Monte Carlo Tools for the Tevatron (& LHC)

Where We Stand

Stephen Mrenna

Fermilab

Les Houches 2005

Not your standard review of generators and packages

- There are many useful, important Monte Carlo tools on the market[a]
  - I will not mention them all
  - A lot of hard work behind them
  - Basic approach understood by everyone in this room
- Rather, I want to focus on 3 questions
  1. How are experiments using Monte Carlo tools?
  2. What types of tools do we have at hand?
  3. What is needed for Run2 and initial running of LHC?
- Nothing on ILC or BSM issues

[a] Sjöstrands CERN academic lectures are an excellent generator reference; see `http://www.thep.lu.se/~torbjorn/`

How experiments use "high-p_T" Monte Carlo tools

To make measurements:

- To calculate acceptances
  - only measure a slice of phase space (p_T, η, ϕ)
  - correct what you see to what was there
  - correct for what you cannot see
- Understand detector response
  - Setting the jet energy scale
    - $\gamma-$jet balancing
  - etc.
- To define physics objects
  - jets: correct for out-of-cone radiation
  - photons, leptons: define isolation
  - b quarks (e.g. mass of displaced tracks)
  - etc.

To relate measurements to physics:

- physics objects ⇒ particles ⇒partons
- Cross sections and shapes for the partons (or particles)
- Typical mode of operation
  1. Correct exclusive back to inclusive level using "LL" tools
  2. Compare to "theory" (NLO, etc.)

Event generators are main workhorses

- This counts CPU-hours, not quality
- Uses parton shower + hardonization model to connect low to high virtuality partons
- Relate hadrons (which experiments observe) to partons
- Ask infrared safe or unsafe questions

The Naked Truth

Experiments don't want to use Monte Carlo tools

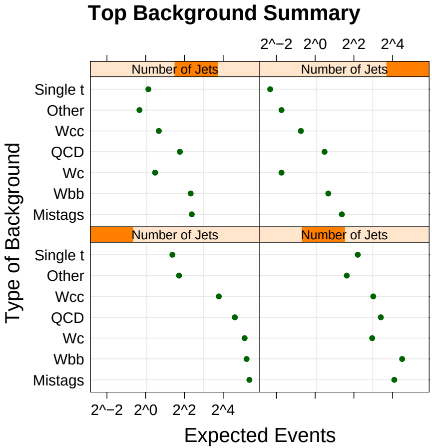
- Try to use data as much as possible
  - *Even this has some assumptions behind it*
  - It is a powerful technique to use predictions of ratios from Monte Carlo
- Perhaps, more demand now by LHC experiments for "sophisticated" tools because there is no data
- Nonetheless, the need for event generators is inescable to ask precision questions or even some simple ones

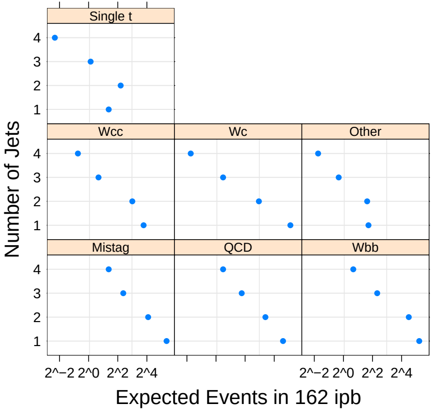
Tools for Top quark discovery
A Counting Experiment

- Primary analysis based on:
  - "LL" event generator `Isajet` without coherence and using Feynman-Field hadronization
  - Tree-level `VECBOS`
  - Supplemented with `Herwig` for cross checks and detailed kinematic analysis of top decays
- Discovery mass $= 176 \pm 8 \pm 10$ GeV, $\sigma = 6.8^{+3.6}_{-2.4}$ pb
- $(S + B)/B = (27/7, 23/15, 6/1.3)$
- Discovery easy, interpretation harder
  - The convincing evidence was the kinematic reconstruction of balancing objects with the same mass
- N.B. we may not be so lucky
  - New physics of EWSB may provide only EW-scale (fb) cross sections or it may not contain *b*-jets or something distinct

Top Properties & Single-Top

Non-Top Cocktail: CDF PRD with 162 fb^{-1}





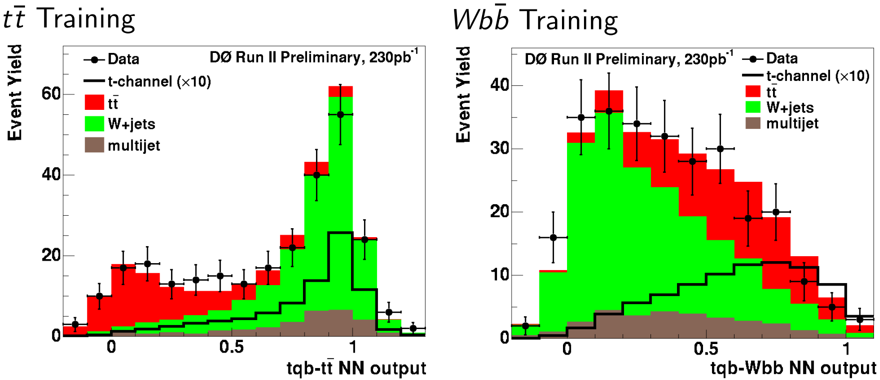
Non-trivial structure from Data+MC: is it right?

95% Confidence Level Expected/Measured Upper Limits
(after final selections, with systematics, using Bayesian statistics)

		s-channel	t-channel
Cut-Based	Electron	11.4/10.8	15.1/17.5
	Muon	13.0/15.2	18.1/13.0
	Combined	9.8/10.6	12.4/11.3
Decision Trees	Electron	6.9/7.9	9.3/13.8
	Muon	7.3/14.8	10.9/7.9
	Combined	4.5/8.3	6.4/8.1
Neural Networks	Electron	7.0/7.3	8.8/7.5
	Muon	7.0/8.7	9.5/7.4
	Combined	4.5/6.4	5.8/5.0

- current state of single-Top is where we will be at the LHC with a few quality fb^{-1}
- NN technique is currently most powerful in setting a limit
- Correlates *Many (11) Kinematic Variables*
  - it challenges our tools

	Signal-Background Pairs			
	tb		tqb	
	Wbb	$t\bar{t}$	Wbb	$t\bar{t}$
Individual object kinematics				
$p_T(\text{jet1}_{\text{tagged}})$	√	√	√	—
$p_T(\text{jet1}_{\text{untagged}})$	—	—	√	√
$p_T(\text{jet2}_{\text{untagged}})$	—	—	—	√
$p_T(\text{jet1}_{\text{nonbest}})$	√	√	—	—
$p_T(\text{jet2}_{\text{nonbest}})$	√	√	—	—
Global event kinematics				
$M_T(\text{jet1,jet2})$	√	—	—	—
$p_T(\text{jet1, jet2})$	√	—	√	—
$M(\text{alljets})$	√	√	√	√
$H_T(\text{alljets})$	—	—	√	—
$M(\text{alljets} - \text{jet1}_{\text{tagged}})$	—	—	—	√
$H(\text{alljets} - \text{jet1}_{\text{tagged}})$	—	√	—	√
$H_T(\text{alljets} - \text{jet1}_{\text{tagged}})$	—	—	—	√
$p_T(\text{alljets} - \text{jet1}_{\text{tagged}})$	—	√	—	√
$M(\text{alljets} - \text{jet}_{\text{best}})$	—	√	—	—
$H(\text{alljets} - \text{jet}_{\text{best}})$	—	√	—	—
$H_T(\text{alljets} - \text{jet}_{\text{best}})$	—	√	—	—
$M(\text{top}_{\text{tagged}}) = M(W, \text{jet1}_{\text{tagged}})$	√	√	√	√
$M(\text{top}_{\text{best}}) = M(W, \text{jet}_{\text{best}})$	√	—	—	—
$\sqrt{\hat{s}}$	√	—	√	√
Angular variables				
$\Delta R(\text{jet1,jet2})$	√	—	√	—
$Q(\text{lepton}) \times \eta(\text{jet1}_{\text{untagged}})$	—	—	√	√
$\cos(\text{lepton}, Q(\text{lepton}) \times z)_{\text{top}_{\text{best}}}$	√	—	—	—
$\cos(\text{lepton}, \text{jet1}_{\text{untagged}})_{\text{top}_{\text{tagged}}}$	—	—	√	—
$\cos(\text{alljets}, \text{jet1}_{\text{tagged}})_{\text{alljets}}$	—	—	√	√
$\cos(\text{alljets}, \text{jet}_{\text{nonbest}})_{\text{all jets}}$	—	√	—	—



- Relies upon the Non-Top Cocktail
- How do we convince ourselves of a signal?
- How can we improve upon the search?

Mixing the Cocktail

Method 2

Monte Carlo ratio
$R = (W + b - jets)/(W + jets)$

- Common factors cancel

Measure $W + jets$ (no b-tag)

data$(W + b - jets) = R \times$data$(W + jets)$

Wcj/Wbb from Monte Carlo

- Several R's

Tools

- Tree-Level
- Parton-shower
- NLO-Level
- Combinations of these

Matrix Elements + Parton Showers

MLM Method

Parton shower and hadronization are essential for studying b-jets

- Parton shower W+Npartons but reject emissions that are too hard (i.e. each post-shower jet should have a pre-shower parton associated with it)
- Build up *inclusive* or *exclusive* samples (i.e. allow or disallow pure PS jets)
- $\delta R/R$ ~25-30%

Heavy Flavor (HF)

Early LEP, Run1 ⇒ PS underestimates HF

(Modifications were made to bound LEP results)

- $P_{qq}(z) = \frac{1}{2}(z^2 + (1-z)^2)$

  no soft ($z \to 0$) enhancement i.e., subleading log in PS
- Use ME with $b\bar{b}$ explicit
- R supplemented by phenomenological factor 1.5 from HF in QCD sample

Method 2 at Tree Level

Madevent (Stelzer and Maltoni)

X-Check

Graph	Cross Sect(fb)
Sum (Wbb)	8.934
Sum (Wjj)	1061.627
$ug \rightarrow e^+ \nu_e dg$	327.810
$u\bar{d} \rightarrow e^+ \nu_e gg$	257.060
$g\bar{d} \rightarrow e^+ \nu_e \bar{u}g$	137.300
$\bar{d}g \rightarrow e^+ \nu_e \bar{u}g$	48.591
$u\bar{u} \rightarrow e^+ \nu_e \bar{u}d$	47.425
$u\bar{d} \rightarrow e^+ \nu_e d\bar{d}$	36.644
$gu \rightarrow e^+ \nu_e dg$	34.445
$u\bar{d} \rightarrow e^+ \nu_e u\bar{u}$	29.816
...	...

$90 < M_{jj} < 110$ GeV, standard jets

$R \times 1.5 = 1.3\%$ (MLM $= 1.4\%$)

- $\langle R \rangle$ roughly the same

Many different topologies

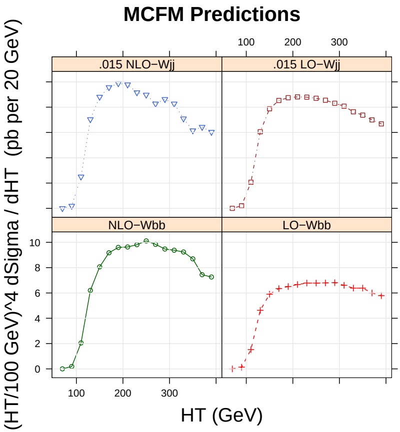
Dominant ones not $q\bar{q}$

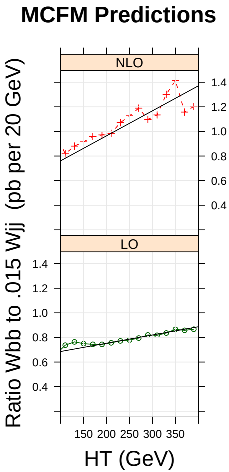
- again, no $z \rightarrow 0$ enhancement

Different topologies parton shower and hadronize differently

Many effects have to be modelled well to have a reliable prediction

Testing Method 2 with other Tools





Significant change in normalization and shapes LO ⇒ NLO

Top Mass Measurement Uncertainties

- CDF has significantly reduced uncertainty on JES
- Precision of our tools is getting important

What is the Uncertainty from Event Generators?

- `Pythia` – `Herwig`?
  - What if one doesn't have a validated model of the underlying event?
- `Madevent` – `Pythia`?
  - Can't run non-showered partons through hadronization
- Turning off/on ISR/FSR?
  - Un-showered partons have poor jet shapes
- Change some scales in parton shower?
  - CDF varied p_T of ISR/FSR emissions (TS,SM)
  - More scientific than previous approaches

- In general, a fair number of parameters are fit to data
- Most of these are related to the phenomenological model of hadronization
- For a given choice of Parton Shower, most sensitive parameters deal with the shower-hadronization transition and scale used in α_s

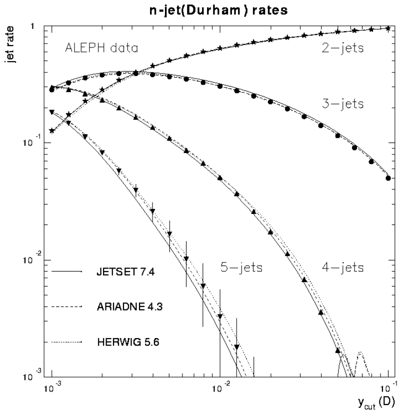
Range of $Q_0, \Lambda_{\rm LLA}$ gives approximate picture of our understanding of FSR in resonance production (numbers in GeV)

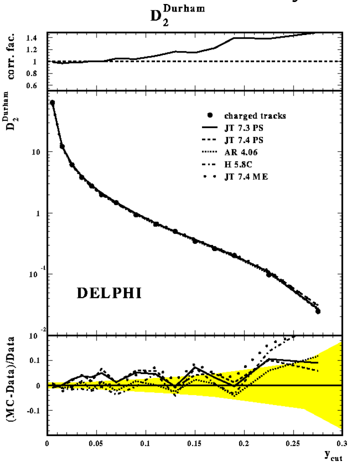
Parameter	Name	Default	Aleph	Delphi	L3	Opal
Shower $\Lambda_{\rm LLA}$	`PARJ(81)`	0.290	0.320	0.297	0.306	0.250
Shower Cutoff Q_0	`PARJ(82)`	1.00	1.22	1.56	1.0	1.9

Good description of complex topologies in e^+e^-

$R_2(y)$

$$D_2 = \frac{R_2(y) - R_2(\Delta y)}{\Delta y}$$





In general, `Ariadne` > `Pythia` > `Herwig`

Quality of fit

Pythia 6.3

Distribution	bins	$\sum \chi^2$ of model PY6.3 $p_\perp$-ord.	PY6.1 mass-ord.
Sphericity	23	25	16
Aplanarity	16	23	168
1−Thrust	21	60	8
Thrust$_{\rm minor}$	18	26	139
jet res. y_3(D)	20	10	22
$x = 2p/E_{\rm cm}$	46	207	151
$p_{\perp \rm in}$	25	99	170
$p_{\perp \rm out} < 0.7$ GeV	7	29	24
$p_{\perp \rm out}$	(19)	(590)	(1560)
x(B)	19	20	68
sum $N_{\rm dof} =$	190	497	765

- $\chi^2/\text{dof} > 1$
- "theory" T should have a systematic error qT
- $\chi^2 = \dfrac{(O-T)^2}{(\sigma_O^2 + (qT)^2)}$

q	0%	0.5%	1%
$\sum \chi^2$	523	364	234

for $q = .01$, $\chi^2/\text{dof} = 234/196 \sim 1 \Rightarrow$ generator good to 1%

except $p_{\perp \rm out} > 0.7$ GeV (10%–20% error)

$p_{\perp \rm out} \equiv$ one-particle inclusive $p_\perp$ spectrum out of the event plane

- problem for all generators

Parton Shower

- overall LEP description very good
- some systematics remain, but nothing serious
- needed improvements to $g \longrightarrow b\bar{b}$

In hadronic environment, PS just now being seriously tested

- Phase space different from LEP
- more serious questions leads to development
- Pythia6.3: p_T-ordered shower $\Leftrightarrow$ Underlying Event Model
- experimentalists want good & flexible tools

Tree Level Tools
Automated Evaluation

MadEvent,Alpgen,CompHep,Grace-based

- powerful tools
- can make quick cross checks
- can test ideas
- incorporate spin correlations
- NOT event generators
- essential to add PS, should make it easy
  - want to write files
  - common format
  - collection of metadata
    - (what will you want to know 10 years from now?)
  - deficiency of LHA

Loop Level Tools
Histogram fillers, Not Event Generators

Experiments will degrade the theory before they correct the data

- operation Theory ⇒ Detector-level may not even have an inverse
- Non-Exclusive predictions require some work for comparison with data
- Jet Physics

  `Jetrad`, `EKS`, `NLOJET++`
  `NLOJET++` flexible for fastNLO calculation
- $\gamma(\gamma)$ Physics

  `diPhox` (NLO frag), `ResBos` (resummed)
  (`Pythia` (PS frag))
- Precision EW

  `Resbos`: ($\theta_{\rm CS}$), `WZGrad`
  NNLO y used for reweighting `Pythia`
- Top Backgrounds

  `MCFM` for ratios of unmeasured b/c, b/j

Marriages of Existing Tools, e.g. MC@NLO

Just brief words (Stefano will cover this)

- Basic idea is to add in and subtract out PS result to NLO
- Asymptotic piece of NLO calculation related to PS limit truncated to the same order

  Think of how NLO calculation ⇒ resummed result in CSS

  $$\frac{d\sigma}{dp_T^2 dy} \sim W_{\mathrm{FT}} + Y$$

  Fixes deficiencies of `Herwig` shower
- Including angular correlations
- Box diagrams? NNLO? NLO $Wjjjj$?
- Real piece may acquire its own NLO correction
- Alternatives

  add ME corrections to PS

  Kramer and Soper (e^+e^-), etc.

Another Example

Matrix Element-Parton Shower Matching

- Originated by MLM for Tevatron physics
- Systemized by CKK&W for e^+e^-, also LLö
- Developed for hadronic colliders by PR, SM

  Further developed for "any" parton shower (SM)
- Currently, several other groups now doing similar things

  `Sherpa` team (analytic Sudakovs and `Pythia` (mass) showers)

  `Ariadne`

  MLM (predates all this and very active)

  - hard-wiring options into `Alpgen`

  SM (second generation with p_T-ordered shower)
- Comparisons in PR/SM showed that all are reasonable and should give similar results

  Differences at the level of the logarithms we do not correctly control

Review of Mr Matching

Pseudo-Shower Method

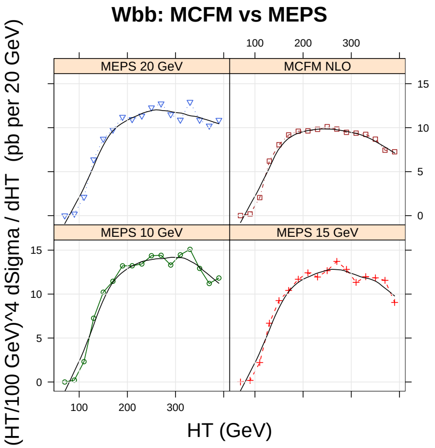
1. Generate $W + N$ parton events, applying a cut ${p_T}^2_{cut}$ on shower p_T^2 (p_T^2 for ISR, $z(1-z)m^2$ for FSR)
2. Form a p_T^2-ordered parton shower history
3. Reweight with $\alpha_s(p_T^2)$ for each emission
4. Add parton shower and keep if no emission harder than $p_{T\,cut}^2$: (save this event)
5. Remove softest of N partons, fix up kinematics, add parton shower and keep if no emission harder than $p_{T\,softest}^2$
6. Continue until no partons remain, or an emission is too hard
7. If not rejected, use the saved event

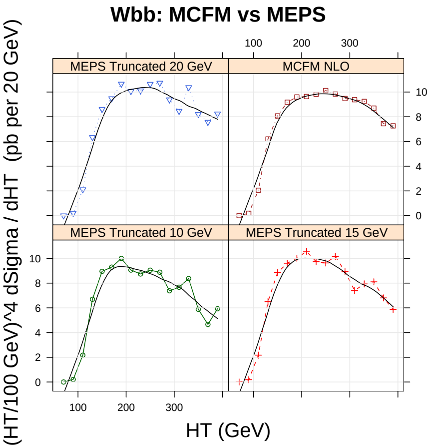
Why it works

- For each N, PS does not add any jet harder than $p^2_{T\,cut}$
- Can safely add different N samples with no double-counting
  - Apply looser rejection on highest N
- Pseudo-showers assure correct PS limit, while retaining hard emissions
  - Interpolates between limits

Why it is necessary

- Suppress unphysical enhancements in tree level calculations from $\alpha_s^n(p_T)\ln^{(2n,2n-1)}\left(\frac{Q}{p_T}\right)$
- Account for many topologies in physical observables, e.g. $H_T = \sum p_T(\text{hard object})$





Matched Datasets have a systematically larger rate and different shape

Truncated Datasets contain only $Wb\bar{b} + Wb\bar{b}j$

HO topologies modify shape

Event Generators in OO Languages

G.Booch OO Analysis & Design 2nd ed. Pag. 13

A complex system that works is found to have evolved from a simple system that worked. A complex system designed from scratch never works and cannot be patched up to make it work. You have to start over, beginning with a working simple system.

- `Pythia` & `Herwig` decided a complete rewrite was necessary in a modern language that CERN would support
- Now beyond the simple (but non-trivial) proof-of-principle stage
  - QCD FSR, QCD ISR, particle decays, etc.
  - Improvements to showers, accounting of particle properties, couplings
- Herwig-f77 frozen, Pythia-f77 evolving: primary tools at Tevatron
- `Herwig++` "will be ready for LHC"; `Pythia8` likely same
- `Sherpa-C++` is released
  - overlap with some `Pythia` physics assumptions
- long road of tuning and validation ahead

LHC Differences from Tevatron

- More statistics for calibrating backgrounds

  $Z \to e^+e^- jj \Rightarrow Z \to e^+e^- b\bar{b} \Rightarrow W \to e\nu b\bar{b}$
- Many EW processes are relevant

  gg-induced backgrounds to Higgs can look more like signals
- Enormous phase space for complicated topologies

  Are all logarithms under control?

  $$\frac{f(x,p_T)}{f(x,v)}\alpha_s(p_T)\ln^2\left(\frac{v}{p_T}\right)$$

  Electroweak (20% of QCD at high p_T?)

  Large x
- I have certainly missed something.

Tools to complete physics program at Run2 and start LHC are here

Better tools will always improve the results

Are There Lessons Here?

- Some (all?) N(N?)LO calculations should be started soon if they are going to happen and be relevant
  Hard to say unequivocally what is “needed”
  Discovery may drive this
  Need to keep fit and train students
- Worth thinking about code structure (more than language!)
  Will it generate events?
  Can it create files of predictions?
  Is it easy to do PDF uncertainty calculations?
- Worth investing time to make theory uncertainties scientific
  Checks for random bugs (*theorist* error) $\neq$ theory uncertainty

Tool Developments I would like to see at LH

- Liberate `MC@NLO` from `Herwig`

  Consensus on alternative approaches

- Marry PS Matching approaches with NLO

  Formulation to deal with files of "regular" NLO predictions?

- Understand importance of EW logarithms

  Real W emissions may cancel effect, but are interesting in themselves

- Fixes to first LHA on event generator interfaces

  Common file format, OO?

- Communication between Experiment & Theory on how tools are made and used