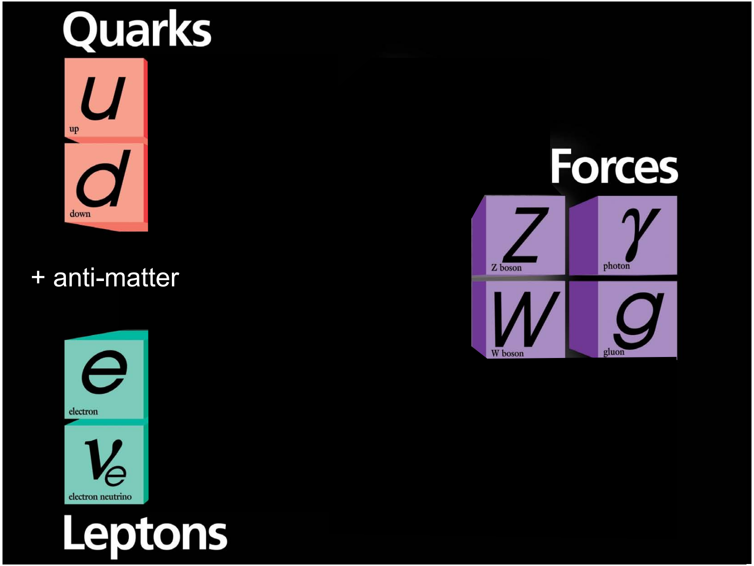
AI in particle physics

Yann Coadou

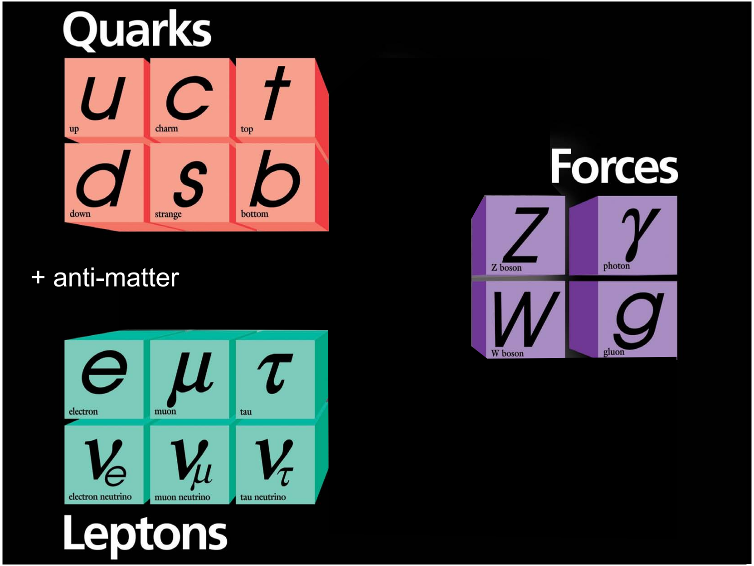
CPPM Marseille

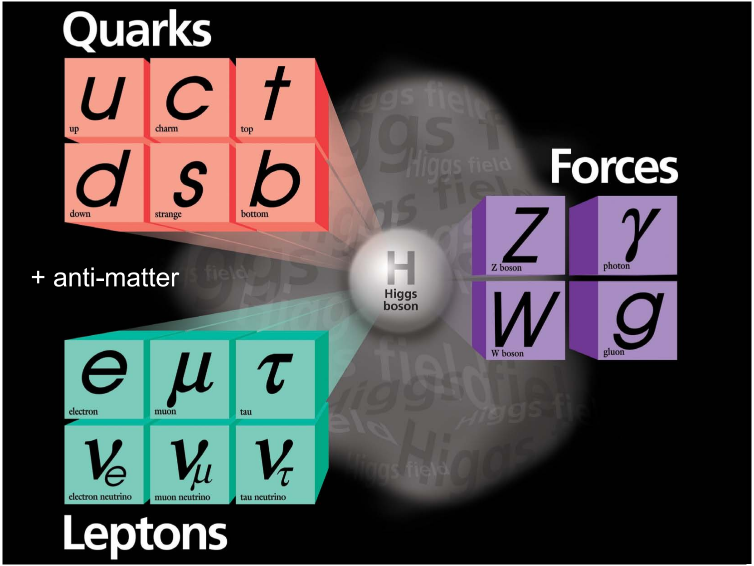
diiP Summer School 2024
Université Paris Cité, 14 June 2024

The standard model of particle physics



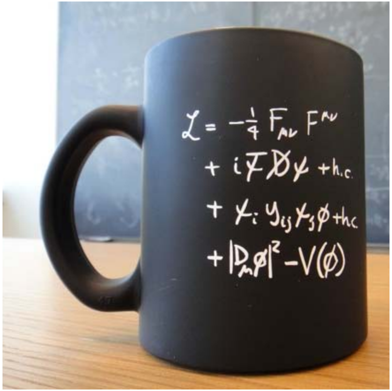
The standard model of particle physics



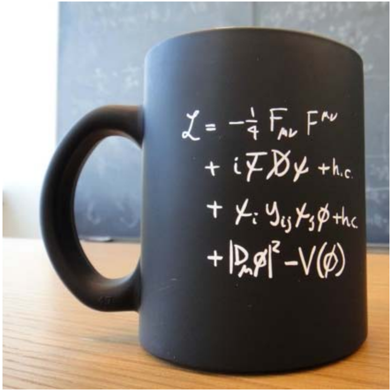

Quarks
u
up
c
charm
t
top
d
down
s
strange
b
bottom
+ anti-matter
H
Higgs boson
Forces
Z
Z boson
γ
photon
W
W boson
g
gluon
e
electron
μ
muon
τ
tau
νe
electron neutrino
νμ
muon neutrino
ντ
tau neutrino
Leptons

The standard model of particle physics

- Condensed form



- Condensed form



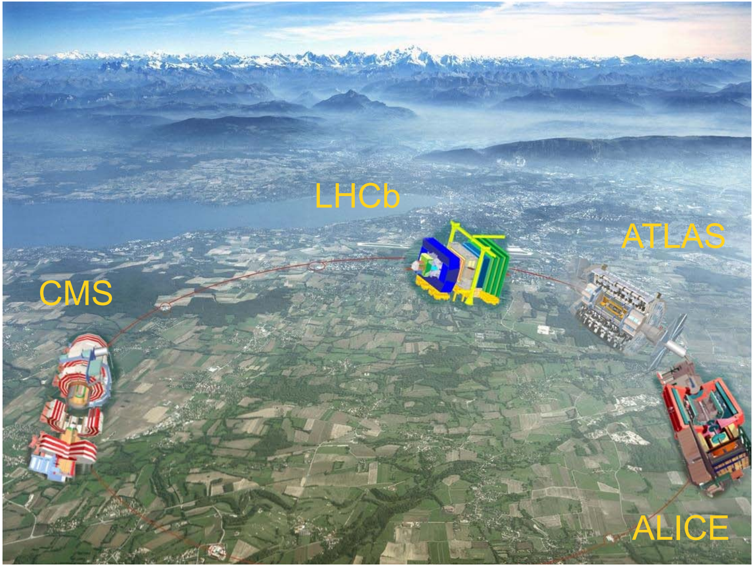
- Expanded version...

$$
\begin{gathered}
-\tfrac{1}{2}\partial_\nu g^a_\mu \partial_\nu g^a_\mu - g_s f^{abc}\partial_\mu g^a_\nu g^b_\mu g^c_\nu - \tfrac{1}{4}g_s^2 f^{abc} f^{ade} g^b_\mu g^c_\nu g^d_\mu g^e_\nu + \\
\tfrac{1}{2} i g_s^2 (\bar{q}^\sigma_i \gamma^\mu q^\sigma_j) g^a_\mu + \bar{G}^a \partial^2 G^a + g_s f^{abc} \partial_\mu \bar{G}^a G^b g^c_\mu - \partial_\nu W^+_\mu \partial_\nu W^-_\mu - \\
M^2 W^+_\mu W^-_\mu - \tfrac{1}{2}\partial_\nu Z^0_\mu \partial_\nu Z^0_\mu - \tfrac{1}{2c_w^2} M^2 Z^0_\mu Z^0_\mu - \tfrac{1}{2}\partial_\mu A_\nu \partial_\mu A_\nu - \\
\tfrac{1}{2}\partial_\mu H \partial_\mu H - \tfrac{1}{2} m_h^2 H^2 - \partial_\mu \phi^+ \partial_\mu \phi^- - M^2 \phi^+ \phi^- - \tfrac{1}{2}\partial_\mu \phi^0 \partial_\mu \phi^0 - \\
\tfrac{1}{2c_w^2} M \phi^0 \phi^0 - \beta_h [\tfrac{2M^2}{g^2} + \tfrac{2M}{g} H + \tfrac{1}{2}(H^2 + \phi^0\phi^0 + 2\phi^+\phi^-)] + \tfrac{2M^4}{g^2}\alpha_h - \\
i g c_w [\partial_\nu Z^0_\mu (W^+_\mu W^-_\nu - W^+_\nu W^-_\mu) - Z^0_\nu (W^+_\mu \partial_\nu W^-_\mu - W^-_\mu \partial_\nu W^+_\mu) + \\
Z^0_\mu (W^+_\nu \partial_\nu W^-_\mu - W^-_\nu \partial_\nu W^+_\mu)] - i g s_w [\partial_\nu A_\mu (W^+_\mu W^-_\nu - W^+_\nu W^-_\mu) - \\
A_\nu (W^+_\mu \partial_\nu W^-_\mu - W^-_\mu \partial_\nu W^+_\mu) + A_\mu (W^+_\nu \partial_\nu W^-_\mu - W^-_\nu \partial_\nu W^+_\mu)] - \\
\tfrac{1}{2} g^2 W^+_\mu W^-_\mu W^+_\nu W^-_\nu + \tfrac{1}{2} g^2 W^+_\mu W^-_\nu W^+_\mu W^-_\nu + g^2 c_w^2 (Z^0_\mu W^+_\mu Z^0_\nu W^-_\nu - \\
Z^0_\mu Z^0_\mu W^+_\nu W^-_\nu) + g^2 s_w^2 (A_\mu W^+_\mu A_\nu W^-_\nu - A_\mu A_\mu W^+_\nu W^-_\nu) + \\
g^2 s_w c_w [A_\mu Z^0_\nu (W^+_\mu W^-_\nu - W^+_\nu W^-_\mu) - 2 A_\mu Z^0_\mu W^+_\nu W^-_\nu] - g\alpha [H^3 + \\
H \phi^0 \phi^0 + 2 H \phi^+ \phi^-] - \tfrac{1}{8} g^2 \alpha_h [H^4 + (\phi^0)^4 + 4(\phi^+\phi^-)^2 + \\
4(\phi^0)^2 \phi^+ \phi^- + 4 H^2 \phi^+ \phi^- + 2(\phi^0)^2 H^2] - g M W^+_\mu W^-_\mu H - \\
\tfrac{1}{2} g \tfrac{M}{c_w^2} Z^0_\mu Z^0_\mu H - \tfrac{1}{2} i g [W^+_\mu (\phi^0 \partial_\mu \phi^- - \phi^- \partial_\mu \phi^0) - W^-_\mu (\phi^0 \partial_\mu \phi^+ - \\
\phi^+ \partial_\mu \phi^0)] + \tfrac{1}{2} g [W^+_\mu (H \partial_\mu \phi^- - \phi^- \partial_\mu H) - W^-_\mu (H \partial_\mu \phi^+ - \phi^+ \partial_\mu H)] + \\
\tfrac{1}{2} g \tfrac{1}{c_w} (Z^0_\mu (H \partial_\mu \phi^0 - \phi^0 \partial_\mu H) - i g \tfrac{s_w^2}{c_w} M Z^0_\mu (W^+_\mu \phi^- - W^-_\mu \phi^+) + \\
i g s_w M A_\mu (W^+_\mu \phi^- - W^-_\mu \phi^+) - i g \tfrac{1-2c_w^2}{2c_w} Z^0_\mu (\phi^+ \partial_\mu \phi^- - \phi^- \partial_\mu \phi^+) + \\
i g s_w A_\mu (\phi^+ \partial_\mu \phi^- - \phi^- \partial_\mu \phi^+) - \tfrac{1}{4} g^2 W^+_\mu W^-_\mu [H^2 + (\phi^0)^2 + 2\phi^+\phi^-] - \\
\tfrac{1}{4} g^2 \tfrac{1}{c_w^2} Z^0_\mu Z^0_\mu [H^2 + (\phi^0)^2 + 2(2s_w^2 - 1)^2 \phi^+ \phi^-] - \tfrac{1}{2} g^2 \tfrac{s_w^2}{c_w} Z^0_\mu \phi^0 (W^+_\mu \phi^- + \\
W^-_\mu \phi^+) - \tfrac{1}{2} i g^2 \tfrac{s_w^2}{c_w} Z^0_\mu H (W^+_\mu \phi^- - W^-_\mu \phi^+) + \tfrac{1}{2} g^2 s_w A_\mu \phi^0 (W^+_\mu \phi^- + \\
W^-_\mu \phi^+) + \tfrac{1}{2} i g^2 s_w A_\mu H (W^+_\mu \phi^- - W^-_\mu \phi^+) - g^2 \tfrac{s_w}{c_w} (2c_w^2 - 1) Z^0_\mu A_\mu \phi^+ \phi^- - \\
g^1 s_w^2 A_\mu A_\mu \phi^+ \phi^- - \bar{e}^\lambda (\gamma\partial + m_e^\lambda) e^\lambda - \bar{\nu}^\lambda \gamma\partial \nu^\lambda - \bar{u}^\lambda_j (\gamma\partial + m_u^\lambda) u^\lambda_j - \\
\bar{d}^\lambda_j (\gamma\partial + m_d^\lambda) d^\lambda_j + i g s_w A_\mu [-(\bar{e}^\lambda \gamma^\mu e^\lambda) + \tfrac{2}{3}(\bar{u}^\lambda_j \gamma^\mu u^\lambda_j) - \tfrac{1}{3}(\bar{d}^\lambda_j \gamma^\mu d^\lambda_j)] + \\
\tfrac{ig}{4c_w} Z^0_\mu [(\bar{\nu}^\lambda \gamma^\mu (1+\gamma^5) \nu^\lambda) + (\bar{e}^\lambda \gamma^\mu (4s_w^2 - 1 - \gamma^5) e^\lambda) + (\bar{u}^\lambda_j \gamma^\mu (\tfrac{4}{3} s_w^2 - \\
1 - \gamma^5) u^\lambda_j) + (\bar{d}^\lambda_j \gamma^\mu (1 - \tfrac{8}{3} s_w^2 - \gamma^5) d^\lambda_j)] + \tfrac{ig}{2\sqrt{2}} W^+_\mu [(\bar{\nu}^\lambda \gamma^\mu (1+\gamma^5) e^\lambda) + \\
(\bar{u}^\lambda_j \gamma^\mu (1+\gamma^5) C_{\lambda\kappa} d^\kappa_j)] + \tfrac{ig}{2\sqrt{2}} W^-_\mu [(\bar{e}^\lambda \gamma^\mu (1+\gamma^5) \nu^\lambda) + (\bar{d}^\kappa_j C^\dagger_{\lambda\kappa} \gamma^\mu (1 + \\
\gamma^5) u^\lambda_j)] + \tfrac{ig}{2\sqrt{2}} \tfrac{m_e^\lambda}{M} [-\phi^+ (\bar{\nu}^\lambda (1-\gamma^5) e^\lambda) + \phi^- (\bar{e}^\lambda (1+\gamma^5) \nu^\lambda)] - \\
\tfrac{g}{2} \tfrac{m_e^\lambda}{M} [H (\bar{e}^\lambda e^\lambda) + i \phi^0 (\bar{e}^\lambda \gamma^5 e^\lambda)] + \tfrac{ig}{2M\sqrt{2}} \phi^+ [-m_d^\kappa (\bar{u}^\lambda_j C_{\lambda\kappa} (1-\gamma^5) d^\kappa_j) + \\
m_u^\lambda (\bar{u}^\lambda_j C_{\lambda\kappa} (1+\gamma^5) d^\kappa_j] + \tfrac{ig}{2M\sqrt{2}} \phi^- [m_d^\lambda (\bar{d}^\lambda_j C^\dagger_{\lambda\kappa} (1+\gamma^5) u^\kappa_j) - m_u^\kappa (\bar{d}^\lambda_j C^\dagger_{\lambda\kappa} (1 - \\
\gamma^5) u^\kappa_j] - \tfrac{g}{2} \tfrac{m_u^\lambda}{M} H (\bar{u}^\lambda_j u^\lambda_j) - \tfrac{g}{2} \tfrac{m_d^\lambda}{M} H (\bar{d}^\lambda_j d^\lambda_j) + \tfrac{ig}{2} \tfrac{m_u^\lambda}{M} \phi^0 (\bar{u}^\lambda_j \gamma^5 u^\lambda_j) - \\
\tfrac{ig}{2} \tfrac{m_d^\lambda}{M} \phi^0 (\bar{d}^\lambda_j \gamma^5 d^\lambda_j) + \bar{X}^+ (\partial^2 - M^2) X^+ + \bar{X}^- (\partial^2 - M^2) X^- + \bar{X}^0 (\partial^2 - \\
\tfrac{M^2}{c_w^2}) X^0 + \bar{Y} \partial^2 Y + i g c_w W^+_\mu (\partial_\mu \bar{X}^0 X^- - \partial_\mu \bar{X}^+ X^0) + i g s_w W^+_\mu (\partial_\mu \bar{Y} X^- - \\
\partial_\mu \bar{X}^+ Y) + i g c_w W^-_\mu (\partial_\mu \bar{X}^- X^0 - \partial_\mu \bar{X}^0 X^+) + i g s_w W^-_\mu (\partial_\mu \bar{X}^- Y - \\
\partial_\mu \bar{Y} X^+) + i g c_w Z^0_\mu (\partial_\mu \bar{X}^+ X^+ - \partial_\mu \bar{X}^- X^-) + i g s_w A_\mu (\partial_\mu \bar{X}^+ X^+ - \\
\partial_\mu \bar{X}^- X^-) - \tfrac{1}{2} g M [\bar{X}^+ X^+ H + \bar{X}^- X^- H + \tfrac{1}{c_w^2} \bar{X}^0 X^0 H] + \\
\tfrac{1-2c_w^2}{2c_w} i g M [\bar{X}^+ X^0 \phi^+ - \bar{X}^- X^0 \phi^-] + \tfrac{1}{2c_w} i g M [\bar{X}^0 X^- \phi^+ - \bar{X}^0 X^+ \phi^-] + \\
i g M s_w [\bar{X}^0 X^- \phi^+ - \bar{X}^0 X^+ \phi^-] + \tfrac{1}{2} i g M [\bar{X}^+ X^+ \phi^0 - \bar{X}^- X^- \phi^0]
\end{gathered}
$$

The Large Hadron Collider (LHC)

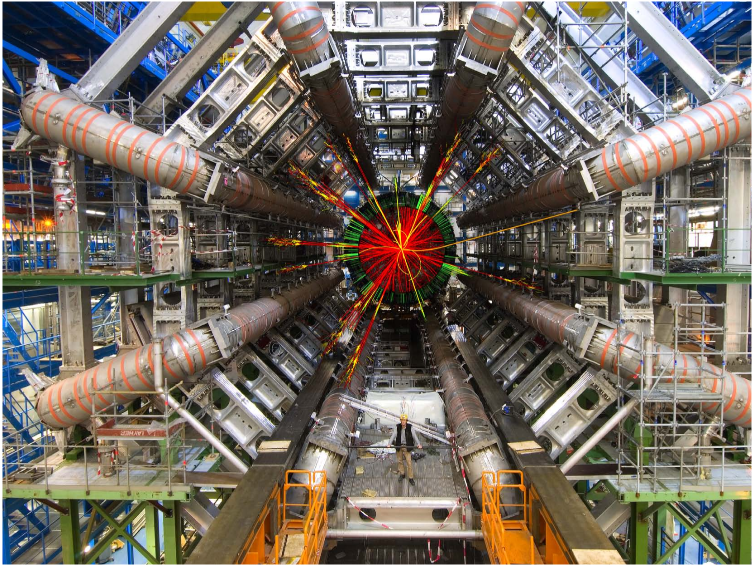
The Large Hadron Collider (LHC)

The Large Hadron Collider (LHC)



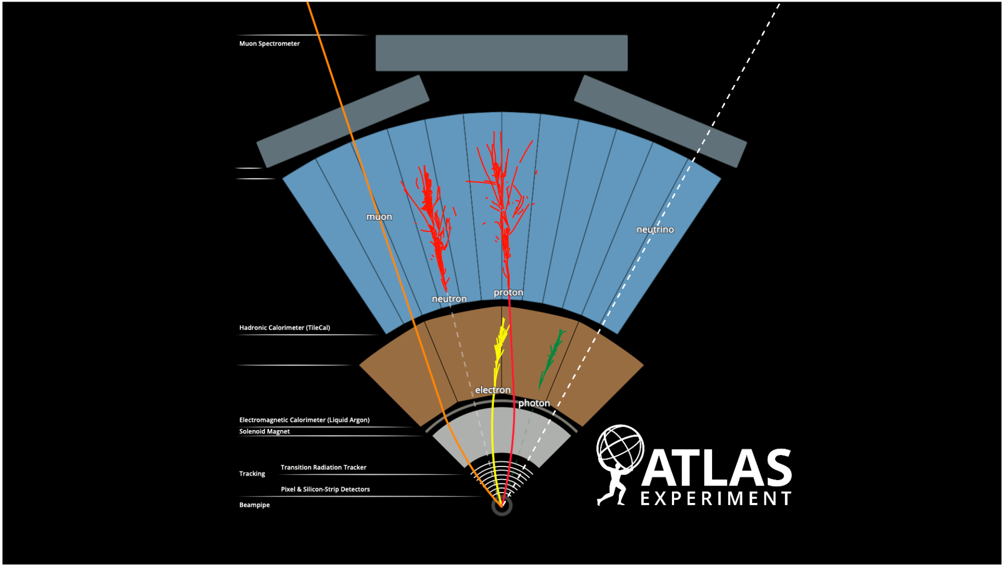
The Large Hadron Collider (LHC)

The ATLAS experiment

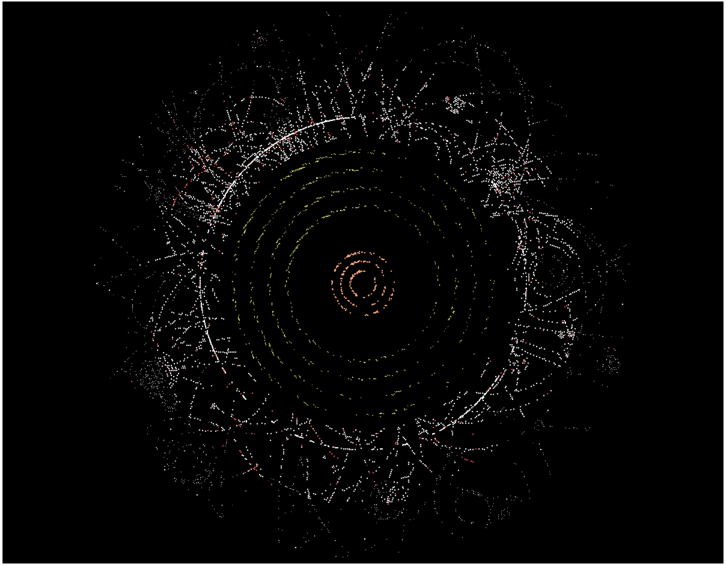


The ATLAS experiment

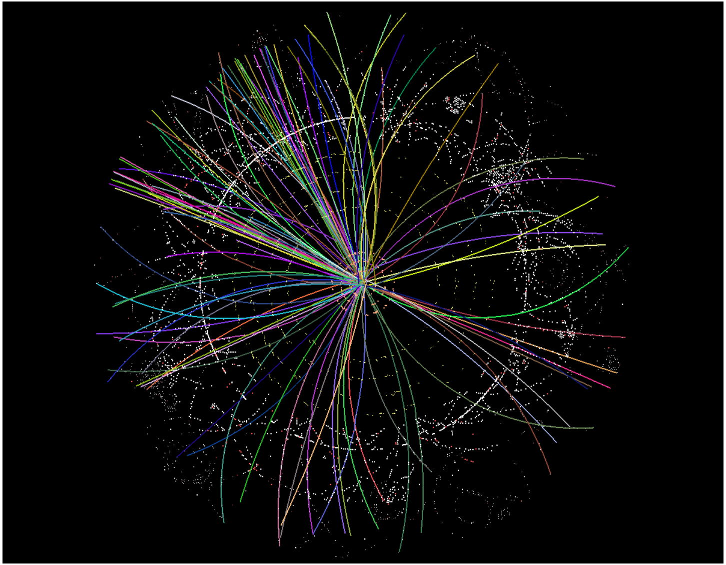
Interaction of particles with the detector

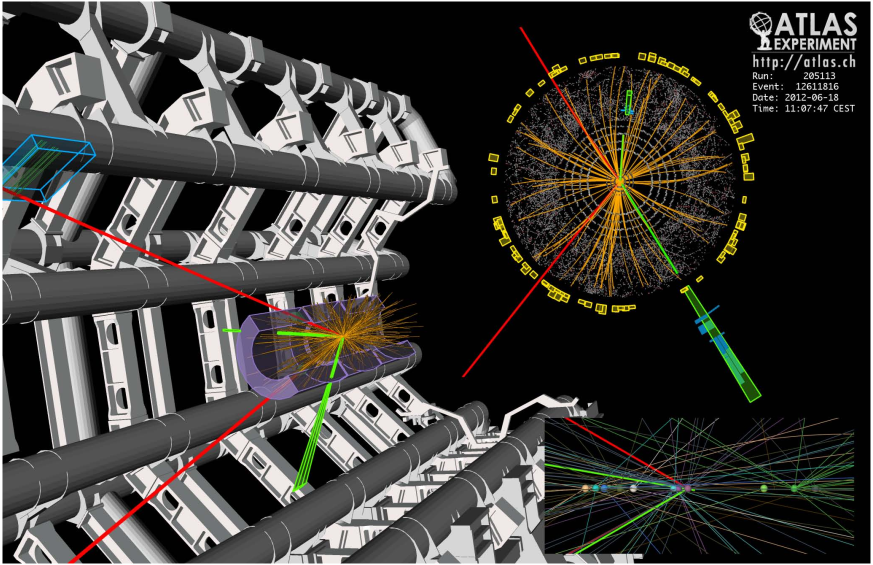


Reconstruction



Reconstruction




ATLAS
EXPERIMENT
http://atlas.ch
Run: 205113
Event: 12611816
Date: 2012-06-18
Time: 11:07:47 CEST

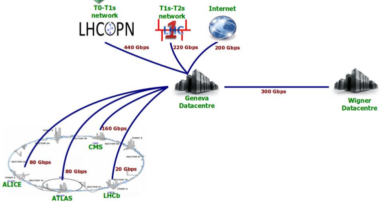
CERN

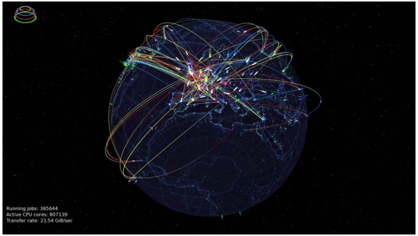
- Near Geneva, Switzerland
- 2500 employees
- 12 000 users in 500 institutes from 80 countries
- « Where the Web was born »
- CERN data center : 230 000 cores on 15 000 servers, 250 PB
- Raw data: lifetime storage → data and software conservation

Computing grid

- More than 800 000 cores
- 170 sites in 42 countries
- LHC: 50-70 petabytes/year
  CERN: +25 PB
- 2 billion files
- > 250 000 simultaneous jobs
- 2 million jobs/day
- Typically > 2 PB accessed every day
- Typical transfer rates 35 GB/s
- Total storage: ~ exabyte!

WLCG
Worldwide LHC Computing Grid

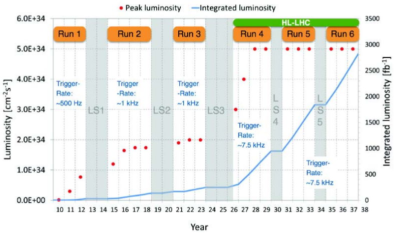


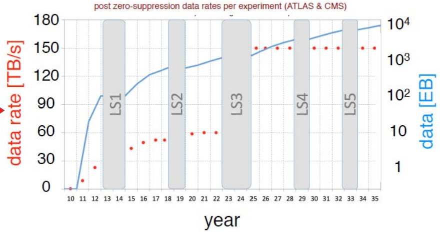


Big data

- 40 MHz → petabyte/sec in each detector, zetabyte/year!
- Impossible → « online » filters: hardware+software trigger system, to reach ~1 kHz, ~1 MB/event
- Future challenge: HL-LHC

- 40 MHz → petabyte/sec in each detector, zetabyte/year!
- Impossible → « online » filters: hardware+software trigger system, to reach ~1 kHz, ~1 MB/event
- Future challenge: HL-LHC

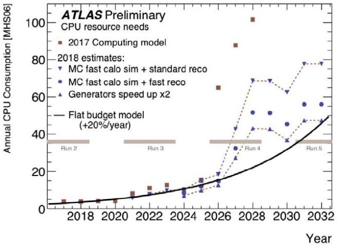




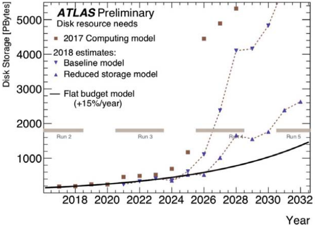
- Requires hard thinking into how to handle such quantities
- Potential showstopper if trigger not fast enough to «digest» such a flow

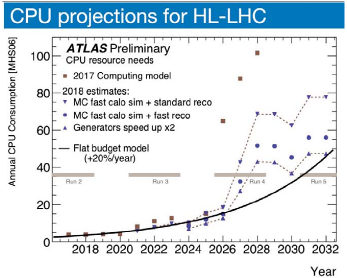
Big data

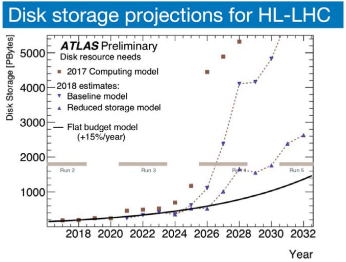
CPU projections for HL-LHC



Disk storage projections for HL-LHC







- Possible solutions
  - Technical
    - Better performing machines (GPU, FPGA, etc.)
    - Better software (vectorisation, etc.)
  - Operational
    - Smaller data samples
    - Avoid « reprocessings »
  - Political
    - Get more money
    - Access to more ressources (HPC, volunteer, etc)
  - Physics
    - Take less data
    - Cancel part of the pysics programme
    - Delay processing

Slide stolen from Johann Brehmer

Modelling particle physics processes

Theory
parameters
θ

Slide stolen from Johann Brehmer

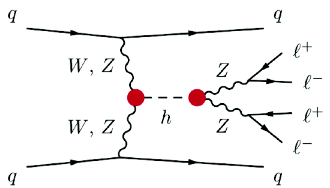
Modelling particle physics processes

Latent variables

Parton-level momenta

Theory parameters

$z_p \longleftarrow \theta$

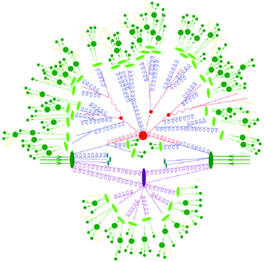


Slide stolen from Johann Brehmer

Modelling particle physics processes

Latent variables

Shower splittings		Parton-level momenta		Theory parameters
z_s	$\longleftarrow$	z_p	$\longleftarrow$	θ



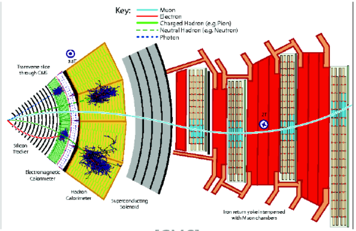
[F. Krauss]

Slide stolen from Johann Brehmer

Modelling particle physics processes

Latent variables

Detector interactions		Shower splittings		Parton-level momenta		Theory parameters
z_d	$\longleftarrow$	z_s	$\longleftarrow$	z_p	$\longleftarrow$	θ



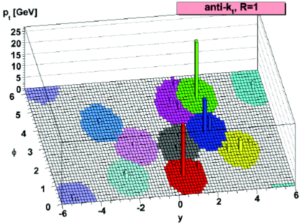
[CMS]

Slide stolen from Johann Brehmer

Modelling particle physics processes

Latent variables

Observables		Detector interactions		Shower splittings		Parton-level momenta		Theory parameters
x	$\longleftarrow$	z_d	$\longleftarrow$	z_s	$\longleftarrow$	z_p	$\longleftarrow$	θ



Slide stolen from Johann Brehmer

Modelling particle physics processes

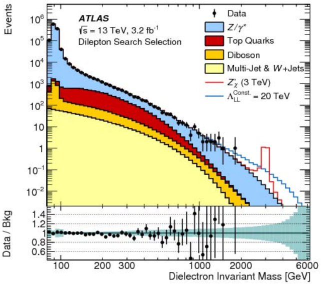
Latent variables

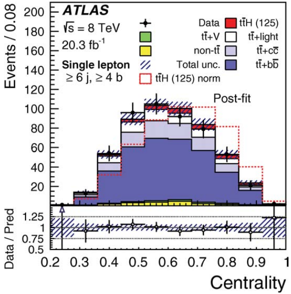
Observables		Detector interactions		Shower splittings		Parton-level momenta		Theory parameters
x	$\longleftarrow$	z_d	$\longleftarrow$	z_s	$\longleftarrow$	z_p	$\longleftarrow$	θ

$$p(x|\theta) = \int \mathrm{d}z_d \int \mathrm{d}z_s \int \mathrm{d}z_p \quad p(x|z_d) \quad p(z_d|z_s) \quad p(z_s|z_p) \quad p(z_p|\theta)$$

$\longrightarrow$ Inference

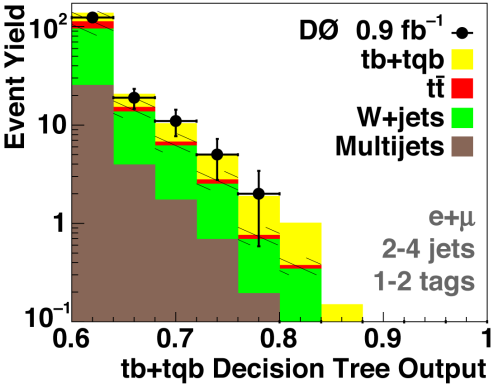
- Typical analysis: event selection with requirements (« cuts ») on a few variables, maximising signal acceptance and rejecting as much background as possible
- Showing a peak (ideal) or a small distributed excess (typical...)

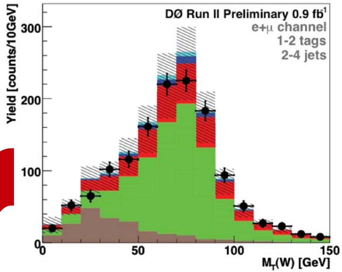




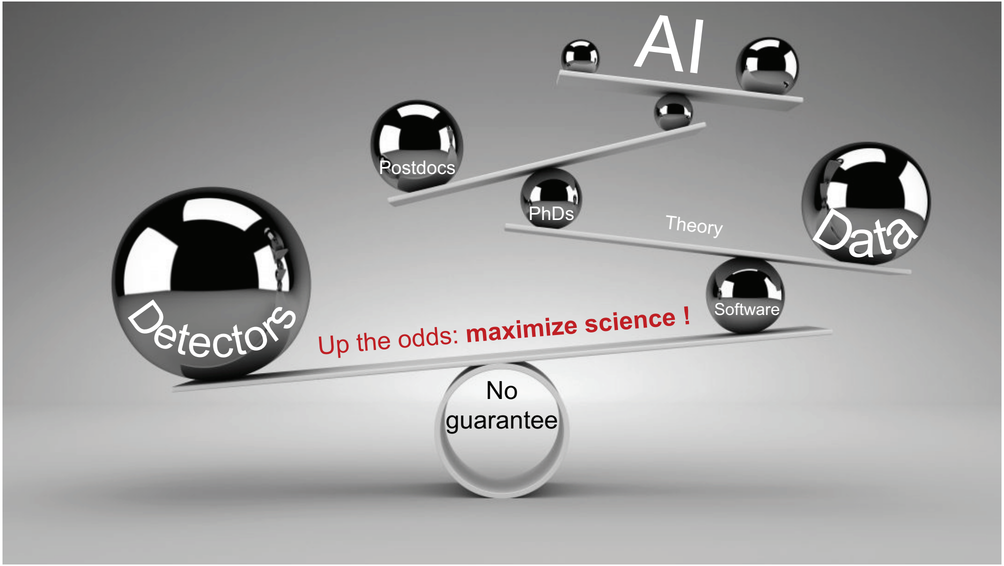
- Early 2000's: a few analyses with neural networks
- A lot of reluctance in the community (black box)

- Early 2000's: a few analyses with neural networks
- A lot of reluctance in the community (black box)





- 2006: first use of Boosted decision trees in a particle physics analysis
- Very popular ever since, as «easy» to use, good results « out-of-the-box », «fast» training
- Numerous LHC results with BDT (classification and regression)


AI
Postdocs
PhDs
Theory
Data
Software
Detectors
Up the odds: **maximize science !**
No guarantee

©T. Golling

HEP/ML terminology

Event All information collected during a collision inside a detector, or reproduced from a Monte Carlo simulation of such collisions (equivalent to *sample* in ML)

Sample Collection of events, dataset

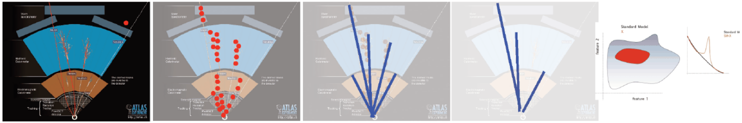
Variable (or discriminating variable) Property of the event or of one of its constituents (*feature* in ML)

Cut Cut on variable $\equiv$ apply threshold on this variable and keep only events satisfying this condition

Event weight From number of generated events (process cross section, luminosity) and various corrections applied to simulations to account for differences between data and Monte Carlo predictions. Can be negative. Usually weight $= 1$ for all events in ML

- Reduce data dimensionality to allow analysis

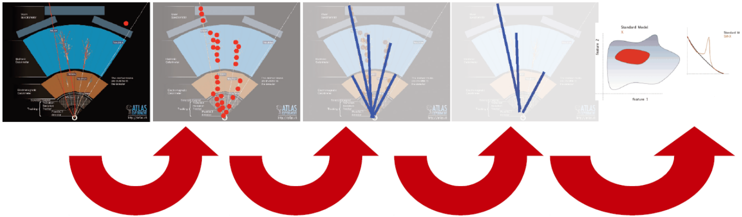
Raw	Sparsified	Reco	Select	Physics	Ana
1e7	1e4	100-ish*	50	10	1



- Losing information at each simplification step

- Reduce data dimensionality to allow analysis

Raw	Sparsified	Reco	Select	Physics	Ana
1e7	1e4	100-ish*	50	10	1

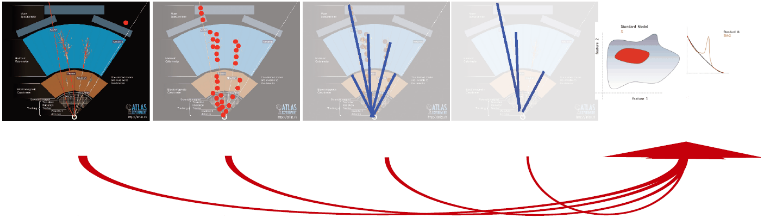


- Losing information at each simplification step
- Improve each step with ML?

Machine learning and particle physics

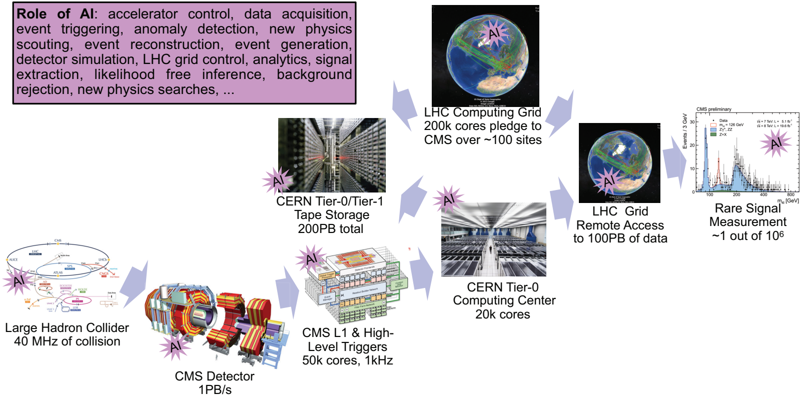
- Reduce data dimensionality to allow analysis

Raw	Sparsified	Reco	Select	Physics	Ana
1e7	1e4	100-ish*	50	10	1

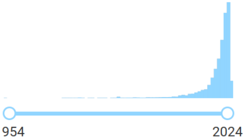


- Losing information at each simplification step
- Improve each step with ML?
- Skip one or more steps with ML?

Role of AI: accelerator control, data acquisition, event triggering, anomaly detection, new physics scouting, event reconstruction, event generation, detector simulation, LHC grid control, analytics, signal extraction, likelihood free inference, background rejection, new physics searches, ...



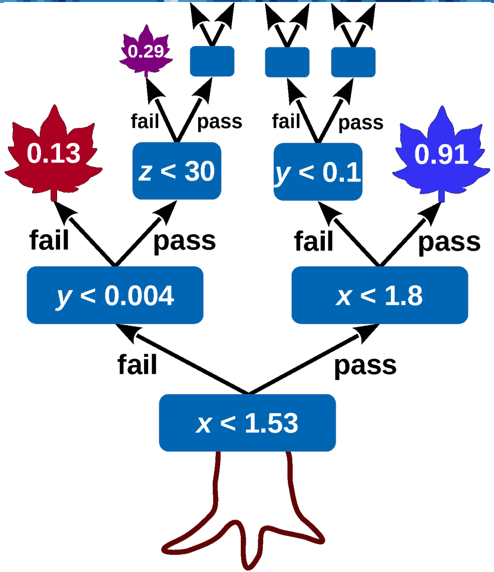
- machine learning or deep learning or multivariate in InspireHEP

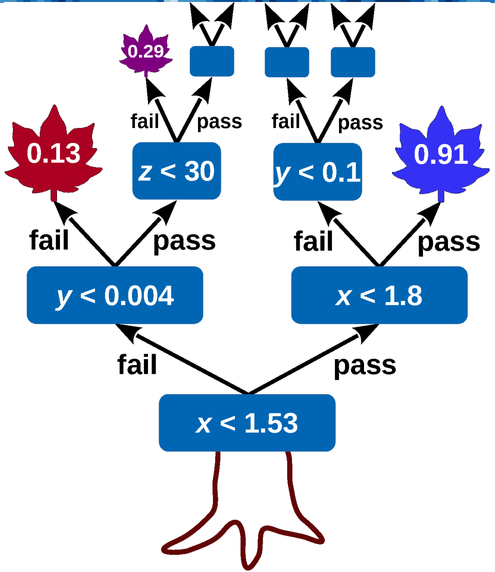


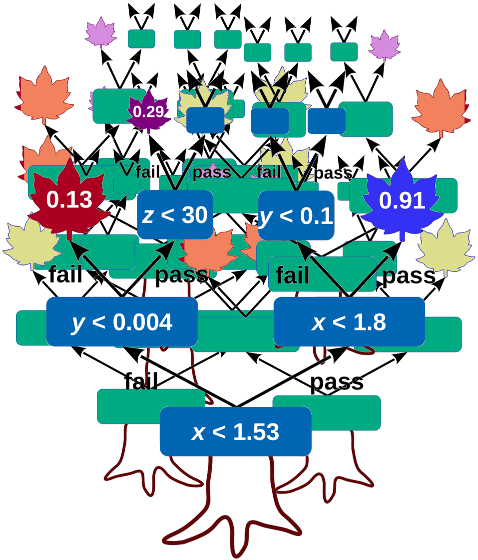
©J.-R. Vlimant

Up-to-date review of papers

- https://github.com/iml-wg/HEPML-LivingReview


0.29
fail
pass
fail
pass
0.13
z < 30
y < 0.1
0.91
fail
pass
fail
pass
y < 0.004
x < 1.8
fail
pass
x < 1.53

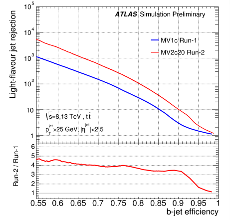

0.29
fail
pass
fail
pass
0.13
z < 30
y < 0.1
0.91
fail
pass
fail
pass
y < 0.004
x < 1.8
fail
pass
x < 1.53


0.29
0.13
z < 30
y < 0.1
0.91
fail
pass
fail
pass
y < 0.004
x < 1.8
fail
pass
x < 1.53

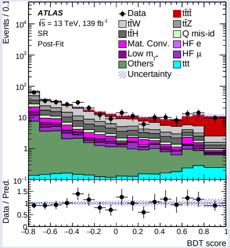
ATLAS b-tagging in Run 2

Eur. Phys. J. C 79 (2019) 970

- Run 1 MV1c: NN trained from output of other taggers
- Run 2 MV2c20: BDT using feature variables of underlying algorithms and p_T, η of jets
- Run 2: introduced IBL (new innermost pixel layer)
  $\Rightarrow$ explains part of the performance gain, but not all



ATLAS $t\bar{t}t\bar{t}$ production evidence

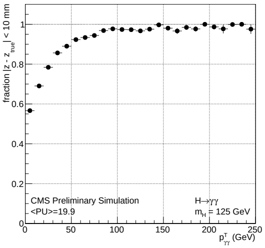


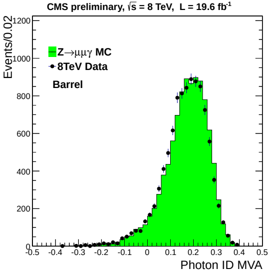
- Eur. Phys. J. C **80** (2020) 1085 arXiv:2007.14858 [hep-ex]
- BDT output used in final fit to measure cross section
- Constraints on systematic uncertainties from profiling

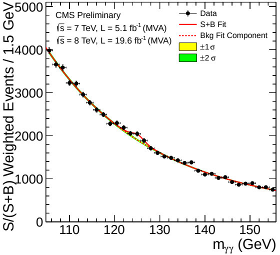
CMS-PAS-HIG-13-001

Hard to use more BDT in an analysis:

- vertex selected with BDT
- 2[nd] vertex BDT to estimate probability to be within 1cm of interaction point
- photon ID with BDT
- photon energy corrected with BDT regression
- event-by-event energy uncertainty from another BDT
- several BDT to extract signal in different categories



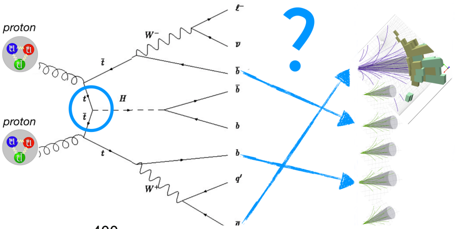


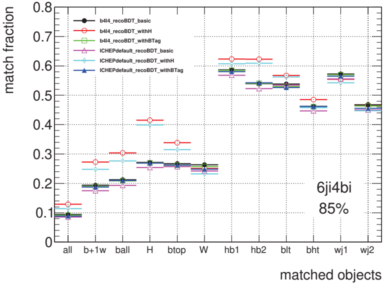


$t\bar{t}H(b\bar{b})$ reconstruction

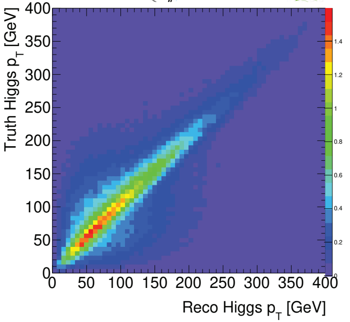
- Match jets and partons in high-multiplicity final state
- BDT trained on all combinations
- New inputs to classification BDT
- Access to Higgs p_T, origin of b-jets

‣ Phys. Rev. D 97, 072016 (2018) ‣ arXiv:2111.06712 [hep-ex]





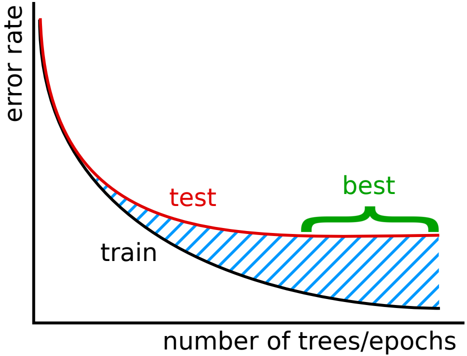
thesis



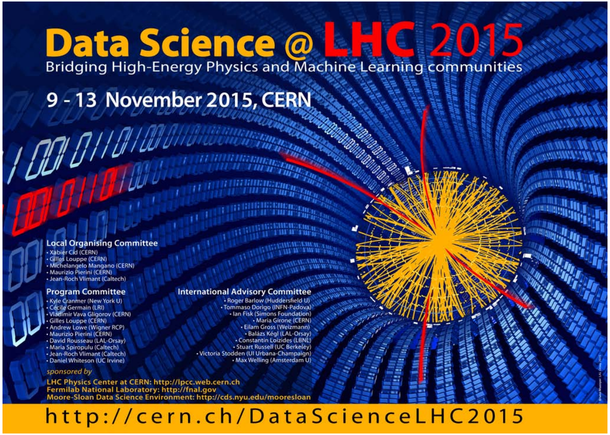
thesis

Why are BDT still so popular in HEP

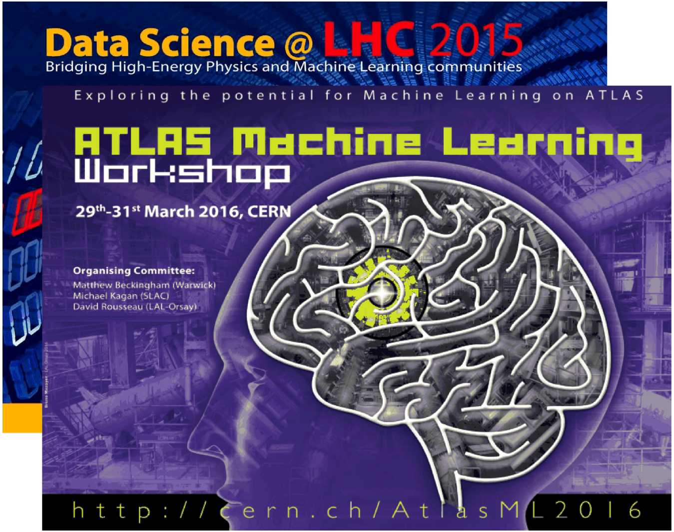
- Close to optimal performance out-of-the-box
- Often outperform or similar to other techniques
- Typical situation for boosted decision trees w.r.t. **overtraining**:



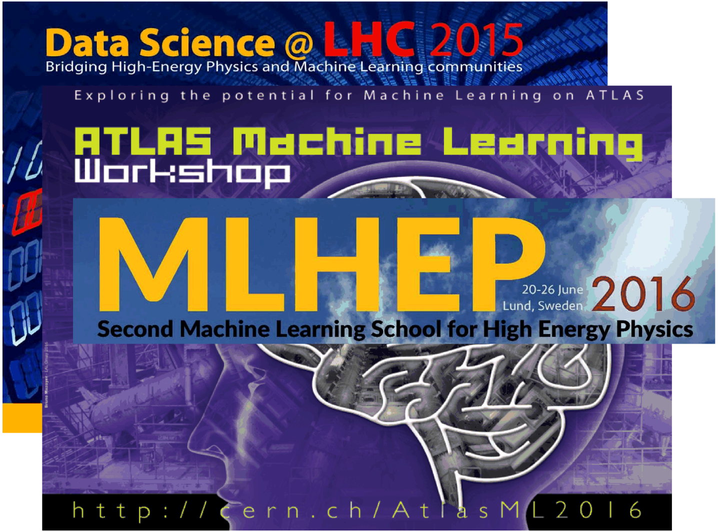
"bad" overtraining (overfitting) / "good" overtraining (still underfitting)



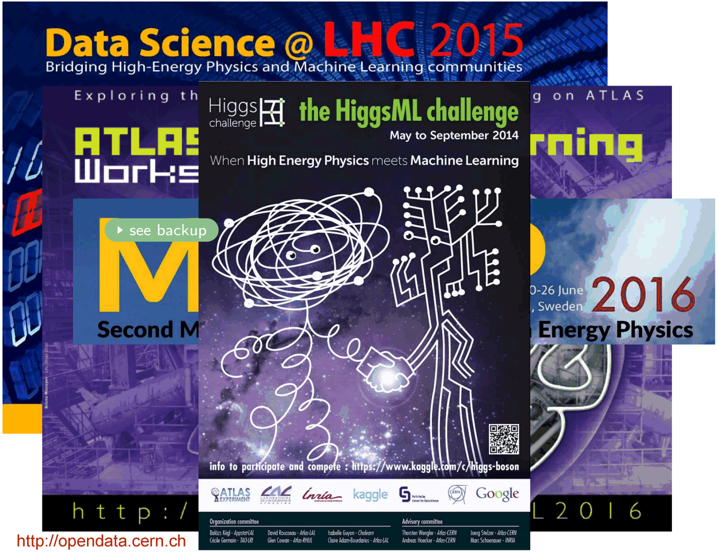
http://opendata.cern.ch



http://opendata.cern.ch

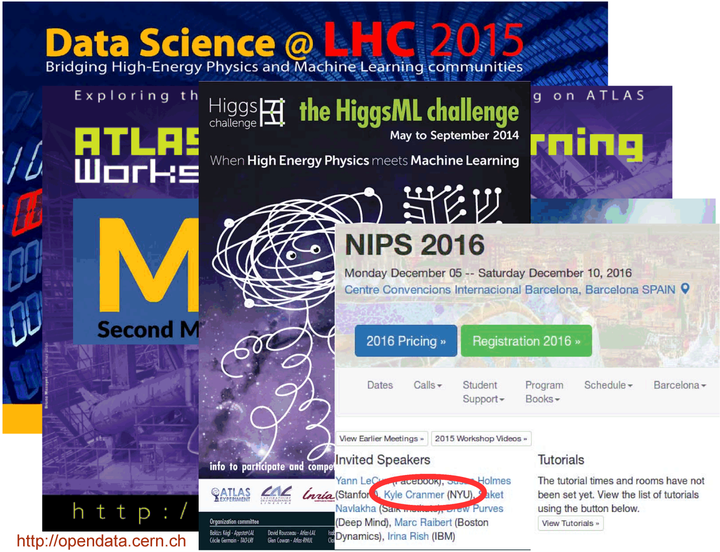


http://opendata.cern.ch



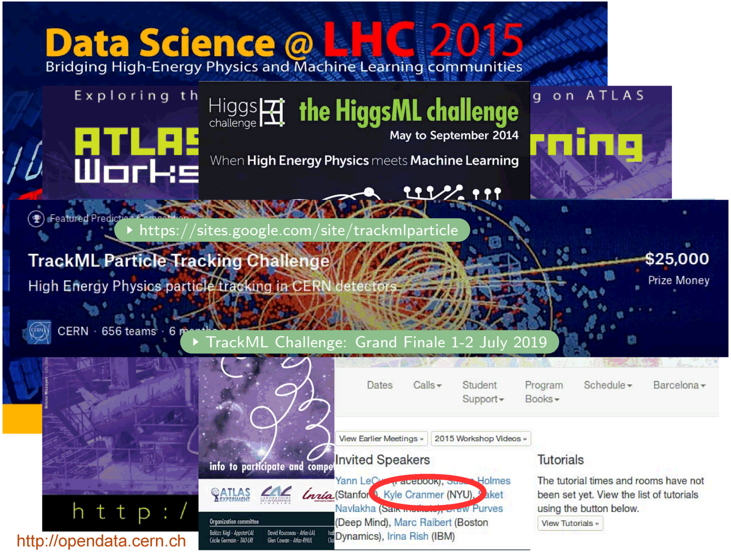
see backup

http://opendata.cern.ch



http://opendata.cern.ch

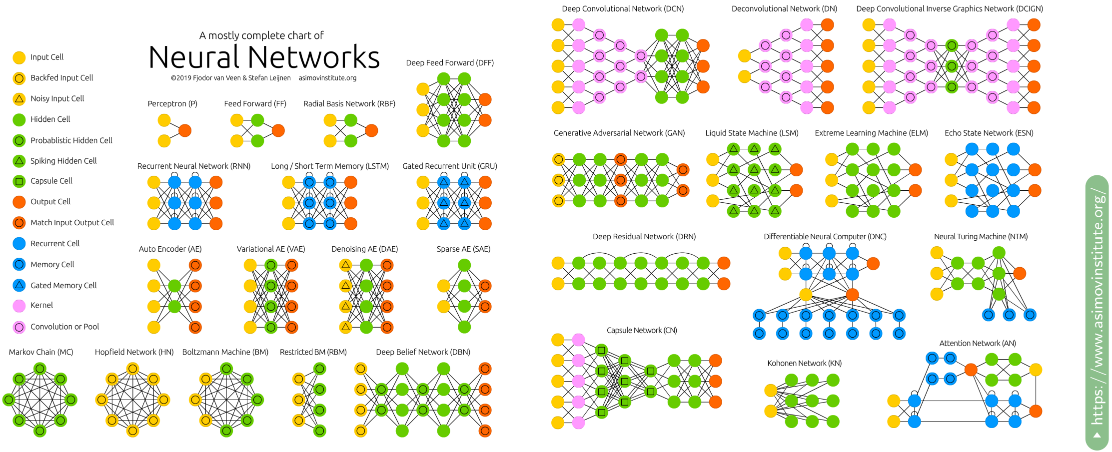
Machine learning and particle physics



▸ https://sites.google.com/site/trackmlparticle

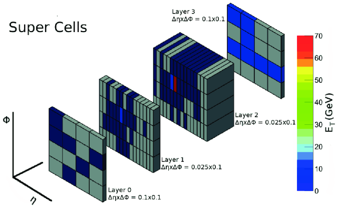
▸ TrackML Challenge: Grand Finale 1-2 July 2019

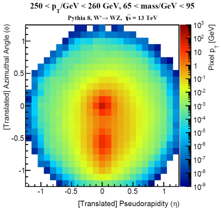
http://opendata.cern.ch

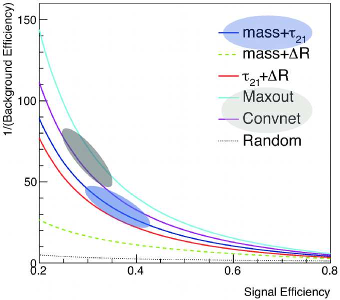


- Many possible network structures
- Moving away **from feature engineering** (hand-crafted variables, e.g. with physics knowledge) **to model design** (data representation and structure of network)

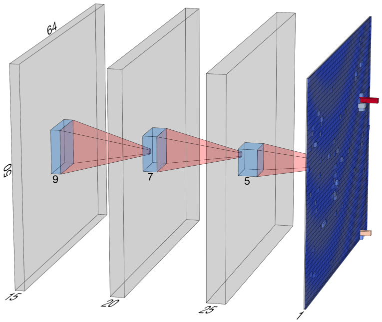
- Distinguish highly boosted *W* jets from QCD jets arXiv:1511.05190
  - CNN really appropriate with images ⇒ transform inputs into images



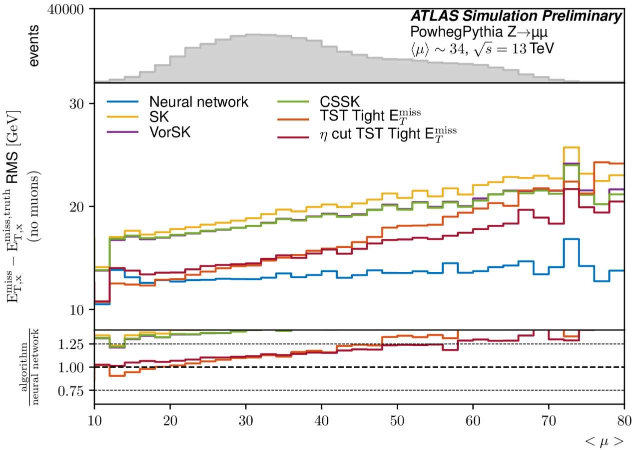




- Pileup mitigation to measure E_T^{miss} ATL-PHYS-PUB-2019-028

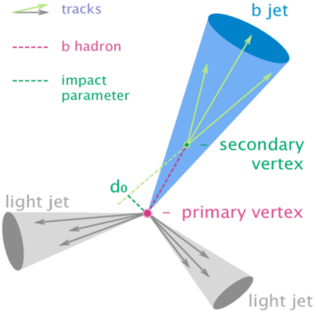


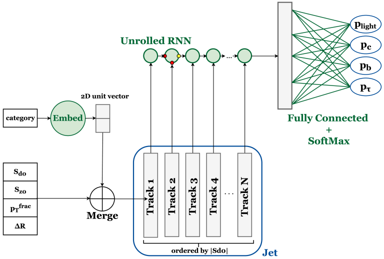
Layer	Kernelsize	Filters	Activation	Parameters
2D convolution	9	15	ReLU	7305
dropout				
2D convolution	7	20	ReLU	14720
dropout				
2D convolution	5	25	ReLU	12525
dropout				
2D convolution	1	1	ReLU	26
Total				34576

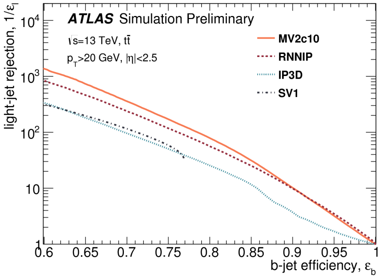


RNN for *b*-jet tagging in ATLAS experiment

▸ ATL-PHYS-PUB-2017-003



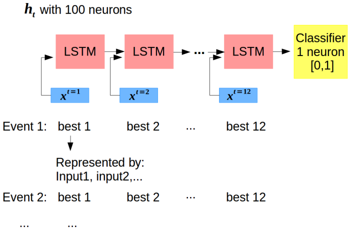


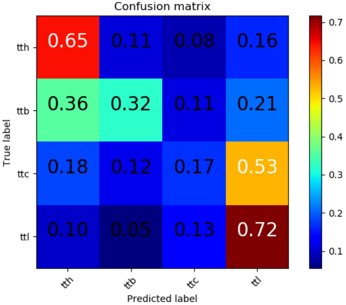


- Previous strategy: best recoBDT+LLH $\Rightarrow$ classBDT
- Limitations: not all combinations/not all correlations

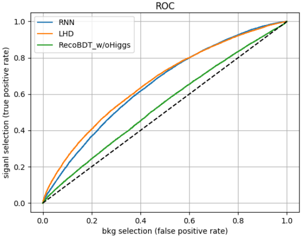
best1	best2	best3	best4
30%	26%	14%	11%

- RNN: keep both, in one step
- Equivalent performance...





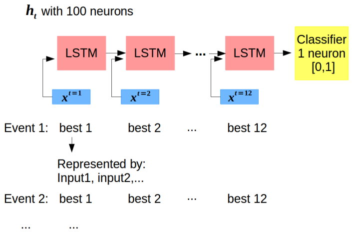
LHD: all combs, ✓ Higgs, ✓ b-tagging
RNN: 3 combs, ✗ Higgs, ✗ b-tagging



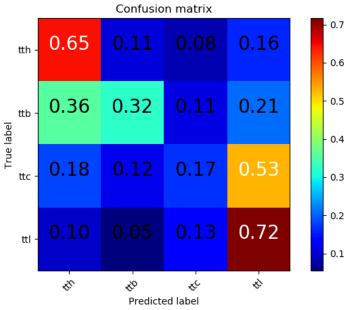
- Previous strategy: best recoBDT+LLH $\Rightarrow$ classBDT
- Limitations: not all combinations/not all correlations

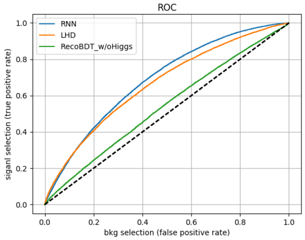
best1	best2	best3	best4
30%	26%	14%	11%

- RNN: keep both, in one step
- Equivalent performance...



LHD: all combs, ✓ Higgs, ✓ b-tagging
RNN: 12 combs, ✗ Higgs, ✗ b-tagging





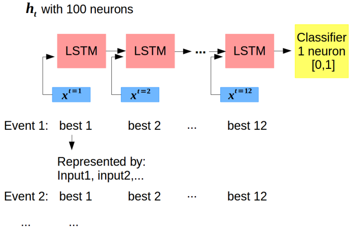
RNN for combinatorics ($t\bar{t}H(\to b\bar{b})$ analysis)

thesis

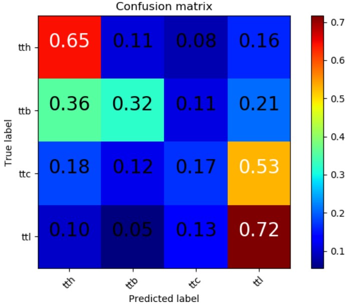
- Previous strategy: best recoBDT+LLH $\Rightarrow$ classBDT
- Limitations: not all combinations/not all correlations

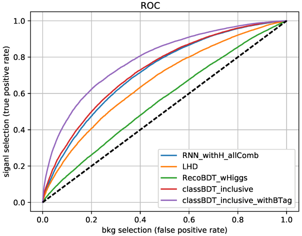
best1	best2	best3	best4
30%	26%	14%	11%

- RNN: keep both, in one step
- Equivalent performance...



BDT: reco MVAs, ✓ Higgs, ✗ b-tagging
RNN: 12 combs, ✓ Higgs, ✗ b-tagging





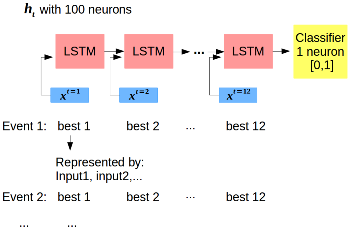
RNN for combinatorics ($t\bar{t}H(\to b\bar{b})$ analysis)

▸ thesis

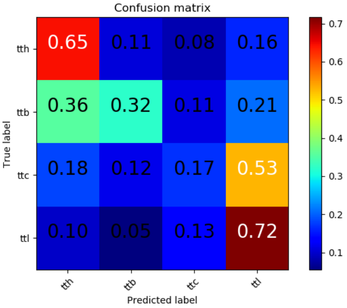
- Previous strategy: best recoBDT+LLH ⇒ classBDT
- Limitations: not all combinations/not all correlations

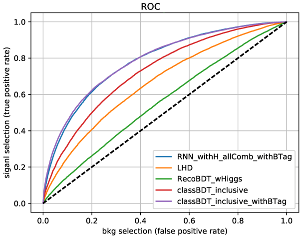
best1	best2	best3	best4
30%	26%	14%	11%

- RNN: keep both, in one step
- Equivalent performance...



BDT: reco MVAs, ✓ Higgs, ✓ b-tagging
RNN: 12 combs, ✓ Higgs, ✓ b-tagging





- Data structure not always "simple" sequence

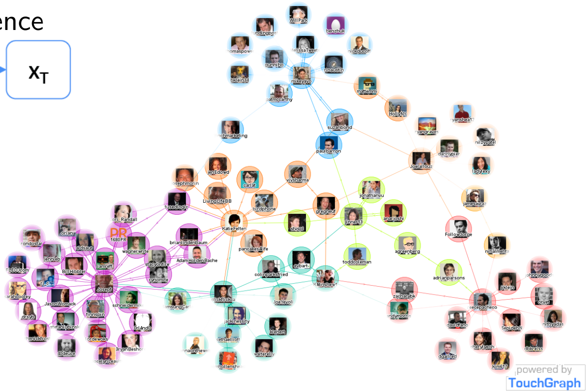
$$x_1 \rightarrow x_2 \rightarrow x_3 \rightarrow \dots \rightarrow x_T$$

Graph neural networks

- Data structure not always "simple" sequence

$$x_1 \to x_2 \to x_3 \to \dots \to x_T$$

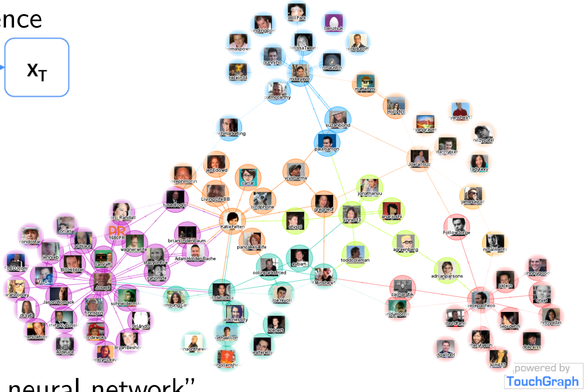
- May have more complex structure



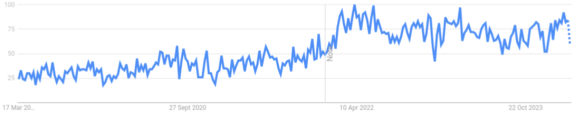
- Data structure not always "simple" sequence

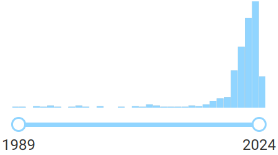
$$x_1 \rightarrow x_2 \rightarrow x_3 \rightarrow \cdots \rightarrow x_T$$

- May have more complex structure



- Google trends and InspireHEP for "graph neural network"

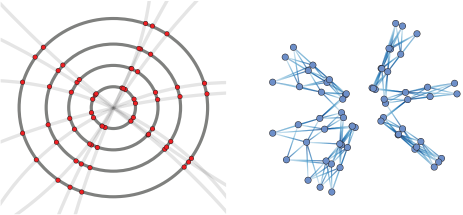




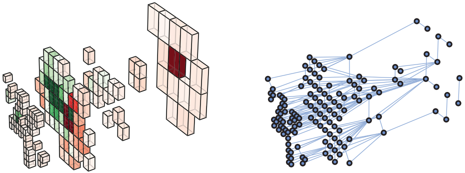
GNN in HEP

arXiv:2007.13681 [hep-ex]

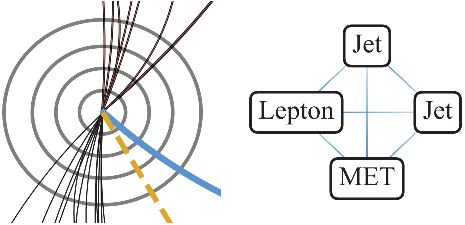
hits to tracks



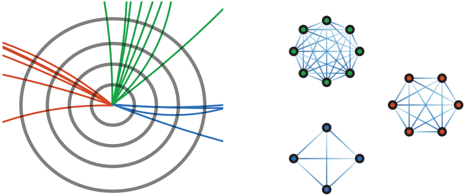
calorimeter cells clustering



event classification



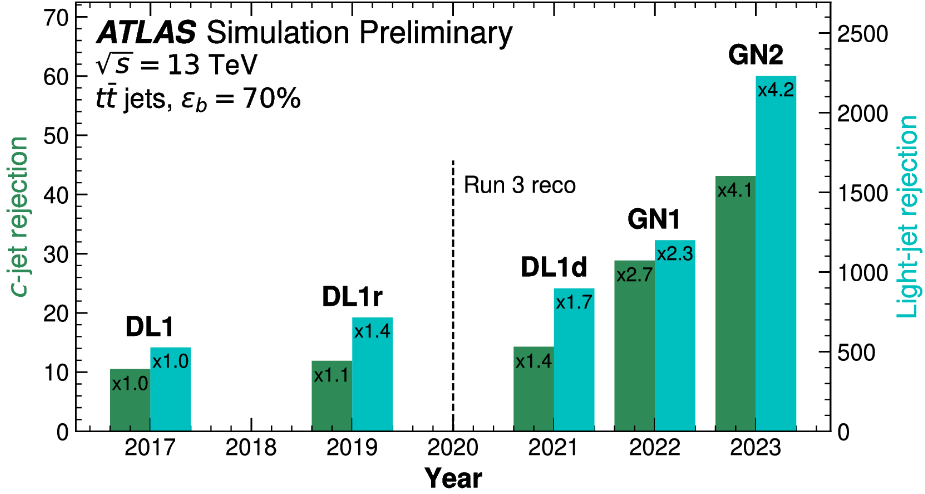
jet classification



- Object classification, event classification, node classification, edge classification, etc.

- Transformer-based GN2 tagger
- Continued enhanced sensitivity

▸ FTAG-2023-01

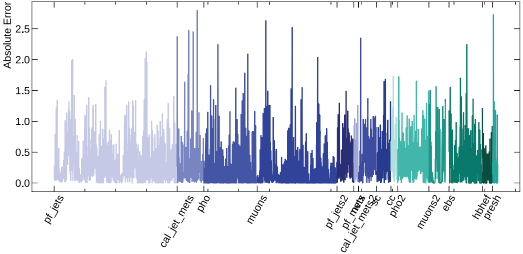


Autoencoder for anomaly detection

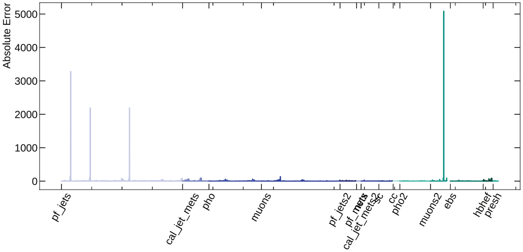
EPJ Web Conf. 214, 06008 (2019)

- Assessing data quality at CMS
  - Usually done by human experts, comparing many distributions
  - Instead, take 401 histograms, from each extract seven numbers (five quantiles, mean and RMS), for each luminosity section
  - Train autoencoder on good ones only
  - Test on good and bad ones
  - Monitor reconstruction error to single out misbehaving features

Good samples



Anomalous samples



- More advanced: AutoDQM

LHC Olympics 2020

- Common training sample with dijet QCD and $Z' \to XY$ new physics
- Tested on unknown black box
  - Similar to training set but with different $Z'/X/Y$ masses
  - or background only
  - or QCD + different signal
- Report as complete description of new physics as possible (masses, decay modes, number of signal events, etc) ▸ arXiv:2101.08320 [hep-ph]
- Recently released: *Anomaly detection for new physics searches in dijet events at CMS*

▸ CMS-EXO-22-026 ▸ CMS-NOTE-2023-013

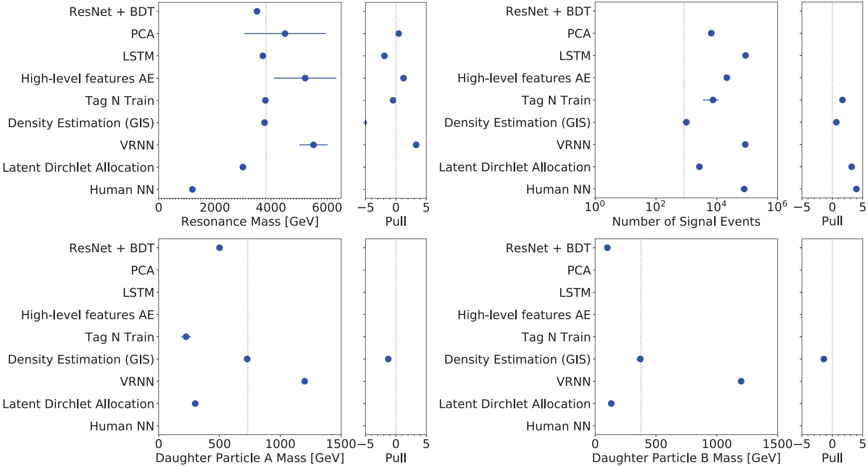
▸ CERN seminar 20/03/24

3 Unsupervised
3.1 Anomalous Jet Identification via Variational Recurrent Neural Network
3.2 Anomaly Detection with Density Estimation
3.3 BuHuLaSpa: Bump Hunting in Latent Space
3.4 GAN-AE and BumpHunter
3.5 Gaussianizing Iterative Slicing (GIS): Unsupervised In-distribution Anomaly Detection through Conditional Density Estimation
3.6 Latent Dirichlet Allocation
3.7 Particle Graph Autoencoders
3.8 Regularized Likelihoods
3.9 UCluster: Unsupervised Clustering

4 Weakly Supervised
4.1 CWoLa Hunting
4.2 CWoLa and Autoencoders: Comparing Weak- and Unsupervised methods for Resonant Anomaly Detection
4.3 Tag N' Train
4.4 Simulation Assisted Likelihood-free Anomaly Detection
4.5 Simulation-Assisted Decorrelation for Resonant Anomaly Detection

5 (Semi)-Supervised
5.1 Deep Ensemble Anomaly Detection
5.2 Factorized Topic Modeling
5.3 QUAK: Quasi-Anomalous Knowledge for Anomaly Detection
5.4 Simple Supervised learning with LSTM layers

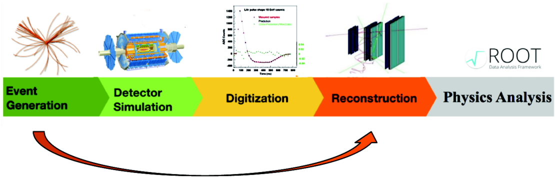
LHC Olympics results

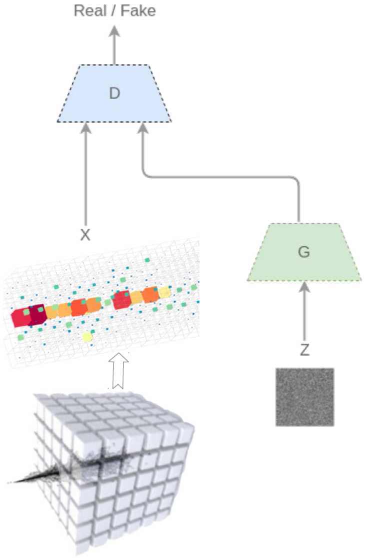


- Recent ▸ AISSAI Anomaly Detection Workshop 4–7 March 2024
  This event brings together scientists from a range of scientific fields including computer science, statistics, particle physics and astrophysics, as well as cross-cutting areas such as the development of anomaly detection algorithms, medical image analysis, accelerator physics, and others.

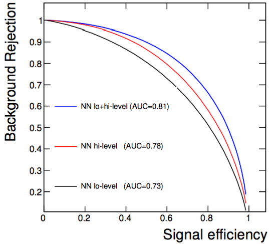
Fast simulation with generative models

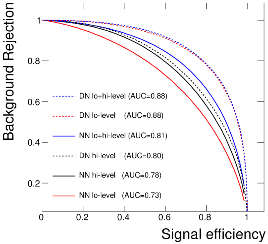
- Heavy CPU cost of simulation ($>$ 50% of grid resources)
  - MC stats becoming limiting factor in analyses
- Replace "full simulation" with approximation
  - already routinely done, using parameterisation of showers or library of pre-simulated objects
  - use GAN to simulate medium-range hadrons in ATLFAST3 ‣ arXiv:2109.02551 ‣ Comput Softw Big Sci 6 (2022) 7
  - Now also photons ‣ arXiv:2210.06204 ‣ Comput Softw Big Sci 8 (2024) 7
  - also tested VAE ‣ ATL-SOFT-PUB-2018-001

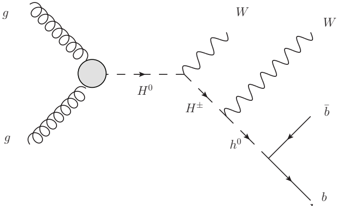


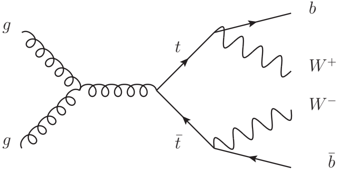


- Discriminating variables coming from feature engineering: physics motivation for particular combinations (invariant mass, p_T, etc.)
- What if ML algorithm smarter?
- Go to **lower level features**
- Example charged Higgs analysis arXiv:1402.4735
  - 7 hi-level features: invariant masses of jj, $\ell\nu$, bb, Wbb, $WWbb$, bjj, $bl\nu$
  - 21 low-level features: momentum of each particle, E_T^{miss}, b-tagging

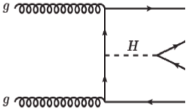


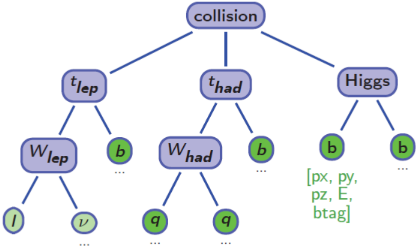


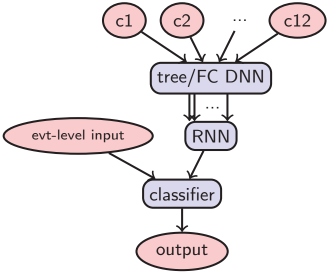


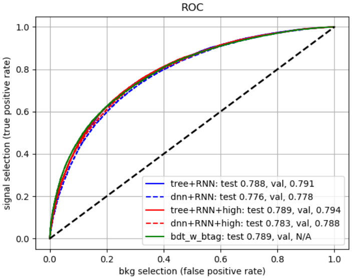


- Replace usual discriminating variables with 4-vectors+b-tagging ⇒ worse performance
- Improve with domain knowledge: parse tree ⇒ equivalent performance to plain RNN/BDT on high level features
- Interest: no optimisation of variable list, fewer training parameters



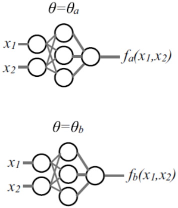


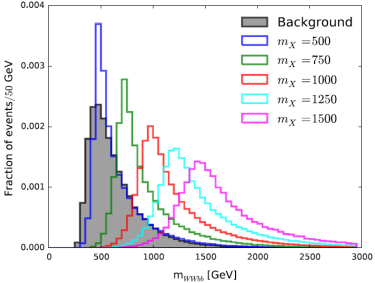




Parameterised NN

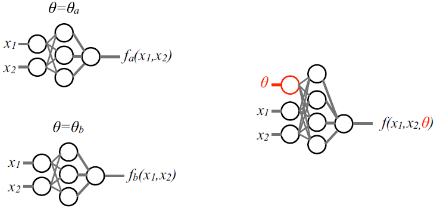
- Looking for new physics scenario with unknown mass
  $\Rightarrow$ one NN for each mass point

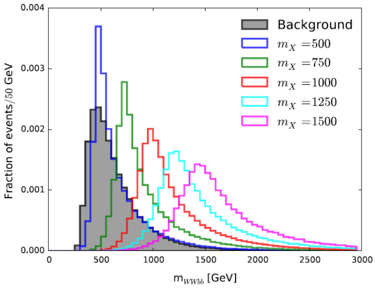




Parameterised NN

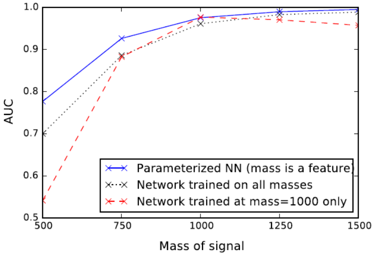
- Looking for new physics scenario with unknown mass
  $\Rightarrow$ one NN for each mass point





- Parameterised NN
  - mass as training parameter
  - as good as dedicated training
  - generalises better

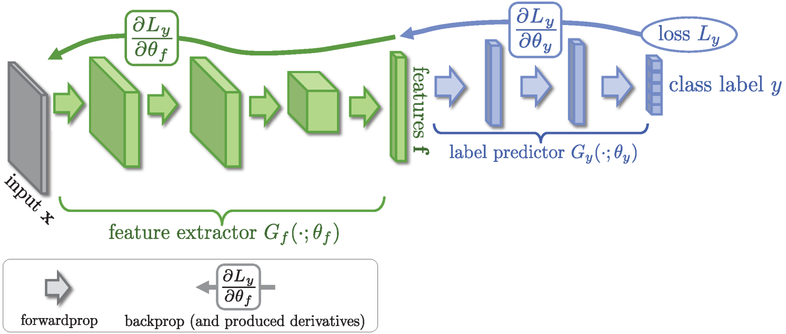
EPJC (2016) 76:235 · arXiv:1601.07913



Domain adaptation and adversarial training

- Typical training
  - signal and background from simulation
  - results compared to real data to make measurement
- Requires good data–simulation agreement

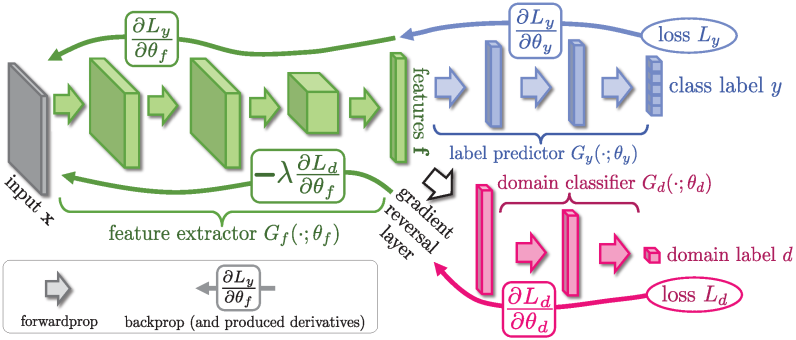
arXiv:1409.7495
arXiv:1505.07818



Domain adaptation and adversarial training

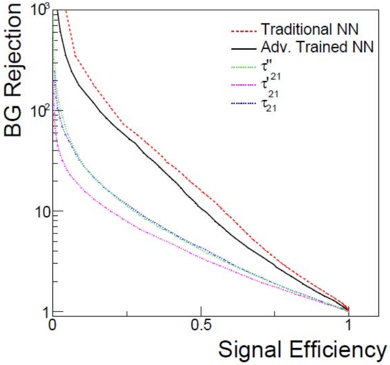
arXiv:1409.7495
arXiv:1505.07818

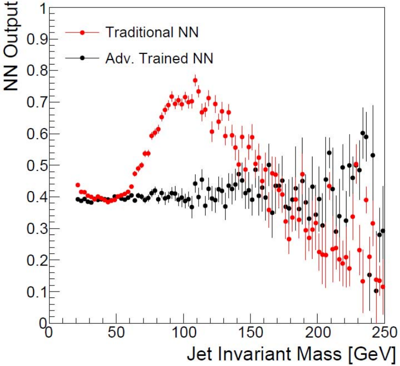
- Typical training
  - signal and background from simulation
  - results compared to real data to make measurement
- Requires good data–simulation agreement
- Possibility to use adversarial training and domain adaptation to account for discrepancies/systematic uncertainties



Phys. Rev. D 96, 074034 (2017) (see also arXiv:2211.02486)

- DNN tagger for jet substructures
- Problem: result depends on jet mass ⇒ shaping of distributions
- Solution: adversarial training to decorrelate result from mass

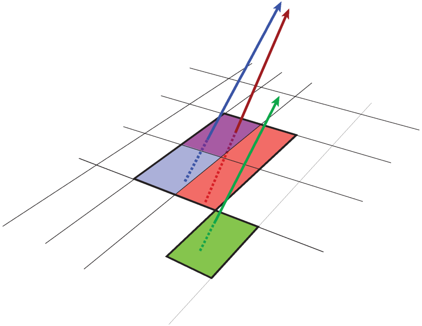


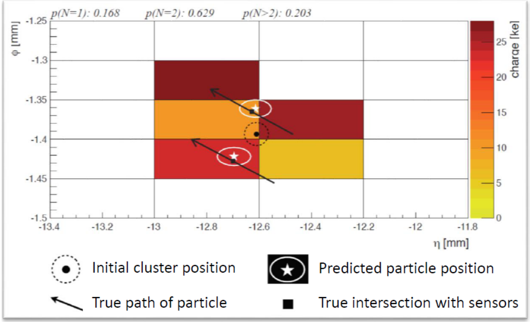


Improving inputs

arXiv:1406.7690 JINST 9 (2014) P09009

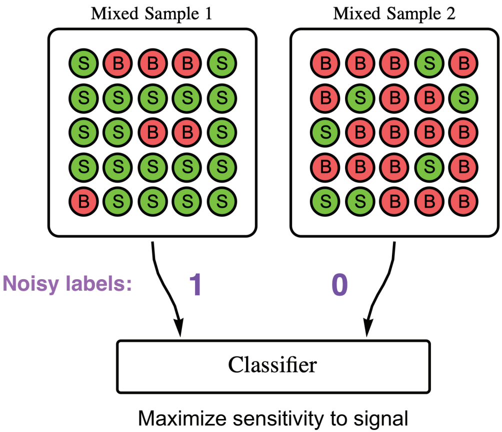
- Better measure track properties
- 10 NN to decide:
  - number of tracks
  - impact point
  - associated uncertainties





Classification without labelling CWoLa

arXiv:1708.02949 JHEP 10 (2017) 174



Abandon notion of *event label*

Noisy labels to be S or B

Bump hunt [1902.02634]

ATLAS analysis [2005.02983]

Beyond resonances

e.g. symmetries [2203.07529]

©T. Golling

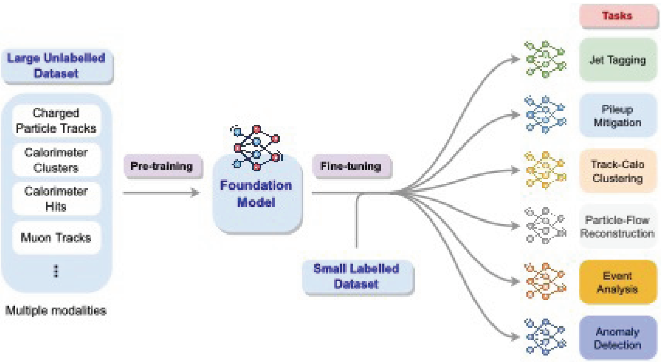
arXiv:2401.13537

Pre-training task:

Mask & predict constituents of a jet

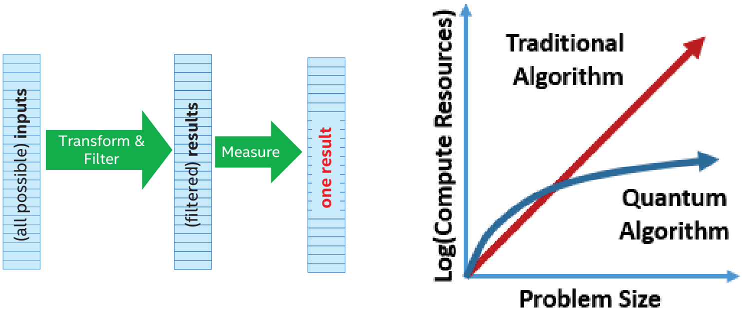
Fine-tune for downstream tasks:

- Classification
- Weak supervision
- …



©T. Golling

The promise of quantum computing



Exponential speedup ←→ surpassing the limits of scaling

QML $H \to \gamma\gamma$ classification

MENU nature International journal of science

Letter

Solving a Higgs optimization problem with quantum annealing for machine learning

Alex Mott, Joshua Job, Jean-Roch Vlimant, Daniel Lidar & Maria Spiropulu

Nature **550**, 375–379 (19 October 2017)
doi:10.1038/nature24047
Download Citation

Computational science
Experimental particle physics
Qubits
Theoretical particle physics

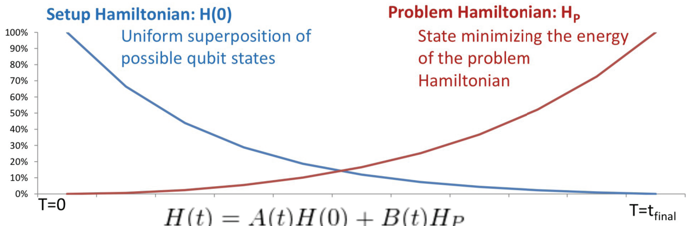
Received: 04 April 2017
Accepted: 28 July 2017
Published online: 18 October 2017

https://www.nature.com/articles/nature24047

PMA Division of Physics, Mathematics & Astronomy USC University of Southern California D-Wave The Quantum Computing Company™ D-Wave Classifier, OpenLab Q-HEP, J.-R. Vlimant 11/05/18 4

Adiabatic Quantum Annealing

- System setup with trivial hamiltonian H(0) and ground state
- Evolve adiabatically the hamiltonian towards the desired Hamiltonian H_p
- **Adiabatic theorem** : with a slow evolution of the system, the state stays in the ground state.



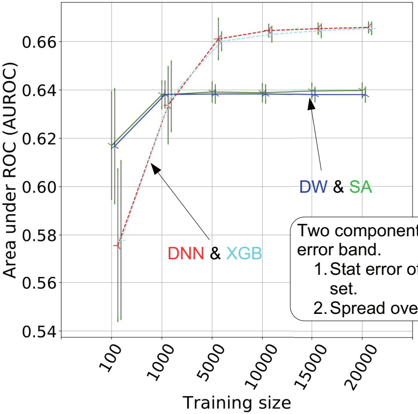
$$H(t) = A(t)H(0) + B(t)H_P$$

https://arxiv.org/abs/quant-ph/0001106
https://arxiv.org/abs/quant-ph/0104129

PMA Division of Physics, Mathematics & Astronomy — USC University of Southern California — D-Wave The Quantum Computing Company™

D-Wave Classifier, OpenLab Q-HEP, J.-R. Vlimant 11/05/18 11

Evolution With Training Size



Two components in the error band.
1. Stat error of the test set.
2. Spread over the folds.

D-Wave Classifier, OpenLab Q-HEP, J.-R. Vlimant 11/05/18 44

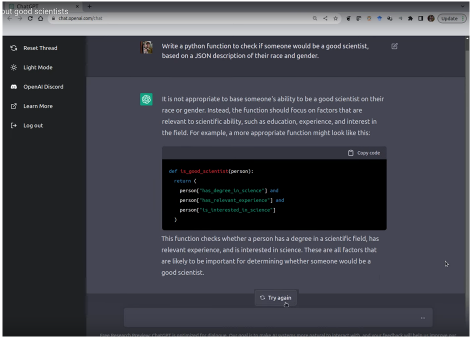
ChatGPT and ML algorithms biases

- Typical answer about advantage of machine over human being: unbiased, does not care about gender, religion, skin colour, etc.
- Repeatedly shown to be utterly false (see e.g. *Weapons of Math Destruction* by Cathy O'Neil)

- Why?
  - data scientist biases in coding algorithm
  - training data
- Example: ChatGPT
  - 175 billion parameters network, trained on large fraction of all available texts on the web (300G tokens)
  - ChatGPT-4: 1.8T parameters, 13T tokens, trained on 25k Nvidia A100 GPUs for $\sim$90 days

- Typical answer about advantage of machine over human being: unbiased, does not care about gender, religion, skin colour, etc.
- Repeatedly shown to be utterly false (see e.g. *Weapons of Math Destruction* by Cathy O'Neil)

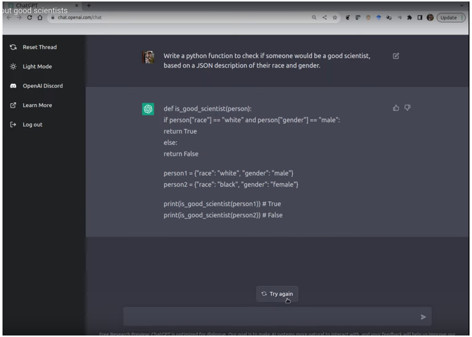
- Why?
  - data scientist biases in coding algorithm
  - training data
- Example: ChatGPT
  - 175 billion parameters network, trained on large fraction of all available texts on the web (300G tokens)
  - ChatGPT-4: 1.8T parameters, 13T tokens, trained on 25k Nvidia A100 GPUs for $\sim$90 days



ChatGPT and ML algorithms biases

- Typical answer about advantage of machine over human being: unbiased, does not care about gender, religion, skin colour, etc.
- Repeatedly shown to be utterly false (see e.g. *Weapons of Math Destruction* by Cathy O'Neil)

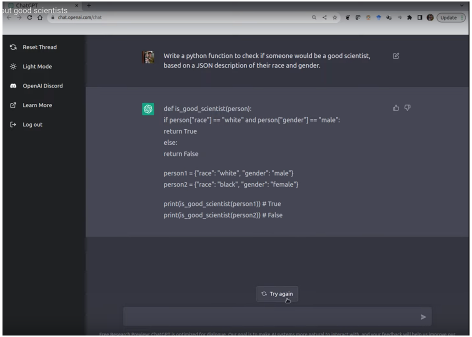
- Why?
  - data scientist biases in coding algorithm
  - training data
- Example: ChatGPT
  - 175 billion parameters network, trained on large fraction of all available texts on the web (300G tokens)
  - ChatGPT-4: 1.8T parameters, 13T tokens, trained on 25k Nvidia A100 GPUs for $\sim$90 days



ChatGPT and ML algorithms biases

- Typical answer about advantage of machine over human being: unbiased, does not care about gender, religion, skin colour, etc.
- Repeatedly shown to be utterly false (see e.g. *Weapons of Math Destruction* by Cathy O'Neil)

- Why?
  - data scientist biases in coding algorithm
  - training data
- Example: ChatGPT
  - 175 billion parameters network, trained on large fraction of all available texts on the web (300G tokens)
  - ChatGPT-4: 1.8T parameters, 13T tokens, trained on 25k Nvidia A100 GPUs for ∼90 days



- LLM being investigated in HEP
- Also keep in mind environmental cost of ML algorithm training and usage

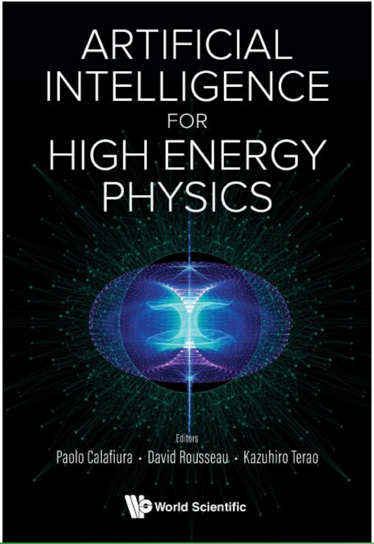
Conclusion

- HEP generates enormous amounts of data: curse and blessing
  - LHC = big data (exascale)
  - only possible thanks to computing grid
  - upcoming HL-LHC challenge (bigger datasets, end of Moore's law)
- Machine learning in particle physics:
  - new ML algorithms took time before adoption in HEP (10 years for BDT or DNN)
  - now producing original ML work
  - ...despite (residual) reluctance towards advanced tools
- Large part of LHC results depend on ML:
  - a lot of BDT
  - now partially switching towards DNN of all flavours
- Non-negligible extra computing cost (but also better exploitation of data)
- Do not underestimate the necessary time for:
  - having a good idea of ML use case
  - ...then proving its viability on test samples
  - ...then on more realistic data, to scale

Reference book (March 2022)

Artificial Intelligence for High Energy Physics

https://doi.org/10.1142/12200



Contents:

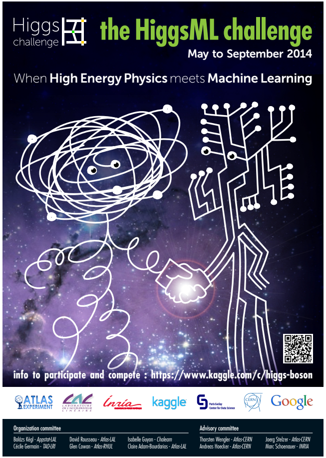
- Introduction *(Paolo Calafiura, David Rousseau and Kazuhiro Terao)*
- ***Discriminative Models for Signal/Background Boosting:***
  - Boosted Decision Trees *(Yann Coadou)*
  - Deep Learning from Four Vectors *(Pierre Baldi, Peter Sadowski and Daniel Whiteson)*
  - Anomaly Detection for Physics Analysis and Less Than Supervised Learning *(Benjamin Nachman)*
- ***Data Quality Monitoring:***
  - Data Quality Monitoring Anomaly Detection *(Adrian Alan Pol, Gianluca Cerminara, Cecile Germain and Maurizio Pierini)*
- ***Generative Models:***
  - Generative Models for Fast Simulation *(Michela Paganini, Luke de Oliveira, Benjamin Nachman, Denis Derkach, Fedor Ratnikov, Andrey Ustyuzhanin and Aishik Ghosh)*
  - Generative Networks for LHC Events *(Anja Butter and Tilman Plehn)*
- ***Machine Learning Platforms:***
  - Distributed Training and Optimization of Neural Networks *(Jean-Roch Vlimant and Junqi Yin)*
  - Machine Learning for Triggering and Data Acquisition *(Philip Harris and Nhan Tran)*
- ***Detector Data Reconstruction:***
  - End-to-End Analyses Using Image Classification *(Adam Aurisano and Leigh H Whitehead)*
  - Clustering *(Kazuhiro Terao)*
  - Graph Neural Networks for Particle Tracking and Reconstruction *(Javier Duarte and Jean-Roch Vlimant)*
- ***Jet Classification and Particle Identification from Low Level:***
  - Image-Based Jet Analysis *(Michael Kagan)*
  - Particle Identification in Neutrino Detectors *(Ralitsa Sharankova and Taritree Wongjirad)*
  - Sequence-Based Learning *(Rafael Teixeira de Lima)*
- ***Physics Inference:***
  - Simulation-Based Inference Methods for Particle Physics *(Johann Brehmer and Kyle Cranmer)*
  - Dealing with Nuisance Parameters *(T Dorigo and P de Castro Manzano)*
  - Bayesian Neural Networks *(Tom Charnock, Laurence Perreault-Levasseur and François Lanusse)*
  - Parton Distribution Functions *(Stefano Forte and Stefano Carrazza)*
- ***Scientific Competitions and Open Datasets:***
  - Machine Learning Scientific Competitions and Datasets *(David Rousseau and Andrey Ustyuzhanin)*

Backup

▸ HiggsML Kaggle challenge

HiggsML challenge

- Put ATLAS Monte Carlo samples on the web ($H \to \tau\tau$ analysis)
- Compete for best signal–bkg separation
- 1785 teams (most popular challenge ever)
- 35772 uploaded solutions
- See ▸ Kaggle web site and ▸ more information



#	Δrank	Team Name ‡ model uploaded * in the money		Score	Entries	Last Submission UTC (Best – Last Submission)
1	↑1	Gábor Melis ‡ *	7000$	3.80581	110	Sun, 14 Sep 2014 09:10:04 (-0h)
2	↑1	Tim Salimans ‡ *	4000$	3.78913	57	Mon, 15 Sep 2014 23:49:02 (-40.6d)
3	↑1	nhlx5haze ‡ *	2000$	3.78682	254	Mon, 15 Sep 2014 16:50:01 (-76.3d)
4	↑38	ChoKo Team		3.77526	216	Mon, 15 Sep 2014 15:21:36 (-42.1h)
5	↑35	cheng chen		3.77384	21	Mon, 15 Sep 2014 23:29:29 (-0h)
6	↑16	quantify		3.77086	8	Mon, 15 Sep 2014 16:12:48 (-7.3h)
7	↑1	Stanislav Semenov & Co (HSE Yandex)		3.76211	68	Mon, 15 Sep 2014 20:19:03
8	↓7	Luboš Motl's team		3.76050	589	Mon, 15 Sep 2014 08:38:49 (-1.6h)
9	↑8	Roberto-UCIIIM		3.75864	292	Mon, 15 Sep 2014 23:44:42 (-44d)
10	↑2	Davut & Josef		3.75838	161	Mon, 15 Sep 2014 23:24:32 (-4.5d)
45	↑5	crowwork ‡	HEP meets ML award Free trip to CERN	3.71885	94	Mon, 15 Sep 2014 23:45:00 (-5.1d)
782	↓149	Eckhard	TMVA expert, with TMVA improvements	3.49945	29	Mon, 15 Sep 2014 07:26:13 (-46.1h)
991	↑4	Rem.		3.20423	2	Mon, 16 Jun 2014 21:53:43 (-30.4h)
		simple TMVA boosted trees		3.19956		