Issues for Jet Algorithms: A Quick Review, (But No Quick Answers)

(Thanks especially to Joey Huston & Matthias Tönnesmann)

The Goal is 1% Strong Interaction Physics (where Run I was ~ 10%)

Using Jet Algorithms we want to *precisely* map

- What we can measure, *e.g.*, $E(y,\phi)$ in the detector

On To

- What we can calculate, *e.g.*, arising from small numbers of partons as functions of E, y,ϕ

➢ We "understand" what happens at the level of partons and leptons, *i.e.*, LO theory is simple, can reconstruct masses, *etc.*

➢ We want to map the observed (hadronic) final states onto a representation that mimics the kinematics of the energetic partons; ideally on a event-by-event basis.

➢ But we know that the (short-distance) partons shower (perturbatively) and hadronize (nonperturbatively), *i.e.*, spread out.

We want to associate "nearby" hadrons or partons into JETS (account for spreading)

- Renders PertThy IR & Collinear Safe

- Nearby in angle – Cone Algorithms, *e.g.*, Snowmass (main focus here)

- Nearby in momentum space – k_T Algorithm

⇒But mapping of hadrons to partons can *never* be 1 to 1, event-by-event!

colored states ≠ singlet states

Think of the algorithm as a "microscope" for seeing the (colorful) underlying structure -

Fundamental Issue

Warning:

We must all use the same algorithm!!
(as closely as inhumanly possible)

In the Beginning - *Snowmass Cone Algorithm*

- **Cone Algorithm – particles, calorimeter towers, partons in cone of size R, defined in angular space, *e.g.*, (η,φ)**
- **CONE center - (η^C,φ^C)**
- **CONE i $\in$ *C* *iff*** $$\sqrt{\left(\eta^i-\eta^C\right)^2+\left(\varphi^i-\varphi^C\right)^2}\leq R$$
- **Energy** $$E_T^C=\sum_{i\subset C}E_T^i$$
- **Centroid** $$\overline{\eta}^C=\sum_{i\subset C}E_T^i*\eta^i/E_T^C\ ;\ \overline{\varphi}^C=\sum_{i\subset C}E_T^i*\varphi^i/E_T^C$$

- *"Flow vector"* $\vec{F}^C = \left(\overline{\eta}^C - \eta^C, \overline{\varphi}^C - \varphi^C\right)$
- **Jet is defined by "stable" cone*:***

$$\eta^J = \eta^C = \overline{\eta}^C \;\; ; \;\; \varphi^J = \varphi^C = \overline{\varphi}^C \;\; ; \;\; \vec{F}^C = 0$$

- **Stable cones found by iteration: start with cone anywhere (and, in principle, *everywhere*), calculate the centroid of this cone, put new cone at centroid, iterate until cone stops "flowing", *i.e.*, stable ⇒ Proto-jets (prior to split/merge)**

⇒ *unique, discrete jets event-by-event* (at least in principle)

k_T Algorithm

- **Combine partons, particles or towers pair-wise based on "closeness" in momentum space, beginning with low energy first.**
- **Jet identification is unique – no merge/split stage**
- **Resulting jets are more amorphous, energy calibration difficult (subtraction for UE?), and analysis can be very computer intensive (time grows like N^3)**

Run I Issues (History):

Cone: **Seeds – only look for jets under brightest street lights,** ***i.e.*****, near very active regions**

⇒ problem for theory, IR sensitive at NNLO

Stable Cones found by iteration (E_T weighted centroid = geometric center) can Overlap,

⇒ require Splitting/Merging

To understand the issues consider Snowmass "Potential"

- In terms of 2-D vector $\vec{r} = (\eta, \varphi)$ or (y, φ) define a potential

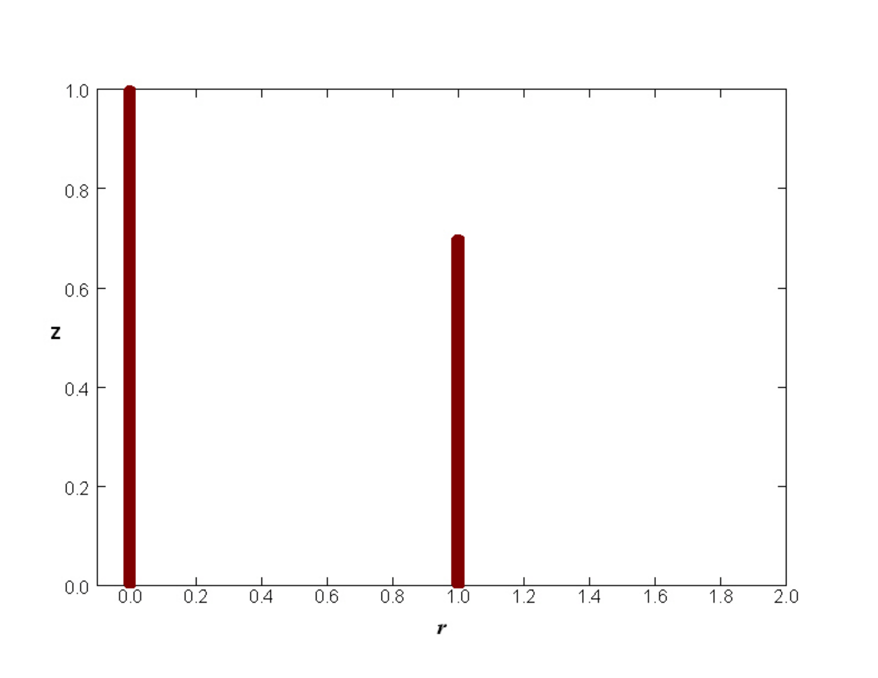
$$V(\vec{r}) \equiv -\frac{1}{2}\sum_i E_T^i \left(R^2 - (\vec{r}^{\,i} - \vec{r})^2\right) \Theta\left(R^2 - (\vec{r}^{\,i} - \vec{r})^2\right)$$

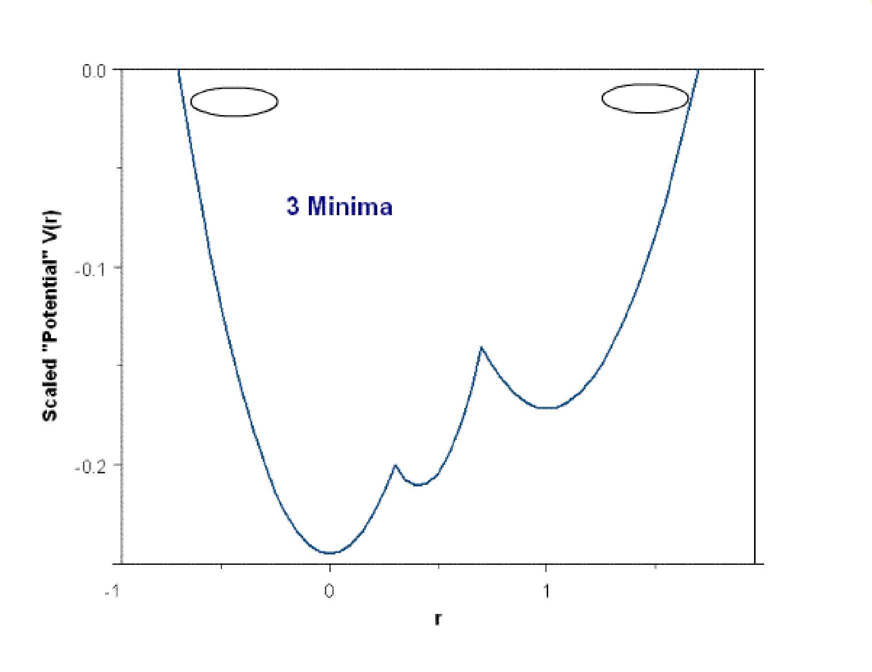
- Extrema are the positions of the stable cones; gradient is "force" that pushes trial cone to the stable cone, *i.e.*, the flow vector

$$\vec{F}(\vec{r}) = -\vec{\nabla} V(\vec{r}) = \sum_i E_T^i (\vec{r}^{\,i} - \vec{r}) \Theta\left(R^2 - (\vec{r}^{\,i} - \vec{r})^2\right)$$

Simple Theory Model - 2 partons (separated by < 2R): yield potential with 3 minima – trial cones will migrate to minima from seeds near original partons

$\Rightarrow$ miss central minimum





$z = p_{min} / p_{max}$, r = separation

Smearing of order R

Run I Issues (History):

Cone: **Seeds – only look for jets under brightest street lights, *i.e.*, near very active regions**

⇒ problem for theory, IR sensitive at NNLO

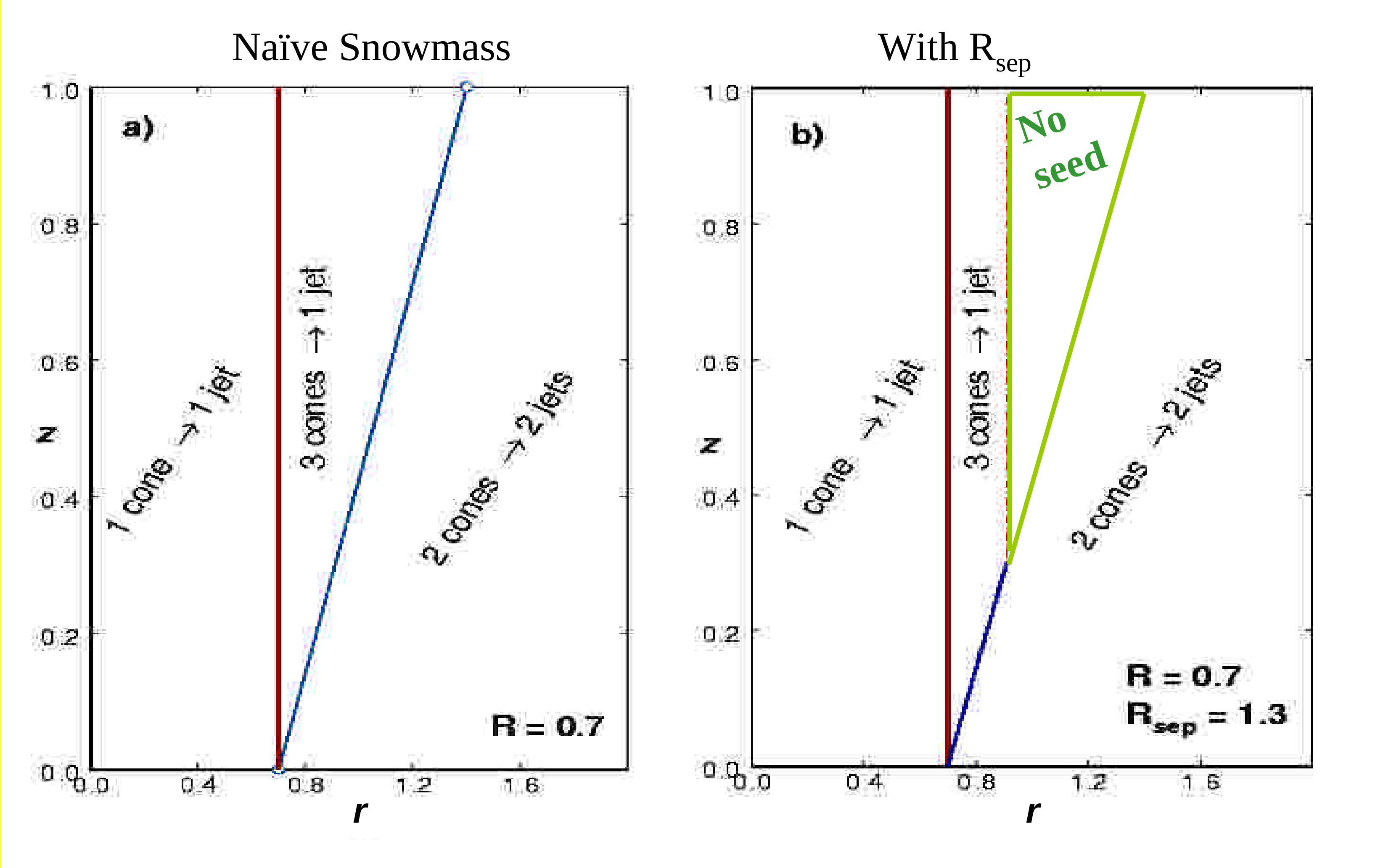
Stable Cones found by iteration (E_T weighted centroid = geometric center) can Overlap,

⇒ require Splitting/Merging scheme
⇒ Different in different experiments

⇒ Don't find "possible" central jet between two well separated proto-jets (partons)

⇒"simulate" with R_{SEP} parameter in theory

NLO Perturbation Theory – r = parton separation, $z = p_2/p_1$
R_{sep} simulates the cones missed due to no middle seed



Run I Issues (History):

Cone: **Seeds – only look for jets under brightest street lights, *i.e.*, near very active regions**

⇒ problem for theory, IR sensitive at NNLO

Stable Cones found by iteration (E_T weighted centroid = geometric center) can Overlap,

⇒ require Splitting/Merging scheme
⇒ Different in different experiments

⇒ Don't find "possible" central jet between two well separated proto-jets (partons)

⇒"simulate" with R_{SEP} parameter in theory

Kinematic variables: $E_{T,Snow} \neq E_{T,CDF} \neq E_{T,4D} = p_T$ – Different in different experiments and in theory

For example, consider 2 partons: $p_1 = zp_2$

$$E_{T,scalar} = E_{T,Snow} = p_1 + p_2$$

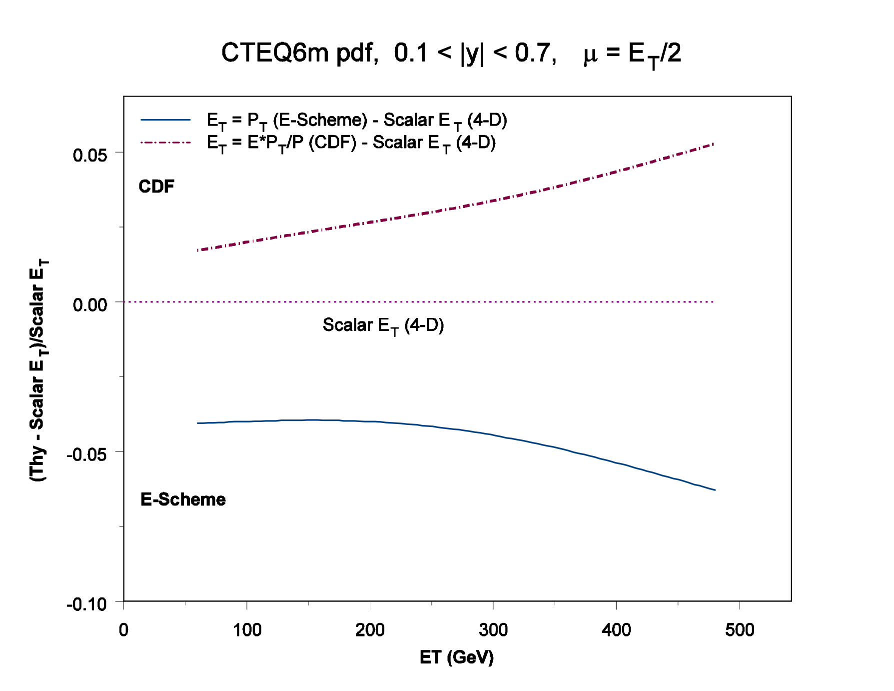
$$E_{T,4D} \equiv \left|\vec{P}_{J,T}\right| = \sqrt{p_1^2 + p_2^2 + 2p_1p_2\cos\Delta\phi}$$

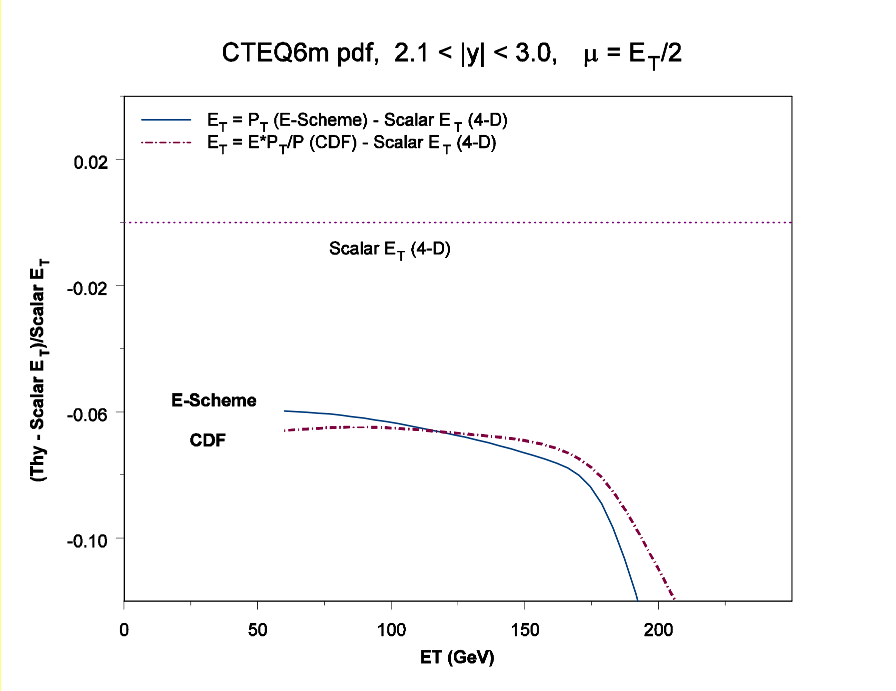
$$= E_{T,Snow}\frac{\sqrt{1+z^2+2z\cos\Delta\phi}}{1+z} \le E_{T,Snow}$$

$$E_{T,CDF} \equiv E_J\sin\theta_J = E_J\frac{\left|\vec{P}_{J,T}\right|}{\left|\vec{P}_J\right|} = E_{T,4D}\frac{\sqrt{\left|\vec{P}_J\right|^2 + M_J^2}}{\left|\vec{P}_J\right|} \ge E_{T,4D}$$

$\Rightarrow$ **mass dependence – the soft stuff**

5% Differences (at NLO) !!





(see later)

"HIDDEN" issues, detailed differences between experiments

- **Energy Cut on towers kept in analysis (*e.g.*, to avoid noise)**
- **(Pre)Clustering to find seeds (and distribute "negative energy")**
- **Energy Cut on precluster towers**
- **Energy cut on clusters**
- **Energy cut on seeds kept**

+ **Starting with seeds find stable cones by iteration, but in JETCLU (CDF), "once in a seed cone, always in a cone", the <u>"ratchet"</u> effect**

<u>Overlap</u>: stable cones must be split/merged

Depends on overlap parameter f_{merge}

Order of operations matters

All of these issues impact the content of the "found" jets

- Shape may not be a cone
- Number of towers can differ, *i.e.*, different energy
- Corrections for underlying event must be tower by tower

Detailed Differences mean Differences in:

- Impact of UE contributions
- Impact of calorimeter info *vs* tracking info
- Impact of Non-perturbative hadronization (& showering) compared to PertThy
- (Potential) Impact of Higher orders in perturbation theory

Fundamental Issue

Warning:

We must all use the same algorithm!!

(as closely as inhumanly possible)

To address these issues, the Run II Study group Recommended

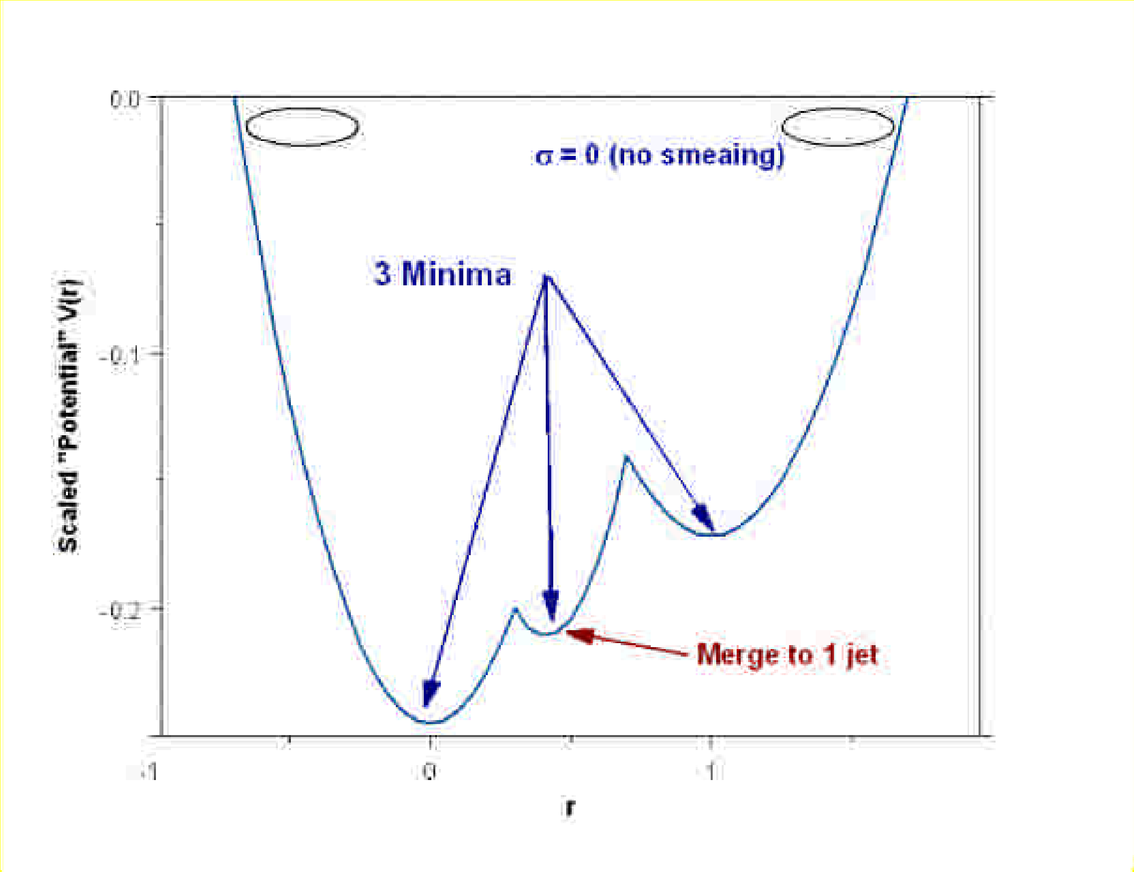
Both experiments use

- **(legacy) Midpoint Algorithm – always look for stable cone at midpoint between found cones**
- **Seedless Algorithm**
- **k_T Algorithms**

- **Use identical versions except for issues required by physical differences (in preclustering??)**

- **Use (4-vector) E-scheme variables for jet ID and recombination**

Consider the corresponding "potential" with 3 minima, expect via MidPoint or Seedless to find middle stable cone



Use common Split/Merge Scheme for Stable Cones

- Process stable cones in decreasing energy order, pair wise

- f_{merge} = 0.50% (< 0.75% in JETCLU);
  Merge if shared energy > f_{merge}, Split otherwise

- Split/Merge is iterative, starting again at top of reordered list after each split/merge event ($\neq$ JETCLU which is a "single-pass" scheme, no reordering)

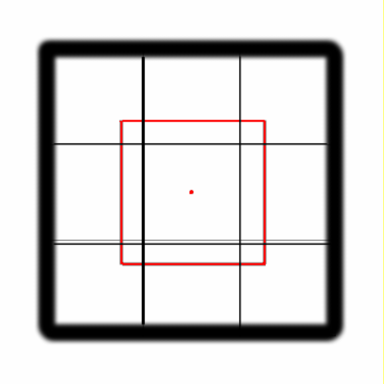
  $\Rightarrow$ Enhance the merging fraction wrt JETCLU (see later)

Streamlined Seedless Algorithm

- **Data in form of 4 vectors in (η,φ)**
- **Lay down grid of cells (~ calorimeter cells) and put trial cone at center of each cell**
- **Calculate the centroid of each trial cone**
- **If centroid is outside cell, remove that trial cone from analysis, otherwise iterate as before**
- **Approximates looking everywhere; converges rapidly**
- **Split/Merge as before**

Run II Issues

- k_T – “vacuum cleaner” effect accumulating “extra” energy – Does it over estimate E_T?
- “Engineering” issue with streamlined seedless – must allow some overlap or lose stable cones near the boundaries (M. Tönnesmann)



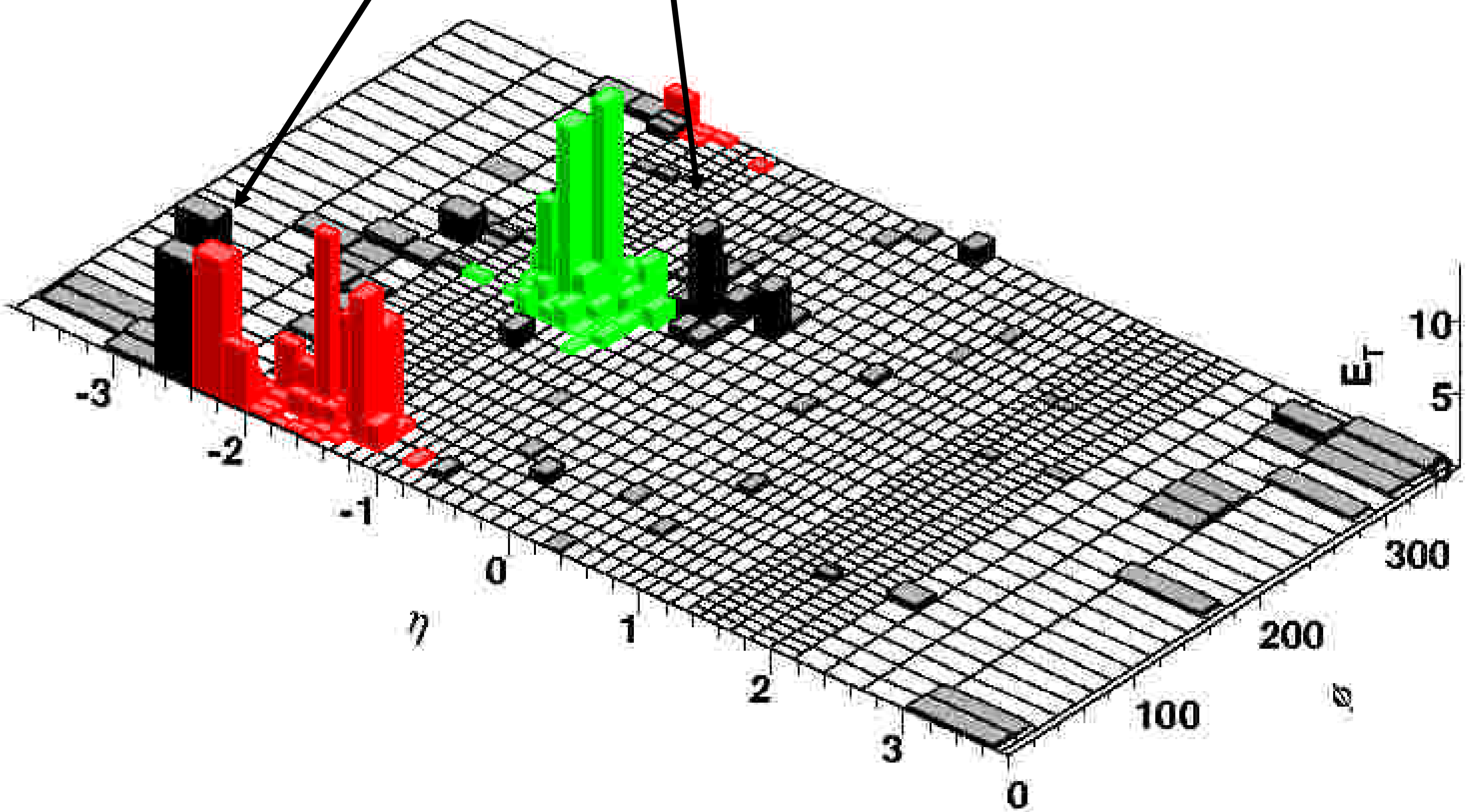
A NEW issue for Midpoint & Seedless Cone Algorithms

- Compare jets found by JETCLU (with ratcheting) to those found by MidPoint and Seedless Algorithms

- "Missed Energy" – when energy is smeared by showering/hadronization do not always find stable cones expected from perturbation theory

  ⇒ 2 partons in 1 cone solutions
  ⇒ or even second cone

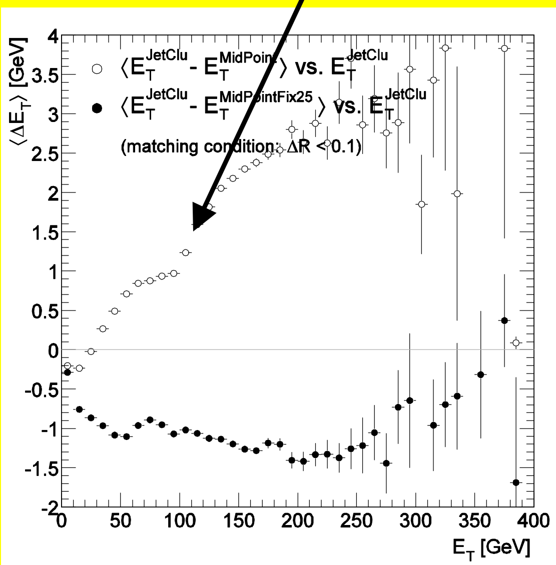
Missed Towers (not in any stable cone) – How can that happen? Does DØ see this?



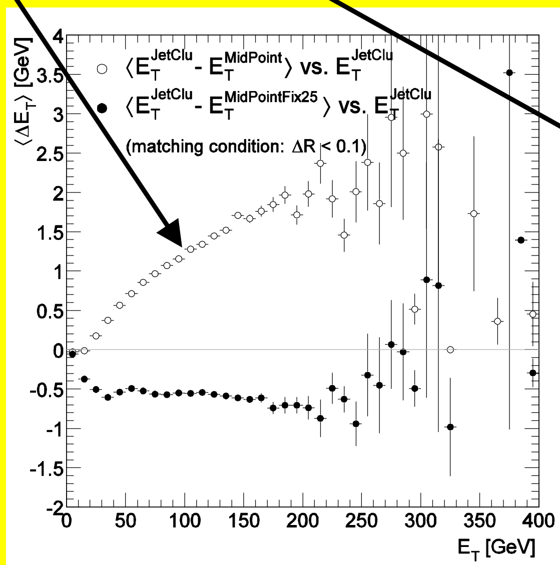
Match jets found by 2 algorithms, Compare E_T Lost Energy $E_T^{MidPoint} < E_T^{JETCLU}$!?

($\Delta E_T/E_T$~1%, $\Delta\sigma/\sigma$~5%)

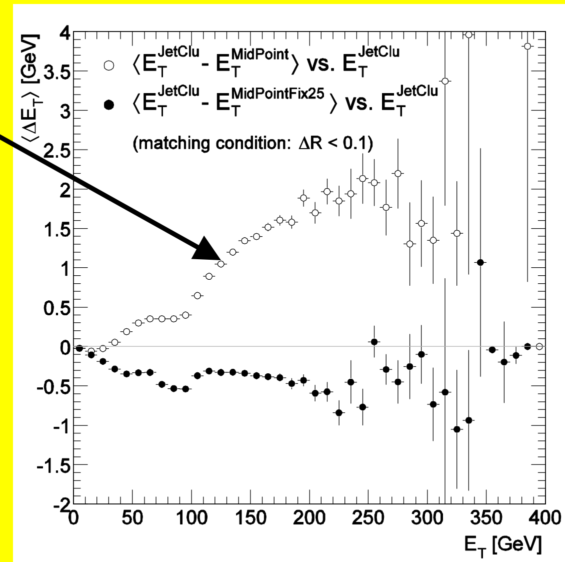
Partons



Calorimeter



Hadrons

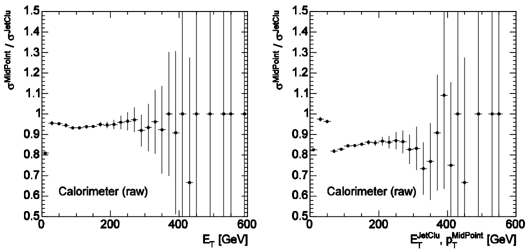


Note Differences between graphs

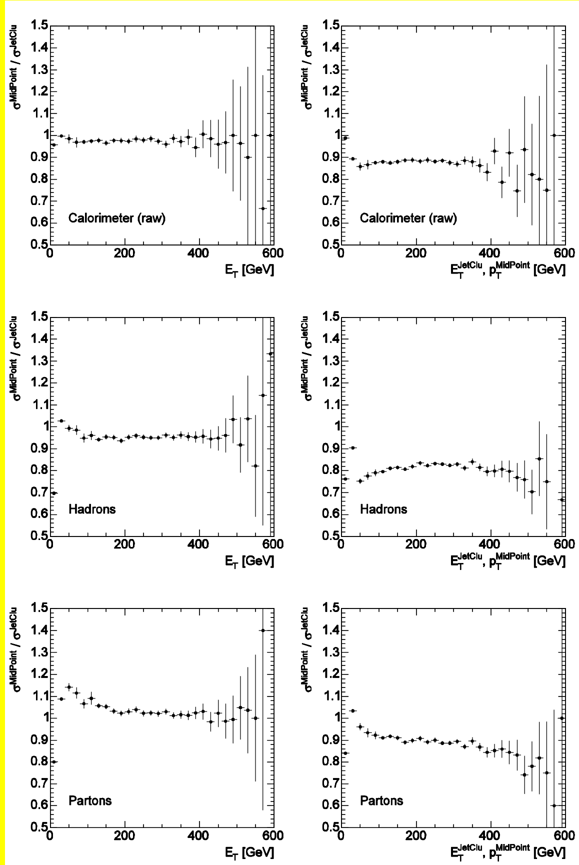
Results from M. Tönnesmann

$\sigma_{MidPoint}/\sigma_{JetClu}$

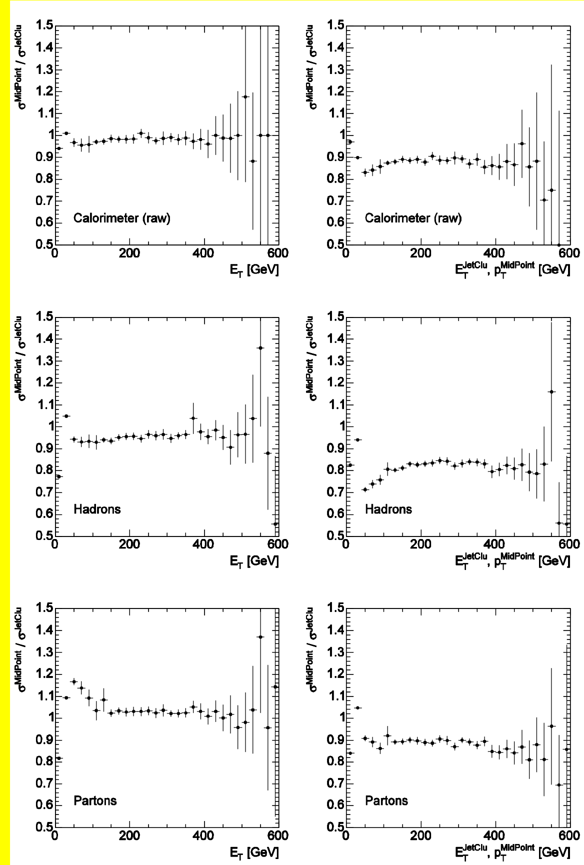
DATA



HERWIG



PYTHIA



Cal

Had

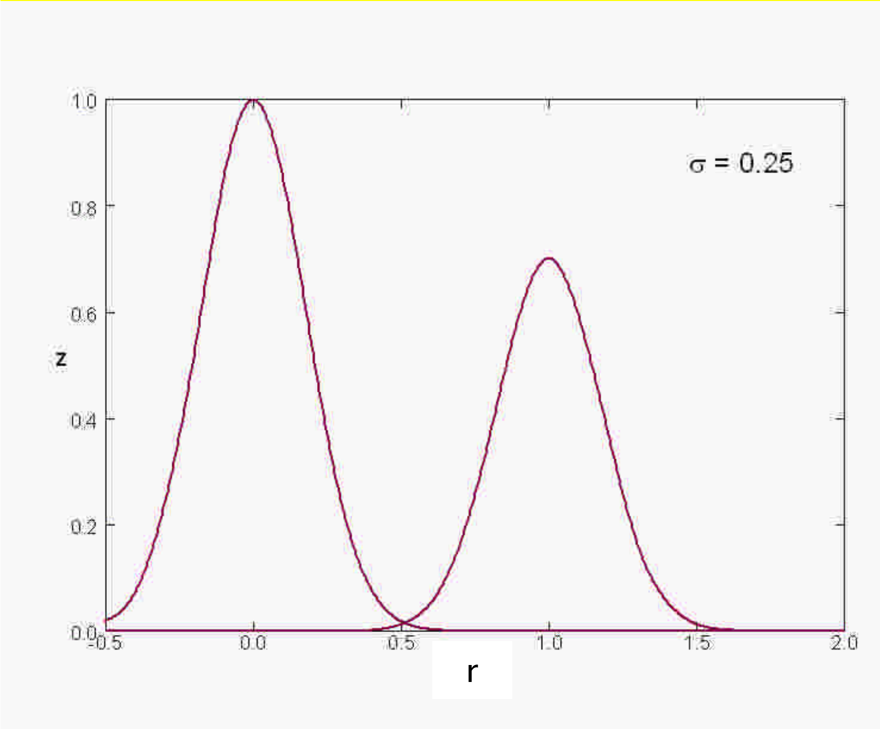
Par

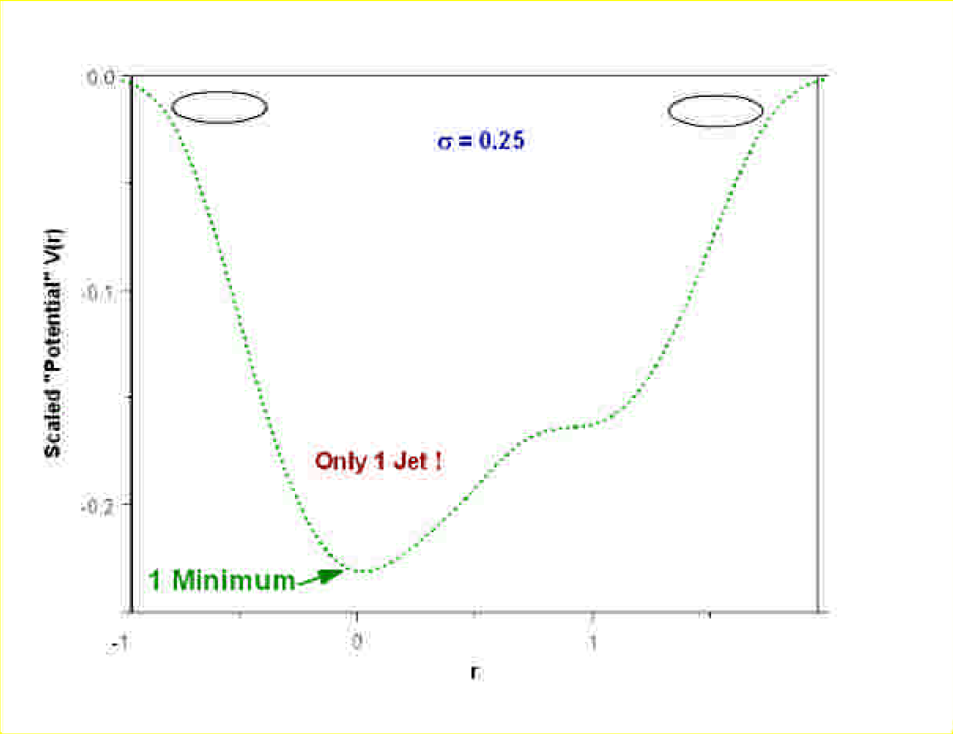
Result depends on choice of variable & MC

Note Differences

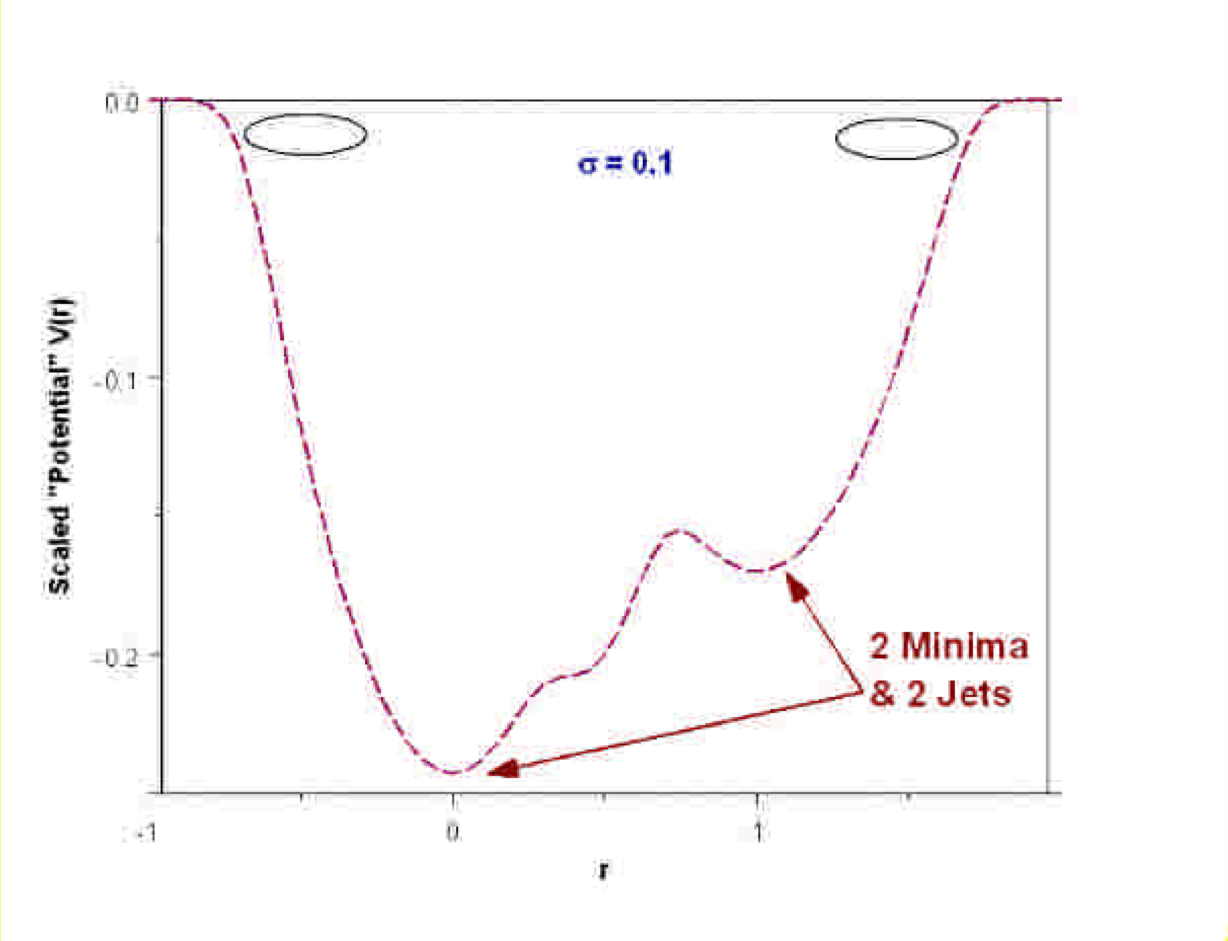
Results from M. Tönnesmann

Include smearing (~ showering & hadronization) in simple picture, find only 1 stable cone





Even if 2 stable cones, central cone can be lost to smearing

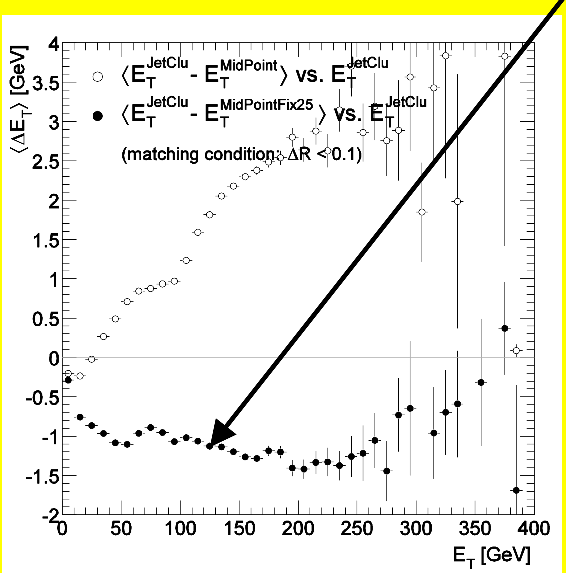


“Fix”

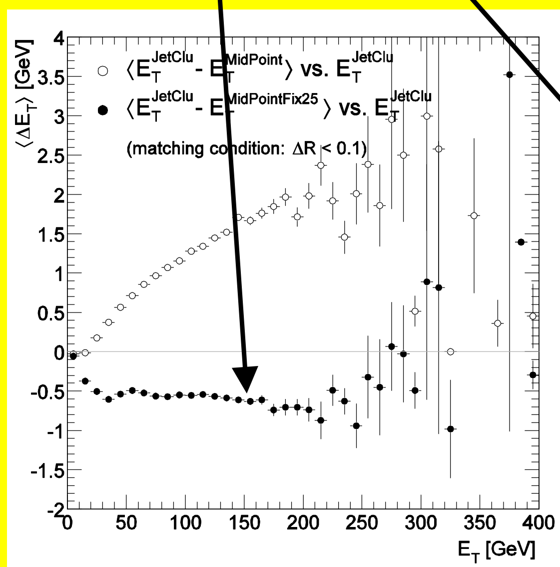
- Consider 2 distinct steps:
  - Find Stable cones
  - Construct Jets (split/merge, add 4-vectors)
- Use $R' < R$, *e.g.*, $R/2$, during stable cone discovery, less sensitivity to smearing, especially energy at periphery $\Rightarrow$ more stable cones
- Use R during jet construction

(Over)Found Energy!? $E_T^{MidPoint} > E_T^{JETCLU}$

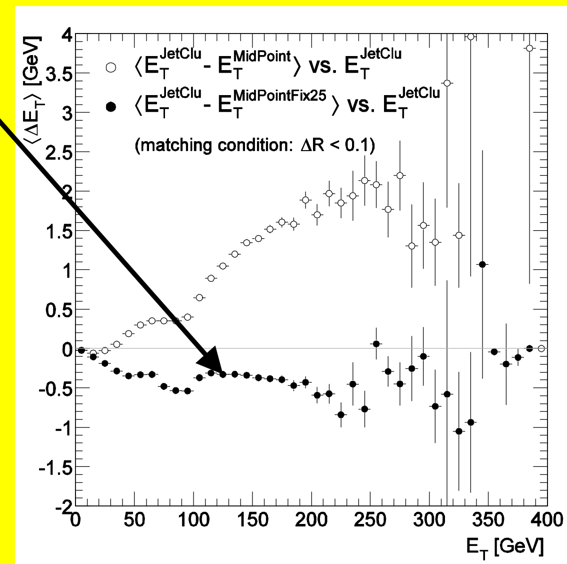
Partons



Calorimeter



Hadrons



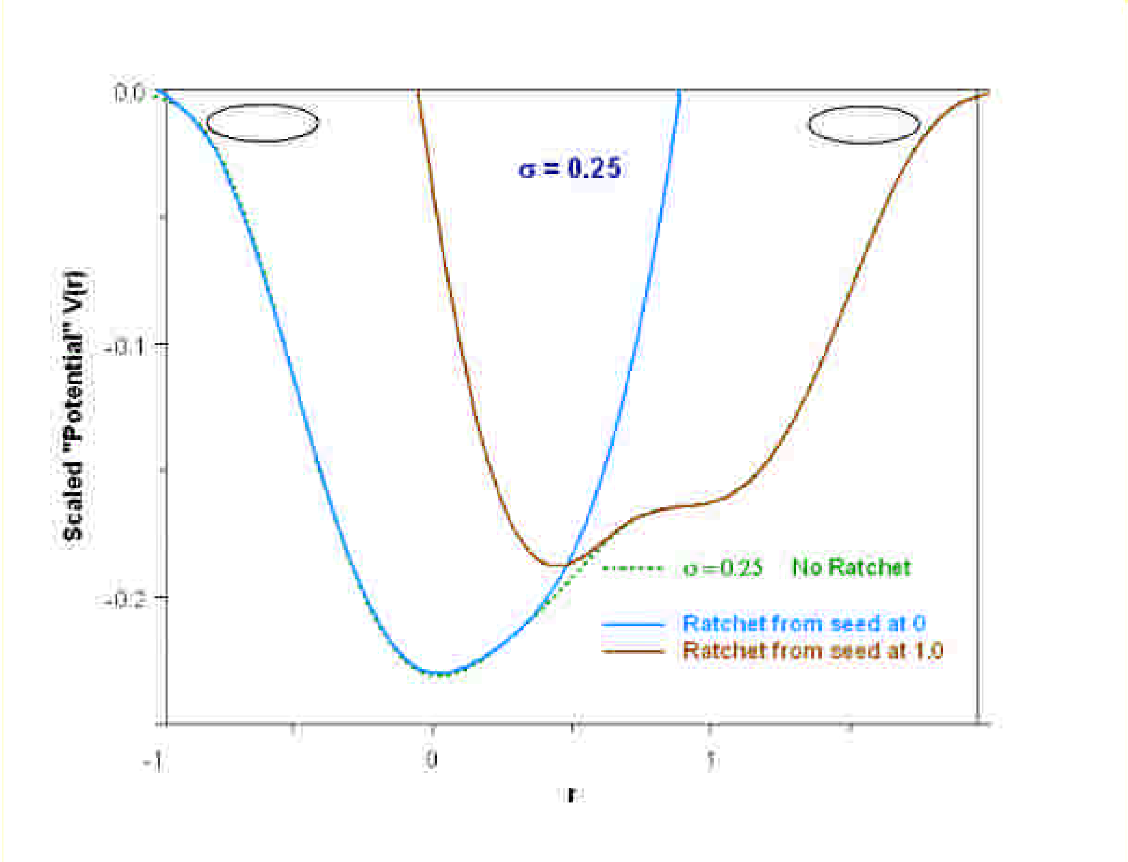
Note Differences

Results from M. Tönnesmann

Racheting – Why did it work?

Must consider seeds and subsequent migration history of trial cones – yields separate potential for each seed

INDEPENDENT of smearing, first potential finds stable cone near 0, while second finds stable cone in middle (even when right cone is washed out)! ~ NLO Perturbation Theory!!

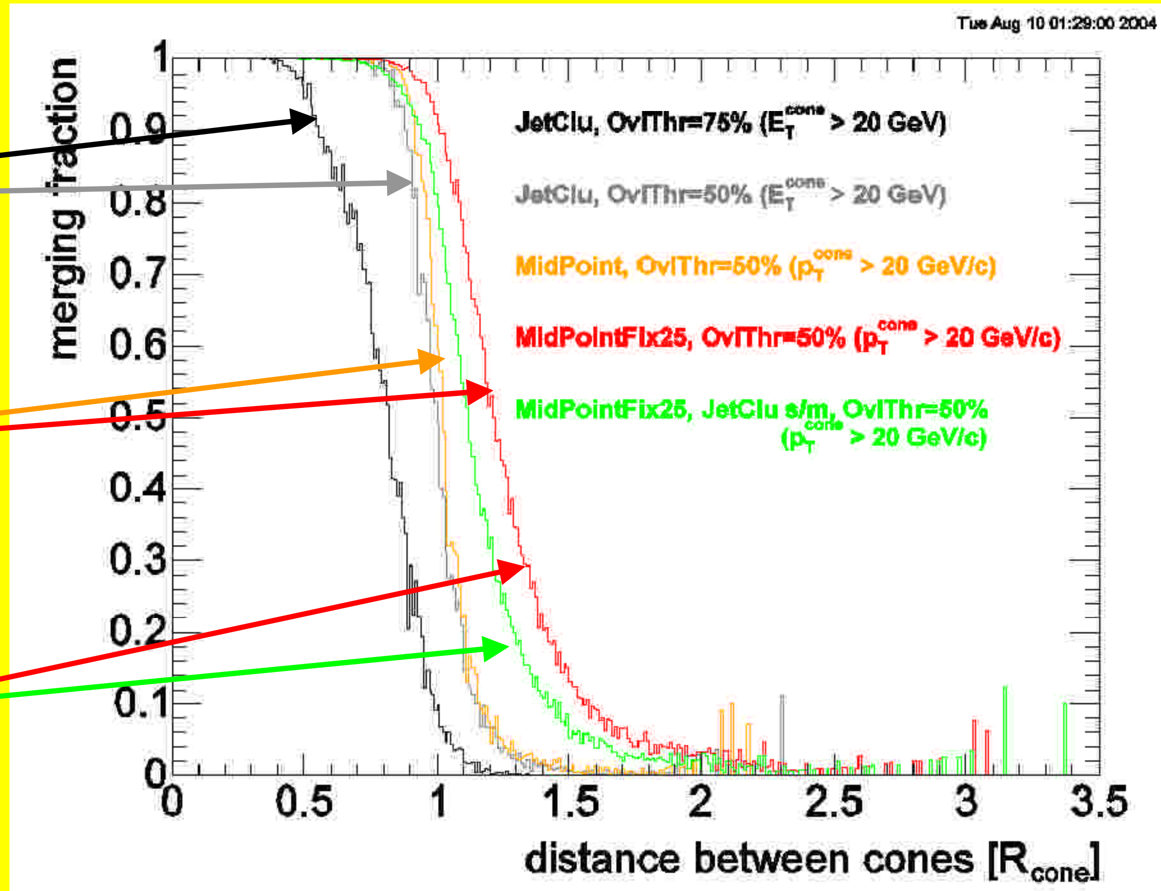


But underlying Structure is Different – Consider Cone Merging Probability

Result depends on f_{merge}

Result depends on "FIX"

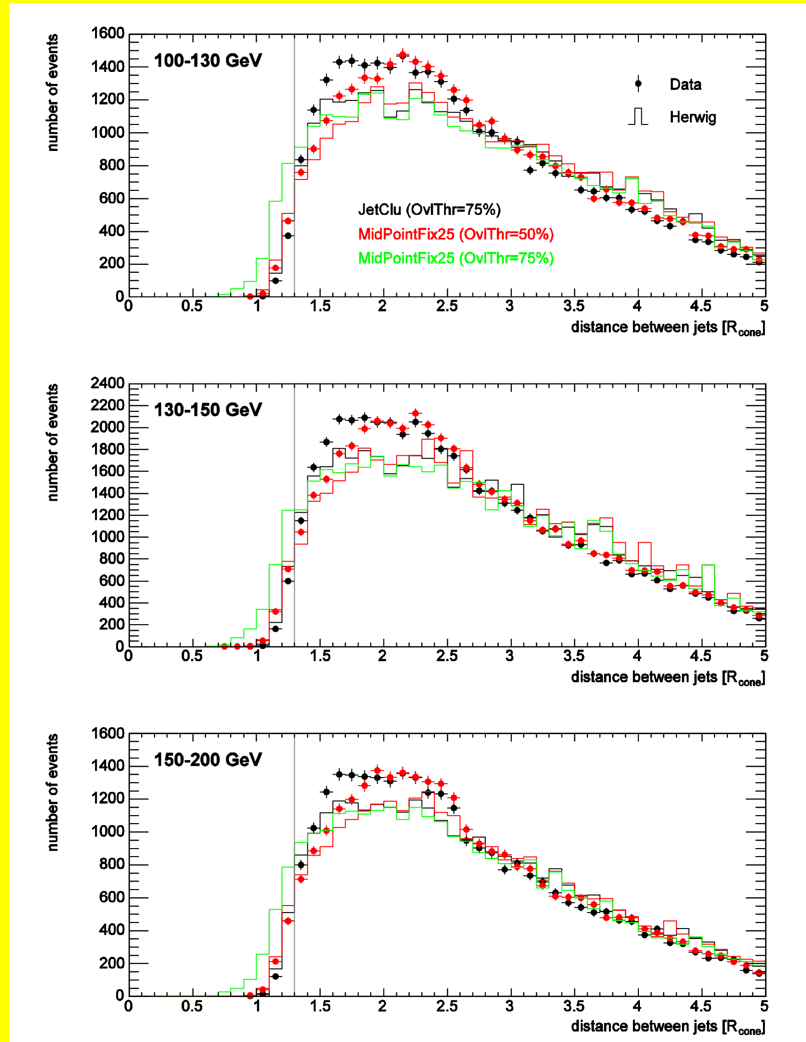
Result depends on ordering in S/M



But found jet separation looks more similar:

Conclude stable cone distributions must differ to match (cancel) the effects of merging

Jet dist ~ (Stable Cone) * (Merge Prob)



But Note – we are "fixing" to match JETCLU which is *NOT* the same as perturbation theory

HW for these Workshops

Can we reach the <u>original</u> goal of precisely mapping experiment onto short-distance theory? Using:

- MidPoint Cone algorithms (with FIX)?
- Seedless Cone Algorithm?
- k_T algorithm?
- Something New & Different, *e.g.*, Jet Energy Flows?

Can we agree to use the **<u>SAME</u>** Algorithm??