Transforming Particle Physics

Tilman Plehn

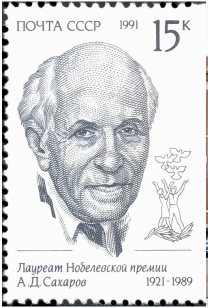
Universität Heidelberg

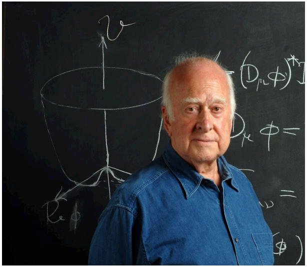
Genova, December 2024

All about LHC physics

Classic motivation

- dark matter?
- matter vs antimatter?
- origin of Higgs boson?





All about LHC physics

Classic motivation

- dark matter?
- matter vs antimatter?
- origin of Higgs boson?

LHC physics

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

Successful past

- measurements of total rates
- analyses inspired by simulation
- model-driven Higgs discovery

All about LHC physics

Classic motivation

- dark matter?
- matter vs antimatter?
- origin of Higgs boson?

LHC physics

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

Successful past

- measurements of total rates
- analyses inspired by simulation
- model-driven Higgs discovery

First-principle, precision simulations

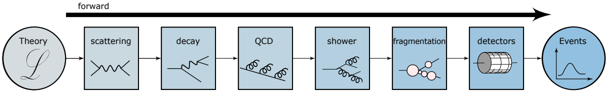
- start with Lagrangian
- calculate scattering using QFT
- simulate collisions
- simulate detectors

$\to$ LHC collisions in virtual worlds

BSM searches

- compare simulations and data
- infer underlying theory [SM or BSM]
- publish useable results

$\to$ Understand LHC data systematically



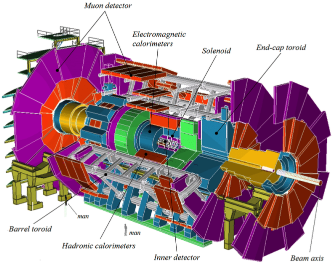
LHC data

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



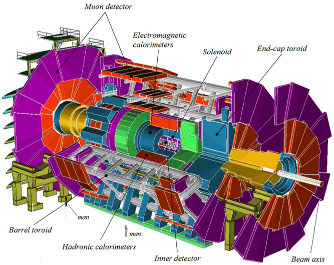
LHC data

Collaborations

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

Detectors

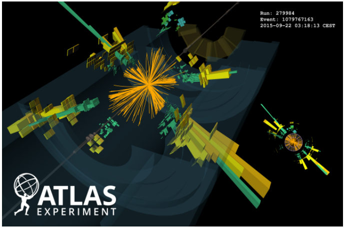
- built around *pp* interaction point
- measuring outgoing particles
- collision rate 40 MHz



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]

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



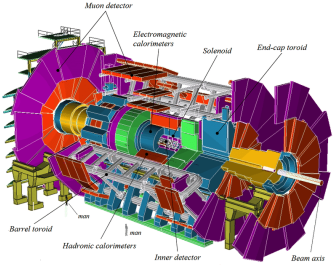
LHC data

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]

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

ML applications

- data selection/compression
- object reconstruction
- object classification
- analysis
- simulation
- theory calculations
- event generation
- inference

→ Everything, faster and better

2024 Nobel prize

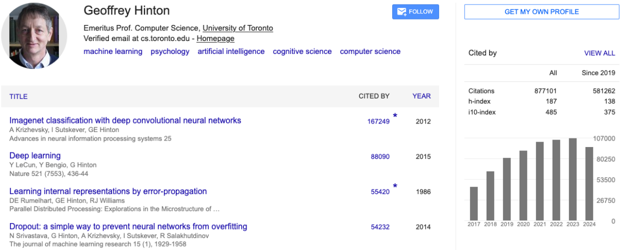
Neural networks using physics

- John Hopfield [physicist in search of problems]
  biological system and spin system with same mathematics
  paper for: physicists, computer scientists, neurobiologists

→ Physics is a point of view

- Geoffrey Hinton [easily passing as a physicist]
  minimal energy with noisy neurons
  log-likelihood maximization

→ Nobel prize in biology/medicine for thermal equilibrium?



2024 Nobel prize

Neural networks using physics

- John Hopfield [physicist in search of problems]
  biological system and spin system with same mathematics
  paper for: physicists, computer scientists, neurobiologists

$\rightarrow$ Physics is a point of view

- Geoffrey Hinton [easily passing as a physicist]
  minimal energy with noisy neurons
  log-likelihood maximization

$\rightarrow$ Nobel prize in biology/medicine for thermal equilibrium?

Physics using neural networks

- applications all over experiment
- regression $x \rightarrow f_\theta(x)$
- classification $x \rightarrow f_\theta(x) \in [0, 1]$
- generation $r \sim \mathcal{N} \rightarrow f_\theta(r)$
- conditional generation $r \sim \mathcal{N} \rightarrow f_\theta(r|x)$

$\rightarrow$ Complexity a feature, not a problem

Formulas for theory self-respect

Encoding a particle energy

- expectation value from probability

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

- internal representation θ

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

- training a generalization of θ-probability

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

Formulas for theory self-respect

Encoding a particle energy

- expectation value from probability

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

- internal representation θ

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

- training a generalization of θ-probability

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

- similarity from minimal KL-divergence

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

$\to$ Ultimate simplification and more...

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

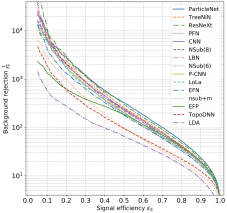
$$\to (E_\theta - E_\text{train})^2 + c(\theta - \theta_0)^2$$

ML in experiment

Top tagging [classification, 2016-today, good old BOOST days with S Marzani]

- ‘hello world’ of LHC-ML
- end of QCD-taggers
- ever-improving [Huilin Qu]

→ Driving NN-architectures

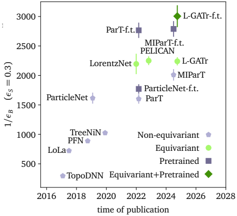


SciPost Physics | Submission

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

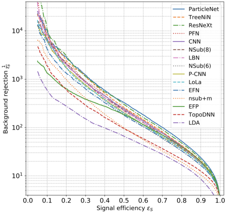


ML in experiment

Top tagging [classification, 2016-today, good old BOOST days with S Marzani]

- ‘hello world’ of LHC-ML
- end of QCD-taggers
- ever-improving [Huilin Qu]

→ Driving NN-architectures



SciPost Physics Submission

The Machine Learning Landscape of Top Taggers

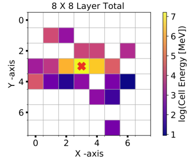
G. Kasieczka (ed)[1], T. Plehn (ed)[2], A. Butter[2], K. Cranmer[3], D. Debnath[4], B. M. Dillon[5], M. Fairbairn[6], D. A. Faroughy[5], W. Fedorko[7], C. Gay[7], L. Gouskos[8], J. F. Kamenik[5,9], P. T. Komiske[10], S. Leiss[1], A. Lister[7], S. Macaluso[3,4], E. M. Metodiev[10], L. Moore[11], B. Nachman,[12,13], 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 [2020-today]

- mother of jet analyses
- combining detectors with different resolution
- optimality the key

→ Modern jet analysis basics



Towards a Computer Vision Particle Flow *

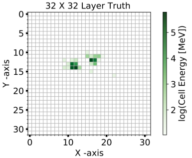
Francesco Armando Di Bello[a,3], Sanmay Ganguly[b,1], Eilam Gross[1], Marumi Kado[3,4], Michael Pitt[2], Lorenzo Santi [3], Jonathan Shlomi[1]

[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

Progress towards an improved particle flow algorithm at CMS with machine learning

Farouk Mokhtar[1], Joosep Pata[2], Javier Duarte[1], Eric Wulff[3], Maurizio Pierini[3] and Jean-Roch Vlimant[4]
(on behalf of the CMS Collaboration)
[1]University of California San Diego, La Jolla, CA 92093, USA
[2]NICPB, Rävala pst 10, 10143 Tallinn, Estonia
[3]European Organization for Nuclear Research (CERN), CH 1211, Geneva 23, Switzerland
[4]California Institute of Technology, Pasadena, CA 91125, USA

E-mail: fmokhtar@ucsd.edu, joosep.pata@cern.ch, jduarte@ucsd.edu

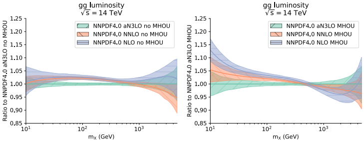


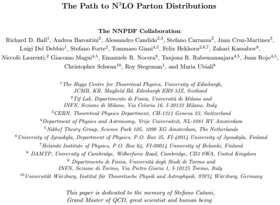
ML in phenomenology

Parton densities [NNPDF, 2002-today]

- pdfs without functional bias and full uncertainties
- precision and calibrated uncertainties

→ Drivers of ML-theory





The Path to N^{3}LO Parton Distributions

The NNPDF Collaboration:
Richard D. Ball[1], Andrea Barontini[2], Alessandro Candido[2,3], Stefano Carrazza[2], Juan Cruz-Martinez[3], Luigi Del Debbio[1], Stefano Forte[2], Tommaso Giani[4,5], Felix Hekhorn[2,6,7], Zahari Kassabov[8], Niccolò Laurenti,[2] Giacomo Magni[4,5], Emanuele R. Nocera[9], Tanjona R. Rabemananjara[4,5], Juan Rojo[4,5], Christopher Schwan[10], Roy Stegeman[1], and Maria Ubiali[8]

[1] *The Higgs Centre for Theoretical Physics, University of Edinburgh, JCMB, KB, Mayfield Rd, Edinburgh EH9 3JZ, Scotland*
[2] *Tif Lab, Dipartimento di Fisica, Università di Milano and INFN, Sezione di Milano, Via Celoria 16, I-20133 Milano, Italy*
[3] *CERN, Theoretical Physics Department, CH-1211 Geneva 23, Switzerland*
[4] *Department of Physics and Astronomy, Vrije Universiteit, NL-1081 HV Amsterdam*
[5] *Nikhef Theory Group, Science Park 105, 1098 XG Amsterdam, The Netherlands*
[6] *University of Jyvaskyla, Department of Physics, P.O. Box 35, FI-40014 University of Jyvaskyla, Finland*
[7] *Helsinki Institute of Physics, P.O. Box 64, FI-00014 University of Helsinki, Finland*
[8] *DAMTP, University of Cambridge, Wilberforce Road, Cambridge, CB3 0WA, United Kingdom*
[9] *Dipartimento di Fisica, Università degli Studi di Torino and INFN, Sezione di Torino, Via Pietro Giuria 1, I-10125 Torino, Italy*
[10] *Universität Würzburg, Institut für Theoretische Physik und Astrophysik, 97074 Würzburg, Germany*

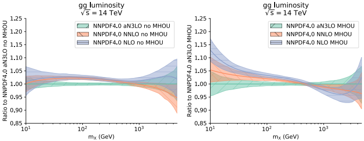
This paper is dedicated to the memory of Stefano Catani, Grand Master of QCD, great scientist and human being

ML in phenomenology

Parton densities [NNPDF, 2002-today]

- pdfs without functional bias and full uncertainties
- precision and calibrated uncertainties

→ Drivers of ML-theory



The Path to N^{3}LO Parton Distributions

The NNPDF Collaboration:
Richard D. Ball[1], Andrea Barontini[2], Alessandro Candido[2,3], Stefano Carrazza[2], Juan Cruz-Martinez[3], Luigi Del Debbio[1], Stefano Forte[2], Tommaso Giani[4,5], Felix Hekhorn[2,6,7], Zahari Kassabov[8], Niccolò Laurenti,[2] Giacomo Magni[4,5], Emanuele R. Nocera[9], Tanjona R. Rabemananjara[4,5], Juan Rojo[4,5], Christopher Schwan[10], Roy Stegeman[1], and Maria Ubiali[8]

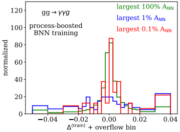
[1] The Higgs Centre for Theoretical Physics, University of Edinburgh, JCMB, KB, Mayfield Rd, Edinburgh EH9 3JZ, Scotland
[2] Tif Lab, Dipartimento di Fisica, Università di Milano and INFN, Sezione di Milano, Via Celoria 16, I-20133 Milano, Italy
[3] CERN, Theoretical Physics Department, CH-1211 Geneva 23, Switzerland
[4] Department of Physics and Astronomy, Vrije Universiteit, NL-1081 HV Amsterdam
[5] Nikhef Theory Group, Science Park 105, 1098 XG Amsterdam, The Netherlands
[6] University of Jyvaskyla, Department of Physics, P.O. Box 35, FI-40014 University of Jyvaskyla, Finland
[7] Helsinki Institute of Physics, P.O. Box 64, FI-00014 University of Helsinki, Finland
[8] DAMTP, University of Cambridge, Wilberforce Road, Cambridge, CB3 0WA, United Kingdom
[9] Dipartimento di Fisica, Università degli Studi di Torino and INFN, Sezione di Torino, Via Pietro Giuria 1, I-10125 Torino, Italy
[10] Universität Würzburg, Institut für Theoretische Physik und Astrophysik, 97074 Würzburg, Germany

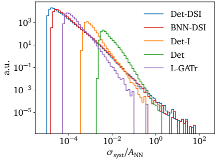
This paper is dedicated to the memory of Stefano Catani, Grand Master of QCD, great scientist and human being

Fast event generation [Sherpa, Madgraph, Badger...]

- 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

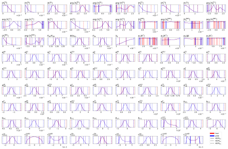
Abstract: Machine learning technology has the potential to dramatically optimise event generation and simulations. We continue to investigate the use of neural networks to approximate matrix elements for high-multiplicity scattering processes. We focus on the case of loop-induced diphoton production 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.

ML in theory

Learned likelihoods [Pierini, Reyes-Gonzales, Torre,...]

- $\mathcal{O}(100)$ physics and nuisance parameters
- learn fast likelihood
- supervised vs unsupervised

$\rightarrow$ Similar to phase space...



The NFLikelihood: an unsupervised DNNLikelihood from Normalizing Flows

Humberto Reyes-González[1,2,3,*] and Riccardo Torre[2,†]

1 Department of Physics, University of Genova, Via Dodecaneso 33, 16146 Genova, Italy
2 INFN, Sezione di Genova, Via Dodecaneso 33, I-16146 Genova, Italy
3 Institut für Theoretische Teilchenphysik und Kosmologie, RWTH Aachen, 52074 Aachen, Germany

* humberto.reyes@rwth-aachen.de
† riccardo.torre@ge.infn.it

Abstract

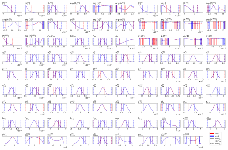
We propose the NFLikelihood, an unsupervised version, based on Normalizing Flows, of the DNNLikelihood proposed in Ref. [1]. We show, through realistic examples, how Autoregressive Flows, based on affine and rational quadratic spline bijectors, are able to learn complicated high-dimensional Likelihoods arising in High Energy Physics (HEP) analyses. We focus on a toy LHC analysis example already considered in the literature and on two Effective Field Theory fits of flavor and electroweak observables, whose samples have been obtained through the HEPFit code. We discuss advantages and disadvantages of the unsupervised approach with respect to the supervised one and discuss a possible interplay between the two.

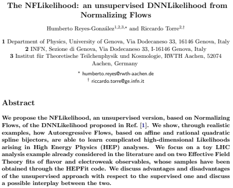
ML in theory

Learned likelihoods [Pierini, Reyes-Gonzales, Torre,...]

- $\mathcal{O}(100)$ physics and nuisance parameters
- learn fast likelihood
- supervised vs unsupervised

$\rightarrow$ Similar to phase space...





The NFLikelihood: an unsupervised DNNLikelihood from Normalizing Flows

Humberto Reyes-González[1,2,3,*] and Riccardo Torre[2,†]

1 Department of Physics, University of Genova, Via Dodecaneso 33, 16146 Genova, Italy
2 INFN, Sezione di Genova, Via Dodecaneso 33, I-16146 Genova, Italy
3 Institut für Theoretische Teilchenphysik und Kosmologie, RWTH Aachen, 52074 Aachen, Germany

* humberto.reyes@rwth-aachen.de
† riccardo.torre@ge.infn.it

Abstract

We propose the NFLikelihood, an unsupervised version, based on Normalizing Flows, of the DNNLikelihood proposed in Ref. [1]. We show, through realistic examples, how Autoregressive Flows, based on affine and rational quadratic spline bijectors, are able to learn complicated high-dimensional Likelihoods arising in High Energy Physics (HEP) analyses. We focus on a toy LHC analysis example already considered in the literature and on two Effective Field Theory fits of flavor and electroweak observables, whose samples have been obtained through the HEPFit code. We discuss advantages and disadvantages of the unsupervised approach with respect to the supervised one and discuss a possible interplay between the two.

Navigating string landscape [reinforcement learning]

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

$\rightarrow$ Model space sampling

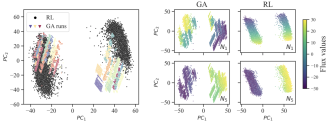
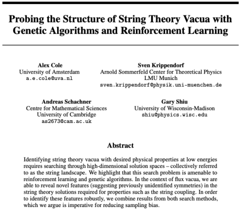


Figure 1: *Left:* Cluster structure in dimensionally reduced flux samples for RL and 25 GA runs (PCA on all samples of GA and RL). The colors indicate individual GA runs. *Right:* Dependence on flux (input) values (N_3 and N_5 respectively) in relation to principal components for a PCA fit of the individual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

Regression — LHC style

Energy calibration with uncertainties [ATLAS + Vogel, 2412.04370]

- interpretable calorimeter phase space x
- learned calibration function

$$\mathcal{R}^{\text{BNN}}(x) \pm \Delta\mathcal{R}^{\text{BNN}}(x) \approx \frac{E^{\text{obs}}(x)}{E^{\text{dep}}(x)}$$

- uncertainties: noise in data
  network expressivity
  data representation ...

Regression — LHC style

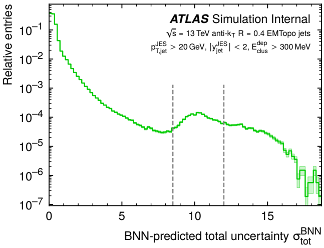
Energy calibration with uncertainties [ATLAS + Vogel, 2412.04370]

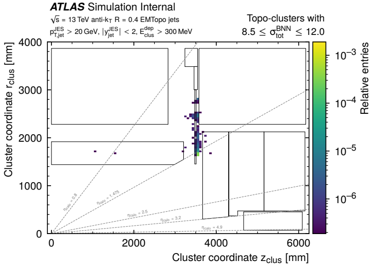
- interpretable calorimeter phase space x
- learned calibration function

$$\mathcal{R}^{\text{BNN}}(x) \pm \Delta\mathcal{R}^{\text{BNN}}(x) \approx \frac{E^{\text{obs}}(x)}{E^{\text{dep}}(x)}$$

- uncertainties: noise in data
  network expressivity
  data representation ...

$\to$ Understand (simulated) detector



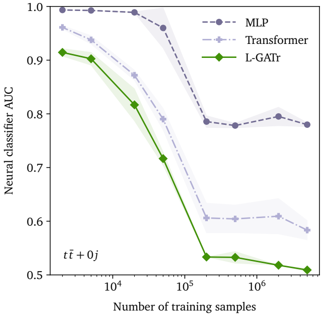


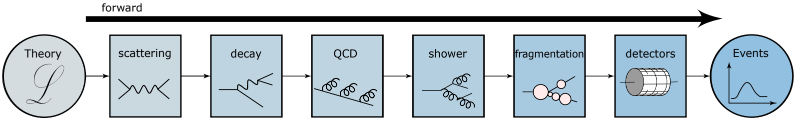
Generative AI

Simulations, MadNIS, calorimeters,... [Milano: Winterhalder]

- learn phase space density
  fast sampling Gaussian $\to$ phase space
- Variational Autoencoder
  $\to$ low-dimensional physics
- Generative Adversarial Network
  $\to$ generator trained by classifier
- Normalizing Flow/Diffusion
  $\to$ (bijective) mapping [ask R Torre]
- JetGPT, ViT
  $\to$ non-local structures
- Equivariant L-GATr
  $\to$ Lorentz symmetry for efficiency

$\to$ Combinations: equivariant transformer CFM...





Generative AI — LHC-style

Compare generated with training data [cf Refereeing the Referees]

- regression accuracy $\Delta = (E_\text{data} - E_\theta)/E_\text{data}$
- harder for generation, unsupervised density
  classify training vs generated events $D(x)$
  learned density ratio [Neyman-Pearson]

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

$\to$ Test ratio over phase space

Generative AI — LHC-style

Compare generated with training data [cf Refereeing the Referees]

- regression accuracy $\Delta = (E_\text{data} - E_\theta)/E_\text{data}$
- harder for generation, unsupervised density
  classify training vs generated events $D(x)$
  learned density ratio [Neyman-Pearson]

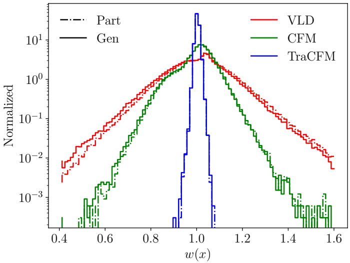
$$w(x_i) = \frac{D(x_i)}{1 - D(x_i)} = \frac{p_\text{data}(x_i)}{p_\text{model}(x_i)}$$

$\rightarrow$ Test ratio over phase space

Progress in NN-generators

- any generative AI task
- compare different architectures
- accuracy from width of weight distribution
- tails indicating failure mode

$\rightarrow$ Systematic performance test



Transforming LHC physics

Number of searches

- optimal inference: signal and background simulations
- CPU-limitation for many signals?

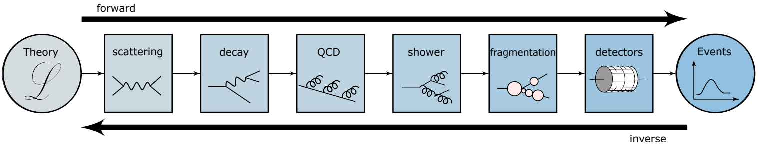
Optimal analyses

- theory limiting many analyses, but continuous progress
- allow for analyses to be updated?

Public LHC data

- common lore:
  LHC data too complicated for amateurs
- in truth:
  hard scattering and decay simulations public
  BSM physics not in hadronization and detector

→ Unfold to suitable level



ML-Unfolding

Basic structure

- four phase space distributions

$$\begin{tikzcd}
p_\text{sim}(x_\text{part}) \arrow[rr, leftrightarrow, "\text{unfolding inference}"] \arrow[d, "p(x_\text{reco}|x_\text{part})"'] & & p_\text{unfold}(x_\text{part}) \\
p_\text{sim}(x_\text{reco}) \arrow[rr, leftrightarrow, "\text{forward inference}"] & & p_\text{data}(x_\text{reco}) \arrow[u, "p(x_\text{part}|x_\text{reco})"']
\end{tikzcd}$$

- two conditional probabilities

$$p(x_\text{part}|x_\text{reco}) = p(x_\text{reco}|x_\text{part}) \times \frac{p_\text{sim}(x_\text{part})}{p_\text{sim}(x_\text{reco})}$$

- forward and inverse generation symmetric [stochastic]
- learnable from paired events $(x_\text{part}, x_\text{reco})$

$\to$ ML for unbinned and high-dimensional unfolding?

ML-Unfolding

Basic structure

- four phase space distributions

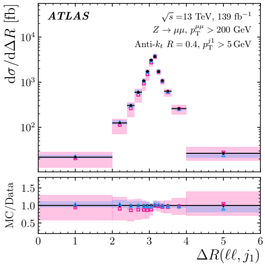
$$\begin{tikzcd}
p_\text{sim}(x_\text{part}) \arrow[rr, leftrightarrow, "\text{unfolding inference}"] \arrow[d, "p(x_\text{reco}|x_\text{part})"'] & & p_\text{unfold}(x_\text{part}) \\
p_\text{sim}(x_\text{reco}) \arrow[rr, leftrightarrow, "\text{forward inference}"] & & p_\text{data}(x_\text{reco}) \arrow[u, "p(x_\text{part}|x_\text{reco})"']
\end{tikzcd}$$

$\to$ ML for unbinned and high-dimensional unfolding?

OmniFold [Andreassen, Komiske, Metodiev, Nachman, Thaler + ATLAS]

- learn $p_\text{sim}(x_\text{reco}) \leftrightarrow p_\text{data}(x_\text{reco})$ [Neyman-Pearson]
- reweight $p_\text{sim}(x_\text{part}) \to p_\text{unfold}(x_\text{part})$

$$\begin{tikzcd}
p_\text{sim}(x_\text{part}) \arrow[rr, "\text{classifier weights}"] \arrow[d, leftrightarrow, "\text{pull/push weights}"'] & & p_\text{unfold}(x_\text{part}) \\
p_\text{sim}(x_\text{reco}) \arrow[rr, leftrightarrow, "\text{classifier weights}"] & & p_\text{data}(x_\text{reco})
\end{tikzcd}$$



- Z+jets in 24D [ATLAS]

$\to$ Driven by (now) established ML-classification

Unfolding by generation

Targeting conditional probability [Winterhalder]

- just like forward ML-generation
- learn inverse conditional probability from $(x_\text{part}, x_\text{reco})$

$$\begin{tikzcd}
p_\text{sim}(x_\text{part}) \arrow[d, leftrightarrow, "\text{paired data}"'] & & p_\text{unfold}(x_\text{part}) \\
p_\text{sim}(x_\text{reco}) \arrow[rr, leftrightarrow, "\text{correspondence}"] & & p_\text{data}(x_\text{reco}) \arrow[u, "p_\text{model}(x_\text{part} | x_\text{reco})"']
\end{tikzcd}$$

Improvements crucial

1. likelihood loss to generate posterior $\to$ cINN
2. make networks more precise $\to$ TraCFM
3. remove training prior

$\to$ Driven by generative networks

Unfolding top decays

A challenge [Favaro, Kogler, Paasch, Palacios Schweitzer, TP, Schwarz]

- first measure m_t in unfolded data
  then unfold full kinematics
- model dependence: simulation m_s vs data m_d

$$\begin{tikzcd}
p_\text{sim}(x_\text{part}|m_s) \arrow[d, "p(x_\text{reco}|x_\text{part})"'] & & p_\text{unfold}(x_\text{part}|m_s, m_d) \\
p_\text{sim}(x_\text{reco}|m_s) \arrow[rr, leftrightarrow, "\text{correspondence}"] & & p_\text{data}(x_\text{reco}|m_d) \arrow[u, "p_\text{model}(x_\text{part}|x_\text{reco}, m_s)"']
\end{tikzcd}$$

Unfolding top decays

A challenge [Favaro, Kogler, Paasch, Palacios Schweitzer, TP, Schwarz]

- first measure m_t in unfolded data
  then unfold full kinematics
- complete training bias $m_d \to m_s$ [too bad to reweight]

$$\begin{tikzcd}
p_\text{sim}(x_\text{part}|m_s) \arrow[d, "p(x_\text{reco}|x_\text{part})"'] & & p_\text{unfold}(x_\text{part}|m_s, \cancel{m_d}) \\
p_\text{sim}(x_\text{reco}|m_s) \arrow[rr, leftrightarrow, "\text{correspondence}"] & & p_\text{data}(x_\text{reco}|m_d) \arrow[u, "p_\text{model}(x_\text{part}|x_\text{reco}, m_s)"']
\end{tikzcd}$$

1. weaken bias by training on m_s-range
2. strengthen data by including batch-wise $m_d \sim M_{jjj} \in x_\text{reco}$

Unfolding top decays

A challenge [Favaro, Kogler, Paasch, Palacios Schweitzer, TP, Schwarz]

- first measure m_t in unfolded data
  then unfold full kinematics
- complete training bias $m_d \to m_s$ [too bad to reweight]

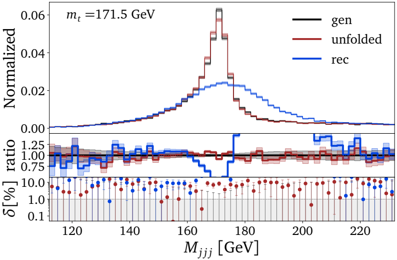
$$
\begin{tikzcd}
p_\text{sim}(x_\text{part} | m_s) \arrow[d, "p(x_\text{reco} | x_\text{part})"'] & & p_\text{unfold}(x_\text{part} | m_s, \cancel{m_d}) \\
p_\text{sim}(x_\text{reco} | m_s) \arrow[rr, leftrightarrow, "\text{correspondence}"] & & p_\text{data}(x_\text{reco} | m_d) \arrow[u, "p_\text{model}(x_\text{part} | x_\text{reco}, m_s)"']
\end{tikzcd}
$$

1 weaken bias by training on m_s-range

2 strengthen data by including batch-wise $m_d \sim M_{jjj} \in x_\text{reco}$

Preliminary unfolding results [TraCFM]

- 4D for calibrated mass measurement



Unfolding top decays

A challenge [Favaro, Kogler, Paasch, Palacios Schweitzer, TP, Schwarz]

- first measure m_t in unfolded data
  then unfold full kinematics
- complete training bias $m_d \to m_s$ [too bad to reweight]

$$\begin{tikzcd}
p_{\text{sim}}(x_{\text{part}}|m_s) \arrow[d, "p(x_{\text{reco}}|x_{\text{part}})"'] & & p_{\text{unfold}}(x_{\text{part}}|m_s, \cancel{m_d}) \\
p_{\text{sim}}(x_{\text{reco}}|m_s) \arrow[rr, leftrightarrow, "\text{correspondence}"] & & p_{\text{data}}(x_{\text{reco}}|m_d) \arrow[u, "p_{\text{model}}(x_{\text{part}}|x_{\text{reco}}, m_s)"']
\end{tikzcd}$$

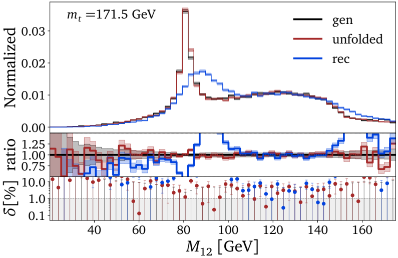
1 weaken bias by training on m_s-range

2 strengthen data by including batch-wise $m_d \sim M_{jjj} \in x_{\text{reco}}$

Preliminary unfolding results [TraCFM]

- 4D for calibrated mass measurement
- 12D published data

$\to$ CMS data next



ML for LHC Theory

Developing ML for the best science

- just another numerical tool for a numerical field
- transformative new language
- driven by money from data science and medical research
- 1000 Einsteins...

  ...improving established tools
  ...developing new tools for established tasks
  ...transforming through new ideas

→ You can be the golden generation!

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]

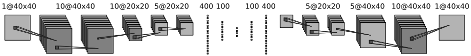
Anomaly searches

Non-resonant searches

- key: bottleneck
  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
  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 $S \leftrightarrow B$?

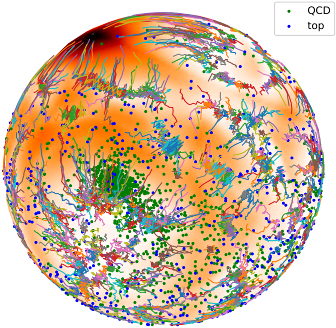
Missing and anomalous features

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

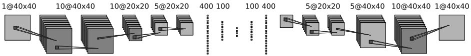
$$p_\theta(x) = \frac{e^{-E_\theta(x)}}{Z_\theta}$$

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

- inducing background metric
- Z_θ from Markov Chain



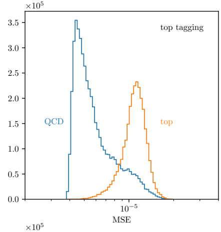
Anomaly searches



Non-resonant searches

- key: bottleneck
  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...