

# Differentiable Matrix Elements with MadJax

**Lukas Heinrich**

Technical University of Munich

E-mail: [lukas.heinrich@cern.ch](mailto:lukas.heinrich@cern.ch)

**Michael Kagan**

SLAC National Accelerator Laboratory

E-mail: [makagan@slac.stanford.edu](mailto:makagan@slac.stanford.edu)

**Abstract.** *MadJax* is a tool for generating and evaluating differentiable matrix elements of high energy scattering processes. As such, it is a step towards a differentiable programming paradigm in high energy physics that facilitates the incorporation of high energy physics domain knowledge, encoded in simulation software, into gradient based learning and optimization pipelines. *MadJax* comprises two components: (a) a plugin to the general purpose matrix element generator *MadGraph* that integrates matrix element and phase space sampling code with the *JAX* differentiable programming framework, and (b) a standalone wrapping code interface for accessing the matrix element code and its gradients, which are computed with automatic differentiation. The *MadJax* implementation and example applications of simulation based inference and normalizing flow based matrix element modeling, with capabilities enabled uniquely with differentiable matrix elements, are presented.

## 1. Introduction

Matrix Element (ME) Monte Carlo generators provide a theoretical description of differential cross sections for high-energy particle scattering processes encapsulated in software. Theoretical calculations are encoded in stochastic simulators to allow for Monte Carlo integration and sampling of scattered particle configurations. Together with parton-shower simulations of QCD radiation processes, and models of hadronization, these programs can be used to sample particle interactions and final state particle kinematic configurations for the experimental analysis of collider data, such as data from the LHC or from future colliders. However, while collider physics data analysis typically relies on large simulated datasets for statistical inference, simulators do not retain event-by-event information on the dependence of simulated predictions on input parameters or variables. Ultimately, this limits the ways in which researchers can use these simulators.

With *MadJax*, access to the dependence of simulated predictions on inputs is enabled through gradients. *MadJax* is a differentiable programming (DP) [1] approach to the simulation of matrix elements, which combines the python-based code generation of the *MadGraph* matrix element generator [2, 3] with the *JAX* DP framework [4]. In DP, software is written in, or transformed into, differentiable code via the use of automatic differentiation (AD) [5], an algorithmic way to efficiently evaluate derivatives of computer programs. This approach is flexible and optimizable; differentiable HEP software and machine learning (ML) tools can be mixed, for instance to use ML surrogates of non-differentiable computations, and can be jointly optimized to improve speed



and prediction accuracy. HEP simulation tools, and the knowledge they encode, can be used as physics prediction engines directly within ML pipelines for developing physics-informed ML tools.

**MadJax** enables the evaluation of MEs and their gradients automatically. **MadJax** also includes a differentiable implementation of the RAMBO algorithm [6], which facilitates transforming particle configurations into *phase space* points on a hypercube for aiding sampling algorithms. An example of the  $e^+e^- \rightarrow h \rightarrow ZZ \rightarrow 4l$  matrix element and **MadJax**-computed gradients with respect to the  $Z$  boson mass is seen in Fig. 1. As expected, the **MadJax**-computed gradients accurately estimate tangents to the ME curve. To the best of our knowledge, **MadJax** is the first differentiable matrix element generator.

Related work on AD / DP in physics are discussed below. Sec. 2 presents the implementation, usage, and performance of **MadJax**. Sec. 3 presents applications of **MadJax** in case studies of simulation-based inference and ME modeling with ML.

### 1.1. Related Work

Automatic differentiation [5, 7], and its use in gradient based optimization, is ubiquitous in ML, statistics, and applied math. AD uses the chain rule to evaluate derivatives of a function that is represented as a computer program. AD takes as input program code and produces new code for evaluating the program and derivatives. AD typically builds a computational graph, or a directed acyclic graph of mathematical operations applied to an input. Gradients are defined for each operation of the graph, and the total gradient can be evaluated from input to output, called forward mode, and from output to input, called reverse mode or backpropagation in ML. AD is backbone of ML / DP frameworks like **TensorFlow** [8], **JAX** [4], and **PyTorch** [9].

In HEP, AD has been used in histogram fitting in **pyHF** [10, 11], and **Neos** [12] builds on **pyHF** to enable joint histogram fitting and analysis optimization. For modeling parton distribution functions (pdf) used by matrix element generators, Ref. [13] implemented the NNPDF software in **TensorFlow** to enable fast evaluation on GPUs. More recently, fitting of neural network pdf models implemented in **TensorFlow** has been explored by the NNPDF collaboration [14]. Concurrent with our work, **MadFlow** [15] has explored using a **TensorFlow** backend for **MadGraph** for GPU acceleration enabled but did not explore AD capabilities.

Outside of HEP, the development of differentiable physics simulators is an active area of research in ML and scientific computing. The growing body of examples includes: Rigid body dynamics [16, 17, 18, 19, 20, 21] and soft-body [22, 23, 24] simulators have been used in reinforcement learning, robotics, optimization, and control problems. Differentiable light transport and Monte Carlo ray tracing is used in rendering and graphics [25, 26, 27] and neural rendering such as NeRF [28]. Differentiable simulation of fluid dynamics, for prediction, control, and optimizing solutions of partial differential equations is an active research area [29, 30, 31], as are differentiable molecular dynamics simulations [32, 33] for computational condensed matter physics.

## 2. Implementation

**MadGraph** is a tool for the computation of matrix elements and event generation at the parton level and is widely used for the description of LHC physics. As a general purpose tool, **MadGraph** receives as input a given physics transition, i.e. the desired input and output particles, and dynamically

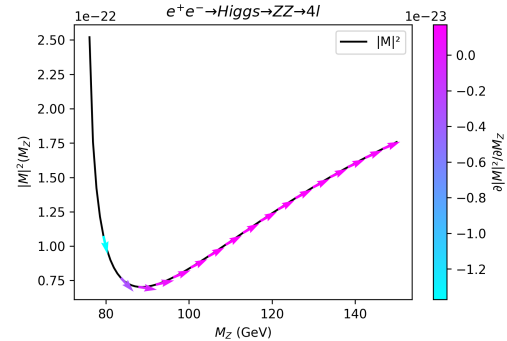


Figure 1: Matrix element  $\mathcal{M}(r|M_Z)$  for the process  $e^+e^- \rightarrow H \rightarrow ZZ^{(*)} \rightarrow ee\mu\mu$  and its gradients evaluated by **MadJax** at the phase space point  $r = [0.2]^8$

1. Generation:

```
generate p p > t t~, t > b udsc udscx , t~ > b~ udsc udscx
output madjax generated_ttbar
set auto_update 0
```

2. Evaluation:

```
import madjax
mj = madjax.MadJax('generated_ttbar')
E_cm = 14000 #GeV
process = 'Matrix_1_gg_ttx_t_budx_tx_bxdux'
matrix_element = mj.matrix_element(E_cm,process)

parameters = {'mass',6): 173.0} #set top mass
phasespace_coords = [0.1]*14 #14D phasespace

val, grad = matrix_element(parameters,phasespace_coords)
grad[('mass', 6)] #gradient wrt top mass
```

Figure 2: Example Usage of the MadJax code generation and evaluation API

compiles a list of hard processes for which matrix elements  $\mathcal{M}(x)$  need to be computed. It then generates code for matrix elements in several possible target languages. The most commonly used ones are C++ and FORTRAN due to their high evaluation performance. A Python output option is also available, but its performance as a dynamically interpreted language is much reduced. The MEs are themselves calculated based on helicity amplitudes as generated by ALOHA [34].

In MadJax, integration with JAX is performed during MadGraph automated matrix element code generation. To do so, several challenges had to be addressed, including: (a) control flow, such as conditional and loop statements, required adaptation to allowable control flow paradigms for the LAX backend of JAX, (b) some components required adaptation to the functional programming paradigm of JAX, and (c) some numerical methods required custom derivative implementation, such as implicit differentiation for root-finding. Code vectorization with JAX's `vmap{...}` functionality was used to increase computational speed and efficiency. For example, loop statements, such as loops over helicity states, were vectorized where possible. Multi-event phase space sampling, and ME and ME gradient evaluation, were enabled with vectorization. MadJax inherits both CPU and GPU compatibility from JAX. However, many of the computations in MadJax generated code do not have optimized GPU kernels and GPU evaluation did not result in significant speed-up.

### 2.1. Usage

MadJax offers an easy-to-use API and transparently connects to the plugin mechanism of MadGraph. It comprises two separate sub-packages: (1) a standard MadGraph plugin to generate matrix element code in the JAX language and must be run within the MadGraph environment, and (2) a standalone wrapping API around the generated code which can be used independently from MadGraph. The first stage is achieved by adding the MadJax plugin to the PLUGINS directory of MadGraph and requesting the MadJax output in the run card via `mg5_aMC --mode=madjax_me_gen run.mg5`. This step produces a directory with generated, differentiable MadJax code, which is then accessible through the MadJax python API. An example listing is shown in Figure 2.

The ME function allows the user to explicitly pass parameter values through a mapping from SLHA block identifiers to parameter values (here: top quark mass). By default the MadJax returns gradients with respect to the parameters, but gradients with respect to the phase-space point can

Evaluation Mode	Performance	Output
JAX no-JIT	5.1s / evaluation	value and gradient
JAX JIT	15.9s / 100k evaluations	value and gradient
JAX JIT + <code>vmap</code>	5.6s / 100k evaluations	value and gradient
Standalone C++	0.8s / 100k evaluations	value only

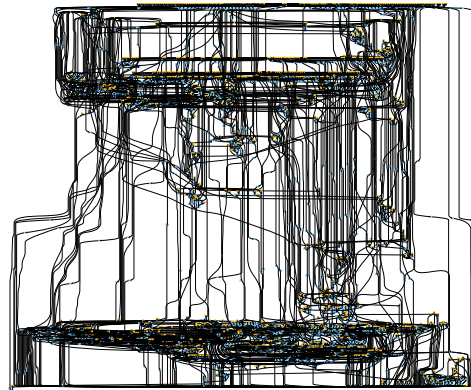
Table 1: Performance Evaluations for a top quark decay example process

be requested as well. The phase space point is identified through a set of hypercube coordinates.

## 2.2. Performance

While the implementation of `MadJax` makes heavy use of the `Python` export option, its performance is significantly better due to the use of built in just-in-time (JIT) compilation features of `JAX`. In Table 1 we report performance results for the gluon-initiated  $t\bar{t}$  production matrix element (`Matrix_1_gg_ttx_t_budx_tx_bxdx`), in which a computation of the ME value and its gradients via reverse-mode differentiation stays within an order of magnitude of the standalone C++ code the `MadGraph` package produces. Here, the JIT compilation and vectorization (`vmap`) capabilities of `JAX` provide a over 10,000x speedup over the non-compiled `JAX` code.

Several challenges were observed when using `JAX` for AD integration. Most notably, the `MadGraph` generated code can contain control flow such as branching conditionals and loop statements which are known to cause inefficiencies during compute graph generation in `JAX` [35]. This can be seen in Fig. 3 where `JAX` generates a massive and complex compute graph even for the few hundred lines of code describing the  $e^+e^- \rightarrow t\bar{t} \rightarrow q\bar{q}bq\bar{q}$  process. The python code generated by `MadGraph` is a near direct translation from the original FORTRAN code, and is not well adapted for JIT compile time code optimization in `JAX`. As a result, even the few hundred lines of code generated by `MadJax` for this process could take 10min or longer to compile.

Figure 3: Compute graph generated by `MadJax` for the  $e^+e^- \rightarrow t\bar{t} \rightarrow 6q$  process

## 3. Applications

Two case studies are presented below that show how the use of ME gradients can improve HEP data analysis techniques. As MEs evaluate the differential cross section of a given particle configuration, they can be interpreted as un-normalized<sup>1</sup> probability density functions  $\tilde{p}(x)$ . With `MadJax`, the gradients  $\nabla_x \tilde{p}(x)$  can be computed, as can gradients of RAMBO transformations.

### 3.1. Mining Gold with Differentiable Simulators

An important use-case for automatic gradients of MEs as provided by `MadJax` is their use in simulation-based inference [36] (SBI) techniques that require knowledge of score  $\nabla_\theta \log p_{\text{ME}}(z|\theta)$  variables, such as the ‘Mining Gold’ technique [37] for likelihood ratio (LR) estimation. Here  $r(x, \theta_0, \theta_1) = p(x|\theta_0)/p(x|\theta_1)$  is approximated by a deep neural network  $\hat{r}_\phi(x, \theta_0, \theta_1)$  parametrized by weights  $\phi$  whose values are optimized through regression. While the true likelihood ratio  $r(x, \theta_0, \theta_1)$  is intractable, which precludes the use of standard regression objectives such as mean squared errors to the true value, it can be shown that regression to the joint likelihood ratio

<sup>1</sup> The total cross section would be the normalizing constant

$r(x, z|\theta_0, \theta_1)$  yields a consistent estimator in the large-data limit. Here the joint LR acts as a “noisy label” which is tractable to compute. This regression on the LR can be dramatically improved by not only utilizing the joint LR, but also its gradient  $t(z, x|\theta_0, \theta_1) = \nabla_{\theta} p(z, x|\theta)/p(z, x|\theta_1)|_{\theta_0}$ , where the use of the joint values as labels yields an additional consistent training objective for the gradient  $\nabla_{\theta_0} r(x, \theta_0, \theta_1)$ . While this technique can work around the intractability of the marginalized likelihood ratio, it still requires computation of the joint quantities, in particular the gradients of the joint probabilistic model  $p(x, z)$ . Without an automatic solution, this approach is thus limited to cases where gradients are derivable in close-form. **MadJax**, can extend the method to arbitrary theory parameters. As a proof-of-concept, Fig. 4 demonstrates a simplified example with the Fermi-constant as the parameter of interest, where the power of gradient-based training data is apparent from the training trajectory. The left figures show a comparison of the histogram-based and learned likelihood ratios, including the samples from the joint ratio and score used for the training. Notably, the score-based training for theory parameters such as  $G_F$  is only possible for the first time with **MadJax**, due its ability to compute gradients with respect to any theory parameter. The middle figure shows the learned likelihood ratio, and the joint ratios used for training, as a function of both phase space and parameter points. The right panel shows the training loss versus training step, wherein the significantly increased efficacy of score-based training approaches enabled uniquely by **MadJax** is clearly seen over training without such gradient information.

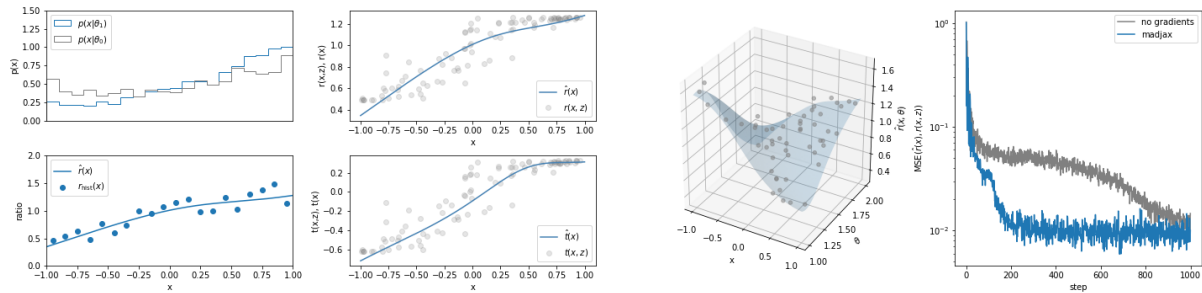


Figure 4: SBI example using **MadJax** for learning ML based likelihood ratios as a function of the theory parameter  $G_F$  in  $e^+e^- \rightarrow Z \rightarrow \mu^+\mu^-$ . Left: histogram-based density estimation of the marginal observable  $x$  (polar angle) and below the learned LR validated against the histogram approach. Middle: the trained marginal LR and its gradient compared to the joint quantities used as noisy label data shown on a 1D slice and in the 2D space of observables and theory parameters. Right: Loss trajectories of LR regression with and without **MadJax** gradient data.

### 3.2. Normalizing Flow Training using Matrix Element Gradients

Normalizing flows [38] (NF), which enable both explicit likelihood estimation and sampling, provide a powerful method for modeling complex MEs in order to increase sampling speed and efficiency (see e.g. [39, 40, 41]). It can be challenging, however, to train NFs to sufficient accuracy, and training can require relatively large datasets sampled from the simulator. ME gradients provide new tools to help reduce the required scale of simulated datasets and improve modeling.

The typical maximum likelihood loss function for NF training has the form:

$$\mathcal{L}_{NLL} = -\mathbb{E}_{x \sim p(x)} \left[ \log q(x; \theta) \right] = D_{KL} \left[ p || q(\cdot; \theta) \right] + \text{const.} \quad (1)$$

where samples are generated by the ME simulator,  $x \sim p(x)$ , and the likelihood under the NF model,  $q(x; \theta)$ , is evaluated and minimized with respect to its parameters  $\theta$ . The Kullback–Leibler (KL) divergence is defined as  $D_{KL}[p||q] \equiv \mathbb{E}_p[\log p - \log q]$ . This loss function has two challenges;

(a) one must generate samples from the simulator for training, which may be computationally costly, and (b) the loss pushes the NF to be *mass-distributing*, i.e. encouraging the NF to assign mass broadly so as not to assign zero density to points in which  $p(x) \neq 0$ . With access to the gradient of the (un-normalized) target density, additional loss terms are available, such as the *reverse KL* loss,  $\mathcal{L}_{RKL}$ , and a *force-matching loss*,  $\mathcal{L}_{FM}$ , which have shown promise in lattice QCD (i.e. [42, 43]) and molecular physics (i.e. [32, 44]). The losses are defined as:

$$\mathcal{L}_{RKL} = \mathbb{E}_{x \sim q(x; \theta)} \left[ \log \frac{q(x; \theta)}{p(x)} \right] = D_{KL} [q(x; \theta) || p(x)] \quad (2)$$

$$\mathcal{L}_{FM} = \mathbb{E}_{x \sim p(x)} \left[ \|\partial_x \log p(x) - \partial_x \log q(x; \theta)\|_2^2 \right] \quad (3)$$

The reverse KL loss leverages much cheaper samples from the NF and exhibits *mode-seeking* behavior; training under this loss pushes the NF to assign zero density where  $p(x)$  is zero and hence makes it concentrate on one of the modes of  $p(x)$ . The force matching pushes the NF model during training to better match the shape (as estimated with gradients) of the target ME distribution. Gradient-based training with both  $\mathcal{L}_{RKL}$  and  $\mathcal{L}_{FM}$  requires differentiation through the matrix element  $p(x)$ , and thus requires differentiable matrix elements as provided by **MadJax**. By mixing the forward KL with the reverse KL and/or force-matching losses, additional information about the ME is provided to the NF during training.

Using **MadJax**, we train multiple NF models of  $e^+e^- \rightarrow t\bar{t} \rightarrow q\bar{q}bq\bar{q}\bar{b}$  matrix elements. We use the loss  $\mathcal{L} = \mathcal{L}_{NLL} + \omega_k \mathcal{L}_{RKL} + \omega_f \mathcal{L}_{FM}$ , with hyperparameters  $\omega_k$  and  $\omega_f$  controlling the impact of the reverse KL and force matching losses, respectively. The NF models comprise four layers of smooth normalizing flows [45], which provide smooth mappings over compact intervals and thus are well adapted to mapping uniform random variables smoothly onto ME distributions defined on phase space hypercubes. Each smooth NF layer uses four bijections and the parameters of the transformation are modeled with a MLP of two layers of 100-neurons. 8000 events are used for training using the Adam [46] optimizer with a learning rate of  $5 \cdot 10^{-5}$  and batch size of 100, and 2000 events are used for testing. 10 initial warm-up epochs, where all models are trained using only the  $\mathcal{L}_{NLL}$ , are used to improve training stability. Fig. 5 shows the evolution (after warm-up epochs) of the likelihood loss (left) and reverse KL loss (right), comparing models trained with and without reverse KL and force matching losses. Variations in initial model performance due to random different model initializations are overcome after epoch  $\sim 75$  and then models trained with  $\mathcal{L}_{RKL}$  are seen to outperform the likelihood-only trained model ( $\omega_k = \omega_f = 0$ ) in both likelihood and reverse KL. The model trained with both reverse KL and force-matching losses resulted in the best likelihood of all models. These results indicate the potential to improve ME modeling using losses enabled uniquely with ME gradients.

#### 4. Conclusion

We present **MadJax**, the first automatically differentiable matrix element code based on **MadGraph**. The gradient information provided by **MadJax** enables previously inaccessible approaches to HEP data analysis, such as highly data-efficient methods for training ML models for high-energy physics or simulation-based inference. **MadJax** is a prototype that serves as proof-of-principle of the utility of such tools for future R&D. The implementation of **MadJax** using **JAX** facilitates easy use with ML pipelines, but **JAX** is not always well-adapted to the structure of matrix element code. Future directions will explore the use of alternative AD tools that may be better suited to matrix element code structures.

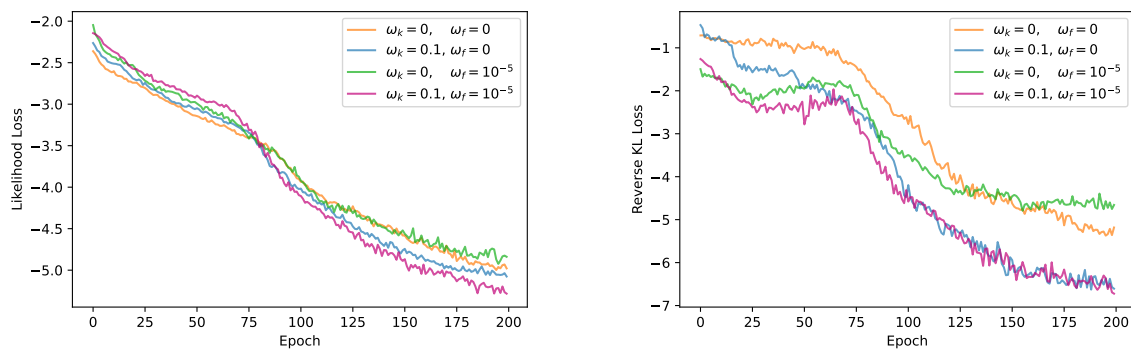


Figure 5: Comparison of training evolution of the likelihood loss (left) and reverse KL loss (right) for models trained with and without reverse KL and force matching losses.

### Acknowledgements

We thank Valentin Hirschi for providing the initial python-MadGraph interface code and thank Matthew Feickert for the software guidance, and thank both for the helpful discussions. We thank Andreas Krämer and Jonas Koehler for supplying their code for Smooth Normalizing Flows. MK is supported by the US Department of Energy (DOE) under grant DE-AC02-76SF00515. LH is supported by the Excellence Cluster ORIGINS, which is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC-2094-390783311.

### References

- [1] Wang F, Decker J, Wu X, Essertel G and Rompf T 2018 *Advances in Neural Information Processing Systems* vol 31 ed Bengio S, Wallach H, Larochelle H, Grauman K, Cesa-Bianchi N and Garnett R (Curran Associates, Inc.)
- [2] Alwall J, Herquet M, Maltoni F, Mattelaer O and Stelzer T 2011 *Journal of High Energy Physics* **2011** ISSN 1029-8479 URL [http://dx.doi.org/10.1007/JHEP06\(2011\)128](http://dx.doi.org/10.1007/JHEP06(2011)128)
- [3] Alwall J, Frederix R, Frixione S, Hirschi V, Maltoni F, Mattelaer O, Shao H S, Stelzer T, Torrielli P and Zaro M 2014 *Journal of High Energy Physics* **2014** 79
- [4] Bradbury J, Frostig R, Hawkins P, Johnson M J, Leary C, Maclaurin D, Necula G, Paszke A, VanderPlas J, Wanderman-Milne S and Zhang Q 2018 JAX: composable transformations of Python+NumPy programs URL <http://github.com/google/jax>
- [5] Bücker H M, Corliss G F, Hovland P D, Naumann U and Norris B (eds) 2005 *Automatic Differentiation: Applications, Theory, and Implementations* Lecture Notes in Computational Science and Engineering (New York, NY: Springer)
- [6] Platzer S 2013 Rambo on diet (*Preprint 1308.2922*)
- [7] Baydin A G, Pearlmutter B A, Radul A A and Siskind J M 2018 *Journal of Machine Learning Research* **18** 1–43 URL <http://jmlr.org/papers/v18/17-468.html>
- [8] Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, Corrado G S, Davis A, Dean J, Devin M, Ghemawat S, Goodfellow I, Harp A, Irving G, Isard M, Jia Y, Jozefowicz R, Kaiser L, Kudlur M, Levenberg J, Mané D, Monga R, Moore S, Murray D, Olah C, Schuster M, Shlens J, Steiner B, Sutskever I, Talwar K, Tucker P, Vanhoucke V, Vasudevan V, Viégas F, Vinyals O, Warden P, Wattenberg M, Wicke M, Yu Y and Zheng X 2015 TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems software available from tensorflow.org URL <https://www.tensorflow.org/>
- [9] Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, Killeen T, Lin Z, Gimelshein N, Antiga L, Desmaison A, Kopf A, Yang E, DeVito Z, Raison M, Tejani A, Chilamkurthy S, Steiner B, Fang L, Bai J and Chintala S 2019 *Advances in Neural Information Processing Systems 32* ed Wallach H, Larochelle H, Beygelzimer A, d'Alché-Buc F, Fox E and Garnett R (Curran Associates, Inc.) pp 8024–8035 URL <http://papers.nips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf>
- [10] Heinrich L, Feickert M and Stark G pyhf: v0.6.3 <https://github.com/scikit-hep/pyhf/releases/tag/v0.6.3> URL <https://doi.org/10.5281/zenodo.1169739>
- [11] Heinrich L, Feickert M, Stark G and Cranmer K 2021 *Journal of Open Source Software* **6** 2823 URL <https://doi.org/10.21105/joss.02823>

- [12] Heinrich L and Simpson N 2020 pyhf/neos: initial zenodo release [Link] URL <https://doi.org/10.5281/zenodo.3697981>
- [13] Carrazza S, Cruz-Martinez J M and Rossi M 2021 *Computer Physics Communications* **264** 107995 ISSN 0010-4655 URL <http://dx.doi.org/10.1016/j.cpc.2021.107995>
- [14] Ball R D, Carrazza S, Cruz-Martinez J, Debbio L D, Forte S, Giani T, Iranipour S, Kassabov Z, Latorre J I, Nocera E R, Pearson R L, Rojo J, Stegeman R, Schwan C, Ubiali M, Voisey C and Wilson M 2021 An open-source machine learning framework for global analyses of parton distributions (*Preprint* 2109.02671)
- [15] Carrazza S, Cruz-Martinez J, Rossi M and Zaro M 2021 Madflow: automating monte carlo simulation on gpu for particle physics processes (*Preprint* 2106.10279)
- [16] Todorov E, Erez T and Tassa Y 2012 *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems* (IEEE) pp 5026–5033
- [17] Lee J, Grey M X, Ha S, Kunz T, Jain S, Ye Y, Srinivasa S S, Stilman M and Liu C K 2018 *Journal of Open Source Software* **3** 500
- [18] Coumans E 2015 *ACM SIGGRAPH 2015 Courses SIGGRAPH '15* (New York, NY, USA: Association for Computing Machinery) ISBN 9781450336345 URL <https://doi.org/10.1145/2776880.2792704>
- [19] de Avila Belbute-Peres F, Smith K, Allen K, Tenenbaum J and Kolter J Z 2018 *Advances in Neural Information Processing Systems* vol 31 ed Bengio S, Wallach H, Larochelle H, Grauman K, Cesa-Bianchi N and Garnett R (Curran Associates, Inc.)
- [20] Qiao Y L, Liang J, Koltun V and Lin M C 2020 *ICML* pp 7847–7856 URL <http://proceedings.mlr.press/v119/qiao20a.html>
- [21] Hu Y, Anderson L, Li T M, Sun Q, Carr N, Ragan-Kelley J and Durand F 2020 *ICLR*
- [22] Hu Y, Liu J, Spielberg A, Tenenbaum J B, Freeman W T, Wu J, Rus D and Matusik W 2019 *Proceedings of IEEE International Conference on Robotics and Automation (ICRA)*
- [23] Huang Z, Hu Y, Du T, Zhou S, Su H, Tenenbaum J B and Gan C 2021 *International Conference on Learning Representations* URL <https://openreview.net/forum?id=xCcdBRQEDW>
- [24] Liang J, Lin M and Koltun V 2019 *Advances in Neural Information Processing Systems* vol 32 ed Wallach H, Larochelle H, Beygelzimer A, d'Alché-Buc F, Fox E and Garnett R (Curran Associates, Inc.) URL <https://proceedings.neurips.cc/paper/2019/file/28f0b864598a1291557bed248a998d4e-Paper.pdf>
- [25] Li T M, Aittala M, Durand F and Lehtinen J 2018 *ACM Trans. Graph. (Proc. SIGGRAPH Asia)* **37** 222:1–222:11
- [26] Bangaru S, Li T M and Durand F 2020 *ACM Trans. Graph.* **39** 245:1–245:18
- [27] Zeltner T, Speierer S, Georgiev I and Jakob W 2021 *Transactions on Graphics (Proceedings of SIGGRAPH)* **40**
- [28] Mildenhall B, Srinivasan P P, Tancik M, Barron J T, Ramamoorthi R and Ng R 2020 *ECCV*
- [29] Holl P, Thurey N and Koltun V 2020 *International Conference on Learning Representations* URL <https://openreview.net/forum?id=HyeSin4FPB>
- [30] Ummerhofer B, Prantl L, Thurey N and Koltun V 2020 *International Conference on Learning Representations*
- [31] Um K, Brand R, Fei Y R, Holl P and Thurey N 2020 *Advances in Neural Information Processing Systems* vol 33 ed Larochelle H, Ranzato M, Hadsell R, Balcan M F and Lin H (Curran Associates, Inc.) pp 6111–6122
- [32] Schoenholz S S and Cubuk E D 2020 *Advances in Neural Information Processing Systems* vol 33 (Curran Associates, Inc.)
- [33] Doerr S, Majewski M, Pérez A, Krämer A, Clementi C, Noe F, Giorgino T and De Fabritiis G 2021 *Journal of Chemical Theory and Computation* **17** 2355–2363 pMID: 33729795 (*Preprint* <https://doi.org/10.1021/acs.jctc.0c01343>) URL <https://doi.org/10.1021/acs.jctc.0c01343>
- [34] de Aquino P, Link W, Maltoni F, Mattelaer O and Stelzer T 2012 *Computer Physics Communications* **183** 2254–2263 ISSN 0010-4655 URL <https://www.sciencedirect.com/science/article/pii/S0010465512001725>
- [35] Jax documentation: Common gotchas URL [https://jax.readthedocs.io/en/latest/notebooks/Common\\_Gotchas\\_in\\_JAX.html#control-flow](https://jax.readthedocs.io/en/latest/notebooks/Common_Gotchas_in_JAX.html#control-flow)
- [36] Cranmer K, Brehmer J and Louppe G 2020 *Proceedings of the National Academy of Sciences* **117** 30055–30062
- [37] Brehmer J, Louppe G, Pavez J and Cranmer K 2020 *Proceedings of the National Academy of Sciences* **117** 5242–5249 ISSN 0027-8424
- [38] Papamakarios G, Nalisnick E, Rezende D J, Mohamed S and Lakshminarayanan B 2021 *Journal of Machine Learning Research* **22** 1–64 URL <http://jmlr.org/papers/v22/19-1028.html>
- [39] Gao C, Höche S, Isaacson J, Krause C and Schulz H 2020 *Phys. Rev. D* **101**(7) 076002
- [40] Bothmann E, Janßen T, Knobbe M, Schmale T and Schumann S 2020 *SciPost Phys.* **8**(4) 69
- [41] Butter A and Plehn T 2020 Generative networks for lhc events (*Preprint* 2008.08558)
- [42] Albergo M S, Kanwar G and Shanahan P E 2019 *Phys. Rev. D* **100**(3) 034515
- [43] Sun X, Detmold W, Luo D and Shanahan P E 2022 Finite-volume pionless effective field theory for few-nucleon

- systems with differentiable programming (*Preprint 2202.03530*)
- [44] Wang J, Olsson S, Wehmeyer C, Pérez A, Charron N E, de Fabritiis G, Noé F and Clementi C 2019 *ACS Central Science* **5** 755–767
- [45] Köhler J, Krämer A and Noe F 2021 *Advances in Neural Information Processing Systems 34 pre-proceedings (NeurIPS 2021)* ed Beygelzimer A, Dauphin Y, Liang P and Vaughan J W
- [46] Kingma D P and Ba J 2015 *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings* ed Bengio Y and LeCun Y