Monte Carlo generators for the LHC

Roberto Chierici
Institut de Physique Nucleaire de Lyon, IN2P3-CNRS, Université Claude Bernard Lyon 1, Villeurbanne, France

Abstract
This contribution briefly reviews the Monte Carlo choices in CMS and ATLAS for the generation of signals and background for Standard Model physics. Emphasis will be given to the generator validation and the Monte Carlo set-up for interpreting the first LHC data.

1 Introduction and desiderata

The year 2009 is crucial for the Monte Carlo (MC) production at the LHC experiments, that will allow interpreting the first data. The experiments are preparing their event generation strategies and are producing large-scale samples of events for training tools and analyses.
In a modern generation setup for physics at the LHC there are certain requirements that need to be fulfilled. They can be summarised as follows:

- an event generator with a description of the hard scattering process with a matrix element (ME) calculation at the highest possible QCD order
- the possibility of interfacing, directly or via intermediate parton level files, to generic tools used for the parton showering (PS) and for parton hadronisation. The most known, and largely used, are PYTHIA [1] and HERWIG [2]
- the presence of models for the description of the underlying event (UE), representing all what is in the event except the primary interaction. PYTHIA and HERWIG already present models for this task
- a coverage, as large as possible, of Standard Model (SM) and Beyond the Standard Model (BSM) processes, with a good flexibility for implementing new physics models in the event generation
- standard output formatting of parton level files, in particular the possibility of outputting events in the Les Houches format [3]

The current article should not be intended as a review of generators, but rather a picture of the current MC set-up chosen by ATLAS and CMS, and of the current validation activities on this subject. I will focus in what follows on generic SM and BSM physics from pp collisions, without discussing generators for heavy ions studies, or dedicated tools for new physics signatures (like black holes generators) or dedicated detector studies (like generator of cosmics, beam halo or beam-gas intercations). These generators remain however essential for the physics programme of ATLAS and CMS.

2 Generators for LHC physics

2.1 Event generators

Both the ATLAS and CMS Collaborations try to use as many event generators as reasonable. The reference generic purpose event generators for SM and BSM physics and beyond are PYTHIA, HERWIG and SHERPA. The first two, whose original version is written in FORTRAN, are now also used in their C++ versions (PYTHIA8, HERWIG++), that will be the only ones maintained in the long term. The main common feature of all generic purpose generators is that they provide a fully hadronised event to be passed directly to the detector simulation. All of them implement models for the description of the radiation, fragmentation and the underlying event. The models in PYTHIA and HERWIG have been extensively tuned to LEP, SLD and Tevatron data for what concerns PS-fragmentation [4] and UE [5]. If PYTHIA and HERWIG include LO descriptions of very many SM and BSM processes, in some cases with the additional corrections to PS for a description of the first QCD emission at NLO, SHERPA also include the possibility of matching PS with ME at higher leading order, for both SM and BSM processes. General interest decay/correction tools, interfacable to all kind of general purpose event generators, are typically used in both Collaborations. Most noticable ones are TAUOLA, for τ decays [6], EvtGen, for hadron decays [7], extensively tuned at the Tevatron and at B-factories, and PHOTOS, for including real QED corrections [6].

If generic purpose event generators represent the 'work-horses' for the MC productions at the LHC, there has been an enormous progress in the last years on implementing ME descriptions of beyond-leading order QCD processes in event generators. This allows to improve the predicitons for observables sensitive to hard QCD emission (multi jet final states, typically). This has been achieved either with techniques matching higher leading order (HLO) ME with PS (examples are given by ALPGEN [8], MadGraph [9], SHERPA [10], HELAC [11]), and by next-to-leading order (NLO) generators (like MC@NLO [12] and POWHEG [13]). The fundamental difference between the two categories of calculations is that the HLO maintains a precision that is typically LO, but more correctly predicts shapes of differential distributions sensitive to real QCD emission, even at several orders beyond the leading, whereas NLO calculations are correct in shape and normalisation at NLO for inclusive variables, but they count on PS for all extra emission beyond the first.

Both CMS and ATLAS have interest in all those generators, and there is already an extensive experience in their use in the collaborations. MadGraph, ALPGEN and MC@NLO are indeed references in the current Monte Carlo productions for physics. The event generators in the HLO and NLO categories remain parton level event generators, and need therefore to be interfaced to PS and hadronisation for use in the experiments. Most of them provide direct interfaces to PYTHIA, or parton level output in the standard Les Houches Accord format [3], that can be input to any hadroniser. Noticeable exception is MC@NLO, directly built on top of HERWIG.

The present list of generators does not exhaust what experiments have used and are using for physics results. Some of them represent useful crosschecks, like AcerMC [14] or TopRex [15] for top physics, or are in place for the description of particular processes, like SingleTop [16] for single top physics or Phantom [17] for the description of full six fermion processes at LO.

2.2 Generators tuning and set-up

A full event generation often implies approximations by use of models, whose parameters need typically to be tuned to data. Examples are the parton showering, fragmentation, the description of the proton PDFs, the modelisation of the underlying event. Without entering a detailed explanation of each topic, I will briefly review the current settings chosen by ATLAS and CMS.
The first essential ingredient, since protons are composite objects, is to describe the probability of the initial state at the hard process scale Q^2 with a certain fraction x of the total proton momentum. Since the Q^2 evolution can be calculated perturbatively in the framework of QCD, PDFs are fitted to a set of heterogeneous data from DIS, Drell-Yan and jet data. Both Collaborations are currently using the LO CTEQ6L1 fit [18] with NLO PDF used only for NLO ME calculations. Errors from the fits, currently only available for NLO fits, are then propagated to the final observables. The scheme adopted at present is likely to change since no one of the generator used is purely LO. There is more and more consensus, in the theory community, for using modified leading order PDFs [19] for all LO calculations, or calculations including LO ME corrections. Modified PDFs are, essentially, LO PDF that relax the partonic momentum sum rule to get predictions artificially closer to NLO.

From parton level four-momenta configurations, initial and final state QCD and QED radiation are produced, via parton showering algorithms down to a certain energy scale: from that scale on fragmentation transforms coloured partons to colourless hadrons according to specific models. Radiation parameters are typically fitted together with the fragmentation parameters, and for the moment both ATLAS and CMS make use of fits from LEP/SLD [4, 20], assuming jet universality. The fragmentation functions chosen for heavy quark fragmentation are the ones better describing LEP/SLD data, namely Bowler [21] and Peterson [22]. With data available, those fits will have to be re-made at the LHC, taking care of the additional complication that initial state radiation at hadron machines contributes to the description of the underlying event as well, so it will be essential to disentangle the two. Moreover, with the use of modern ME-PS matching, tunings of the PS part will have a new meaning with respect to previous tunings.

The underlying event corresponds to what else is present in an event, except the hardest interaction. Multiple parton interaction models turn out to be particularly adequate to describe this kind of physics. Examples of these models are implemented in the general purpose simulation programs PYTHIA, HERWIG/JIMMY [23], and SHERPA. Huge progress in the phenomenological study of the underlying event in jet events have been achieved by CDF [5] using, for the tuning of the models, the multiplicity and transverse momentum spectra of charged tracks in different regions in the azimuth-pseudorapidity space, defined with respect to the direction of the leading jet. The main problem of extrapolating the predictions of the multiple interactions models to the LHC is that some of the parameters are explicitly energy dependent. Some of the tunes, used by ATLAS and CMS [24, 25], have put enphasis in the energy extrapolation by also fitting lower energy data. The results are shown in figure 1, where the predictions of JIMMY and PYTHIA are extrapolated to the LHC energy for the average number of charged tracks and the average p_T sum of tracks in the transverse region (with respect to the leading jet in the event) as a function of the transverse momentum of the leading jet in the event. The curves are compared to CDF data, and it is clear that the extrapolation to CMS energies implies very different shapes compared to Tevatron. Moreover, the extrapolated predictions can differ widely according to the

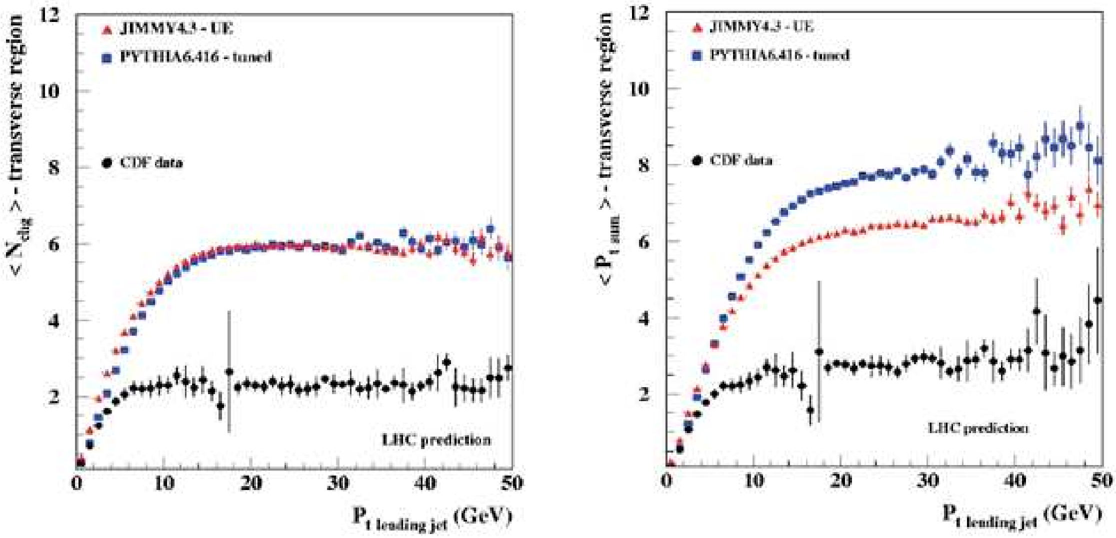

Fig. 1: Average number of charged tracks (left) and average track p_T sum in the transverse region (right) as a function of the transverse momentum of the leading jet in the event. The extrapolated predictions at the LHC are compared to CDF data.

model used, therefore it will be mandatory to use LHC data themselves to validate them.

3 Generator validation

The validation of generators prediction in an experimental framework is an invaluable exercise to gain confidence in the tools being used and to learn about the difference in the physics contents between generators. A few important examples are presented in this section.

3.1 Multiple parton interactions

The presence of multiple parton interactions, i.e. the possibility of having multiple parton-parton interactions overlapping in the same event, has been established already at the Tevatron, as illustrated in figure 2. The left part of the figure shows, for γ+3jets events, the azimuthal distance between the transverse momentum vectors formed by the photon and the most back-to-back jet, and by the other two jets. The MPI component is expected to have a flat behaviour in this variable, and the figure clearly shows that the CDF data can not be described without accounting for it. The most recent PYTHIA version includes MPI interleaved to PS, and it is essential to validate this tool in the experimental framework.

The right-hand part of figure 2 shows a preliminary study by CMS where the prediction of PYTHIA8 with MPI for the same azimuthal variable are compared with PYTHIA6 and HERWIG with the most uptodate UE tune [26, 27], and the same generators without the inclusion of MPI. The plot shows that the newest version of PYTHIA agrees with the default tuned one, and that there are important discrepancies between HERWIG (+JIMMY) and PYTHIA. One more time it is shown that MPI effects are non negligible and should be accounted for.

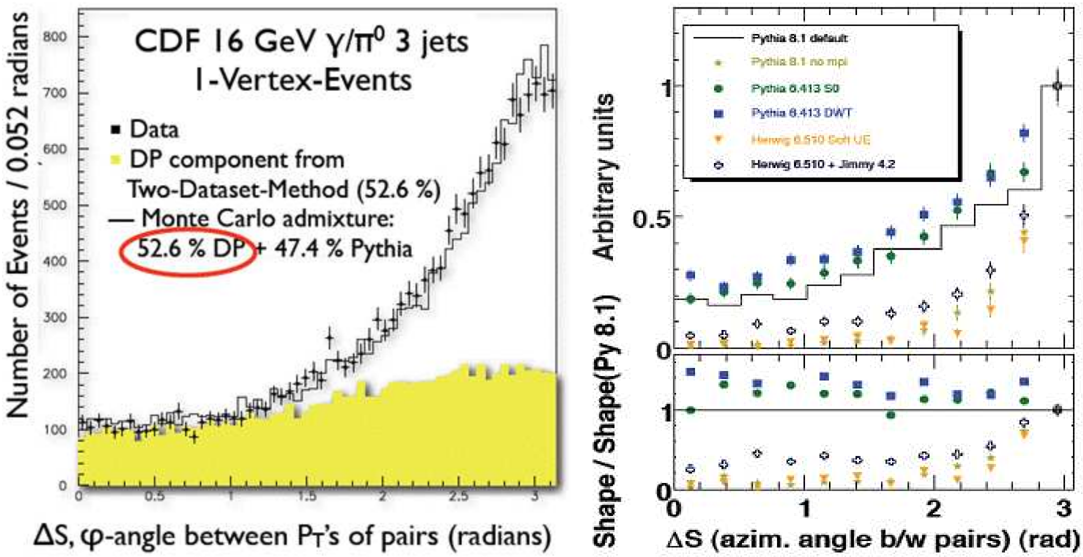

Fig. 2: Azimuthal distance between γ+j and j+j systems in γ+jets events at CDF, comparing data with MC, with or without MPI component (left).Validation of PYTHIA8 with MPI in CMS (right).

3.2 Hard QCD emission in boson production

Recent developments in ME generators allow to describe QCD radiation much more accurately. It is instructive to compare, for high p_T physics, the prediction of those calculations with respect to LO ones for observables that are sensitive to (gluon) radiation. One of such comparison comes from W+jets production. The ATLAS Collaboration compared the transverse momentum of the first four highest p_T jets in the event for ALPGEN and PYTHIA. The results are shown in figure 3, and large difference are observed in the high momentum tails, as expected by a more accurate ME description. Also, the total number of high p_T jets increases very significantly going from a pure LO description to a higher order one with matching to PS.

One important question for the analyses is about the residual uncertainty on total and differential cross-sections when going to high jet multiplicity in the final state. This question addresses the problem of quantifying the confidence on the description of W boson production as background to more complex process like top-pair production, where an associated many-jets production is necessary. To assess this, ATLAS have calculated the predicted cross-sections for all jet multiplicities in W+jets with ALPGEN by varying both the matching scale (from 10 to 40 GeV) and the minimum ΔR ($\sqrt{\Delta\eta^2 + \Delta\phi^2}$) that defines a parton (from 0.3 to 0.7). The result, confirming that the relative importance of the cross-sections at a fixed parton multiplicity varies according to the choice, shows that also the total cross-section, i.e. the sum of all fixed multiplicities contributions, varies quite significantly in the different scenarios, up to around a factor 50%. This is shown in fig. 4, left, where the reconstructed top mass for candidate semileptonic events in signal and W+jets background samples is shown for two choices of the matching scale, 20 and 40 GeV, respectively, at the same parton separation definition of ΔR = 0.7. The event selection is kept very simple with one reconstructed charged electron or muon with

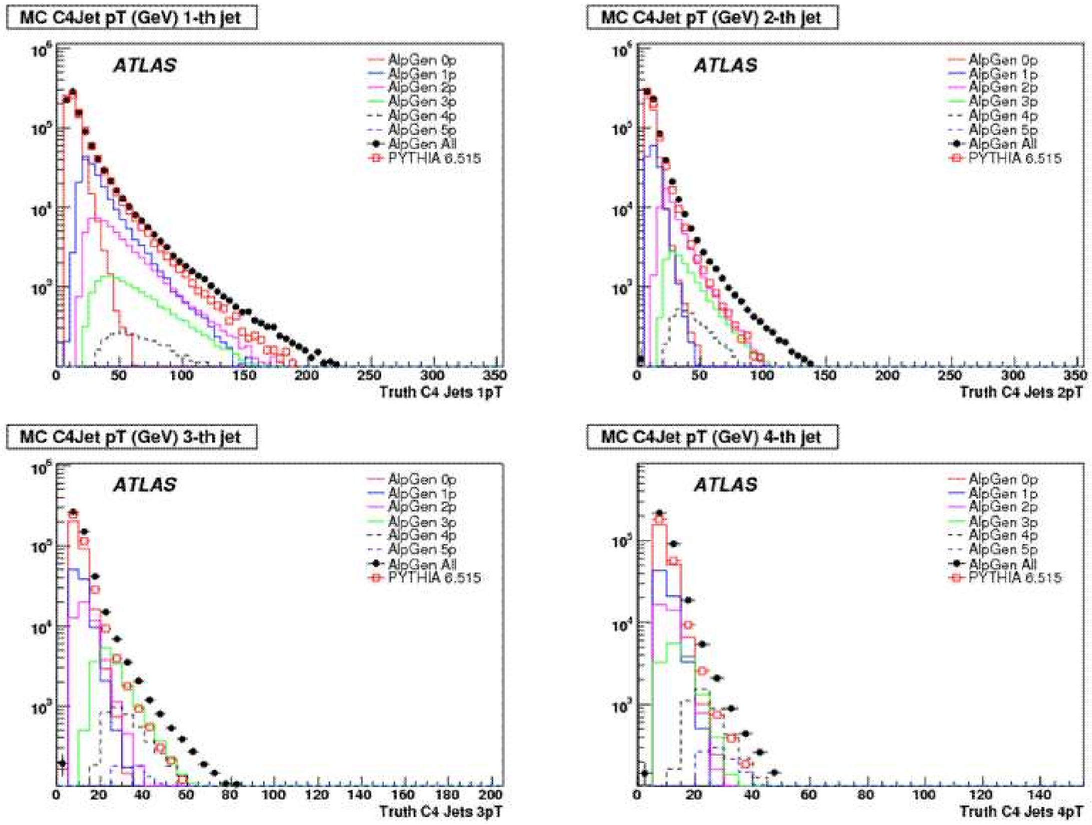

Fig. 3: Transverse momentum of the first four highest p_T jets in W+jets events.

$p_T > 20$ GeV and $|\eta| < 2.5$, missing transverse energy greater than 20 GeV, and at least four reconstructed jets, each with transverse energy of at least 20 GeV and for three of them larger than 40 GeV. Though the shape of the signal is unchanged, the W+jets background scales by a factor 1.5. This reflects an uncertainty of the matching procedure itself that grows as the final parton multiplicity gets higher. Though the matching itself can be constrained using data at the LHC, present comparisons data-MC made at the Tevatron still show an insufficient statistics to constrain such predictions at the LHC. This is shown in fig. 4, right, where the CDF collaboration shows the ratio between data and theory for the inclusive jet multiplicity in W events [28]. As can be seen, the error bands of the matching codes get bigger at high multiplicity and current data is not enough to constrain them significantly.

3.3 Hard QCD emission in top production

A thourough test of the different description of QCD was also made by the CMS Collaboration in the case of top-pair production: differences may manifest themselves in distortions of the top quark angular distributions and transverse variables.

The most spectacular effect is in the transverse momentum of the radiation itself, which equals the transverse momentum of the $t\bar{t}$ system recoiling against it: this is what is shown in fig. 5,

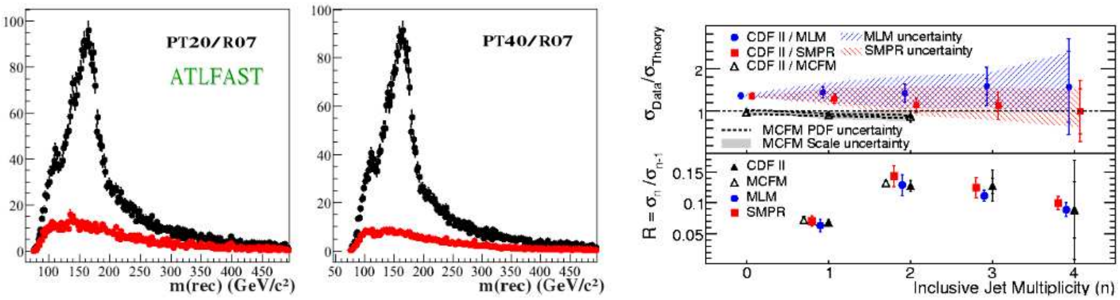

Fig. 4: Reconstructed top mass in ATLAS for $t\bar{t}$ signal and W+jets background (left) and ratio between data and different theory predictions for the inclusive jet multiplicity in W events at the Tevatron (right).

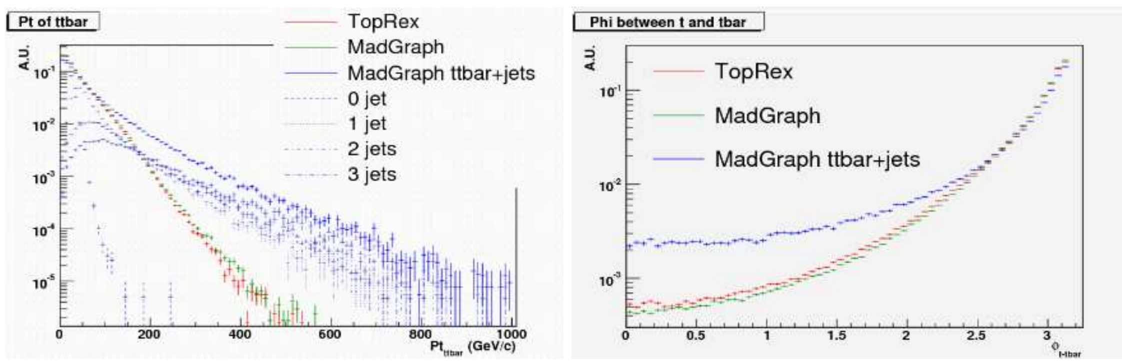

Fig. 5: Transverse momentum of the $t\bar{t}$ system (left), azimuthal angle between the two tops (right). All distributions are normalised to unity.

left, for two leading order generations by MadGraph and TopRex (with PS) in comparison to the ME-PS matching scheme in MadGraph. The contributions to a fixed ME order, ie tt+0jets, tt+1jets, tt+2jets and tt+3jets, are explictly indicated. On the right hand side of the same figure the corresponding distribution of the azimuthal difference between the two tops is also shown. The centre of mass energy is 14 TeV, and it is important to notice that the input parameters settings (cuts, scales, PDFs) of the various generators shown in the figure are kept as uniform as possible to avoid any possible bias in the comparison. From the picture it is evident that gluon production via ME predicts a much harder transverse spectrum. The difference in shape reaches orders of magnitude in the ratio at very high values of p_T. The increased activity in hard gluon emission for the ME-PS matched case also explains a generally decreased azimuthal distance between the two top quarks, which tend to be closer to each other. The distributions confirm the fact that having more ME radiation tends to increase the event transverse activity. The predicted average p_T of the radiation by MadGraph is 62 GeV/c (72 GeV/c with ALPGEN), with a 40% probability of having more than 50 GeV/c as gluon p_T in $t\bar{t}$ events. This large gluon activity will certainly have an impact in the capability of correctly reconstructing top quark events at the LHC, and correctly interpreting radiation as a background for new physics searches.

An important validation step comes from the comparison of the predictions from different

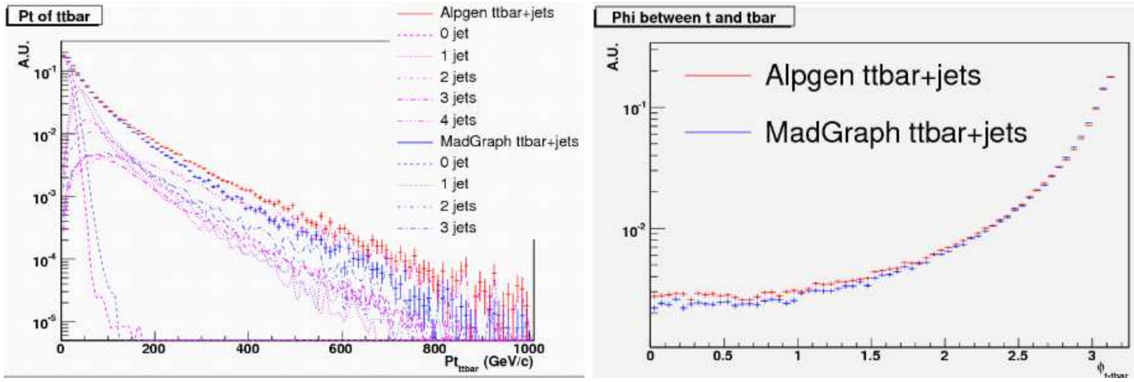

Fig. 6: Transverse momentum of the $\mathrm{t\bar{t}}$ system (left), azimuthal angle between the two tops (right). All distributions are normalised to unity.

ME-PS matched codes. Fig. 6 shows the same distributions of fig. 5, but for ALPGEN and MadGraph with ME-PS matching, respectively. For the $\mathrm{p_T}$ of the $\mathrm{t\bar{t}}$ system the individual parton multiplicity components are also shown. The agreement is more than acceptable, and remarkable for the azimuthal difference between the top quarks. Especially in the tails of the distributions, corresponding to high radiation conditions, the disagreement goes from orders of magnitude of fig. 5, to a maximum discrepancy of 50%. To properly appreciate the difference between the two predictions one should, however, account for the theory errors on them. Scale and PDF dependencies, PS tuning uncertainties could very well account for any residual difference in the tails.

Another important test for the description of radiation in the top-pair production comes from the comparison of matched ME-PS calculations to NLO predictions. This study was made by comparing the previous predictions to MC@NLO. A general very good agreement was found in all distributions, including the transverse ones. In fig. 7 the $\mathrm{p_T}$ of the $\mathrm{t\bar{t}}$ system and the $\mathrm{p_T}$ of the top are shown for ALPGEN, MadGraph and MC@NLO. As can be appreciated from the figure, it is particularly relevant the fact that the tails of the radiation are very well reproduced. The discrepancy in the very soft region is mostly due to the different showering, since MC@NLO is only interfaced to HERWIG whereas the other predictions use PYTHIA as tool for PS and fragmentation.

4 Summary and outlook: towards data

The LHC experiments are preparing their MC production to be ready for the interpretation of the imminent data. There are a few important lessons that have been learned from previous experiments and via the generator validation efforts in ATLAS and CMS, that help planning a winning generation strategy:

- make sure to use the best available tools for the description of the signal and the main backgrounds. For high jet multiplicity signals it is of utmost importance to include higher QCD corrections with now available ME generators.

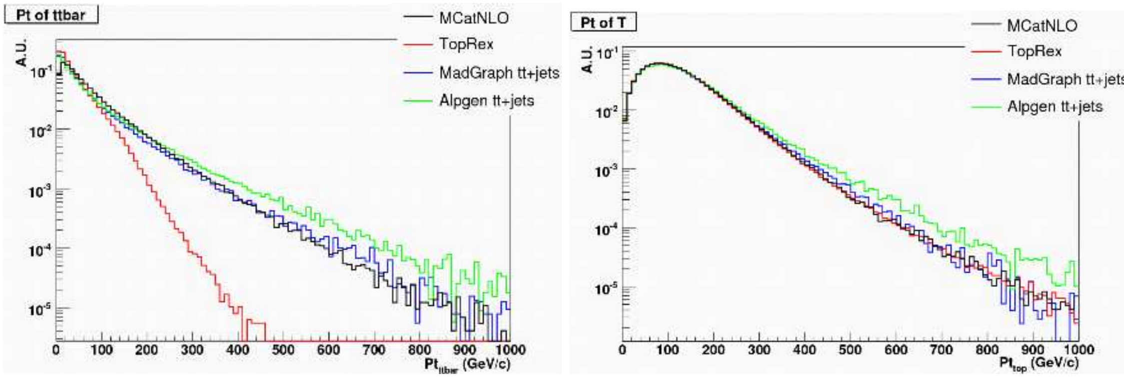

Fig. 7: Transverse momentum of the $t\bar{t}$ system (left), transverse momentum of the top quark (right). All distributions are normalised to unity.

- plan a very accurate MC tuning by using LHC data. All event generators use models for PS, fragmentation and UE/MPI, that need to be tuned. Moreover, interfacing external NLO or HLO generators to more standard PS tools opens new scenarios for the MC tunings. The PDF fits will also be enriched by the use of LHC data at higher value of Q^2
- diversify the event generation and make it redundant, in such a way to compare different tools in the interesting regions of the phase space, or put in place parameter scans to understand possible systematic effects due to theory. Particular attention has to be put to the dependency of the analyses to chosen scales, PDFs and ME-PS matching schemes.
- make the reference SM and BSM generation as much as possible coherent (same input settings and cuts) and consistent (full coverage of phase space). This will help correctly interpreting analyses' results and in shortening the time for any discovery claim

ATLAS and CMS are preparing at their best the start-up of the LHC for what concerns the Monte Carlo set-up and productions. New C++ event generators, as well as more complex HLO/NLO ME tools are used extensively in the analyses, and the level of communication with the theory communities, often a key to success in data interpretation, is constantly increasing. The choices made now will certainly shape the way the collaborations will be doing physics at the start-up, and not only.

References

[1] T. Sjöstrand et al., Comput. Phys. Commun. 135 (2001) 238.

[2] B. Webber et al., Journal of High Energy Physics **0101** (2001) 010.

[3] J.Alwall et al., hep-ph/0609017.
H. Baer et al., hep-ph/0403045.

[4] ALEPH Collaboration, Z. Phys. **C55** (1992) 209.
DELPHI Collaboration, Z. Phys. **C73** (1996) 11.
OPAL Collaboration, Z. Phys. **C69** (1996) 543.
ALEPH Collaboration, Phys. Lett. **B512** (2001) 30.
DELPHI Collaboration, Z. Phys. **C73** (1996) 11.
OPAL Collaboration, Eur.Phys.J.**C29** (2003) 463.
SLD Collaboration, Phys. Rev. D **65**, 092006 (2002).

[5] T. Affolder *et al.*, *Phys. Rev.* D **65** (2002) 092002.
CDF Collaboration, Acta Phys. Polon. B **36** (2005) 167.

[6] *http://wasm.home.cern.ch/wasm/goodies.html*

[7] D.J. Lange et al., Proceedings of the International Conference on Computing in High-Energy Physics: CHEP'98, Chicago (IL), USA, 1998.

[8] M.L. Mangano et al., Journal of High Energy Physics **0307** (2003) 001.

[9] F. Maltoni and T. Stelzer, hep-ph/0208156.

[10] F.Krauss et al., Journal of High Energy Physics **0402** (2004) 056.

[11] A. Kanaki and C. Papadopoulos, Comput. Phys. Commun. 132 (2000) 306.

[12] S. Frixione and B.R. Webber, Journal of High Energy Physics JHEP06 (2002) 029.

[13] S. Frixione, P. Nason, C. Oleari, JHEP **0711:070** (2007).

[14] B.P. Kersevan, E. Richter-Was, hep-ph/0201302.

[15] S.R. Slabospitsky, Comput. Phys. Commun. 148 (2002) 87.

[16] E.Boos et al., Phys. Atom. Nucl. 69, 8 (2006) 1317.

[17] A.Ballestrero et al., in preparation.

[18] D.R. Stump, Prepared for 31st International Conference on High Energy Physics (ICHEP 2002), Amsterdam, The Netherlands, 24-31 Jul 2002. Published in *Amsterdam 2002, ICHEP* 265-267.

[19] A. Sherstnev, R.S. Thorne, arXiv:0807.2132.

[20] P. Bartlaini, R.Chierici, A. de Roeck, CERN-CMS-NOTE-2005-013 (2005).

[21] M.G. Bowler, Z. Phys. **C11** (1981) 169.

[22] C. Peterson et al., Phys. Rev. **D27** (1983) 105.

[23] J.J. Butterworth et al., hep-ph/9601371.

[24] A. Moraes et al., Acta Physica Polonica B 35(1) (2004) 433.

[25] P. Bartalini et al., hep-ph/0003142 (2000) 293.

[26] D. Acosta et al., CERN-CMS-NOTE-2006-067 (2006).

[27] P. Skands and D. Wicke, Eur. Phys. J. C51 (2007) 133-140

[28] CDF Collaboration, arXiv:0711.4044.