New features in Delphes 3

Alexandre Mertens

Centre for Cosmology, Particle Physics and Phenomenology (CP^3),
Université Catholique de Louvain,
Chemin du Cyclotron 2, B-1348 Louvain-la-Neuve, Belgium

E-mail: `alexandre.mertens@uclouvain.be`

Abstract. Delphes is an open source C++ framework to perform the fast simulation of the response of a multipurpose detector. The simulation includes a tracking system, embedded into a magnetic field, calorimeters and a muon system. The framework is interfaced to standard file formats and outputs observables such as isolated leptons, missing transverse energy and collection of jets that can be used for dedicated analyses. The simulation of the detector response takes into account the effect of magnetic field, the granularity of the calorimeters and subdetector resolutions. The program contains parmetrizations for the CMS and ATLAS detectors, based on published performances. Basic parametrizations for the LHCb and FCC detectors are also available. The Delphes framework also includes a simple event display.

Several new features are discussed, such as an emulation of the particle-flow algorithm, pile-up simulation, N-subjettiness and a simple b-tagging algorithm based on counting tracks with large impact parameter.

1. Introduction

High energy particle collisions can produce a large variety of final states. Highly sophisticated detectors are designed in order to detect and precisely measure particles originating from such collisions. Experimental collaborations often rely on Monte-Carlo event generation for designing and optimizing specific analysis strategies. Whenever such studies require a high level of accuracy, the interactions of long-lived particles with the detector matter content are fully simulated with the GEANT package [1], electronic response is emulated by dedicated routines, and final observables are reconstructed by means of complex algorithms. To face the limited computing resources and still allow the use of large samples (for example when scanning parameter spaces), the LHC collaborations have developed fast-simulation techniques [2, 3, 4, 5, 6] which are two to three orders of magnitude faster than the fully GEANT-based simulations.

These procedures require expertise and the deployment of large scale computing resources that can be handled only by large collaborations. For most phenomenological studies, such a level of complexity is not needed and a simplified approach based on the parameterization of the detector response is in general good enough. In 2009, the DELPHES framework [7] was designed to achieve such goal. The DELPHES framework takes as input the most common event generator output and performs a fast and realistic simulation of a general purpose collider detector. To do so, long-lived particles emerging from the hard scattering are propagated to the calorimeters within a uniform magnetic field parallel to the beam direction. The particle energies

are computed by smearing the initial momenta according to the detector resolution. As a result, jets, missing energy, isolated electrons, muons and photons, and taus can be reconstructed.

With respect to its previous incarnation [7], the present version of DELPHES [8] includes an attempt to roughly emulate the particle-flow reconstruction philosophy used in ALEPH [20] and CMS [14], based on the optimally-combined use of the information from all the subdetectors to reconstruct and identify all particles indvidually. While the aim is not to re-implement the particle-flow algorithm in all its complexity (for example, electrons, muons, and photons, are assumed to be perfectly identified and have no fake rate in DELPHES), the simplified approach adopted here is particularly suitable for the treatment of pile-up, as well as for the emulation of b and τ tagging, and is able to reproduce the jet and missing energy resolutions observed in CMS with their complete reconstruction. From a technical perspective, the code structure is now fully modular, providing greater flexibility to the user and allowing the integration of DELPHES routines in other projects.

Among the new features of the DELPHES realease, the particle flow emulation is presented in section 2 and the pile-up implementation in section 3. Other developments such as B-tagging and jet substrure are described in section 4 and section 5.

2. Particle Flow emulation

The implementation of the Particle Flow emulation in DELPHES is described in detail in [8]. To illustrate the method, here are a few simple examples:

- a single charged pion is reconstructed as a track with energy $E_{HCAL,trk}$ and deposits some energy E_{HCAL} in the HCAL. If $E_{HCAL} \leq E_{HCAL,trk}$ only a particle-flow track with energy $E_{HCAL,trk}$ is produced. If $E_{HCAL} > E_{HCAL,trk}$, a particle-flow track with energy $E_{HCAL,trk}$ and a particle-flow tower with energy $E_{HCAL} - E_{HCAL,trk}$ are produced.

- a single photon deposits its energy E_{ECAL} in an ECAL cell. No tracks pointing to the cell is reconstructed. A particle-flow tower is created with energy E_{ECAL}.

- a photon and a charged pion reach the same calorimeter tower, the former deposits some energy E_{ECAL} in ECAL and the latter E_{HCAL} in HCAL. Furthermore, the charged pion is reconstructed as a track, with an energy $E_{HCAL,trk}$. If $E_{HCAL} \leq E_{HCAL,trk}$, a particle-flow track with energy $E_{HCAL,trk}$ and a particle-flow tower with energy E_{ECAL} are created. If $E_{HCAL} > E_{HCAL,trk}$, a particle-flow track with energy $E_{HCAL,trk}$ and a particle-flow tower with energy $E_{ECAL} + E_{HCAL} - E_{HCAL,trk}$ are created.

Defined that way, the particle-flow tracks contain charged particles estimated with a good resolution, while the particle-flow towers contain in general a combination of neutral particles, charged particles with no corresponding reconstructed track and additional excess deposits induced by the positive smearing of the calorimeters, and are characterized by a lower resolution. Besides producing high-resolution inputs for jets and missing transverse energy, the particle-flow approach can be rather useful for addressing pile-up subtraction. While very simple when compared to what is actually required in real experiments, the algorithm illustrated above is shown to reproduce well the performance achieved at LHC [8].

Physics application In modern collider experiments at high energy, $t\bar{t}$ events are among the most copious signatures observed in the detectors. When one top quark decays leptonically and the other hadronically, the signature is characterized by one lepton, missing transverse energy and four jets, two of them originating from the fragmentation of b quarks. Moreover, at the LHC, about 50% of events have extra hard jets coming from initial or final state radiation. Following the semi-leptonic $t\bar{t}$ analysis described in ref. [21] we focus on the mass of the hadronically-decaying top quark.
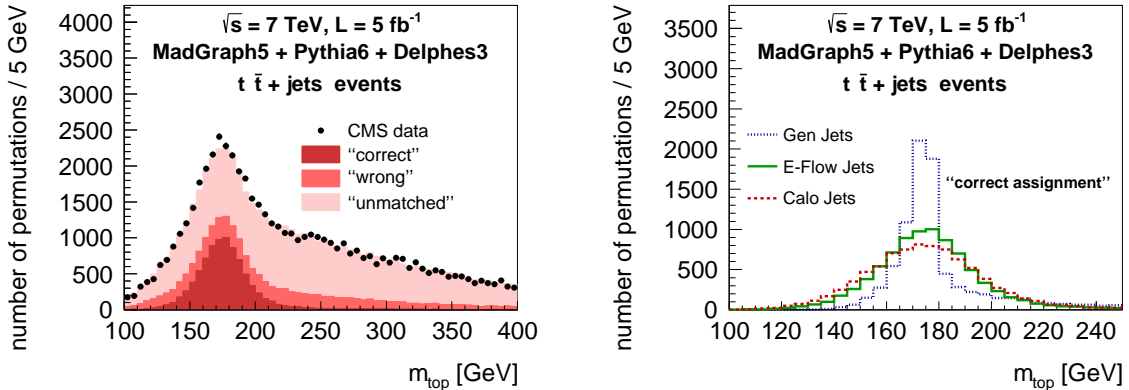
Figure 1. Left: Reconstructed hadronic top mass for all possible assignment categories. The DELPHES distribution is normalized to the CMS yield. The CMS contributions are taken from Ref. [21]. Right: Reconstructed hadronic top mass distribution for the *correct assignments* only. The distribution is shown for *Generated Jets*, *Particle-Flow Jets* and *Calorimeter Jets*.

The $t\bar{t}$+jets sample has been generated with MADGRAPH5 at a centre of mass energy $\sqrt{s} = 7$ TeV and PYTHIA6 was used for parton shower and hadronization. Backgrounds are not considered here. The reconstruction has been performed via DELPHES using the detector configuration designed to mimic the performance of the CMS detector.

Following the CMS approach, we select events with exactly one isolated lepton (electron or muon) with $p_T > 30$ GeV/c and $|\eta| < 2.1$. In addition, we require at least four particle-flow jets with $p_T > 30$ GeV/c and $|\eta| < 2.4$. The anti-k_T [15] algorithm with a parameter $R = 0.5$ was used for jet clustering. Among the selected jets, at least two must be tagged as originating from the hadronization of a b quark (b-tagged) and at least two must be identified as *light jets* (i.e. fail the b-tagging criterion). The b-tagging efficiency parameterization has been extracted from [25]. The signal efficiency for this selection is 2.8%, compared to 2.3% in the CMS analysis, showing a reasonable agreement between DELPHES and CMS. Given the high jet multiplicity, the signal selection is extremely sensitive to changes in requirements that can affect the jet selection.

The top quark mass distributions obtained with DELPHES and CMS for the three permutation categories are shown in figure 1, left. The DELPHES distributions are normalized to the CMS total number of events. Overall the shapes and relative contributions corresponding to the three categories are well reproduced by DELPHES.

For the sake of illustration, the reconstructed hadronic top mass using correct permutations only is shown in figure 1 (right) using three different jet collections: *Generated Jets*, *Calorimeter Jets* and *Particle-Flow Jets*. We observe, as expected, a narrow peak when using *Generated Jets* and wider peaks when using *Particle-Flow Jets* or *Calorimeter Jets*. This illustrates the need for using realistically reconstructed objects rather that hadron-level quantities in prospective phenomenological studies.

3. Pile Up implementation

At the LHC, several collisions per bunch-crossing occur in high luminosity conditions, most of them resulting in a small amount of activity in the detector. Due to the elongated shape of the proton bunches constituting the beams, such additional *pile-up events*, take place in a similarly elongated region (called beam spot) around the nominal interaction point. In DELPHES, pile-up interactions are extracted from a pre-generated low-Q^2 QCD sample. These minimum-bias

interactions are randomly placed along the beam axis according to some longitudinal spread that can be set by the user. The actual number of pile-up interactions per bunch-crossing is randomly extracted from a Poisson distribution.

Pile-up directly affects the performance of jets, E_T^{miss} and isolation. Pile-up interactions are usually identified by means of vertex reconstruction. If such interactions occur far enough from the hard interaction, a precise vertexing algorithm is able to detect them. Combining vertexing and tracking information allows the identification of contaminating charged particles from pile-up. On the other hand, since neutral particles do not produce tracks, neutral pile-up contamination can only be estimated *on average*.

Charged pile-up subtraction In DELPHES we assume that vertices corresponding to pile-up interactions occurring at a distance z from the hard scattering, such that $|z| > \delta Z_{vtx}$ can be reconstructed. The parameter δZ_{vtx} is the spatial vertex resolution of the detector. We assume that pile-up interactions occurring at a coordinate z, such that $|z| < \delta Z_{vtx}$ cannot be disentangled from those originating from the high-Q^2 process. Therefore every charged particle originating from such vertices cannot be subtracted from the event, while every charged particle originating from a vertex positioned at $|z| > \delta Z_{vtx}$ can be identified as originating from pile-up, provided that the corresponding track has been reconstructed. If the particle-flow algorithm is being used, the particle-flow tracks identified as originating from pile-up are removed from the list of 4-vector entering the jet clustering and the isolation procedures.

Residual pile-up subtraction Other techniques are needed in order to extract and remove residual contributions: these include particles that are too close to the hard interaction vertex to be identified as pile-up products with tracking information, charged particles that failed track reconstruction (or outside the tracker volume) and neutral particles. In DELPHES we have opted for the Jet Area method [17, 18]. This approach, widely used in present collider experiments, allows the extraction of an average contamination density ρ on an event-by-event basis. In practice, this is performed in DELPHES with the help of the FASTJET package.

The pile-up density ρ, can then be used to correct observables that are sensitive to the residual contamination, the jet energies and the isolation variable. In the presence of residual pile-up contamination, these two quantities are corrected in the following way:

$$p_{jet} \quad \to \quad p_{jet} - \rho \cdot A_{jet}, \tag{1}$$

$$I(P) \quad \to \quad I(P) - \frac{\rho \cdot \pi R^2}{p_T(P)}, \tag{2}$$

where A_{jet} is the jet area estimated via the FASTJET package, and R is the diameter of the isolation cone.

As a validation of the Pile Up implementation, the resolution in missing transverse energy has been computed as a function of the number of reconstructed primary vertices in the left plot of the Figure 2.

4. B-tagging possibilities

The identification of jets that result from the hadronization of heavy flavour quarks — typically b or c quarks — is important in high energy collider experiments. In DELPHES a purely parametric approach based on Monte-Carlo generator is available as well as a track counting implementation.

Parametric The algorithm for b and τ jet identification proceeds as follows: the jet becomes a potential b(τ) jet if a generated b(τ) is found within some distance $\Delta R = \sqrt{(\eta^{jet} - \eta^{b,\tau})^2 + (\phi^{jet} - \phi^{b,\tau})^2}$ of the jet axis. The probability to be identified as b(τ) depends
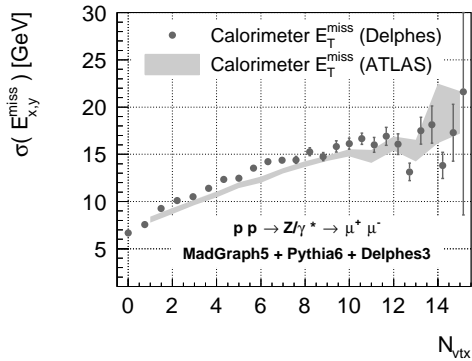
Figure 2. $E_{x,y}^{miss}$ resolution in DELPHES and ATLAS as a function of the number of reconstructed primary vertices [26]. The grey band represents the discrepancy between the ATLAS simulation and data.
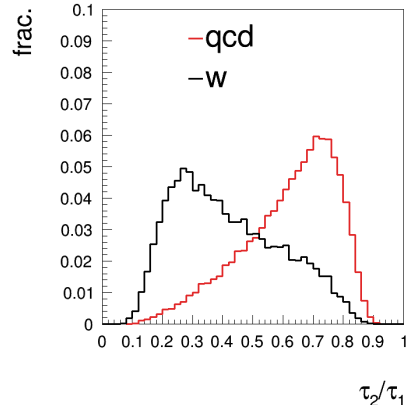

Figure 3. Ratio τ_2/τ_1 computed with N-subjettiness in DELPHES. We observe the discrimination between prompt jets and W-jets.

on user-defined parameterizations of the efficiency. The user can also specify a mis-tagging probability parameterization, that is, the probability that a particle other than b(τ) be wrongly identified. Modularity allows the user to use several b(τ)-tagging algorithms for the same jet collection and to easily implement other tagging algorithms, eventually involving an analysis of the jet constituents.

Track Counting When the Track counting b-tagging is selected, the impact parameter of each track inside all reconstructed jet is smeared according to a gaussian distribution. The width of the gaussian distribution is defined by the user and can be dependent on the p_T and η of the corresponding track. The *Track Counting* is then applied on each jet and require a minimum number of tracks (typically 2 or 3) with a high impact parameter significance for the jet to be tagged. While less precise than the parametric approach, this method provides predictive power needed for future detector design when a full simulation is not yet available.

5. Jet substructure techniques

At high energy, W, Z and Higgs bosons, as well as the top will produce only one jet when they are highly boosted, making the distinction between those objects more difficult. New techniques studying the substructure of the jets have been developed to disentangle boosted topologies from QCD backgrounds. Amongh these methods, the *N-Subjettiness* algorithm [22, 23, 24] from the FastJet-Contribs [13] has been implemented in DELPHES. The method returns the variables τ_1 (1-subjettiness) to τ_5 (5-subjettiness) which are saved as jet members in the DELPHES output. From those variables, it is then possible to build the ratios τ_N/τ_M to discriminate jets originating from N or M-prongs.

An example of application of this method is shown in the Figure 3 where it is possible to discriminate jets comming from a W-boson decay from prompt jets originating from a QCD sample.

Conclusion

Delphes is an advanced parametric simulation program allowing for a gain of four orders of magnitude in CPU time and widely tested by the community. A new version of the framework,

DELPHES 3 has been out since 2013 with major improvements: modularity, particle flow emulation, pile up implementation. Recently other new features such as the implementation of the Track Counting algorithm and jet substructure techniques were added to the framework. Thanks to those new implementations, we have shown that DELPHES is able to reproduce LHC analysis and produce qualitative predictive results.

The aim of DELPHES is to provide a simulation tool that goes beyond the run I for ATLAS and CMS. Thanks to its speed, DELPHES is able to handle the 200 PU scenario for the HL-LHC. Other detectors are also considered. Currently, Future Circular Collider (FCC), LHCb and Alice detector simulations are under preparation.

References

[1] J. Allison, K. Amako, J. Apostolakis, H. Araujo, P. A. Dubois, M. Asai, G. Barrand and R. Capra et al., *Geant4 developments and applications*, IEEE Trans. Nucl. Sci. **53**, 270 (2006).

[2] S. Hamilton, E. Kneringer, W. Lukas, E. Ritsch, A. Salzburger, K. Sliwa, S. Todorova and J. Wetter et al., *The ATLAS Fast Track Simulation Project*, ATL-SOFT-PROC-2011-038.

[3] W. Lukas [ATLAS Collaboration], *Fast Simulation for ATLAS: Atlfast-II and ISF*, J. Phys. Conf. Ser. **396**, 022031 (2012).

[4] R. Rahmat, R. Kroeger and A. Giammanco, *The fast simulation of the CMS experiment*, J. Phys. Conf. Ser. **396**, 062016 (2012).

[5] S. Abdullin *et al.* [CMS Collaboration], *The fast simulation of the CMS detector at LHC*, J. Phys. Conf. Ser. **331**, 032049 (2011).

[6] [CMS Collaboration], *Comparison of the Fast Simulation of CMS with the first LHC data*, CMS-DP-2010-039.

[7] Ovyn S, Rouby X and Lemaitre V 2009 DELPHES, a framework for fast simulation of a generic collider experiment *Preprint* hep-ph/09032225

[8] de Favereau J *et al*, 2013, arXiv:1307.6346 [hep-ex].

[9] Antcheva I *et al* 2009 A C++ framework for petabyte data storage, statistical analysis and visualization *Comput. Phys. Commun.* **180** 2499

[10] Dobbs M *et al* 2001 *Comput. Phys. Commun.* **134** 41

[11] Garren L and Lebrun P 2006 StdHep User Manual `http://cepa.fnal.gov/psm/stdhep/`

[12] Alwall J *et al* 2007 A Standard format for Les Houches event files *Comput. Phys. Commun.* **176** 300

[13] Cacciari M, Salam G P and Soyez G 2012 FastJet User Manual *Eur. Phys. J.* C **72** 1896

[14] CMS Collaboration 2009 Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET CMS-PAS-PFT-09-001

[15] Cacciari M, Salam G P and Soyez G 2008 The Anti-k(t) jet clustering algorithm *JHEP* **0804** 063

[16] Alwall J, Herquet M, Maltoni F, Mattelaer O and Stelzer T 2011 MadGraph 5 : Going Beyond *JHEP* **1106** 128

[17] Cacciari M, Salam G P and Soyez G 2008 The Catchment Area of Jets *JHEP* **0804** 005

[18] Cacciari M, Salam G P and Soyez G 2008 Pileup subtraction using jet areas *Phys. Lett.* B **659** 119

[19] Tadel M 2008 Raw-data display and visual reconstruction validation in ALICE *J. Phys. Conf. Ser.* **119** 032036

[20] D. Buskulic *et al.* [ALEPH Collaboration], *Performance of the ALEPH detector at LEP*, Nucl. Instrum. Meth. A **360**, 481 (1995).

[21] CMS Collaboration, *Measurement of the top-quark mass in tt̄ events with lepton+jets final states in* pp *collisions at* $\sqrt{s} = 7$ *TeV*, JHEP **1212**, 105 (2012) [arXiv:1209.2319 [hep-ex]].

[22] J. Thaler and K. Van Tilburg, JHEP **1103**, 015 (2011) [arXiv:1011.2268 [hep-ph]].

[23] J. Thaler and K. Van Tilburg, JHEP **1202** (2012) 093 [arXiv:1108.2701 [hep-ph]].

[24] A. J. Larkoski, D. Neill and J. Thaler, JHEP **1404** (2014) 017 [arXiv:1401.2158 [hep-ph]].

[25] CMS Collaboration, *b-Jet Identification in the CMS Experiment*, CMS-PAS-BTV-11-004.

[26] ATLAS Collaboration, *Performance of Missing Transverse Momentum Reconstruction in ATLAS with 2011 Proton-Proton Collisions at sqrts = 7 TeV*, ATLAS-CONF-2012-101.