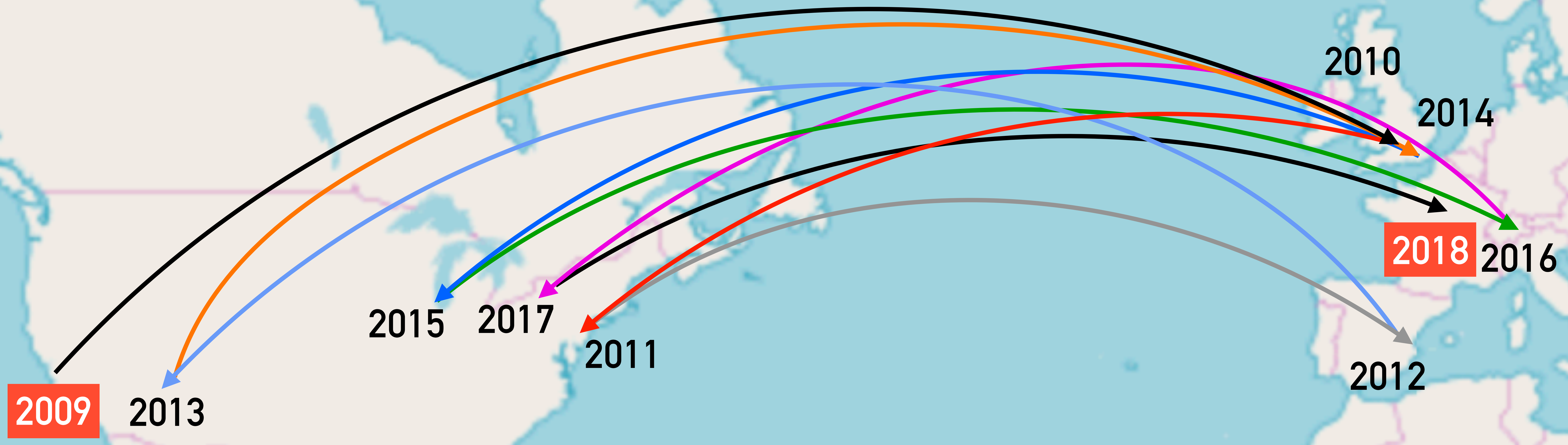
10 years of boost: a theory perspective

*Gavin P. Salam, CERN**

Boost 2018, 19 July 2018, Paris, France



~ 400 (exp+th) talks!

*on leave from CNRS and University of Oxford/Royal Society

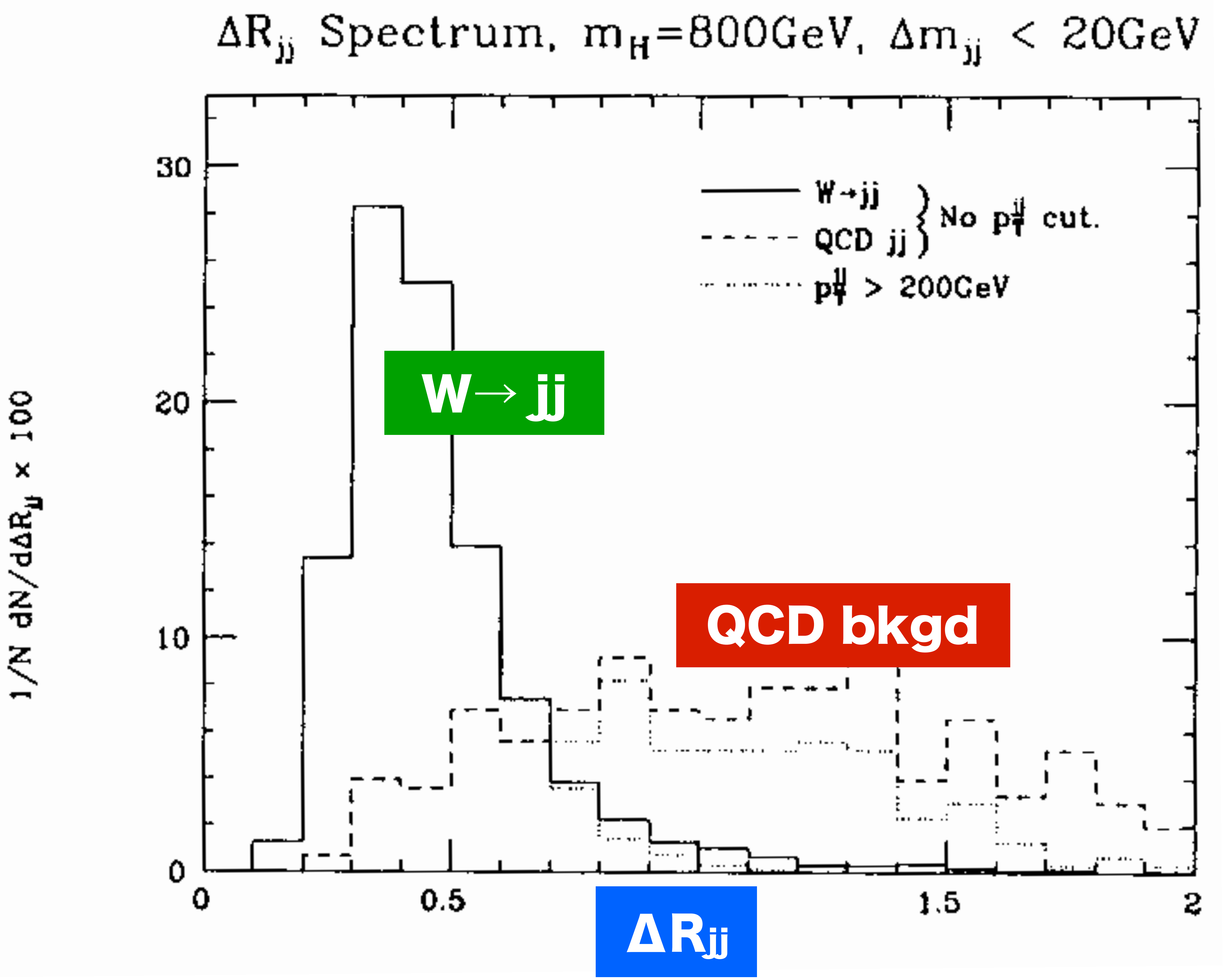
pre-prehistory

1991

Tagging a heavy Higgs boson
M.H. Seymour (Cambridge U.). Jan 22, 1991. 16 pp.
CAVENDISH-HEP-90-25
Talk presented at Conference: C90-10-04 (Aachen ECFA Workshop 1990:0557-569), p.557-569 Proceedings

References | BibTeX | LaTeX(US) | LaTeX(EU) | Harvmac | EndNote
KEK scanned document; CERN Document Server

Detailed record - Cited by 2 records



The W-finder used in this study utilizes this cut by running a jet-finder twice, with cone sizes of $\Delta R = 0.75$ and $\Delta R = 0.25$, and then demands a big jet containing two small jets, with $|m_{jj} - m_W| < 10$ GeV. This set of cuts still leaves the background a factor of ≈ 75 larger than the signal, and without further cuts the situation is hopeless.

1991

New clustering algorithm for multi - jet cross-sections in e+ e- annihilation

S. Catani (Cambridge U. & INFN, Florence), Yuri L. Dokshitzer (St. Petersburg, INP & Lund U.), M. Olsson (Lund U.), G. Turnock, B.R. Webber (Cambridge U.). Jul 1991. 7 pp.

Published in **Phys.Lett. B269 (1991) 432-438**
CAVENDISH-HEP-91-5
DOI: 10.1016/0370-2693(91)90196-W

References | BibTeX | LaTeX(US) | LaTeX(EU) | Harvmac | EndNote
ADS Abstract Service

Longitudinally invariant K_t clustering algorithms for hadron hadron collisions

S. Catani (CERN), Yuri L. Dokshitzer (St. Petersburg, INP & Lund U.), M.H. Seymour (Lund U.), B.R. Webber (CERN). Feb 1993. 37 pp.

Published in **Nucl.Phys. B406 (1993) 187-224**
CERN-TH-6775-93, LU-TP-93-2
DOI: 10.1016/0550-3213(93)90166-M

References | BibTeX | LaTeX(US) | LaTeX(EU) | Harvmac | EndNote
CERN Document Server; CERN Document Server

$$d_{kB} = p_{tk}^2, \quad d_{kl} = \min(p_{tk}^2, p_{tl}^2) R_{kl}^2,$$

$$R_{kl}^2 = (\eta_k - \eta_l)^2 + (\phi_k - \phi_l)^2.$$

1993

...nation jet algorithm for hadron collisions

Stephen D. Ellis (CERN), Davison E. Soper (Oregon U.). May 24, 1993. 29 pp.

Published in **Phys.Rev. D48 (1993) 3160-3166**
CERN-TH-6860-93
DOI: 10.1103/PhysRevD.48.3160
e-Print: **hep-ph/9305266 | PDF**

References | BibTeX | LaTeX(US) | LaTeX(EU) | Harvmac | EndNote
KEK scanned document; CERN Document Server; ADS Abstract Service; OSTI.gov Server; CERN Document Server

2. Find the smallest of all the d_i and d_{ij} and label it $d_{\min}$.

3. If $d_{\min}$ is a d_{ij}, merge protojets i and j into a new protojet k

4. If $d_{\min}$ is a d_i, the corresponding protojet i is "not mergable." Remove it from the list of protojets and add it to the list of jets.

1993

Searches for new particles using cone and cluster jet algorithms: A Comparative study

Michael H. Seymour (Lund U.). Jun 1993. 23 pp.

Published in **Z.Phys. C62 (1994) 127-138**
LU-TP-93-8
DOI: 10.1007/BF01559532

"As a simple example (in fact the only way in which we use sub-jets in this paper), one could cluster the event until there is exactly one jet remaining-this is then the hardest jet. Then **one could recluster only those particles that ended up in the hardest jet until there are exactly two jets**-these are then the sub-jets corresponding to the hardest emission within the hardest jet."

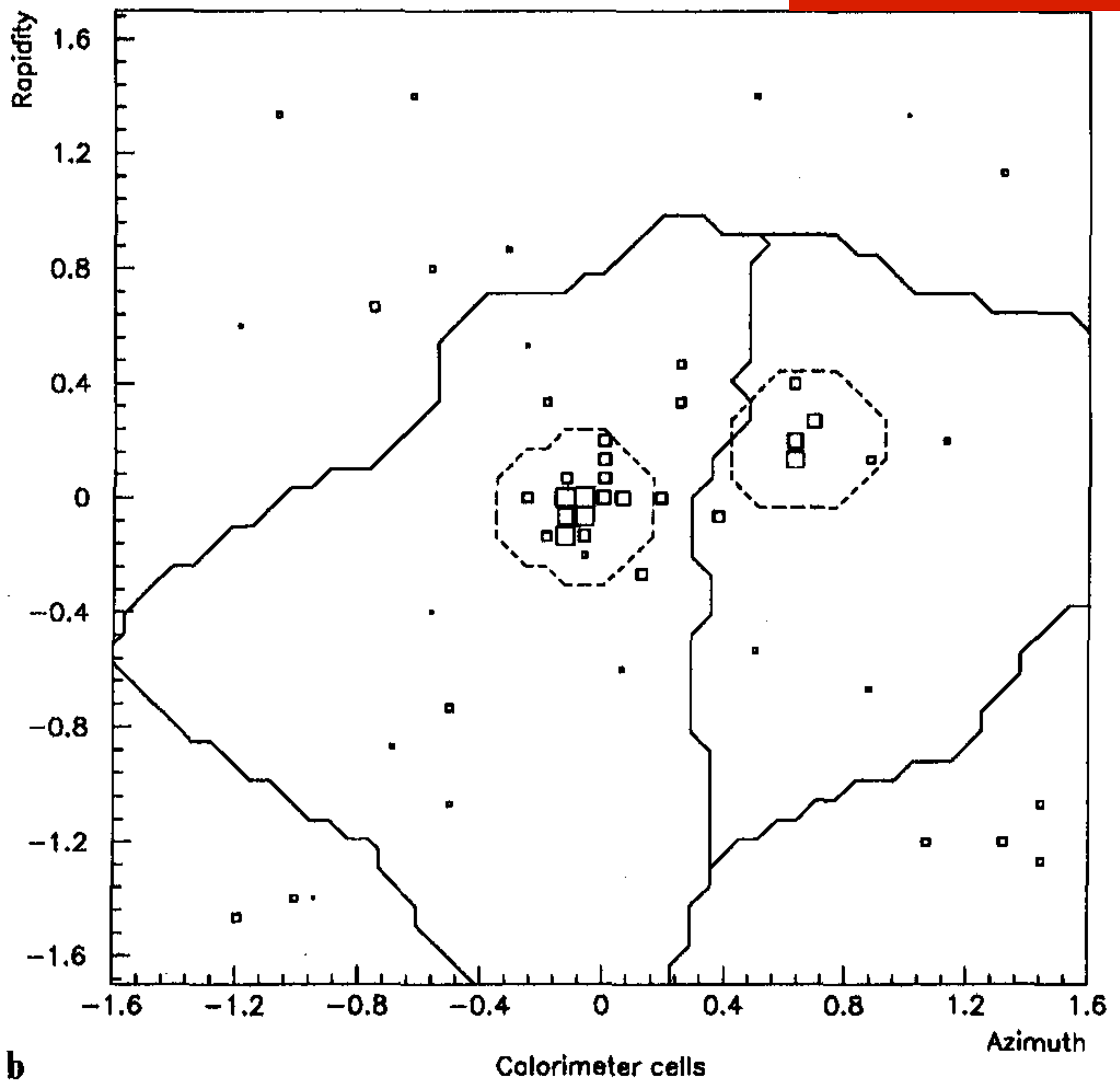


Fig. 2. A hadronic W decay, as seen at calorimeter level, **a** without, and **b** with, particles from the underlying event. Box sizes are logarithmic in the cell energy, lines show the borders of the sub-jets for infinitely soft emission according to the cluster (solid) and cone (dashed) algorithms

"Then we recluster only those particles that ended up in the hardest jet, using a radius $R = \alpha R_{jj}$,"

1997/8 Cambridge/Aachen algorithm

Better jet clustering algorithms
Yuri L. Dokshitzer (Milan U.), G.D. Leder, S. Moretti, B.R. Webber (Cambridge U.). Jul 1997. 33 pp.
Published in **JHEP 9708 (1997) 001**
CAVENDISH-HEP-97-06
DOI: 10.1088/1126-6708/1997/08/001
e-Print: **hep-ph/9707323** | **PDF**

References | BibTeX | LaTeX(US) | LaTeX(EU) | Harvmac | EndNote
ADS Abstract Service

Detailed record - Cited by 890 records 500+

Hadronization corrections to jet cross-sections in deep inelastic scattering
M. Wobisch (Aachen, Tech. Hochsch.), T. Wengler (Heidelberg U.). Apr 1998. 10 pp.
Published in **In *Hamburg 1998/1999, Monte Carlo generators for HERA physics* 270-279**
PITHA-99-16
To be published in the proceedings of Conference: C98-04-27, p.270-279
Proceedings
e-Print: **hep-ph/9907280** | **PDF**

References | BibTeX | LaTeX(US) | LaTeX(EU) | Harvmac | EndNote
ADS Abstract Service

Detailed record - Cited by 477 records 250+

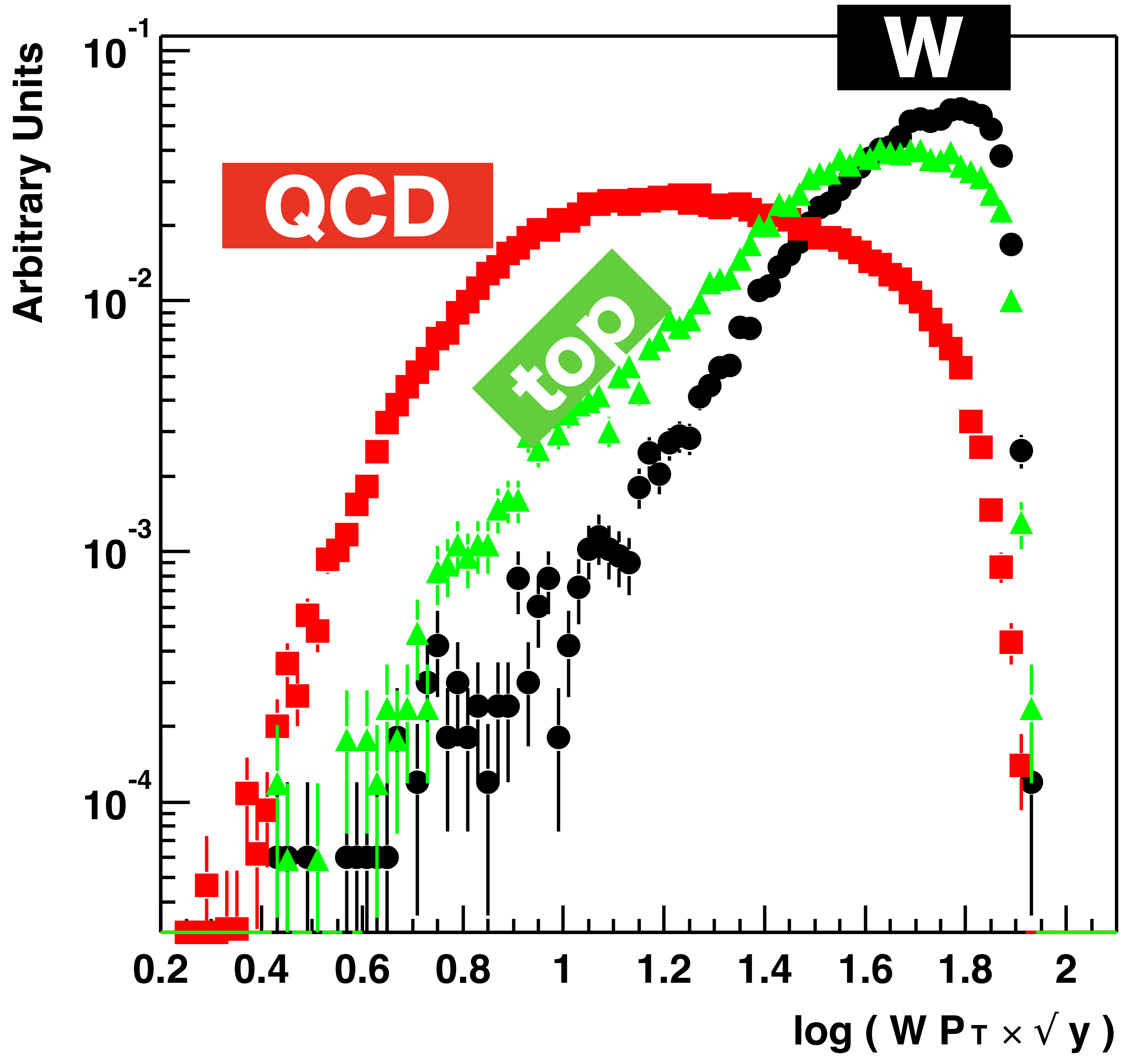
ABSTRACT: We investigate modifications to the $k_\perp$-clustering jet algorithm which preserve the advantages of the original Durham algorithm while reducing non-perturbative corrections and providing better resolution of jet substructure. We find that a simple change in the sequence of clustering (combining smaller-angle pairs first), together with the 'freezing' of soft resolved jets, has beneficial effects.

2002

WW scattering at the CERN LHC

J.M. Butterworth (University Coll. London), B.E. Cox, Jeffrey R. Forshaw (Manchester U.). Jan 2002. 29 pp.

Published in **Phys.Rev. D65 (2002) 096014**
MC-TH-01-13, MAN-HEP-01-05, UCL-HEP-2001-06
DOI: 10.1103/PhysRevD.65.096014
e-Print: hep-ph/0201098 | PDF



$=½log(d_{12})$

this analysis we develop a new technique. The extra pieces of information gained from the subjet decomposition are the y cut at which the subjets are defined and the four-vectors of the subjets. For a genuine W decay the expectation is that the scale at which the jet is resolved into subjets (i.e. yp_T^2) will be $\mathcal{O}(M_W^2)$. The distribution of $\log(p_T\sqrt{y})$ is shown in Figure 12(d). The scale of the splitting is indeed high in the signal and softer in the $W+$ jets background, where the hadronic W is in general a QCD jet rather than a genuine second W. A cut is applied at $1.6 < \log(p_T\sqrt{y}) < 2.0$. The effect of this cut is

prehistory

c. 2005

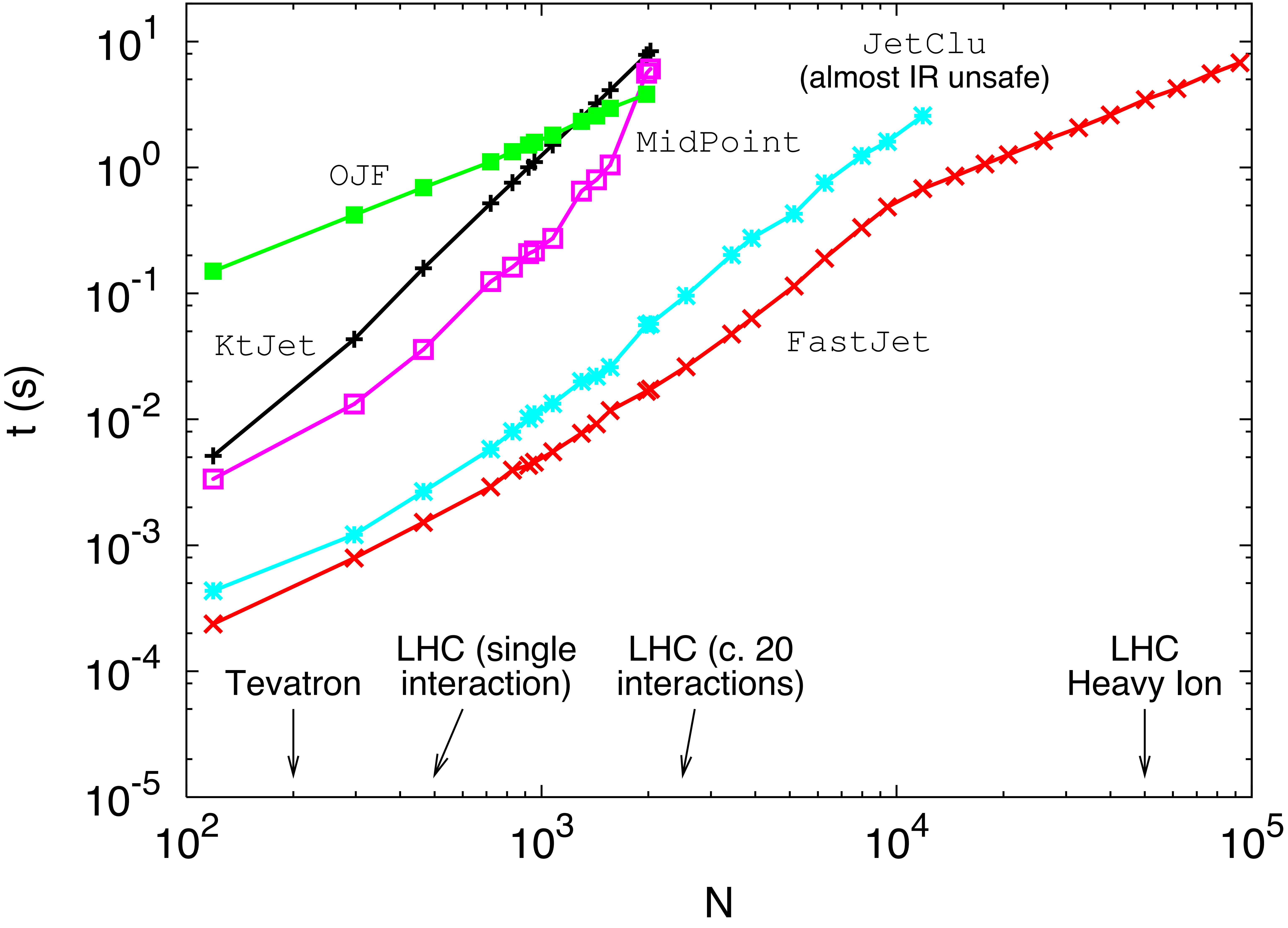
kt adapts to the jet structure
the cone is too rigid
cone has big hadronisation corrections
cone is infrared unsafe
many theorists
the cone gives nice conical jets
kt's a vacuum cleaner
kt's too irregular I can't correct for pileup
kt's too slow
many experimenters

Dispelling the N^3 myth for the k_t jet-finder

Matteo Cacciari and Gavin P. Salam

LPTHE, Universities of Paris VI and VII and CNRS,

arXiv:hep-ph/0512210v2



speed [fastjet]

2005

2007/8

arXiv:0802.1188

arXiv:0707.1378

Pileup subtraction using jet areas

Matteo Cacciari and Gavin P. Salam

LPTHE, Université P. et M. Curie – Paris 6,

Université D. Diderot – Paris 7, CNRS UMR 7589, Paris, France

The Catchment Area of Jets

Matteo Cacciari, Gavin P. Salam

LPTHE

UPMC Université Paris 6,

Université Paris Diderot – Paris 7,

CNRS UMR 7589, Paris, France

Gregory Soyez

Brookhaven National Laboratory, Upton, NY 11973, USA

2007

A practical Seedless Infrared-Safe Cone jet algorithm

Gavin P. Salam and Grégory Soyez*†

LPTHE,
Université Pierre et Marie Curie – Paris 6,
Université Denis Diderot – Paris 7,
CNRS UMR 7589, 75252 Paris cedex 05, France.

arXiv:0704.0292

algorithm part 1

Algorithm 1 A full specification of a modern cone algorithm, governed by four parameters: the cone radius R, the overlap parameter f, the number of passes N_{pass} and a minimum transverse momentum in the split–merge step, $p_{t,\min}$. Throughout, particles are to be combined by summing their 4-momenta and distances are to be calculated using the longitudinally invariant Δy and $\Delta\phi$ distance measures (where y is the rapidity).

1: Put the set of current particles equal to the set of all particles in the event.
2: **repeat**
3: Find *all* stable cones of radius R (see Eq. (1)) for the current set of particles, *e.g.* using algorithm 2, section 4.2.2.
4: For each stable cone, create a protojet from the current particles contained in the cone, and add it to the list of protojets.
5: Remove all particles that are in stable cones from the list of current particles.
6: **until** No new stable cones are found, or one has gone around the loop N_{pass} times.
7: Run a Tevatron Run-II type split–merge procedure [6], algorithm 3 (section 4.3), on the full list of protojets, with overlap parameter f and transverse momentum threshold $p_{t,\min}$.

part 2

Algorithm 2 Procedure for establishing the list of all stable cones (protojets). For simplicity, parts related to the special case of multiple cocircular points (see footnote 7) are not shown. They are a straightforward generalisation of steps 2 to 2.

1: For any group of collinear particles, merge them into a single particle.
2: **for** particle $i = 1 \dots N$ **do**
3: Find all particles j within a distance $2R$ of i. If there are no such particles, i forms a stable cone of its own.
4: Otherwise for each j identify the two circles for which i and j lie on the circumference. For each circle, compute the angle of its centre C relative to i, $\zeta = \arctan \frac{\Delta\phi_{iC}}{\Delta y_{iC}}$.
5: Sort the circles found in steps 2 and 2 into increasing angle ζ.
6: Take the first circle in this order, and call it the current circle. Calculate the total momentum and checkxor for the cones that it defines. Consider all 4 permutations of edge points being included or excluded. Call these the "current cones".
7: **repeat**
8: **for** each of the 4 current cones **do**
9: If this cone has not yet been found, add it to the list of distinct cones.
10: If this cone has not yet been labelled as unstable, establish if the in/out status of the edge particles (with respect to the cone momentum axis) is the same as when defining the cone; if it is not, label the cone as unstable.
11: **end for**
12: Move to the next circle in order. It differs from the previous one either by a particle entering the circle, or one leaving the circle. Calculate the momentum for the new circle and corresponding new current cones by adding (or removing) the momentum of the particle that has entered (left); the checkxor can be updated by XORing with the label of that particle.
13: **until** all circles considered.
14: **end for**
15: **for** each of the cones not labelled as unstable **do**
16: Explicitly check its stability, and if it is stable, add it to the list of stable cones (protojets).
17: **end for**

part3

Algorithm 3 The disambiguated, scalar $\tilde{p}_t$ based formulation of a Tevatron Run-II type split–merge procedure [6], with overlap threshold parameter f and transverse momentum threshold $p_{t,\min}$. To ensure boost invariance and IR safety, for the ordering variable and the overlap measure, it uses of $\tilde{p}_{t,\text{jet}} = \sum_{i\in\text{jet}} |p_{t,i}|$, *i.e.* a scalar sum of the particle transverse momenta (as in a 'p_t' recombination scheme).

1: **repeat**
2: Remove all protojets with $p_t < p_{t,\min}$.
3: Identify the protojet (i) with the highest $\tilde{p}_t$.
4: Among the remaining protojets identify the one (j) with highest $\tilde{p}_t$ that shares particles (overlaps) with i.
5: **if** there is such an overlapping jet **then**
6: Determine the total $\tilde{p}_{t,\text{shared}} = \sum_{k\in i\&j} |p_{t,k}|$ of the particles shared between i and j.
7: **if** $\tilde{p}_{t,\text{shared}} < f\tilde{p}_{t,j}$ **then**
8: Each particle that is shared between the two protojets is assigned to the one to whose axis it is closest. The protojet momenta are then recalculated.
9: **else**
10: Merge the two protojets into a single new protojet (added to the list of protojets, while the two original ones are removed).
11: **end if**
12: If steps 3–3 produced a protojet that coincides with an existing one, maintain the new protojet as distinct from the existing copy(ies).
13: **else**
14: Add i to the list of final jets, and remove it from the list of protojets.
15: **end if**
16: **until** no protojets are left.

2007/8

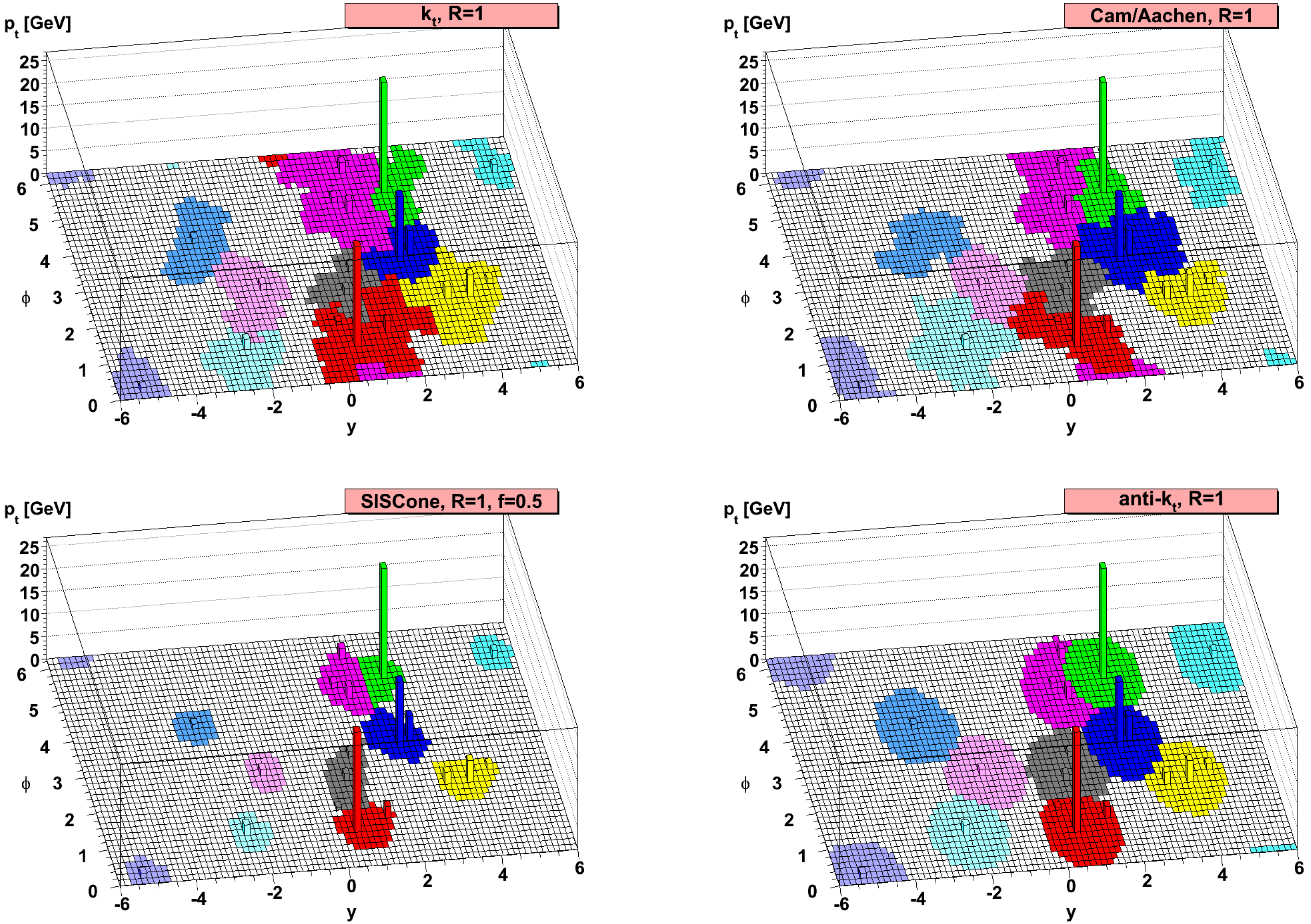
The anti-k_t jet clustering algorithm

Matteo Cacciari and Gavin P. Salam
LPTHE
UPMC Université Paris 6,
Université Paris Diderot – Paris 7,
CNRS UMR 7589, Paris, France

Gregory Soyez
Brookhaven National Laboratory, Upton, NY, USA

arXiv:0802.1189

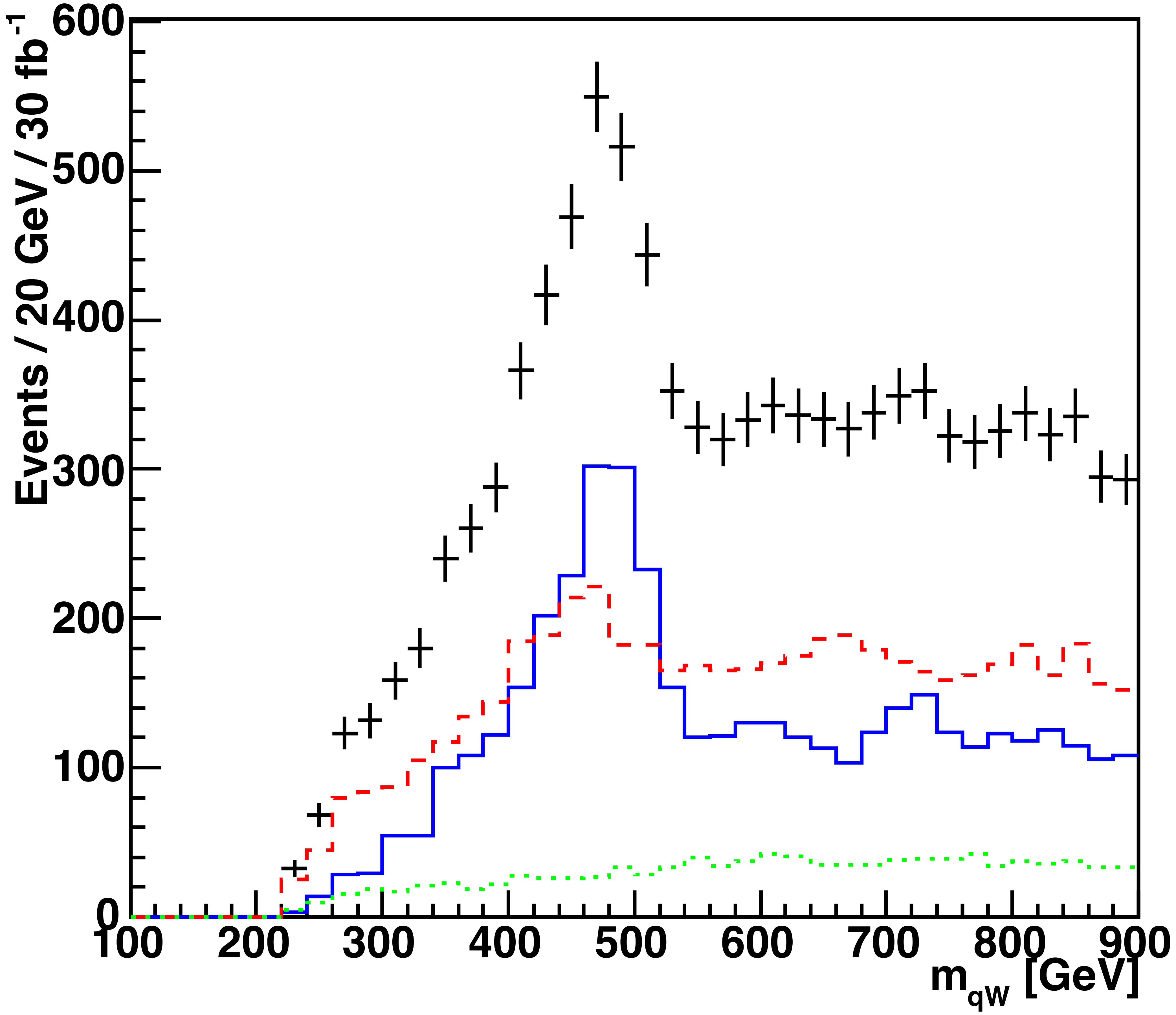
cone-like seq.rec. jets



Reconstructing Sparticle Mass Spectra using Hadronic Decays

J. M. Butterworth[1], John Ellis[2] and A. R. Raklev[2,3]

mass of hadronic boosted W + another quark demonstrating clear kinematic edge

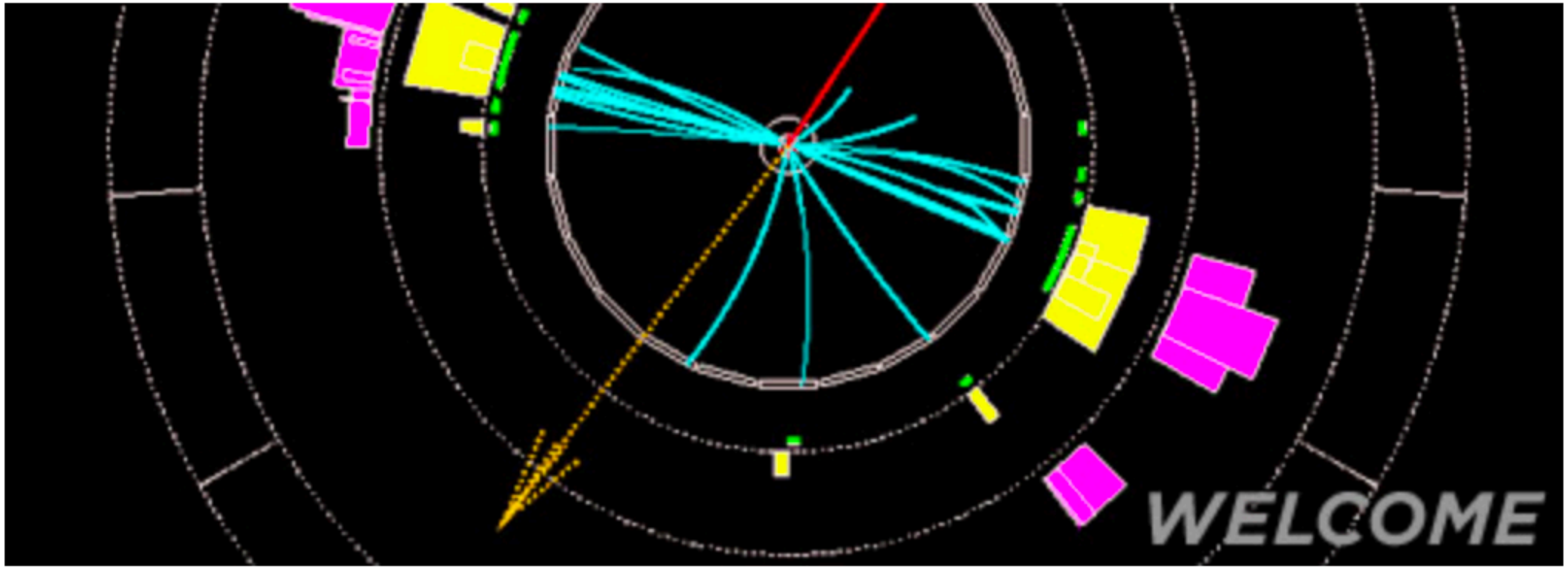


2007

BOOST 2009

GIVING NEW PHYSICS A BOOST — SLAC NATIONA

- Home
- Registration
- Participant List
- Agenda
- Accommodations
- General Information
- Travel and Directions
- Visa Information
- Social Event



Giving New Physics a Boost

Thursday and Friday, July 9-10, 2009 from 8:00 am to 5:00 pm.

Kavli Auditorium

SLAC National Accelerator Laboratory

Menlo Park, California

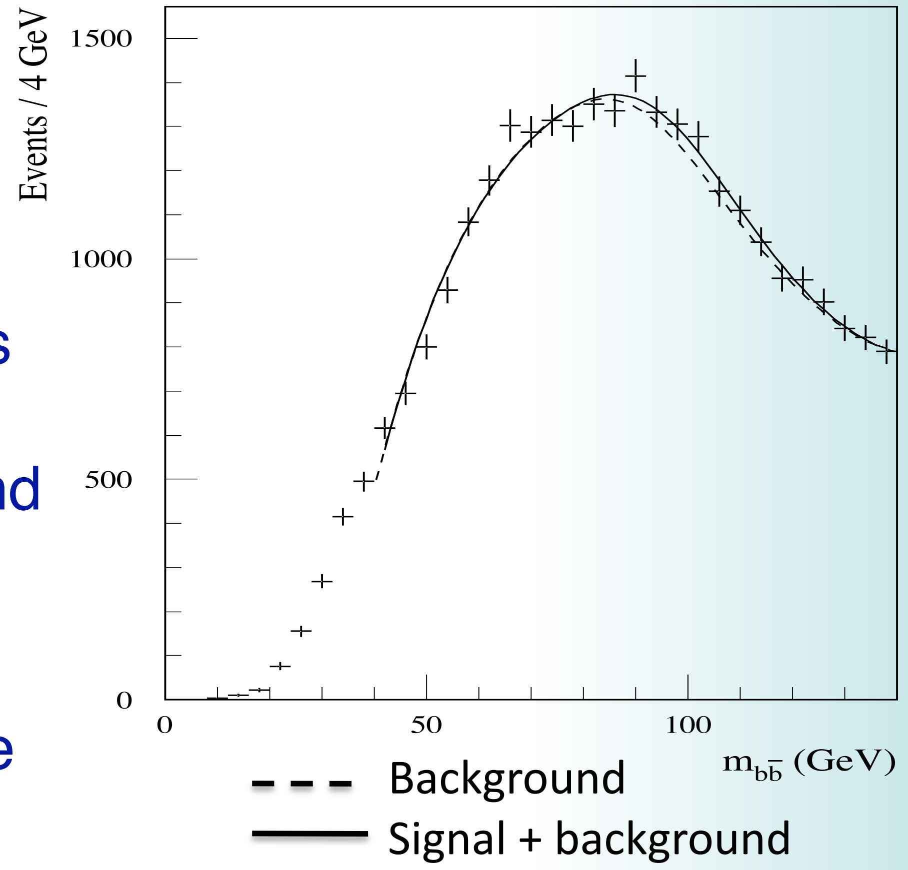
Boost '09

Why and where Jet Substructure Techniques work at the LHC

Jon Butterworth

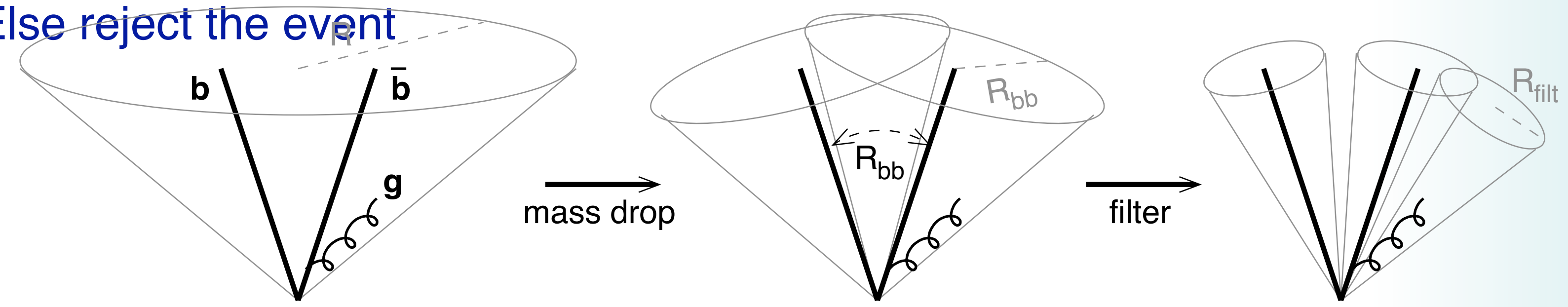
Higgs + (W or Z)

- Example: ATLAS Physics TDR (1999)
  - Poor acceptance
  - Cuts introduce artificial mass scale into the background
  - Top anti-top has a similar mass scale
  - Large combinatorial background
- Signal swamped by backgrounds
  - "very difficult … even under the most optimistic assumptions"

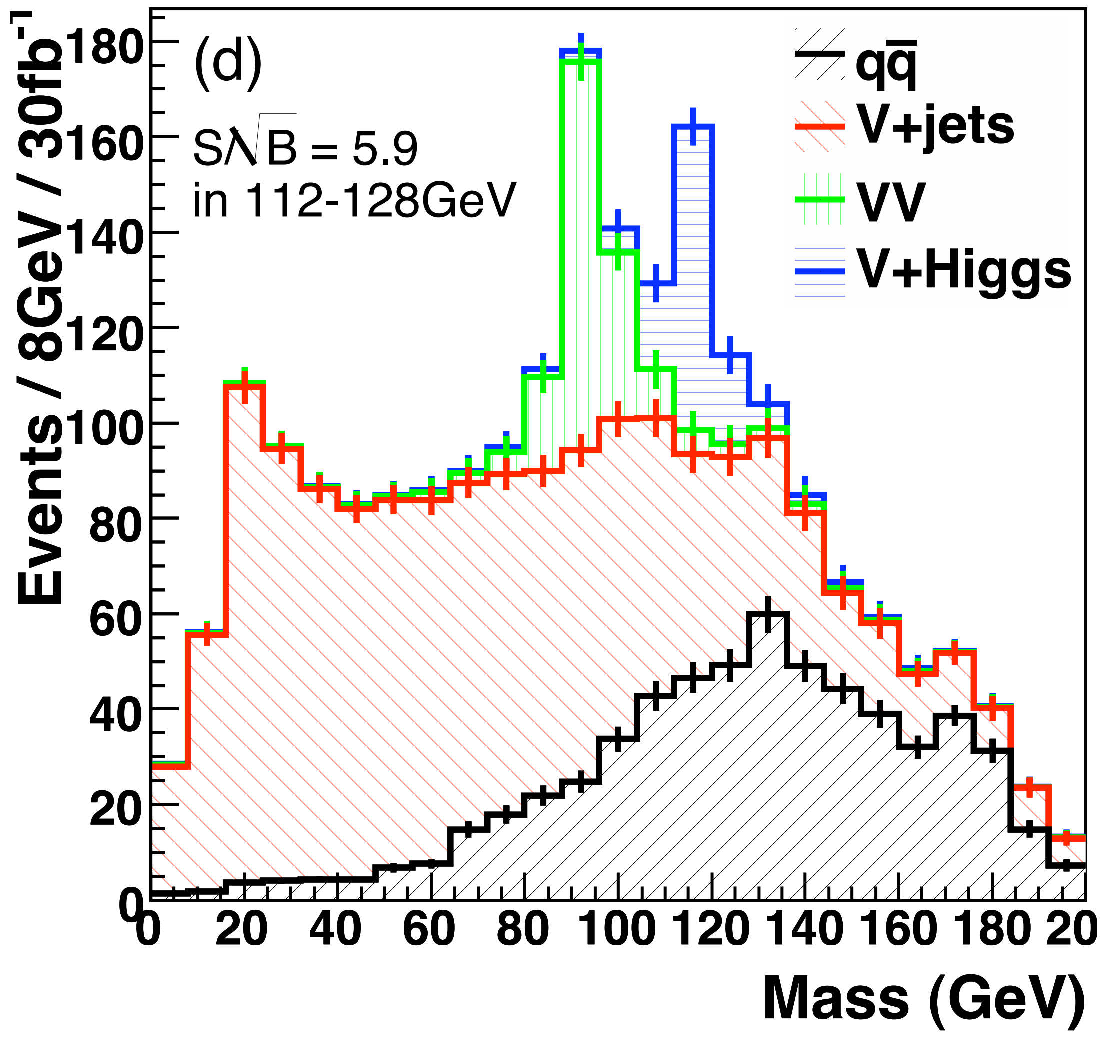


Sub-jet analysis

1. Start with Higgs candidate jet (highest p_T jet in acceptance) with mass m)
2. Undo last stage of clustering (reduce radius to R_{12})
   $J \rightarrow J_1, J_2$
3. If $\max(m_1,m_2) < 2m/3$
   Call this a "mass drop". This fixes the optimal radius for reconstructing the Higgs decay. Keep the jet J and call it the Higgs candidate.
   Else, go back to 2
4. Require $Y_{12} > 0.09$
   Dimensionless rejection of asymmetric QCD splitting
   Else reject the event
5. Require J_1, J_2 to each contain a b-tag
   Else reject the event



Combined particle-level result



- Note excellent Z peak for calibration
- 5.9 σ; potentially very competitive
- bb branching information critical for extracting Higgs properties
  - *"Measuring the Higgs sector" Lafaye, Plehn, Rauch, D.Zerwas, Duhrssen, arXiv:0904.3866 [hep-ph]*
- Studies within ATLAS are promising and nearly public.

Boost '09

Looking for New (BSM) Physics at the LHC with Single Jets: PRUNING

Steve Ellis, Jon Walsh and Chris Vermilion

0903.5081
0907.XXXX

- go to tinyurl.com/jetpruning

Big Picture:

The LHC will be both very exciting and very challenging –

• most of the data will be about hadrons (jets)

• many interesting objects (W's, Z's, tops, SUSY particles) will be boosted enough to appear in single jet

• must be able to ID/reconstruct these jets to find the BSM physcs

Department of Physics
University of Washington

Giving New Physics a Boost - 2009
SLAC 7/09/09

Pruning :

Procedure:

- Start with the objects (e.g. towers) forming a jet found with a recombination algorithm
- Rerun the algorithm, but at each recombination test whether:
  - $z < z_{cut}$ and $\Delta R_{ij} > D_{cut}$
  - $m_J/P_{T,J}$ is IR safe measure of opening angle of found jet

CA: $z_{cut} = 0.1$ and $D_{cut} = m_J/P_{T,J}$

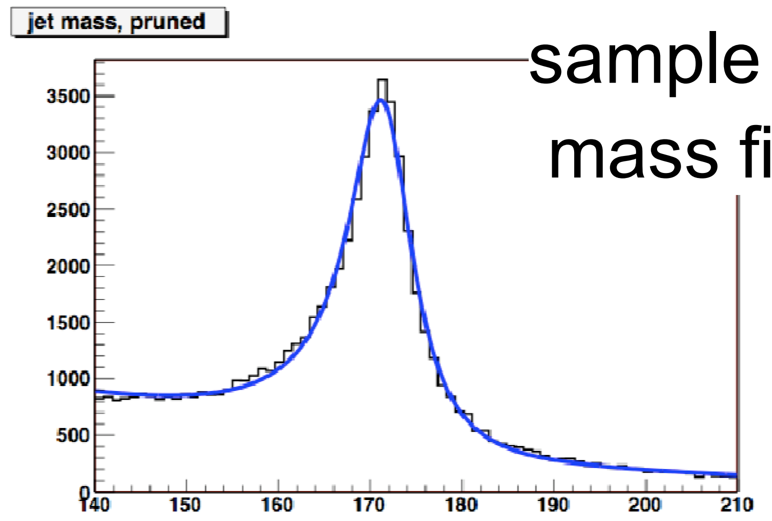
kT: $z_{cut} = 0.15$ and $D_{cut} = m_J/P_{T,J}$

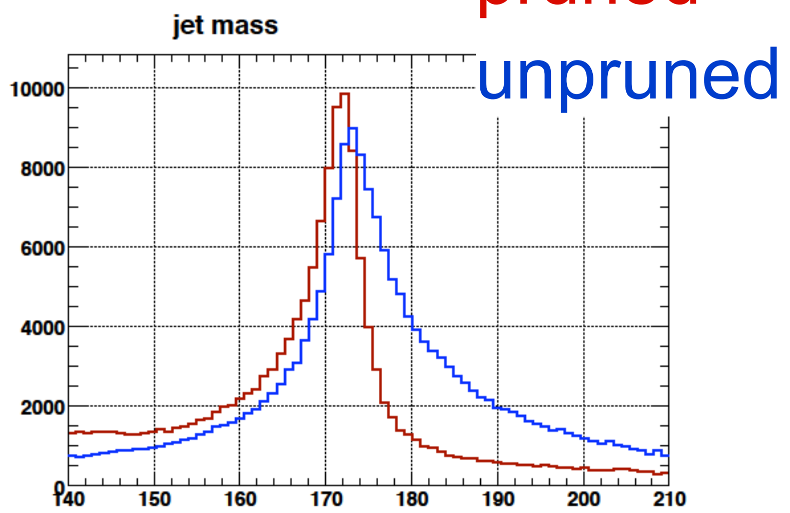
- If true (a soft, large angle recombination), prune the softer branch by NOT doing the recombination and discarding the softer branch
- Proceed with the algorithm

$\Rightarrow$ The resulting jet is the pruned jet

Defining Reconstructed Tops – Search Mode

- A jet reconstructing a top will have a mass within the top mass window, and a primary subjet mass within the W mass window - call these jets top jets
- Defining the top, W mass windows:
  - **Fit** the observed jet mass and subjet mass distributions with (asymmetric) Breit-Wigner plus continuum → widths of the peaks
  - The top and W windows are defined separately for pruned and not pruned - test whether pruning is narrowing the mass distribution





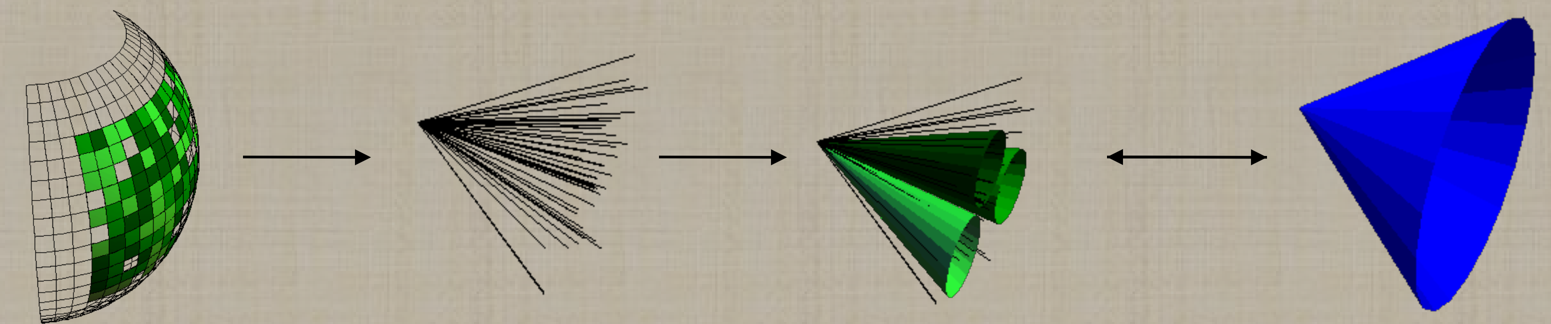
TOP TAGGING

Matthew Schwartz
Harvard University

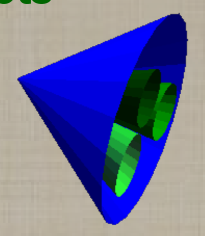
SUBJET DECOMPOSITION

1. Find fat jets first
   - We use geometric Cambridge-Aachen algorithm
   - Fat jet size R= 0.4-0.8

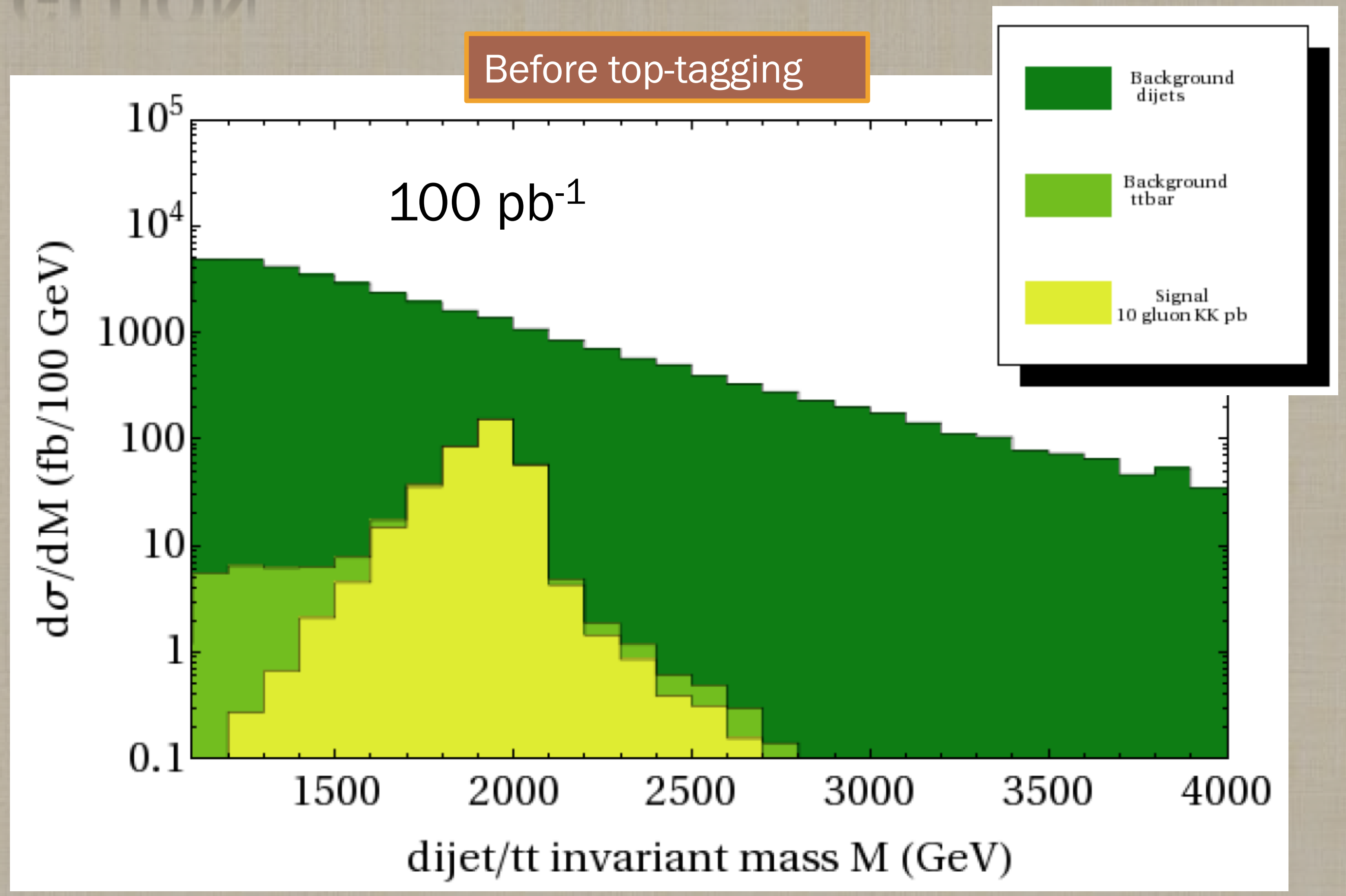
Fat jet size R=0.4-0.8
Varies with jet p_T



2. Reverse clustering steps
   - Clean out soft radiation $\frac{p_T^{(\text{particle})}}{p_T^{(\text{jet})}} > \delta_p \sim 0.05 - 0.1$
   - Clean out collinear radiation $|\Delta\eta| + |\Delta\phi| > \delta_r \sim 0.2$
   - Tops should have a 3 (or 4) subjets

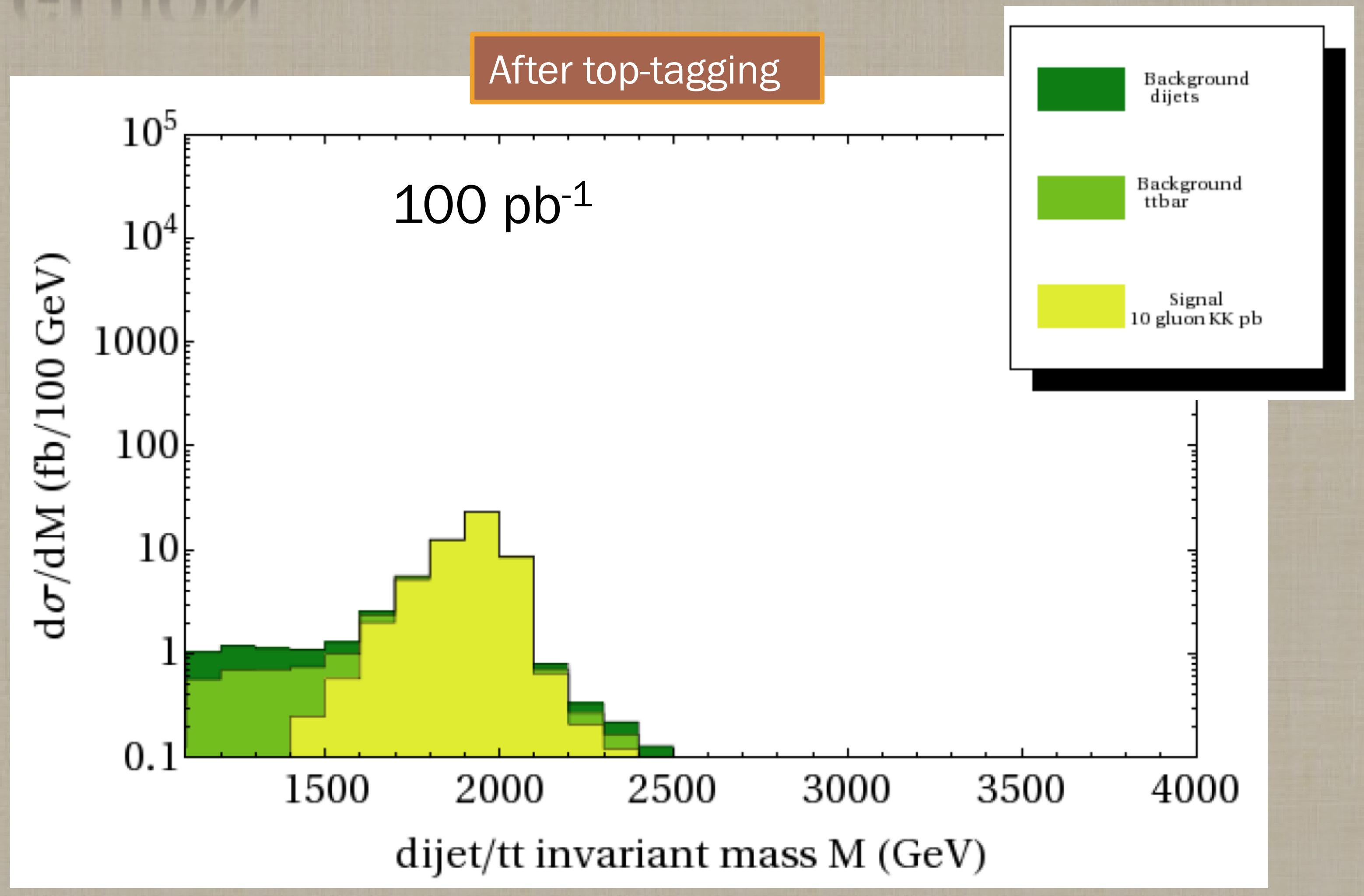


KK GLUON



σ x BR=10 pb

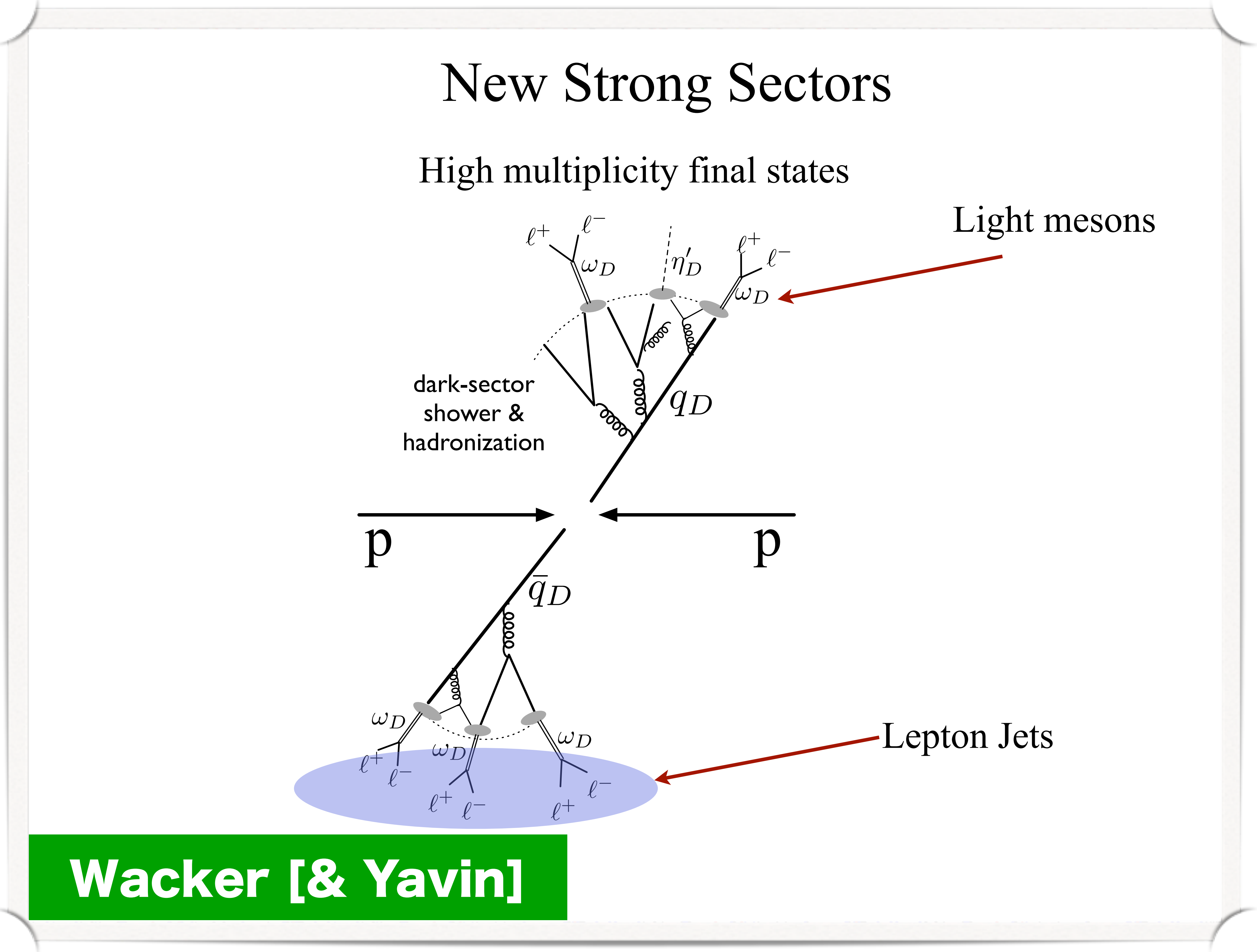
KK GLUON

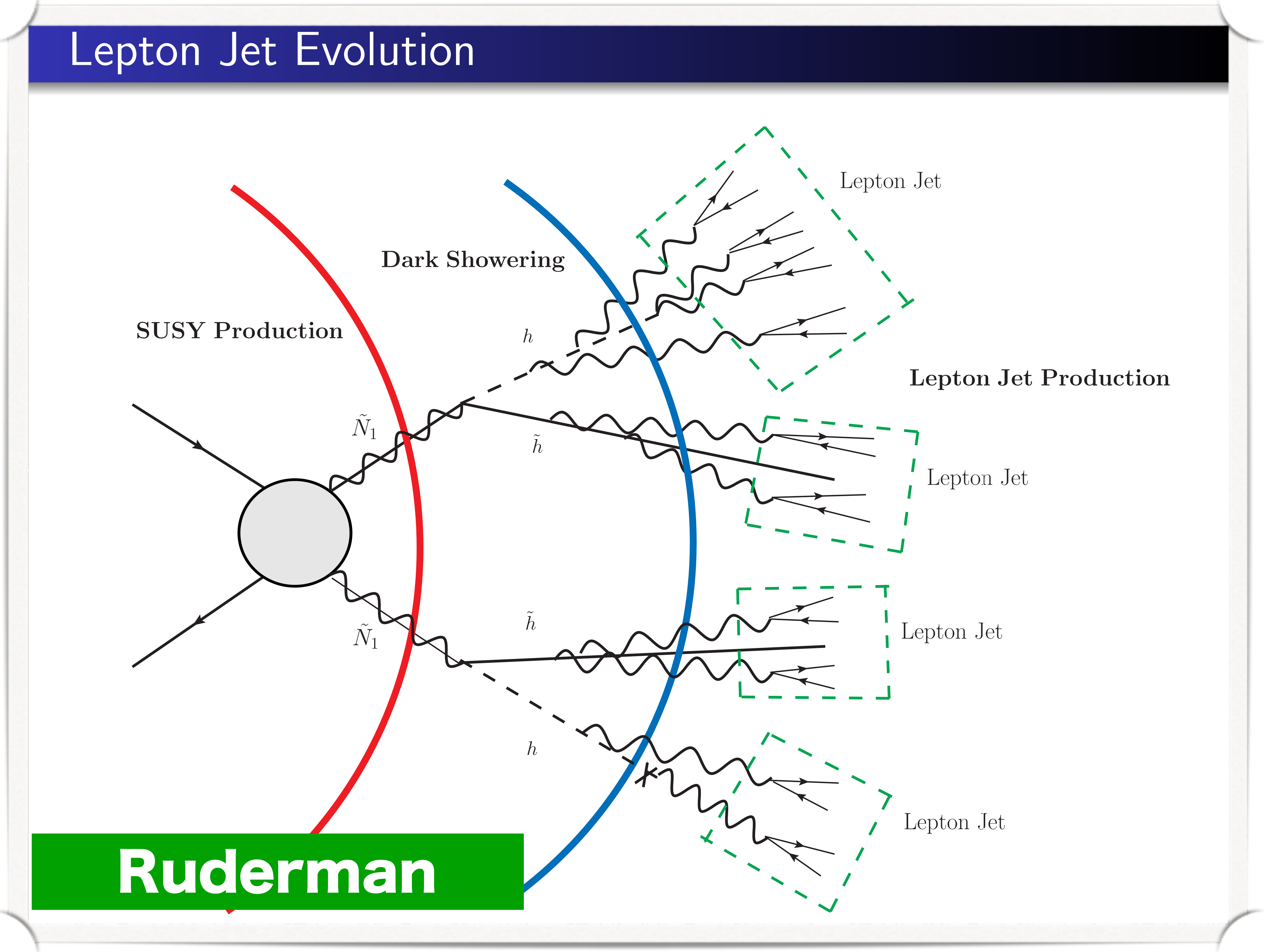


σ =10 pb

lepton jets

boost '09





a call for organising principles

Boost '10

Objectives

Aim (from opening)	Status (my view)
Bring together theorists and experimentalists to study physics behind boosted signatures, how to detect them and which tools to use	Accomplished ☺
Better understanding of what measurements need to be made this year and their efficient feedback to theorists	Tevatron measurements coming in! ☺ Towards LHC, a lot of good ideas exchanged, discussion of measurements/observables, but no concrete recipe for immediate follow-up made. Need a bit "more" and "publicly available" LHC data.
Common tools/standards to increase efficiency in comparing various algorithms in the market to better evaluate their effectiveness	Initiated and aimed to be part of write-up (Hadronic WG). ☺
Better understanding of how well experiments can utilize and calibrate such tools given their specific detectors	Work is progressing and getting there.

Muge Karagoz: closing

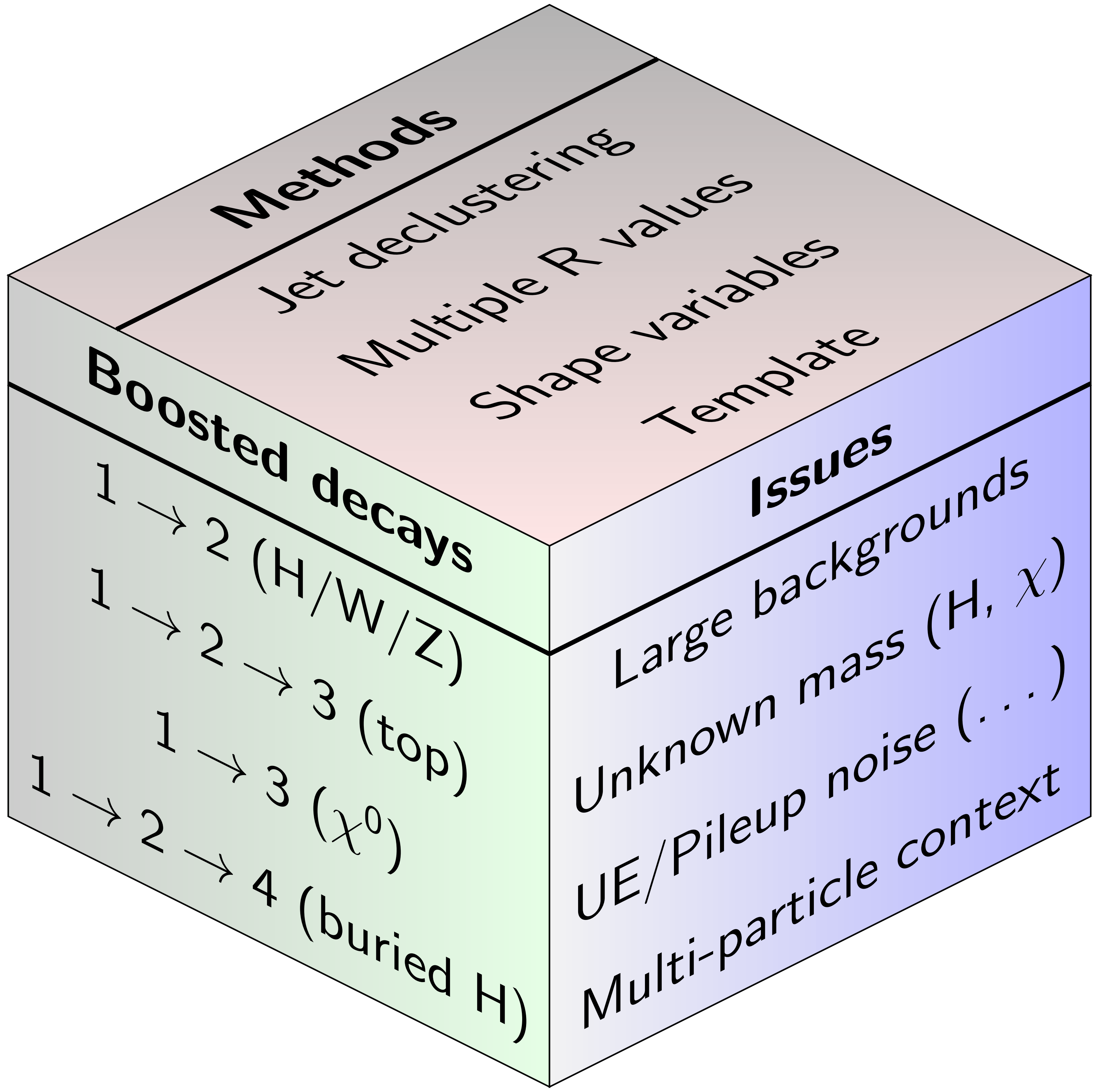
systematic exploration of BSM model space

Boost '10

Jay Wacker: intro

14:00	SUSY	Are Raklev	Dennis Sciama Lecture Theatre, University of Oxford	14:00 - 14:30
	Technicolor Models	Francesco Sannino	Dennis Sciama Lecture Theatre, University of Oxford	14:30 - 15:00
15:00	Extra particles and Extra Dimensions	Jose Santiago	Dennis Sciama Lecture Theatre, University of Oxford	15:00 - 15:30
	Tea Break		Dennis Sciama Lecture Theatre, University of Oxford	15:30 - 15:50
16:00	Black Holes	James Frost	Dennis Sciama Lecture Theatre, University of Oxford	15:50 - 16:20
	Higgs Sector	Mads Toudal Frandsen	Dennis Sciama Lecture Theatre, University of Oxford	16:20 - 16:50
17:00	Dark Sector	Jacob Wacker	Dennis Sciama Lecture Theatre, University of Oxford	16:50 - 17:20
14:00	Boosted Light Higgs from TeV-Scale Resonances: h->\tau\tau.	Dr Andrey Katz	Dennis Sciama Lecture Theatre, University of Oxford	14:00 - 14:15
	Unburied Higgs	David Krohn	Dennis Sciama Lecture Theatre, University of Oxford	14:15 - 14:30
	Discovering MSSM Higgs Bosons with Jet Substructure	Adam Martin	Dennis Sciama Lecture Theatre, University of Oxford	14:30 - 14:45
	Boosting Higgs Searches	Michael Spannowsky	Dennis Sciama Lecture Theatre, University of Oxford	14:45 - 15:00

the space of methods



Boosted objects will undoubtedly be part of the scene for LHC searches.

Anytime you do a search you should keep an eye on substructure

Open questions?

- Mostly, so far, developments have been based on a mixture of inspiration and trial+error. Can we give our methods a more quantitative foundation? Will this be of concrete benefit?

  E.g. flat backgrounds of χ^0 search in Butterworth et al. '09

- There's still wok to be done in comparing tools (quoted numbers not always comparable)

  Public code for all tools would help

Boost '10

Template Overlap Method

- Any region of partonic phase space for the boosted decays, {f}, defines a template
- Ansatz: good (if not best) rejection power using signal distribution for templates
- Define "template overlap" as the maximum functional overlap of j to a state f[j]:

$$Ov(j,f) = \max_{\{f\}} \mathcal{F}(j,f)$$

- Can match arbitrary final states *j* to partonic partners *f[j]* at any given order in PQCD.

Perez

27

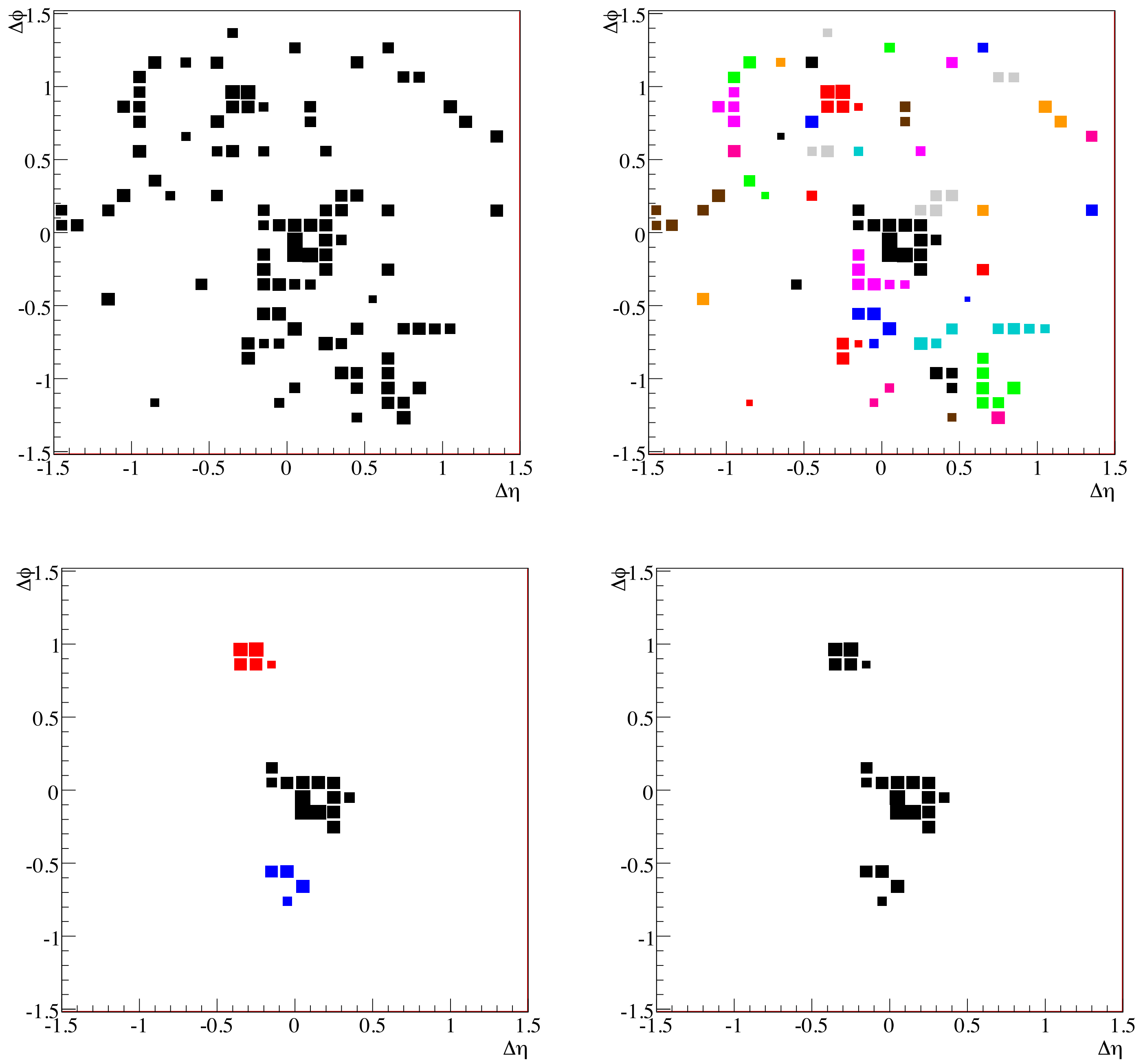
Trimming: not really presented…

not at Boost '10

arXiv:0912.1342

Jet Trimming

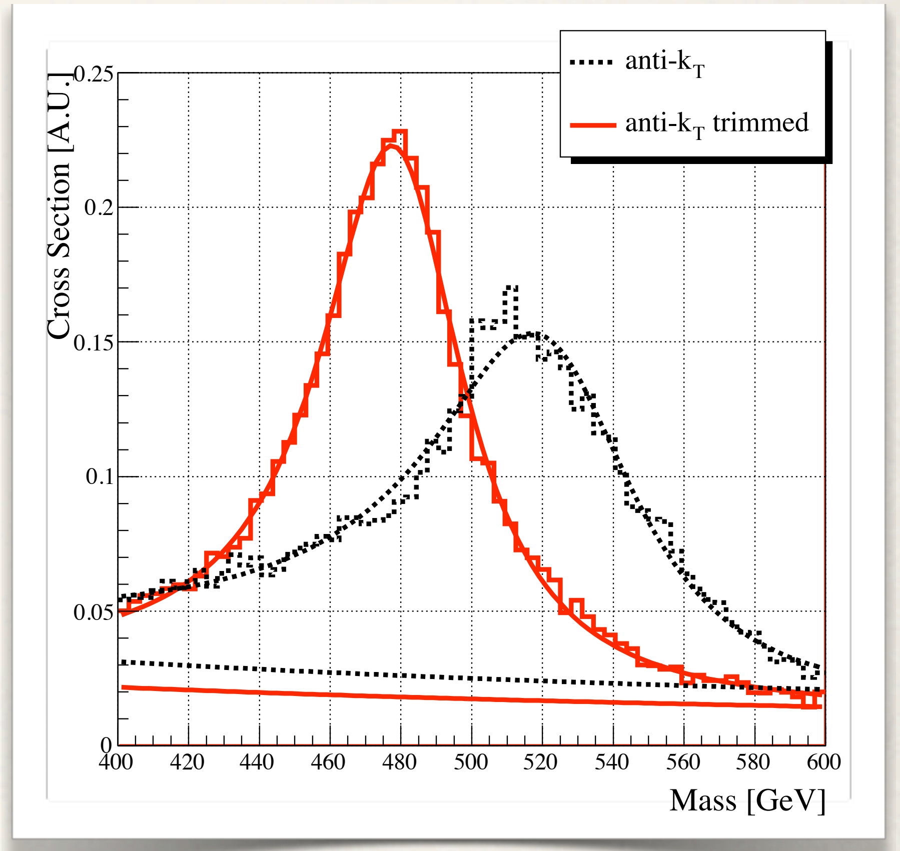
David Krohn[a], Jesse Thaler[b,c], and Lian-Tao Wang[a]



Krohn

Grooming Procedure

- To improve our mass resolution we apply jet trimming to our fat jets
  - Although reconstructing boosted heavy particles was not the original goal of *Jet Trimming*, we find it can be quite effective.
  - In limited testing can be competitive with filtering/pruning (see Soper and Spannowsky).



- *Jet Trimming*, DK, J. Thaler, L. Wang, [arXiv:0912.1342] JHEP 1002 (2010) 084
- *Combining subjet algorithms to enhance ZH detection at the LHC*, D. E. Soper, M. Spannowsky, [arXiv:1005.0417]

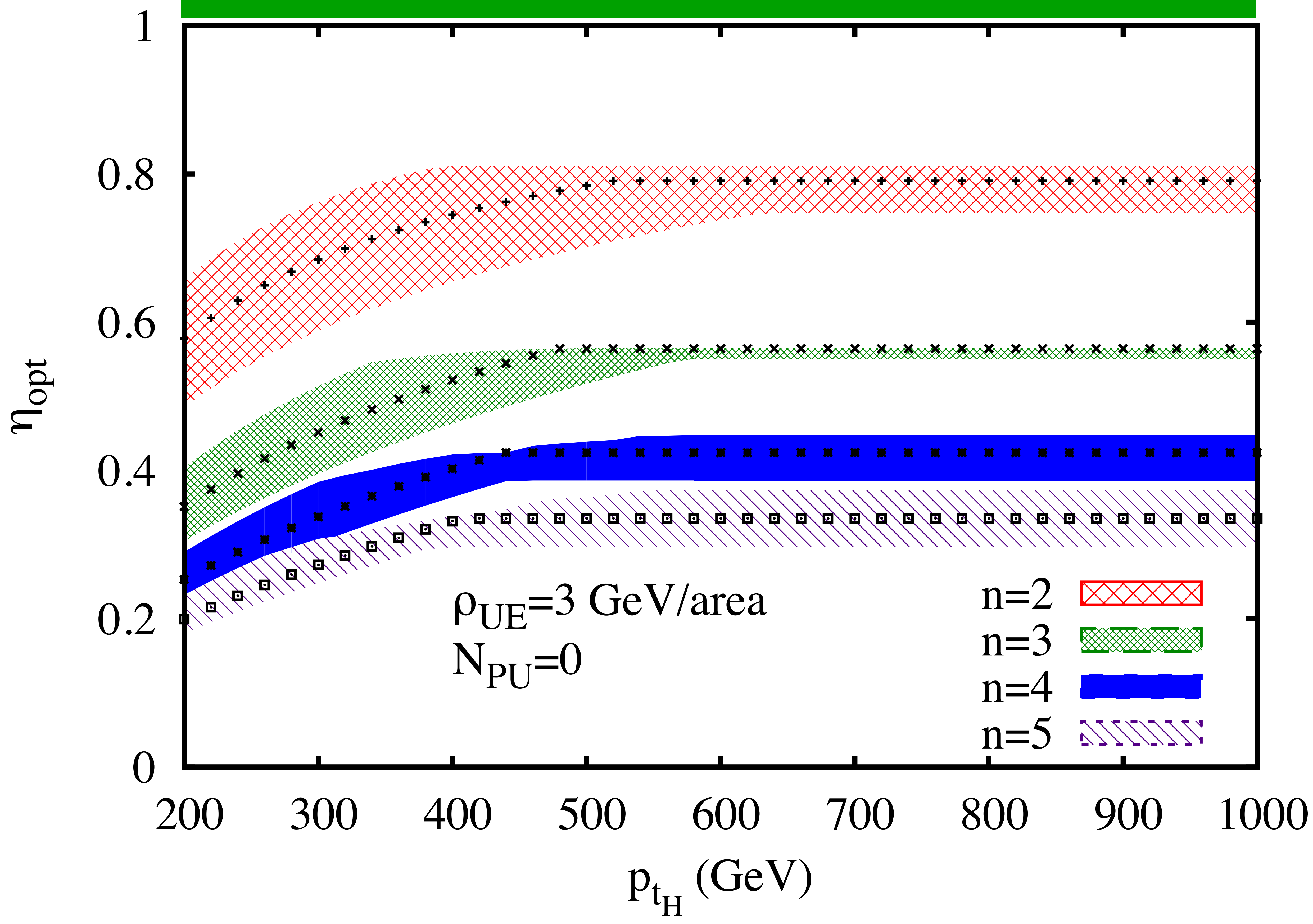
not at Boost '10

Non-Global Logarithms in Filtered Jet Algorithms

Mathieu Rubin
LPTHE
UPMC Univ. Paris 6
CNRS UMR 7589
Paris 05, France.

arXiv:1002.4557

optimal filtereting param. v. p_{tH}



Boost 2011

22-26 May 2011

Princeton

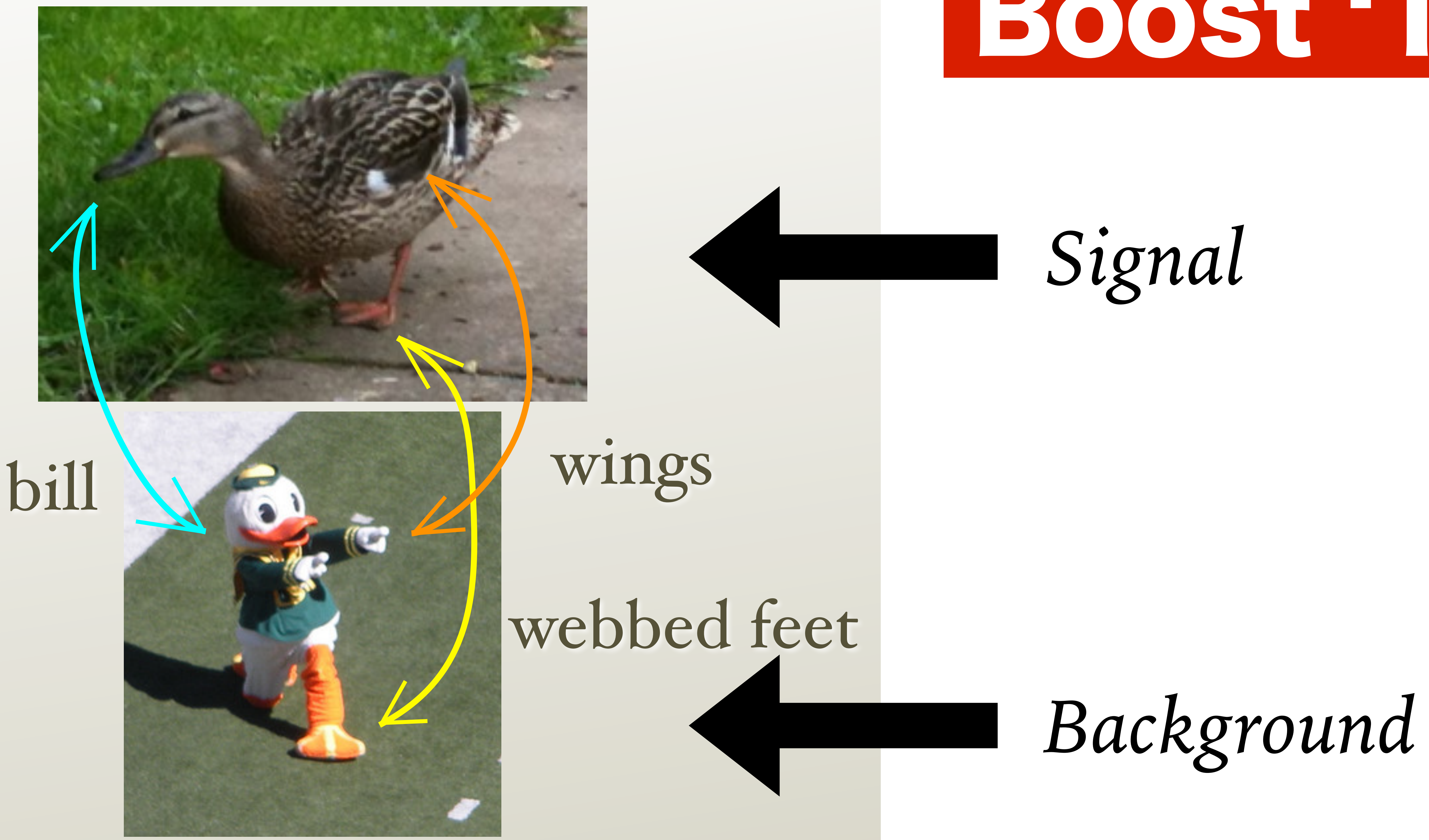
US/Eastern timezone

beyond N hard prongs

(& theorists being silly)

- There are many features in common between signal and background creatures.

- We need to find the differences. (Eg. one of these can’t swim.)



- Unfortunately, background creatures come with a range of mutations that make them sometimes look like a signal creatures.

- Thus we look need a statistical method to tell if there are signal creatures.

"Mirror, mirror on the wall ..."

Do I look like a Higgs jet or do I look like a gluon jet?

Don't answer:

To me you just look like a fat jet

The first observable to efficiently probe radiation flow

Boost '11

Introducing N-subjettiness

"There ... seems to be a rule in physics that the longer you let theorists play with an idea, the more likely it is that they'll give it a silly name."

— Flip Tanedo, USLHC Blog, April 22, 2011

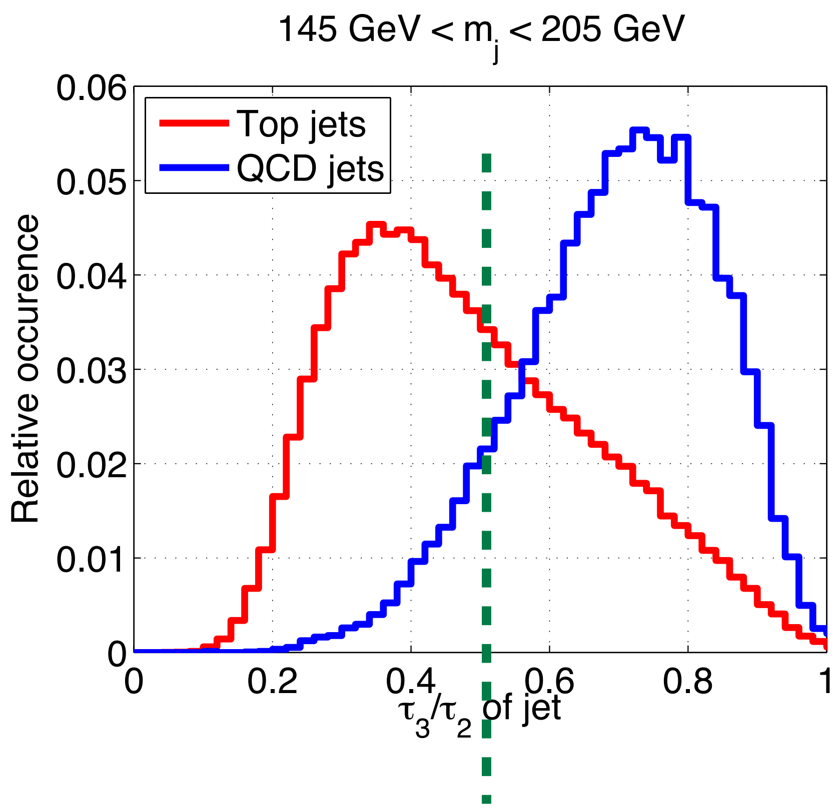
N-subjettiness: Degree to which a jet has N subjets!

$$\tau_N \simeq 0 \quad \Rightarrow \quad \leq N \text{ subjets}$$
$$\tau_N \simeq 1 \quad \Rightarrow \quad > N \text{ subjets}$$

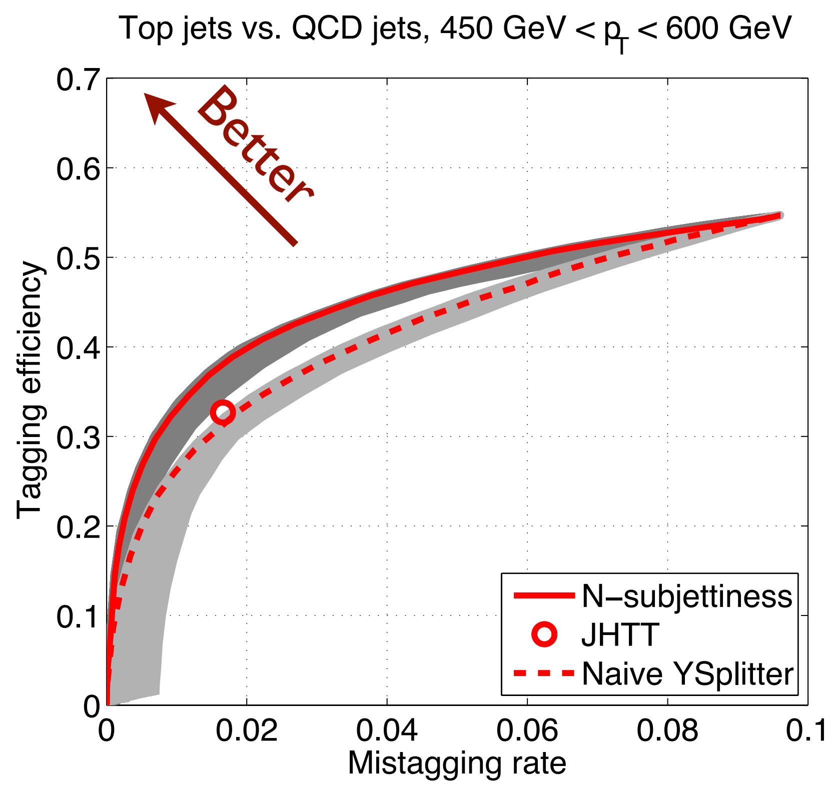
(You prefer "Voronoi-Tessellated Angularities"?)

Adapted from "N-jettiness" (See Iain's talk)
[Stewart, Tackmann, Waalewijn: 1004.2489]

N-subjettiness for Boosted Tops



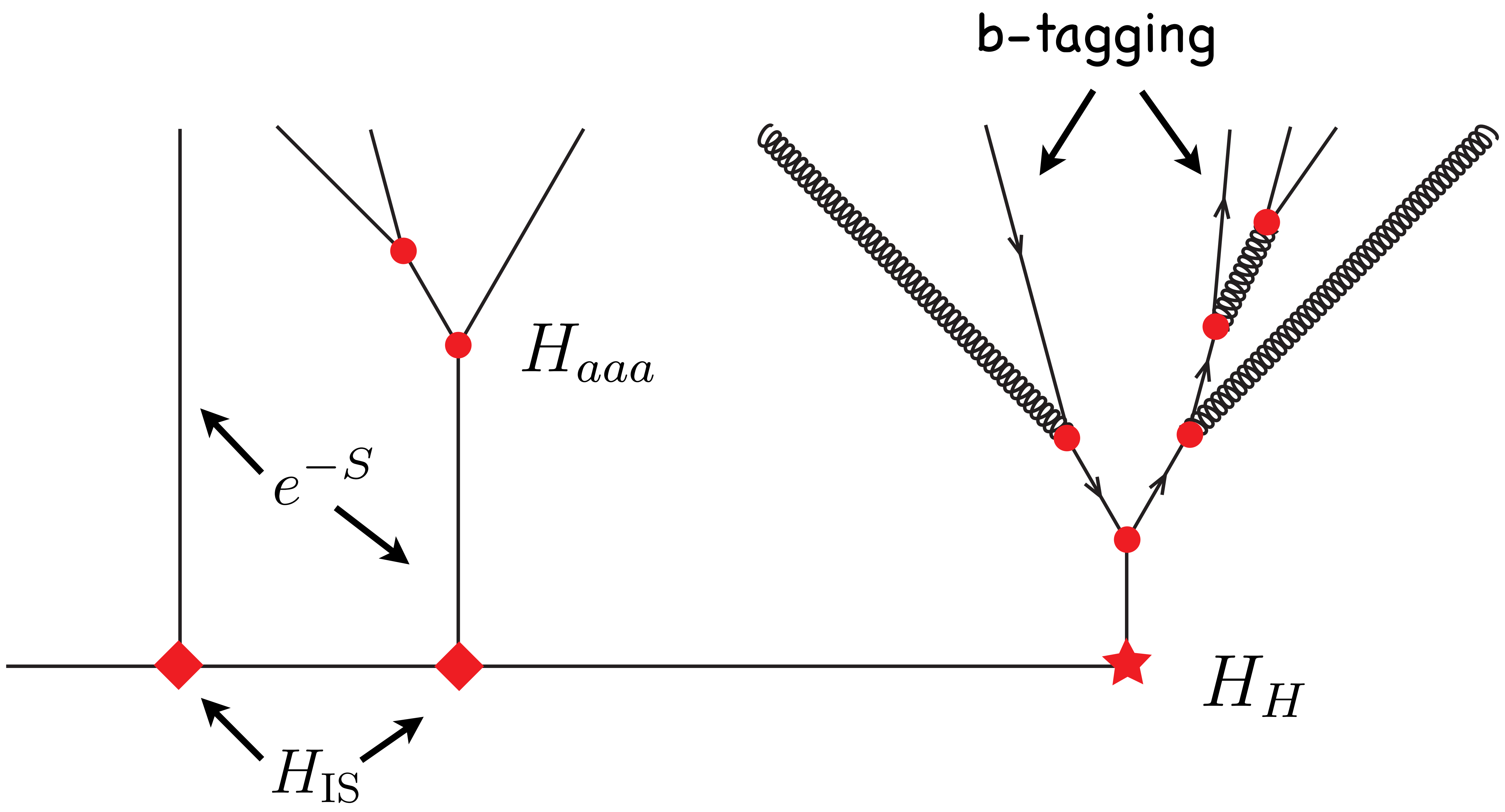
Flexible cut to adjust signal acceptance vs. background rejection



Compares favorably to existing methods!

τ_3/τ_2: Boosted Tops
τ_2/τ_1: Boosted W/Z/H

(But "Theorist Detector". Our study before the Boost2010 Proceedings)



Want to convert the shower history into analytic expression

➡ Feynman Rules for Shower Deconstruction

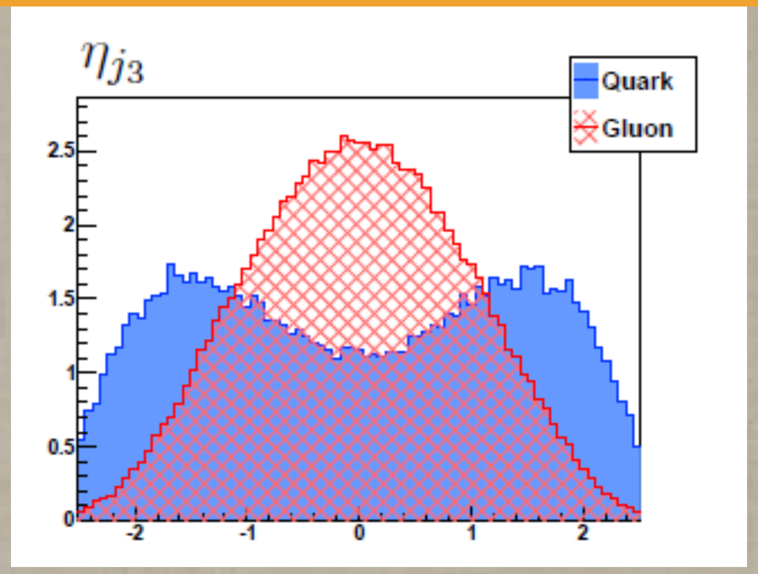
$$\chi(\{p,t\}_N) = \frac{P(\{p,t\}_N|\mathrm{S})}{P(\{p,t\}_N|\mathrm{B})} = \frac{\sum_{\rm histories} H_{ISR} e^{-S_{I1}} \dots \sum_{\rm histories} H_H e^{-S_{s1}} H^s_{bg} e^{-S_{s2}} \dots}{\sum_{\rm histories} H_{ISR} e^{-S_{I1}} \dots \sum_{\rm histories} H_{gbb} e^{-S_{b1}} H^b_{bg} e^{-S_{b2}} \dots}$$

χ is the one analytic function to discriminate signal from background

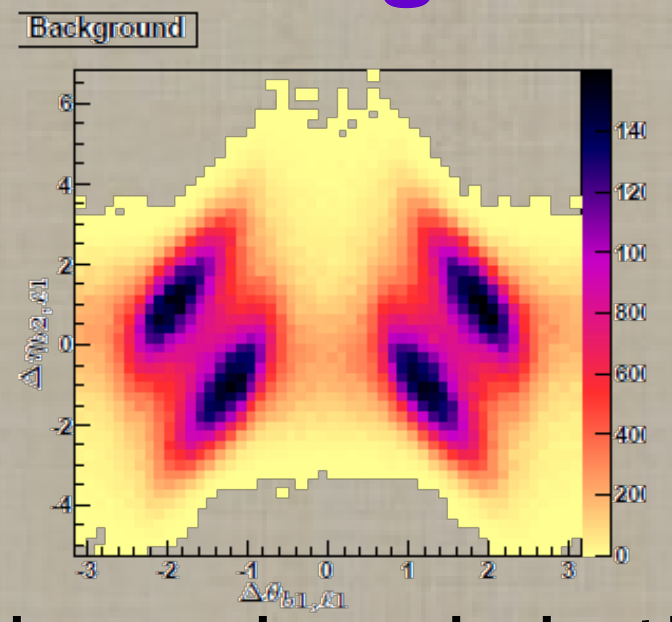
(for more detail see 1102.3480)

shower deconstruction

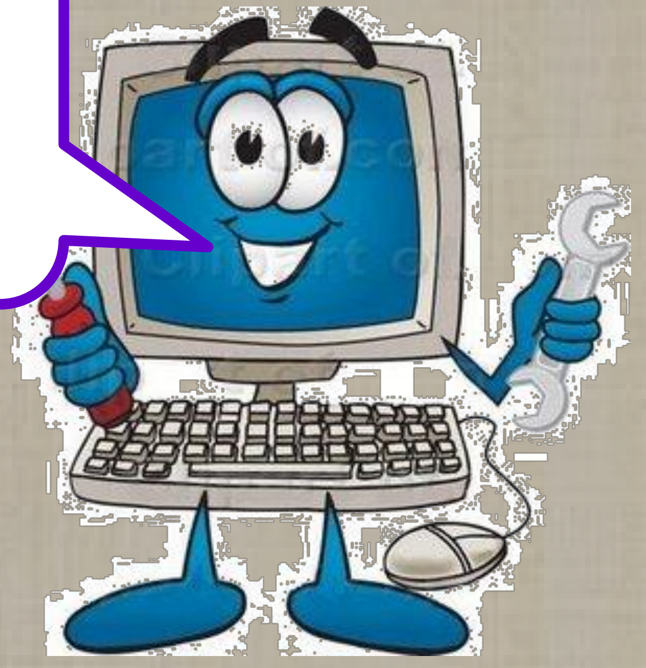
WHY USE A MULTIVARIATE APPROACH?



- We can think about and visualize single variables



- Two variables are harder
- Nobody who thinks in 11 dimensions is in this room!
- There things that computers are just better at.
- Multivariate approach lets you figure out how well you could possibly do



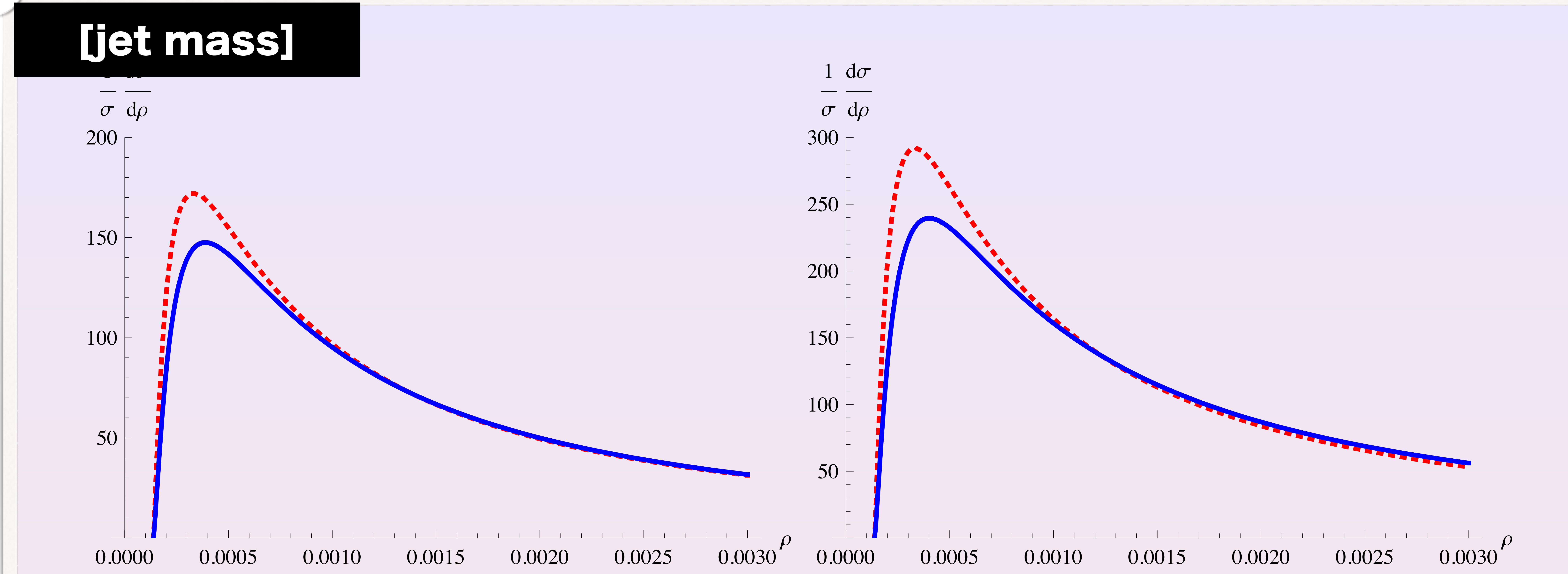
EFFICIENCY
Save you the trouble or looking for good variables (project killer)

FRAMING
See if simple variables can do as well (establishes the goal)

POWER
Sometimes they are really necessary (e.g. ZH)

calculations for "simple observations" get going

Boost '11



Plots are for $p_t = 250$ GeV and $E_0 = 15$ GeV and $E_0 = 60$ GeV

Banfi, MD ,Khelifa Kerfa, Marzani, 2010

Dasgupta

calculations for “simple observations” get going

[jet mass]

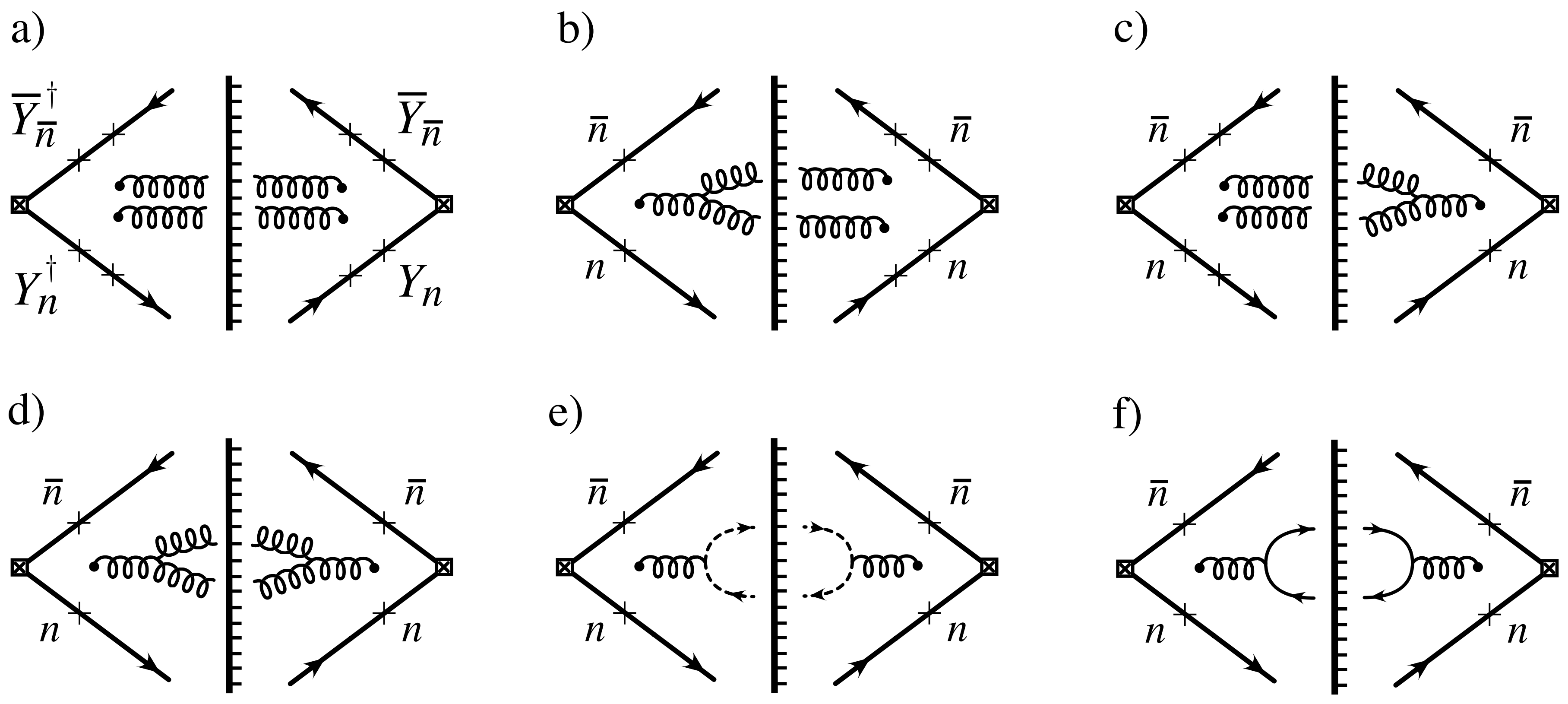
Unfortunately we did not quite get final results in time for the workshop ...

Jouttenus, IS, Tackmann, Waalewijn

Stewart

Wednesday, May 25, 2011 41

Calculating the Dijet Soft Function at $\mathcal{O}(\alpha_s^2)$



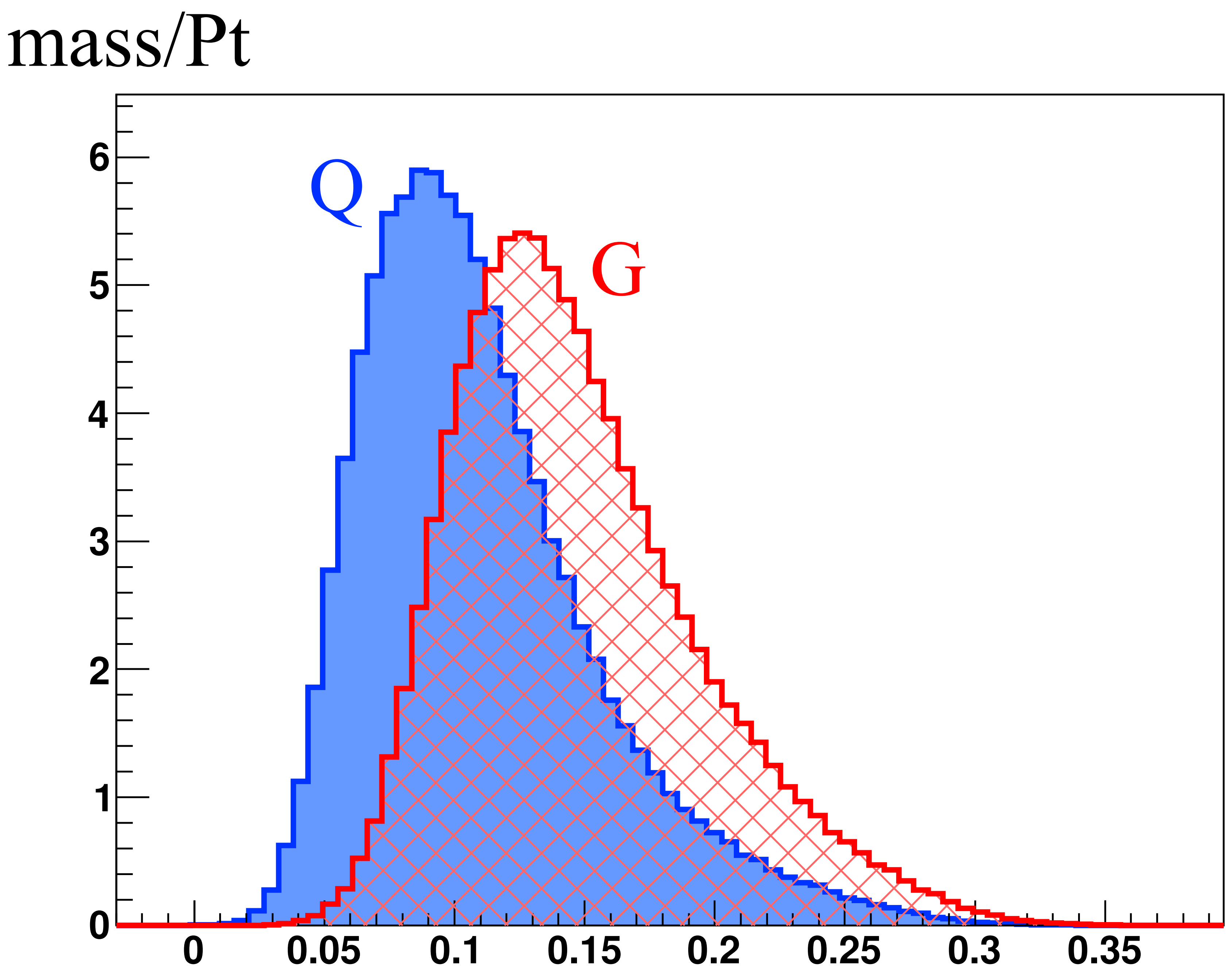
$$C_F^2 : \quad \text{a}$$
$$C_F C_A : \quad \text{a,b,c,d,e}$$
$$C_F T_R n_f : \quad \text{f}$$

Lee

Boost '11

Jet Mass as an Example Observable

- Normalizing by p_T (200 GeV in this sample) generalizes better.



Gallicchio

HEPTopTagger

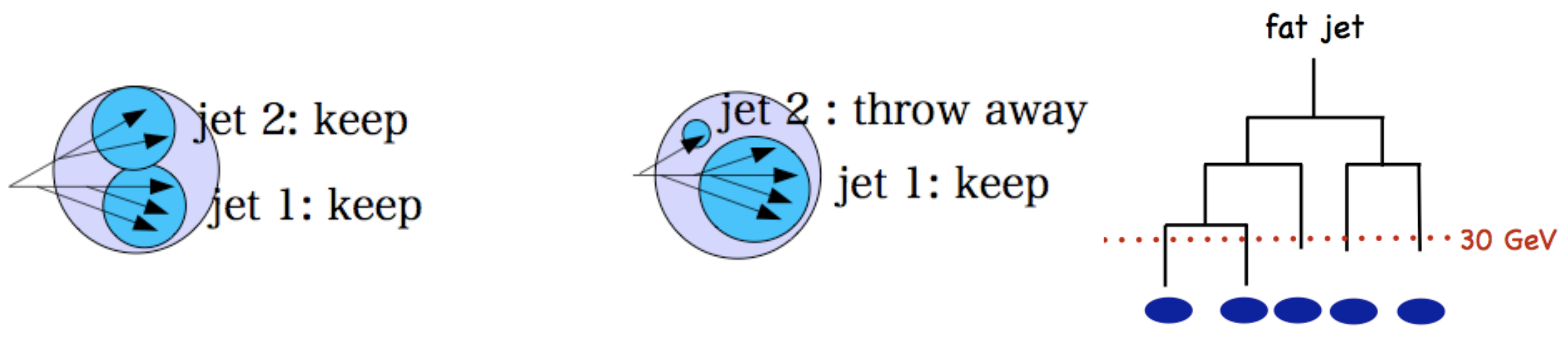
Boost '11

Stop reconstruction with the HEPTopTagger

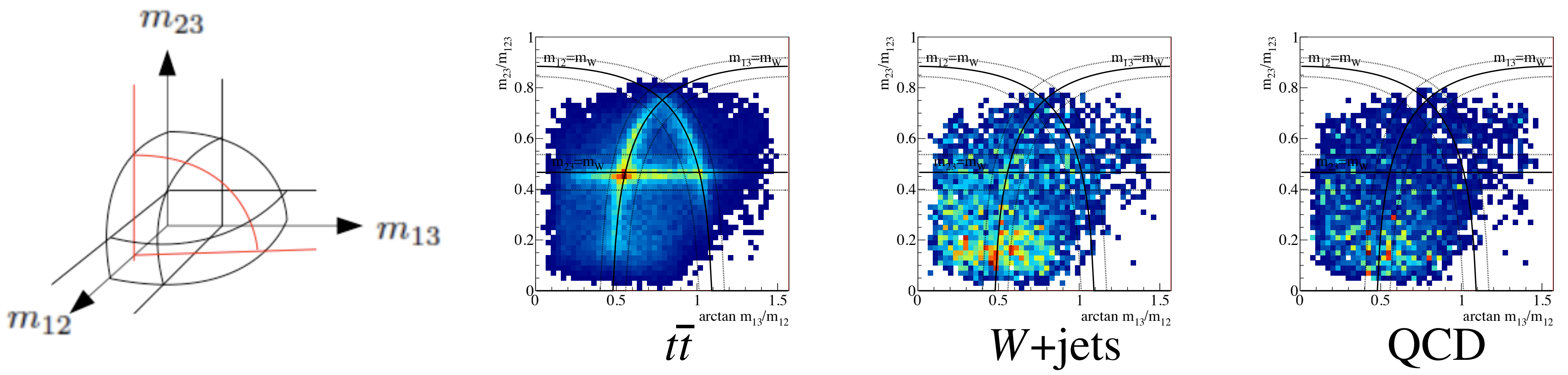
Michihisa Takeuchi (Uni Heidelberg)

HEPTopTagger [JHEP 1010:078,2010. arXiv:1006.2833 [hep-ph] T. Plehn, M. Spannowsky, D. Zerwas, MT]

1. **fat jets** – C/A($R = 1.5$), $p_T^{\text{fatjet}} > 200$ GeV
2. **mass drop criterion**
   – find hard proto-jets $m_j < 30$ GeV, $m_{j1} < 0.8 m_j$ to keep j_1 and j_2

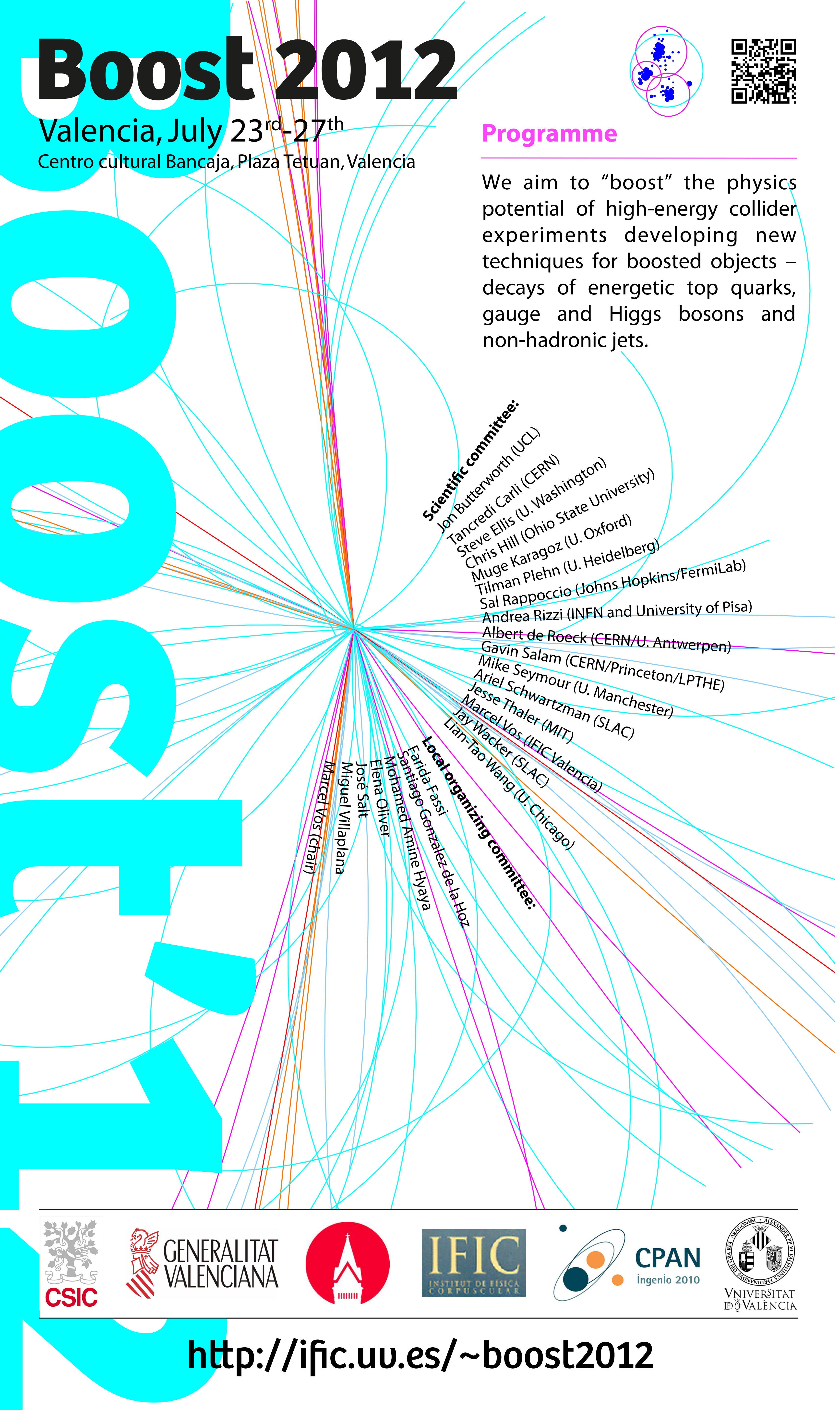


3. **choose 3 hard proto-jets with best filtered mass**
   – $|m_{jjj}^{\text{filt}} - m_t| < 25$ GeV and $p_T^{\text{rec}} > 200$ GeV $\rightarrow$ top candidate
4. **check mass ratios**
   – m_t condition: $m_t^2 = m_{123}^2 = m_{12}^2 + m_{13}^2 + m_{23}^2 \rightarrow$ spherical surface: 2D mass ratios



   – W mass condition, soft-collinear cut $\rightarrow$ tagged top

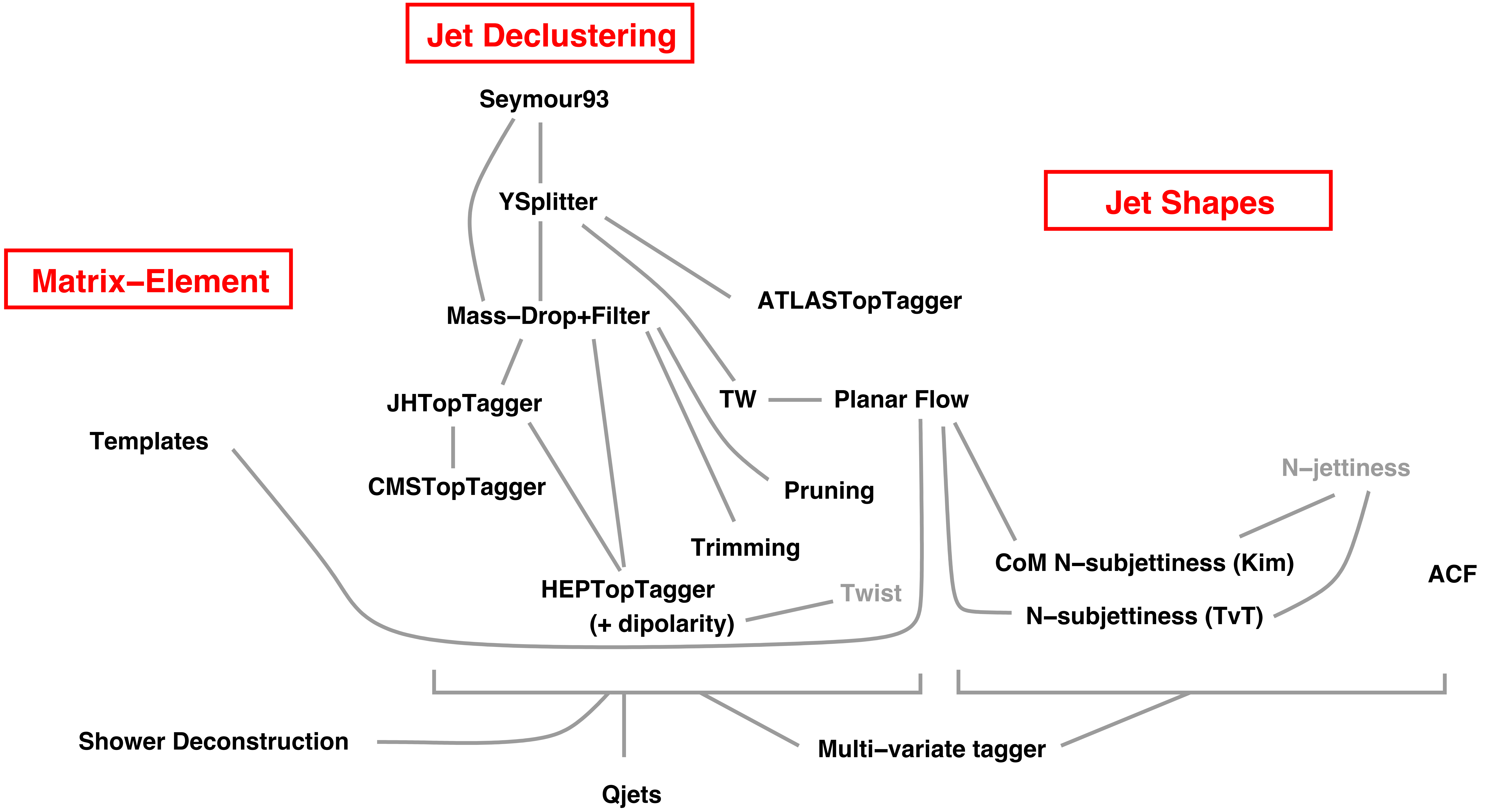
Takeuchi



many tools;
calculations from several groups

Some taggers and jet-substructure observables

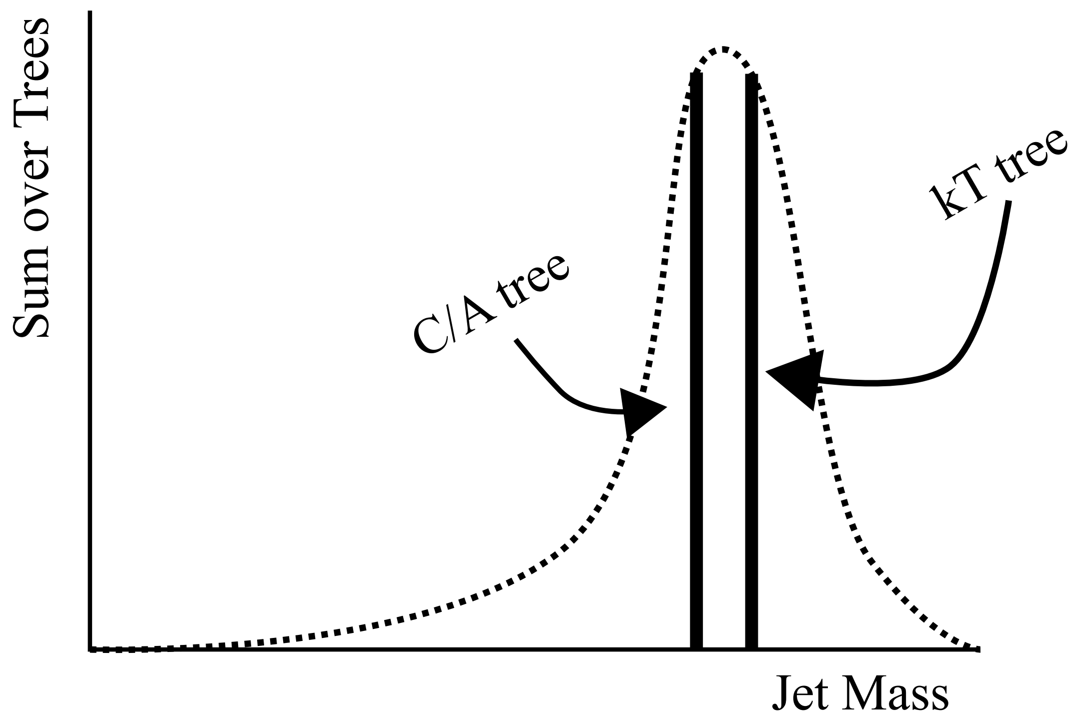
Boost '12



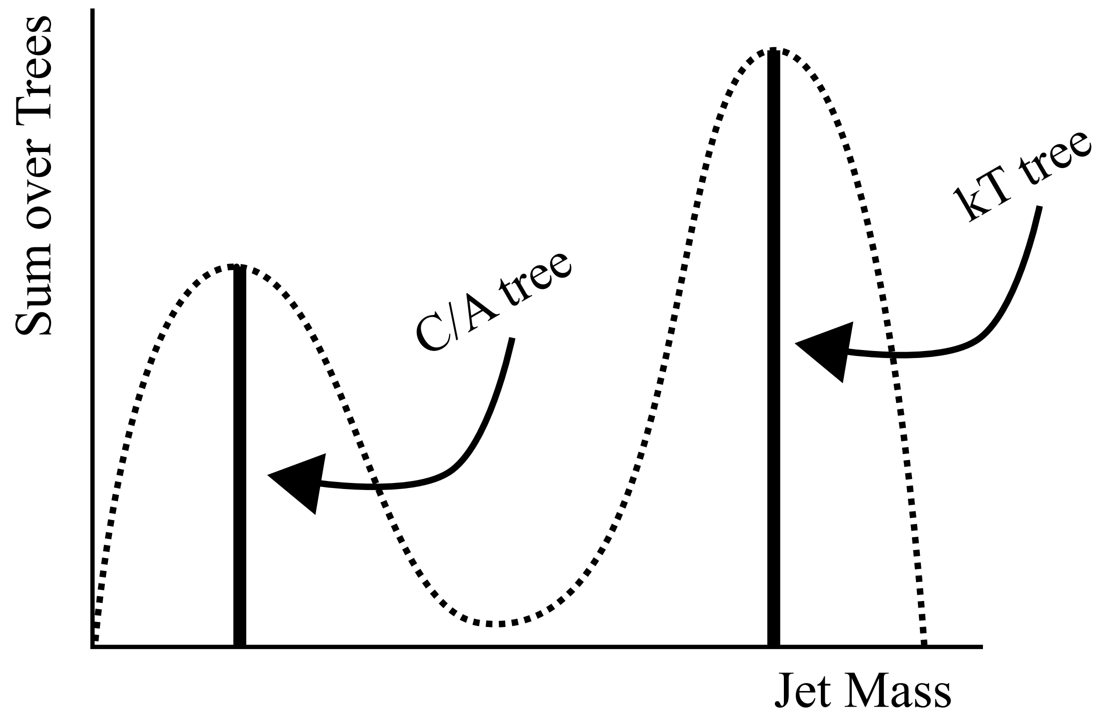
apologies for omitted taggers, arguable links, etc.

Basics of Qjets

- substructure assumes a shower creates trees, and best tree is good enough
- Qjets: take all (or many trees)
- example: apply pruning to the various recombinations allowed within a single jet



VS



Jet shape subtraction

No pileup: $f_{\rm hard}({\rm jet}) = f(\{p_{t,i}\}_{\rm hard})$

With pileup: $f_{\rm full}({\rm jet}) = f(\{p_{t,i}\}_{\rm hard}, \{p_{t,i}\}_{\rm PU})$

- Background has 2 degrees of freedom: ρ and ρ_m
- Assume $\rho \ll p_t$ and expand in series of ρ and ρ_m
- PU$\propto$ghosts $\Rightarrow \partial_\rho \propto \partial_{\rm ghostscale}$

Subtraction:

$$\begin{aligned} f_{\rm sub}({\rm jet}) &= f_{\rm full}({\rm jet}) - \rho\, a_{\rm ghost}\, \partial_{\rm ghostscale} f_{\rm full}({\rm jet}) \\ &+ \frac{1}{2}(\rho\, a_{\rm ghost})^2\, \partial^2_{\rm ghostscale} f_{\rm full}({\rm jet}) + \dots \end{aligned}$$

with additional term for the ρ_m contributions

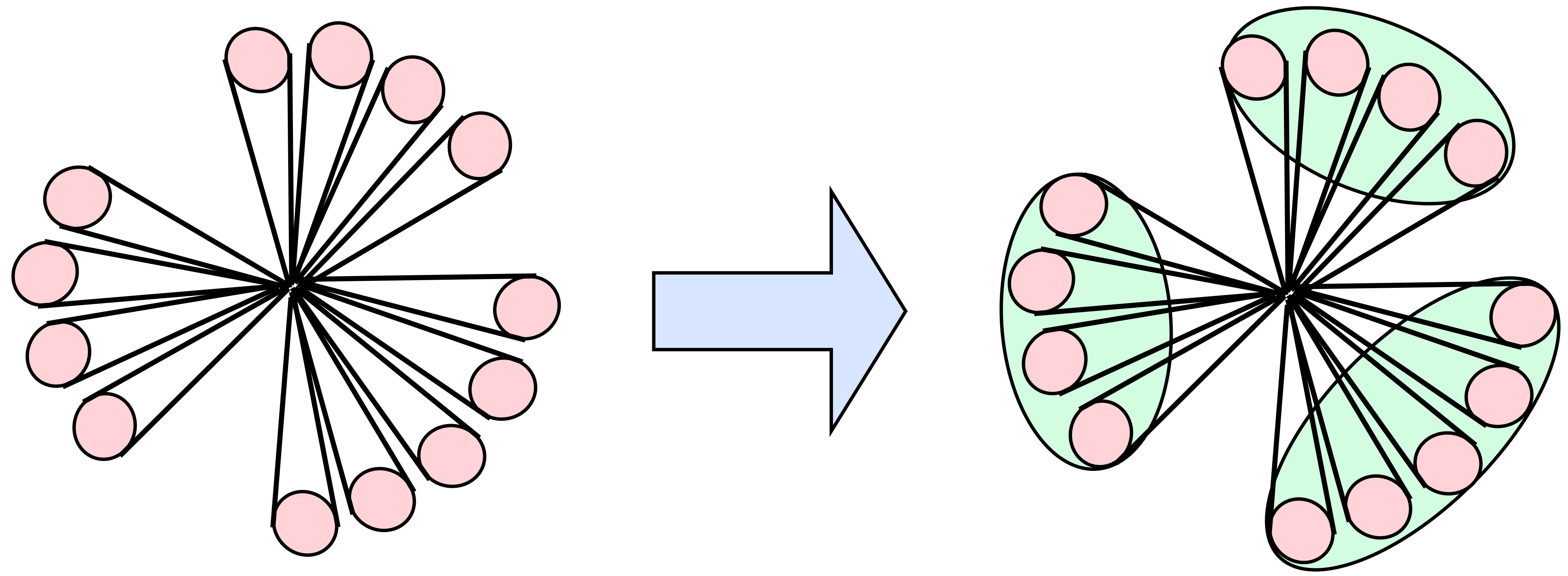
Soyez

Boost '12

Seem to have lost
the single feature that made
these events special

13 Jet Event

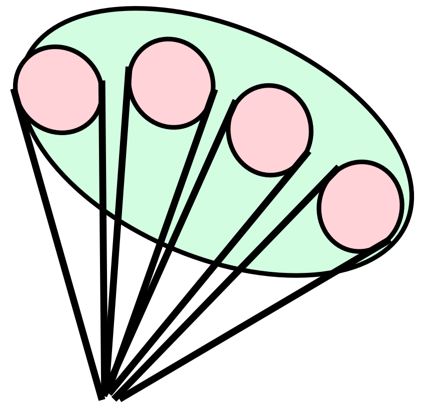
3 Jet Event

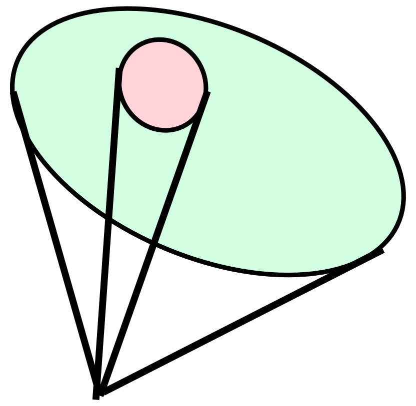


The difference between them is clear

Large Jet Mass

Small Jet Mass





$$\frac{m_j}{p_T} \sim 1$$

$$\frac{m_j}{p_T} \sim 0.3$$

Each jet mass is *approximately* independent for QCD

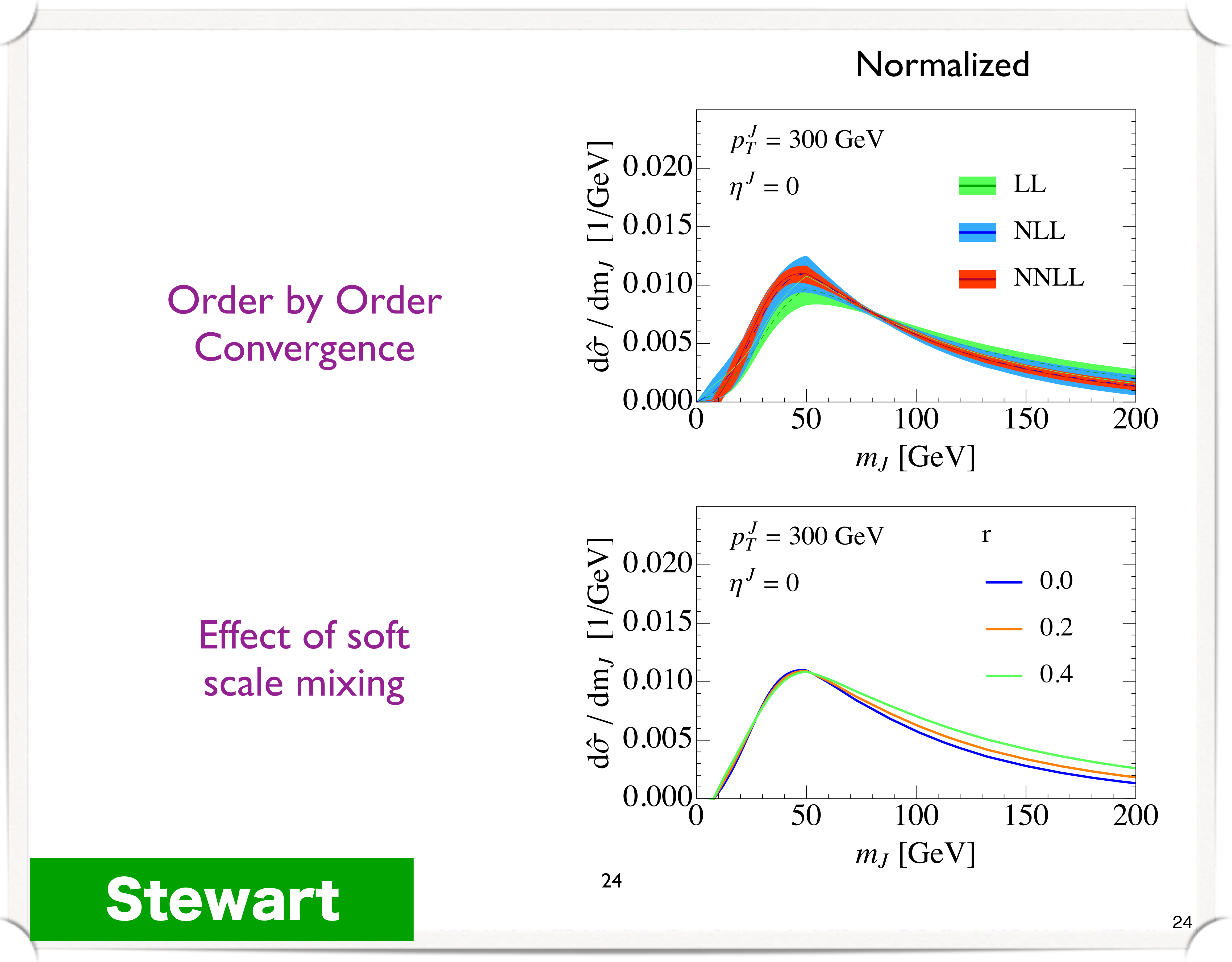
Getting multiple massive jets rare

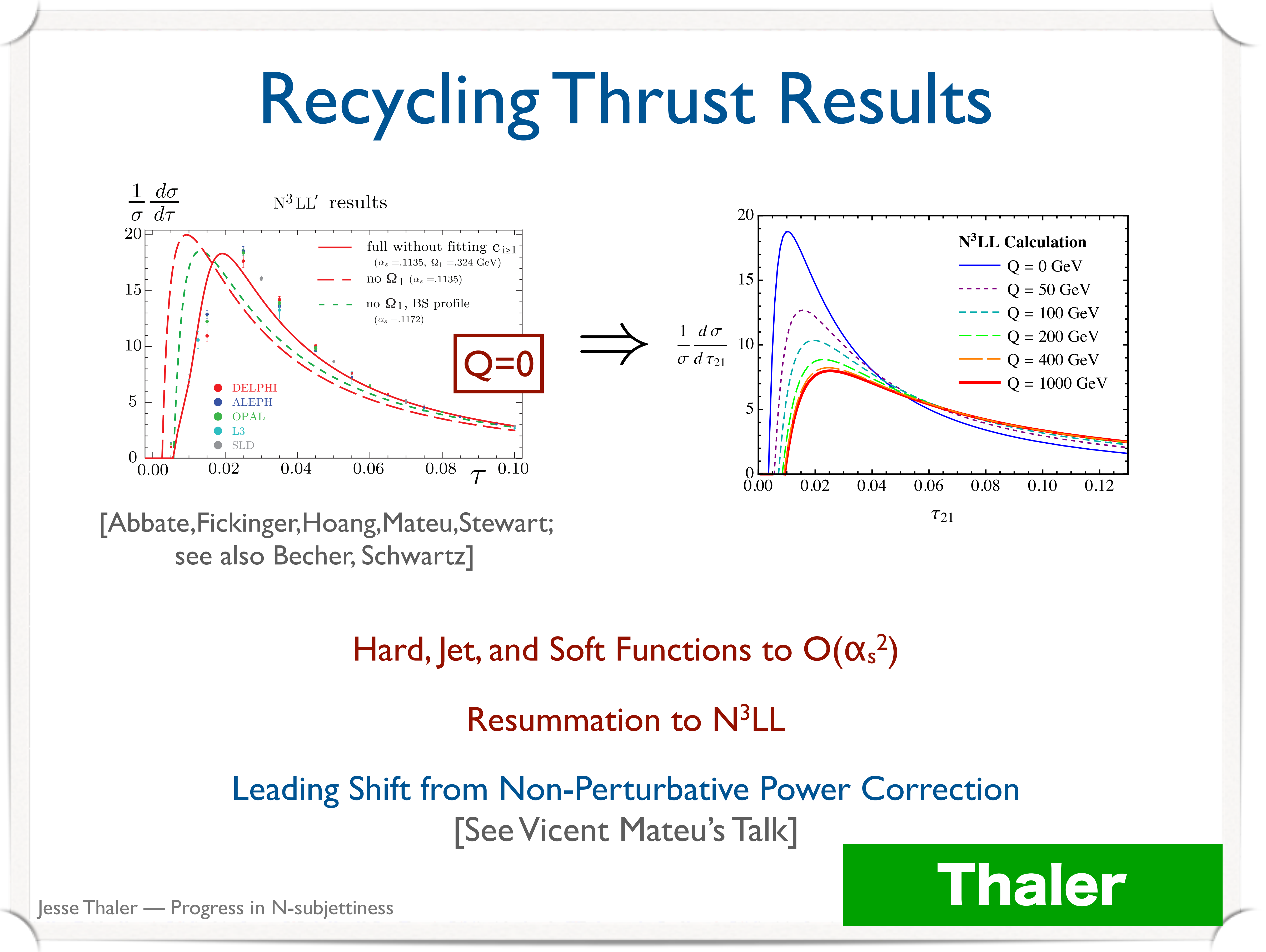
Jet mass correlations never studied bef

Wacker

N-subjettiness and jet mass results from SCET

Boost '12





+ a heated debate about non-global logarithms

but can you calculate things for groomed jets?

Boost '12

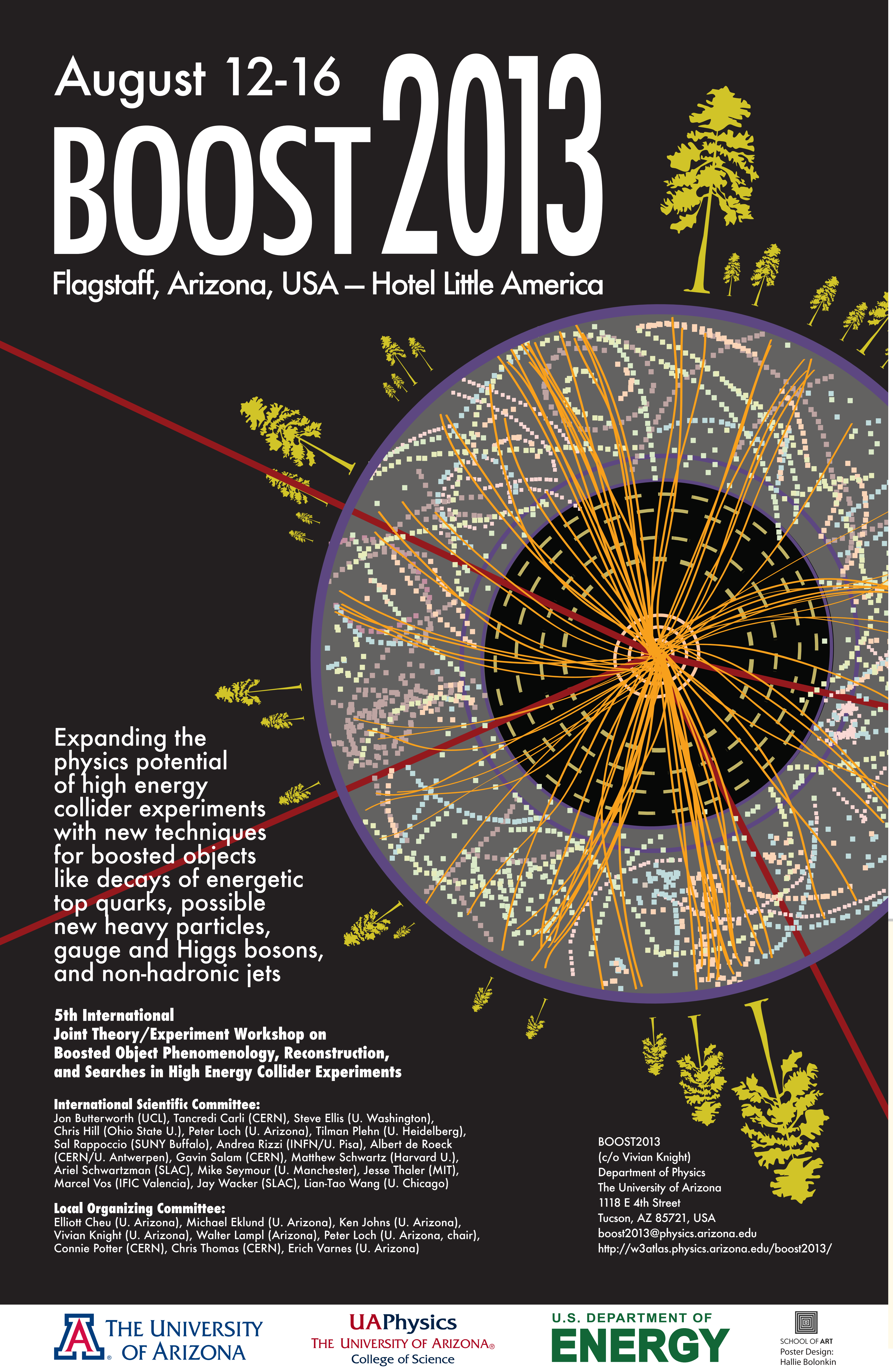
- Grooming is great for removing pileup/UE. No doubt.
  - Bump hunting
  - Jet substructure
  - New physics searches
  - **not** QCD

Alternatives
- Observable specific subtraction
  - See Thaler's talk on n-subjettiness
- Parameterize, fit shapes
- Pileup-insensitive observables

Schwartz

Major changes in FastJet 3

Soyez



amazing posters!
(but not only)

Boost '13

Energy Correlation Functions

$$\mathrm{ECF}(N,\beta) = \sum_{i_1<i_2<\ldots<i_N\in J} \left(\prod_{a=1}^{N} p_{T\,i_a}\right)\left(\prod_{b=1}^{N-1}\prod_{c=b+1}^{N} R_{i_b i_c}\right)^{\beta}$$

$$\mathrm{ECF}(0,\beta) = 1,$$

$$\mathrm{ECF}(1,\beta) = \sum_{i\in J} p_{T\,i},$$

$$\mathrm{ECF}(2,\beta) = \sum_{i<j\in J} p_{T\,i}\,p_{T\,j}\,(R_{ij})^{\beta},$$

Banfi, Salam, Zanderighi 2004
Jankowiak, AL 2011

$$\mathrm{ECF}(3,\beta) = \sum_{i<j<k\in J} p_{T\,i}\,p_{T\,j}\,p_{T\,k}\,(R_{ij}R_{ik}R_{jk})^{\beta},$$

$$\mathrm{ECF}(4,\beta) = \sum_{i<j<k<\ell\in J} p_{T\,i}\,p_{T\,j}\,p_{T\,k}\,p_{T\,\ell}\,(R_{ij}R_{ik}R_{i\ell}R_{jk}R_{j\ell}R_{k\ell})^{\beta}$$

3

Larkoski

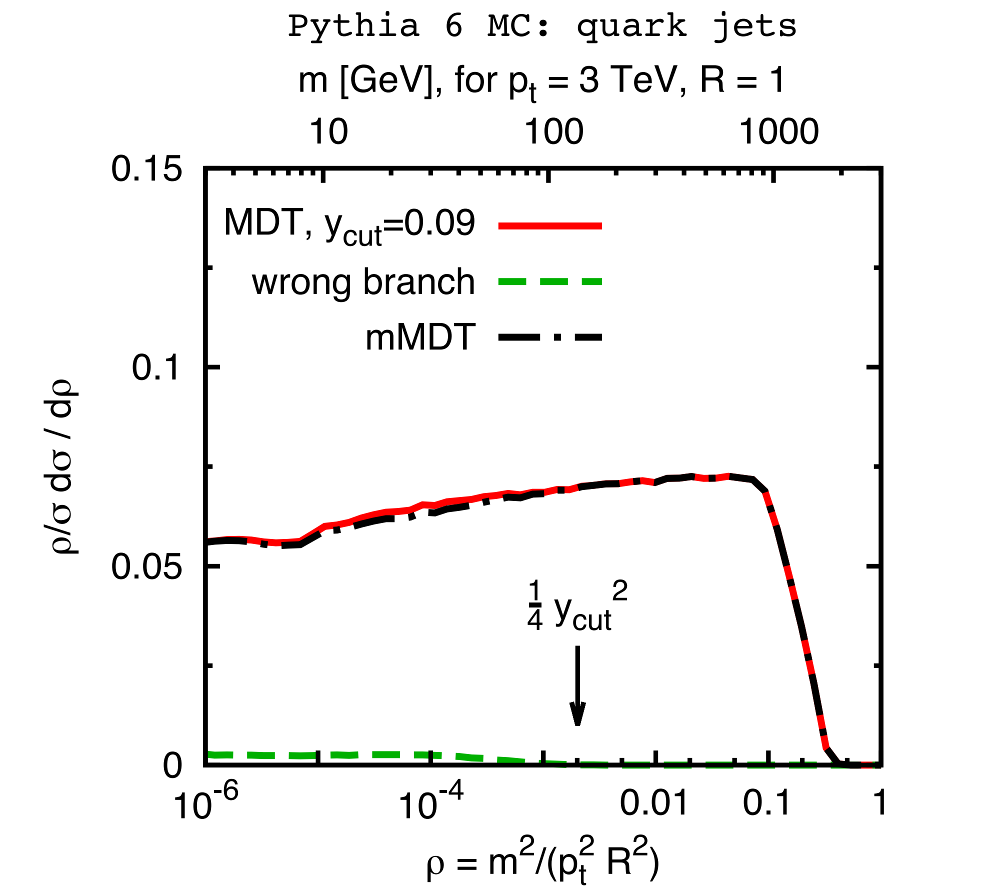
first analytical grooming calculations → mMDT

Boost '13

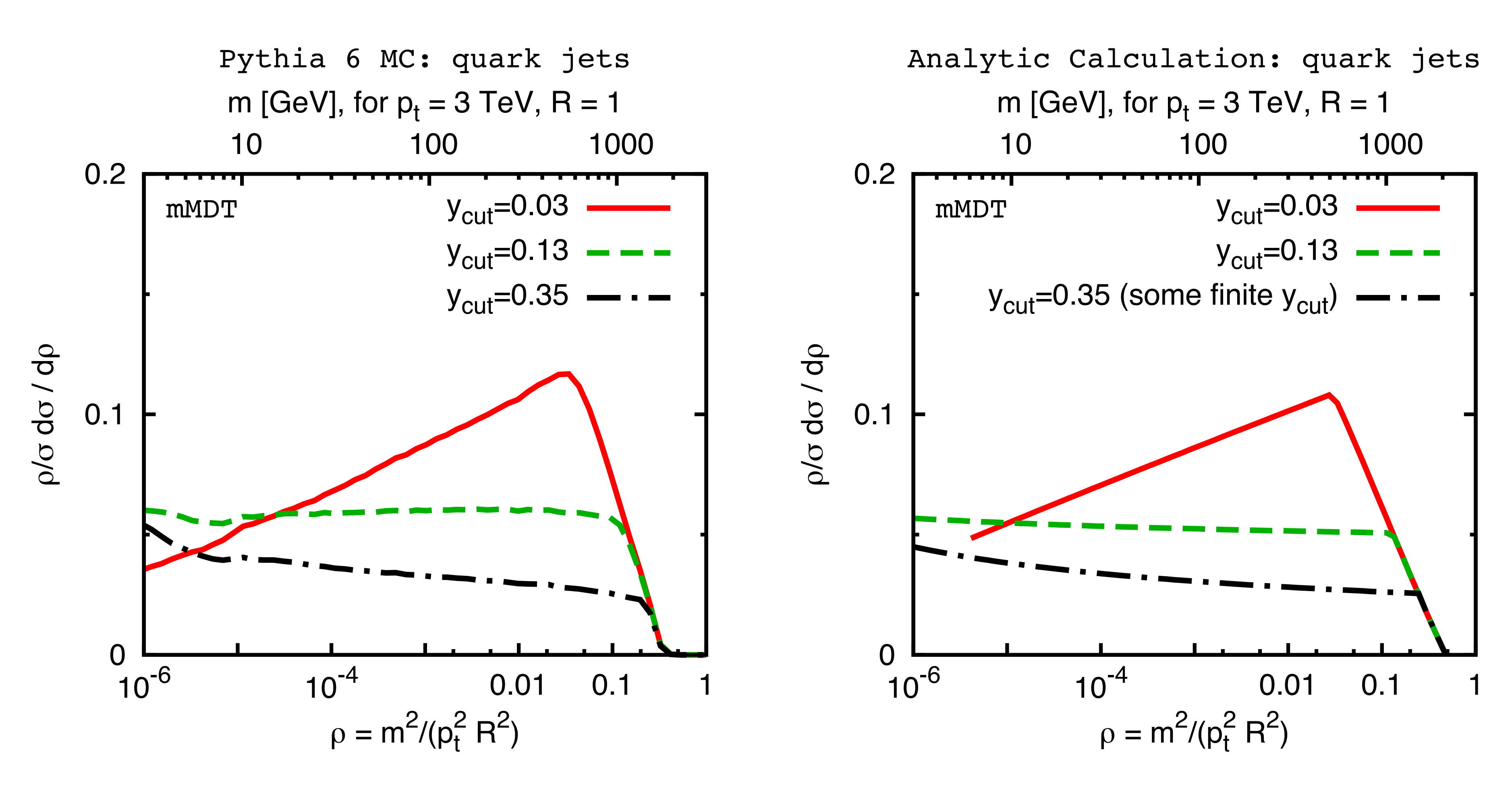
Modified Mass Drop Tagger

1. Undo the last stage of the C/A clustering. Label the two subjets j_1 and j_2 ($m_1 > m_2$)
2. If $m_1 < \mu m$ (mass drop) and the splitting was not too asymmetric ($y_{ij} > y_{cut}$), tag the jet.
3. Otherwise redefine j to be the subjet with highest transverse mass and iterate.

- In practice the soft-branch contribution is very small
- However, this modification makes the all-order structure particularly interesting



Comparison to MC



Remarkable agreement !

Marzani

first sophisticated responses about IR unsafety

Boost '13

Unsafe but Calculable

Ratio Observables in Perturbative QCD

Jesse Thaler

Based on 1307.1699 with Andrew Larkoski

Boost 2013, Flagstaff — August 14, 2013

Jesse Thaler — Unsafe but Calculable

Thaler

first sophisticated responses about IR unsafety

Boost '13

Track Thrust Resummation

- For $\tau \ll 1$ resummation of $\ln \tau$ required
- Focus on exponentation of single soft emission $S^{(1)}(k, \mu)$

$$\bar{S}^{(1)}(\bar{k}, \mu) = \int_0^\infty \mathrm{d}k\, S^{(1)}(k, \mu) \int_0^1 \mathrm{d}x\, T_g(x, \mu)\, \delta(\bar{k} - xk)$$
$$= \delta(\bar{k}) + \frac{\alpha_s C_F}{2\pi} \Big[-\frac{8}{\mu} \Big(\frac{\ln(k/\mu)}{k/\mu} \Big)_+ + \frac{8 g_1^L}{\mu} \frac{1}{(k/\mu)_+}$$
$$+ \Big(\frac{\pi^2}{6} - 4 g_2^L \Big) \delta(\bar{k}) \Big]$$

- Depends on logarithmic moments of T_g

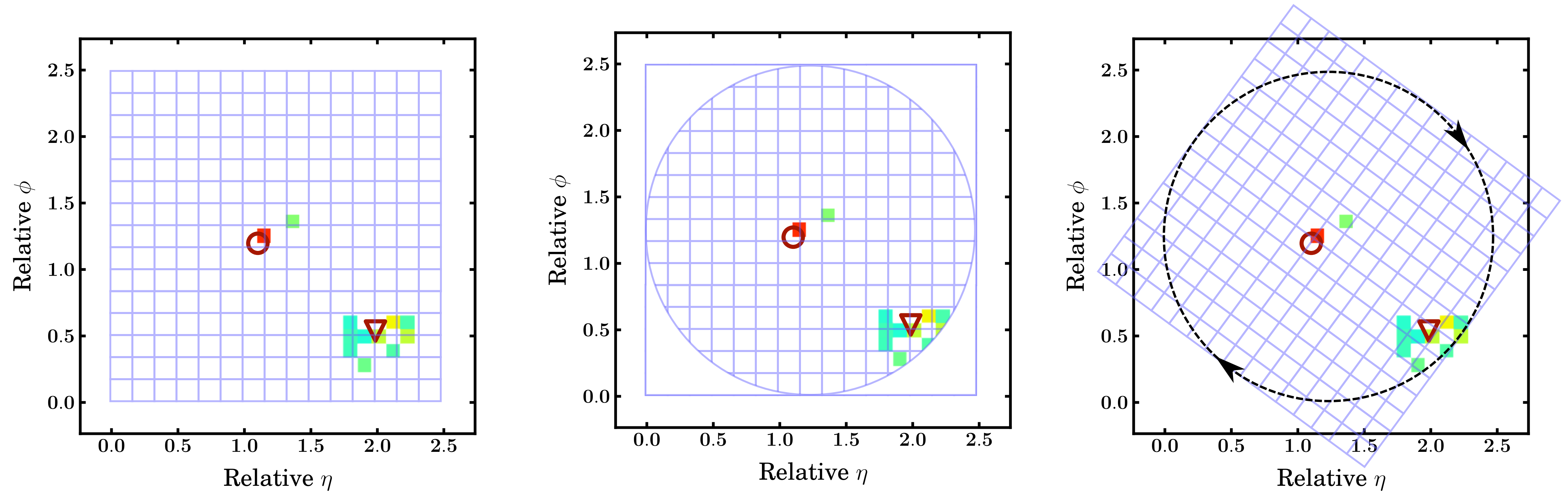
$$g_n^L \equiv \int_0^1 \mathrm{d}x\, T_g(x, \mu) \ln^n x$$

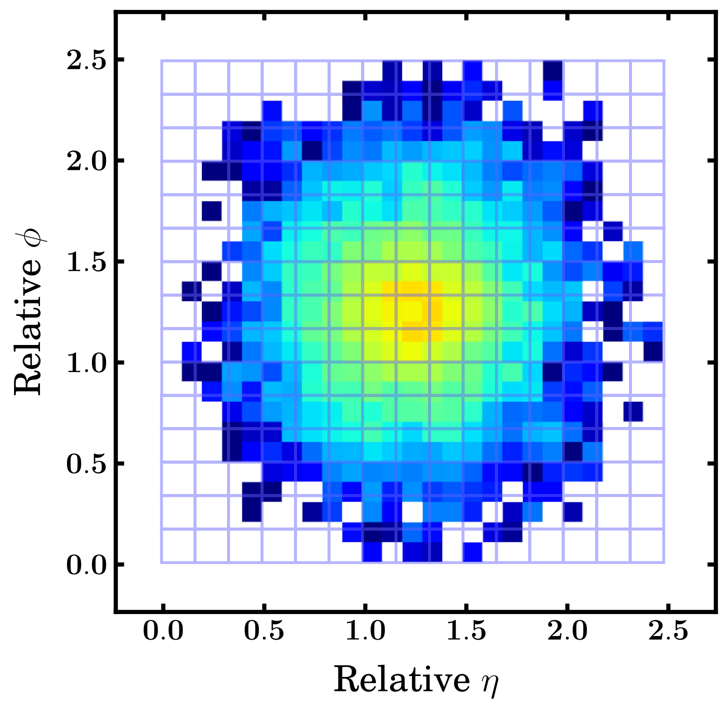
- Interestingly, hadronization effects are exponentiated

Boost ‘13

IMAGE PROCESSING

- Center and rotate jet-images before training MVA
- Introduces small smearing, but huge gain in discrimination!



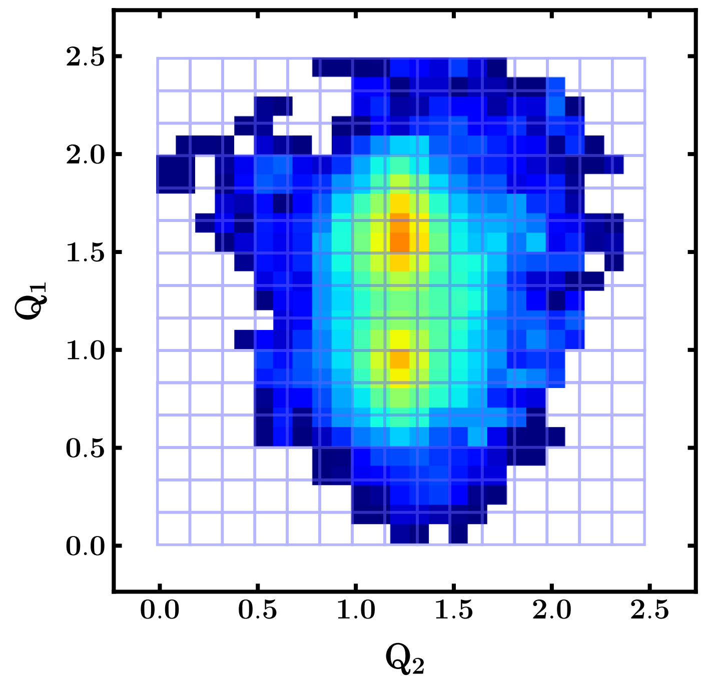


Average of *unrotated* W jet

Not much info!

Average of *rotated* W jet

Much better!



Cogan

Version 1.005 of FastJet Contrib is distributed with the following packages

Package	Version	Information
GenericSubtractor	1.2.0	README NEWS
JetFFMoments	1.0.0	README NEWS
VariableR	1.0.1	README NEWS
Nsubjettiness	1.0.2	README NEWS
EnergyCorrelator	1.0.1	README NEWS
ScJet	1.1.0	README NEWS

BOOST2014, 18-22 August 2014, University College London

6th International Workshop on Boosted Object Phenomenology, Reconstruction & Searches in HEP

UCL

- Overview
- Public Event
- Workshop Agenda
- Abstracts
- Registration
- Payment
- Participants
- Social Events
- Committees
- Previous Workshops
- Contacts

Practical Info

- Accommodation
- Travel and Maps
- London

Local amenities

- Pubs / Bars
- Restaurants

Other links

- UCL HEP

BOOST2014 is the sixth of a series of successful joint theory/experiment workshops which bring together the world leading experts from theory and the Tevatron and LHC experiments to discuss the latest progress and develop new approaches on the reconstruction and use of boosted decay topologies as search tools for new physics. This year, the workshop is hosted by the UCL HEP Group at the heart of London.

the new generation of tools

today's tools see the light: soft drop

Boost '14

Marzani

Soft Drop

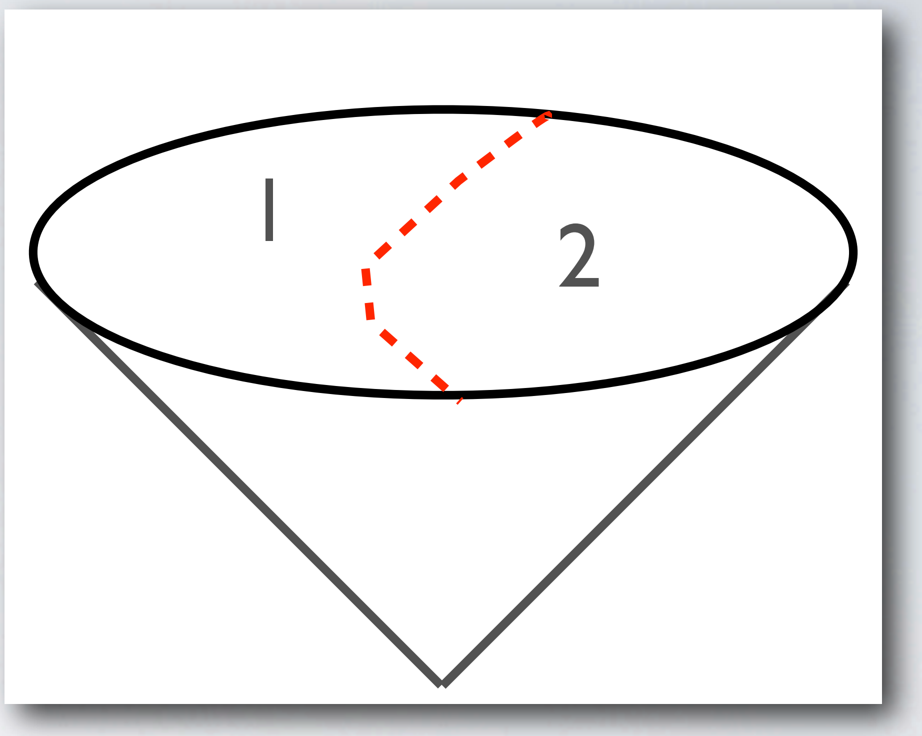
Larkoski, SM, Soyez and Thaler (2014)

1. Undo the last stage of the C/A clustering. Label the two subjets j_1 and j_2 .

2. If

$$\frac{\min(p_{T1}, p_{T2})}{p_{T1} + p_{T2}} > z_{\mathrm{cut}} \left(\frac{\Delta R_{12}}{R_0}\right)^{\beta}$$

then deem j to be the soft-dropped jet.



3. Otherwise redefine j to be the harder subjet and iterate.

1-prong jets can be either kept (grooming mode) or discarded (tagging mode)

- Generalisation of the (modified) Mass Drop procedure
- no mass drop condition (not so important)
- mMDT recovered for β=0
- some inspiration from semi-classical jets

Butterworth, Davison, Rubin and Salam (2008)
Dasgupta, Fregoso, SM and Salam (2013)

Tseng and Evans (2013)

10

today's tools see the light: pileup

Boost '14

BOOST2014

Constituent Subtraction

Peter Berta[1)], Martin Spousta[1)], David Miller[2)], Rupert Leitner[1)]

[1)] Charles University in Prague
[2)] Enrico Fermi Institute, University o

21st Aug 2014

P. Berta (Charles University in Prague) Constituent Subtraction

08/21/14

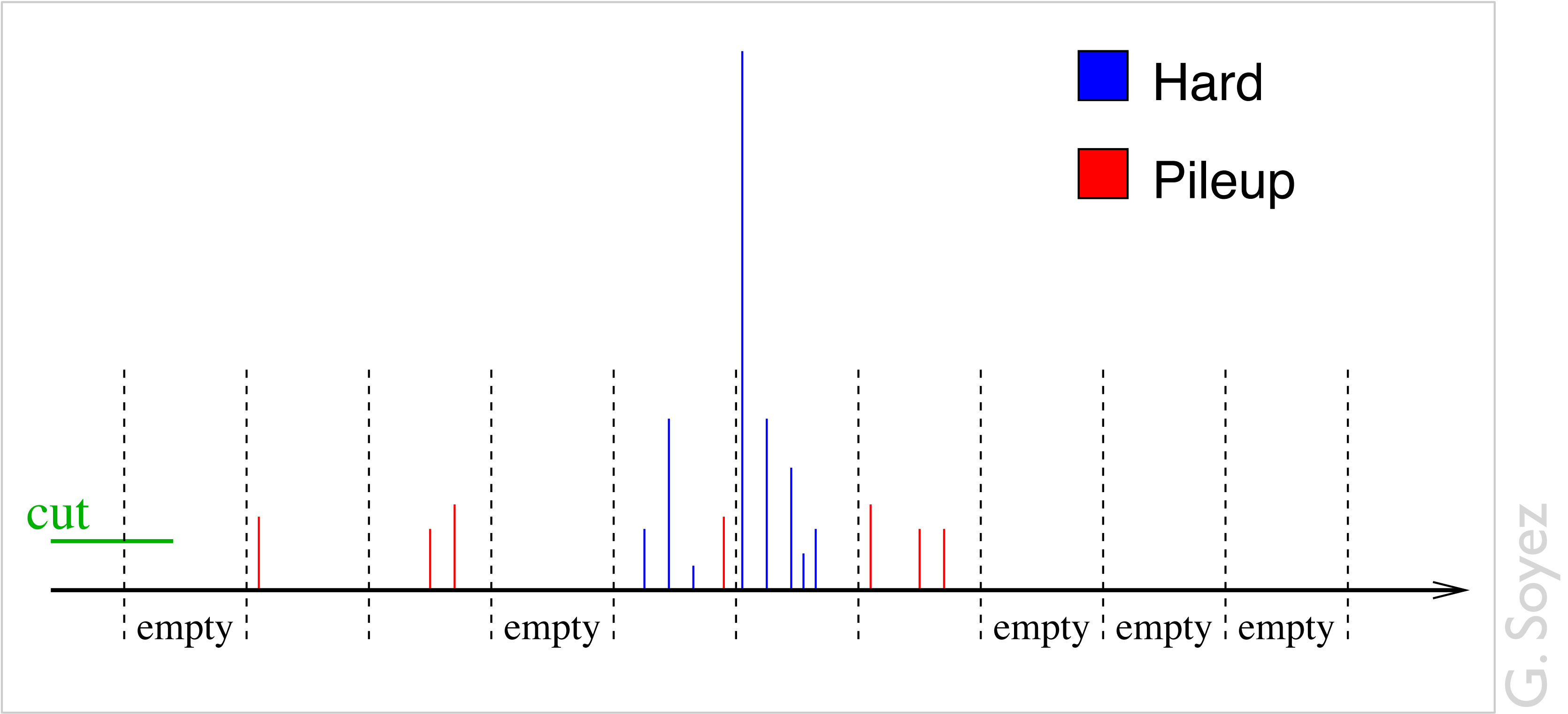
arxi

PileUp Per Particle Id

Philip Harris (CERN) Nhan Tran(FNAL), Daniele Bertolini(MIT), Matthew Low(Chicago)

Soft Killer

$$p_t^{cut} = \underset{i \in \text{patches}}{\text{median}} \{p_{ti}^{max}\}$$



Half of the event is empty ⇒ **ρ = 0** (because it's the median)

NB. SK needs tuning of the size of the patches used to calculate ρ.
0.4 was found to be a good choice for R=0.4 jets

Matteo Cacciari - LPTHE BOOST - London - August 2014 28

calculations for shape-like quantities

Boost '14

Predicting Multi-Differential Jet Cross Sections

Andrew Larkoski
MIT

AJL, I. Moult, D. Neill 1401.4458; AJL, I. Moult 14XX.soon;
AJL, I. Moult, D. Neill 14XX.soon

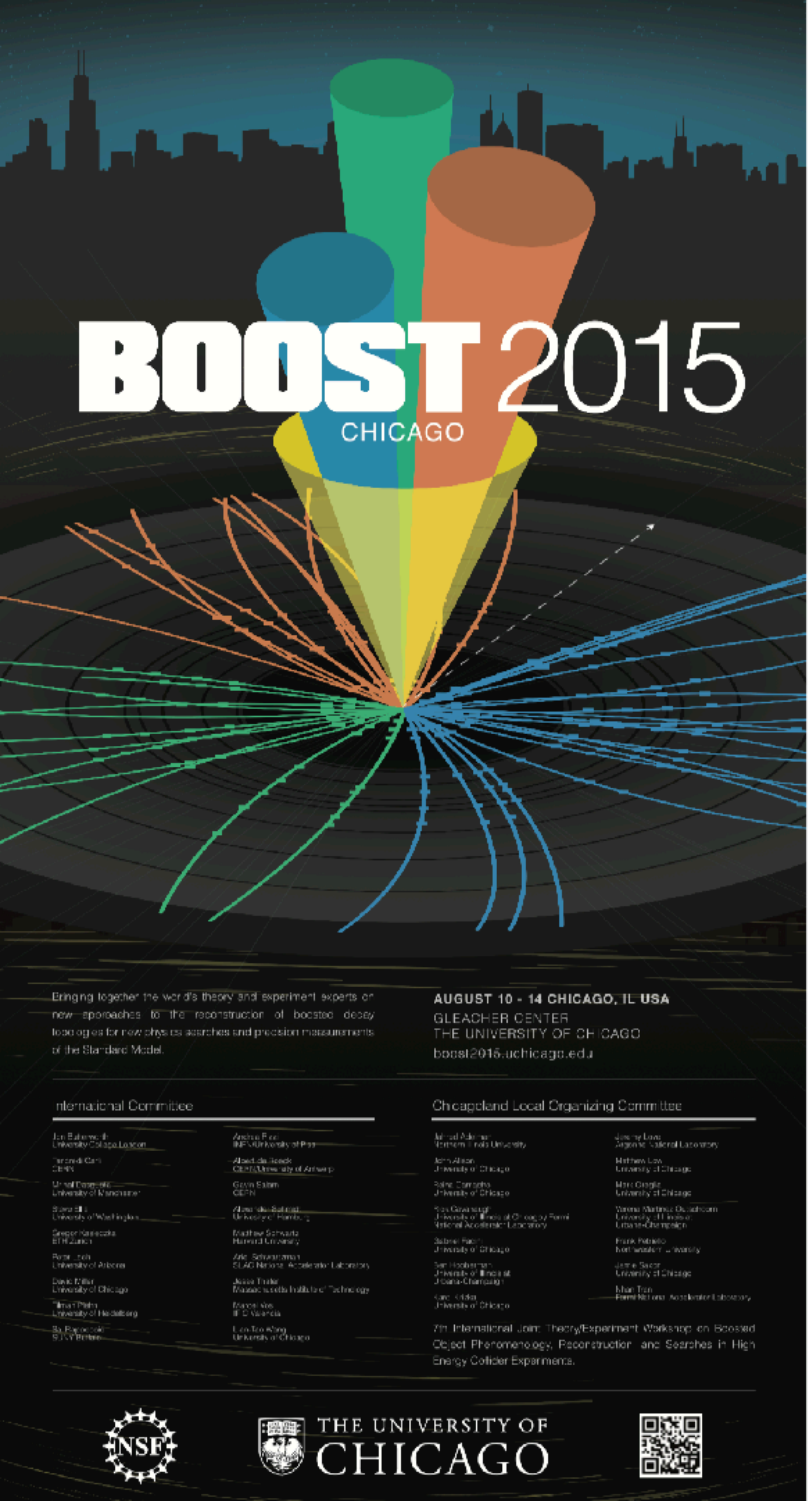
BOOST 2014, August 18, 2014

Q-calculations

Conclusions and Outlook

- Q-thrust:
    - non-deterministic ***but*** energy-flow variable
    - calculable!
    - interesting (important?) effect on NGLs!
    - generalizes naturally to Q-(sub)jettiness
- Outlook:
    - performance & correlations
    - many related observables to study, should exhibit same generic properties (calculability and NGLs)

Hornig


BOOST 2015
CHICAGO
Bringing together the world's theory and experiment experts on new approaches to the reconstruction of boosted decay topologies for new physics searches and precision measurements of the Standard Model.
AUGUST 10 - 14 CHICAGO, IL USA
GLEACHER CENTER
THE UNIVERSITY OF CHICAGO
boost2015.uchicago.edu
International Committee
Chicagoland Local Organizing Committee
7th International Joint Theory/Experiment Workshop on Boosted Object Phenomenology, Reconstruction and Searches in High Energy Collider Experiments.
NSF
THE UNIVERSITY OF CHICAGO

open data

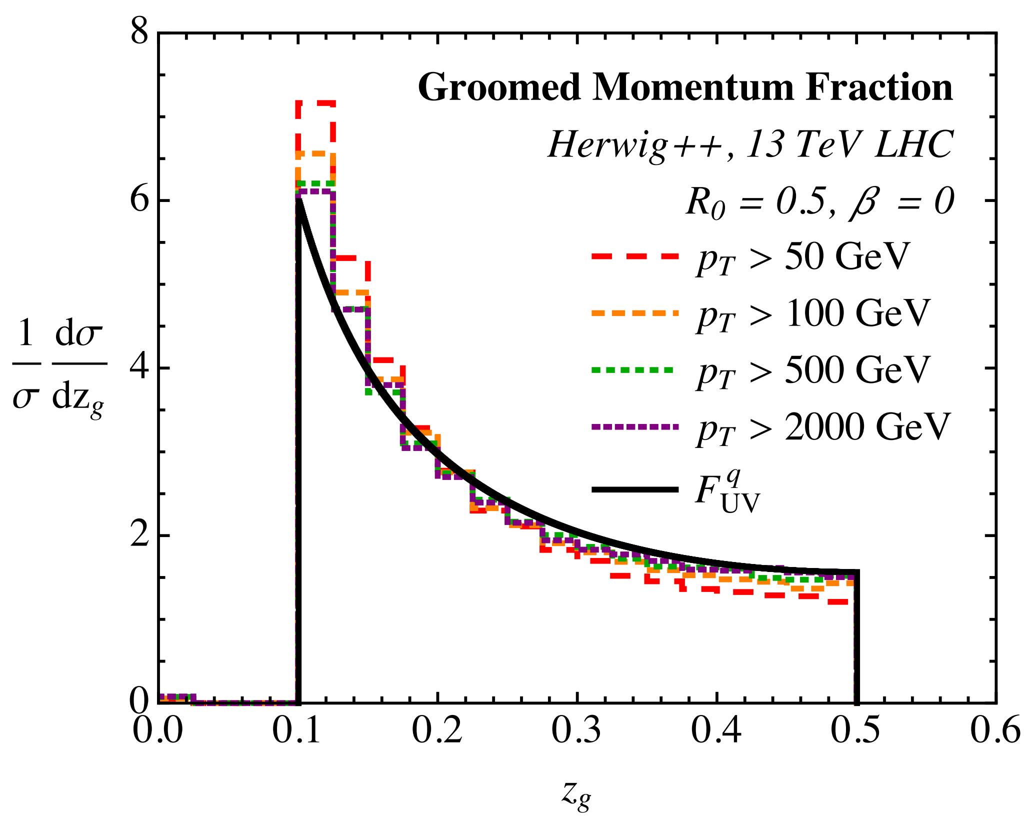
z_g, open data

Boost '15

A Standard Candle for Jets

$z > z_{\text{cut}}$

$$\frac{1}{\sigma}\frac{d\sigma}{dz_g} = \frac{\overline{P}_i(z_g)}{\int_{z_{\text{cut}}}^{1/2} dz\, \overline{P}_i(z)} + \ldots$$



$z_g \Rightarrow P(z)$

≈ independent of α_s (!)
≈ independent of jet p_T and radius
≈ same for quarks and gluons

calculable deviations from universality

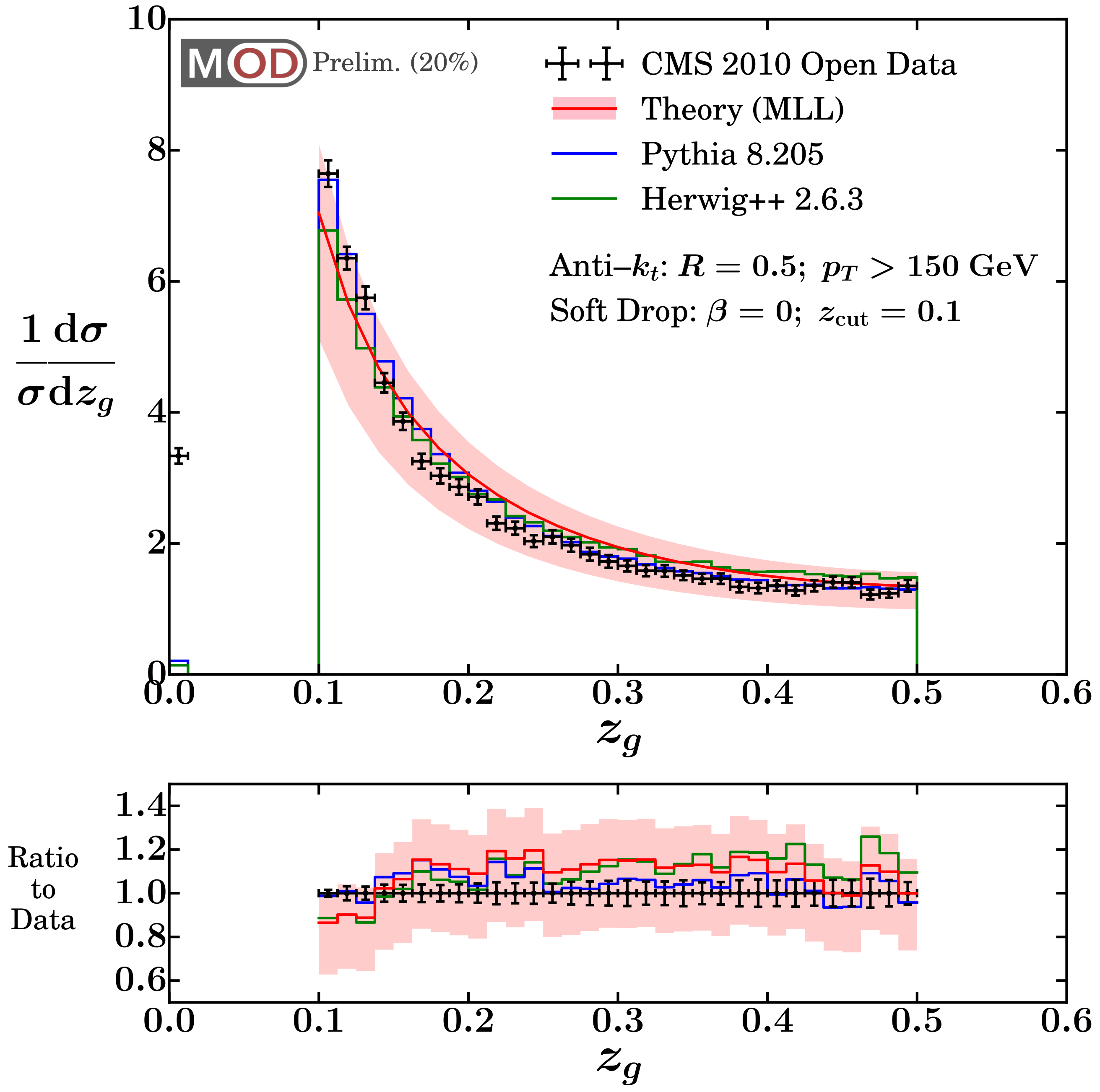
(see backup for $\beta \neq 0$)

[Larkoski, Marzani, JDT, 1502.01719]

15

Thaler

Open Data Analysis



As nature intended:

$p_T > 150$ GeV
$z_{\text{cut}} = 0.1$

CMS Open Data: Jet Primary Data Set with Particle Flow Candidates

Statistical uncertainties only, no unfolding, 58021 events

Using single jet triggers with ≈100% efficiency, AK5 jet energy corrections with area subtraction, no PFC corrections

AOD → MOD format (MIT Open Data project)

More plots in backup slides

[Thanks to Sal Rappoccio, **Aashish Tripathee**, Wei Xue]

Jesse Thaler — Probing the Core of QCD

24

z_g: the path to interaction with the heavy-ion community

+ demonstration of potential of open data

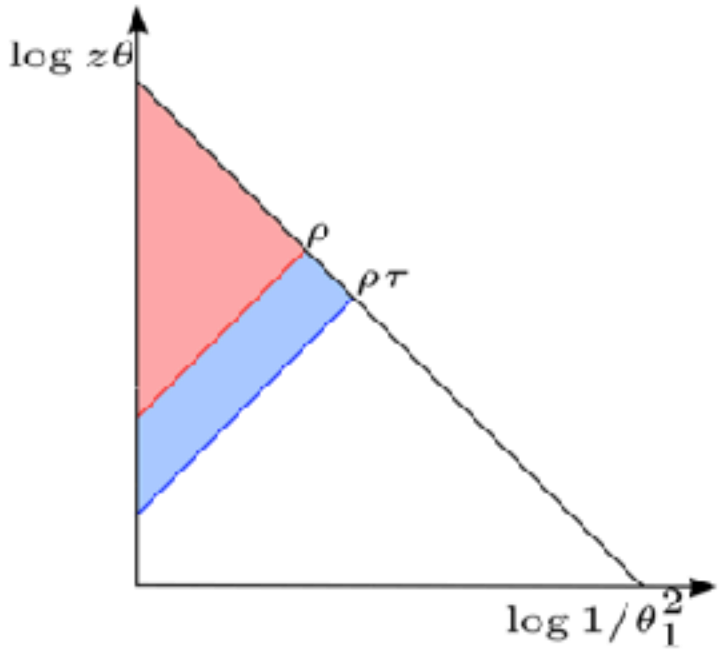
understanding many shapes

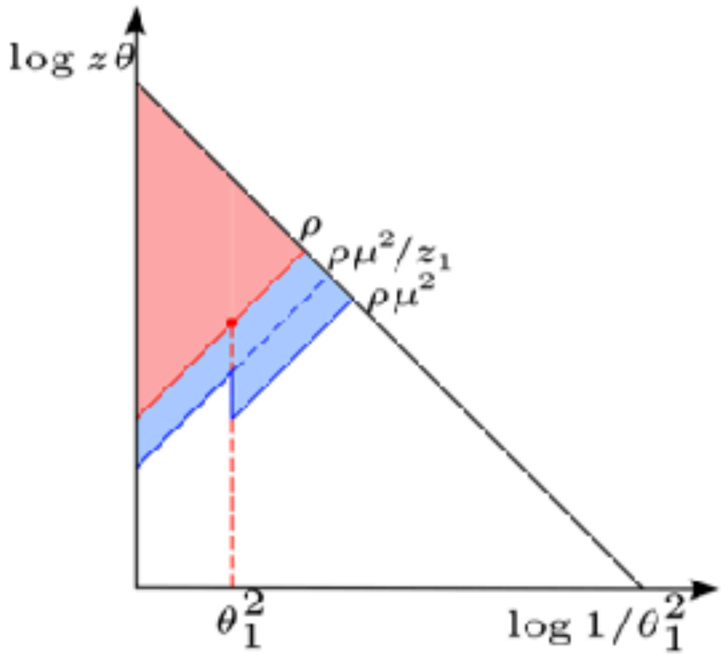
Boost '15

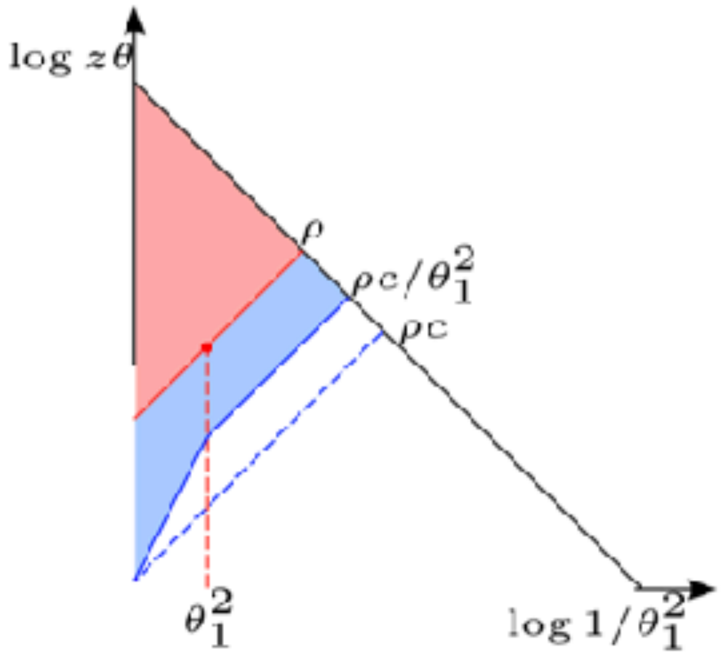
Subjet mass constraints on boosted jets

└ Results for QCD background

N-subjettiness	*Mass drop*	*Energy correlation*







$$R_\tau(z_1) = \frac{\alpha_s C_R}{2\pi}\left[\frac{L_\tau^2}{2} + L_\rho L_\tau\right] + \frac{\alpha_s C_A}{2\pi}\frac{L_\tau^2}{2}$$

$$R_{\mu^2_{1/2}}(z_1) = \frac{\alpha_s C_R}{2\pi}\left[\frac{L_\mu^2}{2} + L_\rho L_\mu\right] - \frac{\alpha_s C_R}{2\pi}\frac{L_\rho}{2}(L_\rho - L_1) + \frac{\alpha_s C_A}{2\pi}\frac{(L_\mu - L_1)^2}{2}$$

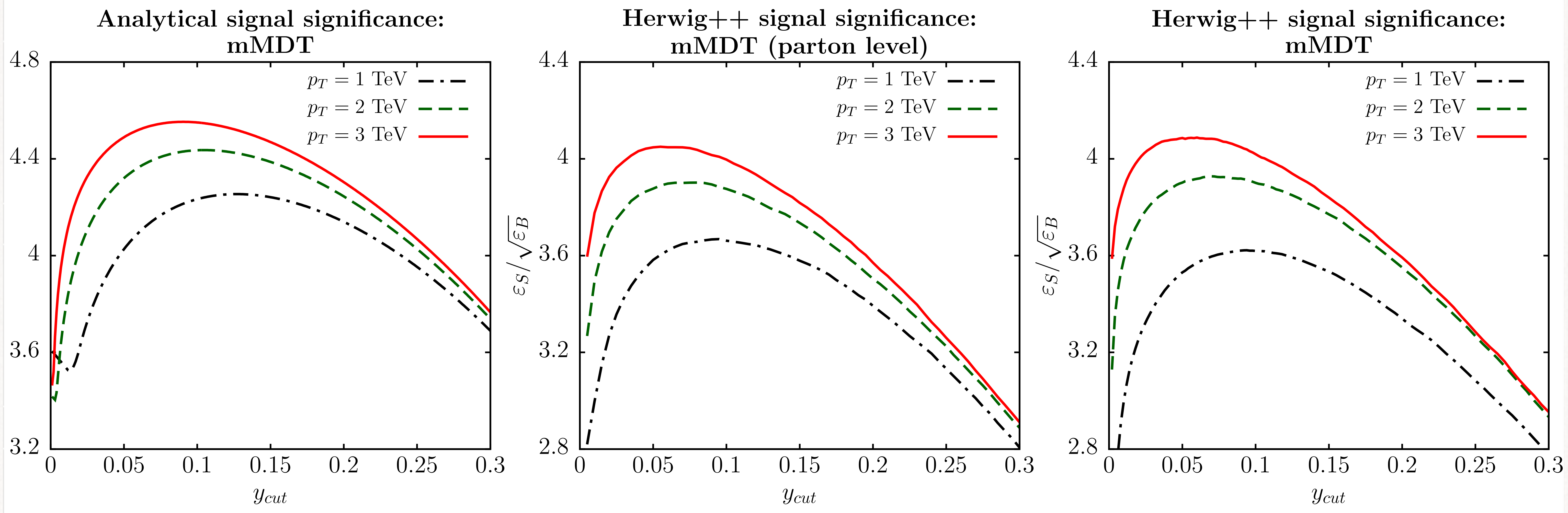
$$R_{C2}(z_1) = \frac{\alpha_s C_R}{2\pi}\left[\frac{L_e^2}{2} + (L_e - L_\rho + L_1)L_1\right] + \frac{\alpha_s C_A}{2\pi}\frac{1}{2}(L_e - L_\rho + L_1)^2$$

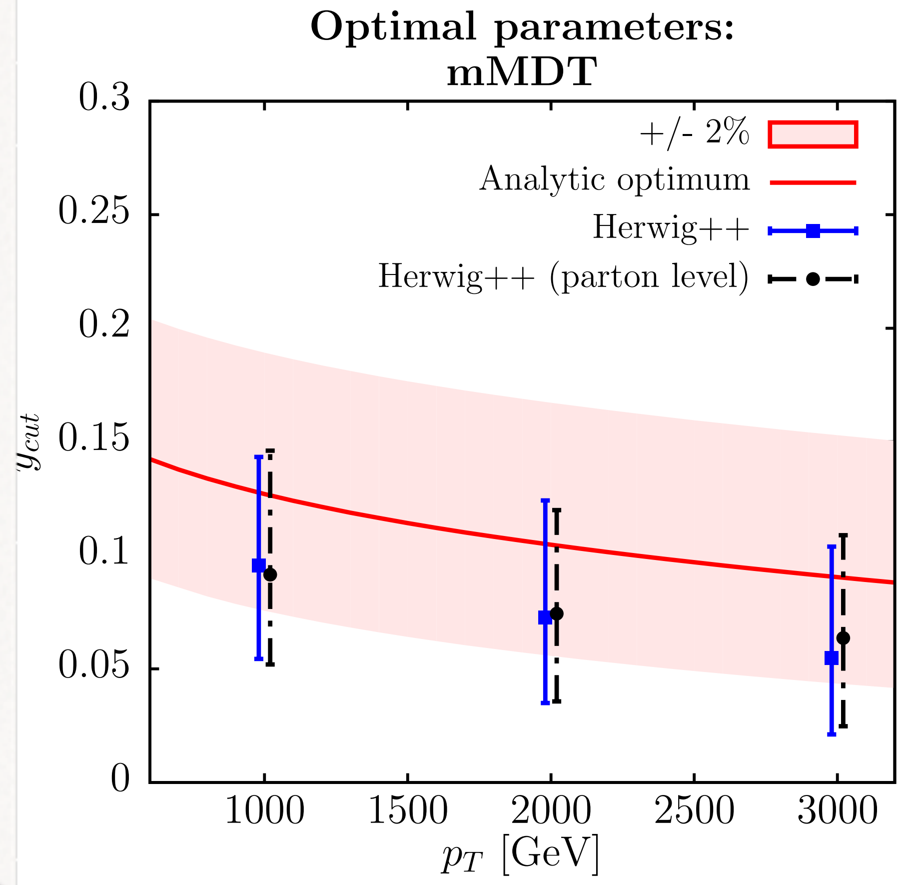
$L_X = \log(1/X)$

Schunk

Siodmok

Optimal values - mMDT





- Herwig parton level agrees with analytics (both the peak positions and the evolution of opt. y_{cut} with p_T).
- hadronisation and UE do not change the picture significantly
- peaks are broad $\Rightarrow$ slightly non-optimal y_{cut} is still ok.
- good degree of overlap within tolerance band between full MC and analytical estimates.

beyond ROCs?
deep learning?

beyond the ROC curve — with shapes

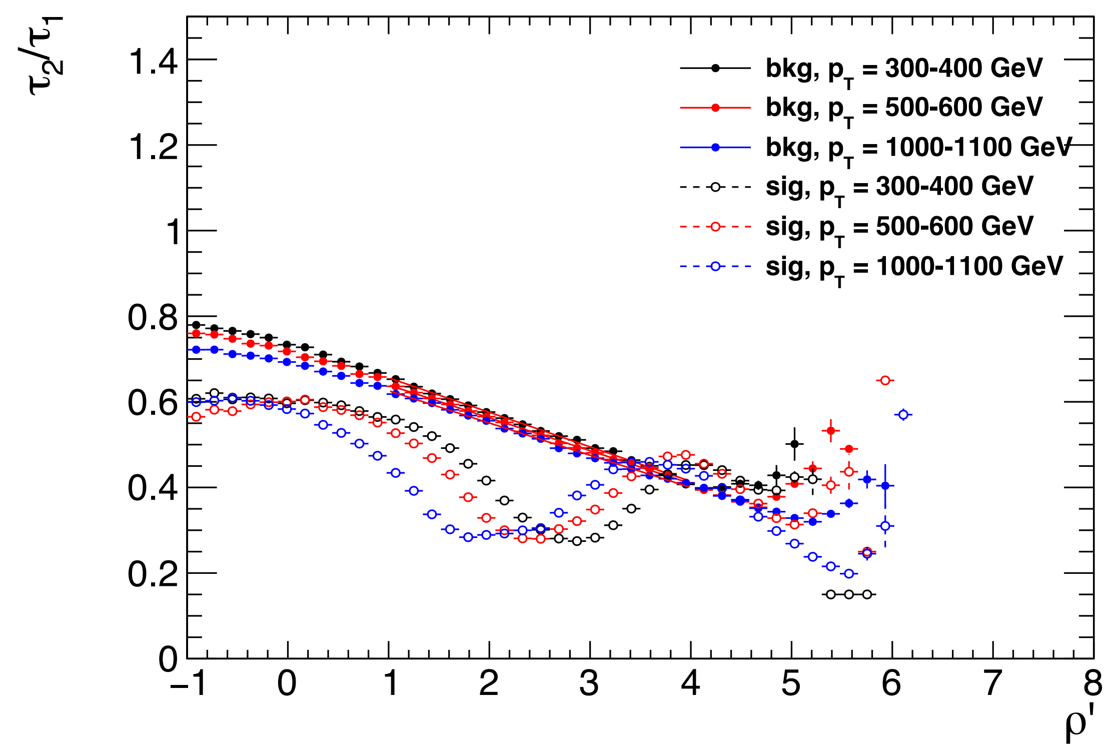
Boost '16

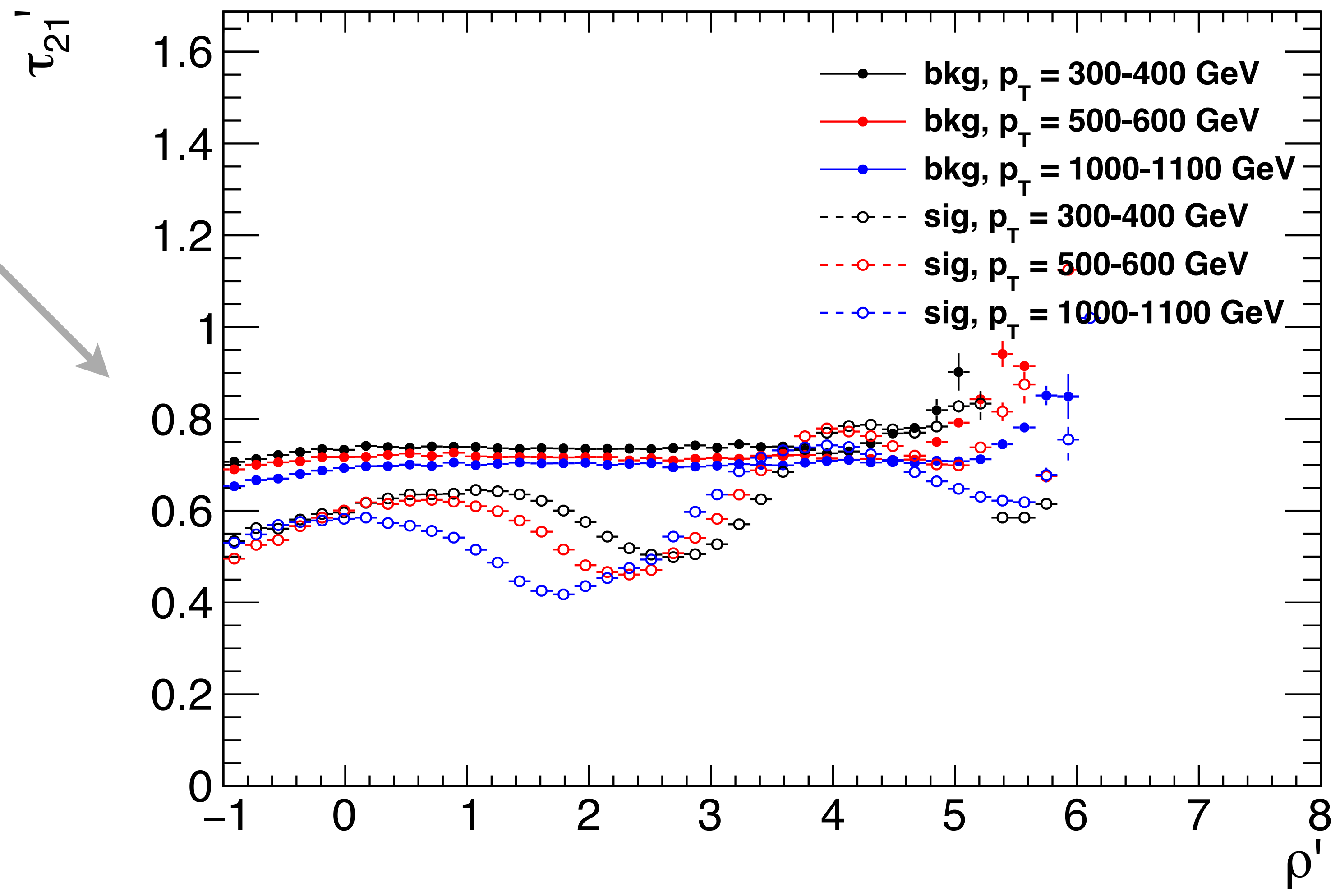
Designing Decorrelated Taggers

Now linear correlation constant in bins of pT!

trivial transformation

$$\tau_{21}{}^{DDT} = \tau^{21} - M \times \rho^{DDT}$$





now $\tau_{21}{}^{DDT}$ is uncorrelated with ρ^{DDT}!
makes background estimation easier!

9

Rappoccio

deep learning hits the market

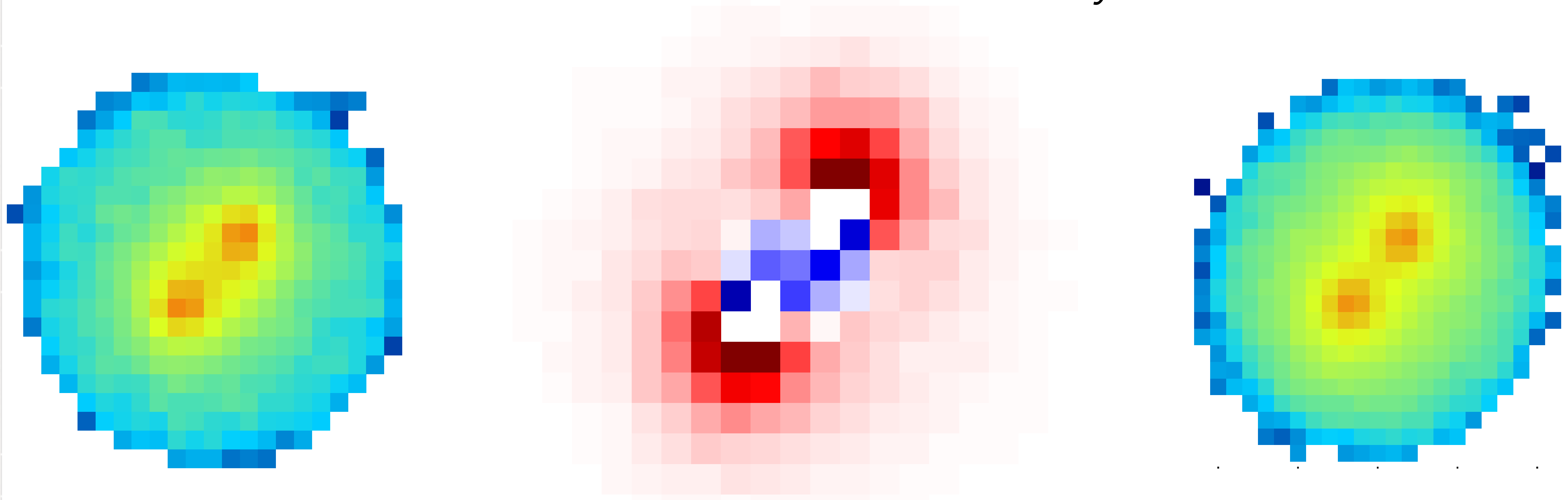
Boost '16

Jet Images with Deep Learning

JHEP, 2016(7), 1-32

Anton Apostolatos, Leonard Bronner, Luke de Oliveira, Michael Kagan, Lester Mackey, Benjamin Nachman, and Ariel Schwartzman

SLAC and Stanford University



BOOST 2016: Zurich, Switzerland

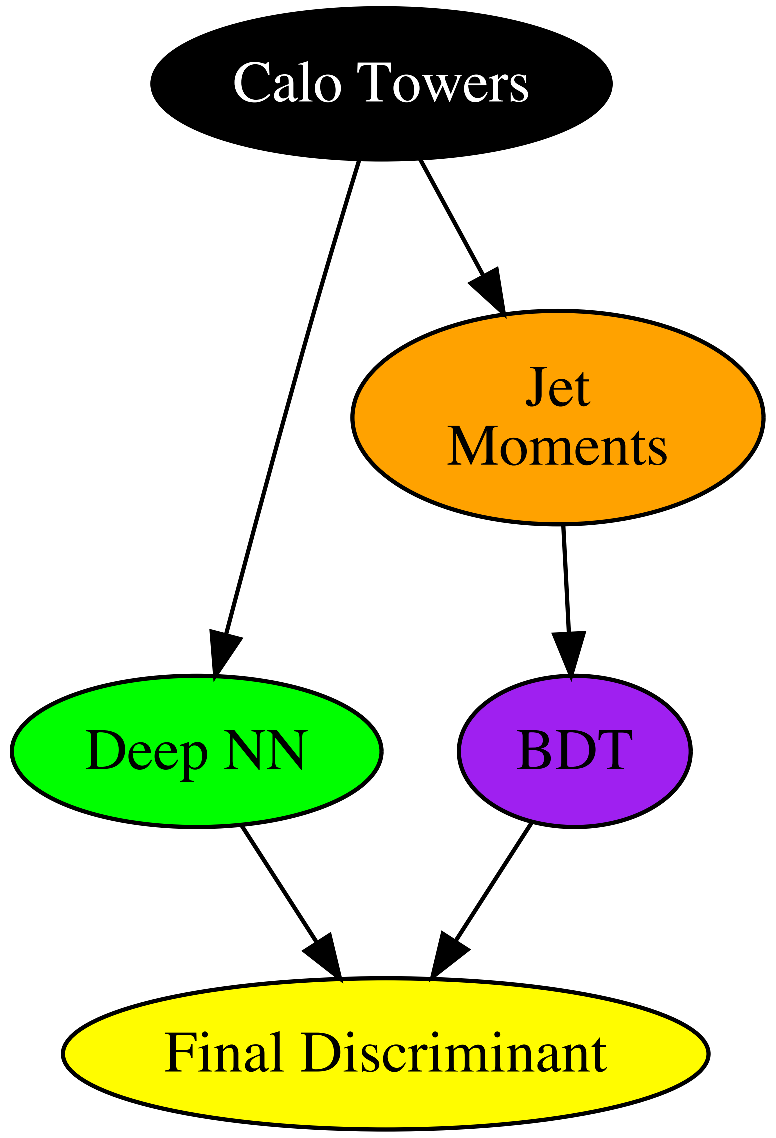
Thursday, July 21, 2016

Nachmann

Guest

Our Setup

- Used Delphes, $\sqrt{s} = 14\,\text{TeV}$, $\langle \mu \rangle = 50$
- Signal: $pp \to WW \to qqqq$
- Background: $pp \to qq, q, gg$
- Anti-k_T jets $\Delta R = 1.2$
- $300\,\text{GeV} < p_T < 400\,\text{GeV}$
- Apply pileup suppression (trimming)



The Question

If we compare

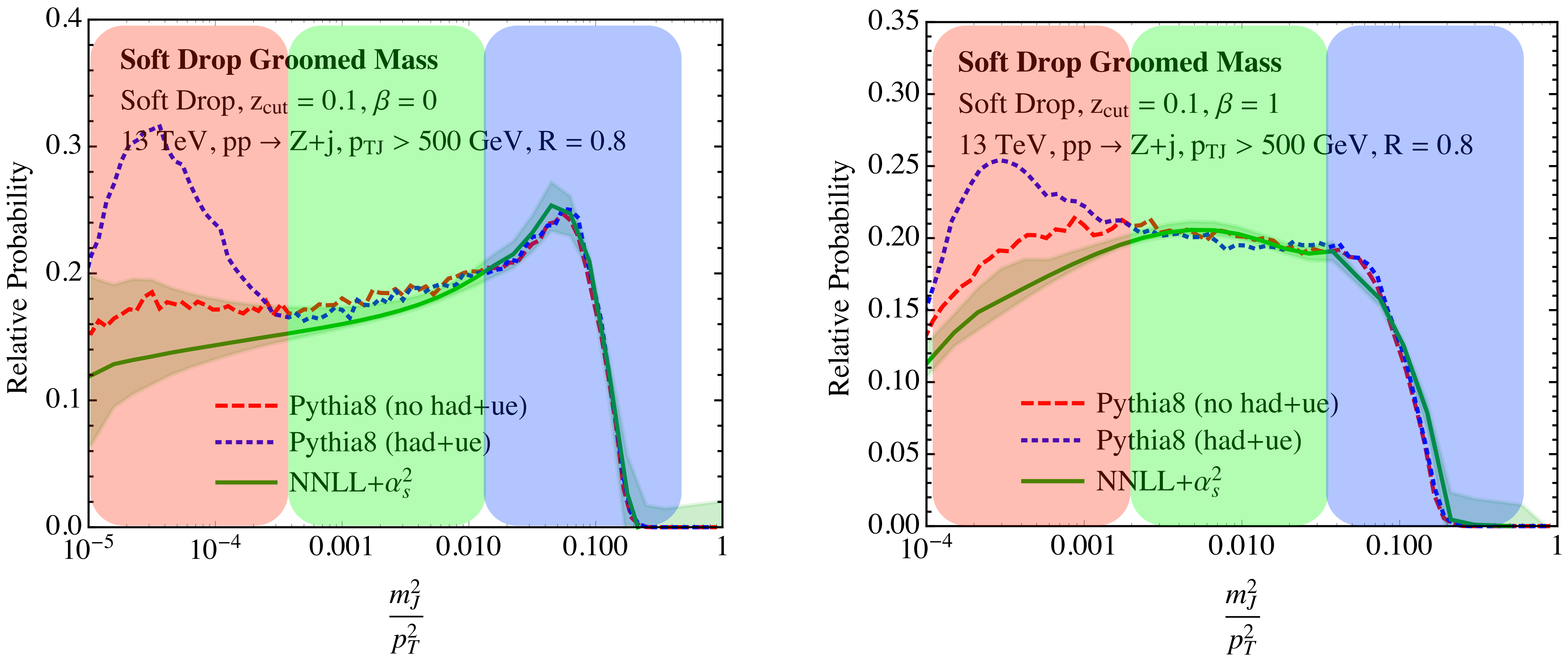
- A BDT on engineered variables, to
- A deep network on the jet image

which one is a better classifier?

mMDT/soft drop meets precision resumation

Boost '16

Results: NNLL+α_s^2 Jet Substructure



NNLL+α_s^2, β = 0 | NNLL+α_s^2, β = 1

Hadronization Regime

Resummation Regime
Fixed-Order Regime } Perturbative Regime

Almost three decades of perturbative control in a single jet distribution!

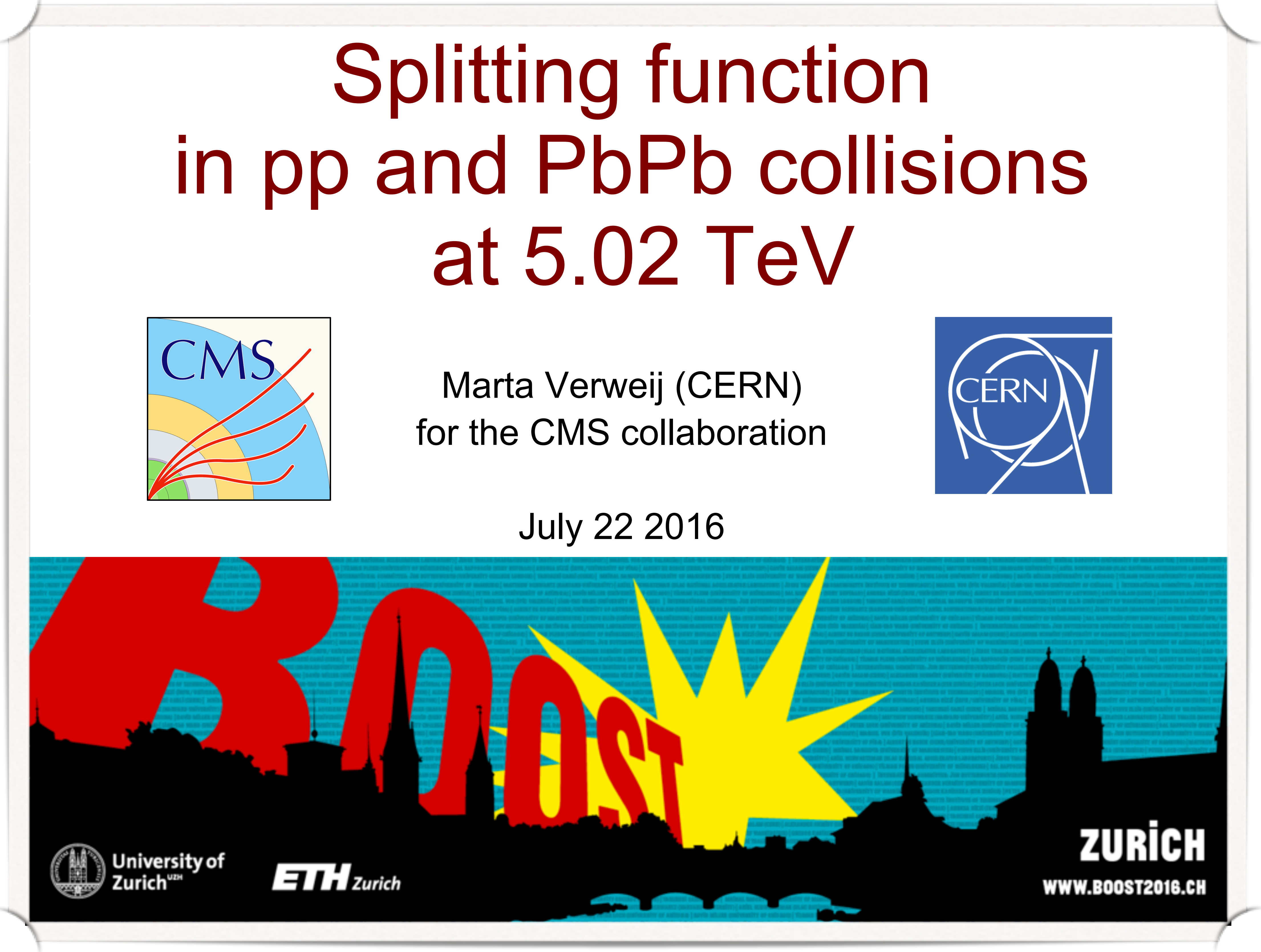
Larkoski

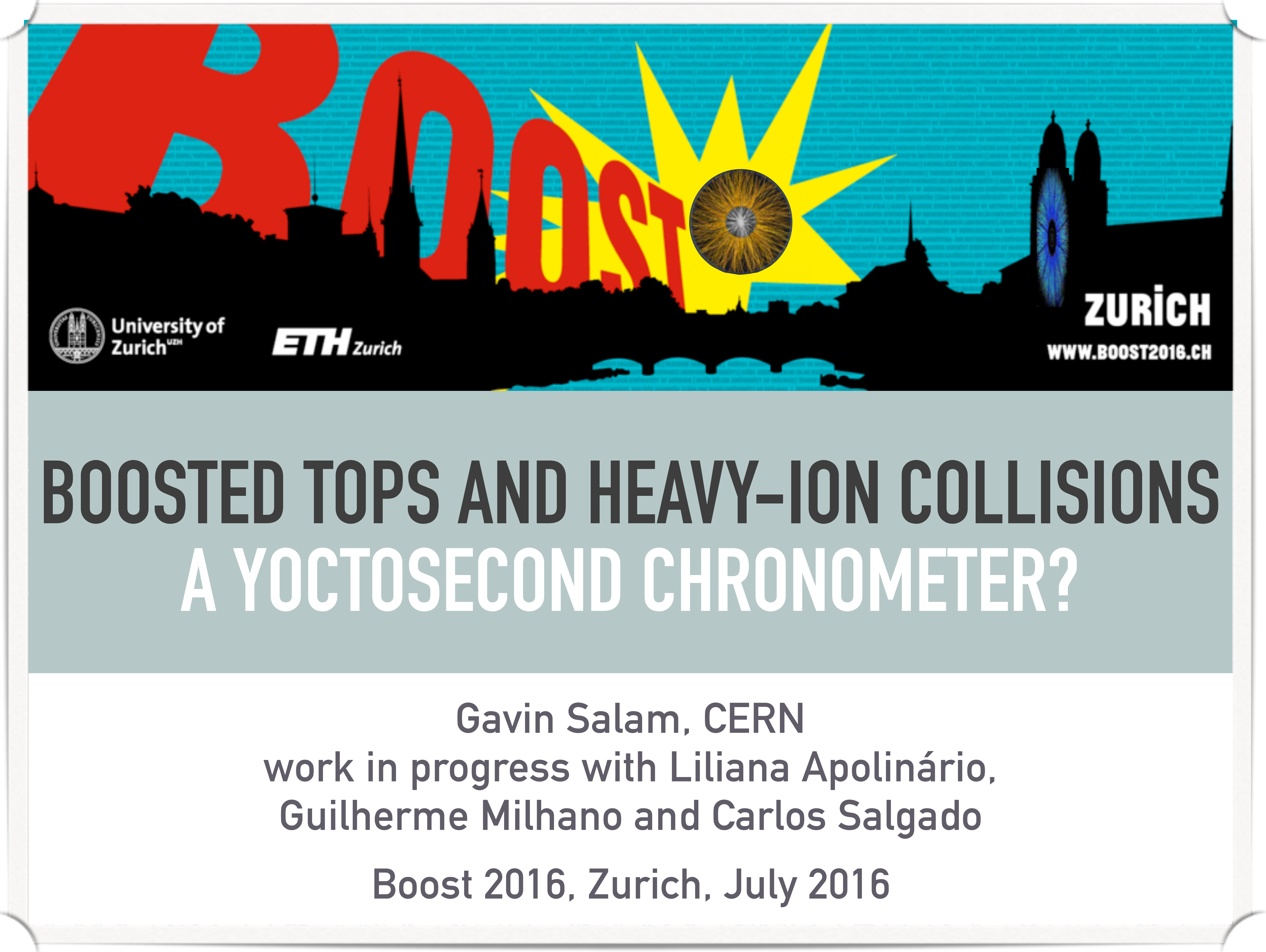
18

Frye, AJL, Schwartz, Yan 2016

theory ⇔ expt link with heavy-ions established

Boost '16





deep learning v. deep thinking

Deep learning & deep thinking

Shows wonderful performance (and likely more to come)

“We’re not concerned by IRC safety (in ATLAS)”
(undisclosed source)

“It’s time to organize and move forward. It’s time for deep thinking, reformation of the Democratic Party”
(K. Vanden Heuvel)

“More is different: Just because you know the QCD Lagrangian doesn’t mean you know all of its physics”
(Andrew’s intro on Monday)

More than “Deep learning v. Deep thinking”,
what about “Deep Understanding”?

Soyez

recursive & iterative soft drop

Boost '18

6

Iterated soft drop

see Frederic Dreyer's talk for "recursive soft drop"

algorithm's parameters: $z_{\rm cut}$, β , $\theta_{\rm cut}$
used to define variables: z_n , θ_n

- begin at trunk of C/A clustering tree with $n = 1$
- at branching into subjets i,j require

  $\theta_{ij} > \theta_{\rm cut}$

  otherwise terminate algorithm
- if soft drop criterion is satisfied

  $z_{ij} > z_{\rm cut} \left(\frac{\theta_{ij}}{R} \right)^{\beta}$

  then

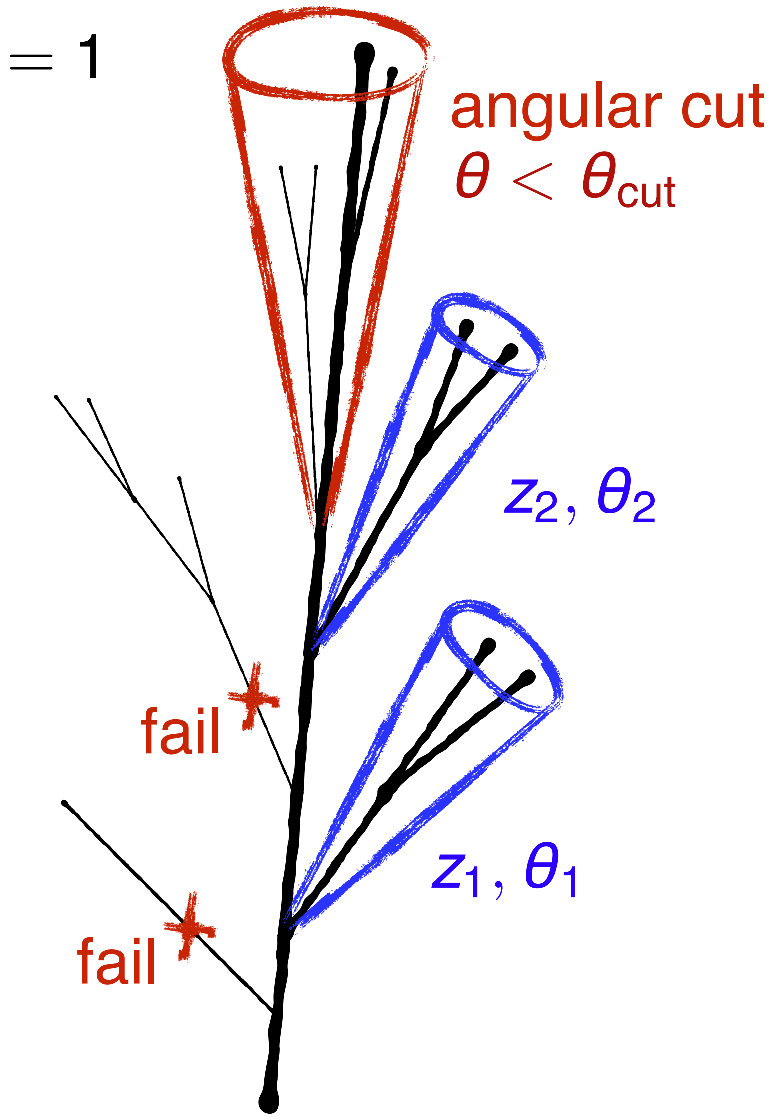
  $z_n = z_{ij}$

  $\theta_n = \theta_{ij}$

  $n \to n + 1$

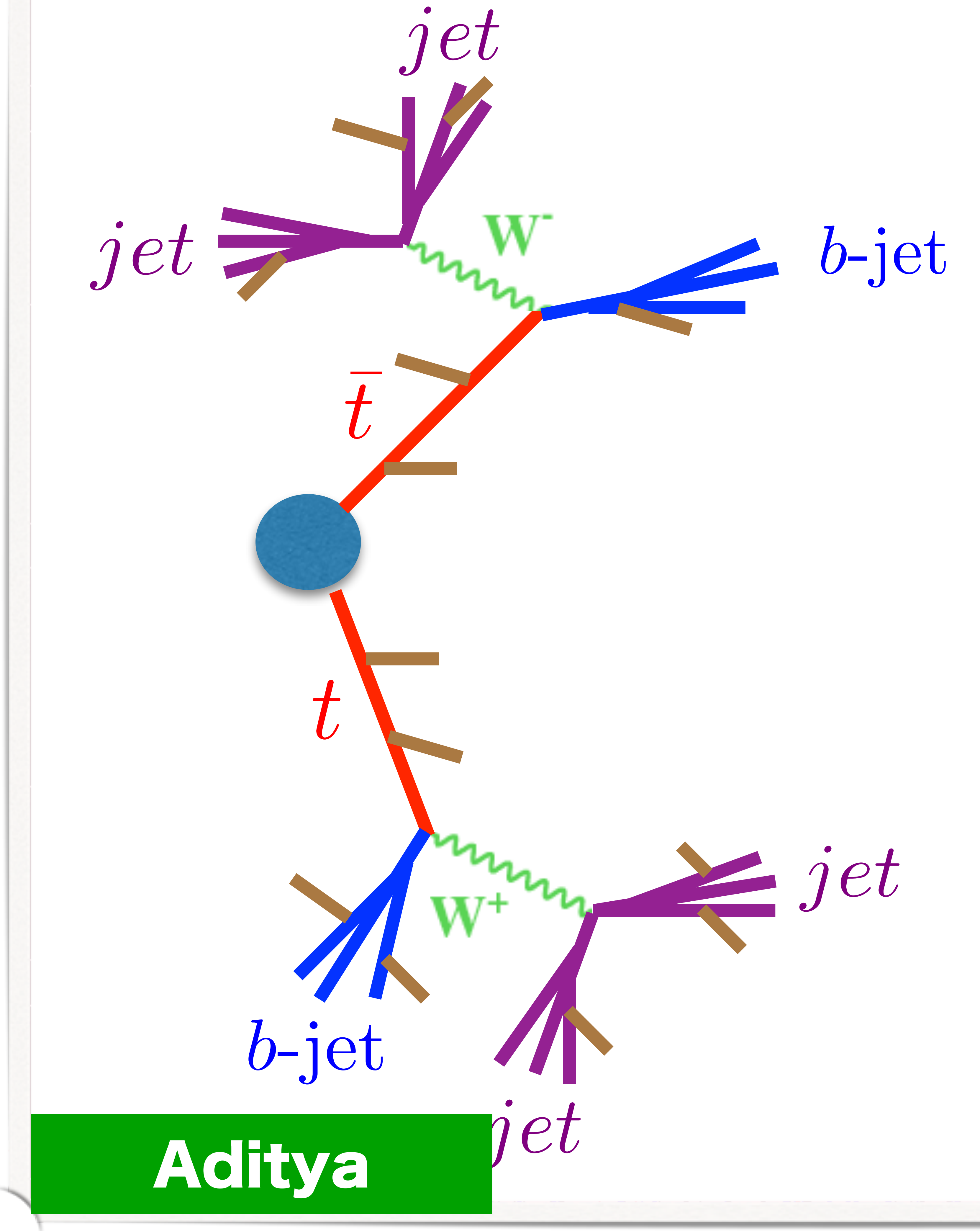
Frye

j and recurse



Larkoski, Marzani, Soyez, Thaler
JHEP 1405 (2014) 146

Dreyer



TOP JET MASS WITH SOFT DROP

Effective Theory for Groomed Top Jets

software: fastjet from python

Boost '18

FastJet 3.3.0

Date: Sun, 16 July 2017

Hi Gregory,
I am sorry to bother you with that, but students - at least undergrads in X - seem to prefer python over c++. I was wondering if there is a pthon wrapper for fastjet, or if there is another way of running fj with python?
Cheers,
***

Date: Wed, 12 July 2017

Release of FastJet 3.3.0
This is a main release which adds a first version of **a Python interface** to FastJet.

Soyez


BOOST 2018
10th International Workshop on Boosted Objects
Phenomenology, Reconstruction and Searches
Paris 16-20 July 2018
Searches using jet substructure
New jet substructure algorithms
Measurements and modelling
First-principles calculations
Machine Learning
Pileup mitigation
Heavy-ion collisions
Future colliders
Local Organising Committee
Matteo Cacciari (LPTHE Paris)
Reina Camacho (LPNHE Paris, CNRS)
Gregory Soyez (IPhT Saclay, CNRS)
International Advisory Committee
Lily Asquith (Sussex)
Ayana Arce (Duke)
Jon Butterworth (UCL)
Mrinal Dasgupta (Manchester)
Robin Erbacher (UC Davis)
Gregor Kasieczka (Hamburg)
Peter Loch (Arizona)
David Miller (Chicago)
Tilman Plehn (Heidelberg)
Sal Rappoccio (Buffalo)
Andrea Rizzi (Pisa)
Gavin Salam (CERN)
Alexander Schmidt (Aachen)
Matthew Schwartz (Harvard)
Ariel Schwartzman (SLAC)
Gregory Soyez (CNRS)
Jesse Thaler (MIT)
Marcel Vos (Valencia)
Lian-Tao Wang (Chicago)
indico.cern.ch/e/boost2018

today

the space of methods

c. 2018

Jet Declustering

Seymour93

YSplitter

ATLASTopTagger

Mass–Drop+Filter

Tagger

Tagger

TW

Planar Flow

Pruning

Trimming

HEPTopTagger (+ dipolarity)

Twist

Jet Shapes

N–jettiness

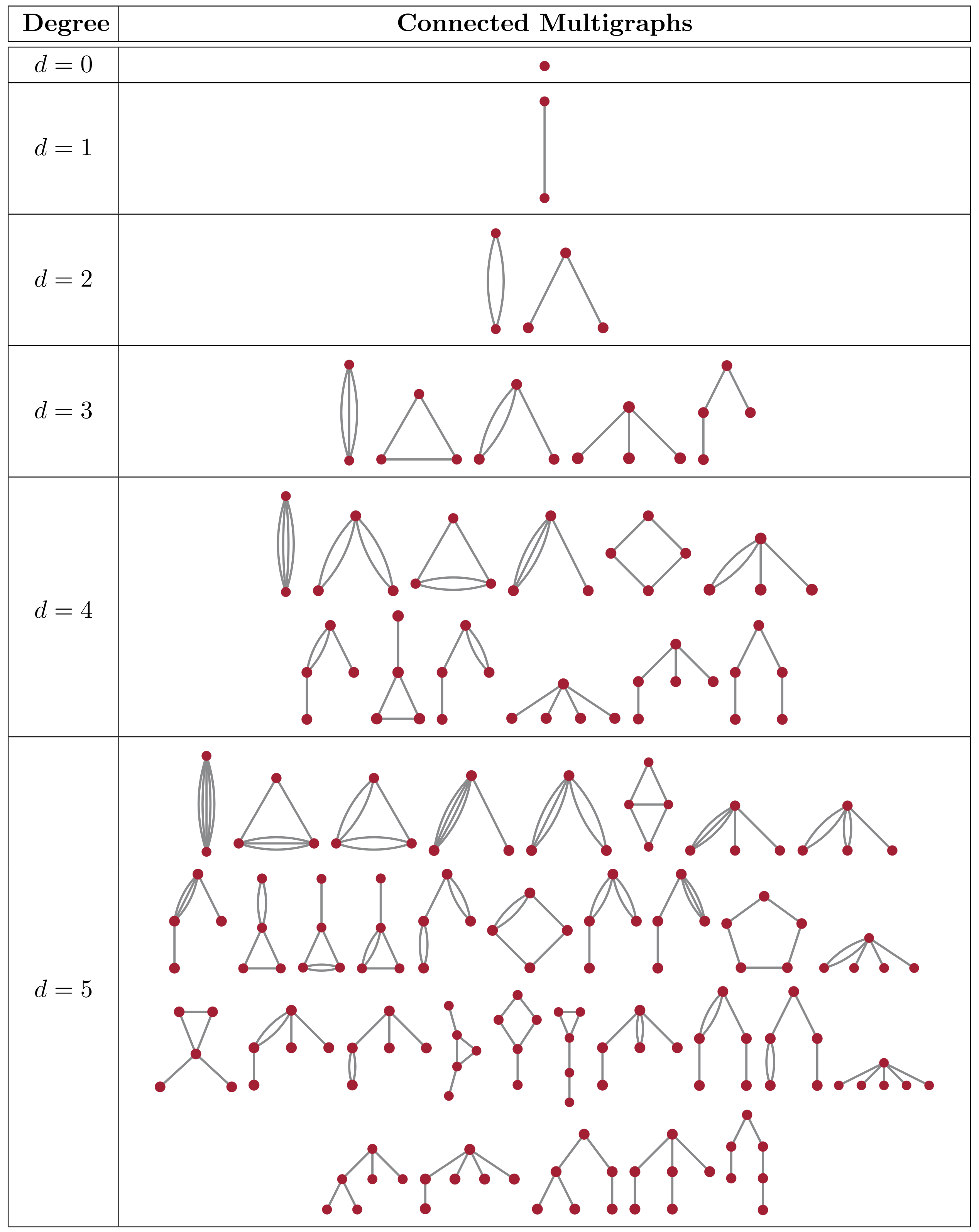
CoM N–subjettiness (Kim)

N–subjettiness (TvT)

EEC

Matrix–Element

modified mass drop
soft drop
iterated soft drop
recursive soft drop



C_n, D_n, ${}_ve_n^{(\beta)}$, M_n, N_n, U_n, EFPs

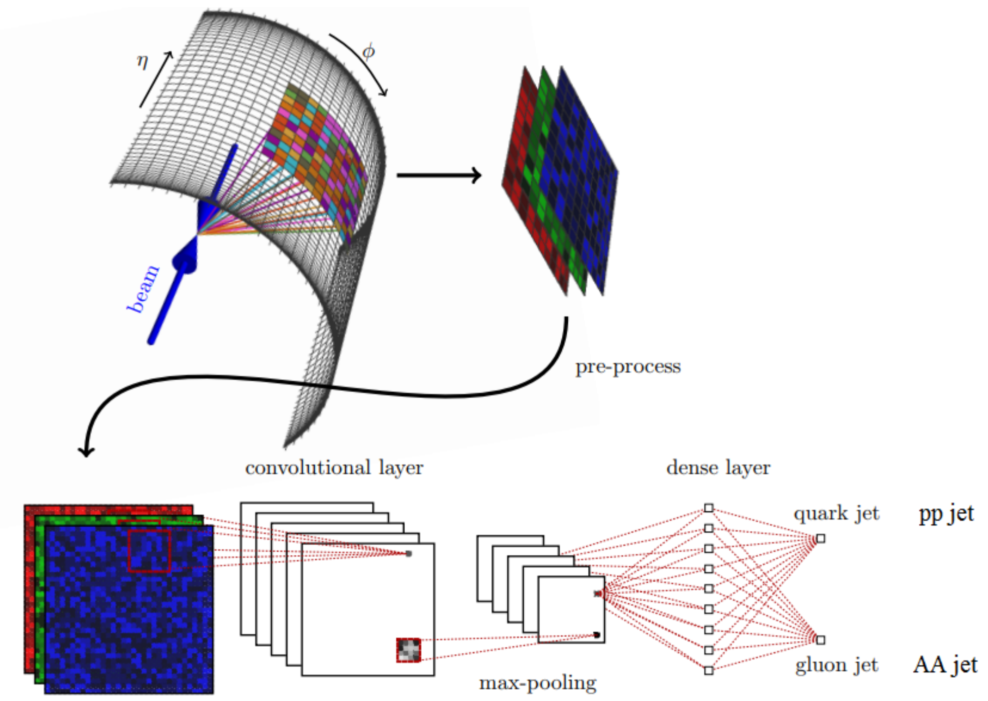
Shower Deconstruction

Qjets

Multi–variate tagger

classification without labels
weak supervision

machine learning
DNN, CNN,
RNN, LSTM, etc



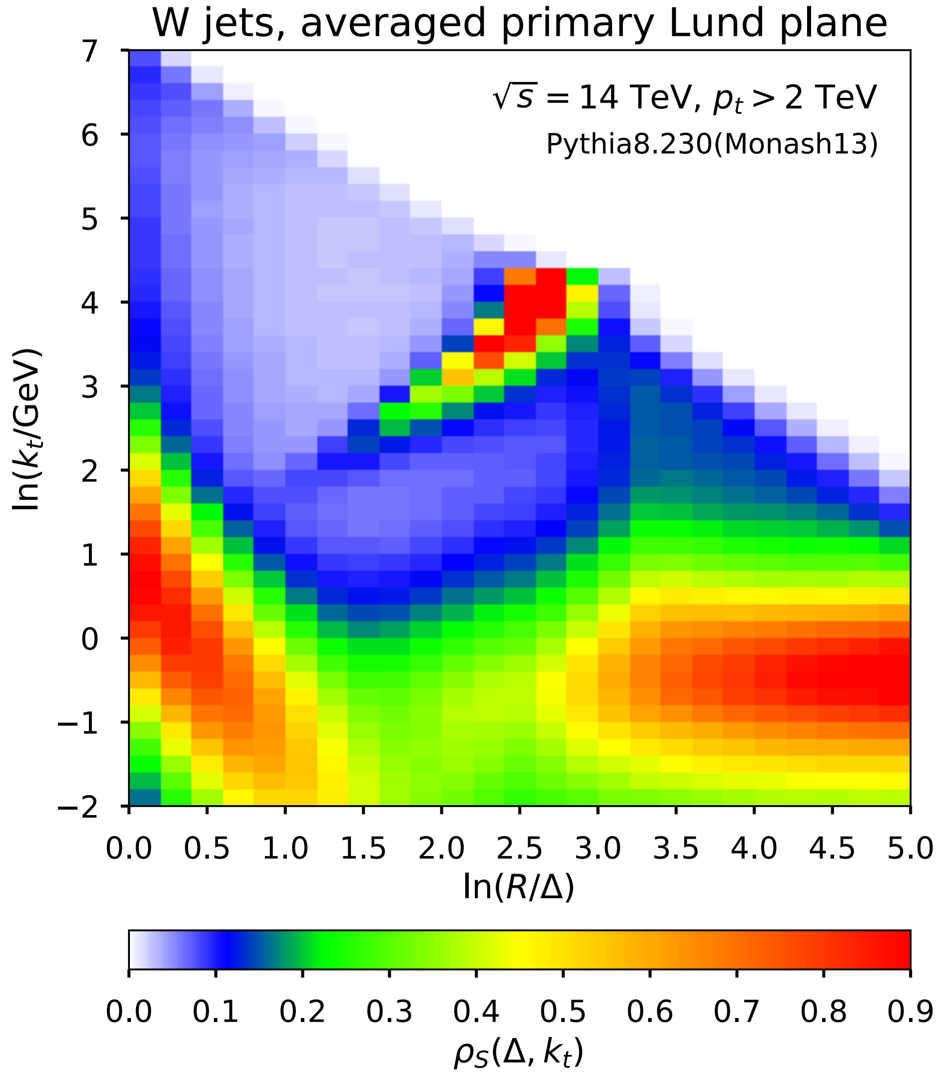
etc.

what info is deep learning using?

Boost '18

Lund images for QCD and W jets

- Hard splittings clearly visible, along the diagonal line with jet mass $m = m_W$.
- Depletion of events around W peak due to shadow cast by leading emission.



Dreyer

Recurrent networks with a Lund plane

- Jets generally associated with a clustering trees, where each node contains similar type of information.
- Particularly well-adapted for recurrent networks, which loop over inputs and use the same weights.
- For each declustering node, we consider the inputs

$$\{\ln(R/\Delta), \ln(k_t/\mathrm{GeV})\}$$

- In practice, we will use Long Short-Term Memory (LSTM) networks, which can retain dependencies over widely separated points.

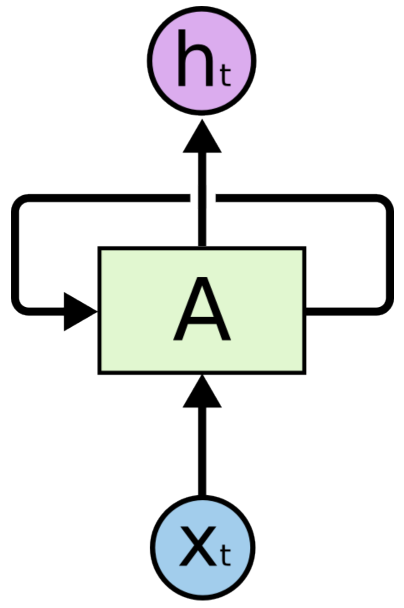
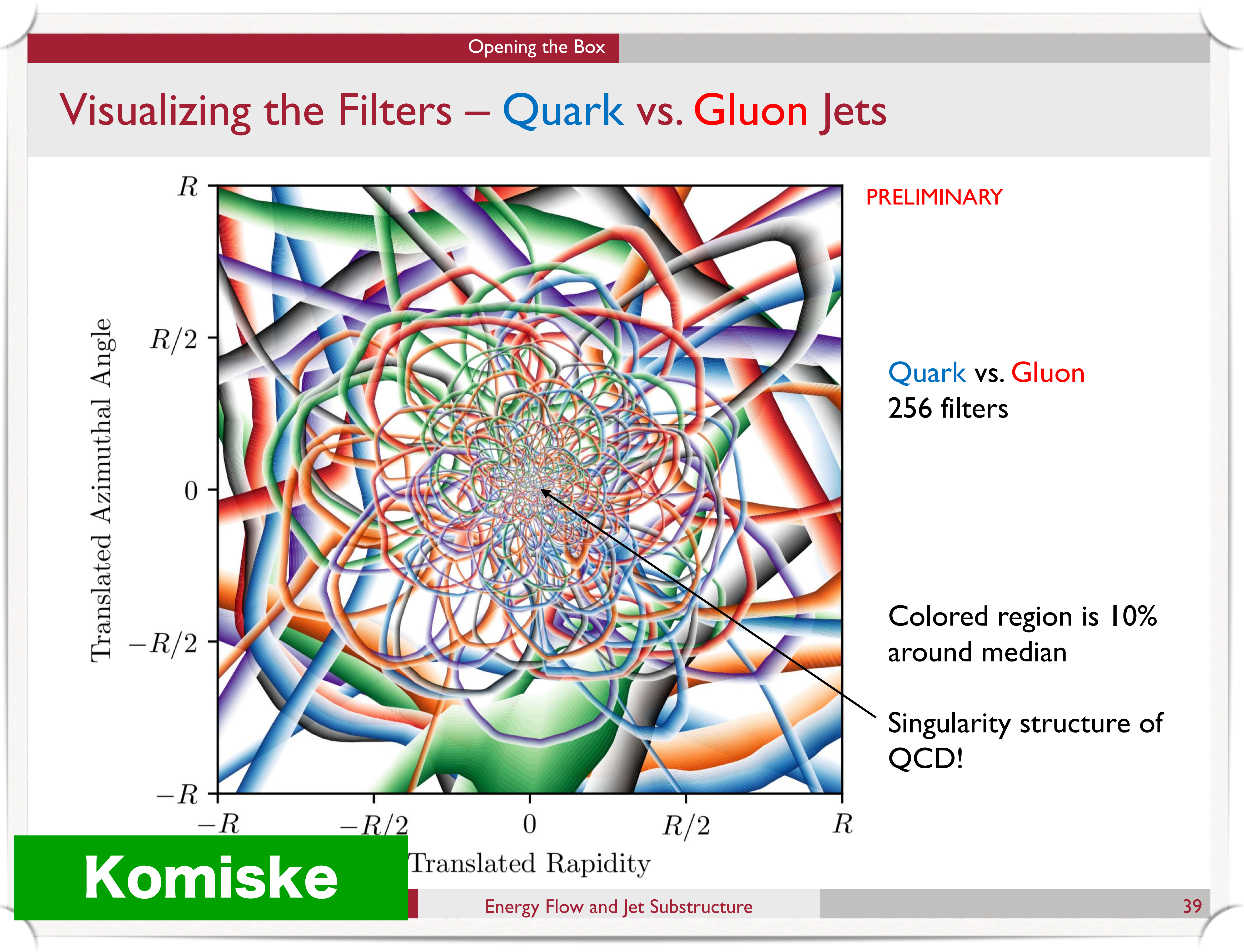


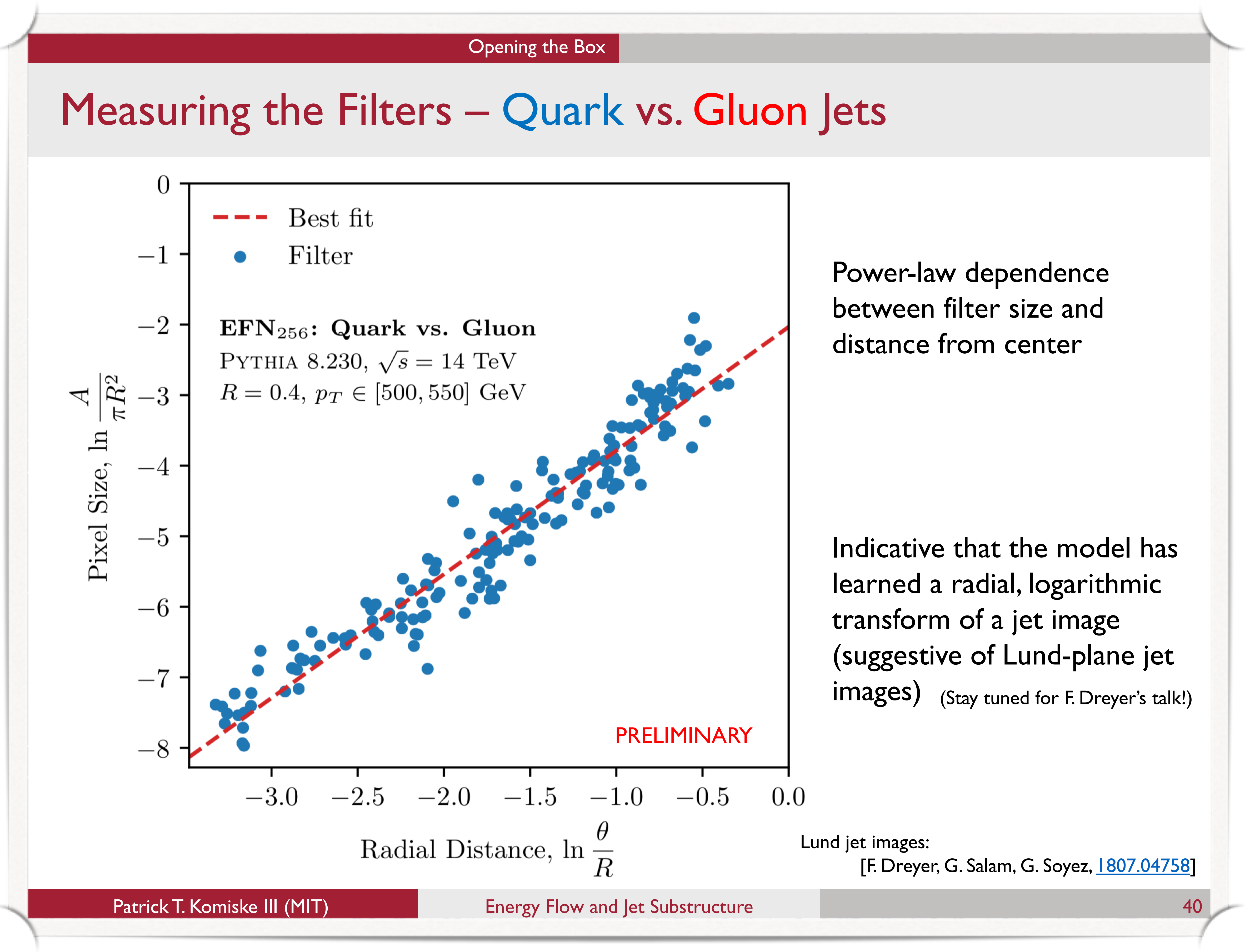
Figure from
http://colah.github.io/posts/2015-08-Understanding-LSTMs/

what info is deep learning using?

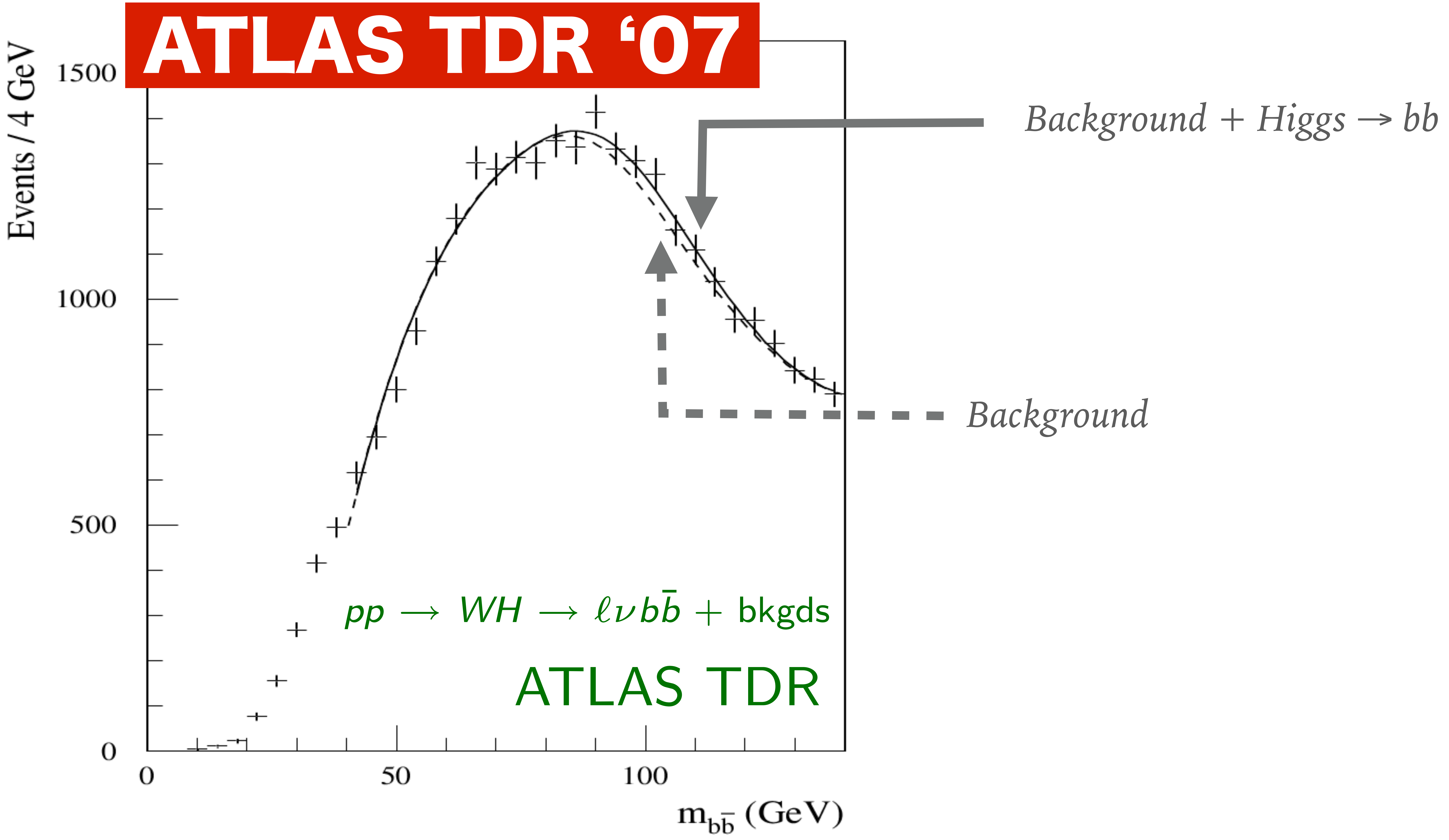
Boost '18



Komiske

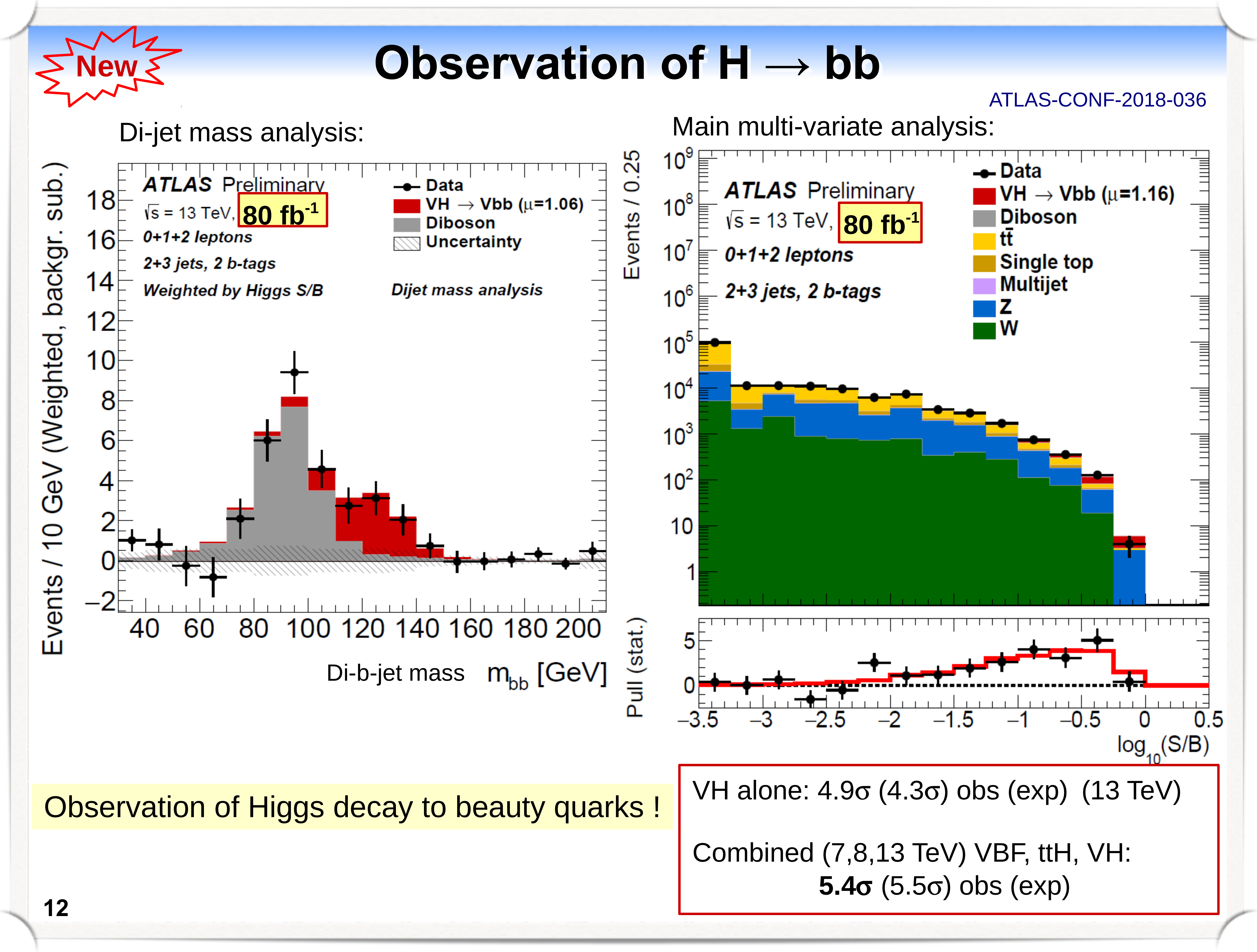


ideas from the field may get used in unforeseen ways



ideas from the field may get used in unforeseen ways

ICHEP '18



Analysis does not use substructure

But it does rely on a selecting a subsample of high-p_t events

(BDRS prediction was that you'd need ~45fb^{-1}; it turned out to be ~80fb^{-1})

Signal regions	0-lepton		1-lepton		2-lepton			
	$p_T^V > 150$ GeV, 2-b-tag		$p_T^V > 150$ GeV, 2-b-tag		75 GeV $< p_T^V <$ 150 GeV, 2-b-tag		$p_T^V > 150$ GeV, 2-b-tag	
Sample	2-jet	3-jet	2-jet	3-jet	2-jet	≥3-jet	2-jet	≥3-jet

what do we want from

BOOSTON 2019

and beyond?