Stefano Frixione

Monte Carlos and Predictive Power
(a contradiction in terms?)

MC4BSM-3, CERN, 11/3/2008

Plan

- Monte Carlos: scope and limitations
- Towards more (perturbatively) accurate tools
- Outlook, or of what will possibly happen while LHC runs

I shall not give any technical details

Parton Shower Monte Carlos

Based on simple ideas with profound implications. Very flexible, essential tools for experimental physics. But:

- ♦ Each emission in a shower is based on a collinear approximation; matrix elements are leading order

Parton Shower Monte Carlos

Based on simple ideas with profound implications. Very flexible, essential tools for experimental physics. But:

- ◆ Each emission in a shower is based on a collinear approximation; matrix elements are leading order
- ◆ No K factors, no hard emissions

Parton Shower Monte Carlos

Based on simple ideas with profound implications. Very flexible, essential tools for experimental physics. But:

- ♦ Each emission in a shower is based on a collinear approximation; matrix elements are leading order
- ♦ No K factors, no hard emissions
- ♦ Very good in peak regions, ie the bulk of the cross section

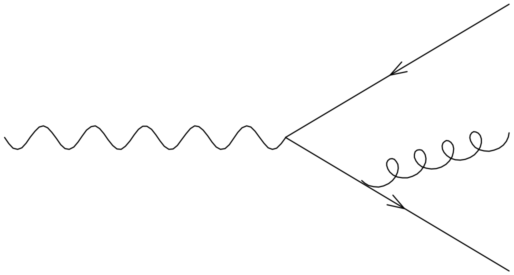
Parton Shower Monte Carlos

Based on simple ideas with profound implications. Very flexible, essential tools for experimental physics. But:

- Each emission in a shower is based on a collinear approximation; matrix elements are leading order
- No K factors, no hard emissions
- Very good in peak regions, ie the bulk of the cross section
- Fairly poor in large-p_T tails, ie rare events

A 30" guide to Monte Carlos

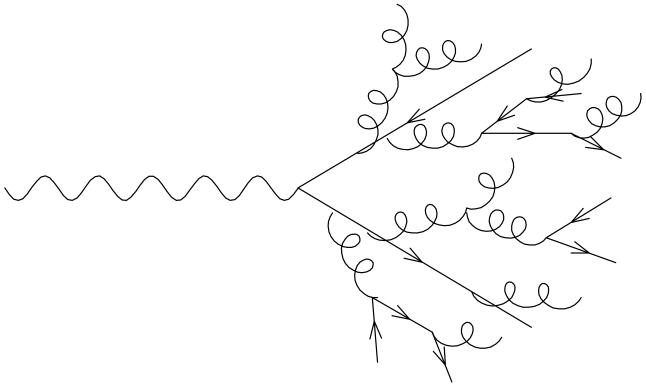
Key observation: collinear emissions factorize



$$d\sigma_{q\bar{q}g} \xrightarrow{t\to 0} d\sigma_{q\bar{q}} \times \frac{\alpha_S}{2\pi}\frac{dt}{t}P_{qq}(z)dz\frac{d\varphi}{2\pi}$$

$$t = (p_q + p_g)^2\,, \qquad z = E_q/(E_q+E_g)$$

Obviously, the process can be iterated as many times as one wants ⟶ parton shower; emissions are exponentiated into a Sudakov form factor



- Shower resums leading logarithmic contributions
- The cross sections are always positive (and at leading order)
- Large final-state multiplicities: fully realistic description of the collision process, including hadronization and underlying event
- Monte Carlos differ in the choice of *shower variables*: z, t

Soft physics

QCD has soft divergences. In MC's they are easy to locate:

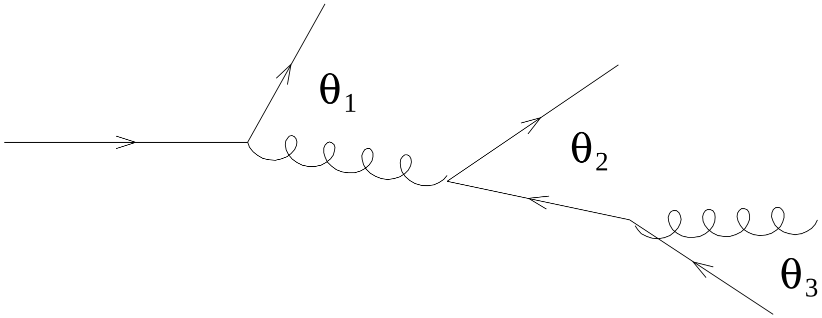
$$z \to 1 \qquad \Longrightarrow \qquad P_{qq},\ P_{gg} \sim \frac{1}{1-z}$$

The choice of shower variables affects the double-log structure

$$t = z(1-z)\theta^2 E^2 \quad (\text{virtuality}) \qquad \Longrightarrow \qquad \frac{1}{2}\log^2\frac{t}{E^2}$$

$$t = z^2(1-z)^2\theta^2 E^2 \quad (p_T^2) \qquad \Longrightarrow \qquad \log^2\frac{t}{E^2}$$

$$t = \theta^2 E^2 \quad (\text{angle}) \qquad \Longrightarrow \qquad \log\frac{t}{\Lambda}\log\frac{E}{\Lambda}$$



The choice that respects colour coherence is angular ordering (Mueller), as in HERWIG:

$$\theta_1 > \theta_2 > \theta_3$$

Implications

- There are large uncertainties in LO+LL QCD: one can go way too far beyond limits of applicability of the MC, without noticing it
- To stretch the theory to fit data may hide some interesting unknown physics

Is this important?

In the pre-LHC era, the basic requirement was the capability of the MCs to fit the data – predictive power was not an (important) issue

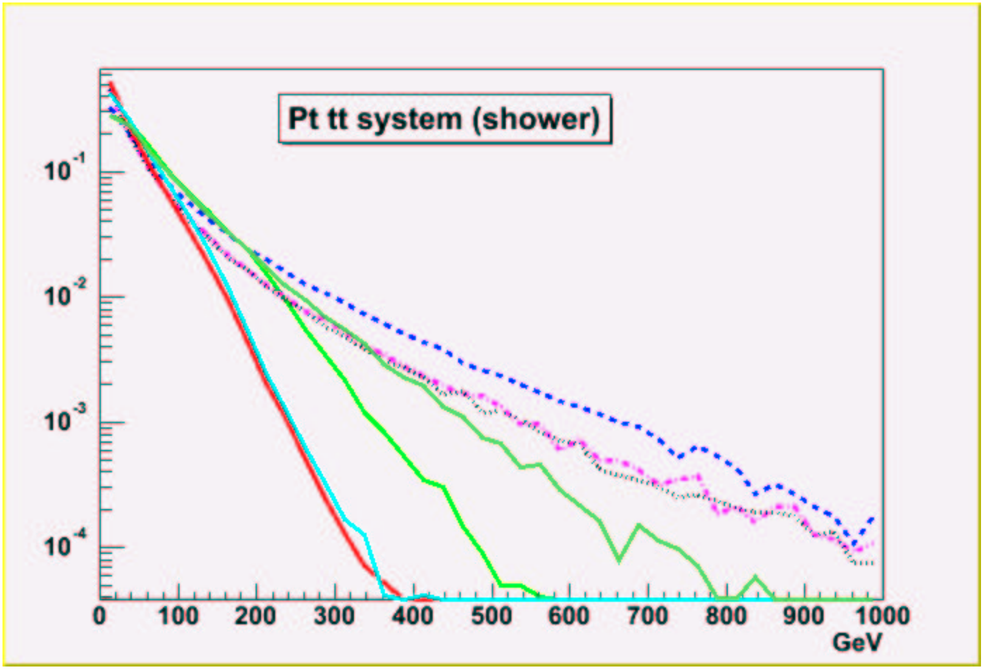
This fact has almost no implications on searches: *evidence* of new physics must be as independent as possible from MC truth

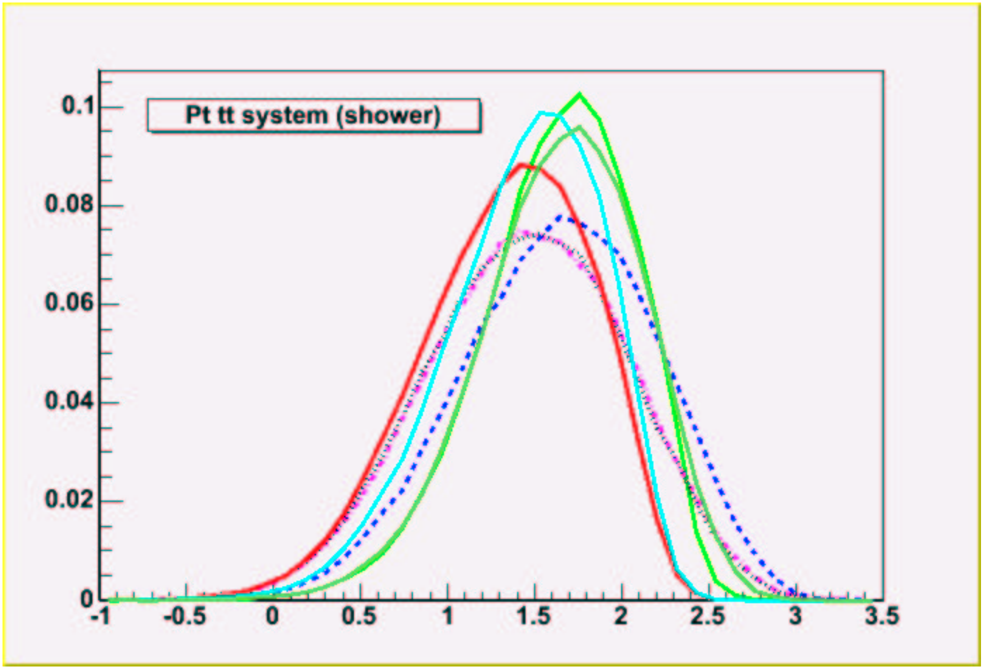
MCs must not be seen as discovery tools

LHC will however pose new problems:

- It may be more complicated to get *unbiased* evidence of new physics, if using vastly inadequate MCs
- A lot of BSM models, sometimes difficult to distinguish: to tell the right from the wrong, it is best to have precise simulations

Predictive power in standard MCs

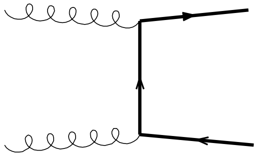




- Good at postdictions, not predictions
- No need to divide by zero to obtain these results. It is sufficient to choose suitable parameters (admittedly, some of them quite extreme)
- Can basically fit any data. On the other hand, many observables are under better control

Another example of too much flexibility in MCs: $b\bar{b}$ production

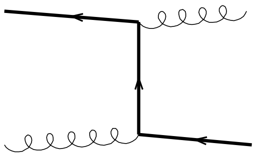
The leading-order picture implies starting from the hard process



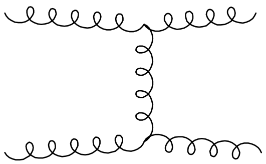
Flavour CReation

Typically, FCR underestimates the rate by a factor of 4, and misses key kinematic features (see R. Field)

So add other sources of b's



Flavour EXcitation



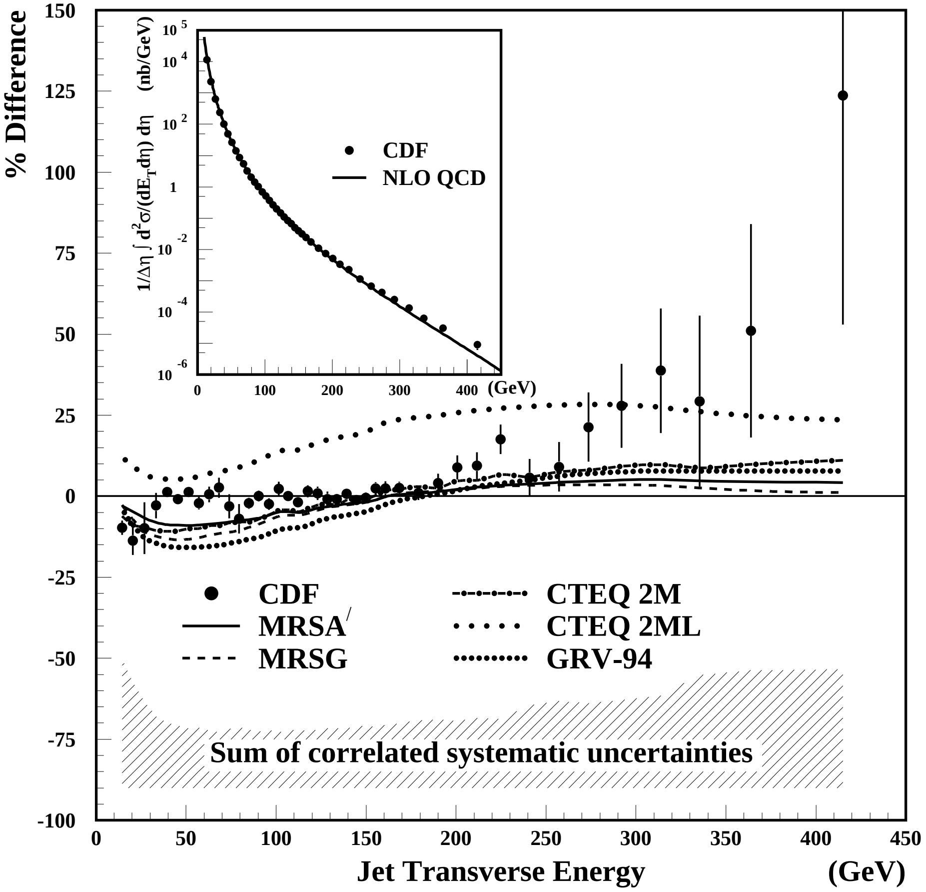
Gluon SPlitting

- In FEX, the missing Q or $\overline{Q}$ results from initial-state radiation
- In GSP, the Q and $\overline{Q}$ result from final-state gluon splitting
- Both are divergent, and must be rendered finite with a cutoff

This structure is fragile

- There are analysis at the Tevatron that combine FCR, FEX, and GSP by multiplying them by constants, which are fitted to data (and do not turn out to be equal)
- This is not what QCD tells one to do

The opposite case



An excess of data over theory had been reported by CDF (PRL77(96)438) for $p_T > 250$ GeV in the single-inclusive jet spectrum. A kinematic region sensitive to new physics

"Evidence" of new physics arose from lack of flexibility. In this case, a re-fit of the PDFs was sufficient for the data to agree with QCD NLO predictions

- Not an MC example, but stresses a general point: it is crucial to carefully assess the uncertainties affecting our (most accurate) predictions

What to take home

- At the LHC, standard MCs are either incapable of describing hard processes, or they do so at the price of rendering it impossible the study of uncertainties

- In the context of a reliable computation, the assessment of the theoretical uncertainty is a well defined procedure

- Hence, the final aim of giving MCs more predictive power is *not that of turning them into discovery tools*, but rather that of
  - Obtain state-of-the-art simulations of backgrounds
  - Reduce theory bias on search strategies
  - Better discriminate among BSM scenarios

How to improve Monte Carlos?

The key issue is to go beyond the collinear approximation

$\Longrightarrow$ use exact matrix elements of order higher than leading

How to improve Monte Carlos?

The key issue is to go beyond the collinear approximation*

$\Longrightarrow$ use exact matrix elements of order higher than leading

Which ones?

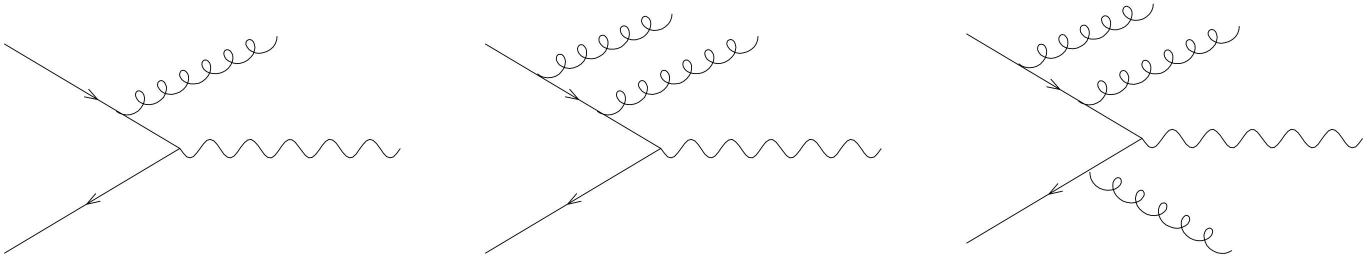
There are two possible choices, that lead to two vastly different strategies:

- Matrix Element Corrections $\longrightarrow$ tree level
- NLOwPS $\longrightarrow$ tree level and loop

* I won't discuss perspectives for Underlying Events – lot of work done (modelling and tuning), but still sort of plug & pray for LHC. Needs deeper theoretical understanding

Matrix Element Corrections

Compute (exactly) as many as possible real emission diagrams before starting the shower. Example: W production



. . . . . .

Problems

- Double counting (the shower can generate the same diagrams)
- The diagrams are divergent

Solution

$\longrightarrow$ Catani, Krauss, Kuhn, Webber (2001), Lonnblad (2002), Mangano (2005)

How to achieve MEC

- Preliminary step: compute the real matrix elements

  Non trivial for high-multiplicities. Problem now fully solved and highly automatized (`AcerMC`, `ALPGEN`, `AMEGIC++`, `CompHEP`, `Grace`, `MadEvent`)

- The strategy: apply a cut δ_{sep} on matrix elements to avoid divergences

  For a fixed multiplicity n, this implies a large, unphysical δ_{sep} dependence

$$\sigma_n \sim \alpha_S^{n-2} \sum_k a_k \alpha_S^k \log^{2k} \delta_{sep}$$

Then reweight ME's and modify the shower to eliminate or reduce the δ_{sep} dependence

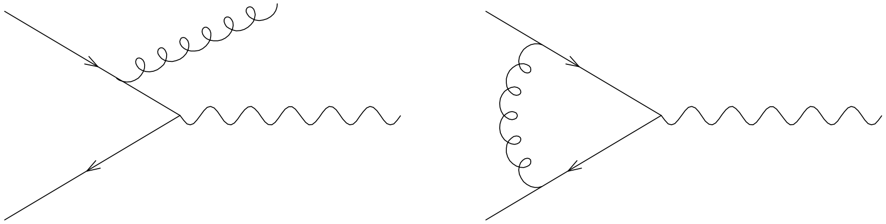
Following CKKW, one gets

$$\sigma_n \sim \alpha_S^{n-2} \sum_k a_k \alpha_S^k \log^{2k} \delta_{sep} \quad \longrightarrow \quad \alpha_S^{n-2} \left(\delta_{sep}^a + \sum_k b_k \alpha_S^k \log^{2k-2} \delta_{sep} \right)$$

NLOwPS

Compute all the NLO diagrams (and only those) before starting the shower.

Example: W production



……

Problems

- Double counting (the shower can generate *some of* the same diagrams)
- The diagrams are divergent

Solution

$\longrightarrow$

Proposals for NLOwPS's

- First working hadronic code (Z): Φ-veto (Dobbs, 2001)
- First correct general solution: MC@NLO (Frixione, Webber, 2002)
- Automated computations of ME's: grcNLO (GRACE group, 2003)
- Absence of negative weights (Nason, 2004; Frixione, Nason, Oleari, 2007) – POWHEG
- Showers with high log accuracy in ϕ^3_6 (Collins, Zu, 2002–2004)
- Proposals for $e^+e^- \to jets$ (Soper, Krämer, Nagy, 2003–2006)
- Shower and matching with QCD antennae (Giele, Kosower, Skands 2007) – VINCIA
- Within Soft Collinear Effective Theory (Bauer, Schwartz, 2006)
- With analytic showers (Bauer, Tackmann, Thaler, 2008) – GenEvA

Some of these ideas have passed the crucial test of implementation. Two codes (MC@NLO and POWHEG) can be used to fully simulate hadronic processes (examples in talk by Webber)

NLOwPS vs Matrix Element Corrections

NLOwPS are vastly different from MEC. MEC lack virtual corrections

This forces the use of an unphysical cutoff δ_{sep} in MEC, upon which physical observables depend $\longrightarrow$ matching systematics

NLOwPS are better than MEC since:

+ There is no δ_{sep} dependence (i.e., no matching systematics)

+ The computation of total rates is meaningful and reliable

NLOwPS are worse than MEC since:

− The number of hard legs is smaller

■ The days of the universal tools are over. Choose the one that best suits your analysis. Typically: small/large number of *extra* legs $\Longrightarrow$ NLOwPS/MEC

Which implies that MEC are the workhorses for searches studies

This is great progress. Multi-jet backgrounds not a matter of science fiction any longer (like in early TDRs...)

However:

- Never forget to check the merging systematics (a $\sim$ 20% effect?)
- These are LO computations: the scale dependence is very large
- For a low-multiplicity process, you should be using an NLOwPS instead

LHC will run for a long time, and will eventually become a “precision” machine. It would be a mistake not to strive for accurate predictions (keep the lesson of LEP in mind)

With NLO corrections

- NLOwPS's are the only way in which K-factors can be embedded into MC's (rescaling is WRONG!)

With NLO corrections

- NLOwPS's are the only way in which K-factors can be embedded into MC's (rescaling is WRONG!)
- The scale dependence of observables is meaningful

With NLO corrections

- NLOwPS's are the only way in which K-factors can be embedded into MC's (rescaling is WRONG!)
- The scale dependence of observables is meaningful
- Realistic hadronization for NLO-accurate predictions

With NLO corrections

- NLOwPS's are the only way in which K-factors can be embedded into MC's (rescaling is WRONG!)
- The scale dependence of observables is meaningful
- Realistic hadronization for NLO-accurate predictions
- Allow a fully-consistent determination of PDF uncertainties (PDF with errors are NLO fits), and of PDFs themselves

With NLO corrections

- NLOwPS's are the only way in which K-factors can be embedded into MC's (rescaling is WRONG!)
- The scale dependence of observables is meaningful
- Realistic hadronization for NLO-accurate predictions
- Allow a fully-consistent determination of PDF uncertainties (PDF with errors are NLO fits), and of PDFs themselves
- Non-trivial dynamics beyond LO ($t-\bar{t}$ asymmetry, FCR vs FEX vs GSP in $b\bar{b}$, $qg\longrightarrow Wq$, $Wt\longleftrightarrow t\bar{t}$ interference, jet algorithms, ...)

Outlook

- NLOwPS and MEC have complementary advantages
- NLOwPS have made it to the implementation stage. MC@NLO is used for SM physics by ATLAS, to a lesser extent by CMS, and for heavy flavour analysis by CDF, D0, and STAR
- Not limited to QCD ($pp \to Z'$ exists in MC@NLO)

Future prospects

- NLOwPS (low mult)+MEC (high mult): soon
- MEC $\longrightarrow$ NLOwPS (high mult): require automated 1-loop computations
- NNLOwPS?

Barring spectacular scenarios, LHC experiments will have to re-discover the SM before searching evidence for new physics

This startup phase will be crucial to learn to make a proper use of MCs of the new generation (MEC and NLOwPS): tune parameters, and *treat theoretical uncertainties seriously*

If this will be accomplished, it will help put search strategies and disentangling of BSM models on a firmer ground