

Application of Kolmogorov–Arnold Networks in High Energy Physics

E. E. Abasov^{1*}, P. V. Volkov¹, G. A. Vorotnikov¹, L. V. Dudko¹,
A. D. Zaborenko¹, E. S. Iudin¹, A. A. Markina¹, and M. A. Perfilov¹

¹Skobeltsyn Institute of Nuclear Physics, Lomonosov Moscow State University, Moscow, 119991 Russia

Received October 1, 2024; revised October 10, 2024; accepted October 20, 2024

Abstract—Kolmogorov–Arnold Networks represent a recent advancement in machine learning, with the potential to outperform traditional perceptron-based neural networks across various domains as well as provide more interpretability with the use of symbolic formulas and pruning. This study explores the application of KANs to specific tasks in high-energy physics. We evaluate the performance of KANs in distinguishing multijet processes in proton–proton collisions and in reconstructing missing transverse momentum in events involving dark matter.

Keywords: top-quark, dark matter, neural networks

DOI: 10.3103/S0027134924702035

1. INTRODUCTION

Neural networks have been playing a significant role in high-energy physics for the past decades. They help improve event reconstruction in detectors, process classification, particle reconstruction, and many other important tasks. Potential architectures range from simple multilayered perceptrons to complex transformer-based models [1]. All of these types of networks, however, are based on the same principle of perceptron [2, 3], which is, in essence, a learnable linear function with some fixed nonlinear activation function. The possibility of their usage for function fitting is based on the Universal Approximation Theorem [4].

Recently, a new approach was developed [5] that uses the Kolmogorov–Arnold representation theorem [6–9] instead. Kolmogorov–Arnold networks utilize learnable splines instead of linear functions and, thus, promise to provide greater level of accuracy than perceptron-based models. In addition, KANs allow for the splines to be fixed automatically or manually to a variety of analytical functions, which could prove instrumental in a greater understanding of neural networks output.

This study explores two potential applications for this new type of networks in the field of top quark physics: the separation of multijet processes, for which a traditional MLP is typically used, described

in Section 2, and the reconstruction of the missing 4-momentum of invisible particles in processes involving dark matter, outlined in Section 3. In addition, a comparison of the original implementation of KAN, further denoted as pyKAN [10], and an optimized but feature-bare Efficient KAN (eKAN) [11] is provided. All neural networks are constructed and trained in the PyTorch framework [12].

2. CLASSIFICATION TASK: MULTIJET PROCESSES SEPARATION

The first task on which KANs were tested is the separation of the multijet QCD background in the

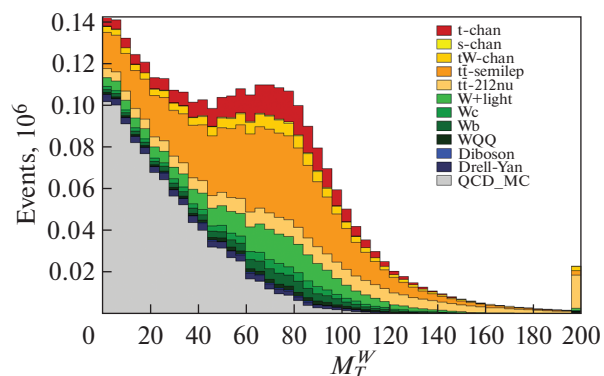


Fig. 1. M_T^W distribution for single-top production analysis. The QCD contribution is shown in grey.

*E-mail: emil@abasov.ru

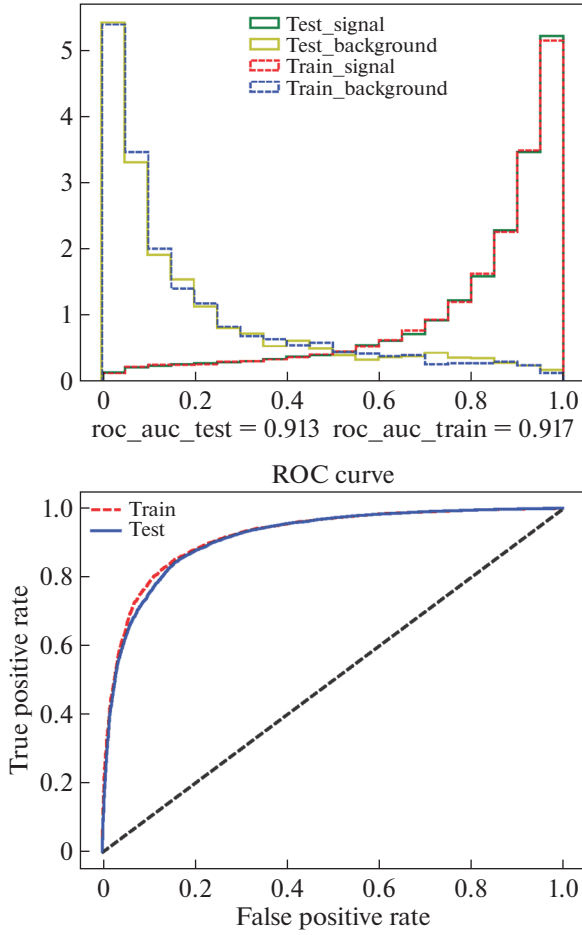


Fig. 2. Output of the MLP for the multijet QCD separation (top) and the ROC curve (bottom).

framework of single top-quark analysis. QCD processes have a very high production rate and represent an important, but difficult-to-model background. Therefore, in order to reduce bias from the QCD normalization uncertainty, this process is usually suppressed.

Both the QCD background and other processes are modelled on the parton level in the POWHEG-BOX [13], CompHEP4.6 [14, 15], and MadGraph5 [16, 17] Monte Carlo generators. The hadronization of these processes is done in Pythia8 [18], detector response is simulated in DELPHES [19] using the CMS detector profile.

The traditional approach of QCD suppression involves cuts on kinematic variables, sensitive to it, such as M_T^W , shown in Fig. 1, a reconstructed transverse mass of the W boson from top-quark decay. However, this approach cuts out a substantial amount of other processes, too, and, as a consequence, reduces the statistics for the next steps of the analysis.

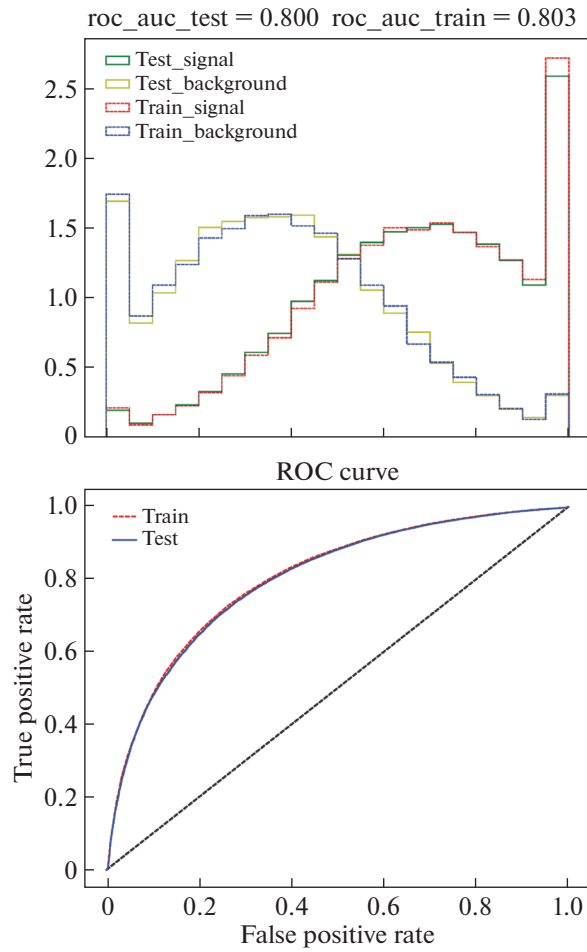


Fig. 3. Output of the KAN for the multijet QCD separation (top) and the ROC curve (bottom). The network was trained on mean squared error as the loss function and clipped to the (0, 1) range.

2.1. Baseline Network

Another approach to this problem, the use of neural networks, has proven to be more successful [20]. As an example of this method, a simple 2-layered perceptron with 100 neurons per layer is trained on the classification of QCD and other events, achieving the ROC AUC score of 0.91. LeakyRELU activation function is used, the network is trained using binary cross-entropy as the loss function. In order to prevent overfitting, dropout of 30% is used in conjunction with L2 regularization. This MLP is taken as a baseline with which all KAN implementations are compared in this section, its performance is shown in Fig. 2.

2.2. KAN Implementation

Firstly, the original pyKAN implementation [10] was tested. Despite the rich functionality, at the time of writing, it is plagued with various bugs, such as

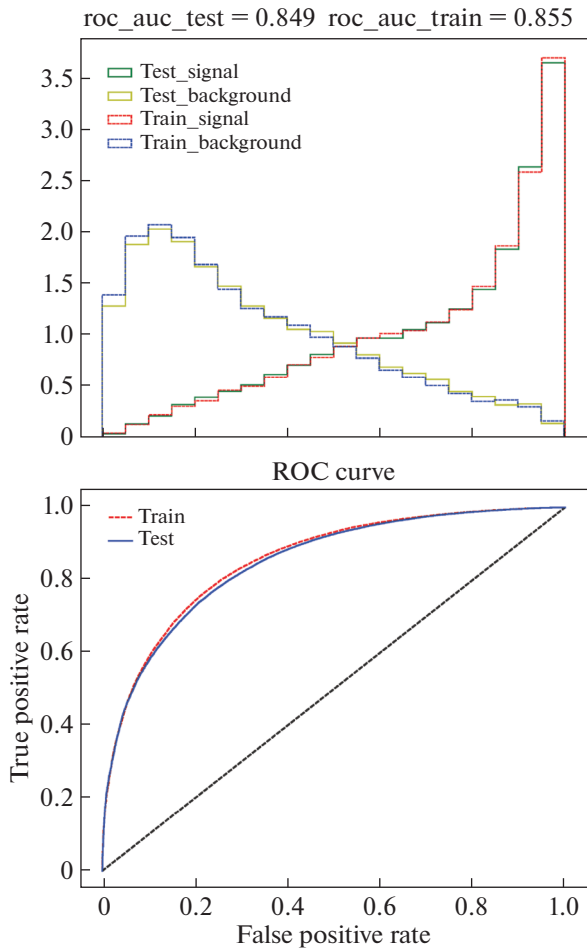


Fig. 4. Output of the Efficient KAN for the multijet QCD separation (top) and the ROC curve (bottom). The network was trained with binary cross-entropy as the loss function.

inability to train the network on GPUs, poor training with cross-entropy loss function, and general instability. Due to that fact, the only working solution for this task was to train the KAN using the mean squared error loss function instead of cross-entropy and then clip the output to the (0, 1) range. Results were provided for two-layered KAN with 64 nodes in the middle layer. KANs utilize local B-spline basis functions, so the configuration of the grid, on which these functions are nonzero, is important. The amount of B-splines for each activation function is controlled by grid size. Cubic splines were used for the fitting, with a grid size of 5. It should be noted that the LBFGS optimizer was used for all KAN applications. As can be seen in Fig. 3, the ROC AUC score, as well as the general shape of the distribution, is worse than the baseline MLP.

It should be noted that the pyKAN is being regularly updated with various bugfixes being implemented.

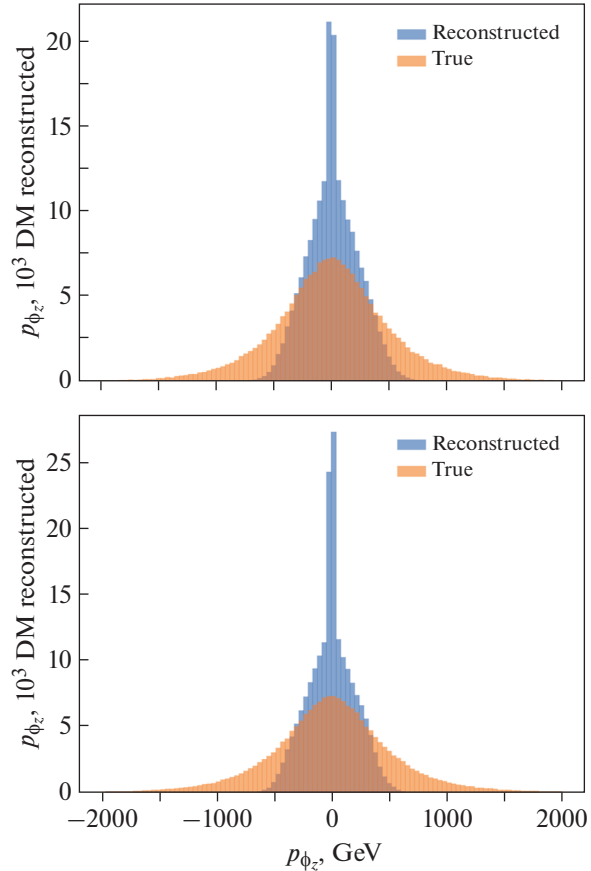


Fig. 5. Reconstruction of the longitudinal component of the dark matter mediator momentum for the Baseline MLP from [25] (top), Efficient KAN (bottom).

The Efficient KAN [11] is a pure bare-bones PyTorch implementation of KAN. The only functionality available in this library is the model initialization, without the ability to plot, prune, or use the symbolic regression. On the other hand, it does not exhibit the aberrant behaviour of the pyKAN.

For the QCD classification, the shape of the network, as well as the spline order and grid are the same as in the pyKAN example. One noticeable difference in this case is the use of binary cross-entropy for the training procedure, which was not possible for pyKAN. The output of the network and its ROC curve are shown in Fig. 4. eKAN performs noticeably better than the pyKAN, mostly due to the availability of the cross-entropy as the loss function, while still not being able to beat the baseline.

In conclusion, in this task the baseline MLP proves to be superior to both KAN implementations, most likely owing to the simplicity of the task, which can be solved even with the simplest neural networks.

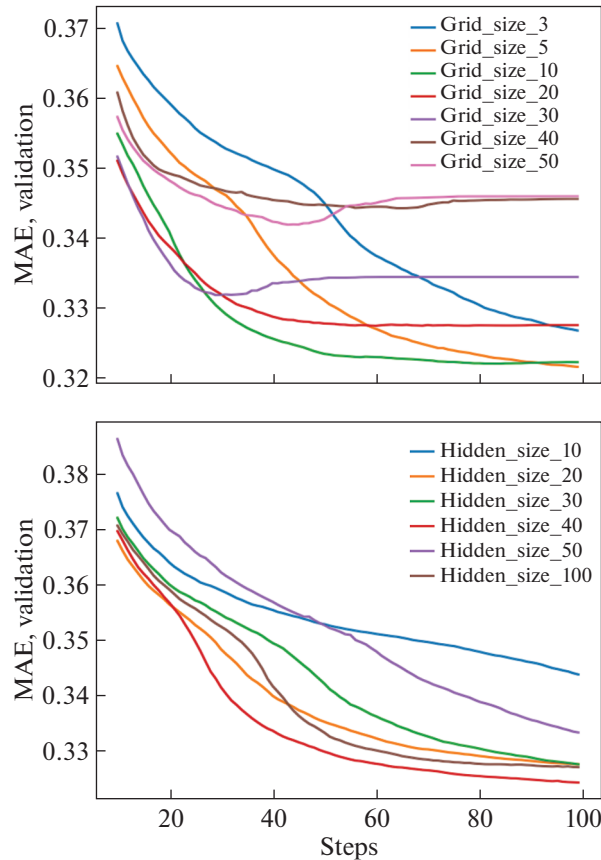


Fig. 6. Hyperparameter optimization: grid size—top, hidden layer size—bottom. Training is shown from the 10th step.

3. REGRESSION TASK: RECONSTRUCTION OF THE MISSING MOMENTUM

Another possible application of Kolmogorov–Arnold networks arises in the study of top-quark production in association with dark matter particles.

The possible interaction of dark matter particles with SM particles can be realized by means of an intermediate particle, the so-called mediator. For simplified DM models [21–24] it is assumed that DM particles interact with SM particles by exchanging one or more particles, called mediators, which have weak coupling with SM particles. This section focuses on single-top quark production with subsequent leptonic decay, which is well studied in the SM. One of the significant properties of such process in SM is correlation of the spin of top-quark and its decay products with the direction of the down-type quark. From that one can get the relation for the cosine of angle between down-type quark and the lepton in a top-quark decay. This relation changes in the presence of heavy scalar mediator, due to the change it makes in calculating the top-quark rest frame, in which the target variable is constructed [25]. Thus, in order to properly reconstruct this variable, it

is required to separate contributions of two invisible particles: neutrino and dark matter mediator.

As the baseline, the multilayered perceptron was used to reconstruct components of the 4-momentum of both particles, however, for the mediator this method did not yield desired results. Based on the same data, the Kolmogorov–Arnold network from the Efficient KAN framework is also trained. In order to compare the results obtained with the MLP baseline from [25], the KAN is trained on the same objective function: mean absolute error (MAE). The MAE score on the test dataset is 0.28 for the MLP and 0.30 for the KAN, which implies similar performance for these networks. This can be further reinforced by comparing the reconstructed momentum with its target value for both approaches (Fig. 5). The KAN was further tuned with respect to the following hyperparameters: grid size, hidden size and depth of the network (Fig. 6). The best performance was achieved with 2 hidden layers with 40 nodes per layer. With the performance of the neutrino and mediator momenta for KAN and MLP being almost identical, neither approach allows for adequate reconstruction of both particles.

4. RESULTS

Kolmogorov–Arnold networks provide a promising new alternative to traditional perceptron-based models. At the time of writing this article neither of available KAN implementations reaches full potential, with original pyKAN suffering from numerous bugs and the Efficient KAN and its derivatives lacking most of the original functionality, such as pruning and symbolic regression. For both of the problems discussed in this article, KANs perform either on par or slightly worse than traditional MLPs. With further improvements to the original pyKAN package, however, it is hoped that KANs could outperform traditional neural networks in these tasks and provide new level of interpretability.

FUNDING

This work is supported by the Russian Science Foundation, project no. 22-12-00152.

CONFLICT OF INTEREST

The authors of this work declare that they have no conflicts of interest.

REFERENCES

1. T. Plehn, A. Butter, B. Dillon, T. Heimel, C. Krause, and R. Winterhalder, arXiv Preprint (2022).
<https://doi.org/10.48550/arXiv.2211.01421>
2. W. S. McCulloch and W. Pitts, *Bull. Math. Biophys.* **5**, 115 (1943).
<https://doi.org/10.1007/bf02478259>
3. F. Rosenblatt, *The Perceptron, a Perceiving and Recognizing Automaton Project Para: A Report* (Cornell Aeronautical Laboratory, 1957).
4. K. Hornik, M. Stinchcombe, and H. White, *Neural Networks* **2**, 359 (1989).
[https://doi.org/10.1016/0893-6080\(89\)90020-8](https://doi.org/10.1016/0893-6080(89)90020-8)
5. Z. Liu, Yi. Wang, S. Vaidya, F. Ruehle, J. Halverson, M. Soljacic, T. Y. Hou, and M. Tegmark, arXiv Preprint (2024).
<https://doi.org/10.48550/arXiv.2404.19756>
6. V. Arnold, in *Collected Works*, Ed. by A. Givental et al., Vladimir I. Arnold—Collected Works, Vol. 1 (Springer, Berlin, 1957), p. 5.
https://doi.org/10.1007/978-3-642-01742-1_2
7. V. Arnold, in *Seventeen Papers on Analysis*, American Mathematical Society Translations: Series 2, Vol. 28 (1963), p. 61.
<https://doi.org/10.1090/trans2/028>
8. A. N. Kolmogorov, *Dokl. Akad. Nauk SSSR* **108**, 179 (1956).
9. A. N. Kolmogorov, *Dokl. Akad. Nauk SSSR* **114**, 953 (1957).
10. Kolmogorov–Arnold networks.
<https://github.com/KindXiaoming/pykan>.
11. An efficient pure-PyTorch implementation of Kolmogorov–Arnold network (KAN).
<https://github.com/Blealtan/efficient-kan>.
12. A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimsheine, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. Devito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, in *Advances in Neural Information Processing Systems*, Ed. by H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (Curran Associates, 2019), Vol. 32.
https://proceedings.neurips.cc/paper_files/paper/2019/file/bdbca288fee7f92f2bfa9f7012727740-Paper.pdf.
13. C. Oleari, *Nucl. Phys. B, Proc. Suppl.* **205–206**, 36 (2010).
<https://doi.org/10.1016/j.nuclphysbps.2010.08.016>
14. E. Boos, V. Bunichev, M. Dubinin, L. Dudko, V. Edneral, V. Ilyin, A. Kryukov, V. Savrin, A. Semenov, and A. Sherstnev, *Nucl. Instrum. Methods Phys. Res., Sect. A* **534**, 250 (2004).
<https://doi.org/10.1016/j.nima.2004.07.096>
15. A. Pukhov, E. Boos, M. Dubinin, V. Edneral, V. Ilyin, D. Kovalenko, A. Kryukov, V. Savrin, S. Shichanin, and A. Semenov, arXiv Preprint (1999).
<https://doi.org/10.48550/arXiv.hep-ph/9908288>
16. R. Frederix, S. Frixione, V. Hirschi, D. Pagani, H.-S. Shao, and M. Zaro, *J. High Energy Phys.* **2018**, 185 (2018).
[https://doi.org/10.1007/JHEP07\(2018\)185](https://doi.org/10.1007/JHEP07(2018)185)
17. J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H.-S. Shao, T. Stelzer, P. Torrielli, and M. Zaro, *J. High Energy Phys.* **2014**, 79 (2014).
[https://doi.org/10.1007/JHEP07\(2014\)079](https://doi.org/10.1007/JHEP07(2014)079)
18. Ch. Bierlich, S. Chakraborty, N. Desai, L. Gellersen, I. Helenius, P. Ilten, L. Lönnblad, S. Mrenna, S. Prestel, Ch. T. Preuss, T. Sjöstrand, P. Skands, M. Uthmeim, and R. Verheyen, *SciPost Physics Codebases*, 8 (2022).
<https://doi.org/10.21468/scipostphyscodeb.8>
19. J. de Favereau, Ch. Delaere, P. Demin, A. Giammanco, V. Lemaitre, A. Mertens, and M. Selvaggi, *J. High Energy Phys.* **2014**, 57 (2014).
[https://doi.org/10.1007/JHEP02\(2014\)057](https://doi.org/10.1007/JHEP02(2014)057)
20. E. E. Abasov, M. I. Belobrova, P. V. Volkov, G. A. Vorotnikov, L. V. Dudko, A. D. Zaborenko, M. A. Perfilov, and E. S. Sivakova, *Phys. At. Nucl.* **85**, 708 (2023).
<https://doi.org/10.1134/s1063778822060023>
21. J. Alwall, P. C. Schuster, and N. Toro, *Phys. Rev. D* **79**, 75020 (2009).
<https://doi.org/10.1103/physrevd.79.075020>

22. J. Abdallah, H. Araujo, A. Arbey, A. Ashkenazi, A. Belyaev, J. Berger, C. Boehm, A. Boveia, A. Brennan, J. Brooke, O. Buchmueller, M. Buckley, G. Busoni, L. Calibbi, S. Chauhan, N. Daci, G. Davies, I. De Bruyn, P. De Jong, A. De Roeck, K. de Vries, D. Del Re, A. De Simone, A. Di Simone, C. Doglioni, M. Dolan, H. K. Dreiner, J. Ellis, S. Eno, E. Etzion, M. Fairbairn, B. Feldstein, H. Flaecher, E. Feng, P. Fox, M.-H. Genest, L. Gouskos, J. Gramling, U. Haisch, R. Harnik, A. Hibbs, S. Hoh, W. Hopkins, V. Ippolito, T. Jacques, et al., *Phys. Dark Universe* **9–10**, 8 (2015).
<https://doi.org/10.1016/j.dark.2015.08.001>
23. D. Abercrombie, N. Akchurin, E. Akilli, J. A. Maestre, B. Allen, B. A. Gonzalez, J. Andrea, A. Arbey, G. Azuelos, P. Azzi, M. Backović, Ya. Bai, S. Banerjee, J. Beacham, A. Belyaev, A. Boveia, A. J. Brennan, O. Buchmueller, M. R. Buckley, G. Busoni, M. Buttignol, G. Cacciapaglia, R. Caputo, L. Carpenter, N. F. Castro, G. G. Ceballos, Ya. Cheng, J. P. Chou, A. C. Gonzalez, Ch. Cowden, F. D'Eramo, A. De Cosa, M. De Gruttola, A. De Roeck, A. De Simone, A. Deandrea, Z. Demiragli, A. DiFranzo, C. Doglioni, T. du Pree, R. Erbacher, J. Erdmann, C. Fischer, H. Flaecher, et al., *Phys. Dark Universe* **27**, 100371 (2020).
<https://doi.org/10.1016/j.dark.2019.100371>
24. D. Alves, N. Arkani-Hamed, S. Arora, Ya. Bai, M. Baumgart, J. Berger, M. Buckley, B. Butler, S. Chang, H.-C. Cheng, C. Cheung, R. S. Chivukula, W. S. Cho, R. Cotta, M. D'Alfonso, S. E. Hedri, R. Essig, J. A. Evans, L. Fitzpatrick, P. Fox, R. Franceschini, A. Freitas, J. S. Gainer, Yu. Gershtein, R. Gray, T. Gregoire, B. Gripaios, J. Gunion, T. Han, A. Haas, P. Hansson, J. Hewett, D. Hits, J. Hubisz, E. Izaguirre, J. Kaplan, E. Katz, C. Kilic, H.-D. Kim, R. Kitano, S. A. Koay, P. Ko, D. Krohn, et al. (LHC New Physics Working Group), *J. Phys. G: Nucl. Part. Phys.* **39**, 105005 (2012).
<https://doi.org/10.1088/0954-3899/39/10/105005>
25. E. Abasov, E. Boos, V. Bunichev, P. Volkov, G. Vorotnikov, L. Dudko, A. Zaborenko, E. Iudin, A. Markina, and M. Perfilov, *Int. J. Mod. Phys. A*, null (2025).
<https://doi.org/10.1142/S0217751X24501264>

Publisher's Note. Allerton Press, Inc. remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

AI tools may have been used in the translation or editing of this article.