3nd ASTERICS-OBELICS Workshop

23-25 October 2018, Cambridge, UK.

H2020-Astronomy ESFRI and Research Infrastructure Cluster
(Grant Agreement number: 653477).

Deep Learning in High Energy Physics

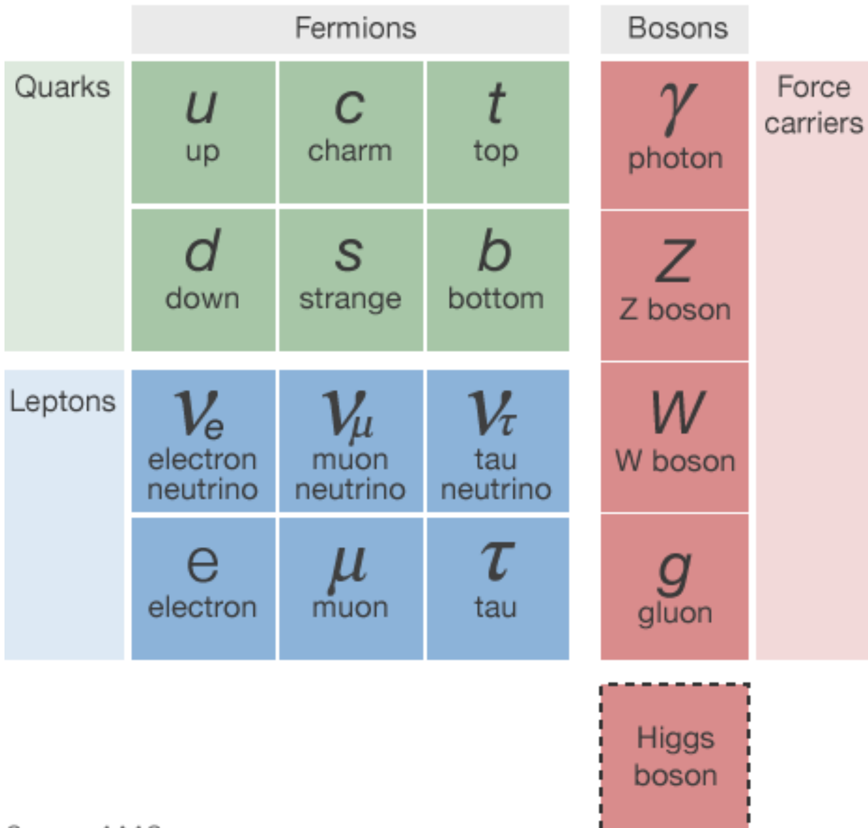
Michael Kagan

SLAC NATIONAL ACCELERATOR LABORATORY

ASTERICS-OBELICS 2018, Cambridge

October 24, 2018

What We Know: The Standard Model



Source: AAAS

$$
\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 + 2H\phi^+\phi^-] - \\
\tfrac{1}{8} g^2 \alpha_h [H^4 + (\phi^0)^4 + 4(\phi^+\phi^-)^2 + 4(\phi^0)^2\phi^+\phi^- + 4H^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}
$$

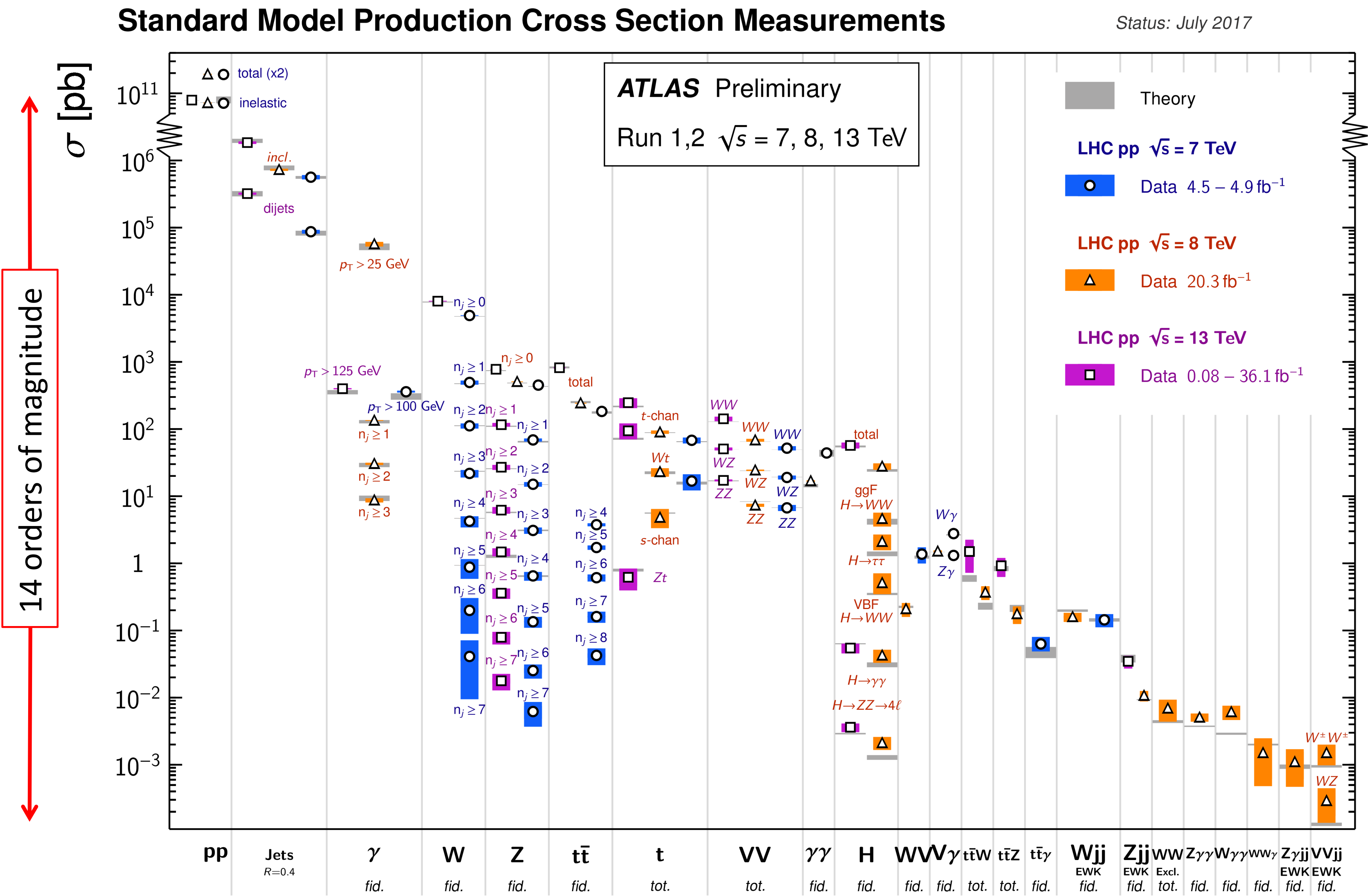
What We Know: The Standard Model

19 parameters

Symbol	Description	Value
m_e	Electron mass	511 keV
m_μ	Muon mass	105.7 MeV
m_τ	Tau mass	1.78 GeV
m_u	Up quark mass	1.9 MeV
m_d	Down quark mass	4.4 MeV
m_s	Strange quark mass	87 MeV
m_c	Charm quark mass	1.32 GeV
m_b	Bottom quark mass	4.24 GeV
m_t	Top quark mass	172.7 GeV
θ_{12}	CKM 12-mixing angle	13.1°
θ_{23}	CKM 23-mixing angle	2.4°
θ_{13}	CKM 13-mixing angle	0.2°
δ	CKM CP-violating Phase	0.995
g_1	U(1) gauge coupling	0.357
g_2	SU(2) gauge coupling	0.652
g_3	SU(3) gauge coupling	1.221
θ_{QCD}	QCD vacuum angle	~0
v	Higgs vacuum expectation value	246 GeV
m_H	Higgs mass	125 GeV

$$
\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}ig_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 - igc_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)] - igs_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^2c_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) - 2A_\mu Z^0_\mu W^+_\nu W^-_\nu] - g\alpha[H^3 + H\phi^0\phi^0 + 2H\phi^+\phi^-] - \\
\tfrac{1}{8}g^2\alpha_h[H^4 + (\phi^0)^4 + 4(\phi^+\phi^-)^2 + 4(\phi^0)^2\phi^+\phi^- + 4H^2\phi^+\phi^- + 2(\phi^0)^2H^2] - \\
gMW^+_\mu W^-_\mu H - \tfrac{1}{2}g\tfrac{M}{c_w^2}Z^0_\mu Z^0_\mu H - \tfrac{1}{2}ig[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) - ig\tfrac{s_w^2}{c_w}MZ^0_\mu(W^+_\mu\phi^- - W^-_\mu\phi^+) + \\
igs_w MA_\mu(W^+_\mu\phi^- - W^-_\mu\phi^+) - ig\tfrac{1-2c_w^2}{2c_w}Z^0_\mu(\phi^+\partial_\mu\phi^- - \phi^-\partial_\mu\phi^+) + \\
igs_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}ig^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}ig^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 + igs_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 + igc_w W^+_\mu(\partial_\mu\bar{X}^0 X^- - \partial_\mu\bar{X}^+ X^0) + igs_w W^+_\mu(\partial_\mu\bar{Y}X^- - \\
\partial_\mu\bar{X}^+ Y) + igc_w W^-_\mu(\partial_\mu\bar{X}^- X^0 - \partial_\mu\bar{X}^0 X^+) + igs_w W^-_\mu(\partial_\mu\bar{X}^- Y - \\
\partial_\mu\bar{Y}X^+) + igc_w Z^0_\mu(\partial_\mu\bar{X}^+ X^+ - \partial_\mu\bar{X}^- X^-) + igs_w A_\mu(\partial_\mu\bar{X}^+ X^+ - \\
\partial_\mu\bar{X}^- X^-) - \tfrac{1}{2}gM[\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}igM[\bar{X}^+ X^0\phi^+ - \bar{X}^- X^0\phi^-] + \tfrac{1}{2c_w}igM[\bar{X}^0 X^-\phi^+ - \bar{X}^0 X^+\phi^-] + \\
igMs_w[\bar{X}^0 X^-\phi^+ - \bar{X}^0 X^+\phi^-] + \tfrac{1}{2}igM[\bar{X}^+ X^+\phi^0 - \bar{X}^- X^-\phi^0]
\end{gathered}
$$

The Standard Model in Action

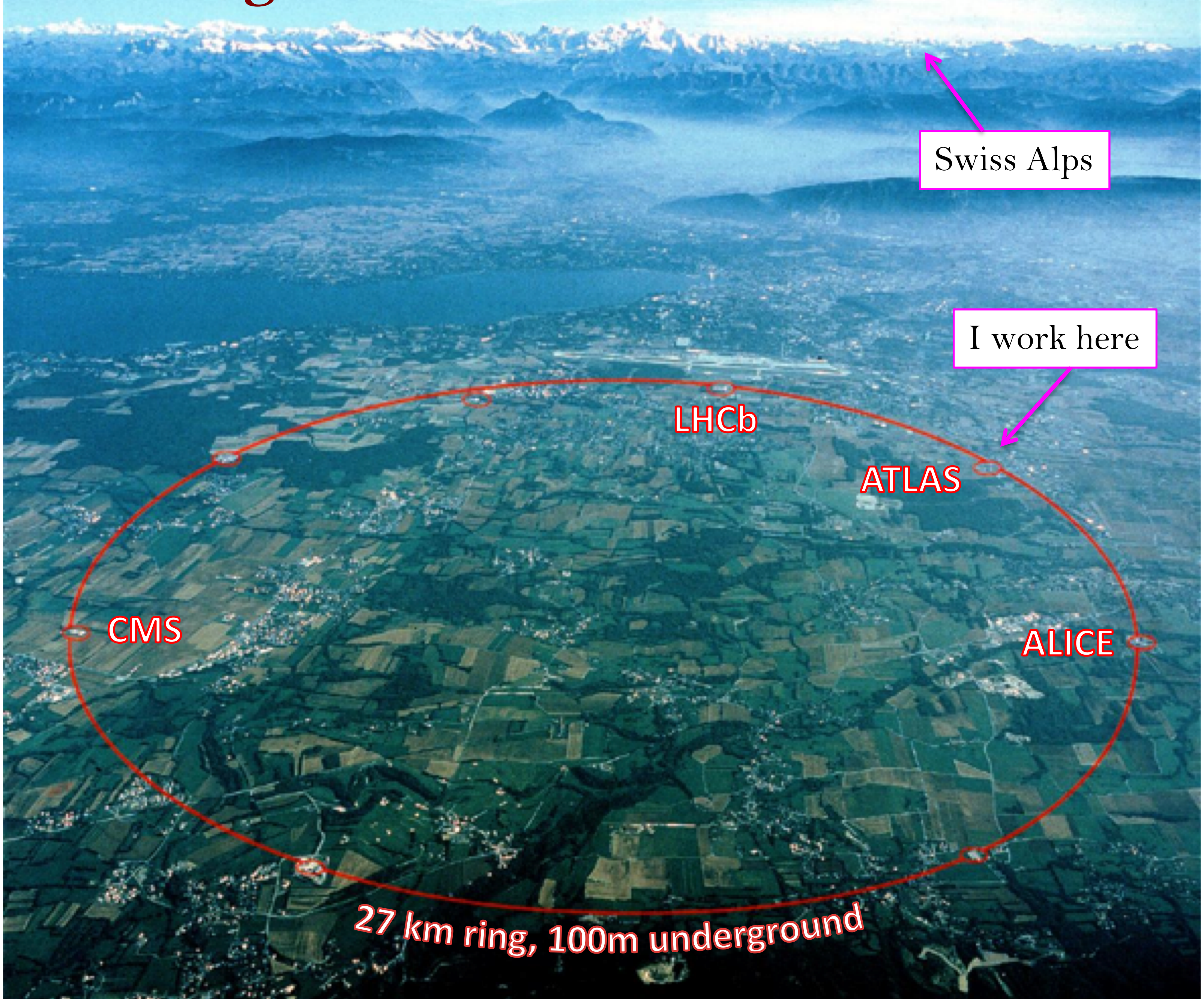


What are we missing?

- Is the Higgs boson connected to New Physics?
- Why Gravity so much weaker then the other forces?
- What are Dark Matter and Dark Energy?
- What gives neutrinos their mass?
- Why is there a matter /anti-matter asymmetry?
- …

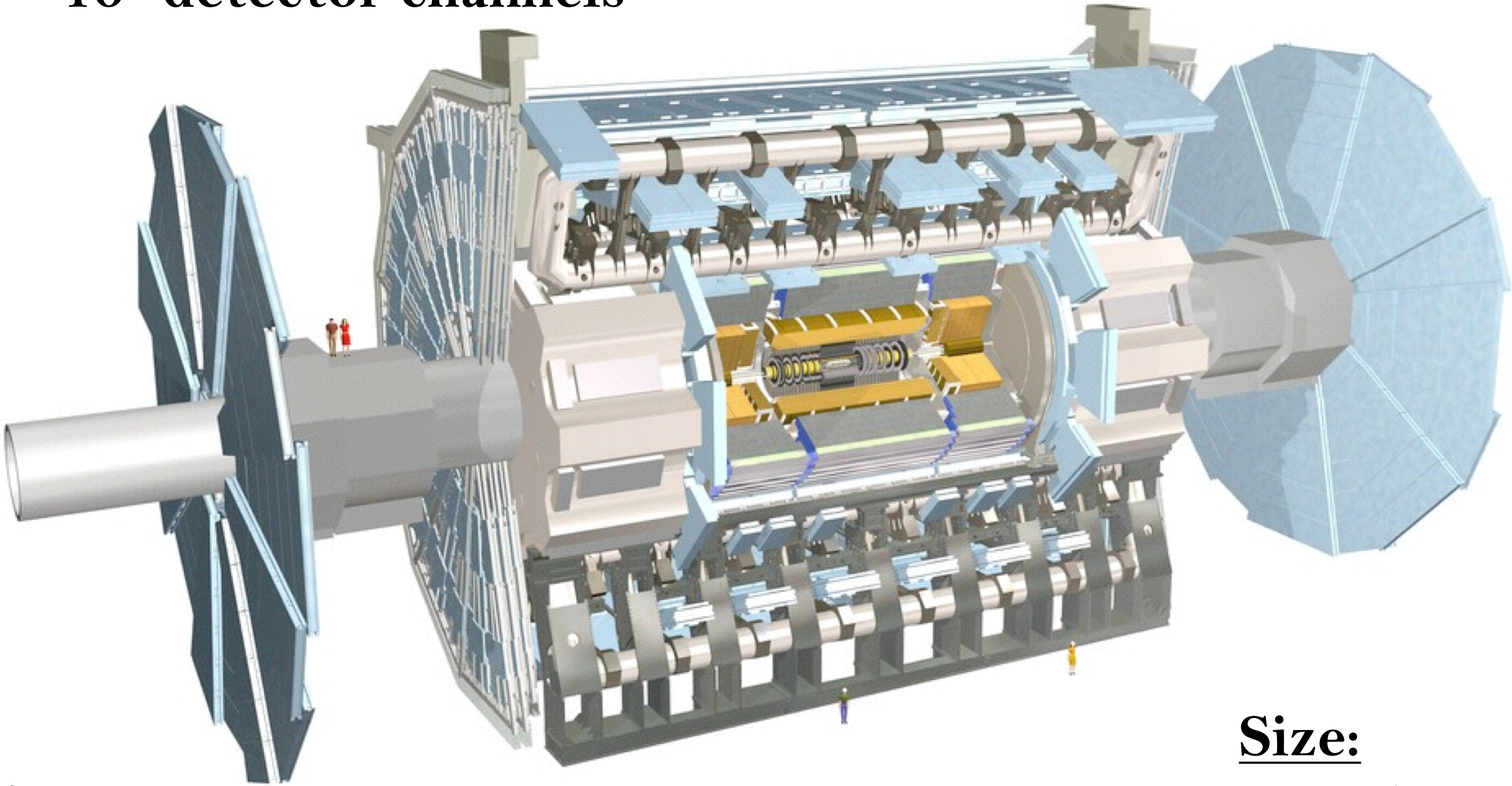
- New forces and heavy particles may have been active during the early universe that explain these phenomena
- We can look for them in high energy physics experiments!
- How can machine learning help?

The Large Hadron Collider at CERN



The ATLAS Experiment

$\sim 10^8$ detector channels



Data:

~300 MB / sec

~3000 TB / year

Weight:

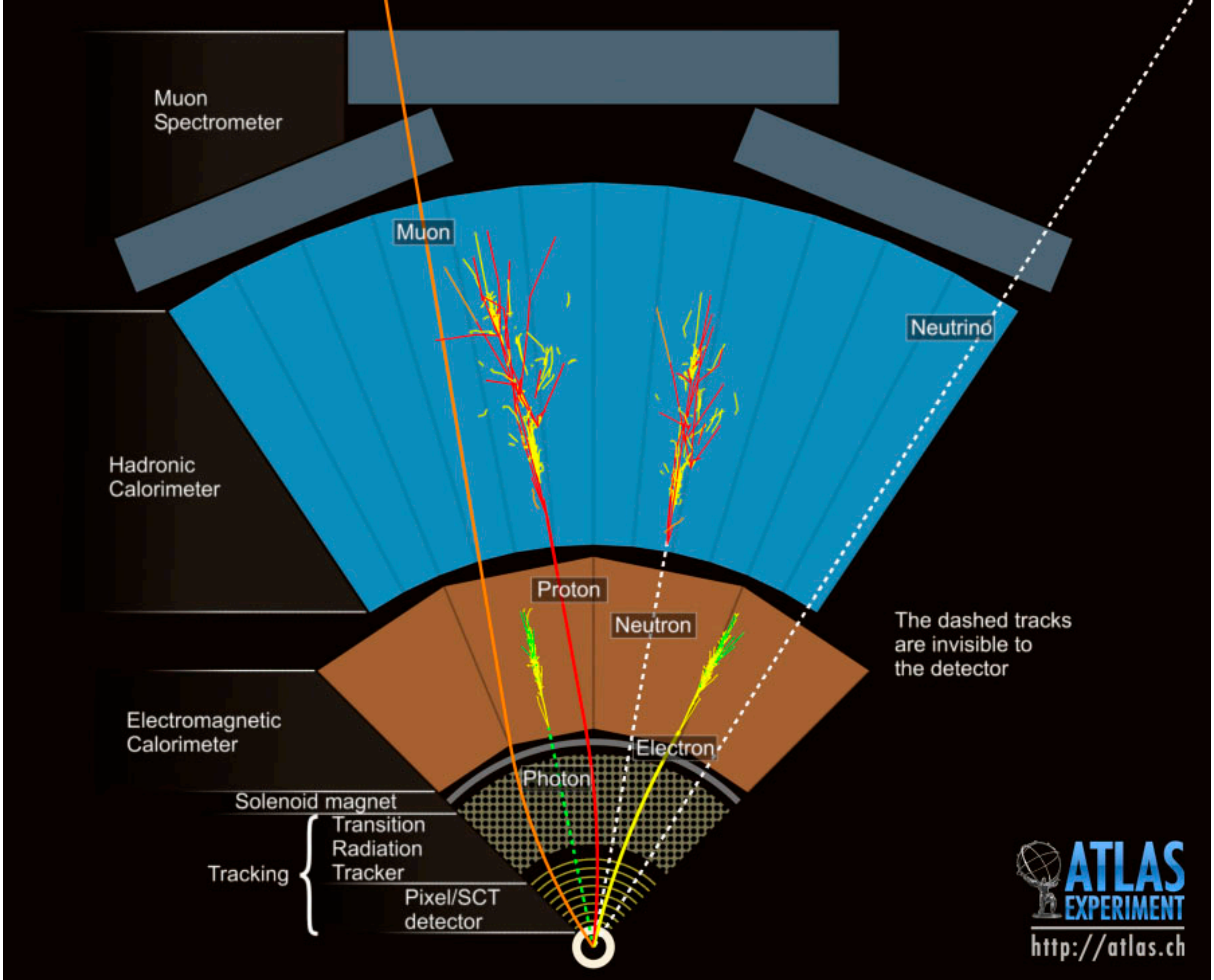
7000 tons

Size:

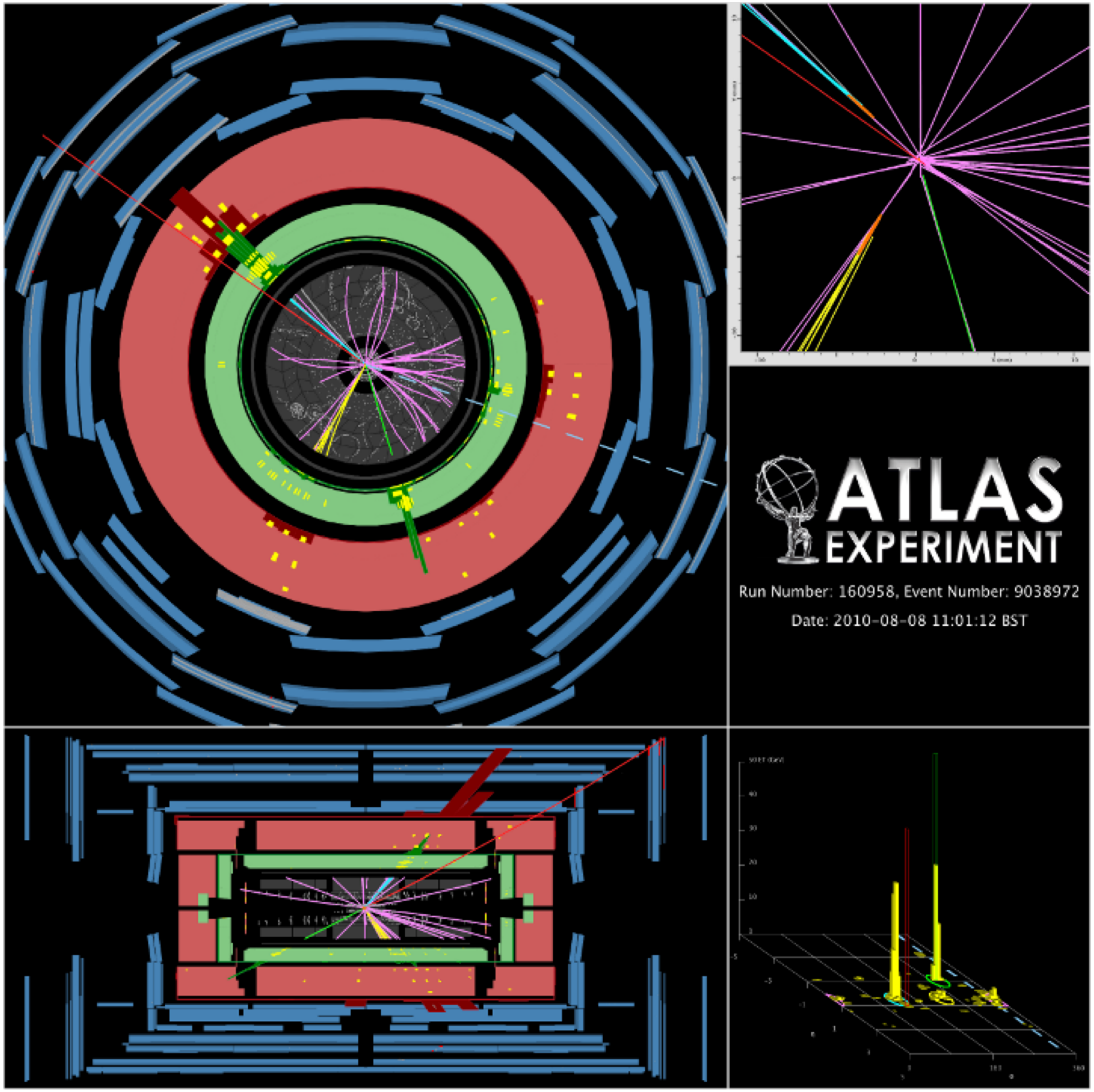
46 m long,

25 m high,

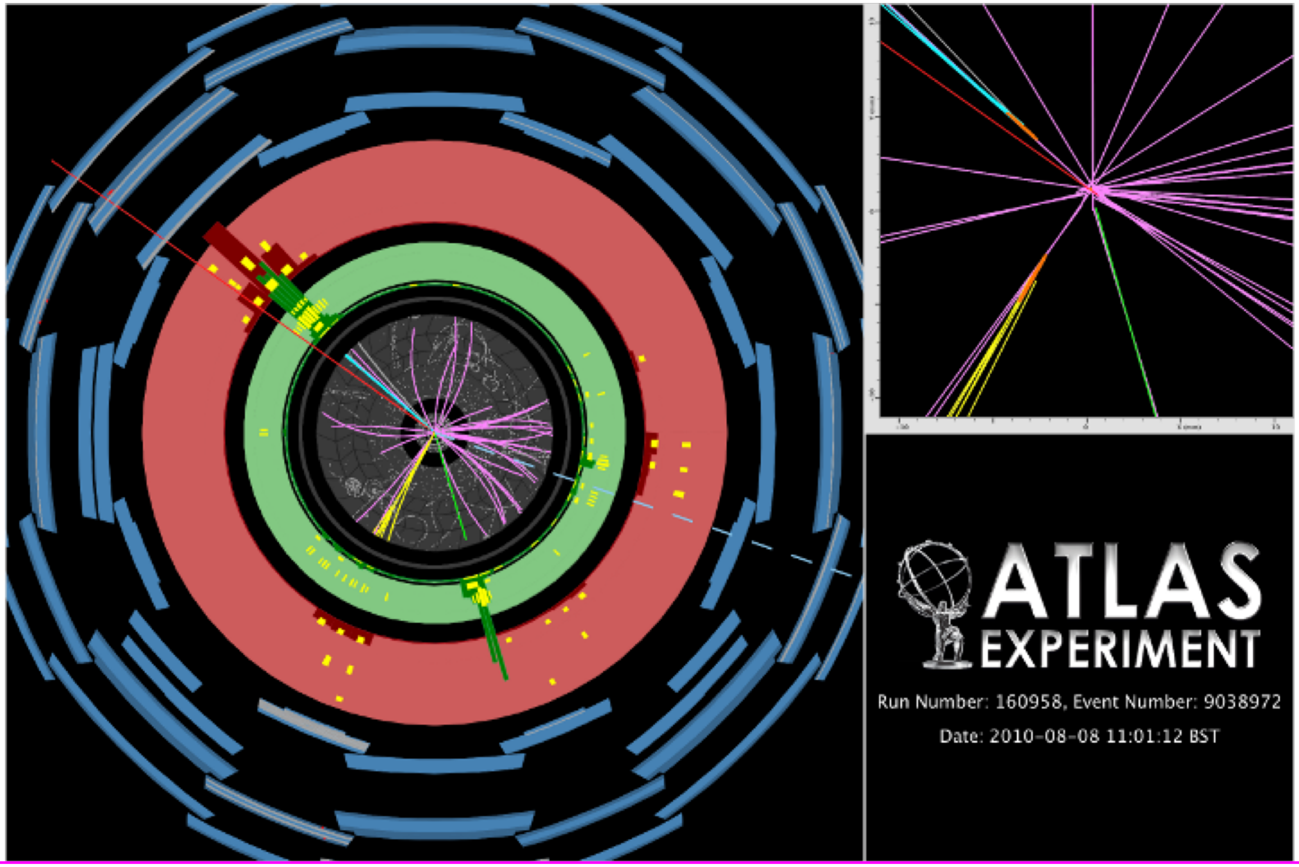
25 m wide


Muon Spectrometer
Muon
Neutrino
Hadronic Calorimeter
Proton
Neutron
The dashed tracks are invisible to the detector
Electromagnetic Calorimeter
Electron
Photon
Solenoid magnet
Transition Radiation Tracker
Tracking
Pixel/SCT detector
ATLAS EXPERIMENT
http://atlas.ch

Studying Collisions



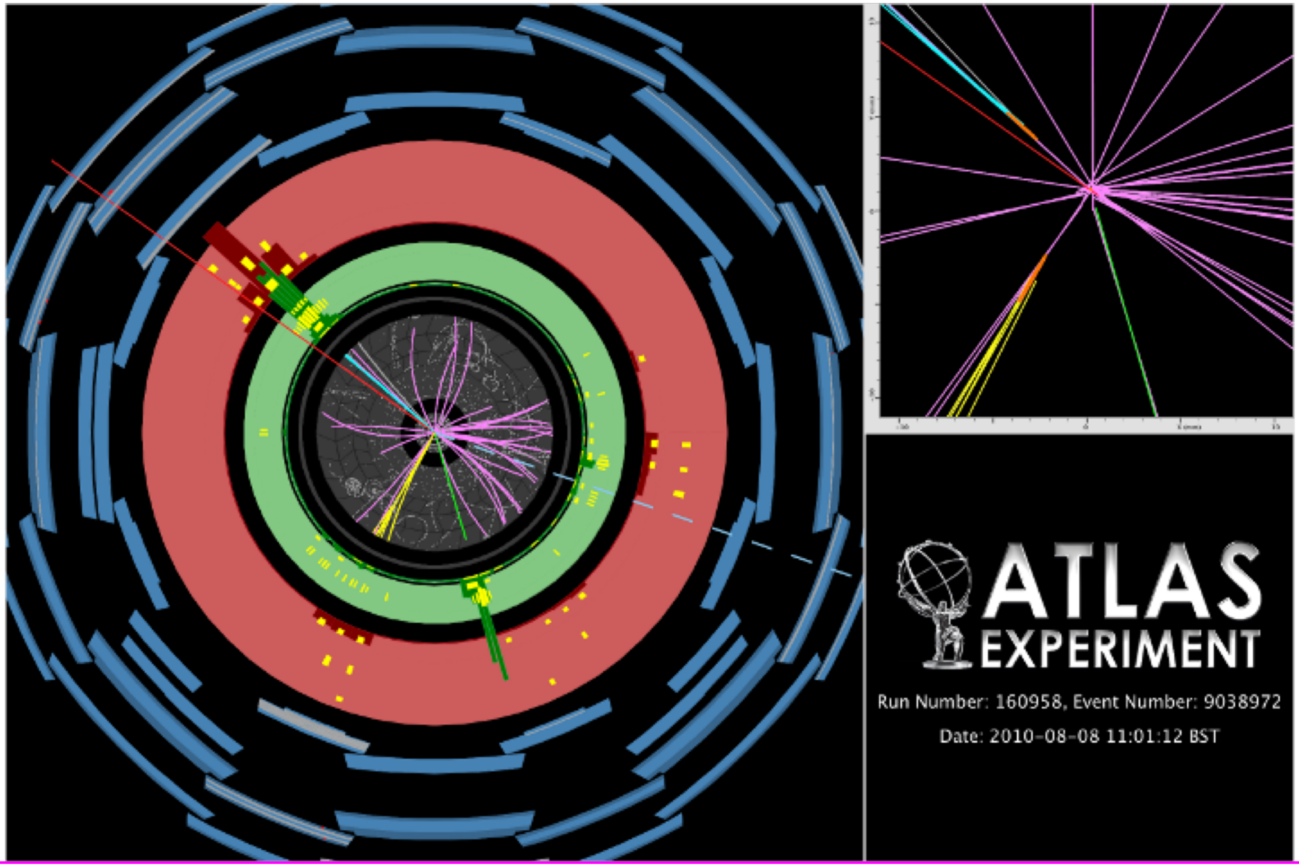
Studying Collisions



- Causal and Compositional Structure

Collision → particle X → "final state" particles → detector data

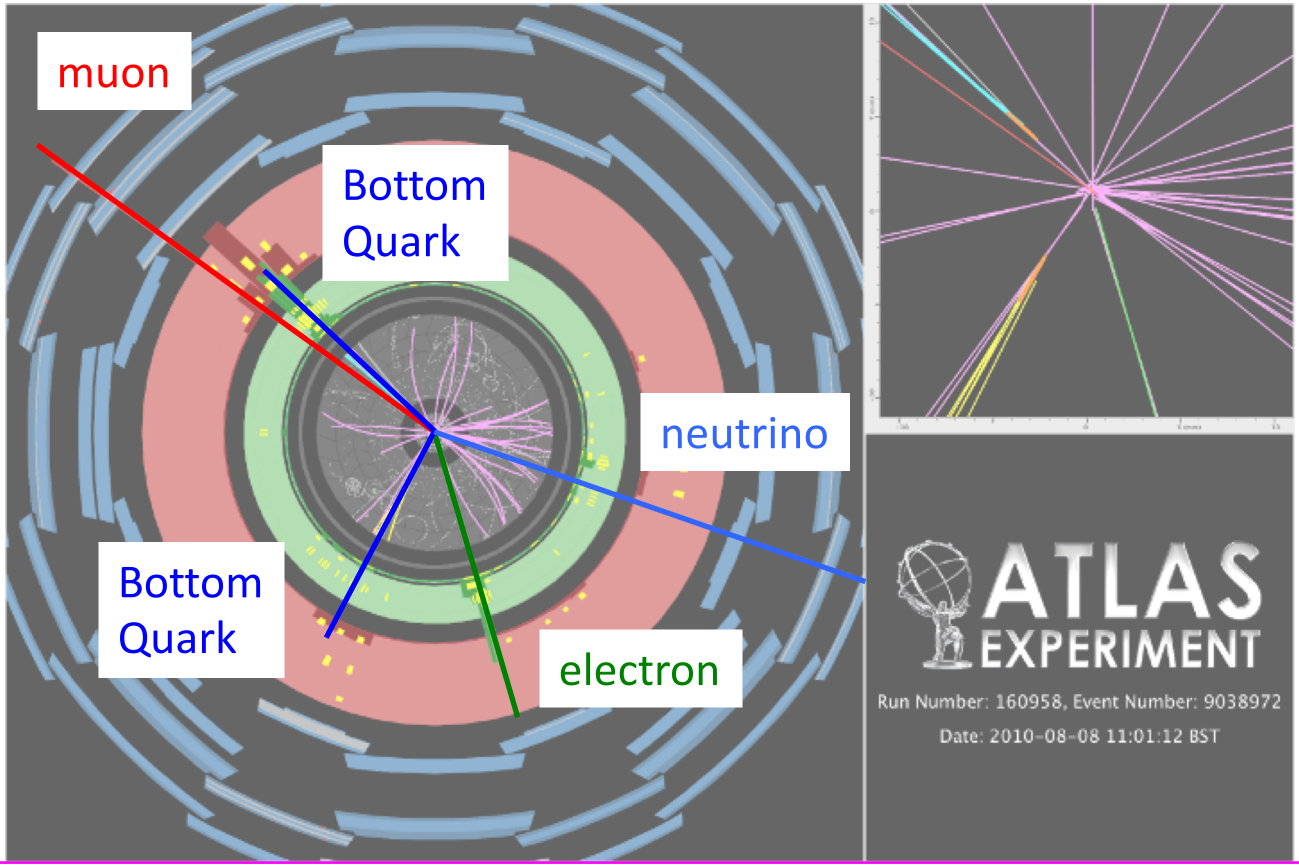
Studying Collisions



- Causal and Compositional Structure

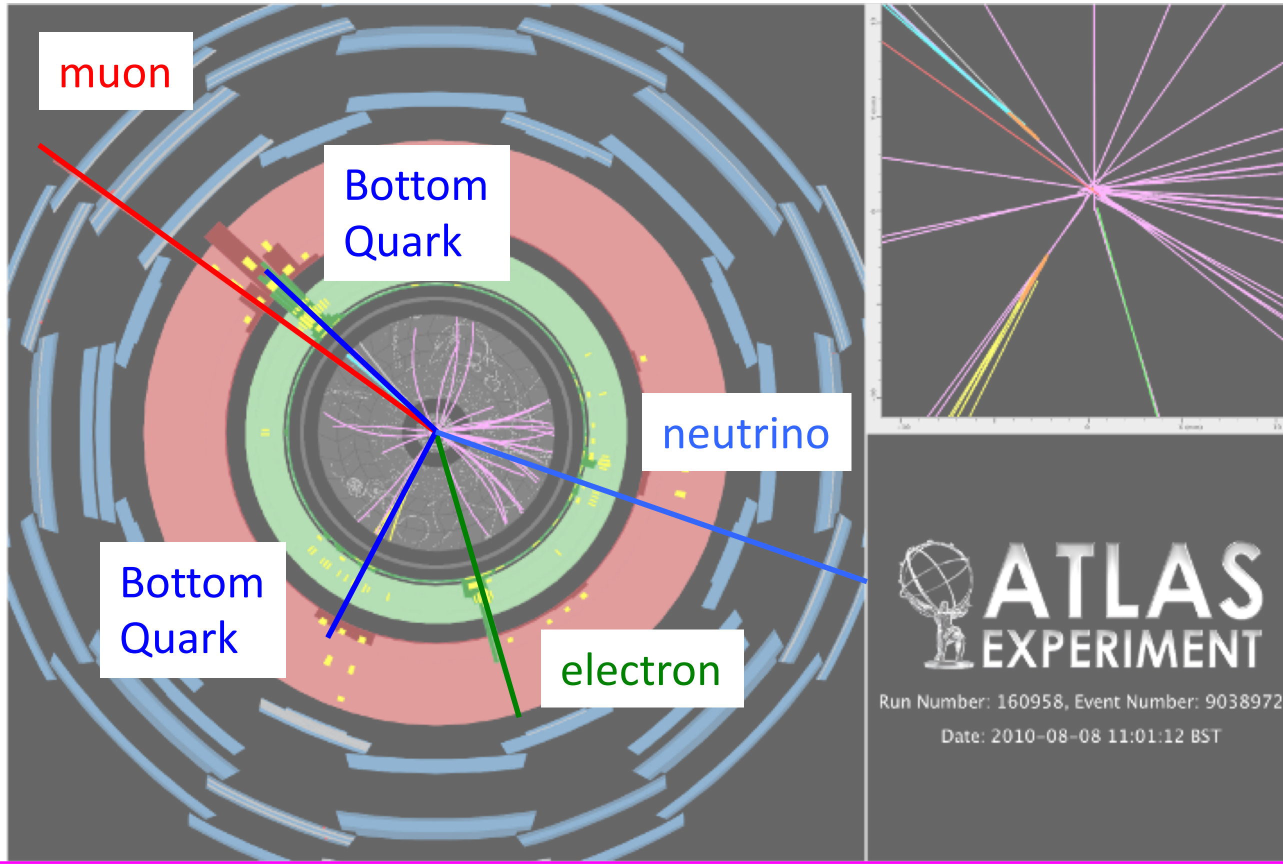
particle X ← "final state" particles ← detector data

Studying Collisions

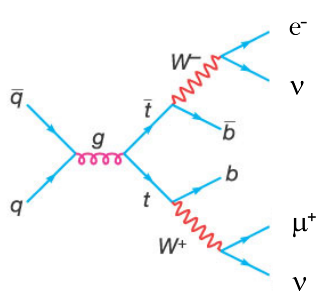


- Reconstruction: Find the "final state" particles in the detector

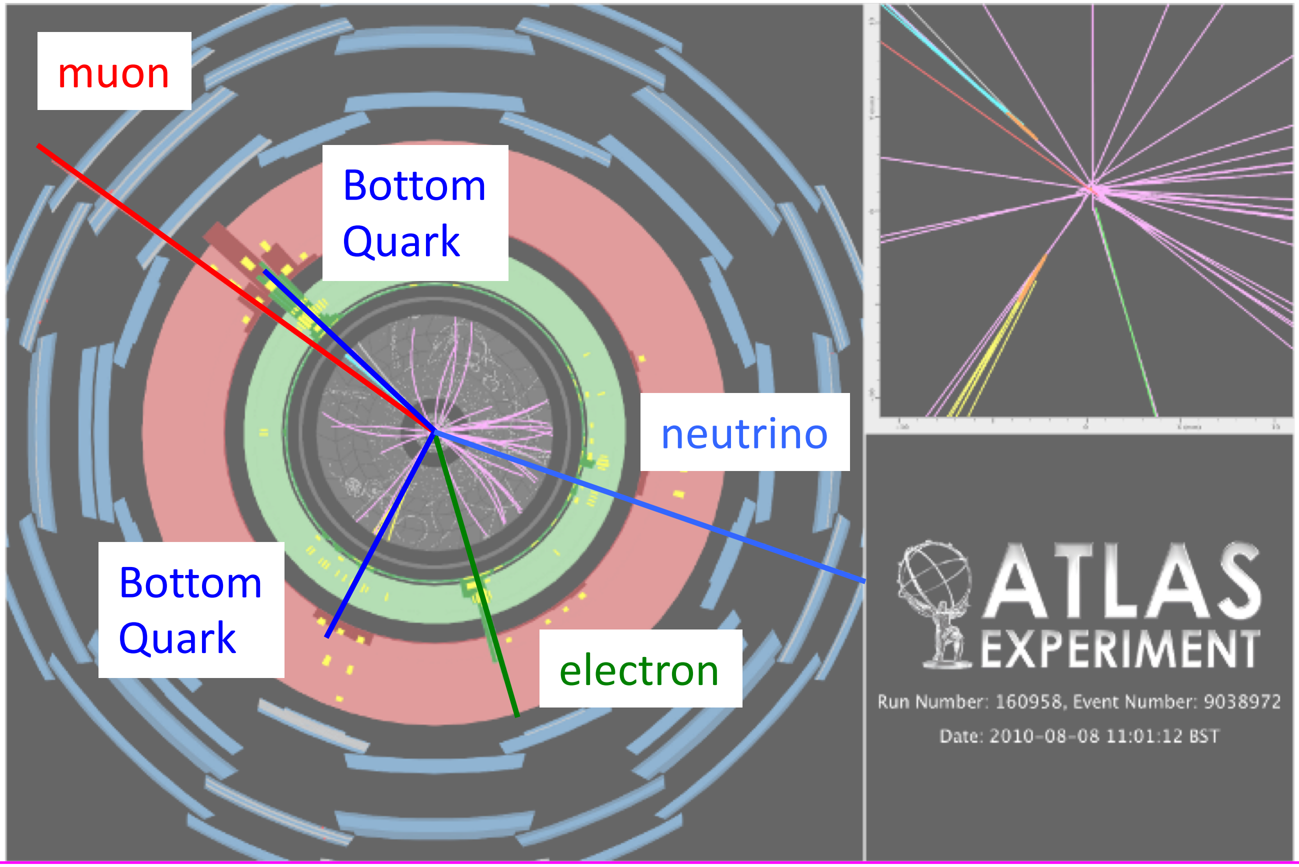
Studying Collisions



- Add them together to study underlying collision



Studying Collisions

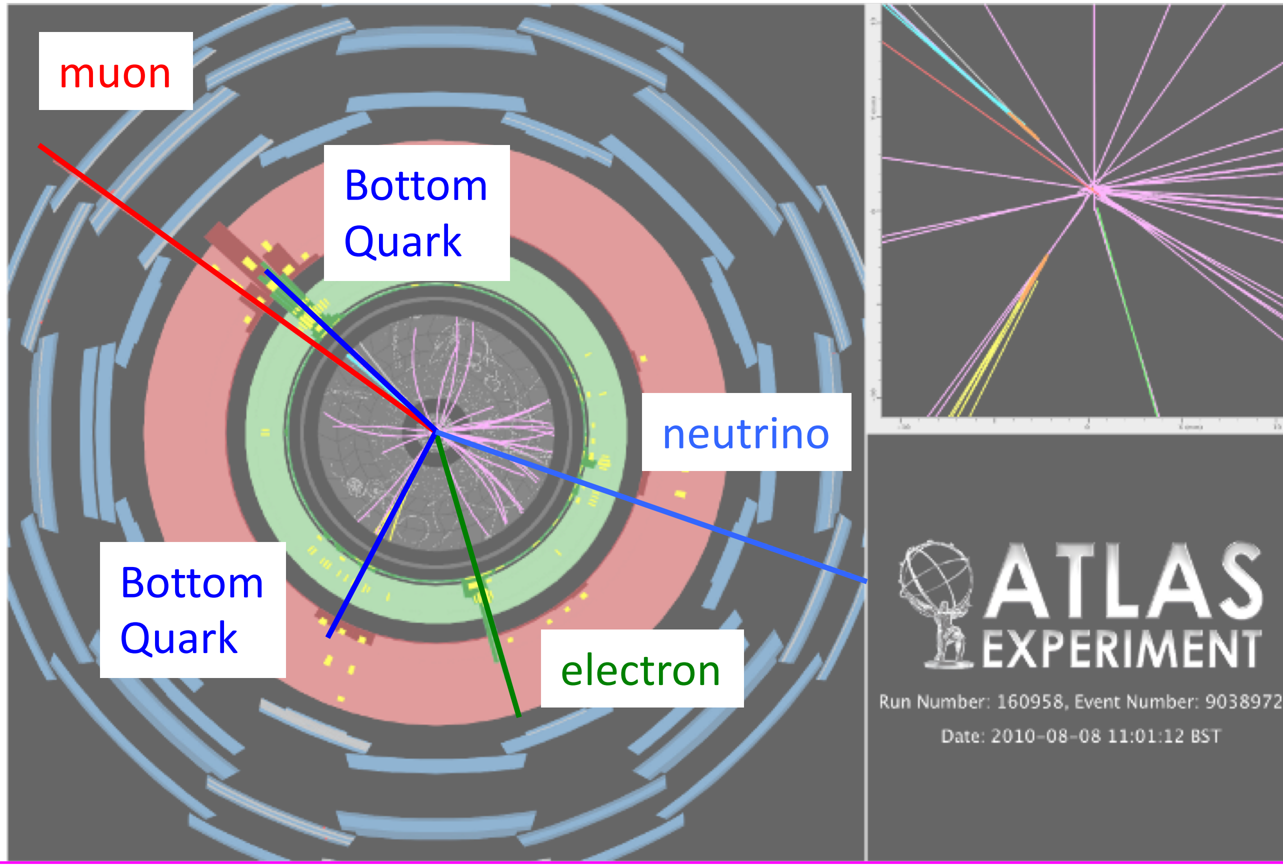


- Works because of energy and momentum conservation:

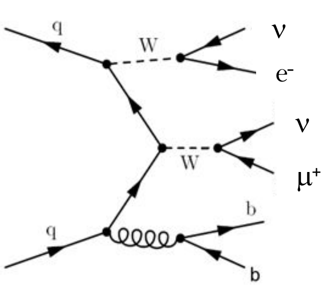
$$(E_\mathrm{X}, \vec{p}_\mathrm{X}) = \sum_{i \in \text{decay products}} (E_i, \vec{p}_i)$$

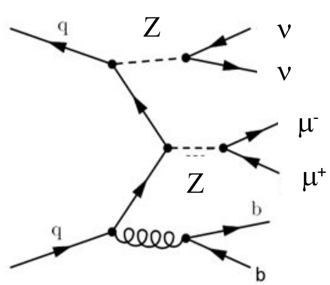
$$M_\mathrm{X} c^2 = \sqrt{E_\mathrm{X}^2 - \vec{p}_\mathrm{X}^{\,2}\, c^2}$$

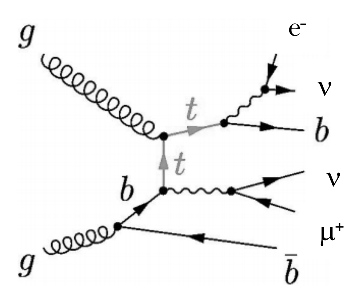
Studying Collisions



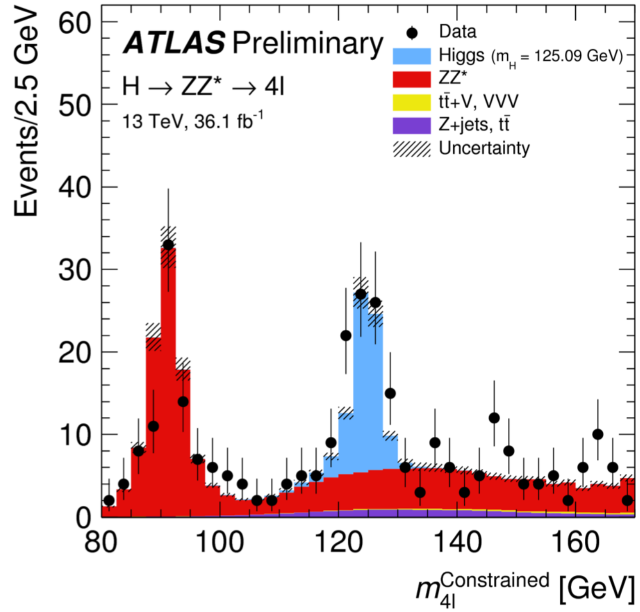
- Multiple processes contribute to same signature





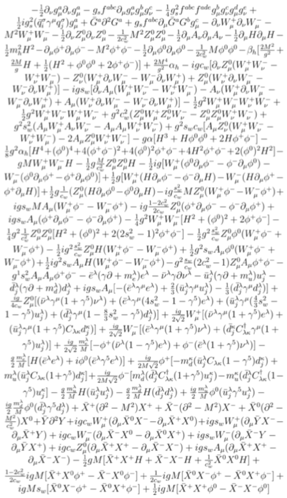


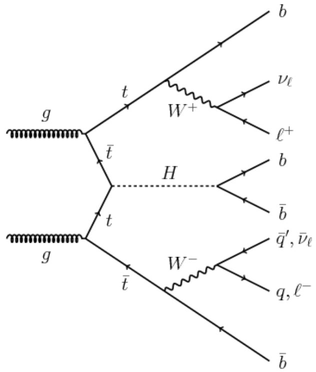
Studying Collisions



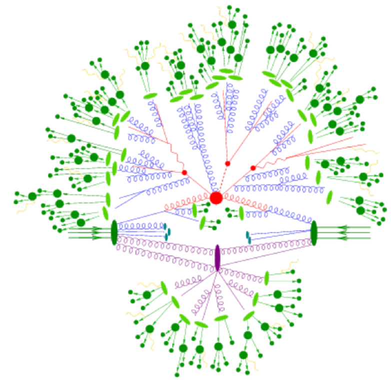
- With a collection of collisions we can perform:
  - Hypothesis testing: new particle present?
  - Measurement: Inference of latent parameters, e.g. Higgs mass
- Extremely accurate simulations + knowledge of the data generating process (i.e. physics) to analyze our data!

From Theory to Experiment

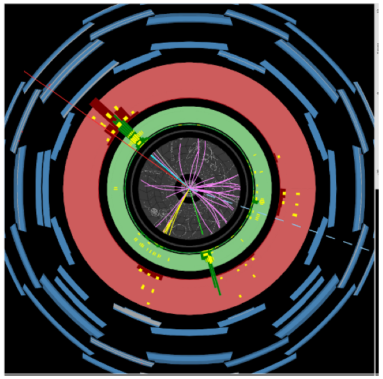




O(10) particles

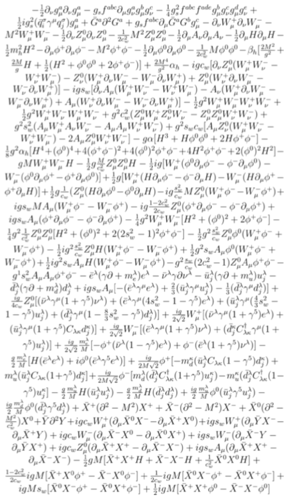


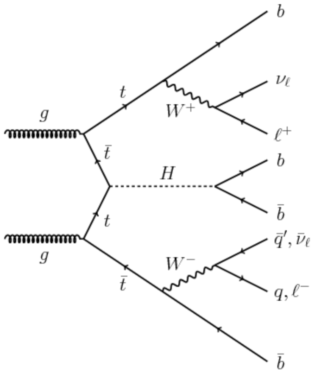
O(100) particles

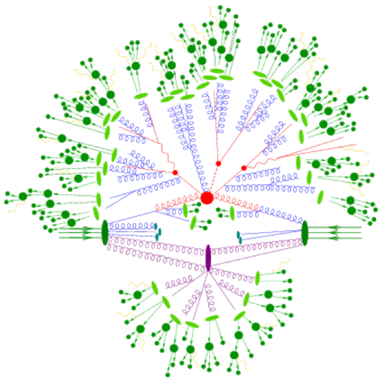


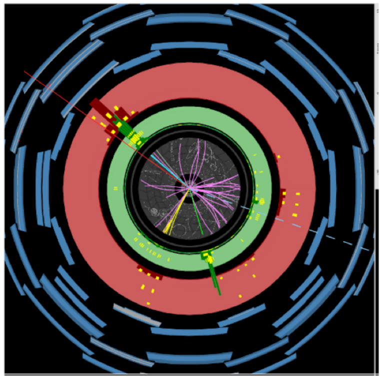
O(10^8) detector elements

From Theory to Experiment









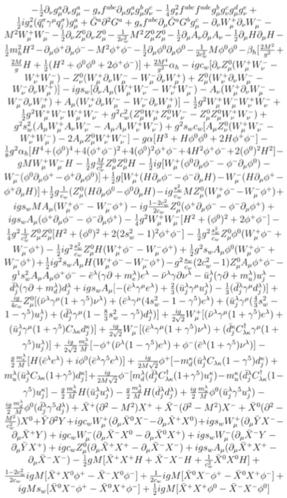
O(10) particles

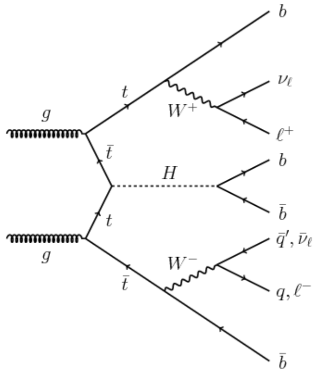
O(100) particles

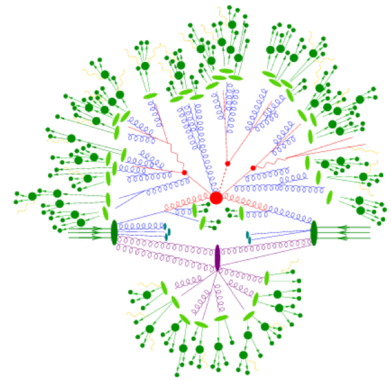
O(10^8) detector elements

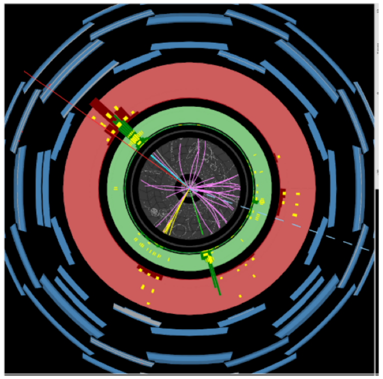
Hypothesis testing / Measurement ← **Event Reconstruction and Selection** ← **Reconstruction of particles** ← **Detector data**

From Theory to Experiment









O(10) particles

O(100) particles

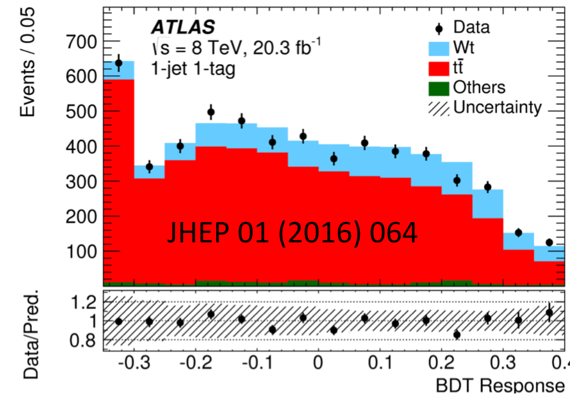
O(10^8) detector elements

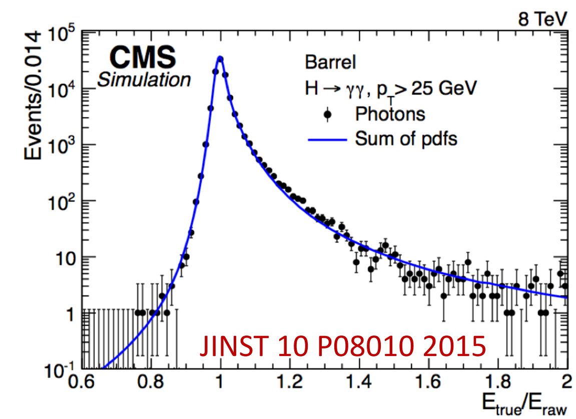
Hypothesis testing / Measurement ← **Event Reconstruction and Selection** ← **Reconstruction of particles** ← **Detector data**

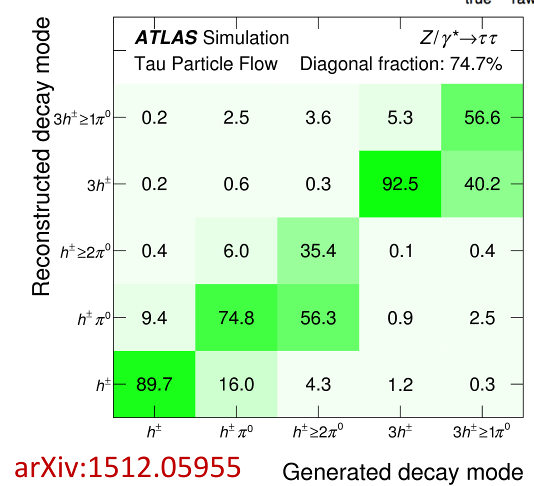
- Build on our knowledge of how the data is created
  - Use our simulation to design and study reconstruction algorithms, and to compare predictions with our experimental data
- Use Machine learning to improve (or rethink) the steps of this process?

Machine Learning Applied Widely in HEP

- **In analysis**:
  - Classifying signal from background, especially in complex final states
  - Reconstructing heavy particles and improving the energy / mass resolution
- **In reconstruction**:
  - Improving detector level inputs to reconstruction
  - Particle identification tasks
  - Energy / direction calibration
- **In the trigger**:
  - Quickly identifying complex final states
- **In computing**:
  - Estimating dataset popularity, and determining needed number and location of dataset replicas

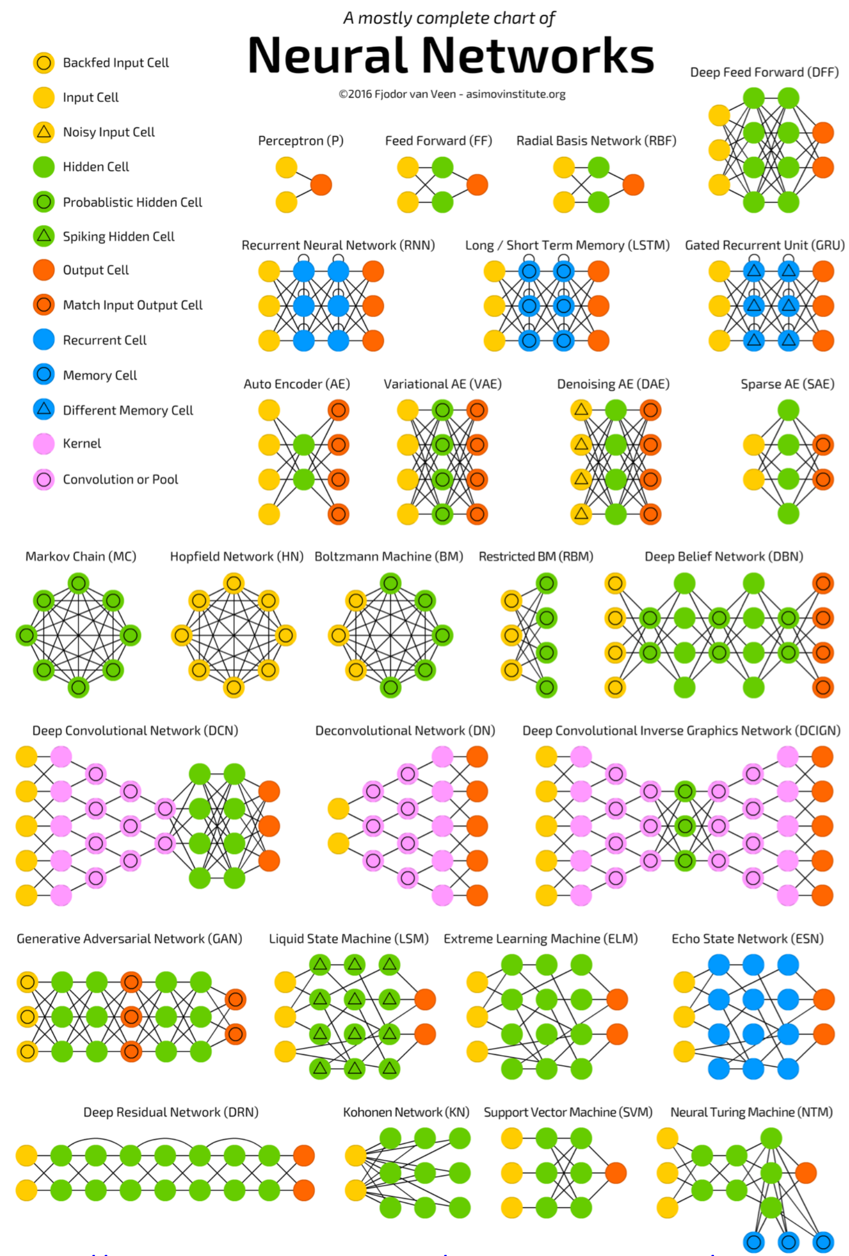






Deep Learning for HEP

- How do we move deep learning advancements into HEP?
  - Translate problems in HEP into problems in ML domain
  - Incorporate HEP domain knowledge when building models
  - How do we extract what is learned?



http://www.asimovinstitute.org/neural-network-zoo/

Deep Learning in the Pipeline

Experiment
Data Collection

Simulation of
physics + detector

Generative Model

Reconstruction
Raw data to final
state paticles

Cluster energy depositions
Classify clusters as particles
Infer / regress properties

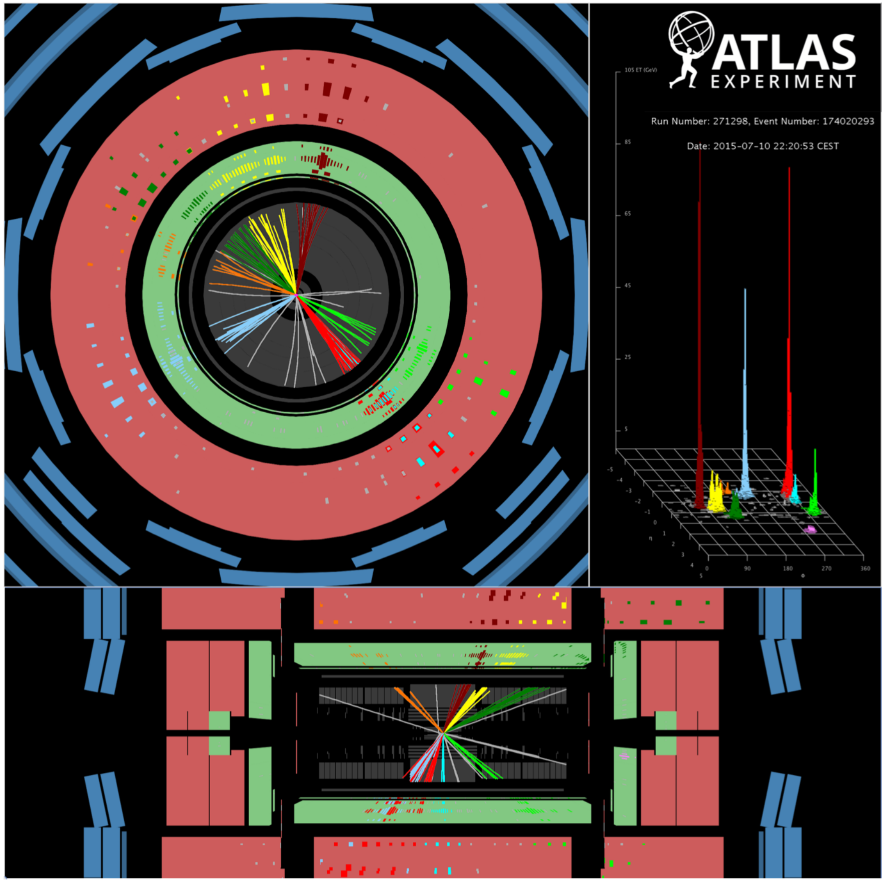
Infer properties of collision
Classification of interesting collisions

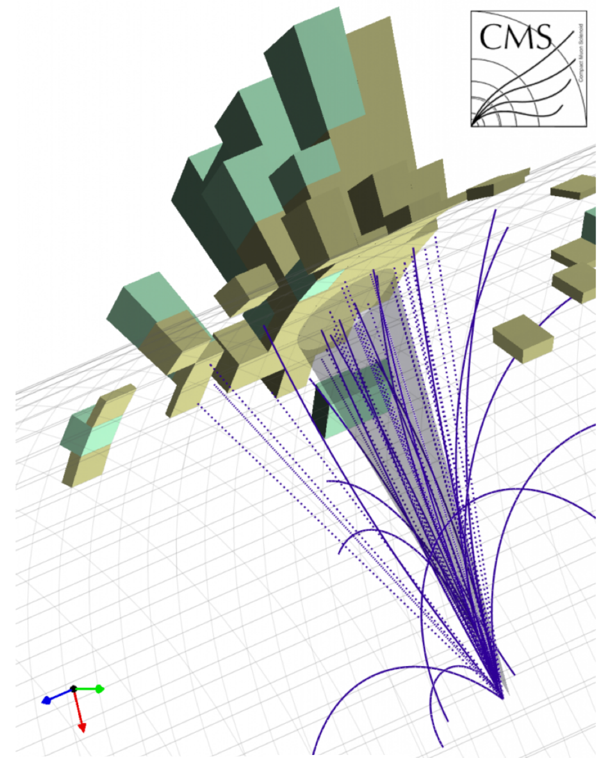
Event Reconstruction and Selection

Hypothesis testing / Measurement

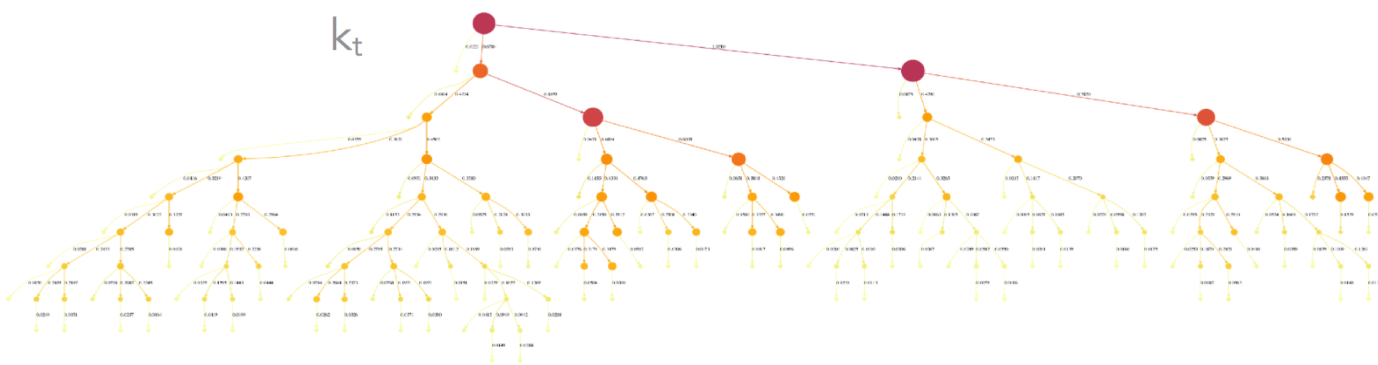
Latent parameter
estimation

Jets at the LHC

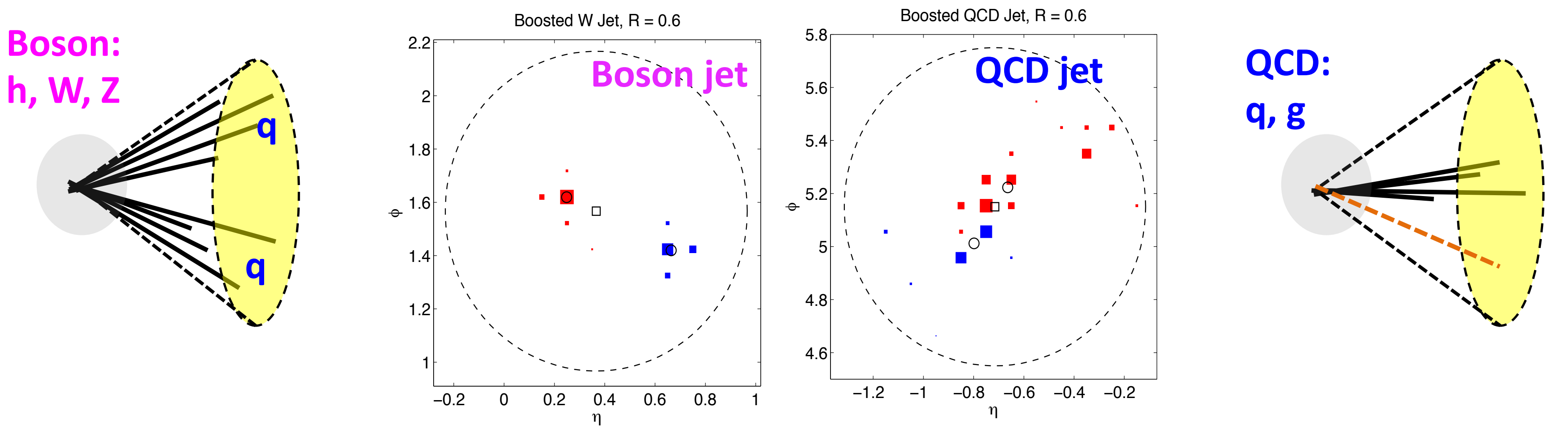




k_t



Machine Learning and Jet Physics



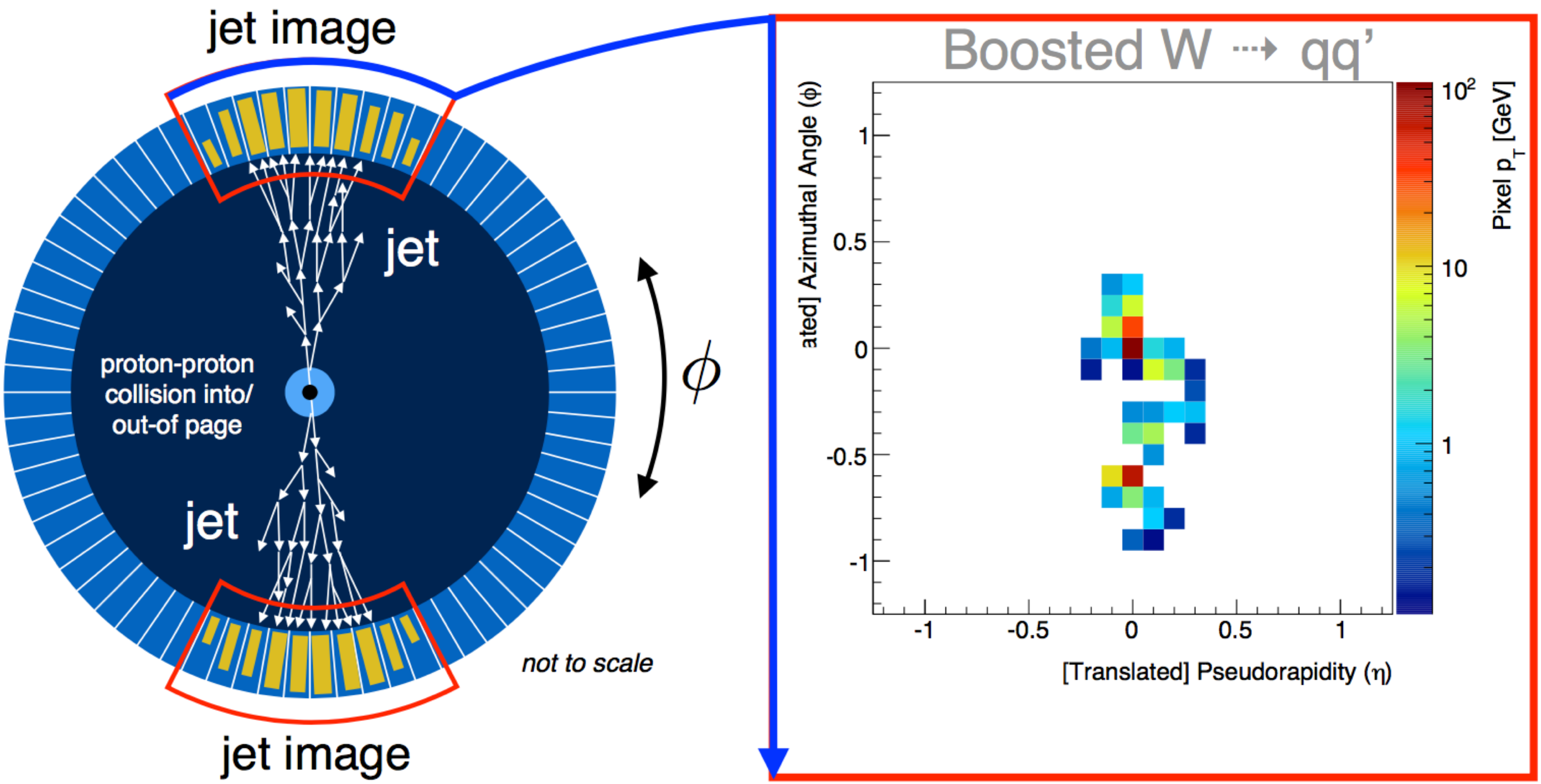
- Can use internal structure of a jet for classification
  - Also known as Jet Substructure

- A wealth of domain expertise has gone in feature engineering

- Can deep learning perform this classification?

Jet Images

Unrolled slice of detector



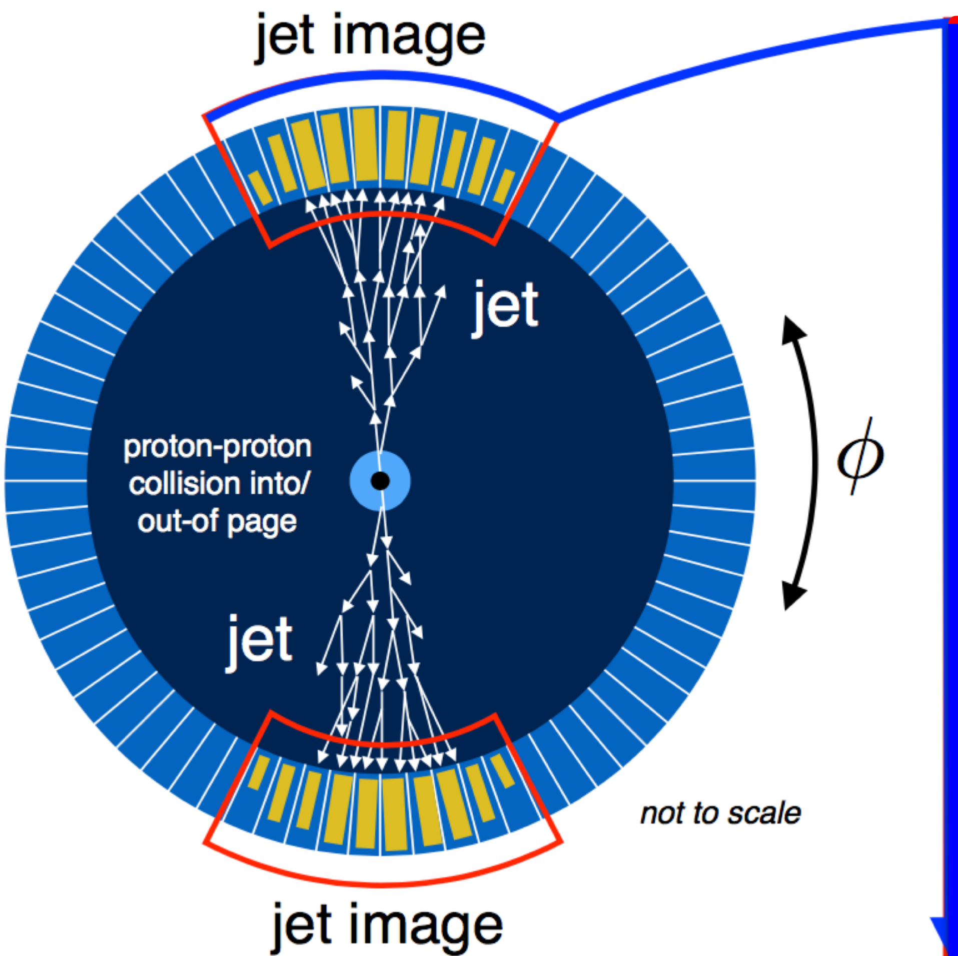
Calorimeter towers as pixels

Energy depositions as intensity

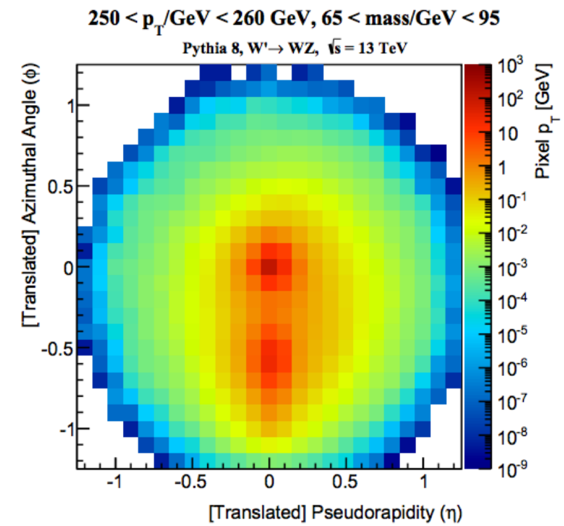
B. Nachman:
https://indico.cern.ch/event/567550/contributions/2656471/

Jet Images

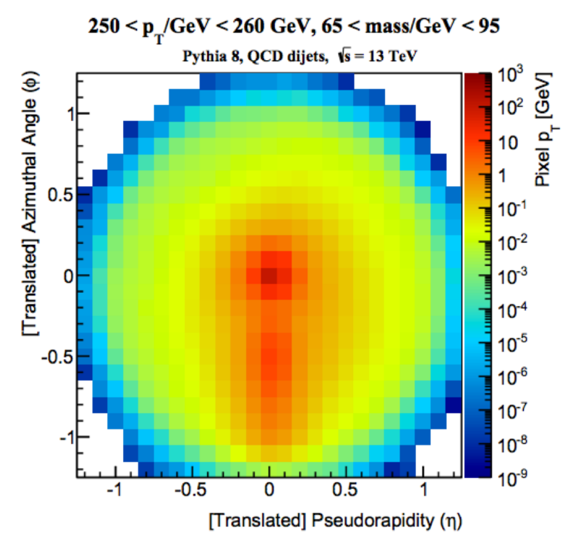
Average of large number of Jet Images



W-jets

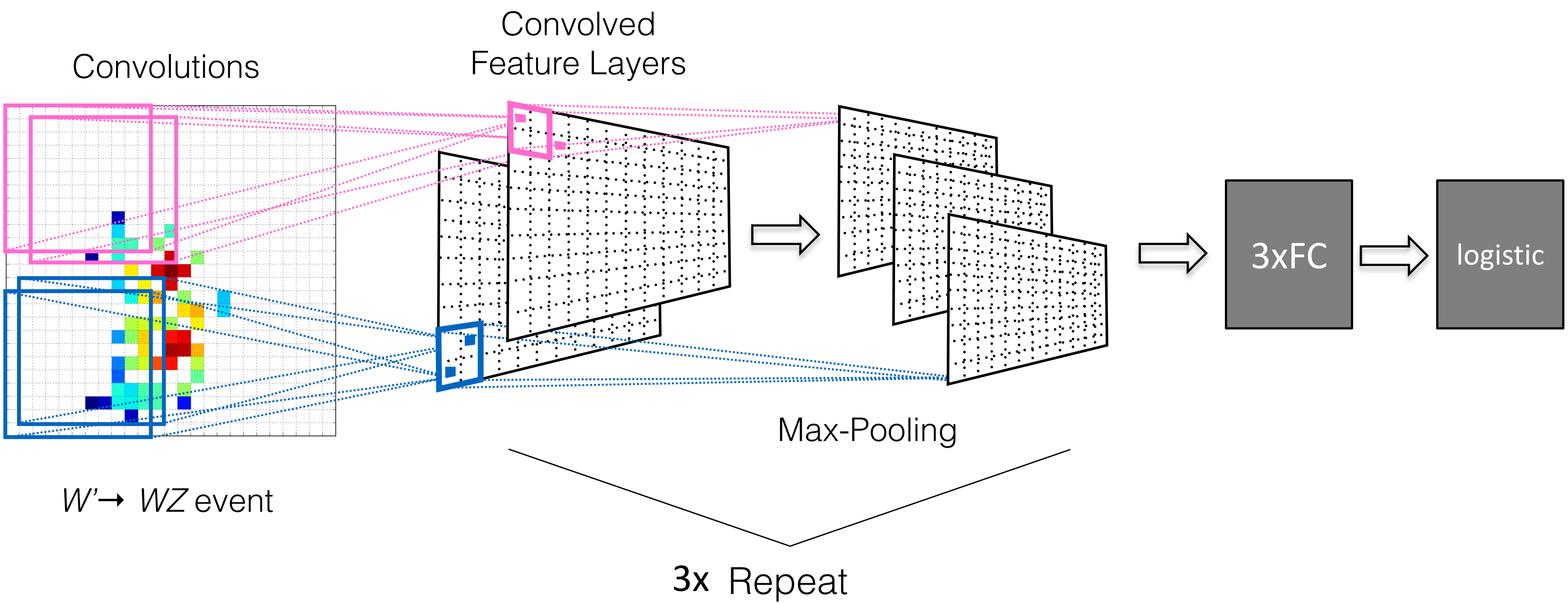


QCD-jets



B. Nachman:
https://indico.cern.ch/event/567550/contributions/2656471/

Deep Jets – Convolutional Neural Networks



<u>Jet-image based developments</u>

P. Baldi et al. 1603.09349 (W-tagging)

J. Barnard et al. 1609.00607 (W-tagging)

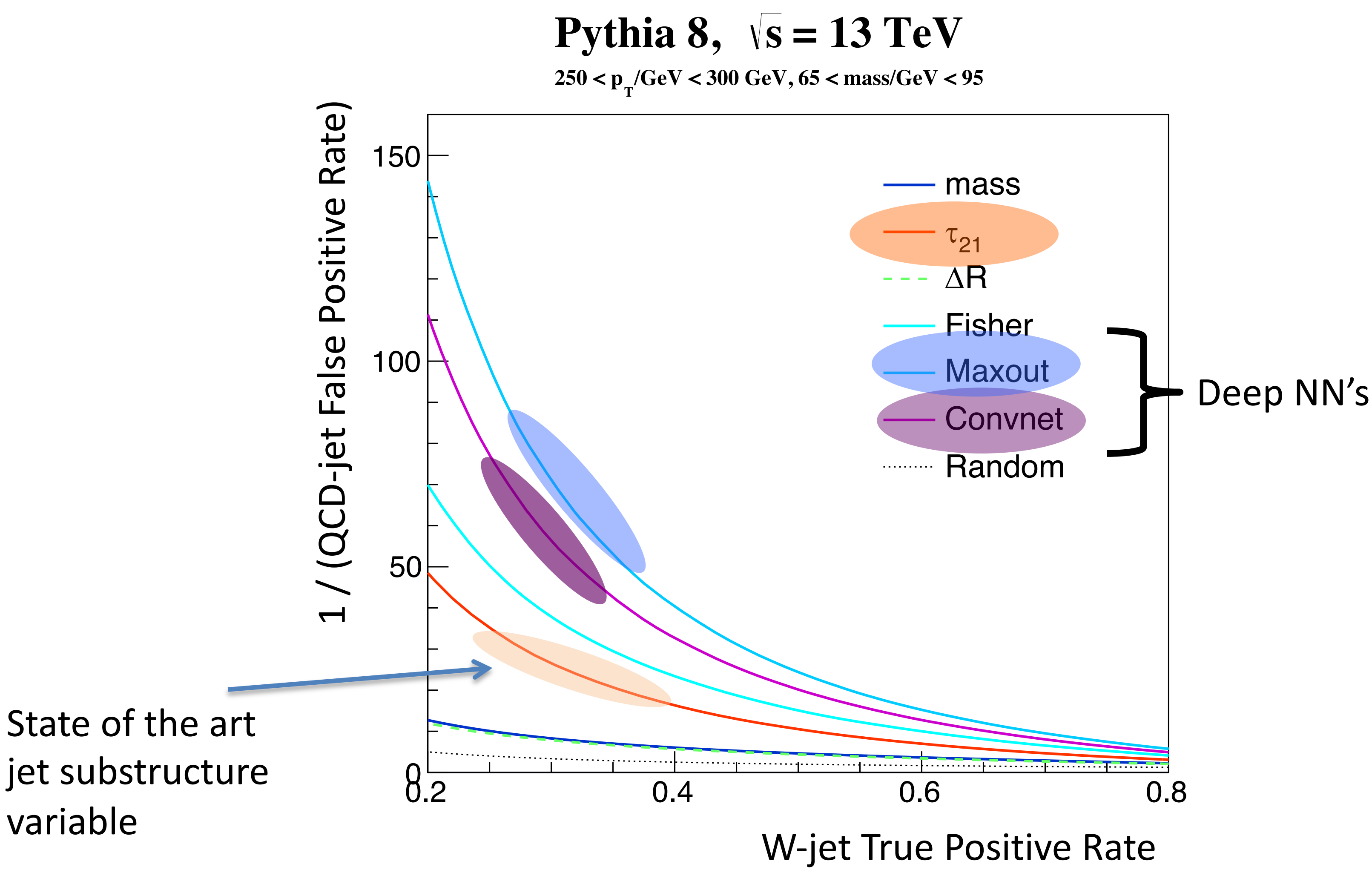
P. Komiske et al. 1612.01551 (q/g-tagging)

L. de Oliveira et. al. 1701.05927 (jet-image GAN)

G. Kasieczka et al. 1701.08784 (top-tagging)

arXiv:1511.05190

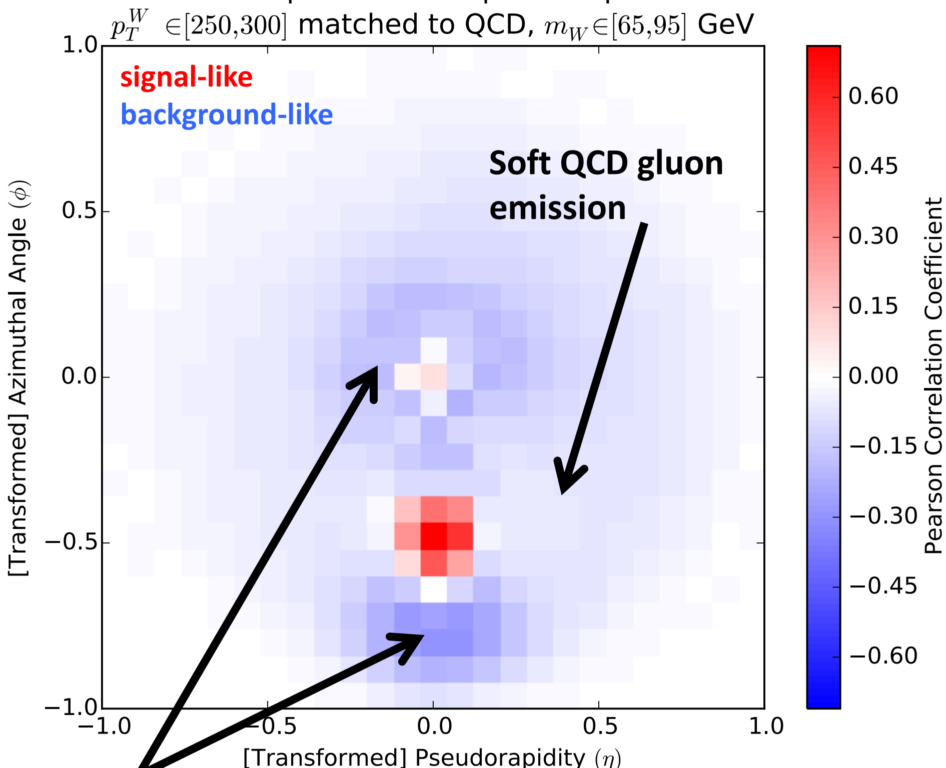
Performance with Deep Neural Networks



Learning from the Machines

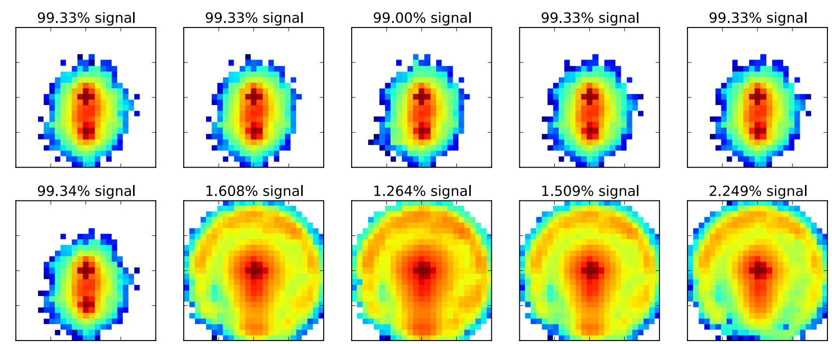
arXiv:1511:05190

Correlation of Deep Network output with pixel activations.

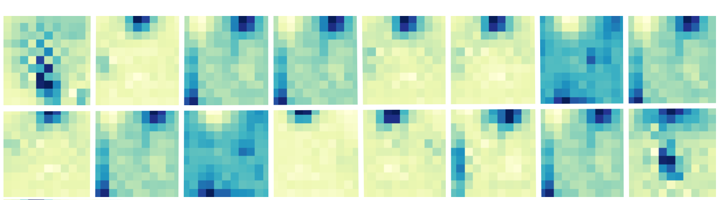


Additional radiation in QCD jets

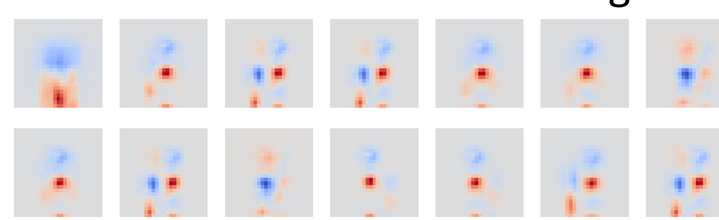
Average of most activating jets for a given neural



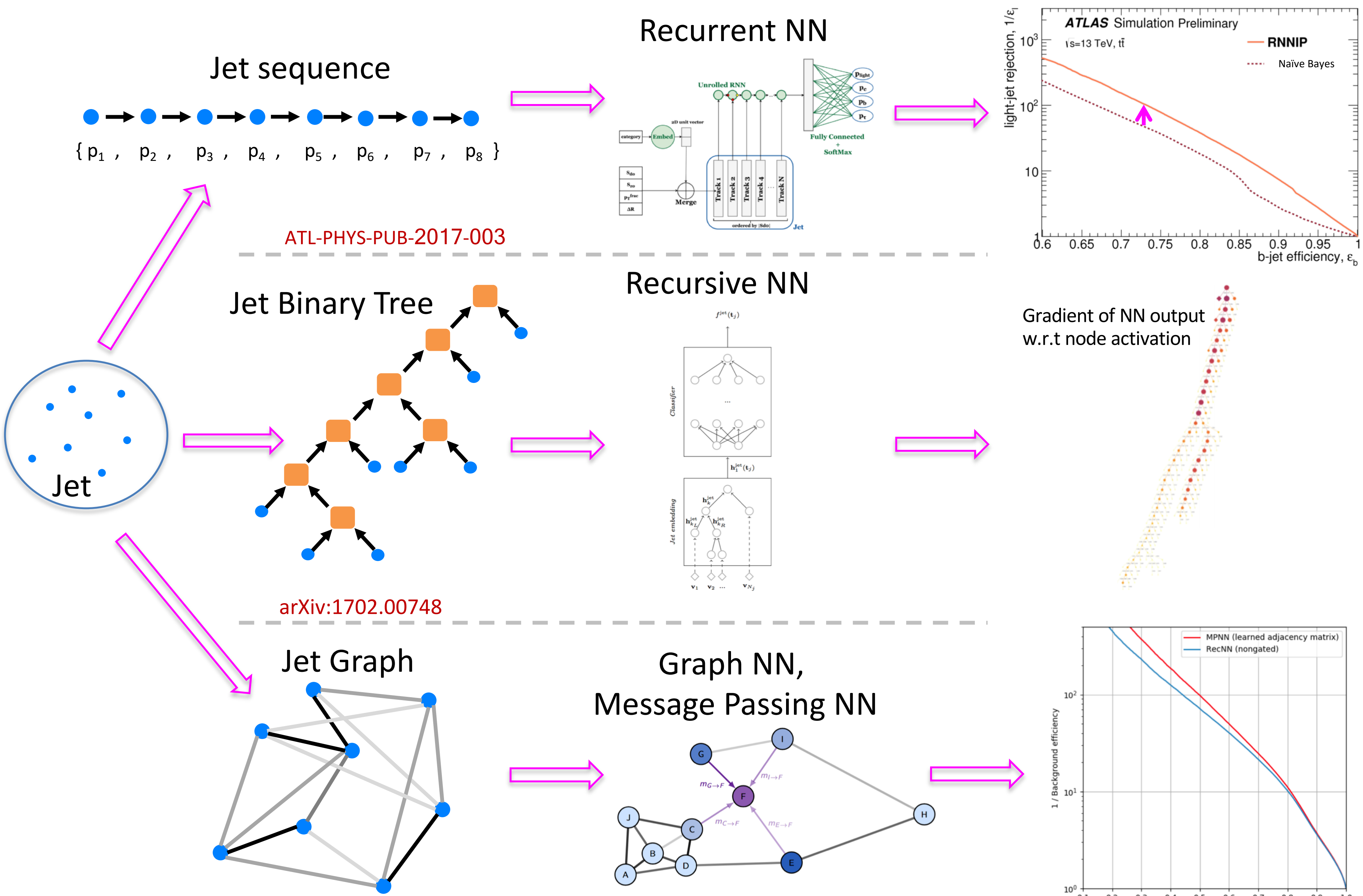
Filters



Filters convolved with images

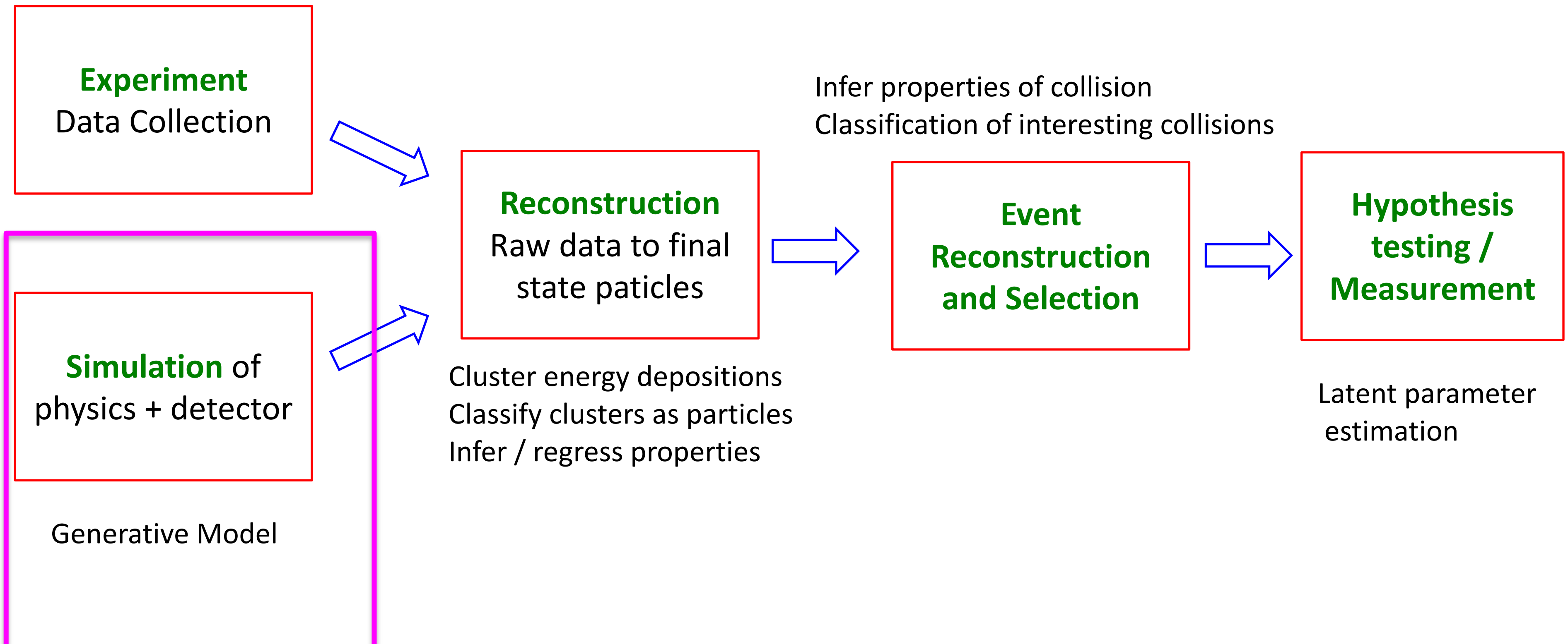


Beyond Images: New representations, models, and applications for deep learning in jet physics

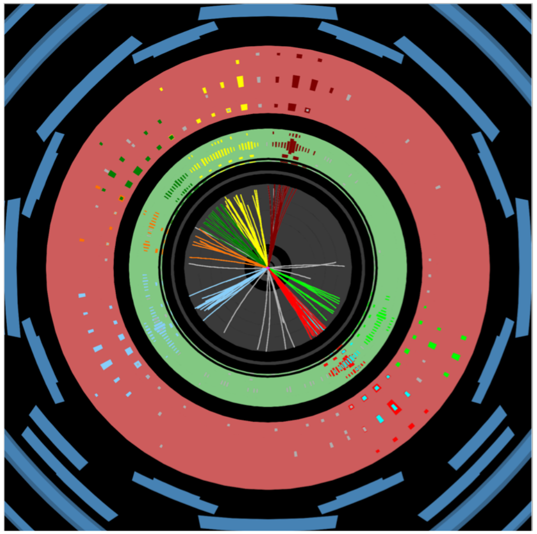


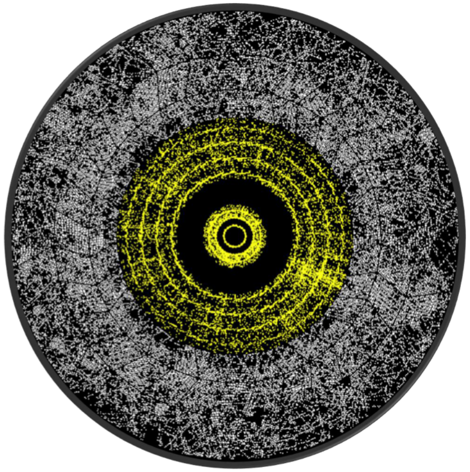
I. Henrion et. al. , presented at NIPS workshop on deep learning for physics sciences

Deep Learning in the Pipeline



Generative Adversarial Networks (GAN)





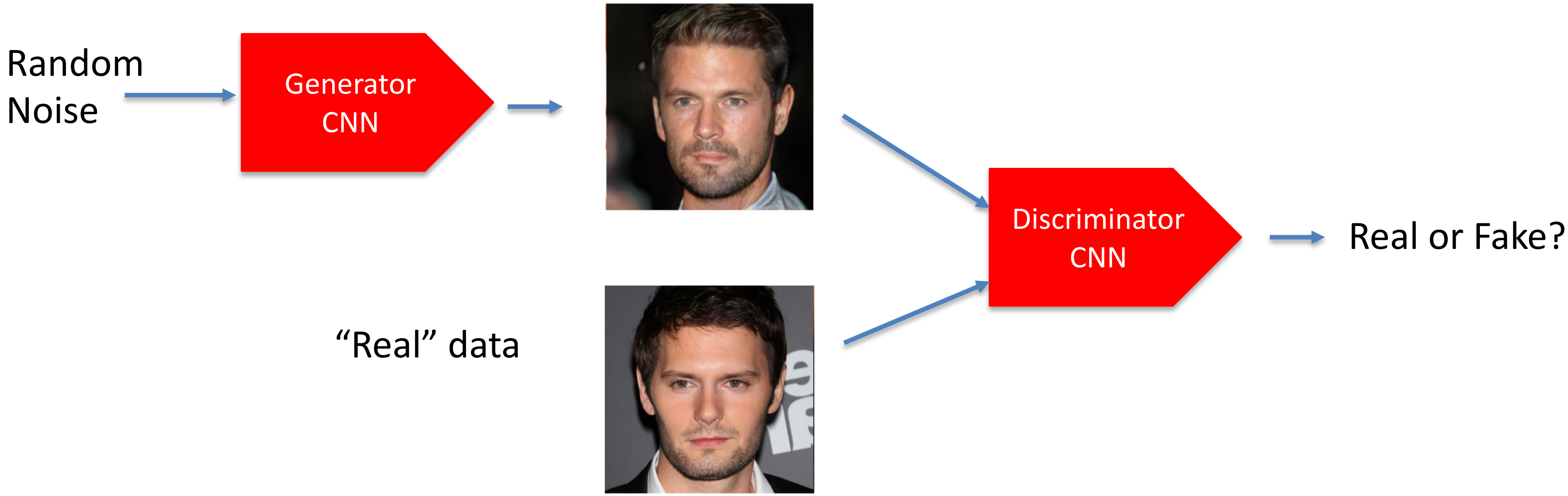
- Full Simulation: accurate simulation of particle interactions with material
  - Computationally very costly
  - Only produce sample, can't compute analytically P(energy deposits | particle)

- Fast Simulation: simplified parametric model of energy deposits

- Generative models to learn data distribution, p(x), and produce samples?
  - Generative Adversarial Networks (GAN)
  - Variational Auto-Encoders (VAE)

GANs

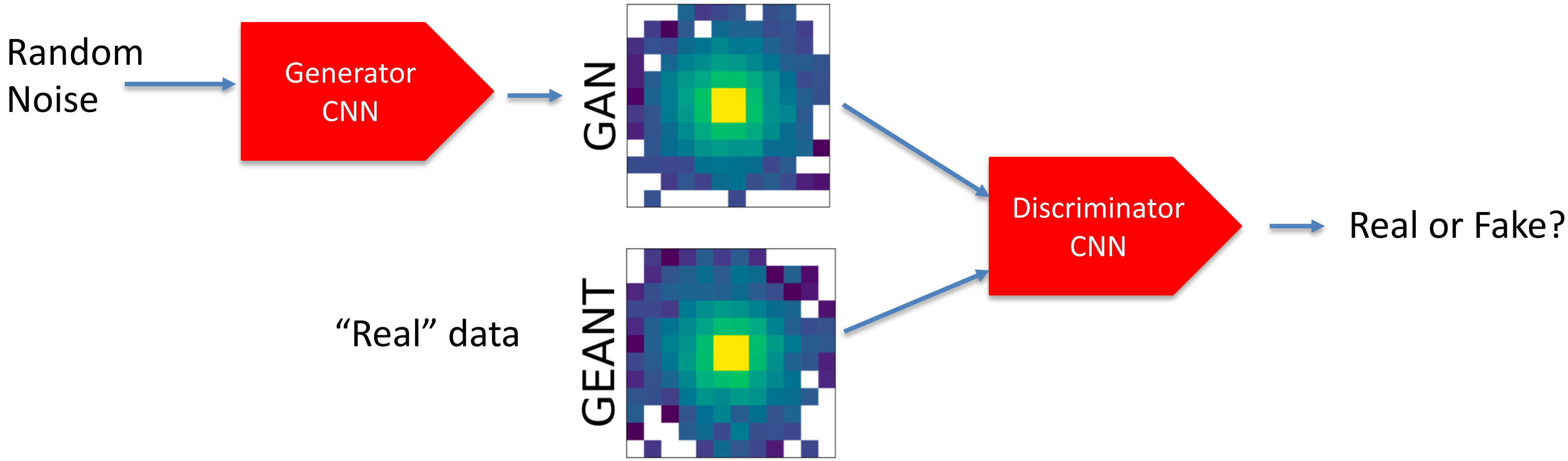
arXiv:1406.2661



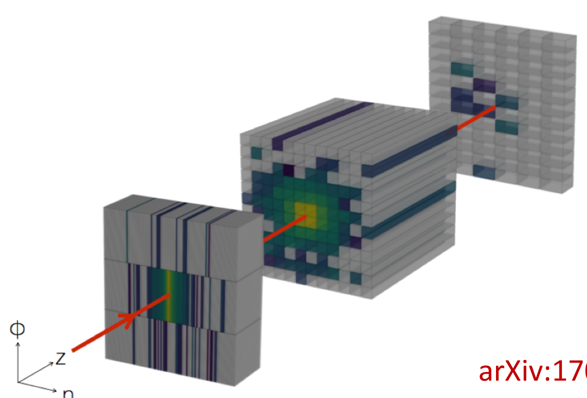
- Generator produces images from random noise and tries to trick discriminator into thinking they are real
- Classifier tries to tell the difference between real and fake images

Images: arXiv:1710.10196

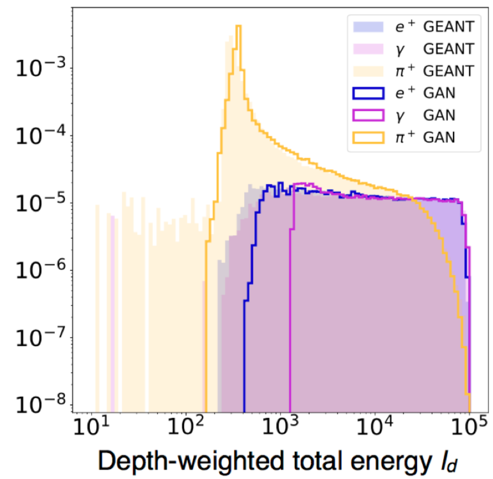
GANs for HEP

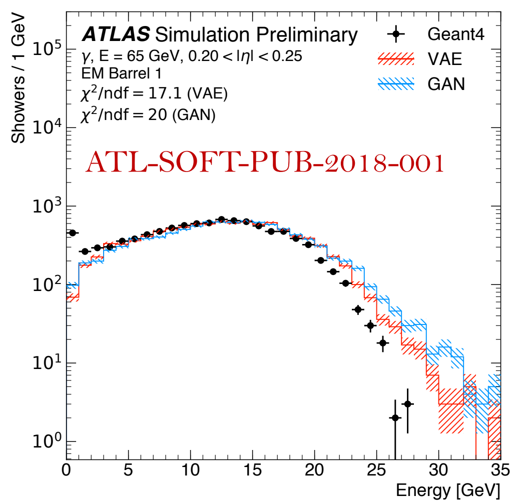


- GANs and VAEs being studied for generating Jet-images, and 3D calorimeter energy depositions in toy simulation and at the LHC experiments!

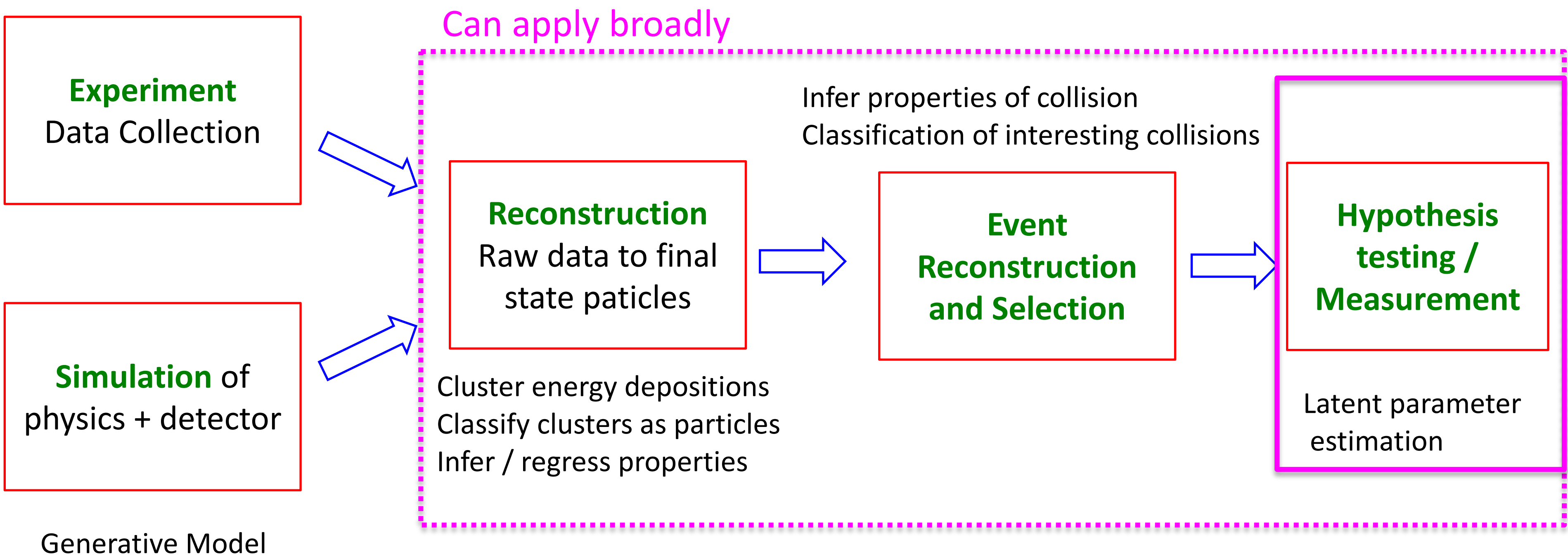


arXiv:1705.02355
arXiv:1701.05927





Deep Learning in the Pipeline

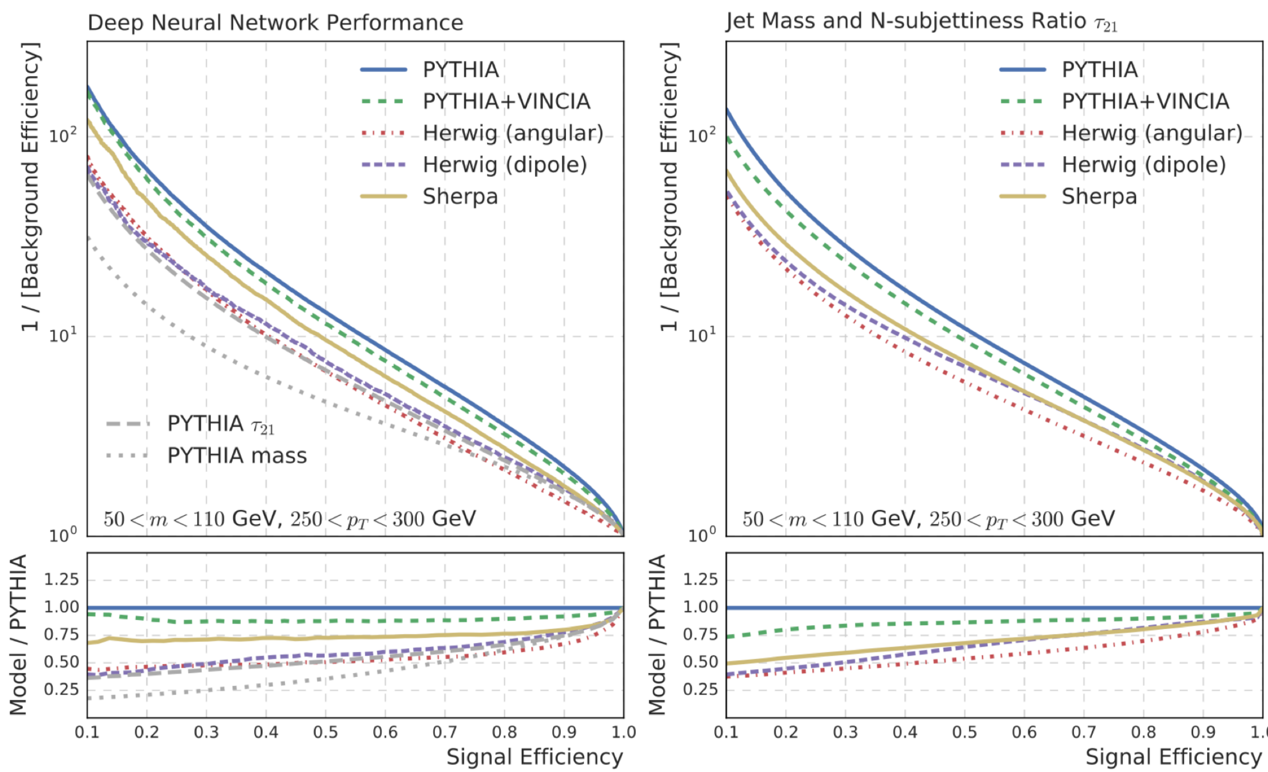


- Developing new ways to train our algorithms. Examples:
  - Parametrized learning [EPJ C76 (2016) no.5, 235]
  - **Adversarial learning to pivot** [NIPS 2017, 1611.01046]
  - Learning from label proportions [1702.00414]

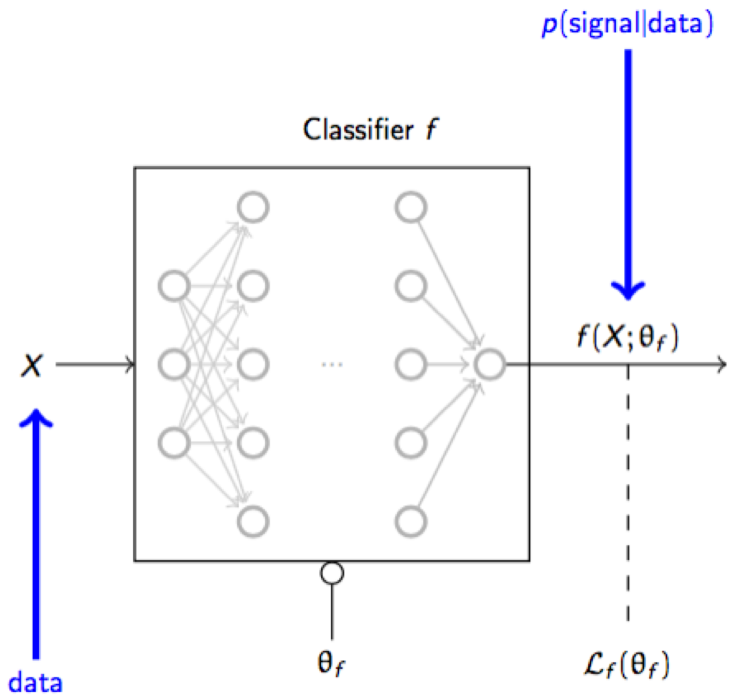
Dealing with Systematic Uncertainties

- Systematic uncertainties encapsulate our incomplete knowledge of physical processes and detectors

- Can we teach a classifier to be robust to these kinds of uncertainties?

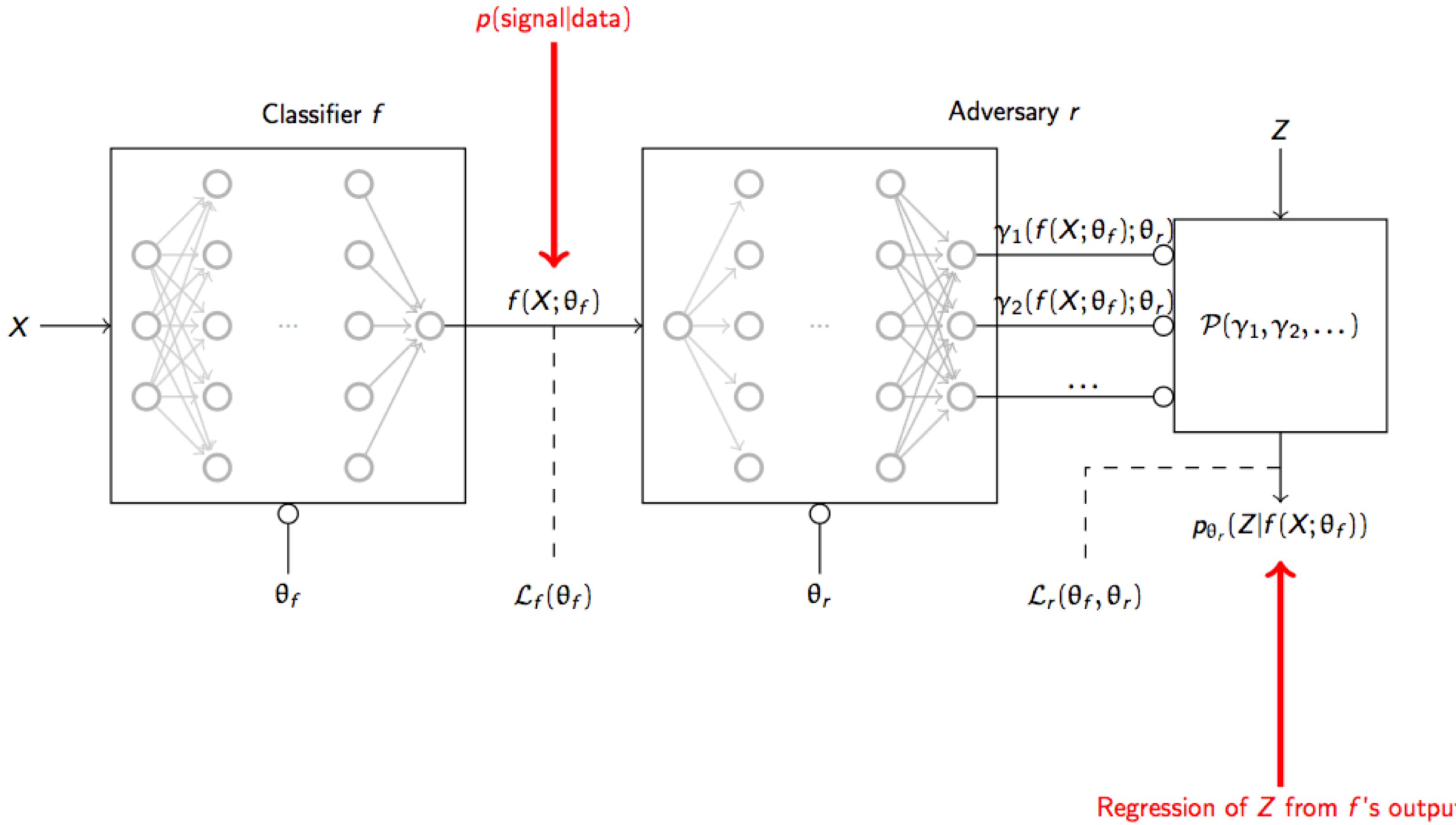


arXiv:1609.00607



- Classifier built to solve problem at hand

Adversarial Networks



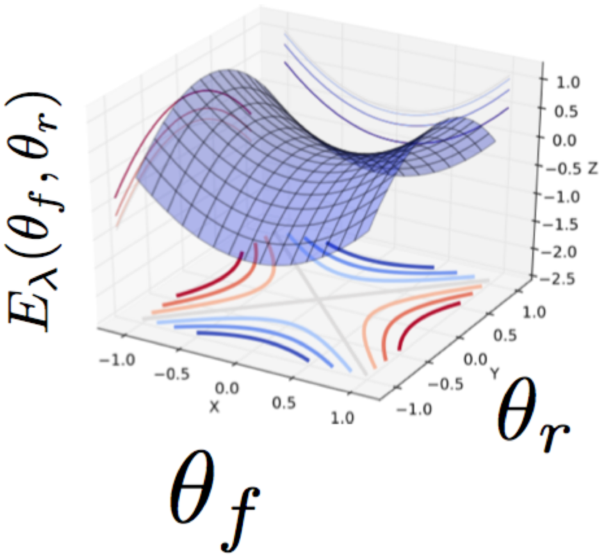
- Systematic uncertainty encoded as nuisance parameters, Z
- Adversary to predict the value of Z given classifier output

Adversarial Networks

G. Louppe, M. K., K. Cranmer, arXiv:1611.01046

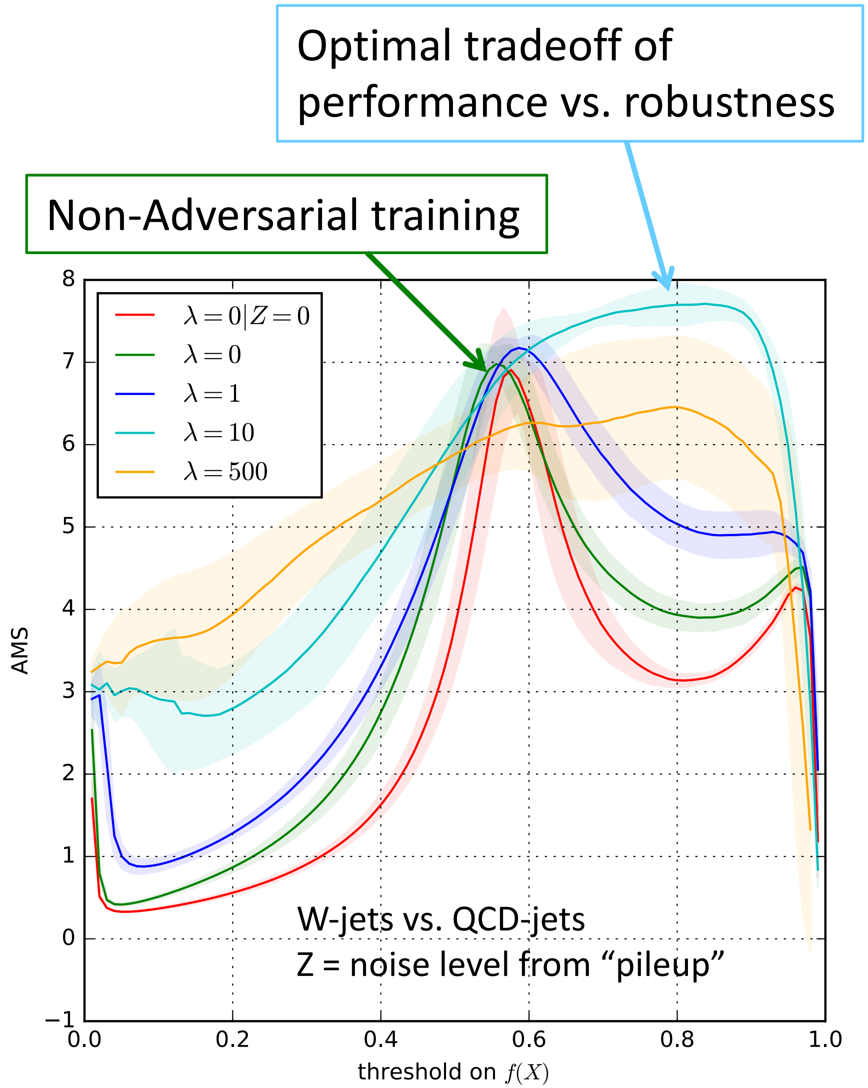
$$\hat{\theta}_f, \hat{\theta}_r = \arg \min_{\theta_f} \max_{\theta_r} E(\theta_f, \theta_r).$$

$$E_\lambda(\theta_f, \theta_r) = \mathcal{L}_f(\theta_f) - \lambda \mathcal{L}_r(\theta_f, \theta_r),$$



- Loss encodes performance of classifier and adversary
  - Classifier penalized when adversary does well at predicting Z

- Hyper-parameter λ controls trade-off
  - Large λ enforces f(…) to be pivotal, e.g. robust to nuisance
  - Small λ allows f(…) to be more optimal



[AMS = Estimate of statistical significance including systematic uncertainty]

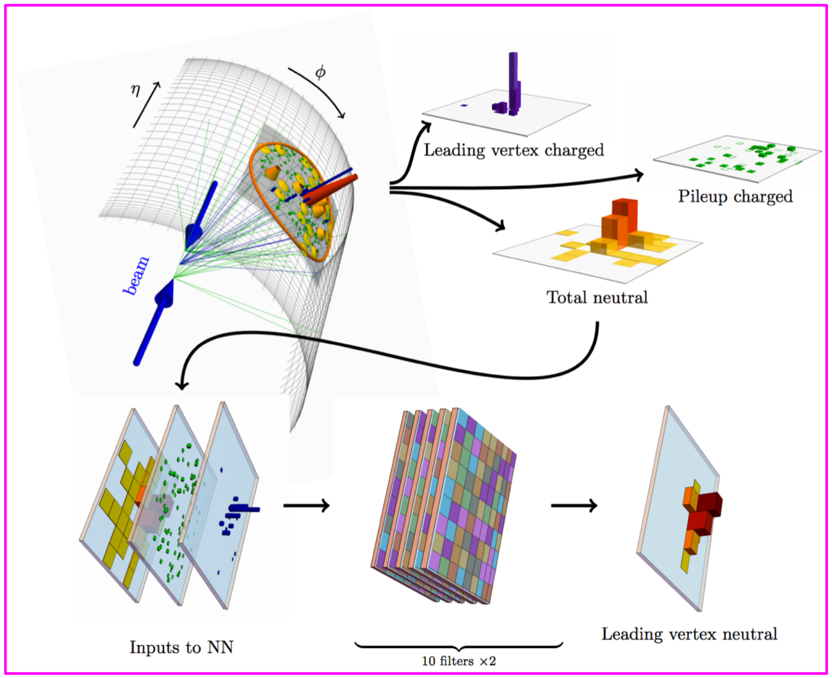
- $\lambda=0$, $Z=0$
  - Standard training with no systematics during training, evaluate systematics after training
- $\lambda=0$
  - Training samples include events with systematic variations, but no adversary used
- $\lambda=10$
  - Trading accuracy for robustness results in net gain in terms of statistical significance

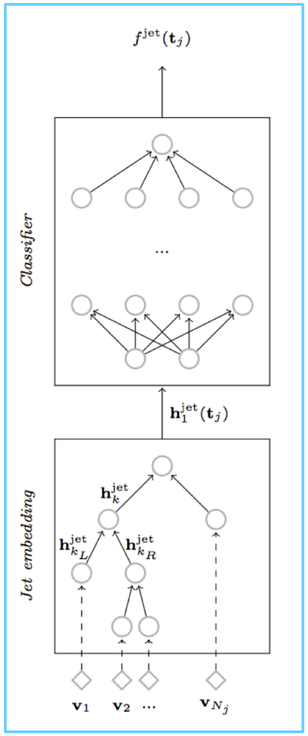
Directions in DL in HEP

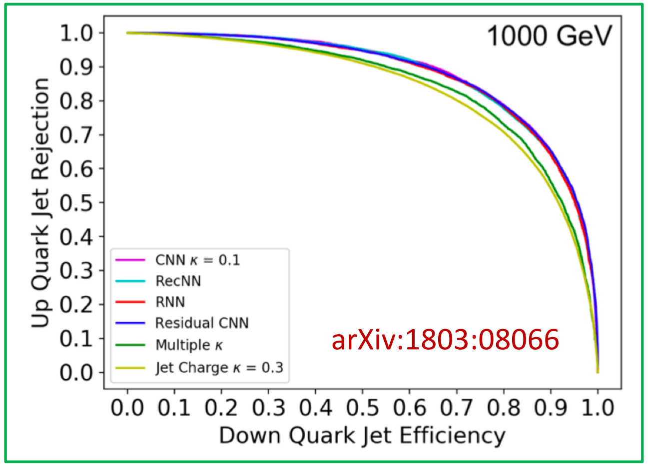
Directions in DL in HEP

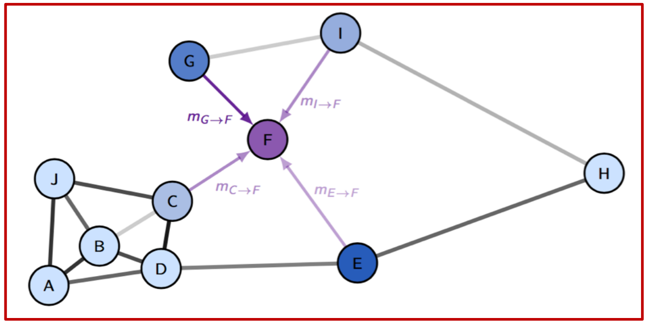
- Better understanding how to use computer vision and natural language processing techniques

- Thinking about new data structures, like trees and graphs, that can be analyzed with Deep Learning









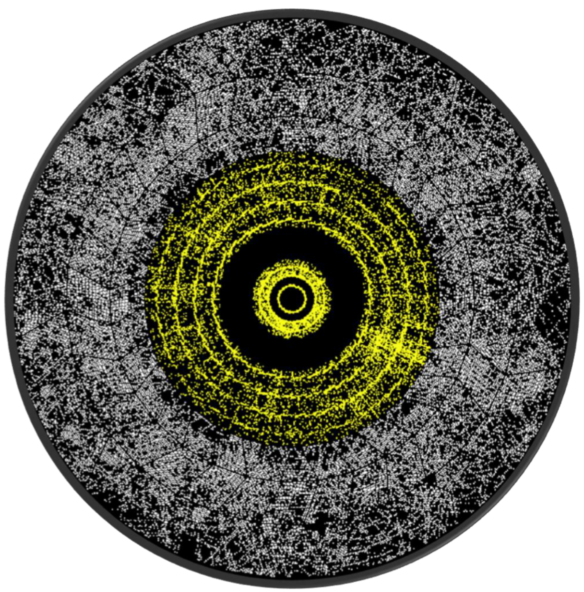
arXiv:1707.08600

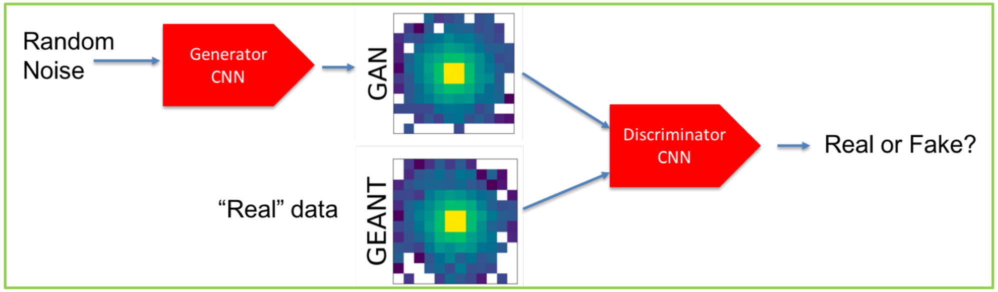
arXiv:1702.00748

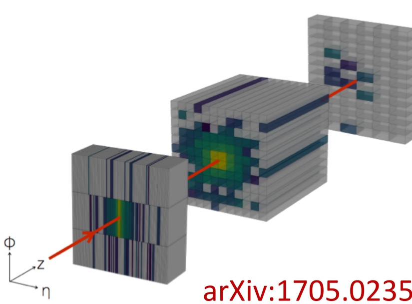
I. Henrion et. al. , presented at NIPS workshop on deep learning for physics sciences

Directions in DL in HEP

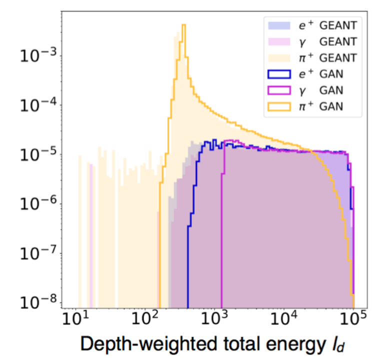
- Better understanding how to use computer vision and natural language processing techniques
- Thinking about new data structures, like trees and graphs, that can be analyzed with Deep Learning
- Can ML help with our most computationally costly problems, like simulation or the combinatorial challenge of tracking?

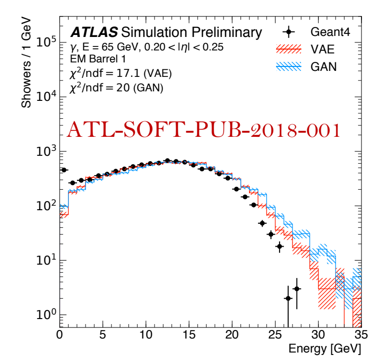






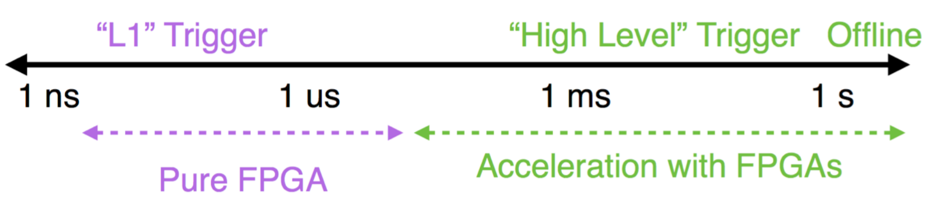
arXiv:1705.02355
arXiv:1701.05927





Directions in DL in HEP

- Better understanding how to use computer vision and natural language processing techniques
- Thinking about new data structures, like trees and graphs, that can be analyzed with Deep Learning
- Can ML help with our most computationally costly problems, like simulation or the combinatorial challenge of tracking?
- Can fast O(ns-μs) NN inference be done with FPGAs to put ML early in the trigger / data acquisition process?



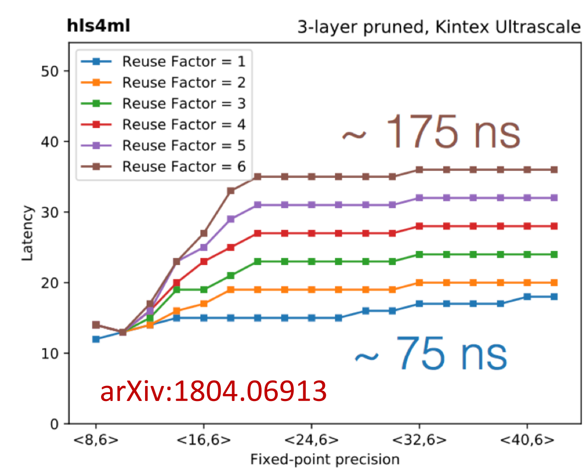


Image from:
https://indico.cern.ch/event/714134/contributions/2960185/attachments/1640629/2620365/20180426_hls4ml_kreis.pdf

Directions in DL in HEP

- Better understanding how to use computer vision and natural language processing techniques

- Thinking about new data structures, like trees and graphs, that can be analyzed with Deep Learning

- Can ML help with our most computationally costly problems, like simulation or the combinatorial challenge of tracking?

- Can fast O(ns-μs) NN inference be done with FPGAs to put ML early in the trigger / data acquisition process?

- Can we design better architectures and training algorithms to tackle our HEP challenges?

- How can we make best use of our simulation for inference without the PDF, i.e. Likelihood Free Inference?

Conclusion

- Just touched the surface of the rapid progress in Machine Learning in HEP

- Deep learning application developing quickly in High Energy Physics, across the whole data acquisition, simulations, and analysis pipeline

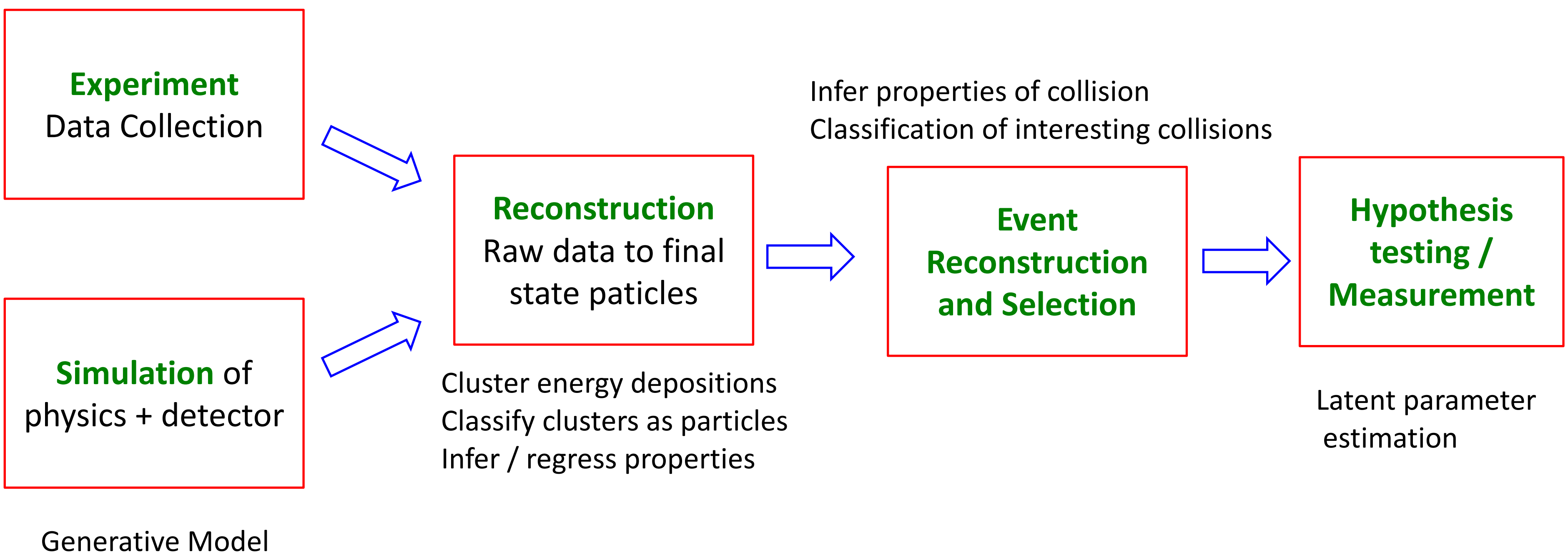
- Many new developments and performance improvements driven by thinking about HEP challenges in completely new ways

Acknowledgement

- H2020-Astronomy ESFRI and Research Infrastructure Cluster (Grant Agreement number: 653477).

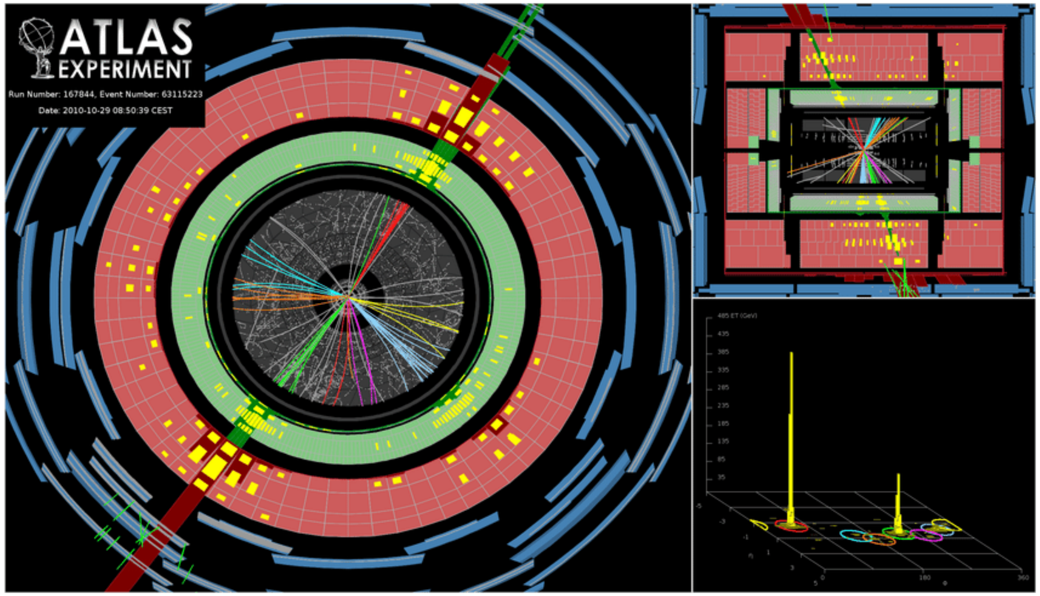
backup

Pipeline



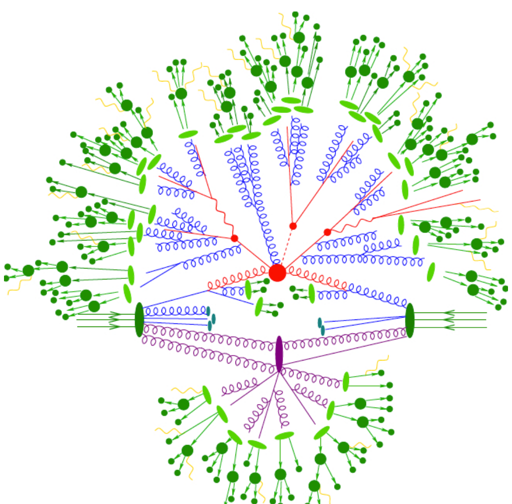
- Build on our knowledge of how the data is created
  - Use our simulation to design and study reconstruction algorithms, and to compare predictions with our experimental data
- Use Machine learning to improve (or rethink) the steps of this process?

Simulation

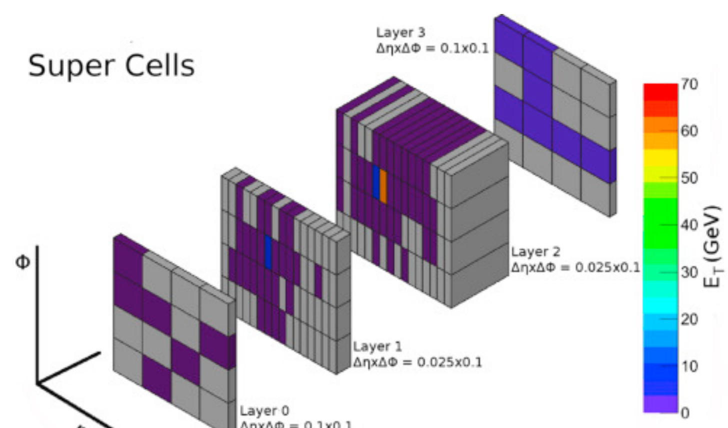


$p(particles \mid interation\ type)$

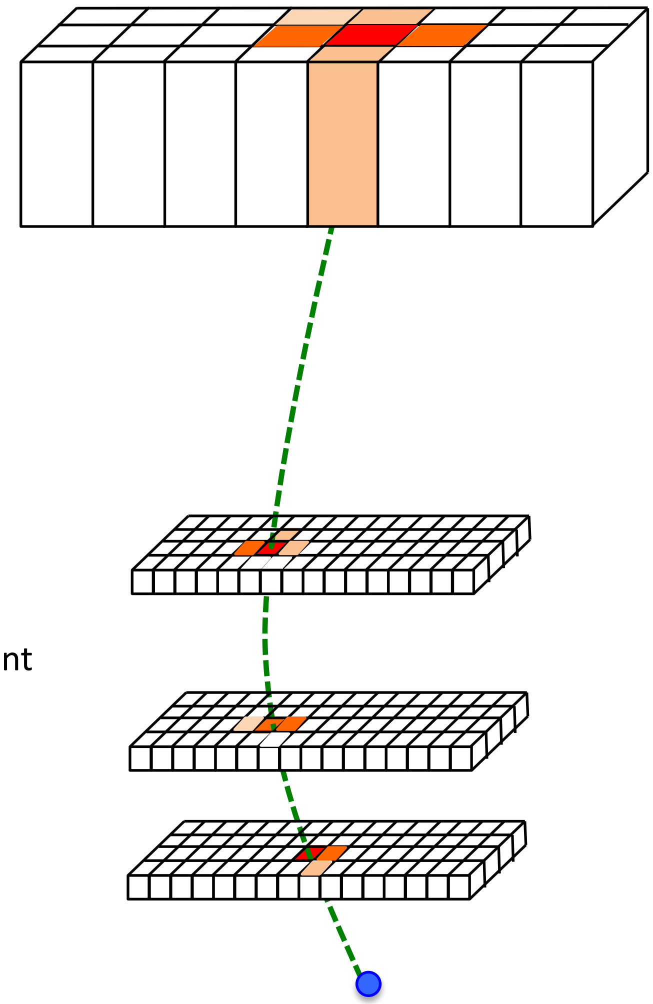
$p(detector\ signature \mid particle)$



[Sherpa]



Reconstructing Particles



Calorimeter:
Stops particle and destructively measure energy / direction

Tracking detector:
Typically Si-pixel detector
Non-destructive space-point measurement

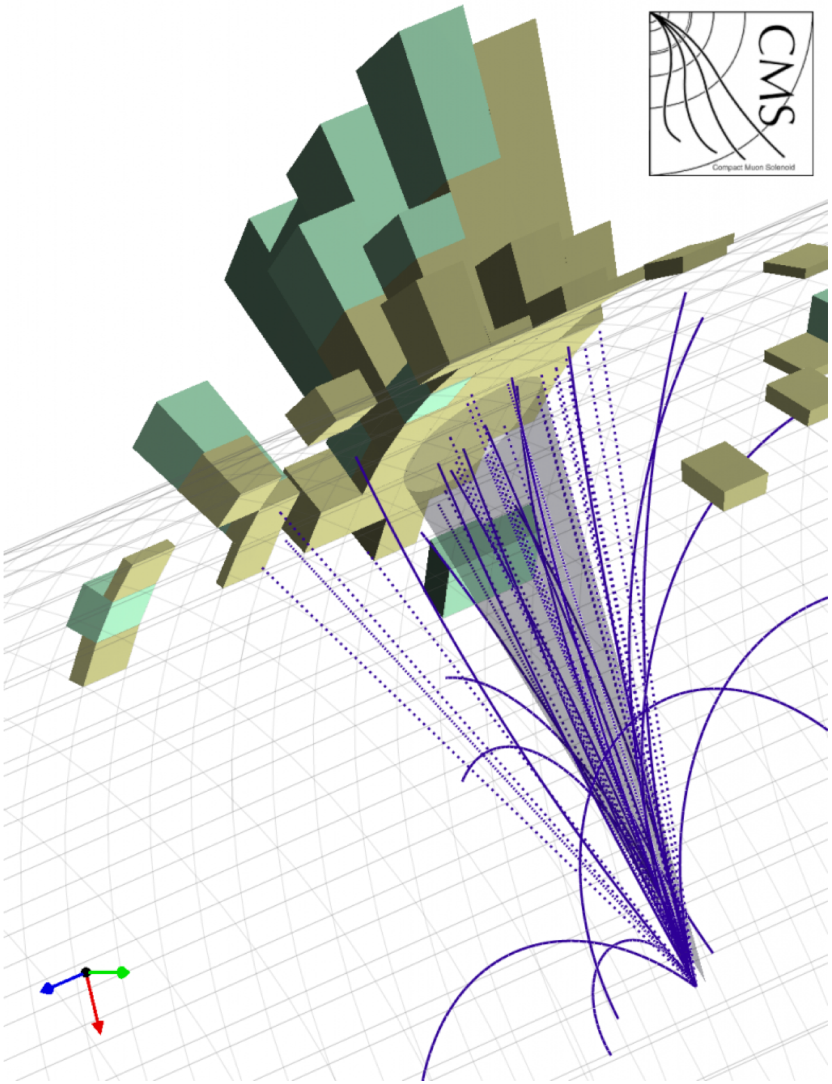
- Particle identification = Classification

$$p(\text{electron} \mid \text{data})$$

- Energy estimation = Inference, regression

$$p(E_{\text{true}}^{\text{electron}} \mid \text{electron data})$$

Understanding Clusters of Particles: Jets



- **Jet**: stream of particles produced by high energy quarks and gluons
  - Clustering algorithms used to find them

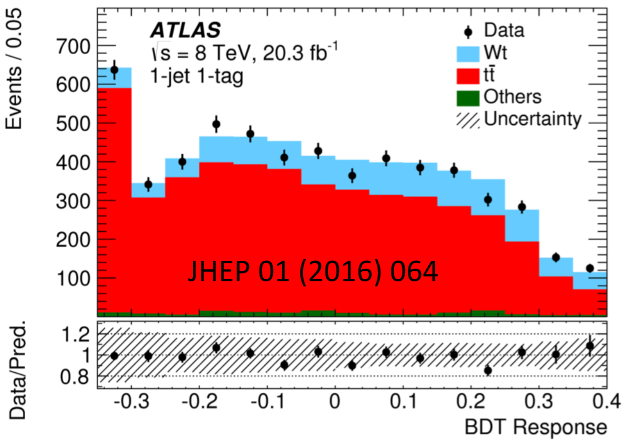
- Jet identification = Classification

$$p(parent\ particle \mid jet\ cluster)$$

- Energy estimation = Inference, regression

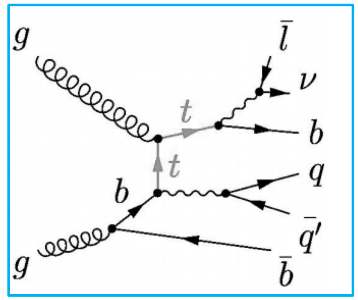
$$p(E_{true}^{jet} \mid jet\ cluster)$$

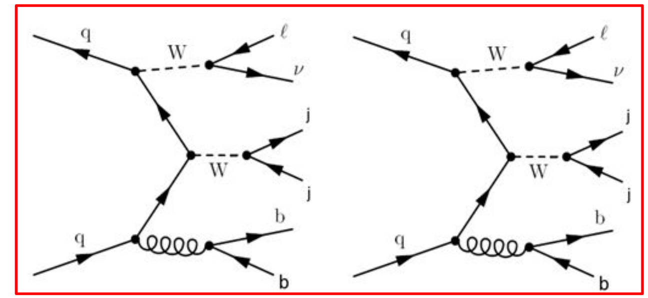
Analyzing Events and Hypothesis Testing



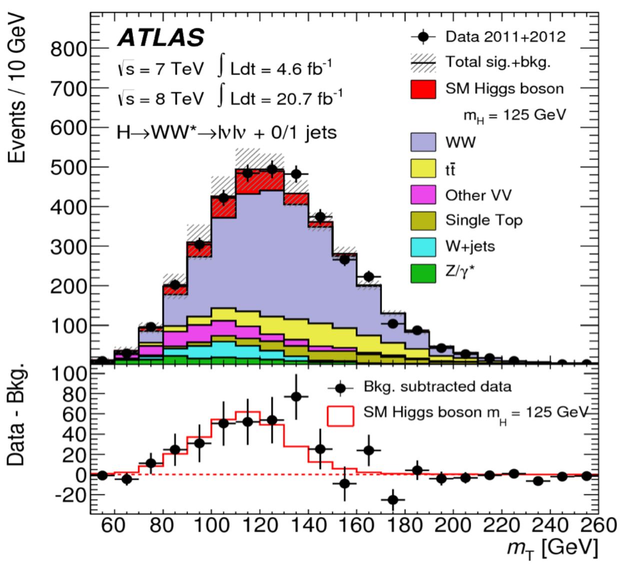
Analyzing Events

Signal vs background



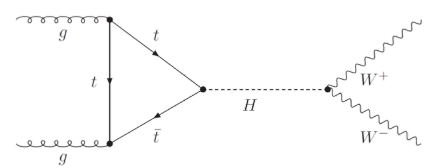


Hypothesis Testing and Parameter Estimation



Phys. Lett. B 726 (2013)

Is there a Higgs?
What is the Higgs mass?



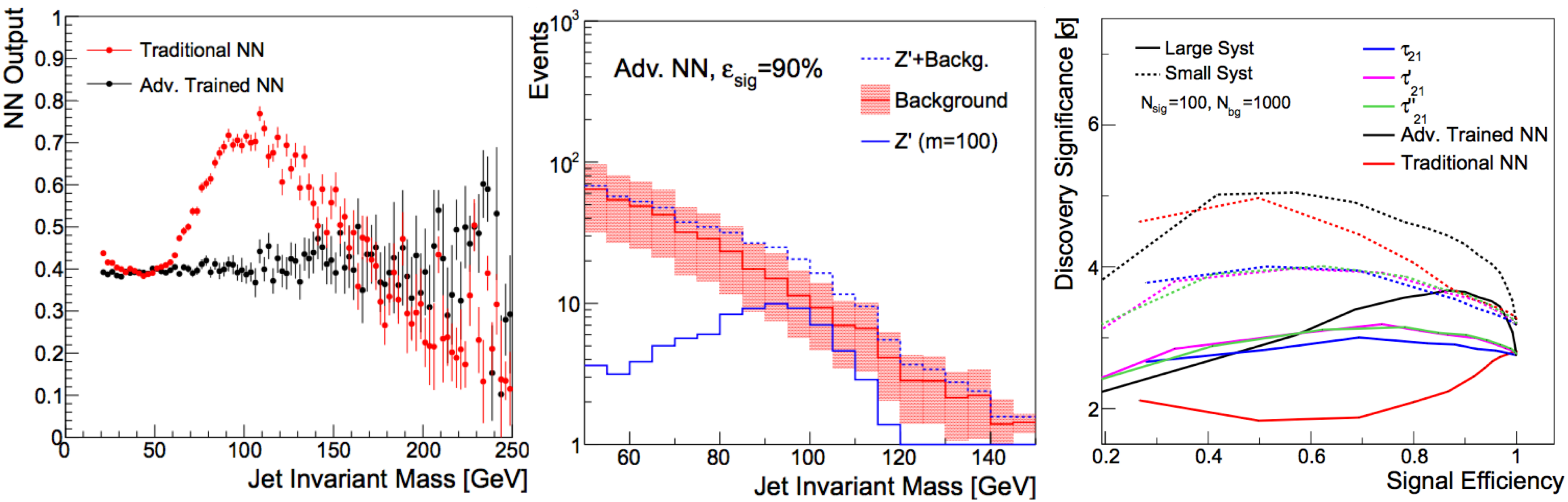
$$\lambda = \prod_{\mathbf{x}\in\mathcal{D}} \frac{p(\mathbf{x}|\text{background})}{p(\mathbf{x}|\text{signal+background})}$$

Decorrelating Variables

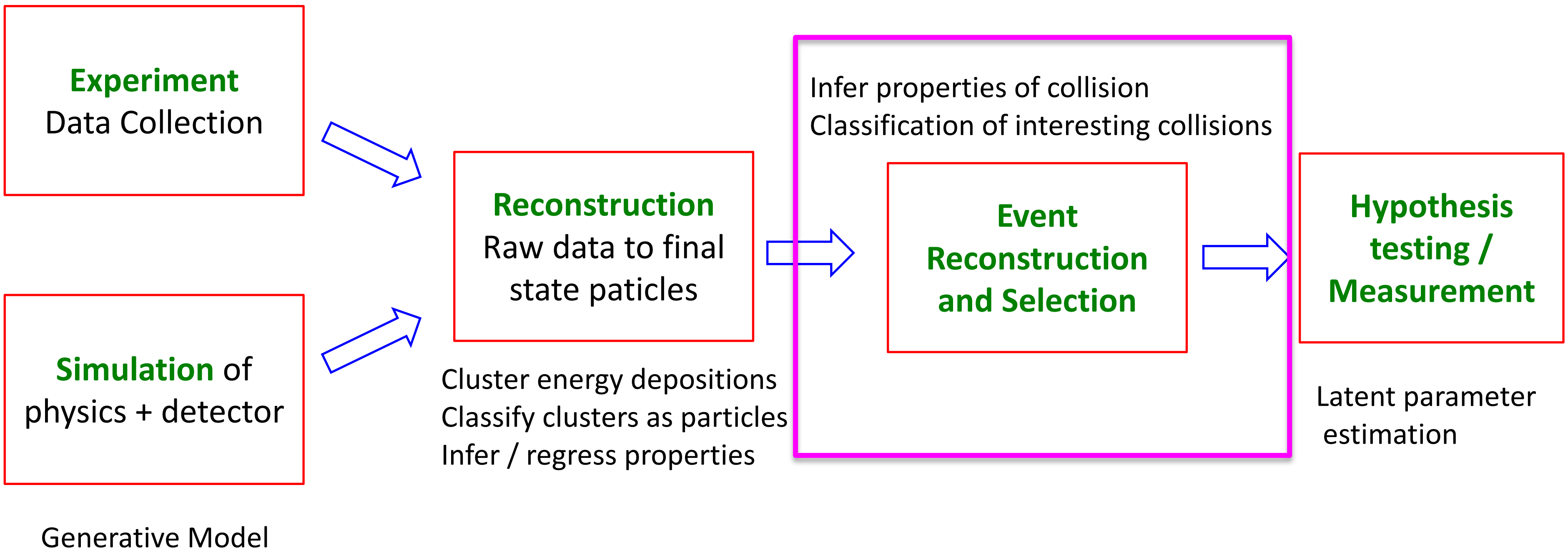
arXiv:1703.03507

- Same adversarial setup can decorrelate a classifier from a chosen variable (rather than nuisance parameter)

- In this example, decorrelate classifier from jet mass, so as not to sculpt jet mass distribution with classifier cut

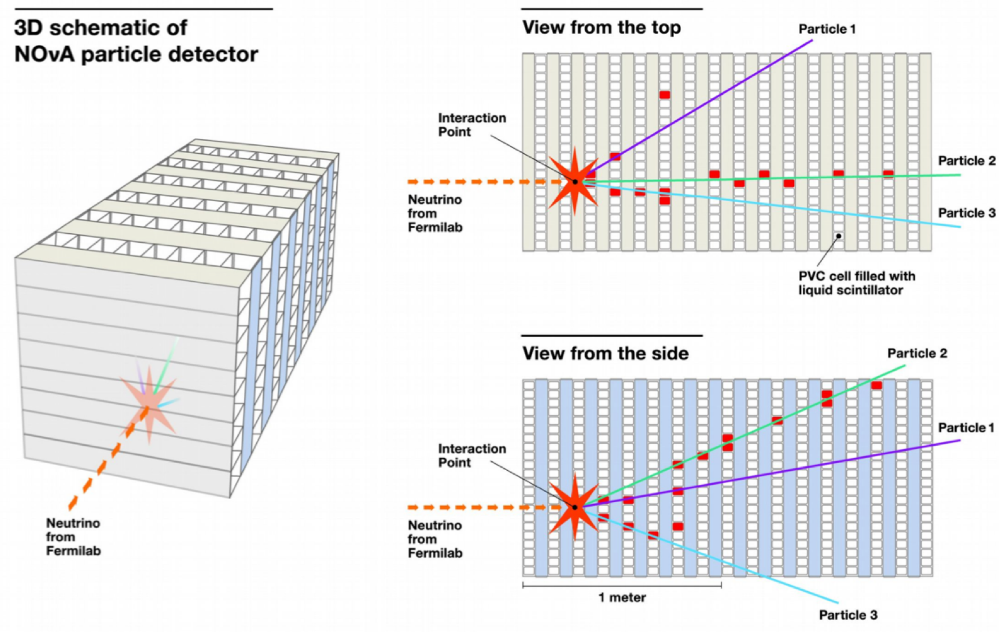


Deep Learning in the Pipeline

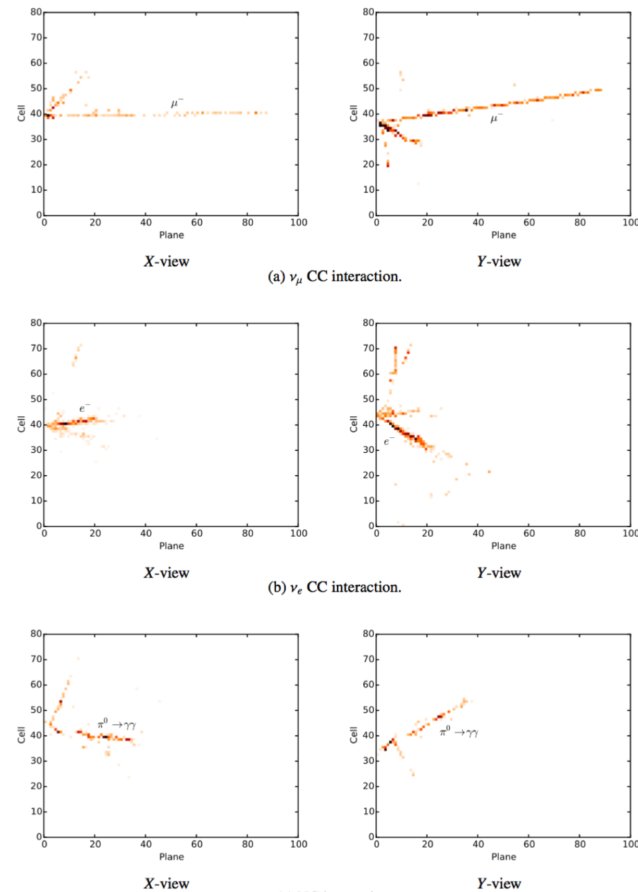


Neutrino Identification at NOvA

arXiv:1604.01444



- Two 2D projections of the interactions
- Goal: discriminate between different neutrino interactions / backgrounds



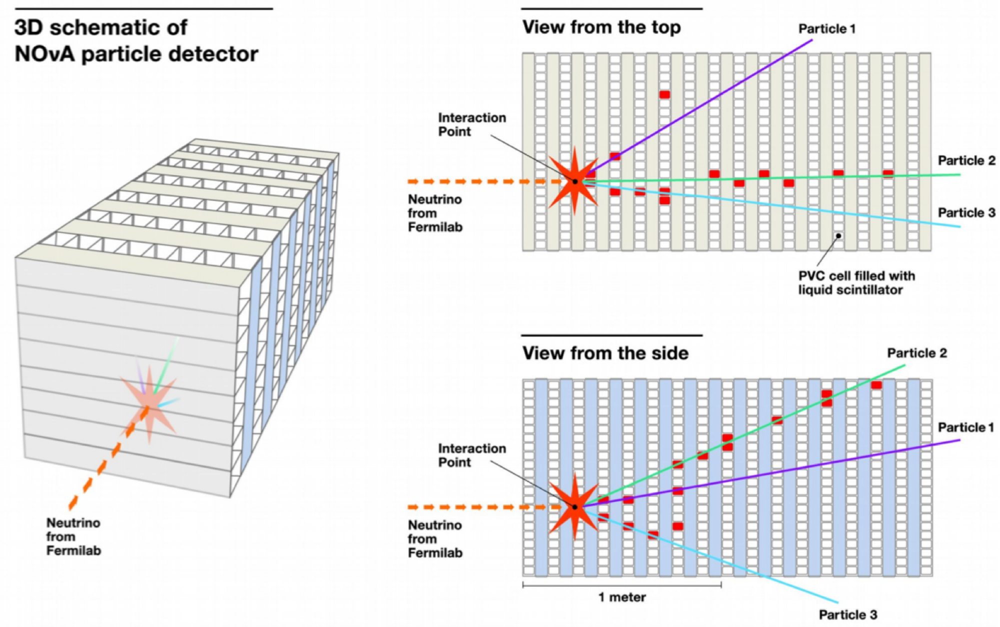
(a) ν_μ CC interaction.

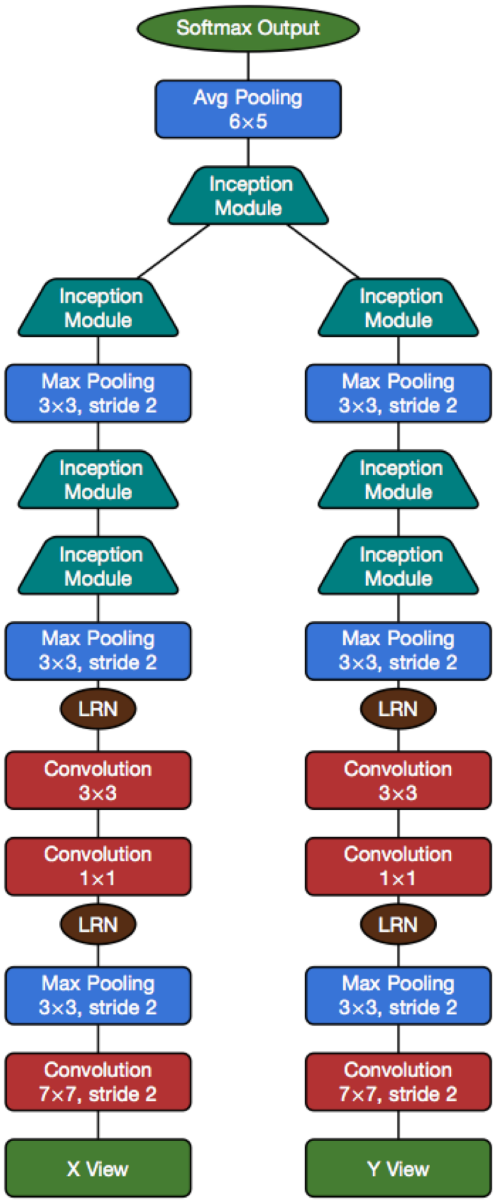
(b) ν_e CC interaction.

(c) NC interaction.

Neutrino Identification at NOvA

arXiv:1604.01444

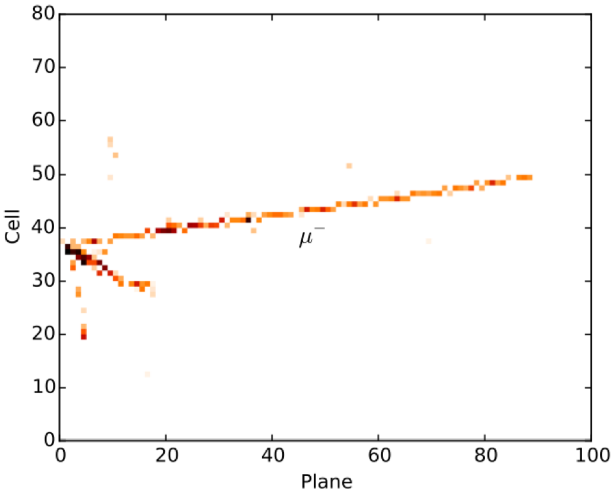


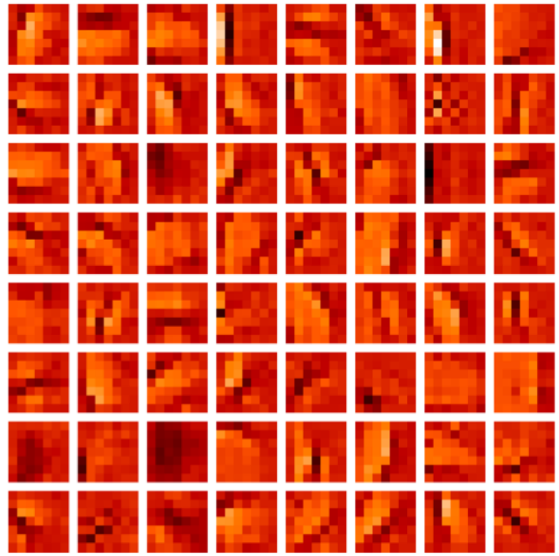


- Two 2D projections of the interactions
- Goal: discriminate between different neutrino interactions / backgrounds
- Make use of powerful computer vision architectures, here GoogLeNet, and adapt to our challenges

Neutrino Identification at NOvA

arXiv:1604.01444





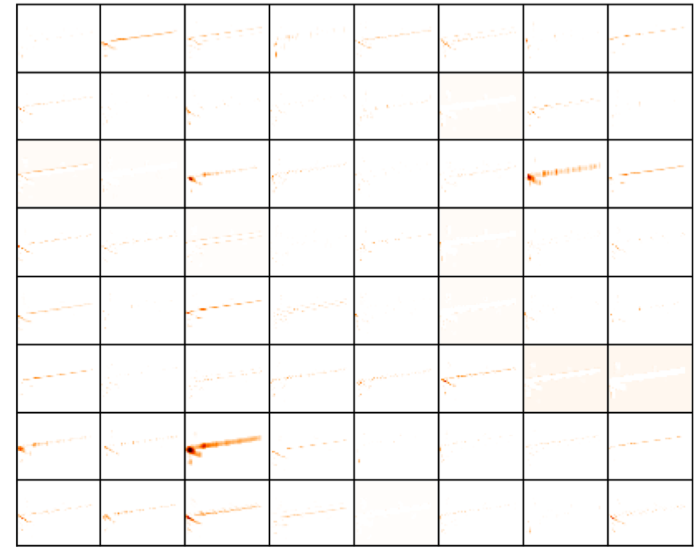


Image Y-view → First CONV layer filters → Output of convolution

- Convolution filters and outputs show interesting features about how the NN is providing discrimination

- Major gains over current algorithms in ν_e-CC discrimination:
  35% → 49% signal efficiency for the same background rejection