New ML-Ideas for LHC Theory

Tilman Plehn

Universität Heidelberg

CERN, April 2022

Modern LHC physics

Classic motivation

- dark matter
- baryogenesis
- Higgs VEV

LHC physics

- fundamental questions
- huge data set
- complete uncertainty control
- first-principle precision simulations

Modern LHC physics

Classic motivation

- dark matter
- baryogenesis
- Higgs VEV

LHC physics

- fundamental questions
- huge data set
- complete uncertainty control
- first-principle precision simulations

Not our future

- seach for BSM models
- measure fiducial rates
- measure couplings

Modern LHC physics

Classic motivation

- dark matter
- baryogenesis
- Higgs VEV

LHC physics

- fundamental questions
- huge data set
- complete uncertainty control
- first-principle precision simulations

Not our future

- seach for BSM models
- measure fiducial rates
- measure couplings

Simulation-based inference

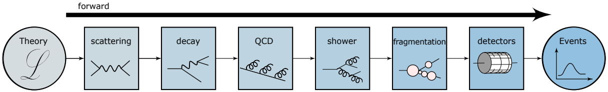
- start with Lagrangian
- calculate scattering using QFT
- simulate events [Sherpa, Madgraph, Pythia]
- simulate detectors [Geant4, Delphes]

$\rightarrow$ LHC events in virtual worlds

Searching for BSM physics

- compare simulations and data
- analyze data systematically [SMEFT]
- publish useable results
- understand LHC dataset [SM or BSM]

$\rightarrow$ With a little help from ML



Shortest ML-intro ever

Fit-like approximation

- approximate known $f(x)$ using $f_\theta(x)$
- no parametrization, just very many values θ
- θ-space the fun space [latent space]

Construction and contol

- define (well-defined) loss function
- minimize loss to find best θ
- compare $x \to f_\theta(x)$ for training/test data

LHC applications

- regression [matrix element over phase space]
- classification [gluon/quark/bottom/top inside jet]
- generation [event generation, detector simulation]
- conditional generation [unfolding, inference]
- ...

$\to$ Transforming numerical science

Generative networks

GANGogh [2017]

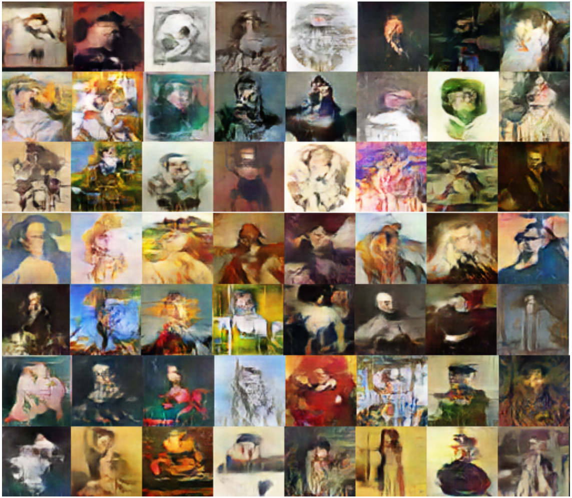
- generation $\quad r \to p_\theta(r)$ sampled $r \sim \mathcal{N}$
- networks to create new pieces of art
- train on 80,000 pictures
- generate flowers

Generative networks

GANGogh [2017]

- generation $\quad r \to p_\theta(r)$ sampled $r \sim \mathcal{N}$
- networks to create new pieces of art
- train on 80,000 pictures
- generate portraits

$\to$ Nowadays INNs

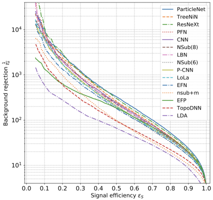


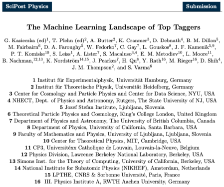
ML-applications for analysis

Top tagging [supervised classification]

- ‘hello world’ of LHC-ML
- different NN-architectures

$\rightarrow$ Just do it right...




SciPost Physics | Submission

The Machine Learning Landscape of Top Taggers

G. Kasieczka (ed)[1], T. Plehn (ed)[2], A. Butter[2], K. Cranmer[3], D. Debnath[4], B. M. Dillon[5], M. Fairbairn[6], D. A. Faroughy[5], W. Fedorko[7], C. Gay[7], L. Gouskos[8], J. F. Kamenik[5,9], P. T. Komiske[10], S. Leiss[1], A. Lister[7], S. Macaluso[3,4], E. M. Metodiev[10], L. Moore[11], B. Nachman,[12,13], K. Nordström[14,15], J. Pearkes[7], H. Qu[8], Y. Rath[16], M. Rieger[16], D. Shih[4], J. M. Thompson[2], and S. Varma[6]

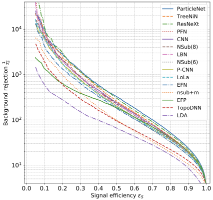
1 Institut für Experimentalphysik, Universität Hamburg, Germany
2 Institut für Theoretische Physik, Universität Heidelberg, Germany
3 Center for Cosmology and Particle Physics and Center for Data Science, NYU, USA
4 NHECT, Dept. of Physics and Astronomy, Rutgers, The State University of NJ, USA
5 Jozef Stefan Institute, Ljubljana, Slovenia
6 Theoretical Particle Physics and Cosmology, King’s College London, United Kingdom
7 Department of Physics and Astronomy, The University of British Columbia, Canada
8 Department of Physics, University of California, Santa Barbara, USA
9 Faculty of Mathematics and Physics, University of Ljubljana, Ljubljana, Slovenia
10 Center for Theoretical Physics, MIT, Cambridge, USA
11 CP3, Universitéxx Catholique de Louvain, Louvain-la-Neuve, Belgium
12 Physics Division, Lawrence Berkeley National Laboratory, Berkeley, USA
13 Simons Inst. for the Theory of Computing, University of California, Berkeley, USA
14 National Institute for Subatomic Physics (NIKHEF), Amsterdam, Netherlands
15 LPTHE, CNRS & Sorbonne Université, Paris, France
16 III. Physics Institute A, RWTH Aachen University, Germany

ML-applications for analysis

Top tagging [supervised classification]

- ‘hello world’ of LHC-ML
- different NN-architectures

$\rightarrow$ Just do it right...



SciPost Physics Submission

The Machine Learning Landscape of Top Taggers

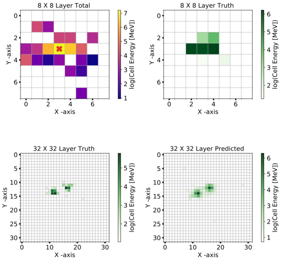
G. Kasieczka (ed)[1], T. Plehn (ed)[2], A. Butter[2], K. Cranmer[3], D. Debnath[4], B. M. Dillon[5], M. Fairbairn[6], D. A. Faroughy[5], W. Fedorko[7], C. Gay[7], L. Gouskos[8], J. F. Kamenik[5,9], P. T. Komiske[10], S. Leiss[1], A. Lister[7], S. Macaluso[3,4], E. M. Metodiev[10], L. Moore[11], B. Nachman,[12,13], K. Nordström[14,15], J. Pearkes[7], H. Qu[8], Y. Rath[16], M. Rieger[16], D. Shih[4], J. M. Thompson[2], and S. Varma[6]

1 Institut für Experimentalphysik, Universität Hamburg, Germany
2 Institut für Theoretische Physik, Universität Heidelberg, Germany
3 Center for Cosmology and Particle Physics and Center for Data Science, NYU, USA
4 NHECT, Dept. of Physics and Astronomy, Rutgers, The State University of NJ, USA
5 Jozef Stefan Institute, Ljubljana, Slovenia
6 Theoretical Particle Physics and Cosmology, King's College London, United Kingdom
7 Department of Physics and Astronomy, The University of British Columbia, Canada
8 Department of Physics, University of California, Santa Barbara, USA
9 Faculty of Mathematics and Physics, University of Ljubljana, Ljubljana, Slovenia
10 Center for Theoretical Physics, MIT, Cambridge, USA
11 CP3, Universitéxx Catholique de Louvain, Louvain-la-Neuve, Belgium
12 Physics Division, Lawrence Berkeley National Laboratory, Berkeley, USA
13 Simons Inst. for the Theory of Computing, University of California, Berkeley, USA
14 National Institute for Subatomic Physics (NIKHEF), Amsterdam, Netherlands
15 LPTHE, CNRS & Sorbonne Université, Paris, France
16 III. Physics Institute A, RWTH Aachen University, Germany

Particle flow [classification, super-resolution]

- mother of jet tools
- combined detector channels

$\rightarrow$ Seriously impressive



Towards a Computer Vision Particle Flow ⋆

Francesco Armando Di Bello[a,3], **Sanmay Ganguly**[b,1], **Eilam Gross**[1], **Marumi Kado**[3,4], **Michael Pitt**[2], **Lorenzo Santi** [3], **Jonathan Shlomi**[1]

[1]Weizmann Institute of Science, Rehovot 76100, Israel
[2]CERN, CH 1211, Geneva 23, Switzerland
[3]Università di Roma Sapienza, Piazza Aldo Moro, 2, 00185 Roma, Italy e INFN, Italy
[4]Université Paris-Saclay, CNRS/IN2P3, IJCLab, 91405, Orsay, France

Fig. 7: An event display of total energy shower (within topocluster), as captured by a calorimeter layer of 8×8 granularity, along with the location of the track, denoted by a red cross (left) and the same shower is captured by a calorimeter layer of 32×32 granularity (middle). The bottom right panel shows the corresponding event predicted by the NN. The figure shows that the shower originating from a $\pi^0 \rightarrow \gamma\gamma$ is resolved by a 32×32 granularity layer.

ML4Theory

Tilman Plehn

LHC physics

ML examples

Uncertainties

Inversion

Symbolic reg

QCD and symmetries

Lund plane representation [input preprocessing]

- QCD-inspired input with cutting-edge networks
- angular separation vs transverse momentum

$\rightarrow$ Understanding data helps

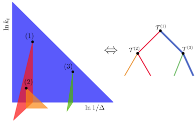


Figure 1. The Lund plane representation of a jet (left) where each emission is positioned according to its Δ and k_t coordinates, and the corresponding mapping to a binary Lund tree of tuples (right). The thick blue line represents the primary sequence of tuples $\mathcal{L}_{\text{primary}}$.

Prepared for submission to JHEP OUTP-20-15P

Jet tagging in the Lund plane with graph networks

Frédéric A. Dreyer,[a] Huilin Qu[b]

[a] *Rudolf Peierls Centre for Theoretical Physics, Clarendon Laboratory, Parks Road, Oxford OX1 3PU, UK*

[b] *CERN, EP Department, CH-1211 Geneva 23, Switzerland*

Abstract: The identification of boosted heavy particles such as top quarks or vector bosons is one of the key problems arising in experimental studies at the Large Hadron Collider. In this article, we introduce LundNet, a novel jet tagging method which relies on graph neural networks and an efficient description of the radiation patterns within a jet to optimally disentangle signatures of boosted objects from background events. We apply this framework to a number of different benchmarks, showing significantly improved performance for top tagging compared to existing state-of-the-art algorithms. We study the robustness of the LundNet taggers to non-perturbative and detector effects, and show how kinematic cuts in the Lund plane can mitigate overfitting of the neural network to model-dependent contributions. Finally, we consider the computational complexity of this method and its scaling as a function of kinematic Lund plane cuts, showing an order of magnitude improvement in speed over previous graph-based taggers.

QCD and symmetries

Lund plane representation [input preprocessing]

- QCD-inspired input with cutting-edge networks
- angular separation vs transverse momentum

→ Understanding data helps

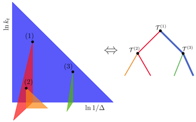


Figure 1. The Lund plane representation of a jet (left) where each emission is positioned according to its Δ and k_t coordinates, and the corresponding mapping to a binary Lund tree of tuples (right). The thick blue line represents the primary sequence of tuples $\mathcal{L}_{\text{primary}}$.

Prepared for submission to JHEP OUTP-20-15P

Jet tagging in the Lund plane with graph networks

Frédéric A. Dreyer,[a] Huilin Qu[b]

[a] Rudolf Peierls Centre for Theoretical Physics, Clarendon Laboratory, Parks Road, Oxford OX1 3PU, UK

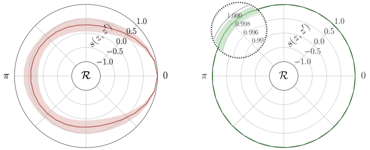
[b] CERN, EP Department, CH-1211 Geneva 23, Switzerland

Abstract: The identification of boosted heavy particles such as top quarks or vector bosons is one of the key problems arising in experimental studies at the Large Hadron Collider. In this article, we introduce LundNet, a novel jet tagging method which relies on graph neural networks and an efficient description of the radiation patterns within a jet to optimally disentangle signatures of boosted objects from background events. We apply this framework to a number of different benchmarks, showing significantly improved performance for top tagging compared to existing state-of-the-art algorithms. We study the robustness of the LundNet taggers to non-perturbative and detector effects, and show how kinematic cuts in the Lund plane can mitigate overfitting of the neural network to model-dependent contributions. Finally, we consider the computational complexity of this method and its scaling as a function of kinematic Lund plane cuts, showing an order of magnitude improvement in speed over previous graph-based taggers.

Self-supervised training [contrastive learning, transformer network]

- rotations, translations, permutations, soft splittings, collinear splittings
- learn symmetries/augmentations

→ Symmetry-aware latent space



SciPost Physics Submission

Symmetries, Safety, and Self-Supervision

Barry M. Dillon[1], Gregor Kasieczka[2], Hans Olischlager[1], Tilman Plehn[1], Peter Sorrenson[3], and Lorenz Vogel[1]

1 Institut für Theoretische Physik, Universität Heidelberg, Germany
2 Institut für Experimentalphysik, Universität Hamburg, Germany
3 Heidelberg Collaboratory for Image Processing, Universität Heidelberg, Germany

August 11, 2021

Abstract

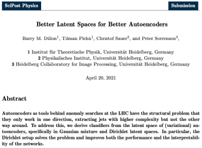
Collider searches face the challenge of defining a representation of high-dimensional data such that physical symmetries are manifest, the discriminating features are retained, and the choice of representation is new-physics agnostic. We introduce JetCLR to solve the mapping from low-level data to optimized observables though self-supervised contrastive learning. As an example, we construct a data representation for top and QCD jets using a permutation-invariant transformer-encoder network and visualize its symmetry properties. We compare the JetCLR representation with alternative representations using linear classifier tests and find it to work quite well.

Non-QCD and parton densities

Anomaly searches [unsupervised training]

- train on QCD-jets, SM-events
- look for non-QCD jets, non-SM events

→ Trigger, searches


SciPost Physics Submission

Better Latent Spaces for Better Autoencoders

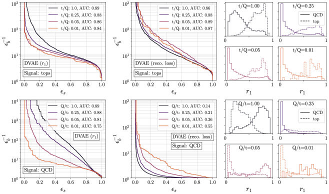
Barry M. Dillon[1], Tilman Plehn[1], Christof Sauer[2], and Peter Sorrenson[3],

1 Institut für Theoretische Physik, Universität Heidelberg, Germany
2 Physikalisches Institut, Universität Heidelberg, Germany
3 Heidelberg Collaboratory for Image Processing, Universität Heidelberg, Germany

April 20, 2021

Abstract

Autoencoders as tools behind anomaly searches at the LHC have the structural problem that they only work in one direction, extracting jets with higher complexity but not the other way around. To address this, we derive classifiers from the latent space of (variational) autoencoders, specifically in Gaussian mixture and Dirichlet latent spaces. In particular, the Dirichlet setup solves the problem and improves both the performance and the interpretability of the networks.



Non-QCD and parton densities

Anomaly searches [unsupervised training]

- train on QCD-jets, SM-events
- look for non-QCD jets, non-SM events

→ Trigger, searches

SciPost Physics Submission

Better Latent Spaces for Better Autoencoders

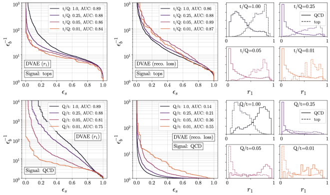
Barry M. Dillon[1], Tilman Plehn[1], Christof Sauer[2], and Peter Sorrenson[3],

1 Institut für Theoretische Physik, Universität Heidelberg, Germany
2 Physikalisches Institut, Universität Heidelberg, Germany
3 Heidelberg Collaboratory for Image Processing, Universität Heidelberg, Germany

April 20, 2021

Abstract

Autoencoders as tools behind anomaly searches at the LHC have the structural problem that they only work in one direction, extracting jets with higher complexity but not the other way around. To address this, we derive classifiers from the latent space of (variational) autoencoders, specifically in Gaussian mixture and Dirichlet latent spaces. In particular, the Dirichlet setup solves the problem and improves both the performance and the interpretability of the networks.



NNPDF/N3PDF parton densities [full blast]

- starting point: pdfs without functional ansatz
- moving on: cutting-edge ML everywhere

→ Leaders in ML-theory

A data-based parametrization of parton distribution functions

Stefano Carrazza[1,2,3], Juan Cruz-Martinez[1], and Roy Stegeman[1]

[1] TIF Lab, Dipartimento di Fisica, Università degli Studi di Milano and INFN Sezione di Milano.
[2] CERN, Theoretical Physics Department, CH-1211 Geneva 23, Switzerland.
[3] Quantum Research Centre, Technology Innovation Institute, Abu Dhabi, UAE.

Received: date / Revised version: date

Abstract. Since the first determination of a structure function many decades ago, all methodologies used to determine structure functions or parton distribution functions (PDFs) have employed a common prefactor as part of the parametrization. The NNPDF collaboration pioneered the use of neural networks to overcome the inherent bias of constraining the space of solution with a fixed functional form while still keeping the same common prefactor as a preprocessing. Over the years various, increasingly sophisticated, techniques have been introduced to counter the effect of the prefactor on the PDF determination. In this paper we present a methodology to remove the prefactor entirely, thereby significantly simplifying the methodology, without a loss of efficiency and finding good agreement with previous results.

PACS. 12.38.-t Quantum chromodynamics – 12.39.-x Phenomenological quark models – 84.35.+i Neural Networks

Events and amplitudes

Speeding up Sherpa [sampling]

- precision simulations limiting factor for Runs 3&4
- unweighting critical

→ Phase space sampling

	$gg \to t\bar{t}ggg$	$ug \to t\bar{t}ggu$	$uu \to t\bar{t}guu$	$u\bar{u} \to t\bar{t}gd\bar{d}$
ϵ_{full}	1.1e−2	7.3e−3	6.8e−3	6.6e−4
$\epsilon_{\text{1st,surr}}$	8.7e−3	5.8e−3	4.7e−3	3.6e−4
$\langle t_{\text{full}}\rangle / \langle t_{\text{surr}}\rangle$	39312	2417	199	64
$x_{\max}^{\text{p.m.}}$	52.03	32.52	69.76	326.19
$\epsilon_{\text{2nd,surr}}^{\text{p.m.}}$	2.4e−2	3.8e−2	2.1e−2	5.6e−3
$\alpha^{\text{p.m.}}$	0.9989	0.9984	0.9994	0.9981
$f_{\text{eff}}^{\text{p.m.}}$	2.21	4.89	1.47	0.19
$x_{\max}^{\text{med}}$	30.40	19.14	27.78	25.34
$\epsilon_{\text{2nd,surr}}^{\text{med}}$	4.3e−2	6.4e−2	5.1e−2	7.1e−2
α^{med}	0.9983	0.9966	0.9943	0.9321
$f_{\text{eff}}^{\text{med}}$	3.90	8.26	3.91	2.22

Table 6: Performance measures for partonic channels contributing to $t\bar{t}$+3 jets production at the LHC.

SciPost Physics Submission

MCNET-21-13

Accelerating Monte Carlo event generation – rejection sampling using neural network event-weight estimates

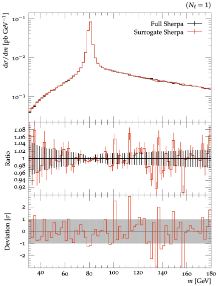
K. Danziger[1], T. Janßen[2], S. Schumann[2], F. Siegert[1]

1 Institut für Kern- und Teilchenphysik, TU Dresden, Dresden, Germany
2 Institut für Theoretische Physik, Georg-August-Universität Göttingen, Göttingen, Germany

September 27, 2021

Abstract

The generation of unit-weight events for complex scattering processes presents a severe challenge to modern Monte Carlo event generators. Even when using sophisticated phase-space sampling techniques adapted to the underlying transition matrix elements, the efficiency for generating unit-weight events from weighted samples can become a limiting factor in practical applications. Here we present a novel two-staged unweighting procedure that makes use of a neural-network surrogate for the full event weight. The algorithm can significantly accelerate the unweighting process, while it still guarantees unbiased sampling from the correct target distribution. We apply, validate and benchmark the new approach in high-multiplicity LHC production processes, including Z/W+4 jets and $t\bar{t}$+3 jets, where we find speed-up factors up to ten.



ML4Theory

Tilman Plehn

LHC physics

ML examples

Uncertainties

Inversion

Symbolic reg

Events and amplitudes

Speeding up Sherpa [sampling]

- precision simulations limiting factor for Runs 3&4
- unweighting critical

→ Phase space sampling

	$gg \to t\bar{t}ggg$	$ug \to t\bar{t}ggu$	$uu \to t\bar{t}guu$	$u\bar{u} \to t\bar{t}gd\bar{d}$
ϵ_{full}	1.1e−2	7.3e−3	6.8e−3	6.6e−4
$\epsilon_{\text{1st,surr}}$	8.7e−3	5.8e−3	4.7e−3	3.6e−4
$\langle t_{\text{full}}\rangle / \langle t_{\text{surr}}\rangle$	39312	2417	199	64
$x_{\text{max}}^{\text{p.m.}}$	52.03	32.52	69.76	326.19
$\epsilon_{\text{2nd,surr}}^{\text{p.m.}}$	2.4e−2	3.8e−2	2.1e−2	5.6e−3
$\alpha^{\text{p.m.}}$	0.9989	0.9984	0.9994	0.9981
$f_{\text{eff}}^{\text{p.m.}}$	2.21	4.89	1.47	0.19
$x_{\text{max}}^{\text{med}}$	30.40	19.14	27.78	25.34
$\epsilon_{\text{2nd,surr}}^{\text{med}}$	4.3e−2	6.4e−2	5.1e−2	7.1e−2
α^{med}	0.9983	0.9966	0.9943	0.9321
$f_{\text{eff}}^{\text{med}}$	3.90	8.26	3.91	2.22

Table 6: Performance measures for partonic channels contributing to $t\bar{t}$+3 jets production at the LHC.

SciPost Physics Submission

MCNET-21-13

Accelerating Monte Carlo event generation – rejection sampling using neural network event-weight estimates

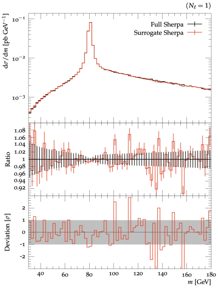
K. Danziger[1], T. Janßen[2], S. Schumann[2], F. Siegert[1]

1 Institut für Kern- und Teilchenphysik, TU Dresden, Dresden, Germany
2 Institut für Theoretische Physik, Georg-August-Universität Göttingen, Göttingen, Germany

September 27, 2021

Abstract

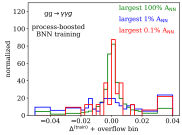
The generation of unit-weight events for complex scattering processes presents a severe challenge to modern Monte Carlo event generators. Even when using sophisticated phase-space sampling techniques adapted to the underlying transition matrix elements, the efficiency for generating unit-weight events from weighted samples can become a limiting factor in practical applications. Here we present a novel two-staged unweighting procedure that makes use of a neural-network surrogate for the full event weight. The algorithm can significantly accelerate the unweighting process, while it still guarantees unbiased sampling from the correct target distribution. We apply, validate and benchmark the new approach in high-multiplicity LHC production processes, including Z/W+4 jets and $t\bar{t}$+3 jets, where we find speed-up factors up to ten.

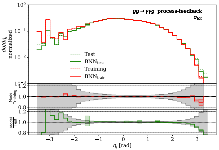


Speeding up amplitudes [precision regression]

- loop-amplitudes expensive
- interpolation standard

→ Network amplitudes





Prepared for submission to JHEP IPPP/20/116

Optimising simulations for diphoton production at hadron colliders using amplitude neural networks

Joseph Aylett-Bullock[a,b] Simon Badger[c] Ryan Moodie[a]

[a] Institute for Particle Physics Phenomenology, Department of Physics, Durham University, Durham, DH1 3LE, United Kingdom
[b] Institute for Data Science, Durham University, Durham, DH1 3LE, United Kingdom
[c] Dipartimento di Fisica and Arnold-Regge Center, Università di Torino, and INFN, Sezione di Torino, Via P. Giuria 1, I-10125 Torino, Italy

E-mail: j.p.bullock@durham.ac.uk, simondavid.badger@unito.it, ryan.i.moodie@durham.ac.uk

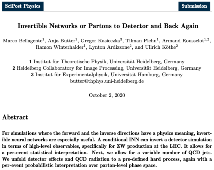
Abstract: Machine learning technology has the potential to dramatically optimise event generation and simulations. We continue to investigate the use of neural networks to approximate matrix elements for high-multiplicity scattering processes. We focus on the case of loop-induced diphoton production through gluon fusion, and develop a realistic simulation method that can be applied to hadron collider observables. Neural networks are trained using the one-loop amplitudes implemented in the NJet C++ library, and interfaced to the Sherpa Monte Carlo event generator, where we perform a detailed study for $2 \to 3$ and $2 \to 4$ scattering problems. We also consider how the trained networks perform when varying the kinematic cuts effecting the phase space and the reliability of the neural network simulations.

Invertible event generation and errors

Unfolding and inversion [conditional normalizing flows]

- shower/hadronization unfolded by jet algorithm
- detector/decays unfolded e.g. in tops
- calibrated inverse sampling

→ Discussed later



SciPost Physics Submission

Invertible Networks or Partons to Detector and Back Again

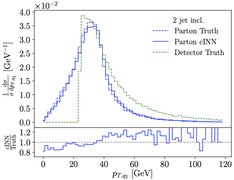
Marco Bellagente[1], Anja Butter[1], Gregor Kasieczka[3], Tilman Plehn[1], Armand Rousselot[1,2], Ramon Winterhalder[1], Lynton Ardizzone[2], and Ullrich Köthe[2]

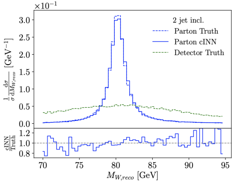
1 Institut für Theoretische Physik, Universität Heidelberg, Germany
2 Heidelberg Collaboratory for Image Processing, Universität Heidelberg, Germany
3 Institut für Experimentalphysik, Universität Hamburg, Germany
butter@thphys.uni-heidelberg.de

October 2, 2020

Abstract

For simulations where the forward and the inverse directions have a physics meaning, invertible neural networks are especially useful. A conditional INN can invert a detector simulation in terms of high-level observables, specifically for ZW production at the LHC. It allows for a per-event statistical interpretation. Next, we allow for a variable number of QCD jets. We unfold detector effects and QCD radiation to a pre-defined hard process, again with a per-event probabilistic interpretation over parton-level phase space.



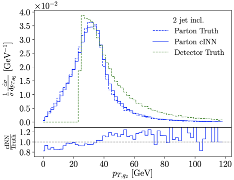


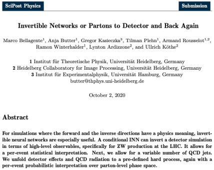
Invertible event generation and errors

Unfolding and inversion [conditional normalizing flows]

- shower/hadronization unfolded by jet algorithm
- detector/decays unfolded e.g. in tops
- calibrated inverse sampling

→ Discussed later




SciPost Physics Submission

Invertible Networks or Partons to Detector and Back Again

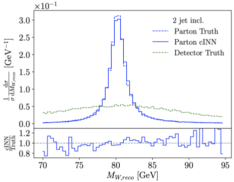
Marco Bellagente[1], Anja Butter[1], Gregor Kasieczka[3], Tilman Plehn[1], Armand Rousselot[1,2], Ramon Winterhalder[1], Lynton Ardizzone[2], and Ullrich Köthe[2]

1 Institut für Theoretische Physik, Universität Heidelberg, Germany
2 Heidelberg Collaboratory for Image Processing, Universität Heidelberg, Germany
3 Institut für Experimentalphysik, Universität Hamburg, Germany
butter@thphys.uni-heidelberg.de

October 2, 2020

Abstract

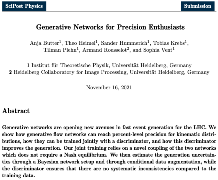
For simulations where the forward and the inverse directions have a physics meaning, invertible neural networks are especially useful. A conditional INN can invert a detector simulation in terms of high-level observables, specifically for ZW production at the LHC. It allows for a per-event statistical interpretation. Next, we allow for a variable number of QCD jets. We unfold detector effects and QCD radiation to a pre-defined hard process, again with a per-event probabilistic interpretation over parton-level phase space.



Generative networks with uncertainties [Bayesian discriminator-flows]

- control through discriminator [GAN-like]
- uncertainties through Bayesian networks

→ Discussed later


SciPost Physics Submission

Generative Networks for Precision Enthusiasts

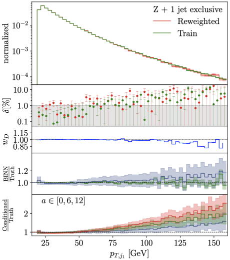
Anja Butter[1], Theo Heimel[1], Sander Hummerich[1], Tobias Krebs[1], Tilman Plehn[1], Armand Rousselot[2], and Sophia Vent[1]

1 Institut für Theoretische Physik, Universität Heidelberg, Germany
2 Heidelberg Collaboratory for Image Processing, Universität Heidelberg, Germany

November 16, 2021

Abstract

Generative networks are opening new avenues in fast event generation for the LHC. We show how generative flow networks can reach percent-level precision for kinematic distributions, how they can be trained jointly with a discriminator, and how this discriminator improves the generation. Our joint training relies on a novel coupling of the two networks which does not require a Nash equilibrium. We then estimate the generation uncertainties through a Bayesian network setup and through conditional data augmentation, while the discriminator ensures that there are no systematic inconsistencies compared to the training data.



String landscape and learned formulas

Navigating string landscape [reinforcement learning]

- searching for viable vacua
- high dimensions, unknown global structure

→ Model space sampling

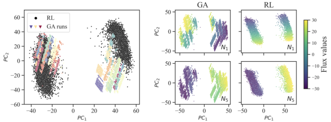


Figure 1: *Left:* Cluster structure in dimensionally reduced flux samples for RL and 25 GA runs (PCA on all samples of GA and RL). The colors indicate individual GA runs. *Right:* Dependence on flux (input) values (N_3 and N_5 respectively) in relation to principal components for a PCA fit of the individual output of GA and RL.

Probing the Structure of String Theory Vacua with Genetic Algorithms and Reinforcement Learning

Alex Cole
University of Amsterdam
a.e.cole@uva.nl

Sven Krippendorf
Arnold Sommerfeld Center for Theoretical Physics
LMU Munich
sven.krippendorf@physik.uni-muenchen.de

Andreas Schachner
Centre for Mathematical Sciences
University of Cambridge
as2673@cam.ac.uk

Gary Shiu
University of Wisconsin-Madison
shiu@physics.wisc.edu

Abstract

Identifying string theory vacua with desired physical properties at low energies requires searching through high-dimensional solution spaces – collectively referred to as the string landscape. We highlight that this search problem is amenable to reinforcement learning and genetic algorithms. In the context of flux vacua, we are able to reveal novel features (suggesting previously unidentified symmetries) in the string theory solutions required for properties such as the string coupling. In order to identify these features robustly, we combine results from both search methods, which we argue is imperative for reducing sampling bias.

String landscape and learned formulas

Navigating string landscape [reinforcement learning]

- searching for viable vacua
- high dimensions, unknown global structure

→ Model space sampling

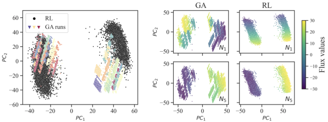


Figure 1: *Left:* Cluster structure in dimensionally reduced flux samples for RL and 25 GA runs (PCA on all samples of GA and RL). The colors indicate individual GA runs. *Right:* Dependence on flux (input) values (N_3 and N_5 respectively) in relation to principal components for a PCA fit of the individual output of GA and RL.

Probing the Structure of String Theory Vacua with Genetic Algorithms and Reinforcement Learning

Alex Cole
University of Amsterdam
a.e.cole@uva.nl

Sven Krippendorf
Arnold Sommerfeld Center for Theoretical Physics
LMU Munich
sven.krippendorf@physik.uni-muenchen.de

Andreas Schachner
Centre for Mathematical Sciences
University of Cambridge
as2673@cam.ac.uk

Gary Shiu
University of Wisconsin-Madison
shiu@physics.wisc.edu

Abstract

Identifying string theory vacua with desired physical properties at low energies requires searching through high-dimensional solution spaces – collectively referred to as the string landscape. We highlight that this search problem is amenable to reinforcement learning and genetic algorithms. In the context of flux vacua, we are able to reveal novel features (suggesting previously unidentified symmetries) in the string theory solutions required for properties such as the string coupling. In order to identify these features robustly, we combine results from both search methods, which we argue is imperative for reducing sampling bias.

Learning formulas [genetic algorithm, symbolic regression]

- approximate numerical function through formula
- example: score/optimal observables

→ Discussed later

compl	dof	function	MSE
3	1	$a\,\Delta\phi$	$1.30 \cdot 10^{-1}$
4	1	$\sin(a\Delta\phi)$	$2.75 \cdot 10^{-1}$
5	1	$a\Delta\phi x_{p,1}$	$9.93 \cdot 10^{-2}$
6	1	$-x_{p,1}\sin(\Delta\phi + a)$	$1.90 \cdot 10^{-1}$
7	1	$(-x_{p,1} - a)\sin(\sin(\Delta\phi))$	$5.63 \cdot 10^{-2}$
8	1	$(a - x_{p,1})x_{p,2}\sin(\Delta\phi)$	$1.61 \cdot 10^{-2}$
14	2	$x_{p,1}(a\Delta\phi - \sin(\sin(\Delta\phi)))(x_{p,2} + b)$	$1.44 \cdot 10^{-2}$
15	3	$-(x_{p,2}(a\Delta\eta^2 + x_{p,1}) + b)\sin(\Delta\phi + c)$	$1.30 \cdot 10^{-2}$
16	4	$-x_{p,1}(a - b\Delta\eta)(x_{p,2} + c)\sin(\Delta\phi + d)$	$8.50 \cdot 10^{-3}$
28	7	$(x_{p,2} + a)(bx_{p,1}(c - \Delta\phi) - x_{p,1}(d\Delta\eta + ex_{p,2} + f)\sin(\Delta\phi + g))$	$8.18 \cdot 10^{-3}$

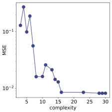


Table 8: Score hall of fame for simplified WBF Higgs production with $f_{W\widetilde{W}} = 0$, including a optimization fit.

SciPost Physics
Submission

Back to the Formula — LHC Edition

Anja Butter[1], Tilman Plehn[1], Nathalie Soybelman[1], and Johann Brehmer[2]

1 Institut für Theoretische Physik, Universität Heidelberg, Germany
2 Center for Data Science, New York University, New York, United States
nathalie@soybelman.de

November 16, 2021

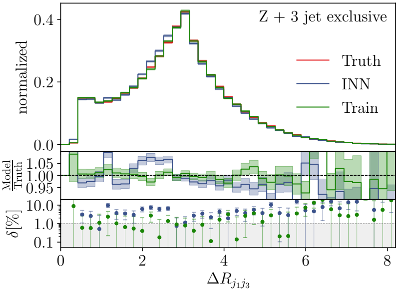
Abstract

While neural networks offer an attractive way to numerically encode functions, actual formulas remain the language of theoretical particle physics. We use symbolic regression trained on matrix-element information to extract, for instance, optimal LHC observables. This way we invert the usual simulation paradigm and extract easily interpretable formulas from complex simulated data. We introduce the method using the effect of a dimension-6 coefficient on associated ZH production. We then validate it for the known case of CP-violation in weak-boson-fusion Higgs production, including detector effects.

Controlled precision generator

ML-event generators

- useful ML-playground
  efficient ways to ship events
  training on combined MC and data
  transferable to detector simulation
- training from event samples
  no detector effects [Fastsim easy to include]
- $Z_{\mu\mu} + \{1, 2, 3\}$ jets [Z-peak, variable jet number, jet-jet topology]



Controlled precision generator

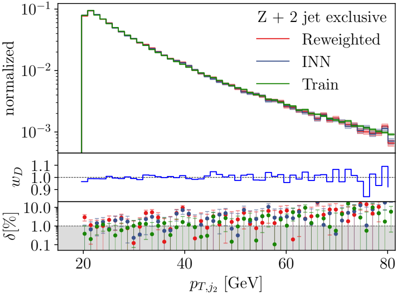
ML-event generators

- useful ML-playground
  efficient ways to ship events
  training on combined MC and data
  transferable to detector simulation
- training from event samples
  no detector effects [Fastsim easy to include]
- $Z_{\mu\mu} + \{1, 2, 3\}$ jets [Z-peak, variable jet number, jet-jet topology]

Control through discriminator

- classification easier than generation
- input $\{p_T, \eta, \phi, M, M_{\mu\mu}, \Delta R\}$
- output $D = 0(\text{generator}), 1(\text{truth})$
- decent generator training $D \approx 0.5$
- additional event weight $w_D = \frac{D}{1-D}$

$\rightarrow$ Control & reweight

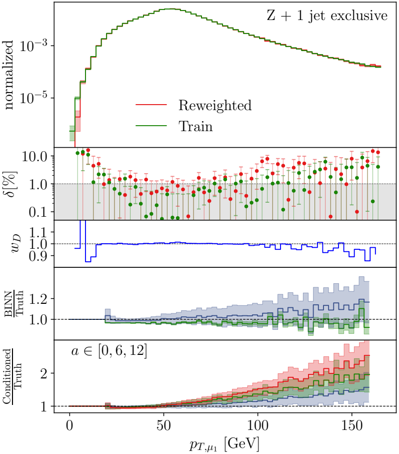


Uncertain precision generator

Uncertainties from Bayesian INN

- learned phase space density
  plus uncertainty over phase space
- useful after control step
- low statistics means large uncertainty

$\rightarrow$ Training-related error bars

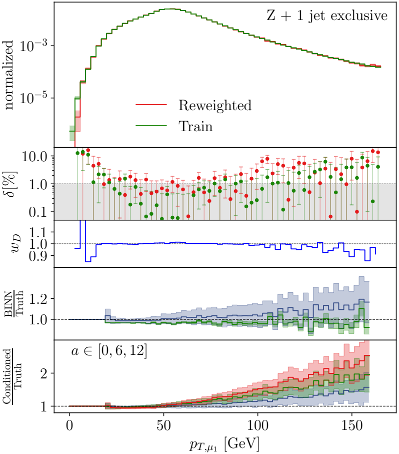


Uncertain precision generator

Uncertainties from Bayesian INN

- learned phase space density
  plus uncertainty over phase space
- useful after control step
- low statistics means large uncertainty

$\rightarrow$ Training-related error bars



Theory uncertainties

- systematics from data augmentation

$$w = 1 + a \left(\frac{p_{T,j_1} - 15\ \text{GeV}}{100\ \text{GeV}} \right)^2$$

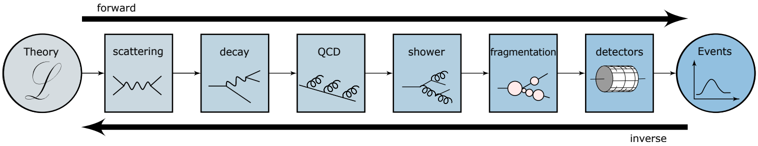
- train conditionally on a
- uncertainty from sampling a
- correlation to all of phase space

$\rightarrow$ Network for LHC standards

Inverse simulation

Invertible ML-simulation

- forward: $r \to$ events trained on model
- inverse: $r \to$ anything trained on model, conditioned on event



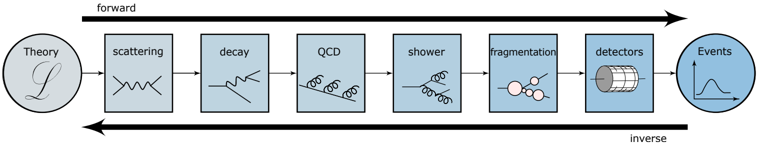
Inverse simulation

Invertible ML-simulation

- forward: $r \rightarrow$ events trained on model
- inverse: $r \rightarrow$ anything trained on model, conditioned on event
- individual steps known problems

  detector unfolding [needed for global analyses]
  unfolding to QCD parton means jet algorithm
  unfolding jet radiation known combinatorics problem
  unfolding to hard process standard in top groups [needed for global analyses]
  matrix element method an old dream
- improved through coherent ML-method

$\rightarrow$ Free choice of data-theory inference point



Inverse simulation

Invertible ML-simulation

- forward: $r \to$ events trained on model
- inverse: $r \to$ anything trained on model, conditioned on event
- individual steps known problems

  detector unfolding [needed for global analyses]
  unfolding to QCD parton means jet algorithm
  unfolding jet radiation known combinatorics problem
  unfolding to hard process standard in top groups [needed for global analyses]
  matrix element method an old dream
- improved through coherent ML-method

$\to$ Free choice of data-theory inference point

Conditional INN

- partonic events from $\{r\}$, given detector event
- maximum likelihood loss

$$L = -\left\langle \log p(\theta|x_p, x_d) \right\rangle_{x_p, x_d}$$
$$= -\left\langle \log p(g(x_p, x_d)) + \log \left| \frac{\partial g(x_p, x_d)}{\partial x_p} \right| \right\rangle_{x_p, x_d} - \log p(\theta) + \text{const.}$$

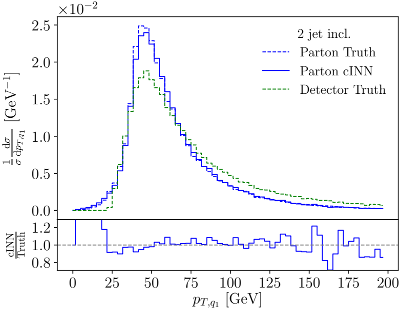
- eventually to be combined with reweighting

$\to$ Stable and statistically calibrated

Inverting to hard process

Undo QCD jet radiation

- nasty jet combinatorics, little information
- detector level: $pp \to ZW$+jets [variable number of objects]
- hard process given, ME vs PS jets from network



Inverting to hard process

Undo QCD jet radiation

- nasty jet combinatorics, little information
- detector level: $pp \to ZW$+jets [variable number of objects]
- hard process given, ME vs PS jets from network

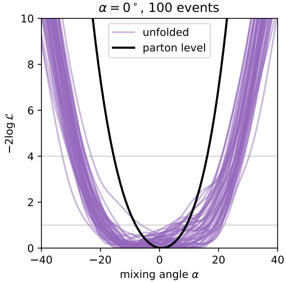
Matrix element method [Butter, Heimel, Martini, Peitzsch, TP (soon)]

- parameter likelihood from parton-level events [think $pp \to tH$ with CPV]

$$\mathcal{L}(\theta) = \prod_{i=1}^{N} p(\vec{x}^{(i)}|\theta) = \prod_{i=1}^{N} \frac{1}{\sigma_\text{fid}(\theta)} \int d^m z \, \frac{d^m\sigma(\theta)}{dz_1 \dots dz_m} \, T(\vec{x}^{(i)}, \vec{z})$$

$$T(\vec{x}, \vec{z}) = p_\text{INN}(\vec{z}|\vec{x})\epsilon(\vec{x}) \qquad \Rightarrow \qquad \mathcal{L}(\theta) = \prod_{i=1}^{N} \frac{\epsilon(\vec{x}^{(i)})}{\sigma_\text{fid}(\theta)} \int d^m z \, \frac{d^m\sigma(\theta)}{dz_1 \dots dz_m} \, p_\text{INN}(\vec{z}|\vec{x}^{(i)})$$

$$= \prod_{i=1}^{N} \frac{\epsilon(\vec{x}^{(i)})}{\sigma_\text{fid}(\theta)} \left\langle \frac{d^m\sigma(\theta)}{dz_1 \dots dz_m} \right\rangle_{\vec{z} \sim p_\text{INN}}$$



Optimal observables

Measure model parameter θ optimally

- single-event likelihood again

$$p(x|\theta) = \frac{1}{\sigma_{\text{tot}}(\theta)} \frac{d^m\sigma(x|\theta)}{dx^m}$$

- expanded in θ around θ_0, define score

$$\log \frac{p(x|\theta)}{p(x|\theta_0)} \approx (\theta - \theta_0)\, \nabla_\theta \log p(x|\theta)\Big|_{\theta_0} \equiv (\theta - \theta_0)\; t(x|\theta_0) \equiv (\theta - \theta_0)\; \mathcal{O}^{\text{opt}}(x)$$

- parton level, as used in ATLAS [CPV]

$$p(x|\theta) \approx |\mathcal{M}|_0^2 + \theta |\mathcal{M}|_{\text{int}}^2 \qquad \Rightarrow \qquad t(x|\theta_0) \sim \frac{|\mathcal{M}|_{\text{int}}^2}{|\mathcal{M}|_0^2} \,,$$

$\Rightarrow$ Easy at parton level, LEP physics...

Optimal observables

Measure model parameter θ optimally

- single-event likelihood again

$$p(x|\theta) = \frac{1}{\sigma_{\text{tot}}(\theta)} \frac{d^m\sigma(x|\theta)}{dx^m}$$

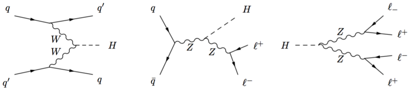
- expanded in θ around θ_0, define score

$$\log \frac{p(x|\theta)}{p(x|\theta_0)} \approx (\theta-\theta_0)\, \nabla_\theta \log p(x|\theta)\Big|_{\theta_0} \equiv (\theta-\theta_0)\, t(x|\theta_0) \equiv (\theta-\theta_0)\, \mathcal{O}^{\text{opt}}(x)$$

- parton level, as used in ATLAS [CPV]

$$p(x|\theta) \approx |\mathcal{M}|_0^2 + \theta|\mathcal{M}|_{\text{int}}^2 \qquad \Rightarrow \qquad t(x|\theta_0) \sim \frac{|\mathcal{M}|_{\text{int}}^2}{|\mathcal{M}|_0^2} \,,$$

$\Rightarrow$ Easy at parton level, LEP physics...



Discrete symmetry

- CPV at dimension-6 in WBF
- unique CP-observable [C-even, P-odd, $\hat{T}$-odd]

$$t \propto \epsilon_{\mu\nu\rho\sigma}\; k_1^\mu\; k_2^\nu\; q_1^\rho\; q_2^\sigma\; \text{sign}\,[(k_1-k_2)\cdot(q_1-q_2)] \xrightarrow{\text{lab frame}} \sin\Delta\phi_{jj}$$

$\Rightarrow$ Computable including prefactor

PySR

Analytic formula for score [M Cranmer (2020)]

- function to approximate $t(x|\theta)$
- order-one phase space parameters $x_p = p_T/m_H, \Delta\eta, \Delta\phi$ [node]
- operators $\sin x, x^2, x^3, x+y, x-y, x*y, x/y$ [node]
- represent formula as tree [complexity = number of nodes]

$\Rightarrow$ figures of merit

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}\Big[g_i(x) - t(x,z|\theta)\Big]^2$$

$$\to \text{MSE} + \text{parsimony} \cdot \text{complexity}$$

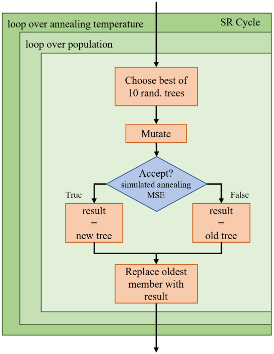
Simulated annealing

- combine trees to populations
- mutate trees exchange, add, delete nodes
- acceptance probability

$$p = \exp\left(-\frac{\text{MSE'}_\text{new} - \text{MSE'}_\text{old}}{\alpha T\ \text{MSE'}_\text{old}}\right)$$

- added: proper fit of pre-factors

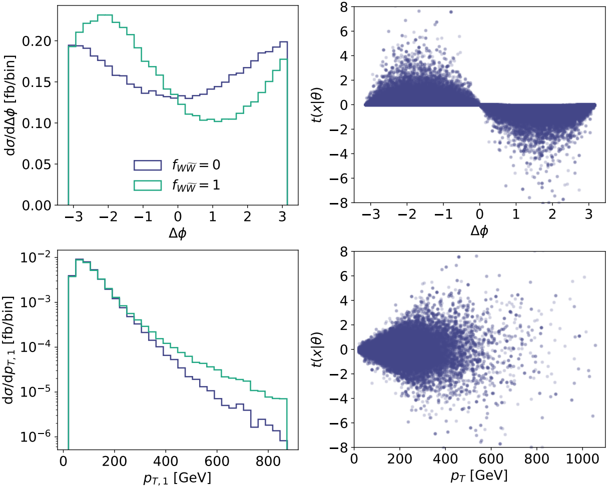
$\Rightarrow$ Hall of Fame: best equation per complexity



Score around Standard Model

Score around Standard Model [Brehmer, Butter, TP, Soybelman]

- shift in distributions, reflected in score [parton level]
  CP-effect in $\Delta\phi_{jj}$
  D6-effect in $p_{T,j}$



Score around Standard Model

Score around Standard Model [Brehmer, Butter, TP, Soybelman]

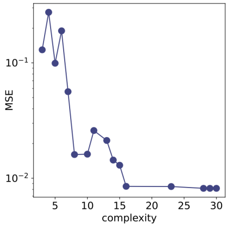
- shift in distributions, reflected in score [parton level]
  CP-effect in $\Delta\phi_{jj}$
  D6-effect in $p_{T,j}$
- best 4-parameter formula including $\Delta\eta$ [without/with detector]

$$t = -x_{p,1}\,(x_{p,2}+c)\,(a - b\Delta\eta)\sin(\Delta\phi + d)$$

with $a = 1.086(11)$ $b = 0.10241(19)$ $c = 0.24165(8)$ $d = 0.00662(32)$
$a = 0.926(2)$ $b = 0.08387(35)$ $c = 0.3542(20)$ $d = 0.00911(67)$

$\Rightarrow$ Mostly expected formula

compl	dof	function	MSE
3	1	$a\,\Delta\phi$	$1.30 \cdot 10^{-1}$
4	1	$\sin(a\Delta\phi)$	$2.75 \cdot 10^{-1}$
5	1	$a\Delta\phi x_{p,1}$	$9.93 \cdot 10^{-2}$
6	1	$-x_{p,1}\sin(\Delta\phi + a)$	$1.90 \cdot 10^{-1}$
7	1	$(-x_{p,1} - a)\sin(\sin(\Delta\phi))$	$5.63 \cdot 10^{-2}$
8	1	$(a - x_{p,1})x_{p,2}\sin(\Delta\phi)$	$1.61 \cdot 10^{-2}$
14	2	$x_{p,1}(a\Delta\phi - \sin(\sin(\Delta\phi)))(x_{p,2} + b)$	$1.44 \cdot 10^{-2}$
15	3	$-(x_{p,2}(a\Delta\eta^2 + x_{p,1}) + b)\sin(\Delta\phi + c)$	$1.30 \cdot 10^{-2}$
16	4	$-x_{p,1}(a - b\Delta\eta)(x_{p,2} + c)\sin(\Delta\phi + d)$	$8.50 \cdot 10^{-3}$
28	7	$(x_{p,2} + a)(bx_{p,1}(c - \Delta\phi) - x_{p,1}(d\Delta\eta + ex_{p,2} + f)\sin(\Delta\phi + g))$	$8.18 \cdot 10^{-3}$



Score around Standard Model

Score around Standard Model [Brehmer, Butter, TP, Soybelman]

- shift in distributions, reflected in score [parton level]
  CP-effect in $\Delta\phi_{jj}$
  D6-effect in $p_{T,j}$
- best 4-parameter formula including $\Delta\eta$ [without/with detector]

$$t = -x_{p,1}\,(x_{p,2} + c)\,(a - b\Delta\eta)\sin(\Delta\phi + d)$$

with $a = 1.086(11)$ $b = 0.10241(19)$ $c = 0.24165(8)$ $d = 0.00662(32)$
$a = 0.926(2)$ $b = 0.08387(35)$ $c = 0.3542(20)$ $d = 0.00911(67)$

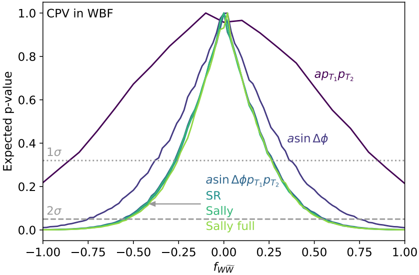
$\Rightarrow$ Mostly expected formula

Using the formula

- expected limits:

  very wrong formula
  wrong formula
  right formula
  MadMiner
- same within statistical limitation

$\Rightarrow$ Unknown optimal observables next



ML for LHC Theory

ML-applications in LHC physics

- just another numerical tool for a numerical field
- driven by money from data science, medical research
- goals are...

  ...improve established tasks
  ...develop new tools for established tasks
  ...transform through new ideas

$\to$ Turn HL-LHC into fun!

Machine Learning and LHC Event Generation

Anja Butter[1,2], Tilman Plehn[1], Steffen Schumann[3] (Editors), Simon Badger[4], Sascha Caron[5, 6] Kyle Cranmer[7,8], Francesco Armando Di Bello[9], Etienne Dreyer[10], Stefano Forte[11], Sanmay Ganguly[12], Dorival Gonçalves[13], Eilam Gross[10], Theo Heimel[1], Gudrun Heinrich[14], Lukas Heinrich[15], Alexander Held[16], Stefan Höche[17], Jessica N. Howard[18], Philip Ilten[19], Joshua Isaacson[17], Timo Janßen[3], Stephen Jones[20], Marumi Kado[9,21], Michael Kagan[22], Gregor Kasieczka[23], Felix Kling[24], Sabine Kraml[25], Claudius Krause[26], Frank Krauss[20], Kevin Kröninger[27], Rahool Kumar Barman[13], Michel Luchmann[1], Vitaly Magerya[14], Daniel Maitre[20], Bogdan Malaescu[2], Fabio Maltoni[28,29], Till Martini[30], Olivier Mattelaer[28], Benjamin Nachman[31,32], Sebastian Pitz[1], Juan Rojo[33,34], Matthew Schwartz[35], David Shih[25], Frank Siegert[36], Roy Stegeman[11], Bob Stienen[5], Jesse Thaler[37], Rob Verheyen[38], Daniel Whiteson[18], Ramon Winterhalder[28], and Jure Zupan[19]

Abstract

First-principle simulations are at the heart of the high-energy physics research program. They link the vast data output of multi-purpose detectors with fundamental theory predictions and interpretation. This review illustrates a wide range of applications of modern machine learning to event generation and simulation-based inference, including conceptional developments driven by the specific requirements of particle physics. New ideas and tools developed at the interface of particle physics and machine learning will improve the speed and precision of forward simulations, handle the complexity of collision data, and enhance inference as an inverse simulation problem.

Submitted to the Proceedings of the US Community Study
on the Future of Particle Physics (Snowmass)

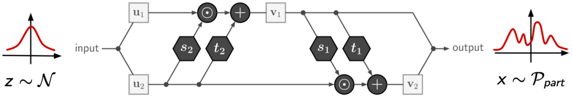
Contents

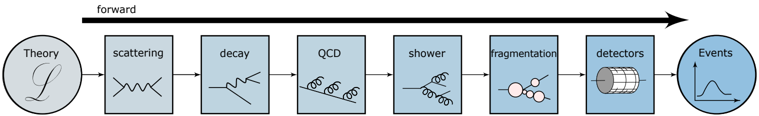
Modern generative networks

Normalizing flows — INN

- Gaussian latent space
- bijective mapping
- known Jacobian
- log-likelihood loss

$\rightarrow$ Better than VAEs and GANs



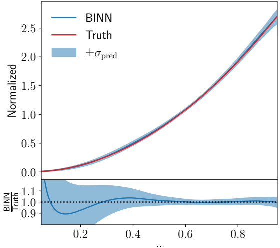


Modern generative networks

Normalizing flows — INN

- Gaussian latent space
- bijective mapping
- known Jacobian
- log-likelihood loss

$\rightarrow$ Better than VAEs and GANs



Bayesian INNs

- network weight distributions [Gal (2016)]
- sample for output [efficient ensembling]
- working for regression, classification
- events with error bars [density & uncertainty maps]
- 2D: wedge ramp, kicker ramp,...

$\rightarrow$ INNs just fits