Machine Learning at the High Energy Frontier

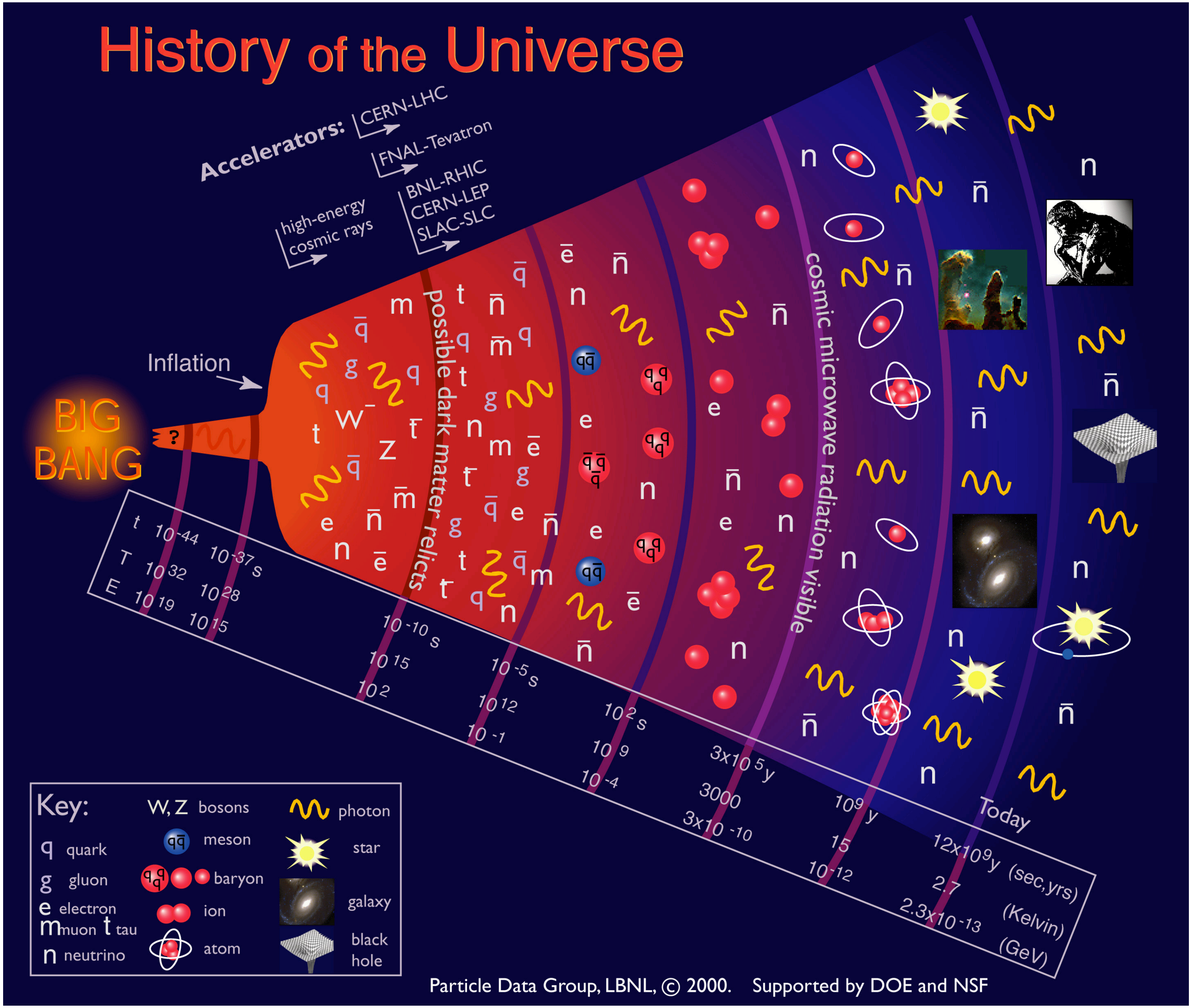
Michael Kagan

SLAC

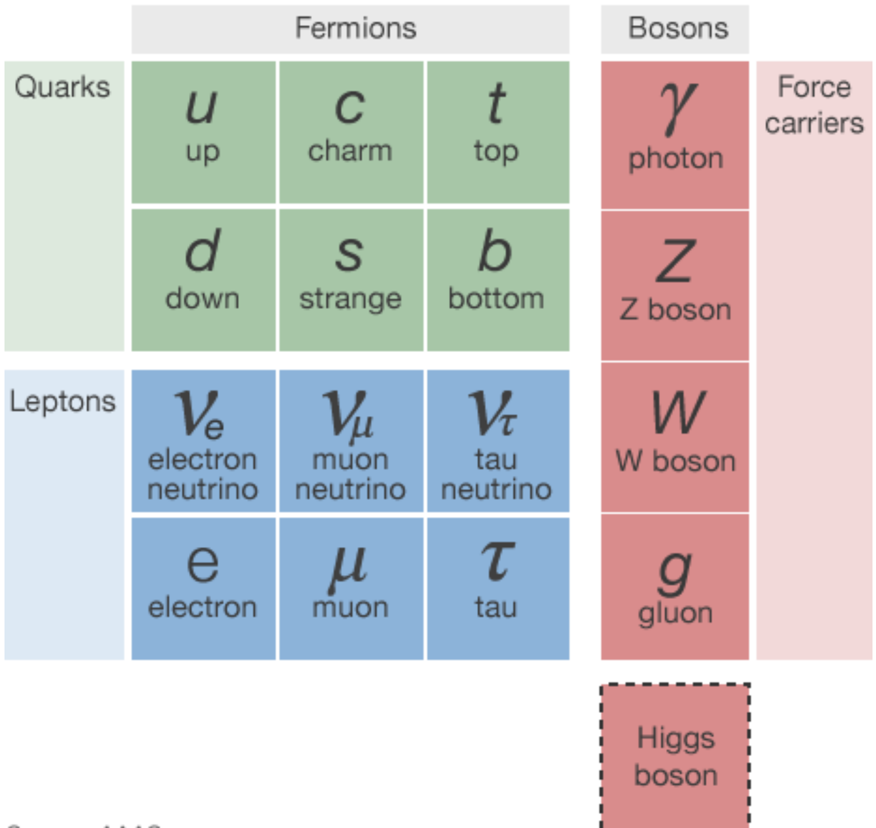
Artificial Intelligence for Science, Industry and Society

October 21, 2019

Particle Physics



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}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^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) - 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)^2 H^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} M Z^0_\mu(W^+_\mu\phi^- - W^-_\mu\phi^+) + \\
igs_w M A_\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}
$$

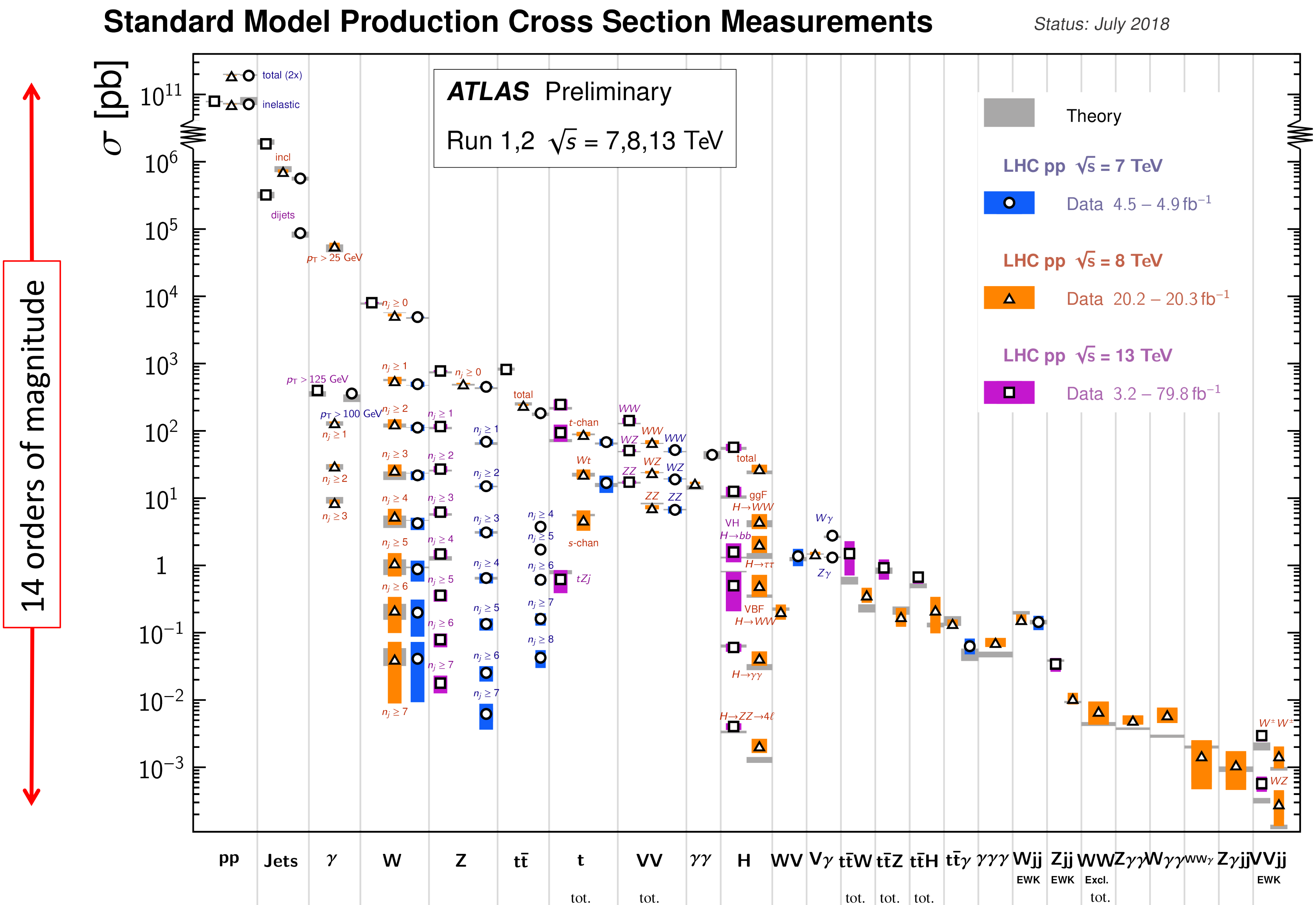
What We Know: The Standard Model

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

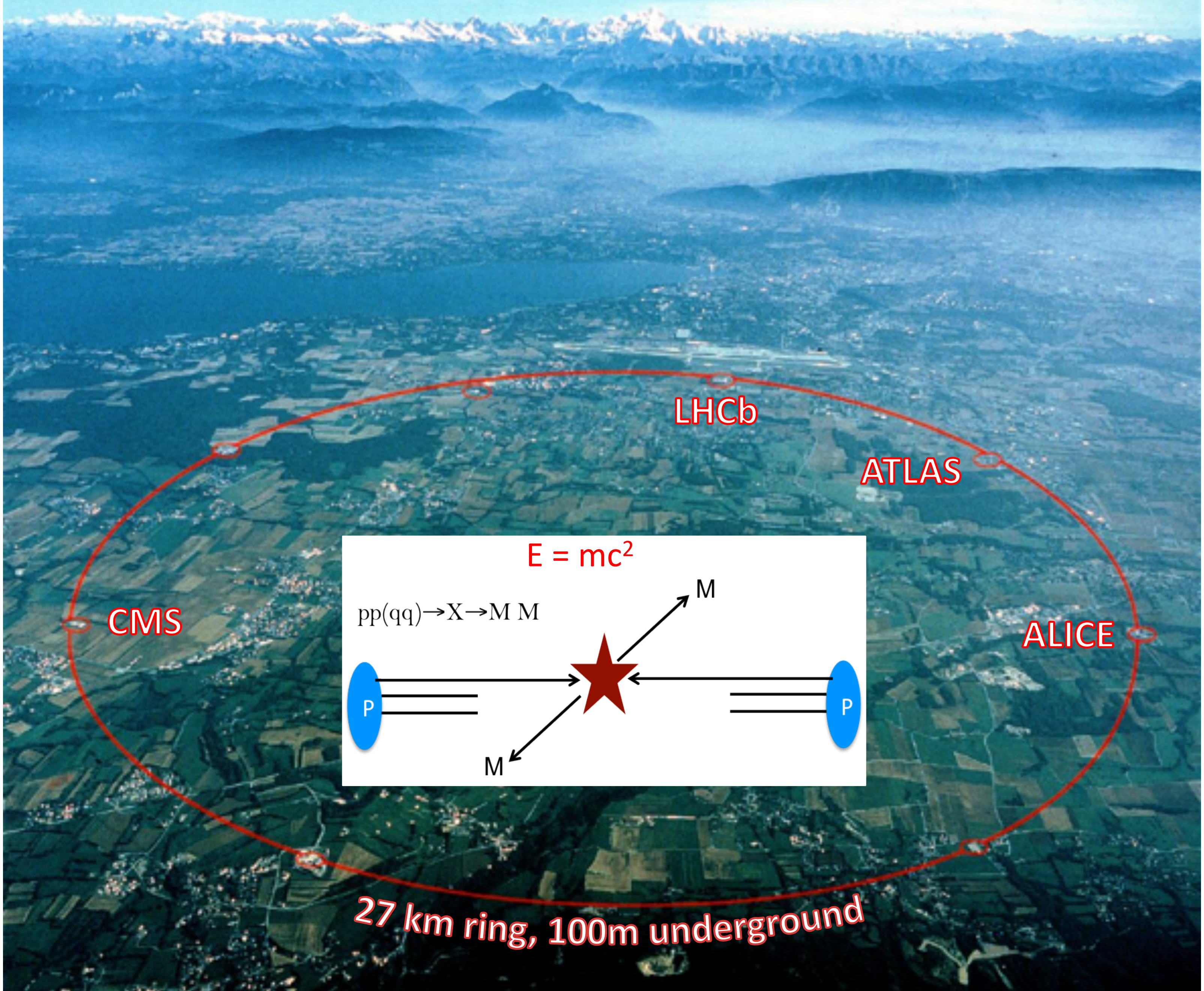


What are we missing?

- Why Gravity so much weaker then the other forces?
- What are Dark Matter and Dark Energy?
- What gives neutrinos their mass?
- Why is there anything at all?
  - Matter and anti-matter should have annihilated in the early universe
- …

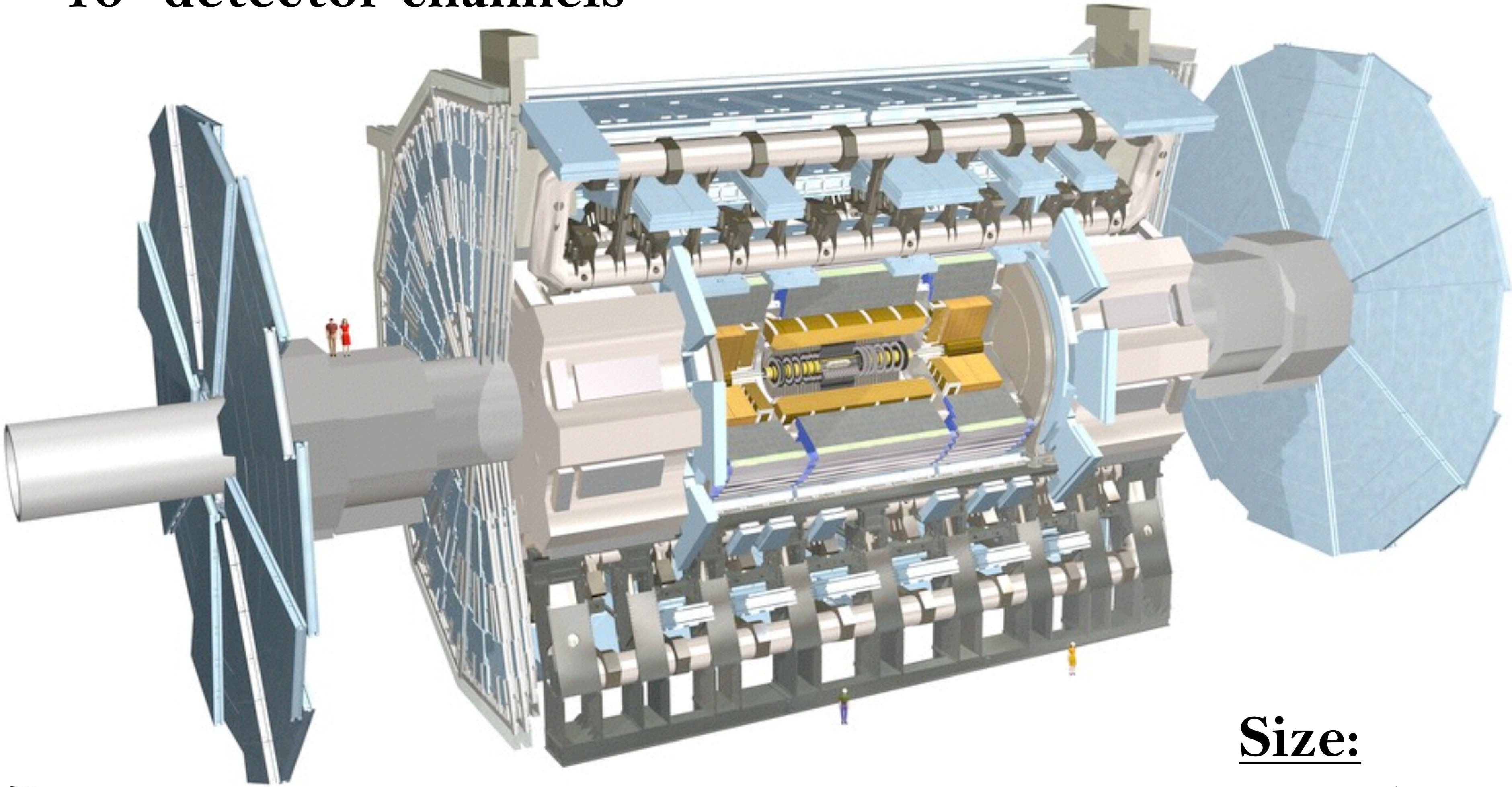
- New forces and heavy particles may have been active during the early universe that explain these phenomena
- We can look for them in high energy physics experiments!

The Large Hadron Collider at CERN



The ATLAS Experiment

$\sim 10^8$ detector channels



Data:

O(1) GB / sec

O(10) PB / year

Weight:

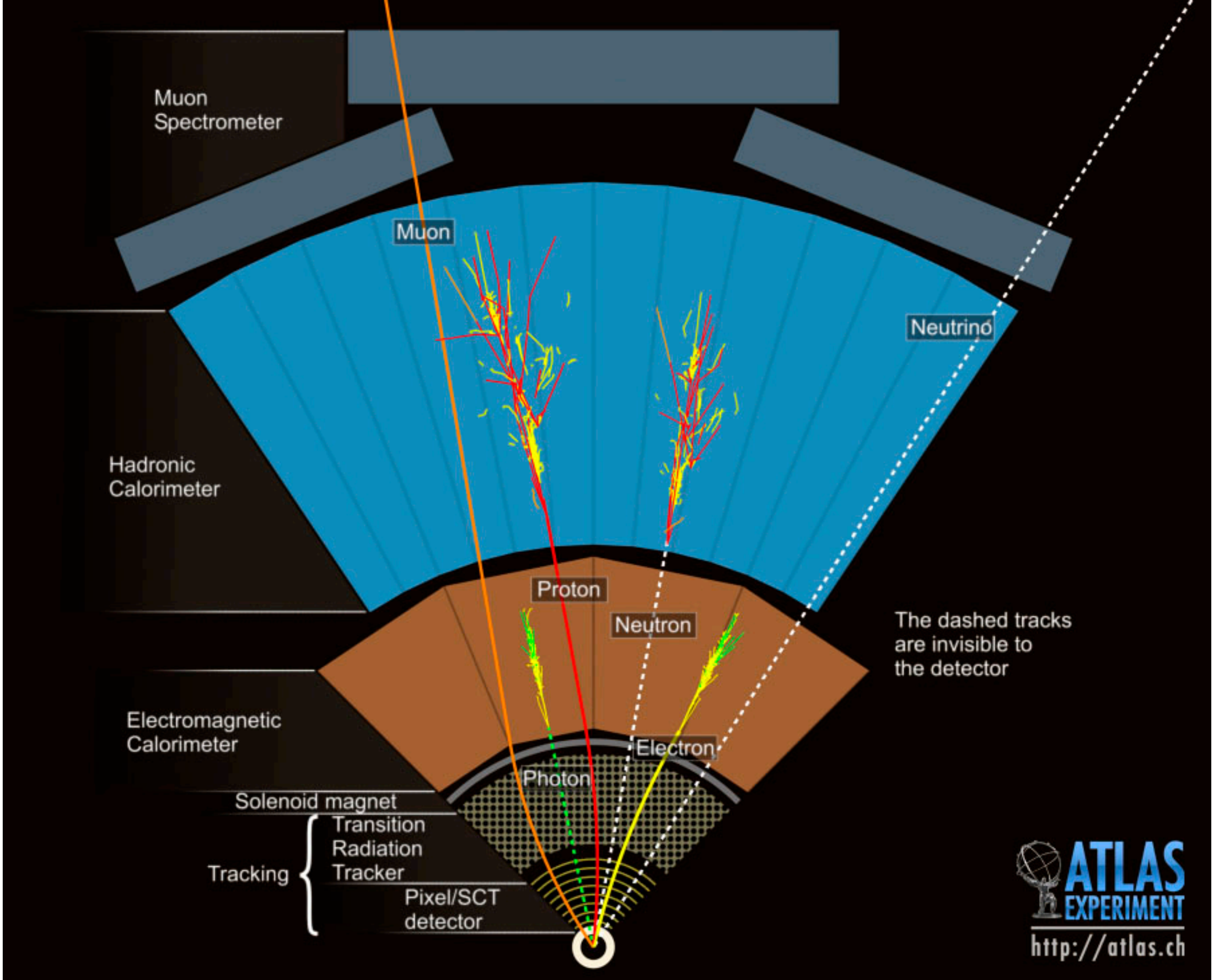
7000 tons

Size:

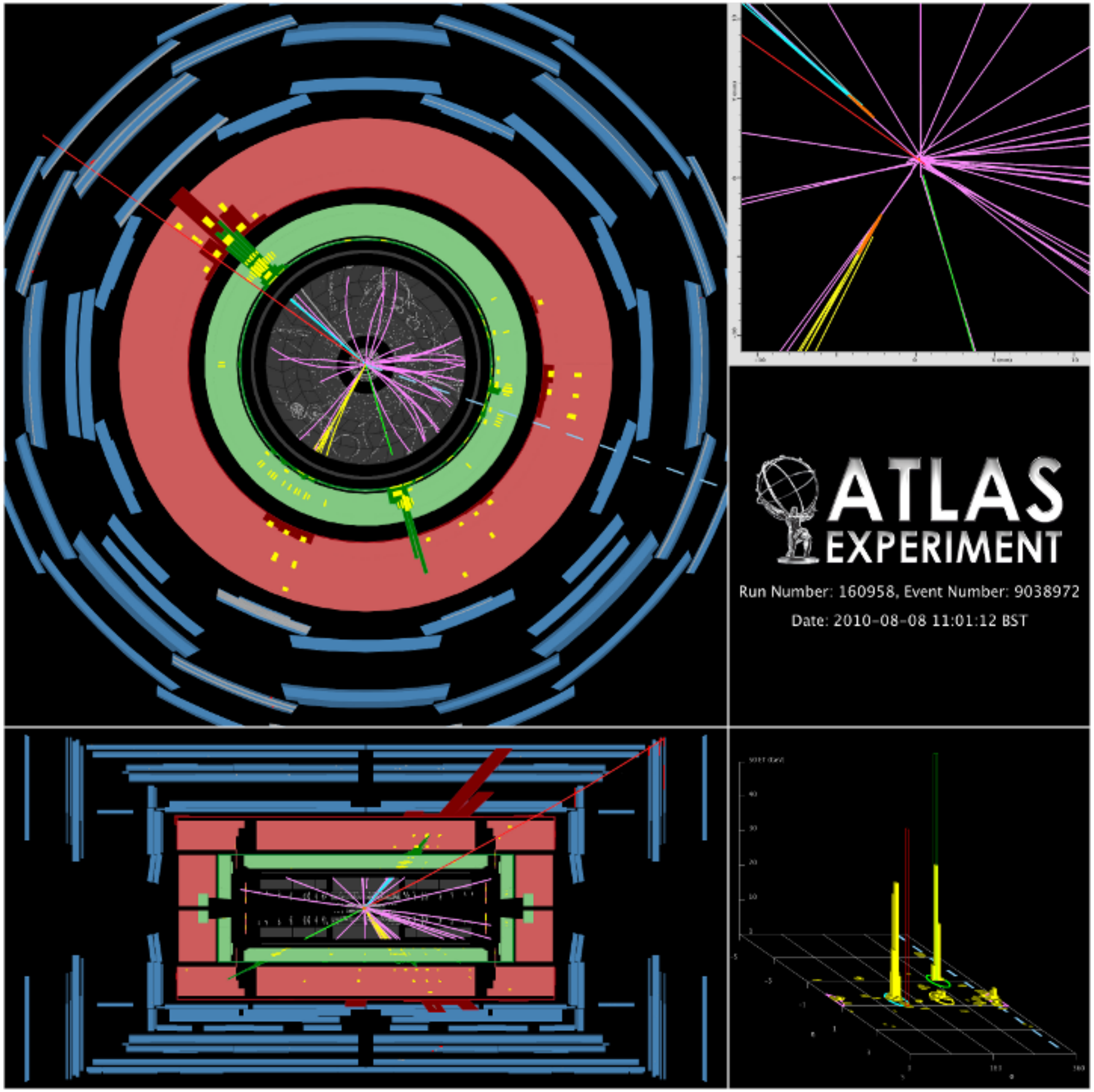
46 m long,

25 m high,

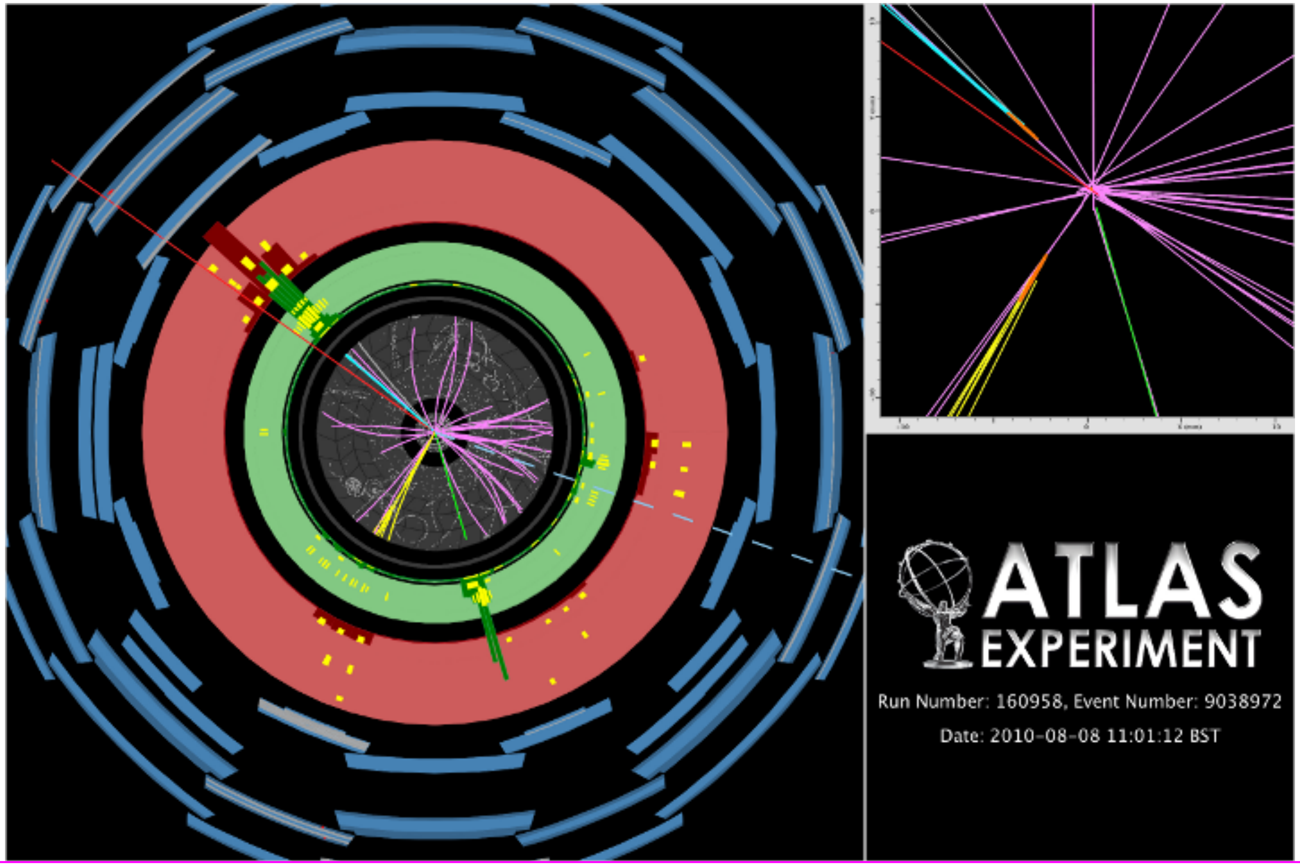
25 m wide


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

Studying Collisions



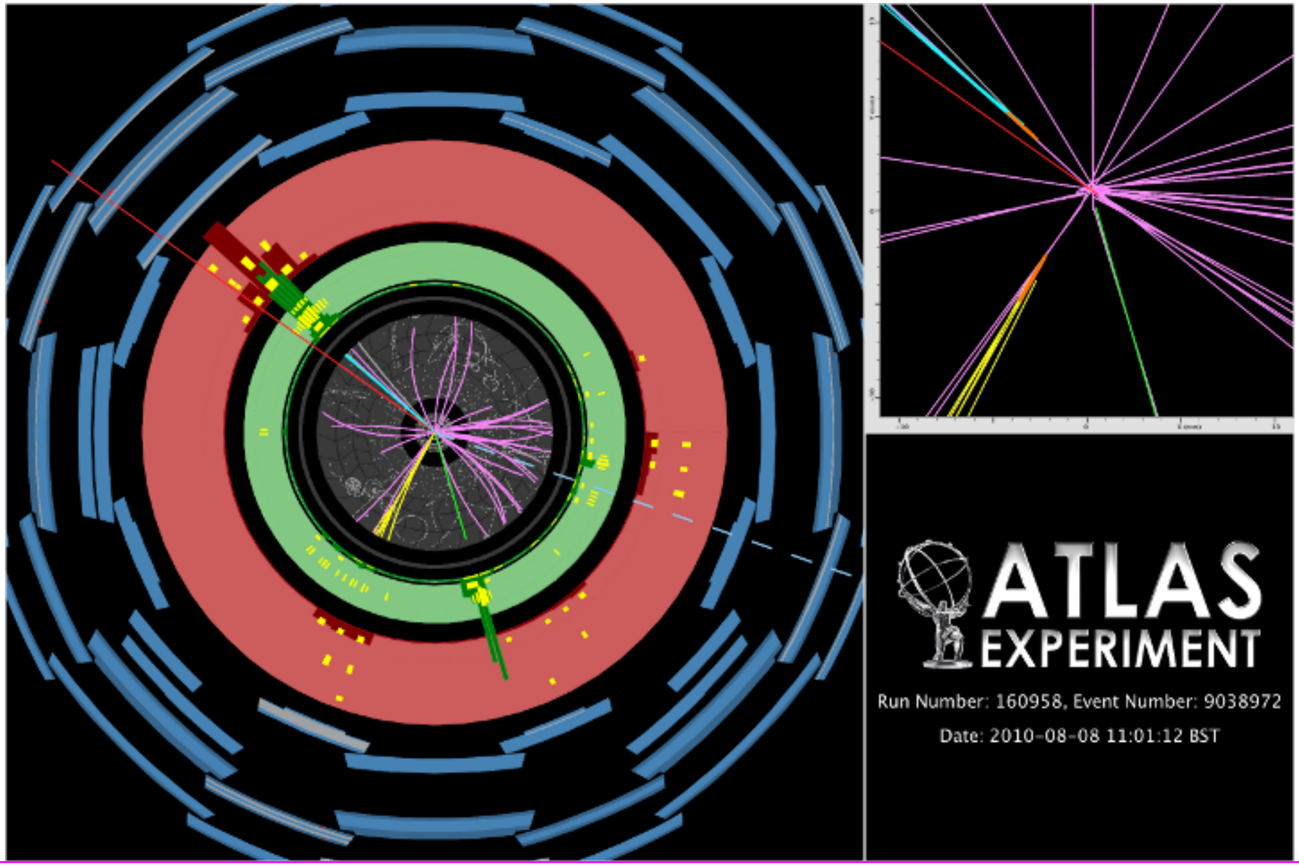
Studying Collisions



- Causal and Compositional Structure

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

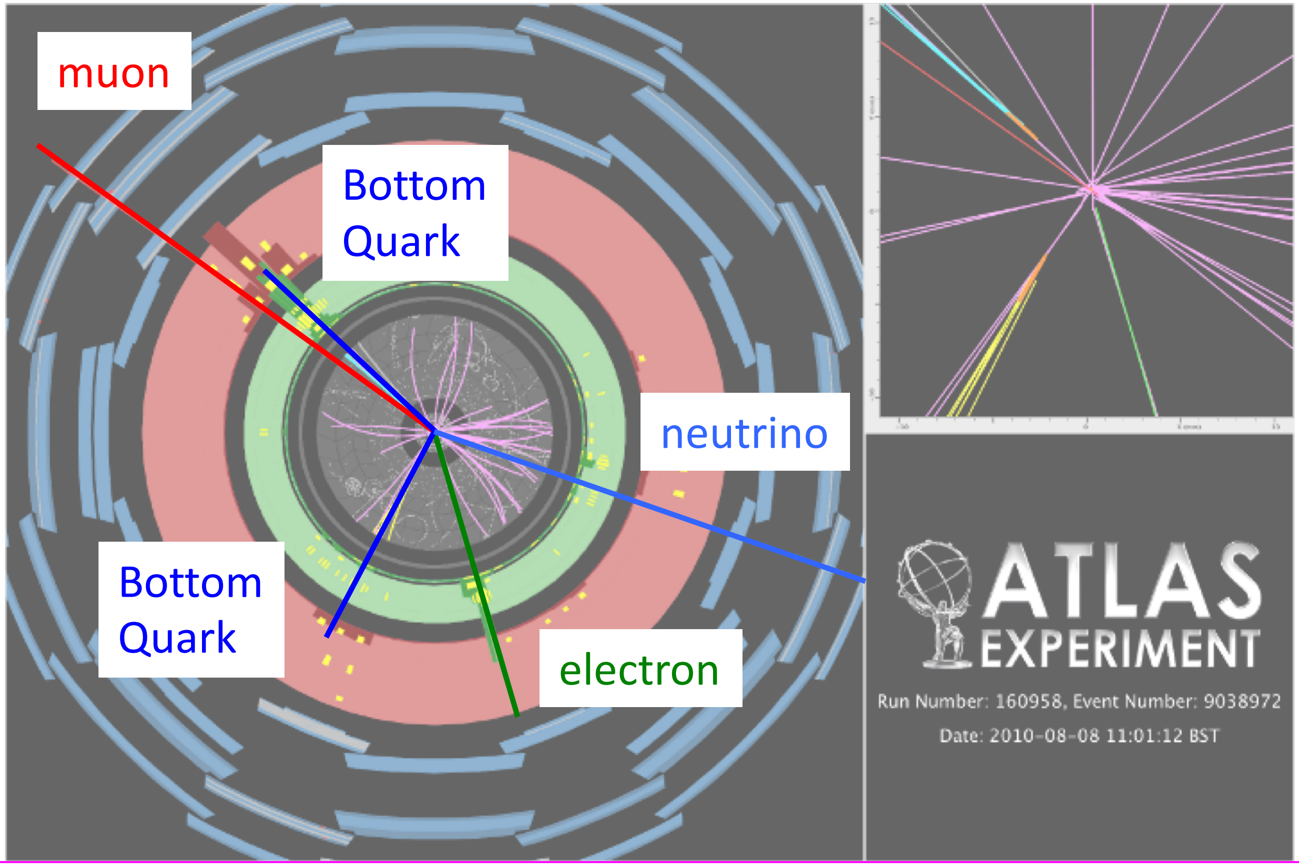
Studying Collisions



- Causal and Compositional Structure

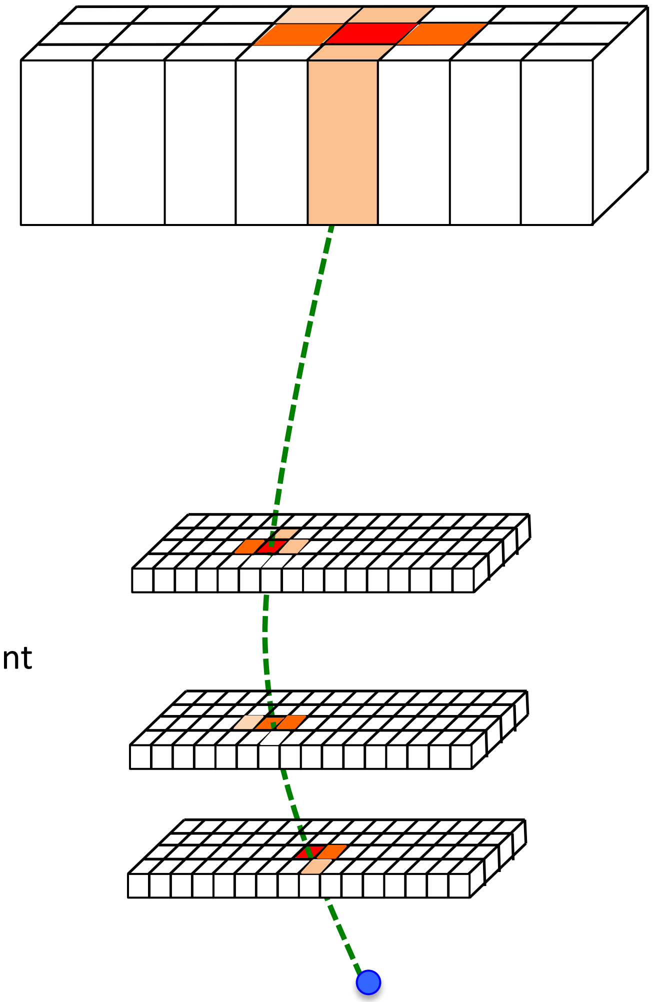
particle X ← "final state" particles ← detector data

Studying Collisions



- Reconstruction: Find the “final state” particles in the detector

Reconstructing Particles



Calorimeter:

Stops particle and destructively measure energy / direction

Tracking detector:

Typically Si-pixel detector

Non-destructive space-point measurement

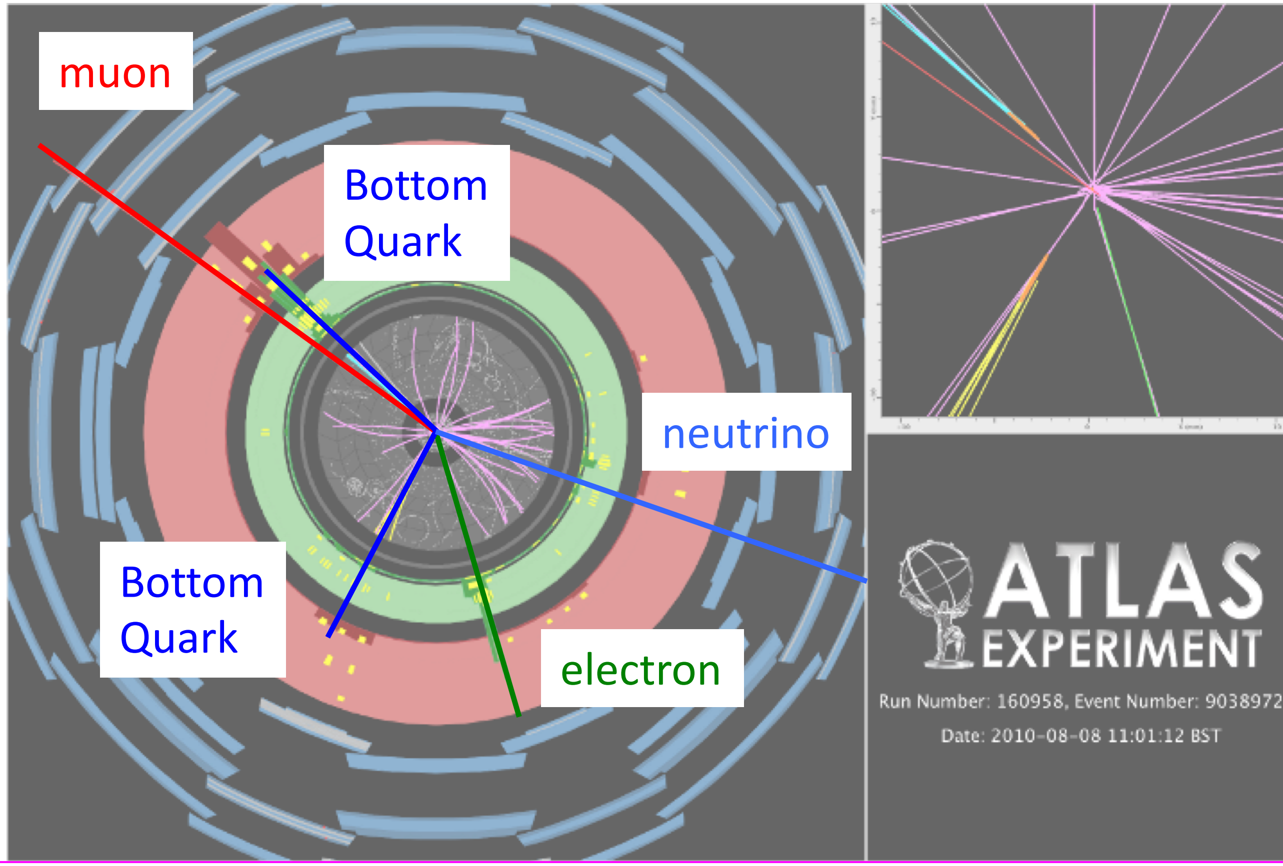
- Particle identification = Classification

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

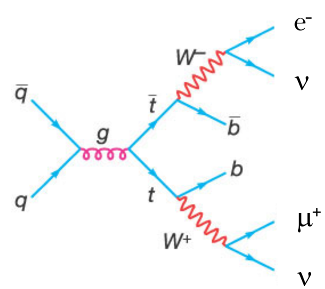
- Energy estimation = Inference, regression

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

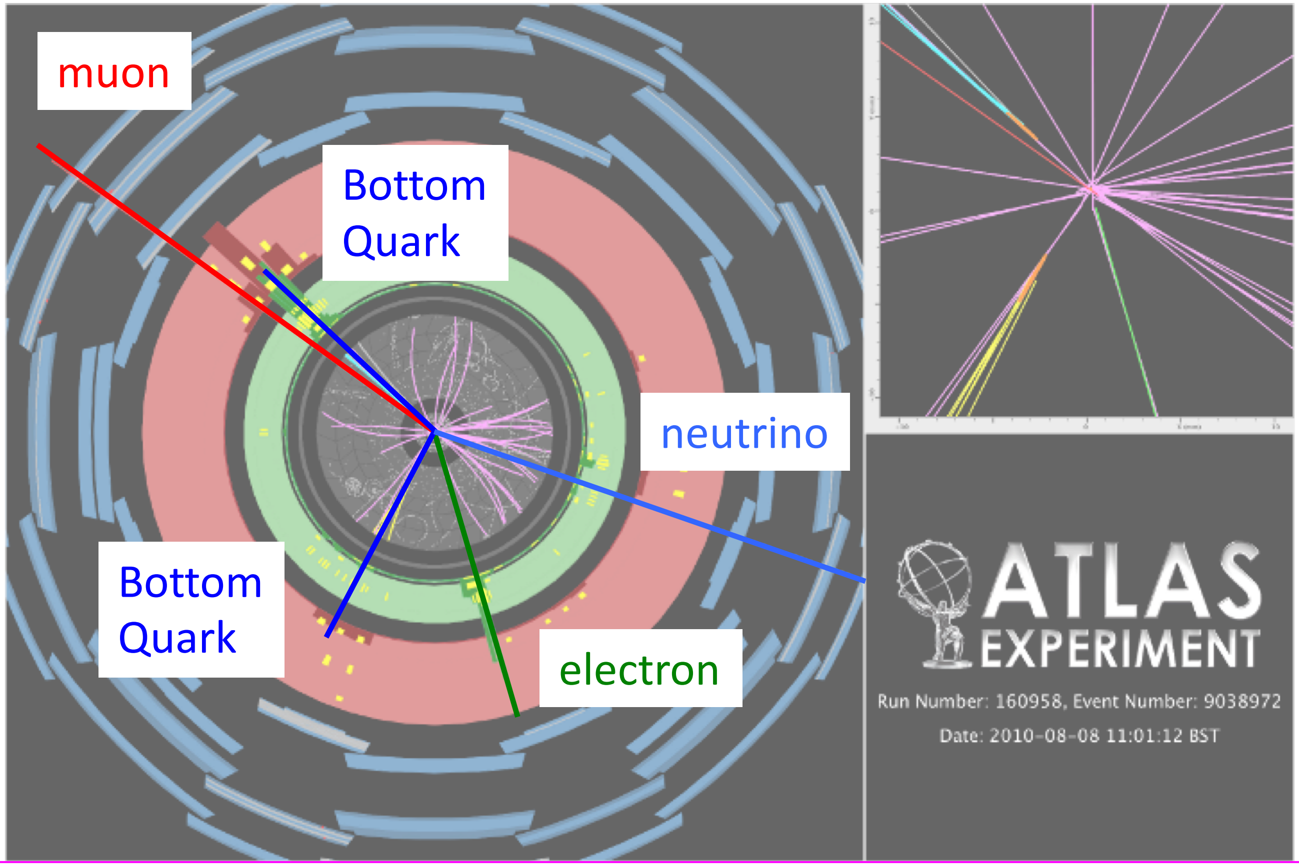
Studying Collisions



- Add them together to study underlying collision



Studying Collisions

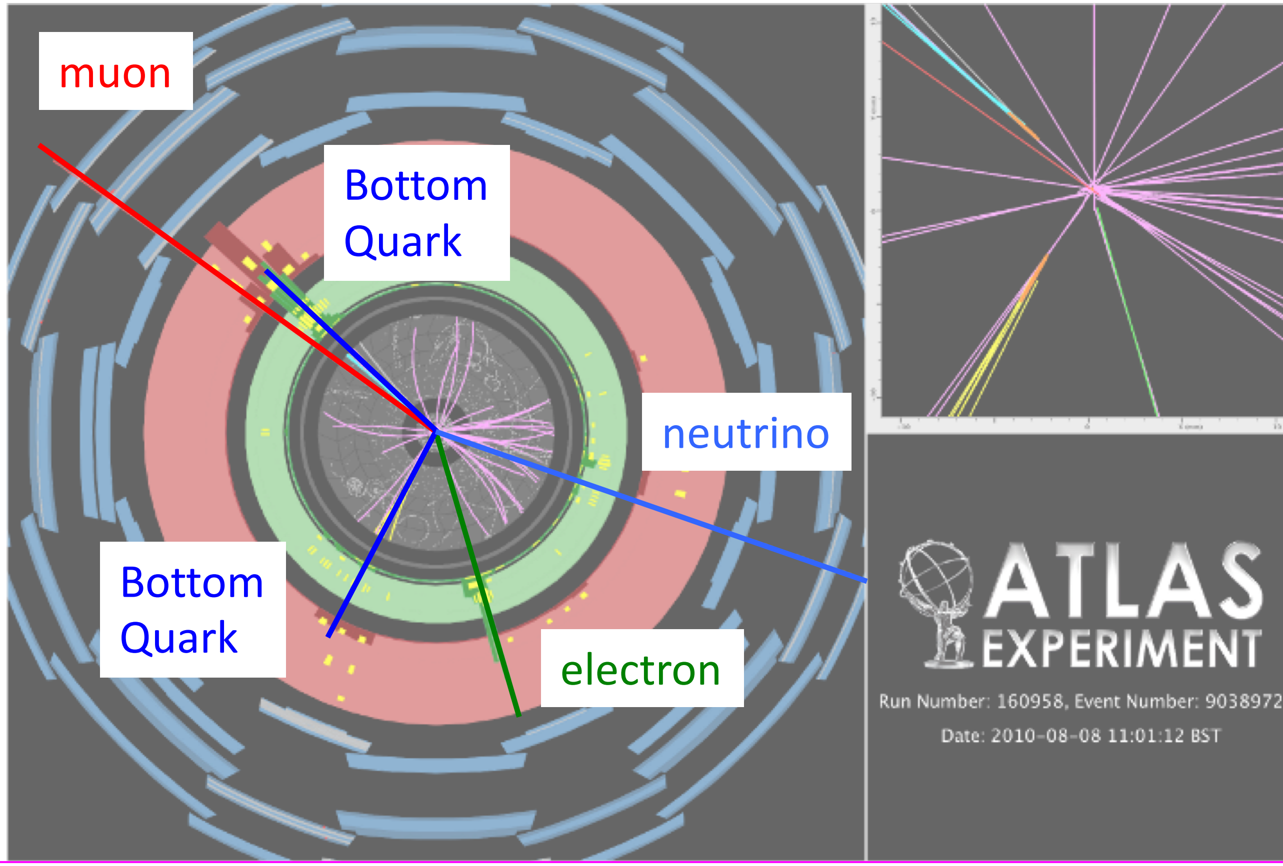


- Works because of energy and momentum conservation:

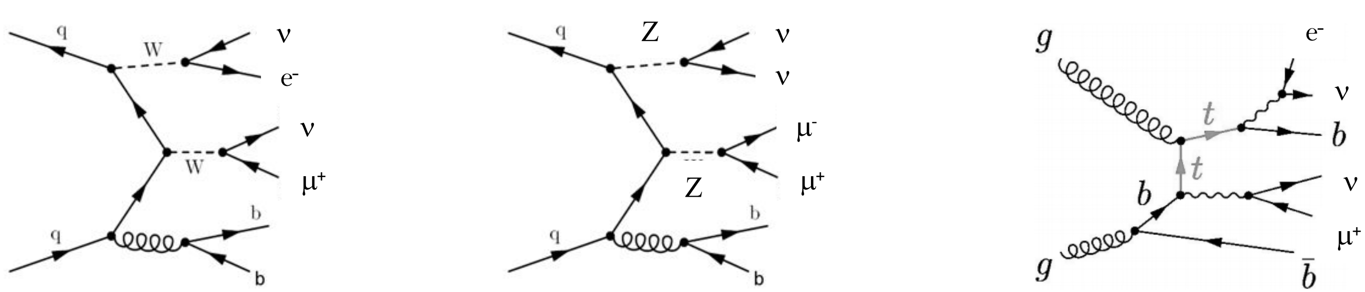
$$(E_X, \vec{p}_X) = \sum_{i \in \text{decay products}} (E_i, \vec{p}_i)$$

$$M_X c^2 = \sqrt{E_X^2 - \vec{p}_X^{\,2}\, c^2}$$

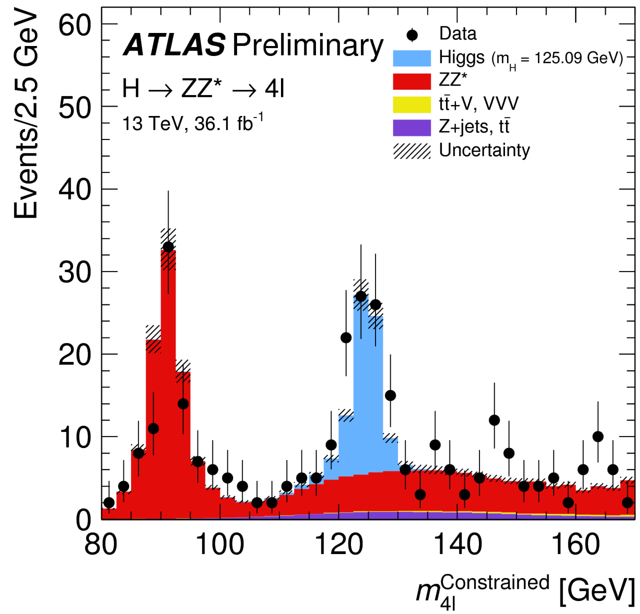
Studying Collisions



- Multiple processes contribute to same signature

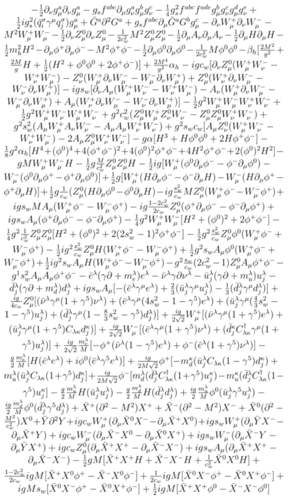


Studying Collisions

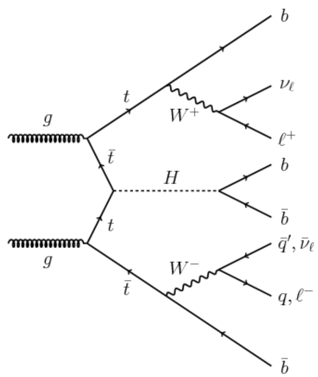


- Use physics knowledge to design sensitive observable
- With a collection of collisions we can perform:
  - Hypothesis testing: new particle present?
  - Measurement: Inference of latent parameters, e.g. Higgs mass

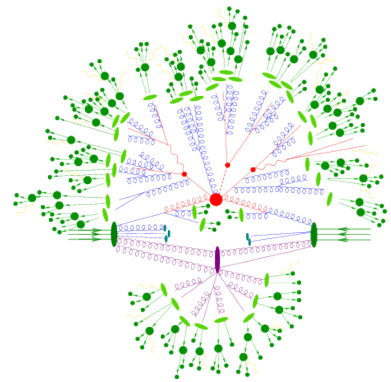
From Theory to Experiment



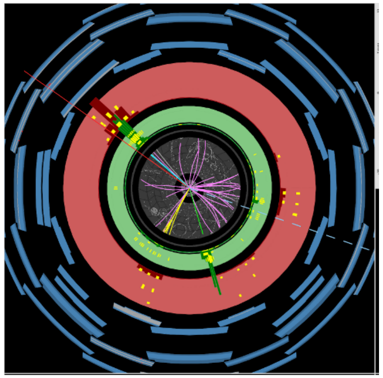
Parameters θ



O(10) particles



O(100) particles



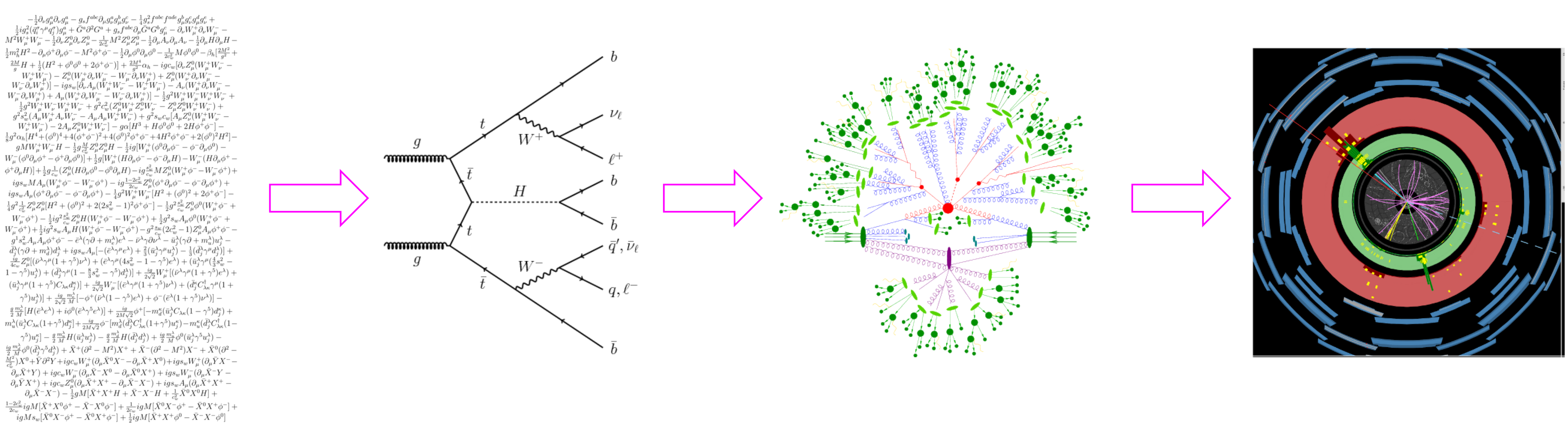
O(10^8) detector elements

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

Simulation Based Science

- Mechanistic understanding of particle evolution and interaction which we encode in a simulator

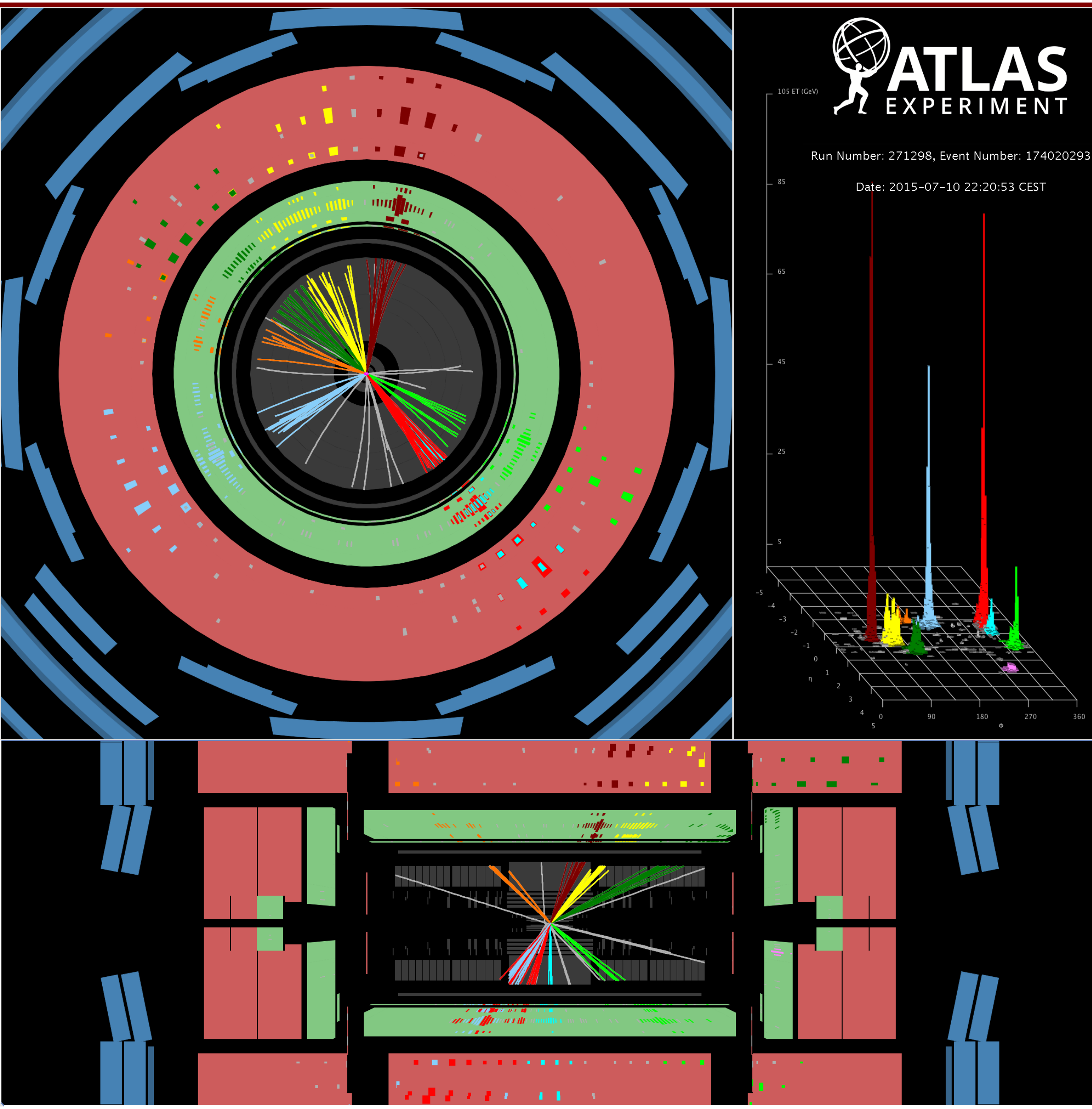
- Likelihood $p(x|\theta) = \int p(x,z|\theta)dz$ is intractable
  - High dimensional data x
  - Many latent random variables z for underlying physics process
  - Non-differentiable control flow in simulation

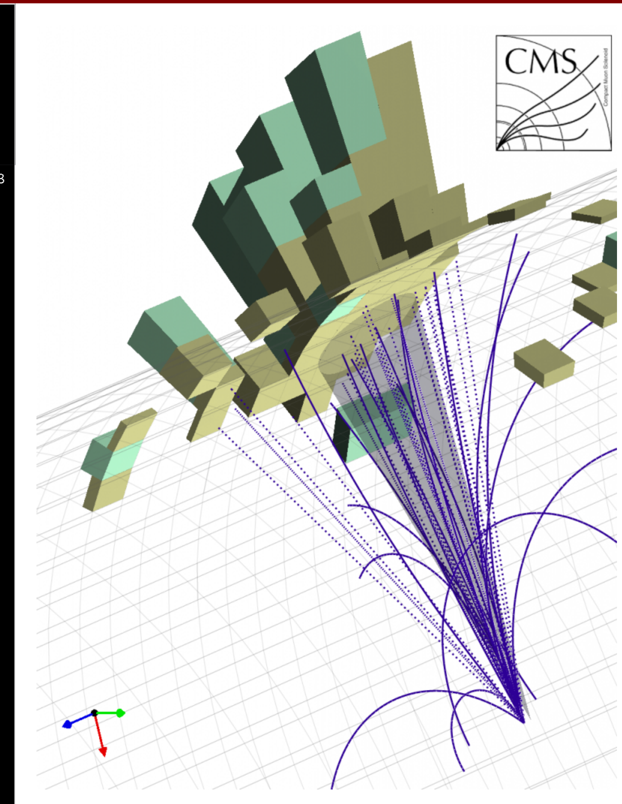


Simulation Based Science

- Mechanistic understanding of particle evolution and interaction which we encode in a simulator
- Likelihood $p(x|\theta) = \int p(x,z|\theta)dz$ is intractable
  - High dimensional data x
  - Many latent random variables z for underlying physics process
  - Non-differentiable control flow in simulation
- Monte Carlo methods allow us to generate samples by running the physics in the *forward* direction
- Can produce vast quantities of labeled data on demand from high fidelity simulation!
- Majority of ML in HEP thus far has focused on supervised learning: classification and regression

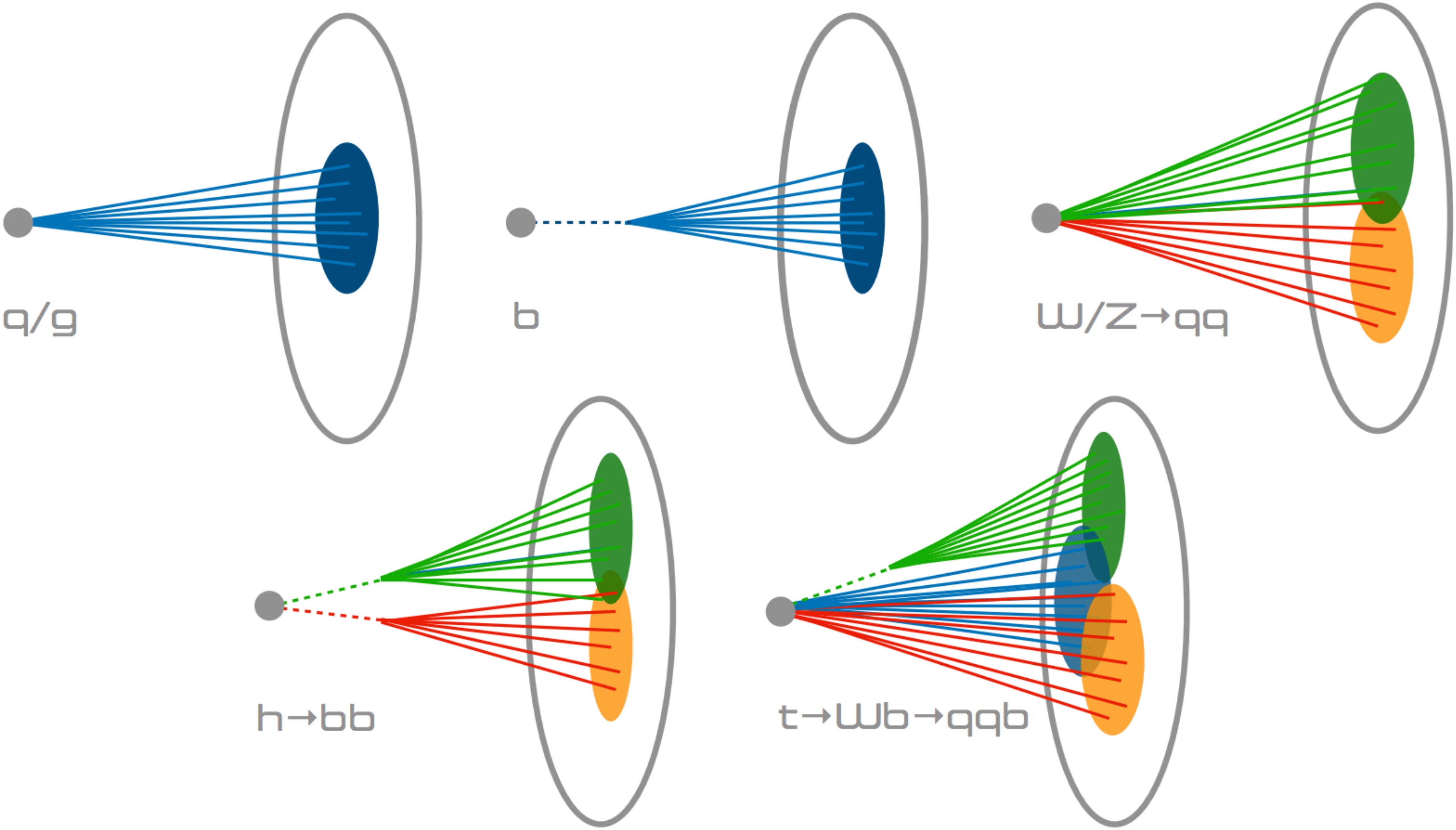
Jets at the LHC





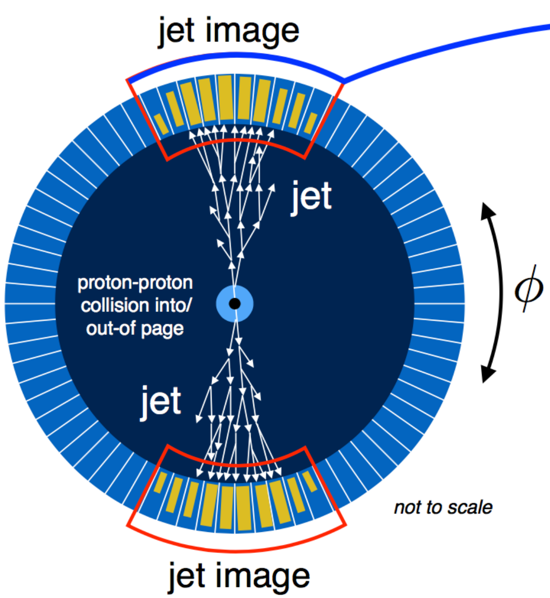
- **Jet**: stream of particles produced by high energy quarks and gluons
  - Clustering algorithms used to find them

Canonical Discrimination Problem: Jet Identification

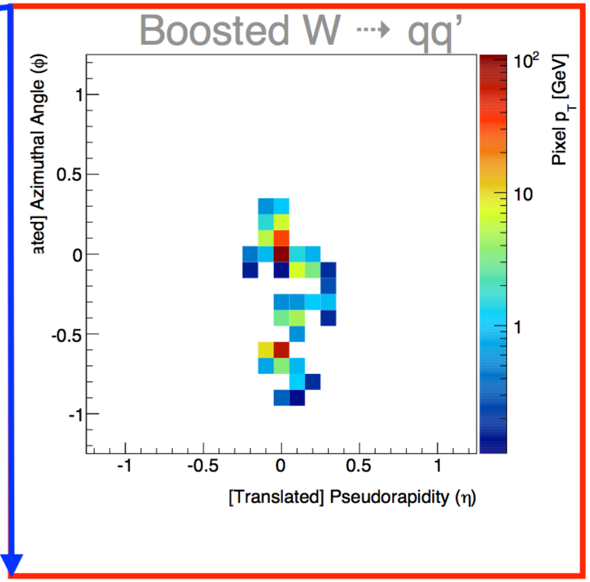


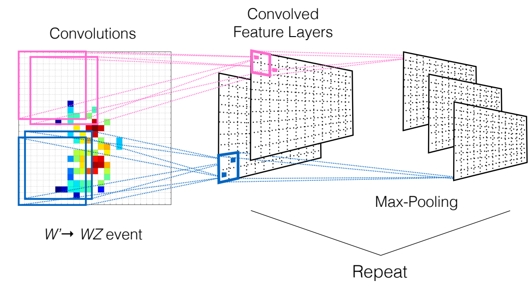
Jet Images

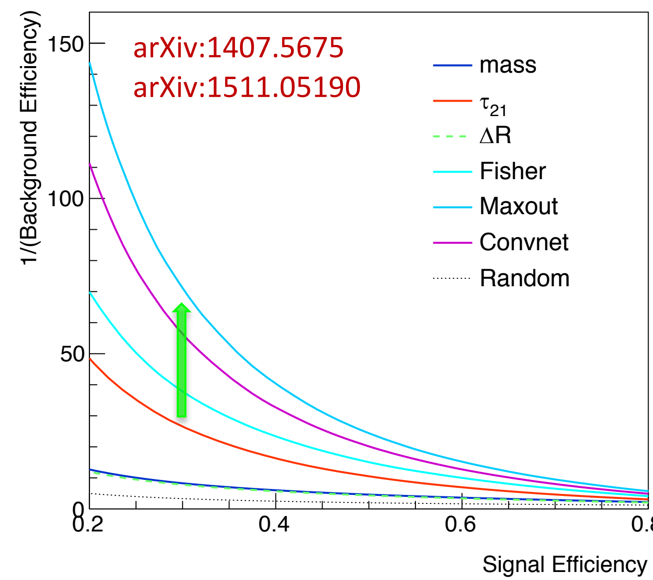
[Image from B. Nachman]

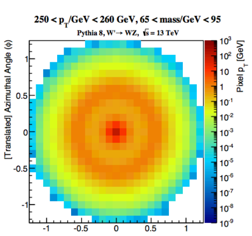


Unrolled slice of detector

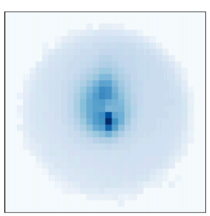




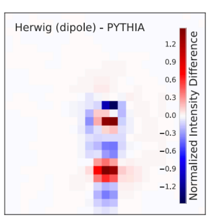




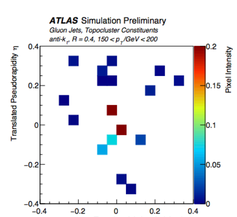
1511.05190



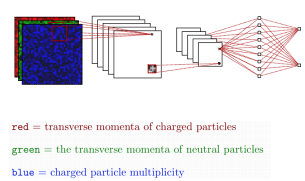
1603.09349



1609.00607

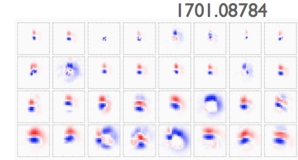


ATL-PHYS-PUB-2017-017

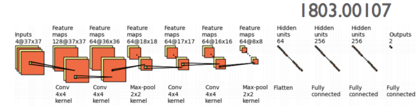


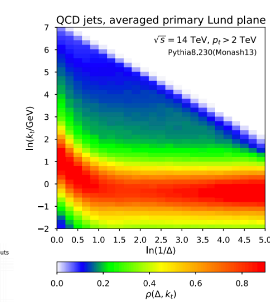
1612.01551

1701.08784



1803.00107





1807.04758

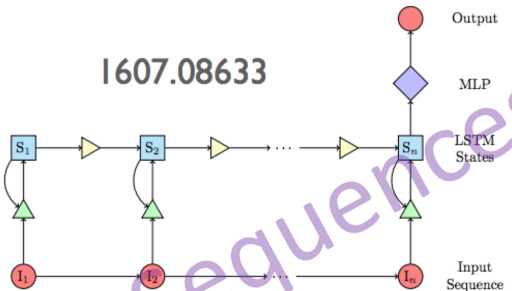
[Based on slides H. Qu]

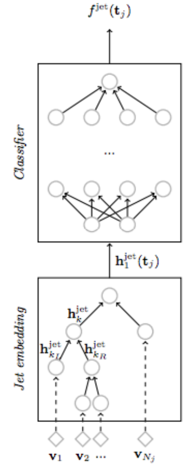
New Ways to Think about Jets

[Based on slides H. Qu]

Sequences

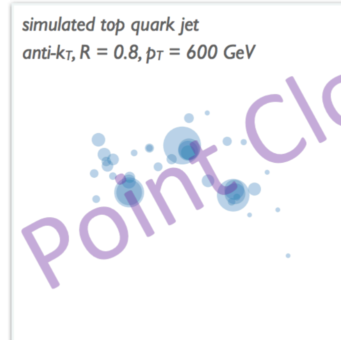
1607.08633

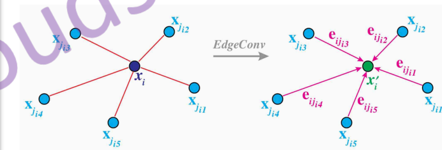




1702.00748

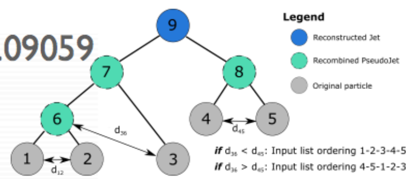
Point Clouds



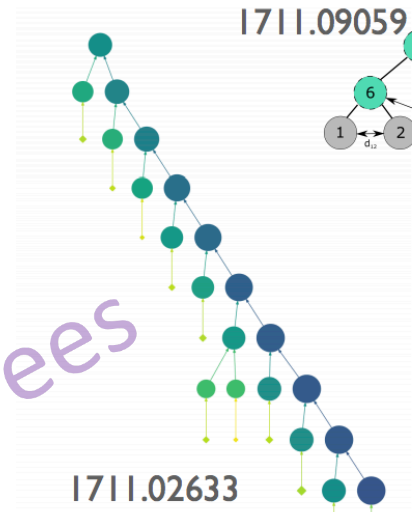


1902.08570

1711.09059

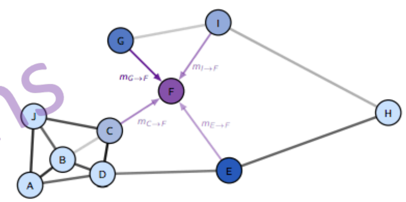


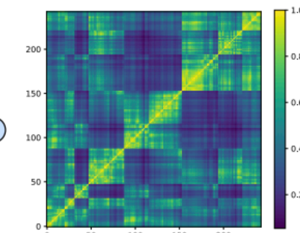
Trees



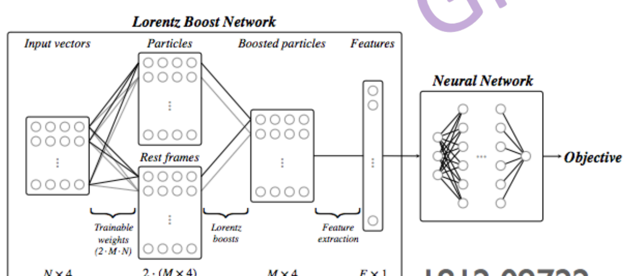
1711.02633

Graphs

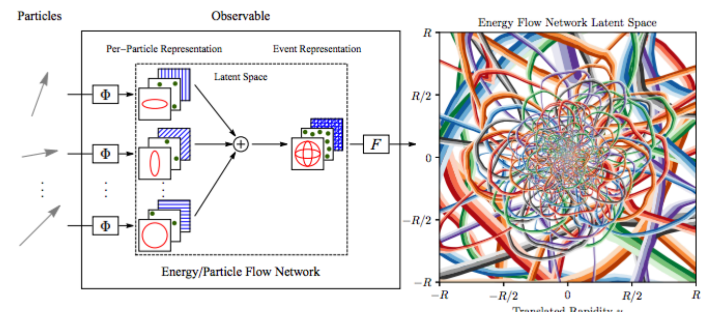




NIPS2017 workshop [http://bit.ly/2AkwYRG]



1812.09722



1810.05165

$$k_{\mu,i} \xrightarrow{\text{CoLa}} \widetilde{k}_{\mu,j} = k_{\mu,i}\, C_{ij}$$

$$\text{with} \quad C = \begin{pmatrix} 1 & 1 & \cdots & 0 & \chi_1 & \cdots & 0 & C_{1,N+2} & \cdots & C_{1,M} \\ \vdots & & \ddots & & & \ddots & & \vdots & \ddots & \vdots \\ 1 & 0 & \cdots & 1 & 0 & \cdots & \chi_N & C_{N,N+2} & \cdots & C_{N,M} \end{pmatrix} .$$

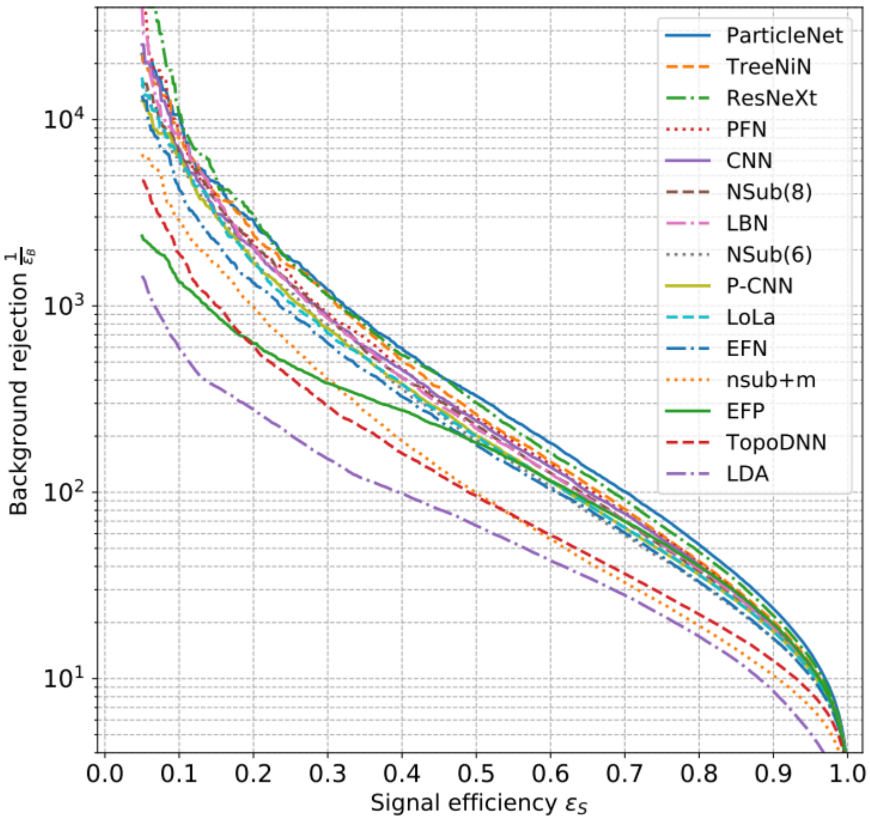
$$\tilde{k}_j \xrightarrow{\text{LoLa}} \hat{k}_j = \begin{pmatrix} m^2(\tilde{k}_j) \\ p_T(\tilde{k}_j) \\ p_T(\tilde{k}_j)\Delta R_{j,\text{jet}} \\ w_{jm}^{(E)}\, E(\tilde{k}_m) \\ w_{jm}^{(d)}\, d_{jm}^2 \\ E_T(\tilde{k}_j)E_T(\tilde{k}_m)(\Delta R_{jm})^{0.2} \end{pmatrix},$$

1707.08966, 1812.09223

Sets

And more…

Comparison of methods



	AUC	Acc	$1/\epsilon_B$ ($\epsilon_S = 0.3$) single	mean	median	#Param
CNN [16]	0.981	0.930	914±14	995±15	975±18	610k
ResNeXt [31]	0.984	0.936	1122±47	1270±28	1286±31	1.46M
TopoDNN [18]	0.972	0.916	295±5	382± 5	378 ± 8	59k
Multi-body N-subjettiness 6 [24]	0.979	0.922	792±18	798±12	808±13	57k
Multi-body N-subjettiness 8 [24]	0.981	0.929	867±15	918±20	926±18	58k
TreeNiN [43]	0.982	0.933	1025±11	1202±23	1188±24	34k
P-CNN	0.980	0.930	732±24	845±13	834±14	348k
ParticleNet [47]	0.985	0.938	1298±46	1412±45	1393±41	498k
LBN [19]	0.981	0.931	836±17	859±67	966±20	705k
LoLa [22]	0.980	0.929	722±17	768±11	765±11	127k
LDA [54]	0.955	0.892	151±0.4	151.5±0.5	151.7±0.4	184k
Energy Flow Polynomials [21]	0.980	0.932	384			1k
Energy Flow Network [23]	0.979	0.927	633±31	729±13	726±11	82k
Particle Flow Network [23]	0.982	0.932	891±18	1063±21	1052±29	82k
GoaT	0.985	0.939	1368±140		1549±208	35k

Deep Image Prior

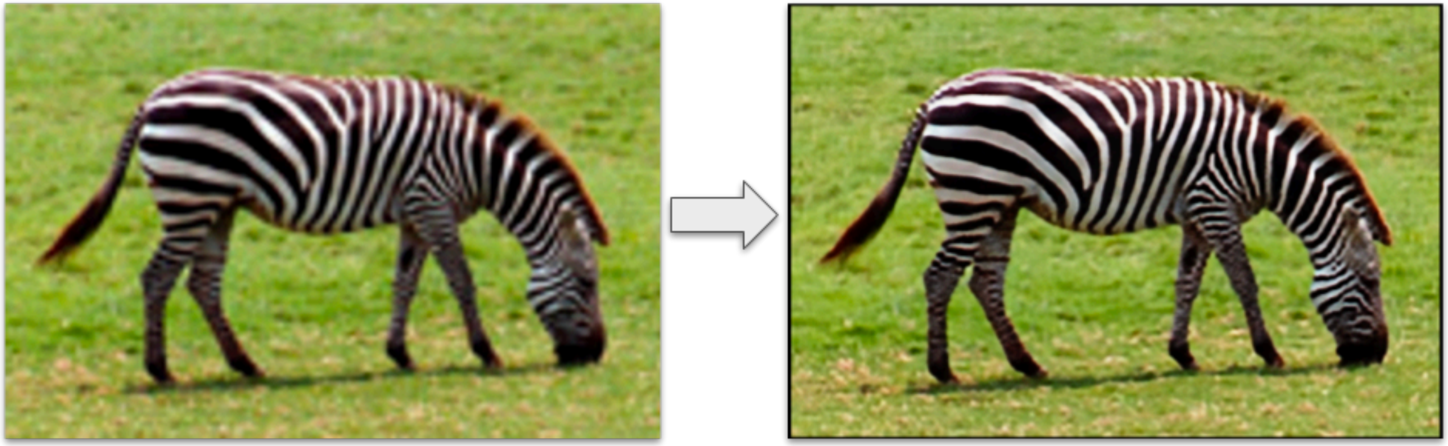
Corrupted

$$\min_x E(x;\ x_0)^2 + R(x)$$

- Image restoration: find "true" image x from corrupted image x_0

Deep Image Prior

arXiv:1711.10925



Corrupted   Deep image prior

- Image restoration: find “true” image x from corrupted image x_0
- Instead of hand designed prior, use a ConvNet that outputs pixel predictions to define x

$$\boxed{\min_{\theta} E(f_\theta(z);\; x_0)}$$

  - For each image: fix random noise z, run optimization
  - Rather than search image space, search network parameter space

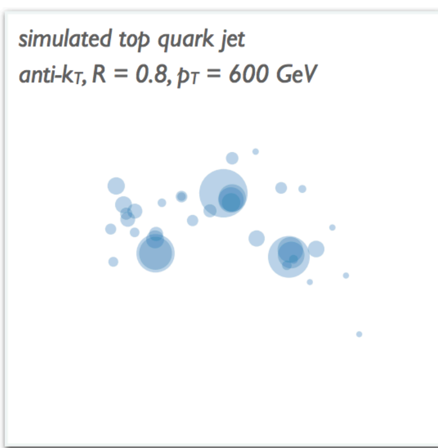
Implicit Priors

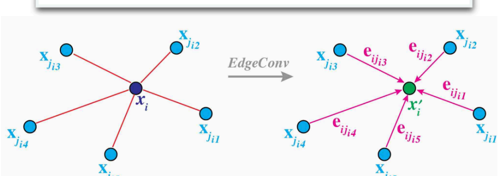
- Structure of ConvNet constrains the set of solutions; It is an *implicit prior* over solutions

- Not all data are images…
  - What is a good implicit prior for physics?
  - What architectures capture good physics solution space?
  - Should we consider the impact of measurement devices in choosing an architecture?
  - Can we use this technique (or super-resolution methods) for reconstructing particles from noisy detectors?

Physics Inductive Bias as a Guide

arXiv:1702.00748
arXiv:1711.02633





arXiv:1902.08570

arXiv:1902.09914

	AUC	Acc	$1/\epsilon_B$ ($\epsilon_S = 0.3$) single	mean	median	#Param
CNN [16]	0.981	0.930	914±14	995±15	975±18	610k
ResNeXt [31]	0.984	0.936	1122±47	1270±28	1286±31	1.46M
TopoDNN [18]	0.972	0.916	295±5	382± 5	378 ± 8	59k
Multi-body N-subjettiness 6 [24]	0.979	0.922	792±18	798±12	808±13	57k
Multi-body N-subjettiness 8 [24]	0.981	0.929	867±15	918±20	926±18	58k
TreeNiN [43]	0.982	0.933	1025±11	1202±23	1188±24	34k
P-CNN	0.980	0.930	732±24	845±13	834±14	348k
ParticleNet [47]	0.985	0.938	1298±46	1412±45	1393±41	498k
LBN [19]	0.981	0.931	836±17	859±67	966±20	705k
LoLa [22]	0.980	0.929	722±17	768±11	765±11	127k
LDA [54]	0.955	0.892	151±0.4	151.5±0.5	151.7±0.4	184k
Energy Flow Polynomials [21]	0.980	0.932	384			1k
Energy Flow Network [23]	0.979	0.927	633±31	729±13	726±11	82k
Particle Flow Network [23]	0.982	0.932	891±18	1063±21	1052±29	82k
GoaT	0.985	0.939	1368±140		1549±208	35k

- Physics inspired models make learning easier and more interpretable
  - The way the jet was clustered
  - The geometry of the particles in the jet

From Discriminative to Generative Models

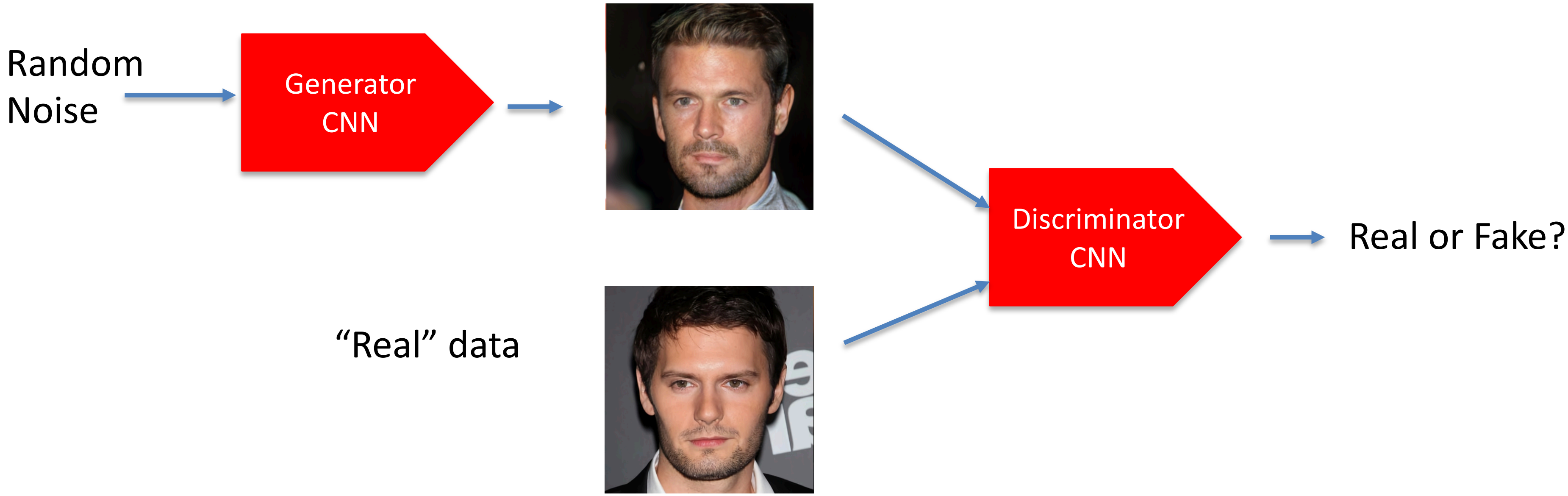
- Discriminative models: $\mathrm{f}(x) \approx \bar{y} = E_{p(y|x)}[y]$

- How do we model uncertainty on predictions, i.e. learn a posterior on likelihood?

- How do we interpret the predictions?

- Physics defines data generating process $y \rightarrow x$
  How can we better inject this knowledge?

Generative Adversarial Networks (GAN) [arXiv:1406.2661]



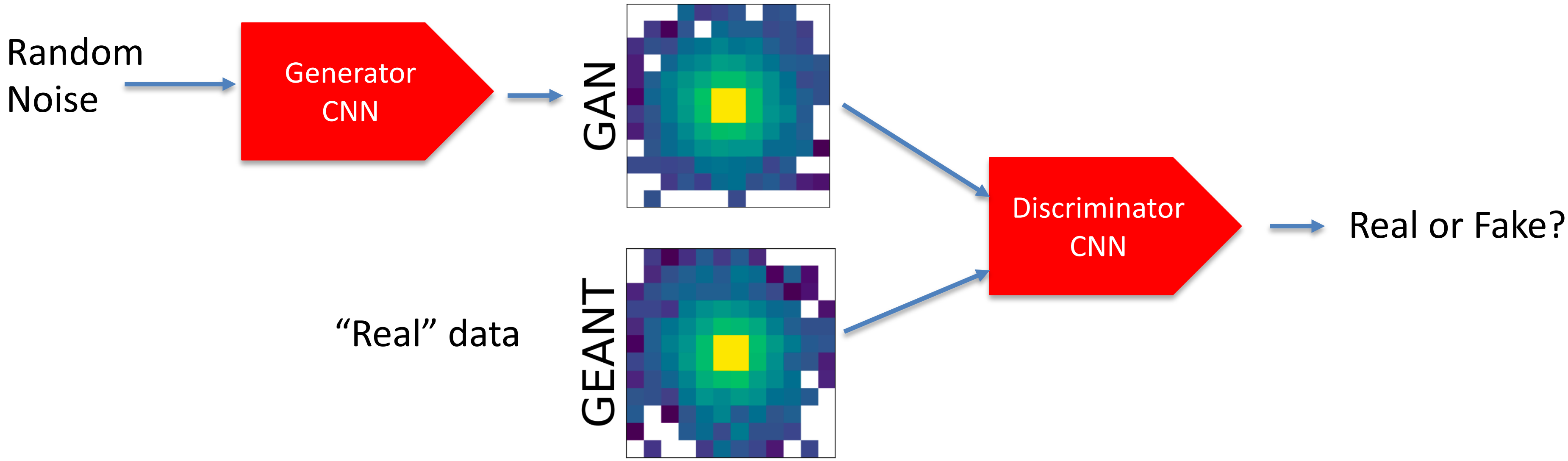
$$\min_G \max_D V(G, D)$$

$$V(G, D) = \mathbb{E}_{p_{data}(x)}[\log D(x)] + \mathbb{E}_{p_z(z)}[\log(1 - D(G(z))]$$

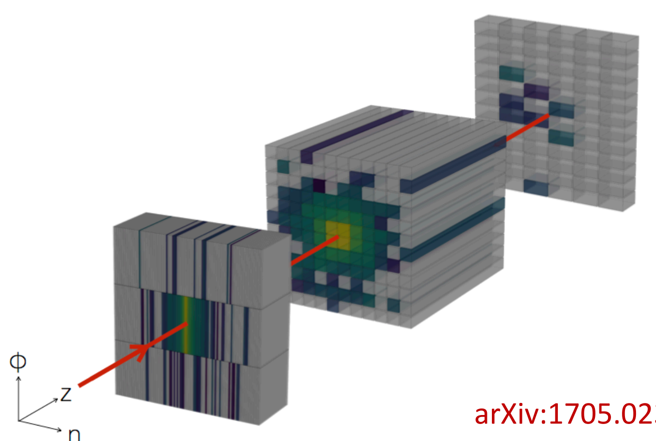
- Generator produces images from random noise and tries to trick discriminator into thinking they are real
- Classifier tries to tell the difference between real and fake images

Images: arXiv:1710.10196

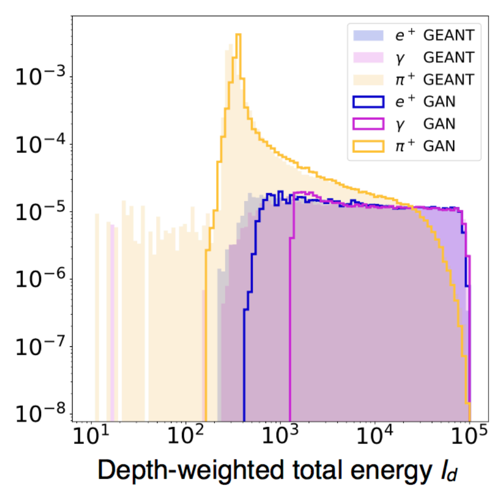
GANs for HEP

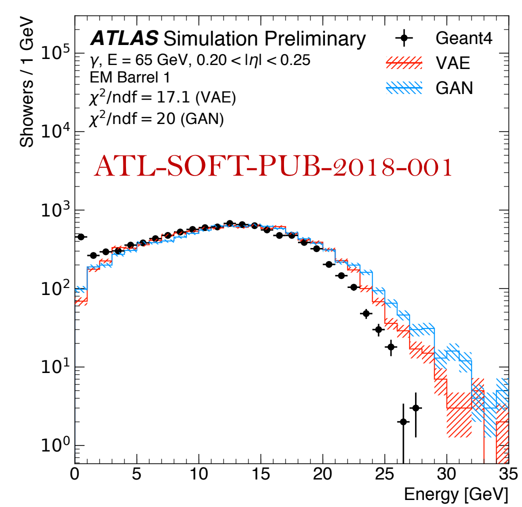


- Once trained, potential to dramatically speed up extremely computationally costly pieces of the simulation



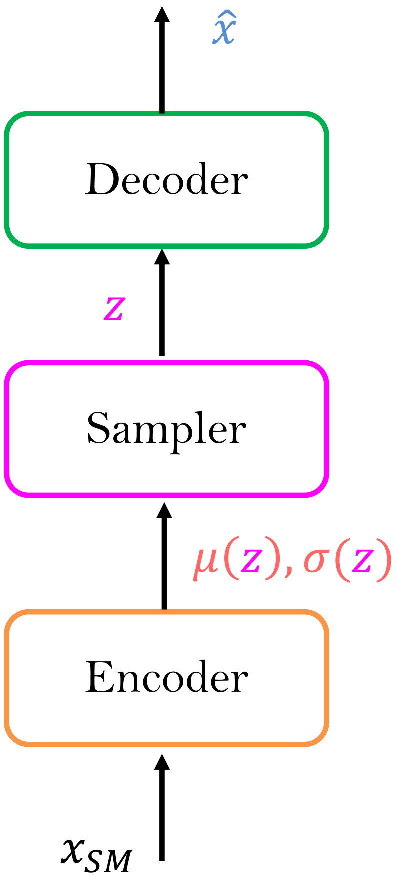
arXiv:1705.02355
arXiv:1701.05927





Variation AutoEncoders (VAE)

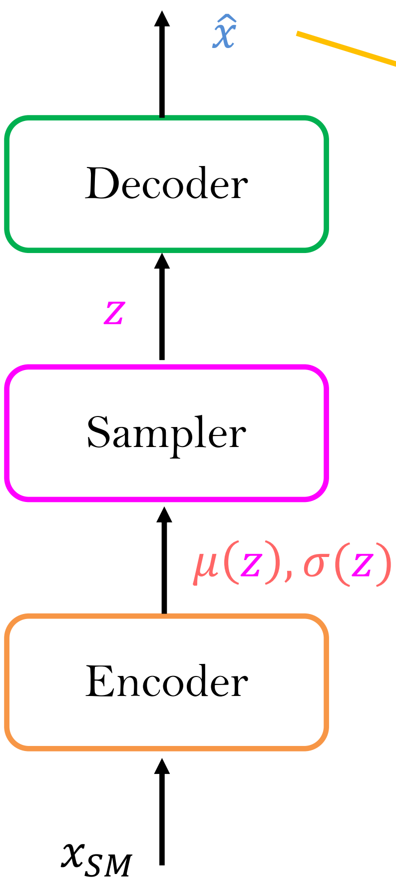
arXiv:1312.6114
arXiv:1401.4082

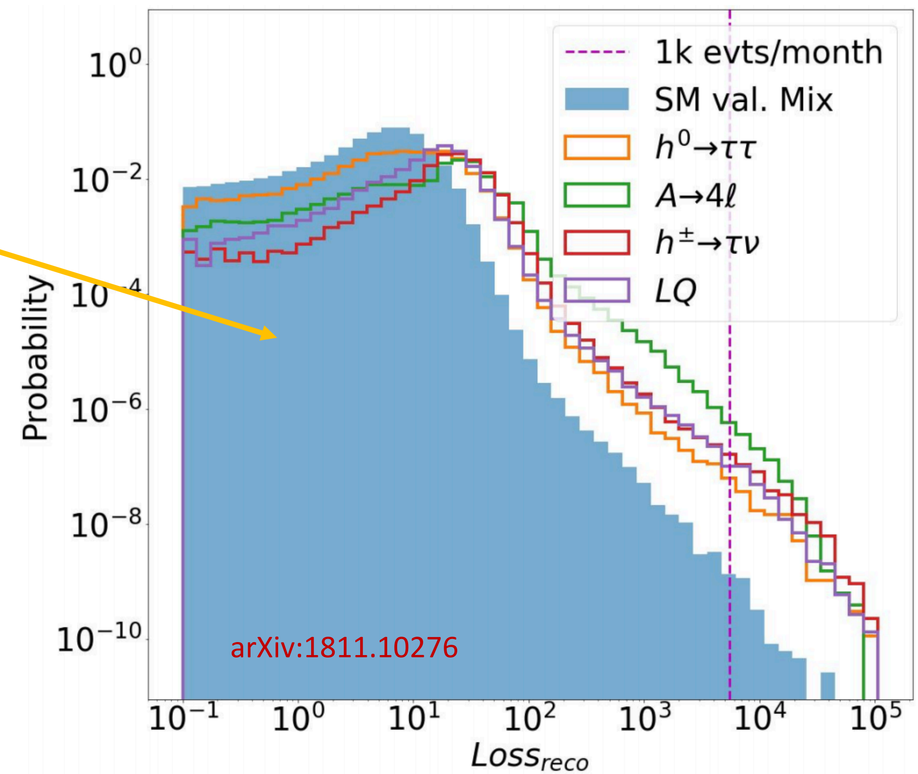


$$\mathcal{L}(x;\theta,\phi) = \mathbb{E}_{q_\phi(z|x)}[\log p_\theta(x|z)] - D_{KL}[q_\phi(z|x) \parallel \mathrm{p}(z)]$$

Anomaly Detection with VAEs

- What if we don't know what a new signal looks like?
  - Goal: Find rare events that differ from "standard"



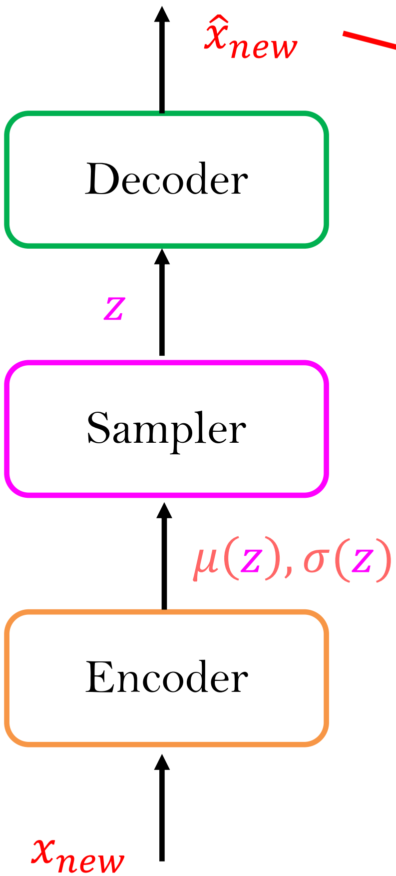


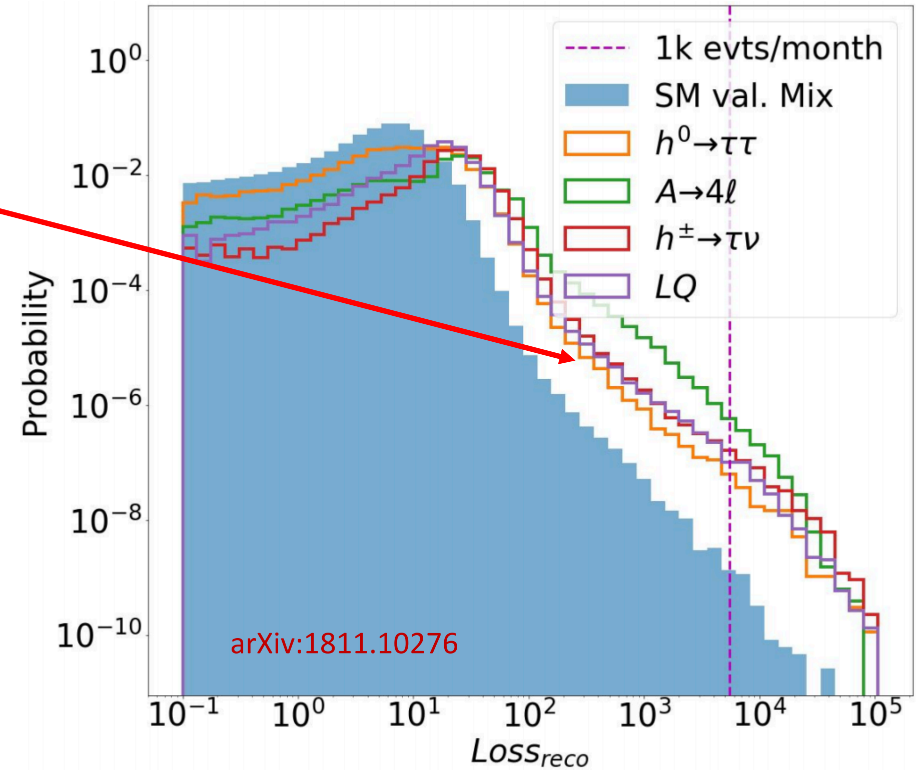
Learn what is "normal" with VAE

$$\mathcal{L}(x;\theta,\phi) = \mathbb{E}_{q_\phi(z|x)}[\log p_\theta(x|z)] - D_{KL}[q_\phi(z|x) \parallel p(z)]$$

Anomaly Detection with VAEs

- What if we don't know what a new signal looks like?
  - Goal: Find rare events that differ from "standard"



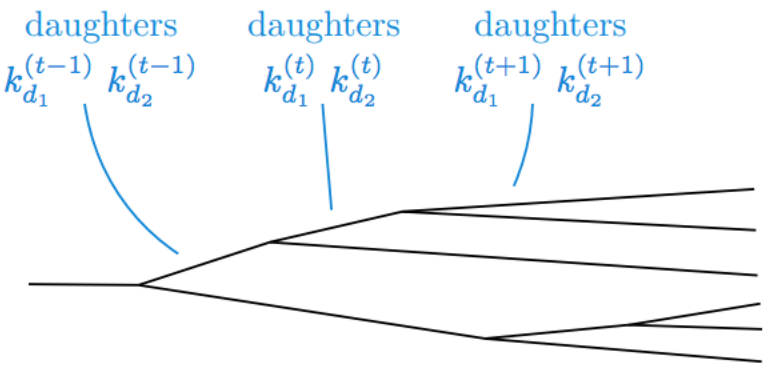


Learn what is "normal" with VAE

Look for anomalies as poorly reconstructed events

$$\mathcal{L}_{reco}(x_{new}; \theta, \phi) = \mathbb{E}_{q_\phi(z|x_{new})}[\log p_\theta(x_{new}|z)]$$

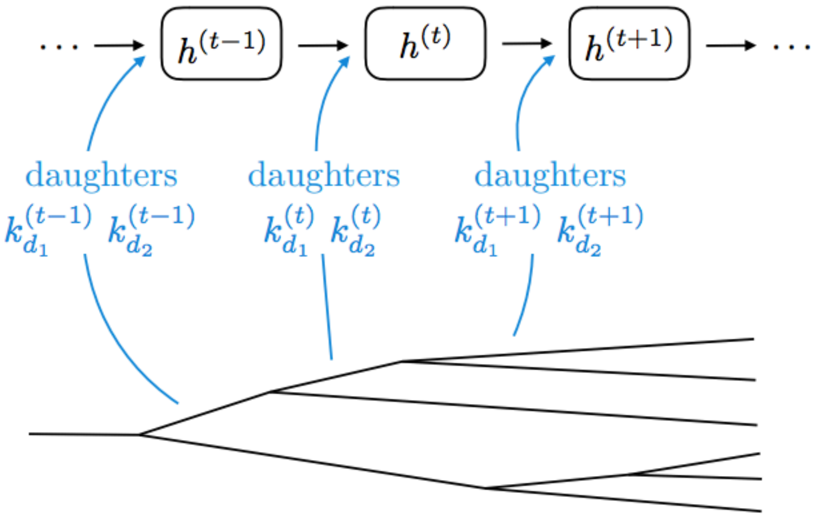
Physics Driven Generative Modeling: JUNIPR

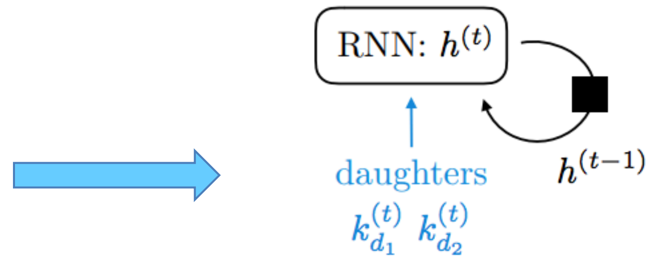


$$P_{\text{jet}}(\{p_1, \ldots p_n\}) = \left[\prod_{t=1}^{n-1} P_t\left(k_1^{(t+1)}, \ldots, k_{t+1}^{(t+1)} \middle| k_1^{(t)}, \ldots, k_t^{(t)}\right) \right] \times P_n\left(\text{end} \middle| k_1^{(n)}, \ldots, k_n^{(n)}\right).$$

arXiv:1804.09720

Physics Driven Generative Modeling: JUNIPR

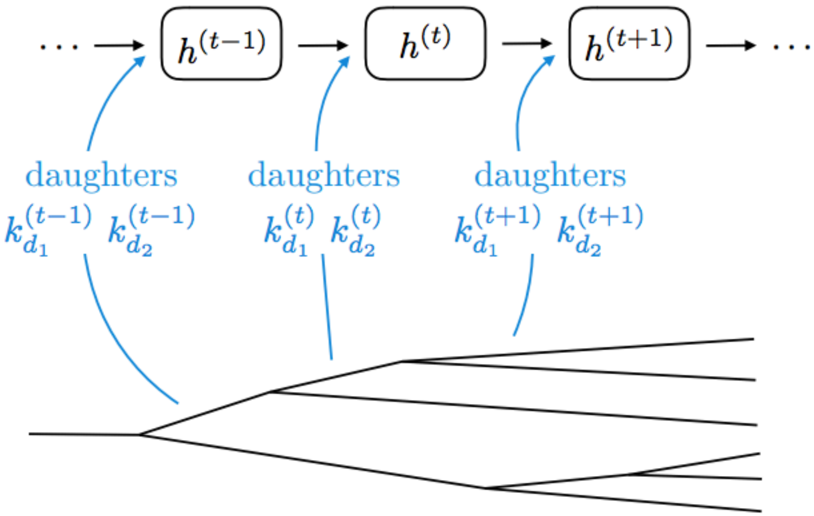


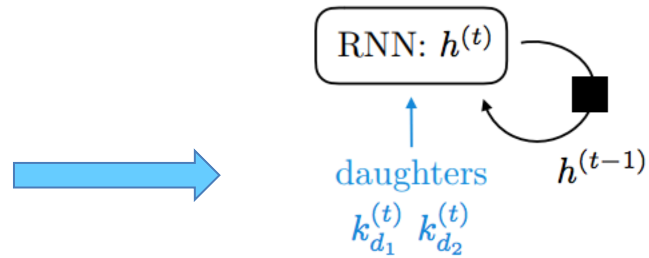


$$P_{\text{jet}}(\{p_1,\ldots p_n\}) = \left[\prod_{t=1}^{n-1} P_t\left(k_1^{(t+1)},\ldots,k_{t+1}^{(t+1)} \middle| k_1^{(t)},\ldots,k_t^{(t)}\right)\right] \times P_n\left(\text{end} \middle| k_1^{(n)},\ldots,k_n^{(n)}\right).$$

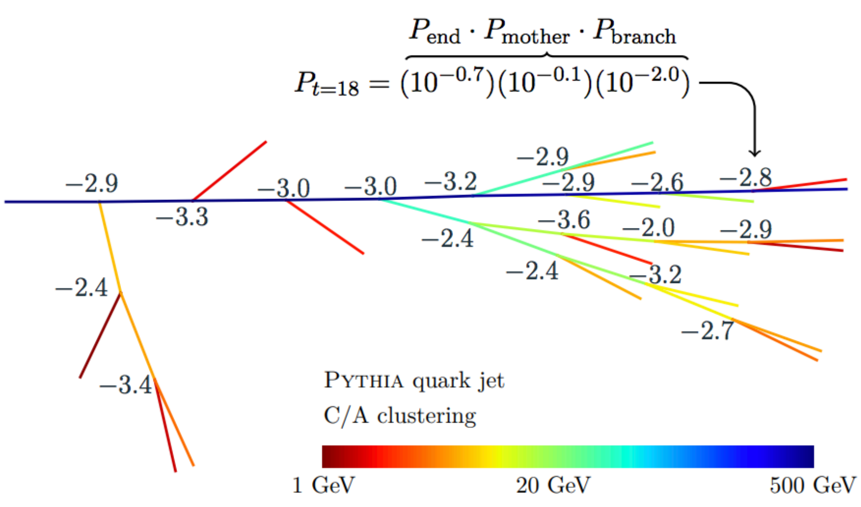
arXiv:1804.09720

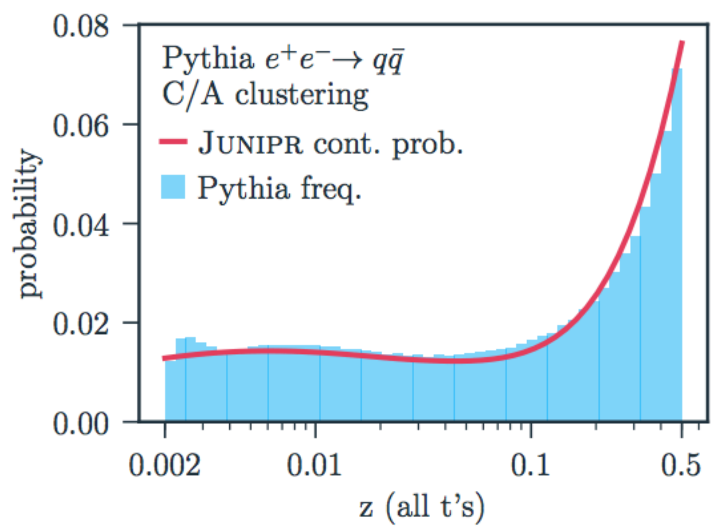
Physics Driven Generative Modeling: JUNIPR





$$P_{\text{jet}}(\{p_1,\ldots p_n\}) = \left[\prod_{t=1}^{n-1} P_t\left(k_1^{(t+1)},\ldots,k_{t+1}^{(t+1)} \middle| k_1^{(t)},\ldots,k_t^{(t)}\right)\right] \times P_n\left(\text{end}\middle|k_1^{(n)},\ldots,k_n^{(n)}\right).$$





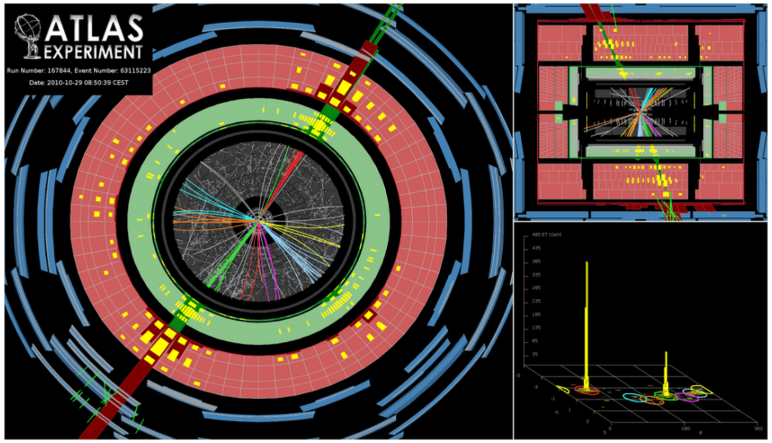
arXiv:1804.09720

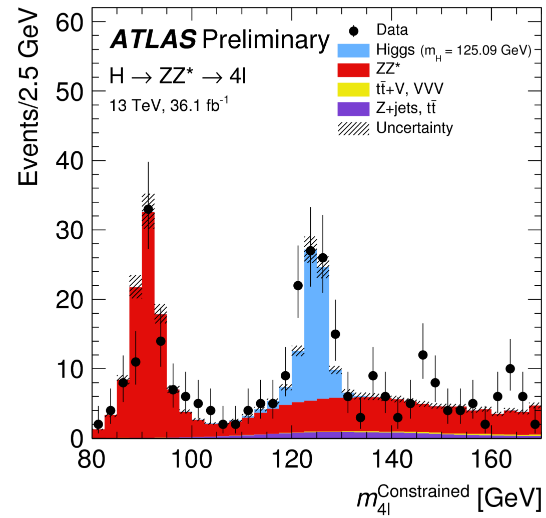
Maximizing the Use of Simulators

- Simulator already encapsulate our physics knowledge → Our best generative model
- How can we better use it in the inference pipeline?
- Can we extract more information for inference?
- Can we learn to approximate it?

Extracting Insights

Want $r(x|\theta) = \dfrac{p(x|\theta_0)}{p(x|\theta_1)}$ for hypothesis testing





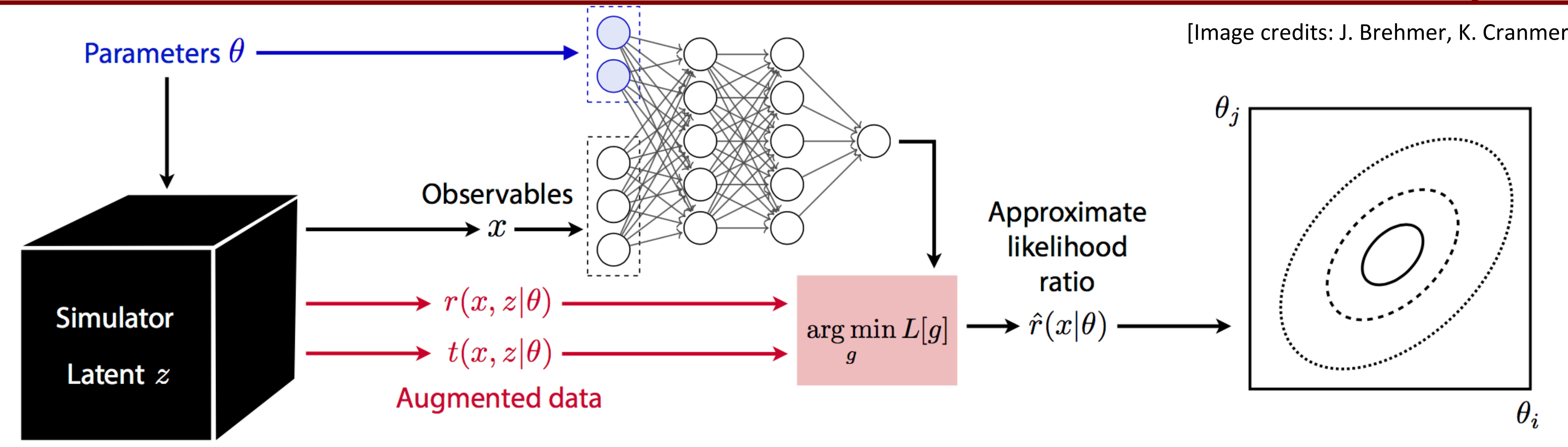
- High dimensional data x
- $p(x|\theta) = \int p(x, z|\theta) dz$ is intractable
- Simulator: can only produce samples from $p(x, z|\theta)$

- Summary statistics based on physical insights
- Density estimation in low dim. i.e. with histograms
- Challenging to design good statistic, often lose information

Mining Gold for Likelihood-Free Inference

arXiv:1805.12244

[Image credits: J. Brehmer, K. Cranmer]

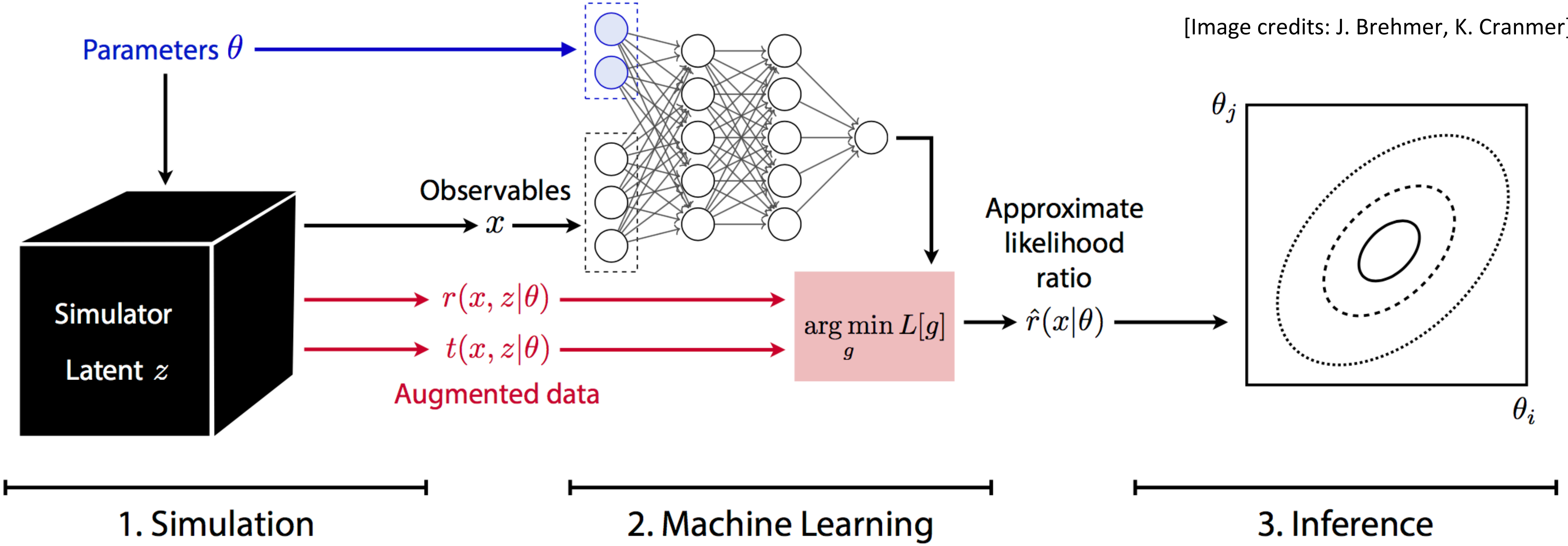


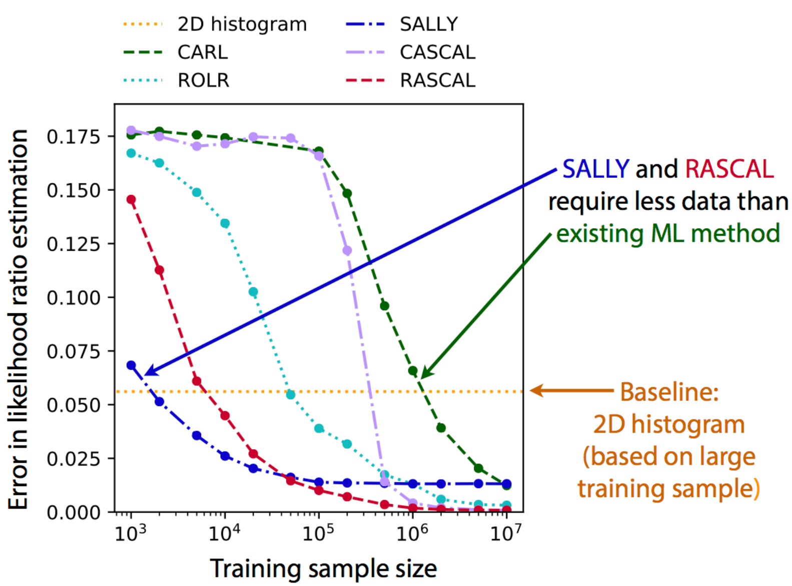
1. Simulation	2. Machine Learning	3. Inference
"Mining gold": Extract additional information from simulator	Use this information to train estimator for likelihood ratio	Limit setting with standard hypothesis tests

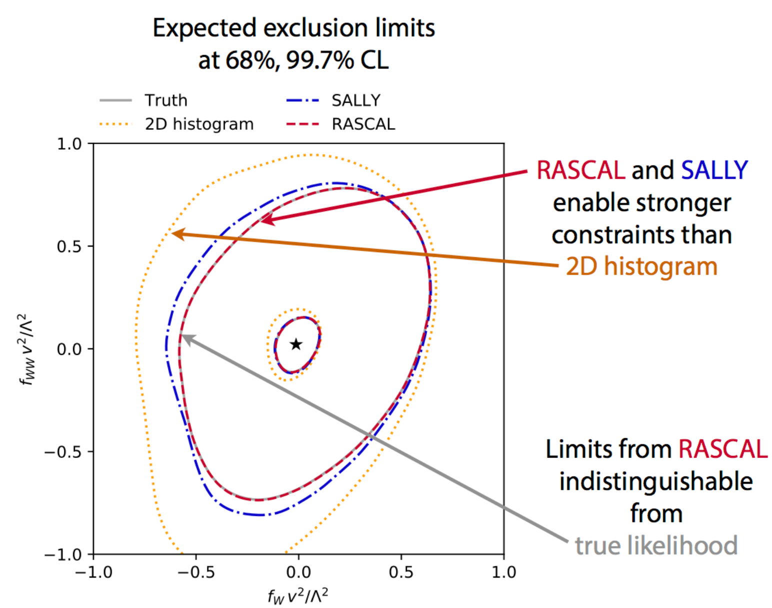
- Idea: extract ratio $r(x, z|\theta) = \dfrac{p(x,z|\theta_0)}{p(x,z|\theta_1)}$
- ML regression can approximate ratio: $\hat{r}(x|\theta)$

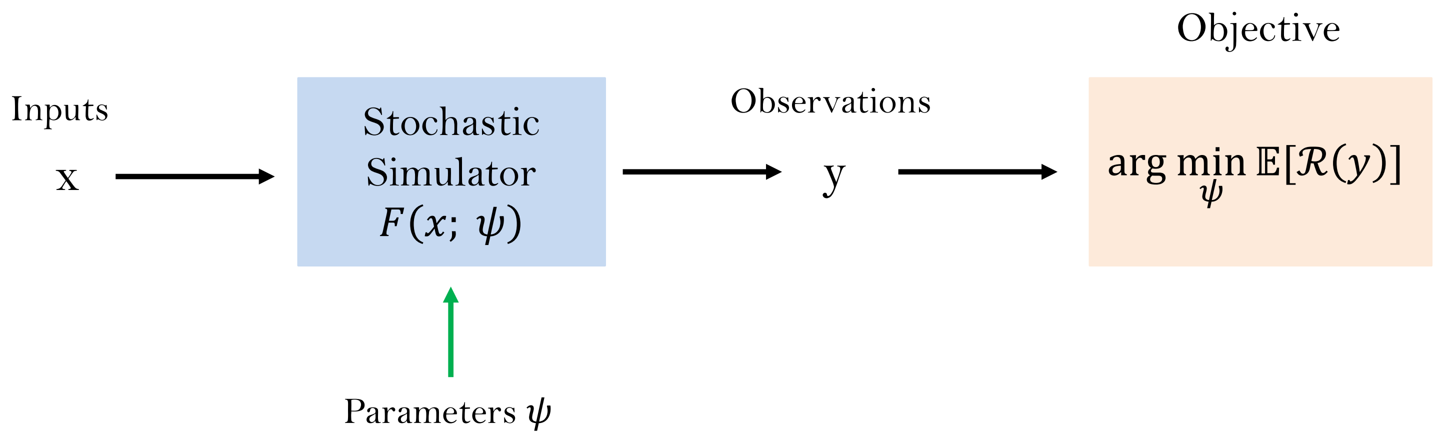
Mining Gold for Likelihood-Free Inference

[Image credits: J. Brehmer, K. Cranmer]



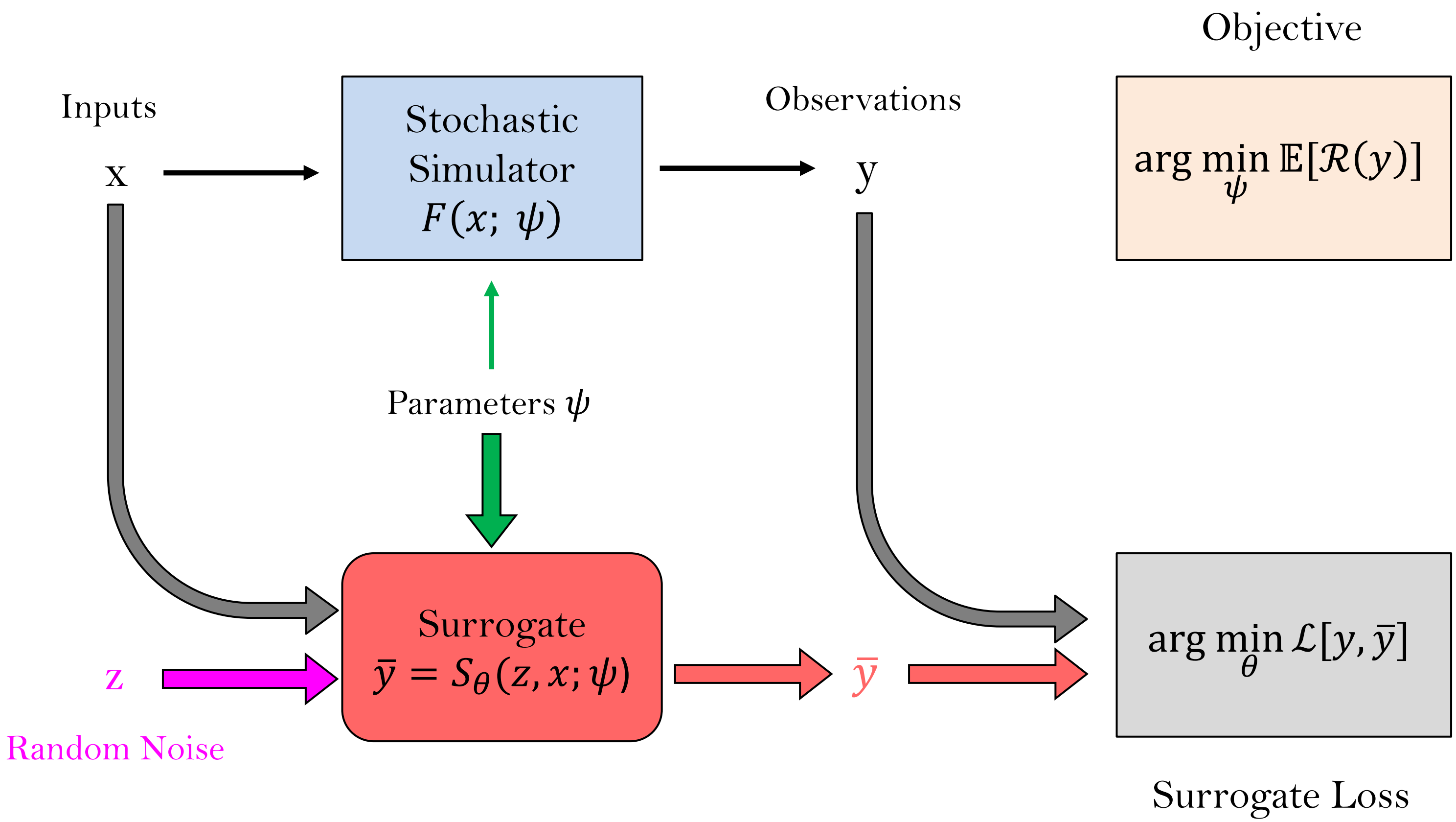






- Goal: Optimize simulator parameters to minimize objective
- Can we approximate the simulator directly?

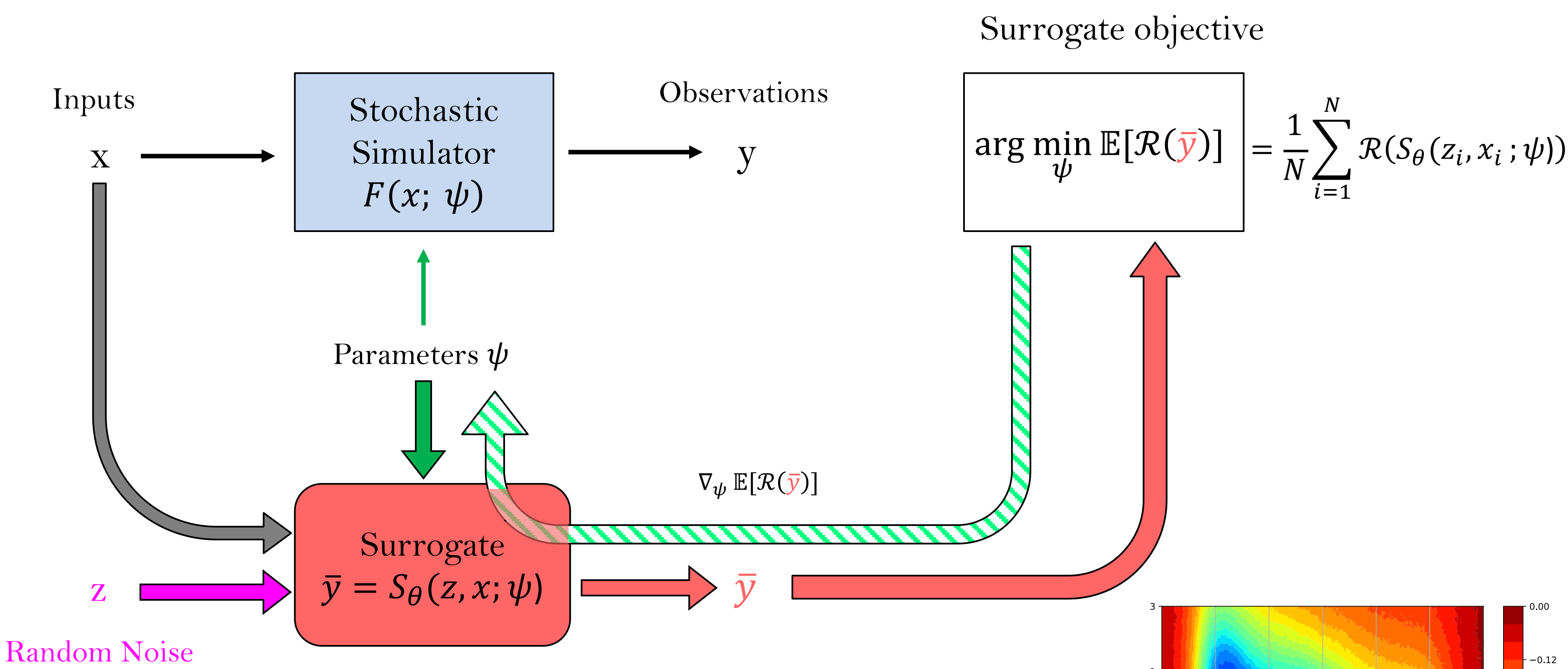
Differentiable Surrogates

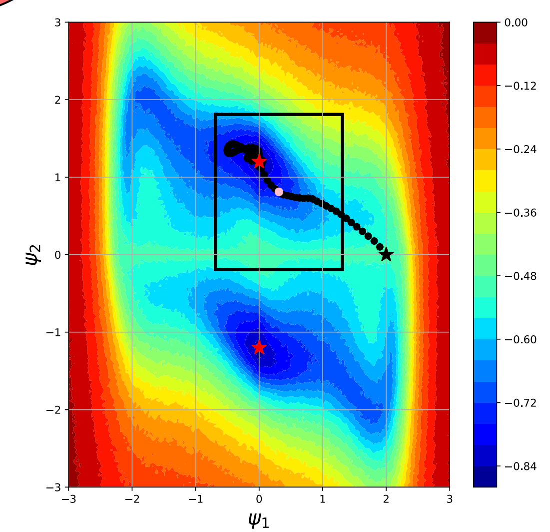


- Train parameterized generative surrogate model S, i.e. GAN or flow, to approximate $F(x;\ \psi)$
  - Can account for stochastic nature of F
  - Samples from surrogate can be differentiated!

Differentiable Surrogates

S. Shirbikov, V. Belavin, M. K. ,
A. Baydin, A. Ustyuzhanin



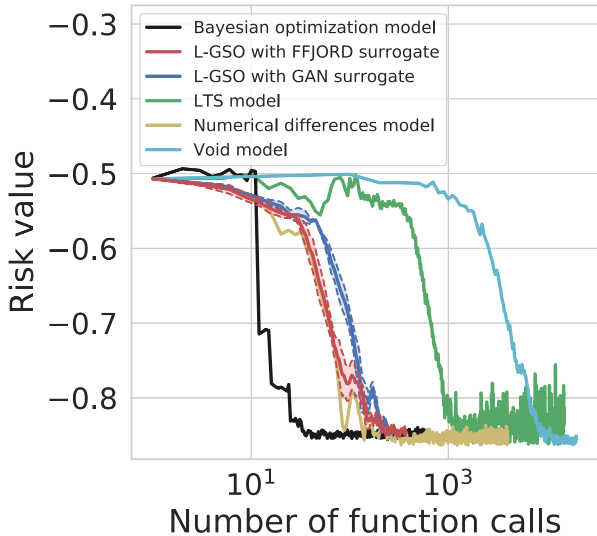


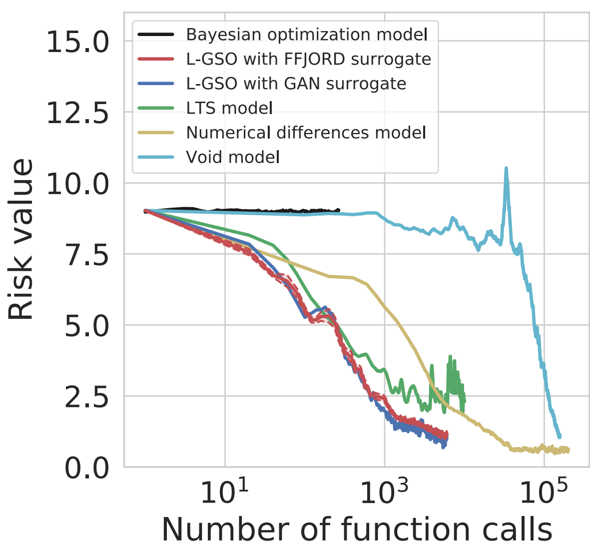
- Optimize objective with gradient descent using trained surrogate to produce differentiable samples

Optimization In Action

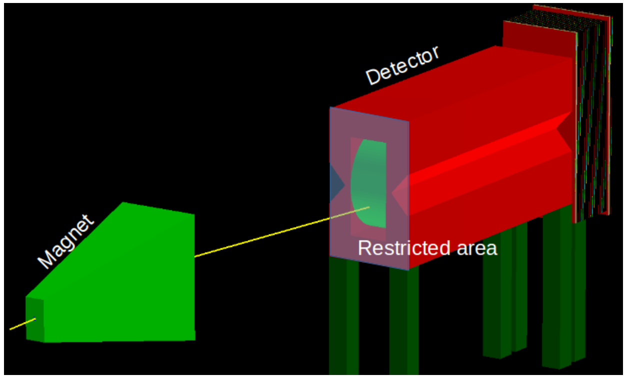
S. Shirbikov, V. Belavin, M. K. ,
A. Baydin, A. Ustyuzhanin

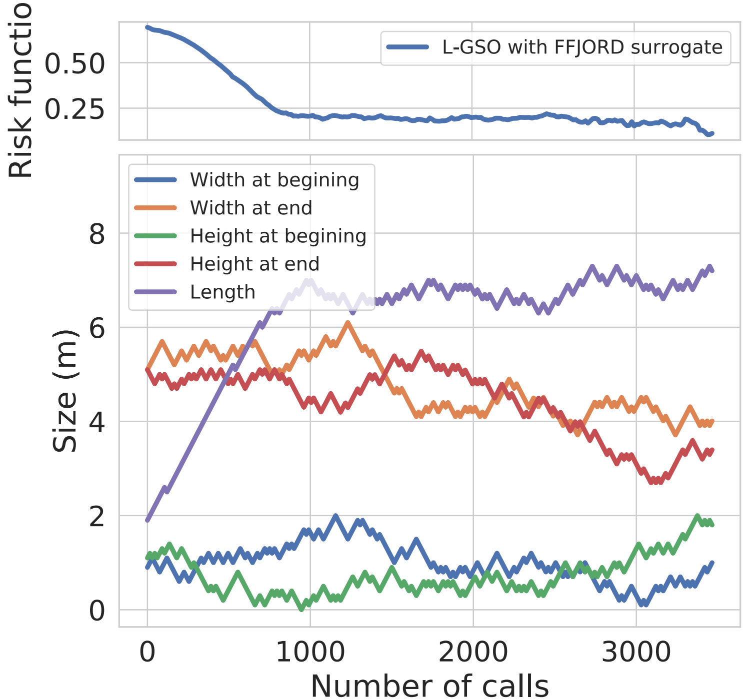
Toy examples

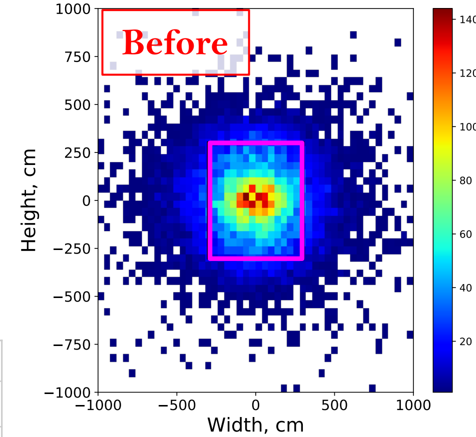


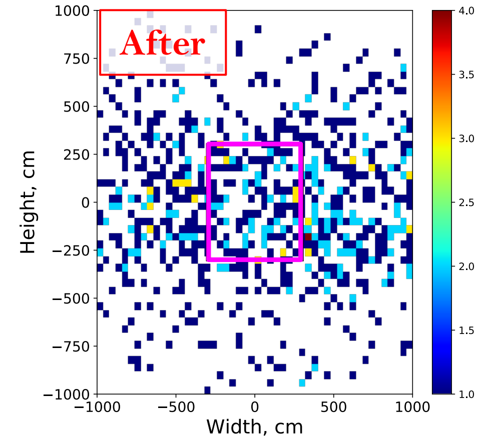


Physics Example









Conclusions

- The HEP data analysis pipeline is greatly advancing with Machine Learning methods, due in large part to Massive datasets and high fidelity simulation

- Incorporating physics insights and making the best use of our simulators in ML models can help push performance, ease the learning tasks, and increase interpretability

- What is the best way to introduce these physics insights into our models?

- How can we best use the "intractable" physics knowledge we have put into the high fidelity simulators?

Backup

Pipeline

Low-latency decision making

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

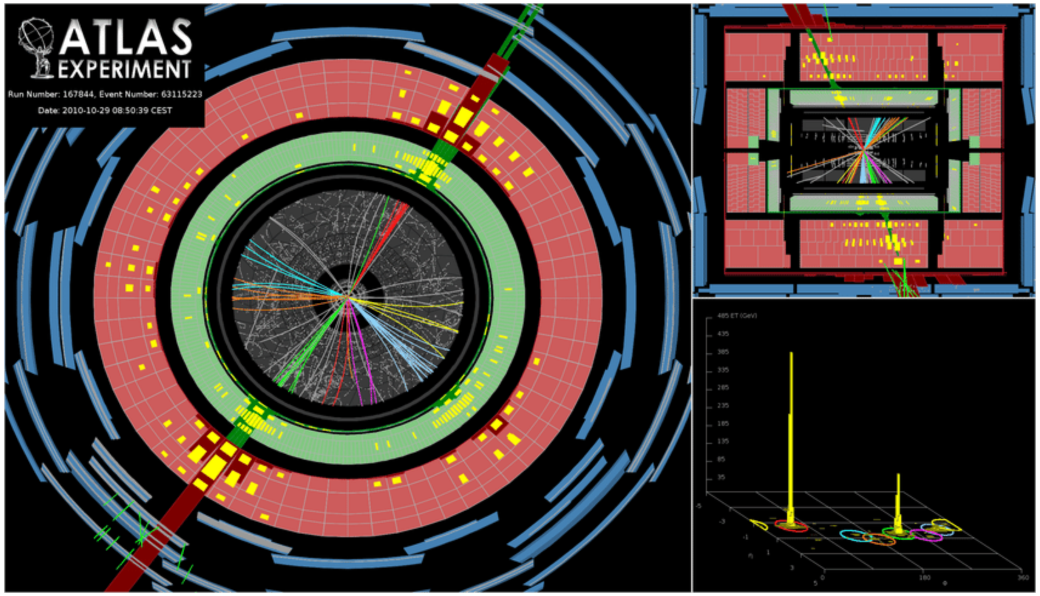
Infer properties of collision
Classification of interesting collisions

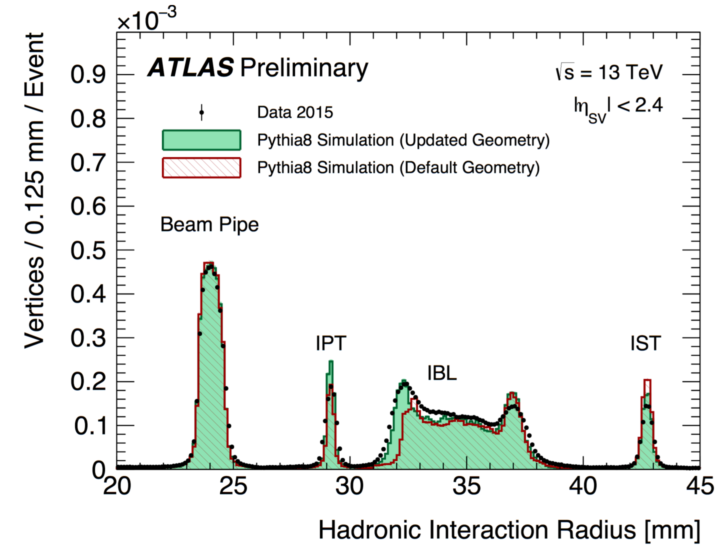
Event Reconstruction and Selection

Hypothesis testing / Measurement

Latent parameter
estimation

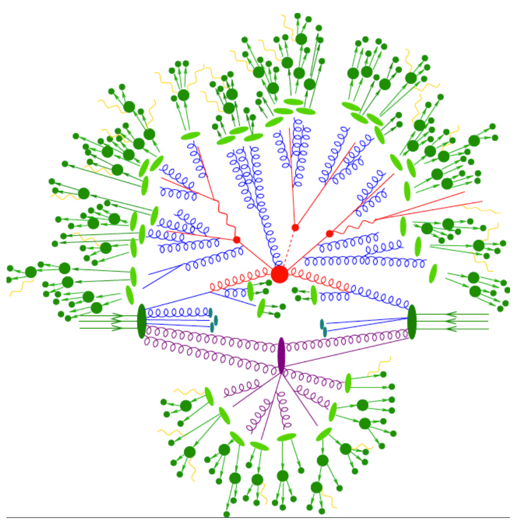
High Fidelity Simulation



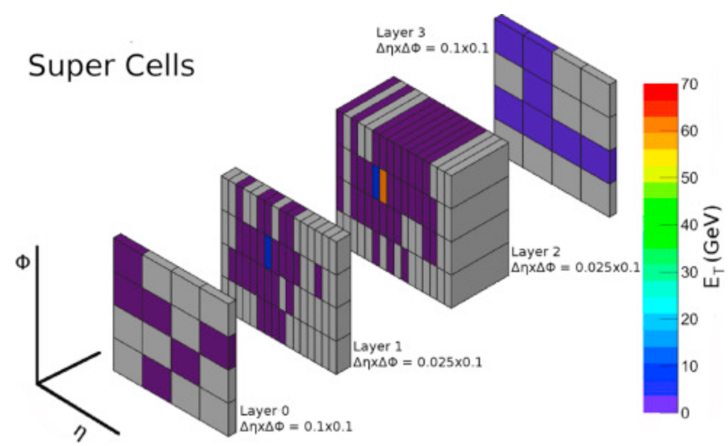


$p(particles \mid interation\ type)$

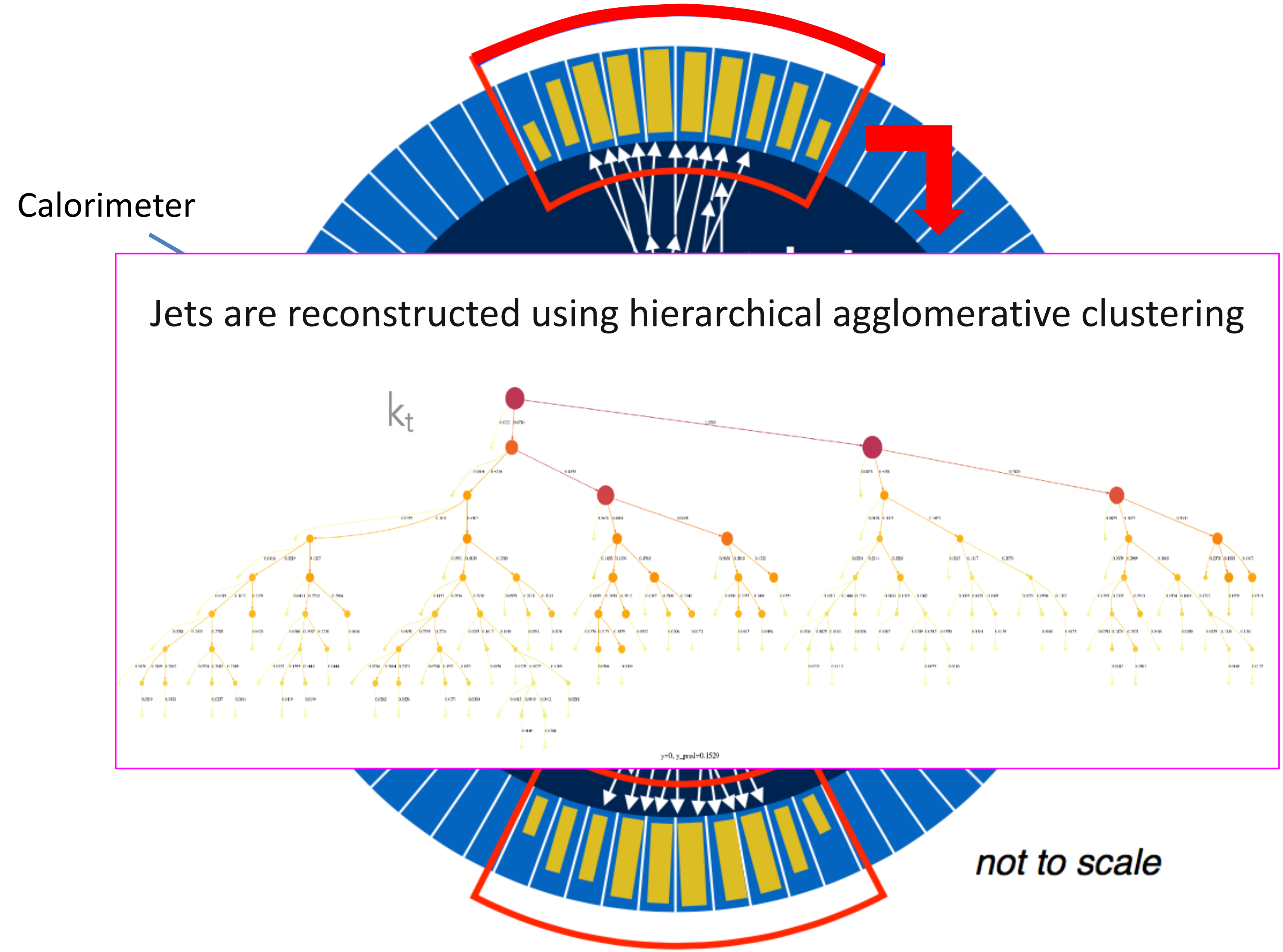
$p(detector\ signature \mid particle)$



[Sherpa]



Jets at the LHC

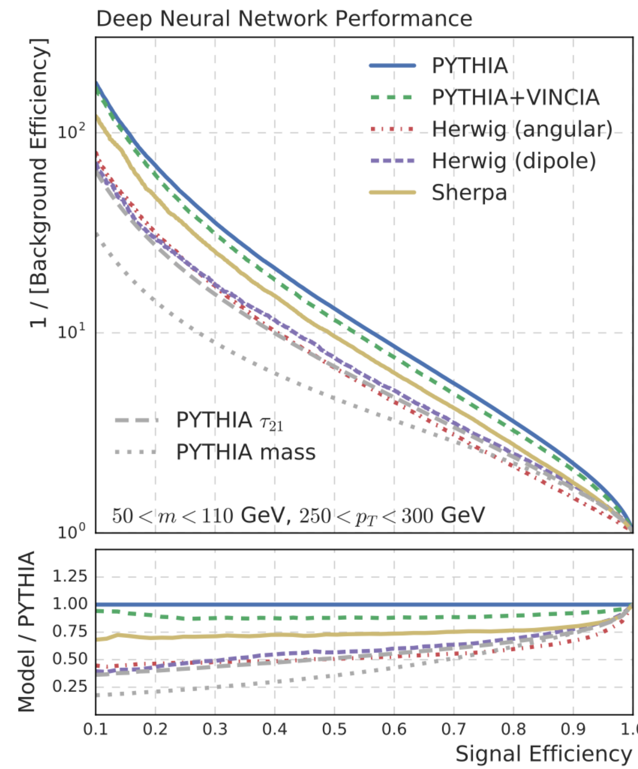


Jets are reconstructed using hierarchical agglomerative clustering

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

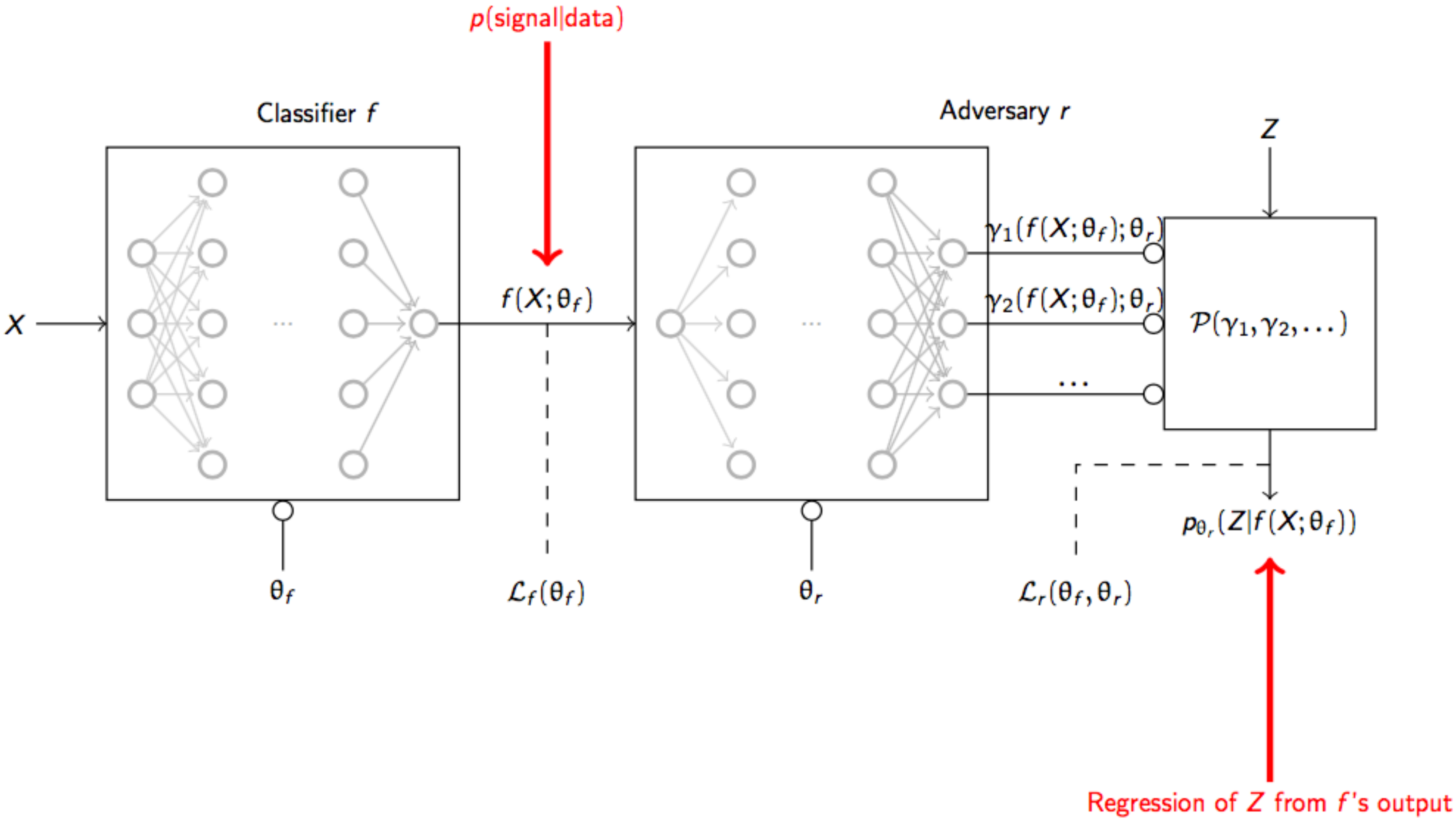
Enforcing Invariance

- Must consider ***Domain Shift***: Simulations don't perfectly capture data generating process
  - Hard to control how models learn and utilize information
  - Unwanted sensitivity to poorly modeled aspects of simulation
  - Potentially unwanted sculpting of key physics distributions like mass



- ***Idea***: Augment training of classifier to enforce invariance to changes in a variable Z

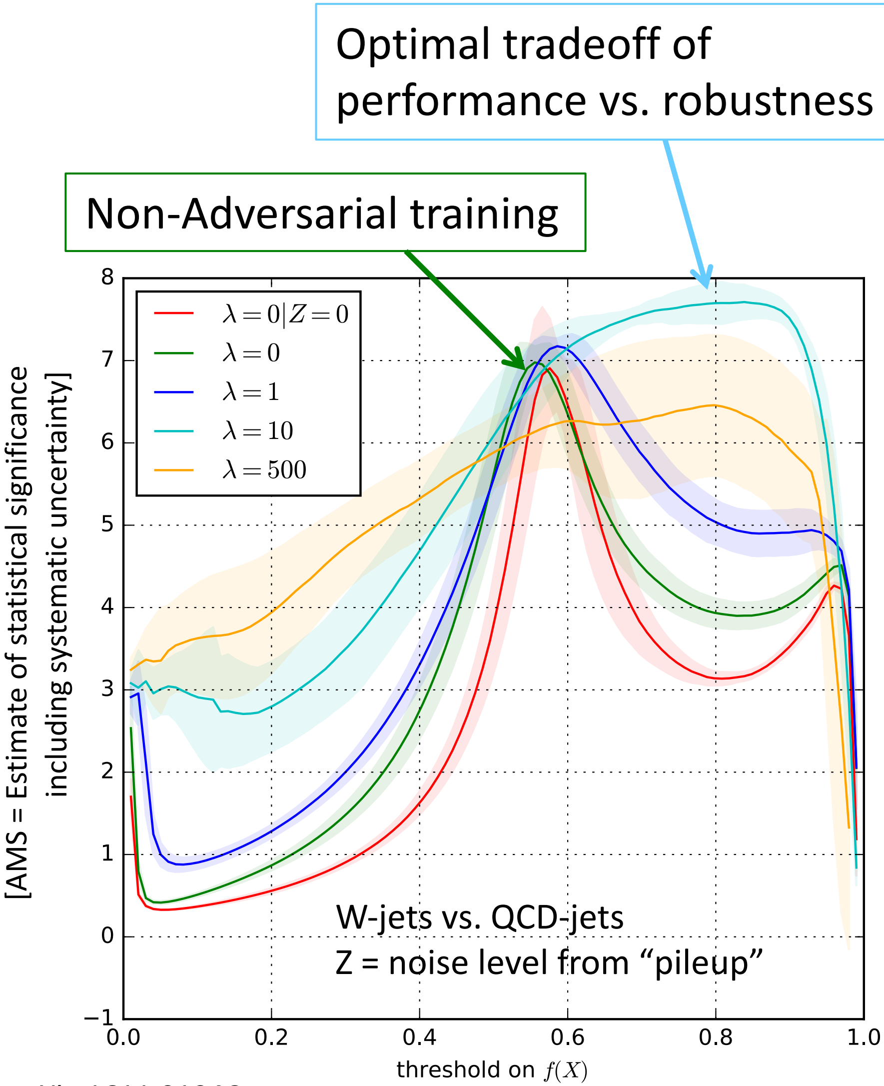
Adversarial Learning



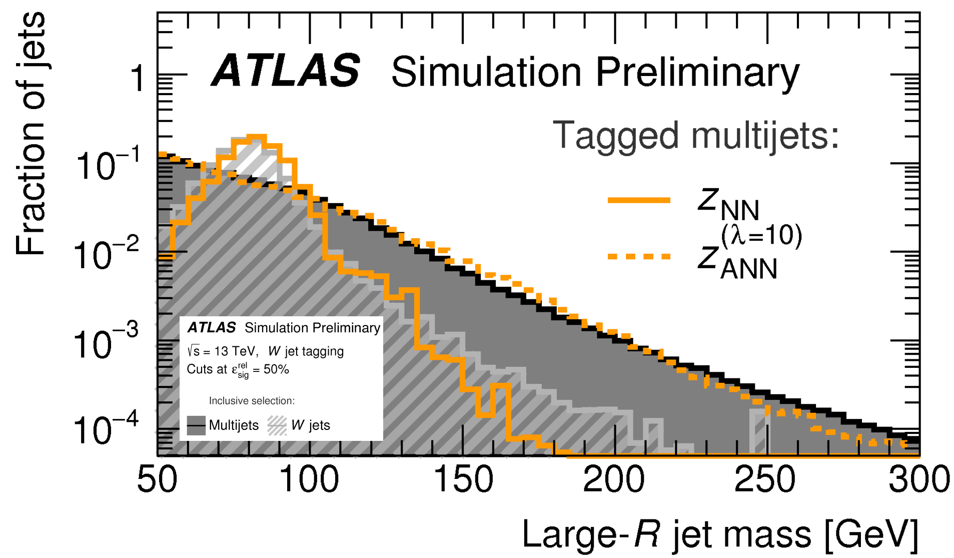
- Adversarial Approach:
  - Build loss that encodes performance of a classifier and and adversary
  - Classifier penalized when adversary does well at predicting Z

Learning to Pivot: Physics Examples

- Optimize statistical significance in presence of systematic uncertainty



arXiv:1611.01046



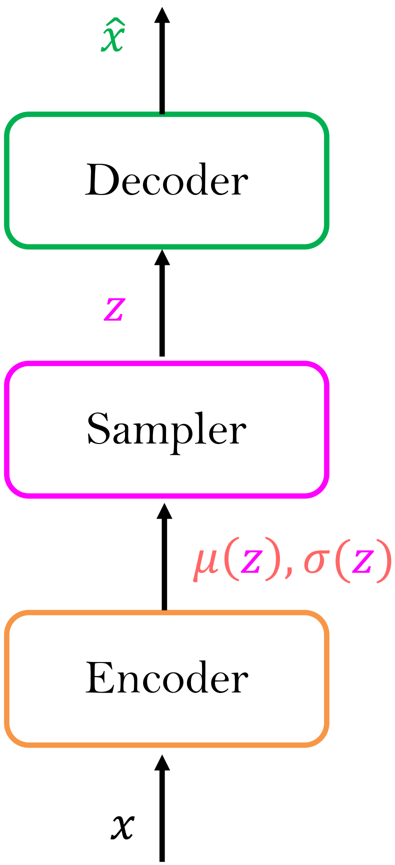
- Decorrelate a classifier from a chosen variable
  - For example, to not sculpt jet mass distribution with classifier

[arXiv:1703.03507]

Generative Models

[Images motivated by A. Canziani, R. Ng]

Variational AutoEncoder

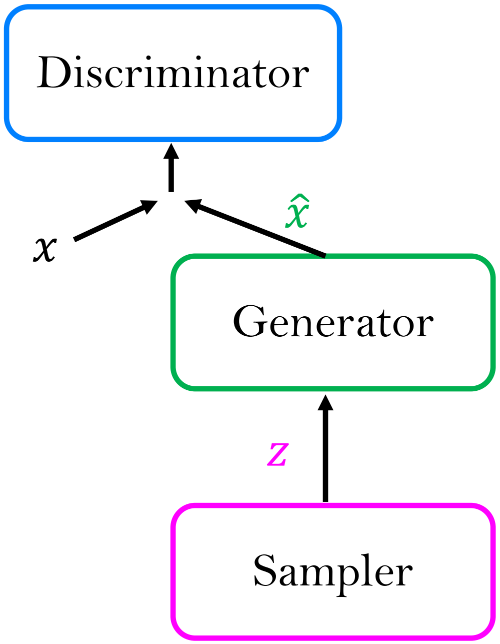


$$\mathcal{L}(x;\theta,\phi) = \mathbb{E}_{q_\phi(z|x)}[\log p_\theta(x|z)] - D_{KL}[q_\phi(z|x) \parallel \mathrm{p(z)}]$$

arXiv:1312.6114

arXiv:1401.4082

Generative Adversarial Network

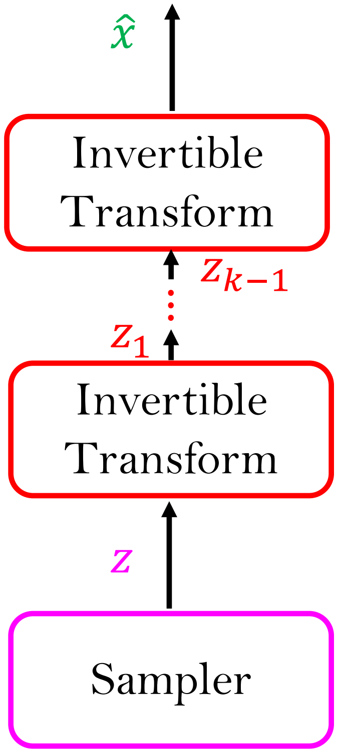


$$\min_G \max_D V(G,D)$$

$$V(G,D) = \mathbb{E}_{p_{data}(x)}[\log D(x)] + \mathbb{E}_{p_z(z)}[\log(1 - D(G(z)))]$$

arXiv:1406.2661

Flow



$$z_1 = f_1(z) \quad \rightarrow \quad p(z_1) = p(z) \left| det \frac{df_1}{dz} \right|^{-1}$$

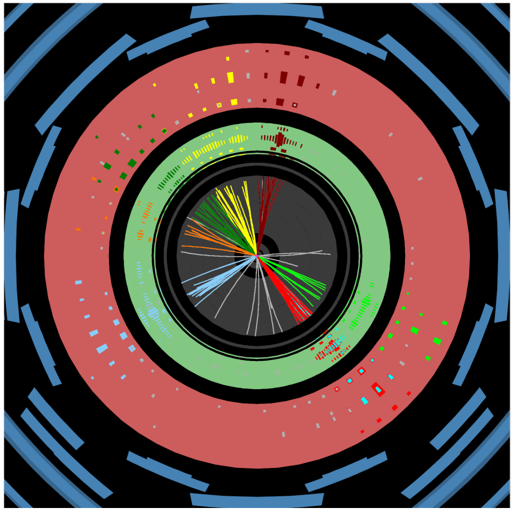
$$x = f_k \circ f_{k-1} \circ \cdots \circ f_1\ (z)$$

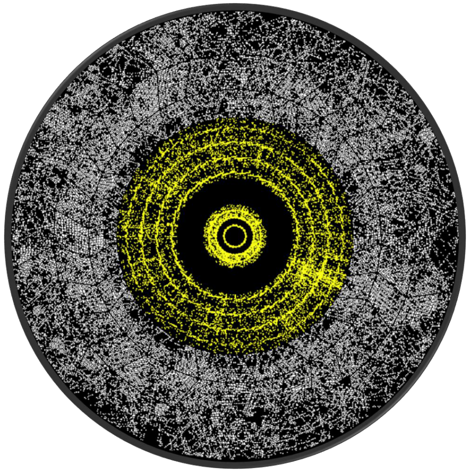
$$\log p(x) = \log p(z) - \sum_{i=1}^{k} \log \left| det \frac{df_i}{dz_i} \right|$$

arXiv:1410.8516

arXiv:1505.05770

Resource Constraints





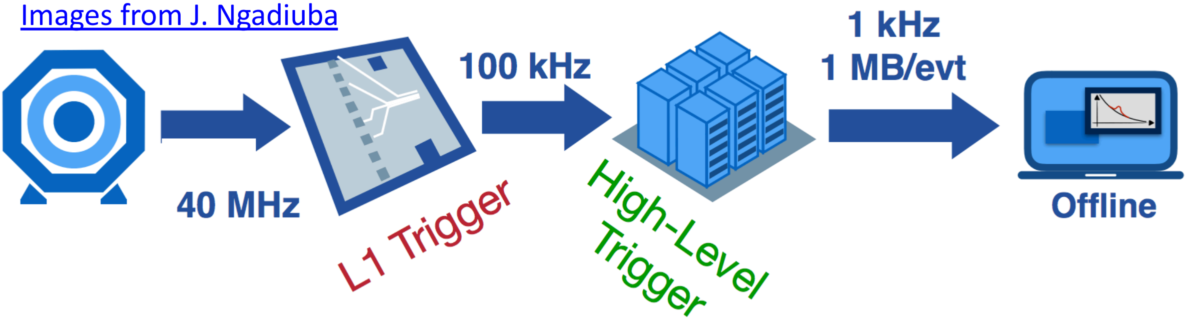
- Increased pileup at HL-LHC will push boundaries of our computational capabilities
  - Major challenges in triggering, large scale simulation, and high multiplicity tracking
  - New tools and developments in ML may help address some of these challenges

- Simulation
  - Accurate but often costly simulation of particle interactions with material, that produces sample and not analytic P(energy deposits | particle)
  - *ML approach*: Generative models to learn data distribution, p(x), and produce samples?

- Trigger
  - High performance algorithms early in trigger to reduce backgrounds for key signals?

Fast Data Acquisition with ML on FPGA

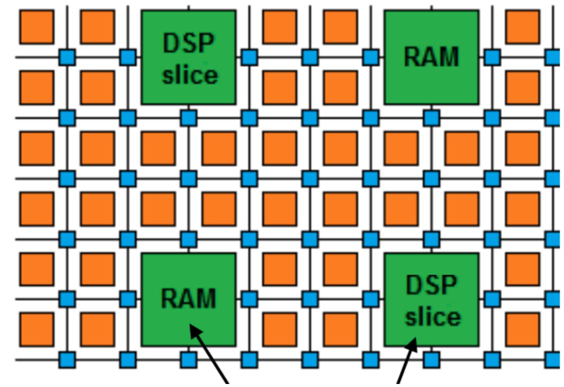
Images from J. Ngadiuba



Absorbs 100s Tb/s
Trigger decision to be made in O(μs)
Latencies require all-FPGA design

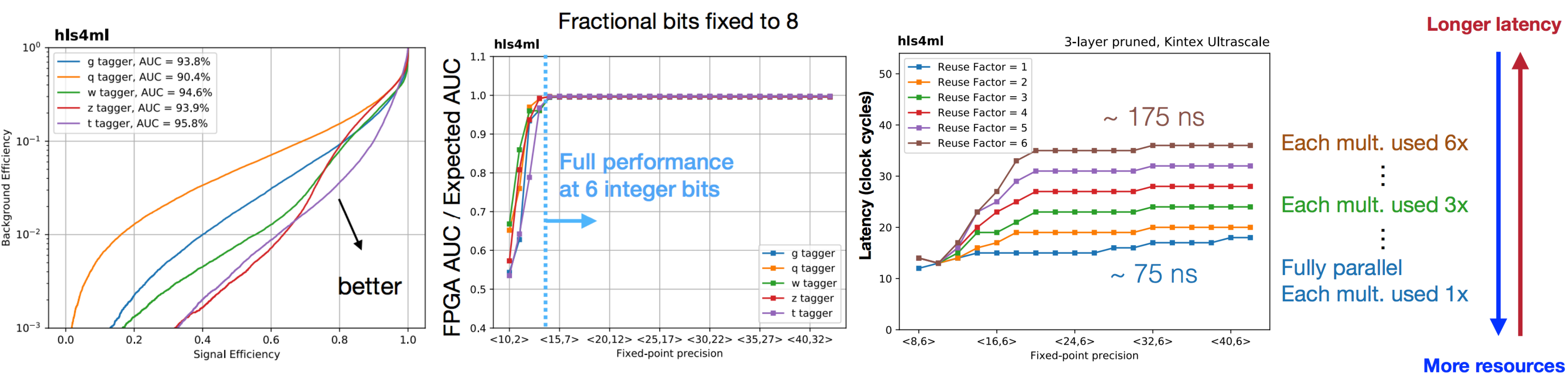
Computing farm for detailed analysis of the full event
Latency O(100 ms)

FPGA Diagram



- FPGAs are high speed, low power, and highly parallelizable
- Dedicated SW needed to efficiently and effectively port ML algorithms to FPGA
- Tuning resource usage, data precision, and model pruning needed to hit timing needs

- Example: Boosted jet tagging with HLS4ML software approach [arXiv:1804.06913]



Optimization of Toy Problems

S. Shirbikov, V. Belavin, M. K. ,
A. Baydin, A. Ustyuzhanin