ÖAW AUSTRIAN ACADEMY OF SCIENCES

HEPHY INSTITUTE OF HIGH ENERGY PHYSICS

COMETA WG2 Activities

— 1st COMETA General Meeting, Izmir, Türkiye —

Claudius Krause

Institute of High Energy Physics (HEPHY), Austrian Academy of Sciences (OeAW)

March 1, 2024

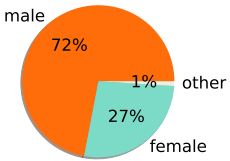
Working Group 2 of COMETA

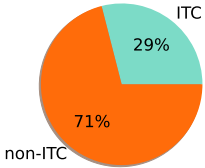
Technological Innovation in Data Analysis

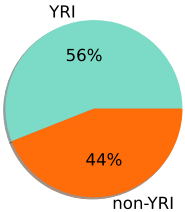
goals:

- Create a discussion forum for the HEP and ML communities, enabling the development of innovative tools that will improve future multi-boson measurements.
- Establish a long-lasting, mutually beneficial cooperation between the HEP and ML communities.

Currently, Working Group 2 has 100 members:







ITC: Inclusiveness Target Countries

YRI: Young Researchers and Innovators (below age of 40)

Leaders of Working Group 2

Alessandra Cappati

(experiment)

- MSCA Fellow in CMS (Lab. Leprince-Ringuet Palaiseau)
- background: multiboson physics and HZZ convenor
- now: Higgs and AI-aided reconstruction

Riccardo Finotello

(industry)

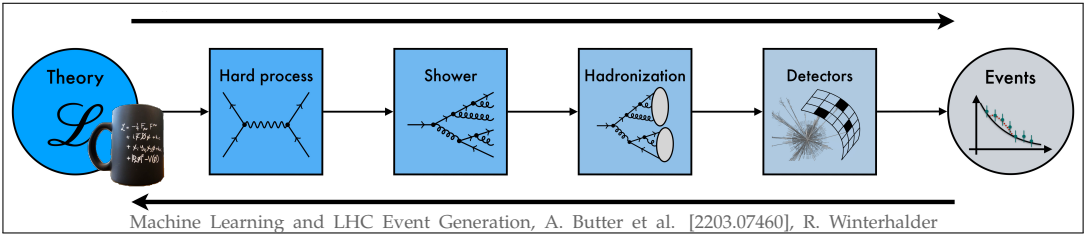
- Research Engineer at CEA (Paris-Saclay)
- background: string theory and AI for algebraic geometry
- now: computer vision for hyperspectral images

Claudius Krause

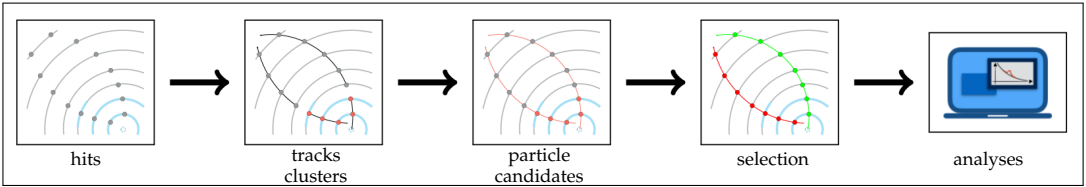
(theory)

- tenure-tracker in ML4HEP (HEPHY Vienna)
- background: Higgs and EFT
- now: application of generative models

Data analysis is very complex in HEP …



Machine Learning and LHC Event Generation, A. Butter et al. [2203.07460], R. Winterhalder

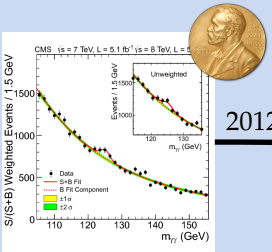


…and machine learning can help in every single step!

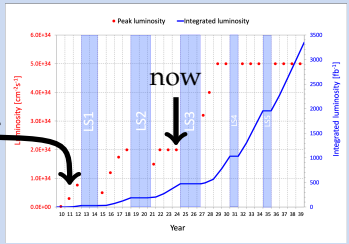
pics based on [2105.01160]

Why is machine learning so popular for high-energy physics?

Data volume



2012

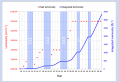


https://lhc-commissioning.web.cern.ch/schedule/HL-LHC-plots.htm

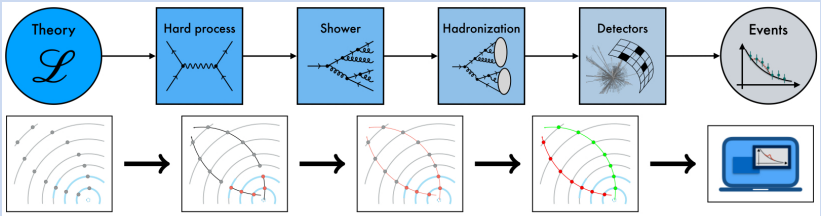
Large amounts of labeled (simulation) and unlabeled (experiment) data.
$\Rightarrow$ ML works best with lots of data

Why is machine learning so popular for high-energy physics?

Data volume



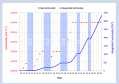
Data complexity



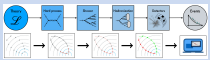
High-dimensional & highly correlated data.
$\Rightarrow$ ML can handle that well

Why is machine learning so popular for high-energy physics?

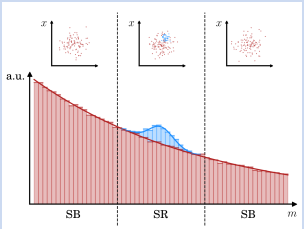
Data volume



Data complexity



Signal detection

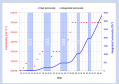


Hallin et al. [2109.00546]

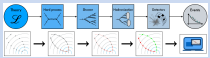
Rare and elusive signals among large backgrounds.
$\Rightarrow$ ML has high sensitivity

Why is machine learning so popular for high-energy physics?

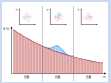
Data volume



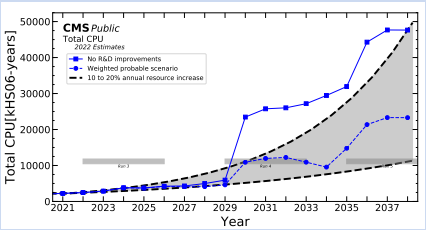
Data complexity



Signal detection



Computing budget



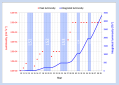
https://twiki.cern.ch/twiki/bin/view/CMSPublic/CMSOffline ComputingResults

Simulation & analysis are computationally expensive.

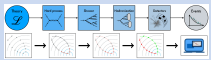
$\Rightarrow$ ML is fast

Why is machine learning so popular for high-energy physics?

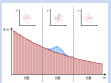
Data volume



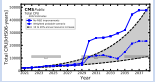
Data complexity



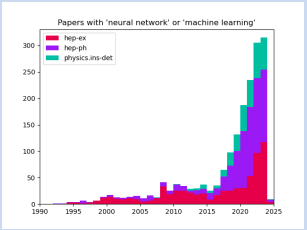
Signal detection



Speed



Increasing interest



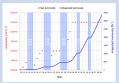
via "The INSPIRE REST API"

We see about 300 papers / year.

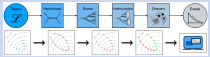
$\Rightarrow$ ML is everywhere, as there are no off-the-shelf solutions

Why is machine learning so popular for high-energy physics?

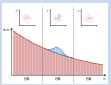
Data volume



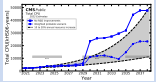
Data complexity



Signal detection



Speed



Interest

ML is fun

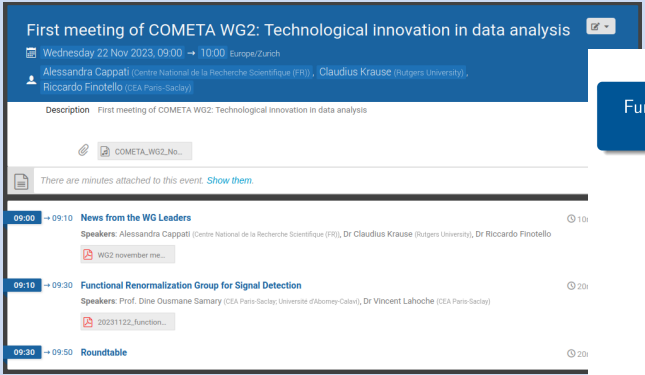
via midjourney: “Albert Einstein smiling while having fun coding”

⇒ Like Galileo Galilei looking through the telescope for the first time!

Additionally: Big players develop and maintain python packages, which makes research on our end much easier

Past Highlight: 1st (online) meeting

22 November 2023 https://indico.cern.ch/event/1349057/



Current Highlight: 1st general meeting I.

Current Highlight: 1st general meeting I.

12:00

Taggers for boosted HH searches within the ATLAS experiment

Education Hall, İzmir, Türkiye

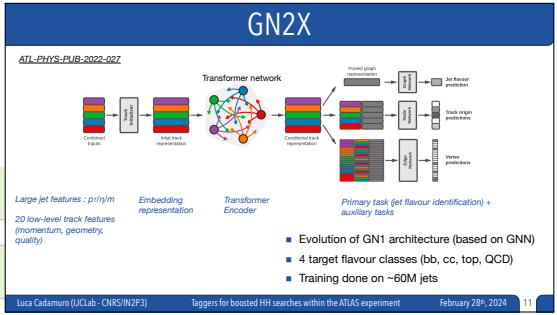
Machine Learning for EFT interpretation

Education Hall, İzmir, Türkiye

13:00

Kernel methods for new physics searches *Dr Marco Letizia*

Education Hall, İzmir, Türkiye 13:00 - 13:20

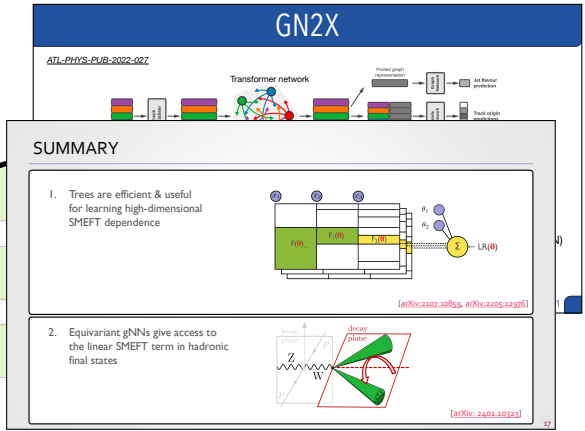


Current Highlight: 1st general meeting I.

12:00 **Taggers for boosted HH searches within the ATLAS experiment**
Education Hall, İzmir, Türkiye

Machine Learning for EFT interpretation
Education Hall, İzmir, Türkiye

13:00 **Kernel methods for new physics searches**
Education Hall, İzmir, Türkiye

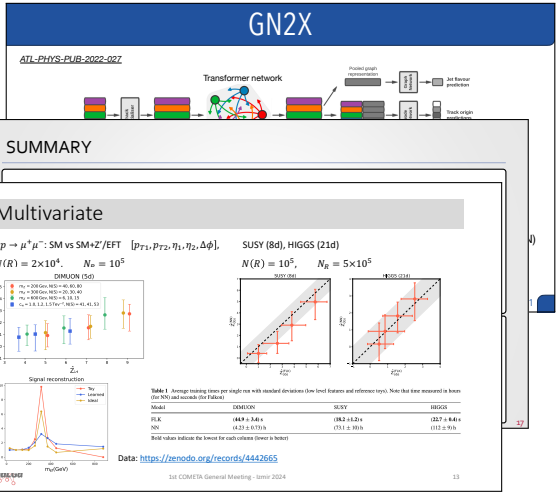


Current Highlight: 1st general meeting I.

12:00 **Taggers for boosted HH searches within the ATLAS experiment**
Education Hall, İzmir, Türkiye

Machine Learning for EFT interpretation
Education Hall, İzmir, Türkiye

13:00 **Kernel methods for new physics searches**
Education Hall, İzmir, Türkiye



Current Highlight: 1st general meeting II.

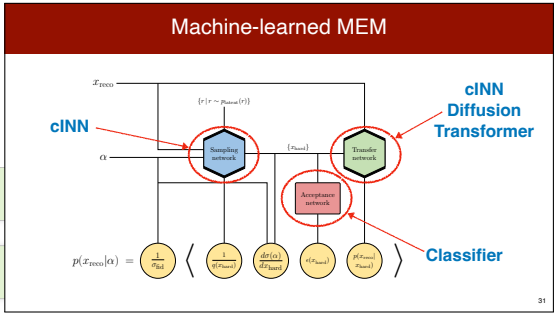
15:00	**Precision machine learning approaches**	*Nathan Huetsch*	
	Education Hall, İzmir, Türkiye	15:00 - 15:20	
	Machine Learning based multivariate observables for global SMEFT fits	*Jaco ter Hoeve*	
	Education Hall, İzmir, Türkiye	15:30 - 15:50	
16:00	**Normalizing flows for lattice field theory**	*Alessandro Nada*	
	Education Hall, İzmir, Türkiye	16:00 - 16:20	

Current Highlight: 1st general meeting II.

15:00 **Precision machine learning approaches**
Education Hall, İzmir, Türkiye

Machine Learning based multivariate observables for global SMEFT fits
Education Hall, İzmir, Türkiye

16:00 **Normalizing flows for lattice field theory** *Alessandro Nada*
Education Hall, İzmir, Türkiye 16:00 - 16:20



Current Highlight: 1st general meeting II.

15:00 **Precision machine learning approaches**
Education Hall, İzmir, Türkiye

Machine Learning based multivariate observables for global SMEFT fits
Education Hall, İzmir, Türkiye

16:00 **Normalizing flows for lattice field theory**
Education Hall, İzmir, Türkiye

Machine-learned MEM

A combined framework

Best of two worlds?

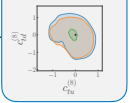
Theory
SM: (N)NLO QCD + NLO EW
EFT: NLO QCD, linear and quadratics, with SMEFTiNLO
NNPDF4.0 no top

Data
447 measurements from Higgs, top, diboson and EWPO
Full experimental correlations

Projections
HL-LHC, FCCee, CEPC
Automatic projection module

SMEFiT — External likelihoods

ML4EFT



Fitting framework
Nested Sampling: Bayesian
Monte Carlo Replica Method (NNPDF)
External likelihoods

Validation
Extensive toolbox:
PCA, Fisher info, closure tests,...
Full posterior distribution available

- The ultimate global EFT fit combines **binned** and **multivariate unbinned** ML observables
- We need a framework that connects them

Jaco ter Hoeve - 1st COMETA General meeting - 29/01/24 27

Current Highlight: 1st general meeting II.

15:00 **Precision machine learning approaches**
Education Hall, İzmir, Türkiye

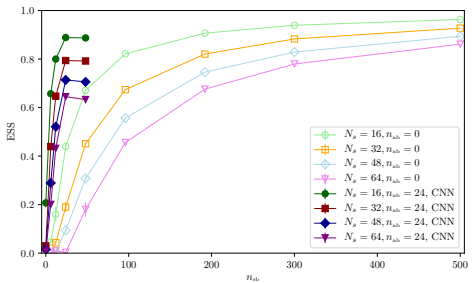
Machine Learning based multivariate observables for global SMEFT fits
Education Hall, İzmir, Türkiye

16:00 **Normalizing flows for lattice field theory**
Education Hall, İzmir, Türkiye

Machine-learned MEM

A combined framework

Training length: 10^4 epochs for all volumes. Saturates fast



Current Highlight: 1st general meeting II.

Machine-learned MEM

A combined framework

Saturates fast

15:00 **Precision machine learning**
Education Hall, İzmir, Türkiye

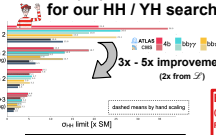
Machine Learning based mu
Education Hall, İzmir, Türkiye

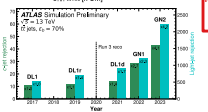
16:00 **Normalizing flows for lattice**
Education Hall, İzmir, Türkiye

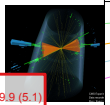
Conclusion

ML is transforming the way we do physics and the H(bb) results that are now key driver for our HH / YH search programs

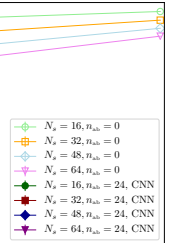
- early Run 2: Shallow NNs / BDTs
- full Run 2: Deep NNs (RNNs / Sets)
- late Run 2 / early Run 3: Graph NNs
- State of the art: Transformers







Boosted:
Obs (exp): 9.9 (5.1)
Resolved:
Obs (exp): 5.4 (8.1)
Obs (exp): 3.8 (7.8)



from Nicole Hartman

Future Highlight: 2nd (online) meeting

"Normalizing Flows for Particle Physics" date tbd: `https://lettucemeet.com/l/OnmW6`
One afternoon of talks of 20min + questions, and overall discussion session

preliminary agenda:

- Sofia Palacios-Schweitzer on unfolding
- Kim Nicoli on reweighting and importance sampling for lattices
- Ramon Winterhalder on multichannel importance sampling for MadGraph
- Claudius Krause on calorimeter shower simulation

"Short-term" future: Next steps

Deliverable for 1st year:

Definition of common benchmarks to develop algorithms for COMETA

my suggestion: 1 ATLAS / 1 CMS (?)

possible topics:

- jet/event topology tagging
- polarization
- simulation
- jet substructures
- track reconstruction
- (rare) signal detection

In any case: We need realistic, public data!
Short-term associations are still an entry barrier, especially from outside HEP!

"Mid-term" future

Upcoming COMETA topical workshops:

- on Effective Field Theory in Multiboson Production, Padova, Italy, 10–11 June 2024

⇒ `https://indico.cern.ch/event/1358085/`

- on Vector-Boson Polarisations, Toulouse, France, 23–24 September 2024

⇒ `https://indico.cern.ch/event/1371888/`

- on Boosted Hadronic Bosons, Vienna, Austria, Winter 2024/2025

⇒ stay tuned!

Once we have the benchmark challenges:

- advertize inside and outside of HEP
- organize hackathlons

COMETA WG2 — Technological Innovation in Data Analysis

- If you have not already, please join COMETA and WG2

⇒ `https://www.cost.eu/actions/CA22130/`

- Discuss ideas / requests / questions with us!
- We should start discussing a set of well-defined problems (and their datasets) that we want to tackle!
- The pace of ML industry is incredible ⇒ we will try to pull that into HEP and multiboson physics!