Notes on recasting the "ATLAS-EXOT-2019-23" search for pairs of displaced hadronic jets in the ATLAS calorimeter

Louie Dartmoor Corpe[1], Andreas Goudelis[1], Thomas Chehab[1]

[1] *Laboratoire de Physique de Clermont (UMR 6533), CNRS/IN2P3, Univ. Clermont Auvergne, 4 Av. Blaise Pascal, F-63178 Aubière Cedex, France*

Version 1.0

***Abstract*— This note describes the validation procedure for the provided material allowing the reinterpretation of an ATLAS search for decays of pair-produced neutral long-lived particles decaying in the hadronic part of the calorimeter, or at the edge of the electromagnetic calorimeter, using the full Run-2 ATLAS dataset. The reinterpretation material includes an efficiency map linking truth-level kinematic information (position, transverse momentum and decay products of the LLPs) to the probability of the reconstructed event being selected in the analysis signal region. In this document, we describe how the map was used to recover the limits presented in the ATLAS publication using events generated with `MadGraph5_aMC@NLO` and hadronized using `Pythia8`, and identify for which lifetime ranges this recasting procedure is reliable for limit extraction.**

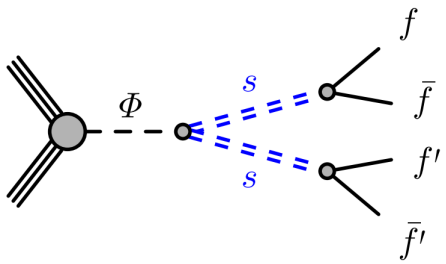


Fig. 1. Schematic diagram of the process $\Phi \to ss \to f\bar{f}f'\bar{f}'$ decay used as a benchmark model. The LLPs couple to SM fermions in a Yukawa-like manner, via their mixing with Φ, and therefore decay primarily to heavy quarks [1].

I. Introduction

The decay of a long-lived particle (LLP) produced in proton–proton collisions in the ATLAS detector could give rise to one of a variety of highly unconventional signatures, depending on the properties of the LLP and the region of the detector where the decay took place. Searches for promptly-decaying particles often have very low sensitivity to such signatures and, hence, the study of LLPs requires dedicated analyses.

In particular, Ref. [1] presents a search sensitive to neutral LLPs decaying in the calorimeters of the ATLAS detector. The analysis is capable of probing lifetime values ranging between a few centimeters up to a few tens of meters. This search set limits on a hidden-sector benchmark model in which a pair of long-lived scalars s with masses between 5 GeV and 475 GeV is produced through a scalar mediator Φ with a mass ranging from 60 GeV to 1 TeV, as represented in Fig. 1.

Depending on their nature, the fermions that come from the LLP decay may result in jets that can be far from the interaction point. Thus, they will lead to displaced jets and vertices. Ref. [1] focuses on decays taking place in the calorimeters: either towards the edge of the electromagnetic calorimeter (ECal) or in the hadronic calorimeter (HCal). In such cases, the LLP decay products can be reconstructed as a single jet which is typically trackless, narrower and with a higher proportion of its energy deposited in the HCal than in the ECal compared with their SM counterparts. Pair-produced LLPs are considered, thus this analysis requires two such non-standard jets, with two selections targeting different LLP kinematic regimes. One is optimized for models with $m_\phi > 200$ GeV (High-E_T) and one for models with $m_\phi < 200$ GeV (Low-E_T), where E_T stands for the transverse energy. Dedicated techniques, heavily reliant on machine learning and the detailed response of the detector, were developed for the reconstruction of displaced jets produced by LLPs that decay hadronically in the ATLAS HCal. The resulting search presented selection efficiencies for various parameter choices in this benchmark model.

Two additional important aspects should be kept in mind about the analysis when considering reinterpretation: first, there is an explicit requirement that the sum of missing hadronic energy in the event should be less than 0.6 times the total hadronic energy. Secondly, there is a further requirement that the jets in the event must be "trackless" (in practice this means that they should not have tracks within a cone of radius 0.2).

The background was evaluated using a data-driven ABCD method. In a plane of two uncorrelated variables,

four regions A, B, C, D are defined. The signal is chiefly found in region A while the background contribution in that region should be obtained from a simple relation ($A = B \times C/D$) of the yields in the other regions. Since no excess is seen, a simultaneous signal-plus-background fit is used to produce upper limits on the cross-section times branching fraction for these models.

But what if one wanted to evaluate the impact of this analysis on a model with slightly different kinematics? Re-running the analysis or manually reproducing the cutflow is impracticable for an external user, given the highly detector-dependent selection and heavy use of machine-learning. However, the HEPData analysis record [2] did provide a new form of efficiency map linking the detector response and selection probability to truth-level information about the LLPs, folding in detector effects. In this note, we seek to validate this efficiency map. The note is accompanied by the code we used to generate the benchmark events, read the map, evaluate the efficiency as a function of lifetime, and set limits.

II. RE-INTERPRETATION MATERIAL

The efficiency map is provided in `.yaml` format in the HEPData record for the search [2]. This record also provides some template code (under `Resources > Python file`) to read in the map and return per-event selection probabilities. The map takes as input truth-level kinematic information about the LLPs and returns a probability that the event would be selected in region A for the analysis, at the detector level. It can in some respects be thought of as a "folding matrix" and relies on the fact that the detector should not care what the details of the model were, but should have the same probability of selecting a neutral LLP decay in any given section of the detector, for given transverse momentum p_T and PDGID decay product particles. The map exists in two versions, one for the High-E_T selection and one for the Low-E_T selection. They are shown in Fig. 2. [The maps assume that a new model being evaluated would pass the two other important selections: the tracklessness condition and the limit on missing hadronic energy.]

A. *Structure*

Each bin of the map represents the probability of an event to be selected in region A for a pair of LLPs given their decay positions (in the transverse direction L_{xy} for the barrel or the longitudinal direction L_z for the endcap, depending on the pseudorapidity η). The map is also binned in the LLPs' transverse momenta and decay products' PDGIDs. Each bin outputs a number between 0 and 1, which is the probability that an event in that bin would be selected in Region A of the search. The sum of the output across a sample is an approximation of the total number of events passing the selection. To obtain the efficiency of a given sample, one should sum the output values for each event as extracted from the map and divide by the total number of events in the sample (adequately weighted if appropriate). As we can see in Fig. 2, the map is symmetric between the two LLPs, The choice of LLP 1 and 2 is arbitrary. This map folds in all detector and analysis effects, including triggering and machine-learning discriminants.

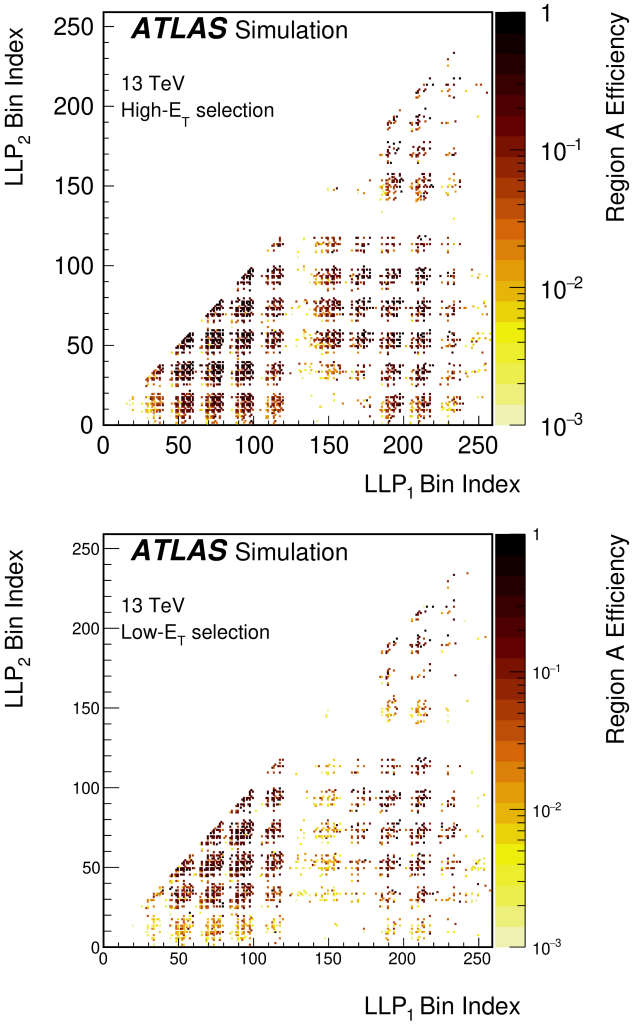


Fig. 2. The two re-interpretation maps provided by the ATLAS analysis [1]. The High-E_T map is shown on top, the Low-E_T map is shown at the bottom. The definition of the "Bin Index" is given in the main body of the text.

The LLP transverse momentum (p_T) is binned in the range of $[0, 50, 100, 200, 400, 1600]$ GeV (5 bins), the LLP transverse decay position L_{xy} is binned in $[0, 1.5, 2, 2.5, 3, 3.5, 3.9, \infty]$m (7 bins), and the longitudinal decay position L_z in $[0, 3.6, 4.2, 4.8, 5.5, 6, \infty]$m (13 bins). These decay position bins are concatenated, with the L_{xy} bins used for decays in the barrel ($|\eta| < 1.5$), and L_z bins used for decays in the endcap ($|\eta| > 1.5$). **In the forward region, the map does not assign a probability of zero to events which have one or more LLPs outside of acceptance ($|\eta| > 2.5$): the proportion of outside-of-acceptance events is in a sense folded into the proba-**

bilities returned by the map. But if the η-distribution of LLPs in any new model varies significantly from those of the HS, this could have an effect. This could be corrected post-hoc however in any re-interpretation study. Four decay modes are considered by the map, namely pairs of c, b, t, τ in bins [0,1,2,3] respectively. These were, indeed, the principle decay products which appeared in the benchmark models of the original analysis, but it seems reasonable that decays into u,d,s quarks would have a similar efficiency as decays into c-quarks. Other decay channels (for example into electrons and muons) should be assigned a probability of 0. The final efficiency is presented as a function named "Bin Index" :

Bin Index = decay position bin index × (number of pT bins
× number of decay type bins)
+ pT bin index × (number of decay type bin)
+ decay type bin index

The bin index does not need to be calculated manually: the helper code includes a function to do so. From the four regions previously introduced (A,B,C,D), just one is approximated by the map, namely region A. **This leads to another assumption in to bear in mind in future re-interpretation works: one would need to assume that most signal events fall in Region A rather than B.**

B. *Map uncertainties*

The description of the maps provided in [1] includes an estimate of their associated uncertainty, which was derived from internal closure tests on the benchmark samples (without lifetime extrapolation). The following estimates are quoted:

- For the High-E_T map, for overall efficiencies above 0.5%, the results are typically accurate at the 25% level, but below this the efficiency can be overestimated and therefore these values should not be used for re-interpretation.
- For the Low-E_T map, for efficiencies above 0.15%, results are typically accurate at a level of 33%. Below this the efficiency is typically accurate up to factor of 3.
- The map is less accurate for the Low-E_T samples because the values of efficiency tend toward zero.

III. COMPUTATIONAL FRAMEWORK

The code which we have used to validate the map (generating events, drawing random lifetimes, evaluating the efficiency and calcualting new limits) is available at [3]. The instruction to install the dependencies for this code are found in: `CalRatioDisplacedJet/READ_ME`

A. *Benchmark model*

The original ATLAS analysis presented in [1] was performed using the so-called "Hidden Abelian Higgs Model" (HAHM) [4, 5] as a benchmark model. In this scenario, the role of the mediator Φ (*cf* Fig. 1) is played by the SM Higgs boson, whereas that of the LLP by the "dark Higgs" that is responsible for the breaking of a hidden $U(1)_X$ gauge symmetry. The masses of both particles are taken to vary as described in Section I. Since here we are mainly interested in recovering the ATLAS results, we will be using the same model for event generation. A `FeynRules` [6] implementation of the HAHM was presented in [7]. Note that in this implementation the mediator Φ is produced primarily through gluon fusion, which is described by an effective ggh vertex.

B. *Event generation*

In order to generate our event samples we used `MadGraph5_aMC@NLO` (MG) [8] employing the *NNPDF2.3Lo* set of Parton Distribution Functions. Showering and hadronization were simulated with `PYTHIA 8` [9] with all settings chosen at their default values.

In Fig. 3 we show, for comparison, the jet p_T distributions without (top panel) and including (bottom panel) hadronization effects for an example set of mediator and LLP masses. It is clear that running `PYTHIA 8` is important when evaluating the efficiency using the map, since the kinematic spectra which are predicted (and are inputs to the map) can differ significantly after hadronization.

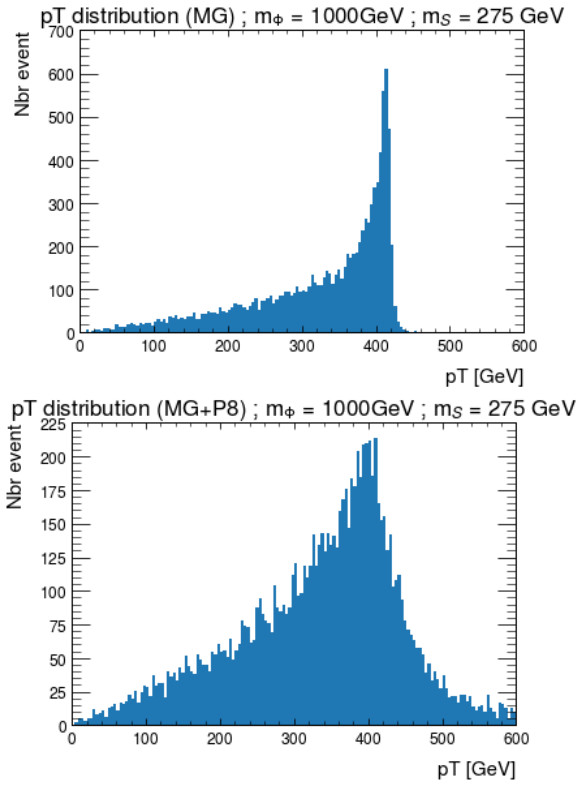


Fig. 3. Comparison of distribution for the transverse momenta without and with the hadronization.

All in all, we generated samples for six different benchmark points, which are presented in Table I. Note that in

principle, the LLP lifetime is a predicted quantity in the HAHM model, once all couplings are fixed. In practice, in the analysis performed in [1] the mean proper lifetime of the LLP scalars is treated as a free parameter and the event-per-event LLP rest frame lifetime is drawn randomly from an exponentially decaying distribution.

All the parameters and run cards that were used in our analysis can be found in [3] in the function: `CalRatioDisplacedJet/Writing_Scripts_MG+P8.py`.

C. *Instructions for generating events and evaluating efficiencies*

After installing the code dependencies, including the event generators, as in `CalRatioDisplacedJet/READ_ME`, one can generate datacards and scripts to run the event generation as follows.

```
python Writing_Scripts_MG+P8.py        (1)
```

This function will create six scripts named `Script_mHXX_mSXX`, then, launch them in MG and add directly the hadronization. MG will output a LHE (Les Houches Events) file that Pythia8 Will hadronize and output a HEPMC file.

Next, the code uses the map to recast the obtained data. To recover the information and parse the LHE/HEPMC files, one can use the script:

```
python Computation_Map.py        (2)
```

This script makes use of additional `python` packages which contain helper functions and parsers:

- `lhe_parser.py` and `hepmc_parser.py`: to parse the LHE and HEPMC files and extract the kinematics needed for decay length and efficiency calculations;
- `Computation_function.py`: this has various utilities to, for example, sample the decay length for a given proper lifetime for each event, and evaluate the map for a range of lifetime values;
- `ReadMapNew.py`: a version of the map=reading code provided on HEPData which has been optimised to evaluate efficiencies more quickly.

Efficiency and limit plots will be returned, named `Efficiency_comparison_mHXX_mSXX.png` and `Cross_section_mHXX_mSXX.png`. Both will be saved in the current directory. Results that we obtained with this method are shown in Figs. 4, 5 and 6.

The code also produces limits: for this, a single-bin fit was performed for the observed and expected number of events in Region A from the original paper, with a dummy signal strength for an arbitrary number of predicted signal events The limit on the signal strength is then rescaled by the number of expected signal events from the map, at each lifetime value, to produce limits on the model cross-section times branching fraction as a function of $c\tau$.

Mass of ϕ in GeV	Mass of the LLP in GeV	Nbr events
1000	275	10000
600	150	10000
400	100	10000
200	50	50000
125	55	50000
60	5	50000

TABLE I

MASS OF THE DARK HIGGS AND OF THE LLP GENERATED WITH MG AND THE NUMBER OF EVENTS FOR EACH SAMPLE.

IV. VALIDATION RESULTS

A. *Efficiency results*

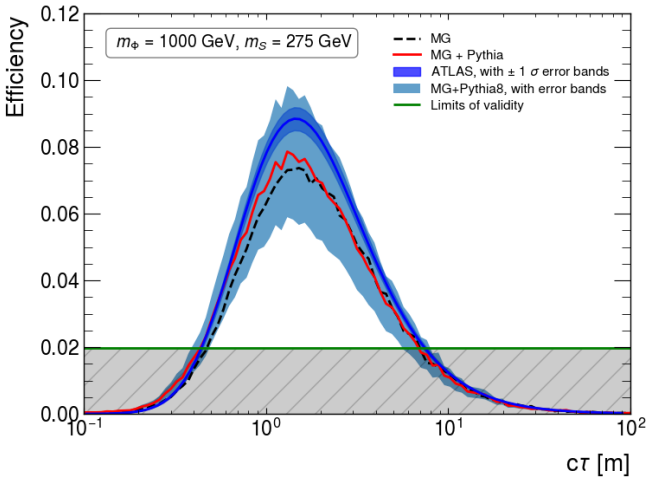
In order to validate the procedure, we compare the efficiency values obtained by passing our generated event samples through the map with the results presented in [1] by the ATLAS collaboration. All the results shown in Fig. 4 have been obtained by generating 10 000 events, whereas for the ones shown in Fig. 5 we generated 50 000 events. The results from ATLAS are shown in dark blue with the uncertainty bands from the search. The results using the efficiency maps are shown in dashed black form MG only and solid reg for MG with Pythia for hadronization. This last line also has the map uncertainties and validity limits described in Sec. II-B applies, shown by the light blue shading and green line with hatched region. Our results are shown in Figs. 4 and 5

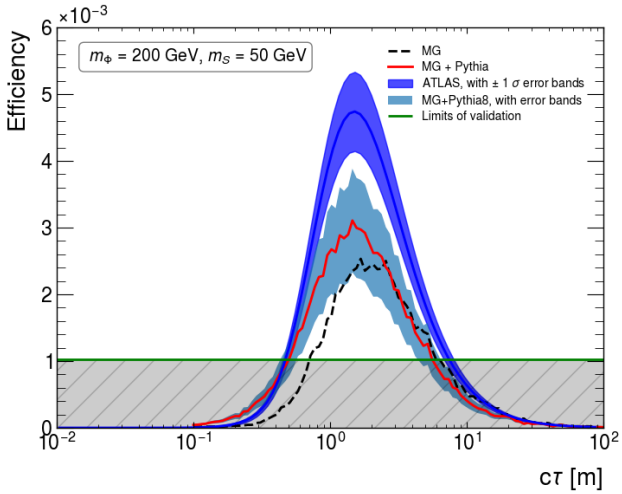
We can see that for our High-E_T benchmarks (Fig. 4), the results obtained through the map reproduce fairly well the ones obtained through a full-blown analysis, once all uncertainties are properly taken into account. In this case, the regults from MG-only and MG+Pythia agree fairly well: the differences in kinematics due to hadornization only have a small effect on the final outcome.

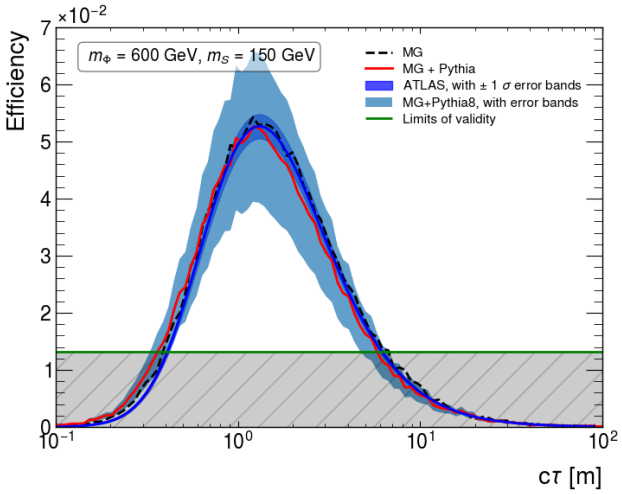
On the other hand, the procedure becomes less reliable once we switch to the low-E_T samples (Fig. 5). As a side-comment, let us also point out that according to our findings, even though the showering/hadronization effects have little to no impact on the efficiency values obtained for the high-mass benchmarks, they become crucial for the low-mass ones. This can be understood physically because without the recoils introduced by hadronization, the p_T thresholds for the trigger and analysis are impossible to achieve for the LLPs given the mass of the parent mediator. Note that the values obtained for the sample with m_ϕ = 60 GeV and m_s = 5 GeV are too low to obtain usable results. Nevertheless, within the region of validity of the maps, the results do agree within order of magnitude.

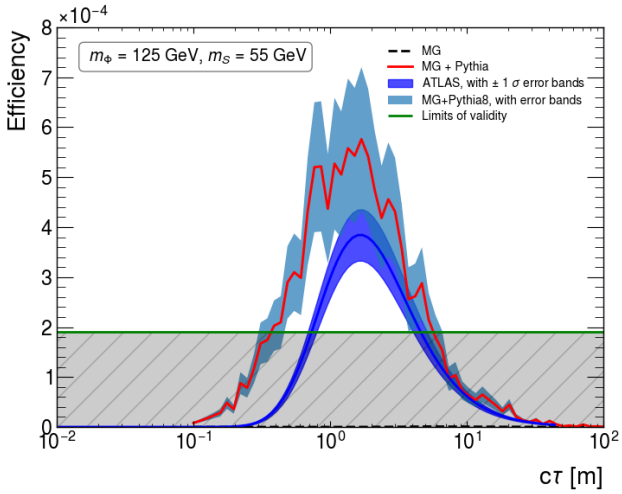
B. *Cross section limits results*

As for the efficiencies, in order to validate the procedure we have plotted the limits that we can set on the cross section by passing our generated event samples through the map and by comparing the results to the one presented in [1] by the ATLAS collaboration. Our results are shown in Figs. 6 and 7









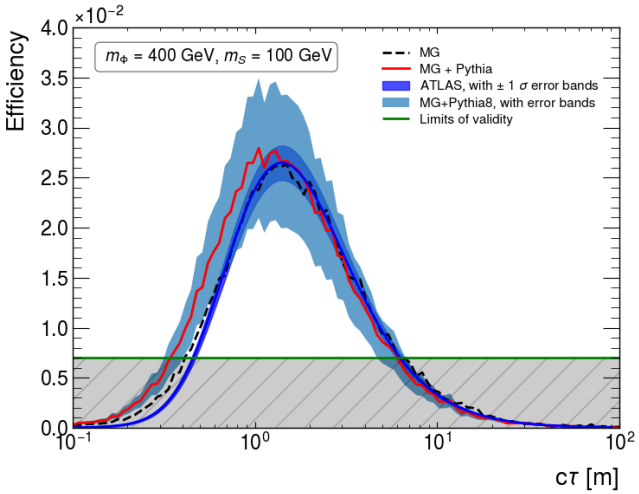


Fig. 4. Samples with m_ϕ = 1000,600,400 GeV; m_s = 275,150,100 GeV; High-E_T.

Fig. 5. sample m_ϕ = 200,125 GeV; m_s = 50,55 GeV; Low-E_T.

As before, we can see that for the limits sets on High-E_T benchmarks (Fig. 6) fir pretty well to the ones obtained through a full-blown analysis, apart from the low lifetimes values.

V. Lesson from validation

All in all, our validation procedure allows us to draw several useful conclusions concerning on one hand the map itself, as well as the process of utilizing it as an external user. In this Section we swiftly present some comments which we find to be the most relevant.

A. *Lesson on the map*

Let us first comment on a few issues concerning the re-interpretation map, the validity of some of the approximations it relies on and the results it allows a user to obtain.

- The full ATLAS search used a simultaneous signal-plus-background fit over all ABCD regions, whereas the present treatment only relied on region A results : according to our findings, this is a valid approximation. But for a new model, this may possibly break down if a large fraction of the signal ended up in

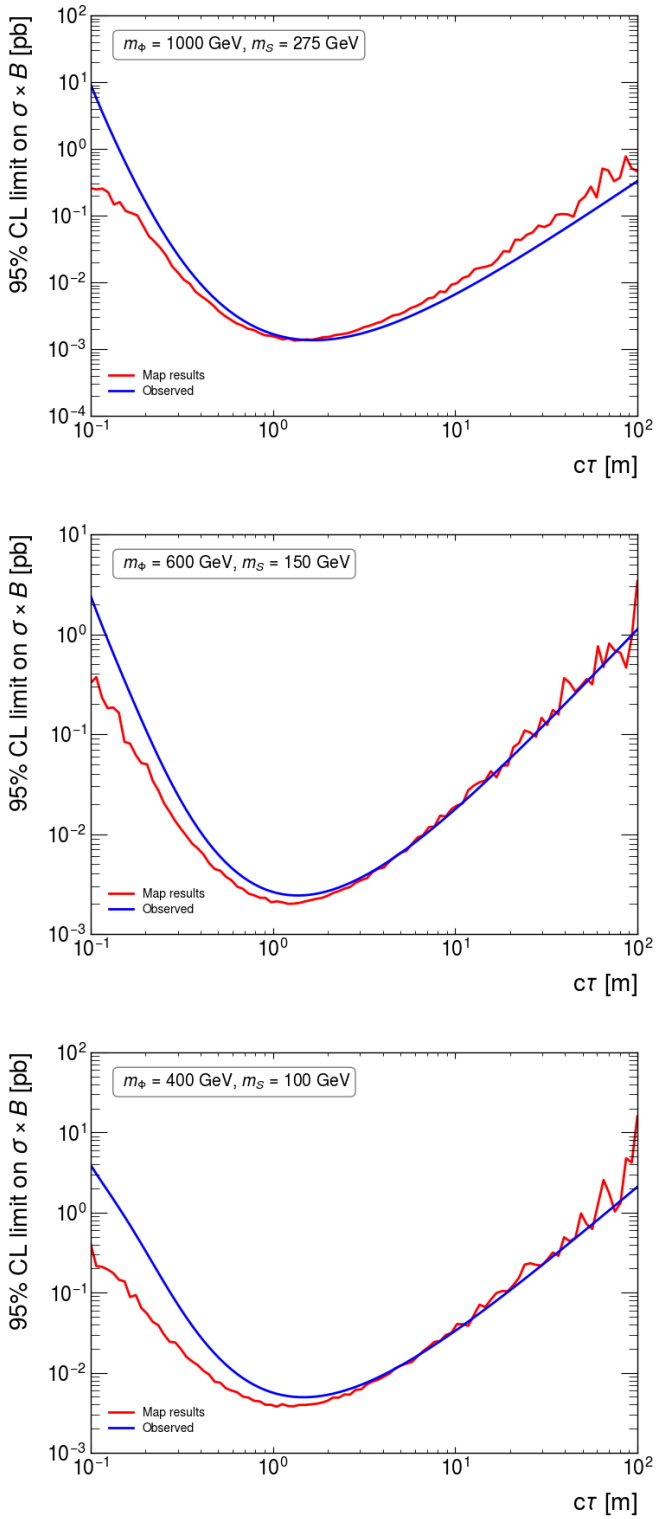


Fig. 6. Limits for the samples m_ϕ = 1000,600,400 GeV; m_s = 275,150,100 GeV; High-E_T.

another region. This seems likely to be a small effect, but it's not immediately clear how this can be tested.

- The limits set by relying on the map are fairly realistic for large enough lifetime values but become too aggressive at lower values. One could argue that the map is only safe to use for $c\tau$ values above about 5cm.
- The map does not account for the fact that some models may have different η distributions leading to more particles lost in the out-of-acceptance region beyond the tracker acceptance. The map implicitly assumes that the η distribution of the LLPs would resemble those from a HS model. This could potentially lead to an over-estimate of the efficiency, so care should be taken. An ad-hoc fix could be to compare the fraction of LLPs in the high-η region compared to a HS model, and derive a scaling factor from that.
- The map similarly incorporates selections to do with tracklessness and the fraction of missing hadronic energy int he event. Implicitly it assumes a new model would pass these selections, so again care should be taken to check this for any new model being re-interpreted.
- Hadronization effects are essential during event gen-

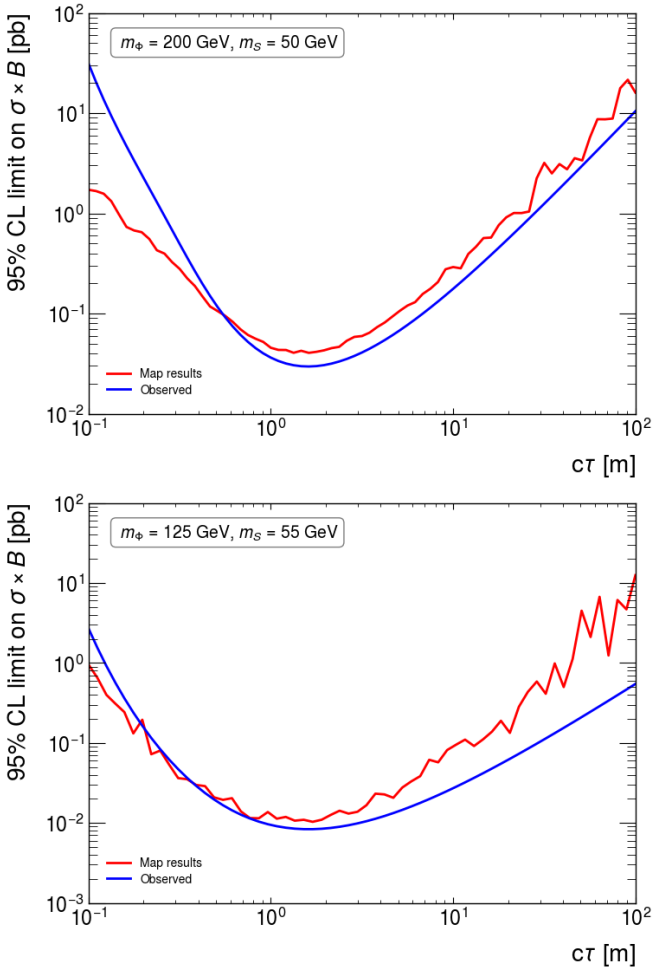


Fig. 7. Limits for the samples m_ϕ = 200,125 GeV; m_s = 100,55 GeV; Low-E_T.

eration, especially when the entire mass spectrum is shifted towards lower values (in Fig, 8 we can see that if hadronization is ignored, the transverse momentum distribution collapses close to $m_s/2$ ($\approx$ 27 GeV), which is close to the experimental limit for a jet to pass selection cuts).
- The obtained results become less precise as the mass of the particles involved decreases. Moreover, in order to avoid wildly oscillating results, a larger number of events needs to be generated.

B. *Event generation are hard for an external user :*

We below point out a few issues encountered during the validation which hindered a comparison of the map results with ATLAS results:

- The benchmark model employed in the ATLAS analysis is not explained adequately and essentially no links to the necessary infrastructure (`FeynRules` implementation etc) are provided.
- Some key issues concerning the usage of the model are not stated explicitly. In particular, as we already alluded to, even though the LLP mean proper lifetime is, in principle, a quantity which can be *predicted* in the HAHM model once all couplings and masses are fixed, in the actual analysis it was treated as a free parameter.
- Lastly, several other internal parameters of the run are changed by hand and not documented (masses, width calculation, cuts).

It would be useful to have an official recipe from ATLAS to generate the benchmark models, because here, as an external user it was very hard.

VI. Conclusion

After briefly describing the materials provided for ATLAS-EXOT-2019-23, which contained a powerful tool for reinterpretation: a map that only needs a few truth-level parameters in input and only one region of interest to work. It is an object in six dimensions where each bin outputs a probability to select an event in their analysis. To check it, we have generated our samples with MadGraph and added Pythia and we have then compared the result with the ATLAS public data. To finish, we have made a summary of the difficulties encountered during the study for a user that aims to re-interpret data and gave comments on the re-casting.

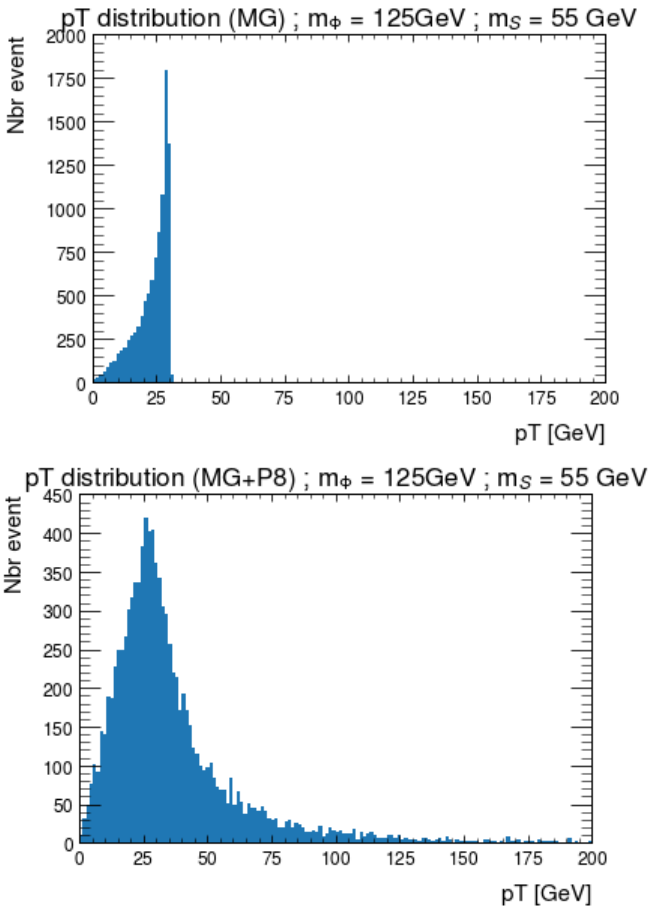


Fig. 8. Distribution comparison without and with Pythia for the Low-E_T sample of m_ϕ = 125GeV; m_s = 55 GeV.

Appendix

References

[1] Georges Aad et al. "Search for neutral long-lived particles in pp collisions at $\sqrt{s}$ = 13 TeV that decay into displaced hadronic jets in the ATLAS calorimeter". In: *JHEP* 06 (2022), p. 005. DOI: 10.1007/JHEP06(2022)005. arXiv: 2203.01009 [hep-ex].

[2] *ATLAS-EXOT-2019-23 HEPdata record, howpublished = https://www.hepdata.net/record/ins2043503, note = Accessed: 11 Jul 2023.*

[3] *ATLAS-EXOT-2019-23 Recasting Code, howpublished = https://github.com/ThomasChehab/recastingCodes/tree/master/CalRatioDisplacedJet, note = Accessed: 27 Jul 2023.*

[4] Shrihari Gopalakrishna, Sunghoon Jung, and James D. Wells. "Higgs boson decays to four fermions through an abelian hidden sector". In: *Phys. Rev. D* 78 (2008), p. 055002. DOI: 10.1103/PhysRevD.78.055002. arXiv: 0801.3456 [hep-ph].

[5] James D. Wells. "How to Find a Hidden World at the Large Hadron Collider". In: (Mar. 2008). Ed. by Gordon Kane and Aaron Pierce, pp. 283–298. arXiv: 0803.1243 [hep-ph].

[6] Adam Alloul et al. "FeynRules 2.0 - A complete toolbox for tree-level phenomenology". In: *Comput. Phys. Commun.* 185 (2014), pp. 2250–2300. DOI: 10.1016/j.cpc.2014.04.012. arXiv: 1310.1921 [hep-ph].

[7] David Curtin et al. "Illuminating Dark Photons with High-Energy Colliders". In: *JHEP* 02 (2015), p. 157. DOI: 10.1007/JHEP02(2015)157. arXiv: 1412.0018 [hep-ph].

[8] J. Alwall et al. “The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations”. In: *JHEP* 07 (2014), p. 079. DOI: 10.1007/JHEP07(2014)079. arXiv: 1405.0301 [hep-ph].

[9] Christian Bierlich et al. “A comprehensive guide to the physics and usage of PYTHIA 8.3”. In: (Mar. 2022). DOI: 10.21468/SciPostPhysCodeb.8. arXiv: 2203.11601 [hep-ph].