

An update on the LHC monojet excess

**Pouya Asadi,^a Matthew R. Buckley,^a Anthony DiFranzo,^a Angelo Monteux^{a,b}
and David Shih^a**

^a*NHETC, Department of Physics and Astronomy, Rutgers, The State University of New Jersey,
136 Frelinghuysen Rd, Piscataway, NJ 08854 U.S.A.*

^b*Department of Physics and Astronomy, University of California, Irvine,
4129 Frederick Reines Hall, Irvine, CA 92697-4575, U.S.A.*

E-mail: pa304@physics.rutgers.edu, mbuckley@physics.rutgers.edu,
anthony.difranzo@rutgers.edu, amonteux@physics.rutgers.edu,
shih@physics.rutgers.edu

ABSTRACT: In previous work, we identified an anomalous number of events in the LHC jets+MET searches characterized by low jet multiplicity and low-to-moderate transverse energy variables. Here, we update this analysis with results from a new ATLAS search in the monojet channel which also shows a consistent excess. As before, we find that this “monojet excess” is well-described by the resonant production of a heavy colored state decaying to a quark and a massive invisible particle. In the combined ATLAS and CMS data, we now find a local (global) preference of 3.3σ (2.5σ) for the new physics model over the Standard Model-only hypothesis. As the signal regions containing the excess are systematics-limited, we consider additional cuts to enhance the signal-to-background ratio. We show that binning finer in H_T and requiring the jets to be more central can increase S/B by a factor of ~ 1.5 .

KEYWORDS: Phenomenological Models

ARXIV EPRINT: [1712.04939](https://arxiv.org/abs/1712.04939)

As the LHC reaches a phase of stable running, it is important to re-examine our search strategies for new physics. Without large increases in energy or luminosity, it becomes less and less likely that new physics will suddenly appear with large statistical significance in a low-background channel. Instead, we expect new physics at the LHC to appear only gradually, starting with small deviations from the Standard Model predictions. As the searches for new physics at the LHC grow in sophistication and complexity (especially on the CMS side), it can become increasingly difficult to separate out statistically-meaningful deviations from random noise. This is exacerbated by the increasing reliance on “simplified models” to interpret the data. While simplified models are well-suited for limit-setting, they are too few in number (and of too limited variety) to populate more than a small subset of the hundreds of signal regions across all of the LHC searches, so that relying exclusively on simplified models to characterize the data can greatly bias the search for new physics.

In a previous work [1], we developed a “rectangular aggregation” technique which attempted to overcome these biases by combining signal regions in a more model-independent way. This was based on the simple observation that any signal can populate multiple neighboring bins, and therefore aggregating signal regions within larger kinematic ranges can extract information about underlying excesses without making assumptions about a specific signal model. As a proof of principle, we applied our aggregation technique to the CMS jets+ \cancel{E}_T searches [2] and [3] (hereafter referred to as CMS033 and CMS036, respectively). While originally motivated by supersymmetry, these searches are broadly sensitive to new physics, owing to the fact that they each consist of hundreds of exclusive signal regions, defined by number of jets, number of b -tagged jets, and transverse energy variables such as H_T , missing transverse momentum \cancel{E}_T , and/or M_{T2} .

Through our method of rectangular aggregations, we identified a number of interesting $\sim 3\sigma$ excesses within these searches. The most interesting one was consistent between both searches. We dubbed this the “monojet excess” because it is characterized by low jet multiplicity, no b -jets, and low \cancel{E}_T and H_T . We found that the anomaly’s kinematic distributions could not be well-fit by supersymmetry-like pair production of colored particles, or in simplified models for dark matter pair production [1]. Instead, a good fit was obtained using a colored scalar ϕ , resonantly produced through couplings to quarks, and decaying to an invisible massive Dirac fermion ψ and a Standard Model quark (the “mono- ϕ ” model), see figure 1.

To avoid decays of the ψ back to visible states, its Dirac partner ψ' can be coupled to invisible states N and \tilde{N} . The interaction Lagrangian for the minimal model is [1]

$$-\mathcal{L} \supseteq g\phi^*q_i^c\psi + \lambda\phi q_i^c q_j^c + m_\psi\psi\psi' + m_\phi^2|\phi|^2 + g'\psi'N\tilde{N} + \text{h.c.} \quad (1)$$

Here, the q_i are the right-handed quarks. The scalar ϕ is a color-triplet, and its charge can be $+\frac{2}{3}$ or $-\frac{1}{3}$. For a given ϕ mass, the resonance production cross section is set by λ , while the branching ratios of ϕ to $q\psi$ versus qq are set by both λ and g . The ϕ resembles a squark in R -parity violating supersymmetry, though in order to avoid baryon-number-violating decays the ψ cannot be identified with a Majorana neutralino [4].

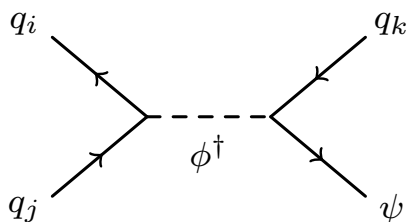


Figure 1. The “mono- ϕ ” simplified model that fits well the monojet excess in the CMS and ATLAS searches.

We also found further hints for the same anomaly in the ATLAS 2-6 jets+ \cancel{E}_T search [5] (ATLAS022) which, owing to high m_{eff} and \cancel{E}_T thresholds, was not as sensitive to this signal as it could have been. The anomaly is in some tension with null results from the CMS dark matter+jets exotica search [6] (CMS048), but a production cross section on the order of 0.5 pb can evade the 95%CL limits from that search, while still maintaining a local 3σ preference for signal over background in CMS036 and ATLAS022.¹

This letter serves as an update to the original analysis, containing three new points concerning the monojet excess:

1. We include the newly released ATLAS monojet search [7] (ATLAS060). This is a search for a dark matter mediator in events with missing energy and at least one high- p_T jet ($p_T^{j_1} > 250$ GeV), with ten exclusive bins in the \cancel{E}_T variable, starting at $\cancel{E}_T = 250$ GeV. This search has a better sensitivity to the mono- ϕ model than ATLAS022, where the most sensitive signal region had much higher thresholds, requiring two jets and $m_{\text{eff}} \equiv \cancel{E}_T + \sum_j p_T^j > 1200$ GeV. In the ATLAS060 data, we find a 2–2.5 σ preference for this model in the same region of parameter space preferred by the CMS searches. The previous ATLAS022 analysis had only a 1–1.5 σ preference.
2. We perform a joint statistical analysis of the CMS and ATLAS data, including the look-elsewhere-effect. Combining the ATLAS060 and CMS036 searches, the local significance for this model reaches 3.5 σ in a region of parameter space in the $m_\phi = 1200$ –1800 GeV range and mass splitting $m_\phi - m_\psi = 300$ –400 GeV, which is lowered to 3.3 σ when requiring the signal cross section to be allowed at 95%CL by CMS048. Using 10,000 pseudoexperiments, we estimate that this corresponds to a 2.5 σ global significance.
3. We suggest additional cuts to enhance the experimental sensitivity to this signal. As we will describe, the experimental errors for the signal regions containing the observed excess are systematics-dominated. Therefore, additional data may not appreciably increase the overall significance of the anomaly, even if it is due to new physics. With the production mode in figure 1, the signal H_T distribution is peaked at the mass

¹As the data sets between the searches from a single collaboration are overlapping, we cannot statistically combine multiple CMS or ATLAS analyses, and must confine ourselves to a single CMS and a single ATLAS result.

| Search | N_j | N_b | H_T | M_{T2}, \cancel{E}_T | Obs. | Bg. (pre-fit) | Bg. (post-fit) | Best-fit Signal |
|----------|----------|-------|------------|------------------------|--------|-------------------|-------------------|-----------------|
| CMS036 | 1–3 | 0 | 250–450 | 200–300 | 145144 | 137256 ± 8159 | 140391 ± 1524 | 4753 |
| CMS033 | 2–4 | 0 | 300–500 | 300–500 | 58138 | 54550 ± 2246 | 55976 ± 780 | 2162 |
| ATLAS060 | ≥ 1 | -- | ≥ 250 | 350–700 | 74686 | 72645 ± 1140 | | 2041 |

Table 1. Regions of kinematic space containing the excess, and relative observed and background event counts, as well as number of signal events after the fit. For the background, we quote both the pre- and post-fit values for the expected counts and the relative errors.

difference between ϕ and ψ , while the background is smoothly falling. In addition, the signal jet tends to be more centrally produced than the background. Therefore, we find the most effective way to increase sensitivity is to define finer H_T bins and require a tighter cut on the leading jet pseudo-rapidity, in particular $|\eta| \lesssim 0.5$. We find that S/B can be increased by a factor of ~ 1.5 compared to the current CMS036 analysis, to an overall level of $S/B \sim 8\%$.

In table 1, we show the range of kinematic parameters within the various ATLAS and CMS searches containing the anomalous events which we have identified as the monojet excess. For the CMS033 and CMS036 searches, the anomaly is spread out over a number of signal regions. The uncertainties on the expected number of events in these signal regions are highly correlated, and we make use of simplified covariance matrices provided by CMS [3] (no correlations were provided for ATLAS060). In table 1, we report both “pre-fit” and “post-fit” background predictions. (We are following standard CMS terminology, see e.g. [8]) The pre-fit backgrounds are the simple aggregation of the background counts and the sum of covariance matrix for the bins populated by signal (as detailed in [1]). The post-fit values refer to a combined fit of all the signal regions in the presence of a new-physics signal which only populates a specific subset of all the bins. Due to the high degree of correlation between bins populated by the signal and those bins where no signal events fall, post-fit errors are reduced: effectively, the bins not populated by signal act as additional control regions and lower the uncertainty in the bins of interest.² In the signal regions of interests, with pre-fit errors of order 4–5%, this procedure results in post-fit uncertainties at the 1% level. As background correlations were not released with the ATLAS060 search, there is no difference between pre-fit and post-fit for that search.

The signal regions of table 1 identify the “core” of the identified excess, and are independent of any particular new physics model. However, a full fit — including all signal regions of each search — requires both a model and a recasting of the experimental search sensitivity for that model. Scanning over the (m_ϕ, m_ψ) mass plane, we generated mock-LHC data for the mono- ϕ model using MADGRAPH5 [9], PYTHIA8 [10] for showering and hadronization, and a tuned implementation of DELPHES3 [11] for detector simulation. Events were generated without jet matching, though comparison with matched samples demonstrated that the effect was minimal. Full details of our recasting procedure and cross-checks can be found in [1]. For each ATLAS or CMS search [2, 3, 5, 6], we calcu-

²We thank Claudio Campagnari for emphasizing this point in our procedure.

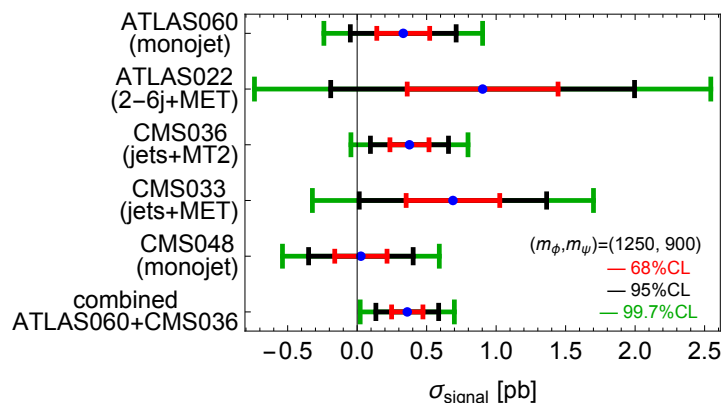


Figure 2. Values of the signal cross sections favored at 1,2 and 3 σ by each individual search considered, and by the combination of ATLAS060 and CMS036.

late the statistical preference for the signal+background hypothesis over background-only using the profile likelihood method [12, 13], treating the cross section times branching ratio at each mass point as a free parameter in the fit. The results are indicated in figure 2, where we show the best-fit confidence intervals for $\sigma \times \text{BR}$ of a reference mass point $(m_\phi, m_\psi) = (1250, 900)$ GeV, for each of the ATLAS and CMS searches of interest. As can be seen, the anomaly seen in ATLAS060 is broadly consistent with that previously identified in the CMS033, CMS036, and ATLAS022 data, and at higher significance than the previous ATLAS search. While the CMS monojets search CMS048 did not see any evidence for new physics, its confidence intervals are entirely consistent with the size of the excess seen by the other searches.

Although we cannot combine all of these searches to produce an overall best fit cross section, we can pick one from CMS and one from ATLAS for a joint fit. Choosing CMS036 and ATLAS060 as being the two that are most sensitive to our signal, the resulting significance plot is shown in figure 3. To take into account the non-observation of signal from CMS048, we require that the best-fit cross section be less than the 95%CL upper limit from that search.³ Even after this, the combined fit finds a local preference for signal at the 3.3 σ level for $m_\phi \sim 1200\text{--}1800$ GeV and $m_\phi - m_\psi \sim 300\text{--}400$ GeV, with $\sigma \times \text{BR} \sim 0.3$ pb. This represents an increased significance from the 3 σ result reported in [1].

To additionally illustrate the compatibility between the excesses in CMS036 and ATLAS060, in figure 4 we show the residuals (observed minus expected) from ATLAS060 and from the $N_j = 1$ bins of CMS036 as a function of \cancel{E}_T , along with the distribution predicted by the mono- ϕ model for the specific mass point $(m_\phi, m_\psi) = (1250, 900)$ GeV, using the best-fit cross sections. The kinematics are similar to other good-fit mass points with $m_\phi - m_\psi \approx 300$ GeV.

We further update the analysis of [1] to include the global significance of the combined fit to CMS036 and ATLAS060. We generate 10,000 pseudoexperiments of ATLAS060 and CMS036 data, drawing from the background-only distributions (as in the combined

³As discussed in [1], this model also gives a correlated signature in the dijet resonance channel, however the exact signal strength depends on additional couplings not determined by this fit. Here we assume that the couplings are always chosen such that the $\sigma \times \text{BR}$ into dijet resonances is consistent with current bounds.

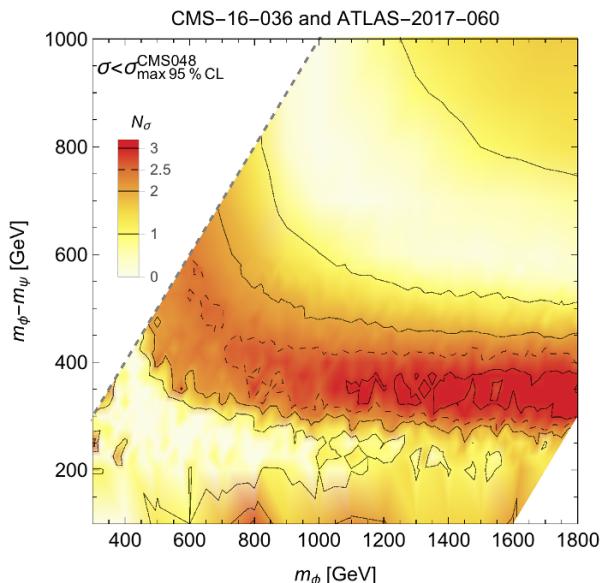


Figure 3. Best-fit significance for the model in the $m_\phi/m_\phi - m_\psi$ mass plane, obtained combining the CMS jets+ M_{T2} search [3] and the ATLAS monojet search [7]. The corresponding best-fit cross section is $\mathcal{O}(0.35 \text{ pb})$ in the region with highest significance.

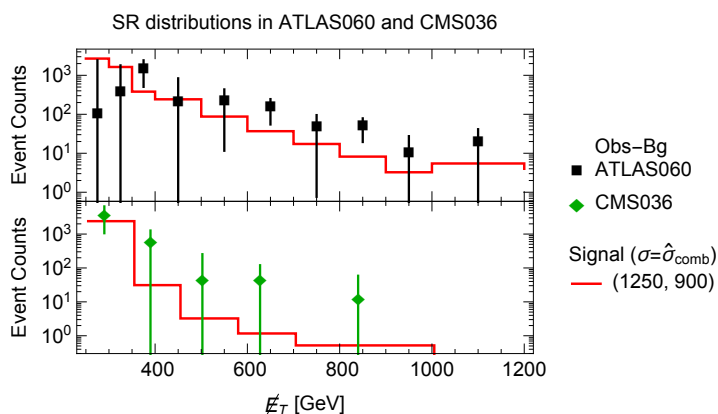


Figure 4. Difference between observed and background counts with relative error bars for ATLAS060 (black) and the CMS036 $N_j = 1$ bins (green), to be compared with the \cancel{E}_T distribution of the signal for $(m_\phi, m_\psi) = (1250, 900)$, respectively in solid and dashed red, given the production cross section set by the joint fit to ATLAS060 and CMS036.

fit, we neglect possible correlations between systematics of ATLAS and CMS). For each pseudoexperiment, we perform a combined fit and count the number of pseudoexperiments for which the significance for the mono- ϕ model is greater than observed in the data (3.3σ), scanning over the mass plane. The fraction of pseudoexperiments where the background mimics the signal at the 3.3σ level or more is 0.0128. We therefore conclude that our 3.3σ local excess in the combined dataset corresponds to a global 2.5σ anomaly in the context of the mono- ϕ model.

A 2.5σ global excess is potentially interesting, but certainly not definitive proof of physics beyond the Standard Model. Since the quoted significance is dominated by systematic errors (see the quoted pre- and post-fit errors in table 1, which are significantly larger than \sqrt{N}), the situation might not necessarily improve with more data. Instead, one must either reduce the systematic errors, or identify further cuts that can enhance the new physics scenario over the Standard Model background. Since the former is something only the experimentalists can do, here we focus on the latter.

For specificity, we consider the search with the greatest signal significance (CMS036), though a similar analysis could be performed with ATLAS060. We simulate the primary backgrounds in the relevant signal regions: $(Z \rightarrow \nu\nu)+\text{jets}$ and $(W \rightarrow \ell\nu)+\text{jets}$, where the lepton is missing. The events are generated in the same MADGRAPH5, PYTHIA8, and DELPHES3 chain used for our signal, matched up to four jets. We normalize each background sample by reweighting them in each exclusive signal region to match the expected pre-fit background rates reported in CMS036.

Since the signal contains one parton from the hard process, we focus on the monojet bins of CMS036, which is defined by the criteria $N_j = 1$, $N_b = 0$, $H_T \geq 250$ GeV. With only one jet in each event, the only kinematic variables to cut on are the jet p_T and pseudo-rapidity. As the signal is produced from the decay of a resonantly produced massive scalar, while the background is produced from t -channel scale-invariant QCD, we expect the signal to be peaked in p_T and more central than the background. This is confirmed by the η and H_T histograms shown in figure 5. There we show the background multiplied by a conservative estimate of the reported pre-fit systematic error ($\sim 5\%$, as can be seen in table 1).

In figure 5, we further show the H_T distributions before (center panel) and after (right panel) a tighter η cut on the jet. It can be seen that by requiring the jet to be more central, the peak of the signal distribution goes up by a factor of ~ 1.5 as compared to all events. Moreover, signal and background peak at different values of H_T , and so this distribution is possibly robust against systematic errors near threshold. Note that this difference can only be seen if the events are binned in sufficiently small ranges of H_T . If, after the inclusion of this additional η cut, the $\sim 5\%$ pre-fit systematic errors can still be reduced to the previously-achieved $\sim 1\%$ level post-fit, then this factor of 1.5 boost in signal over the background could potentially increase the local statistical significance from the single CMS036 search up to $\sim 3\sigma \times 1.5 \sim 4.5\sigma$. Similar improvements in signal sensitivity could presumably be performed with the ATLAS data.

In summary, we have updated the analysis described in [1] with the latest monojet search from ATLAS, which appears to have an excess in the same place and is consistent with the excess found in [1]. As in [1], we find that a resonantly-produced color-triplet scalar decaying to jet plus missing energy fits all the data well. Performing a joint fit, the local significance of the monojet excess grows to 3.3σ local (2.5σ global). Finally, we show that by binning more finely in H_T and putting a simple cut on the centrality of the jet, we can enhance signal over background by a factor of at least ~ 1.5 . This could greatly increase the significance of the anomaly, as well as providing more confidence that it is not due to systematic errors.

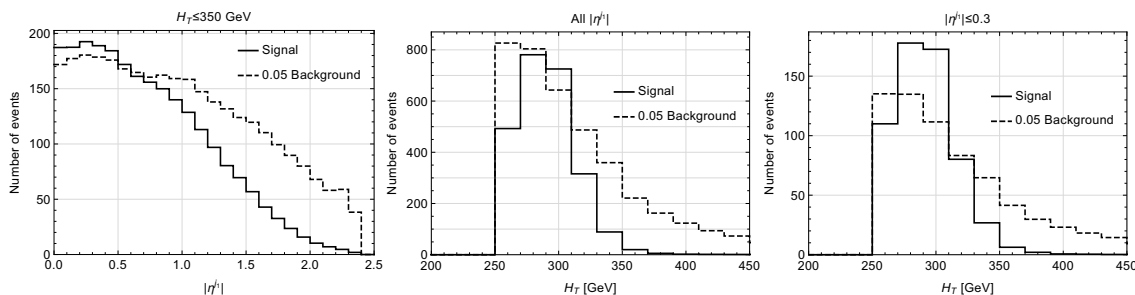


Figure 5. Distributions of signal (solid) and our estimate of the pre-fit systematic error (5% of the background events) in the $N_j = 1$, $N_b = 0$ SRs of CMS036. Left: distribution with respect to $|\eta|$. Center: distribution with respect to H_T without a cut on $|\eta|$. Right: distribution with respect to H_T requiring jet $|\eta| < 0.3$.

Acknowledgments

We thank Claudio Campagnari, Elliot Lipeles and Keisuke Yoshihara for helpful discussions. The work of P.A. and D.S. is supported by DOE grant DE-SC0010008. M.R.B. is supported by DOE grant DE-SC0017811. The work of A.M. was supported in part by NSF Grant No. PHY-1620638 and in part by Simons Investigator Award #376204.

Open Access. This article is distributed under the terms of the Creative Commons Attribution License ([CC-BY 4.0](https://creativecommons.org/licenses/by/4.0/)), which permits any use, distribution and reproduction in any medium, provided the original author(s) and source are credited.

References

- [1] P. Asadi, M.R. Buckley, A. DiFranzo, A. Monteux and D. Shih, *Digging Deeper for New Physics in the LHC Data*, *JHEP* **11** (2017) 194 [[arXiv:1707.05783](https://arxiv.org/abs/1707.05783)] [[INSPIRE](https://inspirehep.net/literature/1707057)].
- [2] CMS collaboration, *Search for supersymmetry in multijet events with missing transverse momentum in proton-proton collisions at 13 TeV*, *Phys. Rev. D* **96** (2017) 032003 [[arXiv:1704.07781](https://arxiv.org/abs/1704.07781)] [[INSPIRE](https://inspirehep.net/literature/1704077)].
- [3] CMS collaboration, *Search for new phenomena with the M_{T2} variable in the all-hadronic final state produced in proton-proton collisions at $\sqrt{s} = 13$ TeV*, *Eur. Phys. J. C* **77** (2017) 710 [[arXiv:1705.04650](https://arxiv.org/abs/1705.04650)] [[INSPIRE](https://inspirehep.net/literature/1705046)].
- [4] R. Barbieri and A. Masiero, *Supersymmetric Models with Low-Energy Baryon Number Violation*, *Nucl. Phys. B* **267** (1986) 679 [[INSPIRE](https://inspirehep.net/literature/1705046)].
- [5] ATLAS collaboration, *Search for squarks and gluinos in final states with jets and missing transverse momentum using 36 fb^{-1} of $\sqrt{s} = 13$ TeV pp collision data with the ATLAS detector*, [arXiv:1712.02332](https://arxiv.org/abs/1712.02332) [[INSPIRE](https://inspirehep.net/literature/1712023)].
- [6] CMS collaboration, *Search for new physics in final states with an energetic jet or a hadronically decaying W or Z boson and transverse momentum imbalance at $\sqrt{s} = 13$ TeV*, [arXiv:1712.02345](https://arxiv.org/abs/1712.02345) [[INSPIRE](https://inspirehep.net/literature/1712023)].

- [7] ATLAS collaboration, *Search for dark matter and other new phenomena in events with an energetic jet and large missing transverse momentum using the ATLAS detector*, *JHEP* **01** (2018) 126 [[arXiv:1711.03301](#)] [[INSPIRE](#)].
- [8] CMS, ATLAS collaborations, *Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s} = 7$ and 8 TeV*, *CMS-PAS-HIG-15-002* (2015).
- [9] 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*, *JHEP* **07** (2014) 079 [[arXiv:1405.0301](#)] [[INSPIRE](#)].
- [10] T. Sjöstrand et al., *An Introduction to PYTHIA 8.2*, *Comput. Phys. Commun.* **191** (2015) 159 [[arXiv:1410.3012](#)] [[INSPIRE](#)].
- [11] DELPHES 3 collaboration, J. de Favereau et al., *DELPHES 3, A modular framework for fast simulation of a generic collider experiment*, *JHEP* **02** (2014) 057 [[arXiv:1307.6346](#)] [[INSPIRE](#)].
- [12] CMS collaboration, *Simplified likelihood for the re-interpretation of public CMS results*, *CMS-NOTE-2017-001* (2017).
- [13] G. Cowan, K. Cranmer, E. Gross and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, *Eur. Phys. J. C* **71** (2011) 1554 [*Erratum ibid.* **C 73** (2013) 2501] [[arXiv:1007.1727](#)] [[INSPIRE](#)].