1 Scientific Comments

1.1 PAGE 1

1.1.1 Abstract

Lines 1 and 2: "the DELPHES fast-simulation framework is presented" makes the reader think that the content of the article is purely related to the software technicalities, while the paper is instead mostly about the physics content of the fast simulation - if one excepts Section 5. [See related comment about Section 5 below.]

Suggestion : Replace "fast-simulation framework" by "fast simulation", and remove the second sentence, which is out of place in an Abstract.

Comment addressed.

Lines 5 to 9: The description of DELPHES in the abstract is too software oriented. The journal to which the preprint is submitted is called "Journal of High-Energy Physics", not "Journal of High-Energy Software". The casual reader does not really care that the program produces "collections", he cares instead about the physics content of the simulation. The suggestion here would be to rephrase the end of the abstract to indicate the simulation was enhanced with new features, needed for the simulation of the LHC detectors in the coming period (e.g., additional pile-up interactions, which will become crucial in the coming decade, or particle-flow reconstruction, which has become a salient feature in the first years of the LHC for one of the two multi-purpose detectors), and that the program simulates "physics objects" used for data analysis at hadron colliders such as "jets", "taus", "missing energy", "electrons", "muons", "photons", "isolation", "pile-up mitigation", etc.

The abstract has been revisited by removing software specific statements and referring only to the simulation aspects (with the exception of the mentioning of the "modular design" which we want to keep since it is a new feature).

It is indeed important that the concept of "hadron collider" appears clearly in the abstract. Because the simulation is analysis-oriented, it would take a number of important modifications to make it useable for analysis of e+e- collisions, for example.

We do not agree with the fact that Delphes is hadron collider specific. Delphes would process e+e- events just fine. Therefore, it would be incorrect to write in the abstract, or elsewhere, that it is a simulation for hadron colliders. At most, one could argue that it is not general enough to be able to simulate asymmetric detectors like ALICE or LHCb (which are experiments at hadron colliders...), but we are clearly stating that we want to simulate "multipurpose detectors". It is true, on the other hand, that some aspects of the default reconstruction sequence (e.g. focus on transverse quantity in the isolation module) are targeting hadron colliders. However, thanks to the modular design, it is easy to adapt high-level routine to other needs. We now make that clear in a few places in the text.

1.2 PAGE 2

1.2.1 Introduction

First (and second) paragraph:
A reference should also be made to the "fast simulations" developed by CMS and ATLAS, which are two to three orders of magnitude faster than the GEANT-based simulations, while minimizing the loss of accuracy, and allowing the "complex reconstruction algorithms" to be used.

Comment addressed.

Third paragraph:
L4: It's not enough to say that the magnetic field is uniform (as this adjective qualifies only the absolute value in Tesla). The word "axial" and the expression "along the beam direction" could be used to the sake of clarity.

Comment addressed.

L5: The energy smearing applies also to all other particles, not only to photons and electrons. Why are these two particles singled out here ?

This is correct, "photons and electrons" have been replaced by "long-lived visible particles"

L6-7: "Jets and missing energy can be computed with the particle-flow algorithm." is an incorrect sentence. First "the particle-flow algorithm" would probably need to be somehow defined, or at least given a reference. Second, "the particle-flow algorithm" does not deliver jets and missing energy, it delivers a list of reconstructed and identified particle candidates. Jets and missing (transverse) energy can then be obtained from either calorimeter deposits, or from these particle candidates. Both approaches are conceptually and technically identical for jets and missing energy. This misconception of the PF reconstruction appears in several places in the article, and most likely in several aspects of the simulation implementation too. A number of the comments that follow are related to this aspect.

We agree that the particle-flow algorithm should be defined. We call our approach "energy-flow" in the new draft: we don't aim at re-implementing the PF algorithm itself, but at emulating its effects. This would make more clear why we only apply it to jets and MET: there is no PF-like approach that we can follow for muons, electrons, etc., as those are already perfectly identified objects in our simulation. The new version of the paper make this more clear and references have been added.

The suggestion regarding the last two comments is to explain that all particle energies are smeared according to parameterized detector resolutions, and that physics objects used in physics analyses (isolated leptons, isolated photons, jets, missing transverse energy, taus) are derived from these smeared energies. The concept of calorimeter and particle flow algorithm is to be kept for the next paragraph.

Comment addressed.

Paragraph 4:
L1: Add "which was only simulating energy deposits in the calorimeters" or anything closer to the truth, after "predecessor".

Comment addressed.

L3: Add "to deliver a list of reconstructed and identified particles as close as possible to the true (generated) list." after "sub-detectors".

Not added, since the energy-flow algorithm in DELPHES simply focuses on the reconstruction of jets and missing energy.

L6: "fully modular" would need some more explanation for the reader to understand it. But is it so important for a JHEP article ?

It is important to mention the modularity since it is a crucial improvement with respect to the prior version. The modular aspects of Delphes are explained in the technical description part, which was moved in the appendix section. At this stage it is enough for the non-software expert reader to know that modularity implies greater flexibility to the user.

Paragraph 5:
L2/L3: Propose to drop. The software implementation is out of context.

The software description in the introduction has been dropped since the whole software related section has now been moved to the Appendix.

1.3 PAGE 3

Par 2:
L5/6/7: This sentence starts with "As for the tracking efficiency", but nothing was said about the tracking efficiency settings prior to this sentence. A mention of the fact that this efficiency can be user-defined should appear at the beginning of the paragraph, and replace "(good)".

Comment addressed.

1.3.1 Section 2.2

Par 2:
L3: "in the transverse direction" is at best redundant, but has actual very little meaning here. Transverse to what? Suggest to drop "in the transverse direction" with no loss of information.

Comment addressed.

L4: It is bad to assume the same granularity for ECAL and HCAL, as it is in general not the case in HEP detectors, for very good physics reasons (in the sense that (i) electromagnetic showers are much more compact than hadron showers, and (ii) it has important consequences for the performance of the particle-flow reconstruction.). The assumption might actually be the origin of the some of the disagreements seen in the particle-flow performance later in the article. The "computational reasons" are not spelt out, but it is difficult to understand why computing limitations would force anyone to make such an unrealistic assumption.

Showers are not produced in Delphes since the particle momenta are simply smeared according to the relevant calorimeter resolution. The discreteness of the calorimeter has simply an impact on the angular resolution of the final observables. As it is, the pessimistic angular resolution assumed in ECAL only affects photons, since for electrons we have the track information and we assume infinite resolution. The "computational reasons" are that with the current state of the implementation, the combinatorics are reduced, and this has an impact on the the jet clustering procedure, especially in the presence of pile-up. Improvements with this respect are considered for future releases.

Par 3:
L3: "Neutral pions" do not leave their energy in ECAL, as they decay promptly to two photons (i.e., they are not "long-lived particles reaching the calorimeters". Suggest to drop "and neutral pions".

Comment addressed.

L3-7: "while charged pions and other neutral hadrons deposit all their energy in HCAL. Long-lived particle such as kaons, pions and Lambdas are considered stable by most event generators. In Delphes, such particles are assumed to deposit a fixed fraction of their energy both in ECAL and HCAL. By default, fECAL is set to 30% and fHCAL to 70% according to their expected decay products ... etc." There are several problems with the logic of the above:
1. The two sentences contradict themselves : do charged pions deposit all their energy in HCAL, or in both ECAL and HCAL?
2. In the second sentence, one should be more specific and write e.g. charged pions instead of pions (pi0s are not long lived).
3. Are "such particles" stable in Delphes? If yes why is there a discussion about their

decay products? Are fECAL and fHCAL fixed, or do they depend on the decay products (if any)? How exactly do fECAL and fHCAL depend on the decay products? What does "according to their decay products" mean ? What are the decay products of charged pions in DELPHES ?
All in all, the whole paragraph needs substantial work, and the DELPHES implementation might need serious revision if what is currently described is indeed implemented.

The whole paragraph has been re-written for better clarity. The "according to their decay products" expression was referring only to lambda and k-shorts, whose decay products are considered, as an approximation, on average 30% electromagnetically interacting, and 70% strongly interacting. On the other hand charged pions deposit 100% in HCAL if it has not been decayed by the event generator. These are set by default in the configuration file, but they can, as explained in the text, easily be changed by the user.

Eq 2.1 : It is not clear whether the same resolution is used for ECAL and HCAL. It is not clear either whether the ECAL resolution is different for photons and for those hadrons that leave 30% of their energy in the ECAL. It would be very useful for the reader's understanding to have a table of the values of S, N and C used to reproduce the CMS and ATLAS performance. Are these values compatible with the actual CMS and ATLAS resolution, or do they have to be tuned to reproduce the performance ? Along the same line, the calorimeter granularities and the tracker resolutions used for the two detectors would need to be spelt out and compared to the actual values.

For producing the plots we use the nominal resolutions from CMS and ATLAS. The resolution is different for ECAL and HCAL (the text has been changed to make this clear). We are not in favour of quoting the CMS and ATLAS resolutions, since, at this level, we want to keep the discussion general and not give the impression that Delphes is limited to these two experiments. We would like to emphasize that Delphes can be used with completely different parameters, corresponding to any generic (symmetric) detector. Of course, actual values of the resolution function used are given in the publicly available configuration file such that the Delphes plots could in principle be reproduced by anybody. We also believe that adding the content of the steering file in the Appendix would not be appropriate.

Eq 2.2. : Several problems here too.

It is not clear whether the shower energy is or is not distributed over several towers. Neither Eq 2.2 nor the text seems to mention that. I seem to understand that the energy of each particle is concentrated in a single tower from the algorithm described later on, but the casual reader will certainly miss this subtlety.

Comment addressed.

ECAL and HCAL are undefined, even though the casual reader may go as far as guessing that they are defined by equation 2.1 (?)

Comment addressed.

What is the physics motivation for doing a log-normal instead of a Gaussian smearing?

*The lognomal distribution ressembles to a gaussian when mean > 6*sigma, that is for most values at high energy, but has the advantage at low energy of being always positive. This ensures to avoid the positive bias in the effective mean and the s.d induced by having a truncated gaussian. In that sense, this is a purely ad-hoc choice.*

To define a log-normal distribution, one usually gives the mean and sigma of the logarithm of the distribution, which is normal. Here, are the authors talking about the mean and variance of the log-normal distribution? I guess so, but it would be good to clarify.

We are talking about the mean and the variance of the lognormal, that have a pretty complicated expression in terms of the mean and variance of the underlying normal variable, which, as the referee correctly remarked, are the ones that are usually given. However, it is clear from the text that "m", and "s" are the mean and variance of the log-normal distribution and not the mean and variance of the underlying normal distribution.

1.4 PAGE 4

L2/3: It is difficult to understand why one would want to "avoid having to deal with discrete tower positions". "Discrete tower positions" are actually happening in CMS and ATLAS, and are being dealt with without difficulties. The authors may want to be more explicit about their motives here.

If the detector is not highly granular and cell size are irregular, spikes can appear in angular distributions, if proper binning is not chosen. In Delphes since the same cell sizing is chosen for HCAL and ECAL spikes can appear in the forward region where the detector is not highly granular. Therefore we apply uniform smearing for purely cosmetical reasons, and has indeed no impact whatsoever on the angular resolution of calorimeter towers.

1.4.1 Section 2.3

Par 1:
L2: "reconstructing the event" → "reconstructing all the individual par- ticles in the event".

As said in the comments in the introduction, we decided to call the algorithm "energy-flow" in Delphes. We believe it is more appropriate since in Delphes, such algorithm is *aimed at optimizing the performances of jets and missing energy. Particle identification is perfect by construction in Delphes, since it is based on MC truth. Following the referee's suggestions, the energy-flow algorithm has been sensibly re-visited, as well as the section 2.3 of the paper.*

Par 2:

L5: Drop "if particle-flow is switched on" are it is obvious in the context of Section 2.3 "Particle Flow Reconstruction". (Two occurrences.)

Comment addressed.

L5: "We assume it is always convenient to estimate charged particle mo- menta via the the tracker." This is a wrong assumption. As the transverse momentum or the pseudo-rapidity increases, the transverse momentum reso- lution becomes worse than the calorimeter resolution. This assumption may be the reason of the disagreement between DELPHES and CMS in Fig.5 (left), where the jet pT resolution is significantly pessimistic at high pT. This caveat must be mentioned in the text, either here, or when discussing Fig. 5, or (better) in both places. The same comment applies to the last sentence of PAGE 4 and the first sentence of PAGE 5.

It has already been mentionned in the text that this assumption is true only up to some energy threshold, and that we adopt it nevetheless for any energy. After having re-implemented the energy-flow algorithm and documented in this new version of the paper, the agreement is now found to be good at high energy (see Figure. 3, left).

Par 3.
There seems to be here again an overall misunderstanding of what a particle-flow algorithm is for. The authors seem to believe that it is aimed at reconstructing jets and missing energy. The particle-flow algorithm aims at reconstructing all individual particles in the event with an optimal resolution by making use of the identification capabilities of a detector. It can therefore provide a list a photon, charged leptons, and charged/neutral hadrons, that can be later used to define all sorts of physics objects - not limited to jets and missing energy.

See previous answer about our decision to talk about "energy-flow". We don't want to perform a real reconstruction, and therefore a real particle-id algorithm, but to emulate its effects. The gains from PF are larger on jets and MET than on other high-level objects, and therefore the need for an emulation of PF effects is stronger for jets and MET.

First bullet: "Hits" are not defined, and it is difficult to understand the concept of a "hit that originate from a particle". The expression "at least one among fECAL and fHCAL is non-zero" carries little meaning. The footnote content is not related whatsoever to the information in this bullet.

We agree and now the comment not relevant anymore, since the whole section has been re-written.

1.5 PAGE 5

The two bullets here contain quite involved a logic, which is difficult to follow even by experts. The suggestion is to work the text out and come with a clearer version.

Comment not relevant anymore.

First bullet :
L6: Add "and he corresponding hits are dropped." after "such tracks get stored as particle-flow tracks".

Comment not relevant anymore.

L8/9: The energy smearing was already addressed earlier in the text. Why repeating it here ?

Comment not relevant anymore.

L9/10: Do I understand properly that when a charged pion and a pho- ton leave energy in the same tower, the PF algorithm is assumed to be smart enough to find the photon, irrespectively of the ECAL granularity and the photon energy ? That's overly optimistic, and it does not allow the DELPHES user to make studies about the relevance of a better calorimeter granularity, for example. On the other hand, the assumption that, when an electron and a neutron point to the same HCAL tower, the e ID is smart enough to detect it, is almost correct for most detector designs.

Although this is an un-realistic assumption, it shows to correctly reproduce the performance. However, as the referee correctly points out, this assumptions restricts the variety of detector studies that can be performed with Delphes. We point out, however, that Delphes is primarily a tool used by phenomenologists, who do not need to make advanced detector studies. We believe that these should be done inside experimental collaborations by means of detector-specific simulation tools (based on Geant or not).

Second bullet:
L6/10 : "The resolution will be exactly the same. It is therefore useless ... the full calorimeter tower." This logic is incorrectly representing that of a sound particle flow algorithm. While it is true that the resolution (and actually the value) of the energy would be (not "will be") the same, replacing a charged hadron + a neutral hadron by the sole calorimetric energy deposit has several drawbacks for data analysis. First, it artificially reduces the reconstructed charged multiplicity - which may be precious, e.g., when determining the charged isolation of a particle. Second, it reduces the ability

of pile-up mitigation (mentioned in the next paragraph), by losing the origin vertex information. Third, it worsens the angular resolution of the jets, that become limited by the tower granularity. Fourth, it does not follow the particle-flow philosophy that aims at reconstructing all particles in an event. Also, the logic of the two bullets misses an important point : when the calorimetric energy is compatible (within a small number of st. deviations) with the track momentum, no neutral hadrons is created even if there is one; and when the calorimetric energy is in excess of the track momentum, a neutral hadron is always created, even if there is none. The current implementation DELPHES misses both aspects, which tends to explain the too good resolution of jet pT at low pT.

The version of the energy-flow algorithm presented here, was designed specifically to address these concerns by the referee. All the reconstructed tracks are now stored as energy-flow tracks. Fake-neutral hadrons are now created when the calorimeter energy excesses that of the tracks. The old scheme was a good approximation in the absence of PU, but that indeed it was noticed to be insufficient when PU is introduced, as the referee pointed out.

Last par, last line : It is not true that the emulation of the PF algorithm reproduces the performance of, e.g., CMS, even for jets. (See related com- ments later.) Again, it would be interesting for the reader to understand the resolution parameters used in DELPHES to get to this performance.

The disagreement in Figure 5 has been sensibly reduced with the present energy-flow implementation. A discrepancy is still observed in the low energy range (20-30 GeV). We are satisfied with the present performance, since with increasing luminosity (hence higher trigger rates, increasing pile-up), most physics analyses will reject such low energy jets anyways.

1.6 PAGE 6

1.6.1 Section 3.1.1

Par 1, L3/4 : "while leptonic decays can be indirectly studied from the decay products when processing the DELPHES output" is very vague a sentence, with no real meaning. Suggestion is to drop it. The rest of the section is extremely verbose, and could be replaced by one sentence stating that electron and muon energy/momentum is smeared with resolutions parameterized as a function of pT and eta (and mention that these parameterizations can be changed by the user?). The convoluted explanations about what a typical collider experiment is actually doing is of no interest for the reader, as DELPHES does not do the same anyway. This text could be replaced by a figure showing the resolutions used to reproduce the CMS and ATLAS performance for electrons and muons, as well as a comparison with the actual detector resolutions.

The verbosity of the paragraph has been reduced as suggested by the referee. The resolutions used to reproduce ATLAS and CMS resolutions, are the ATLAS and CMS *resolutions themselves and they are shown in the validation section. In addition, they would be out of context here, since the discussion is supposed to stay general at this level.*

1.6.2 Section 3.1.2

Last line : "Neutral pions are automatically classified as photons". As al- ready mentioned, neutral pions are not long-lived particles in any generator. Instead they decay promptly to two photons, which obviously are classified as photons. The authors might want to drop the comments about neutral pions.

Comment addressed.

1.7 PAGE 7

1.7.1 Section 3.2.2

L2/3: "as these might indicate the production of heavy unstable particles" carries little meaning is the context of this article (and probably in a wider context too) without additional explanation. Suggestion is to drop it.

Comment addressed.

1.8 PAGE 8

1.8.1 Section 3.2.2

Par 1: L1/2 : It is not useful to indicate what is done in real experiments. Instead, it is important to describe what is done in DELPHES.

Comment addressed.

1.8.2 Section 3.1.3

As a general comment, the isolation definition chosen here is very much hadron collider biassed. One would not do the same in e+e- collisions. This comment supports the initial request that the abstract includes a sentence stating that DELPHES is aimed at simulating hadron collider experiments (so far).

That is correct, but one of the reasons why we refactored Delphes to make it fully modular is to simplify the task for any user who wants to code his/her own definition of isolation, therefore we don't want to send the message that Delphes is for hadron colliders: the default definition of isolation is appropriate for most of the users, because *most of the users nowadays are interested in simulating the LHC conditions, but users interested in ILC, LEP3 (or even LEP1/2) studies only need a very minor amount of work to rewrite the isolation module in Delphes. Actually isolation was one of the most mentioned use cases for modularizing the code, as very different definitions may be used within the very same experiment. A sentence has been added to clarify that the default isolation definition is hadron collider specific.*

1.9 PAGE 9

"Charged pile-up subtraction"
There are several problems in this paragraph.
It is not stated if this paragraph is specific to PF or not. If it's not, the procedure to remove charged particles from the event in which purely calorimetric jets are reconstructed needs to be spelt out clearly.

Two sentences have been added at the end of the paragraph to clarify this point.

Even if the paragraph is PF specific, it is not clear what "subtracted from the event: means. For example, are the pile-up charged hadrons removed from the event before computing the missing transverse en- ergy ? To the best of my knowledge, it is not what is done in LHC experiments.

Charged particles are not removed for calculating the missing energy since this would degrade the overall missing Et resolution. The last sentence should now clarify what we mean by "subtracted from the event".

The criterion "a distance $|z| > Zvtx$" is obscure. What if the hard interaction is produced with $|z| > Zvtx$? and what is the definition of this "distance" ? If is is the distance between the vertex of a PU interaction with respect to the vertex of the hard interaction, the use of "z" instead of "z" (with the proper definition in the text) is in order.

A sentence was added at the beginning of the paragraph to clarify that in Delphes, every hard scattering occurs at (0,0,0) coordinates. The notation of "Zvtx" has been changed as suggested.

"Residual pile-up subtraction"
It is not clear how "rho" is obtained in DELPHES.

A sentence stating that we use FastJet to estimate "rho" has been added.

"mainly the jet energies and the isolation" : is "rho" used for anything else ? if yes, it should be stated. If not, "mainly" should be removed.

Comment addressed.

1.10 PAGE 10

Par 2:
L4: It would be useful to mention one of the advantages of the PF recon- struction in the context of PU mitigation, namely the fact that the calorime- ter energy deposits associated to PU charged hadrons are "automatically" removed. In DELPHES, however, this does not happen when a charged hadron and a neutral deposit fall in the same calorimeter tower, because the charged hadron is then ignored by DELPHES.

Thanks to the new implementation of the energy-flow algorithm, every charged particle originating from pile-up, as well as its corresponding calorimeter deposit can now be removed prior to the jet clustering. The sentence has been changed in order to account for this comment.

1.10.1 Section 5

The suggestion is to remove this section altogether. It is not detailed enough to be useful to a potential user (in contrast to a user manual), there is also not enough information to understand how things are truly implemented, but it still contains details irrelevant for a publication in JHEP.

The section 5 has been moved to the appendix as it has been done for other software related papers published in JHEP such as MadGraph5 (JHEP 1106 (2011) 128). We believe that it is important that the information contained in this section, although very basic, appears in the official Delphes public publication. For the reader it important to know which dataformat can be used as an input to Delphes, how the data-flow works, and finally, what are the basic performances in terms of speed and memory usage, particularly important for a "fast-simulation software".

1.11 PAGE 11

Figure 1 is a striking illustration of the previous comment. This diagram is both too detailed for a JHEP publication, but too simplistic and not readable.
For example:
there is no mention of PF, which is an important addition in DELPHES3

Comment addressed.

there is no mention of PU subtraction, another important addition in DELPHES3,

Comment addressed.

the workflow is neither specific and accurate

Comment addressed.

the figure is cut on its right part

Comment addressed.

the caption lacks a complete discussion

Comment addressed.

1.12 PAGE 12

Figure 2 is yet another illustration of the same problem:

there is no explanation of what is the grey band in the right plot, there is not explanation of why there is no grey band in the left plot.

The grey band has been replaced by more usual vertical error bars.

the caption lacks a discussion

added discussion in the caption.

As the suggestion is to remove the section altogether, no further comment is made in this report about it.

moved to appendix and expanded

1.12.1 Section 6.1

Last line: What does "alternatively" mean ? If one choice is for electrons and the other for muons, the authors should state it clearly.

the last sentence has been changed to make the statement more clear.

Footnote: this statement deserves a complete section, with the list of parameters used in the default CMS and ATLAS configurations, as well as possible explanations as to why this parameter choice was made, and a comparison with the actual CMS and ATLAS resolutions and granularities.

We disagree with this comment. The reasons are:

- *the resolutions in the Delphes CMS and ATLAS cards are taken directly from the cited papers, and it would be redundant to quote them here.*

- *the only difference is in the calorimeters granularity. Both the CMS and ATLAS configurations in Delphes use the granularity of the HCAL detector. As said in the calorimeter section, the ECAL granularity is exactly the same as the HCAL granularity in Delphes. This comment has now been added. A table with the actual HCAL granularity of the LHC experiments (already public in the relevant technical design reports of CMS and ATLAS) would be of poor interest to the reader and redundant.*
- *We now clearly state that the validation plots obtained in the publication are obtained with version 3.0.11 and the reader can easily read the specifications of the detector from the relevant CMS and ATLAS cards that are distributed with DELPHES v3.0.11.*

1.13 PAGE 13

1.13.1 Figure 3

It is not clear what the grey bands are in this plot. Shouldn't they be removed? In CMS, they are supposed to cover differences between the simulation and the data in CMS, not the difference between Delphes and CMS. This comment is valid for all plots. Strangely enough, the ATLAS red band width is way smaller than that for CMS. Does it represent the same thing ?

The grey bands mean different things in the left and right plots. The caption has been extended in order to explain the details required by the referee.

For all plots, it is important to have the statistical uncertainty bars indicated, or to state that they are covered by the size of the markers. In the latter case, an explanation is needed for the apparent scatter of the DELPHES points, and to compare this scatter with the input resolution function.

The apparent scatter is just due to the fact the the parametrisation of the resolution is binned, and has been chosen to match approximately that of CMS and ATLAS. The choice was made to adopt round values which may result in the apparent scatter. As these plots are just an illustration of a parametrisation which is correct by construction, we believe that no further explanation is needed.

For all plots, label, legends, etc... are way too small to be readable.

This comment has been addressed for all the plots.

1.13.2 Figure 4

Looking at Ref. [20], and in particular its slide 33, my impression is that the DELPHES resolution is a factor 2 optimistic with respect to the CMS resolution. It seems that DELPHES parametrized the Gaussian width of the core of the CMS resolution, rather than the effective 68% width.

the referee is correct, the electron resolution was parametrized with the Gaussian width, since the resolution in Delphes is Gaussian by construction. By proceeding this way we are voluntarily neglecting the tails effects in Delphes. This detail has been added in the caption.

1.14 PAGE 14

1.14.1 Figure 5

On the CMS side, it would be nice to have an explanation for the following effects (already alluded to above):
why does the calo jet resolution curve saturates at low jet pT ?
why does the PF jet resolution curve of Delphes departs from CMS at high pT? (it's probably because the tracker pT resolution is used to determine the PF track pT) And at low pT ? Actually, the DELPHES jet pT resolution is almost independent with pT: it goes from 10 to 7% when varying the pt from 30 to 500 GeV/c, while the actual CMS PF varies from 14% to 5%.
why does the PF curve show a discontinuity between the 1st and 2nd points ?

This plot has been re-done. The ECAL/HCAL calorimeter resolutions have been set to the actual CMS resolution, and the energy-flow implementation is as explained in Section 2.3. As a result we have a perfect agreement at all medium and high pt values. At low pt there is a discrepancy for $20 < pt < 30$ *which is not understood. However we believe this discrepancy to have very low impact on physics analyses, that most often consider jets with* $pt > 30$*. The small discrepancy observed in the* $30 > pt > 40$ *GeV bin is 1% in resolution. These comments have been added in the text.*

Once the effects are understood, it would be important to fix the imple- mentation in DELPHES. These differences are important for data analysis, and may to DELPHES user draw wrong conclusions from his DELPHES studies (of particular importance if DELPHES is used to define the upgrade strategy of the expensive LHC detectors).

Comment addressed. However, the referee, as well as the LHC experiments, should be aware that any study performed with Delphes should be understood as preliminary. We made this point clear when consulted by the LHC collaborations. Delphes is designed as a pheno tool, not as a replacement of fast simulation tools from the collaborations. A Delphes based study should be perfomed after a pure parton-level and before a study *with Geant or with the specific fast simulations of the collaborations. As a result, we are perfectly happy with an agreement with a few percents discrepancy in the physics object resolutions.*

1.14.2 Section 6.3

Par 2:
It is not clear if the CMS study is made with of without pile-up. As pile-up and its particle-flow mitigation are an important adds-on to DELPHES 3.0, it would be nice to have an illustration of their performance here.

We do not want to create to much imbalance between CMS and ATLAS (the only exception is electrons, but we did not find the relevant plot for ATLAS), so we have decided to produce a real MET validation plot for CMS and the fake validation for ATLAS. This choice was simply driven by the fact that the real MET validation plot was not found in the ATLAS. In addition, pile-up mitigation for MET has not been addressed in this note, since it is not done. As a matter of fact PU mitigation on the MET relies on complex multi-variate algorithms in CMS which are out of scope in Delphes. The only advantage in using the energy-flow objects instead of calorimeter tower for computing MET in the presence of pile-up is the superior resolution of track in comparison to towers. PU tracks cannot simply be removed (in contrast to what is done for jets).

1.15 PAGE 16

Par 1:
L3: The anti-kT algorithm has no "cone" - and "a cone R = 0.5" has no meaning even for a cone algorithm. (Note: the same mistake occurs in Section 7.2)

Comment addressed.

L9/11: The slight difference of efficiency is a large difference (20%), which might the DELPHES user draw incorrect conclusions from the abilities of his/her analysis. The explanation given here should be checked, and the cul- prit (jet energy correction or b tagging efficiency should be fixed in DELPHES.

We totally disagree with the referee here. The examples given here a purely illustrative and it is out of scope to fine-tune them. The purpose was precisely to show the opposite: without particular fine-tuning Delphes gives very reasonable agreement with the CMS analysis. We insist that a 20% difference is an acceptable difference since very often results (rates, efficiencies) obtained with full geant based simulation give larger discrepancies than 20% with respect to data. Most LHC analysis nowadays make use of signal-free regions to normalize backgrounds, and scale factors in excess of 1.2 are not unusual. More generally, private communications from CMS collaborations recently *confirmed that an excellent agreement is observed when comparing Delphes to internal studies made with full simulation. Unfortunately, these are not results that we are allowed to show. As a side comment, the efficiency was re-computed after the change in the energy-flow algorithm and the result was found to be the same.*

Par 2:
Bullet 1: "any parton from the top quark decay" ! "any parton from the decay of either of top quarks".

Comment addressed.

I find the definition of unmatched rather unnatural. If the three jets from the "hadronically-decaying top" were matched, I fail to understand why the event is classified as unmatched, even if the other b for the other top is unmatched.

The categories definitions were chosen in the cited CMS paper. We agree with the referee, however, for the sake of comparing to the CMS results we have decided to adopt the same definitions used in the CMS paper. We have cross-checked with the CMS authors to make sure that this is indeed the definition that was used.

Par 3:
L5: Why are the distributions not normalized to the number of events ? Is it because the number of permutations/event is vastly different in DELPHES and in CMS ? If it is the case, shouldn't DELPHES be fixed ?

The Delphes distributions are normalized to the total number of events in CMS (to account for the 20% in the selection efficiency). The total number of permutation is proportional to the event yield. The purpose of this example is indeed to show, as Table 1 illustrates, that we get the correct fraction of each permutation category in Delphes.

L9/10: "Pile-up, not considered in the present study, can degrade the jet energy resolution." : A strong emphasis is put on the ability of Delphes 3.0 to simulate pile up, and to simulate its mitigation procedures based on the PF reconstruction. Since Delphes is fast, I would assume that it would take no time to redo this study with pile-up. It is disappointing for the reader not to see this validation in the paper. It actually casts doubts on the DELPHES ability to accurately simulate pile-up and its mitigation, which is surely not what the authors aim at.

The whole analysis has been re-done with the new implementation of the energy-flow algorithm. The discrepancy in the top mass resolution is not present anymore, the reason being that the resolution of the new energy-flow jets matches almost exaclty that of the CMS particle-flow jets at the energy of interest (60 GeV). We attribute therefore the small discrepancy that was previously observed to the imperfect implementation of the previous particle-flow algorithm rather than to the absence of pile-up simulation. The whole sentence mentioning the discrepancy has been dropped, since the discrepancy is not present anymore.

1.16 PAGE 17

1.16.1 Figure 7

The bottom inserts of all plots are difficult to understand. The label "rel. diff." makes the reader guess that they show the relative difference (i.e., the ratio - 1) of the two distributions, but a quick look at the distributions leads the reader to doubt about it. For exampe, the right plot shows a 20difference between the two distributions around the maximum, which is not visible in the "rel. diff." plot. Maybe the wrong scale was chosen for the bottom inserts ?

The range has been changed to (-0.5,0.5) for "y" axis. The 10% (rather than 20%) difference the referee refers to is now visible.

1.17 PAGE 18

Par 4: The reader is again disappointed to see that, in the search for VBF- produced Higgs boson with pile-up, for which it is said that pile-up has a pretty large and negative impact, the authors decided to use calorimeter jets instead of particle-flow reconstruction, aimed exactly at mitigating pile-up effects. Again, it casts doubts on the ability of DELPHES to simulate pile-up in particle reconstruction, and to simulate its mitigation with particle-flow reconstruction. The paper has therefore the effect opposite to what the authors are aiming at.

We have followed the referee's suggestion and re-done the study considering energy-flow jets and energy-flow based pu mitigation.

Criterion 2.
The pT cut used to count light jets between j1 and j2 ought to be given.

It is said in the text that the requirement refers to the two leading jets non b-tagged jets. The pt threshold are defined by requirement 1.

Are these four cuts used in the CMS analysis which the authors are using for comparison ?

These cuts are probably similar to those used in any VBF analysis performed in ATLAS or CMS. However, the goal here is not to compare to any existing analysis, but rather to give a simple example of utilisation of Delphes.

1.18 PAGE 20

Par 2:
L4: "accurate productions in high pile-up scenarios should solely rely on full simulation tools" is very strong, and probably incorrect statement, and should probably be removed (or seriously rephrased). First, it casts again large doubts on the pertaining DELPHES ability to simulate pile up interactions. Second, the fact that, maybe, DELPHES cannot deal with high PU en- vironment does not mean that other fast simulation tools cannot do. For example, it seems the CMS fast simulation deals pretty well with the high pile-up produced by LHC in 2012.

"High pile-up scenarios" has been replaced by "extreme pile-up scenarios", which is what we actually meant. To our knowledge there is no evidence that fast-simulation can cope to >100 simultaneous interactions environments, simply since these did not occur in any hadron collider yet. We have reformulated the relevant sentence saying that Delphes has not yet been compared to fullsim at extreme PU. Indeed, once done, it may become more quantitative in that region too. Still, by the time we reach such pileup conditions, experimental collaborations may find ways to cope with pileup that are not foreseen in Delphes.

2 Language/Style comments

2.1 PAGE 1

2.1.1 Abstract

Line 4: "hadronic calorimeter" ! "hadron calorimeter". The authors are invited to look for all instances of "hadronic calorimeter" and modify ac- cordingly. [This comment is not repeated for other occurrences.]

Comment addressed.

Line 5: "allows to produce collections" is not proper english. Replace either by "allows the easy production of collections" or "allows collections to be easily produced". The authors are invited to look for all instances of the verb "to allow" in the text, as it most often incorrectly used throughout. [This comment is not going to be repeated for the other sections.]

Comment addressed when needed. It is however correct to say: "allows the user to store collections .."

Line 5: The word "collections" appears out of context here, and carries no meaning to the casual reader.

"collections" changed to "physics objects".

2.2 PAGE 2

2.2.1 Introduction

Second paragraph:
L3: ”parametrisation” has the wrong spelling. How about using ”param- eterization”, which is the proper word. The authors are invited to look for all instances of ”parametris...” in the text, as this comment is not going to be repeated for the other sections.

Comment addressed.

Third paragraph:
L1: A sentence must not start with an acronym (here ”DELPHES”). The authors are invited to look for all instances of DELPHES, and rephrase when at a start of a sentence.

DELPHES is neither an acronym nor an abbreviation; it is therefore fine to start a sentence with it.

L1: No hyphen between ”data” and ”format”. Actually, ”data-format” can be dropped altogether without losing information.

Comment addressed.

L4: ”final observables” of photons and leptons is jargon, and carries lit- tle meaning to the casual reader. The authors probably mean ”measured energy” ?

Comment addressed.

L5: ”the relevant sub-detectors” is also very much unclear. How about simply ”the detector resolution”

Comment addressed.

L6: ”High-level reconstructed quantities such as” carries little meaning to the casual reader. The sentence would read better without it.

Comment addressed.

L7: Why are ”calorimeter deposits” qualified as ”simple” ? Suggest to drop it. Why is the ”particle-flow algorithm” qualified as ”so-called” ? Sug- gest to drop it.

Comment addressed.

Fourth paragraph:
L4: ”b” and all particle names ought to be written in roman style. (True for the whole paper.) No hyphen between ”tau” and ”tagging”.

Comment addressed.

Fifth paragraph:
L2: "will be described/given" ! "is described/given in Sections xxx" (two occurrences). The authors are invited to look for all instances of "will", and replace it by the present tense in most cases.

Comment addressed.

2.2.2 Section 2

L3: I am not sure of "symmetric along the beam axis", especially the word "along". How about "with a cylindrical symmetry around the beam axis" ?

Comment addressed.

2.2.3 Section 2.1

L2: What is the meaning of "The magnetic field is applied" ? Applied to what ? Maybe the authors want to say "is assumed to be localized" ?

Comment addressed.

2.3 PAGE 3

L2: The magnetic field is not "solenoidal", it is "axial" (because it is pro- duced by a solenoid).

Comment addressed.

Par 2:
L1: What is the meaning of "(good)" ? Maybe the authors want to say "in general high" ? "Seen" ! "reconstructed"

Comment addressed.

L2: "which provide a direct measurement of their momentum": I am not sure of what "provide(s)" this measurement in this sentence. How about starting a new sentence indicated that the measured curvature to the recon- structed trajectory and the magnetic field intensity allow the momentum to be measured.

Comment addressed.

L3: "The angular resolution is assumed excellent" ! "A perfect angular resolution is assumed"

Comment addressed.

2.3.1 Section 2.2

Par 1:
L3: ”strongly interacting particles” ! ”long-lived charged and neutral hadrons”

Comment addressed.

Par 2:
L5: ”calorimeter object” is an undefined concept here. How about ”calorime- ter energy deposit (later called hit)” ? ”center” ! ”centre” (unless the au- thors decide to write in american english).

Comment addressed.

2.4 PAGE 4

2.4.1 Section 2.3

Par 1:
L3: ”several collaborations” : Ref. [3] points to the CMS Collaboration. What are the other(s) ?

Added reference to energy-flow at LEP

L3: ”intrinsically” has little meaning here. Suggest to drop.

Comment addressed.

Par 2:
L1: ”higher” ! ”better”

Comment addressed.

2.5 PAGE 5

2.5.1 Section 2.3

Last paragraph, L1 : ”consists in” ! ”consists of”

Comment addressed.

2.6 PAGE 6

2.7 Section 3.1.3

L1: Why is "isolated" in italics ? "surrounding" ! "surroundings", or better "vicinity"

Comment addressed.

L2: Starting a sentence with "Requiring" (or any verb in "ing") is bad English, as there is nobody in the sentence to do the action of "requiring". The authors are inviting to look for all instances of the "... ing" form, and check if there is somebody or something in the sentence to do the action of "...ing" (and otherwise fix the sentence).

Comment addressed.

2.8 PAGE 7

2.8.1 Section 3.2.2

Title : no hyphen between "tau" and "jets"

Comment addressed.

L1: "Identifying" is bad. "tau-lepton" ! "tau". "flavor" ! "flavour".

Comment addressed.

L2: No hyphen between "c" and "quarks". "Crucial" ! "important". "experiment" ! "experiments"

Comment addressed.

2.9 PAGE 8

2.9.1 Section 3.2.2

Par 2:
L3: It is not clear what "parton" refers to in the Delta R formula. Maybe the authors want to replace it by "b, tau" ?

Comment addressed.

L6: "is wrongly" ! "be wrongly" (subjunctive is in order here).

Comment addressed.

2.9.2 Section 3.3

L(N-1) : "particle-flow candidates" is an undefined concept at this level.

Comment addressed.

2.10 PAGE 10

Par 2:
L1: "residual" ! "neutral"

Comment addressed.

L3: "automatically" : it is not clear what the meaning of "automatically".

Comment addressed.

2.11 PAGE 15

2.11.1 Section 7.1

Par 1:
L1 : "High Energy" ! "high energy"; What is the meaning of "most common" ? How about "most copious" ?

Comment addressed.

L4: "two jets originating from one b quark" is possible, but is probably not what the authors want to say.

Comment addressed.

L6: "the hadronic top-mass" 1) is jargon, 2) has one hyphen too many. How about "the mass of the hadronically-decaying top quark" ? "We will reconstruct" is bad too.

Comment addressed.

Par 2:
L1: "center" ! "centre"

Comment addressed.

L5: I am not sure that the "DELPHESANALYSIS" package is relevant for the clarity of the paper.

Comment addressed.

2.12 PAGE 16

Par 1:
L4: "b-tagged" is jargon. "tagged as originating from the hadronization of a b quark" would be better.

Comment addressed.

L5: "criteria" ! "criterion".

Comment addressed.

L8: While I have no doubt that the signal selection is sensible, I guess that the authors want to say "sensitive".

Comment addressed.

L9: Drop "slight".

Comment addressed.

L10: "efficiencies value" ! "efficiency"

2.13 PAGE 17

Figure 7, 8, ... : "ref." ! "Ref."

Comment addressed.

2.13.1 Section 7.2

L1: "Searching the Higgs particle produced via VBF, and decaying to a bbbar pair" is the climax of improper use of English. 1) "Searching" is bad because there is nobody to do the action of searching in the sentence. 2) The correct use of "to search" is "to search for". 3) "decaying" is bad, because the readers thinks that it applies to the same missing person that does the action of "searching [for]", and that later "decays". I would like to re-iterate the suggestion to ask a native English speaker to read and fix the paper throughout.

Comment addressed.

2.14 PAGE 19

Par 2;
L1: "figure 10" ! "Fig. 10"

Not addressed, since jhep manual says abbreviation on "figure" should be avoided.