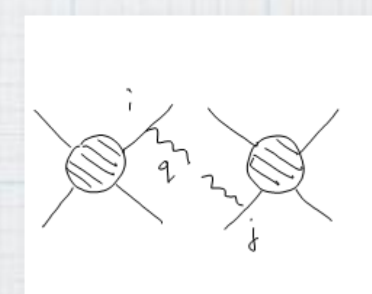


$$d\sigma_{n+1} = d\sigma_n \frac{\alpha_s}{2\pi}\frac{dq^2}{q^2} dz\ P_{ba}(z)$$

* soft limit (eikonal rules)



$$= 2 g p_\mu \delta_{\lambda\lambda'} T^a_{ij}$$



$$d\sigma_{n+1} = d\sigma_n \frac{\alpha_s}{2\pi}\frac{dE}{E}\frac{d\Omega}{2\pi}\sum_{ij} C_{ij} E^2 \frac{p_i \cdot p_j}{p_i \cdot q\, p_j \cdot q}$$

estimating p_t shifts

- we can use soft emission kinematics to estimate the changes in p_t from the hard parton to the measured quantities

- assume a finite coupling in the IR

PT radiation:

$$q: \quad \langle \Delta p_t \rangle \simeq \frac{\alpha_s C_F}{\pi} p_t \ln R$$

Hadronisation:

$$q: \quad \langle \Delta p_t \rangle \simeq -\frac{C_F}{R} \cdot 0.4 \text{ GeV}$$

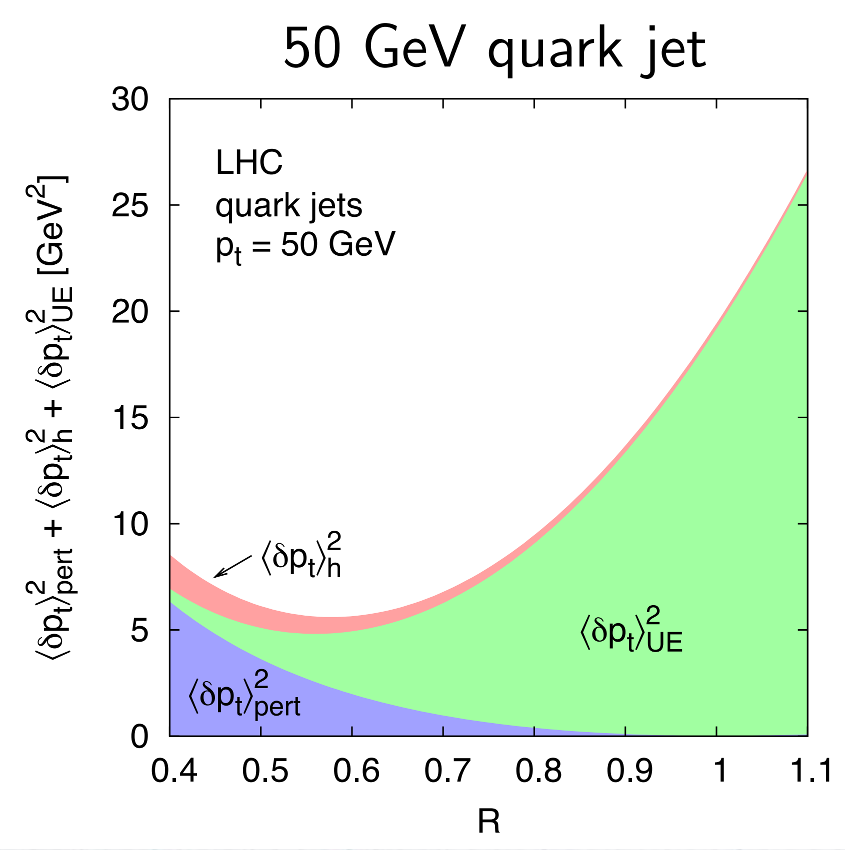
Underlying event:

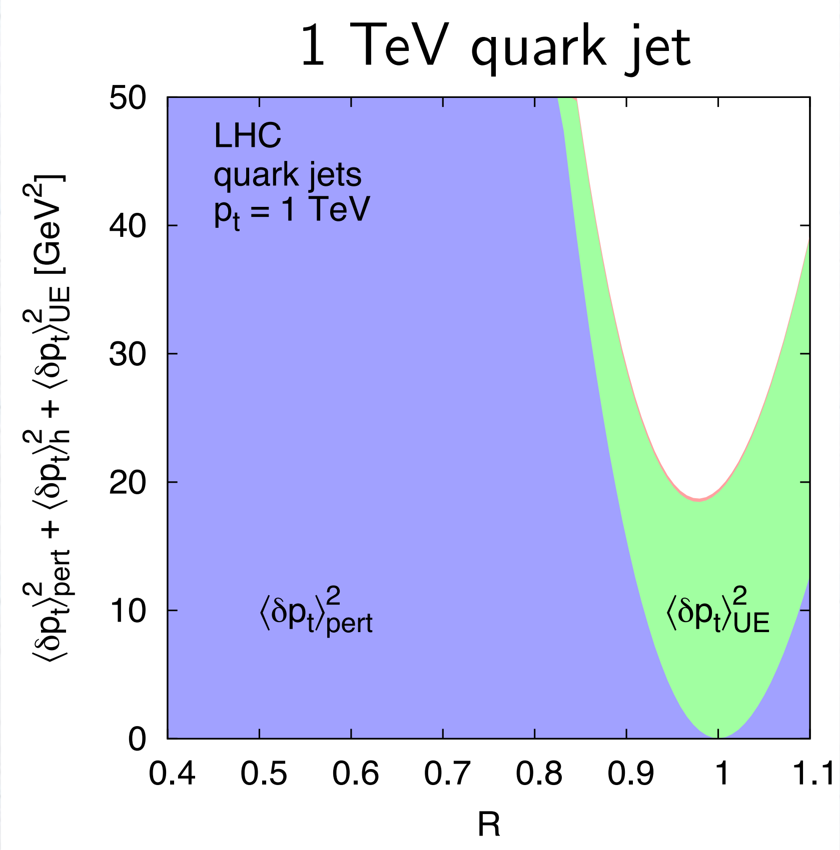
$$q, g: \quad \langle \Delta p_t \rangle \simeq \frac{R^2}{2} \cdot 2.5\text{--}15 \text{ GeV}$$

Dasgupta, Magnea, Salam (2007)

calculation

what is the optimal R?

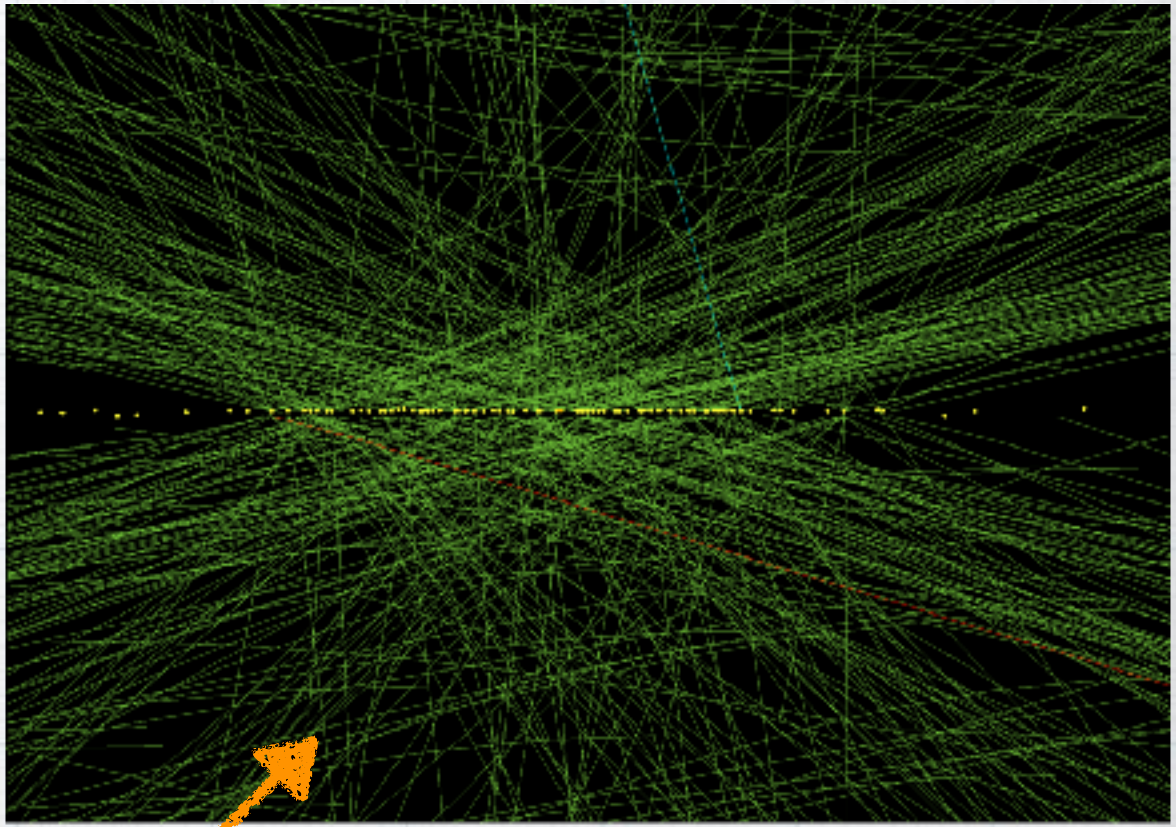




* at low p_t small R (0.4-0.6) reduces the impact of UE
* at high p_t perturbative effects dominate (see lecture 2)
* at high p_t R=1 seems excellent (good also for boosted-object, see lecture 3)

pile-up

* pile-up can deposit several tens of GeV (or even hundreds, in a heavy ion collision) into a medium-sized jet
* it's a direct consequence of the desired high luminosity
* it hampers how ability of extracting useful information about the hard scatters

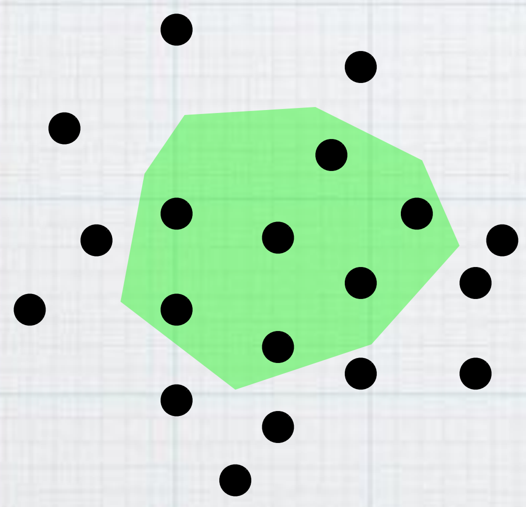


a 78-vertices event from CMS

hard jets and pile-up

* susceptibility measures how much background is picked up (jet area)

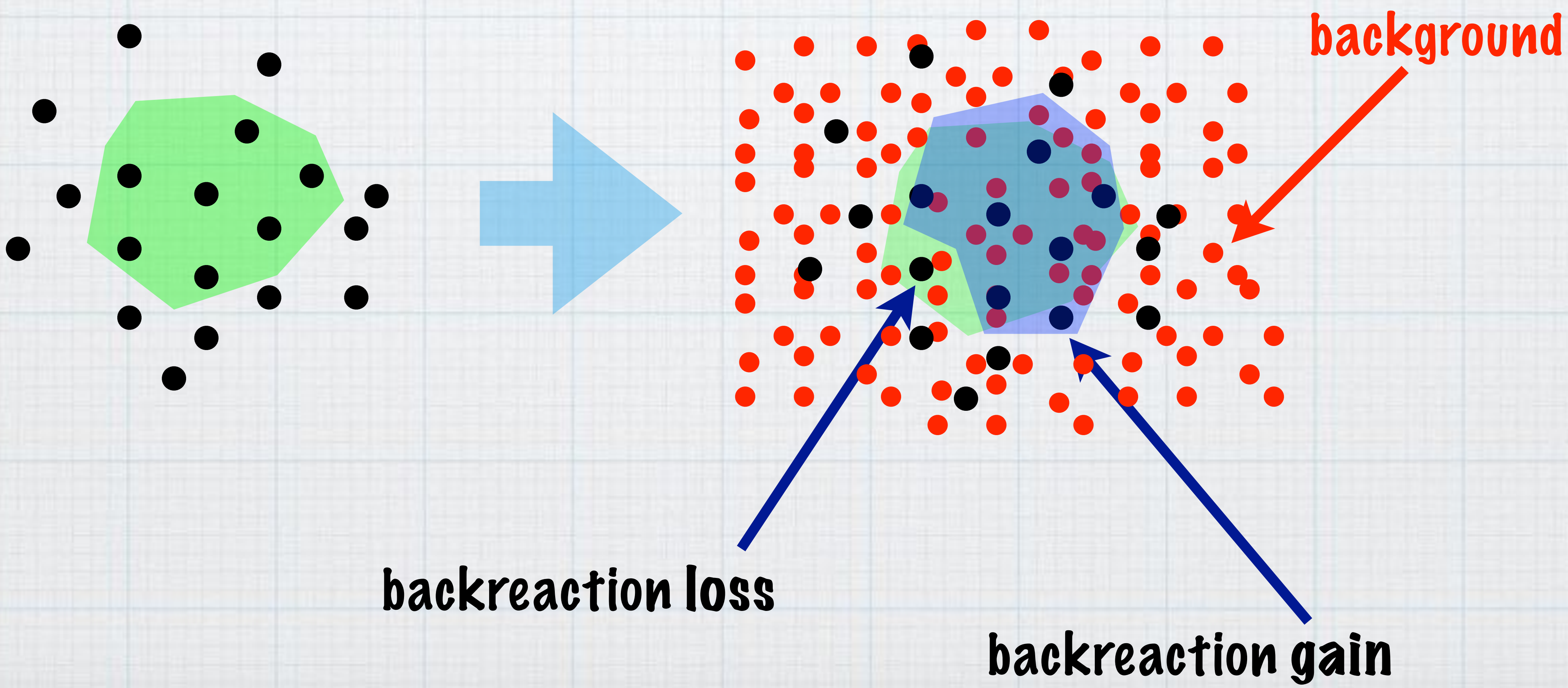
* resiliency measures how much the original jet is modified (backreaction)



hard jets and pile-up

* susceptibility measures how much background is picked up (jet area)

* resiliency measures how much the original jet is modified (backreaction)



hard jets and pile-up

* **susceptibility** measures how much background is picked up (jet area)

* **resiliency** measures how much the original jet is modified (backreaction)

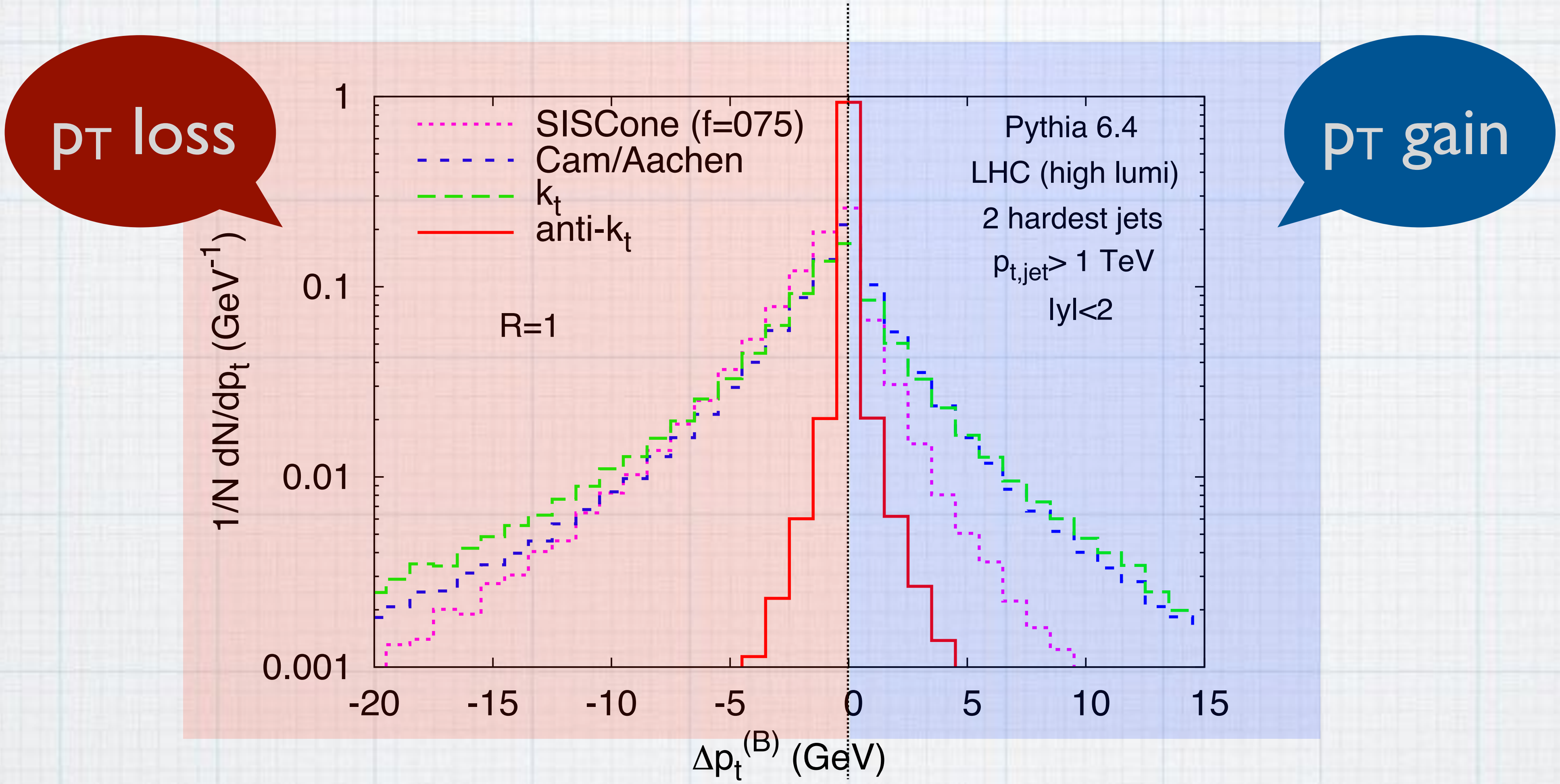
$$\Delta p_t = \rho A \pm \left(\sigma\sqrt{A} + \sigma_\rho A + \rho\sqrt{\langle A^2\rangle - \langle A\rangle^2}\right) + \Delta p_t^{BR}$$

background momentum density (per unit area)

background 'susceptibility'

backreaction 'resiliency'

resiliency



- anti-k_t jets are much more resilient to changes from background immersion
- their regular shape makes them easier to correct for detector effects
- default choice for LHC collaborations

mitigating pile-up

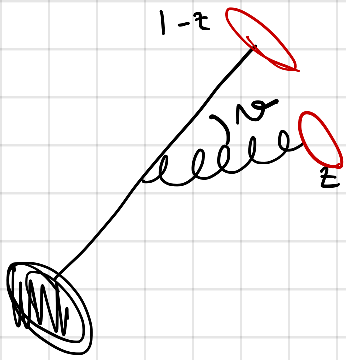
* Jet-based
    * Cluster the full event, determine the event-specific (ρ) and jet-specific (A) quantities, and subtract the relevant contamination from a given observable
    * Pros: largely unbiased subtraction
    * Cons: slow, potentially large(er) residual uncertainty
    * Examples: `jet area/median' in FastJet, GenericSubtractor for jet shapes, JetFFMoments for fragmentation functions, ...
* Particle-based
    * Produce a reduced event, by dropping some of the particles. Cluster this reduced event, and calculate from it the observables
    * Pros: fast, often small(er) residual uncertainty
    * Cons: not natively unbiased, can depend on choice of parameters
    * Examples: ConstituentSubtractor, SoftKiller, PUPPI, ...

for a complete review see G. Soyez, "Pile-up mitigation at the LHC: a theorist's view (2018)

summary of lecture 1

* jet definitions and jet algorithm
* the generalised k_t family
* the issue of pile-up

SHIFTs in p_T



① average pert. loss of a quark initiated jet

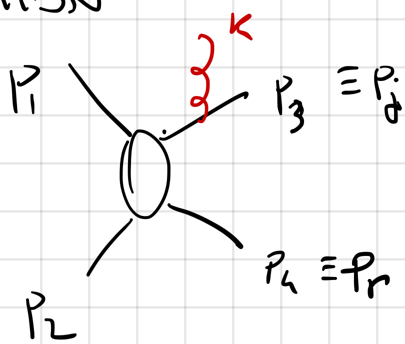
$$\langle \delta p_T \rangle_{pert} = \int \frac{d\vartheta^2}{\vartheta^2} \int dz \, P_{gq}(z) \left[p_T \max(z, 1-z) - p_T \right] \Theta(\vartheta - R)$$

$$= -C_F \ln \frac{1}{R^2} p_T \int dz \, \frac{\alpha_s}{\pi} \left(\frac{1}{z} - 1 + \frac{z}{2} \right) \min(z, 1-z)$$

$$= C_F \ln R^2 \, p_T \frac{\alpha_s}{\pi} \left(\ln 2 - \frac{3}{16} \right)$$

$$\frac{\langle \delta p_T \rangle_{pert}}{p_T} = \frac{\alpha_s C_F}{\pi} \ln R \left(2\ln 2 - \frac{3}{8} \right) \quad p_T \text{ loss.}$$

② HADRONISATION



$$p_1 = \frac{\sqrt{s}}{2}(1,0,0,1)$$

$$p_2 = \frac{\sqrt{s}}{2}(1,0,0,-1)$$

$$p_3 = p_T(1,1,0,0)$$

$$p_4 = p_T(1,-1,0,0)$$

$$p_T = \sqrt{s}/2$$

k is recombined in the jet

$$p_j = \left(\sqrt{p_T^2 + m_j^2}, p_T, 0, 0 \right)$$

$$p_r = (p_T, -p_T, 0, 0)$$

$$\sqrt{p_T^2 + m_j^2} + p_T = \sqrt{s}$$

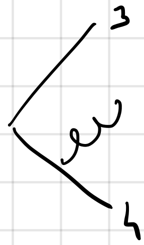
$$\cancel{p_T^2} + m_j^2 = s + \cancel{p_T^2} - 2 p_T \sqrt{s}$$

$$p_T = \frac{s - m_j^2}{2\sqrt{s}} = \frac{\sqrt{s}}{2}\left(1 - \frac{m_j^2}{s}\right)$$

$$\delta p_T^{+} = -\frac{m_j^2}{2\sqrt{s}} = -\frac{p_3 \cdot k}{\sqrt{s}}$$

if k is not recombined, the recoiling system acquires a mass.

$$\delta p_t^{-} = -\frac{p_4 \cdot k}{\sqrt{s}}$$



$$\langle \delta p_T \rangle = C_{34} \int [dk] \frac{p_3 \cdot p_4}{p_3 \cdot k \; p_4 \cdot k} \delta p_T$$

$$\delta p_T = \delta p_\perp^{+} \Theta_{IN} + \delta p_T^{-} \Theta_{OUT} =$$

$$\delta p_T^{-} + (\delta p_T^{+} - \delta p_T^{-}) \Theta_{IN}$$

global term. focus on this

$$[dk] = d^2 k_T \, d\eta \, \frac{d\phi}{2\pi}$$ collinear limit Θ_{IN}

$$\langle \delta p_T \rangle = C_{34} \int dk_T \, k_T \int d\eta \int \frac{d\phi}{2\pi} \frac{p_3 \cdot p_4}{p_3 \cdot k \; p_4 \cdot k} (\delta p_T^{+} - \delta p_T^{-})$$

$\frac{\alpha_s}{2\pi}(C_{34})$

where $k = k_T(\operatorname{ch}\eta, \cos\phi, \sin\phi, \operatorname{sh}\eta)$

we're interested in the NP region of this contribution

$\mathcal{K} < \mu_{NP}$

$$A(\mu_{NP}) = \frac{1}{\pi}\int_0^{\mu_{NP}} d\mathcal{K} \left[\alpha_s^{NP}(\mathcal{K}) - \alpha_s^{PT}(\mathcal{K})\right]$$

↑ physical non-pert. finite coupling

$$P_3 \cdot P_4 = 2P_T^2 = s/2$$

$$k \cdot P_3 = P_T k_T(\operatorname{ch}\eta - \cos\phi)$$

$$k \cdot P_4 = P_T k_T(\operatorname{ch}\eta + \cos\phi)$$

$$\delta P_T^{\pm} = -k_T P_T \frac{\operatorname{ch}\eta \mp \cos\phi}{2P_T^2} = -\frac{k_T}{2}(\operatorname{ch}\eta \pm \cos\phi)$$

$$\mathcal{K}_{i,34}^2 = 2\frac{P_3 \cdot k \; P_4 \cdot k}{P_3 \cdot P_4} = k_T^2\left[\operatorname{ch}^2\eta - \cos^2\phi\right]$$

$$\langle \delta P_T \rangle_4 = C_{34} A \int d\eta \frac{d\phi}{2\pi} \frac{\cos\phi}{[\operatorname{ch}^2\eta - \cos^2\phi]^{3/2}} \Theta(\eta^2 + \phi^2 < R^2)$$

$\eta = r\cos\alpha$

$\phi = r\sin\alpha$

$$d\eta\, d\phi = r\,dr\,d\alpha$$

$$= C_{34} A \int_0^R dr\, r \int_0^{2\pi} \frac{d\alpha}{2\pi} \frac{1 - \frac{1}{2}r^2\sin^2\alpha}{[r^2 + \dots]^{3/2}}$$

$$= C_{34} A \int_0^R dr \frac{r}{r^3} = C_{34} A \left[-\frac{1}{R} + \frac{1}{R_{cut}}\right]$$

singularity cancels against global term.