Deep Learning in High Energy Physics

Improving the Search for Exotic Particles

P.Baldi, P. Sadowski, and D. Whiteson

Department of Computer Science
Department of Physics
Center for Machine Learning and Intelligent Systems

Deep Learning in HEP

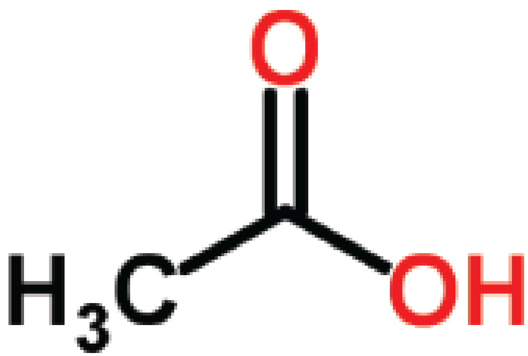
Peter Sadowski

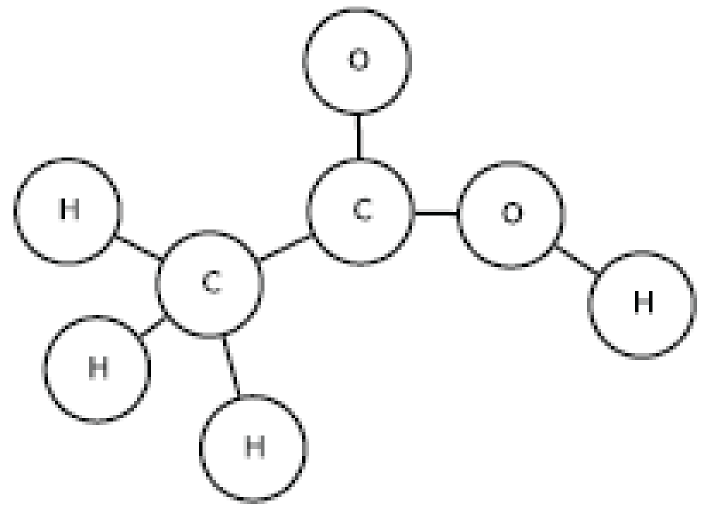
Daniel Whiteson

Machine Learning (DL) in the Natural Sciences

- Physics
  - -**HEP**
  - -QM
  - -Astronomy
- Chemistry
  - -Prediction of Molecular or Material Properties
  - -Prediction of Chemical Reactions
- Earth Sciences
- Biology
  - -Prediction of Protein Structures
  - -Prediction of gene Expression
  - -Biomedical imaging

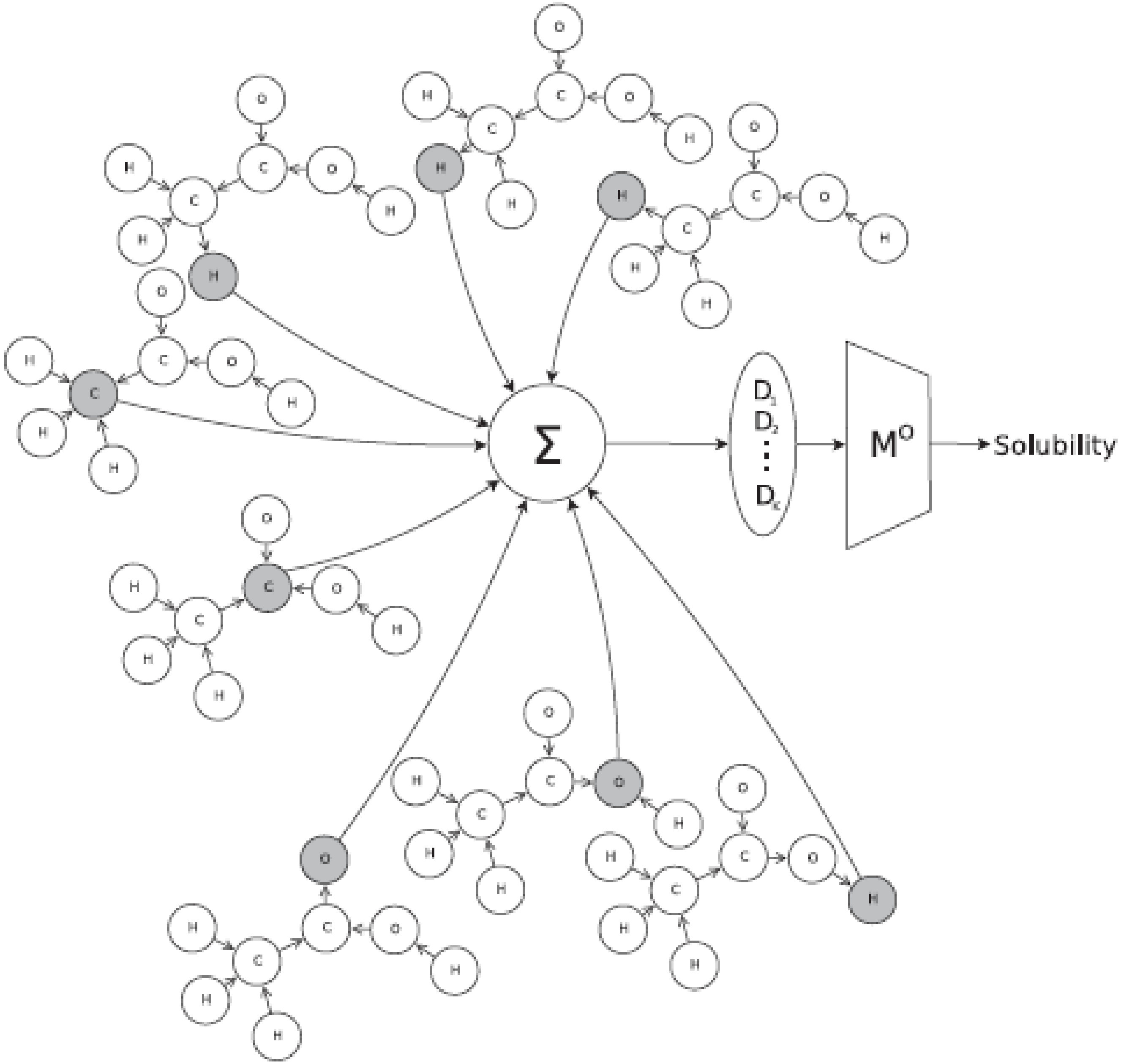
Deep Learning in Chemistry





CC(=O)O

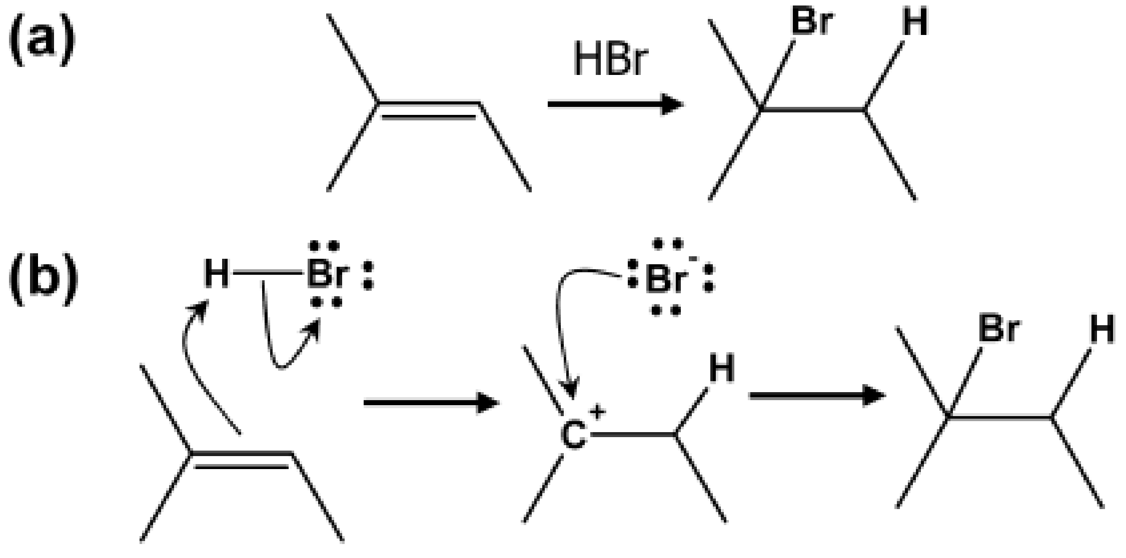
Acetic Acid



A. Lusci, G. Pollastri, and P. Baldi. Deep Architectures and Deep Learning in Chemoinformatics: the Prediction of Aqueous Solubility for Drug-Like Molecules. *Journal of Chemical Information and Modeling*, 53, 7, 1563–1575, (2013).

Deep Learning Chemical Reactions

$RCH{=}CH_2 + HBr \rightarrow RCH(Br){-}CH_3$



Deep Learning Chemical Reactions

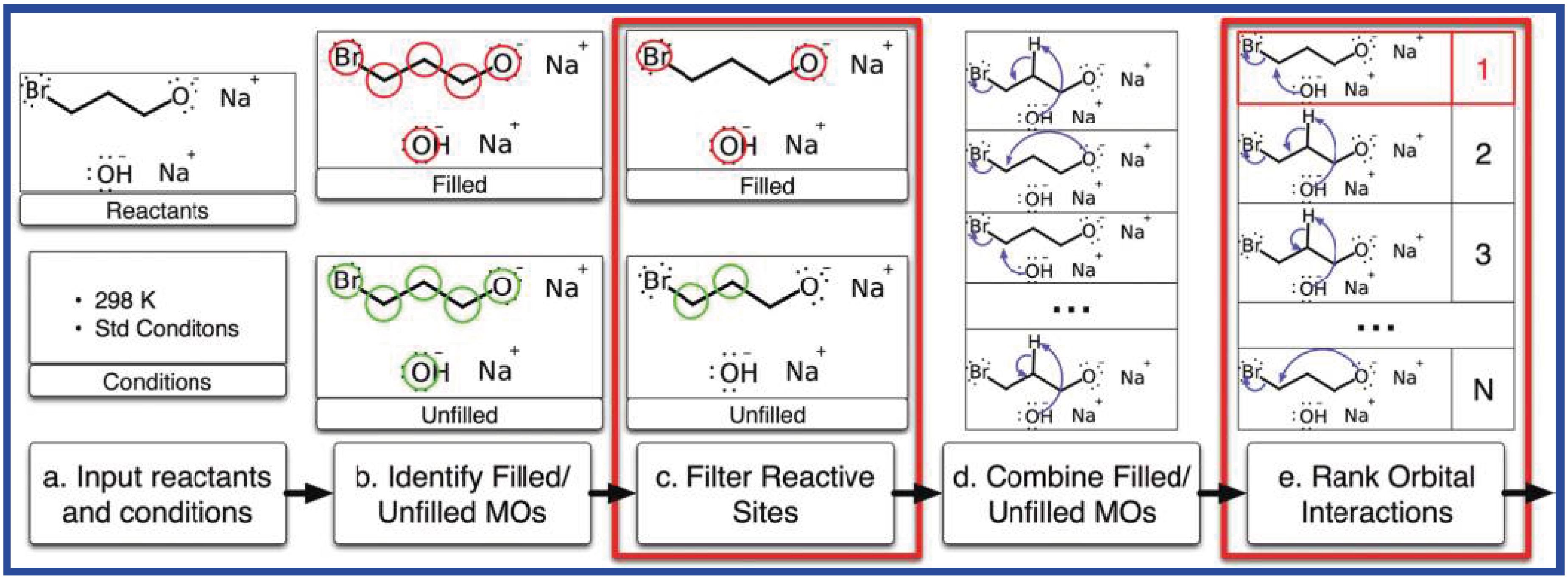
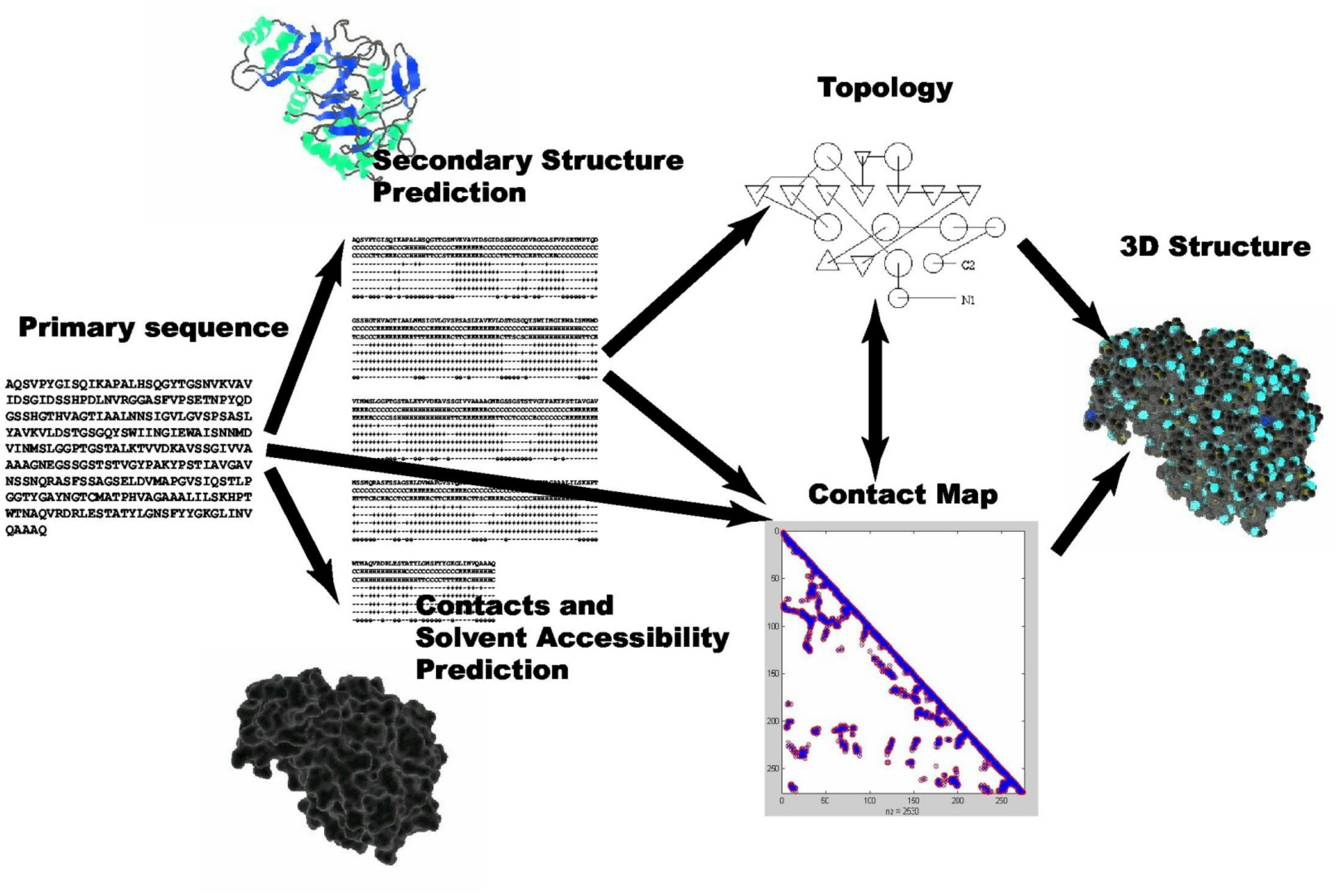


Figure 2. Overall reaction prediction framework: (a) A user inputs the reactants and conditions. (b) We identify potential electron donors and acceptors using coarse approximations of electron-filled and -unfilled MOs. (c) Highly sensitive reactive site classifiers are trained and used to filter out the vast majority of unreactive sites, pruning the space of potential reactions. (d) Reactions are enumerated by pairing filled and unfilled MOs. (e) A ranking model is trained and used to order the reactions, where the best ranking one or few represent the major products. The top-ranked product can be recursively chained to a new instance of the framework for multistep reaction prediction.

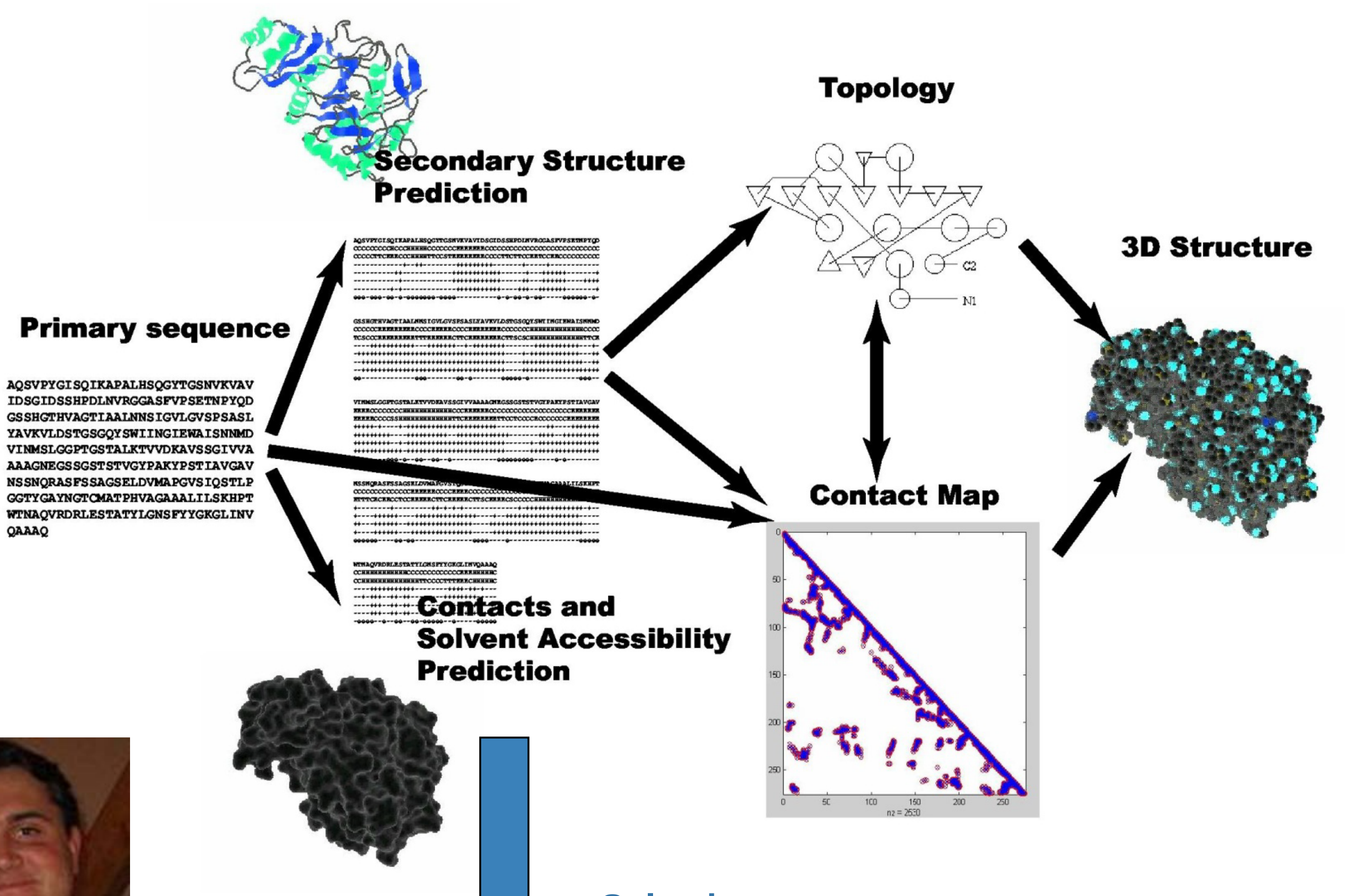
M. Kayala, C. Azencott, J. Chen, and P. Baldi. Learning to Predict Chemical Reactions. *Journal of Chemical Information and Modeling*, 51, 9, 2209–2222, (2011).

M. Kayala and P. Baldi. ReactionPredictor: Prediction of Complex Chemical Reactions at the Mechanistic Level Using Machine Learning. *Journal of Chemical Information and Modeling*, 52, 10, 2526–2540, (2012).

Deep Learning in Biology



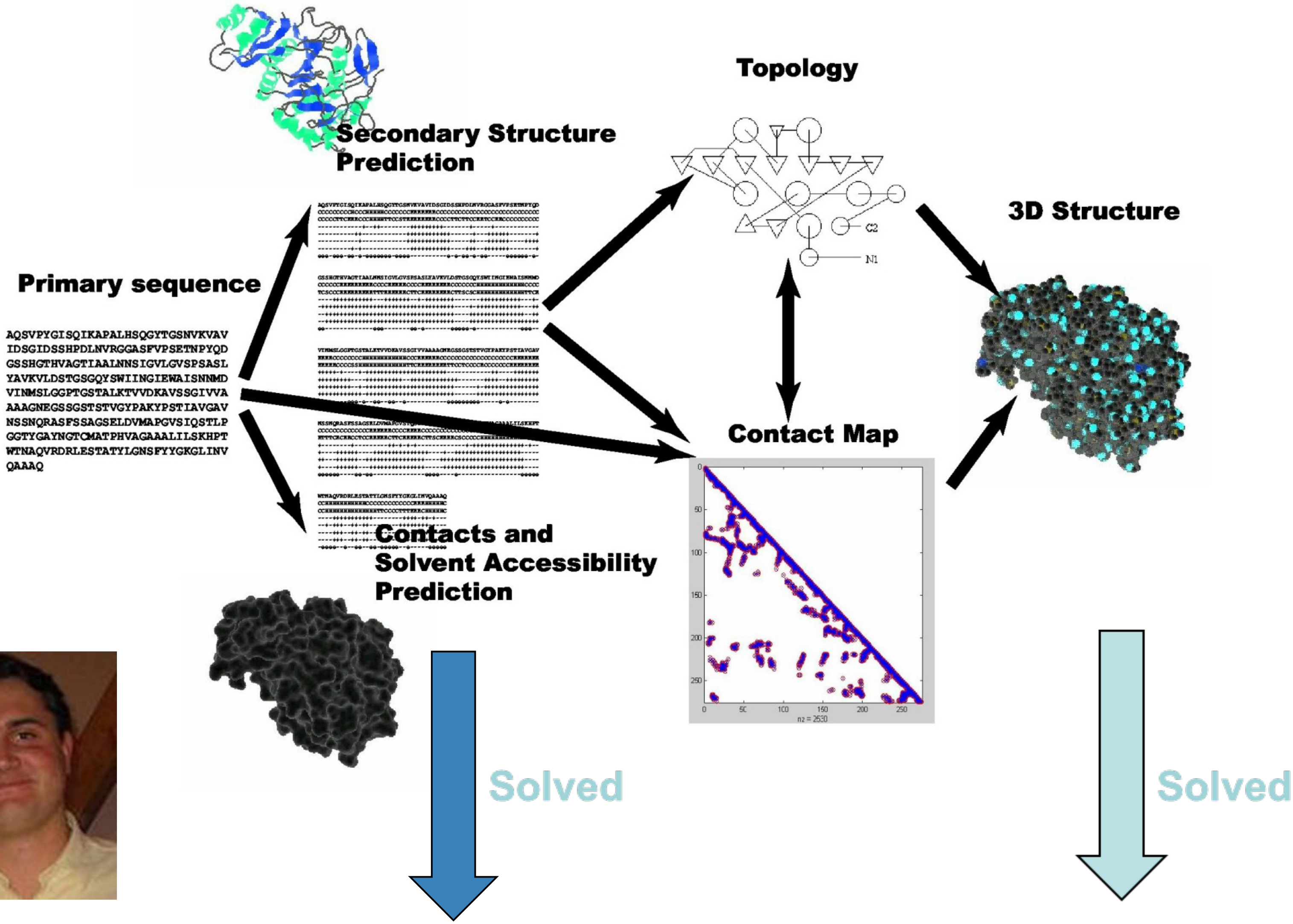
Deep Learning in Biology: Mining Omic Data



Solved

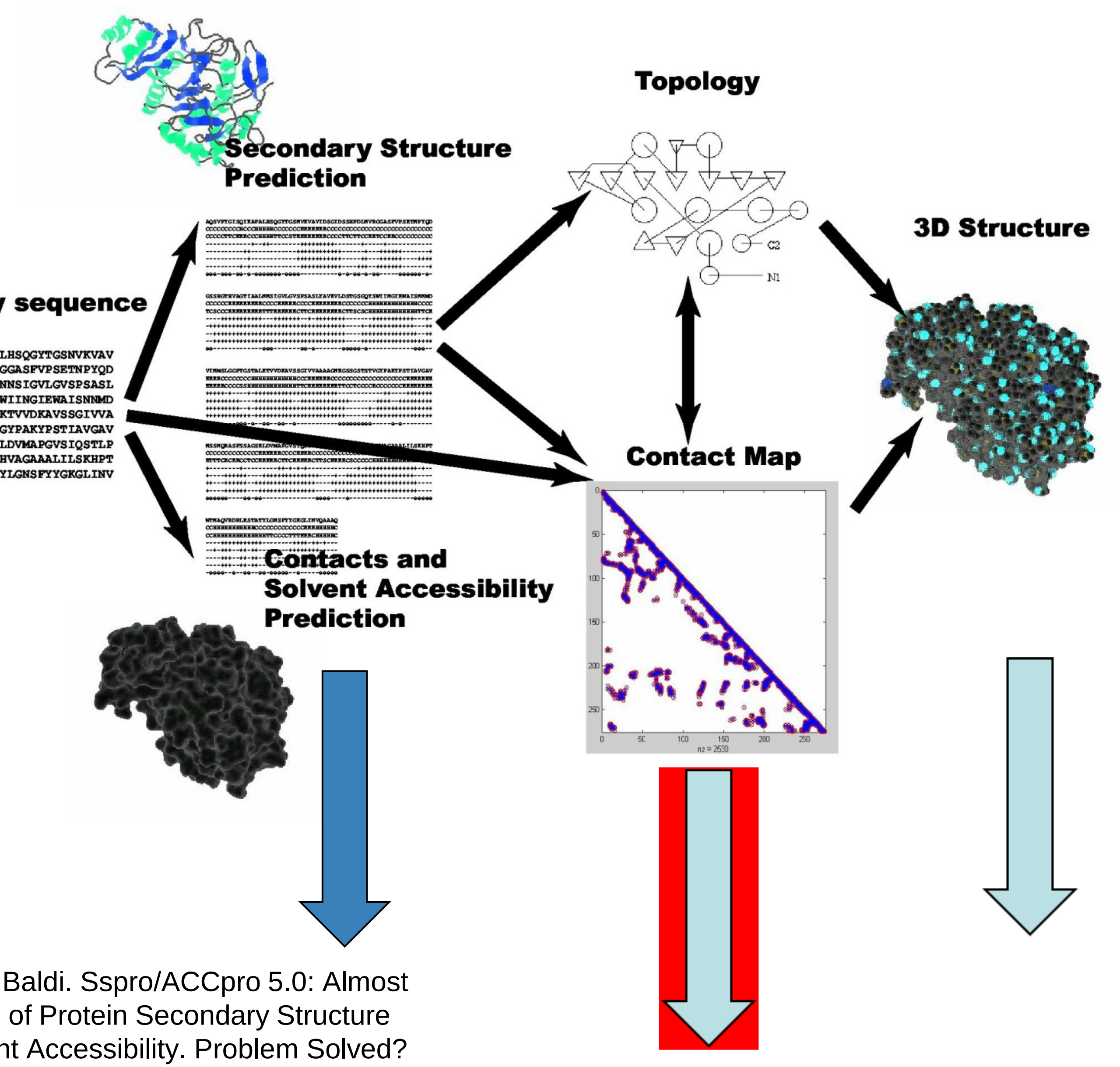
C. Magnan and P. Baldi. Sspro/ACCpro 5.0: Almost Perfect Prediction of Protein Secondary Structure and Relative Solvent Accessibility. Problem Solved? *Bioinformatics*, (advance access June 18), (2014).

Deep Learning in Biology: Mining Omic Data



C. Magnan and P. Baldi. Sspro/ACCpro 5.0: Almost Perfect Prediction of Protein Secondary Structure and Relative Solvent Accessibility. Problem Solved? *Bioinformatics*, (advance access June 18), (2014).

Deep Learning in Biology: Mining Omic Data



C. Magnan and P. Baldi. Sspro/ACCpro 5.0: Almost Perfect Prediction of Protein Secondary Structure and Relative Solvent Accessibility. Problem Solved? *Bioinformatics*, (advance access June 18), (2014).

10th Community Wide Experiment on the
Critical Assessment of Techniques for Protein Structure Prediction

RR Analysis

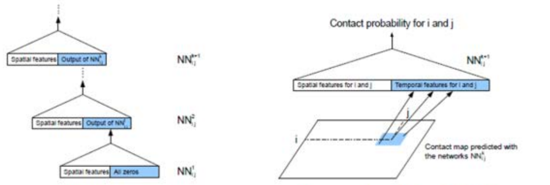
The table summarizes the evaluation of predictions in 'RR' category.
The analysis was performed at per domains basis; only predictions for domains classified as "FM", "TBM/FM", "TBM hard" were considered.
The groups were ranked according to sum of average Z-scores for two measures Acc and Xd.
The per target Z-scores were recalculated from the "cleaned" distributions, where the outlier predictions (below mean - 2 std dev) were eliminated.

- **Domain classification:**
  - FM
  - TBM/FM
  - TBM hard (max gdt_ts < 50)
- **Contact Range:** long
- **List Size:** L/5

#	GR#	GR Name	Count domains	AVG Acc	AVG Zscore Acc	AVG Xd	AVG Zscore Xd	Zsco... Acc + Zscore Xd
1.	222 s	MULTICOM-CONSTRUCT	14	19.41	0.58	12.08	0.77	1.35
2.	305 s	IGBteam	15	19.22	0.72	10.19	0.58	1.30
3.	424 s	MULTICOM-NOVEL	14	20.39	0.50	10.32	0.72	1.22
4.	125 s	MULTICOM-REFINE	14	21.35	0.51	10.29	0.70	1.21
5.	413 s	ZHOU-SPARKS-X	12	12.26	0.62	8.26	0.59	1.21
6.	113 s	SAM-T08-server	11	16.13	0.72	9.44	0.47	1.19
7.	358 s	RaptorX-Roll	8	12.07	0.58	8.23	0.55	1.13
8.	314 s	ProC_S4	14	17.91	0.59	9.76	0.47	1.05
9.	087 s	Distill_roll	15	13.97	0.60	8.57	0.36	0.96
10.	489	MULTICOM	14	12.96	0.43	8.19	0.40	0.83
11.	184 s	ICOS	14	17.03	0.40	9.72	0.39	0.78
12.	396 s	ProC_S5	14	16.51	0.36	9.10	0.36	0.72
13.	381 s	SAM-T06-...	10	10.98	0.37	7.94	0.31	0.68

Deep Learning


P. Di Lena, K. Nagata, and P. Baldi. Deep Architectures for Protein Contact Map Prediction. *Bioinformatics*, 28, 2449-2457, (2012)



(a) DST-NN architecture (b) Temporal input features for NN^k_{ij}

Figure 1: DST-NN architecture. (a) Overview. Each NN^k_{ij} represents a feed-forward neural network trainable by back-propagation. (b) For a pair of residues (i, j), the temporal inputs into NN^{k+1}_{ij} consist of the contact probabilities produced by the network at the previous level over a neighborhood of (i, j).

Machine Learning (DL) in the Natural Sciences

- Physics
  - -**HEP**
  - -QM
  - -Astronomy
- Chemistry
  - -Prediction of Molecular or Material Properties
  - -Prediction of Chemical Reactions
- Earth Sciences
- Biology
  - -Prediction of Protein Structures
  - -Prediction of gene Expression
  - -Biomedical imaging

Deep Learning in HEP

- Higgs Boson Detection
- Supersymmetry
- Higgs To Tau Tau Decay

nature COMMUNICATIONS

ARTICLE

Received 19 Feb 2014 | Accepted 4 Jun 2014 | Published 2 Jul 2014

DOI: 10.1038/ncomms5308

Searching for exotic particles in high-energy physics with deep learning

P. Baldi[1], P. Sadowski[1] & D. Whiteson[2]

Collisions at high-energy particle colliders are a traditionally fruitful source of exotic particle discoveries. Finding these rare particles requires solving difficult signal-versus-background classification problems, hence machine-learning approaches are often used. Standard approaches have relied on 'shallow' machine-learning models that have a limited capacity to learn complex nonlinear functions of the inputs, and rely on a painstaking search through manually constructed nonlinear features. Progress on this problem has slowed, as a variety of techniques have shown equivalent performance. Recent advances in the field of deep learning make it possible to learn more complex functions and better discriminate between signal and background classes. Here, using benchmark data sets, we show that deep-learning methods need no manually constructed inputs and yet improve the classification metric by as much as 8% over the best current approaches. This demonstrates that deep-learning approaches can improve the power of collider searches for exotic particles.

Deep Learning in HEP

- Higgs Boson Detection (NC 2014)
- Supersymmetry (NC 2014)

- Higgs to Tau Tau Decay (NIPS 2014)

Deep Learning in HEP

- Higgs Boson Detection
- Supersymmetry
- Higgs Decay

- Common Features and Results:

  -dozens of features: raw + human-derived

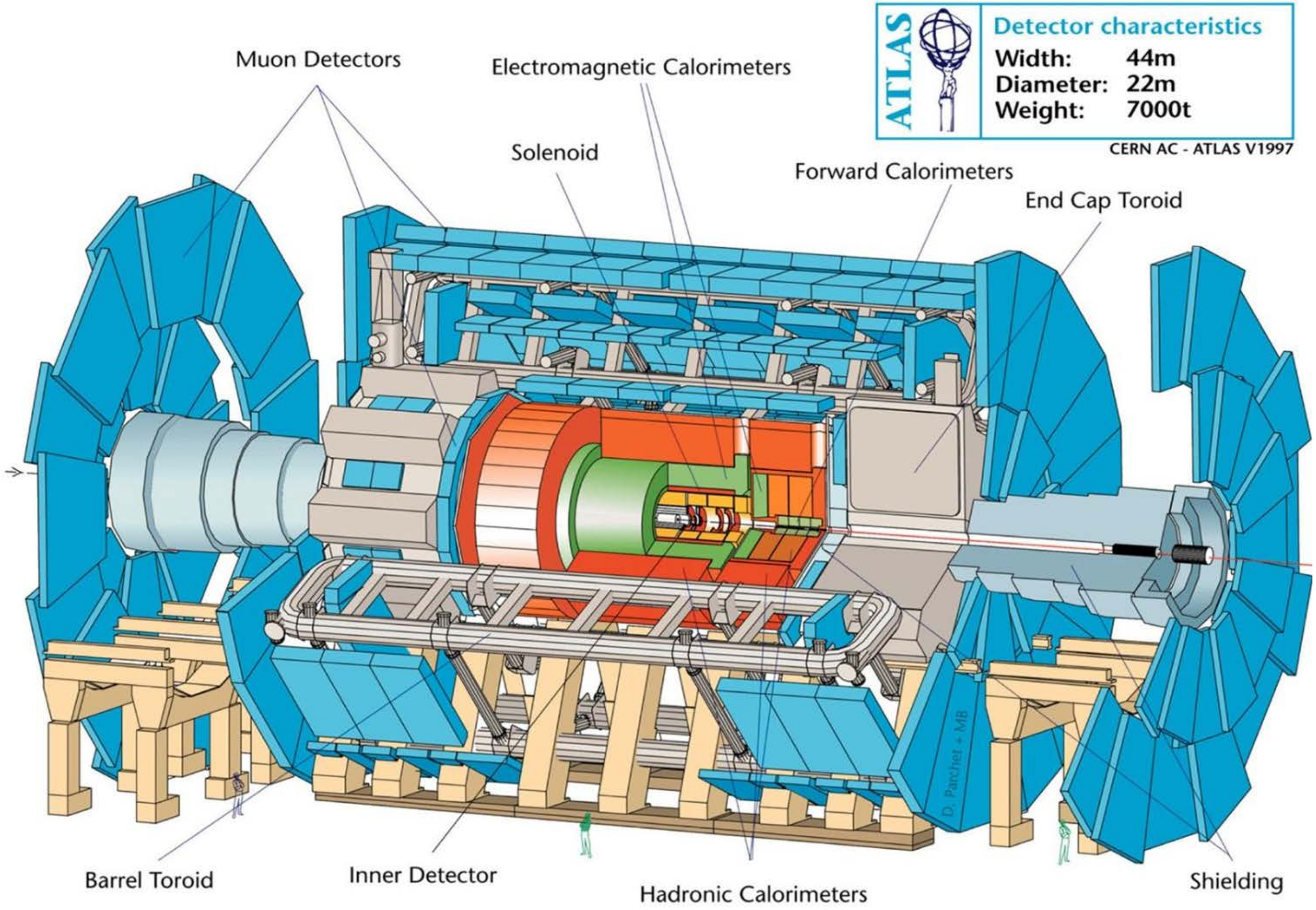
  -millions of examples

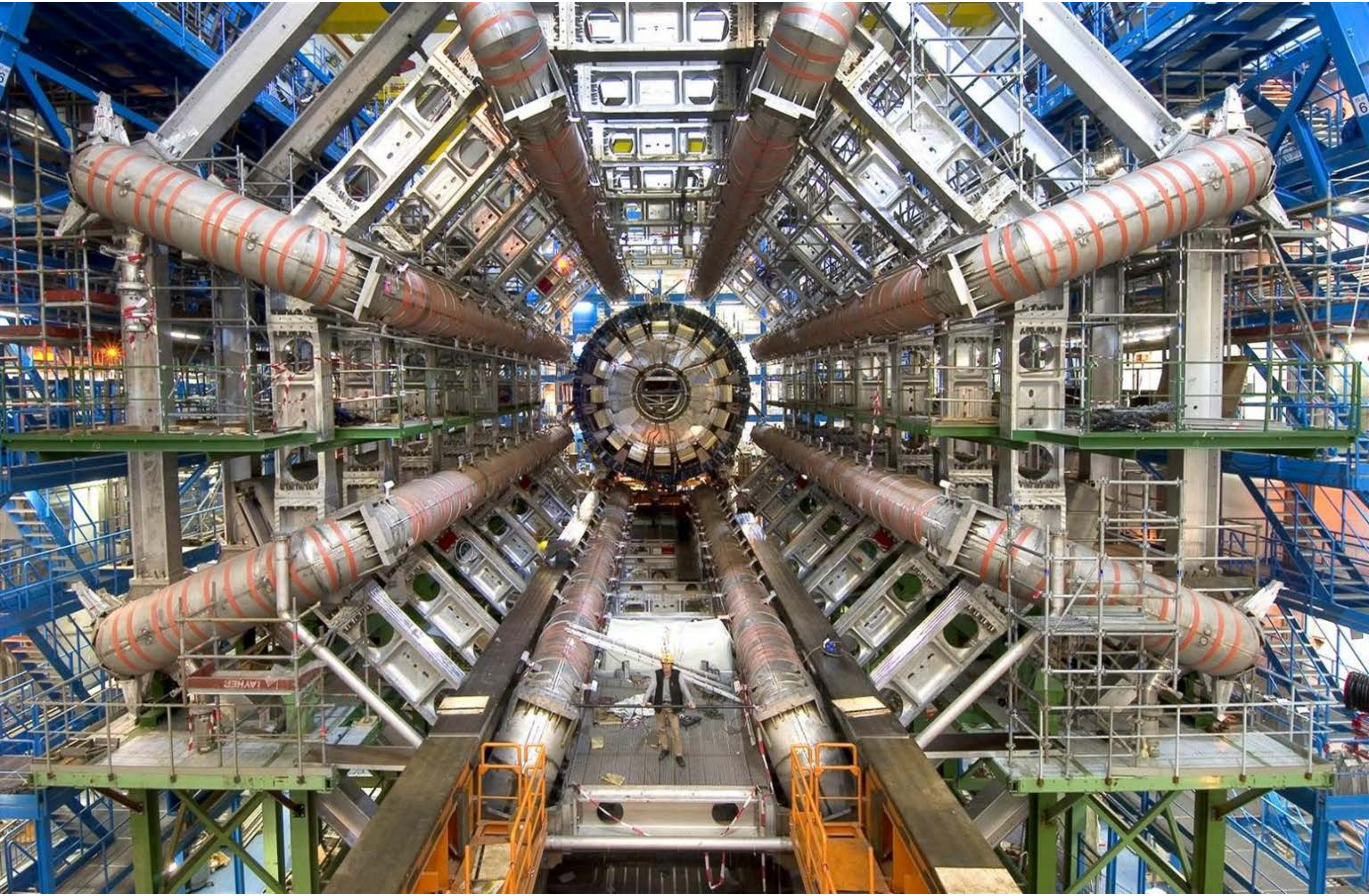
  -classification problems

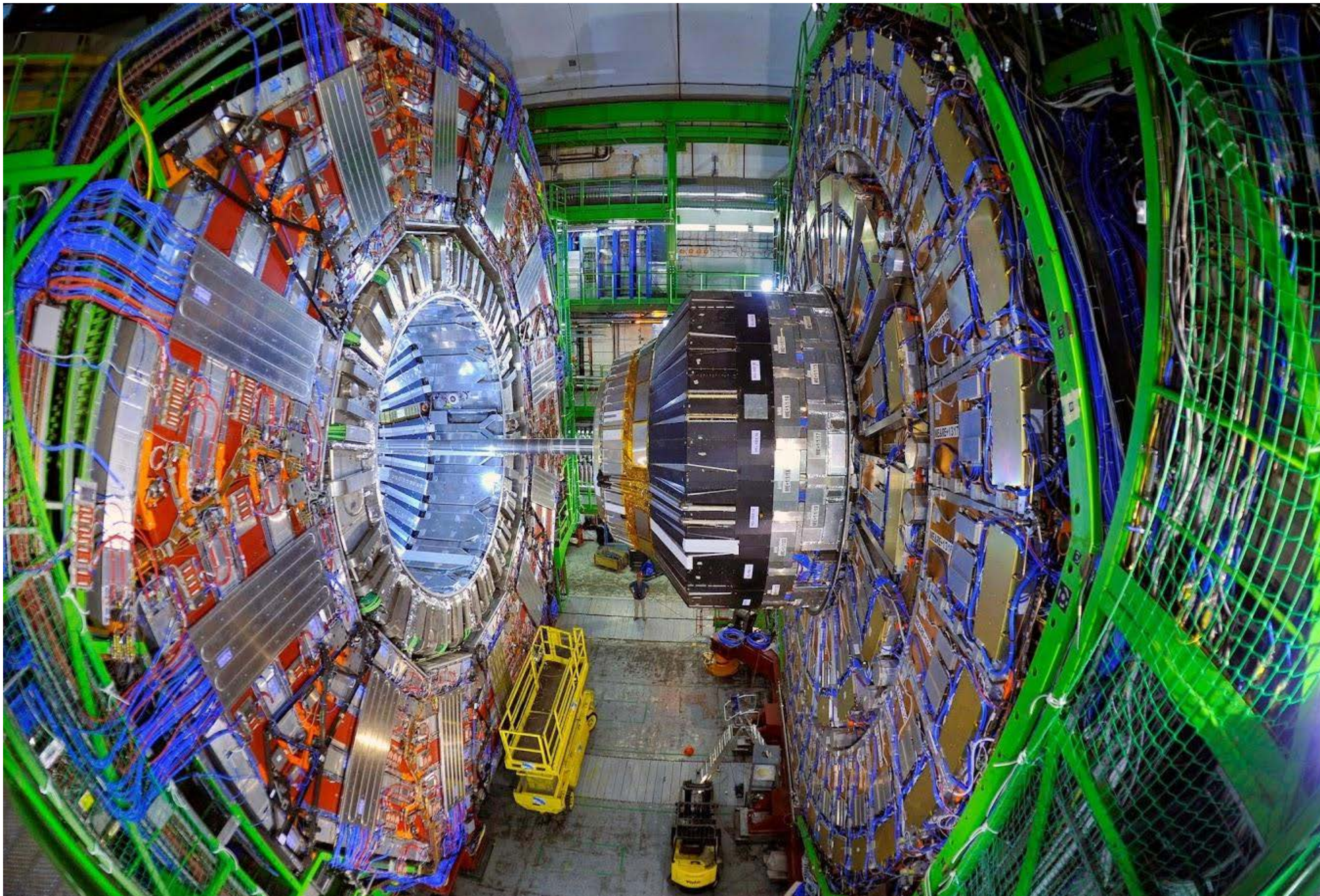
  -deep learning outperforms current methods (e.g AUC)

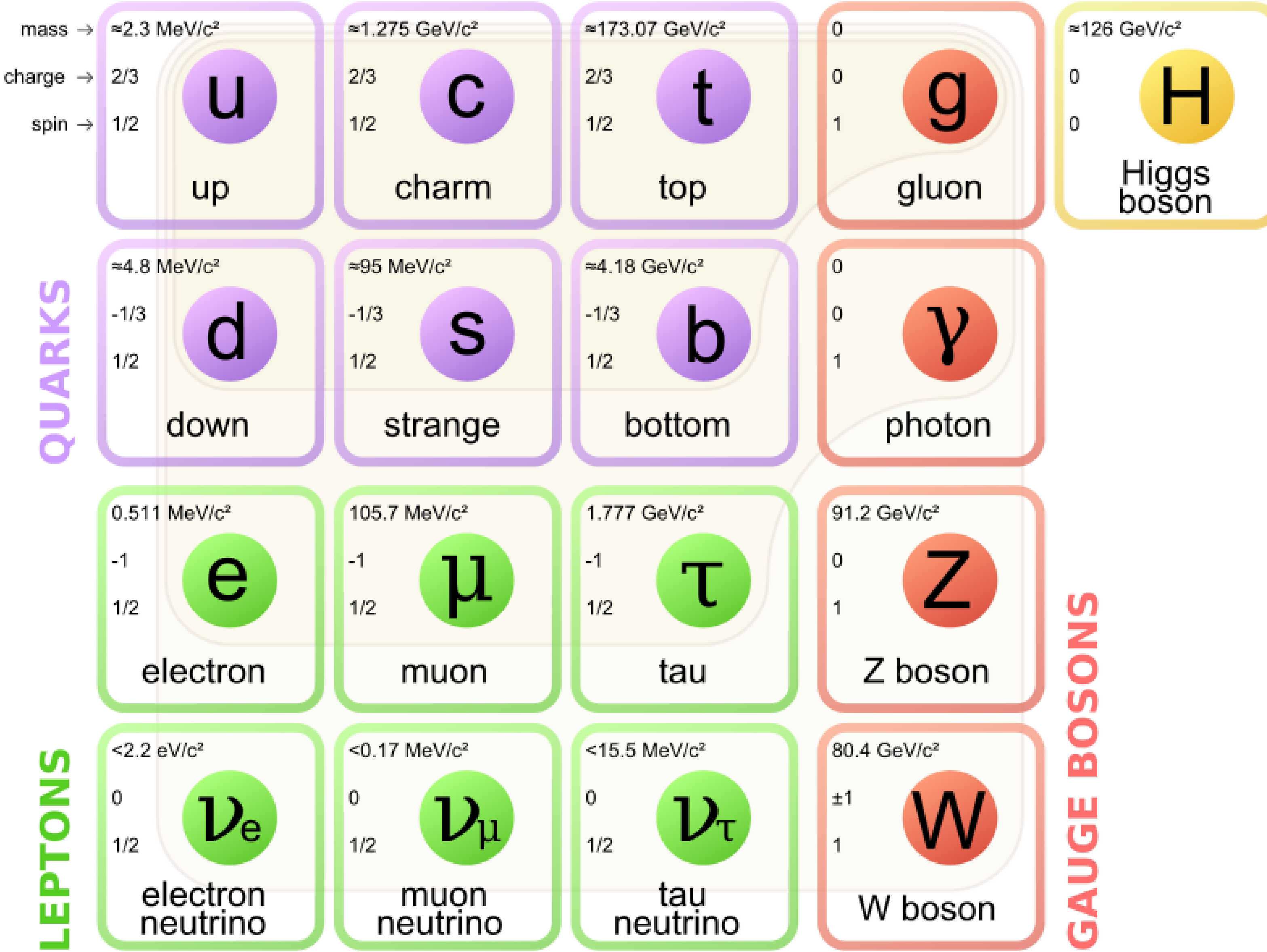
  - deep learning can work without human-derived features

Machine Learning in HEP



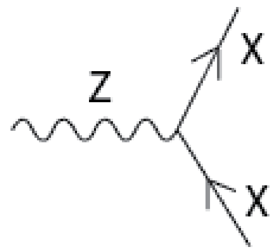




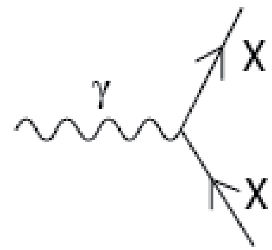


mass →
charge →
spin →
QUARKS
LEPTONS
GAUGE BOSONS
≈2.3 MeV/c²
2/3
1/2
u
up
≈1.275 GeV/c²
2/3
1/2
c
charm
≈173.07 GeV/c²
2/3
1/2
t
top
0
0
1
g
gluon
≈126 GeV/c²
0
0
H
Higgs boson
≈4.8 MeV/c²
-1/3
1/2
d
down
≈95 MeV/c²
-1/3
1/2
s
strange
≈4.18 GeV/c²
-1/3
1/2
b
bottom
0
0
1
γ
photon
0.511 MeV/c²
-1
1/2
e
electron
105.7 MeV/c²
-1
1/2
μ
muon
1.777 GeV/c²
-1
1/2
τ
tau
91.2 GeV/c²
0
1
Z
Z boson
<2.2 eV/c²
0
1/2
νe
electron neutrino
<0.17 MeV/c²
0
1/2
νμ
muon neutrino
<15.5 MeV/c²
0
1/2
ντ
tau neutrino
80.4 GeV/c²
±1
1
W
W boson

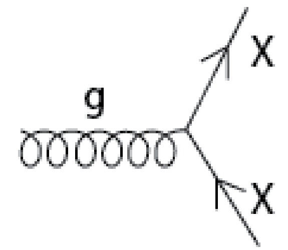
Standard Model Interactions
(Forces Mediated by Gauge Bosons)



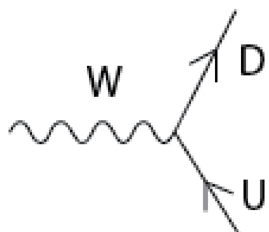
X is any fermion in the Standard Model.



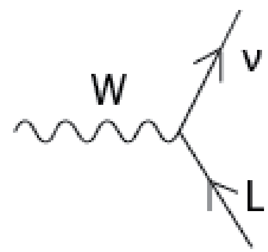
X is electrically charged.



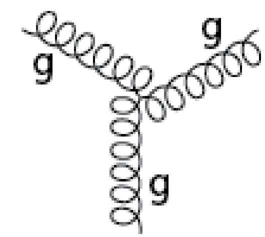
X is any quark.

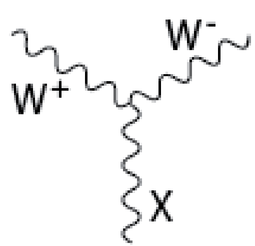


U is a up-type quark;
D is a down-type quark.

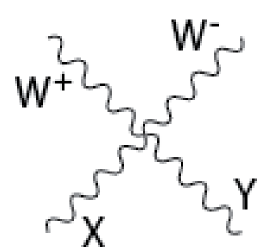


L is a lepton and ν is the corresponding neutrino.

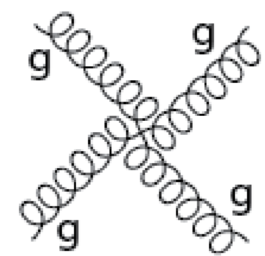




X is a photon or Z-boson.

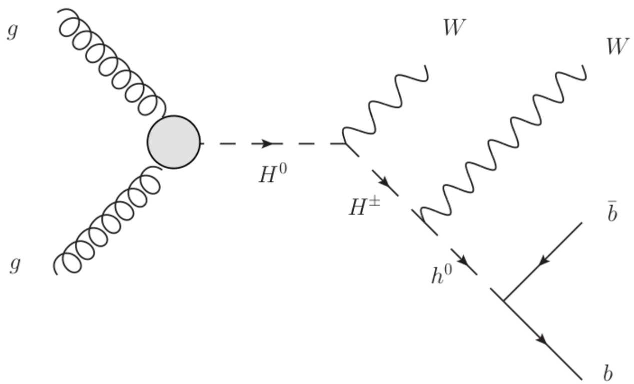


X and Y are any two electroweak bosons such that charge is conserved.

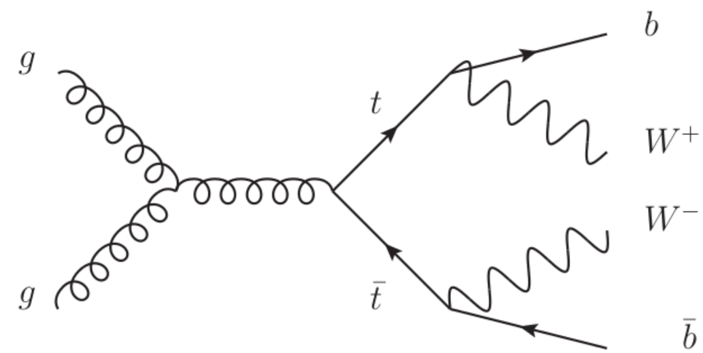


Higgs Boson Detection

Higgs boson decay signal



Background process



Simulation tools:

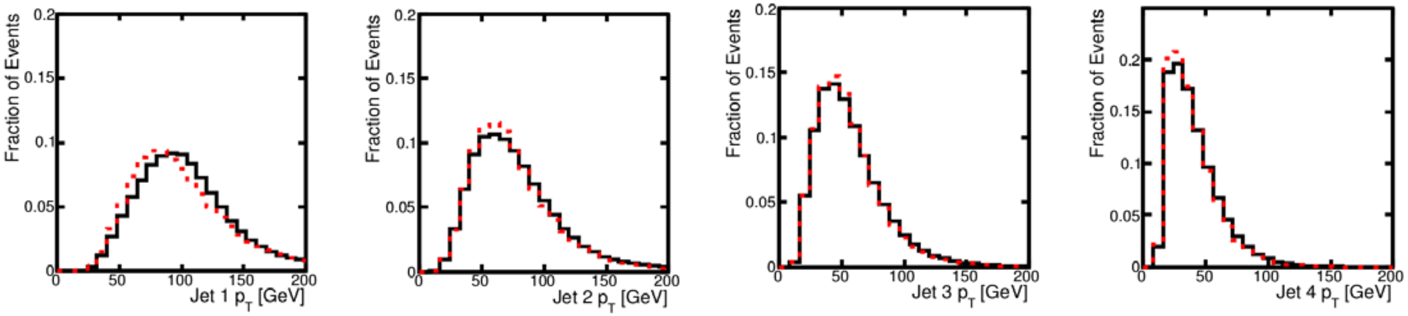
- MadGraph (collisions)
- PYTHIA (showering and hadronization)
- DELPHES (detector response)

11 M examples

Higgs Boson Detection

Supervised learning problem:

- Two classes
- 11 million training examples (roughly balanced)
- 28 features
  - 21 low-level features (momenta of particles)
  - 7 high-level features derived by physicists



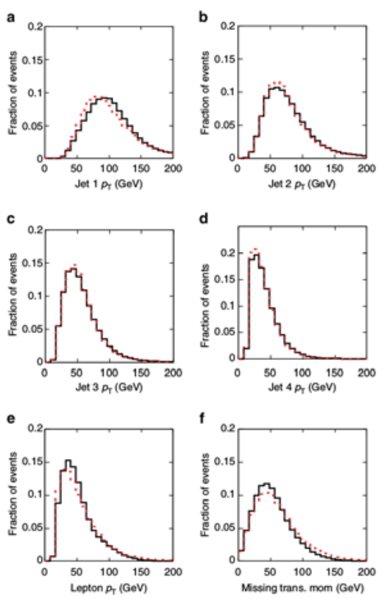
Signal (black) vs background (red)

Data available at archive.ics.uci.edu/ml/datasets/HIGGS

Higgs Boson Detection

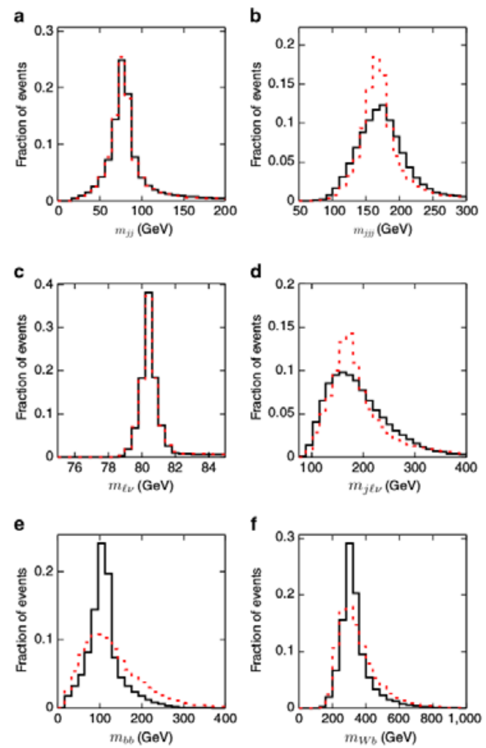
21 low-level features:

- 3D momentum for observed particles
- Missing transverse momentum
- Jets and *b*-tagging information



7 high-level features:

- Reconstruction of invariant masses for each particle subset.



Higgs Boson Detection

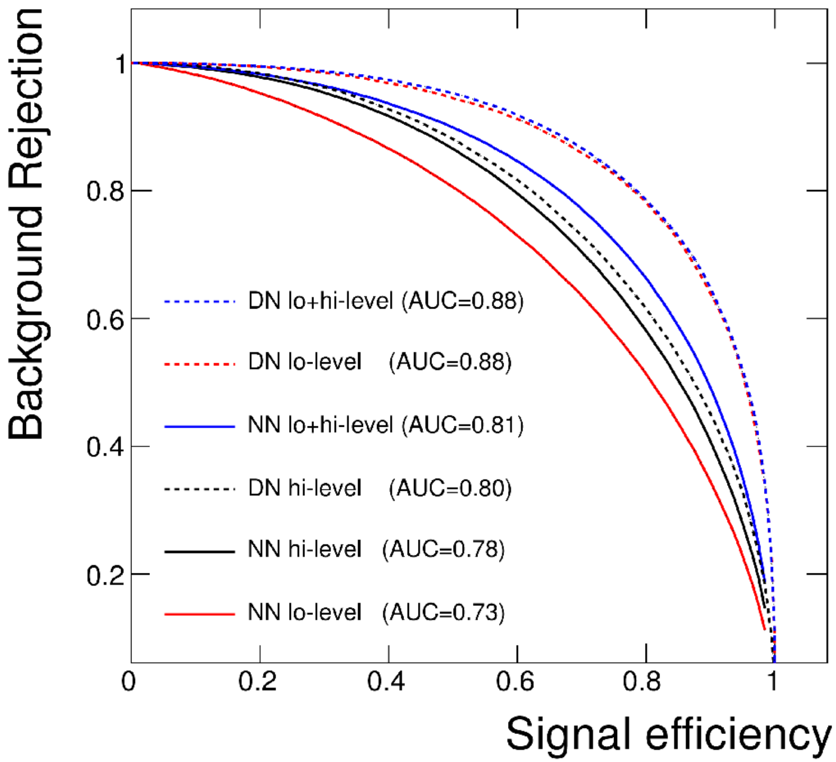
Tuning deep neural network architectures.

Hyper parameters	Choices
Depth	2,3,4,5,6 layers
Hidden units per layer	100,200,300,500
Learning rate	0.01, 0.05
Weight decay	0, 0.00001
Pre-training	none, autoencoder
	multi-task autoencoder
Input features	low-level, high-level
	complete set

Best:

- 5 hidden layers
- 300 neurons per layer
- Tanh hidden units, sigmoid output
- No pre-training
- Stochastic gradient descent
- Mini batches of 100
- Exponentially-decreasing learning rate
- Momentum increasing from .5 to .99 over 200 epochs
- Weight decay = 0.00001

Higgs Boson Detection



	AUC		
Technique	Low-level	High-level	Complete
BDT	0.73	0.78	0.81
NN	0.733 (0.007)	0.777 (0.001)	0.816 (0.004)
DN	0.880 (0.001)	0.800 (< 0.001)	0.885 (0.002)

Deep network improves AUC by 8%

BDT= Bayesian Decision Trees in TMVA package

Nature Communications, July 2014

Higgs Boson Detection

	AUC		
Technique	Low-level	High-level	Complete
NN 300-hidden	0.733	0.777	0.816
NN 1000-hidden	0.788	0.783	0.841
NN 2000-hidden	0.787	0.788	0.842
NN 10000-hidden	0.790	0.789	0.841
DN 3 layers	0.836	0.791	0.850
DN 4 layers	0.868	0.797	0.872
DN 5 layers	0.880	0.800	0.885
DN 6 layers	0.888	0.799	0.893

DNs have 300 tanh units in each hidden layer.

Deeper networks perform better and performance continues to improve after publication ….

Higgs Boson Detection

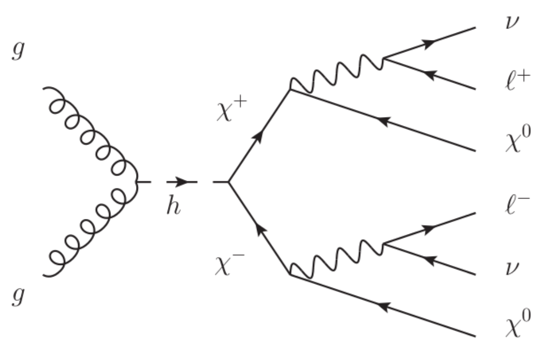
Experiment: regression on 7 high-level features

Technique	Feature Regression MSE
Linear Regression	0.1468
NN	0.0885
DN 3 layers	0.0821
DN 4 layers	0.0818
DN 5 layers	0.0815
DN 6 layers	0.0812

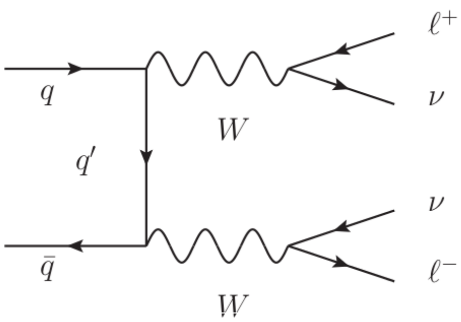
Deeper networks better at learning high-level features.

Supersymmetry (SUSY)

Signal



Background



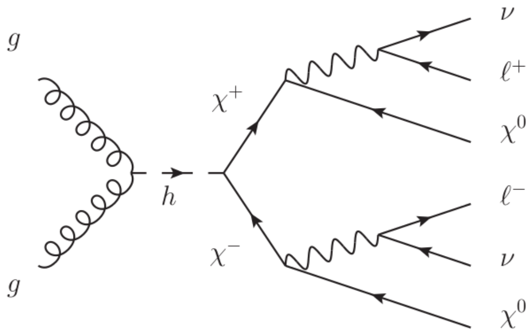
Detect the production of new supersymmetric particles

6 M examples

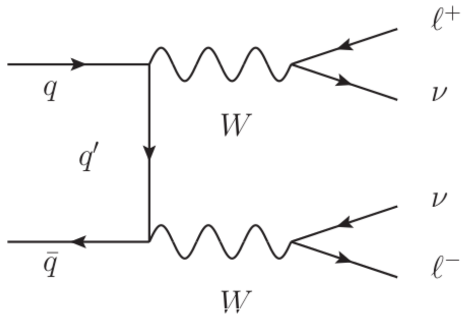
Data available at archive.ics.uci.edu/ml/datasets/SUSY

Supersymmetry (SUSY)

Signal



Background



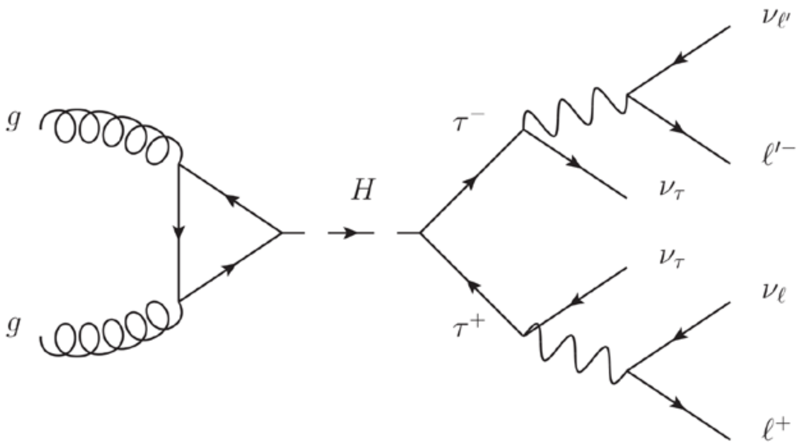
AUC

Technique	Low-level	High-level	Complete
BDT	0.850 (0.003)	0.835 (0.003)	0.863 (0.003)
NN	0.867 (0.002)	0.863 (0.001)	0.875 (< 0.001)
$\mathrm{NN}_{dropout}$	0.856 (< 0.001)	0.859 (< 0.001)	0.873 (< 0.001)
DN	0.872 (0.001)	0.865 (0.001)	0.876 (< 0.001)
$\mathrm{DN}_{dropout}$	0.876 (< 0.001)	0.869 (< 0.001)	0.879 (< 0.001)

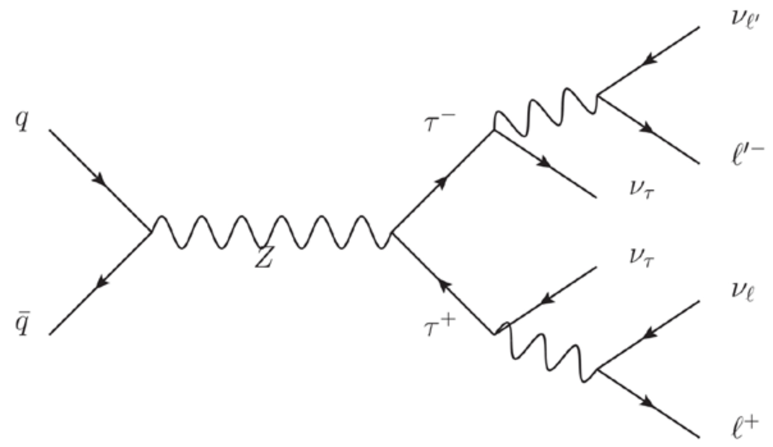
Deep networks again lead to significant gains.

Data available at archive.ics.uci.edu/ml/datasets/SUSY

Higgs to Tau Tau Decay



Signal

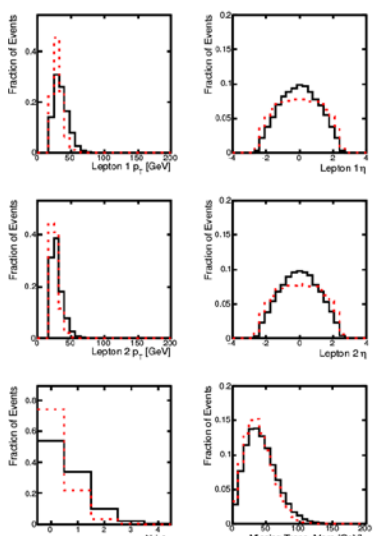


Background

Higgs to Tau Tau Decay

10 low-level features:

- The 3D momenta, p, of the charged leptons
- The imbalance of momentum ($\not{p}_T$) in the final state transverse to the beam direction, due to unobserved or mismeasured particles
- The number and momenta of particle 'jets' due to radiation of gluons or quarks



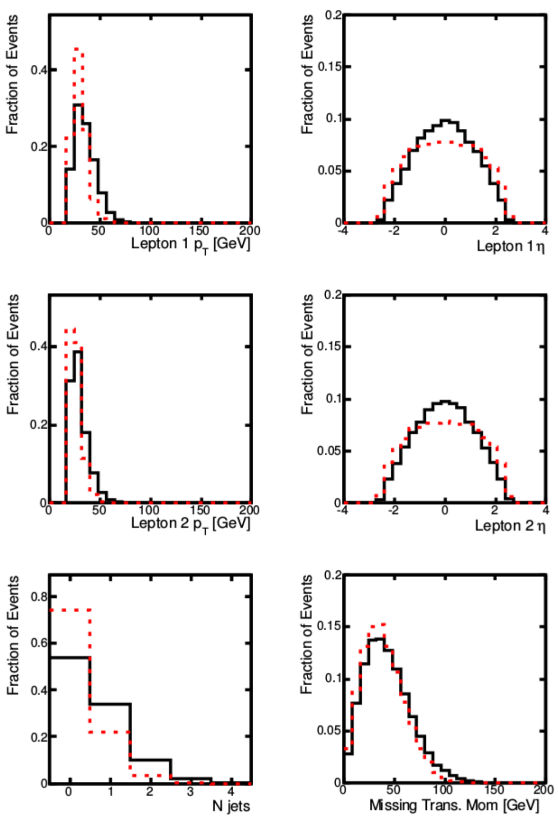
15 high-level:

- Axial missing momentum, $\not{p}_T \cdot p_{\ell^+\ell^-}$
- Scalar sum of the observed momenta, $|p_{\ell^+}| + |p_{\ell^-}| + |\not{p}_T| + \sum_i |p_{\mathrm{jet}_i}|$
- Angular distance between leptons, $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$
- Missing mass, m_{MMC}
- Sphericity and transverse sphericity
- Relative missing momentum, $\not{p}_T$ if $\Delta\phi(p, \not{p}_T) \geq \pi/2$, and $\not{p}_T \times \sin(\Delta\phi(p, \not{p}_T)$ if $\Delta\phi(p, \not{p}_T) < \pi/2$, where p is the momentum of any charged lepton or jet;
- Difference in lepton azimuthal angles, $\Delta\phi(\ell^+, \ell^-)$
- Difference in lepton polar angles, $\Delta\eta(\ell^+, \ell^-)$
- Invariant mass of the two leptons, $m_{\ell^+\ell^-}$
- Invariant mass of all visible objects (leptons and jets).

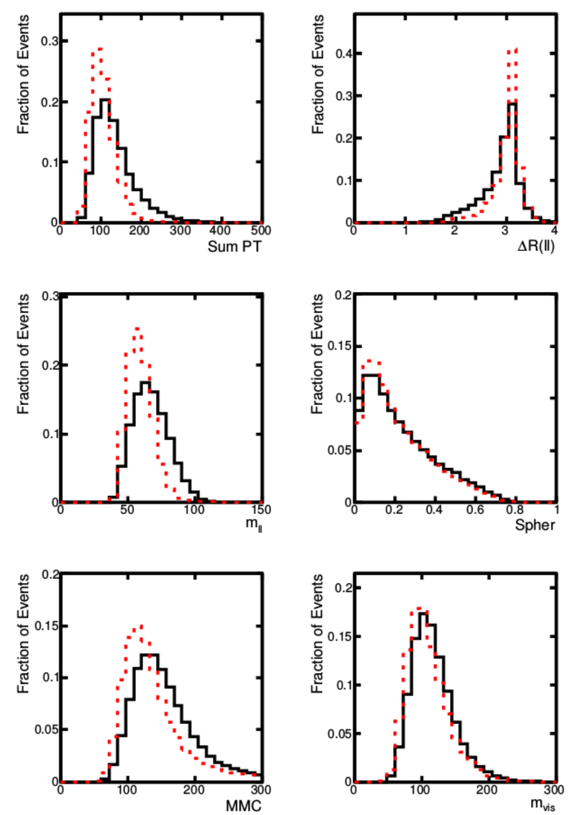
80M examples

Higgs to Tau Tau Decay

10 low-level features:



15 high-level:



Higgs to Tau Tau Decay

Hyper-parameters optimized with Spearmint:

- 100 deep networks trained
- 40M training examples, 100 epochs
- Hyperparameters include depth and width
- Best network:
  - Deepest network available (8 layers)
  - Rectified linear hidden units
  - ~300 units per layer

Spearmint chooses the deepest network.

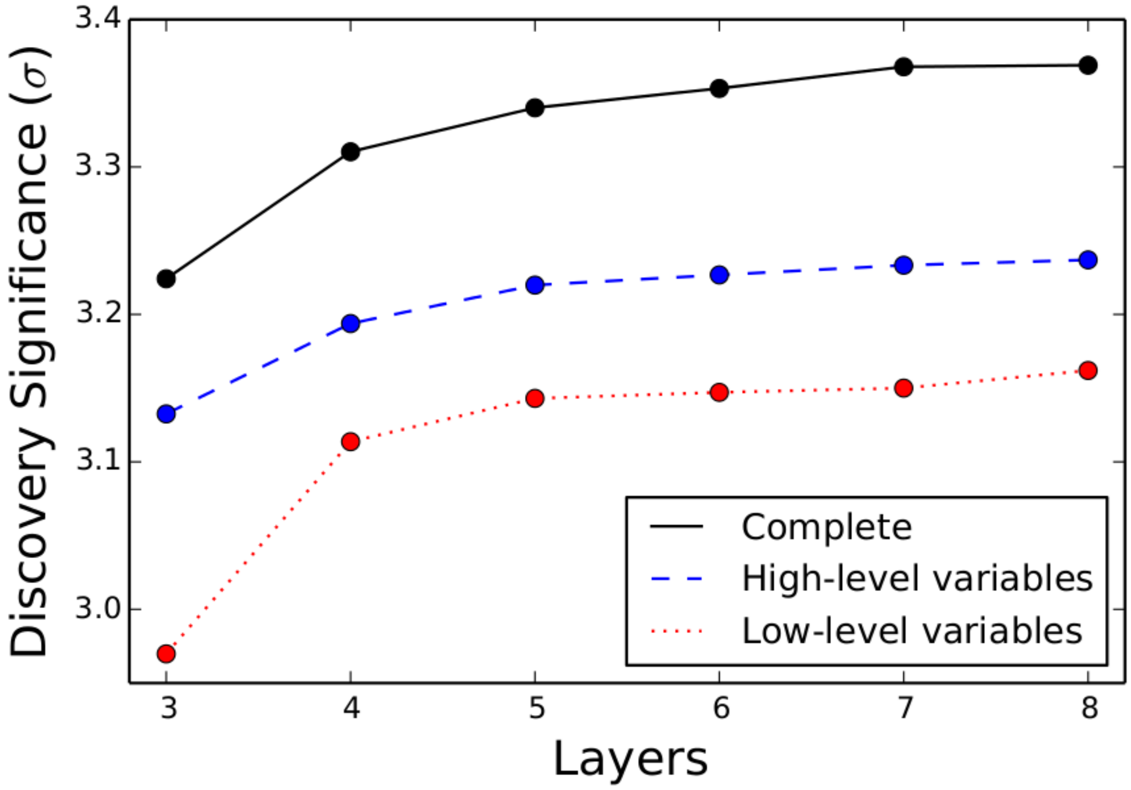
Higgs to Tau Tau Decay

Optimized shallow net vs optimized deep nets:

	AUC		
Technique	Low-level	High-level	Complete
NN	0.789 (0.0010)	0.792 (0.0002)	0.797 (0.0004)
NN ensemble	0.791	0.793	0.798
DNN	0.798 (0.0001)	0.798 (0.0001)	0.802 (0.0001)
DNN ensemble	0.798	0.798	0.803

(1) DNN give significant performance boost
(2) Ensembles give small boost
(3) Slight gap with respect to high-level features remains (the mass of the lepton is included in the high-level features…)

Higgs to Tau Tau Decay



Improvement translates to 20% reduction in data needed for discovery.

Many Open Directions

- Apply ML to earlier stages of processing (“trigger”)
- Model detector signals
- Improve performance
- Apply ML to other exotic particles and theories
- Transfer Learning
- ……Other Natural Sciences
- …….ML

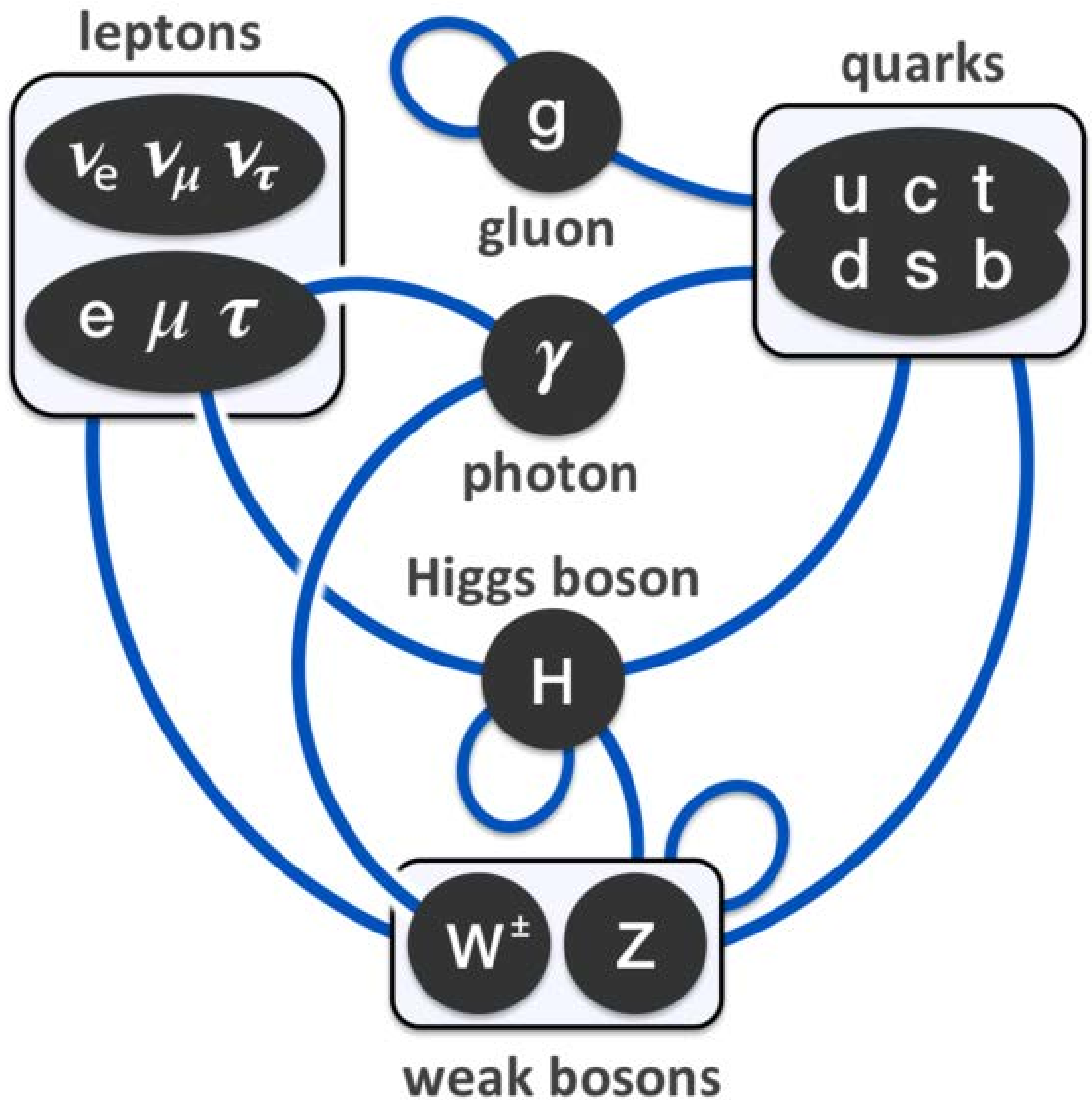
THANK YOU

Higgs to Tau Tau Decay

Optimized shallow net vs optimized deep nets:

	Discovery significance		
Technique	Low-level	High-level	Complete
NN	2.57σ (0.006)	2.92σ (0.006)	3.02σ (0.008)
NN ensemble	2.61σ	2.96σ	3.06σ
DNN	3.16σ (0.003)	3.24σ (0.003)	3.37σ (0.003)
DNN ensemble	3.18σ	3.26σ	3.39σ

(1) DNN give significant performance boost,
(2) Ensembles give small boost
(3) Slight gap with respect to high-level features remains (Mass of lepton is contained in the high-level features)


leptons
ν_e ν_μ ν_τ
e μ τ
g
gluon
γ
photon
quarks
u c t
d s b
Higgs boson
H
$W^\pm$
Z
weak bosons

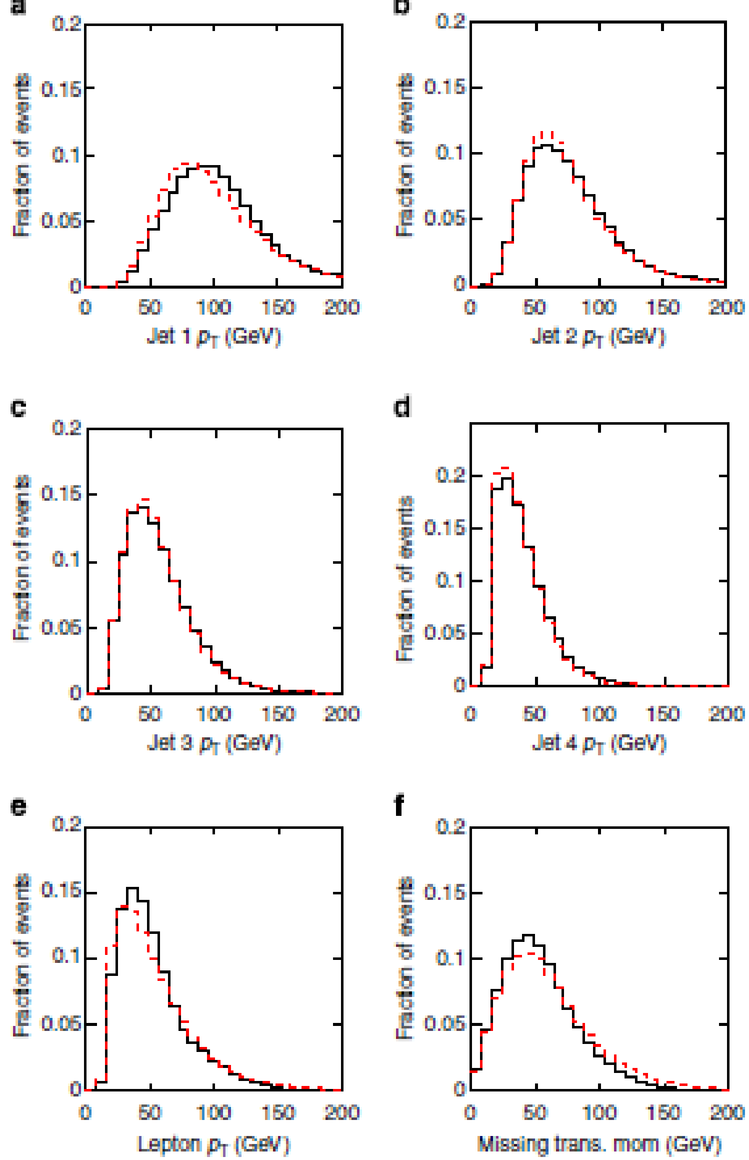
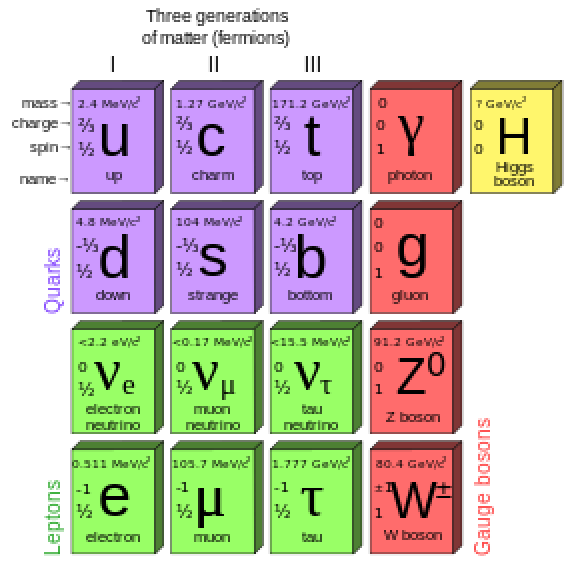
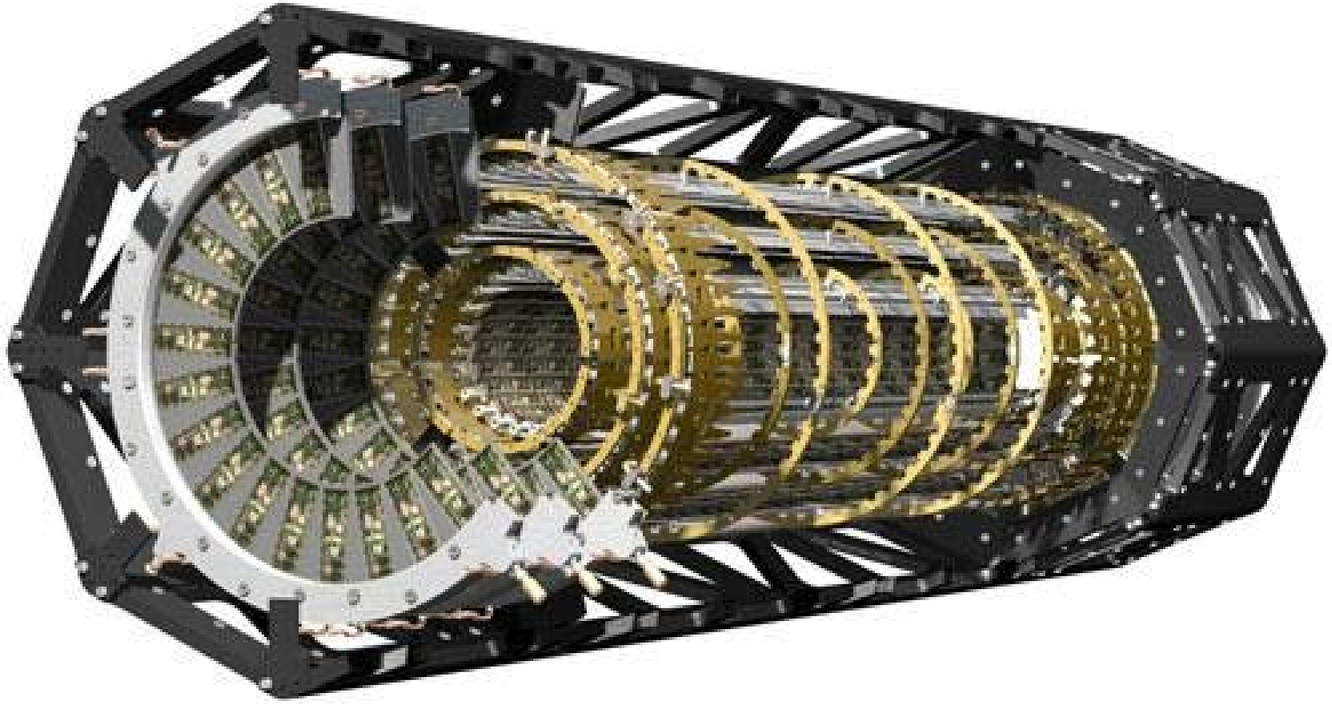


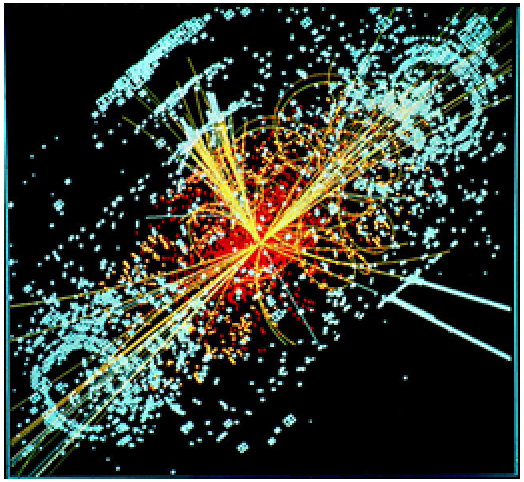
Figure 2 | Low-level input features for Higgs benchmark. Distributions in $\ell\nu jjb\bar{b}$ events for simulated signal (black) and background (red) benchmark events. Shown are the distributions of transverse momenta (p_T) of each observed particle (**a–e**) as well as the imbalance of momentum in the final state (**f**). Momentum angular information for each observed particle is also available to the network, but is not shown, as the one-dimensional projections have little information.



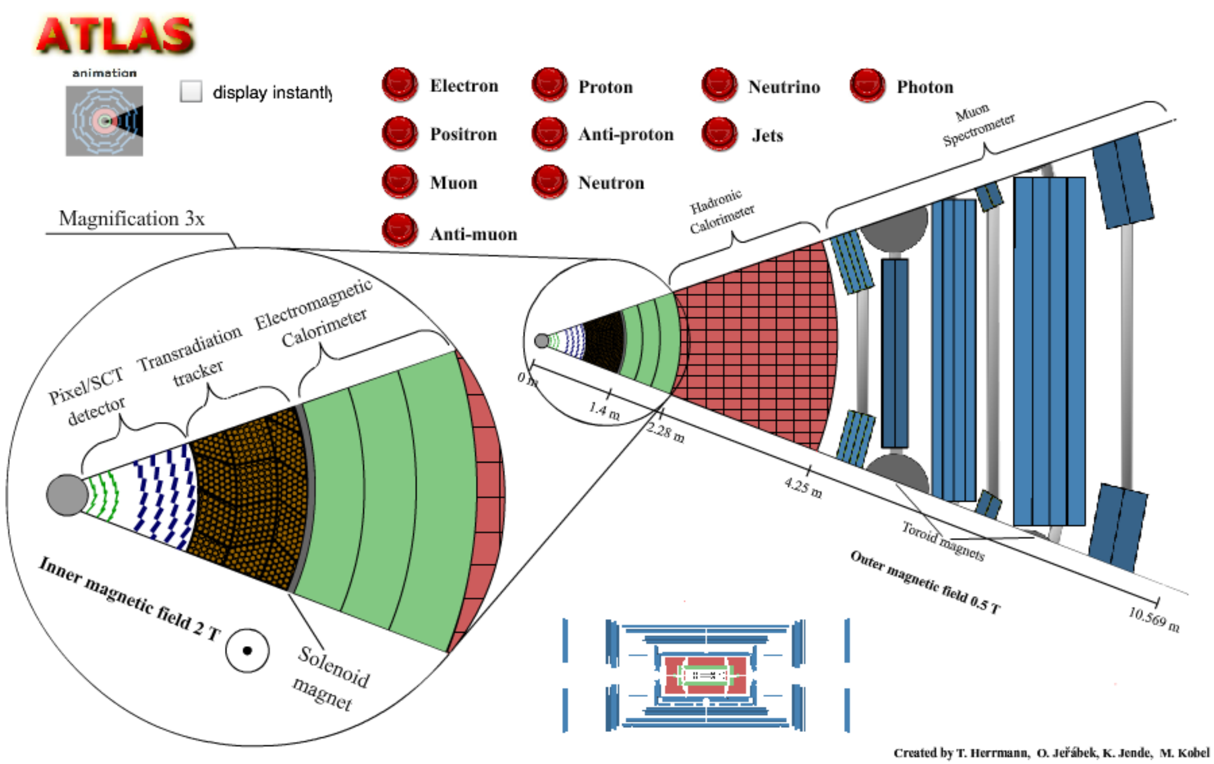
Three generations of matter (fermions)
I
II
III
mass
charge
spin
name
2.4 MeV/c²
⅔
½
u
up
1.27 GeV/c²
⅔
½
c
charm
171.2 GeV/c²
⅔
½
t
top
0
0
1
γ
photon
? GeV/c²
0
0
H
Higgs boson
4.8 MeV/c²
-⅓
½
d
down
104 MeV/c²
-⅓
½
s
strange
4.2 GeV/c²
-⅓
½
b
bottom
0
0
1
g
gluon
<2.2 eV/c²
0
½
νe
electron neutrino
<0.17 MeV/c²
0
½
νμ
muon neutrino
<15.5 MeV/c²
0
½
ντ
tau neutrino
91.2 GeV/c²
0
1
Z0
Z boson
0.511 MeV/c²
-1
½
e
electron
105.7 MeV/c²
-1
½
μ
muon
1.777 GeV/c²
-1
½
τ
tau
80.4 GeV/c²
±1
1
W±
W boson
Quarks
Leptons
Gauge bosons



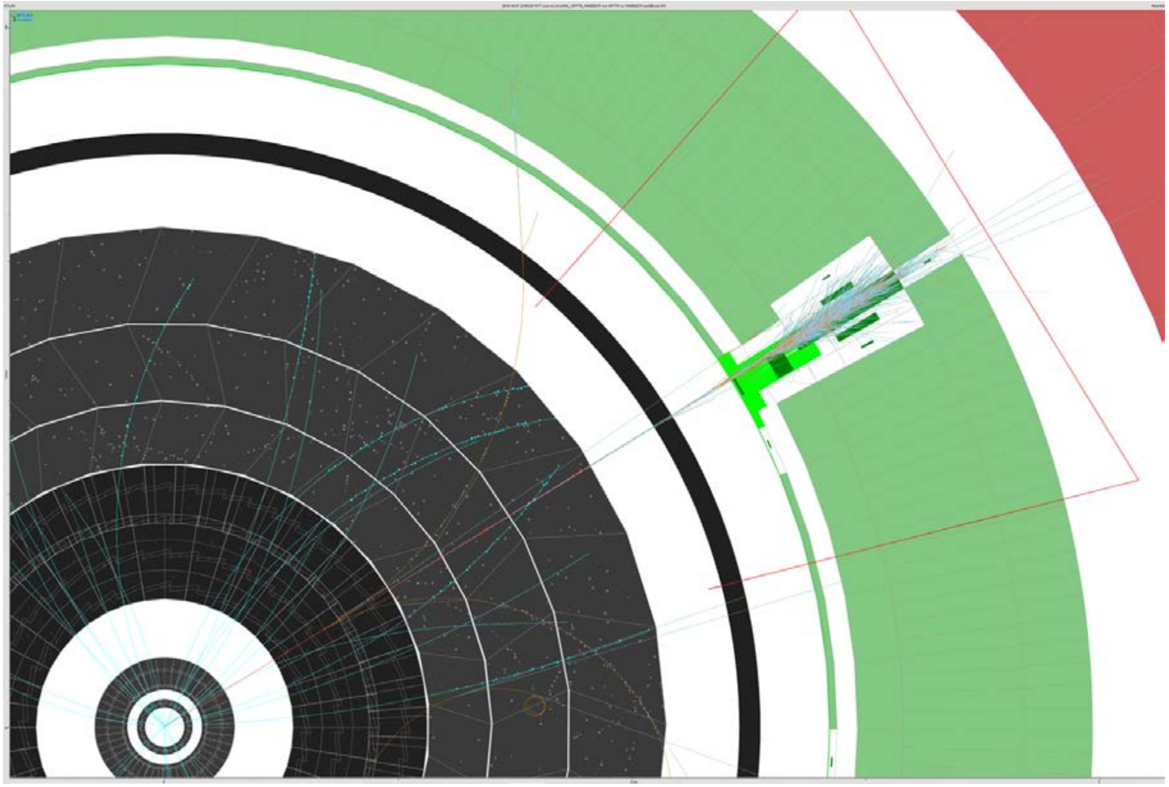
The ATLAS Pixel Detector provides a very high granularity, high precision set of measurements as close to the interaction point as possible. The system provides three precision measurements over the full acceptance, and mostly determines the impact parameter resolution and the ability of the Inner Detector to find short lived particles such as B-Hadrons. The system consists of three barrels at average radii of ~ 5 cm, 9 cm, and 12 cm (1456 modules), and three disks on each side, between radii of 9 and 15 cm (288 modules). Each module is 62.4 mm long and 21.4 mm wide, with 46080 pixel elements read out by 16 chips, each serving an array of 18 by 160 pixels. The 80 million pixels cover an area of 1.7 m^2. The readout chips must withstand over 300 kGy of ionising radiation and over 5x10^14 neutrons per cm^2 over ten years of operation. The modules are overlapped on the support structure to give hermetic coverage. The thickness of each layer is expected to be about 2.5% of a radiation length at normal incidence. Typically three pixel layers are crossed by each track. The pixel detector can be installed independently of the other components of the ID. In the starting phase, only two of the three layers planned for will be installed.



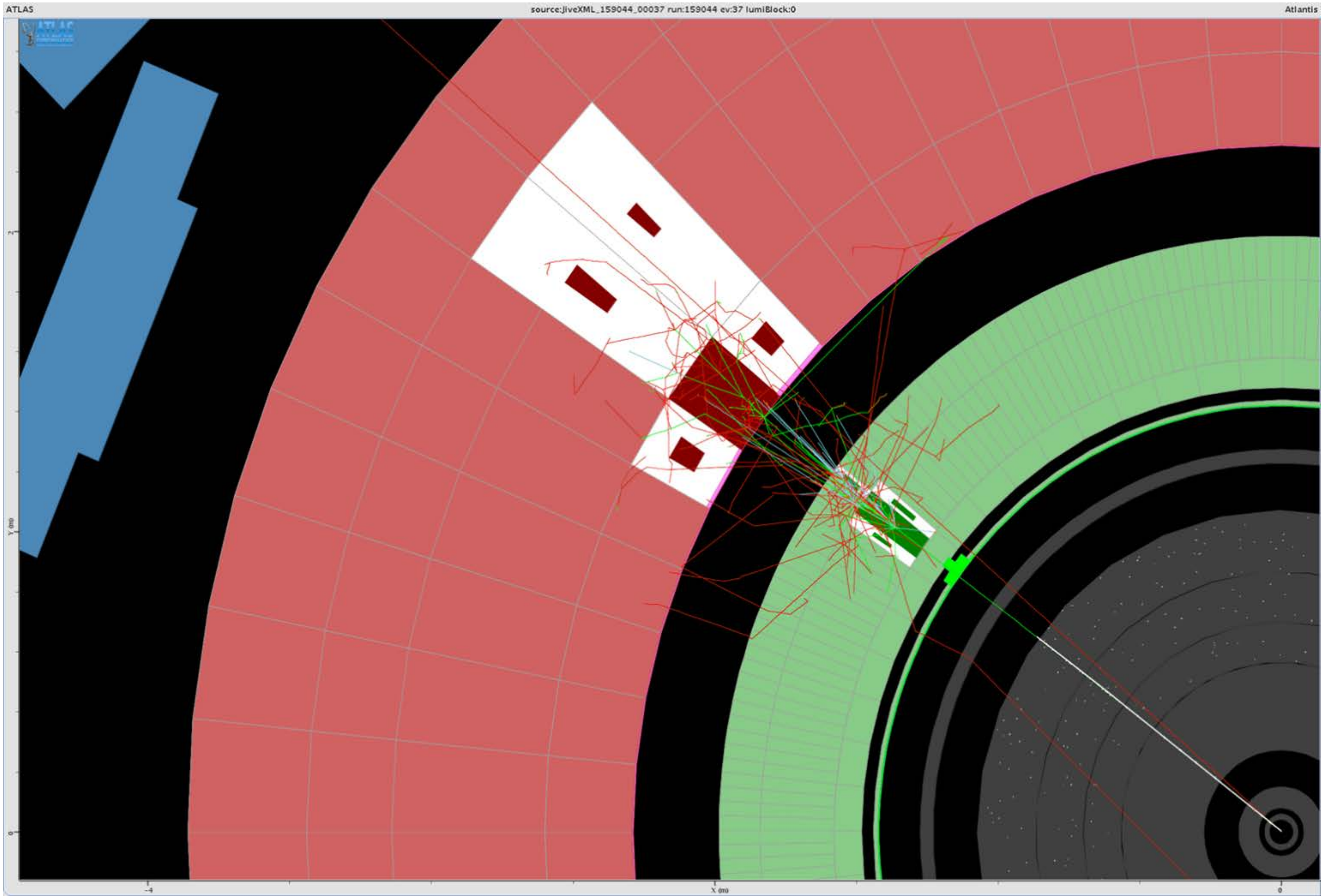
Large Hadron Collider

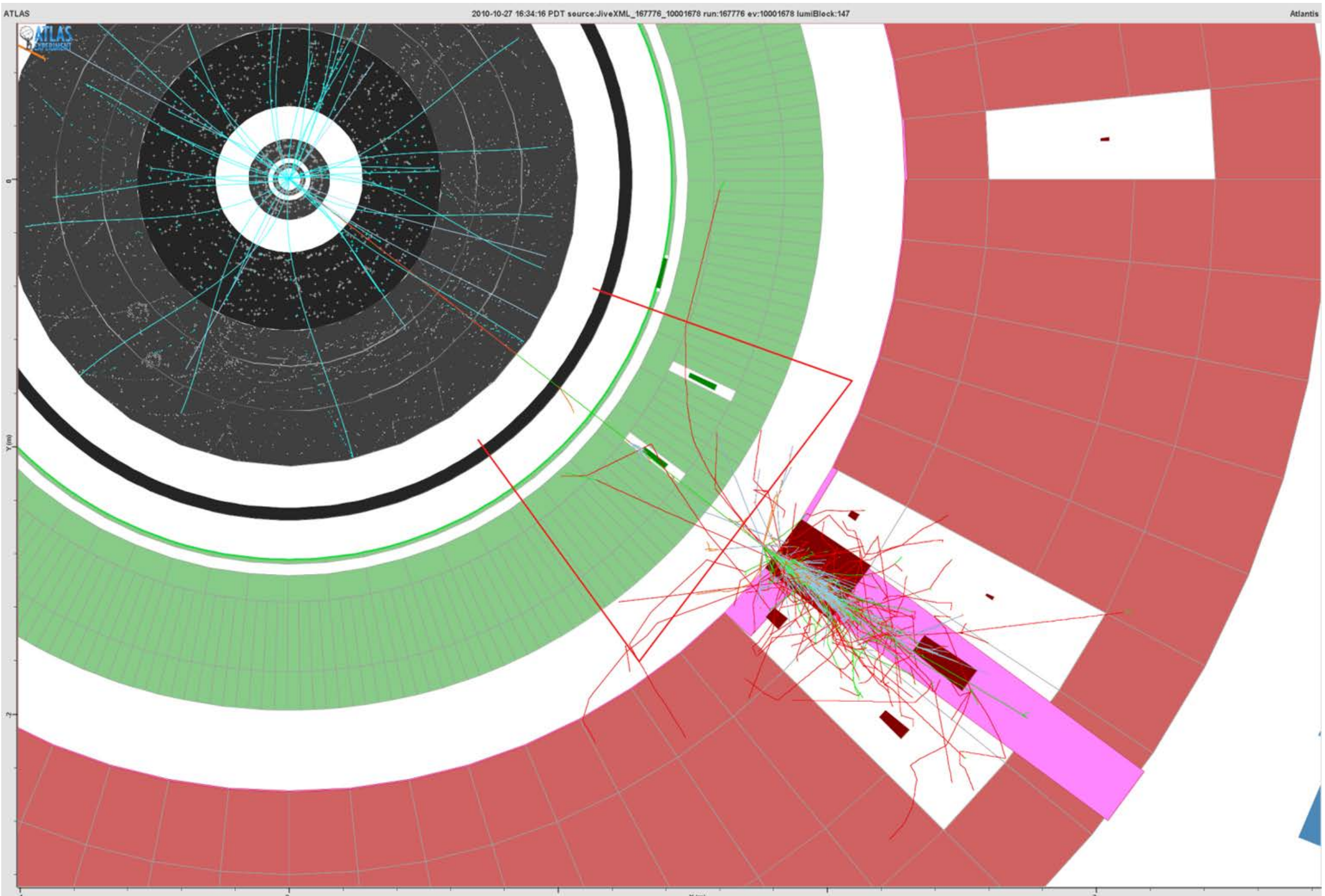


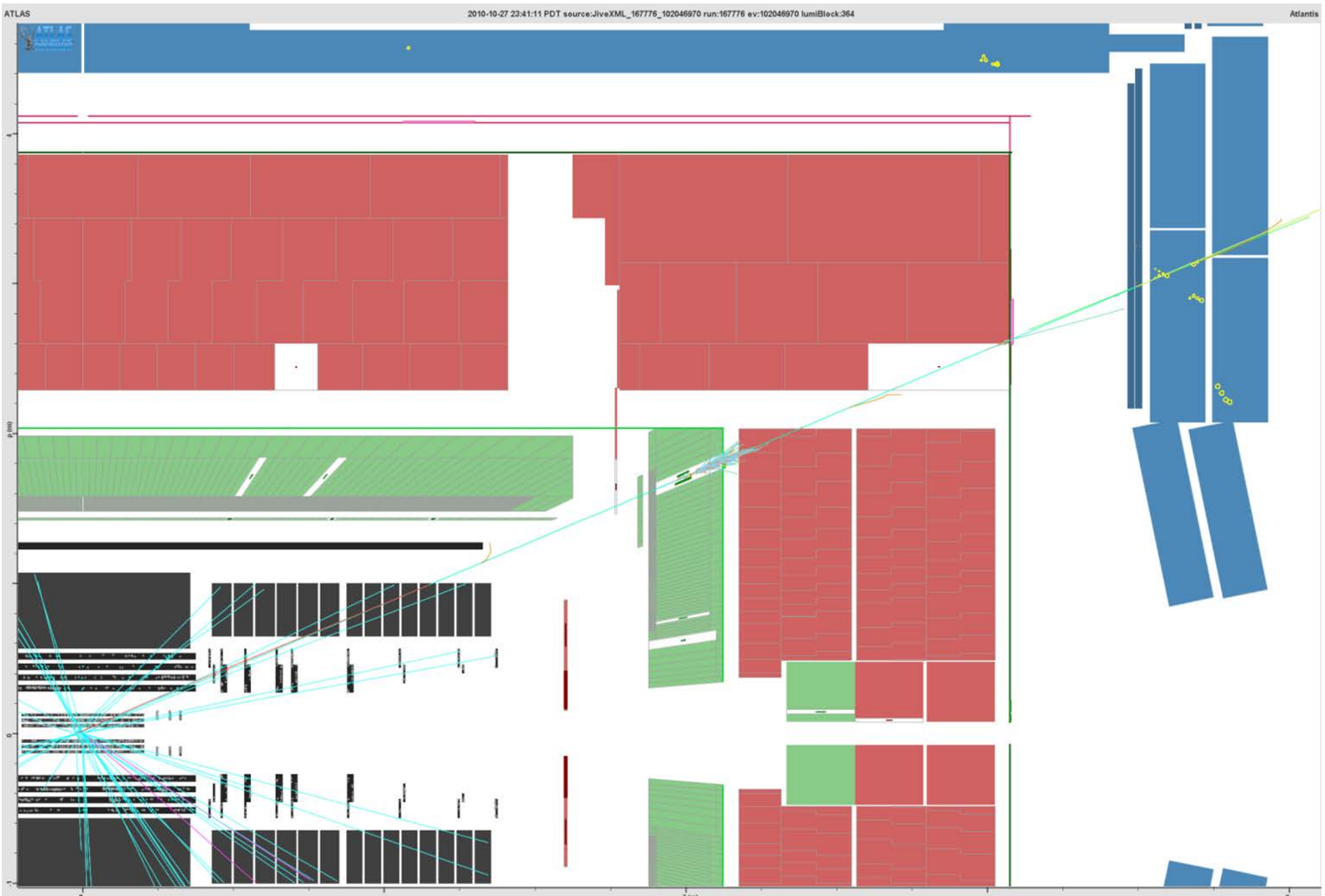
Large Hadron Collider



Large Hadron Collider




ATLAS
2010-10-27 16:34:16 PDT source:JiveXML_167776_10001678 run:167776 ev:10001678 lumiBlock:147
Atlantis
ATLAS EXPERIMENT
Y (m)
X (m)
0
-2
-1
0
3


ATLAS
2010-10-27 23:41:11 PDT source:JiveXML_167776_102046970 run:167776 ev:102046970 lumiBlock:364
Atlantis
ρ (m)
Z (m)

Large Hadron Collider