

Solving key challenges in collider physics with foundation models

Vinicius Mikuni^{1,*} and Benjamin Nachman^{2,3,†}

¹*National Energy Research Scientific Computing Center, Berkeley Lab, Berkeley, California 94720, USA*

²*Physics Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA*

³*Berkeley Institute for Data Science, University of California, Berkeley, California 94720, USA*



(Received 21 May 2024; accepted 25 February 2025; published 17 March 2025)

Foundation models are neural networks that are capable of simultaneously solving many problems. Large language foundation models like ChatGPT have revolutionized many aspects of daily life, but their impact for science is not yet clear. In this paper, we use a new foundation model for hadronic jets to solve three key challenges in collider physics. In particular, we show how experiments can (1) save significant computing power when developing reconstruction algorithms, (2) perform a complete uncertainty quantification for high-dimensional measurements, and (3) search for new physics with model agnostic methods using low-level inputs. In each case, there are significant computational or methodological challenges with current methods that limit the science potential of deep learning algorithms. By solving each problem, we take jet foundation models beyond proof-of-principle studies and into the toolkit of practitioners.

DOI: [10.1103/PhysRevD.111.L051504](https://doi.org/10.1103/PhysRevD.111.L051504)

Deep learning is enabling researchers to analyze high-dimensional data holistically without first projecting into low-dimensional summary statistics. New methods in areas such as parameter estimation [1], deconvolution [2,3], and anomaly detection [4–6] are enabling data analyses that were unimaginable before deep learning. We are also starting to see the first results with real data [7–17]. While the machine learning methods and computing exist to train these methods, there are still critical challenges to deploying them in practice and scaling them to the full phase space. The state-of-the-art machine learning architectures are data hungry and require (and can leverage [18]) massive datasets for training. In many cases, we simply cannot generate datasets large enough—either because the full simulation of a detector is too slow or because running an experiment longer is too expensive. Without new meta methods, we will not be able to harness the power of modern machine learning to make the most of our precious data.

Foundation models have emerged from industry as meta methods capable of enabling a diverse set of downstream deep learning tasks. These models are trained with enormous and diverse datasets that can then be finetuned to other use cases with little or no data. Most of the excitement

around foundation models has been in the area of text, where large language models (LLMs) have clearly demonstrated paradigm-shifting capabilities for the way we interact with the world around us. LLMs are trained on large corpuses of text to fill in the blanks from masked out words or sentences. There have been some proposals in particle physics to use this and other types of *self-supervised representation learning* for datasets composed of particles [19–24]. An alternative approach is to promote transfer learning to a foundation model through *supervised representation learning*. In this case, a model is trained on one dataset and then applied elsewhere. A number of particle physics studies have used transfer learning of this kind [25–28], but only consider one downstream task and so are not yet foundation models. Across all of these promising studies, there has not yet been a demonstrated foundation model that has been used to solve open scientific challenges.

Our goal is to show that foundation models for particle physics can be used to address outstanding challenges with realistic datasets. We make use of a new foundation model called OMNILEARN based on supervised representation discussed in detail in Ref. [29]. OMNILEARN was trained on a large dataset and can provide state-of-the-art results for accuracy, precision, and/or speed of many downstream tasks involving hadronic final states across detectors and collision systems and for both classification and generation. Interestingly, by aligning the training with the inference goals, OMNILEARN is a compact model with less than 2M parameters. This model size fits comfortably in a single graphical processing unit (GPU) and empowers scientists with restricted computing resources to benefit.

*Contact author: vmikuni@lbl.gov

†Contact author: bpnachman@lbl.gov

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/). Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Funded by SCOAP³.

Despite its small size, we consider OMNILEARN to be a foundation model that can address some of the most important challenges in machine learning for collider physics. Here, we show results for the most promising applications starting with a simulation challenge. Full detector simulations at the LHC are computationally expensive, and generating large enough datasets to train state-of-the-art models is becoming a limiting factor for many analyses. OMNILEARN was trained using data from a public fast simulation—can such a model reduce the number of fully simulated events required to achieve cutting edge performance? Second, we address a training challenge. A growing number of machine learning-based methods require the estimation of likelihood ratios. Uncertainties on this estimate require rerunning the learning thousands or tens of thousands of times [11,13], which is prohibitive for full phase-space inference. Can OMNILEARN reduce the time required to train these models so that the full uncertainty quantification can be achieved? Third, we address an anomaly detection challenge. Anomaly detection methods that are trained directly on data are fundamentally limited by the training dataset size. As a result, high-dimensional methods struggle to obtain the relevant sensitivity to rare signals in high-dimensional feature spaces [30]. Can OMNILEARN push the sensitivity of such methods to find signals that would not have been found before?

The details of OMNILEARN can be found in Ref. [29] but are briefly summarized in the following. The backbone network of OMNILEARN leverages modern developments for jet representation, using a combination of attention mechanisms [31] and dynamic convolutional operations [32] to improve both the global and local description of particles clustered inside jets. This model is named a point-edge transformer (PET), consisting of a shared representation (PET body) and two task specific heads, used for the tasks of classification and particle generation. The introduction of tasks specific network components makes the model modular and able to discard irrelevant heads during downstream tasks, thus reducing even further the model size during adaptation. The pretrained PET model is then referred to as OMNILEARN. OMNILEARN is trained using the JetClass dataset [33], consisting of 100 M jets and featuring 10 different jet classes. Each jet is represented as a set of constituent particles.

Jets are ubiquitous in high-energy collider physics, and identifying the origin of a jet (“jet tagging”) is a key component of a multitude of direct searches for new phenomena and precision measurements. Improvements in jet tagging performance directly translate into improved sensitivity of many analyses. State-of-the-art jet tagging models require many tens of millions of jets for training. As full detector simulations are expensive, it is prohibitive to generate large enough datasets for every tagging task and for every time there are changes in the simulation set (e.g., better particle-level description or different operational

conditions). Our hypothesis is that we can adapt a foundation model trained on less accurate fast simulation with a small sample of realistic simulations to achieve competitive performance on the full realistic problem.

We exemplify this idea showing the results obtained by OMNILEARN adapted to the publicly available ATLAS Top tagging dataset [34]. In this dataset, 40M events are generated with PYTHIA8 using the NNPDF2.3LO [35] set of parton distribution functions and the A14 [36] set of tuned parameters. Pileup effects are simulated by overlaying inelastic interactions on top of the underlying hard scattering process based on the 2017 data taking period. Hadronic boosted top quarks are obtained in simulated events containing the decay of a heavy Z' boson with mass of 2 TeV. Background jets are obtained from simulations of generic dijet events. Unified flow objects [37] are used to determine the jet constituents. Jets are clustered using anti- k_T algorithm [38–40] with $R = 1.0$ with additional pileup mitigation algorithms [41–43] applied. The soft-drop algorithm [44] is also applied to remove soft and wide-angle radiation. Note that unlike the JetClass dataset, the ATLAS top tagging dataset features full detector simulation, event reconstruction, and pileup particles.

We consider the scenario where we use the entire training data for the adaptation of OMNILEARN and when only 10% of the data is used (4M events). Results are reported in Table. I for four metrics: the area under curve (AUC), accuracy for a fixed threshold of 0.5, and the inverse background efficiency at two fixed values of the signal efficiency. We observe that OMNILEARN excels the performance of all previously reported benchmarks in this dataset and is able to match the previously best performing model

TABLE I. Comparison between the performance reported for different classification algorithms on the ATLAS top tagging dataset. Bold results represent the algorithm with highest performance. All results besides OMNILEARN and PET are taken from [34]. Five PET and OMNILEARN models are trained, and the error of multiple predictions is at the last decimal place and hence not shown.

	AUC	Acc	$1/\epsilon_B$	
			$\epsilon_S = 0.5$	$\epsilon_S = 0.8$
ResNet 50	0.885	0.803	21.4	5.13
EFN	0.901	0.819	26.6	6.12
hIDNN	0.938	0.863	51.5	10.5
DNN	0.942	0.868	67.7	12.0
PFN	0.954	0.882	108.0	15.9
ParticleNet	0.961	0.894	153.7	20.4
PET classifier (4M)	0.959	0.890	146.5	19.4
OMNILEARN (4M)	0.961	0.894	172.1	20.8
PET classifier (40M)	0.964	0.898	201.4	23.6
OMNILEARN (40M)	0.965	0.899	207.3	24.1

using only 10% of the data, thus requiring significantly fewer examples to achieve state-of-the-art performance.

Correcting physics measurements for detector distortions enables efficient comparisons between measurements and theory predictions. This technique is known in collider physics as *unfolding*. Machine learning greatly increases the flexibility and potential of unfolding by allowing the simultaneous correction of multiple distributions without the use of histograms [2]. The OMNIFOLD algorithm [45,46] introduced an iterative approach for unfolding based on learned classifiers that use the data collected by experiments to determine the unfolded distributions. Since the methodology relies on the data availability, physics processes with limited data can severely restrict the precision the algorithm can achieve. In contrast, the general representation of jets learned by OMNILEARN can compensate for the limited data. We verify this observation adapting OMNILEARN for complete unfolding using all available features in the dataset introduced in Ref. [45]. The dataset consists of proton-proton collisions producing a Z boson, generated at a center-of-mass energy of $\sqrt{s} = 14$ TeV. A sample used as the “data” representative is simulated using particle collisions with the default tune of Herwig 7.1.5 [47–49]. A second dataset, representative of the “simulation” we want to correct, is simulated using PYTHIA8 with Tune 21 [50]. Detector distortions are simulated with DELPHES and the CMS tune that uses a particle flow reconstruction. Jets are clustered using the anti- k_T algorithm with radius parameter $R = 0.4$ as implemented in FASTJET3.3.2 package [39]. The Z bosons are required to have $p_T > 200$ GeV in order to mitigate acceptance effects. Note that the corresponding jets are much lower momentum than those in JetClass (between 500–1000 GeV).

A key challenge with OMNIFOLD is that it requires training $2n$ neural networks every time the algorithm is run, typically with $n = 4$ iterations. When ensembling to improve precision and bootstrapping to estimate uncertainties, this quickly adds up to tens of thousands of networks that must be trained. In Fig. 1, we show the validation loss obtained in an OMNIFOLD step using OMNILEARN and starting the classification task from scratch. OMNILEARN is able to converge twice as fast and reach a lower minimum comparing to training from scratch. The overall speed up time from OMNILEARN is also close to a factor of 2 since the most time-consuming iteration is the first one with consecutive iterations starting from the previous iteration checkpoint and requiring few epochs for convergence. This lower validation loss directly translates into better performance on physics metrics, calculated as the triangular distance between different physics observables introduced in [45] and summarized in Table II. The unfolding is performed at the level of all particle momenta and yet OMNILEARN outperforms the classical OMNIFOLD using all particles with a DEEPSSETS architecture, which outperforms the classical bin-based, dedicated unfolding (IBU [51]).

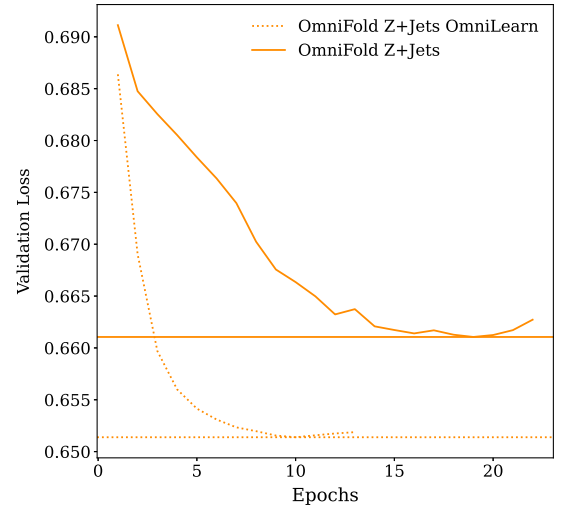


FIG. 1. Validation loss curves obtained in the first iteration and first step of the OMNIFOLD dataset. The OMNILEARN validation loss is compared with the PET classifier trained from scratch.

Like OMNIFOLD, many machine learning-based anomaly detection protocols require training directly on data. One particularly powerful approach with asymptotic guarantees of optimality [55] is to train a classifier to distinguish data from a background-only reference dataset [56]. If the reference is accurate and the classifier is powerful, any nontrivial performance could be an indication of something new. Approaches largely differ in how they generate the reference dataset. In the case of resonant anomalies, sideband information can be used to build the reference. A state-of-the-art approach is CATHODE [57], which trains a conditional generative model in the sideband, interpolates the model to the signal region, samples a reference dataset, and then trains a classifier to distinguish the synthetic data from the real data in the signal region. While originally developed for high-level features, CATHODE was recently studied for processing entire jets [30]. While it was able to discover signals injected into a background-only dataset,

TABLE II. Comparison of the triangular discriminator [52–54] between different algorithms for unfolding. Uncertainties from PET and OMNILEARN are taken from 100 histogram variations within the statistical uncertainty of the prediction. Quantities in bold represent the method with best performance. IBU and OMNIFOLD results with DEEPSSETS is taken from [45].

Metric	IBU	OMNIFOLD		
		DEEPSSETS	PET classifier	OMNILEARN
Jet mass	9.31	2.77	2.8 ± 0.9	2.6 ± 0.8
N	1.51	0.33	0.50 ± 0.15	0.34 ± 0.1
Jet width	0.11	0.10	0.09 ± 0.02	0.07 ± 0.01
$\log \rho$	0.71	0.35	0.23 ± 0.07	0.14 ± 0.03
τ_{21}	1.10	0.53	0.13 ± 0.03	0.05 ± 0.01
z_g	0.37	0.68	0.19 ± 0.03	0.21 ± 0.04

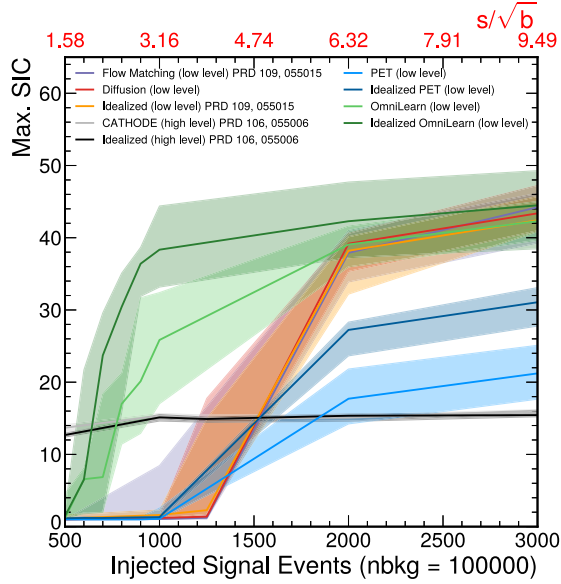


FIG. 2. Maximum values of the SIC curve evaluated over different values of injected signal. OMNILEARN and PET classifier results are compared with other algorithms used for the same task.

the amount of signal had to be so high that it would have also been discoverable without doing any machine learning.

We reexamine the high-dimensional CATHODE approach using OMNILEARN. For this purpose, we use the R&D dataset from the LHC Olympics data challenge [4,58]. The background consists of generic dijet events, while the signal is a resonant boson production $A \rightarrow B(\rightarrow qq')C(\rightarrow qq')$ with masses $m_A, m_B, m_C = 3.5, 0.5, 0.1$ TeV, respectively. Signal and background events are generated with PYTHIA8 interfaced with DELPHES3.4.1 for detector simulation. Jets are defined using the anti- k_T algorithm as implemented in FASTJET with $R = 1$. We focus on the two leading jets in transverse momentum space and require the leading jet to have $p_T > 1.2$ TeV. After selection, we save all particles associated to the two most energetic jets, resulting in a maximum particle multiplicity of 279 particles per jet. We need to use OMNILEARN twice: once to generate the background and once to distinguish the resulting synthetic samples from the data. Since there are two jets, we use a DEEPSETS [59] to mix the two jets into a full event model [29].

The performance of OMNILEARN on the high-dimensional anomaly detection task is shown in Fig. 2. Our figure of merit is the maximum significance improvement characteristic curve (SIC). This is the multiplicative factor by which the significance improves with an ideal selection on the classifier. Values above unity indicate value added and when multiplied by the upper, horizontal axis, the combined number indicates the final significance. The results are presented for OMNILEARN, PET, the simpler model of Ref. [30], and an idealized scenario where the background is a statistically identical sample of true background events in the signal region (no learning). OMNILEARN shows

non-negligible signal sensitivity for signals injections above 600 events, corresponding to an initial significance $S/\sqrt{B} \sim 2$, representing a large improvement in sensitivity compared to previous results that required signal injections above 1400 ($S/\sqrt{B} \sim 4$). Conversely, training the model from scratch shows lower performance compared to results reported in [30]. The reason for this difference is due to the limited amount of data in the signal region (around 100k). Even when using a larger generated background of 350k events, the dataset size of this application is at least two times smaller than all previous datasets investigated so far.

In summary, we have demonstrated that our pretrained transformer approach [29] can be considered as a foundation model for jet physics and is ready to join the toolkit of particle physicists. This is demonstrated using the OMNILEARN foundation model [29] in the context of three potentially transformative applications:

- (1) For tasks with *expensive simulations*, we have shown that OMNILEARN can effectively amplify the training statistics to improve the performance. For example, on a full detector simulation dataset for top tagging from ATLAS [34], we show that we can achieve (or exceed) state-of-the-art performance with only 10% of the training dataset. This could lead to significant computational savings for the experiments when developing new taggers. It opens up the possibility of building analysis-specific taggers without generating large training datasets for every analysis.
- (2) For full phase space *unfolding*, OMNILEARN is not only more precise, but also much faster, converging in about half the time as a dedicated approach. This is critical for practical applications of unfolding, where ensembling and statistical uncertainties require the training of a computationally expensive number of models [11–15,17]. By improving the accuracy, it also will enhance the diversity of downstream analysis tasks that can benefit from high-dimensional, unbinned data.
- (3) For *anomaly detection*, we have shown for the first time that full phase space methods are capable of nontrivial discoveries. Previous work in resonant anomaly detection [30] had shown that full phase-space methods could significantly amplify injected signals, but only if their starting significance was well above 2. With OMNILEARN, this is pushed down to 2.

We are excited to explore (many) other applications of such approaches; all of the code and data to reproduce these results are also public to facilitate other researchers to do the same. We also envision a number of extensions, including the addition of even more training data for the foundation model. The way we train machine learning models has forever changed; instead of starting from scratch, it will be more effective to start with a foundation

model. If the data are sufficiently different that OMNILEARN is not applicable, then the tools used to train it can be readily adapted to create a library of foundation models to enable all tasks in particle physics and beyond.

Acknowledgments. We thank F. Dreyer, Ming Fong, R. Grabarczyk, K. Greif, P. F. Monni, and D. Whiteson for interesting discussions related to transfer learning. We also thank J. Birk, A. Hallin, and G. Kasieczka for comments on the manuscript. Additionally, we thank our colleagues from the H1 Collaboration for allowing us to use the simulated MC event samples. We also thank DESY-IT and the MPI für Physik for providing computing infrastructure and

supporting the data preservation project of the HERA experiments. V. M. and B. N. are supported by the U.S. Department of Energy (DOE), Office of Science under Contract No. DE-AC02-05CH11231. This research used resources of the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231 using NERSC Award No. HEP-ERCAP0021099 and No. HEP-ERCAP0028249.

Data availability. The code for this paper can be found at [60].

-
- [1] K. Cranmer, J. Brehmer, and G. Louppe, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 30055 (2020).
 - [2] M. Arratia *et al.*, *J. Instrum.* **17**, P01024 (2021).
 - [3] H. Du, C. Krause, V. Mikuni, B. Nachman, I. Pang, and D. Shih, [arXiv:2404.18992](https://arxiv.org/abs/2404.18992).
 - [4] G. Kasieczka *et al.*, *Rep. Prog. Phys.* **84**, 124201 (2021).
 - [5] C. Balázs *et al.* (DarkMachines High Dimensional Sampling Group), *J. High Energy Phys.* **05** (2021) 108.
 - [6] G. Karagiorgi, G. Kasieczka, S. Kravitz, B. Nachman, and D. Shih, [arXiv:2112.03769](https://arxiv.org/abs/2112.03769).
 - [7] ATLAS Collaboration, *Phys. Rev. Lett.* **125**, 131801 (2020).
 - [8] G. Aad *et al.* (ATLAS Collaboration), *Phys. Rev. Lett.* **132**, 081801 (2024).
 - [9] ATLAS Collaboration, *Phys. Rev. D* **108**, 052009 (2023).
 - [10] CMS Collaboration, Model-agnostic search for dijet resonances with anomalous jet substructure in proton-proton collisions at $\sqrt{s} = 13$ TeV, Technical Report No. CMS-PAS-EXO-22-026, CERN, Geneva, 2024, <https://cds.cern.ch/record/2892677>.
 - [11] H1 Collaboration, *Phys. Rev. Lett.* **128**, 132002 (2021).
 - [12] H1 Collaboration, H1prelim-22-031 (2022), <https://www-h1.desy.de/h1/www/publications/htmlsplit/H1prelim-22-031.long.html>.
 - [13] H1 Collaboration, *Phys. Lett. B* **844**, 138101 (2023).
 - [14] H1 Collaboration, H1prelim-23-031 (2023), <https://www-h1.desy.de/h1/www/publications/htmlsplit/H1prelim-23-031.long.html>.
 - [15] LHCb Collaboration, *Phys. Rev. D* **108**, L031103 (2023).
 - [16] P. T. Komiske, S. Kryhin, and J. Thaler, *Phys. Rev. D* **106**, 094021 (2022).
 - [17] Y. Song (STAR Collaboration), [arXiv:2307.07718](https://arxiv.org/abs/2307.07718).
 - [18] J. Batson and Y. Kahn, [arXiv:2312.02264](https://arxiv.org/abs/2312.02264).
 - [19] B. M. Dillon, G. Kasieczka, H. Olschlager, T. Plehn, P. Sorrenson, and L. Vogel, *SciPost Phys.* **12**, 188 (2021).
 - [20] B. M. Dillon, R. Mastandrea, and B. Nachman, *Phys. Rev. D* **106**, 056005 (2022).
 - [21] B. M. Dillon, L. Favaro, F. Feiden, T. Modak, and T. Plehn, *SciPost Phys. Core* **7**, 056 (2024).
 - [22] L. Heinrich, T. Golling, M. Kagan, S. Klein, M. Leigh, M. Osadchy, and J. A. Raine, *Mach. Learn. Sci. Tech.* **5**, 035074 (2024).
 - [23] P. Harris, M. Kagan, J. Krupa, B. Maier, and N. Woodward, [arXiv:2403.07066](https://arxiv.org/abs/2403.07066).
 - [24] B. Hashemi, Deep generative models for ultra-high granularity particle physics detector simulation: A voyage from emulation to extrapolation, Other thesis, 2024, [arXiv:2403.13825](https://arxiv.org/abs/2403.13825).
 - [25] M. P. Kuchera, R. Ramanujan, J. Z. Taylor, R. R. Strauss, D. Bazin, J. Bradt, and R. Chen, *Nucl. Instrum. Methods Phys. Res., Sect. A* **940**, 156 (2019).
 - [26] A. Chappell and L. H. Whitehead, *Eur. Phys. J. C* **82**, 1099 (2022).
 - [27] F. A. Dreyer, R. Grabarczyk, and P. F. Monni, *Eur. Phys. J. C* **82**, 564 (2022).
 - [28] H. Beauchesne, Z.-E. Chen, and C.-W. Chiang, *J. High Energy Phys.* **02** (2024) 138.
 - [29] Simultaneous PRD Submission.
 - [30] E. Buhmann, C. Ewen, G. Kasieczka, V. Mikuni, B. Nachman, and D. Shih, *Phys. Rev. D* **109**, 055015 (2024).
 - [31] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, CoRR abs/1706.03762 (2017).
 - [32] Y. Wang, Y. Sun, Z. Liu, S. E. Sarma, M. M. Bronstein, and J. M. Solomon, CoRR abs/1801.07829 (2018).
 - [33] H. Qu, C. Li, and S. Qian, [arXiv:2202.03772](https://arxiv.org/abs/2202.03772).
 - [34] ATLAS Collaboration, Constituent-based top-quark tagging with the ATLAS detector, Technical Report No. ATL-PHYS-PUB-2022-039, CERN, Geneva, 2022, <https://cds.cern.ch/record/2825328>.
 - [35] R. D. Ball *et al.*, *Nucl. Phys.* **B867**, 244 (2013).
 - [36] A. Buckley, in *Proceedings of the 6th International Workshop on Multiple Partonic Interactions at the LHC* (2014), p. 29.
 - [37] G. Aad *et al.* (ATLAS Collaboration), *Eur. Phys. J. C* **81**, 334 (2021).
 - [38] M. Cacciari and G. P. Salam, *Phys. Lett. B* **641**, 57 (2006).

- [39] M. Cacciari, G. P. Salam, and G. Soyez, *Eur. Phys. J. C* **72**, 1896 (2012).
- [40] M. Cacciari, G. P. Salam, and G. Soyez, *J. High Energy Phys.* **04** (2008) 063.
- [41] P. Berta, M. Spousta, D. W. Miller, and R. Leitner, *J. High Energy Phys.* **06** (2014) 092.
- [42] P. Berta, L. Masetti, D. W. Miller, and M. Spousta, *J. High Energy Phys.* **08** (2019) 175.
- [43] M. Cacciari, G. P. Salam, and G. Soyez, *Eur. Phys. J. C* **75**, 59 (2015).
- [44] A. J. Larkoski, S. Marzani, G. Soyez, and J. Thaler, *J. High Energy Phys.* **05** (2014) 146.
- [45] A. Andreassen, P. T. Komiske, E. M. Metodiev, B. Nachman, and J. Thaler, *Phys. Rev. Lett.* **124**, 182001 (2020).
- [46] A. Andreassen, P. T. Komiske, E. M. Metodiev, B. Nachman, A. Suresh, and J. Thaler, *arXiv:2105.04448*.
- [47] M. Bahr *et al.*, *Eur. Phys. J. C* **58**, 639 (2008).
- [48] J. Bellm *et al.*, *Eur. Phys. J. C* **76**, 196 (2016).
- [49] J. Bellm *et al.*, *arXiv:1705.06919*.
- [50] ATLAS Collaboration (2014), <https://cds.cern.ch/record/1966419>.
- [51] G. D’Agostini, *Nucl. Instrum. Methods Phys. Res., Sect. A* **362**, 487 (1995).
- [52] F. Topsoe, *IEEE Trans. Inf. Theory* **46**, 1602 (2000).
- [53] P. Gras, S. Höche, D. Kar, A. Larkoski, L. Lönnblad, S. Plätzer, A. Siódmok, P. Skands, G. Soyez, and J. Thaler, *J. High Energy Phys.* **07** (2017) 091.
- [54] S. Bright-Thonney and B. Nachman, *J. High Energy Phys.* **03** (2019) 098.
- [55] E. M. Metodiev, B. Nachman, and J. Thaler, *J. High Energy Phys.* **10** (2017) 174.
- [56] J. H. Collins, K. Howe, and B. Nachman, *Phys. Rev. D* **99**, 014038 (2019).
- [57] A. Hallin, J. Isaacson, G. Kasieczka, C. Krause, B. Nachman, T. Quadfasel, M. Schlaffer, D. Shih, and M. Sommerhalder, *Phys. Rev. D* **106**, 055006 (2021).
- [58] G. Kasieczka, B. Nachman, and D. Shih, Official Datasets for LHC Olympics 2020 Anomaly Detection Challenge (Version v6) [Data set], [10.5281/zenodo.4536624](https://zenodo.org/record/4536624) (2019).
- [59] M. Zaheer, S. Kottur, S. Ravanbakhsh, B. Póczos, R. Salakhutdinov, and A. Smola, *arXiv:1703.06114*.
- [60] V. Mikuni and P. Hebbar, ViniciusMikuni/OmniLearn: Zenodo entry (1.1) (Zenodo, 2025), [10.5281/zenodo.14977181](https://zenodo.org/record/14977181).