

Estimation of the chances to find new phenomena at the LHC in a model-agnostic combinatorial analysis

S/V.Chekanov*
(Dated: May 17, 2024)

In this paper, we estimate the number of event topologies that have the potential to be produced in pp collisions at the Large Hadron Collider (LHC) without violating kinematic and other constraints. We use numerical calculations and combinatorics, guided by large-scale Monte Carlo simulations of Standard Model (SM) processes. Then, we set the upper limit on the probability that new physics may escape detection, assuming a model-agnostic approach. The calculated probability is unexpectedly large, and the fact that the LHC has not found new physics until now is not entirely surprising. We argue that the optimal direction for maximizing the chances of finding new physics is to use unsupervised machine learning for anomaly detection or algorithms designed for event classification.

I. INTRODUCTION

A simple answer to the question of why the Large Hadron Collider (LHC) did not discover new physics can be rather short: there are no new TeV-scale phenomena that can be discovered using the LHC data, and perhaps this might be the most obvious outcome of the LHC for many years to come. Another way to answer this question is to argue that not all event classes have been explored so far. Physics beyond the Standard Model (BSM) can still produce unusual events with small production cross sections, but such events have not been found yet. Many make this last argument, but putting this discussion on solid quantitative footing requires removing the constraints imposed on experimental research by model builders. In the past, the LHC had too significant a focus on setting limits on proposed BSM models. Looking at this problem from a much wider angle by adopting a model-agnostic view of BSM searches could reveal many unexpected features of the LHC data.

The goal of this paper is to calculate the number of unique event classes produced at the LHC. We define an exclusive event class (or event topology) as a group of events with exactly the same number of identified particles and reconstructed jets. We use numerical calculations based on combinatorics, guided by large-scale Monte Carlo (MC) simulations of the Standard Model (SM). The latter sets the necessary kinematic constraints and boundary conditions for our calculations. Then we estimate the chances that new phenomena might have escaped detection at the LHC. These calculations are fully model-agnostic since they are not guided by BSM models.

II. NUMBER OF UNIQUE TOPOLOGIES FROM COMBINATORICS

Let us estimate how many unique event topologies are expected to be produced in pp collisions at $\sqrt{s} = 13$ TeV. Assume LHC collisions produce events with light-flavor jets (j), jets associated with b -quarks (b -jets), electrons (e), muons (μ), tau (τ) leptons and photons (γ). In addition, neutrinos can lead to missing transverse energy, referred to by the acronym “MET” (denoted in the numerical calculation by the letter m). All events can be grouped into exclusive classes denoted as

$$(Nm, Nj, Nb, Ne, N\mu, N\tau, N\gamma),$$

where N is an integer number that defines the number of objects of a certain type, $v = j, b, e, \mu, \tau, \gamma$, in a collision event. In the following, the word “object” will be used for jets, b -jets, leptons, and photons. In the case of MET, Nm is either $0m$ (no significant MET) or $1m$ (when MET is above 200 GeV). Thus, an event class marked with $1m$ corresponds to one (or several) produced neutrinos. Using this notation, $(1m, 2j, 1b, 0e, 0\mu, 0\tau, 1\gamma)$ represents a class of events with large MET ($1m$), two light-flavor jets ($2j$), one b -jet ($1b$) and one photon (1γ).

To estimate how many unique topologies are expected from the SM, we have produced a sample of PYTHIA8 version 8.307 [1] MC events with pp collisions at $\sqrt{s} = 13$ TeV after enabling all SM processes of this generator. Similar to [2], the simulation used 44 physics sub-processes at leading-order QCD, such as light-flavor dijet production, all top production, weak single and double boson production, prompt photons and all Higgs SM processes. The cut on the two-body matrix elements in PYTHIA8 was set to 100 GeV. The total integrated luminosity of the simulation was 154 fb^{-1} , i.e. larger than the LHC Run2 data sample of 140 fb^{-1} . Since light-flavor QCD dijets cannot create the required complexity of the event classes, the generation of such events was relatively suppressed compared to other sub-processes with lower cross sections. The total number of generated events was 0.53 billion. Stable particles with a lifetime larger than $3 \cdot 10^{-10}$ seconds were considered, while neutrinos were

* HEP Division, Argonne National Laboratory, 9700 S. Cass Avenue, Argonne, IL 60439, USA.

excluded from consideration. The NNPDF 2.3 LO [3] parton density function, interfaced with PYTHIA8 via the LHAPDF library [4], is used. A detector simulation was not applied. The object reconstruction was the same as in [2]. Hadronic jets are reconstructed using the anti- k_T algorithm [5] with a distance parameter of $R = 0.4$ implemented in the FASTJET package [6]. The transverse momenta (p_T) of the jets must be greater than 20 GeV, and the pseudorapidity (η) must satisfy $|\eta| < 2.5$. A jet is classified as a b -jet if its four-momentum matches the momentum of a b -quark, and the b -quark contributes more than 50% of the total jet energy. Leptons and photons are required to be isolated. A cone of size 0.2 in azimuthal angle (ϕ) and η is defined around the true direction of the lepton. Then, all energies of particles inside this cone were summed up. A lepton is considered isolated if it carries more than 90% of the cone energy. The transverse momentum cut and the η cut were the same as for jets. The requirement for MET was 200 GeV, i.e. when $0m$ becomes $1m$ in the symbolic calculation.

According to the above SM MC simulation, the number of non-identical event topologies was 3,537. The maximum number of observed objects was $17j$, $8b$, $4e$, 4μ , 4τ and 4γ . We did not observe more than 20 objects per event. The total number of light-flavor b -jets was always less than 19. In addition, the total number of leptons was never larger than 5. All such restrictions can be called the “boundary” condition, which limits the number of possible event classes. They are summarized below:

$$\begin{aligned} Nm < 2, \quad Nj < 18, \quad Nb < 9, \quad Nj + Nb < 19, \\ Ne < 5, \quad N\mu < 5, \quad N\tau < 5, \quad N\gamma < 5, \\ Nj + N\ell < 19, \quad Nj + N\gamma < 19, \\ Nb + N\ell < 9, \quad Nb + N\gamma < 9, \\ N\ell < 6, \quad N\ell + N\gamma < 6, \\ Nj + Nb + N\ell + N\gamma < 21, \end{aligned} \quad (1)$$

where $N\ell = Ne + N\mu + N\tau$ is the total number of leptons. Events with $9j$ and $5b$ (but with transverse momenta larger than $p_T > 20$ GeV used in this paper) and four-lepton events have been recently studied by the LHC [7, 8], thus our MC simulation should be a reasonable representation of the reality.

We will keep a conservative view that a BSM phenomenon does not violate the boundary condition; otherwise, it can easily be found by looking at inclusive single-particle distributions of identified particles or jets. For example, an observation of events with five muons could alarm the observer in the past and thus such high-multiplicity events cannot represent the experimental challenge for their detection. But new phenomena can be “hidden” in exclusive combinatorial combinations, which are more intricate to discover experimentally. We will come back to the discussion of this point later.

Let us calculate how many combinatorial combinations are expected by preserving the SM boundary condition Eq. (1). First, we set the maximum number of objects to be observed to $Nmax = 20$. There are up to $n = 7$ objects per event (where MET is counted as an additional

“object”). The total number of combinations, where items can be repeated more than once and the ordering of items is not important, is

$$\sum_{r=2}^{Nmax} \frac{(n+r-1)!}{(n-1)!r!}. \quad (2)$$

Thus, the total number of unique combinations is 888022. Imposing the boundary condition Eq. (1) from the SM MC simulation is not straightforward using analytic calculation. However, such a calculation can be obtained numerically as shown in Appendix A. The obtained answer is 19,497 combinations.

The difference between the MC prediction (3,537) for SM processes and what, potentially, can be expected for the number of event classes from combinatorics (19,497) demonstrates that the MC event sample does not include all possible event classes. For example, event topologies such as

$$\begin{aligned} (0m, 2j, 2b, 2e, 3\mu, 0\tau, 0\gamma), \\ (1m, 4j, 0b, 3e, 1\mu, 1\tau, 1\gamma), \\ \text{etc.} \end{aligned} \quad (3)$$

have never been seen in the generated event sample for SM processes. The author does not know which BSM scenario can lead to such event classes. Note that the event topologies are defined in the restricted phase space, i.e. in the limited kinematic region defined by the transverse momentum and pseudorapidity selection. Thus, such event topologies cannot violate charge, lepton number, energy-momentum conservation, and other constraints. The difference of about 5 between the number of event classes predicted by the PYTHIA8 generator and by the numeric combinatorics can be an indication that the event generation may require more events. In addition, not all physics processes are included in the event generator. For example, next-to-leading order QCD effects may be in play. It should be noted that PYTHIA8 agrees well with alternative MC simulations up to six jets [9], but the other event topologies need to be verified too. We will put this question aside and assume that the total number of possible event classes is 19,497, as derived in the numeric computation with the PYTHIA8 boundary condition Eq. (1), but not what has been predicted by PYTHIA8 itself for the number of event classes. More realistic event generators may reduce the discrepancy between our numeric estimate and the generator predictions for the number of unique event classes, but they cannot change the conclusion of this paper, which does not rely on MC simulations.

How can we be sure that previous LHC studies were able to explore all such event topologies? According to the publication record of the ATLAS and CMS experiments, pp collision events have been studied in about 600 publicly available results using 140 fb^{-1} of data. For the sake of argument, let us assume that 5 non-identical

event classes were scrutinized in each paper[10], and they were found to match the SM predictions. This gives 3,000 investigated event classes. Note that one ATLAS publication [11] contains the studies of more than 700 event classes, but that analysis used a small fraction of the LHC Run2 data, and these event classes are expected to be a subset of the 3,000 event topologies assumed before. Therefore, the number of unexplored event topologies, out of 19,497 expected, is close to 81%.

If one considers fully reconstructed (identified) SM heavy particles, such as Z , W and top quarks, the number of event classes will increase. This can easily be checked by adding these additional particles in our numeric notation after reducing the maximum multiplicities of leptons from 4 to 3 (i.e. considering $W \rightarrow \ell\nu$ decays) and reducing the number of jets (b -jets) by one. We should also require that the total number of W , Z and top quarks cannot be greater than 3; the latter boundary condition makes this example more realistic. In this case, the number of event topologies will increase to more than 140,000.

III. DISCUSSION

When discussing the coverage of the event classes by the LHC studies, it is assumed that new phenomena predominantly contribute to a single event class, rather than to many event classes. The latter assumption, keeping in mind our model-agnostic approach, should be quite reasonable considering the fact that we do not know much about what can be expected from BSM physics.

From the standpoint of QCD, even if a BSM model is characterized by a very specific event class (say, with a fixed n_j number jets plus some fixed number of leptons and photons), additional event classes with extra jets can be produced due to the parton shower. The event rate of the events with n_j+1 jets is suppressed with respect to events with n_j jets by the strong coupling constant α_S (times the number of jets). However, ignoring softer jets involves an additional supposition that there is nothing interesting in high-jet multiplicity events since they originate from the QCD parton showering. This assumption is incompatible with the general search strategy since it must involve an undefined cutoff parameter that limits the number of jets and the entire scope of model-agnostic searches. This is why an exclusive approach to jet multiplicity has been adopted by the ATLAS [11] general searches.

From an experimental perspective, it is not unreasonable to think that some inclusive measurements may have certain sensitivity to the 19,497 event classes reported for the condition Eq. (1). This is because many studied distributions at the LHC are a “mix” of different event classes. In our view, inclusive measurements cannot effectively pinpoint a specific event topology produced with a small cross-section. Generally, searches in events with exclusive definitions of jets and particles, where any event

class with a fixed jet multiplicity is treated as a unique hadronic-final signature, are better motivated. For example, it is difficult to understand how an inclusive 2-jet measurement can ping point event class with additional 2 jets and a few leptons shown in Eq. (3), which may have a cross-section by several orders of magnitude smaller than the inclusive 2-jet measurements. Thus, it is necessary to carry out dedicated measurements focusing on such exclusive event topologies.

IV. EXPERIMENTAL SCENARIO

We also verified a scenario after changing the PYTHIA8 boundary condition Eq. (1) to:

$$\begin{aligned} Nm < 2, \quad Nj < 7, \quad Nb < 5, \quad Nj + Nb < 10, \\ Ne < 4, \quad N\mu < 4, \quad N\tau < 4, \quad N\gamma < 4, \\ N\ell < 5, \quad N\ell + N\gamma < 5, \\ Nj + Nb + N\ell + N\gamma < 13. \end{aligned} \quad (4)$$

These restrictions are within the reach of the current LHC studies since they correspond to the realistic reconstruction of multiple hadronic jets [7, 8]. The obtained number of event classes from the numeric calculation is 6,676. PYTHIA8 predicts 2,485 event classes, thus the condition in Eq. (4) reduces the discrepancy between our expectation and the generator. Assuming 3,000 investigated event classes at the LHC, the number of unexplored event topologies is about 55%.

The trigger thresholds used at the LHC are another noteworthy point for consideration. The minimum cut on the transverse momenta $p_T > 20 - 30$ GeV of jets, which can lead to large jet multiplicities, is rather low for effective trigger selection and for high purity in the reconstruction of jets. However, it should be pointed out that the main triggers to be used in such searches do not need to be based on jet triggers: Since almost every event contains e , μ , or γ , the events can effectively be triggered by using these particles, for which a 20 GeV requirement is not unusual. From the point of view of triggers, the most difficult categories for detection are the events with very few jets without associated production of electromagnetic particles or significant MET. But the number of such classes is very small compared to the overall number of combinations, thus it should not change much our conclusion regarding the total number of expected event classes.

V. HARD-QCD SCENARIOS

The previous consideration deals with situations when the numbers of jets can go up to 6. This can be a realistic scenario for some supersymmetric extensions of the Standard Model, where jets stem from exotic particles. However, many BSM models can only be characterized with several hard- p_T jets (say, up to 4), while the other

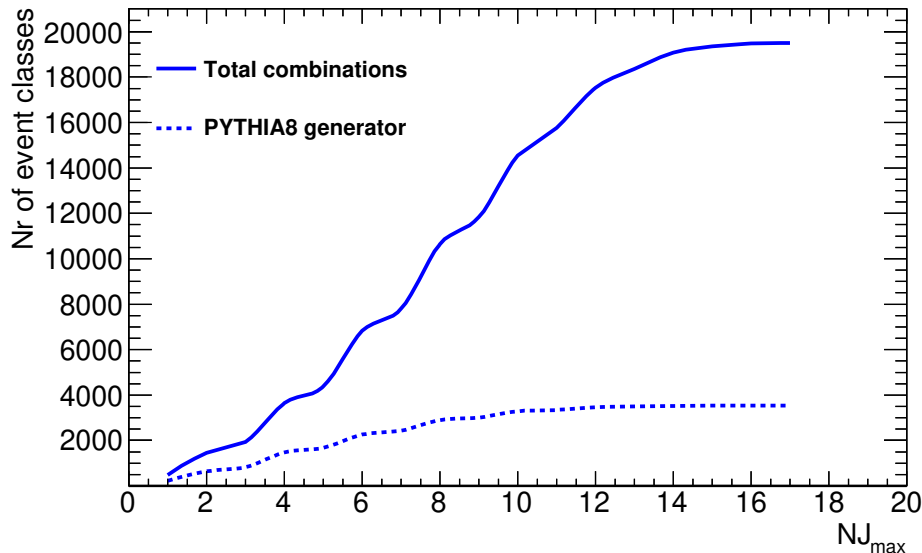


FIG. 1. The number of expected event classes from the combinatorial analysis and PYTHIA8 as a function of the maximum number of jets (NJ_{max}). It is assumed that the number of b -jet should be less than 50% of the total number of jets.

jets are produced due to the parton shower, where the event rate of the events with $Nj+1$ jets is suppressed with respect to the events with Nj jets by the strong coupling constant α_S (times the number of jets).

Therefore, we will consider a more inclusive definition of event classes using Eq. 4, but with the $Nj < 5$ and $Nb < 3$ restrictions. Electromagnetic particles do not shower; therefore, we preserve the original boundary condition for e , μ , τ and γ . This implies that we ignore events with higher jet multiplicities since they originate from the parton shower of quarks and gluons (but α_S suppressed) of the same process, thus they do not carry the new information about such BSM events. The obtained number of event classes from the numeric calculation is 3172. About 40% of such classes were not found in PYTHIA8. Such a strong reduction in event classes compared to events with larger jet multiplicities may also lead to conclusion that many of such events have been studied at the LHC.

Figure 1 demonstrates the dependence of the number of possible event classes as a function of the maximum number of jets (NJ_{max}). The calculations are presented for the numerical combinatorial analysis and for the PYTHIA8 simulation. During the calculation, we limit the number of b -jets to 50% of the NJ_{max} value. It can be concluded that the discrepancies between the expected number of event classes and PYTHIA8 become smaller for small values of NJ_{max} .

VI. HIGH-STATISTICS SCENARIO

Now we will come back to the original condition Eq. (1), and ask the following question: what is the number of classes with more than 9 events. This would correspond to about 3σ evidence for an observation of such event class. This will address the low-statistics problem since this requirement removes low-statistics events, which do not have the potential to lead to "evidences" for experimental observations.

Such estimates cannot be done in a model-independent way since we do not have the information about cross-sections of event classes which are not in PYTHIA8 (or any Monte Carlo simulation). The obtained number of combinations in PYTHIA8 with more than 9 events is 1,958 (out of the total 3,537). Thus, the total number of events which have too low statistics for any observational evidence is about 45%.

In reality, we cannot perform such estimates for hypothetical BSM processes with unknown cross sections. But if we assume that the same fraction of low-statistics events holds for 19,497 classes using the condition Eq. (1), we arrive to the number 10,723 for event classes with more than 9 events in each. Assuming 3,000 investigated event classes at the LHC, the number of unexplored event topologies is about 72%. Of course, this number is very speculative since we do not have any prior knowledge of the cross-sections of such hypothetical BSM processes.

VII. KINEMATIC CONSIDERATION

It is more difficult to understand the kinematic side of the argument beyond the object-multiplicity combinatorics. So far we assumed that all objects are produced in any detector region, following some density distributions expressed in terms of p_T , η and ϕ , and only a composition of their multiplicities can separate one event topology from the other. It is natural to expect that some BSM phenomena may be distinguished from the SM events by their distinct kinematics too. For example, heavy particles can predominately decay into two other jets/particles in the central detector region, while other BSM models may “prefer” to populate the forward detector regions.

We will use a simple toy consideration to calculate the number of possible kinematic features using combinatorics with substitution. Assume that all objects in 19,497 distinct event classes populate the detector phase space according to the SM expectations. We can define a new phenomenon if two objects approach close to each other, i.e. they are the decay products of low-mass states[12]. Such objects are still counted as two separate objects, but they form ensembles of kinematically unique events, and their production rate should be larger than that obtained from pure statistical noise around the SM-defined densities.

Let us count how many such unique kinematic topologies can exist by grouping jets and particles. For example, consider the event topology with one jet, one b -jet and one electron, such as:

$$(1j, 1b, 1e),$$

where we shorten the notation after removing $0m$, 0μ , 0τ and 0γ . This event topology creates 3 kinematically-distinct classes:

$$(0j, 1j1b, 1e), \quad (1j, 0b, 1b1e), \quad (1j1e, 1b, 0e),$$

where the four-character strings, $1j1b$, $1b1e$ and $1j1e$, represent three two-body groups with a certain dynamic correlation between the objects in each group. For example, such objects can be close to each other for a statistically significant number of events, since they stem from exotic low-mass states. Experimentally, these three combinations can be viewed as invariant masses of jet+(b -jet), e +(b -jet) and jet+ e with associated production of other objects produced anywhere in a detector following the SM single-particle densities. Now we can ask this question: how many such sub-classes of events exist out of 19,497 total combinations? The obtained number using numeric combinatorics is 159,674 (see Appendix A).

As before, now we need to estimate how many two-body distributions have been analyzed at the LHC. We assume that, for each of the 3,000 event classes studied at the LHC, at least one relevant two-body kinematic distribution (such as an invariant mass) has been inspected, and no deviations from the SM have been found. Therefore, for the expected 159,674 event classes with

two-body correlation, the chances that the LHC will encounter one of these topologies, which may have an excess over the SM background, are about 2%. This assumes that such events with correlations are explored uniformly across all the event topologies. This estimate can only be used as a conservative guide, or an upper limit on the actual LHC coverage of new phenomena, since this calculation does not consider charge topologies, correlations beyond the two-prong decays, known heavy SM particles, and other possibilities.

For the boundary condition Eq. (4), which is motivated by the recent LHC studies, the calculated number of two-particle sub-classes is 53,108. This leads to $\simeq 6\%$ kinematic distributions potentially explored at the LHC.

VIII. CONCLUSION

The modern approach to searches for new physics at the LHC is usually based on event signatures proposed by model builders. It is quite clear that the LHC has good coverage of event topologies with low jet/particle multiplicities and hard-QCD jets. But for events with large multiplicities, where jets are treated exclusively, the experimental coverage of the LHC is not large.

Nature can be more unpredictable, and more model-agnostic approaches can also be useful for discovering new physics in the LHC data. Our numeric analysis, guided by the large-scale SM simulations, reveals that the non-observation of new phenomena at the LHC is not unsurprising. If a BSM signal with unusual two-particle correlations can equally be found in any of the event classes discussed in this paper, then the chance that the LHC could detect such a new phenomenon is rather small, that is, about 2% (or 6%), depending on the boundary condition used in the numeric calculation. If we are only interested in jet/particle multiplicities, then the number of unexplored event classes is 81% (55%), leading to the probability of 19% (45%) for the observation of a new event topology at the LHC. These estimates assume that a BSM phenomenon can equally be found in any of those unexplored event classes and jets are treated excursively. These values represent the upper bounds on the probability of finding new phenomena because only the lightest identified particles were taken into account, and the calculation of kinematically distinct event classes includes only one feature (i.e. two-particle correlations). Despite the approximate nature of our calculations, they represent the first quantitative estimates obtained under the assumptions proposed in this paper. Thus, the LHC is still at the beginning of the journey to discover new physics.

In order to tackle the problem of searches for new phenomena in the vast number of possible event topologies reported in this paper, novel methods of data analysis, which rely less on expectations from BSM models, should be widely used. For example, unsupervised machine-learning methods can automatically label unusual event

classes as anomalies. Then such anomalous events can be compared with the SM predictions. Only very recently the LHC [13, 14] has started its physics program of using anomaly detection and fully unsupervised machine learning for complete event kinematics. For studies of multiplicities of event classes, one can train a neural network to reproduce the shapes of rates of event class as a function of their multiplicities using a small fraction of data or some control region. Comparing such shapes with the actual data would provide a useful tool to understand the “missing information” problem. We hope this or similar approaches will receive the needed attention, and the general searches with machine learning will realize their full potential to discover new physics in the near future.

ACKNOWLEDGMENTS

The submitted manuscript has been created by UChicago Argonne, LLC, Operator of Argonne National Laboratory (“Argonne”). Argonne, a U.S. Department of Energy Office of Science laboratory, is operated under

Contract No. DE-AC02-06CH11357. The U.S. Government retains for itself, and others acting on its behalf, a paid-up nonexclusive, irrevocable worldwide license in said article to reproduce, prepare derivative works, distribute copies to the public, and perform publicly and display publicly, by or on behalf of the Government. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan. <http://energy.gov/downloads/doe-public-access-plan>. Argonne National Laboratory’s work was funded by the U.S. Department of Energy, Office of High Energy Physics under contract DE-AC02-06CH11357. We gratefully acknowledge the computing resources provided by the Laboratory Computing Resource Center at Argonne National Laboratory.

Appendix A: Code availability

The PYTHIA8 settings, the generator code and the output data used in this study can be accessed via [15]. The numeric code for the combinatorial analysis is implemented in PYTHON 3 without external dependencies.

-
- [1] T. Sjostrand, S. Mrenna, and P. Z. Skands, A Brief Introduction to PYTHIA 8.1, *Comput. Phys. Commun.* **178**, 852 (2008), arXiv:0710.3820 [hep-ph].
 - [2] S. V. Chekanov, J. T. Childers, J. Proudfoot, D. Frizzell, and R. Wang, Precision searches in dijets at the HL-LHC and HE-LHC, *JINST* **13** (05), P05022, arXiv:1710.09484 [hep-ex].
 - [3] R. D. Ball *et al.* (NNPDF), Parton distributions for the LHC Run II, *JHEP* **04**, 040, arXiv:1410.8849 [hep-ph].
 - [4] A. Buckley, J. Ferrando, S. Lloyd, K. Nordström, B. Page, M. Rüfenacht, M. Schönherr, and G. Watt, LHAPDF6: parton density access in the LHC precision era, *Eur. Phys. J. C* **75**, 132 (2015), arXiv:1412.7420 [hep-ph].
 - [5] M. Cacciari, G. P. Salam, and G. Soyez, The anti- k_T jet clustering algorithm, *JHEP* **04**, 063, arXiv:0802.1189 [hep-ph].
 - [6] M. Cacciari, G. P. Salam, and G. Soyez, FastJet User Manual, *Eur. Phys. J. C* **72**, 1896 (2012), <http://fastjet.fr/>, arXiv:1111.6097 [hep-ph].
 - [7] ATLAS Collaboration (ATLAS), Search for phenomena beyond the Standard Model in events with large b -jet multiplicity using the ATLAS detector at the LHC, *Eur. Phys. J. C* **81**, 11 (2021), [Erratum: *Eur. Phys. J. C* **81**, 249 (2021)], arXiv:2010.01015 [hep-ex].
 - [8] CMS Collaboration (CMS), Search for vector-like leptons in multilepton final states in proton-proton collisions at $\sqrt{s} = 13$ TeV, *Phys. Rev. D* **100**, 052003 (2019), arXiv:1905.10853 [hep-ex].
 - [9] ATLAS Collaboration (ATLAS), Multijet simulation for 13 TeV ATLAS Analyses (2019), ATL-PHYS-PUB-2019-017.
 - [10] In our view, this number of studied event classes is very optimistic. Typically, an experimental paper has only a few such distributions, and most of them are too inclusive to pinpoint a specific event class. In addition, ATLAS and CMS often repeat measurements using similar final states. Due to the lack of exact statistics on published distributions, we prefer to consider the most optimistic scenario.
 - [11] ATLAS Collaboration (ATLAS), A strategy for a general search for new phenomena using data-derived signal regions and its application within the ATLAS experiment, *Eur. Phys. J. C* **79**, 120 (2019), arXiv:1807.07447 [hep-ex].
 - [12] In this discussion, “low-mass states” mean BSM particles with masses close to the sum of the masses of the daughter particles. One can also consider an alternative kinematic situation when two objects are back-to-back in the central region of a detector if they originate from the decay products of a high-mass state. However, the production rates of such events are expected to be significantly more suppressed due to large masses, compared to low-mass resonances leading to collimated production of jets/particles.
 - [13] ATLAS Collaboration (ATLAS), Search for New Phenomena in Two-Body Invariant Mass Distributions Using Unsupervised Machine Learning for Anomaly Detection at $\sqrt{s}=13$ TeV with the ATLAS Detector, *Phys. Rev. Lett.* **132**, 081801 (2024), arXiv:2307.01612 [hep-ex].
 - [14] CMS Collaboration (CMS), *Model-agnostic search for di-jet resonances with anomalous jet substructure in proton-proton collisions at $\sqrt{s} = 13$ TeV*, Tech. Rep. (CERN, Geneva, 2024).

- [15] S. V. Chekanov, GitHub repository for the numeric code and data used in HEP-ANL-186383, <https://github.com/chekanov/HEP-ANL-186383> (2023).