Matching Matrix Elements with Parton Showers

with PYTHIA *and* HERWIG

Stephen Mrenna

Fermilab, Computing Division, Physics Simulations Group

`mailto:mrenna@fnal.gov`

⇒ Work with P. Richardson (`HERWIG`) ⇐

- Why do we need Monte Carlo?
- How will we do this?

`Disclaimer: based on my (limited) experience in RunII`

- MC will not be used to discover

  It will play a part, but not indispensable

- It will be used to understand

  It will give confidence

  Can be used for interpretation

Most asked MC questions

1. How can I estimate the "theoretical" systematic uncertainty in a MC prediction?
   - Setting a limit, measuring a physics quantity
2. How can I add different "exclusive" MC samples to make a more inclusive one?
   - i.e. 3 jets + PS $\oplus$ 4 jets + PS
   - setting a better limit, making a more powerful discovery

These are not unrelated!

Prime goal for LHC physics

Q1: Estimating "Theoretical" Uncertainty

Dissection of a MC Prediction

$$d\sigma \sim \sigma_0 H(Q) \exp\left\{-\int_{C_2 Q_0^2}^{C_1 Q^2} \frac{d\mu}{\mu}\left(A(\alpha_s)\ln\left(\frac{C_1 Q^2}{\mu^2}\right)+B(\alpha_s)\right)\right\} F_{\text{NP}}[C_1, C_2] \ \text{SGA}$$

$$+\left(\text{Fixed Order} - \text{Asymptotic}\right)[C_1, C_2] \ \text{HGC}$$

C_1, C_2 set the infrared cutoff and hard scale

SGA $\equiv$ Soft Gluon Approximation; HGC $\equiv$ Hard Gluon Correction

- "Standard" Practice is to turn off ISR (FSR) to evaluate uncertainty
  - $C_1 \rightarrow C_2(Q_0/Q)^2$ everywhere

1. HGC missing except for special (simple) cases
2. Refitting F_{NP} is no easy task (could be automated)

Ask the right questions

i. Given a physics description, how much can it reasonably vary?

ii. What is inherently lacking in the description? What approximations were made?

What theoretical uncertainty isn't

ISR on vs. ISR off

PYTHIA vs. HERWIG

How are we doing better?

MC@NLO (matching a NLO calculation to HERWIG)

Tree Level-Parton Shower Matching

Q2: Adding different MC samples

W+3 partons + PS $\oplus$(?) W+4 partons + PS

W+3 hard jets + b-tags $\equiv$ B

W+4 hard jets + b-tags $\equiv$ S

How much of “top” is W+4 hard jets? Can we use W+3 hard jets?

How do I add without over/under counting?

- In PS, (continuous) variation of topologies comes from Sudakov Form Factor (probability for no emission)
- Matrix Element calculation can be “mapped” into a PS history and reweighted with Sudakov FFs

in soft/collinear limit, recover SGA

in hard limit, apply HGC

Merging ME and PS: I

We want to use both in a consistent way

- ME gives hard/wide angle emissions
- PS gives soft/collinear emission
- Want smooth matching between the two
  - limit sensitivity to where matching occurs
- No double counting of emissions
- No under counting of emissions

X Exact NLO corrections are another story

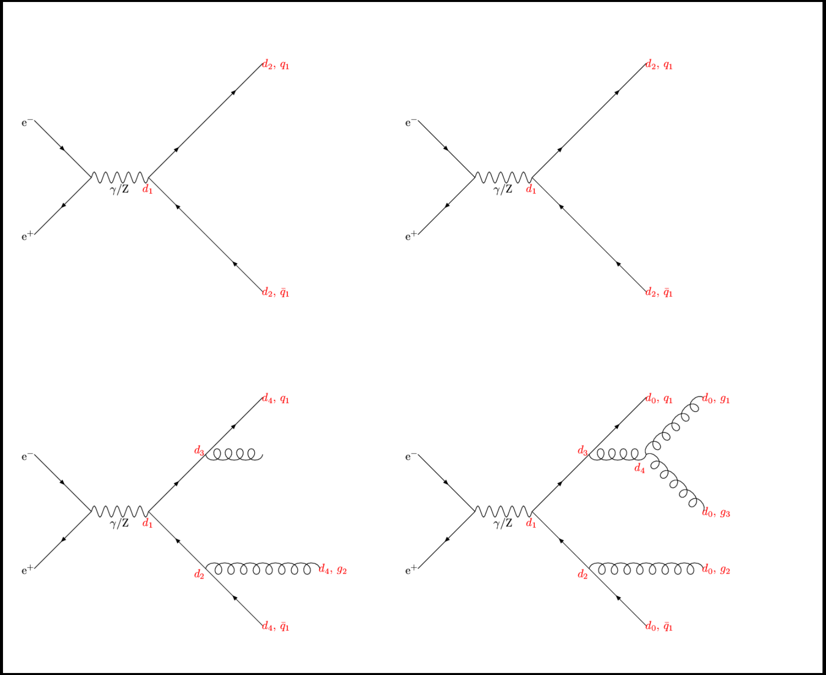
Merging ME and PS: II

- There have been a number of attempts to do this
- Hard emission corrections for relatively simple cases
  - $e^+e^- \longrightarrow q\bar{q}$
  - DIS
  - $\gamma^*/W/Z \longrightarrow$ leptons
  - Top Decay
  - PYTHIA (Sjöstrand, et al)+HERWIG (Seymour, et al)
  - Basic Idea:
    1. Rewrite (simple) ME^2 in terms of shower variables
    2. Reweight first emission to get this expression
- Only hardest (or first) emission correctly described
- Leading order normalization retained

Recent Developments

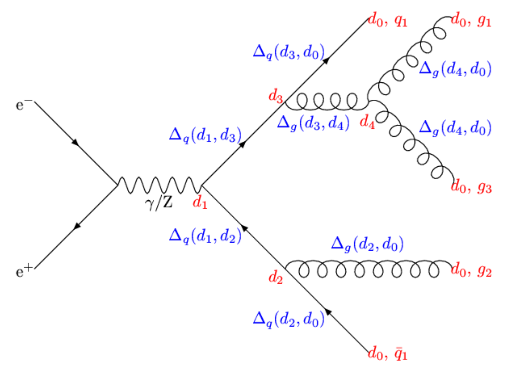
- NLO Simulation (Frixione/Webber)
    - NLO normalization of the cross section
    - Shower unchanged, but gives the correction expansion to NLO
    - Passes negative weights (but total rate is positive)
- Multijet Leading Order (Catani/Kuhn/Krauss/Webber; Lönnblad)
    - LO + NLL
    - Generalizes to many hard emissions
- Rest of talk on 2^{nd} approach

(PS) Anatomy of a Final State



- Resolve more structure as virtualities are lowered

(PS) Weight of the Final State



- Nodal values d_i represent decreases in virtuality
- Sudakov form factors $\Delta_{q,g}$ are probabilities for no emission
- $\alpha_s(d_i)P(z)$ at each splitting
- Shower is stopped at scale $d_0 \sim \Lambda_{QCD} \Rightarrow$ hadronization

The Correction Procedure of CKKW

1. Calculate tree level differential σ_n using $|\mathcal{M}_n|^2$ for $n = 0, N$
    - k_T cluster 4-vectors and require $k_{T\mathrm{min}} > d_0$
2. Use all k_T-values to give resolution values $d_1 > d_2 \ldots > d_n > d_0$
    - defines a parton shower history
3. Weight by $\alpha_S(d_1)\alpha_S(d_2)\cdots\alpha_S(d_n)$ as in PS
4. Apply a NLL Sudakov weight factor $\Delta(d_k, d_j)$ on each internal line
5. Add parton shower vetoing all radiation with $d > d_0$. Starting scale of each PS is the scale at which the particle was created.

- PS result in soft-collinear limit
- ME result in hard limit
- interpolation in between

Practical Application

- Want to do this with PYTHIA and HERWIG
  - tested, trusted, integrated
- PYTHIA and HERWIG are not k_T-ordered showers
- Sudakovs are different/numerical/conserve energy-momentum
- Kinematics within shower not the same as at the end
- Ordering in virtuality sometimes in conflict with k_T

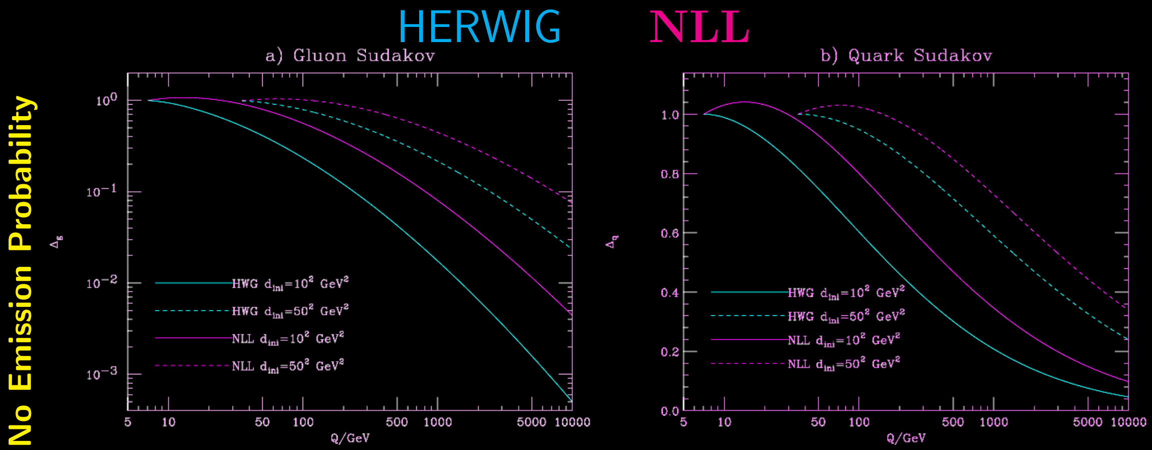
PR has continued development along CKKW lines

- Tries several scales, prefactors, minimum values to achieve stable results

I have developed an approach tailored to each generator

- Less freedom (choices made by generator)

Sudakovs

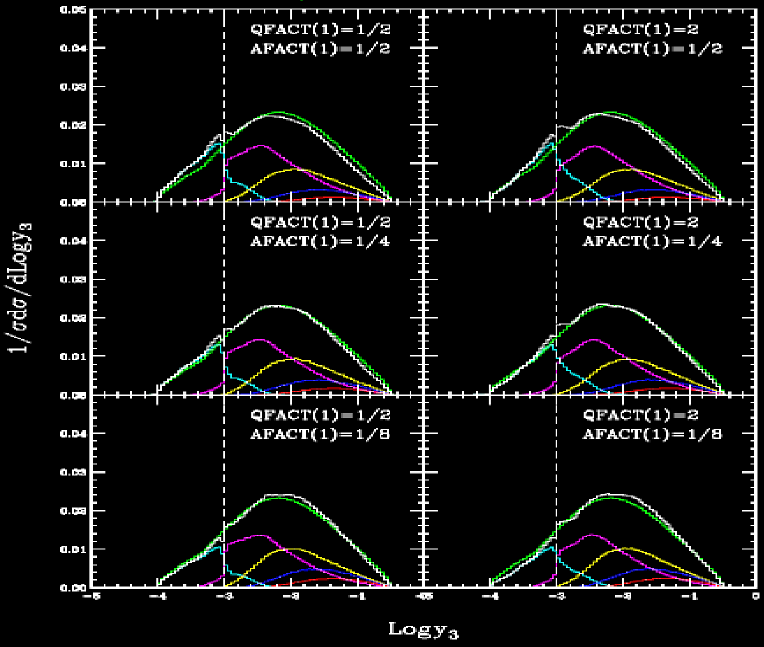


Solid: $d_0 = 10^2\ \mathrm{GeV}^2$ Dashed: $d_0 = 50^2\ \mathrm{GeV}^2$

- HERWIG has energy-momentum conservation
- HERWIG also has NLL α_S
- NLL expressions > 1

$e^+e^- \to Z \to$ **jets using** HERWIG-CKKW

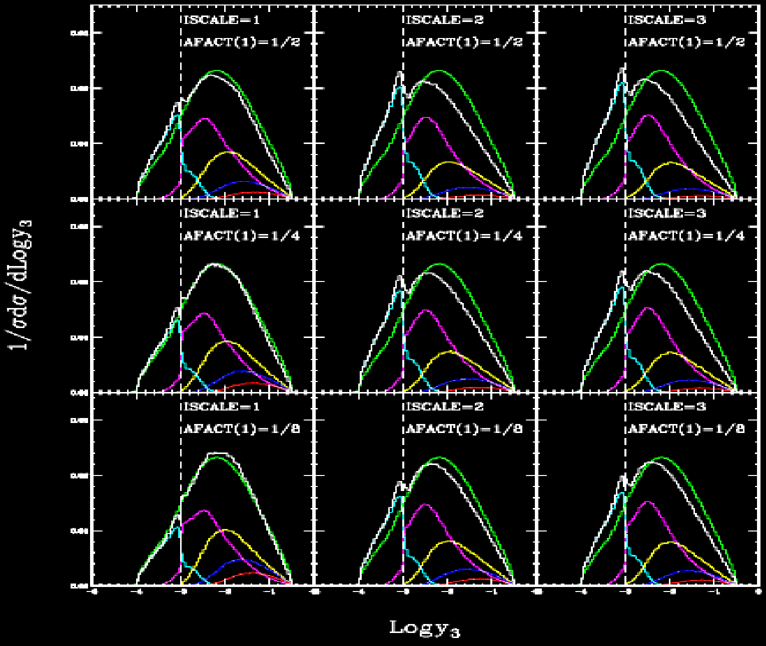
$Y_3 : \mathtt{Q}_0^2 = 2.88^2\ \mathrm{GeV}^2$



HERWIG 2 jets 3 jets 4 jets 5 jets 6 jets

Varying prefactors for scale in Sudakov form factors and α_S

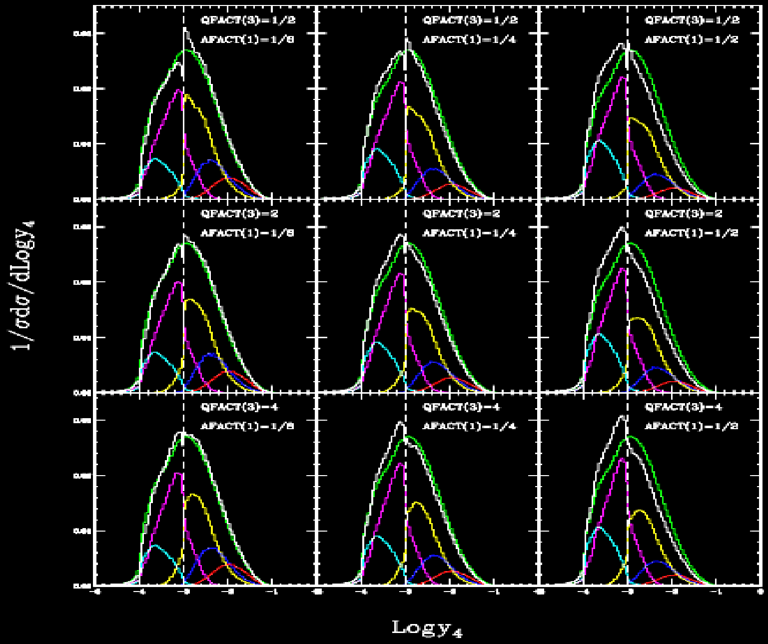
$$k_T^2 \quad p_i \cdot p_j \quad (p_i + p_j)^2$$



HERWIG 2 jets 3 jets 4 jets 5 jets 6 jets

Varying starting scale for showers

Minimum scale



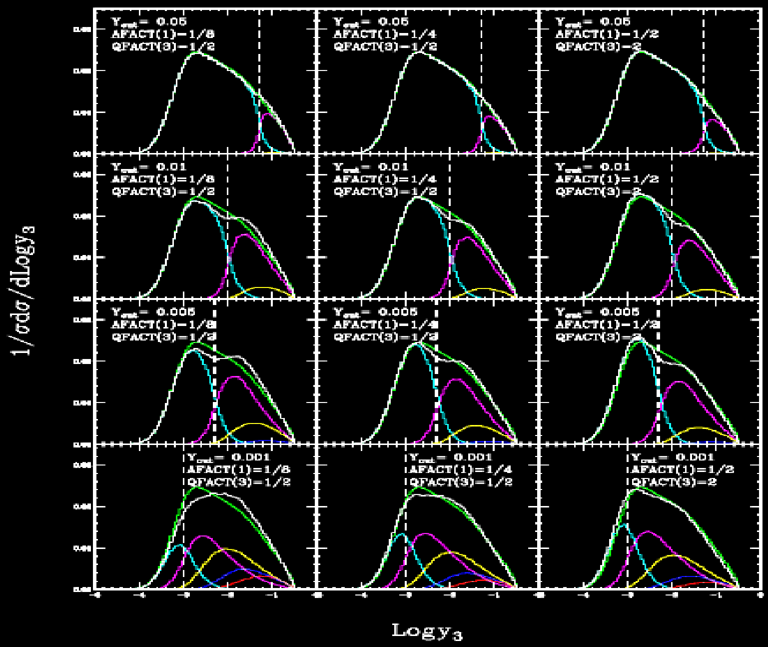
HERWIG 2 jets 3 jets 4 jets 5 jets 6 jets

Fix minimum starting scale for showers

mc4
run2

MC4Run2

HERWIG-CKKW (Hadron Level)



HERWIG 2 jets 3 jets 4 jets 5 jets 6 jets

Retuning at hadron level

HERWIG-CKKW **Summary**

- HERWIG shower is not a NLL k_T shower
    - Sudakov weights not matched
- k_T values not preserved by shower
    - Events migrate above/below cutoffs
- Need different factors for hadron observables
    - Hadronization model should be married to shower

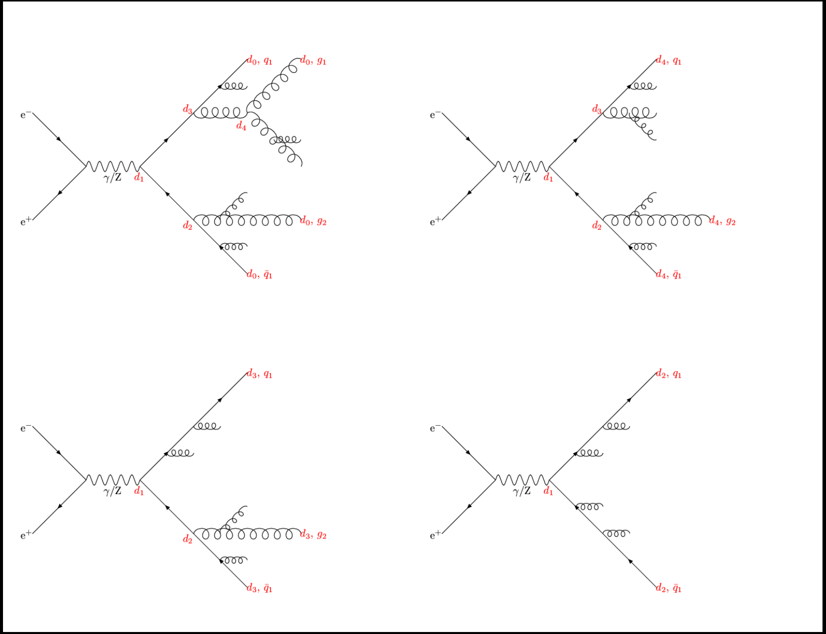
Same holds for PYTHIA

Pseudo–Shower Procedure

- LUCLUS Clustering
  - $d_{ij} = 2\left(\frac{E_i E_j}{E_i + E_j}\right)^2 (1 - \cos\theta_{ij}) = z(1-z)m^2$
- Sudakov form factor
  1. Cluster k partons using p_T scheme to get $\tilde{d}_i$
  2. PS k partons, vetoing emissions with $d > \tilde{d}_k$.
  3. Cluster again and throw away if $d_{k+1} > \tilde{d}_k$
  4. Use PS history to replace the 2 partons at scale $\tilde{d}_k$ with mother
  5. Continue until rejected or no partons left
- Choice of Scales
  - PYTHIA $= Q^2$, HERWIG $= p_i \cdot p_j$

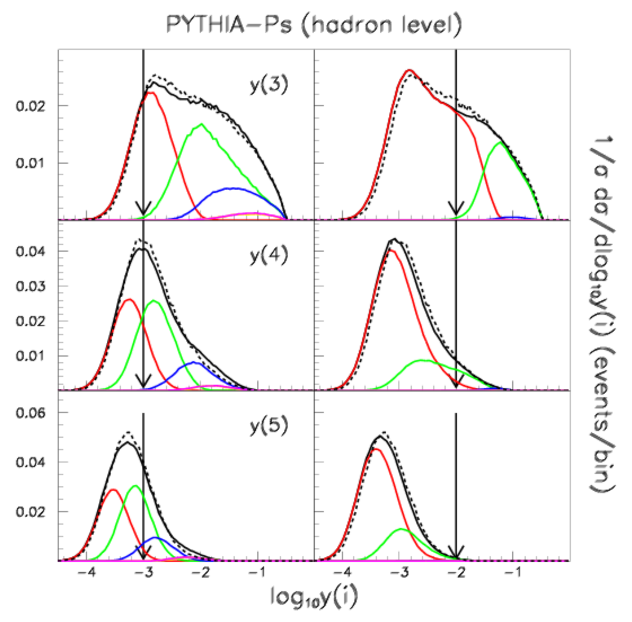
Pseudo–Showers and Sudakov Weight

Rerun the PS history and reject events with “bad” emissions



Reweighting allows smooth matching with lower topology

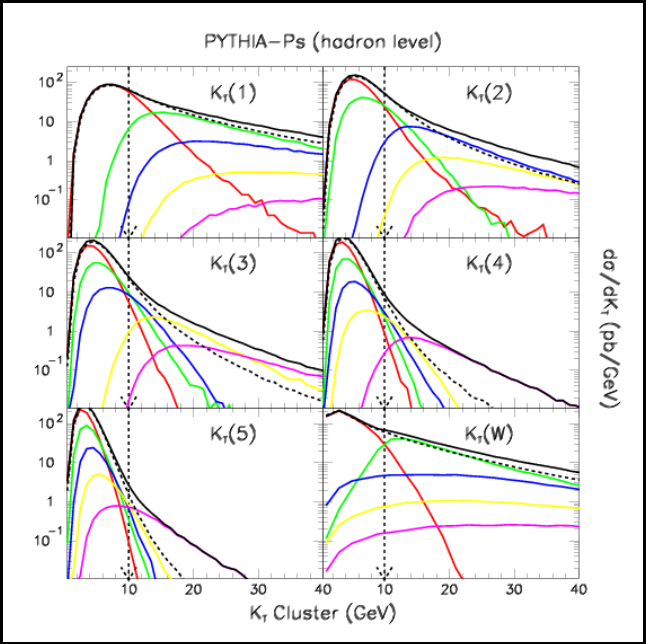
$e^+e^- \to Z \to$ jets using Ps-Sh



The matching scales: $10^{-3} \sim (2.88)^2$ GeV^2 and $10^{-2} \sim (9.12)^2$ GeV^2

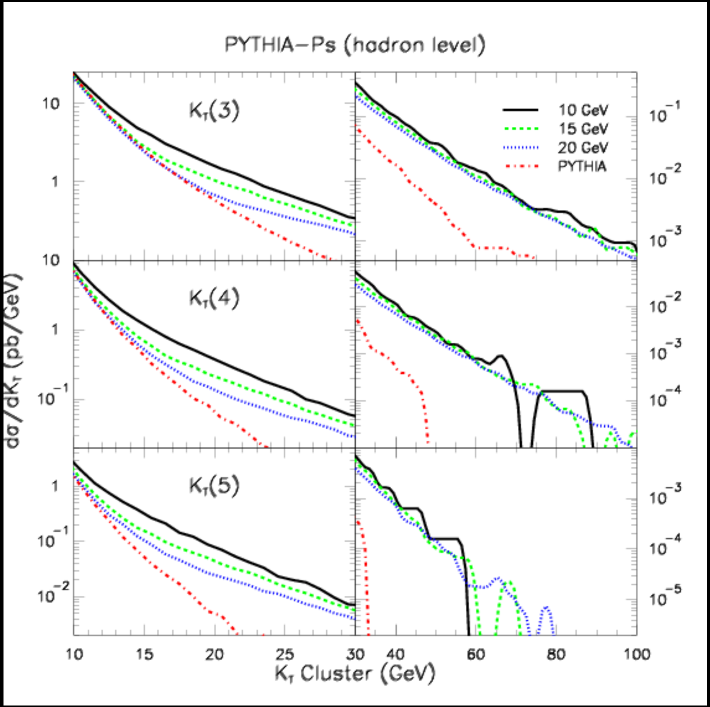
$$y = k_T^2/\hat{s}$$

W+0 $\oplus \cdots \oplus$ W+4 hard partons

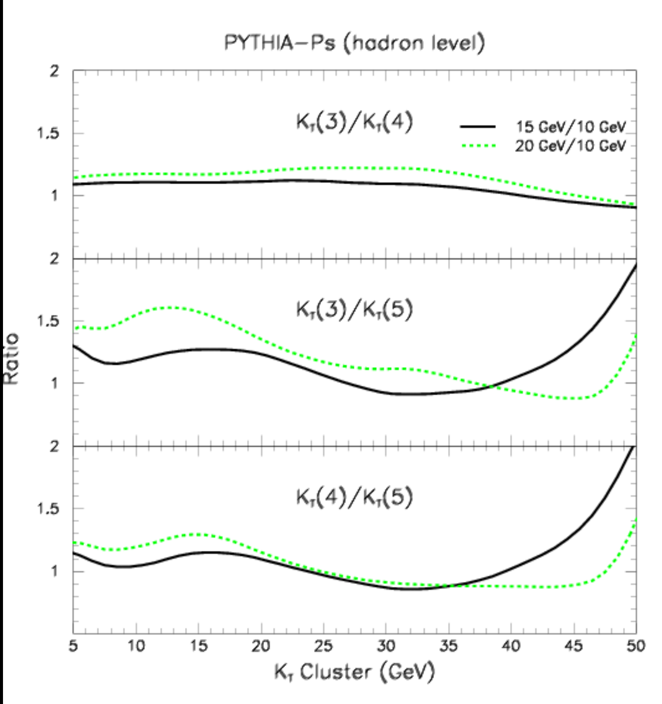


$$k_T^2 = 2\min(E_i, E_j)^2(1-\cos\theta_{ij}) \sim \min(E_i/E_j, E_j/E_i)m^2$$

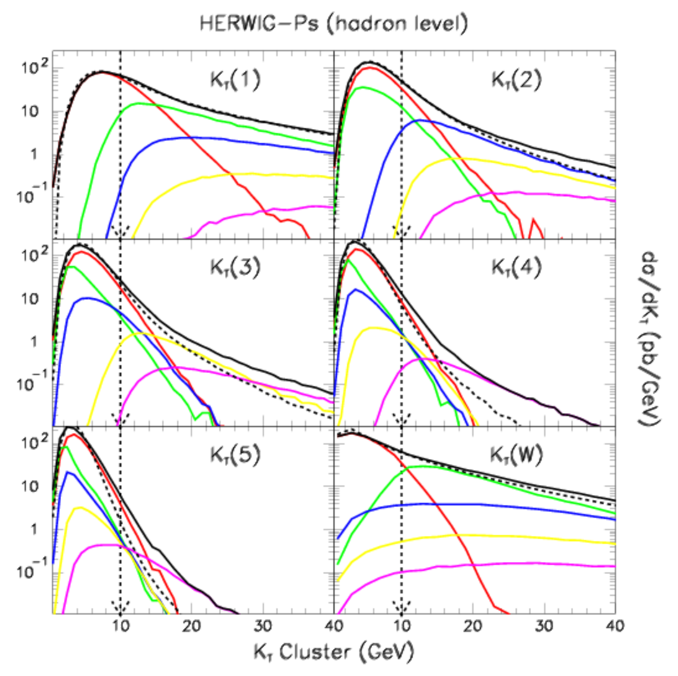
Variation with Cutoff



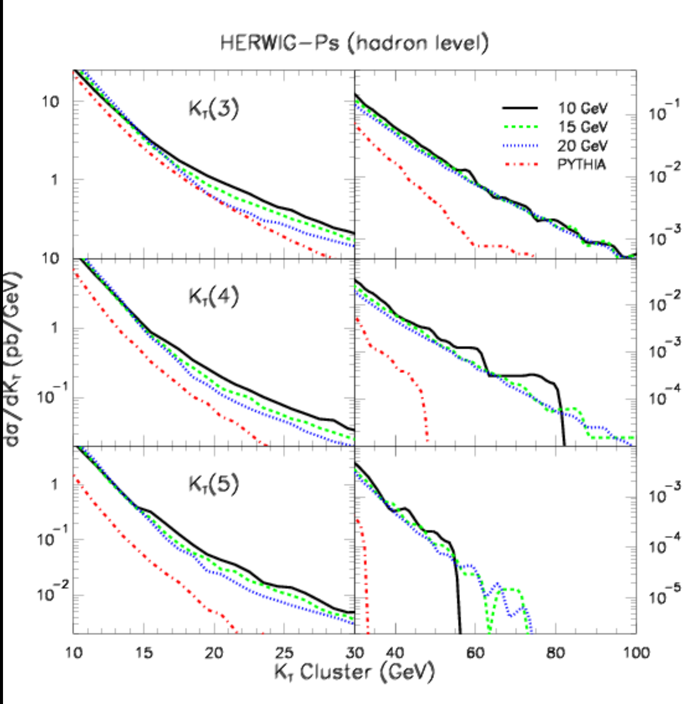
Ratios of Distributions



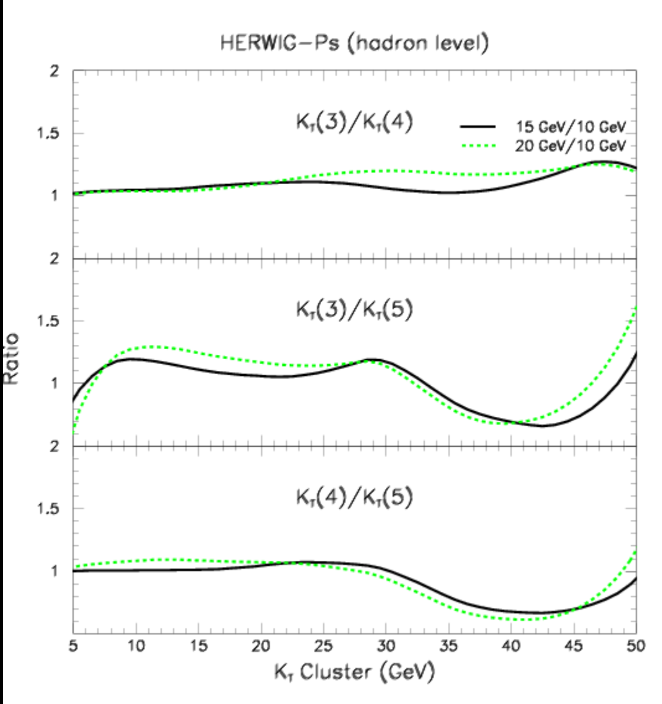
W^+ (Tevatron) (HERWIG)



Variation with Cutoff (HERWIG)



Ratios of Distributions (HERWIG)



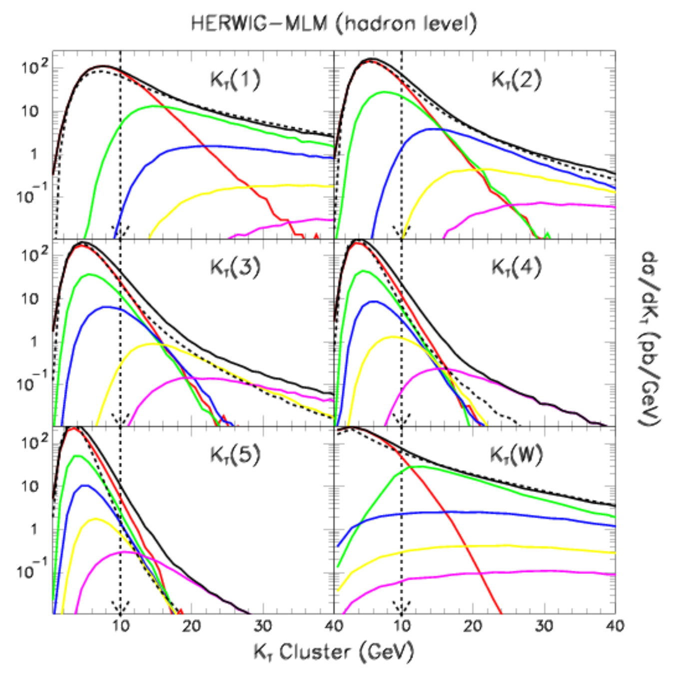
MLM method

1. Generate $W + n$ parton events of uniform weight
   - cuts on $|\eta^i| < \eta^{\max}$, $E_T^i > E_T^{\min}$, and $\Delta R_{ij} > R^{\min}$
2. Apply a PS using HERWIG
   - default scale is $\sqrt{p_i \cdot p_j}$, where i and j are color–connected partons.
3. Showered partons are clustered into N jets using a cone algorithm with parameters $E_T^{\min}$ and $R^{\min}$.
4. If $N < n$, the event is reweighted by 0. If $N \geq n$ (*inclusive*), the event is reweighted by 1 if each of the original n partons is *uniquely* contained within a reconstructed jet. Otherwise, the event is reweighted by 0.

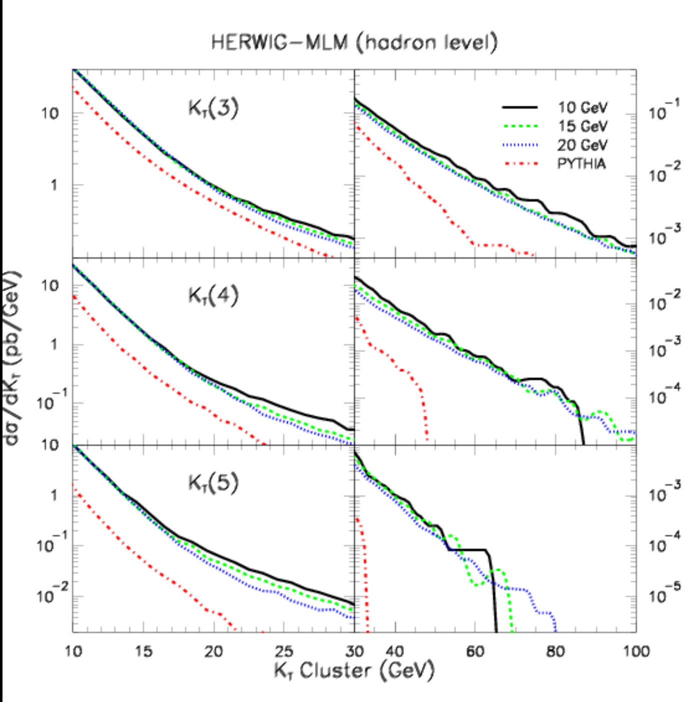
Comments

- Well motivated. It aims to prevent a PS from generating a gluon emission that is harder than any emission already contained in the "hard" matrix-element calculation.
- The cuts on E_T and ΔR play the role of cuts on k_T or p_T
- Rejection of events is like 1^{st} pseudo-shower
  - No internal Sudakovs
  - $\Delta(Q_h, Q_l)\ \alpha_s(q_T) \sim 1?$
- To make a direct comparison:
  - replace cone variables with k_T
  - Rejection replaced by $k_T^{n+1} < \tilde{k}_T^n$
  - Add together different N's

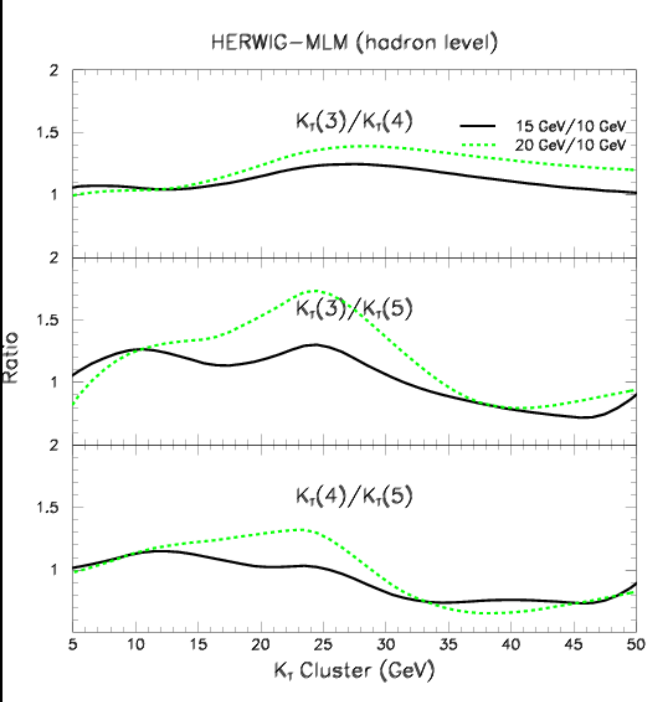
W^+ (Tevatron) (MLM-HERWIG)



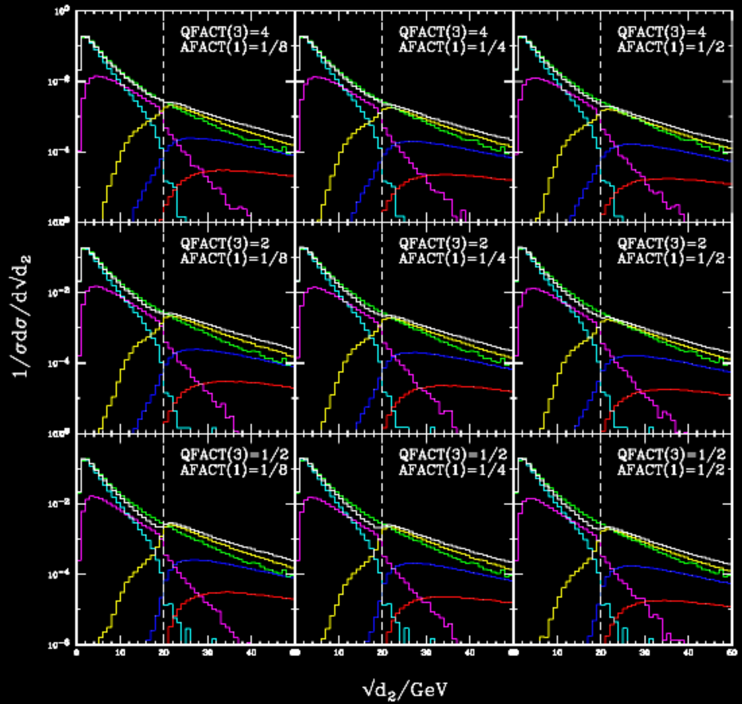
Variation with Cutoff (MLM-HERWIG)



Ratios of Distributions (MLM-HERWIG)



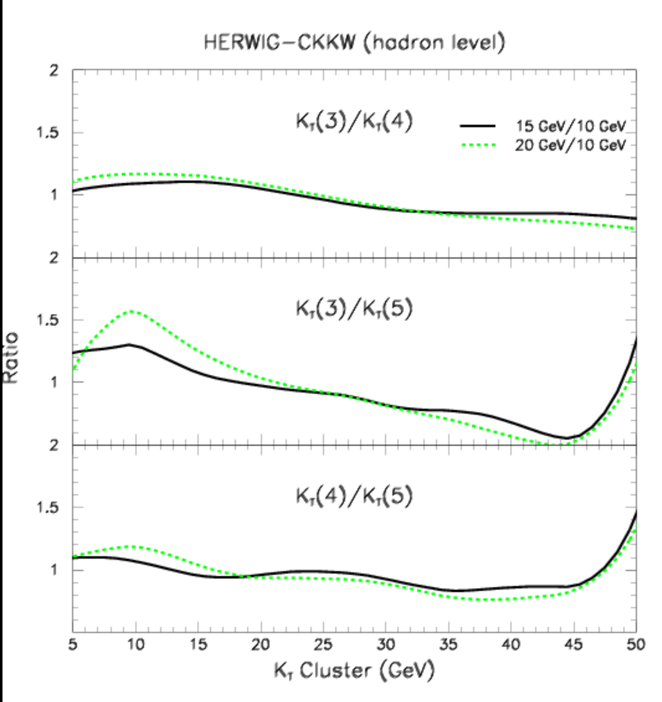
W^+ **(Tevatron)** HERWIG-CKKW



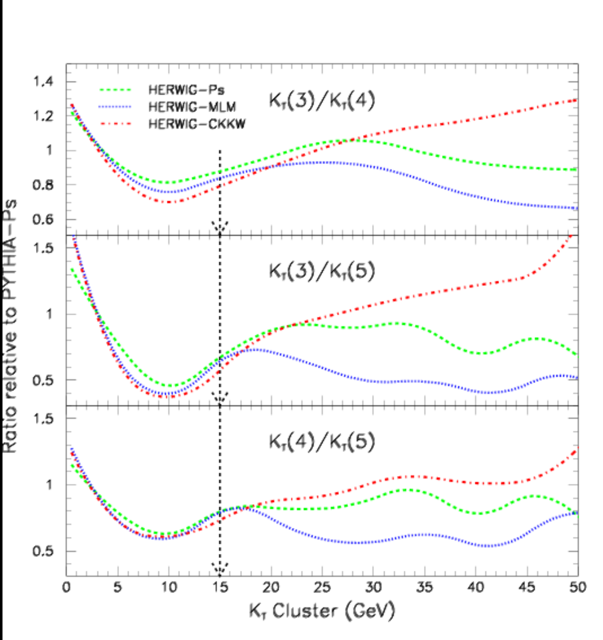
HERWIG 2 jets 3 jets 4 jets 5 jets 6 jets

(Parton Level) $\sqrt{d_2} = k_{T2}$

Ratios of Distributions (**CKKW**-HERWIG)

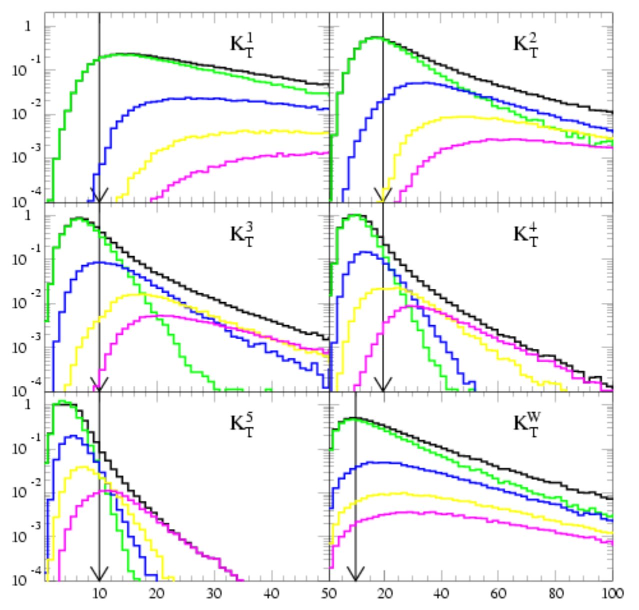


Variation of Scheme yields a Theory Error



- Variation with hard parton cutoff is also relevant
- Must test on specific observable

First Pass at $Wb\bar{b}$



- No cuts on $b\bar{b}$ (reweighted to $\alpha_s(\frac{1}{4}m^2)$)

Lessons

- These calculations are not trivial

  My conviction is that experts should do expert work

  Theory/Phenos need to carry work through to where the experiment can take over

- Those who do this work are necessary and must be supported

  given resources (computing farms, mass storage, etc.)

- Nature of these calculations begs for databases and interface with experimental software

- mass storage
- standard format for files
- writeable from a computing farm
- searchable
- reasonably safe/secure
- easy to access by experiments, theorists
- files downloadable on hits

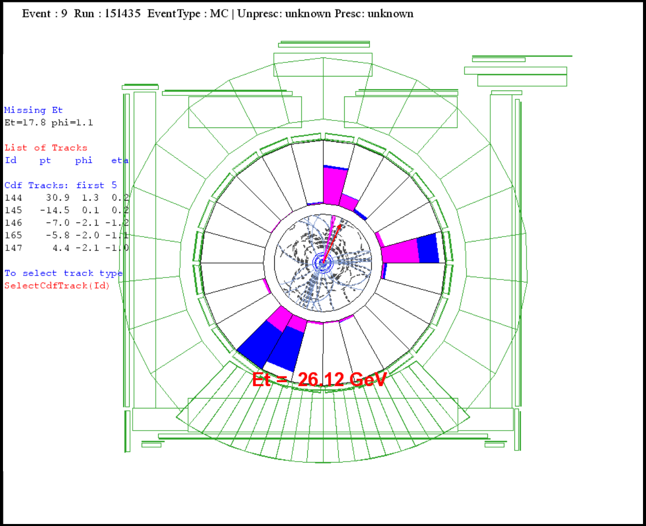
Patriot at FNAL

Physics Analysis Tools Required to Investigate Our Theories

- 1 TB Enstore repository
  - STDHEP + extra information + MCFIO
- Several different generators
  - Herwig, Pythia
  - Madgraph, Gr@ppa, CompHep, Alpgen
- Several different levels of generation
  - partons ↔ showered partons ↔ hadron level
- interface to SAM through disk cache
  - SAM ≡ Sequential data Access via Meta-data
  - Oracle database (mcdb → Oracle)

Processed Events

Theorist ⇒ Patriot ⇒ Experiment



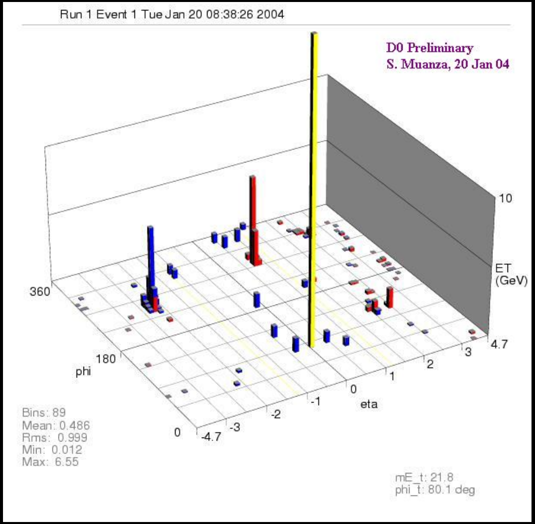
```
Event : 9  Run : 151435  EventType : MC | Unpresc: unknown Presc: unknown

Missing Et
Et=17.8 phi=1.1

List of Tracks
Id    pt    phi   eta

Cdf Tracks: first 5
144    30.9  1.3  0.2
145   -14.5  0.1  0.2
146    -7.0 -2.1 -1.2
165    -5.8 -2.0 -1.1
147     4.4 -2.1 -1.0

To select track type
SelectCdfTrack(Id)
```



Predictions

- “Theory” databases will play an important role in LHC analyses
  - new and developing MC predictions from theorists
  - quality not quantity
- Tricks will be developed to fully exploit them
  - e.g., look tables from fully simulated events to allow a quick scan of different theory predictions
- Theory/Pheno types will organize more along the lines of experimental collaborations
  - ensure that calculations are performed, legacy is maintained
- calculations will be done differently
  - Effective field theory more suitable for parton showers will be used in HO calculations

For LHC physics analyses, MC must:

- ··· give a reasonable estimate of theoretical uncertainty
- ··· be improved beyond the present level of approximation

Important for:

Setting limits
Qualifying an anomaly
Quantifying a measurement

Progress has been made

In RunII, we are learning what we need to do this
Ideas, farms, databases