Transforming Particle Theory with Modern ML

Tilman Plehn

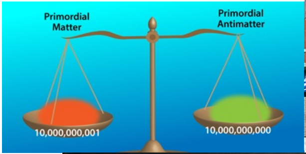
Universität Heidelberg

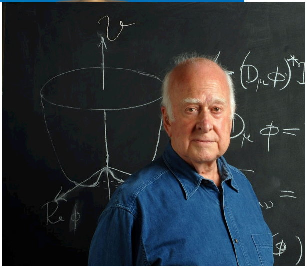
Rutgers, March 2024

Modern LHC physics

Classic motivation

- dark matter?
- baryogenesis?
- origin of Higgs field?





Modern LHC physics

Classic motivation

- dark matter?
- baryogenesis?
- origin of Higgs field?

Defining LHC physics

- fundamental motivation
- first-principle simulations
- huge data set
- uncertainty control

Modern LHC physics

Classic motivation

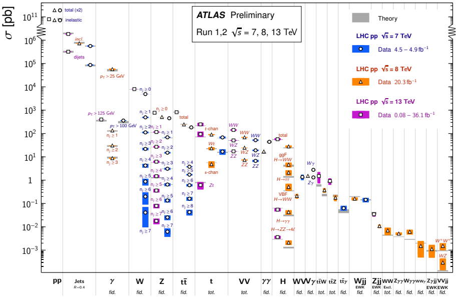
- dark matter?
- baryogenesis?
- origin of Higgs field?

Defining LHC physics

- fundamental motivation
- first-principle simulations
- huge data set
- uncertainty control

Glorious past

- measurements of event counts
- model-driven Higgs discovery
- vast analysis landscape



Modern LHC physics

Classic motivation

- dark matter?
- baryogenesis?
- origin of Higgs field?

Defining LHC physics

- fundamental motivation
- first-principle simulations
- huge data set
- uncertainty control

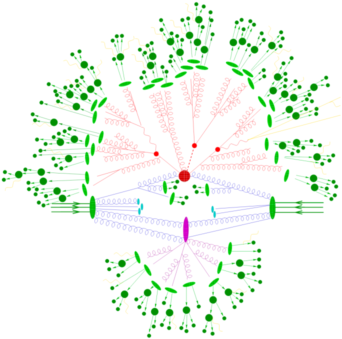
Glorious past

- measurements of event counts
- model-driven Higgs discovery
- vast analysis landscape

First-principle simulations

- Lagrangian/Hamiltonian
- quantum field theory calculation
- simulated collisions
- simulated detectors

→ LHC collisions in virtual worlds



Modern LHC physics

Classic motivation

- dark matter?
- baryogenesis?
- origin of Higgs field?

Defining LHC physics

- fundamental motivation
- first-principle simulations
- huge data set
- uncertainty control

Glorious past

- measurements of event counts
- model-driven Higgs discovery
- vast analysis landscape

First-principle simulations

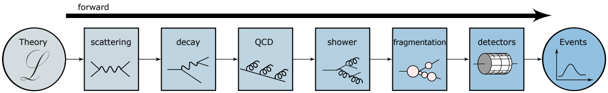
- Lagrangian/Hamiltonian
- quantum field theory calculation
- simulated collisions
- simulated detectors

→ LHC collisions in virtual worlds

Infer underlying theory

- simulations vs data
- symmetries the key
- phase space interpretable
- SM or BSM?

→ ML-case obvious [2203.07460]



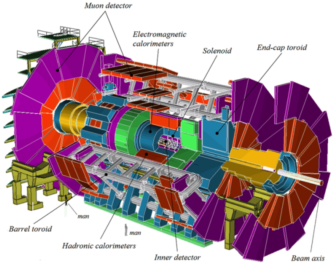
LHC data

Collaborations

- ATLAS & CMS general purpose
  LHCb, ALICE, FASER specialized
- 1000s of scientists per experiment

Detectors

- built around *pp* interaction point
- measuring outgoing particles
- collision rate 40 MHz



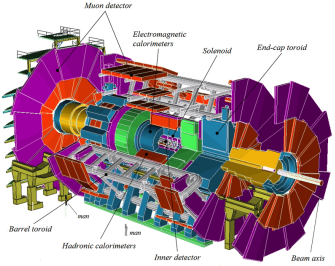
LHC data

Collaborations

- ATLAS & CMS general purpose
  LHCb, ALICE, FASER specialized
- 1000s of scientists per experiment

Detectors

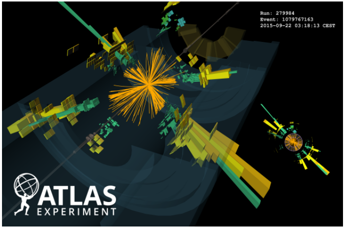
- built around *pp* interaction point
- measuring outgoing particles
- collision rate 40 MHz



Event format

- ATLAS event size 1.6 MB
  data stream 3 PB/s
- measure:
  energy, momentum, charge, etc
- electrons, muons easy
  quarks, gluons as jets [20-50 particles]

→ Event: 100+ ntuples $(E, \vec{p}, Q...)$



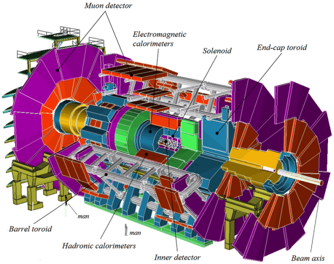
LHC data

Collaborations

- ATLAS & CMS general purpose
  LHCb, ALICE, FASER specialized
- 1000s of scientists per experiment

Detectors

- built around *pp* interaction point
- measuring outgoing particles
- collision rate 40 MHz



Event format

- ATLAS event size 1.6 MB
  data stream 3 PB/s
- measure:
  energy, momentum, charge, etc
- electrons, muons easy
  quarks, gluons as jets [20-50 particles]

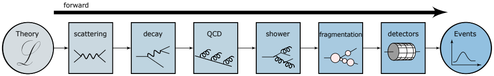
$\to$ Event: 100+ ntuples $(E, \vec{p}, Q...)$

ML applications

- data selection/compression
- object reconstruction
- object classification
- calibration
- analysis preprocessing

$\to$ Everything, faster and better

LHC Theory

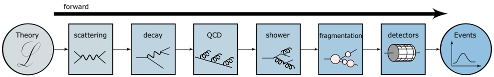


Turning data to knowledge

- Quantum Field Theory
  start with Lagrangian [Hamiltonian]
  generate Feynman diagrams
- compute hard scattering
  compute decays
  compute gluon radiation
- partons inside protons
  hadron-level QCD

$\rightarrow$ Simulations, not modeling

LHC Theory



Turning data to knowledge

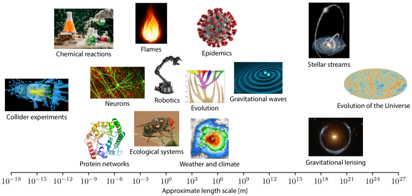
- Quantum Field Theory
  start with Lagrangian [Hamiltonian]
  generate Feynman diagrams
- compute hard scattering
  compute decays
  compute gluon radiation
- partons inside protons
  hadron-level QCD

$\rightarrow$ Simulations, not modeling

HL-LHC: inference with 20$\times$ more data

- SBI starts with Simulation...
- statistical improvement $\sqrt{20} = 4.5$
- rate over phase space to $< 0.1\%$
- theory to follow

$\rightarrow$ precision = QFT * CPU



Ask a data physicist

LHC questions

- How to get from 3 PB/s to 300 MB/s?
  Data compression [Netflix]

Ask a data physicist

LHC questions

- How to get from 3 PB/s to 300 MB/s?
  Data compression [Netflix]
- How to analyze events as ntuples?
  Graph neural networks [Tesla]

Ask a data physicist

LHC questions

- How to get from 3 PB/s to 300 MB/s?
  Data compression [Netflix]
- How to analyze events as ntuples?
  Graph neural networks [Tesla]
- How to incorporate symmetries?
  Covariant networks [Qualcomm]

Ask a data physicist

LHC questions

- How to get from 3 PB/s to 300 MB/s?
  Data compression [Netflix]
- How to analyze events as ntuples?
  Graph neural networks [Tesla]
- How to incorporate symmetries?
  Covariant networks [Qualcomm]
- How to incorporate more symmetries?
  Contrastive learning [Google]

Ask a data physicist

LHC questions

- How to get from 3 PB/s to 300 MB/s?
  Data compression [Netflix]
- How to analyze events as ntuples?
  Graph neural networks [Tesla]
- How to incorporate symmetries?
  Covariant networks [Qualcomm]
- How to incorporate more symmetries?
  Contrastive learning [Google]
- How to combine different detectors?
  Super-resolution [Gaming]

Ask a data physicist

LHC questions

- How to get from 3 PB/s to 300 MB/s?
  Data compression [Netflix]
- How to analyze events as ntuples?
  Graph neural networks [Tesla]
- How to incorporate symmetries?
  Covariant networks [Qualcomm]
- How to incorporate more symmetries?
  Contrastive learning [Google]
- How to combine different detectors?
  Super-resolution [Gaming]
- How to learn correlations?
  Transformers [OpenAI]

Ask a data physicist

LHC questions

- How to get from 3 PB/s to 300 MB/s?
  Data compression [Netflix]
- How to analyze events as ntuples?
  Graph neural networks [Tesla]
- How to incorporate symmetries?
  Covariant networks [Qualcomm]
- How to incorporate more symmetries?
  Contrastive learning [Google]
- How to combine different detectors?
  Super-resolution [Gaming]
- How to learn correlations?
  Transformers [OpenAI]
- How to look for new physics?
  Autoencoders [SAP]

Ask a data physicist

LHC questions

- How to get from 3 PB/s to 300 MB/s?
  Data compression [Netflix]
- How to analyze events as ntuples?
  Graph neural networks [Tesla]
- How to incorporate symmetries?
  Covariant networks [Qualcomm]
- How to incorporate more symmetries?
  Contrastive learning [Google]
- How to combine different detectors?
  Super-resolution [Gaming]
- How to learn correlations?
  Transformers [OpenAI]
- How to look for new physics?
  Autoencoders [SAP]

$\rightarrow$ Uncertainties, Inference, Understanding? [Physics-xAI]

Shortest ML-intro ever

Fit-like approximation

- approximate $f_\theta(x) \approx f(x)$
- no parametrization, just very many θ
- new representation/latent space θ

Training

- minimize loss to find best θ
- back propagation for gradient

Applications

- regression $x \to f_\theta(x)$
- classification $x \to f_\theta(x) \in [0,1]$
- generation $r \sim \mathcal{N} \to f_\theta(r)$

Architecture

- optimized input and data format
- structures, like symmetries or locality
- mostly, images vs language

$\to$ Transforming numerical science and everything

Learning by minimizing

Learning energy $E(x)$

- maximize probability of θ-encoding

$$p(\theta|E_\text{train}) = \frac{p(E_\text{train}|\theta)\; p(\theta)}{p(E_\text{train})} \approx p(\theta|E_\text{train})\; p(\theta)$$

- assume Gaussians and neglect error

$$\mathcal{L} \equiv -\log p(\theta|E_\text{train}) = \frac{(E_\theta - E_\text{train})^2}{2\sigma_E^2} + \frac{(\theta-\theta_0)^2}{2\sigma_\theta^2} \approx (E_\theta - E_\text{train})^2$$

$\to$ MSE loss mimimization

Learning by minimizing

Learning energy $E(x)$

- maximize probability of θ-encoding

$$p(\theta|E_\text{train}) = \frac{p(E_\text{train}|\theta)\; p(\theta)}{p(E_\text{train})} \approx p(\theta|E_\text{train})\; p(\theta)$$

- assume Gaussians and neglect error

$$\mathcal{L} \equiv -\log p(\theta|E_\text{train}) = \frac{(E_\theta - E_\text{train})^2}{2\sigma_E^2} + \frac{(\theta-\theta_0)^2}{2\sigma_\theta^2} \approx (E_\theta - E_\text{train})^2$$

$\to$ MSE loss mimimization

Physics: energy probability [Gal (2016)]

- expectation value from (learnd) probability

$$\langle E \rangle = \int dE\; E\; p(E)$$

- complete internal representation θ

$$\langle E \rangle = \int dE\; E \int d\theta\; p(E|\theta)\; p(\theta|E_\text{train})$$

- maximum probability $\to$ latent probability

$$\int d\theta\; p(E|\theta)\; p(\theta|E_\text{train}) \approx \int d\theta\; p(E|\theta)\; q(\theta)$$

Learning by minimizing

Physics: energy probability [Gal (2016)]

- expectation value from (learnd) probability

$$\langle E \rangle = \int dE \; E \; p(E)$$

- complete internal representation θ

$$\langle E \rangle = \int dE \; E \int d\theta \; p(E|\theta) \; p(\theta|E_\text{train})$$

- maximum probability $\to$ latent probability

$$\int d\theta \; p(E|\theta) \; p(\theta|E_\text{train}) \approx \int d\theta \; p(E|\theta) \; q(\theta)$$

- similar for minimal KL-divergence [optimal transport]

$$\begin{aligned} D_\text{KL}[q(\theta), p(\theta|E_\text{train})] &\equiv \int d\theta \; q(\theta) \; \log \frac{q(\theta)}{p(\theta|E_\text{train})} \\ &= \int d\theta \; q(\theta) \; \log \frac{q(\theta) p(E_\text{train})}{p(E_\text{train}|\theta) p(\theta)} \\ &= - \int d\theta \; q(\theta) \; \log p(E_\text{train}|\theta) + \int d\theta \; q(\theta) \log \frac{q(\theta)}{p(\theta)} + \cdots \end{aligned}$$

$\to$ Bayesian NN: sampling θ for uncertainty

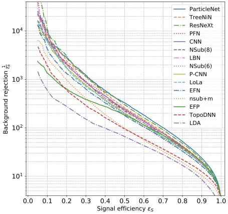
$$\mathcal{L} = - \int d\theta \; q(\theta) \; \log p(E_\text{train}|\theta) + D_\text{KL}[q(\theta), p(\theta)]$$

ML in experiment

Top tagging [classification, 2016-2019]

- 'hello world' of LHC-ML
- end of QCD-taggers
- different NN-architectures

$\rightarrow$ Non-NN left in the dust...



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]

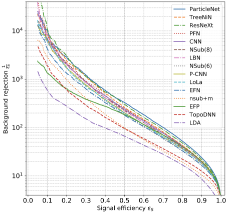
1 Institut für Experimentalphysik, Universität Hamburg, Germany
2 Institut für Theoretische Physik, Universität Heidelberg, Germany
3 Center for Cosmology and Particle Physics and Center for Data Science, NYU, USA
4 NHECT, Dept. of Physics and Astronomy, Rutgers, The State University of NJ, USA
5 Jozef 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 in experiment

Top tagging [classification, 2016-2019]

- ‘hello world’ of LHC-ML
- end of QCD-taggers
- different NN-architectures

$\rightarrow$ Non-NN left in the dust...



SciPost Physics | Submission

The Machine Learning Landscape of Top Taggers

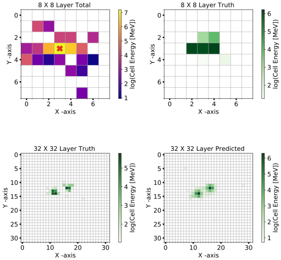
G. Kasieczka (ed)[1], T. Plehn (ed)[2], A. Butter[2], K. Cranmer[3], D. Debnath[4], B. M. Dillon[5], M. Fairbairn[6], D. A. Faroughy[5], W. Fedorko[7], C. Gay[7], L. Gouskos[8], J. F. Kamenik[5,9], P. T. Komiske[10], S. Leiss[1], A. Lister[7], S. Macaluso[3,4], E. M. Metodiev[10], L. Moore[11], B. Nachman,[12,13], 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 [2003.08863]

- mother of jet tools
- combined detector channels
- similar studies in CMS

$\rightarrow$ Modern jet-analysis tool



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

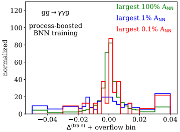
Fig. 7: An event display of total energy shower (within topocluster), as captured by a calorimeter layer of 8×8 granularity, along with the location of the track, denoted by a red cross (left) and the same shower is captured by a calorimeter layer of 32×32 granularity (middle). The bottom right panel 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.

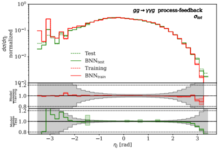
ML in phenomenology

Speeding up amplitudes [regression, Sherpa, Madgraph,...]

- loop-amplitudes expensive
- training fit or interpolation

$\to$ Precision NN-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.

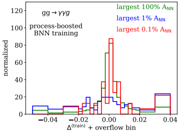
ML in phenomenology

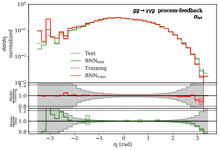
Speeding up amplitudes [regression, Sherpa, Madgraph,...]

· loop-amplitudes expensive

· training fit or interpolation

→ Precision NN-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.

NNPDF/N3PDF parton densities [Forte etal, since 2002]

· starting point: pdfs without functional ansatz

· moving on: cutting-edge ML everywhere

→ Leaders in ML-theory

N3PDF Machine Learning • PDFs • QCD

Home About Team Jobs Research Deliverables Documents For the public

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.

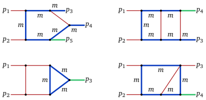
PACS. 12.38.-t Quantum chromodynamics – 12.39.-x Phenomenological quark models – 84.35.+i Neural Networks

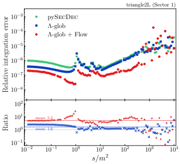
ML in theory

Optimizing integration paths [invertible networks]

- find optimal integration paths
- learn variable transformation

→ Theory-integrator





SciPost | SciPost Phys. 12, 129 (2022)

Targeting multi-loop integrals with neural networks

Ramon Winterhalder[1,2,3], Vitaly Magerya[4], Emilio Villa[4], Stephen P. Jones[5], Matthias Kerner[4,6], Anja Butter[1,2], Gudrun Heinrich[2,4] and Tilman Plehn[1,2]

1 Institut für Theoretische Physik, Universität Heidelberg, Germany
2 HEiKA - Heidelberg Karlsruhe Strategic Partnership, Heidelberg University, Karlsruhe Institute of Technology (KIT), Germany
3 Centre for Cosmology, Particle Physics and Phenomenology (CP3), Université catholique de Louvain, Belgium
4 Institut für Theoretische Physik, Karlsruher Institut für Technologie, Germany
5 Institute for Particle Physics Phenomenology, Durham University, UK
6 Institut für Astroteilchenphysik, Karlsruher Institut für Technologie, Germany

Abstract

Numerical evaluations of Feynman integrals often proceed via a deformation of the integration contour into the complex plane. While valid contours are easy to construct, the numerical precision for a multi-loop integral can depend critically on the chosen contour. We present methods to optimize this contour using a combination of optimized, global complex shifts and a normalizing flow. They can lead to a significant gain in precision.

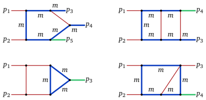
ML in theory

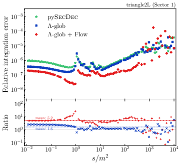
Optimizing integration paths [invertible networks]

· find optimal integration paths

· learn variable transformation

→ Theory-integrator





SciPost | SciPost Phys. 12, 129 (2022)

Targeting multi-loop integrals with neural networks

Ramon Winterhalder[1,2,3], Vitaly Magerya[4], Emilio Villa[4], Stephen P. Jones[5], Matthias Kerner[4,6], Anja Butter[1,2], Gudrun Heinrich[2,4] and Tilman Plehn[1,2]

1 Institut für Theoretische Physik, Universität Heidelberg, Germany
2 HEiKA - Heidelberg Karlsruhe Strategic Partnership, Heidelberg University, Karlsruhe Institute of Technology (KIT), Germany
3 Centre for Cosmology, Particle Physics and Phenomenology (CP3), Université catholique de Louvain, Belgium
4 Institut für Theoretische Physik, Karlsruher Institut für Technologie, Germany
5 Institute for Particle Physics Phenomenology, Durham University, UK
6 Institut für Astroteilchenphysik, Karlsruher Institut für Technologie, Germany

Abstract

Numerical evaluations of Feynman integrals often proceed via a deformation of the integration contour into the complex plane. While valid contours are easy to construct, the numerical precision for a multi-loop integral can depend critically on the chosen contour. We present methods to optimize this contour using a combination of optimized, global complex shifts and a normalizing flow. They can lead to a significant gain in precision.

Navigating string landscape [reinforcement learning]

· searching for viable vacua

· high dimensions, unknown global structure

→ Model space sampling

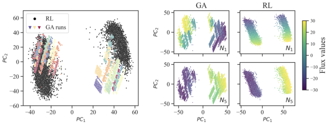


Figure 1: *Left:* Cluster structure in dimensionally reduced flux samples for RL and 25 GA runs (PCA on all samples of GA and RL). The colors indicate individual GA runs. *Right:* Dependence on flux (input) values (N_3 and N_5 respectively) in relation to principal components for a PCA fit of the individual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

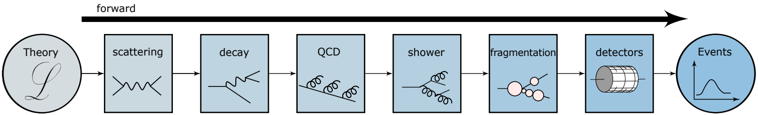
Abstract

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.

Generative networks for optimal inference

Forward simulations [Butter, Hütsch, Palacios Schweitzer, TP, Spinner]

- learn phase space density
  sample Gaussian $\rightarrow$ phase space
- Variational Autoencoder
  $\rightarrow$ low-dimensional physics, high-dimensional objects
- Generative Adversarial Network [Berkeley-Heidelberg]
  $\rightarrow$ generator trained by classifier
- Normalizing Flow/INN/Diffusion [Rutgers-Heidelberg]
  $\rightarrow$ bijective mapping
- JetGPT
  $\rightarrow$ learning all structures



Generative networks for optimal inference

Forward simulations [Butter, Hütsch, Palacios Schweitzer, TP, Spinner]

- learn phase space density
  sample Gaussian → phase space
- Variational Autoencoder
  → low-dimensional physics, high-dimensional objects
- Generative Adversarial Network [Berkeley-Heidelberg]
  → generator trained by classifier
- Normalizing Flow/INN/Diffusion [Rutgers-Heidelberg]
  → bijective mapping
- JetGPT
  → learning all structures

Use case and fundamental questions

- train on first-principle simulations [training on data: David]

  speed up generation/simulaton
  efficient way to ship data
  bridge simulation-reality gap
- GANplification [Berkeley-Hamburg-Heidelberg]

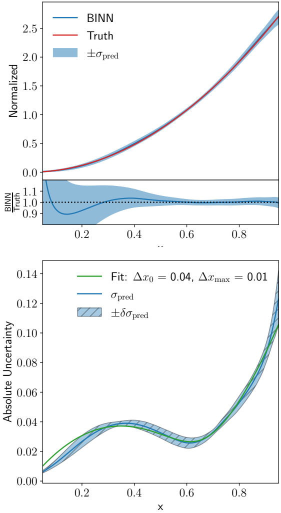
  initial data reproducing training sample
  more data from fit/interpolation
  too much data reproducing statistical fluctiations?

Generative networks with uncertainties

Bayesian generative networks [Bellagente, Haussmann, Luchmann, TP]

- network weight distributions for density
- sampling phase space
  events with error bars on weights
- learned density & uncertainty
  reflecting network learning?

$\rightarrow$ INNs like fitted functions



Generative networks with uncertainties

Bayesian generative networks [Bellagente, Haussmann, Luchmann, TP]

- network weight distributions for density
- sampling phase space
  events with error bars on weights
- learned density & uncertainty
  reflecting network learning?

$\rightarrow$ INNs like fitted functions

LHC events with uncertainties [Heimel, Vent...]

- ntuples for two muons and 1-3 jets
- classifier weight [check and reweight]

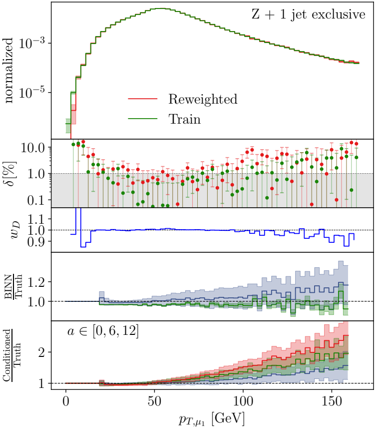
$$w_D(x_i) = \frac{D(x_i)}{1 - D(x_i)} = \frac{p_{\text{data}}(x_i)}{p_{\text{model}}(x_i)}$$

- systematics in training data

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

- sampling a through conditional INN

$\rightarrow$ Precision and uncertainty control

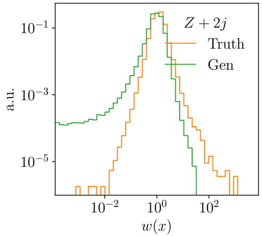


Generative networks with uncertainties

Bayesian generative networks [Bellagente, Haussmann, Luchmann, TP]

- network weight distributions for density
- sampling phase space
  events with error bars on weights
- learned density & uncertainty
  reflecting network learning?

$\rightarrow$ INNs like fitted functions



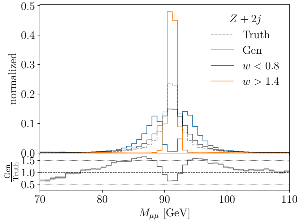
Quality control [Das, Favaro, Heimel, Krause, TP, Shih]

- classifier easier to train
- training vs generated

$$w(x_i) = \frac{D(x_i)}{1 - D(x_i)} = \frac{p_{\text{train}}(x_i)}{p_{\text{gen}}(x_i)}$$

- $w(x_i) \gg 1$ too little generated
  $w(x_i) \ll 1$ too much generated
- precision from width of distribution

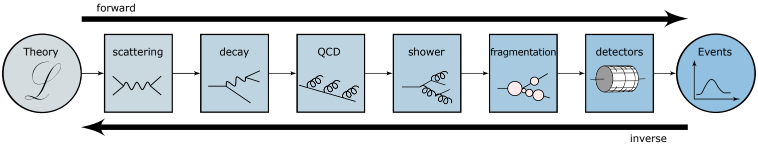
$\rightarrow$ Systematic benchmarking



Inverse simulation for optimal inference

Invertible ML-simulation [Bellagente, Butter, Kasieczka, TP, Winterhalder]

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

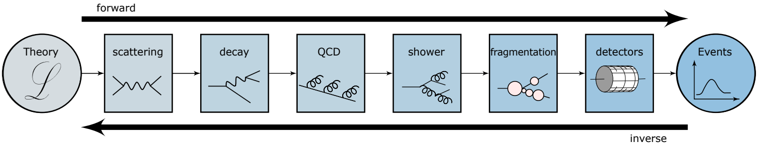


Inverse simulation for optimal inference

Invertible ML-simulation [Bellagente, Butter, Kasieczka, TP, Winterhalder]

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

detector unfolding
unfolding to QCD partons - jet algorithm
unfolding jet radiation - jet combinatorics
unfolding to hard process - top analyses
matrix element method an old dream



Inverse simulation for optimal inference

Invertible ML-simulation [Bellagente, Butter, Kasieczka, TP, Winterhalder]

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

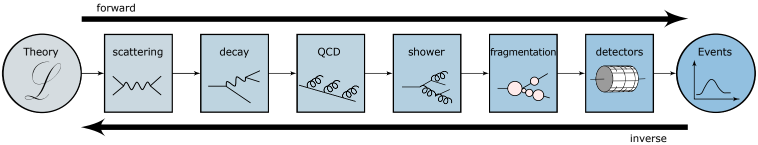
  detector unfolding
  unfolding to QCD partons - jet algorithm
  unfolding jet radiation - jet combinatorics
  unfolding to hard process - top analyses
  matrix element method an old dream

1- reweighting [Omnifold]

2- distribution mapping [Schrödinger bridge, Direct Diffusion]

3- generative unfolding [cINN, CFM]

$\to$ Transformative progress for HL-LHC

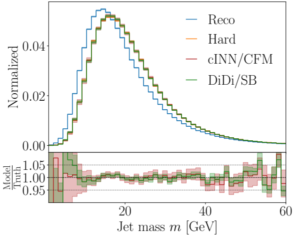


Unfolding

Detector unfolding [Heidelberg-Berkeley-Irvine]

- compare to theory without detector
- analyse data with public tools
- example: quark/gluon jets
- measure QCD splittings and α_s search for light dark matter

$\to$ All methods at per-cent level

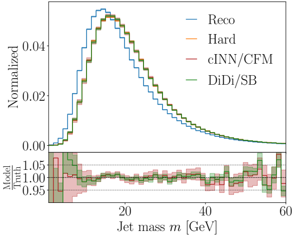


Unfolding

Detector unfolding [Heidelberg-Berkeley-Irvine]

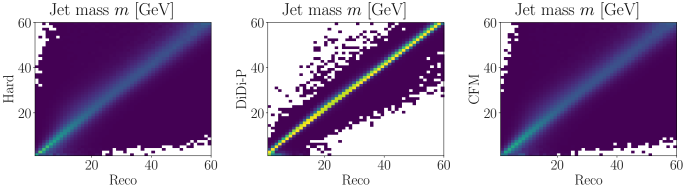
- compare to theory without detector
- analyse data with public tools
- example: quark/gluon jets
- measure QCD splittings and α_s
  search for light dark matter

$\rightarrow$ All methods at per-cent level



Kinematic migration [Hütsch, Villadamigo]

- forward detector simulation as reference
- learned mapping from DiDi
  learned mapping from generative unfolding

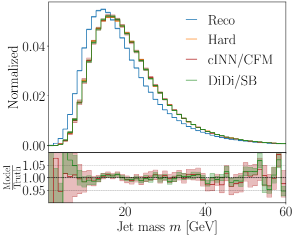


Unfolding

Detector unfolding [Heidelberg-Berkeley-Irvine]

- compare to theory without detector
- analyse data with public tools
- example: quark/gluon jets
- measure QCD splittings and α_s
  search for light dark matter

$\rightarrow$ All methods at per-cent level



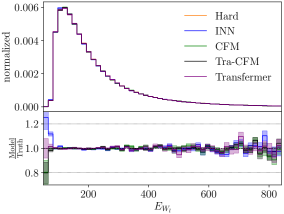
Kinematic migration [Hütsch, Villadamigo]

- forward detector simulation as reference
- learned mapping from DiDi
  learned mapping from generative unfolding

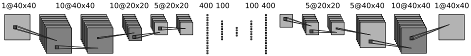
Unfolding to partons

- event kinematics in SMEFT
- example: $t\bar{t}$ production
- search for new particles in kinematics

$\rightarrow$ All methods at per-cent level



Anomaly searches

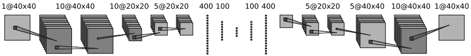


Non-resonant searches

- key: bottleneck [Rutgers-Heideberg]
  training on background
  minimize reconstruction-MSE
  unknown signal from bad MSE
- reconstruct QCD jets $\rightarrow$ top jets hard to describe
- reconstruct top jets $\rightarrow$ QCD jets just simple top-like jet

$\rightarrow$ Symmetric performance $S \leftrightarrow B$?

Anomaly searches



Non-resonant searches

- key: bottleneck [Rutgers-Heideberg]
  training on background
  minimize reconstruction-MSE
  unknown signal from bad MSE
- reconstruct QCD jets $\rightarrow$ top jets hard to describe
- reconstruct top jets $\rightarrow$ QCD jets just simple top-like jet

$\rightarrow$ Symmetric performance $S \leftrightarrow B$?

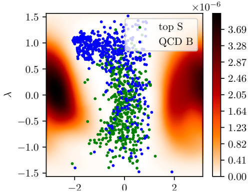
Missing and anomalous features

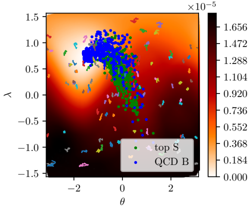
- compact latent space: sphere
- energy-based model
  normalized Boltzmann mapping [E_θ =MSE]

$$p_\theta(x) = \frac{e^{-E_\theta(x)}}{Z_\theta}$$

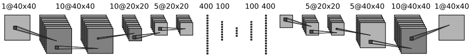
$$\mathcal{L} = -\left\langle \log p_\theta(x) \right\rangle = \left\langle E_\theta(x) + \log Z_\theta \right\rangle$$

- inducing background metric
- Z_θ from Markov Chain





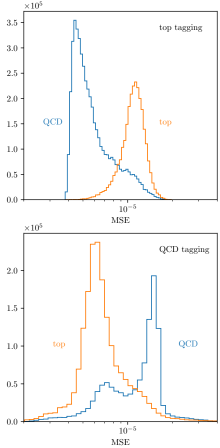
Anomaly searches



Non-resonant searches

- key: bottleneck [Rutgers-Heideberg]
  training on background
  minimize reconstruction-MSE
  unknown signal from bad MSE
- reconstruct QCD jets $\rightarrow$ top jets hard to describe
- reconstruct top jets $\rightarrow$ QCD jets just simple top-like jet

$\rightarrow$ Symmetric performance $S \leftrightarrow B$?



Missing and anomalous features

- compact latent space: sphere
- energy-based model
  normalized Boltzmann mapping [E_θ =MSE]

$$p_\theta(x) = \frac{e^{-E_\theta(x)}}{Z_\theta}$$

$$\mathcal{L} = -\Big\langle \log p_\theta(x) \Big\rangle = \Big\langle E_\theta(x) + \log Z_\theta \Big\rangle$$

- inducing background metric
- Z_θ from Markov Chain

$\rightarrow$ Proper anomaly search, at last [For more, ask David!]

ML for LHC Theory

ML-applications

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

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

  ...precision control
  ...uncertainties
  ...phase space
  ...symmetries
  ...formulas

$\rightarrow$ Opportunities!!

arXiv:2211.01421v2 [hep-ph] 17 Mar 2024

Modern Machine Learning for LHC Physicists

Tilman Plehn[a*], Anja Butter[a,b], Barry Dillon[a],
Theo Heimel[a], Claudius Krause[c], and Ramon Winterhalder[d]

[a] Institut für Theoretische Physik, Universität Heidelberg, Germany
[b] LPNHE, Sorbonne Université, Université Paris Cité, CNRS/IN2P3, Paris, France
[c] HEPHY, Austrian Academy of Sciences. Vienna, Austria
[d] CP3, Université catholique de Louvain, Louvain-la-Neuve, Belgium

March 19, 2024

Abstract

Modern machine learning is transforming particle physics fast, bullying its way into our numerical tool box. For young researchers it is crucial to stay on top of this development, which means applying cutting-edge methods and tools to the full range of LHC physics problems. These lecture notes lead students with basic knowledge of particle physics and significant enthusiasm for machine learning to relevant applications. They start with an LHC-specific motivation and a non-standard introduction to neural networks and then cover classification, unsupervised classification, generative networks, and inverse problems. Two themes defining much of the discussion are well-defined loss functions and uncertainty-aware networks. As part of the applications, the notes include some aspects of theoretical LHC physics. All examples are chosen from particle physics publications of the last few years.[1]