Jesse Thaler

①

"Two Lectures on Jet Substructure"

PiTP Summer School

July 23-24, 2013

Outline:

- → Overview of Jet Substructure
- → Quark vs. Gluon Discrimination
- → Boosted Objects
- → Jet Grooming

Part 1: Overview of Jet Substructure

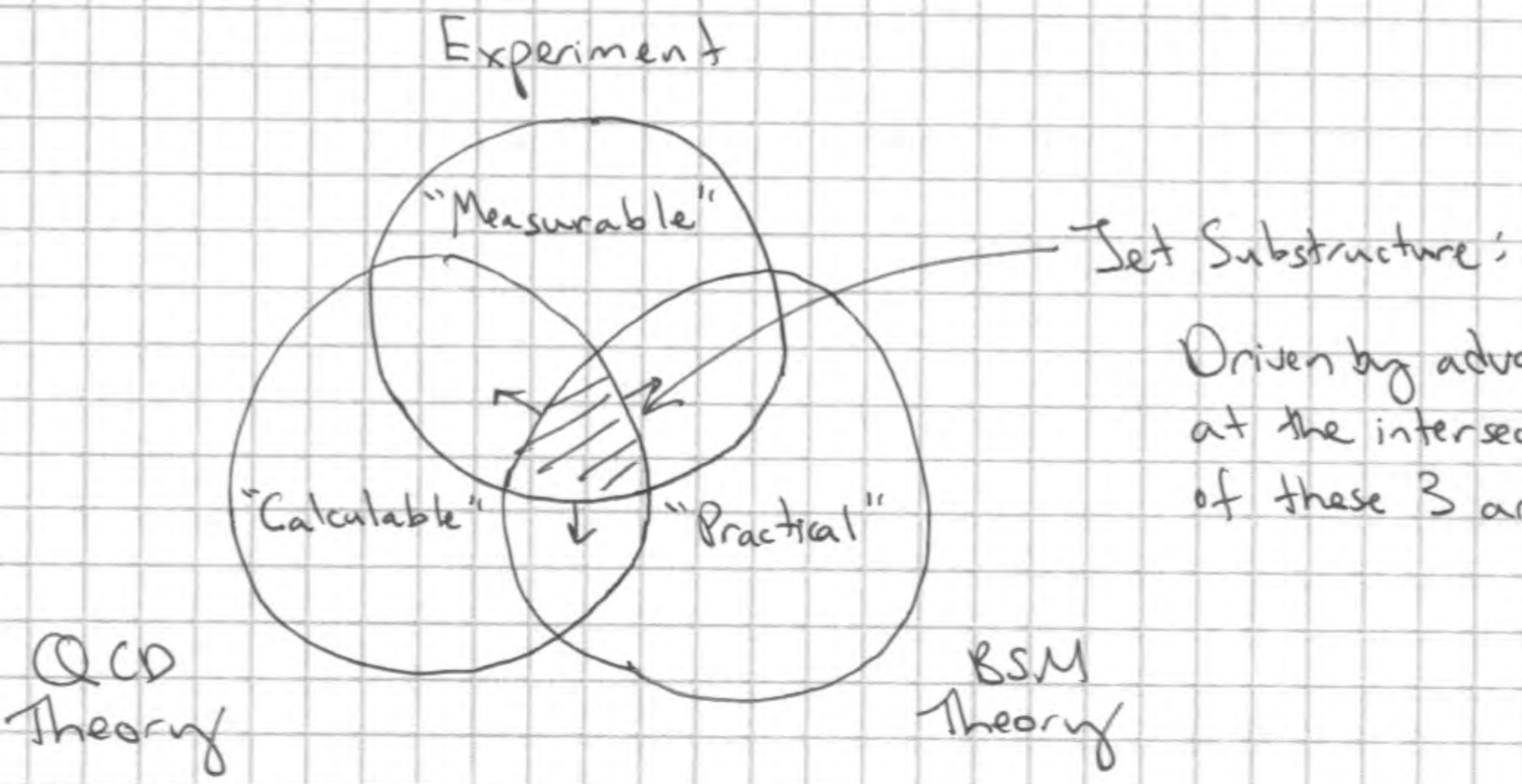
A Renaissance in Jet Physics!

→ First jet algorithm: 1970s (Sterman-Weinberg)

→ First jet algorithm that
experimentalists & theorists
could agree on : 2008 (anti-k_T)

For me, 5 amazing years learning about structure of QCD, relevance for BSM physics, realities of experimental methods.

Convergence of 3 communities



Driven by advances at the intersection of these 3 areas.

Experiment: Fantastic Performance of LHC
& Granularity of ATLAS/CMS

Compared to Tevatron:
$\approx \times 3.5-7$ more energy
$\approx \times 10-20$ more luminosity
$\approx \times 5$ better segmentation

Ability to Resolve Individual Hadrons!

Increased Sophistication for Jet Studies.

QCD Theory: Automated LO (& NLO) Jet Cross Sections

Sophisticated Monte Carlo (ME/PS matching)

New Approaches / Applications of
Factorization & Resummation.

Some Control of Non-perturbative Effects.

BSM Theory: Realization of Importance of Boosted Regime

$$\sqrt{\hat{s}} \gg m_{top}, m_{W/Z}, m_H, m_{\text{RPV gluino}} \ldots$$

Novel Jet Observables Customized
for New Physics Searches

New Approaches to Cascade Decays
& Multi-Jet Final States

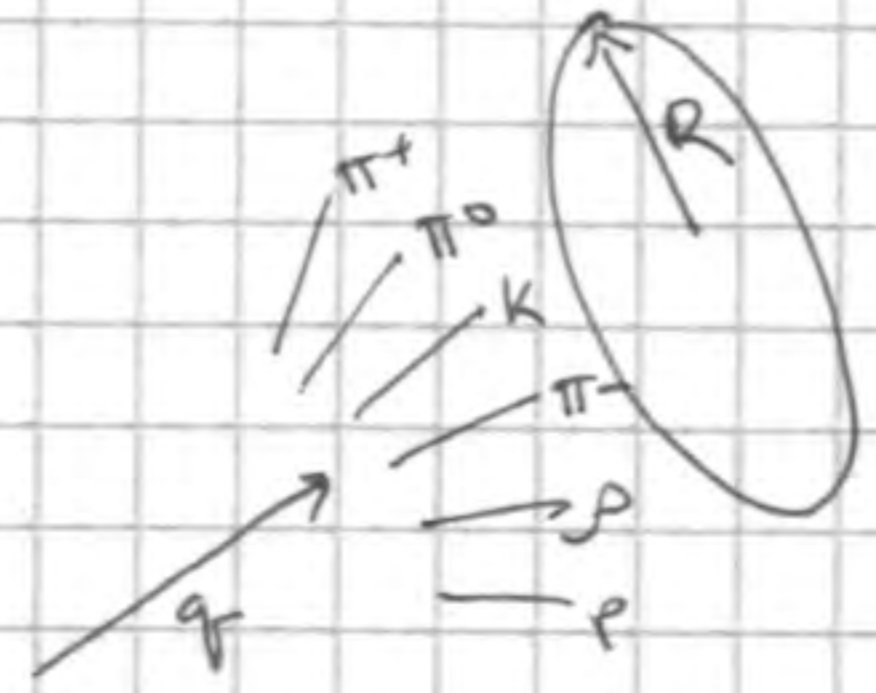
These Lectures: 3 examples of Jet Substructure in Action.

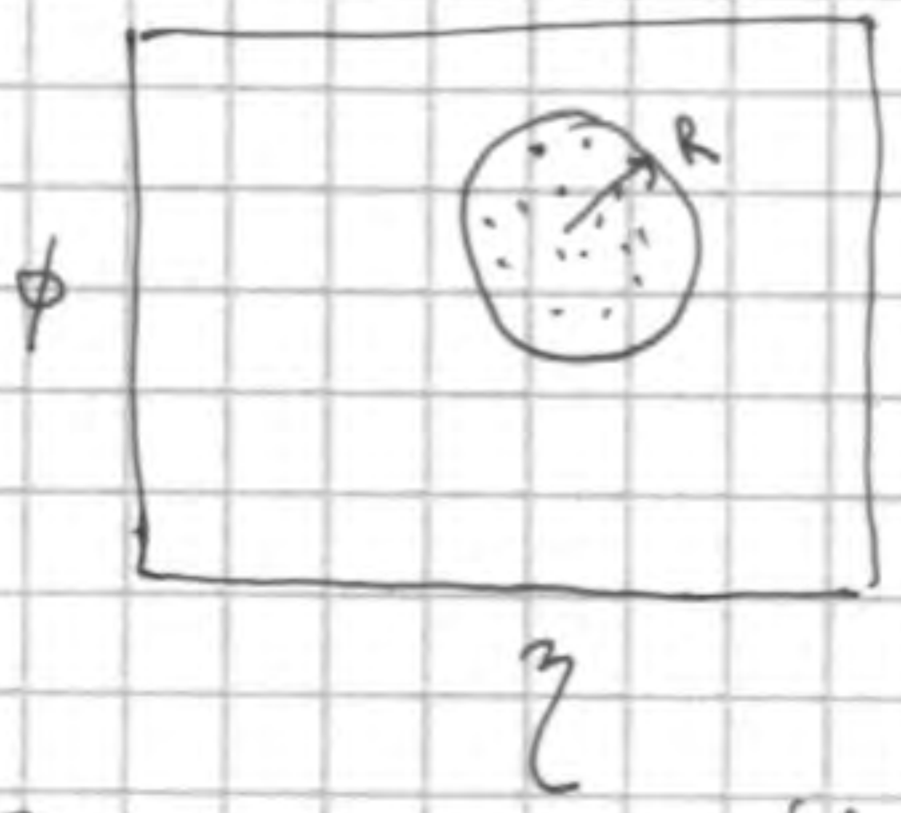
- → Quark vs. Gluon
- → Boosted W/Z/H/top
- → Jet Grooming (time permitting)

Apologies: I will be a bit biased toward my own work, but I'll try to be pedagogical.

Crucial Topic Not Covered: Jet Algorithms

For my purposes: "A jet is a collimated spray of hadrons with a jet radius R."





Ignoring Subtlety that quarks/gluons have color charge, but bound hadrons are color singlets.

(Fundamental Ambiguity in Defining Jets, typically shows up at $\mathcal{O}\left(\frac{\Lambda_{QCD}}{p_T^{jet}}\right)$.)

Ignoring subtle effects of jet boundaries, jet splitting/merging.

(As a community, we've defaulted to anti-k_T to address these issues.)

Part 2: Quark vs. Gluon Jets

A jet is a jet is a jet?

$$\text{Light Quarks} \left\{ \begin{matrix} u \\ \bar{u} \\ d \\ \bar{d} \\ s \\ \bar{s} \end{matrix} \right. \Longrightarrow \begin{matrix} \text{Indistinguishable unless} \\ \text{some kind of jet charge} \\ \text{information is used} \end{matrix}$$

$$\text{Heavy Quarks} \left\{ \begin{matrix} c \\ \bar{c} \\ b \\ \bar{b} \end{matrix} \right. \Longrightarrow \begin{matrix} \text{Can tag because of lifetime/decays} \\ \text{of resulting } D \text{ and } B \text{ mesons} \end{matrix}$$

$$\text{Top Quark} \left\{ \begin{matrix} t \\ \bar{t} \end{matrix} \right. \Longrightarrow \begin{matrix} \text{Totally different beast,} \\ \text{see next lecture. } (t \to bW) \end{matrix}$$

$$\text{Gluon} \left\{ g \right. \Longrightarrow \begin{matrix} \text{Different Color Charge} \\ \text{from Quarks. Should give} \\ \text{rise to different looking jets.} \end{matrix}$$

This lecture: Build up observable sensitive to C_F (quarks) vs. C_A (gluon)

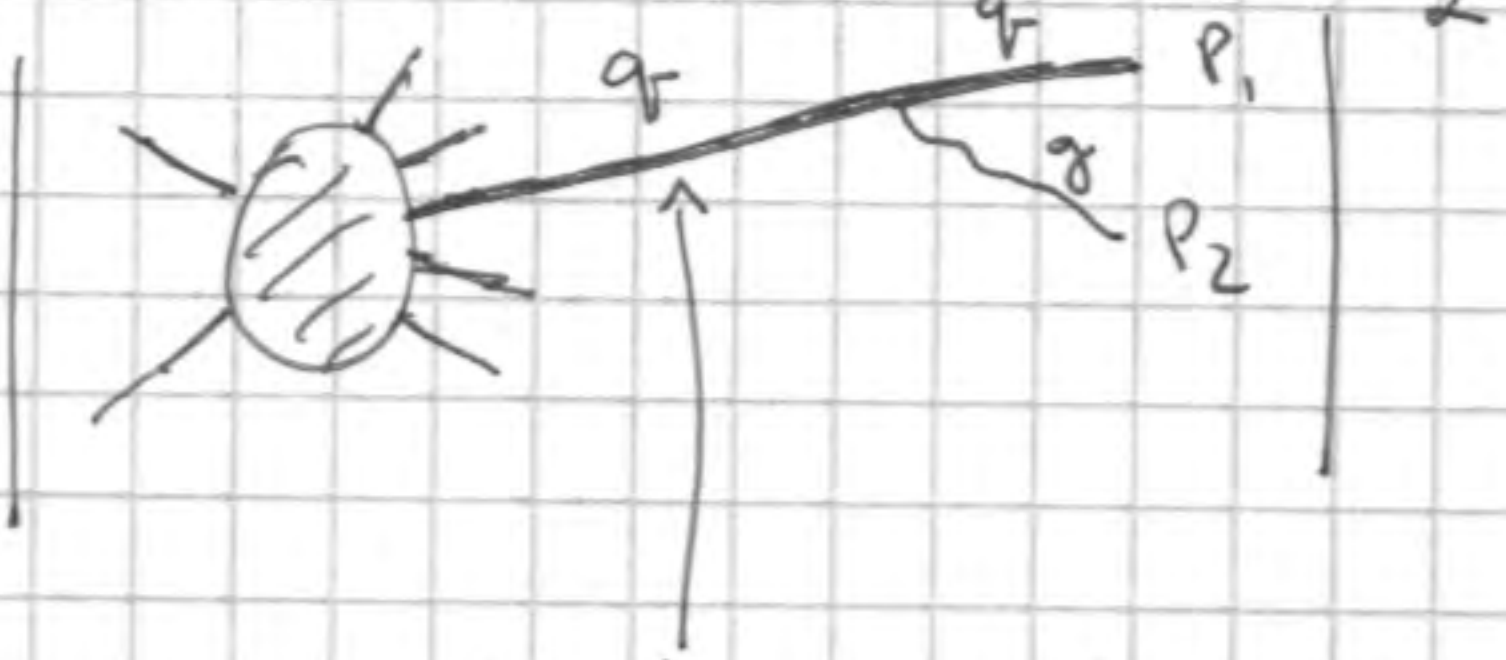
Punchline: In the "strongly-ordered" limit

Quark efficiency: x

Gluon mistag: $x^{C_A/C_F} = x^{9/4}$

First, why do jets form?

1) Soft-Collinear Singularities in QCD



singular propagator when this goes on shell

$p = p_1 + p_2$ with $p_1^2 = p_2^2 = 0$

Tendancy for jets to form

When is $p^2 = 0$? when $\vec{p}_1 \parallel \vec{p}_2$ (Collinear Limit)

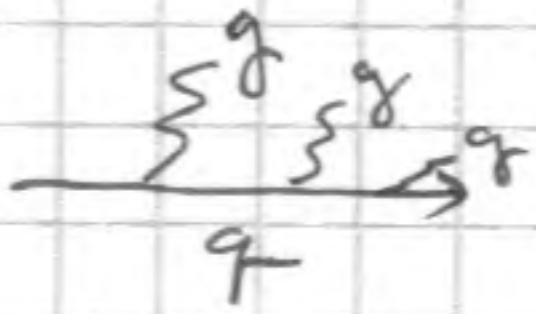
when $|\vec{p}_2| \to 0$ (Soft Limit)

Annoyance. Sensitivity to precise jet definition

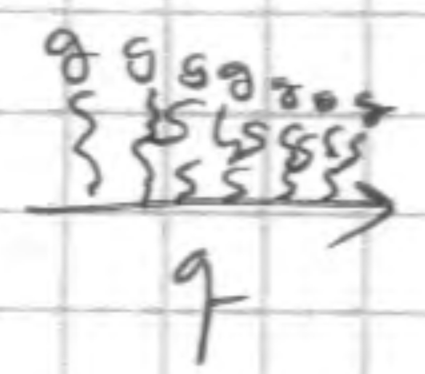
2) At sufficiently high energies, α_s is small.

$$\alpha_s(m_Z) \simeq 0.12$$

Otherwise, many soft-collinear emissions start to look like ball (instead of cone) of radiation.

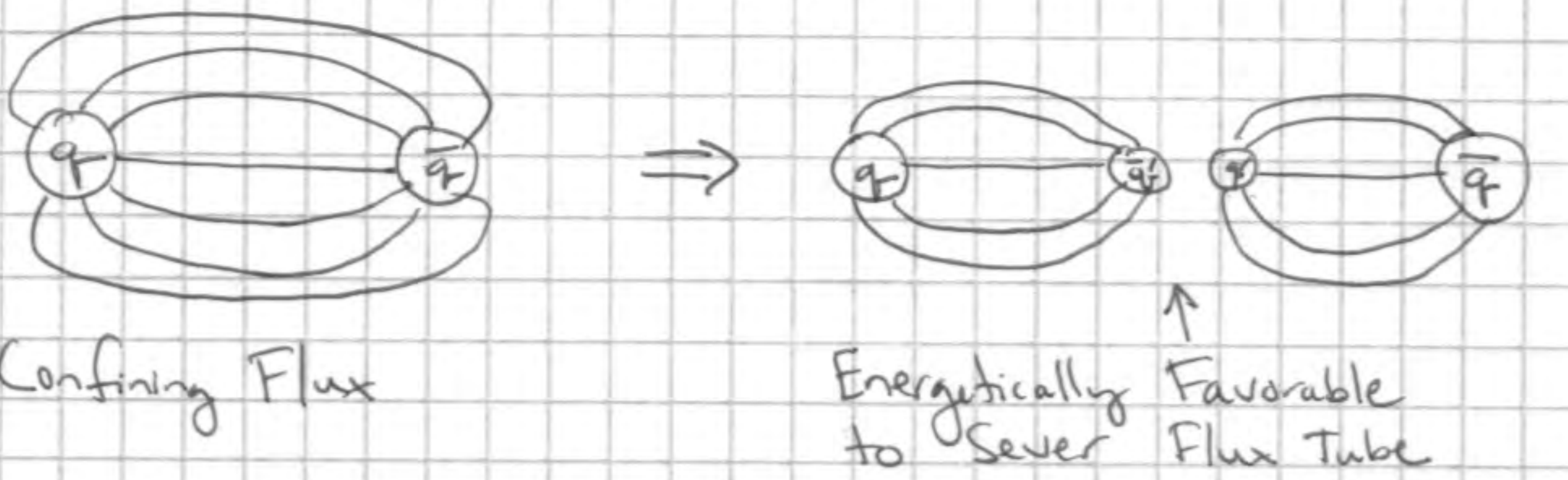


vs.



3) Color "strings" break

$\Rightarrow$ Direction of $q/g \approx$ Direction of Jet



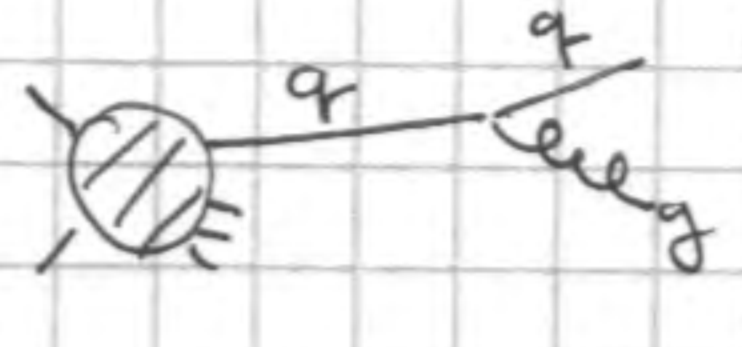
Otherwise, you would get excited hadron states instead of jets.

QCD in the Soft-Collinear Limit

Energy Flow of Partons $\approx$ Energy Flow of Hadrons

Allows us to study jets at partonic level (up to Λ_{QCD} non-perturbative effects)

Key Exercise: Show that (in soft $\underline{\underline{\&}}$ collinear limit)

$\sum_{\text{polar}} \int d\Pi_{n+1} \Big|$  $\Big|^2$ with q and g in the same jet

$\approx \sum_{\text{polar}} \int d\Pi_n \Big|$  $\Big|^2$

$$\times \int_0^1 dz \int_0^R d\theta \; \frac{2\alpha_s}{\pi} C_F \underset{\uparrow}{\frac{1}{z}} \underset{\uparrow}{\frac{1}{\theta}}$$

Swap $C_F \to C_A$ for gluon jets

soft / collinear singularities

Energy Fraction $\quad z = \frac{E_{\text{gluon}}}{E_{\text{jet}}} \quad$ $z \to 0$ soft limit

Splitting Angle $\quad \theta = \theta_{qg} \quad$ $\theta \to 0$ collinear limit

Color Factors:

$SU(N) \longrightarrow SU(3)$

Quarks: $\sum_a t^a_{ij} t^a_{jk} = C_F \delta_{ik} \qquad C_F = \frac{N^2-1}{2N} \longrightarrow 4/3$

Gluons: $\sum_{a,b} f^{abc} f^{abd} = C_A \delta^{cd} \qquad C_A = N \longrightarrow 3$

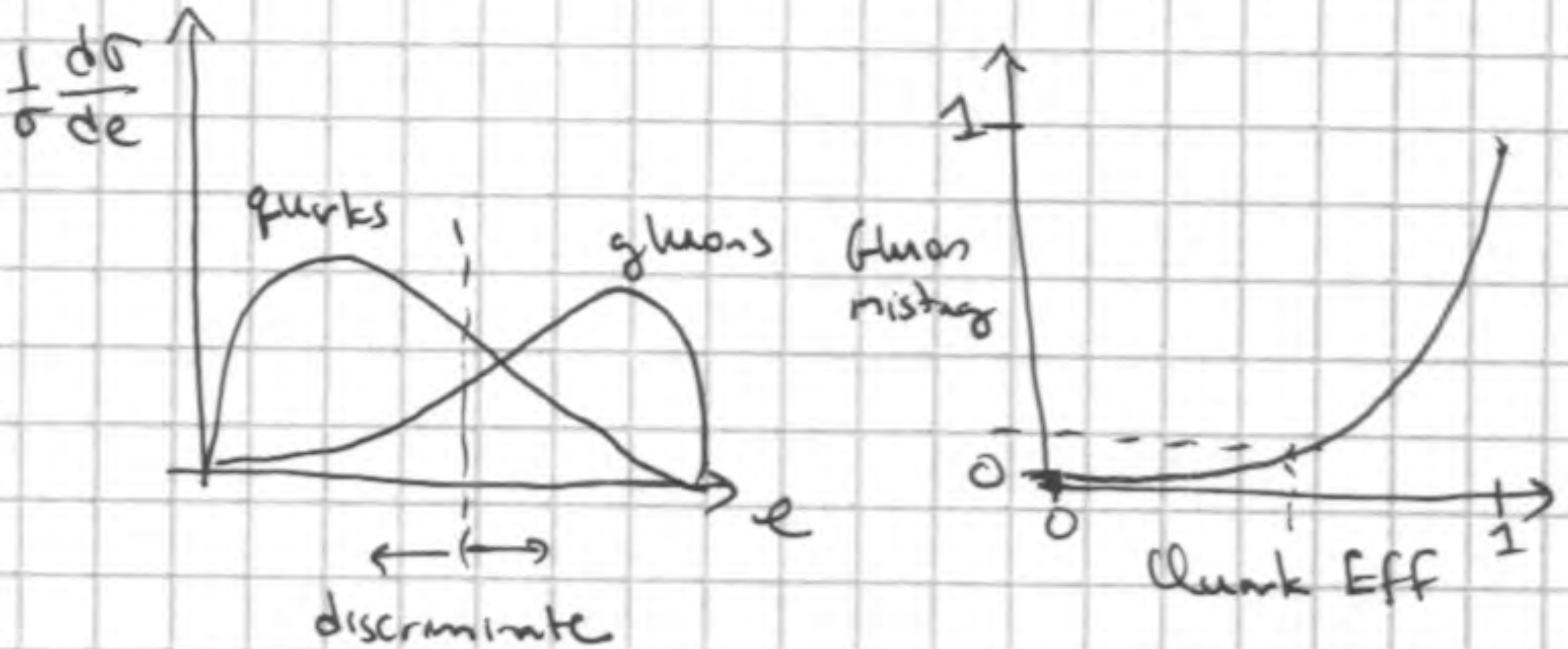
(In just collinear limit, AP splitting functions
In just soft limit, antenna functions.
Above formula for soft <u>and</u> collinear limit only.)

Quark vs. Gluon Jets?

$C_A > C_F$, so gluon jets should be "fatter"

Need observable sensitive to this difference.

Goal:



(Ideally, <u>predict</u> distributions in QCD,
<u>validate</u> in Monte Carlo,
<u>test</u> in LHC data.)

Pause for a moment: Why do we care?

Experiment: Different calibrations for different kinds of jets.

QCD Theory: Build observables sensitive to color structure of QCD

BSM Theory: Many BSM signals are quark-rich, while backgrounds are gluon-rich.

Crucial Feature of an Observable
(if you want to calculate in perturbative QCD)

Infrared | Collinear Safety: (IRC safety)

Observable should be insensitive to infinitely soft or infinitely collinear radiation.

Why?

Technically, needed so IR divergences cancel between real and virtual diagrams. (KLN theorem)

Physically, very soft & collinear radiation controlled by Λ_{QCD}, and you want your observable to be as insensitive as possible to nonperturbative effects.

IRC Safe Observables: Accurately described in partonic language.

Example: Energy-Energy Correlation Function (2013)

($p_T \to E$, $R \to \theta$ for simplicity)

$$C_1^{(\beta)} = \frac{\sum_{i \neq j} E_i E_j \theta_{ij}^{\beta}}{\left(\sum_i E_i\right)^2}$$

$\beta > 0$

Sum over particles in a single jet

Soft safe? Yes, C_1 doesn't change if $E_i \to 0$

Collinear safe? Yes, C_1 is additive, so $E \to E_1 + E_2$ has no effect (for $\beta > 0$)

Q: What is quark/gluon discrimination for C_1?

Hard to answer in general, but we'll go to "strongly-ordered" limit (sometimes called "leading log" or "double log" limit)

Hard quark/gluon emits hierarchical pattern of soft/collinear emissions

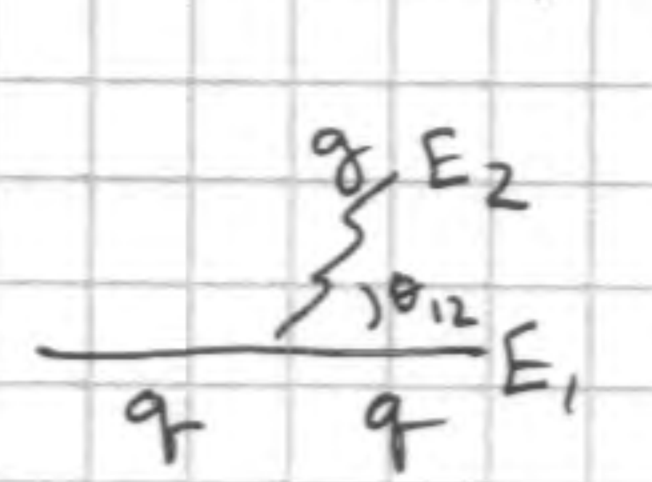
$$P_{\text{emit}}(z,\theta)\,dz\,d\theta = \frac{2\alpha_s}{\pi} C_{F,A} \frac{dz}{z}\frac{d\theta}{\theta}$$

Uniform emissions in $(\log 1/\theta, \log 1/z)$ plane.

Strongly-ordered: Observable dominated by hardest emission(s)

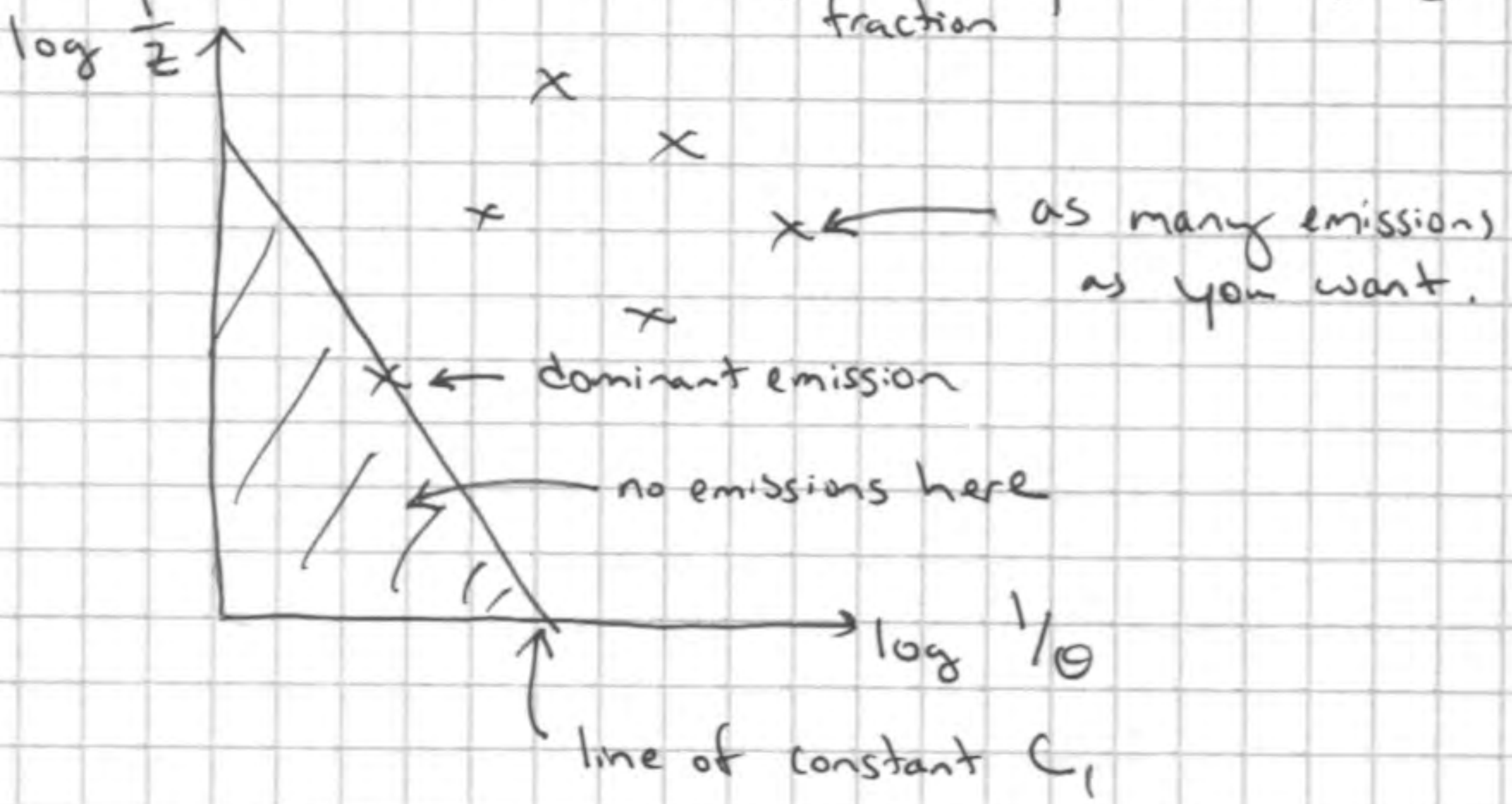
⑫

What is C_1 in strongly-ordered limit?



$$C_1 = \frac{E_1 E_2 \theta_{12}^{\beta}}{(E_1+E_2)^2} \qquad E_2 \ll E_1 \qquad z = \frac{E_2}{E_1+E_2}$$

$$\approx \underset{\substack{\uparrow \\ \text{gluon energy fraction}}}{z}\ \underset{\substack{\uparrow \\ \text{quark/gluon splitting angle}}}{\theta^{\beta}}$$



$$\log \frac{1}{C_1} = \log \frac{1}{z} + \beta \log \frac{1}{\theta}$$

Key Exercise: Probability to get a value of C_1 less than C_1^{max}:

$$\Sigma_q(C_1^{max}) = e^{-\frac{2\alpha_s}{\pi} C_F \left(\text{area under } C_1^{max} \text{ curve}\right)}$$

$$\left(= \underset{\text{(no emissions}}{1} - \underset{\mathcal{O}(\alpha_s)}{◺} + \underset{\mathcal{O}(\alpha_s^2)}{\tfrac{1}{2}\,◺^2} + \cdots \right)$$

$$\Sigma_q\left(C_1^{max}\right) = e^{-\frac{\alpha_s}{\pi}\frac{C_F}{\beta}\log^2\frac{R^\beta}{C_1^{max}}}$$

$$\Sigma_g\left(C_1^{max}\right) = e^{-\frac{\alpha_s}{\pi}\frac{C_A}{\beta}\log^2\frac{R^\beta}{C_1^{max}}}$$

← called a Sudakov form factor

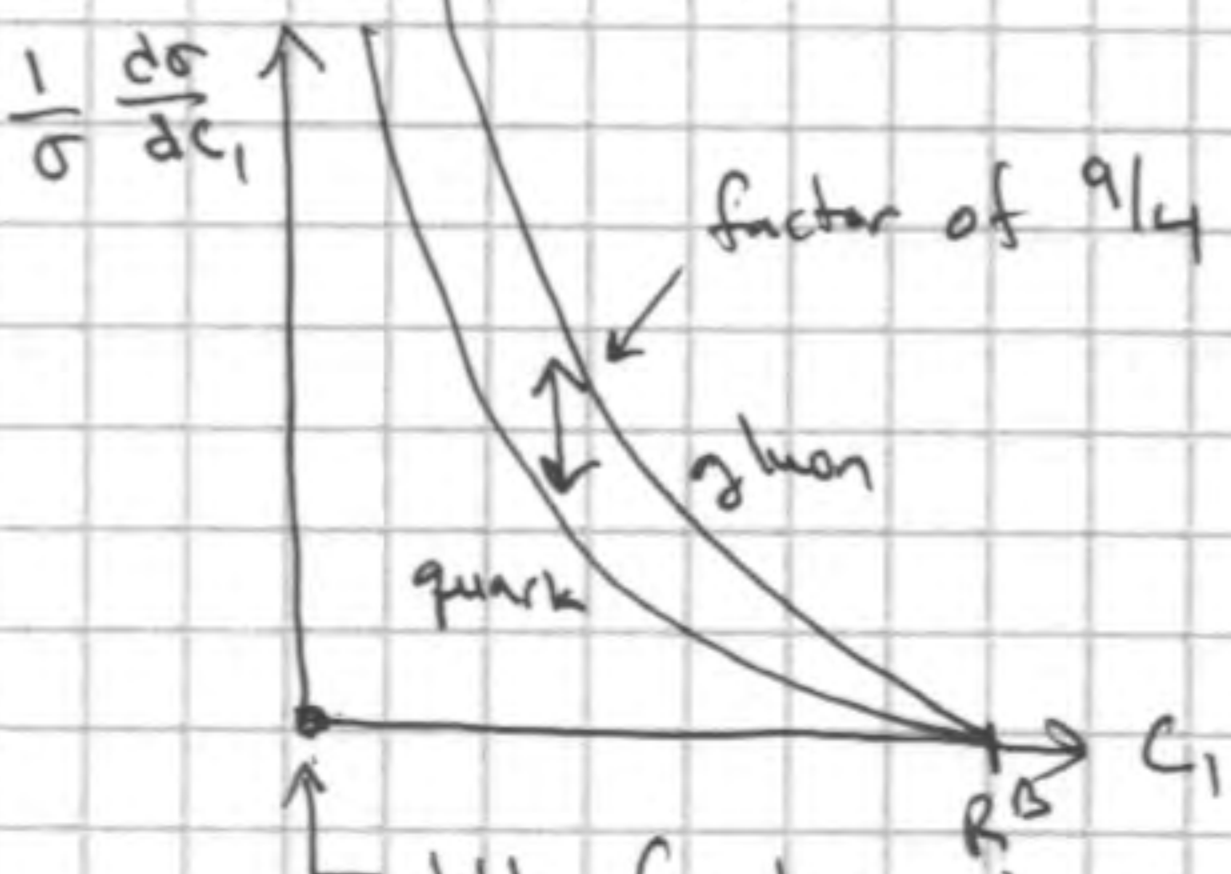
Time to interpret this result!

1) What is the cross section?

$$\frac{1}{\sigma}\frac{\partial\sigma_q}{\partial C_1} = \frac{\partial}{\partial C_1}\Sigma_q(C_1)$$

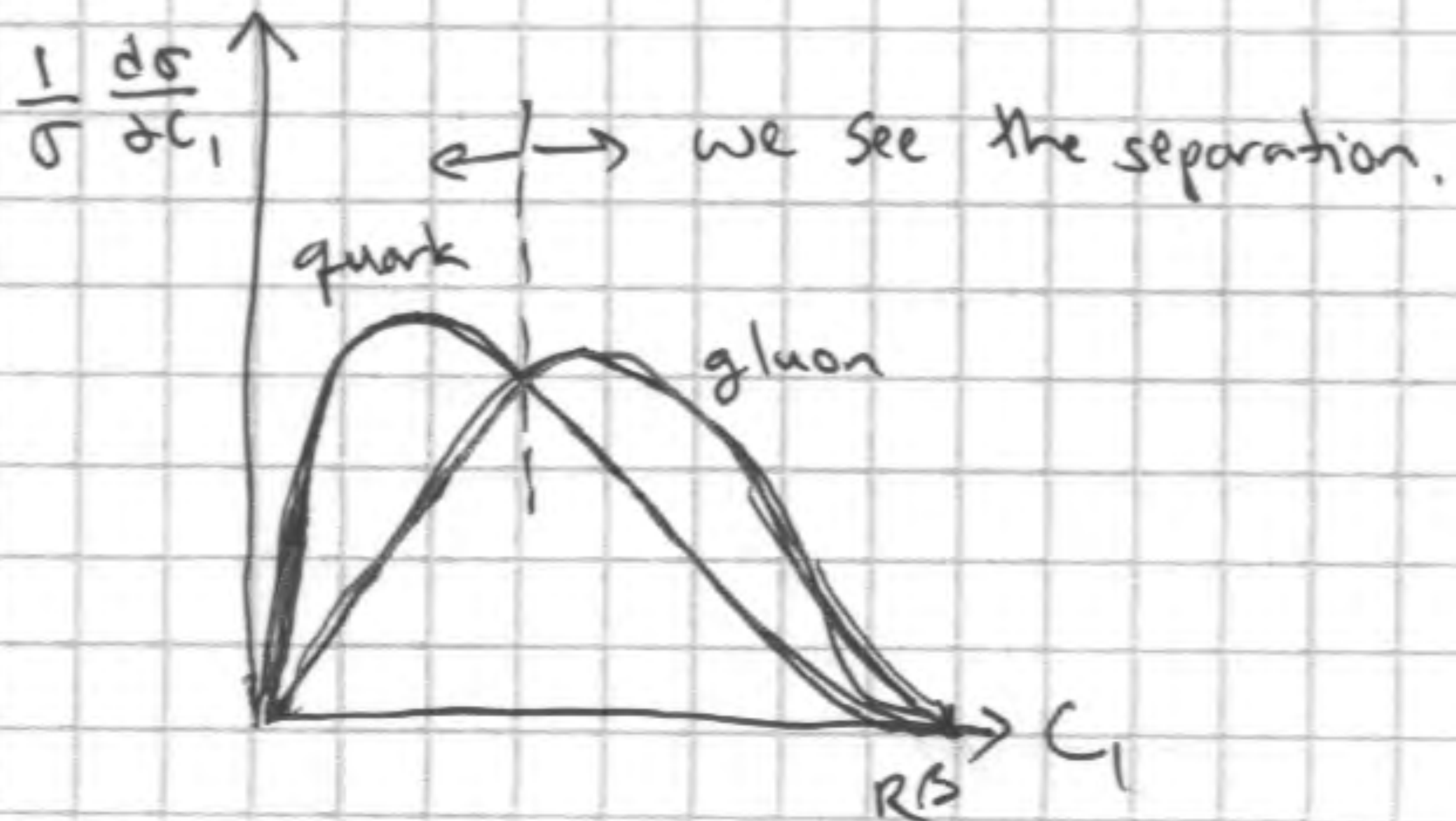
$$= \frac{2\alpha_s}{\pi}\frac{C_F}{\beta}\frac{1}{C_1}\log\frac{R^\beta}{C_1}\, e^{-\frac{\alpha_s}{\pi}\frac{C_F}{\beta}\log^2\frac{R^\beta}{C_1}}$$

At fixed order $\mathcal{O}(\alpha_s)$



delta function at origin from virtual diagrams to make sure distribution is normalized.

In strongly ordered limit



2) What is the discrimination power?

Place a cut C_1^{cut}

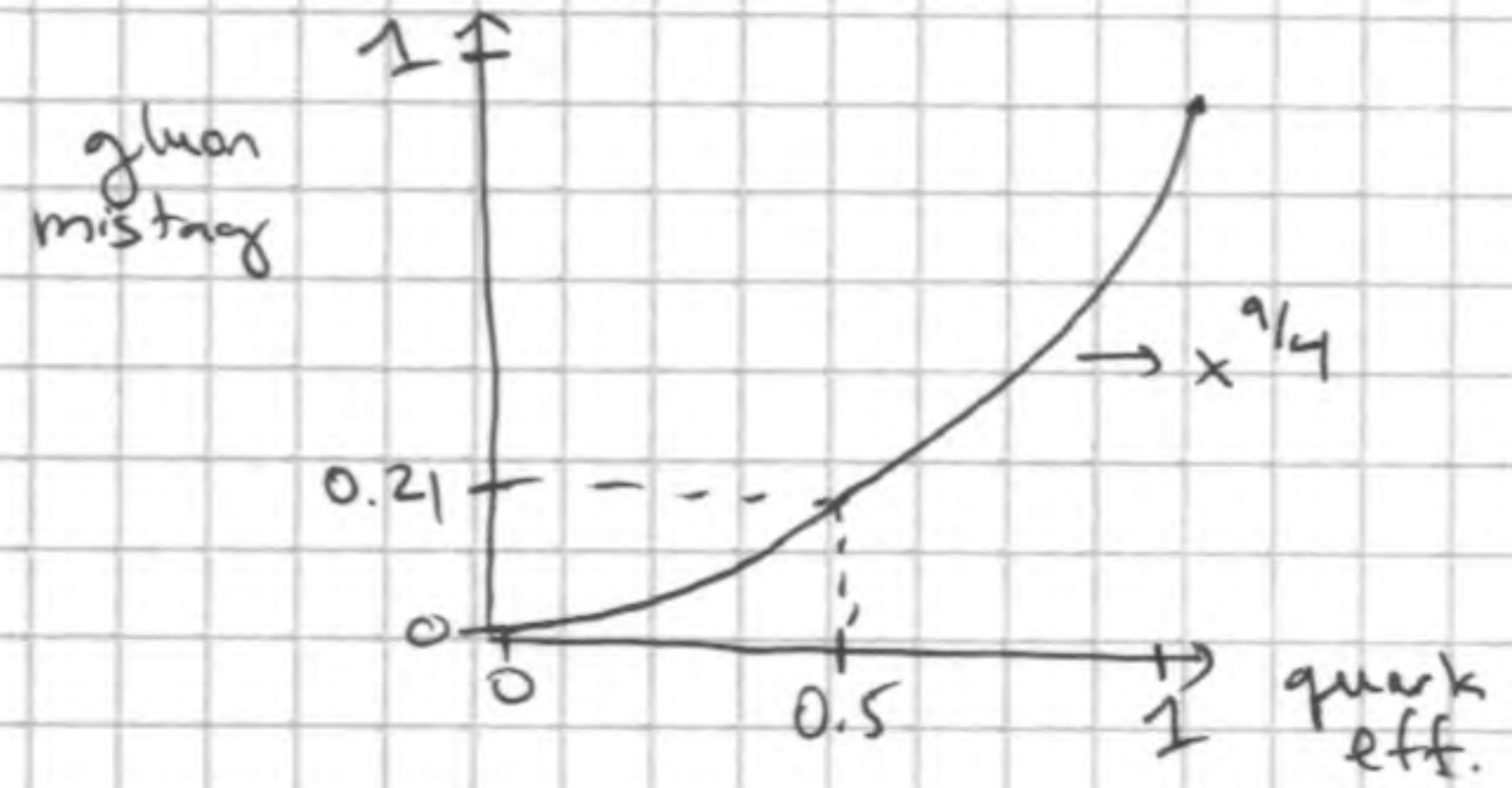
$C_1 < C_1^{cut} \Rightarrow$ "quark"

$C_1 > C_1^{cut} \Rightarrow$ "gluon"

Want to know efficiency $\Sigma_q(C_1^{cut})$

vs. mistag $\Sigma_g(C_1^{cut})$

Key: $$\Sigma_g = \left(\Sigma_q\right)^{C_A/C_F \to 9/4}$$



In strongly-ordered limit, independent of β!

(i.e. at this order, only difference between quarks and gluons is C_F vs. C_A.

Have to work much harder to optimize q vs. g discrimination.)

Higher Order Effects

- → Multiple Emissions
- → Subleading Terms in Splitting Functions
- → Fixed Order Corrections
- → Running α_s

} Advantages to small β, beyond the scope of these lectures.

=

General Strategy for Jet Substructure

→ Figure out your goal

("Discriminate Quarks vs. Gluons")

→ Identify Underlying Physics

("C_F vs. C_A")

→ Construct (clever?) observable to probe that physics

$$\left(\text{"} C_1 = \frac{\sum_{i<j} E_i E_j \theta_{ij}^{\beta}}{\left(\sum_i E_i\right)^2} \text{"} \right)$$

→ Use analytic / Monte Carlo methods to determine distributions / discrimination power

("strongly-ordered limit")

→ Convince experimentalists to apply to data.

=

Next time: Multi-Prong Jets & Jet Grooming
(More complicated observables) (Dealing with contamination)

Part 3: Boosted Objects

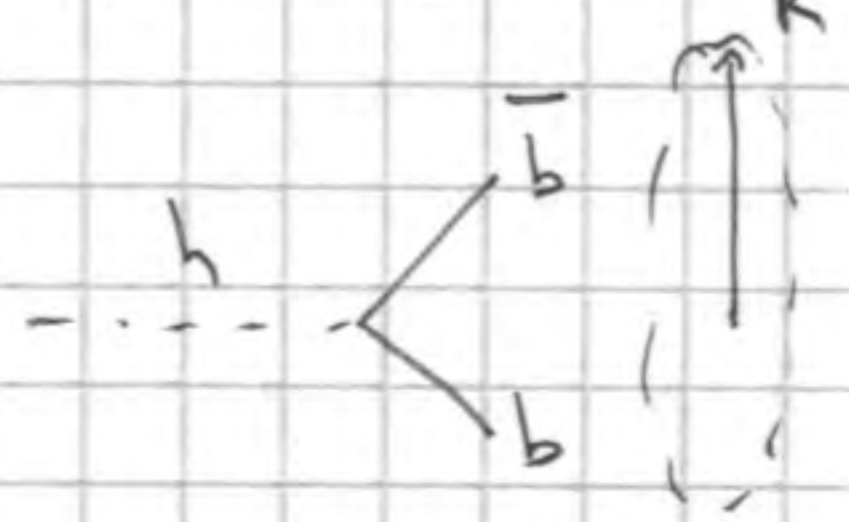
Last time: Quarks vs. Gluons Intrinsically Jet-Like

Used C_F vs. C_A to discriminate
Analytically tractable in strongly-ordered limit

This time: { W / Z / Higgs Intrinsically 2 Jet-Like
top quark Intrinsically 3 Jet-Like

Active research ongoing to get analytic control. This lecture will be more heuristic.

Key point: In extreme kinematic circumstances, W / Z / H / top can look like single jet.



$$m_H^2 \simeq p_{Tb}\, p_{T\bar{b}}\, \Delta R_{b\bar{b}}^2 \qquad p_{TH} \simeq p_{Tb} + p_{T\bar{b}}$$

For $p_{TH} \gg \frac{2 m_H}{R_{jet}}$

boosted Higgs looks like a single jet.

"But Higgs-jet has a mass of 126 GeV whereas quarks & gluons are massless."

Wrong! Quark and gluon jets are not massless.

Key Exercise: In soft-collinear limit show that

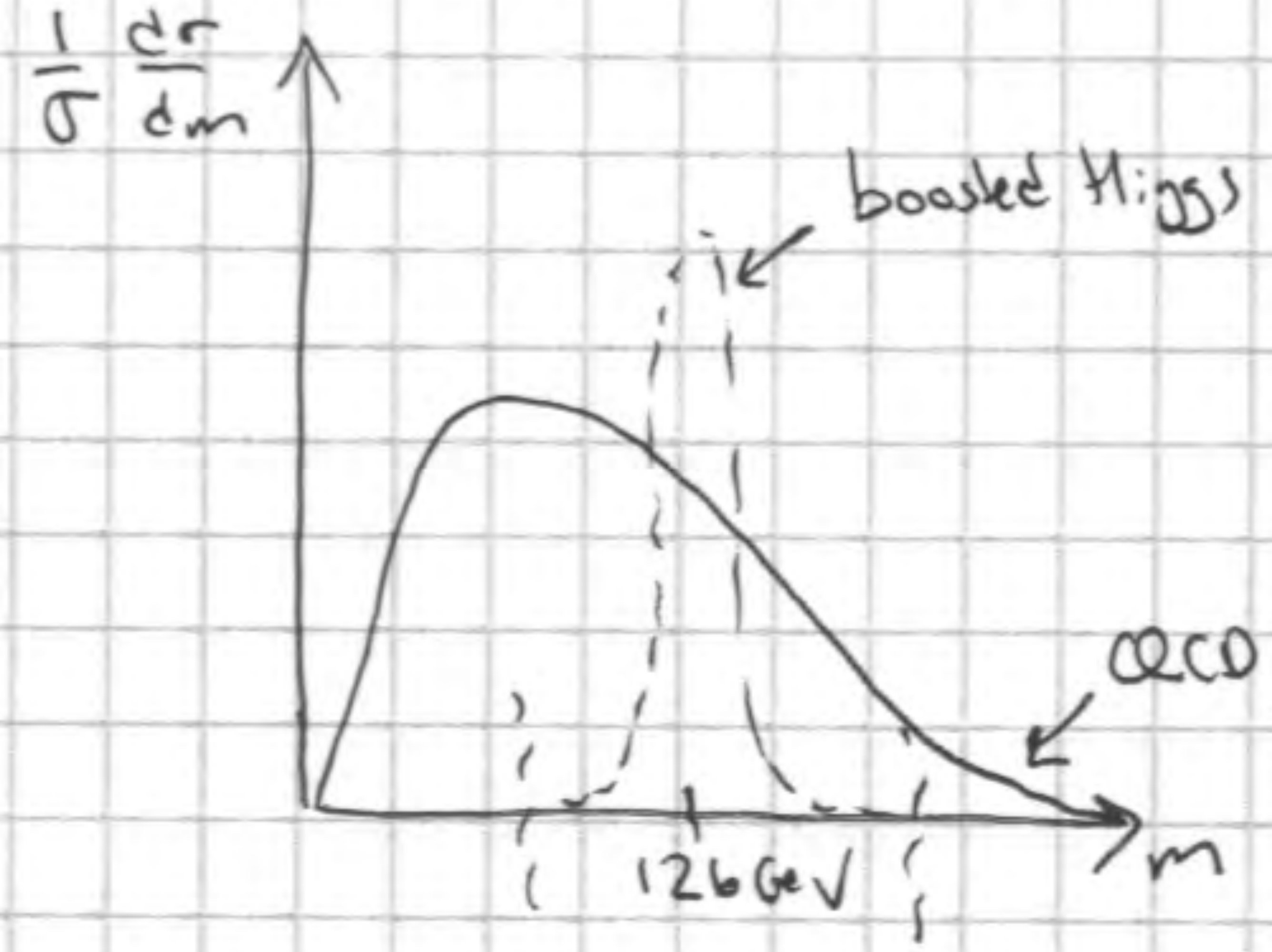
$$\langle m_{jet}^2 \rangle \approx \frac{\alpha_s C_{F,A}}{\pi} p_{T\,jet}^2 R^2$$

vs.

$$m_H^2 \approx \frac{p_{T\,jet}^2}{2} R_{b\bar{b}}^2$$

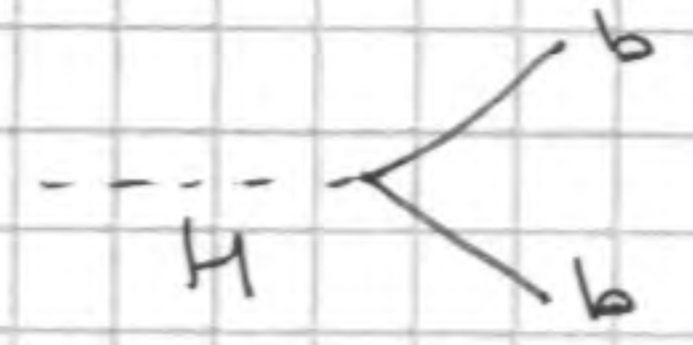
Need observables to distinguish these two cases.

Obvious choice: Jet mass

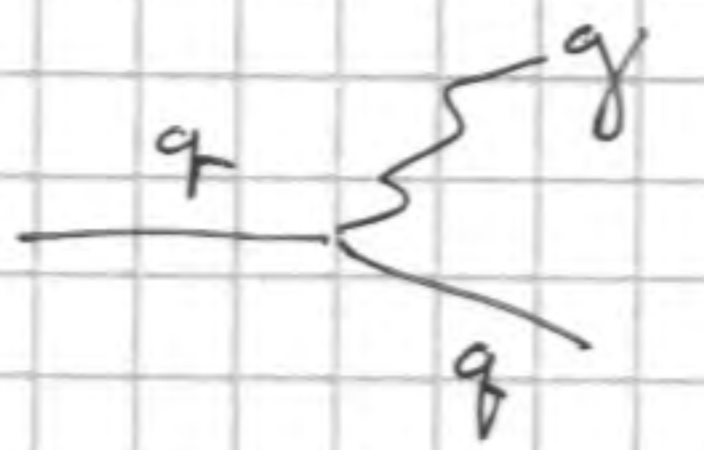


But want other observables to test for 2-prong nature of Higgs-jet.

Leading Order Structure



vs.



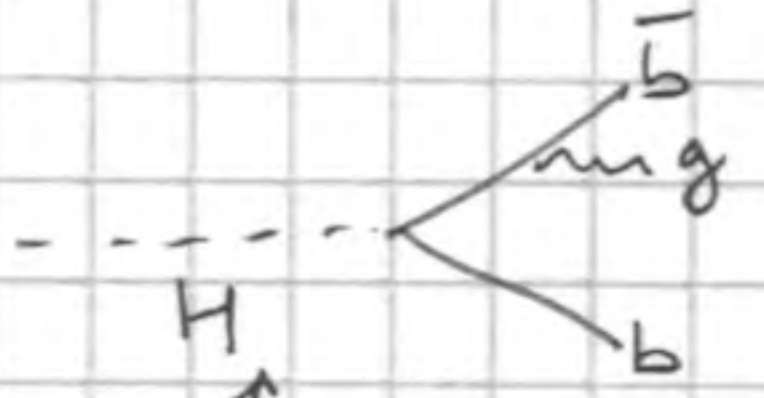
Democratic energy sharing

Flat distribution in $z = \frac{E_b}{E_H}$

Soft singularity

Expect $1/z$-like behavior.

Subleading Structure



↑
color singlet
radiation confined to $b\bar{b}$ dipole

↑
quark/gluon color-connected to rest of event

=

Generic Boosted Object Strategies

→ Toss out wide-angle soft radiation (see jet grooming)
(Large contribution to mass, but unlikely to come from hard subjet.)

→ Identify democratic energy sharing (i.e. measure "z")

→ Identify N-prong nature of jet ($N=2$ W/Z/H, $N=3$ top)

→ Probe color structure of jet (singlet vs. non-singlet)

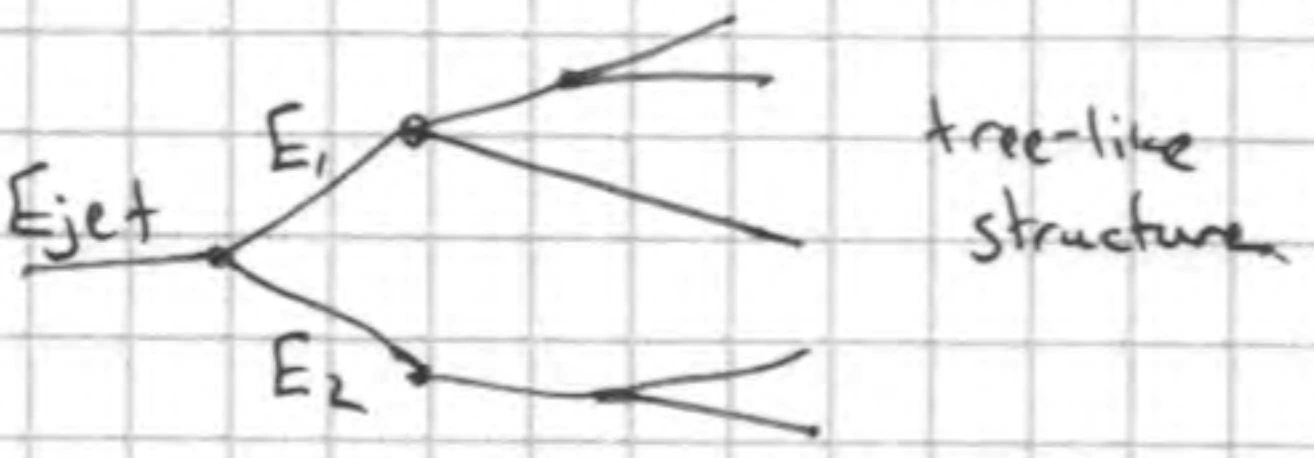
Algorithms / Observables for each of those strategies

(No way I can talk about all of them,
so I'll mention just two.)

BDRS (2008): Started* whole jet substructure industry for $pp \to ZH$, $H \to b\bar{b}$

*Actually, earlier (forgotten) work by Mike Seymour. (1994)
(1991)

1) Recursively cluster a jet using "CA algorithm" (join nearest neighbors until you reach jet radius R.)



2) Un-wind the jet, testing for symmetry and mass drop

$$(E_1 > E_2) \quad \text{Require } \frac{E_2}{E_1} > y_{cut}, \quad \frac{\max(m_1, m_2)}{m_{jet}} < \mu_{cut}$$

3) If test fails, throw out softer subjet, continue recursion on harder subjet.

If test succeeds, you've found a candidate Higgs jet.

Why does BDRS work?

y_{cut} ⇒ good test for democratic energy sharing

μ_{cut} ⇒ QCD jets get mass from many soft wide-angle emissions, whereas Higgs has a hard $H \to b\bar{b}$ splitting

Can be generalized to boosted top quarks (3-prong jets)

N-subjettiness (2010): Jet shape sensitive to N-prong structure.

(based on event shape N-jettiness)

First applied to boosted tops.

$$\tau_N^{(\beta)} = \frac{1}{\underset{\text{irrelevant normalization factor}}{\underset{\uparrow}{d_0}}} \sum_i p_{T_i} \min\{\underbrace{\Delta R_{i,1}, \Delta R_{i,2}, \dots \Delta R_{i,N}}_{\text{N axes chosen by some method, e.g. by minimizing } \tau_N^{(\beta)}}\}^\beta$$

$\tau_N^{(\beta)}$:

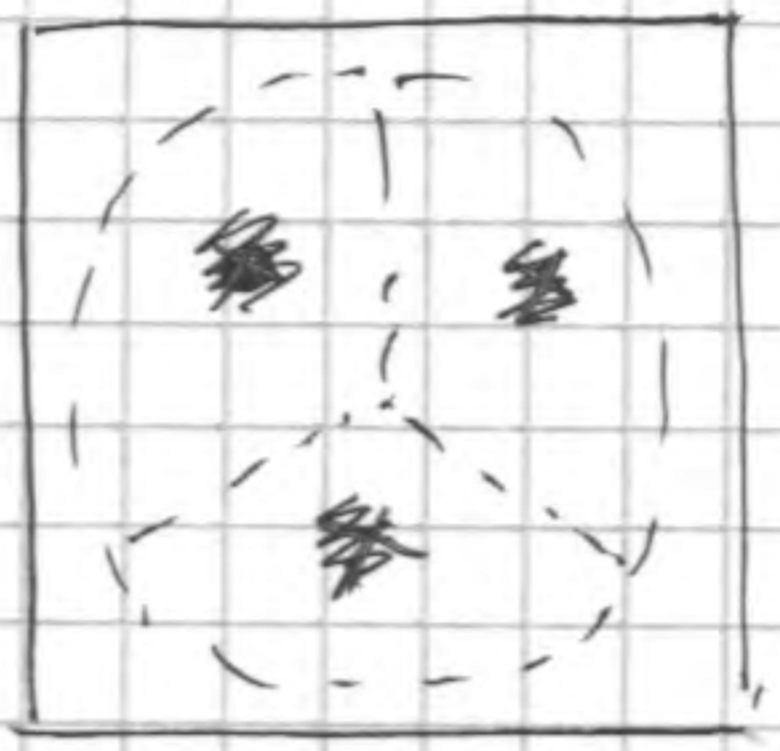
Consider a boosted top $\left(t \to b\underset{\hookrightarrow jj}{W}\right)$

vs. QCD jet with $m_{jet} \approx 170$ GeV

Top jet

QCD jet

τ_3



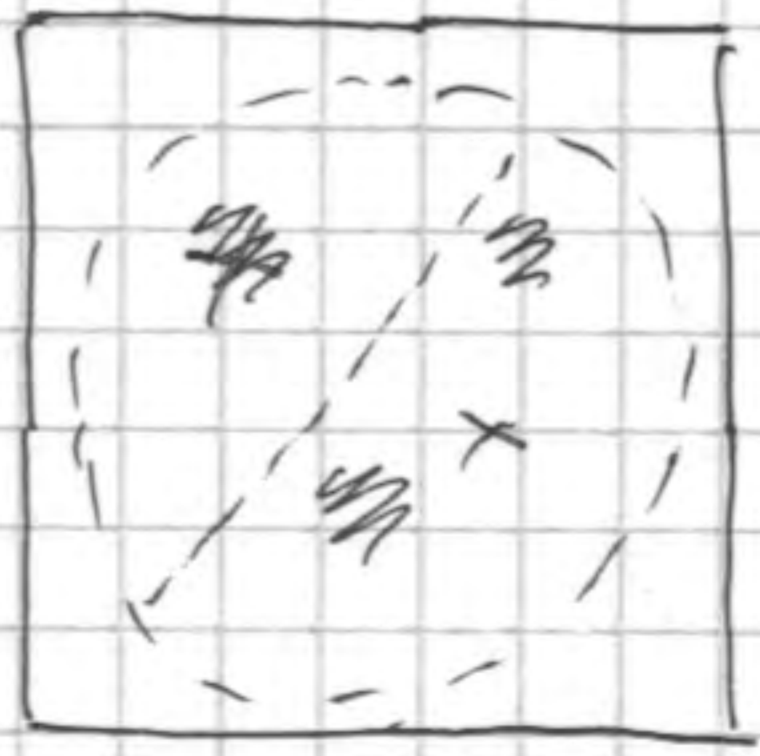
Axes "lock-on" to correct substructure

τ_3 is small

Indeterminate substructure

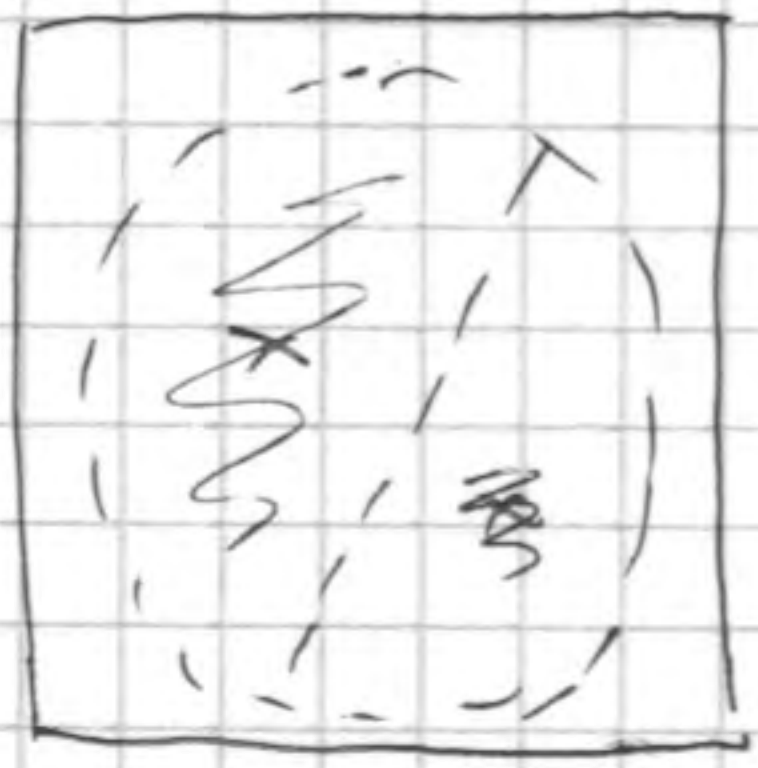
τ_3 is ambiguous.

τ_2



Not very 2-subjet-like

τ_2 is large



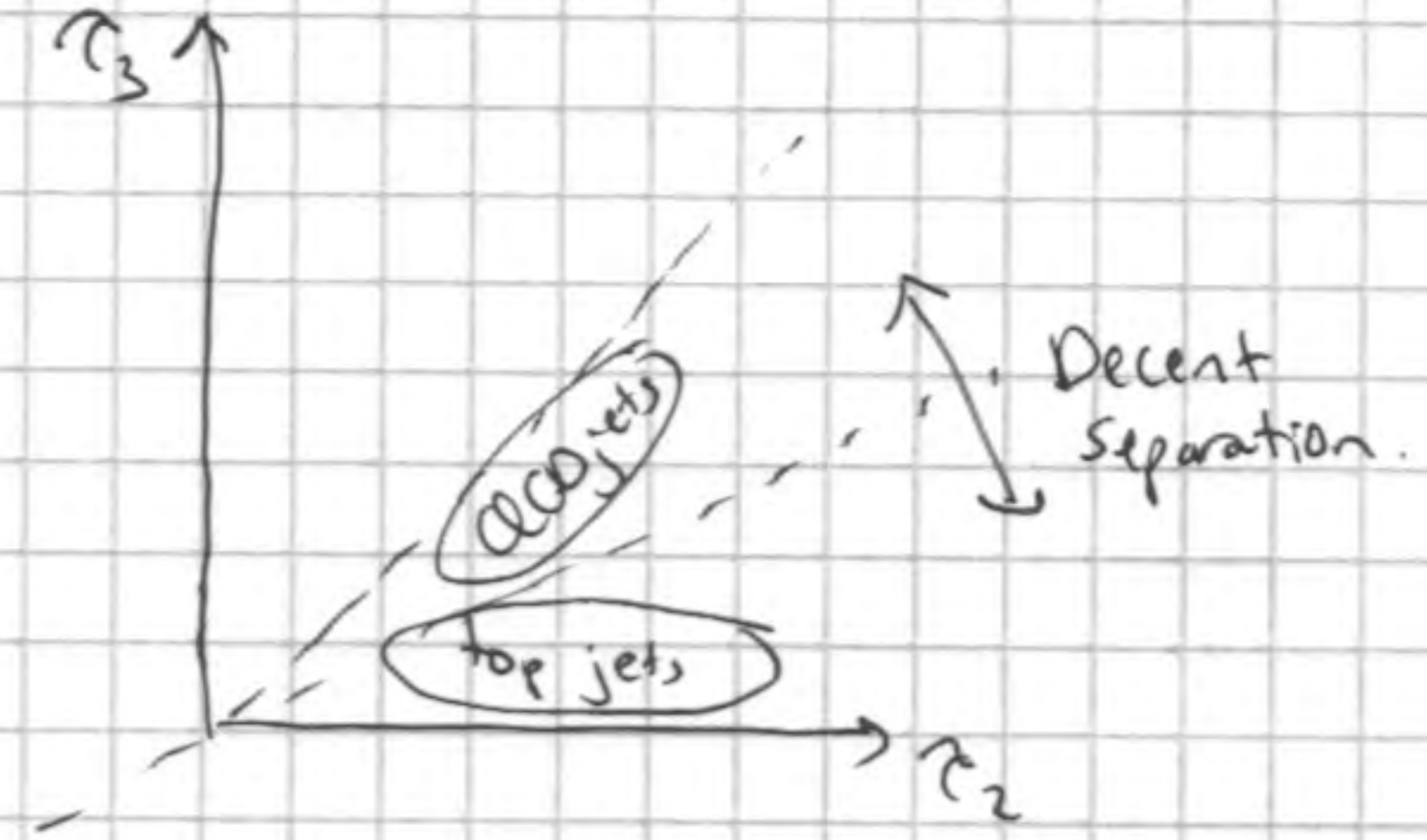
Still indeterminate substructure

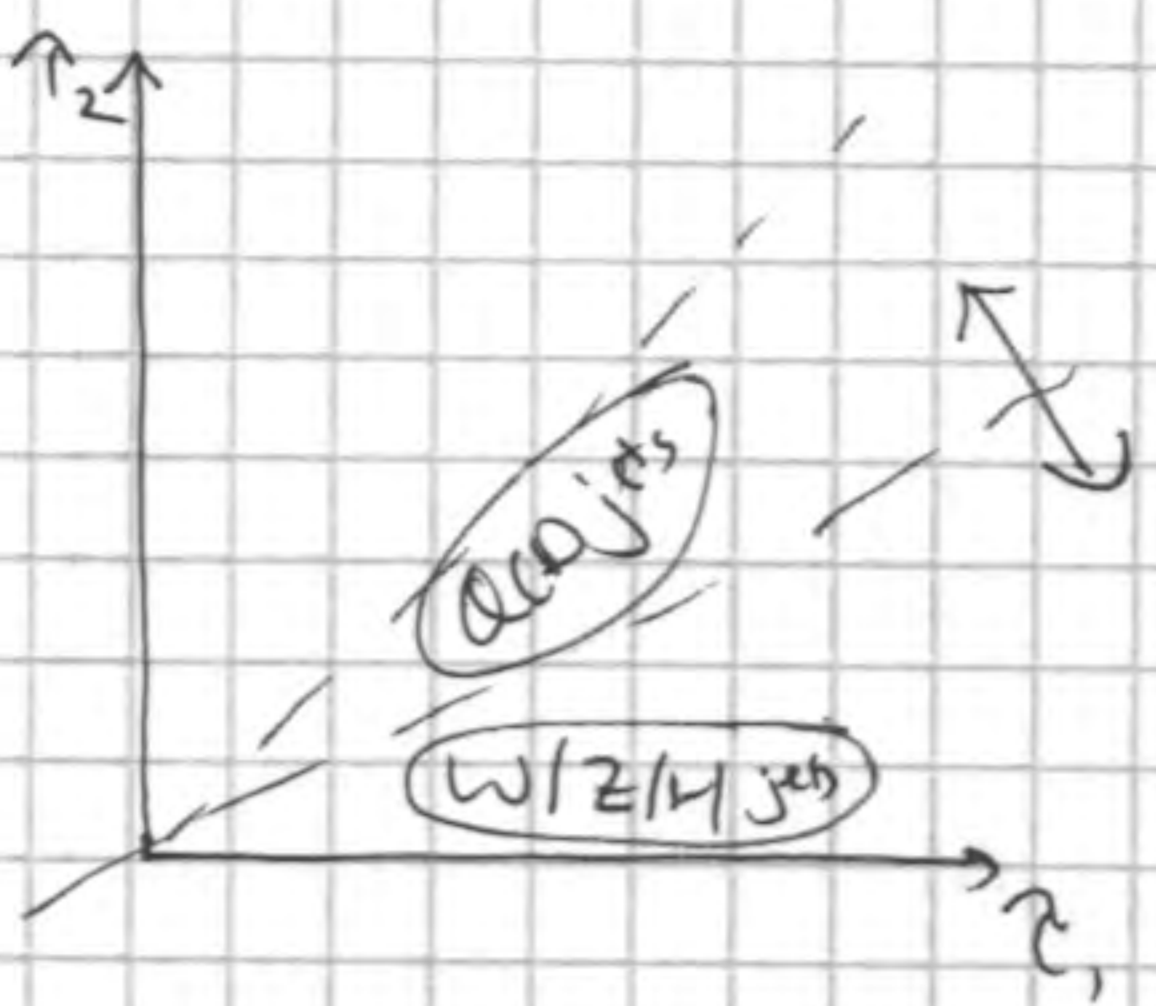
$\tau_2 \approx \tau_3$ is ambiguous.

(22)

Generally, for an N-prong jet $\tau_N \ll \tau_{N-1}$

τ_N / τ_{N-1} is a good discriminant for $N=3$ tops
$N=2$ W/Z/H.





Why does N-subjettiness work?

→ Boosted Object already has mass, but from prongs, not from gluon radiation

$$\tau_N \ll \tau_{N-1}$$

→ QCD jets get mass from energetic core with wide angle emissions

$$\tau_N \approx \tau_{N-1}$$

=

Both BDRS & N-subjettiness used in Higgs studies

BDRS: $h \to b\bar{b}$

τ_2/τ_1: $\underset{\substack{\uparrow \\ \text{heavier Higgs}}}{H} \to \underset{\substack{\uparrow \\ \text{boosted!}}}{WW} \to \underset{\substack{\uparrow \\ \text{help identify hadronic W.}}}{(jj)}(\ell\nu)$

(Aside: Can you combine best features of BDRS and τ_2/τ_1?

BDRS: Does not explicitly test for prongs

τ_2/τ_1: Does not explicitly test for symmetry

Generalized Energy Correlation Functions (2013)

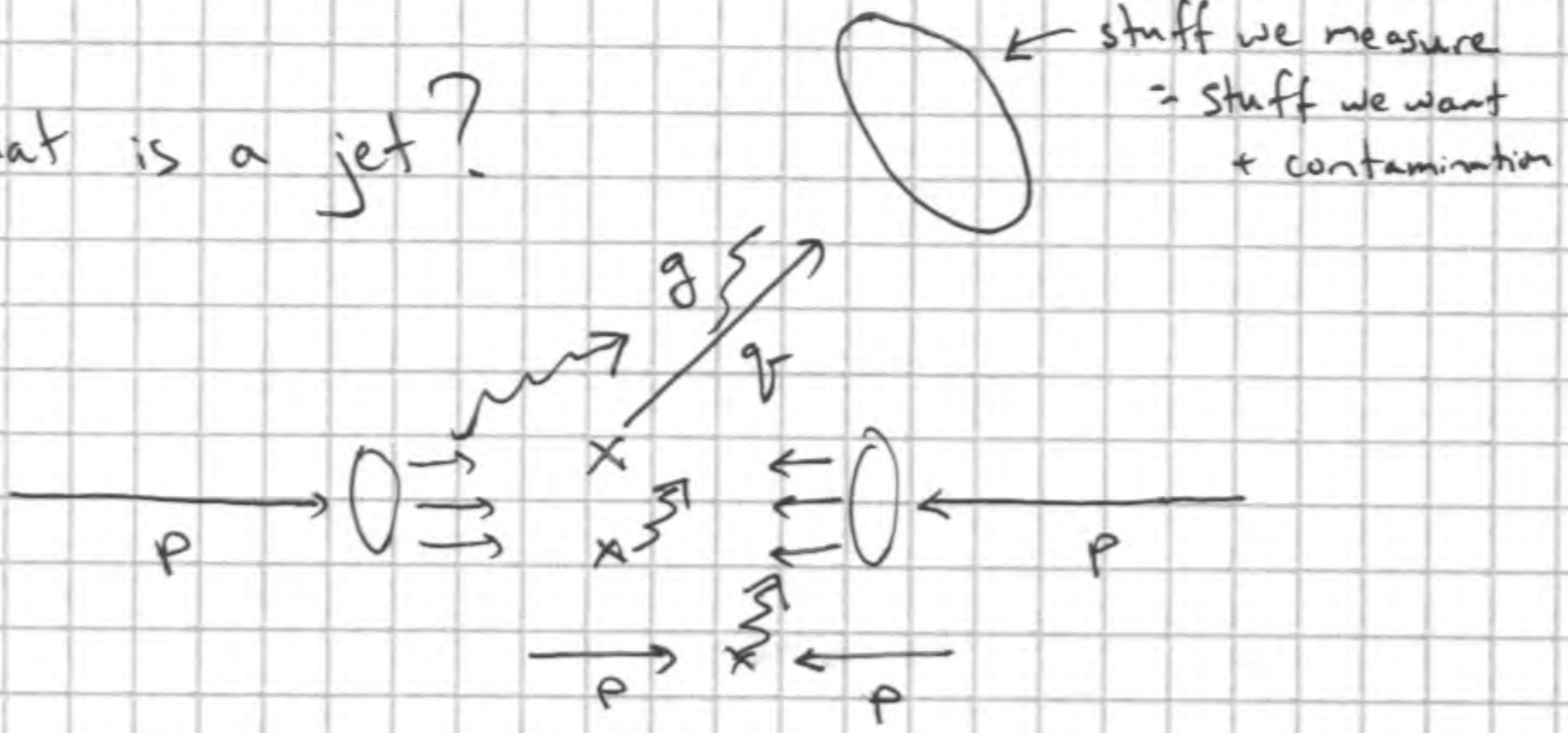
$$C_2^{(\beta)} = \frac{\sum_{ijk} E_i E_j E_k \left(\theta_{ij}\theta_{jk}\theta_{ik}\right)^{\beta} \sum_i E_i}{\left(\sum_{ij} E_i E_j \theta_{ij}^{\beta}\right)^2}$$

Combines features of BDRS & τ_2/τ_1 for 2-prong searches (W/Z/H)

On-going work to analytically understand these various 2-prong methods.)

Part 4 : Jet Grooming

What is a jet?

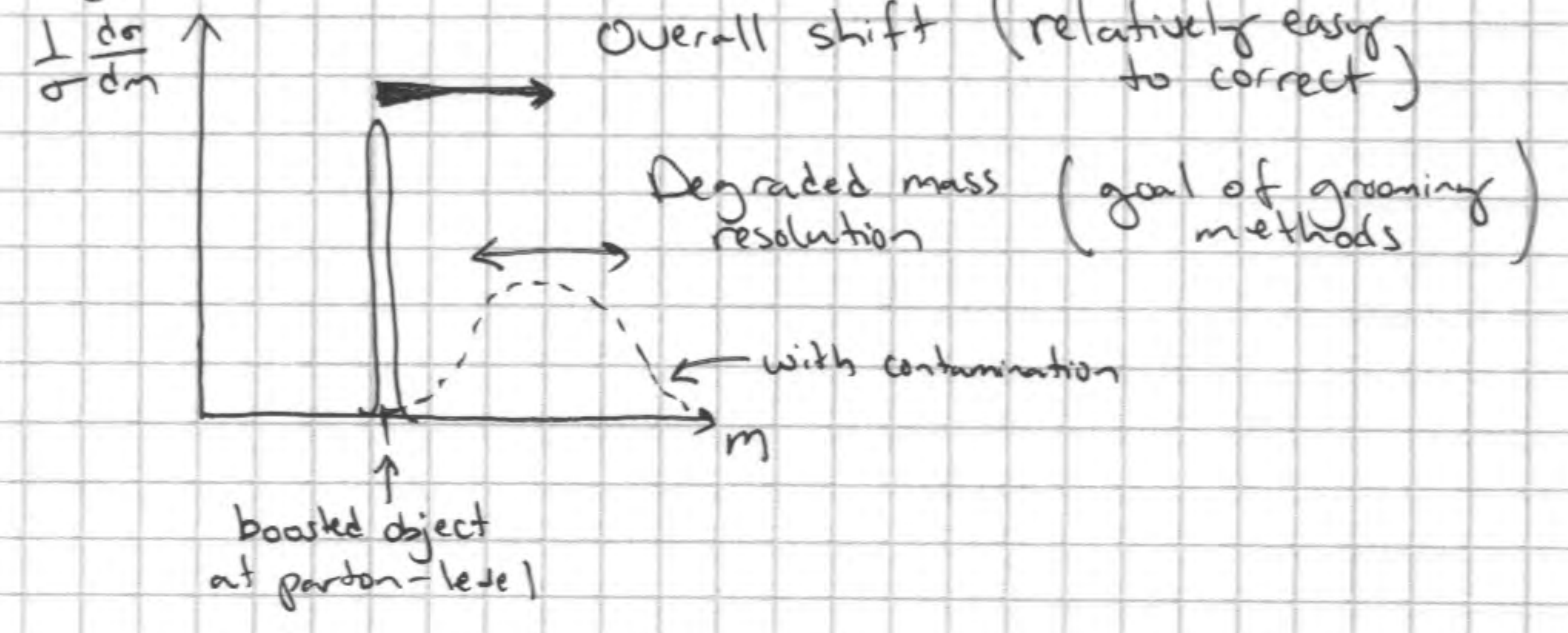


Sources of Contamination

Irreducible
- ISR (soft radiation from initial state)
- Underlying Event (secondary parton collisions)

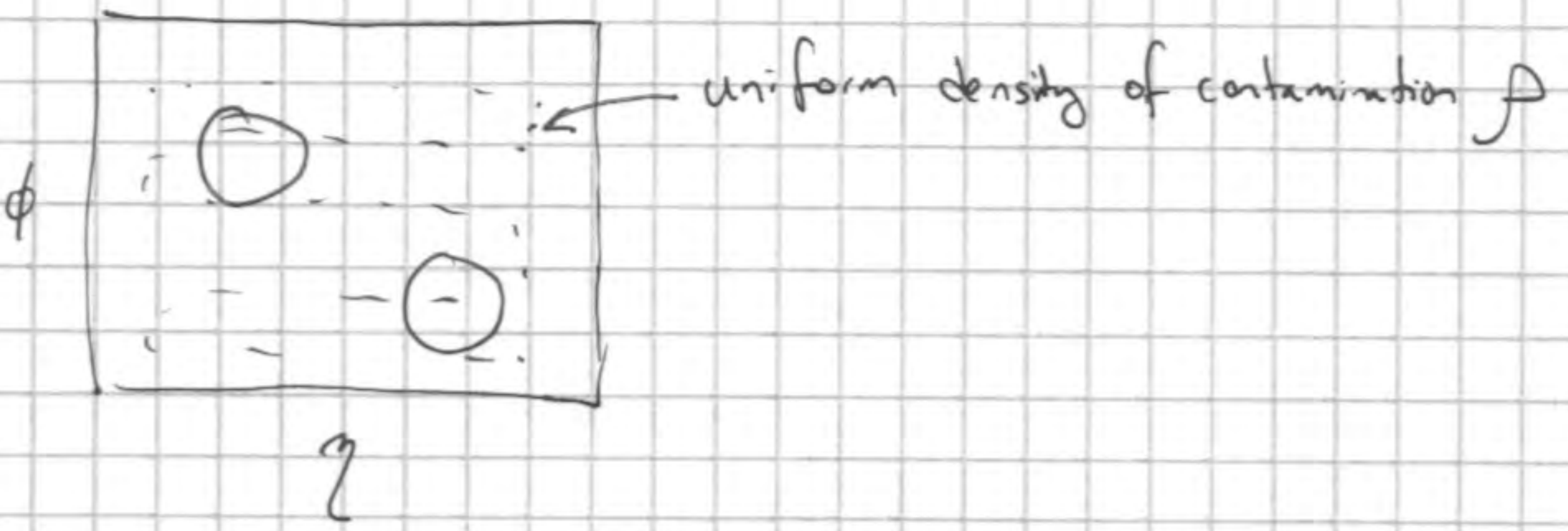
Reducible, but huge issue for high luminosity LHC.
- Pileup (secondary hadron collisions

Why does it matter?



Many methods have been proposed ..
(e.g. trimming, filtering, pruning, mass drop, ...)
... but I'll mention just two here.

Area Subtraction (2008): Specifically for Pileup.



0-th order assumption: (can refine) Pileup is uniform in η-ϕ plane

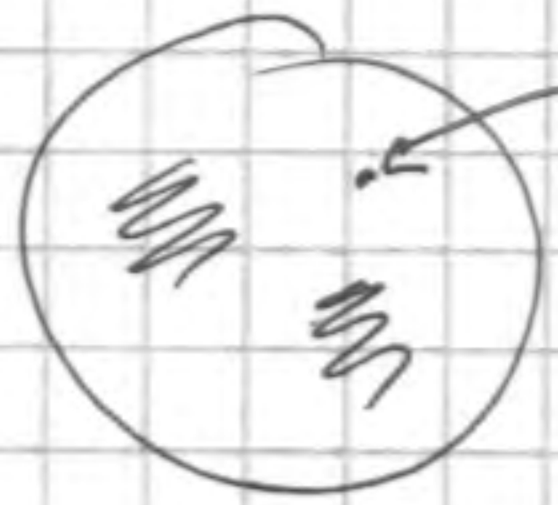
Estimate pileup density ρ
(calculate median contamination in mini-jets)

Add uniform spray of ghost particles with energy density $-\rho$ across entire event.

Effectively deals with event-by-event fluctuations of pileup (but not fluctuations within an event)

Jet Grooming Methods aim to generically remove soft, wide angle emissions.

What is soft?



Test whether E_T is small? No, not IRC safe

Need to define "soft" in IRC safe way.

(Recent paper on "Modified Mass Drop", which I suspect will be somehow the future of jet grooming. I need to understand it better, so focus on a grooming method more familiar to me.)

$\Longrightarrow$

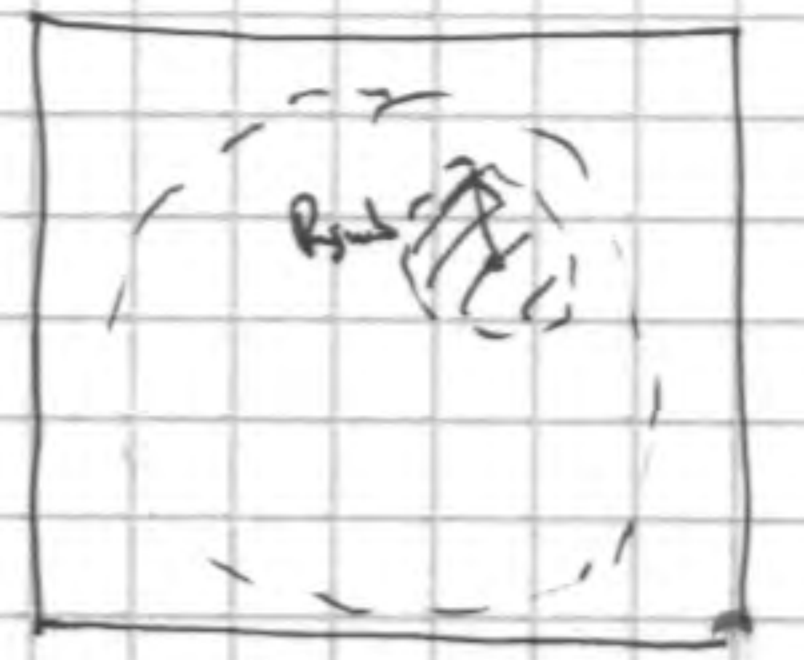
Jet Trimming (2009)

Ordinary jet $$p^\mu_{jet} = \sum_{i \in jet} p_i^\mu$$

Trimmed jet $$t^\mu_{jet} = \sum_{i \in jet} p_i^\mu \underset{\uparrow \text{ weight}}{w_i} \qquad w_i = \begin{cases} 1 & \text{if "hard"} \\ 0 & \text{if "soft"} \end{cases}$$

Originally weights were defined through jet algorithms, but a bit easier to understand through analytic formula. (upcoming work)

$$w_i = \Theta\left(\frac{\sum_{j\in \text{jet}} p_{T_j}\,\theta(R_{ij} < R_{sub})}{\sum_{j\in \text{jet}} p_{T_j}} > f_{cut} \right)$$



Measure energy in a region of size R_{sub}

Keep if energy fraction is above a threshold.

"Active" removal of jet contamination, both pileup and ISR/UE. Recent ATLAS study showing that there is approximately flat response as N_{PV} increases.

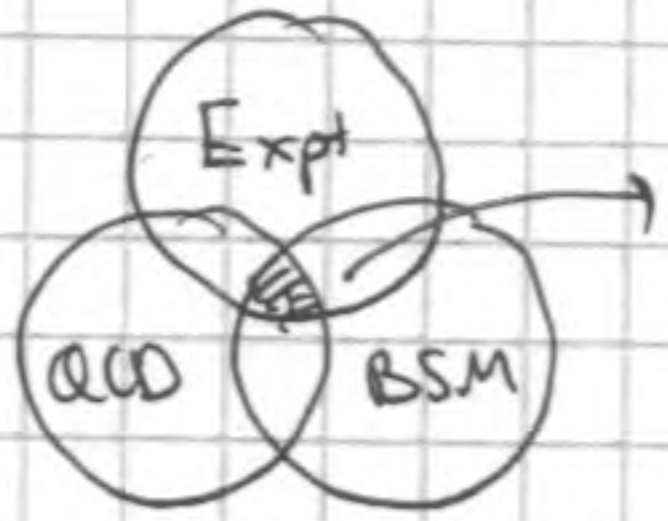
Trimming is a "generic" method, agnostic as to whether the jet is quark/gluon/boosted object. Other grooming methods optimized for boosted objects.

I've given you a brief tour of jet substructure. Much more information in BOOST reports.

Shelton's TASI lectures a good place to start for more information/references.

The Future of Jet Substructure.

An active field (cf. BOOST workshops)
with room for new ideas.



how will we push the boundaries?

Experiment: Near Term: Apply q vs. g, boosted objects, jet grooming to wide variety of searches

Data driven validation methods.

Long Term: How to define/trigger jets in high pileup environment.

QCD Theory: Near Term: Analytic methods for "2-prong" taggers

Gain understanding of differences between methods, validity of Monte Carlo

Long Term: Tackle question of impact of non-perturbative physics

BSM Theory: Near Term: Apply to as many BSM searches as possible

Jet substructure vs. jet superstructure (jet vs. event)

Long Term: Find jet blind spots, especially with high luminosity

Other opportunities in boosted regime?