Jet Theory: Benchmarks for the Tevatron and the LHC

Benchmark = point of reference for a measurement

or to assess performance

Steve Ellis
(with advice from many)

Big Picture:

For the next decade the focus of particle physics phenomenology will be on the Tevatron and the LHC. This activity will be both very exciting and very challenging -

- addressing a wealth of essential scientific questions
- with new (wonderfully precise) detectors
- operating at high energy *and* high luminosity (eventually)
- most of the data will be about hadrons (jets).

Theory and Experiment must work together to make the most of the data.

Department of Physics
University of Washington

Standard Model Benchmarks at the Tevatron and LHC

11/19/10

Outline

- Why jets are essential?
- Old and New lessons for Cone and Recombination (kT) jets*
- How well does theory work for jets?
- Intro to understanding jet masses & jet substructure

*For a recent summary see “Jetography” by Gavin Salam, 0906.1833 (but I recall some history differently!)

Why Jets & Algorithms?

- Focus on large energy exchange processes
(to make interesting stuff)
⇒ resolve the partons within beam protons!

- Partons (q’s and g’s) are colored and radiate when isolated in phase space ⇒ (~collinear) showers of colored partons
⇒ “jets” at parton level

- Long distance dof are color-less hadrons
⇒ jets of hadrons in the detector,
but cannot (strictly) arise from single parton

⇒ ID “jets” with algorithm (a set of rules) applied to both hadrons and partons

Defining Jets – No Unique/Correct Answer

- Map the observed (hadronic) final states onto the (short-distance) partons by summing up all the approximately collinear stuff (shower), ideally on an event-by-event basis.

- Need rules for summing ⇒ jet algorithm
  Start with list of partons/particles/towers
  End with list of jets (and stuff not in jets)

E.g.,

- Cone Algorithms, based on geometry – "non-local" sum over core of shower

  Simple, "well" suited to hadron colliders with Underlying Events (UE)

- Recombination (or kT) Algorithm, based on "local" pair-wise merging of local objects to "undo" shower

  Tends to "vacuum up" soft particles, "well" suited to e+e- colliders

ALGORITHM Benchmarks

- *Fully Specified: including defining in detail any preclustering, merging, and splitting issues*
- *Theoretically Well Behaved: the algorithm should be infrared and collinear safe (and insensitive) with no ad hoc clustering parameters (all orders in PertThy)*
- *Detector Independence: there should be no dependence on cell type, numbers, or size*
- *Level Independence: The algorithms should behave equally at the parton, particle, and detector levels.*
- *Uniformity: everyone uses the same algorithms (theory and experiment, different experiments)*

Historically never entirely true! Do better at the LHC!

Jet issues can arise from -

- Systematics of specific algorithm
- Higher - order perturbative contributions (jet not just single parton)
- Showering – sum of all orders (leading-log, soft-collinear) emissions - smears energy distribution – “splash-out”
- Hadronization – nonperturbative re-organization into color singlet hadrons (confinement) – “splash-out”
- “Uncorrelated” contributions from rest of collision (UE) – “splash-in”
- Uncorrelated contributions of overlapping collisions (PU) “splash-in”
- Poor communication between theory and experiment

“Lesson” about Jet Systematics –

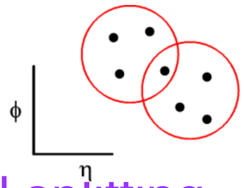
Cone Algorithm – focus on the core of jet (1990 Snowmass)

- Jet = “stable cone” ⇒ 4-vector of cone contents || cone direction
- Well studied – several issues

- **Cone Algorithm – particles, calorimeter towers, partons in cone of size R, defined in angular space, *e.g.*, (*y*,*φ*),**
- **CONE center -** $\left(y^C, \varphi^C\right)$
- **CONE i ∊ *C iff*** $\Delta R^i \equiv \sqrt{\left(y^i - y^C\right)^2 + \left(\varphi^i - \varphi^C\right)^2} \le R$
- **Cone Contents ⇒ 4-vector** $P_\mu^C = \sum_{i \subset C} p_\mu^i$
- **4-vector direction** $\overline{y}^C = 0.5 \ln\left[\frac{P_0^C + P_z^C}{P_0^C - P_z^C}\right] \; ; \; \overline{\varphi}^C = \arctan\left[\frac{P_y^C}{P_x^C}\right]$
- **Jet = stable cone** $\left(\overline{y}^C, \overline{\varphi}^C\right) = \left(y^C, \varphi^C\right)$

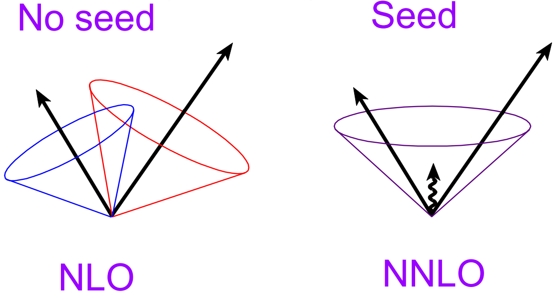
Find by iteration, *i.e.*, put next trial cone at $\left(\overline{y}^C, \overline{\varphi}^C\right)$

Cone – a (bad) Benchmark



1) Stable Cones can and do Overlap: need rules for merging and splitting, but not the same for D0 and CDF

2) Seeds – experiments only look for jets around active regions (save computer time)



⇒ problem for theory, IR sensitive (Unsafe?) at NNLO

This is a BIG deal philosophically – but not a big deal numerically (in data)
⇒ Use SEEDLESS version (SISCone) at the LHC

Remember this lesson at the LHC – talk to theorists !!

3) Splash-out from smearing of energetic parton at edge of cone – can be quantitatively relevant (the R_{sep} thing)

4) Dark towers – secondary showers may not be clustered in any jet

Recombination Algorithm – focus on undoing the shower pairwise, $\Rightarrow$ Natural definition of substructure

Merge partons, particles or towers pairwise based on “closeness” defined by minimum value of k_T, *i.*e. make list of metric values (rapidity *y* and azimuth ϕ, p_T transverse to beam)

$$\text{Pair } ij: k_{T,(ij)} \equiv \mathrm{Min}\left[\left(p_{T,i}\right)^{\alpha},\left(p_{T,j}\right)^{\alpha}\right]\frac{\sqrt{\left(y_i-y_j\right)^2+\left(\phi_i-\phi_j\right)^2}}{D} \equiv \mathrm{Min}\left[\left(p_{T,i}\right)^{\alpha},\left(p_{T,j}\right)^{\alpha}\right]\frac{\Delta R_{ij}}{D},$$

$$\text{Single } i: k_{T,i} = \left(p_{T,i}\right)^{\alpha}$$

If $k_{T,(ij)}$ is the minimum, merge pair (add 4-vectors), replace pair with sum in list and redo list;

If $k_{T,i}$ is the minimum → *i* is a jet! (no more merging for *i*, it is isolated by *D*),

1 angular size parameter *D* *(NLO, equals Cone for D = R,* R_{sep} *= 1),* plus

α *= 1,* ordinary k_T, recombine soft stuff first

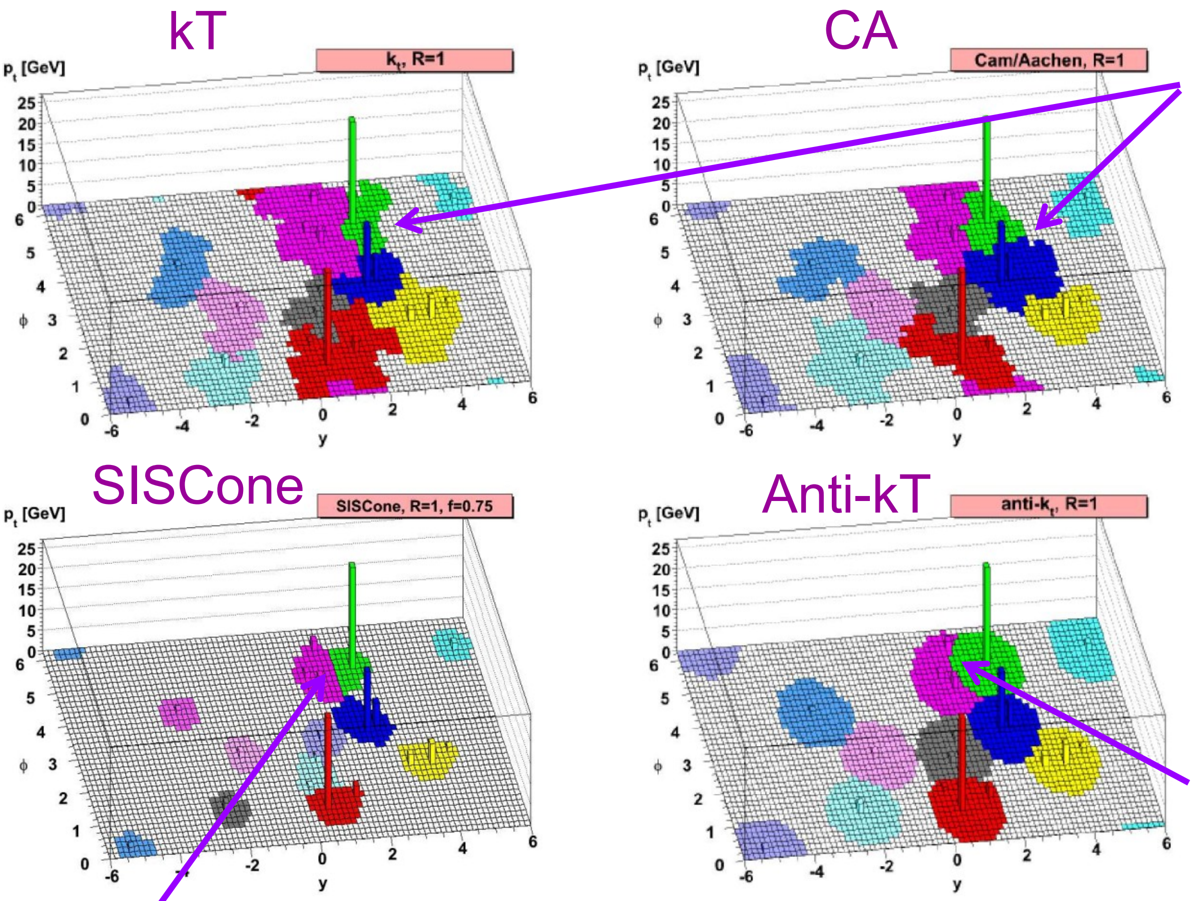
α *= 0, Cambridge/Aachen* (CA)*,* controlled by angles only

α *= -1, Anti-*k_T*,* just recombine stuff around hard guys – cone-like (with seeds)

Recombination Lessons:

- 👍 Jet identification is unique – no merge/split stage
- 👍 “Everything (interesting) in a jet”, no Dark Towers
- 👎 Resulting jets are more amorphous for $\alpha \geq 0$, energy calibration more difficult (subtraction for UE + PU?)
- 👍 But for $\alpha < 0$, Anti-kT (Carriari, Salam & Soyez), jet area seems stable and geometrically regular * - the “real” cone algorithm (but large pT jets take a bite out of small pT one)

Jet Areas – from Salam & Carriari, Salam & Soyez


kT

CA

SISCone

Anti-kT

Amorphous edges

Anti-kT very regular leading jets,

But bites others

S/M effect

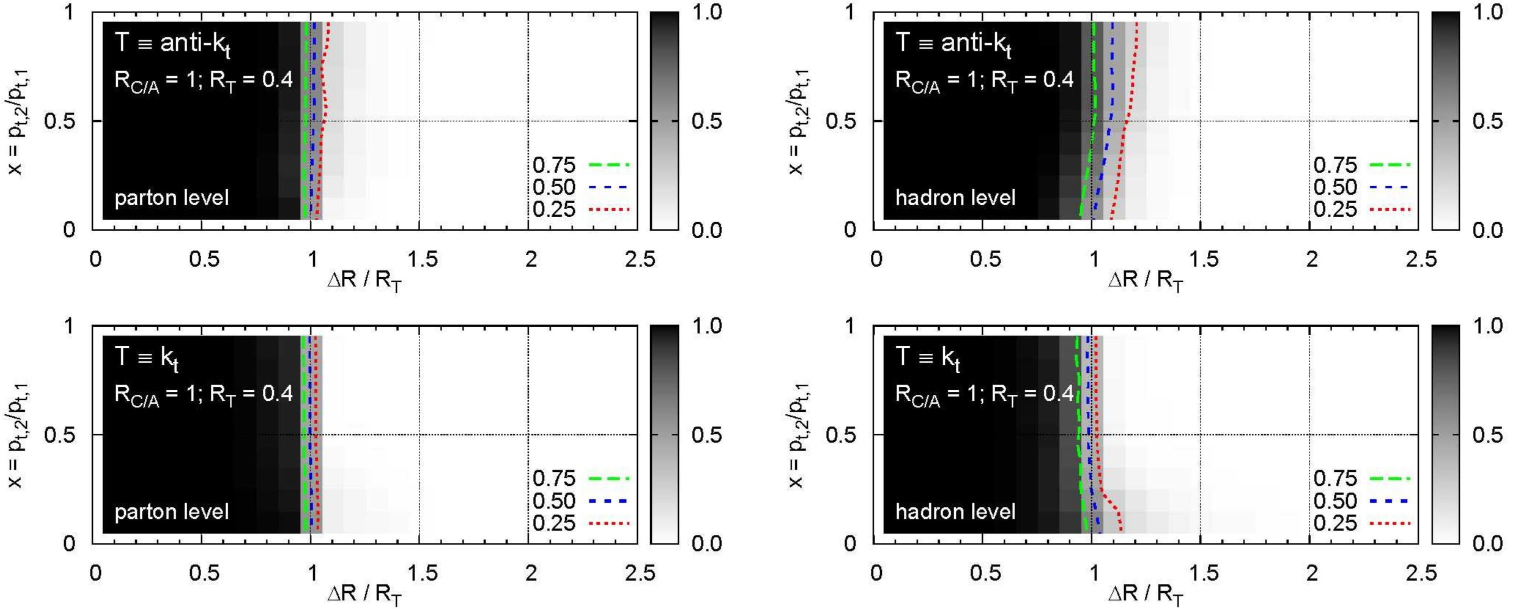
Verify this behavior in data?

Recombination Lessons:

- 👍 Jet identification is unique – no merge/split stage
- 👍 “Everything (interesting) in a jet”, no Dark Towers
- 👎 Resulting jets are more amorphous for $\alpha \geq 0$, energy calibration more difficult (subtraction for UE + PU?)
- 👍 But for $\alpha < 0$, Anti-kT (Carriari, Salam & Soyez), jet area seems stable and geometrically regular * - the “real” cone algorithm (but large pT jets take a bite out of small pT one)
- 👍 Anti-kT - Energetic partons seldom near edge of jets – merging matches Pert Theory, showering less of a issue – suggests more “stable” (reliable) results (easier to approximate) in higher order Pert Theory and SCET (all orders of some logs)

Recombination & showering (Salam)

Find energetic subjets (~last merging) with CA, see what anti-kT and kT find – do they still merge insider R – YES



(Shower only)

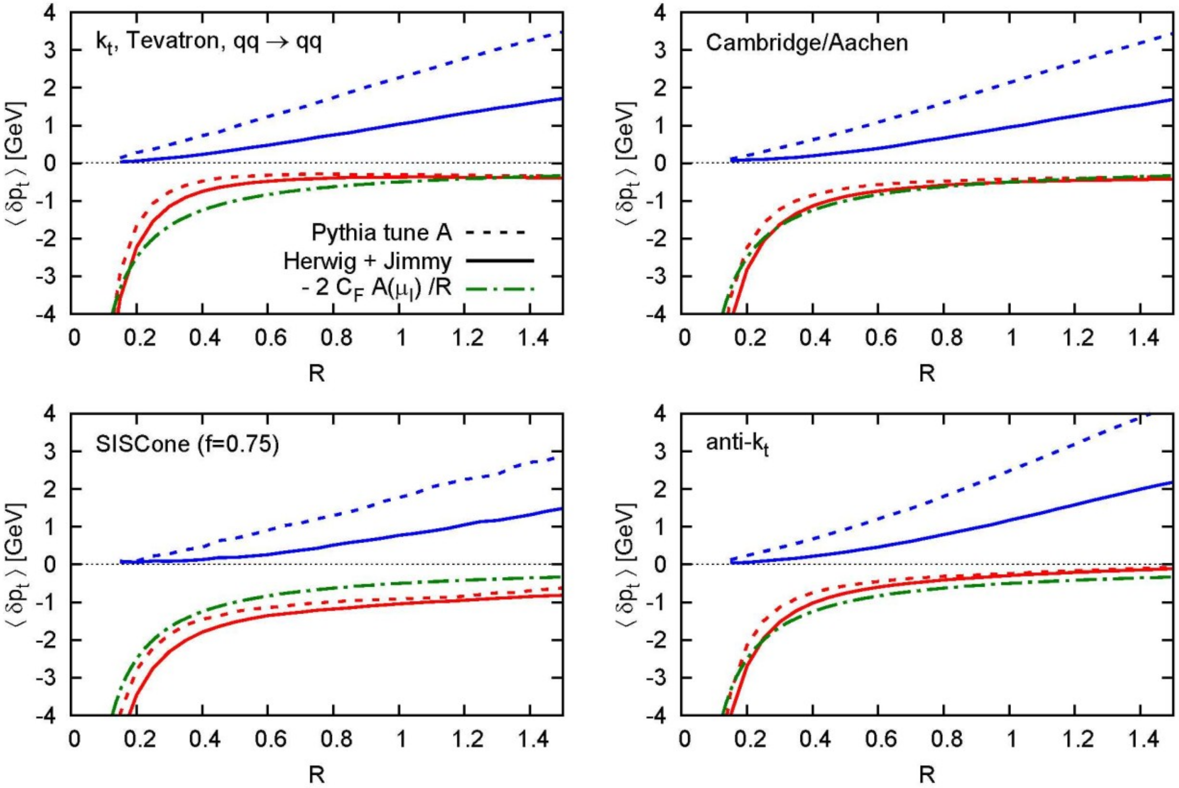
(Shower + had'n, more smearing)

Recombination Lessons:

- 👍 Jet identification is unique – no merge/split stage
- 👍 "Everything" in a jet, no Dark Towers
- 👎 Resulting jets are more amorphous for $\alpha \geq 0$, energy calibration more difficult (subtraction for UE + PU?)
- 👍 But for $\alpha < 0$, Anti-kT (Carriari, Salam & Soyez), jet area seems stable and geometrically regular * - the "real" cone algorithm (but large pT jets take a bite out of small pT one)
- 👍 Anti-kT - Energetic partons seldom near edge of jets – merging matches Pert Theory, showering less of a issue – suggests more "stable" (reliable) results (easier to approximate) in higher order Pert Theory and SCET (all orders of some logs)
- 👎 Analysis can be very computer intensive (time grows like N^3, recalculate list after each merge)
- 👍 New version FASTJet (Salam & Soyez) goes like N^2 or N ln N ($\alpha \geq 0$), plus scheme for finding areas (and UE correction)

Hadronization (splash-out – upper curves) & UE (Splash-out – lower curves) issues:

- Measure and Correct jet properties – MC & analytic Tevatron - $pJ/\Delta R^2 \sim 0.5$ GeV for UE



UE (MC's don't agree)

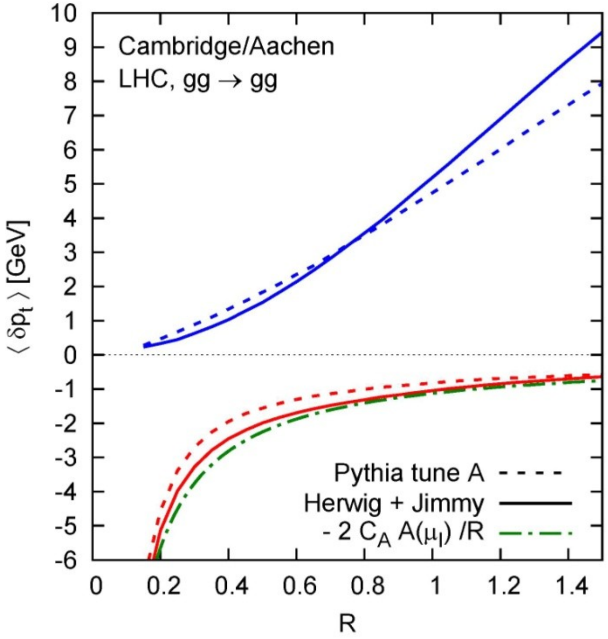
hadronization

Similar for different algorithms and approx. cancel for R ~ 0.7

Studies by Dasgupta, Magnea and Salam

At the LHC, expect - pJ/$\Delta R^2 \sim 1.5$ GeV

- Here gluons!,
  better UE agreement
  Still approx. cancel for
  R ~ 0.6



- Essential to establish data -driven Benchmarks here!

- If “groom” jets (pruning, trimming, filtering), may lose cancellation!

How reliable is Jet Perturbation Theory ?

Depends on *arbitrary* scales μ_{UV}, μ_{CO} -

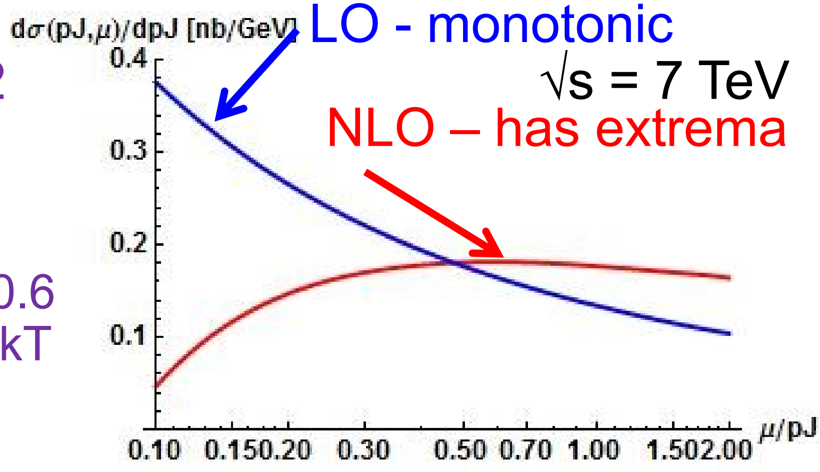
How to we choose values?

How do we vary scales to *estimate* the uncertainty?

(How large a pT is large enough?)

- The History is choose $\mu_{UV} = \mu_{CO} = pJ/2$ based on
  - NLO ~ μ “independent” there
  - NLO ~ LO there
- The Pert Thy tells us the scale!

D = 0.6
Anti-kT



LO - monotonic

$\sqrt{s}$ = 7 TeV

NLO – has extrema

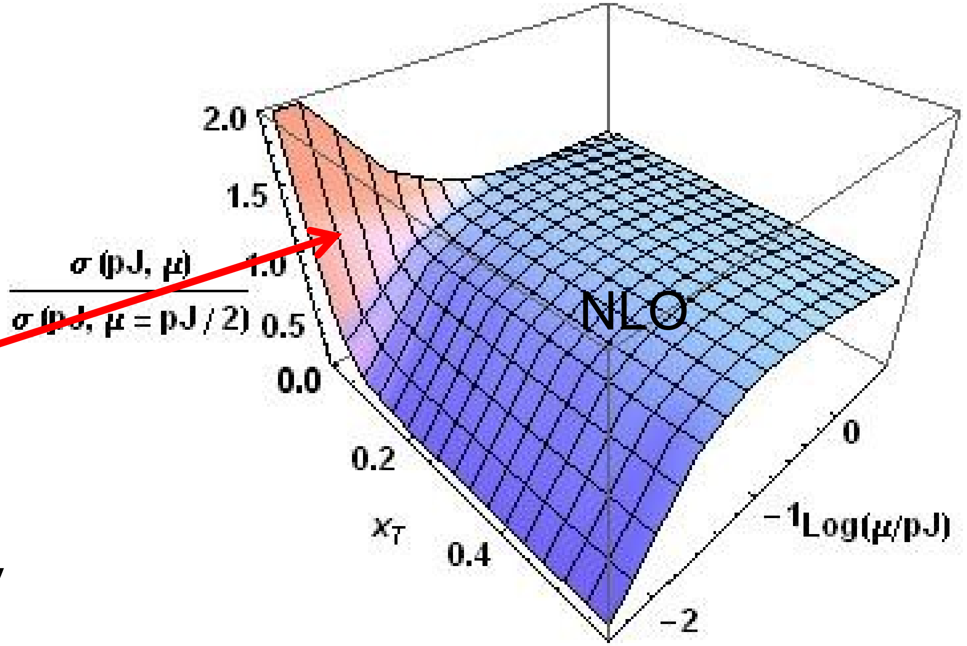
pJ = 200 GeV

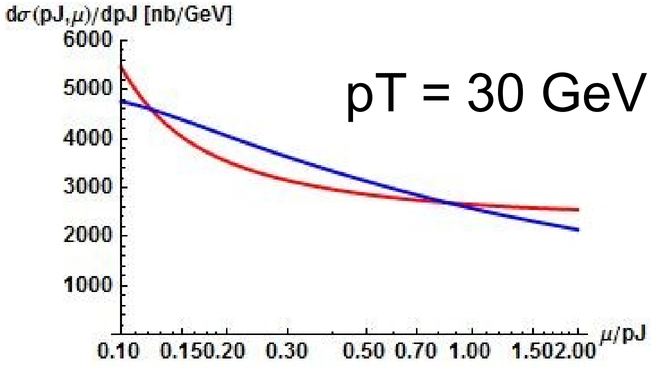
So (some) answers come from Pert Thy calculation itself – ask your house theorist to show you her μ dependence plot!

- But – the answer changes with rapidity and s (and other scales in problem). Due to logs of ratios of scales a large pT at the Tevatron may not be a large pT at the LHC

- Consider xT = 2pT/√s < 0.03, *i.e.*, large ln(xT)

NLO becomes monotonic again at small pJ
pJ < 100 GeV at 7 TeV
NLO not reliable?

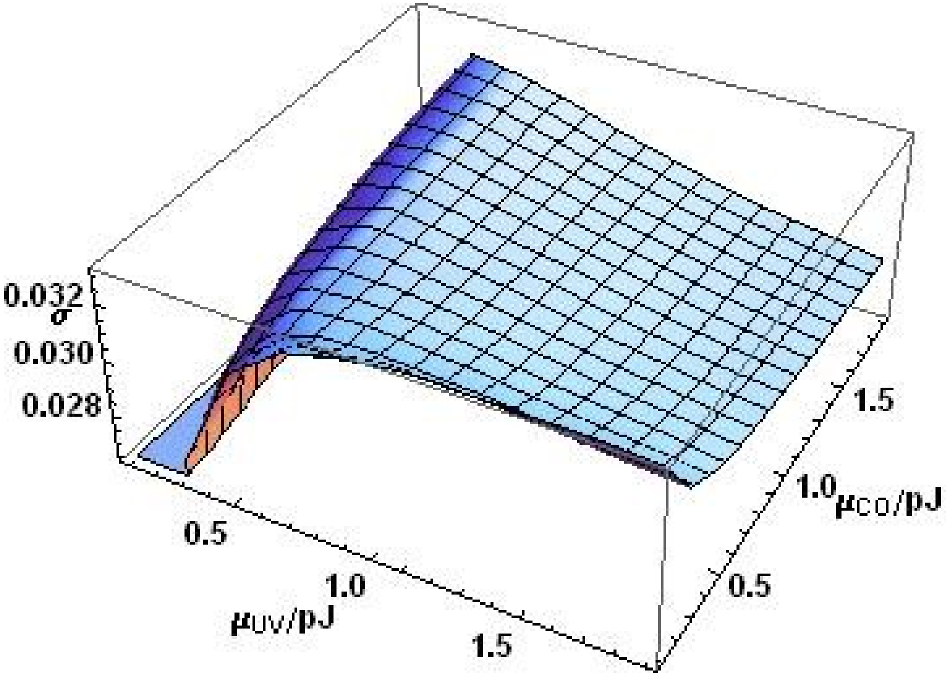




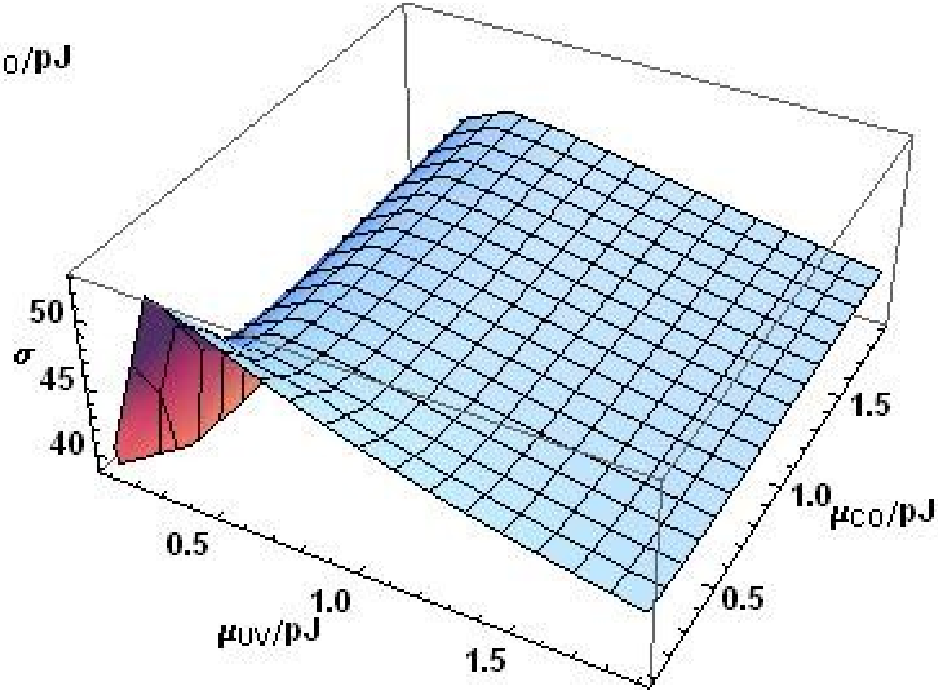
pT = 30 GeV

LO and NLO both monotonic

More detail in μ_{UV}, μ_{CO} plane



pJ = 270 GeV



pJ = 70 GeV

Large enhancement for $\mu_{CO} \to 0$

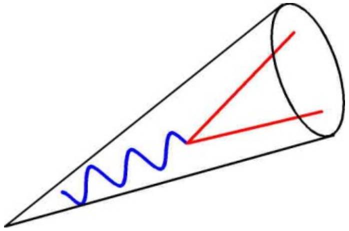
How reliable are theory "NO jet" rates, *i.e.*, jet vetos?

- E.g., in Higgs → j j and veto 3rd jet to reduce bkgs
- Calculate inclusive 2 jet – inclusive 3 – j, where 3rd jet has LOW pT MIN cut (the cut parameter in the veto.
- Can get large logs of ratios like $pT_1/pT_{3\ MIN}$, $\sqrt{s}/pT_{3\ MIN}$
- Theory uncertainties may be larger than previously expected

  See SCET analysis by Berger, Tackmann, Stewart, Waalewijn and Marcatonini – 1011.soon
- Need Theory Benchmarks for uncertainties (1 % goal)

Goals at LHC Different ⇒ Different Role for Jets!

- Find Physics Beyond the Standard Model (BMS), need < 10% precision!
- BSM Event structure likely different from QCD, more jets? Different structure within jets? Must be able to reconstruct masses from multi-jets & also from *single jets*
- Want to select events/jets by non-QCD-ness



- Highly boosted SM and non-SM particles –
  W, Z, top, Higgs, SUSY ⇒ *single* jet instead of
  2 or 3 jets, focus on masses and substructure of jets
- Much recent progress, but lots of work still to be done!!

Jet Masses in QCD: A Brief Review

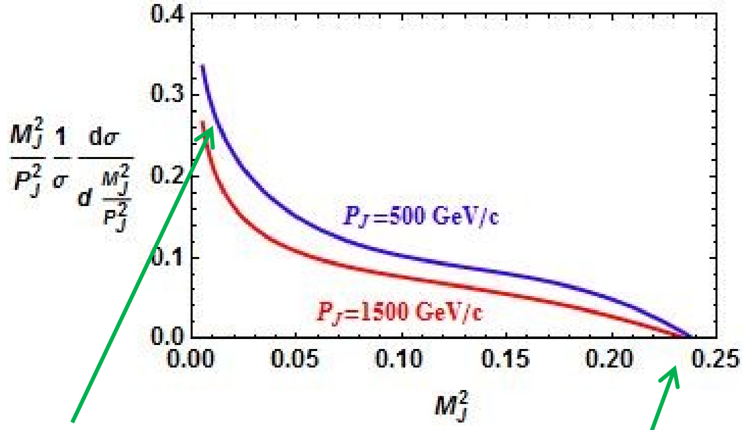
- In NLO PertThy

$$\sqrt{p_{J,\mu} p_J^{\mu}} \Rightarrow \sqrt{\left\langle M^2 \right\rangle_{NLO}} = f\left(\frac{p_J}{\sqrt{s}}\right)\sqrt{\alpha_s(p_J)}\,p_J D$$

Phase space from pdfs,
f ~ 1 & const

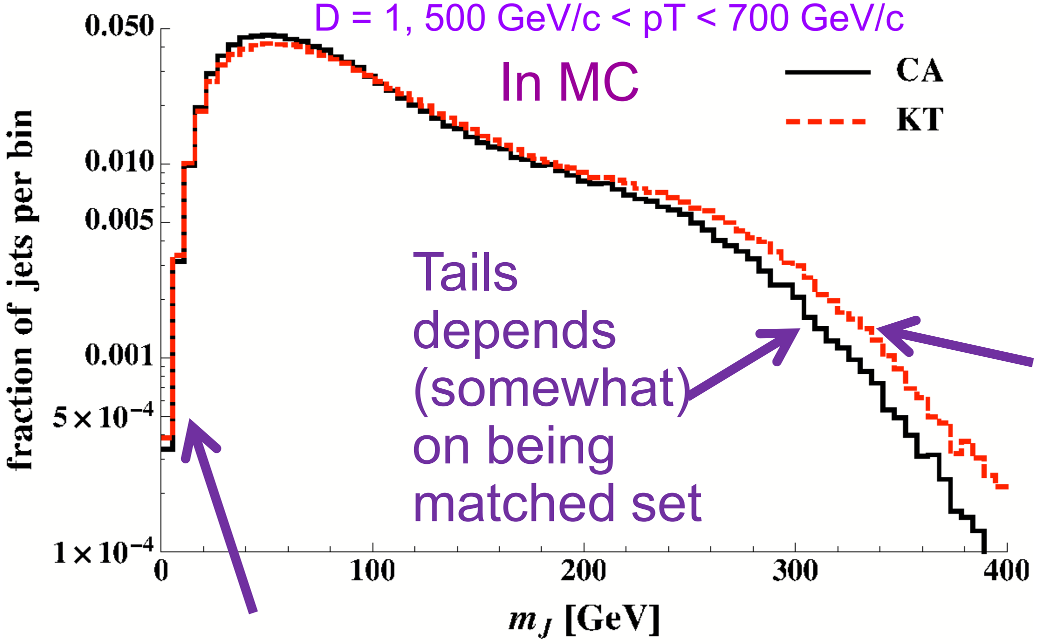
Dimensions

Jet Size, *D* = *R* ~ $\Delta\theta$, determined by jet algorithm



Peaked at low mass (log(m)/m behavior),

cuts off for $(M/P)^2 > 0.25 \sim D^2/4$ (M/P > 0.5) large mass can't fit in fixed size jet, QCD suppressed for M/P > 0.3 (~ $\gamma < 3$)



In MC

Tails depends (somewhat) on being matched set

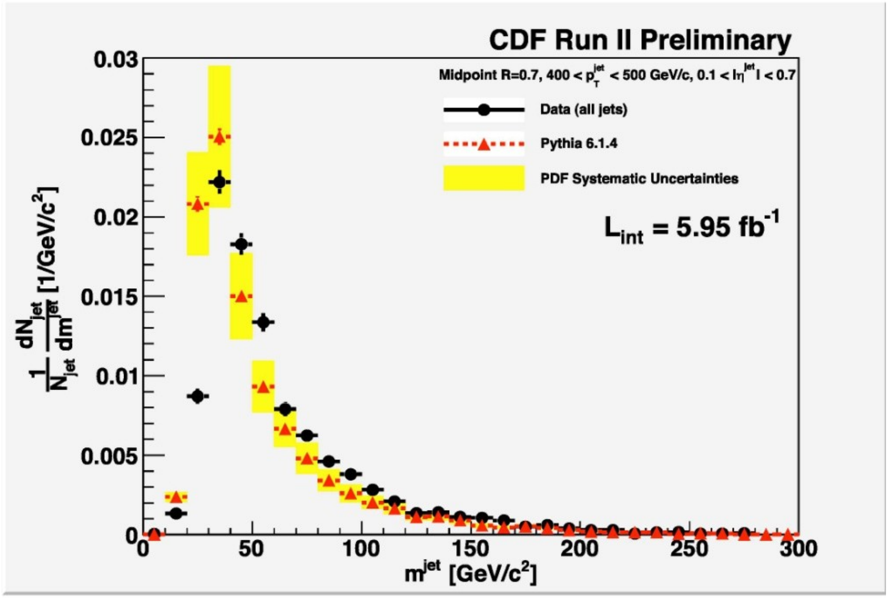
Algorithm matters

Turns over (Sudakov)

Benchmark this behavior in data!

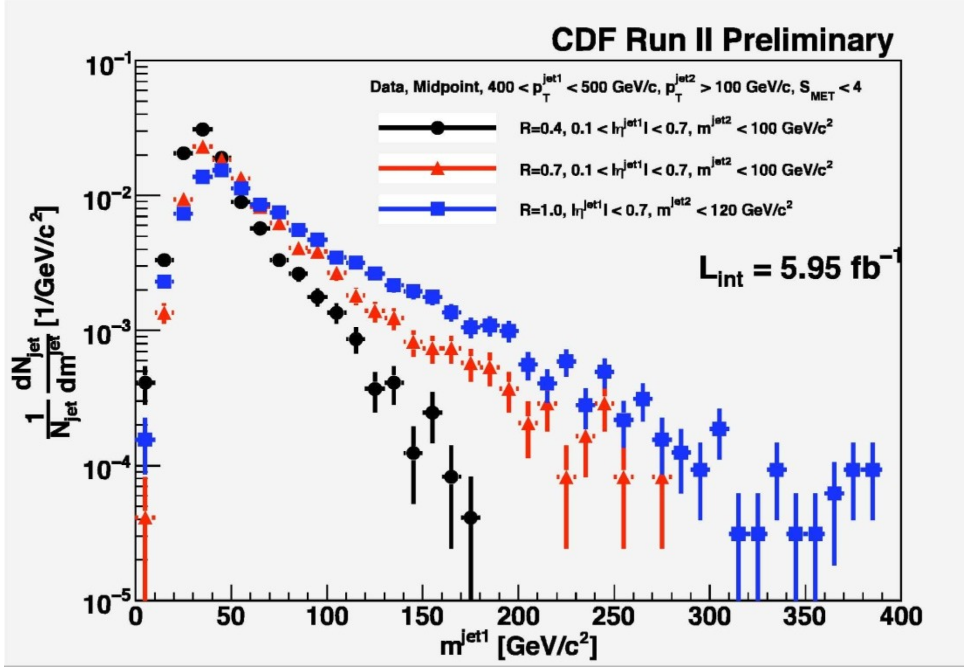
Jet Mass – CDF Data

(CDF/PUB/JET/PUBLIC/10199 7/19/10)



At least qualitatively the expected shape – masses slightly larger than MC – need the true hard emissions (as in matched sets)

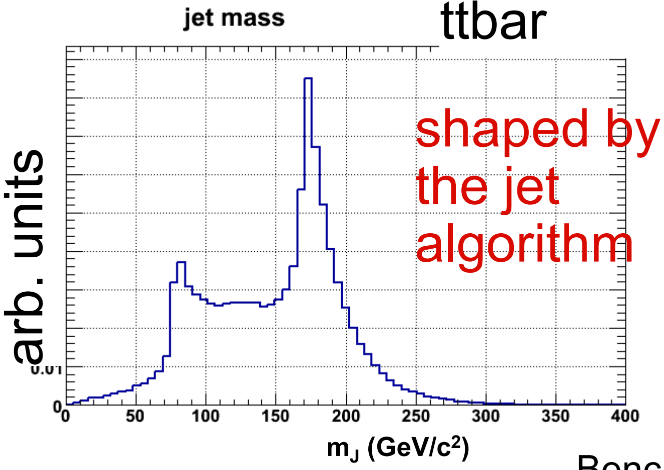
Large mass tail grows, as expected, with jet size parameter in the algorithm - You find what you look for!



Want to do Heavy Particles Searches with Single Jets -

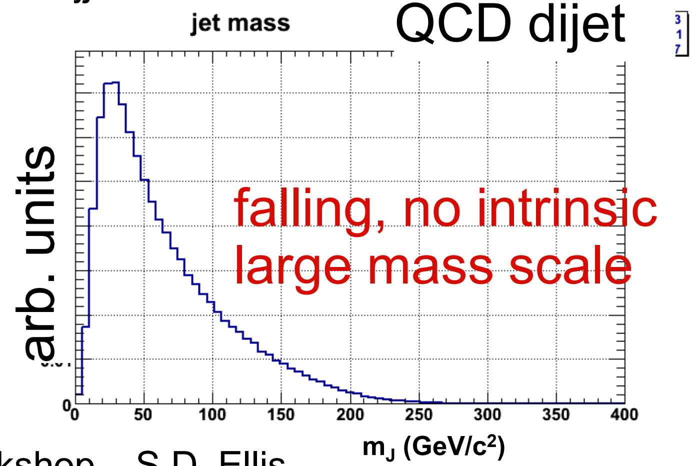
- QCD multijet production rate >> production rate for heavy particles
- In the jet mass spectrum, production of non-QCD jets may appear as local excesses (bumps!) but must be enhanced using analyses
- Use jet substructure as defined by **recombination algorithms** ($\alpha \geq 0$, not anti-kT) to refine jets
- Algorithm will systematically shape distributions
- Example - top quark as surrogate new particle.

$$\sigma_{ttbar} \approx 10^{-3}\sigma_{jj}$$



ttbar

shaped by the jet algorithm



QCD dijet

falling, no intrinsic large mass scale

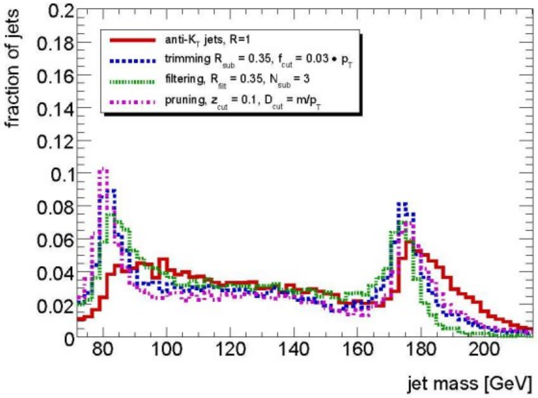
Jet Substructure – Need to Benchmark at the Tevatron and LHC –

- Jet Grooming – “cleanup” jet to make any inherent mass scale more apparent (bump in mass distribution) – also reduces impact of UE, PU and algorithm details (pruning, trimming, filtering)
- Jet Tagging – select for specific substructure characteristic of search target, certify with top quark, W/Z
- More inclusive jet shape measures like jet angularities, jet shape templates, N-subjettiness

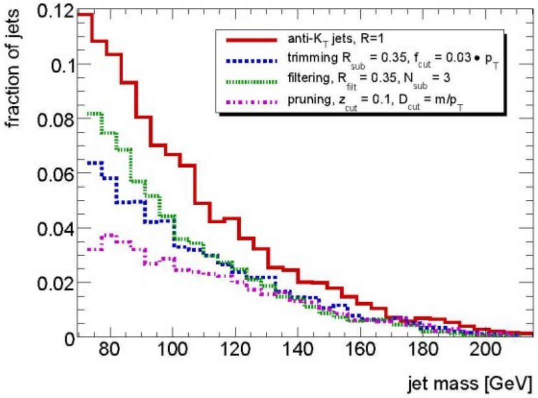
Results from Boost 2010 (to appear)

mass distribution in pT bins

300-
400



(a) $t\bar{t}$, 300–400 GeV

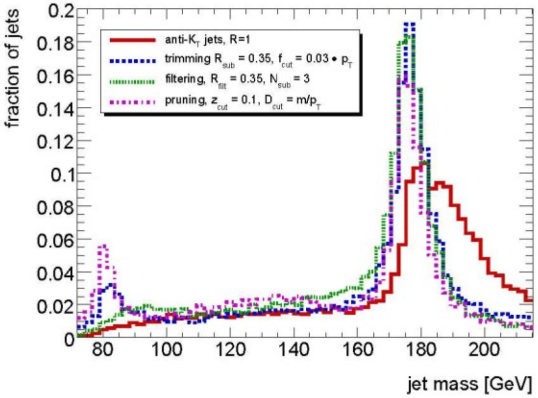


(b) dijets, 300–400 GeV

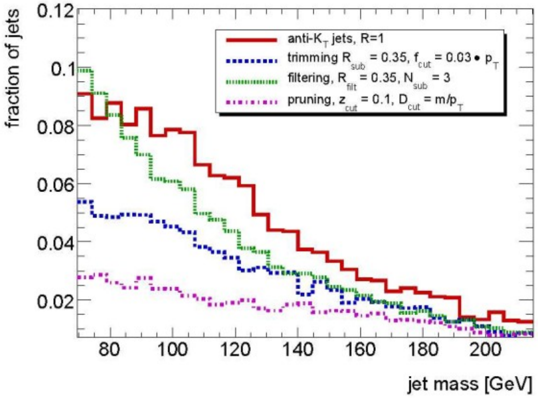
Narrower
peaks

Suppressed
background

500-
600

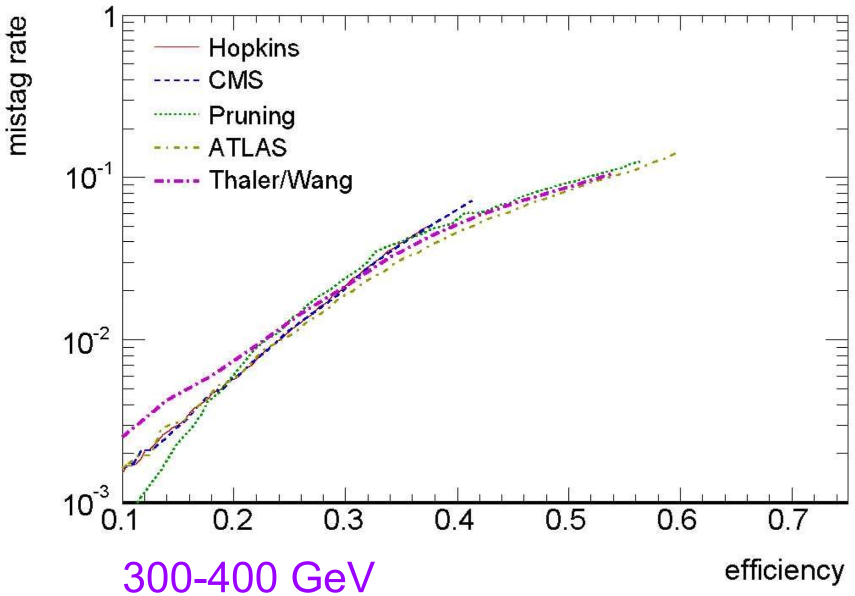


(c) $t\bar{t}$, 500–600 GeV

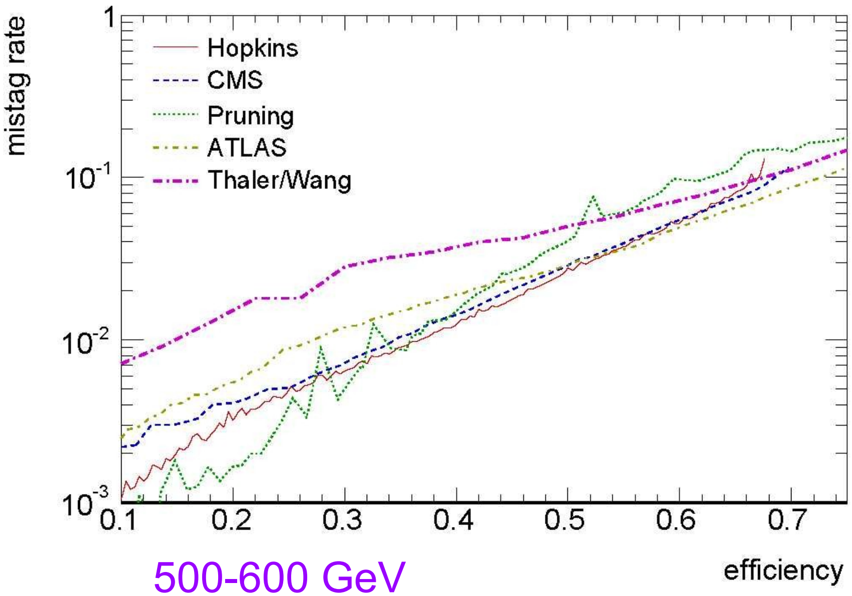


(d) dijets, 500–600 GeV

Compare top taggers



300-400 GeV



500-600 GeV

Similar exponential correlation for all “optimized” taggers (lowest mistag rate for given efficiency)

- ATLAS, T/W use similar kT reclustering and no grooming
- CMS, Hopkins use similar C/A reclustering and filtering

Summary/Conclusions:

- It will take time to understand the SM at the LHC, but we understand jets much better now than we did at the beginning of Run I
- It is essential to test and validate a variety of jet algorithms (*i.e.*, define Benchmarks) – the familiar ones like cones, whose issues we need to re-confirm, and the less familiar ones like Anti-kT, whose issues we need to uncover – different algorithms find (slightly) different jets and will likely have different uses
- It is essential that the different Collaborations document the algorithms they use – and try to use the same ones some of the time
- It is essential to study and understand the role of the Underlying Event and Pile-Up (splash-in) and Showering and Hadronization (splash-out) in jets at the LHC

Summary/Conclusions:

- In comparing to perturbative QCD results, it is important to let the calculation define the appropriate scale. When logs are large, it will be important to sum them. Soft Collinear Effective Theory (SCET) techniques will likely be useful!

- It is essential to study and understand the properties of jets – masses and substructure – validate *Taggers* and *Groomers* by IDing top jets, W/Z jets in Tevatron & LHC data

  Study other shape information like angularities

  ⇒ single jets and their substructure will play a role in the search for BSM physics, along with heavy flavor tags, correlations with other jets (pair production), MET, *etc*.

Experimenters & theorists need to work together!!

Extra Detail Slides

Looking for the hidden truth – a jet performs!!

But it is never that simple!!!

Seeds and Sensibility – An Important Lesson for Theory – Experiment interaction

- Tension between desire

  To Limit analysis time (for experiments) with seeds

  To Use identical algorithms in data and perturbation theory

- Seeds are intrinsically IR sensitive (MidPoint Fix only for NNLO, not NNNLO)

⇒ DON'T use seeds in perturbation theory, correct for them in data analysis, or better, USE SEEDLESS Algorithm (SISCone)!!

In the theory they are a big deal – IR UNsafety (Yikes)!!!!!!

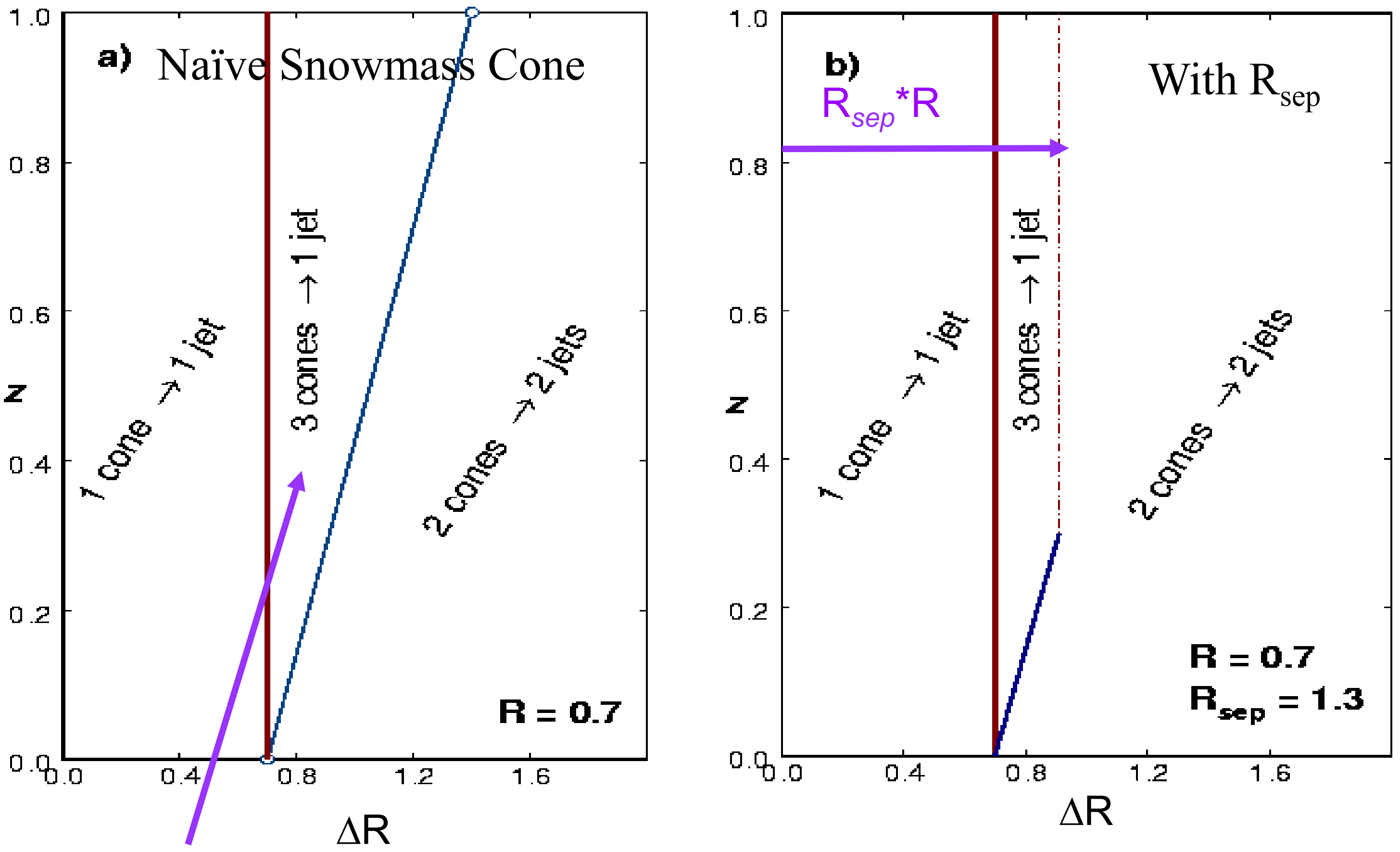
In the data seeds vs seedless is a few % correction (e.g., lower the Seed p_T threshold) and this is small compared to other corrections

Remember this lesson at the LHC!!

2 Parton (NLO) Phase Space

- Seeds can mean missed configurations with 2 partons in 1 Jet, NLO Perturbation Theory – ΔR = parton separation, $z = p_2/p_1$,,

  Simulate the missed middle cones with R_{sep} (not very helpful at higher orders)

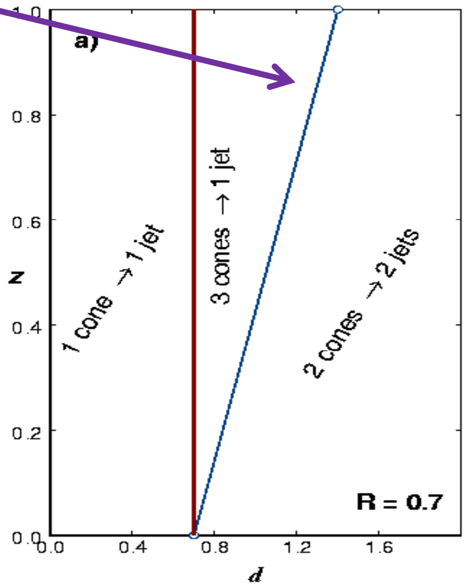


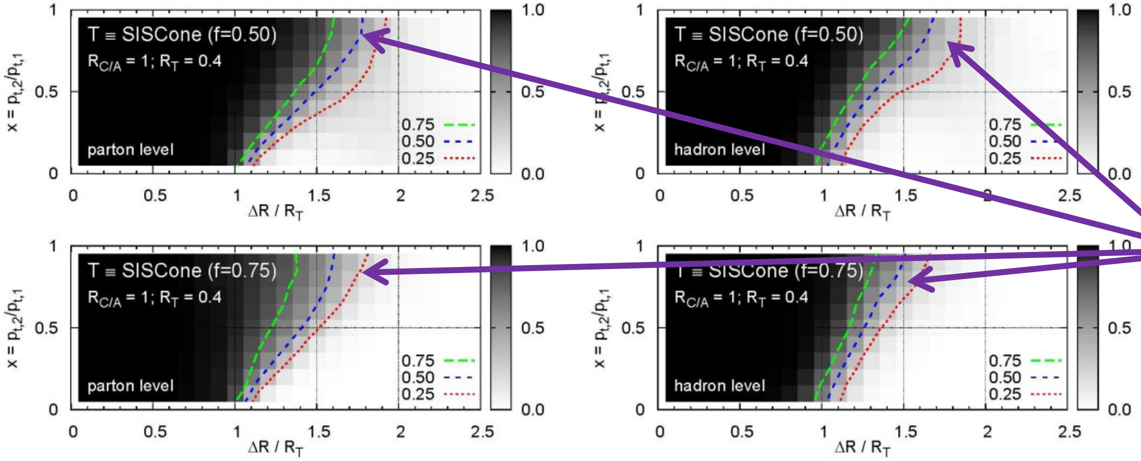
~10% of cross section here

Cone Lesson - (Even if Seedless)

3) Splash-out from smearing of energetic parton at edge of cone – can be quantitatively relevant

- Study by G. Salam* – find 2 (sub)jets with a different algorithm and see when SISCone merges them





(Shower only)

(Shower + Had-n)

Seldom merge in the corner, another reason to cut out corner (R_{sep})

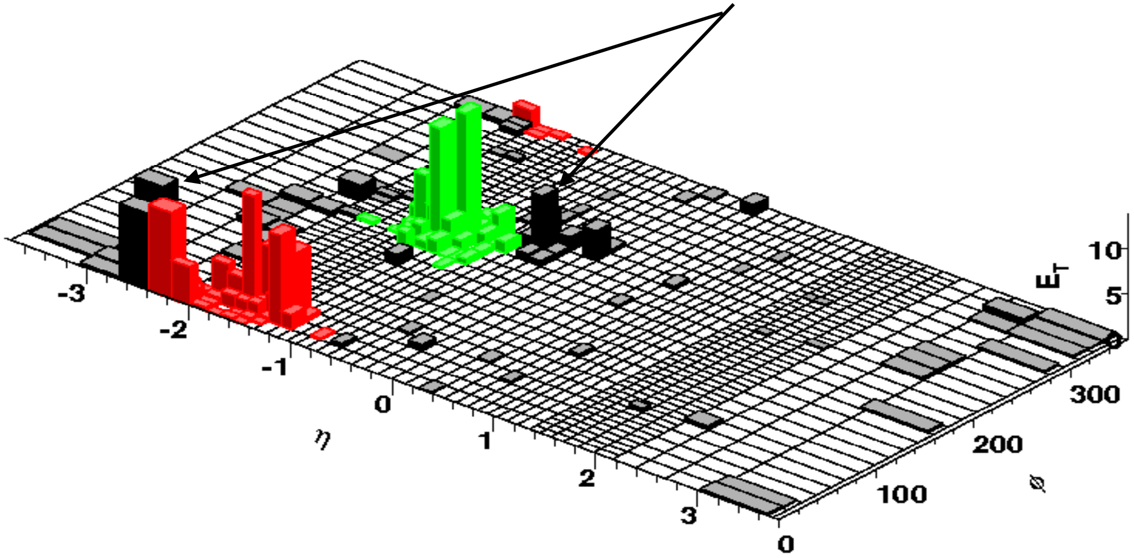
*Similar to earlier CDF study

(Related) Cone Lesson: (Even if Seedless)

4) Dark Towers - Energy in secondary showers may not be clustered in any jet

- Expected stable cone *not* stable due to smearing from showering/hadronization (compared to PertThy)

- Under-estimate E_T (~ 5% effect for jet cross section)



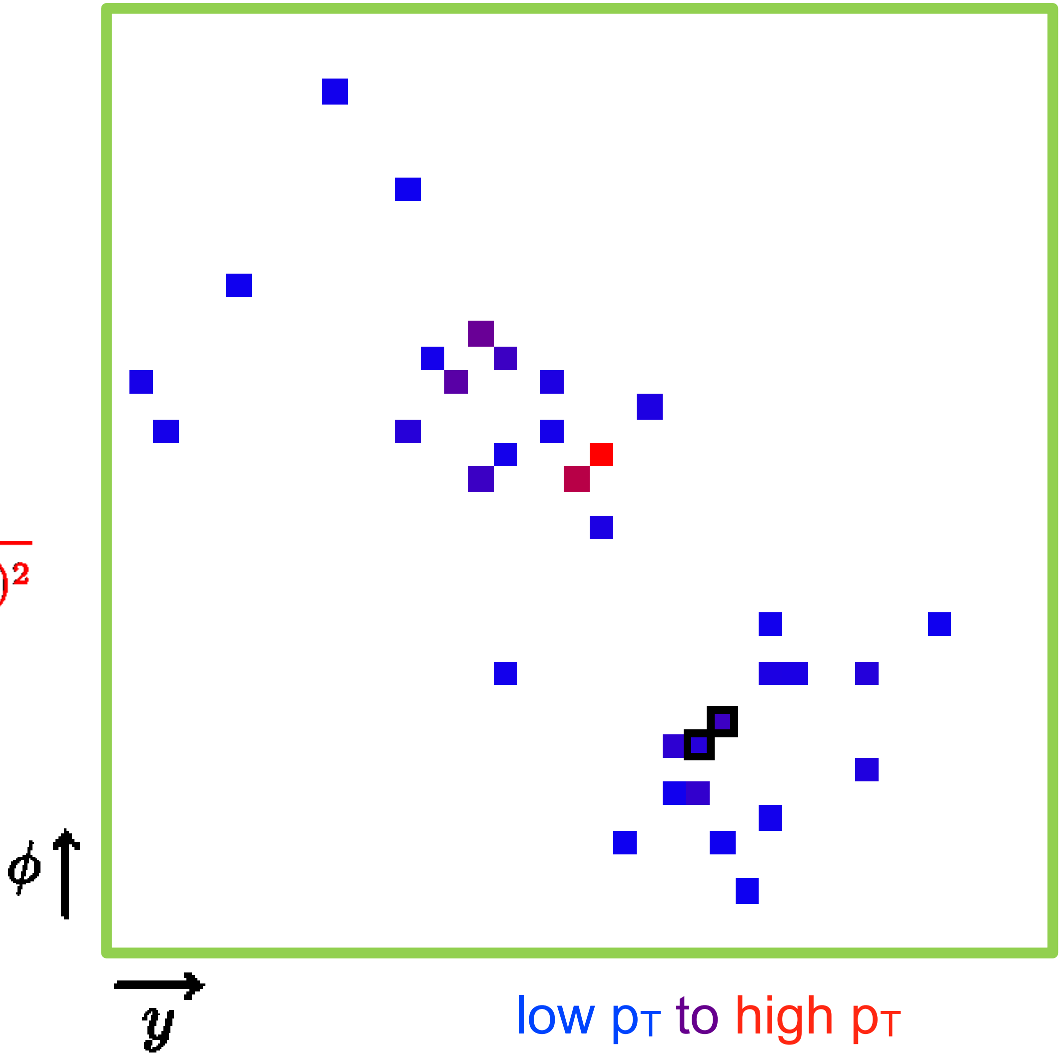
Using Recombination Algorithms at the LHC – Here CA algorithm in action – “natural” substructure at each merging!

Think of starting with calorimeter cells, recombine “closest” pair at each step leading to larger p_T

For CA close in quantity

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

(0.05 x 0.05) Cells with E > 1 GeV



low p_T to high p_T

Recall Jets History at Hadron Colliders

- JETS I – Cone style jets applied to data at the SpbarpS, and Run I at the Tevatron to map final state hadrons onto LO (or NLO) hard scattering, initially 1 jet ⇔1 parton (test QCD)

Little attention paid to masses of jets or the internal structure, except for energy distribution within a jet

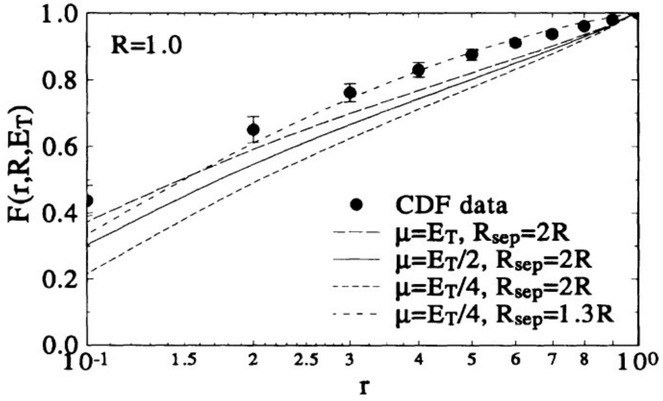


FIG. 2. $F(r,R,E_T)$ vs r for $R=1.0$, $\sqrt{s}=1800$ GeV, $E_T=100$ GeV, and $0.1<|\eta|<0.7$ with $\mu=E_T/4$, $E_T/2$, E_T compared to data from CDF [7]; the dot-dashed curve is explained in the text.

- JETS II – Run II & LHC, starting to look at structure of jets: masses and internal structure – a jet renaissance

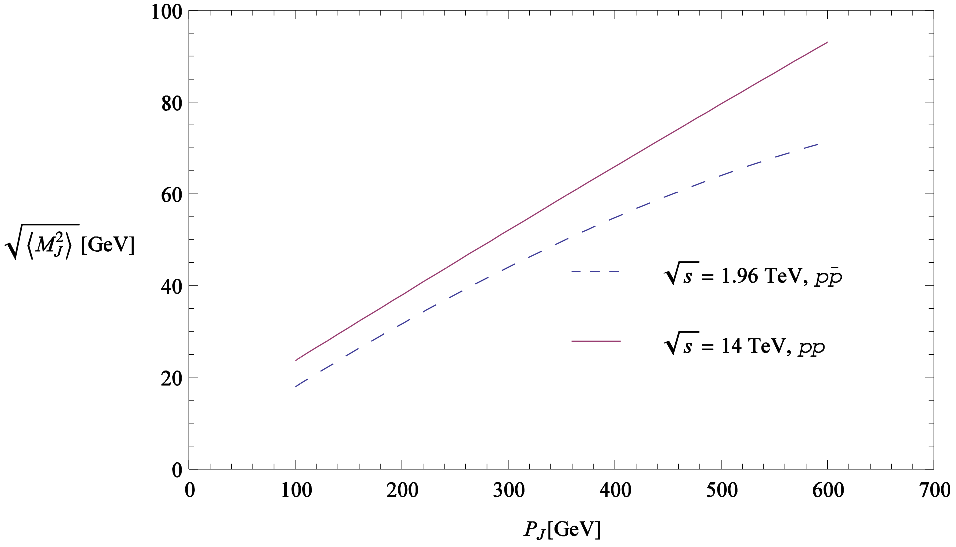
Jet Masses in pQCD:

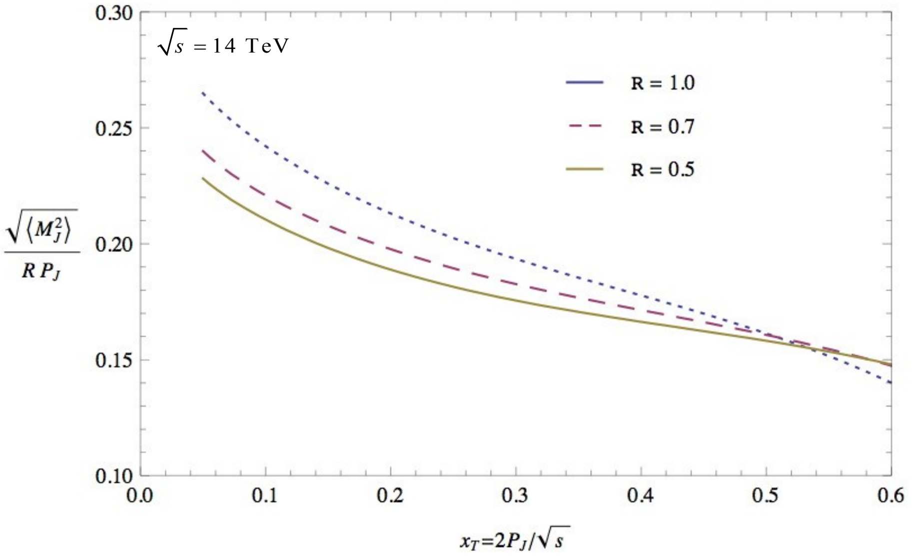
Dimensions

- In NLO PertThy $\sqrt{p_{J,\mu} p_J^{\mu}} \Rightarrow \sqrt{\left\langle M^2 \right\rangle_{NLO}} = f\left(\frac{p_J}{\sqrt{s}}\right)\sqrt{\alpha_s(p_J)}\,p_J R$

Phase space from dpfs, *f* ~ 1

Jet Size, *R*, *D* ~ $\Delta\theta$, determined by jet algorithm



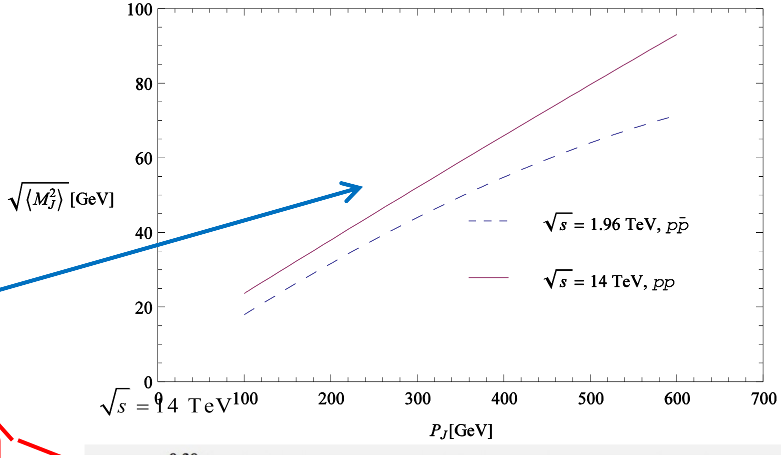


Useful QCD "Rule-of-Thumb" $\Rightarrow \sqrt{\left\langle M^2 \right\rangle_{NLO}} \sim 0.2\, p_J R$

Jet Masses in QCD: A Brief Review

- In NLO PertThy

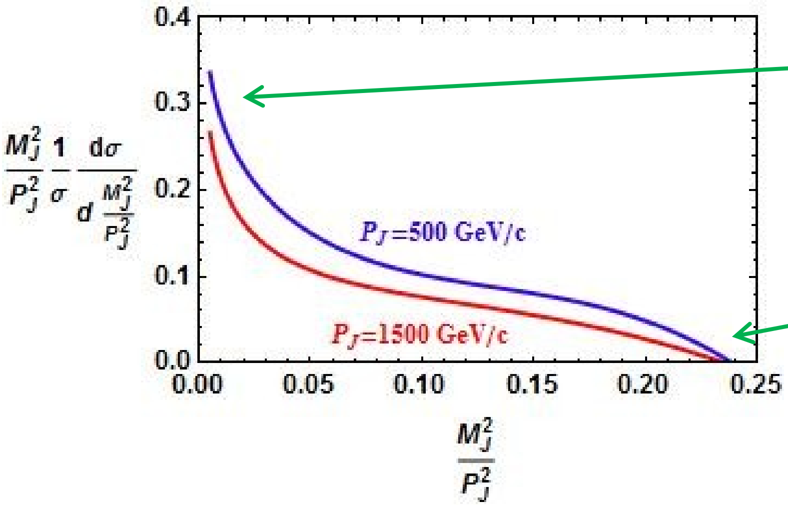
$$\sqrt{p_{J,\mu}p_J^{\mu}} \Rightarrow \sqrt{\left\langle M^2\right\rangle_{NLO}} = f\left(\frac{p_J}{\sqrt{s}}\right)\sqrt{\alpha_s\left(p_J\right)}\,p_J D$$



Phase space from pdfs,
f ~ 1 & const

Dimensions

Jet Size, *D* = *R* ~ $\Delta\theta$, determined by jet algorithm



Peaked at low mass (log(m)/m behavior),

cuts off for $(M/P)^2 > 0.25 \sim D^2/4$ (M/P > 0.5) large mass can't fit in fixed size jet, QCD suppressed for M/P > 0.3 (~ $\gamma < 3$)

Want heavy particle boosted enough to be in a jet (use large-ish D ~1), but not so much to be QCD like (~ $2 < \gamma < 5$)

Soft – Collinear pole version

Useful $$-\ln\left(1-\sqrt{1-4\frac{M_J^2}{P_J^2}\frac{1}{D^2}}\right)\Theta\left(\frac{1}{4}-\frac{M_J^2}{P_J^2}\frac{1}{D^2}\right)_{,}^{-} \sim 0.2\,p_J D\left(1\pm 0.25\right)$$