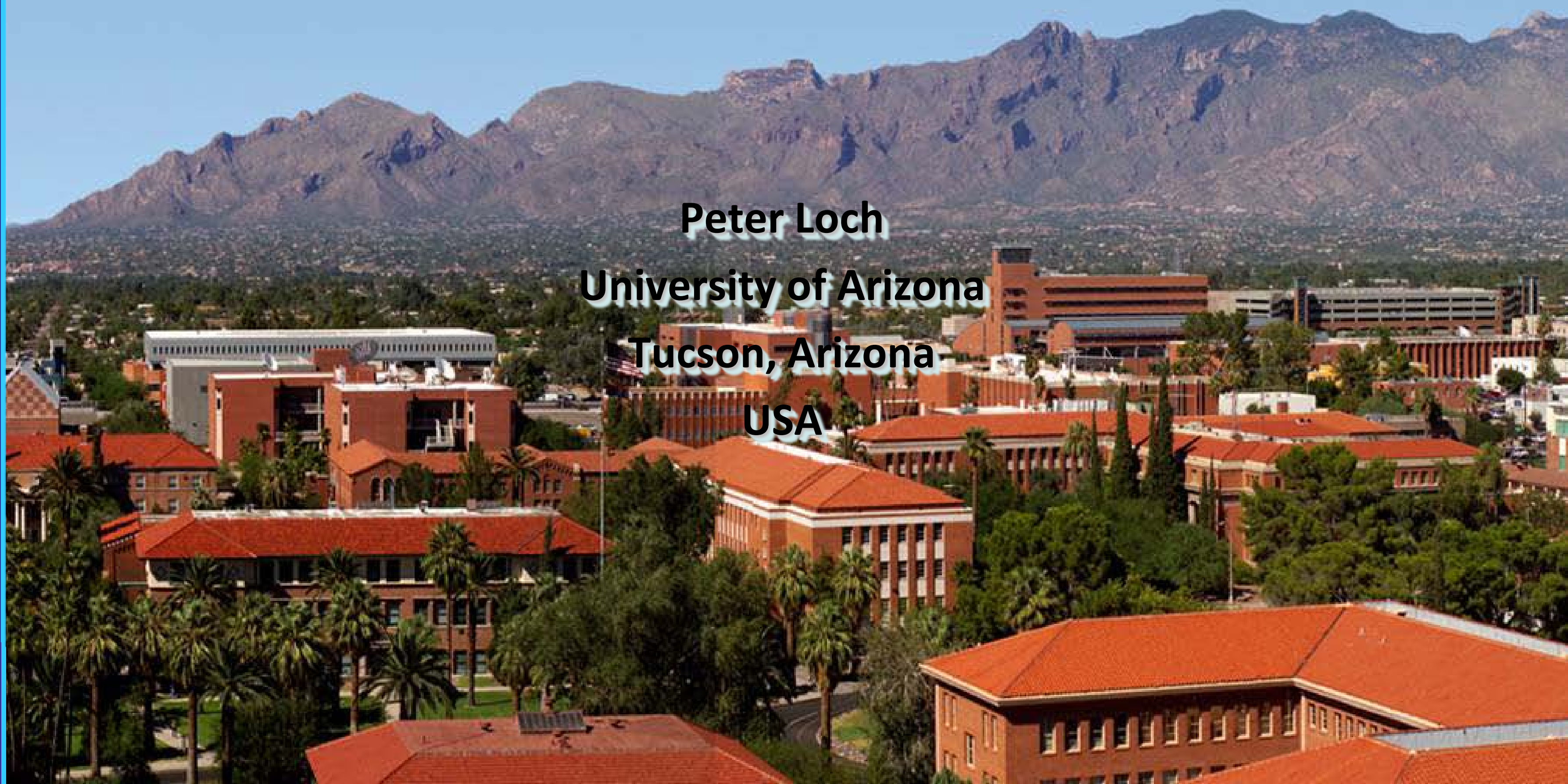
THE UNIVERSITY OF ARIZONA.

Introduction to Hadronic Final State Reconstruction in Collider Experiments (Part VII & VIII)

Peter Loch
University of Arizona
Tucson, Arizona
USA



Validity of Jet Algorithms

P. Loch
U of Arizona
March 19, 2010

Need to be valid to any order of perturbative calculations

- Experiment needs to keep sensitivity to perturbative infinities
  - Jet algorithms must be infrared safe!
  - Stable for multi-jet final states

Clearly a problem for classic (seeded) cone algorithms

- Tevatron: modifications to algorithms and optimization of algorithm configurations
  - Mid-point seeded cone: put seed between two particles
  - Split & merge fraction: adjust between 0.5 – 0.75 for best "resolution"

LHC: need more stable approaches

- Multi-jet context important for QCD measurements
  - Extractions of inclusive and exclusive cross-sections, PDFs
- Signal-to-background enhancements in searches
  - Event selection/filtering based on topology
  - Other kinematic parameters relevant for discovery

Among consequences of IR unsafety:

	Last meaningful order			
	JetClu, ATLAS cone [IC-SM]	MidPoint [IC_{mp}-SM]	CMS it. cone [IC-PR]	Known at
Inclusive jets	LO	NLO	NLO	NLO ($\rightarrow$ NNLO)
W/Z + 1 jet	LO	NLO	NLO	NLO
3 jets	**none**	LO	LO	NLO [nlojet++]
W/Z + 2 jets	**none**	LO	LO	NLO [MCFM]
m_{jet} in $2j + X$	**none**	**none**	**none**	LO

NB: \$30 – 50M investment in NLO

Validity of Jet Algorithms

P. Loch
U of Arizona
March 19, 2010

Need to be valid to any order of perturbative calculations

- Experiment needs to keep sensitivity to perturbative infinities
  - Jet algorithms must be infrared safe!
  - Stable for multi-jet final states

Clearly a problem for classic (seeded) cone algorithms

- Tevatron: modifications to algorithms and optimization of algorithm configurations
  - Mid-point seeded cone: put seed between two particles
  - Split & merge fraction: adjust between 0.5 – 0.75 for best "resolution"

LHC: need more stable approaches

- Multi-jet context important for QCD measurements
  - Extractions of inclusive and exclusive cross-sections, PDFs
- Signal-to-background enhancements in searches
  - Event selection/filtering based on topology
  - Other kinematic parameters relevant for discovery

Starts to miss cones at next order!

Among consequences of IR unsafety:

	Last meaningful order			
	JetClu, ATLAS cone [IC-SM]	MidPoint [IC_{mp}-SM]	CMS it. cone [IC-PR]	Known at
Inclusive jets	LO	NLO	NLO	NLO ($\rightarrow$ NNLO)
$W/Z + 1$ jet	LO	NLO	NLO	NLO
3 jets	**none**	LO	LO	NLO [nlojet++]
$W/Z + 2$ jets	**none**	LO	LO	NLO [MCFM]
m_{jet} in $2j + X$	**none**	**none**	**none**	LO

NB: \$30 – 50M investment in NLO

THE UNIVERSITY OF ARIZONA.

Midpoint Seeded Cone

P. Loch
U of Arizona
March 19, 2010

Attempt to increase infrared safety for seeded cone

- Midpoint algorithm starts with seeded cone
  - Seed threshold may be 0 to increase collinear safety
- Place new seeds between two close stable cones
  - Also center of three stable cones possible
- Re-iterate using midpoint seeds
  - Isolated stable cones are unchanged

Still not completely safe!

- Apply split & merge
  - Usually split/merge fraction 0.75

Find midpoints for stable cones within

$$\Delta R = \sqrt{\Delta y^2 + \Delta \varphi^2} \leq 2R_{\text{cone}}$$

Midpoint Seeded Cone

Attempt to increase infrared safety for seeded cone

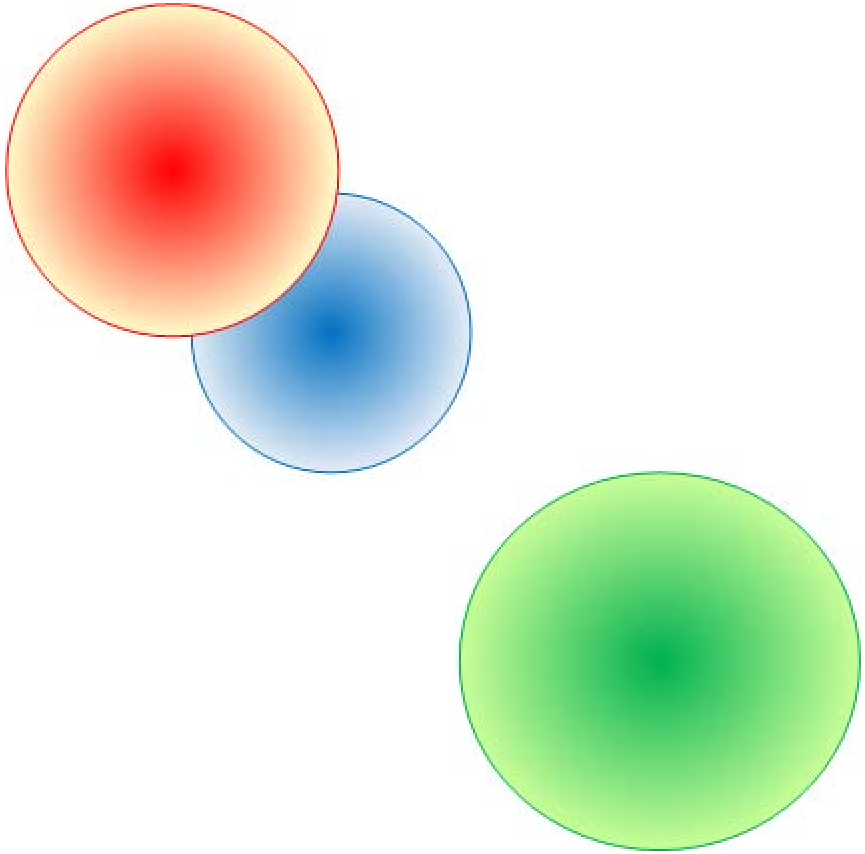
- Midpoint algorithm starts with seeded cone
  - Seed threshold may be 0 to increase collinear safety
- Place new seeds between two close stable cones
  - Also center of three stable cones possible
- Re-iterate using midpoint seeds
  - Isolated stable cones are unchanged

Still not completely safe!

- Apply split & merge
  - Usually split/merge fraction 0.75

Find midpoints for stable cones within

$$\Delta R = \sqrt{\Delta y^2 + \Delta\varphi^2} \leq 2R_{\text{cone}}$$



THE UNIVERSITY OF ARIZONA.

Midpoint Seeded Cone

P. Loch
U of Arizona
March 19, 2010

Attempt to increase infrared safety for seeded cone

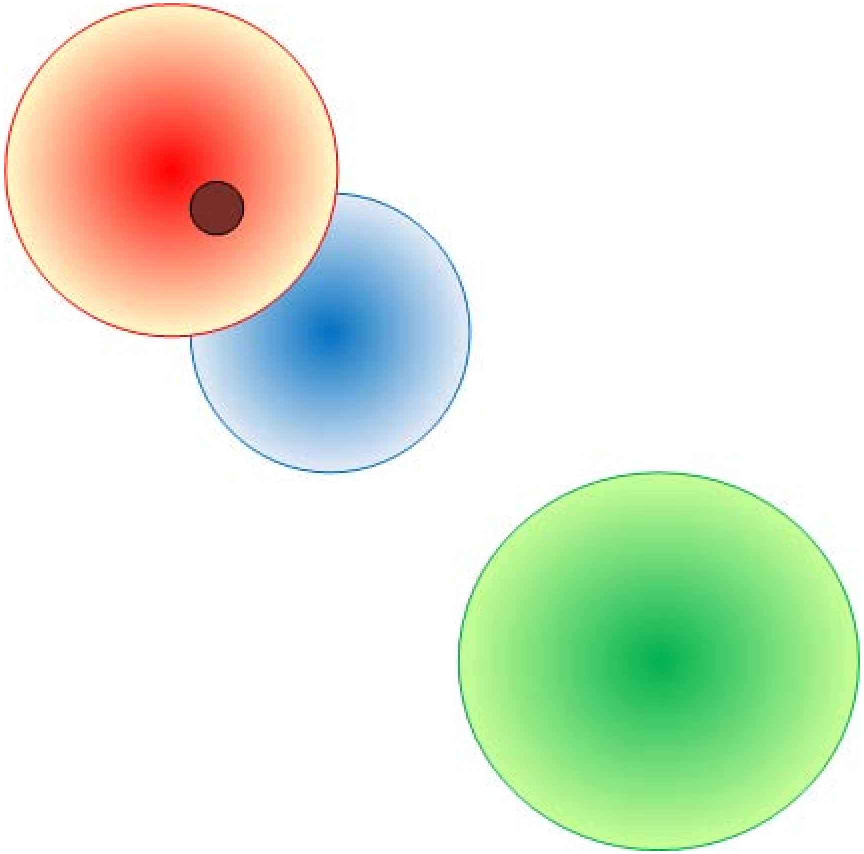
- Midpoint algorithm starts with seeded cone
  - Seed threshold may be 0 to increase collinear safety
- Place new seeds between two close stable cones
  - Also center of three stable cones possible
- Re-iterate using midpoint seeds
  - Isolated stable cones are unchanged

Still not completely safe!

- Apply split & merge
  - Usually split/merge fraction 0.75

Find midpoints for stable cones within

$$\Delta R = \sqrt{\Delta y^2 + \Delta \varphi^2} \leq 2R_{\text{cone}}$$



Midpoint Seeded Cone

Attempt to increase infrared safety for seeded cone

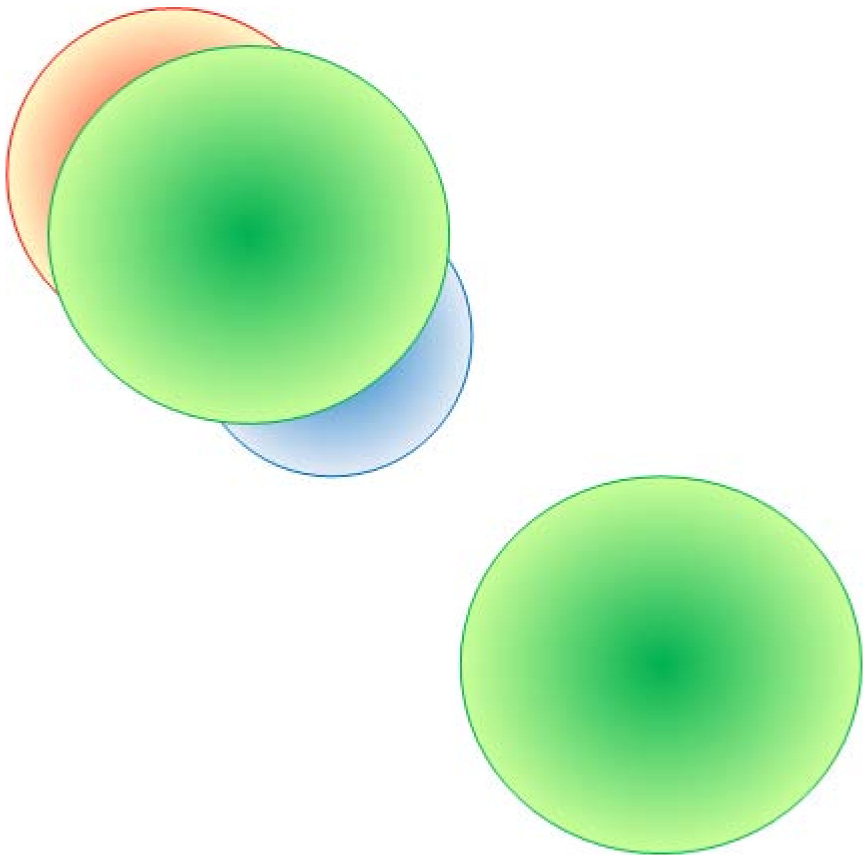
- Midpoint algorithm starts with seeded cone
  - Seed threshold may be 0 to increase collinear safety
- Place new seeds between two close stable cones
  - Also center of three stable cones possible
- Re-iterate using midpoint seeds
  - Isolated stable cones are unchanged

Still not completely safe!

- Apply split & merge
  - Usually split/merge fraction 0.75

Find midpoints for stable cones within

$$\Delta R = \sqrt{\Delta y^2 + \Delta\varphi^2} \leq 2R_{\text{cone}}$$



THE UNIVERSITY OF ARIZONA

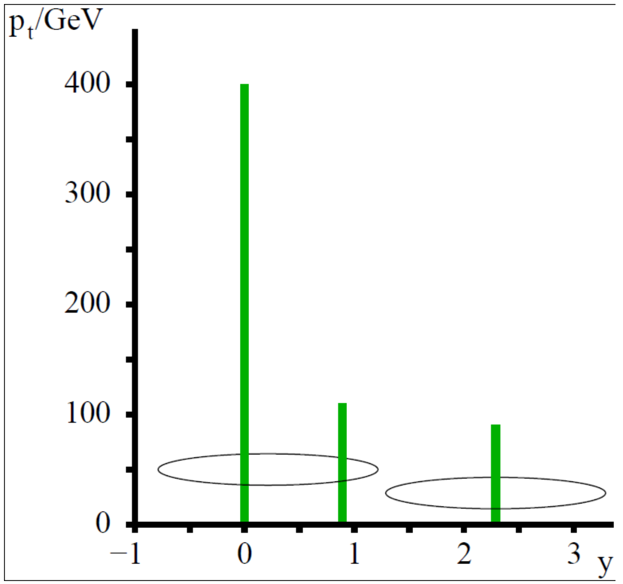
Midpoint Seeded Cone

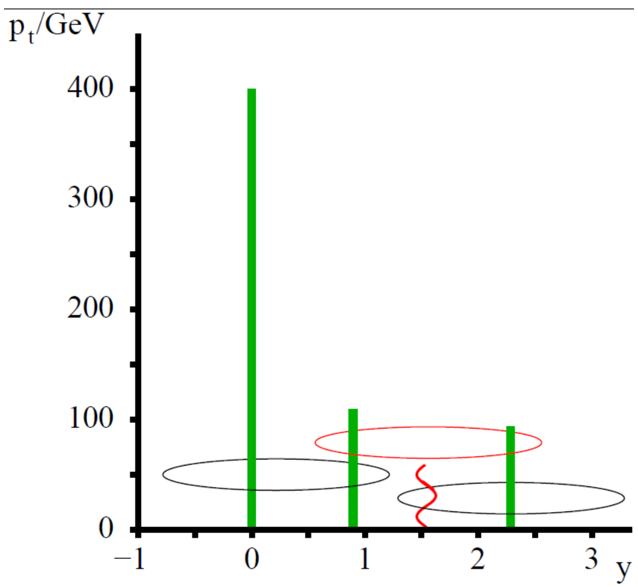
Attempt to increase infrared safety for seeded cone

- Midpoint algorithm starts with seeded cone
  - Seed threshold may be 0 to increase collinear safety
- Place new seeds between two close stable cones
  - Also center of three stable cones possible
- Re-iterate using midpoint seeds
  - Isolated stable cones are unchanged

Still not completely safe!

- Apply split & merge
  - Usually split/merge fraction 0.75





(from G. Salam & G. Soyez, **JHEP 0705:086,2007**)

ATLAS

Seedless Fixed Cone

P. Loch
U of Arizona
March 19, 2010

Improvements to cone algorithms: no seeds

- All stable cones are considered
  - Avoid collinear unsafety in seeded cone algorithm
- Avoid infrared safety issue
  - Adding infinitively soft particle does not lead to new (hard) cone
- Exact seedless cone finder
  - Problematic for larger number of particles
- Approximate implementation
  - Pre-clustering in coarse towers
  - Not necessarily appropriate for particles and even some calorimeter signals

Exact seedless cone for N particles:

$O(N \cdot 2^N)$ operations

N	# operations	remark
4	64	fixed order parton level
10	10240	very low multiplicity final state
100	$\sim 1.3 \cdot 10^{32}$	low multiplicity LHC final state
1,000	$\sim 1.6 \cdot 10^{153}$	typical LHC final state
10,000	∞	LHC high luminosity final state

Approximate seedless cone ($\Delta\eta \times \Delta\varphi = 0.2 \times 0.2$):

N	# operations	remark
40	$\sim 4.4 \cdot 10^{13}$	surviving bins with two narrow jets
70	$\sim 8.3 \cdot 10^{22}$	surving bins with two wide jets

Seedless Fixed Cone

Improvements to cone algorithms: no seeds

- All stable cones are considered
  - Avoid collinear unsafety in seeded cone algorithm
- Avoid infrared safety issue
  - Adding infinitively soft particle does not lead to new (hard) cone
- Exact seedless cone finder
  - Problematic for larger number of particles
- Approximate implementation
  - Pre-clustering in coarse towers
  - Not necessarily appropriate for particles and even some calorimeter signals

Exact seedless cone for N particles:

$O(N \cdot 2^N)$ operations

N	# operations	remark
4	64	fixed order parton level
10	10240	very low multiplicity final state
100	$\sim 1.3 \cdot 10^{32}$	low multiplicity LHC final state
1,000	$\sim 1.6 \cdot 10^{153}$	typical LHC final state
10,000	∞	LHC high luminosity final state

Approximate seedless cone ($\Delta\eta \times \Delta\varphi = 0.2 \times 0.2$):

N	# operations	remark
40	$\sim 4.4 \cdot 10^{13}$	surviving bins with two narrow jets
70	$\sim 8.3 \cdot 10^{22}$	surving bins with two wide jets

ATLAS

Seedless Fixed Cone

Improvements to cone algorithms: no seeds

- All stable cones are considered
  - Avoid collinear unsafety in seeded cone algorithm
- Avoid infrared safety issue
  - Adding infinitively soft particle does not lead to new (hard) cone
- Exact seedless cone finder
  - Problematic for larger number of particles
- Approximate implementation
  - Pre-clustering in coarse towers
  - Not necessarily appropriate for particles and even some calorimeter signals

Exact seedless cone for N particles:

$O(N \cdot 2^N)$ operations

N	# operations	remark
4	64	fixed order parton level
10	10240	very low multiplicity final state
100	$\sim 1.3 \cdot 10^{32}$	low multiplicity LHC final state
1,000	$\sim 1.6 \cdot 10^{153}$	typical LHC final state
10,000	∞	LHC high luminosity final state

Approximate seedles cone ($\Delta\eta \times \Delta\varphi = 0.2 \times 0.2$):

N	# operations	remark
40	$\sim 4.4 \cdot 10^{13}$	surviving bins with two narrow jets
70	[illegible]	surviving bins [illegible] wide jets

Note: 100 particles need ~10^{17} years to be clustered!

Seedless Fixed Cone

P. Loch
U of Arizona
March 19, 2010

Improvements to cone algorithms: no seeds

- All stable cones are considered
  - Avoid collinear unsafety in seeded cone algorithm
- Avoid infrared safety issue
  - Adding infinitively soft particle does not lead to new (hard) cone
- Exact seedless cone finder
  - Problematic for larger number of particles
- Approximate implementation
  - Pre-clustering in coarse towers
  - Not necessarily appropriate for particles and even some calorimeter signals

Exact seedless cone for *N* particles:

$O(N \cdot 2^N)$ operations

N	# operations	remark
4	64	fixed order parton level
10	10240	very low multiplicity final state
100	$\sim 1.3 \cdot 10^{32}$	low multiplicity LHC final state
1,000	$\sim 1.6 \cdot 10^{153}$	typical LHC final state
10,000	∞	LHC high luminosity final state

Approximate seedless cone ($\Delta\eta \times \Delta\varphi = 0.2 \times 0.2$):

N	# operations	remark
40	$\sim 4.4 \cdot 10^{13}$	surviving bins with two narrow jets
70	$\sim 8.3 \cdot 10^{22}$	surving bins with two wide jets

Seedless Infrared Safe Cone

SISCone (Salam, Soyez 2007)

- Exact seedless cone with geometrical (distance) ordering
  - Speeds up algorithm considerably!
- Find all distinctive ways on how a segment can enclose a subset of the particles
  - Instead of finding all stable segments!
- Re-calculate the centroid of each segment
  - E.g., pT weighted re-calculation of direction
  - "E-scheme" works as well
- Segments (cones) are stable if particle content does not change
  - Retain only one solution for each segment
- Still needs split & merge to remove overlap
  - Recommended split/merge fraction is 0.75
- Typical times
  - $N^2 \ln N$ for particles in 2-dim plane

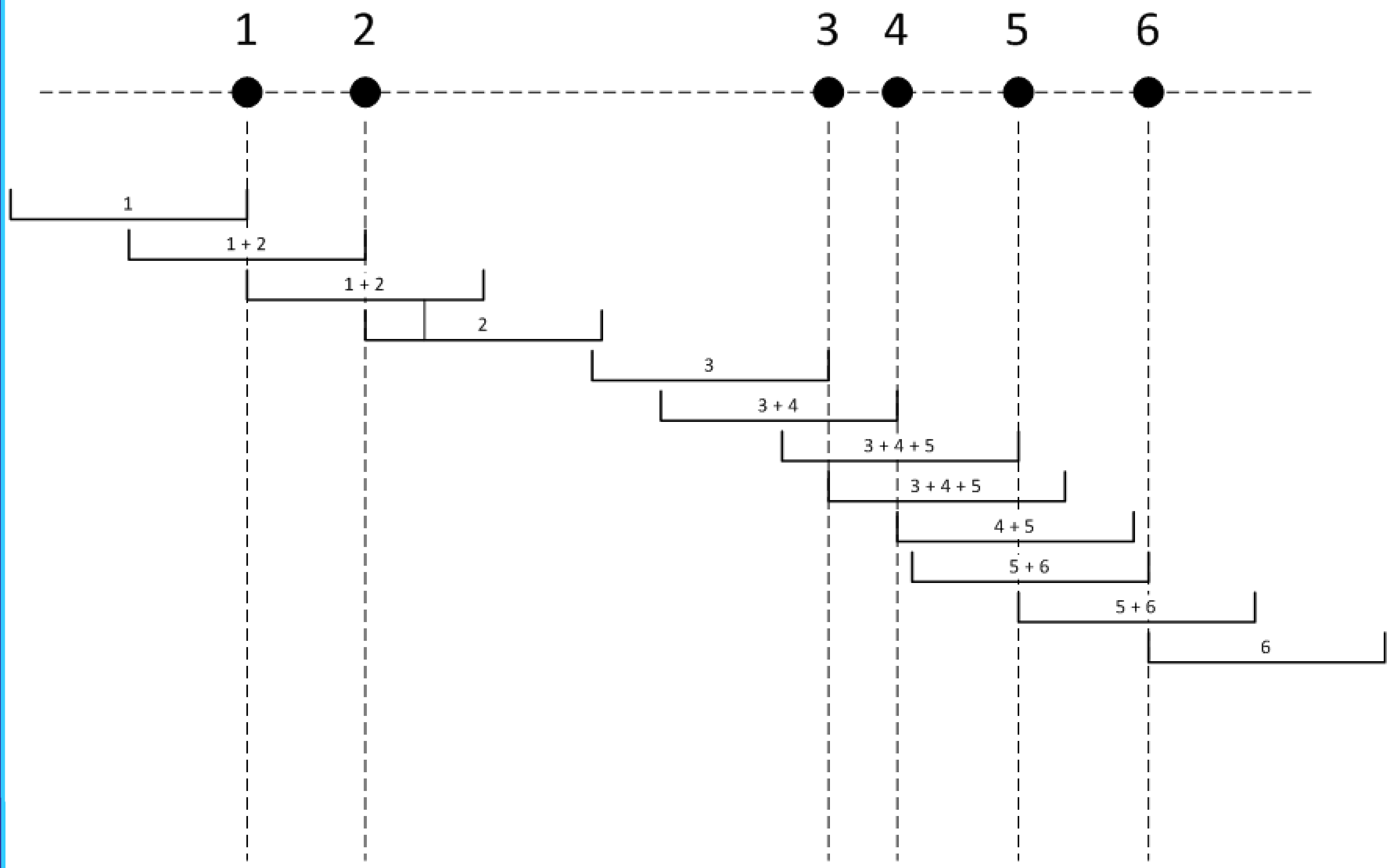
1-dim example:

- See following slides!

(inspired by G. Salam & G. Soyez, **JHEP 0705:086,2007**)

SISCone Principle (1-dim!)

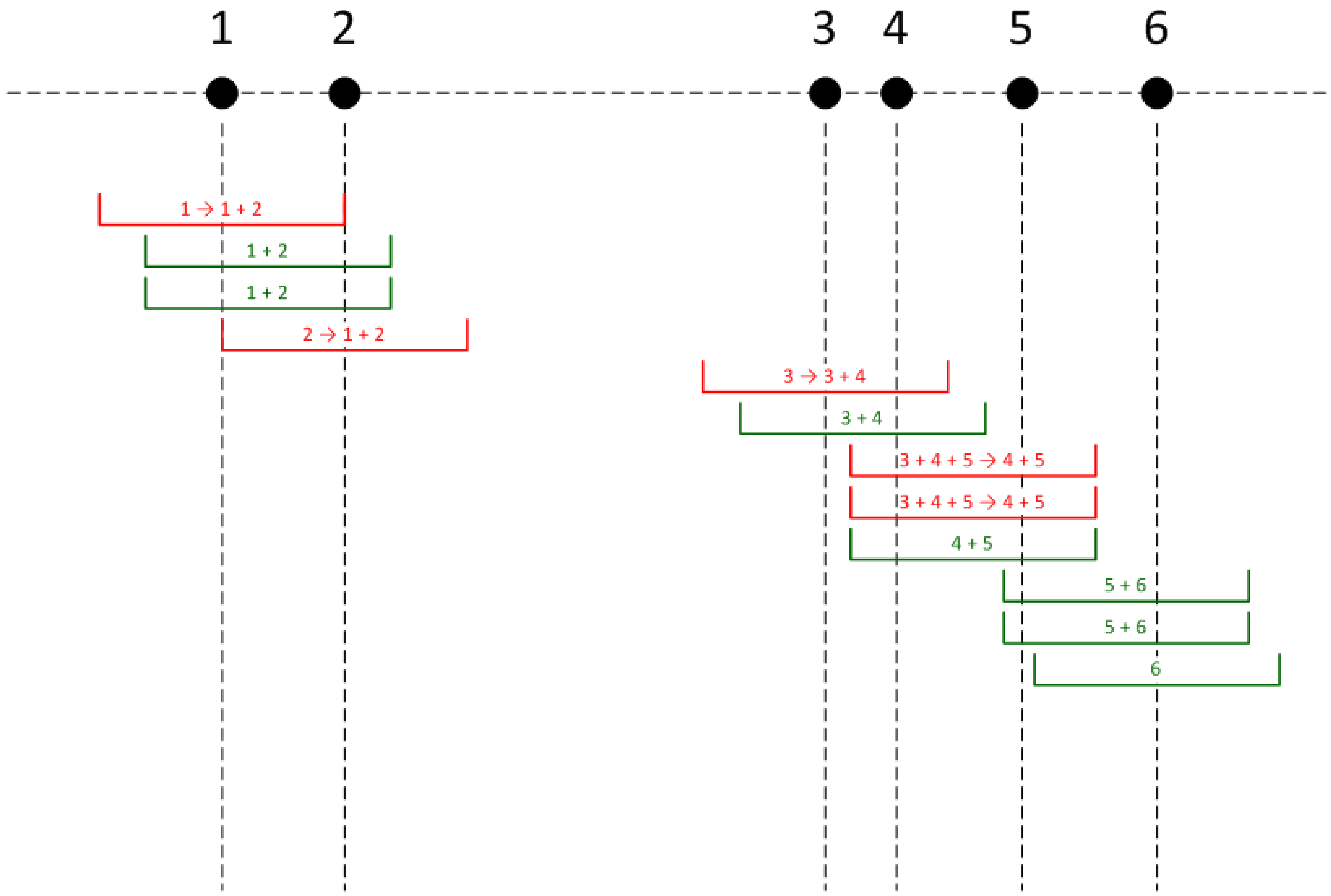
Find all distinctive segments of size $2R_{cone}$ ($O(N)$ operations in 1-dim)



THE UNIVERSITY OF ARIZONA.

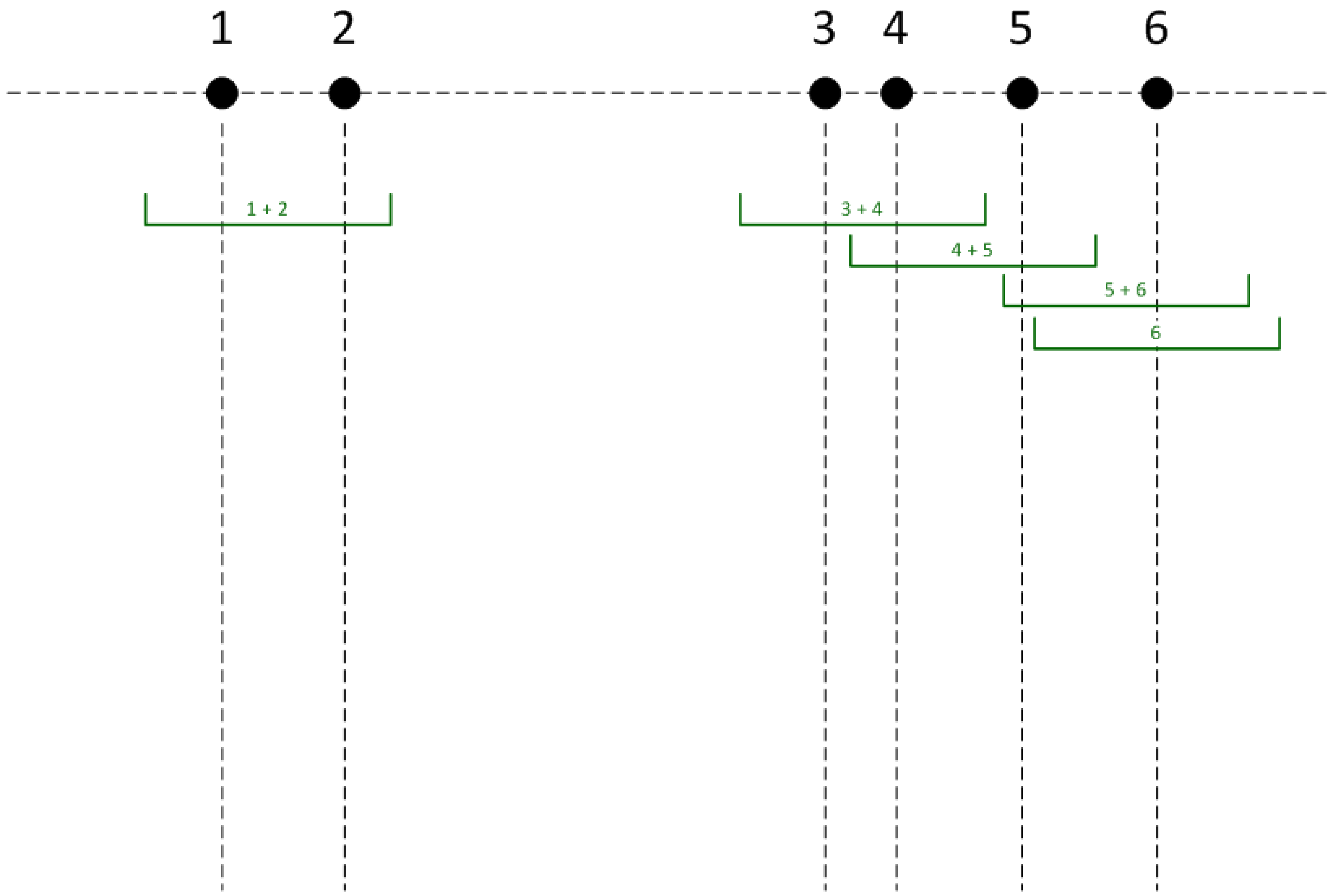
SISCone Principle (1-dim!)

Reposition segments to centroids (green - unchanged, red - changed)



SISCone Principle (1-dim!)

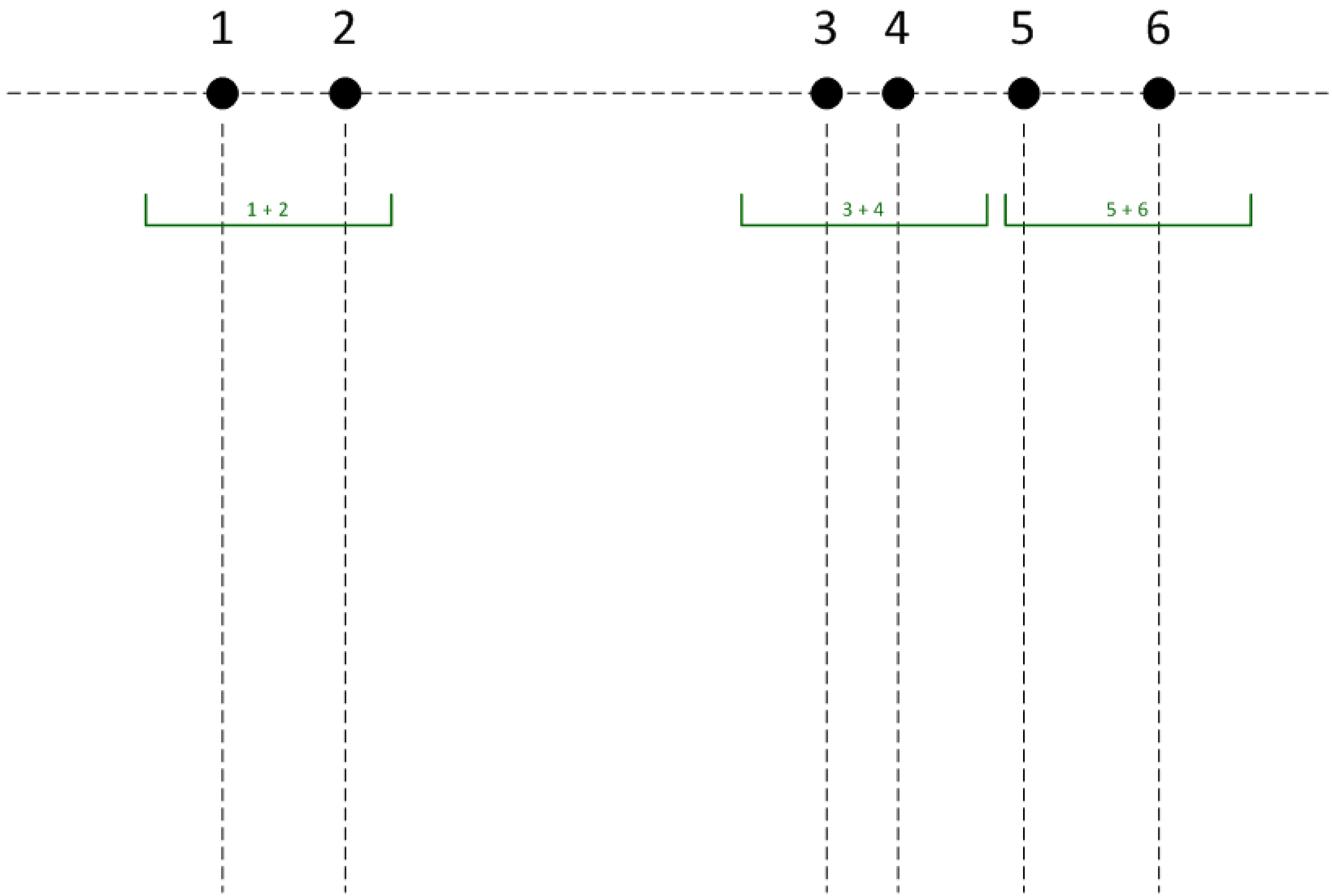
Retain only one stable solution for each segment



P. Loch
U of Arizona
March 19, 2010

SISCone Principle (1-dim!)

Apply split & merge



Similar ordering and combinations in 2-dim

Use circles instead of linear segments

(from G. Salam & G. Soyez, **JHEP 0705:086,2007**)

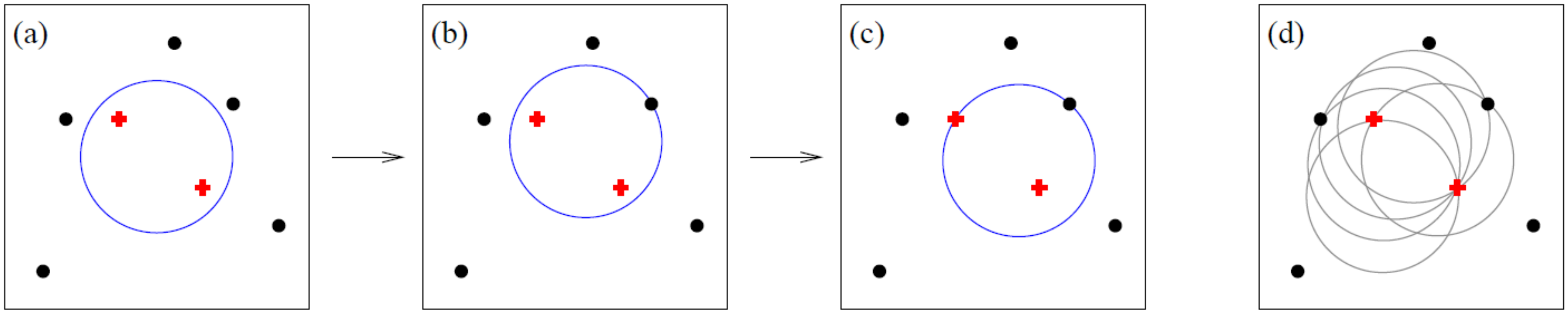


Figure 3: (a) Some initial circular enclosure; (b) moving the circle in a random direction until some enclosed or external point touches the edge of the circle; (c) pivoting the circle around the edge point until a second point touches the edge; (d) all circles defined by pairs of edge points leading to the same circular enclosure.

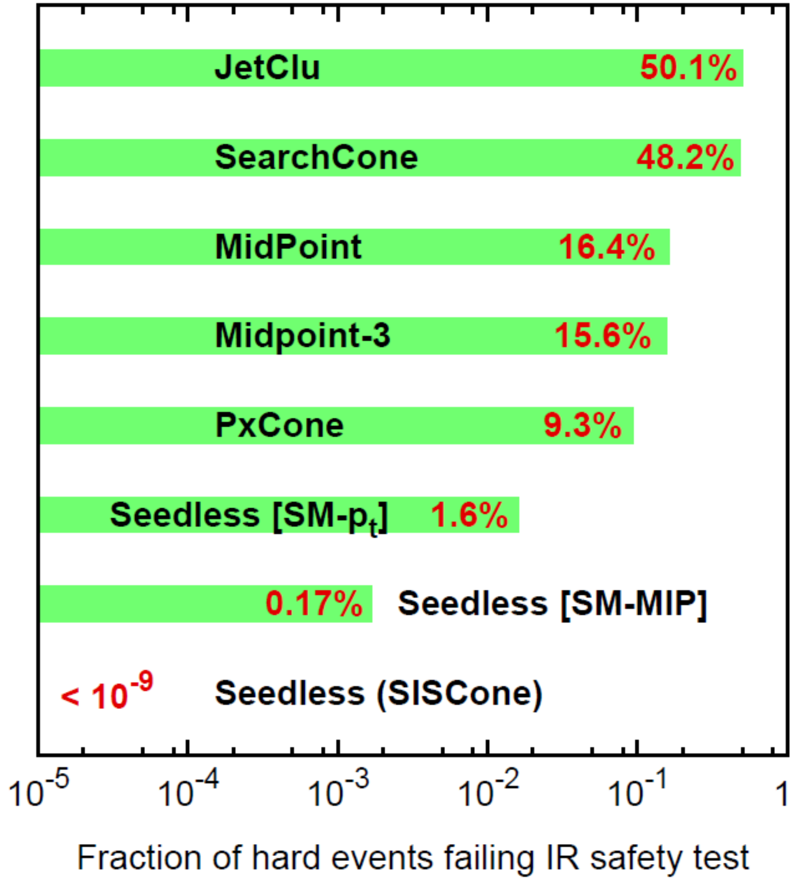
Still need split & merge

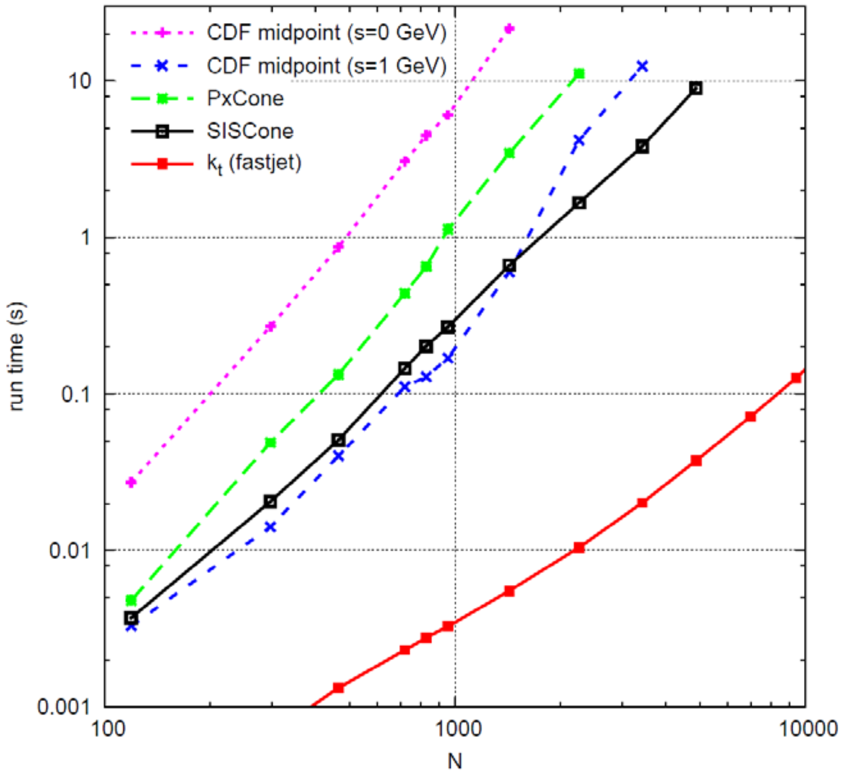
- One additional parameter outside of jet/cone size
  - Not very satisfactory!
- But at least a practical seedless cone algorithm
  - Very comparable performance to e.g. Midpoint!

Infrared safety failure rates

Computing performance

(from G. Salam & G. Soyez, **JHEP 0705:086,2007**)





Recursive Recombination (kT)

Computing performance an issue

- Time for traditional kT is ~N^3
  - Very slow for LHC
- FastJet implementations
  - Use geometrical ordering to find out which pairs of particles have to be manipulated instead of recalculating them all!
- Very acceptable performance in this case!

LHC events (*pp* collisions):

N	# operations	time [s]*
10	10^3	0.05
100	10^6	0.50
1,000	10^9	5.00

LHC events (heavy ion collisions):

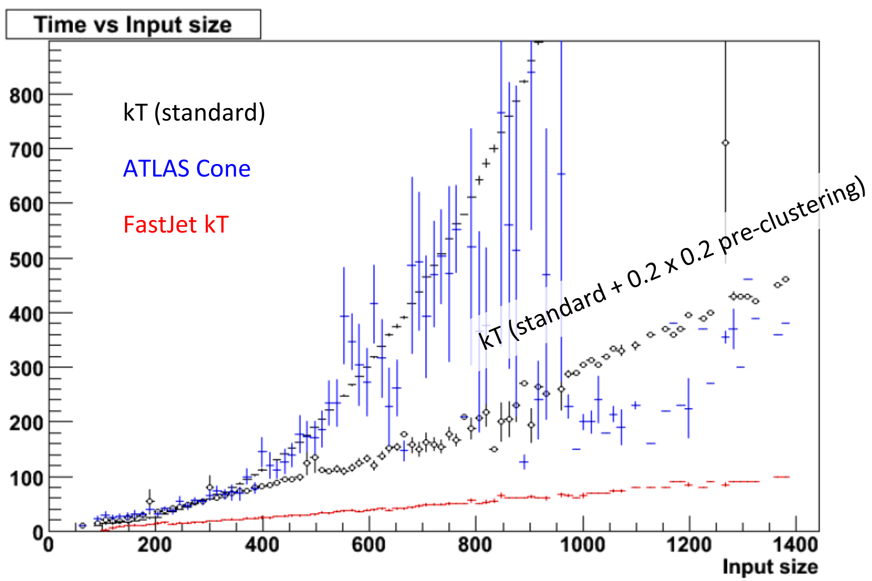
N	# operations	time [s]*
10,000	10^{12}	$5\cdot10^3$
50,000	$1.25\cdot10^{14}$	$6.25\cdot10^5$

*on a modern computer (3 GHz clock)

Recursive Recombination (kT)

Computing performance an issue

- Time for traditional kT is ~N^3
  - Very slow for LHC
- FastJet implementations
  - Use geometrical ordering to find out which pairs of particles have to be manipulated instead of recalculating them all!



FastJet implementations:

kT & Cambridge/Aachen $\sim N \ln N$

N	# operations	time [s]*
10	24	$0.1 \cdot 10^{-6}$
100	460	$2 \cdot 10^{-6}$
1,000	6,900	$35 \cdot 10^{-6}$
10,000	92,000	$0.5 \cdot 10^{-3}$
50,000	541,000	$3 \cdot 10^{-3}$

Anti-kT $\sim \sqrt{N^3}$

N	# operations	time [s]*
10	32	$0.2 \cdot 10^{-6}$
100	1,000	$5 \cdot 10^{-6}$
1,000	32,000	$0.2 \cdot 10^{-3}$
10,000	1,000,000	$5 \cdot 10^{-3}$
50,000	11,200,000	$56 \cdot 10^{-3}$

ATLAS

Address the search approach

- Need to find minimum in standard kT
  - Order N^3 operations
- Consider geometrically nearest neighbours in FastJet kT
  - Replace full search by search over (jet, jet neighbours)
- Need to find nearest neighbours for each proto-jet fast
  - Several different approaches: ATLAS (Delsart 2006) uses simple geometrical model, Salam & Cacciari (2006) suggest Voronoi cells
- Both based on same fact relating d_{ij} and geometrical distance in ΔR
  - Both use geometrically ordered lists of proto-jets

Find minimum for N particles in standard kT:

$$\left\{ d_{ij} = \min(d_i, d_j) \Delta R_{ij} / R, d_i = p_{T,i}^2 \right\},\ i,j = 1,\ldots,N$$

$O(N^2)$ searches, repeated N times $\rightarrow O(N^3)$

FastJet kT uses nearest neighbours search:

$$d_{ij} = \min \ \wedge \ p_{T,i} < p_{T,j}$$

$$\Rightarrow R_{ij} < R_{ik}\ \forall\ k \neq j, \text{ i.e. } (i,j) \text{ geometrical nearest neighbours in } (y,\varphi) \text{ plane}$$

Proof:

Assume an additional particle k exists with geometrical distance R_{ik} to particle i:

$$d_{ik} = \min(d_i, d_k) R_{ik} / R \leq d_i R_{ik} / R$$

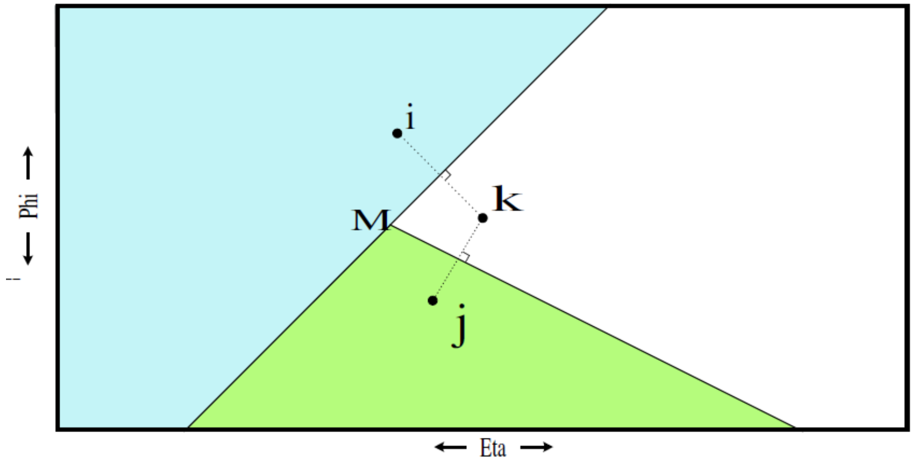
$$> \min = d_{ij} = d_i R_{ij} / R$$

works only for $R_{ik} > R_{ij}$

Fast kT (ATLAS – Delsart)

Possible implementation (P.A. Delsart, 2006)

- Nearest neighbour search
  - Idea is to only limit recalculation of distances to nearest neighbours
- Try to find all proto-jets having proto-jet *k* as nearest neighbour
  - Center pseudo-rapdity (or rapdity)/azimuth plane on *k*
  - Take first proto-jet *j* closest to *k* in pseudo-rapidity
  - Compute middle line L_{jk} between *k* and *j*
  - All proto-jets below L_{jk} are closer to *j* than *k* → *k* is not nearest neighbour of those
- Take next closest proto-jet *i* in pseudo-rapidity
  - Proceed as above with exclusion of all proto-jets above L_{ik}
  - Search stops when point below intersection of L_{jk} and L_{ik} is reached, no more points have *k* as nearest neighbour



Complexity estimate:

Assume *N* proto-jets are uniformly distributed in (η,φ) plane (rectangular with fintie size, area *A*)

Average number of proto-jets in circle with radius *R*:

$$\bar{N} = N\frac{\pi R^2}{A}$$

If *R* is mean distance between two proto-jets:

$$\bar{N} \approx 1 \Rightarrow R \approx \sqrt{\frac{A}{\pi N}}$$

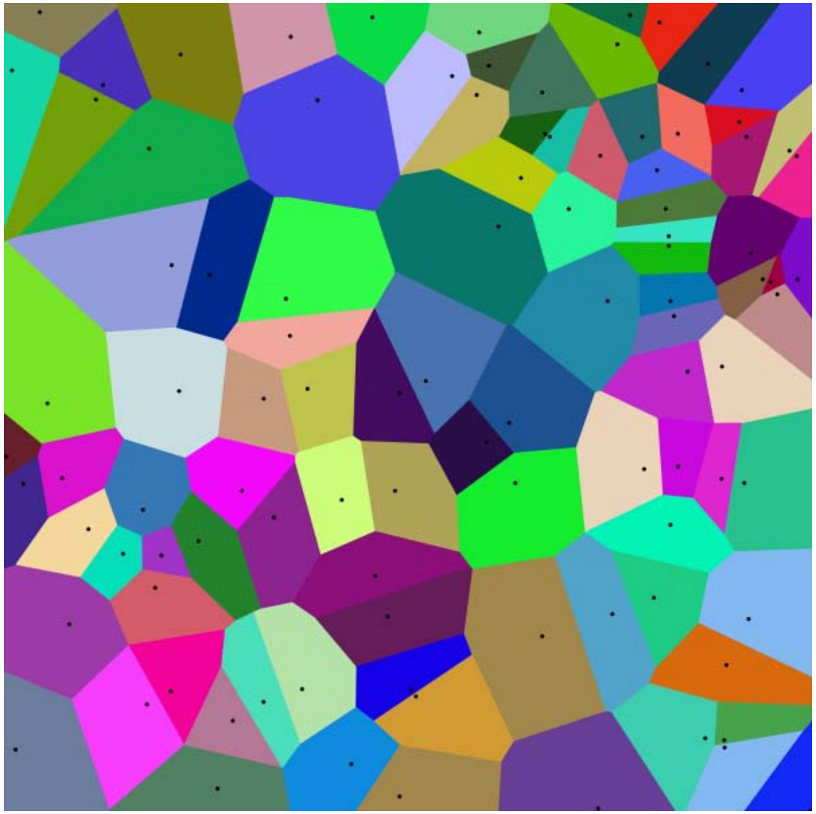
Computation of proto-jet *k*'s nearest neighbours is restricted to

$$\eta \approx [\eta_k - R, \eta_k + R] \mapsto \approx N \cdot 2R \propto \frac{N}{\sqrt{N}} = \sqrt{N} \text{ operations for } k$$

$\Rightarrow N\sqrt{N}$ total complexity (estimate)

Apply geometrical methods to nearest neighbour searches

- Voronoi cell around proto-jet *k* defines area of nearest neighbours
  - No point inside area is closer to any other protojet
  - Apply to protojets in pseudo-rapdity/azimuth plane
- Useful tool to limit nearest neighbour search
  - Determines region of re-calculation of distances in kT
  - Allows quick updates without manipulating too many long lists
- Complex algorithm!
  - Read G. Salam & M. Cacciari, Phys.Lett.B641:57-61 (2006)



(source **http://en.wikipedia.org/wiki/Voronoi_diagram**)

Complexity estimate (Monte Carlo experiment):

$N \ln N$ total complexity

Jet Algorithm Performance

P. Loch
U of Arizona
March 19, 2010

Various jet algorithms produce different jets from the same collision event

- Clearly driven by the different sensitivities of the individual algorithms
  - Cannot expect completely identical picture of event from jets
  - Different topology/number of jets
  - Differences in kinematics and shape for jets found at the same direction
- Choice of algorithm motivated by physics analysis goal
  - E.g., IR safe algorithms for jet counting in $W + n$ jets and others
  - Narrow jets for W mass spectroscopy
  - Small area jets to suppress pile-up contribution

Measure of jet algorithm performance depends on final state

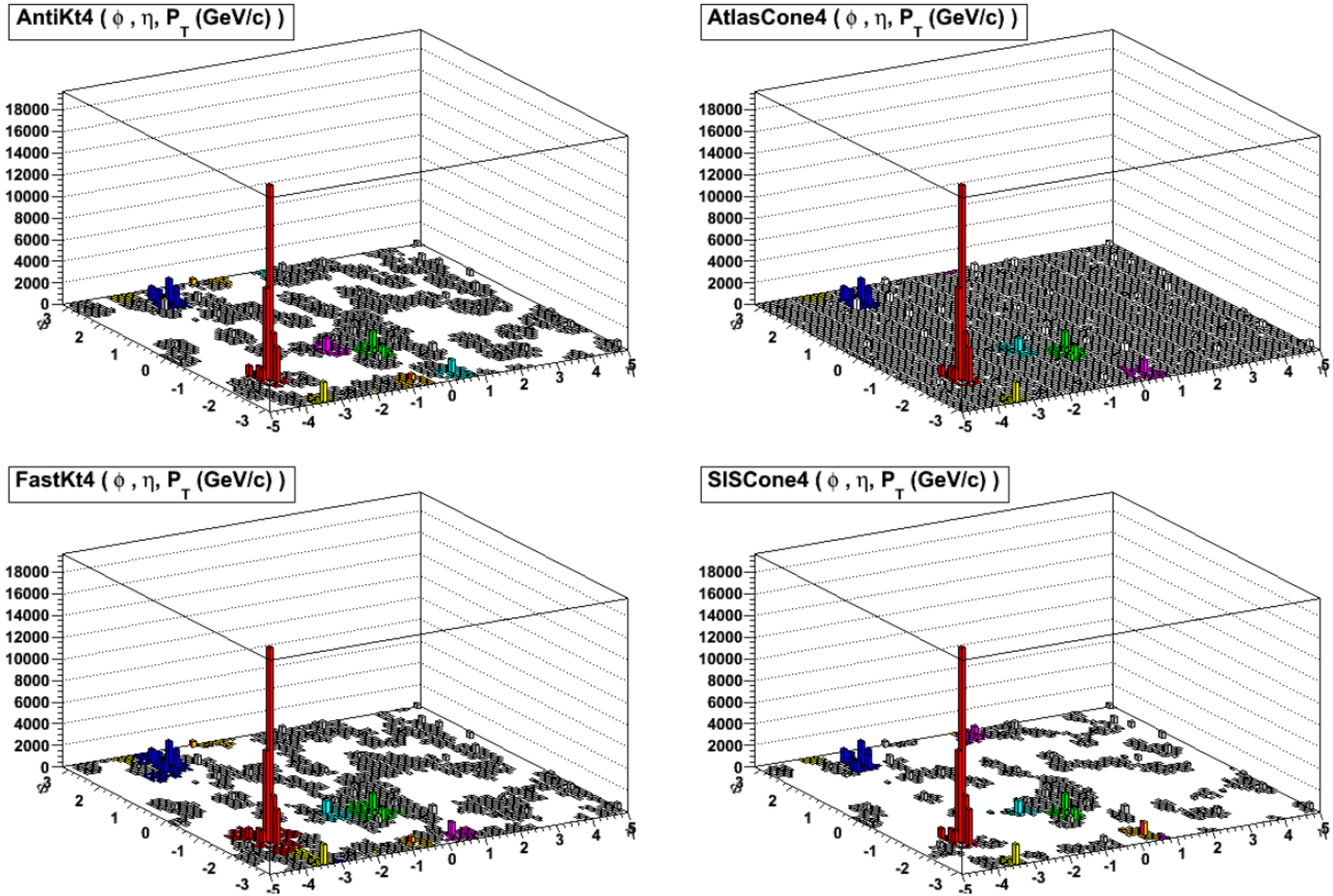
- Cone preferred for resonances
  - E.g., 2 – 3…n prong heavy particle decays like top, Z', etc.
  - Boosted resonances may require jet substructure analysis – need kT algorithm!
- Recursive recombination algorithms preferred for QCD cross-sections
  - High level of IR safety makes jet counting more stable
- Pile-up suppression easiest for regularly shaped jets
  - E.g., Anti-kT most cone-like, can calculate jet area analytically even after split and merge

Measures of jet performance

- Particle level measures prefer observables from final state
  - Di-jet mass spectra etc.
- Quality of spectrum important
  - Deviation from Gaussian etc.

Jet Shapes (1)

P. Loch
U of Arizona
March 19, 2010

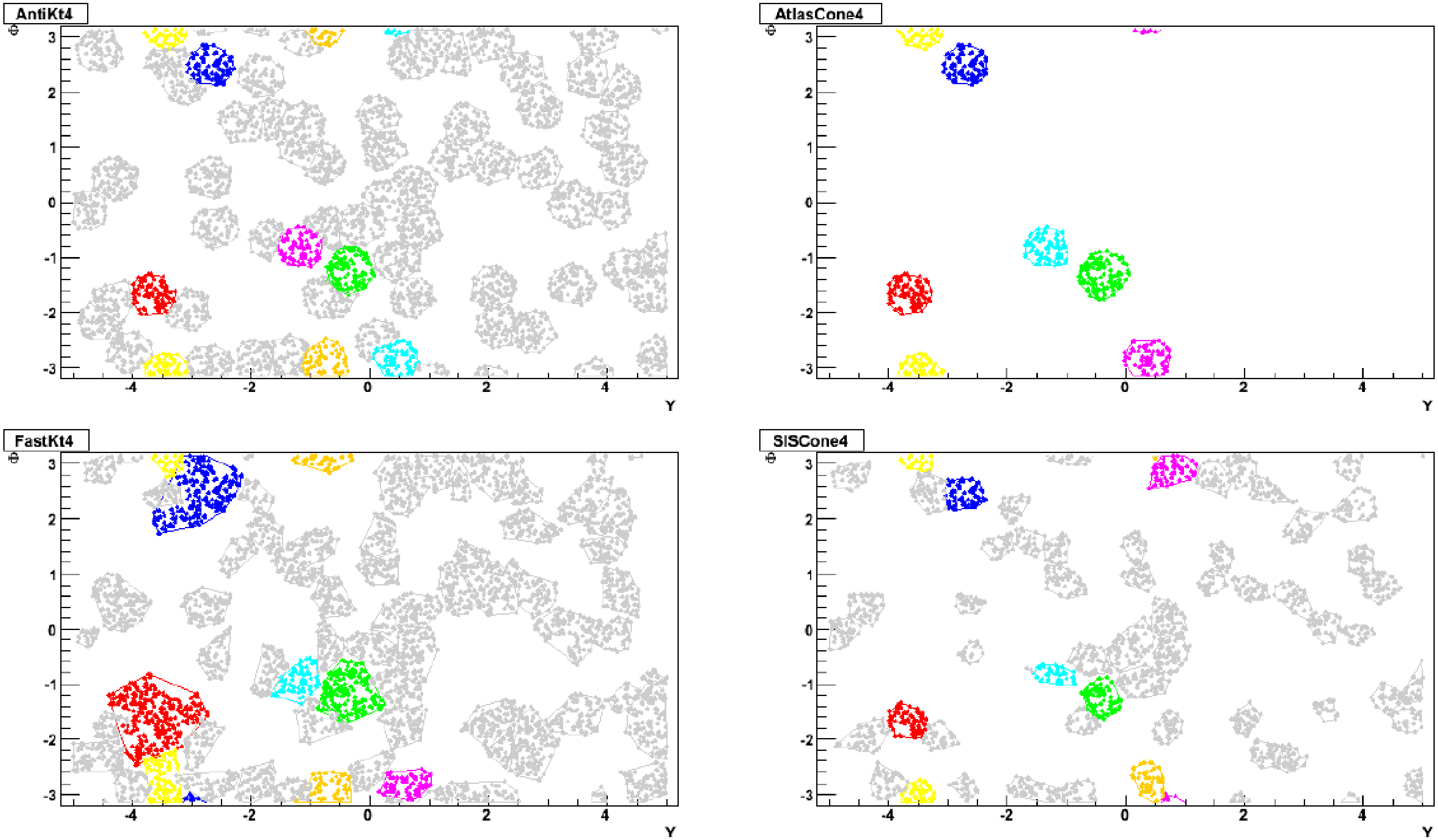


(from P.A. Delsart)

THE UNIVERSITY OF ARIZONA

Jet Shapes (2)

P. Loch
U of Arizona
March 19, 2010

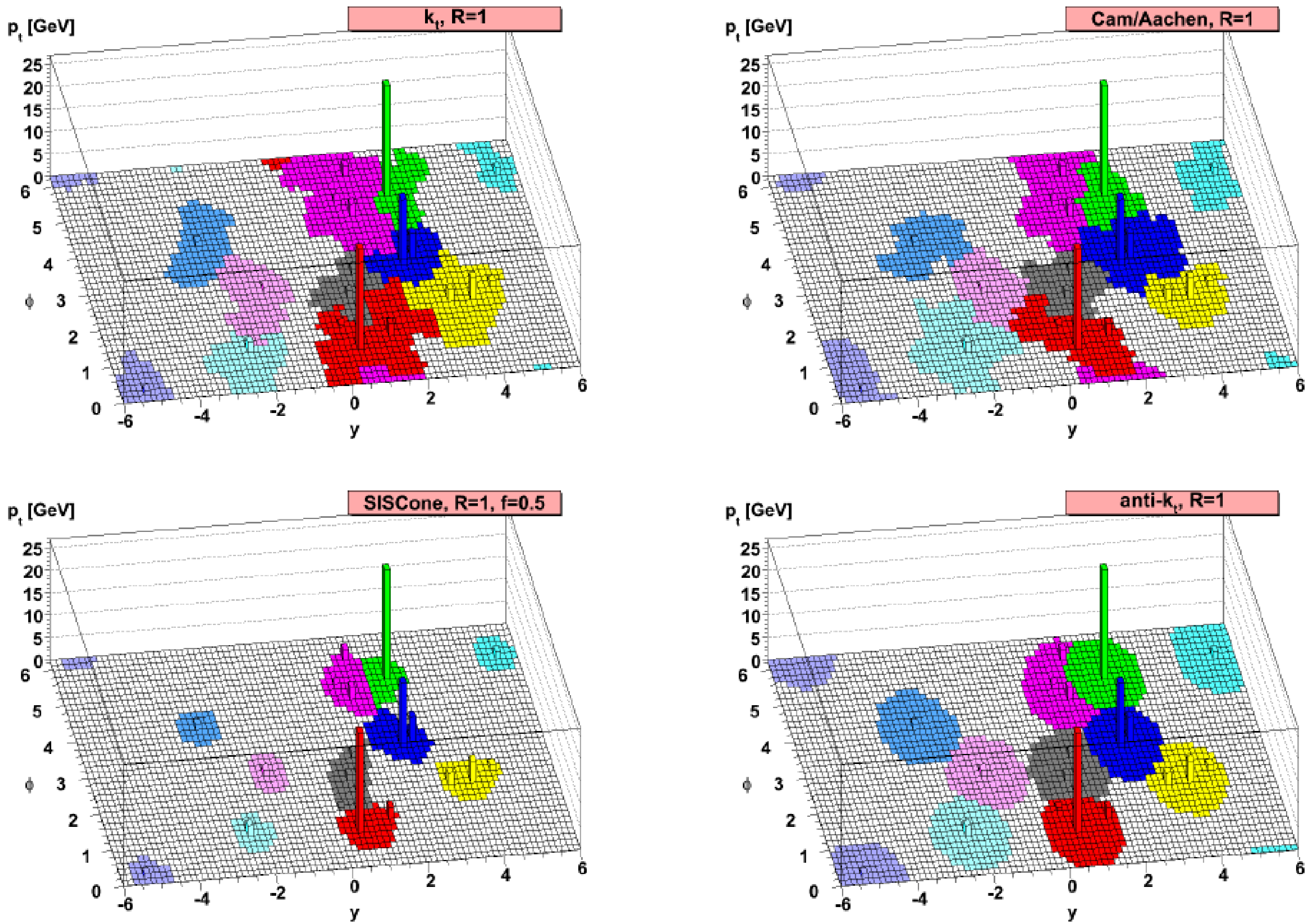


(from P.A. Delsart)

THE UNIVERSITY OF ARIZONA

Jet Shapes (3)

P. Loch
U of Arizona
March 19, 2010



(from G. Salam's talk at the ATLAS Hadronic Calibration Workshop Tucson 2008)

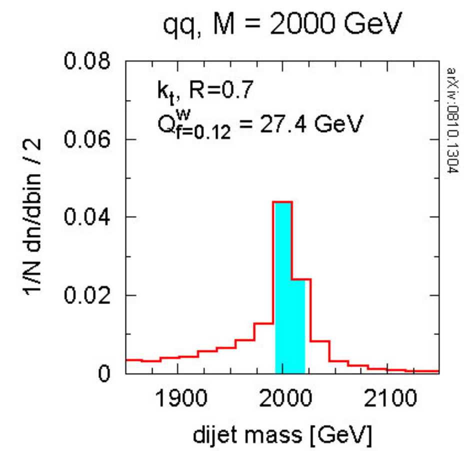
THE UNIVERSITY OF ARIZONA.

Jet Reconstruction Performance (1)

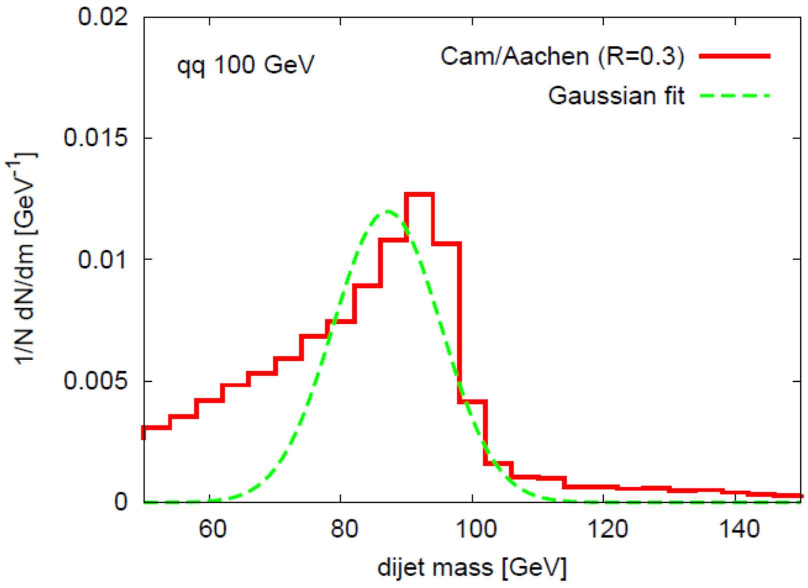
P. Loch
U of Arizona
March 19, 2010

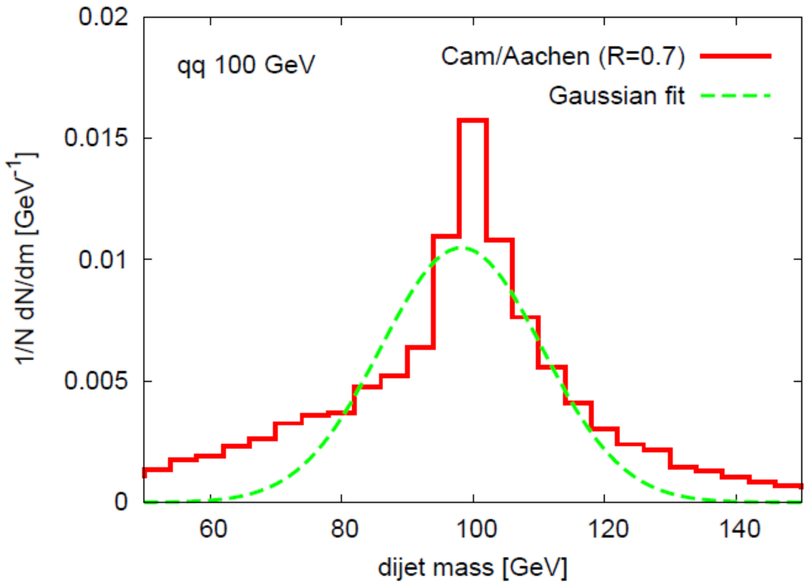
Quality estimator for distributions

- Best reconstruction: narrow Gaussian
  - We understand the error on the mean!
- Observed distributions often deviate from Gaussian
  - Need estimators on size of deviations!
  - Should be least biased measures
- Best performance gives closest to Gaussian distributions
  - List of variables describing shape of distribution on next slide
- Focus on unbiased estimators
  - E.g., distribution quantile describes the narrowest range of values containing a requested fraction of all events
  - Kurtosis and skewness harder to understand, but clear message in case of Gaussian distribution!



(from Salam ,Cacciari, Soyez, http://quality.fastjet.fr)



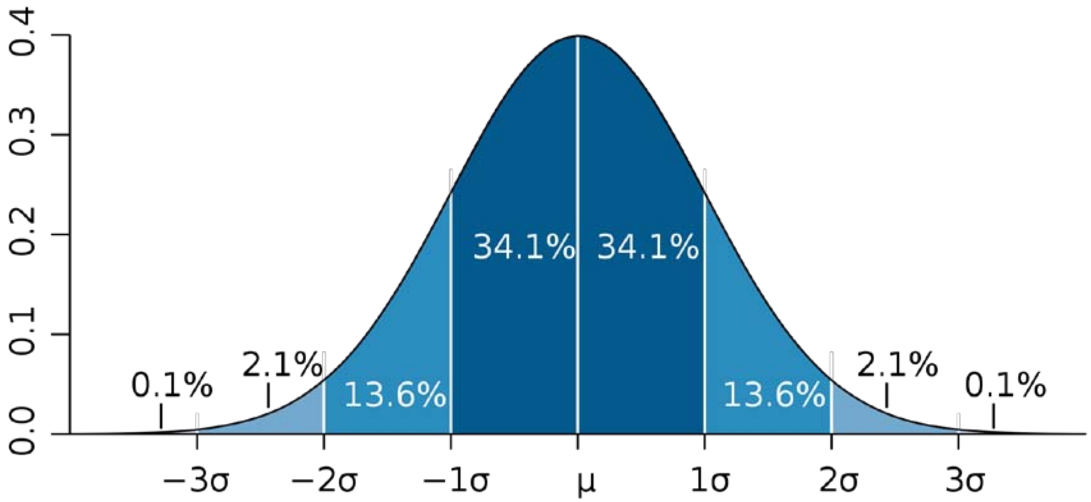


ATLAS

P. Loch
U of Arizona
March 19, 2010

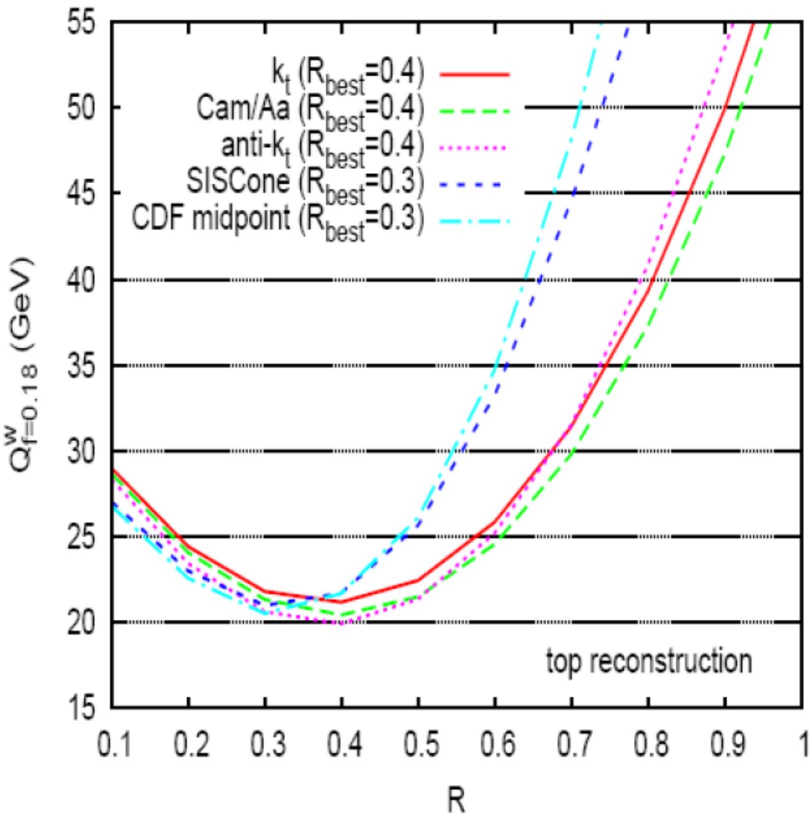
Jet Reconstruction Performance Estimators

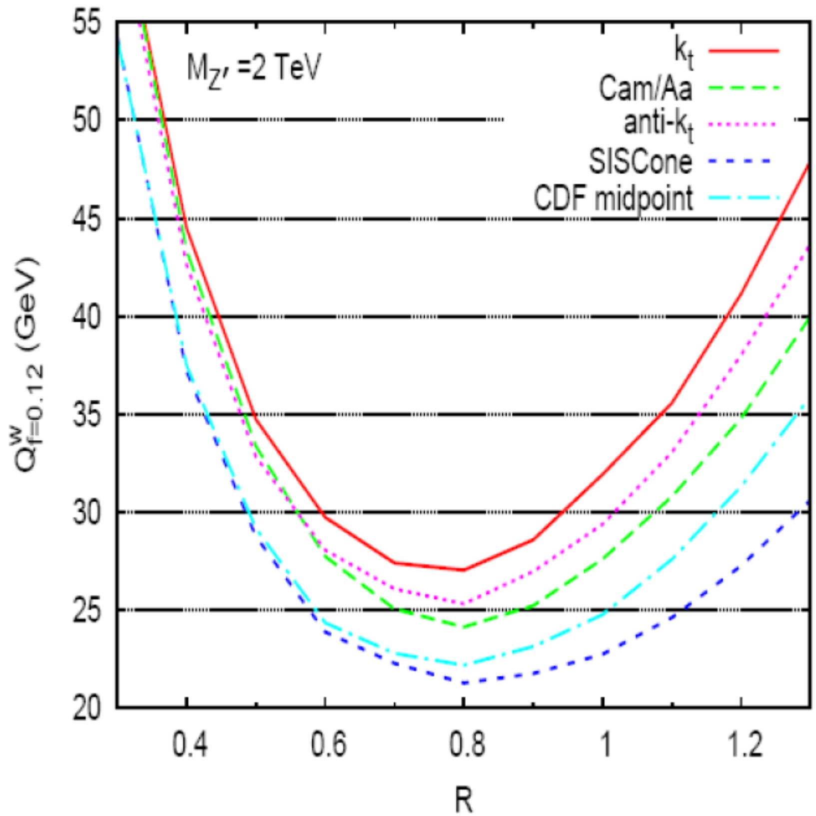
Estimator	Quantity	Expectation for Gaussian
$\langle R \rangle$	statistical mean	$\mu = \langle R \rangle = R_{\mathrm{mop}} = R_{\mathrm{median}}$
R_{median}	median	
R_{mop}	most probable value	
$RMS = \sqrt{\langle R^2 \rangle - \langle R \rangle^2}$	standard deviation	$\sigma = RMS$
$\gamma_3 = \frac{\sum_{i=1}^{N} \left(R_i - \langle R \rangle\right)^3}{N\sigma^3}$	skewness/left-right asymmetry	0
$\gamma_4 = \frac{\sum_{i=1}^{N} \left(R_i - \langle R \rangle\right)^4}{N\sigma^4} - 3$	kurtosis/"peakedness"	0
Q_f^w	quantile	$Q^w_{f \approx 68\%} = 2\sigma$



Quality of mass reconstruction for various jet finders and configurations

- Standard model – top quark hadronic decay
  - Left plot – various jet finders and distance parameters
- BSM – *Z'* (2 TeV) hadronic decay
  - Right plot – various jet finders with best configuration

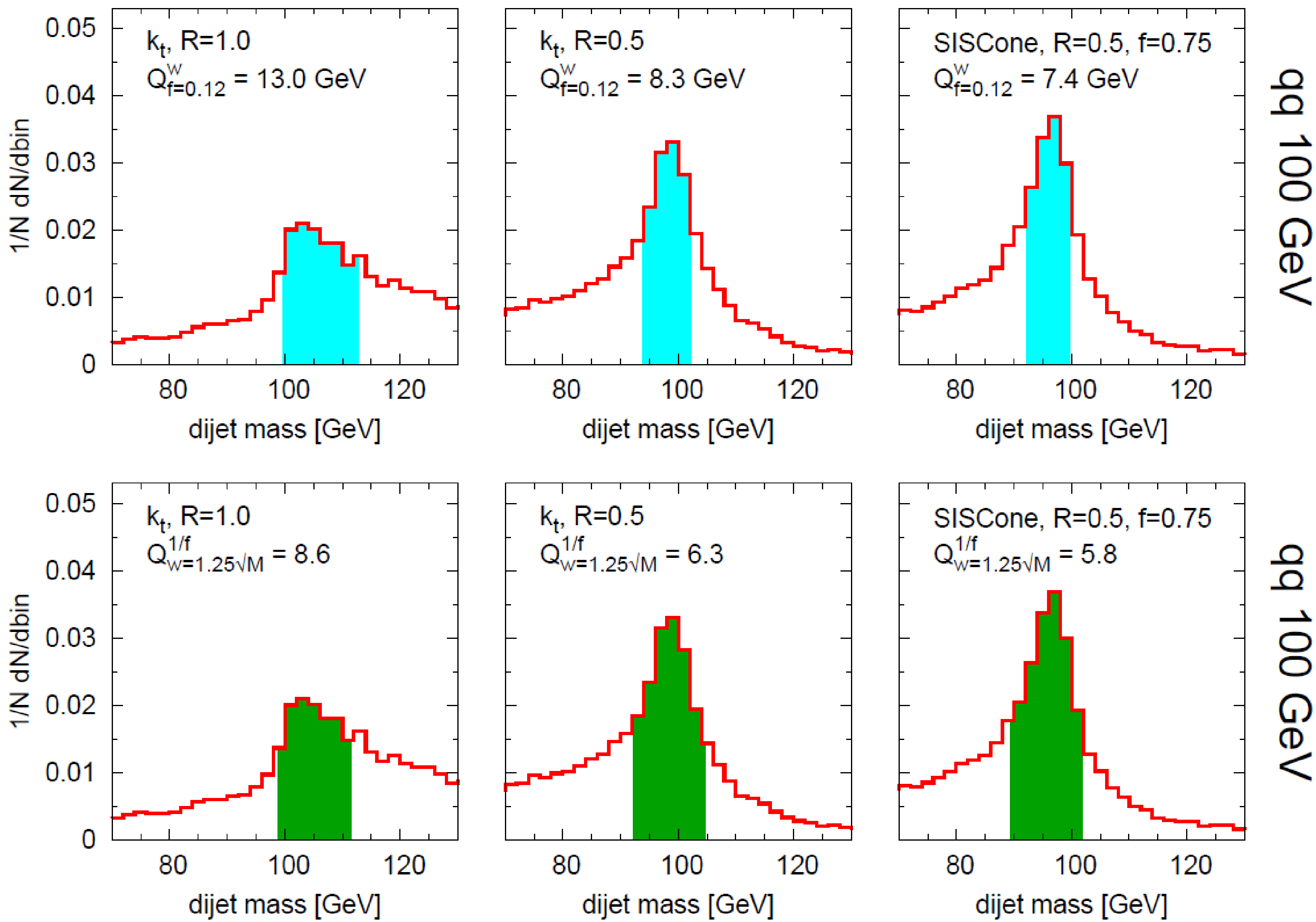




Jet Performance Examples (1)

P. Loch
U of Arizona
March 19, 2010

$$Q^{1/f}_{w=x\sqrt{M}} \equiv \left(\frac{\text{Max \# reco. massive objects in window of width } w = x\sqrt{M}}{\text{Total \# generated massive objects}}\right)^{-1}$$



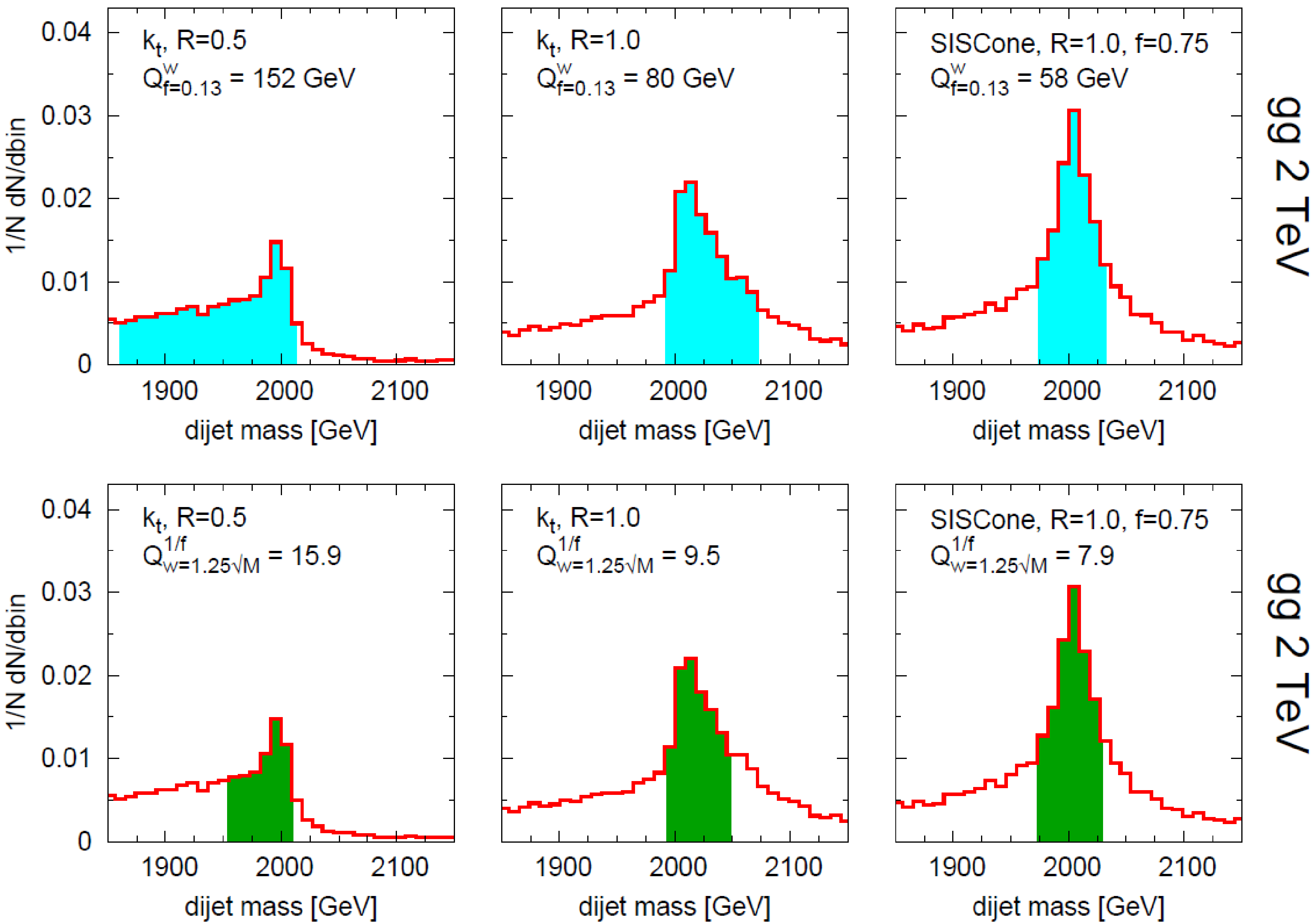
(from Cacciari, Rojo, Salam, Soyez, **JHEP 0812:032,2008**)

THE UNIVERSITY OF ARIZONA.

P. Loch
U of Arizona
March 19, 2010

Jet Performance Examples (2)

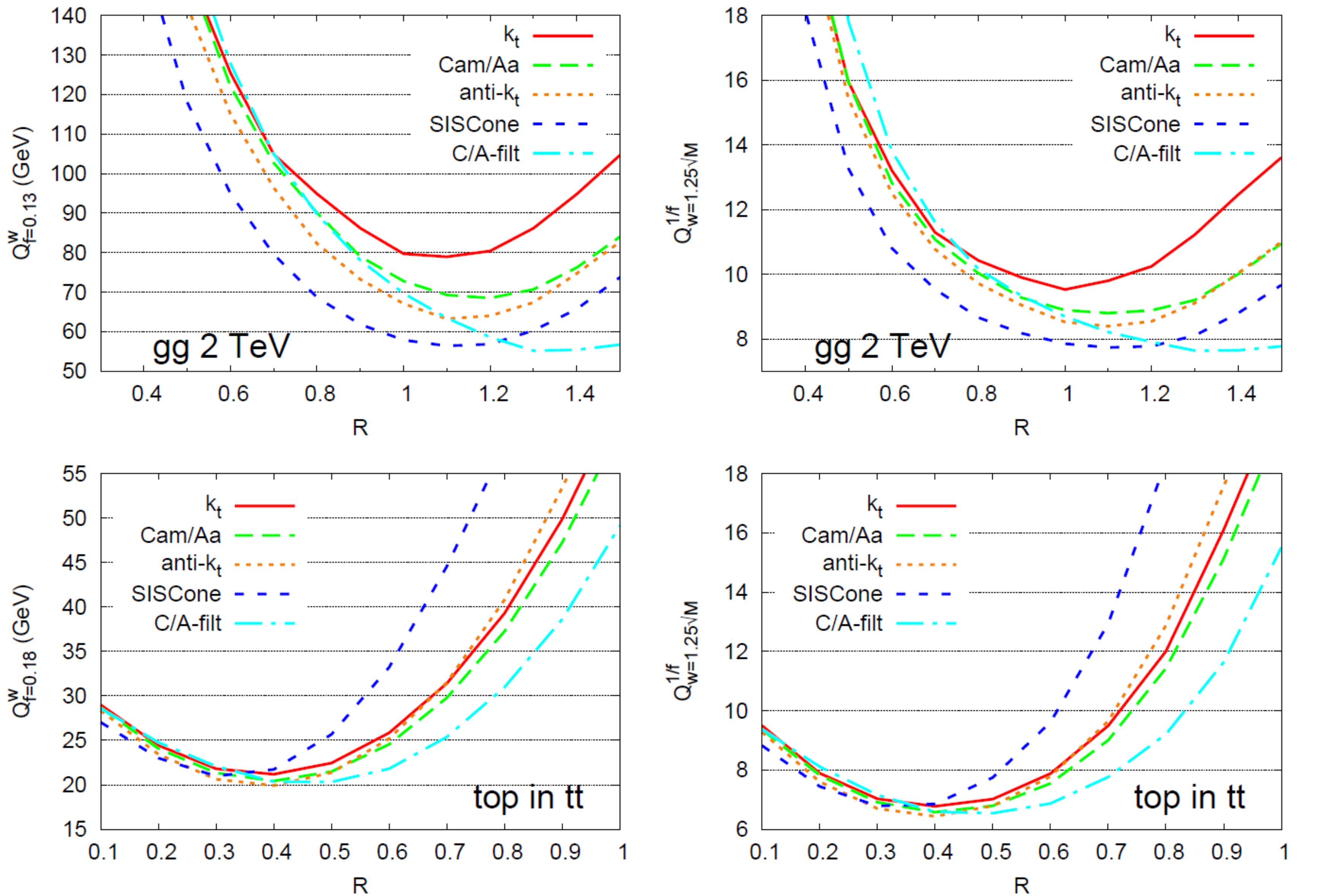
$$Q^{1/f}_{w=x\sqrt{M}} \equiv \left(\frac{\text{Max \# reco. massive objects in window of width } w = x\sqrt{M}}{\text{Total \# generated massive objects}}\right)^{-1}$$



(from Cacciari, Rojo, Salam, Soyez, **JHEP 0812:032,2008**)

THE UNIVERSITY OF ARIZONA.

Jet Performance Examples (3)

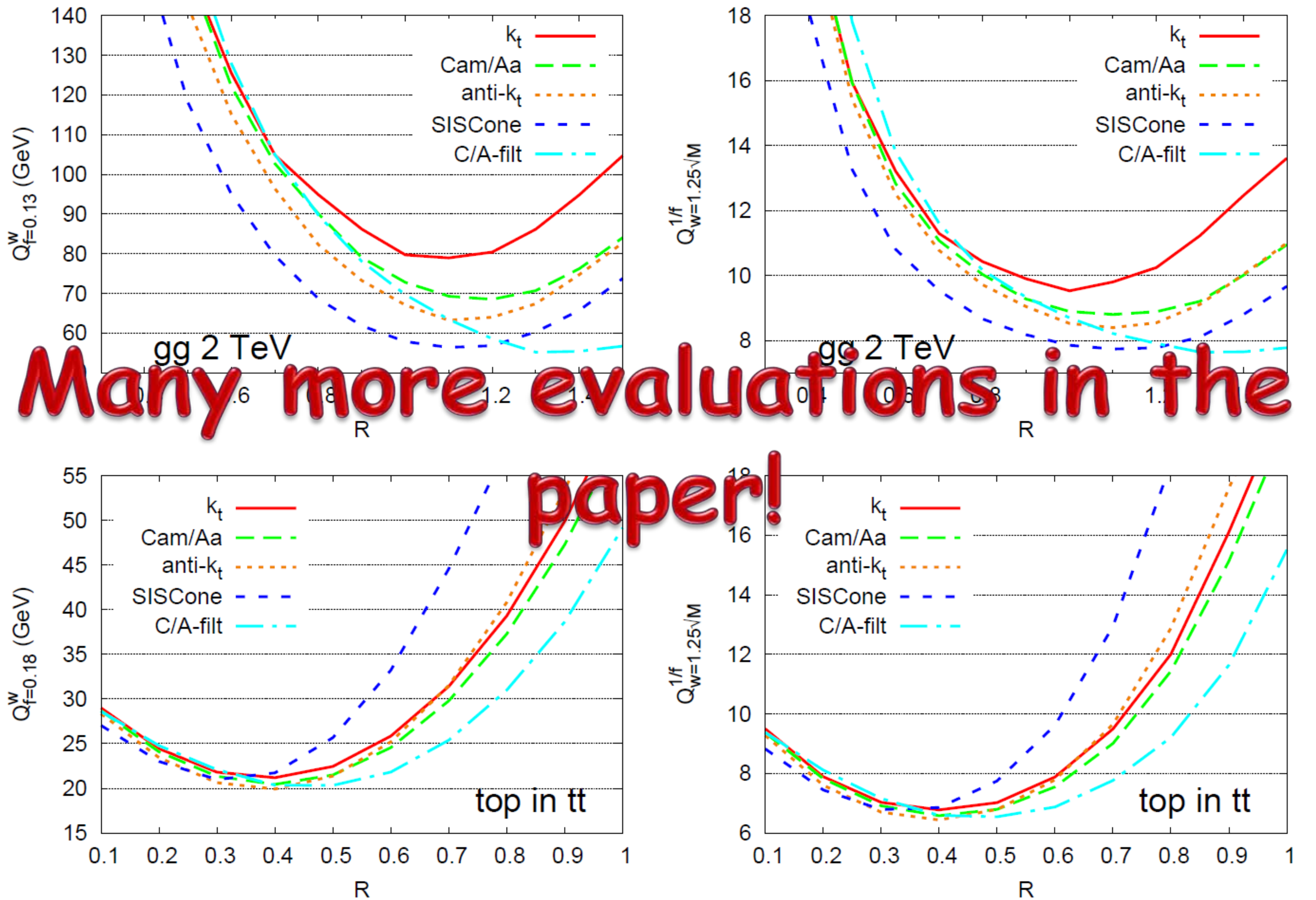


(from Cacciari, Rojo, Salam, Soyez, **JHEP 0812:032,2008**)

THE UNIVERSITY OF ARIZONA.

Jet Performance Examples (3)

P. Loch
U of Arizona
March 19, 2010



(from Cacciari, Rojo, Salam, Soyez, **JHEP 0812:032,2008**)

ATLAS

Interactive Tool

Web-based jet performance evaluation available

http://www.lpthe.jussieu.fr/~salam/jet-quality