Part 2: Jet Algorithms

Last time: Clusters of partons $\underset{\uparrow}{\approx}$ Clusters of hadrons

up to Λ_{QCD} effects

Jets are proxies for underlying quarks / gluons
Essential for new physics searches
Fundamentally ambiguous, but opportunity to optimize.

Today: Jet Algorithms as one way to study QCD processes

(Not only way! E.g. event shapes, probabilistic methods)

Offers unique event interpretation.

Generic Jet Algorithm

$$\{p_1, p_2, p_3 \cdots p_n\}_{\text{hadrons}} \Rightarrow \{p_1, p_2 \cdots p_N\}_{\text{jets}}$$

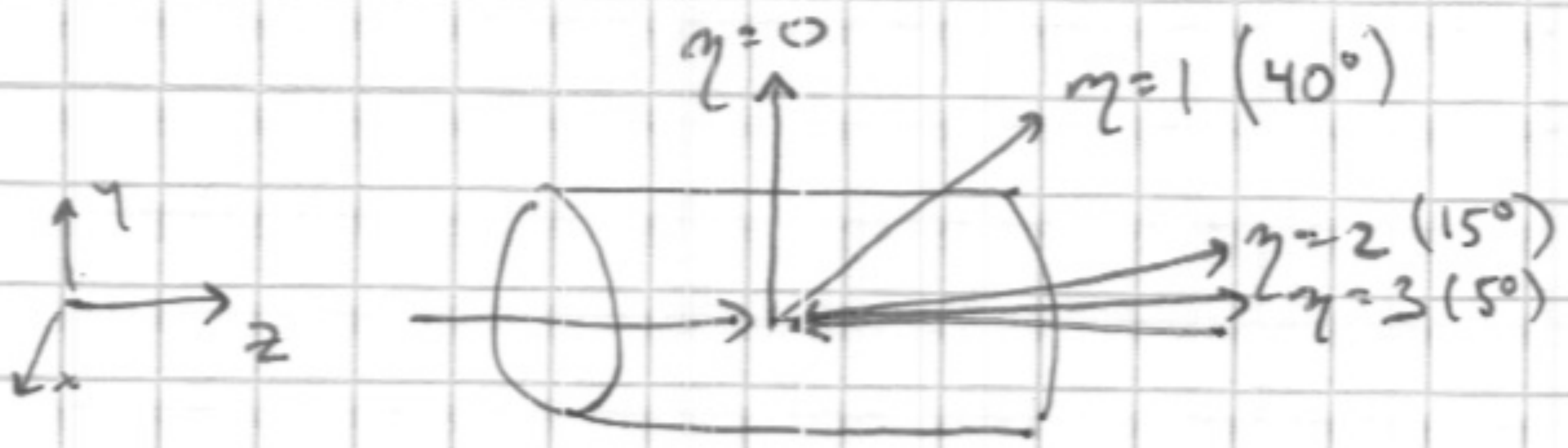
Usually: $p_{jet} = \sum_{i \in jet} p_i$ (E-scheme recombination)

Note: $p_{jet}^2 = \left(\sum_i p_i\right)^2 \geq \sum_i m_i^2$

Jets are full fourvectors with non-zero mass.

Mass has to be included for correct resonance reconstruction. (e.g. $Z' \to$ dijets)

Coordinate system for hadron colliders.



Typically:
longitudinally
boost invariance

$$p_T = \sqrt{p_x^2 + p_y^2} \qquad \phi = \arctan \frac{p_y}{p_x} \qquad \eta = \operatorname{arctanh} \frac{p_z}{|\vec{p}|}$$

$$y = \frac{1}{2} \ln \frac{E + p_z}{E - p_z}$$

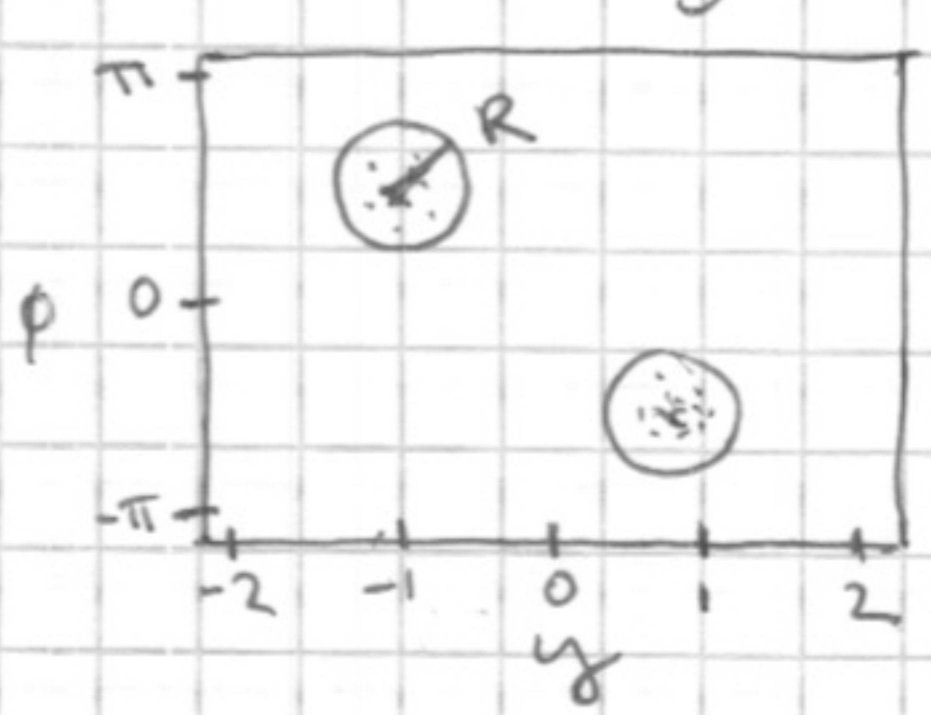
true rapidity

pseudorapidity
(only boost invariant difference for massless particles)

$$= -\ln \tan \frac{\theta}{2}$$

(longitudinal)
Boost-invariant distances $\quad \Delta R = \sqrt{(\Delta\phi)^2 + (\Delta y)^2}$

Jets from cone algorithms?



Punch holes of radius R.

Immediate challenges:
- Which holes?
- Split / Merge ambiguity

Cone algorithms are intuitive: flow of energy in a given direction within a given angle

Subtleties encountered in Tevatron era (e.g. IRC unsafe seeds)

Should probably be revisited (more later)

What is ΔR? (Beyond being just boost invariant)

Standard opening angle: $d\Omega^2 = d\theta^2 + \sin^2\theta \, d\phi^2$

Exercise: Using $\sin\theta = \frac{1}{\cosh\eta}$ for massless particles

show: $d\Omega^2 = \frac{1}{\cosh^2\eta} \underbrace{\left(d\eta^2 + d\phi^2 \right)}_{= dR^2}$

So ΔR is just like opening angle, rescaled by factors of $\cosh\eta$. $(\cosh(\eta=0) = 1)$

For massless particles: $E = p_T \cosh\eta$

$\Rightarrow$ $E\Delta\Omega \approx p_T \Delta R$ (nice thing to remember)

$\Rightarrow$ Jets are smaller in forward (large $|\eta|$) direction.

Q: Why $p_T, \Delta R$ instead of $E, \Delta\Omega$?

$\underbrace{p_T, \Delta R}_{\text{longitudinally boost invariant}}$ $\underbrace{E, \Delta\Omega}_{\text{used extensively in } e^+e^- \text{ colliders}}$

Ok answer: PDFs mean we don't know longitudinal momentum fraction of colliding quark/gluons.

Better answer: Rapidity Plateau: in "minimum bias" collisions, number of particles roughly constant as a function of η.

Q: Why fixed R?

Actually, there is interest in variable R, dynamical R, multiple R...

Right now, $R = 0.4$ anti-k_t jets very standard

↳ ($\sim 22°$ half angle, $\sim 45°$ full angle)

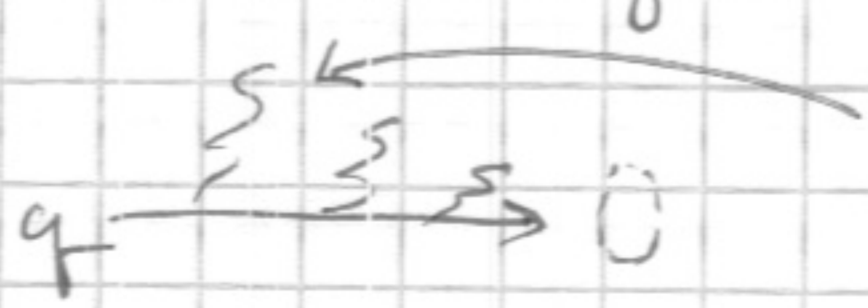
vs. 0.1×0.1 HCAL granularity

Fixed R $\Rightarrow$ roughly uniform jet contamination as a function of η.

Q: Why not very small R?

Perturbative $\log R$ resummation needed

Sensitivity to hadronization, though



"out of cone" corrections.

The way I think of jet algorithms: Optimization Problem

What do you want to optimize? (many choices...)

→ Hard partons should be clustered into jets

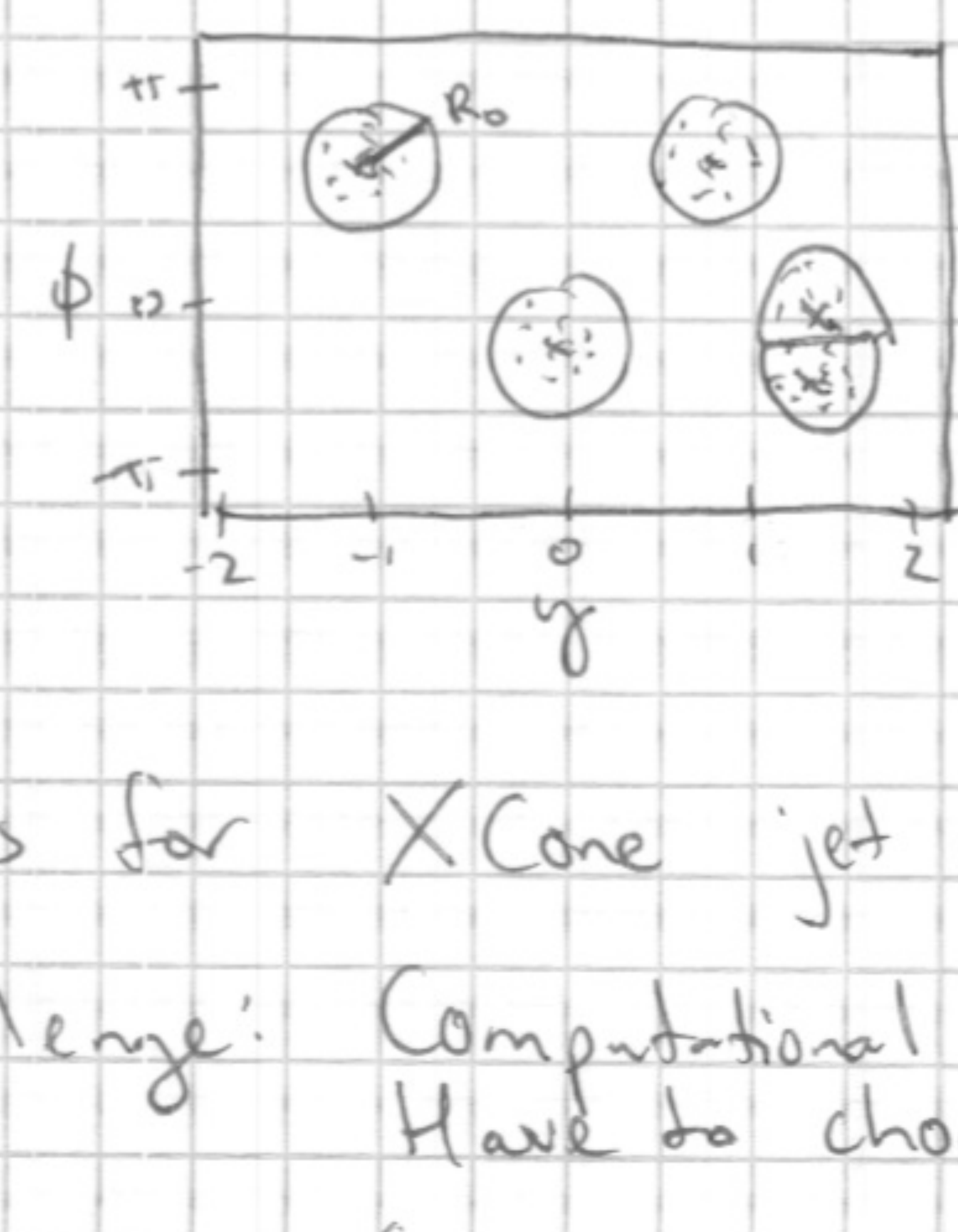
→ Jet axis ≈ jet momentum direction

→ Roughly equal area (for contamination)

All of these can / should be revisited.

E.g. I want exactly N jets with axes: $\hat{n}_1, \hat{n}_2 \cdots \hat{n}_N$

Optimize this:
$$\underbrace{\tau_N}_{\text{N-jettiness penalty function}} = \sum_{i \in \text{event}} \min \Big\{ \underbrace{p_{T_i}}_{\text{not in a jet}}, \underbrace{p_{T_i} \left(\frac{\Delta R_{i,1}}{R_0}\right)^2}_{\text{in first jet}}, \cdots, \underbrace{p_{T_i}\left(\frac{\Delta R_{i,N}}{R_0}\right)^2}_{\text{in N-th jet}} \Big\}$$



Minimize τ_N

⇒ axes that align with hardest (by p_T) clusters

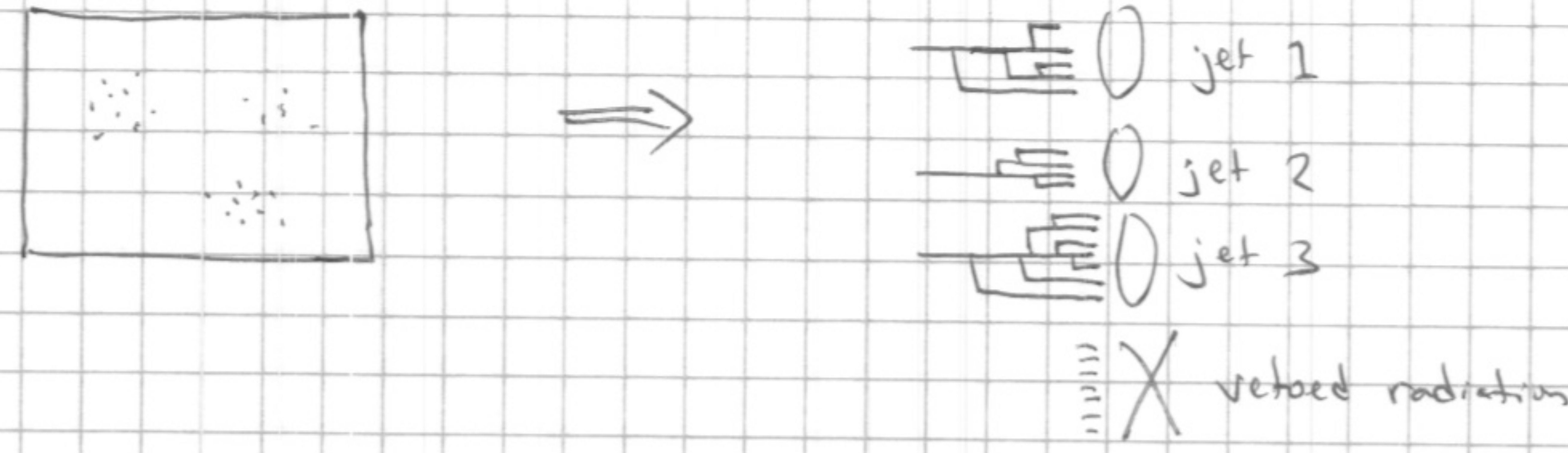
Basis for XCone jet algorithm.

Challenge: Computational expensive to fully optimize
Have to choose N in advance.

(Standard problems for cluster optimization in computer science.)

Workhorse at the LHC: Recursive Jet Algorithms

(a.k.a hierarchical clustering trees)



Instead of global optimization, "local" optimization
↳ in tree-space.

Should I combine two protojets together or
should I declare a protojet finished?

Recursively calculate and take <u>smallest</u> $\underbrace{d_{ij}}_{\text{Merge}}$ $\underbrace{d_i}_{\text{Stop}}$ (Long history, going back to JADE in 1980s)

Nice properties: Works the same on partons / hadrons / calo cells

Computationally very efficient ($n \log n$)

Defined perturbatively to all α_s orders

(Mostly) pathology-free.

Note: $p_{ij} = p_i + p_j$ in E-scheme

Let's make anti-k_t! (Almost every LHC jet is anti-k_t)

Goal 1: Prioritize hardest jets

$$\boxed{d_i = \frac{1}{p_{T_i}^2}}$$

First remove jets of highest p_T

Goal 2: Jets should soak up all radiation within radius R_0.

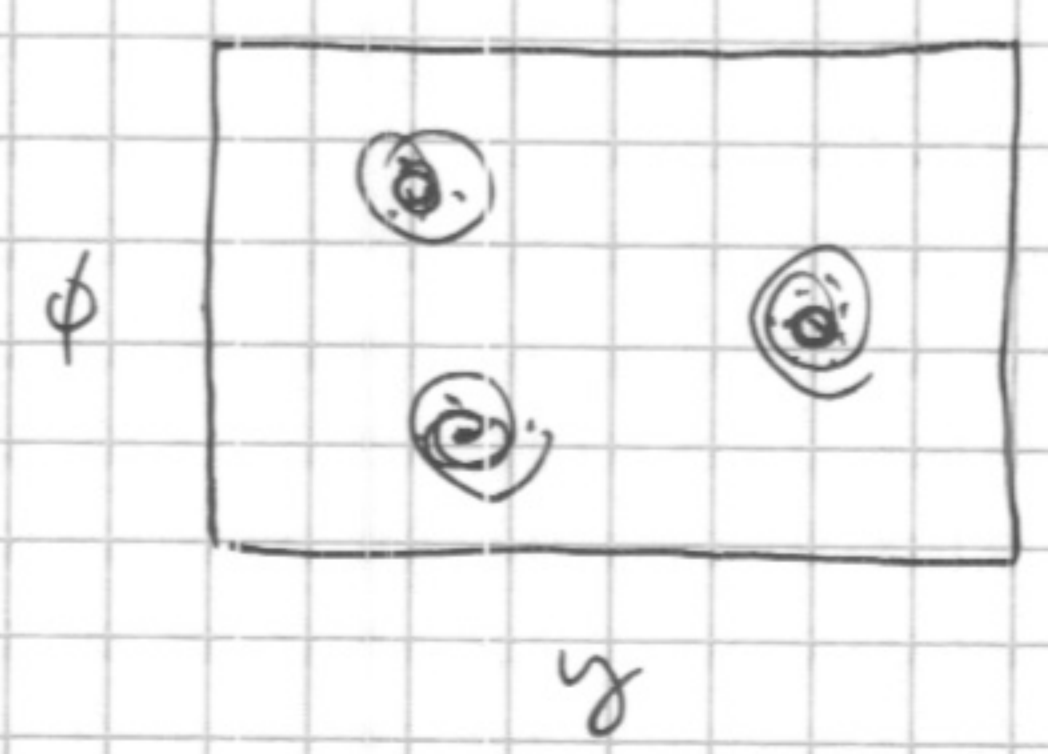
Want $d_{ij} < d_i$ if $R_{ij} < R_0$

$$d_{ij} = \min\{d_i, d_j\}\left(\frac{R_{ij}}{R_0}\right)^2$$

$$\boxed{d_{ij} = \frac{1}{\max\{p_{T_i}^2, p_{T_j}^2\}}\left(\frac{R_{ij}}{R_0}\right)^2}$$

Done! That's anti-k_T

Anti-k_t: Builds jets from hard to soft



Roughly: take hardest particle and merge with nearest neighbor until R_0

For me: A revelation when I read this paper in 2008 (Why did it take so long?)

Different goals?

k_t $\quad d_{ij} = \min\{p_{T_i}^2, p_{T_j}^2\}\left(\frac{R_{ij}}{R_0}\right)^2 \quad d_i = p_{T_i}^2$

Cluster according to singularity structure of QCD

C/A $\quad d_{ij} = \frac{R_{ij}^2}{R_0^2} \quad d_i = 1$

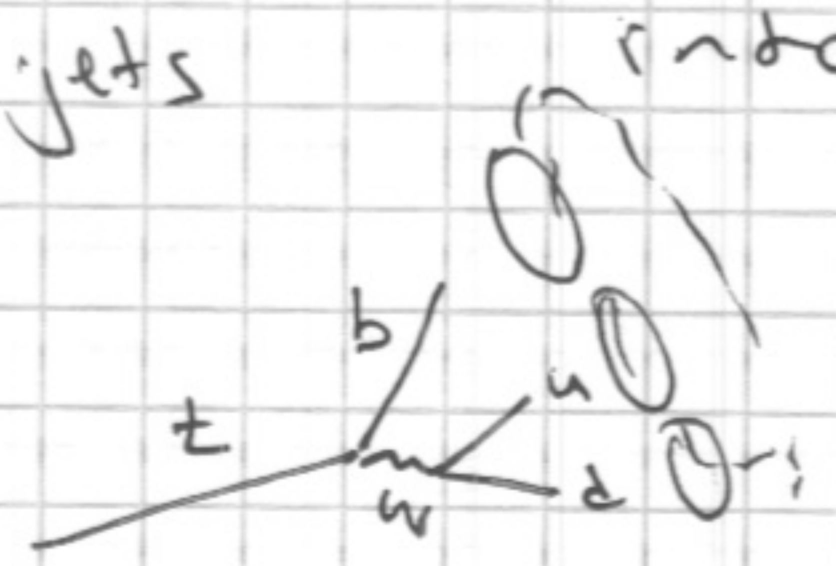
Cluster by angle alone

Both of these yield "amoeba jets" with irregular shapes

Still, very important for jet substructure.

Choice of jet algorithm depends on physics of interest!

E.g. Anti-k_T $R=1.0$ jets $\Rightarrow$ reclustered into $\Rightarrow$ 3 k_T subjets

for 

boosted top identification.

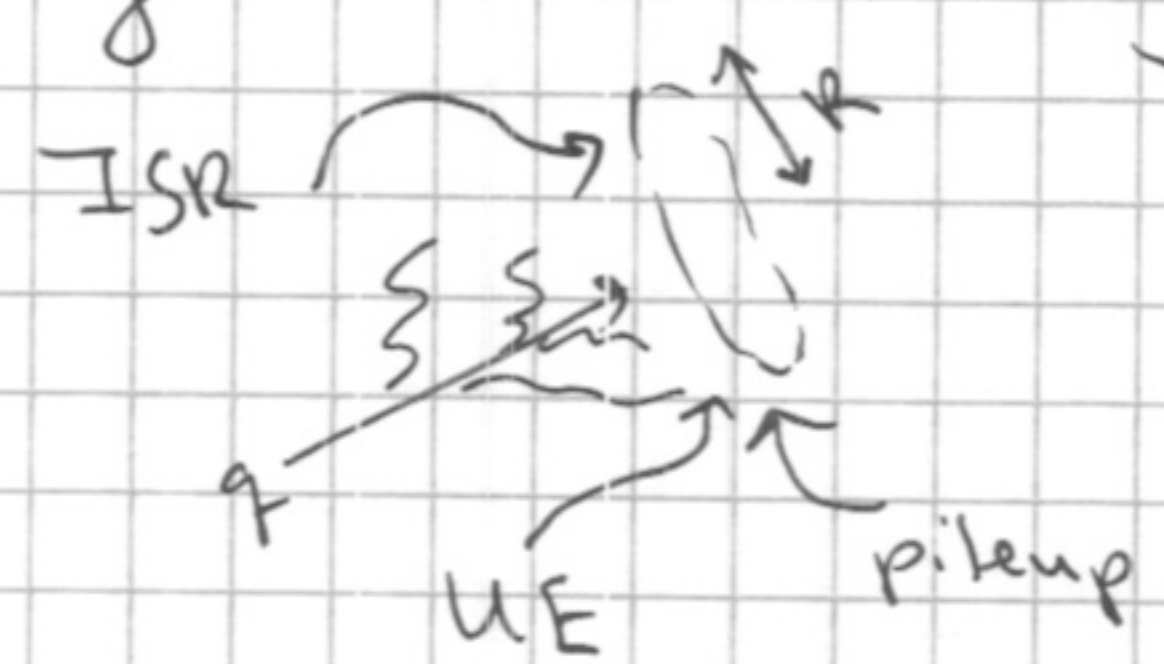
E.g. Anti-k_t $R=0.4$ $\Rightarrow$ Good balance between in-and out-of cone corrections.
Well defined jet area (usually)

Periodically useful to revisit standard jet lore

- → Longitudinal boost invariance?
- → Manifest energy/momentum conservation? ($p_{12} = p_1 + p_2$)
- → Fixed jet radius?
- → Seedless jet algorithms?
- → Unique assignment of hadrons to clusters?
- → Unique jet reconstruction per event?

Your job is to question your elders!

E.g. Well-defined jet area?



Want to subtract contamination from pileup

$$p_{jet}^{corr} = p_{jet} - \rho A_{jet}$$

A_{jet}: jet area

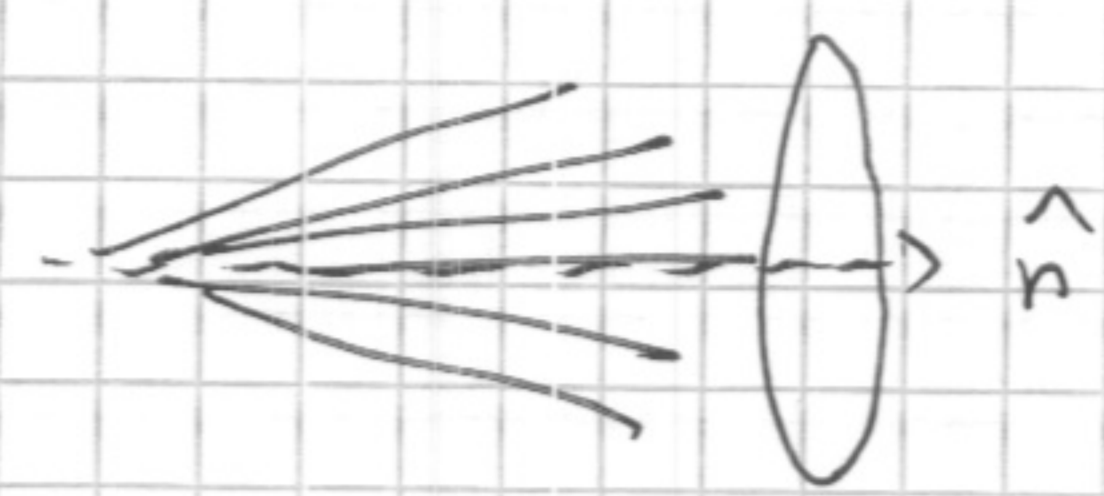
ρ: contamination per unit η, ϕ. (rough 1 GeV per unit per PV)

But why not build an algorithm that refuses to cluster soft particles to begin with?

"Zero Area Jet Algorithms?"

↳ under active investigation

E.g. Jet axis parallel to jet momentum?



$$\hat{n}_{jet} = \frac{\vec{p}_{jet}}{|\vec{p}_{jet}|}$$

In terms of optimization, very natural:

$$\tau_1 = \sum_i p_{T_i} \Big(\underbrace{(\phi_i - \phi_A)^2 + (y_i - y_A)^2}_{\to R_{iA}^2} \Big)$$

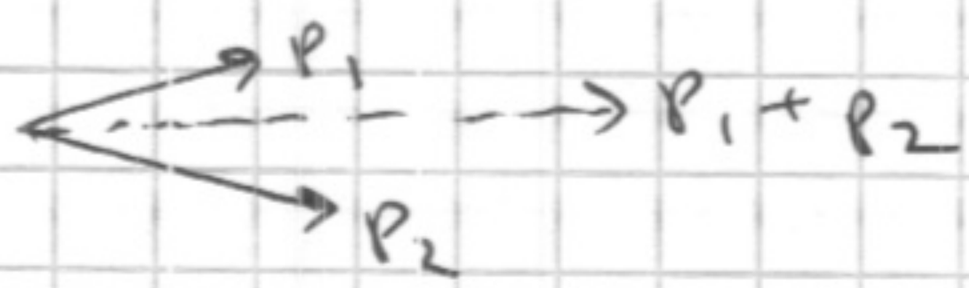
$$\frac{\partial \tau_1}{\partial \phi_A} \overset{?}{=} 0 = \sum_i 2\, p_{T_i} (\phi_i - \phi_A)$$

$\hookrightarrow$ solved for $\phi_A = \dfrac{\sum_i p_{T_i} \phi_i}{\sum_i p_{T_i}}$ mostly the same.

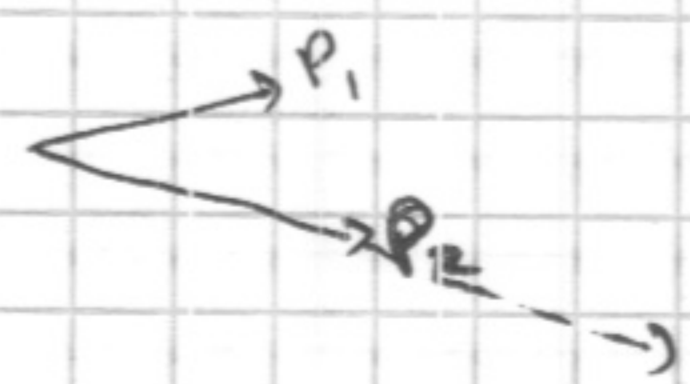
Same as finding "mean"

Why not "median"? $\tau_1 = \sum_i p_{T_i} R_{iA}^1$

In language of clustering:



E-scheme axis ("mean")



$$\frac{\vec{p}_2}{|\vec{p}_2|}\left(|\vec{p}_1| + |\vec{p}_2|\right)$$

Winner-take-all axis ("median")

Like median being robust to outliers,
WTA axis robust to contamination.

One lore that you challenge at your peril!

Infrared / Collinear Safety (more next lecture)

(By the way, I've broken IRC safety with some success. Still, beware!)

Why IRC Safety?

Key success of QCD:

Fixed-order (and resummed) calculations for jet cross sections.

Multi-loop, multi-leg, multi-log. (A whole separate lecture series.)

Famously: Loop diagrams have UV divergences (fix with renormalization)

and IR divergences (cancel against real emission diagrams)

$$\frac{d\sigma}{de} = \sum_N \int d\Phi_N \, \frac{d\sigma}{d\Phi_N} \, \delta\left(e - \hat{e}(\Phi_N)\right)$$

Finite if e is IRC safe observable

N-body phase space

IR divergences at every α_s order

measurement function.

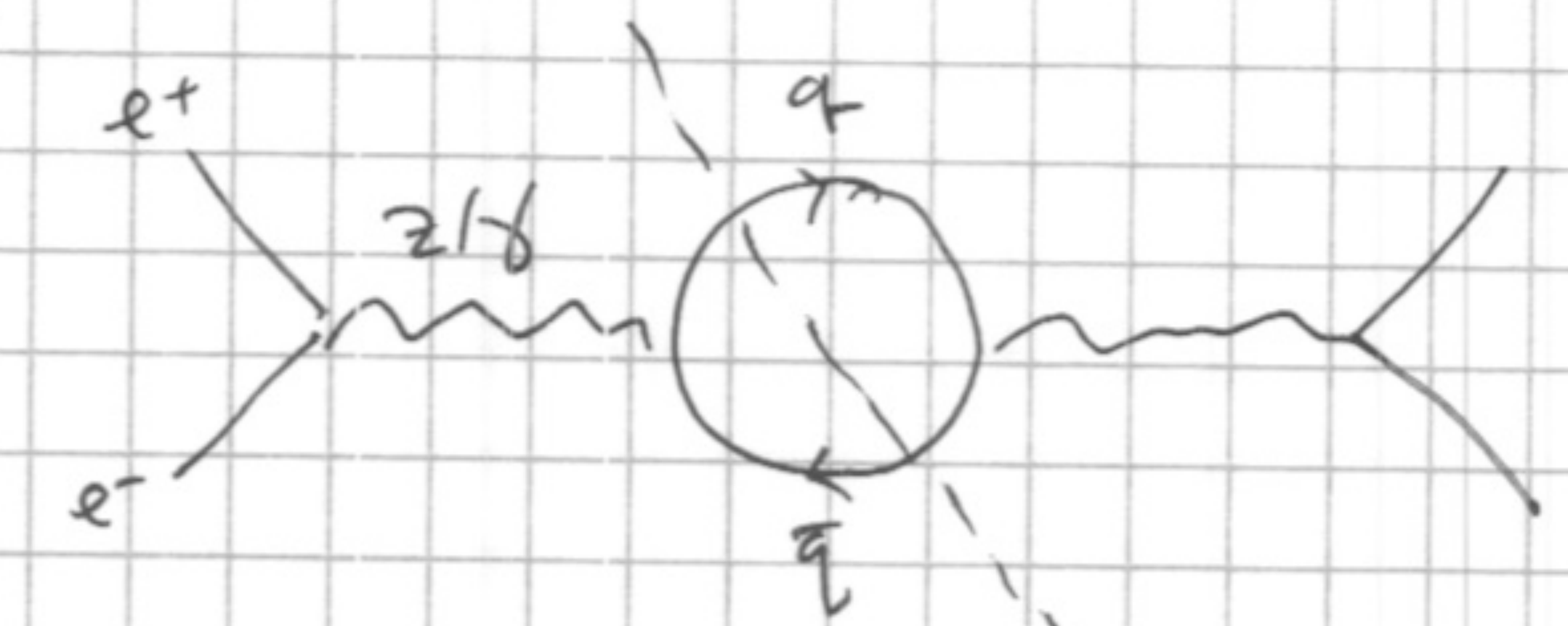
IRC Safety required for sensible perturbative calculations.

↳ well, not quite (e.g. Sudakov safety)

For total cross sections, IRC safety easy to understand by optical theorem.

$$\sigma_{2\to\text{anything}} \Longleftrightarrow \text{Im}\, \mathcal{M}_{\text{forward}}.$$

e.g. $e^+e^- \to$ hadrons $\Longleftrightarrow$ $e^+e^- \to e^+e^-$



$\mathcal{O}(\alpha_s^0)$

Imaginary part from cutting rules.

$\mathcal{O}(\alpha_s^1)$

$e^+e^- \to q\bar{q}g$ (tree)

$e^+e^- \to q\bar{q}$ (one-loop)

singularities have to cancel!

Somewhat counterintuitively:

$2\to 2$ tree-level cross section is good approximation to <u>total</u> cross section (not exclusive 2 jet xsec)

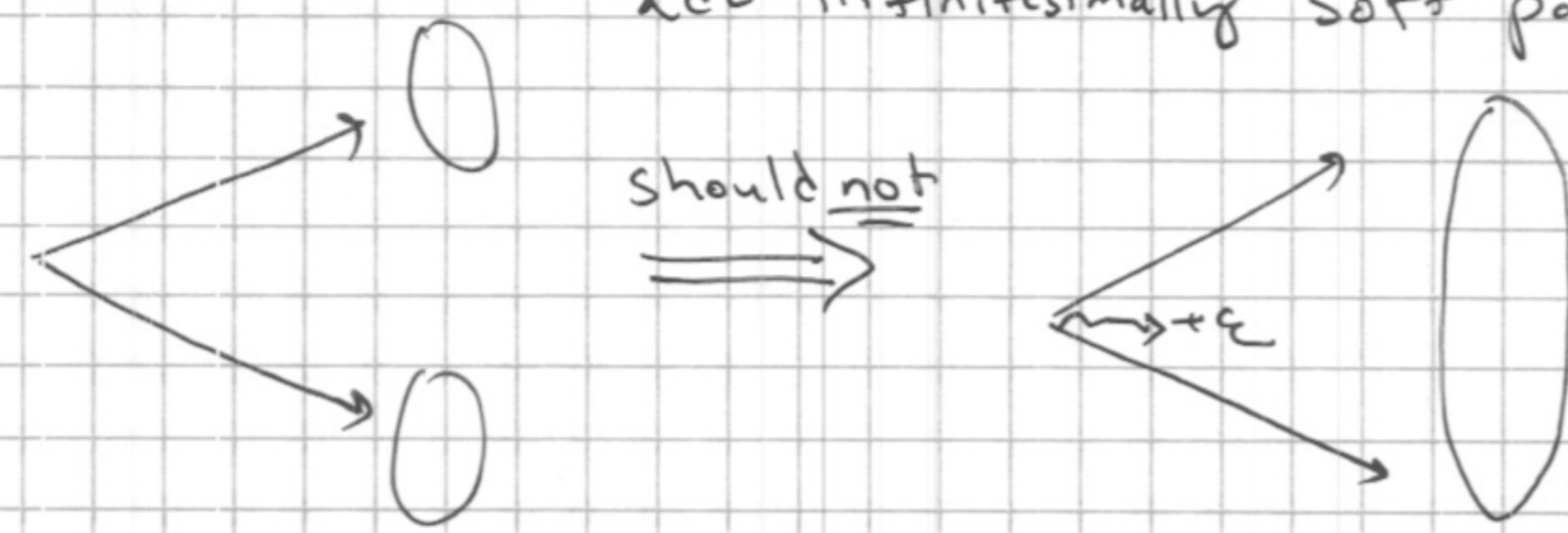
$\underset{\text{loop}}{2\to 2}$ + $\underset{\text{tree}}{2\to 3}$ largely cancel.

For jet cross sections, IRC safety is a bit non-trivial, and people have gotten in trouble over the years.

Heuristic for IRC Safety (almost always works)

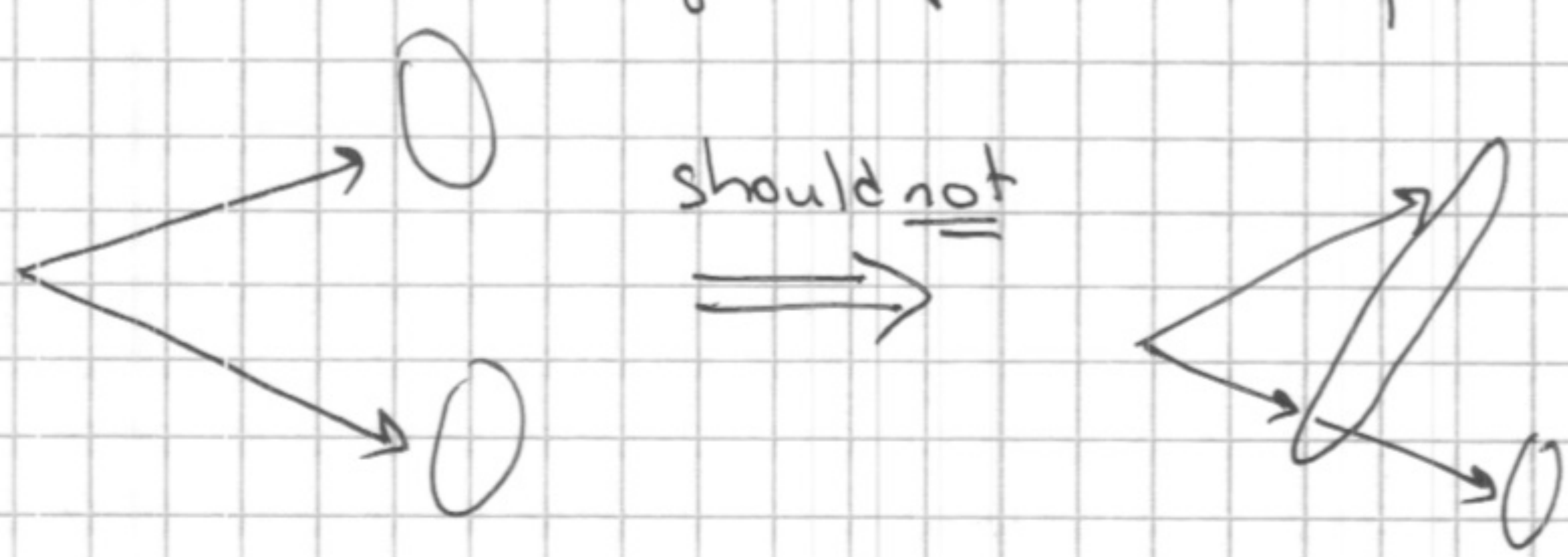
Soft safety: Answer should not change if you add infinitesimally soft particle

E.g.



Collinear safety: Answer should not change if you split a particle exactly in two

E.g.



Among jet afficionados, IRC Safety is the first thing we check about any new idea.

Not safe? Not calculable in fixed-order perturbative QCD!

(Perhaps other strategies available, though)

Is anti-k_t safe?

$$d_{ij} = \frac{1}{\max\{p_{T_i}^2, p_{T_j}^2\}} \frac{R_{ij}^2}{R_0^2} \qquad d_i = \frac{1}{p_{T_i}^2}$$

Soft emission? $d_i^{soft} \to \infty$ (no effect)

d_{ij}^{soft} could be important!
but infinitesimally soft particle
does not change jet p_T

Depends on question!
Anti-k_t jet p_T? Soft safe
Anti-k_t jet multiplicity of constituents? Soft unsafe.

Collinear split? $d_i^{collinear}$ changes!
But $d_{ij}^{collinear} = 0$, so first thing you
do is put split back together.

So yes, anti-k_t is IRC safe.
NNLO jet cross sections being actively developed.

Next time: What exactly are soft/collinear
singularities, and can we use
them to understand quark/gluon tagging.