ATLAS, data reduction, and epistemology:

a tour of some statistical claims in particle physics

Ryan Reece

ryan.reece@cern.ch

https://ucsc.academia.edu/RReece

http://reece.scipp.ucsc.edu

UNIVERSITY OF CALIFORNIA SANTA CRUZ

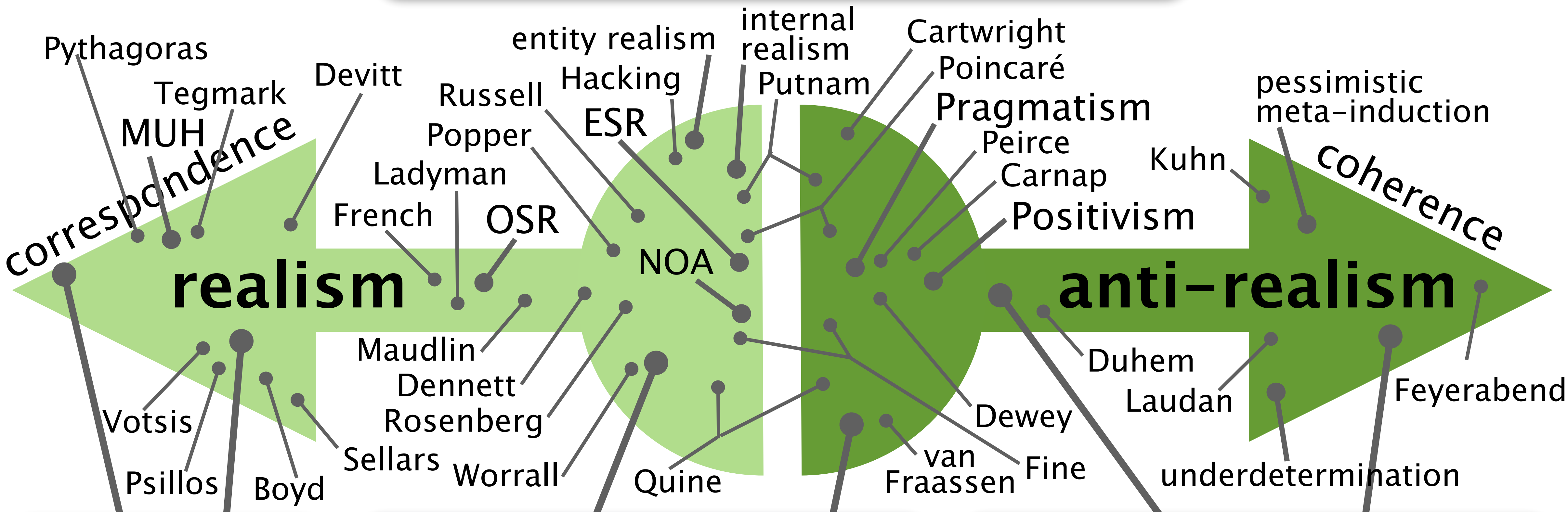
High performance computing centre in Stuttgart (HLRS)

November 29, 2016

Outline

1. Preface on scientific realism
2. Introduction to particle physics
3. Statistical inference and ATLAS data reduction
4. Implications of machine learning
5. Summary

philosophy of science



Naive Realism
The world I see is real. What are you all arguing about?

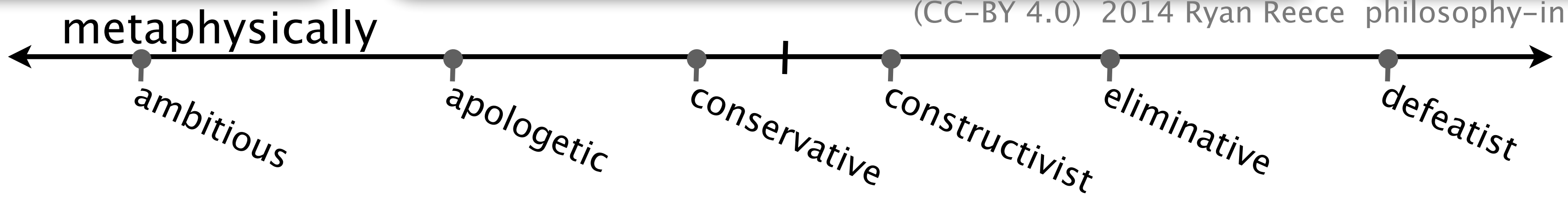
Structural Realism
Science has identified real patterns, relationships, and structures (at least within a regime) in nature.

Instrumentalism
Theoretical concepts may have use in predicting observations, but we have no ontological commitments to them.

Scientific Realism
Science makes real progress in describing real features of the world.

Constructive Empiricism
Science aims to give us theories which are empirically adequate, but does not justify metaphysical claims about reality.

Relativism
Social constructivism. Epistemological anarchism.



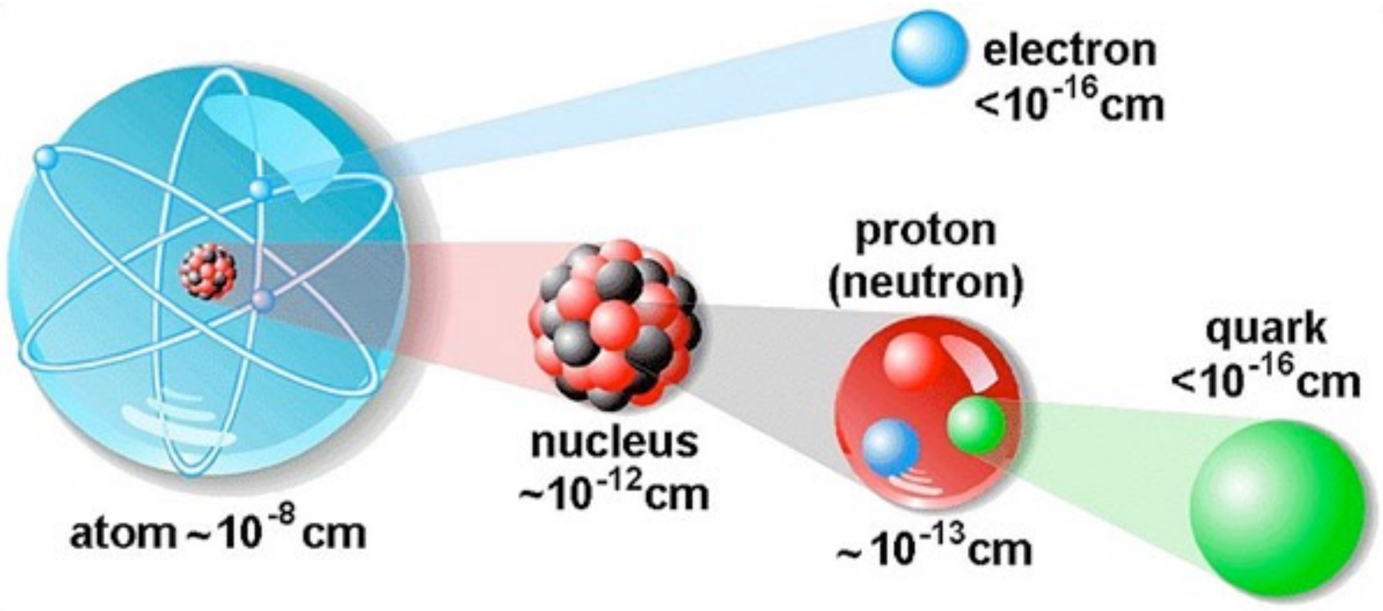
from: http://philosophy-in-figures.tumblr.com/

Introduction to particle physics

Particle Physics

Fundamental questions
of particle physics:

1. *What is matter?*
2. *How does it interact?*

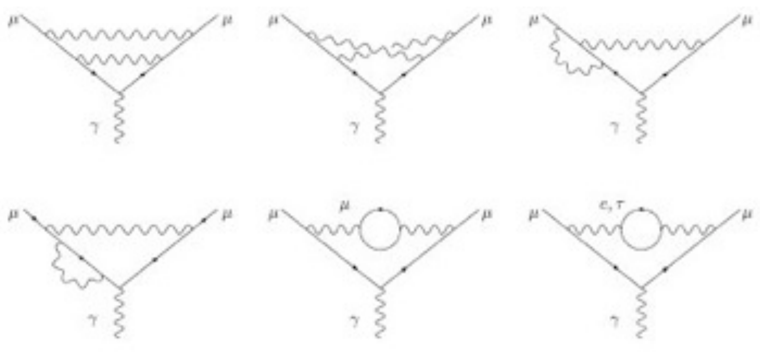


Four fundamental forces at low energies:

1. **Gravity** - very weak, no complete quantum theory
2. **Electromagnetism** - binds atoms, chemistry
3. **Strong force** - nuclear range, binds nuclei
4. **Weak force** - nuclear range, radioactivity, solar fusion

Quantum Field Theory (QFT)

- Every type of matter/energy has a corresponding field.
- In QFT, *fields* are (effectively) what is fundamental, and particles are quantized and often localized excitations in the fields.
- To satisfy relativity, they are the representation of the Poincare group: scalars, vectors, spinors, tensors.
- Non-trivial aspects of QFT have been tested to better than a part per million, e.g. the anomalous magnetic moments of electrons and muons.
- Very impressively, empirically adequate: arguably best tested science.



a_μ (exp) = 11 659 208 (6) × 10^{-10} (0.5 ppm)

analogy from Kyle Cranmer (NYU)

The Standard Model

- ***Gauge symmetries*** determine the character of the forces between fermion fields through exchanging gauge bosons.
- Bosons and chiral fermions develop mass terms that still preserve the gauge symmetries of the Lagrangian through the ***Higgs mechanism*** (proposed in 1964).
- The SM gauge group is

$$\mathbf{SU(3)_C \times SU(2)_L \times U(1)_Y}$$

Electroweak force

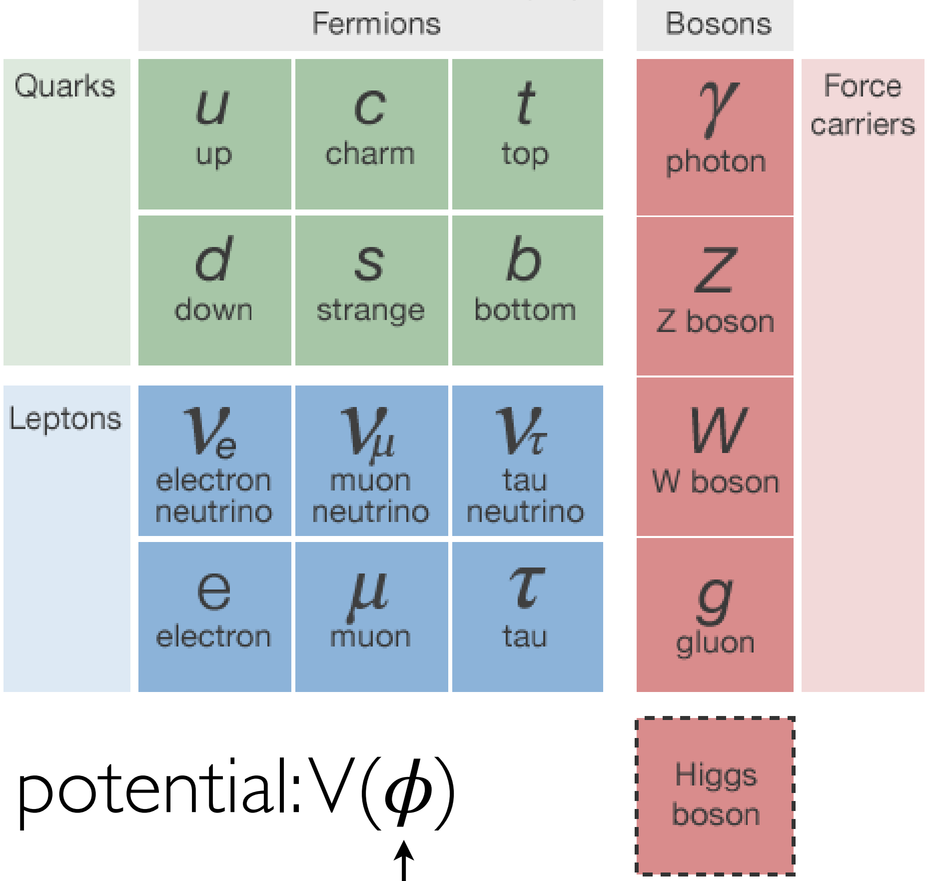
⇓

Strong force

⇓ *Higgs mechanism, EW symmetry breaking*

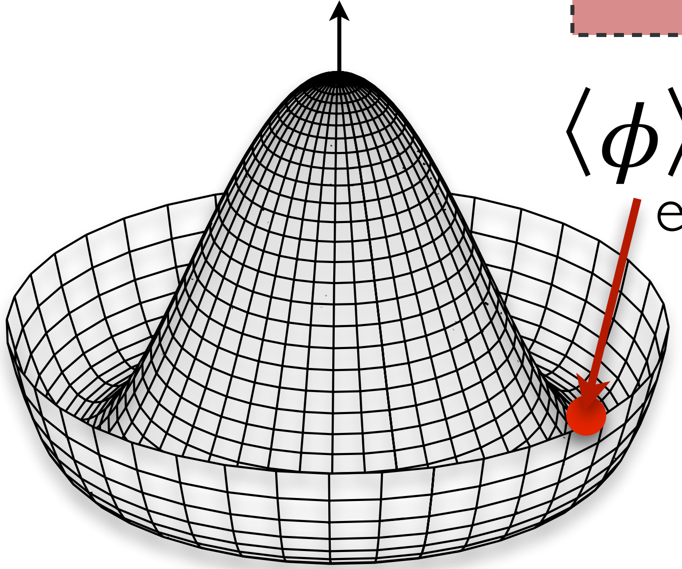
Electromagnetic + weak forces

Glashow, Salam, Weinberg (Nobel Prize 1979)



	Fermions			Bosons	
Quarks	u up	c charm	t top	γ photon	Force carriers
	d down	s strange	b bottom	Z Z boson	
Leptons	ν_e electron neutrino	ν_μ muon neutrino	ν_τ tau neutrino	W W boson	
	e electron	μ muon	τ tau	g gluon	
				Higgs boson	

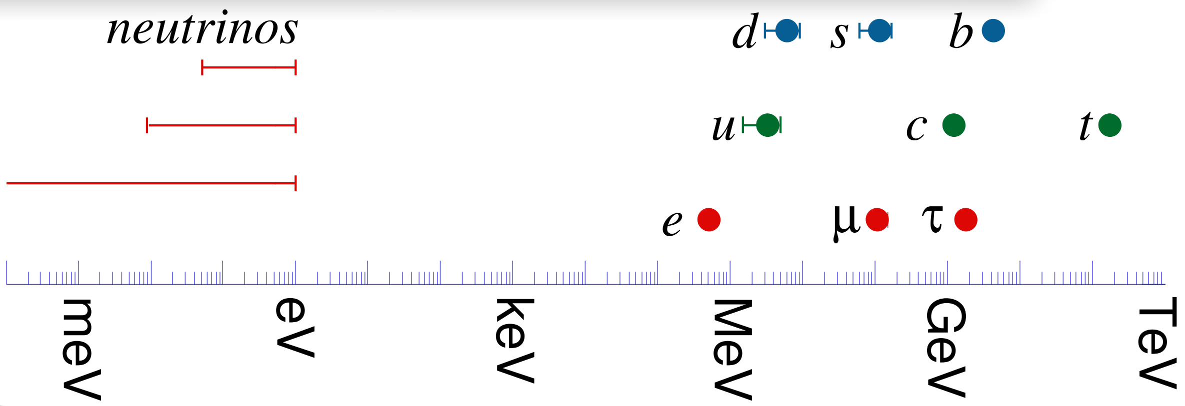
Higgs potential: $V(\phi)$

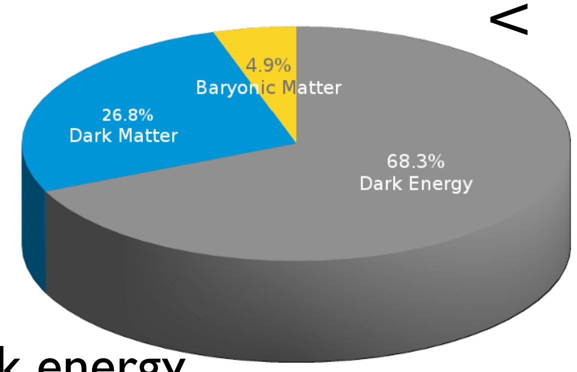


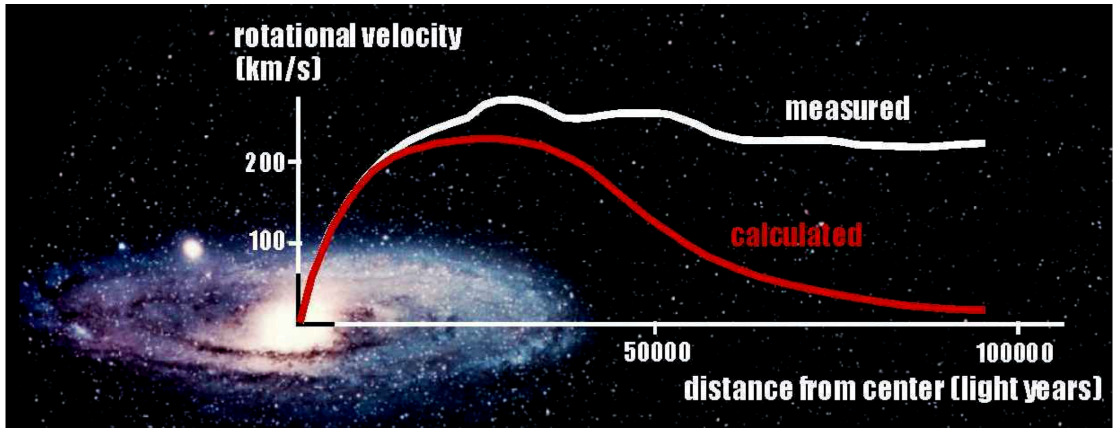
$\langle \phi \rangle$ vacuum expectation value

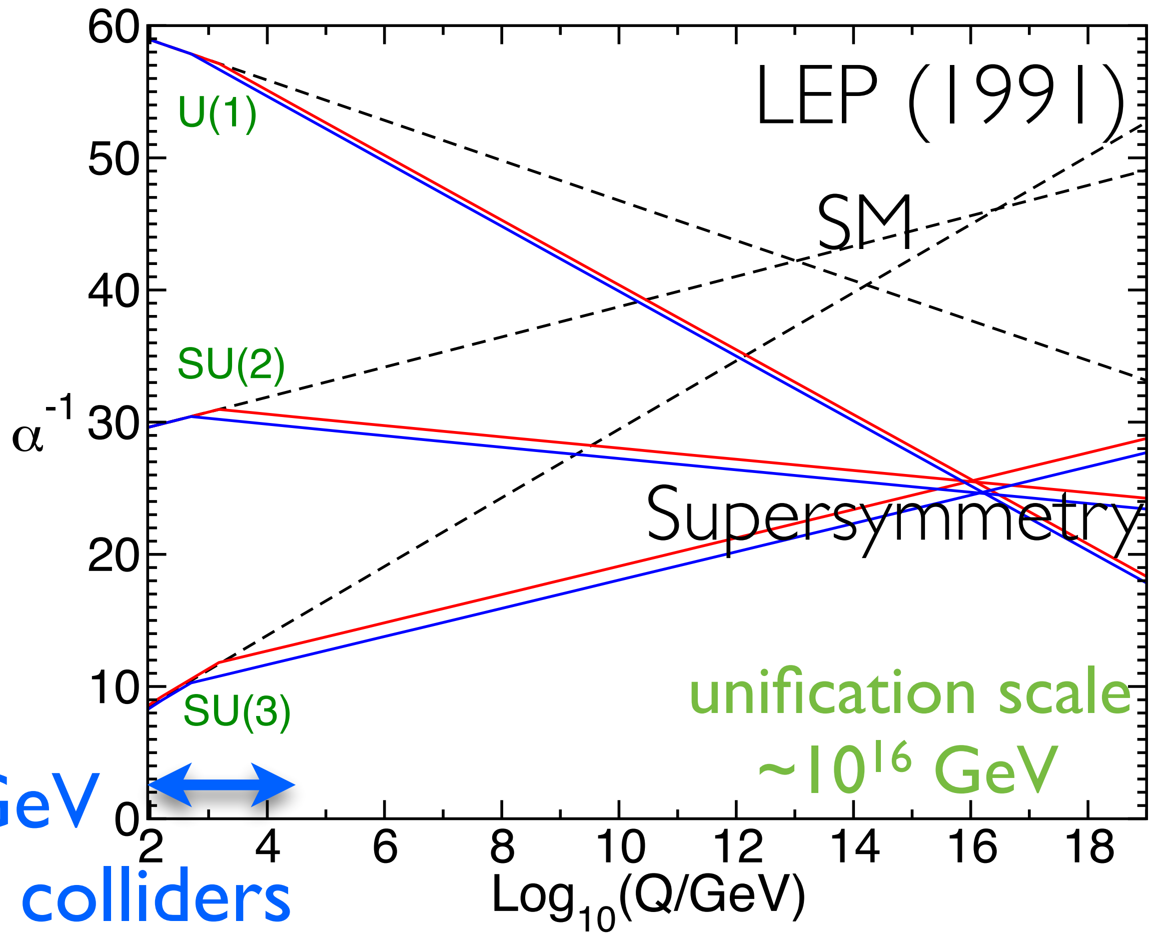
Unanswered problems in particle physics

- Ad hoc features
    - Why SU(3)xSU(2)xU(1) ?
    - Neutrino mixing and masses (Dirac or Majorana)
    - Matter-antimatter asymmetry
    - Strong CP-problem
- Dark matter and dark energy
    - 5% SM, 27% dark matter, 68% dark energy
- Hierarchy problem(s)
    - m_{Higgs} vs m_{Planck},
    - quark masses range: 10^5, leptons: 10^9
- Fine-tuning:
    - EW-scale, flatness problem, vacuum stability, etc.
- Unification? Supersymmetry?
- Why did the early universe have such low entropy?





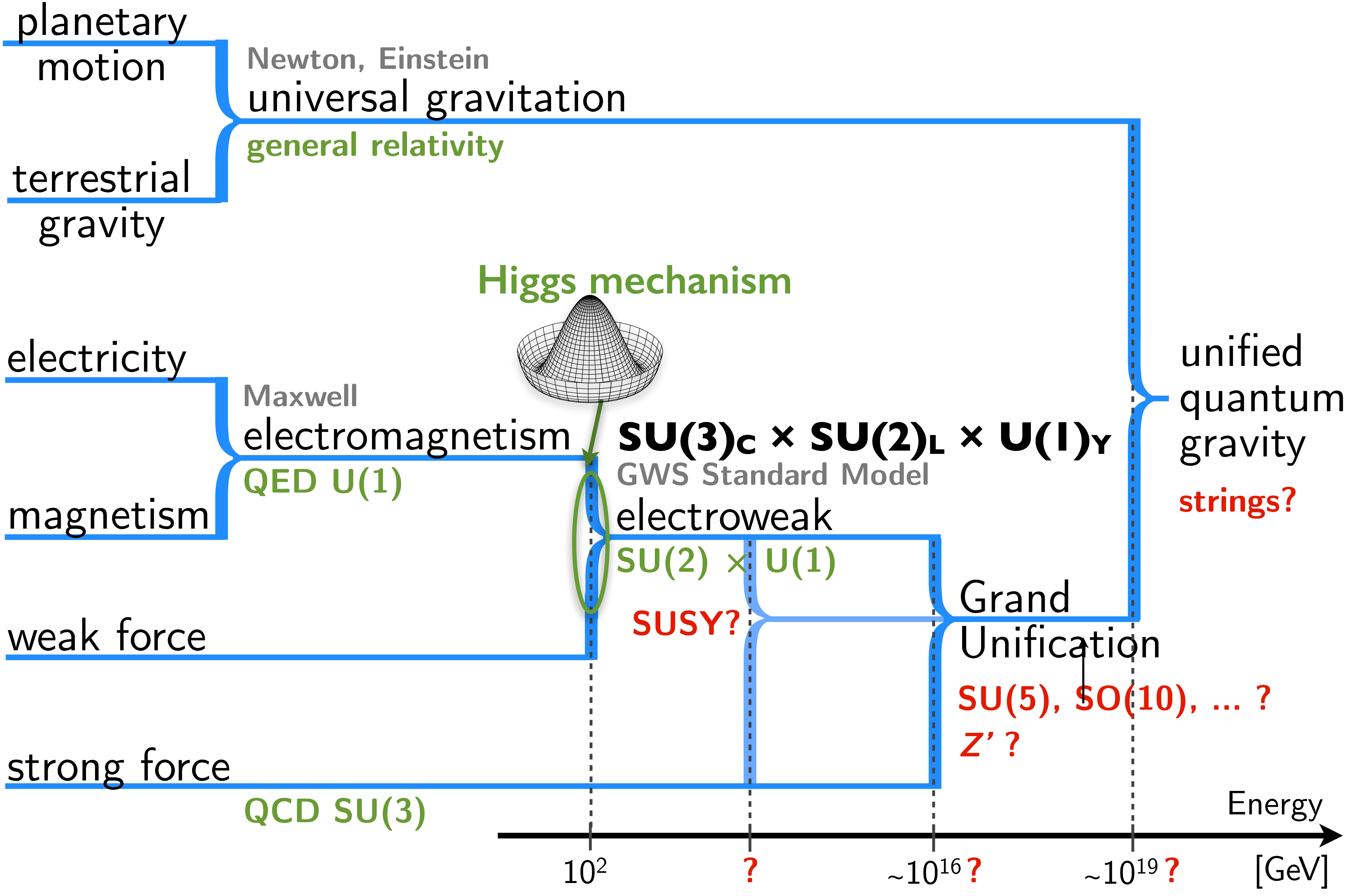




unification scale $\sim 10^{16}$ GeV

$\sim 10^4$ GeV current colliders

Unification?



planetary motion

terrestrial gravity

Newton, Einstein

universal gravitation

general relativity

Higgs mechanism

electricity

magnetism

Maxwell

electromagnetism

QED U(1)

weak force

SU(3)$_C$ × SU(2)$_L$ × U(1)$_Y$

GWS Standard Model

electroweak

SU(2) × U(1)

SUSY?

strong force

QCD SU(3)

Grand Unification

SU(5), SO(10), ... ?

Z' ?

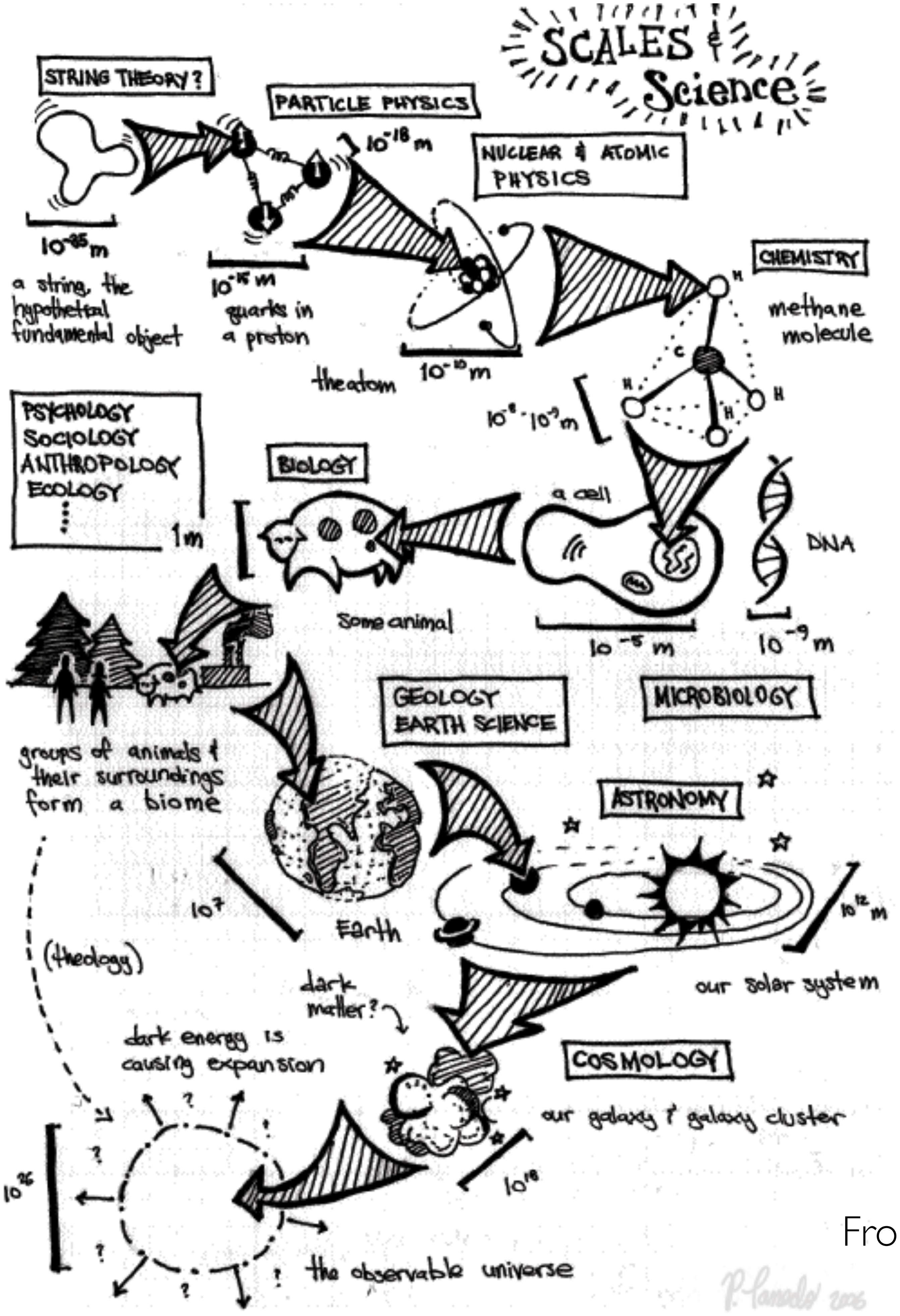
unified quantum gravity

strings?

Energy [GeV]

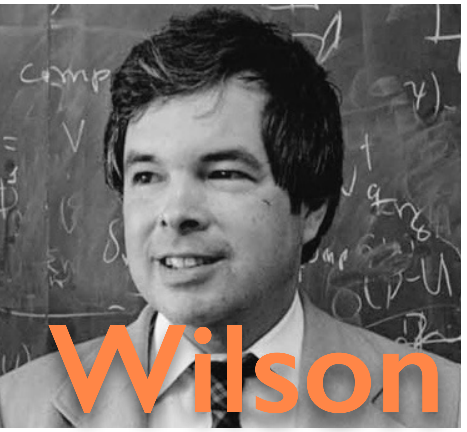
10^2 ? $\sim 10^{16}$? $\sim 10^{19}$?

Effective Theories



Effective theories emerge at different scales and nest into different *regimes* which have some autonomy of description.

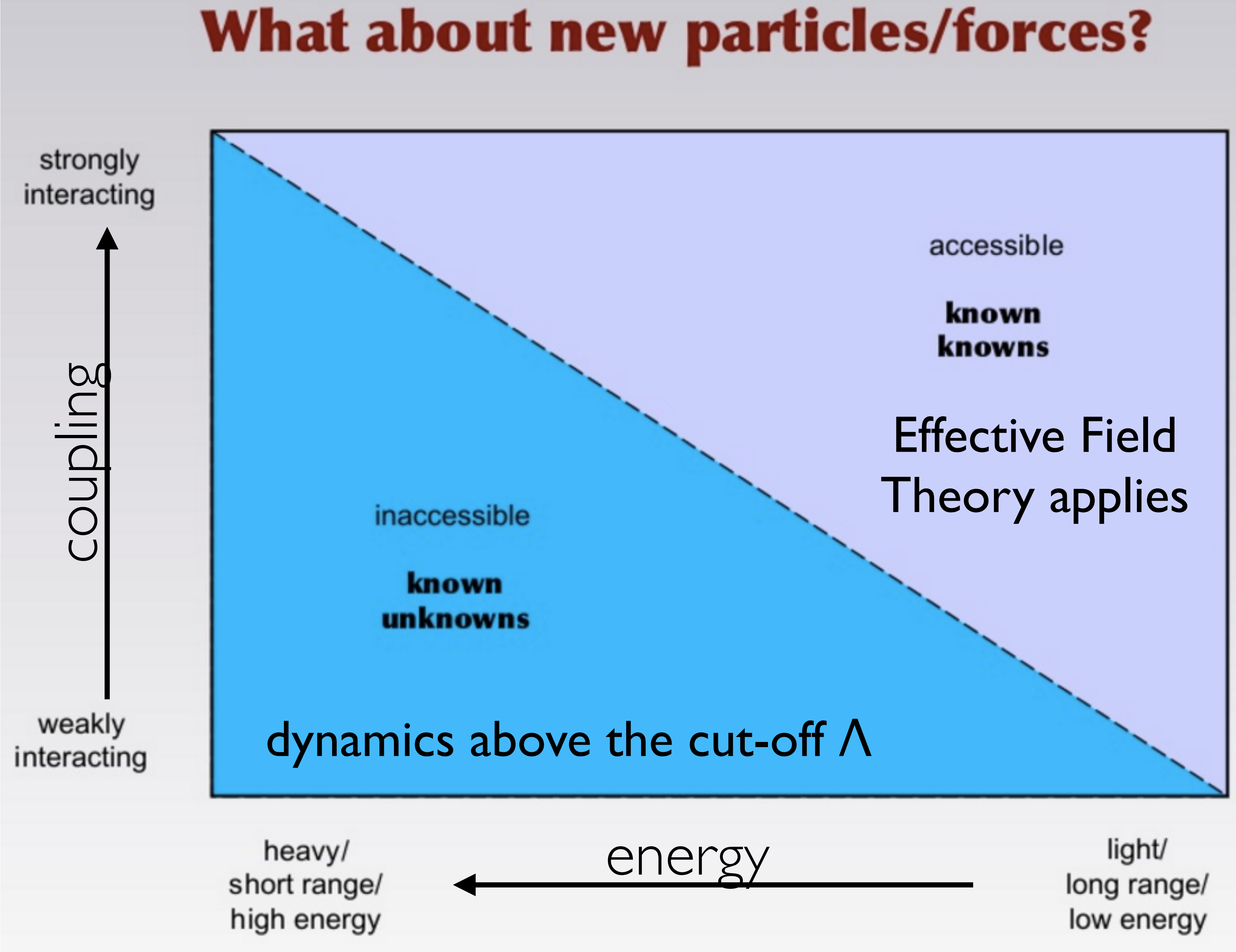
From: Flip Tanado (2009). Quantum Diaries blog: "My research [Part 2] effective theories."



Effective Field Theories have a regime of applicability: below a high-energy cut-off, Λ.

<u>Donald Rumsfeld's</u>

- known knowns
- known unknowns
- unknown unknowns (violations of QFT itself)



What about new particles/forces?

strongly interacting

coupling

weakly interacting

accessible

known knowns

Effective Field Theory applies

inaccessible

known unknowns

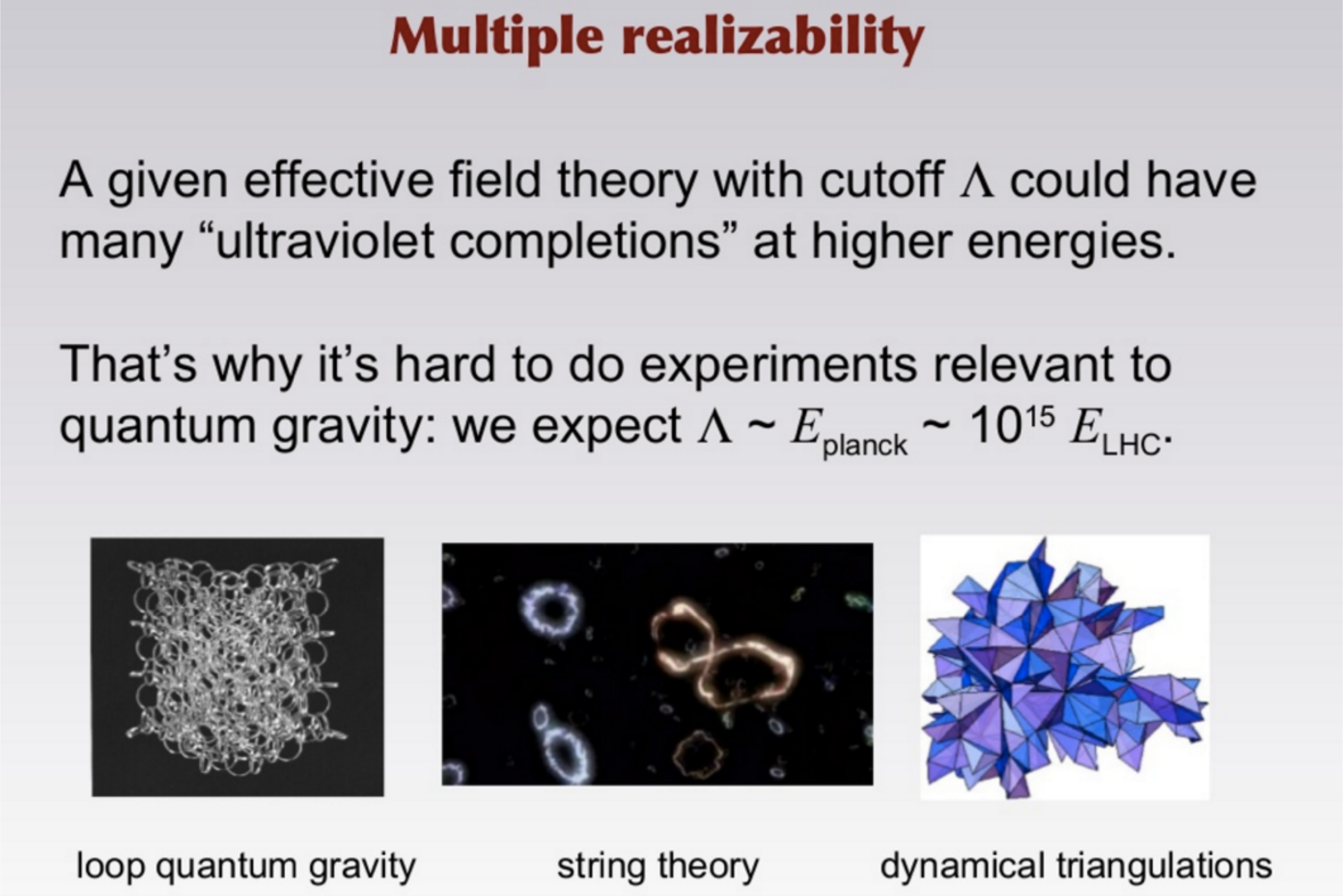
dynamics above the cut-off Λ

heavy/ short range/ high energy

energy

light/ long range/ low energy

Slide from Sean Carroll:
<u>"Quantum Field Theory and the Limits of Knowledge"</u>



Multiple realizability

A given effective field theory with cutoff Λ could have many "ultraviolet completions" at higher energies.

That's why it's hard to do experiments relevant to quantum gravity: we expect $\Lambda \sim E_{\text{planck}} \sim 10^{15}\ E_{\text{LHC}}$.

loop quantum gravity

string theory

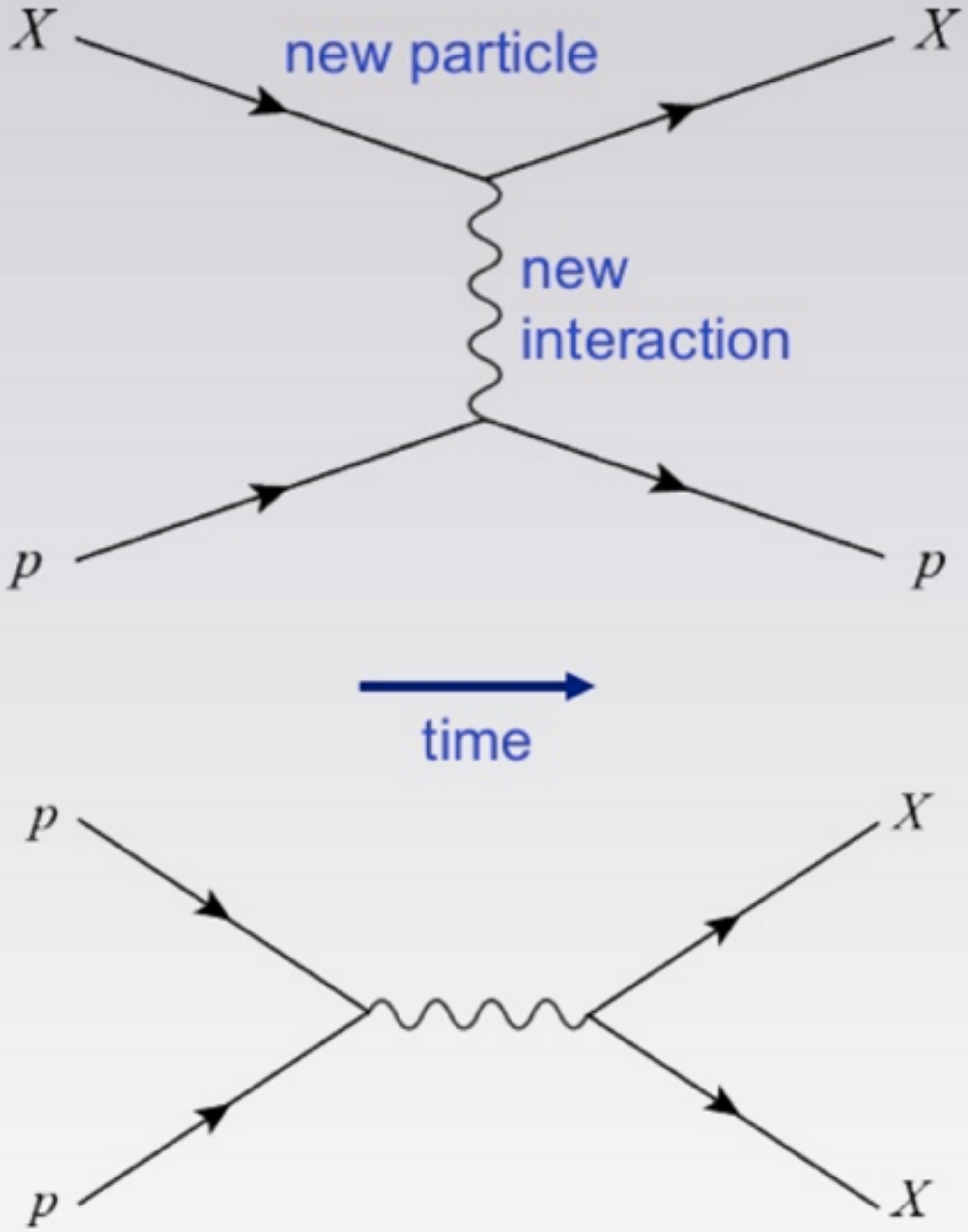
dynamical triangulations

Accepting the *empirical adequacy* or *structural realism* of QFT in a regime does not commit one to any "fundamental" ontology.

Slide from Sean Carroll:
"Quantum Field Theory and the Limits of Knowledge"

QFT puts very tight constraints on new phenomena.

<u>If</u> a new particle can interact with ordinary particles:



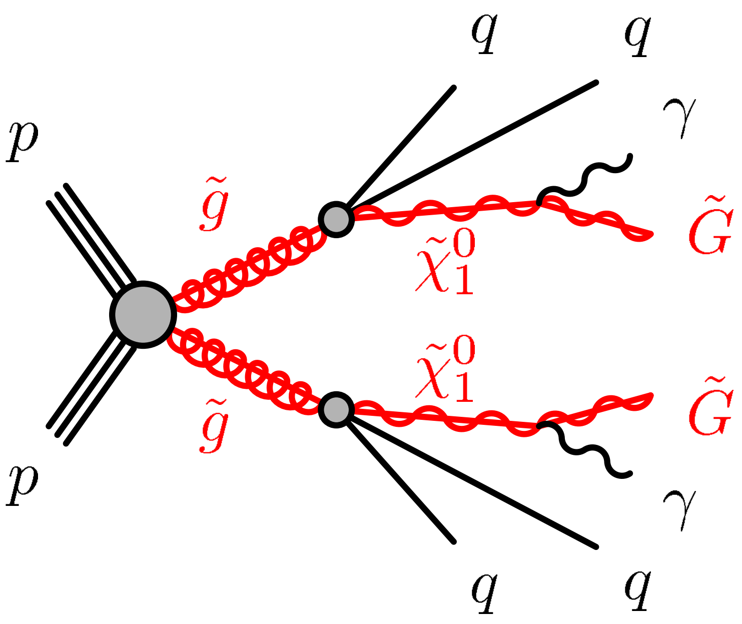
<u>Then</u> that particle can be created in high-energy collisions.

“Crossing symmetry.”

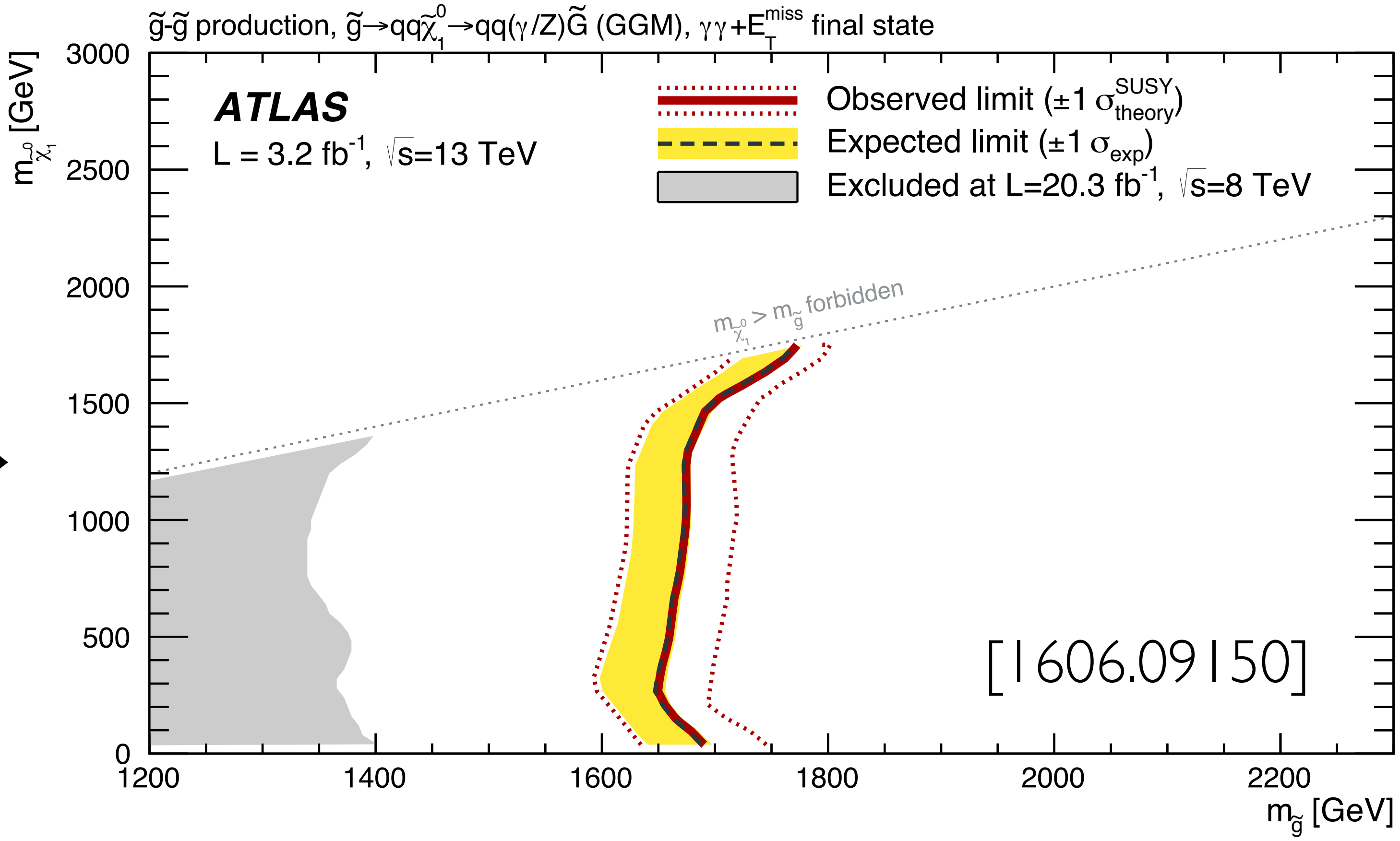
Slide from Sean Carroll:
[“Quantum Field Theory and the Limits of Knowledge”]

Example limits from ATLAS

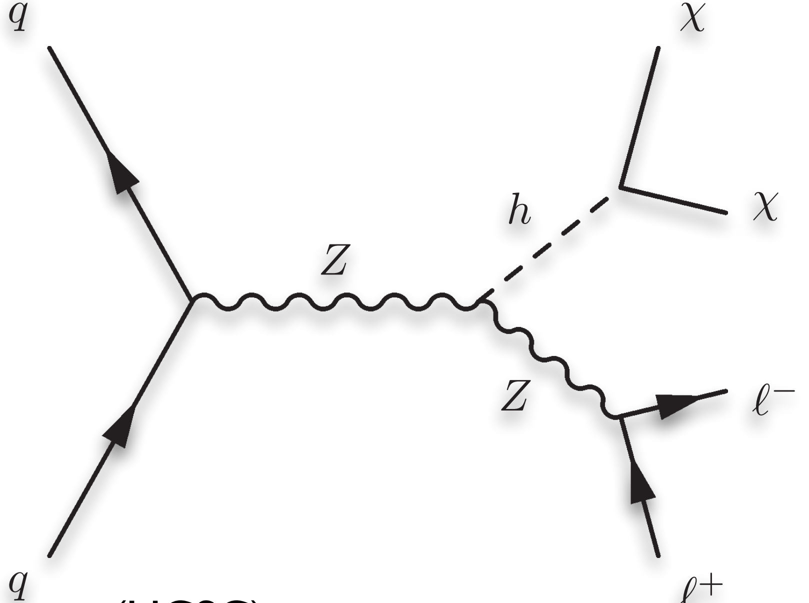
Search for SUSY gluino-neutralino decays to diphoton + missing energy



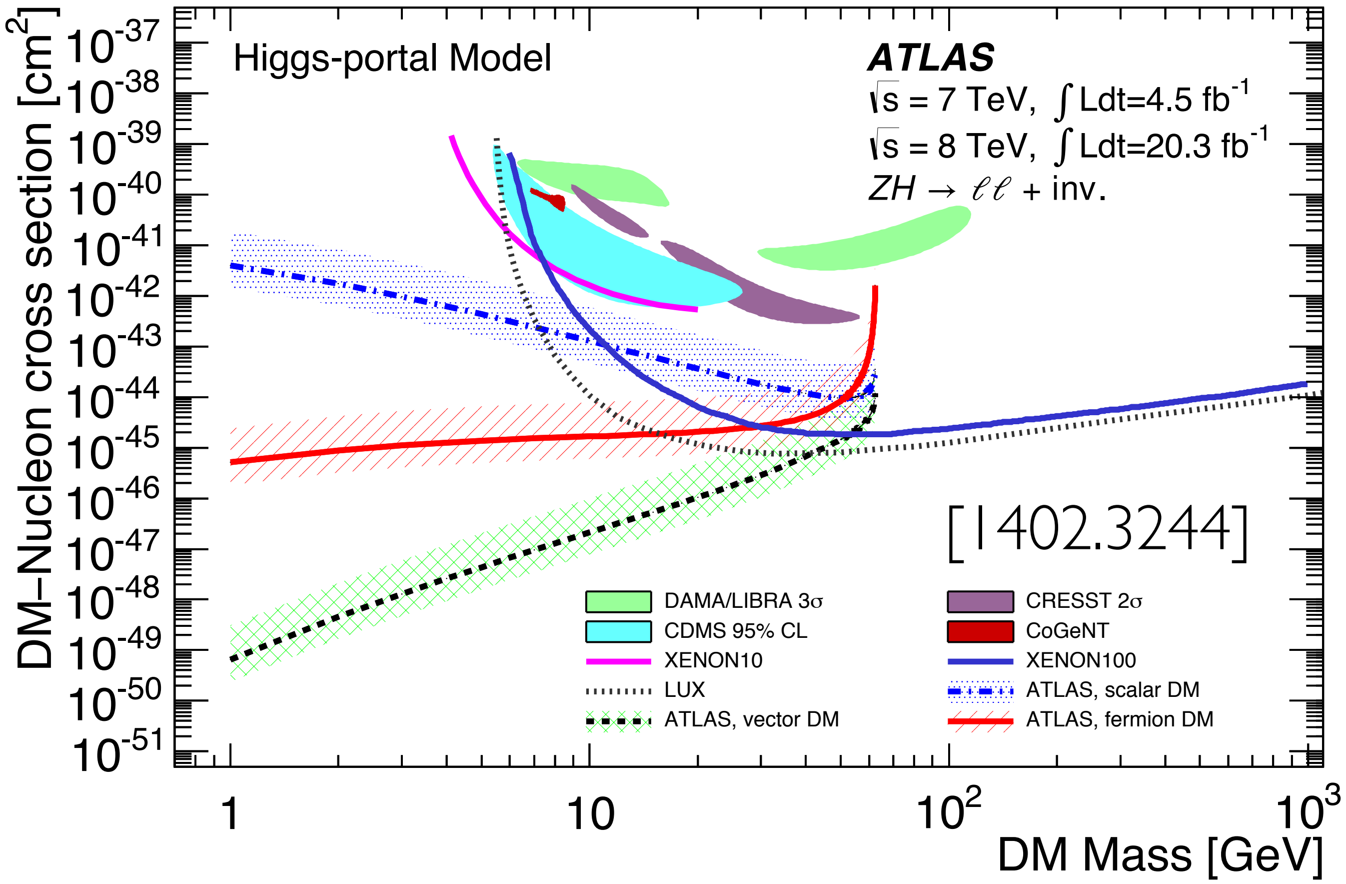
⇒



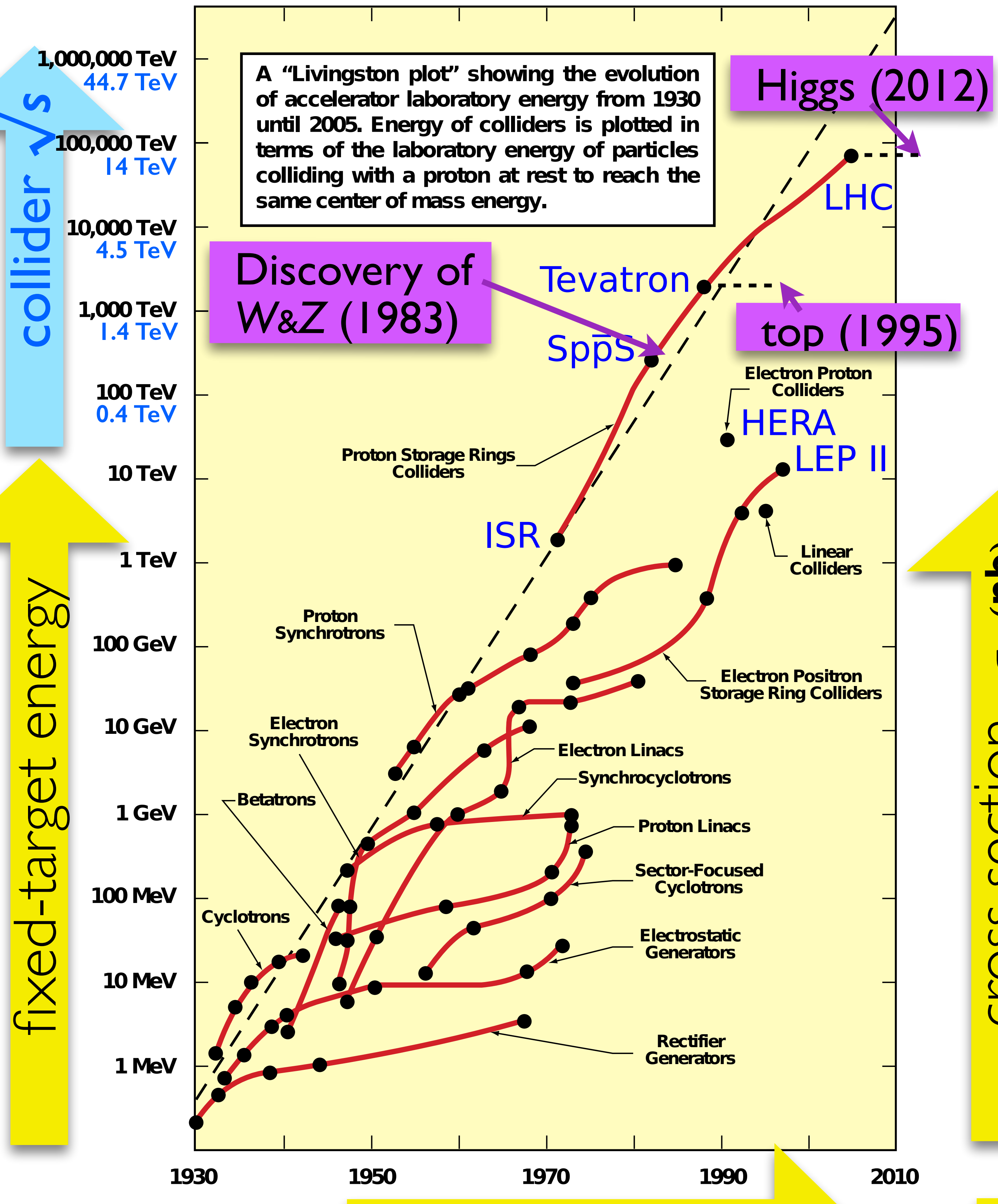
Search for Higgs decaying to additional invisible modes: Higgs portal to dark matter?



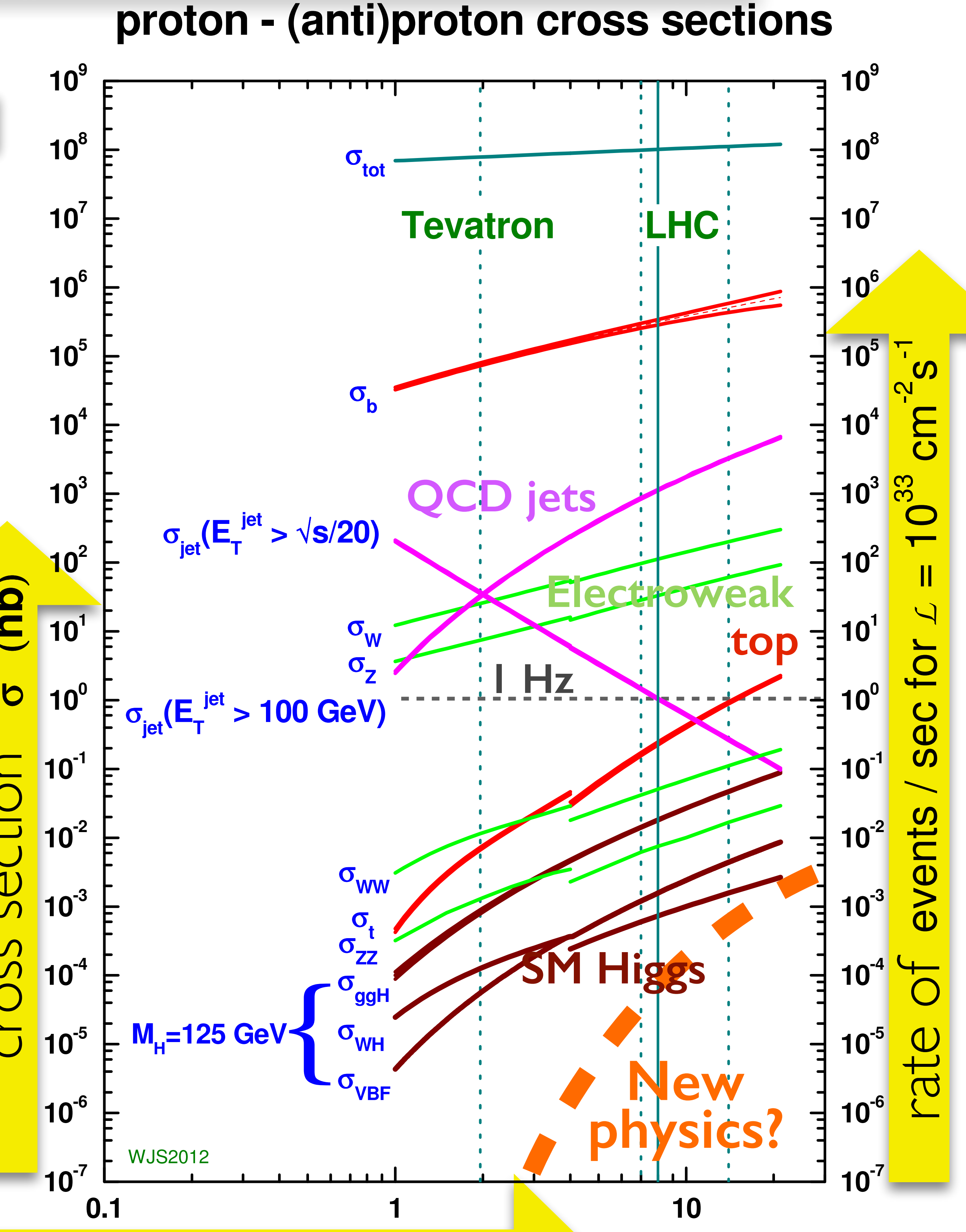
⇒



We need high energies

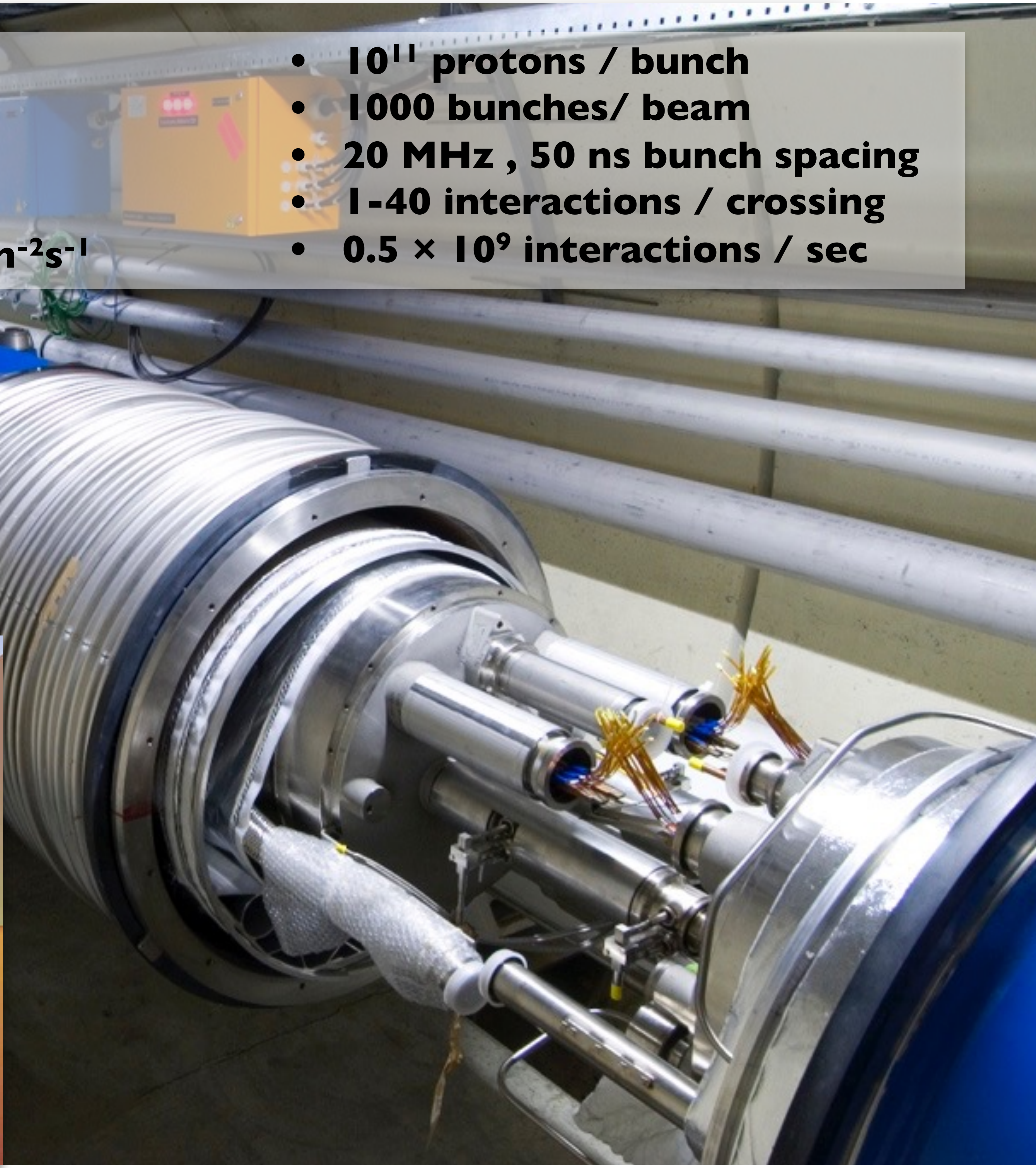


A "Livingston plot" showing the evolution of accelerator laboratory energy from 1930 until 2005. Energy of colliders is plotted in terms of the laboratory energy of particles colliding with a proton at rest to reach the same center of mass energy.



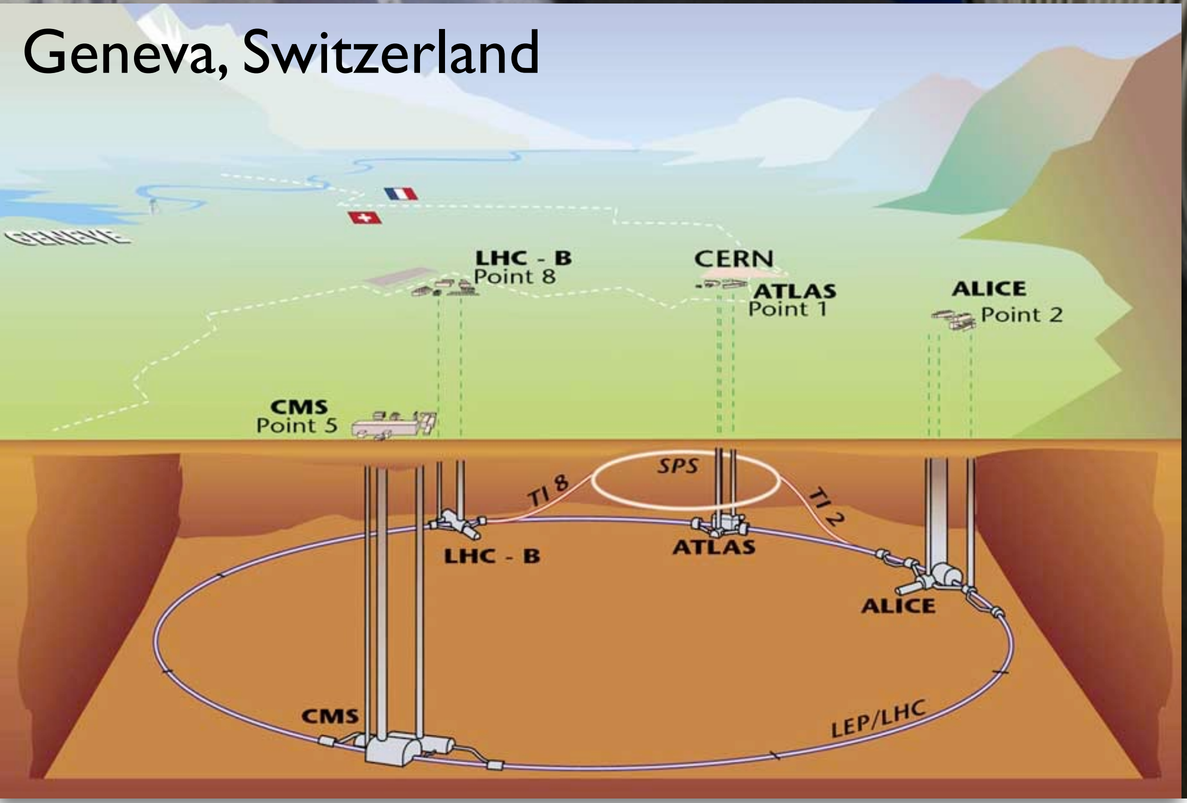
proton - (anti)proton cross sections

Large Hadron Collider



- 27 km circumference
- 1232 dipoles: 15 m , 8.3 T
- 100 tons liquid He, 1.9 K
- p-p collisions at $\sqrt{s}$ = 7-8 TeV
- inst. luminosity = 10^{32}-10^{34} $cm^{-2}s^{-1}$

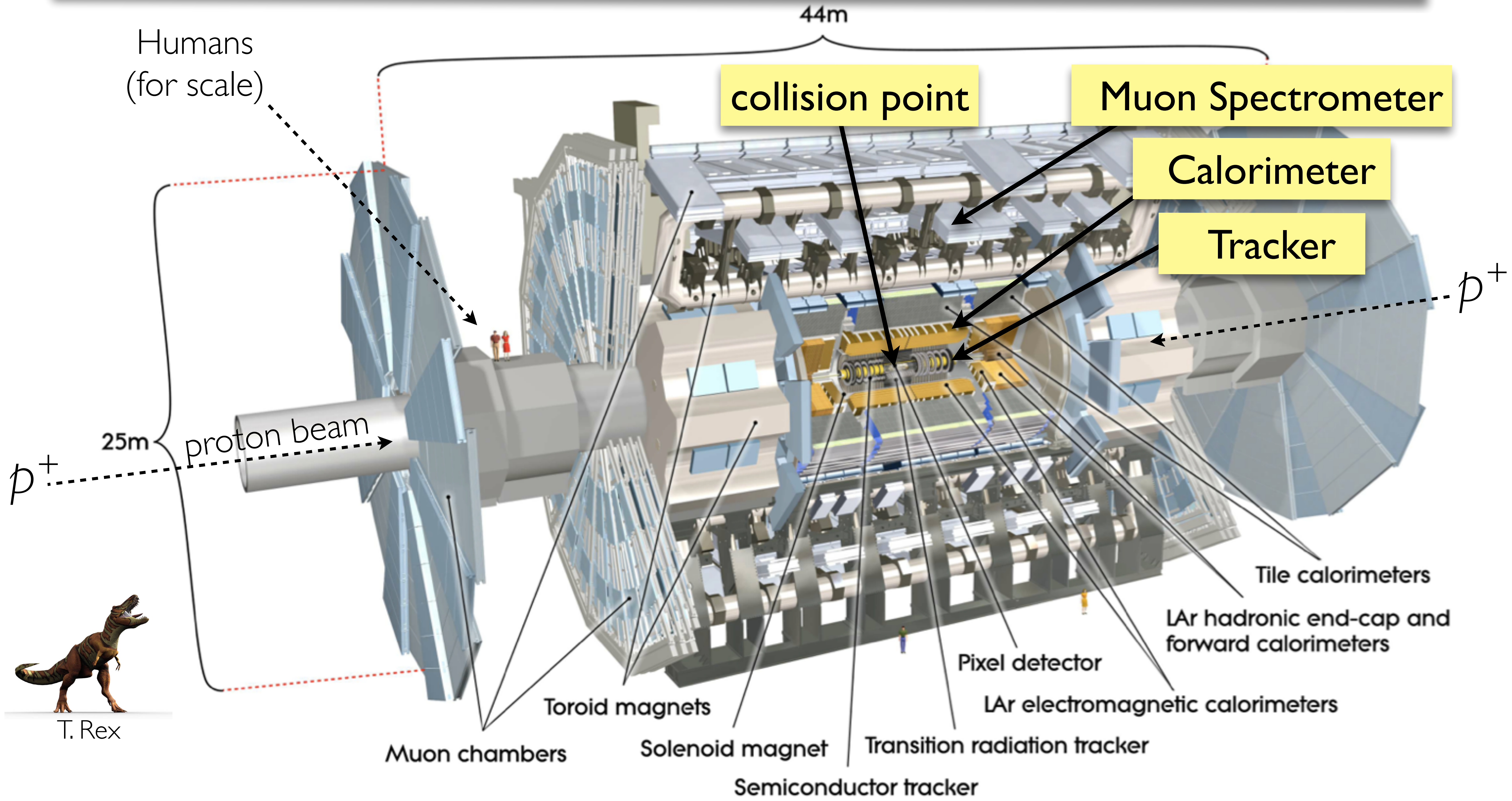
- 10^{11} protons / bunch
- 1000 bunches/ beam
- 20 MHz , 50 ns bunch spacing
- 1-40 interactions / crossing
- 0.5 × 10^{9} interactions / sec



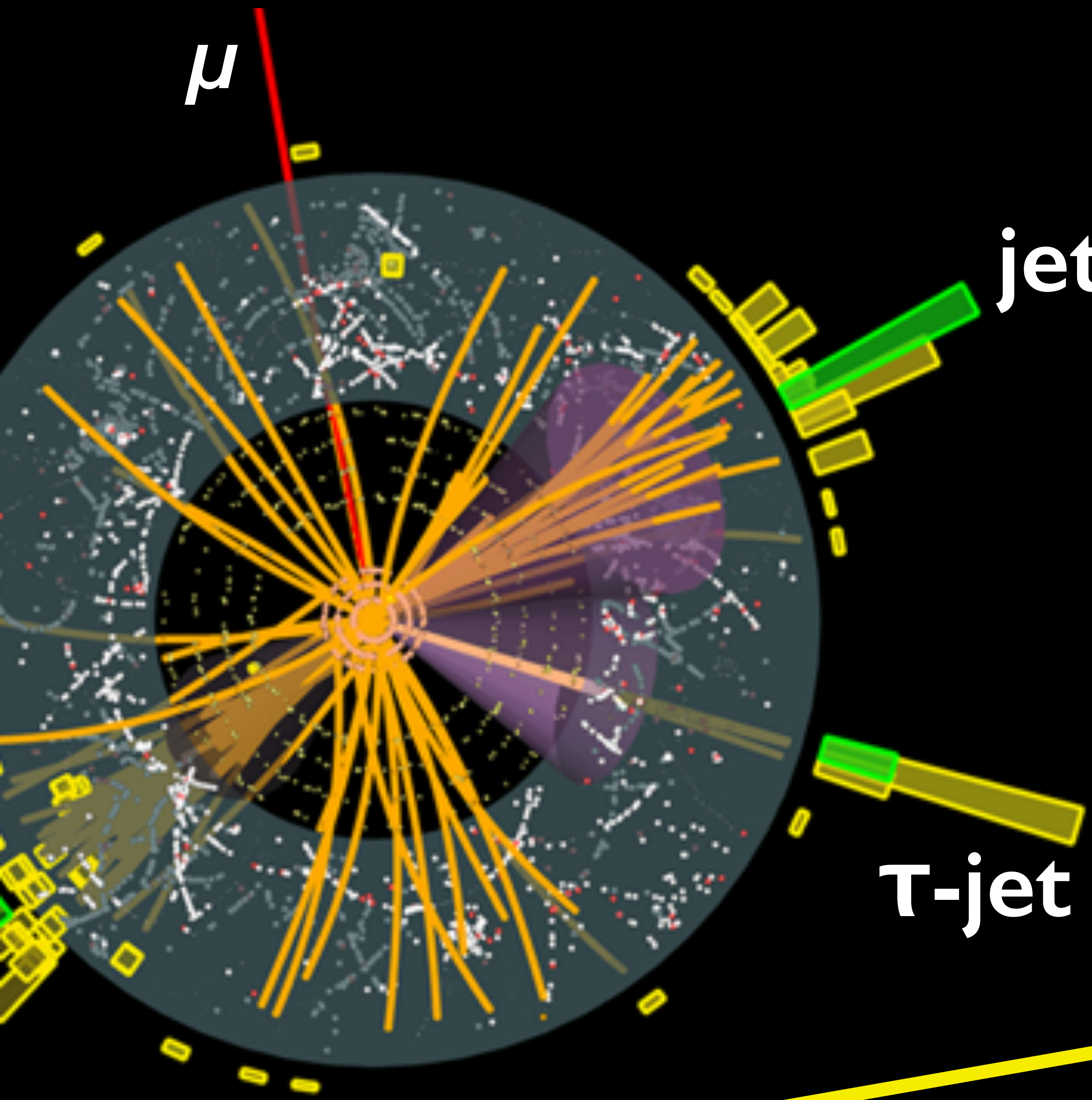
Geneva, Switzerland

ATLAS Detector

ATLAS is a 7 story tall, 100 megapixel "camera", taking 3-D pictures of proton-proton collisions 40 million times per second, saving 10 million GB of data per year, using a world-wide computing grid with over 100,000 CPUs. The collaboration involves more than 3000 scientists and engineers.



What do we reconstruct?



- muons
- electrons & photons
- jets of hadrons
- τ- and *b*-tagged jets
- missing energy

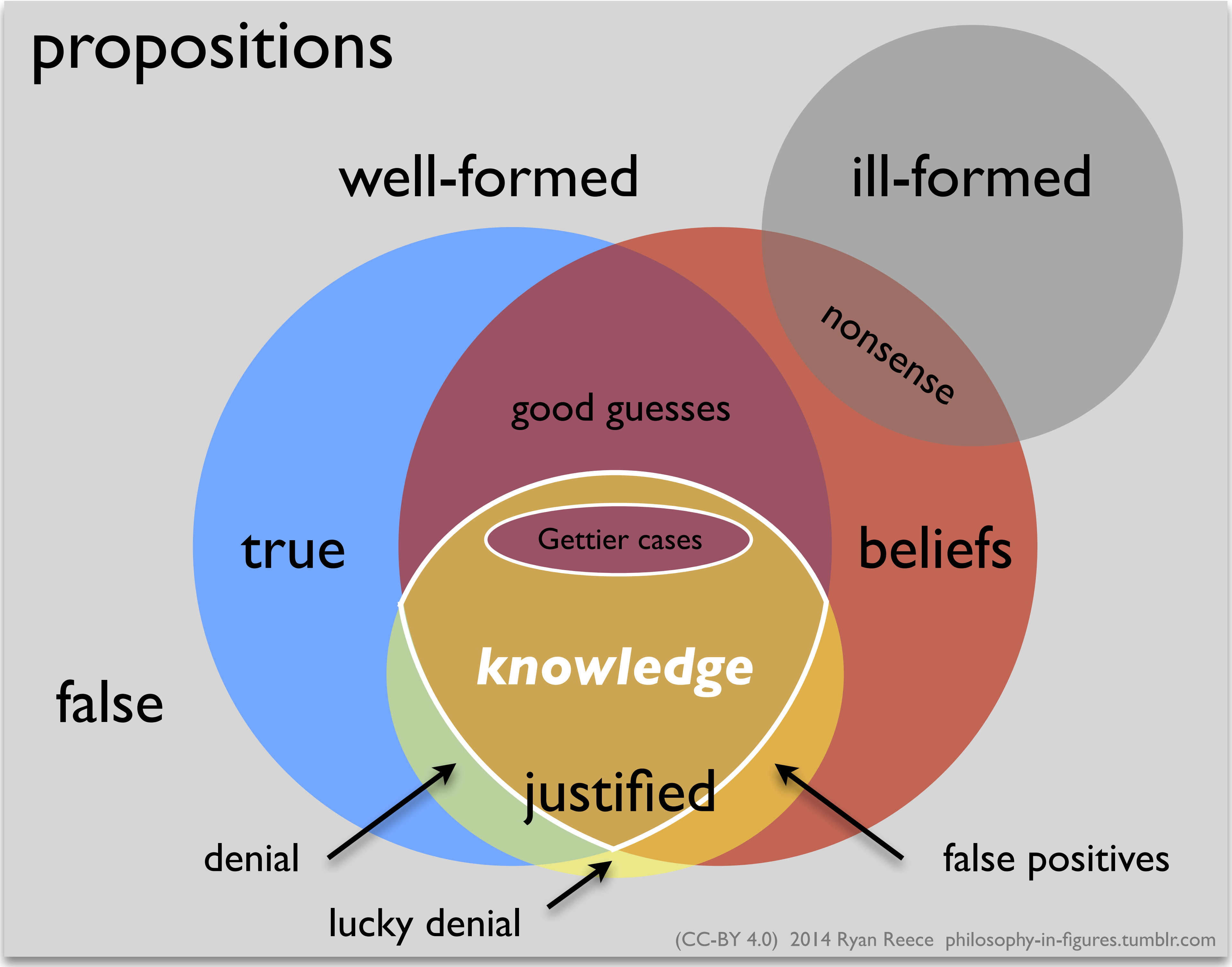
How do we search?

- ATLAS Physics Groups
  - **SM** *W*, *Z*, top,...
  - **Higgs** *H*→γγ, *ZZ*, *WW*, ...
  - **SUSY** *l*+jets,γ+jets, ...
  - **Exotics** *Z'*,*W'*, ...

Currently ATLAS has published 579+ papers

Statistical inference

Knowledge = JTB-G



from: http://philosophy-in-figures.tumblr.com/

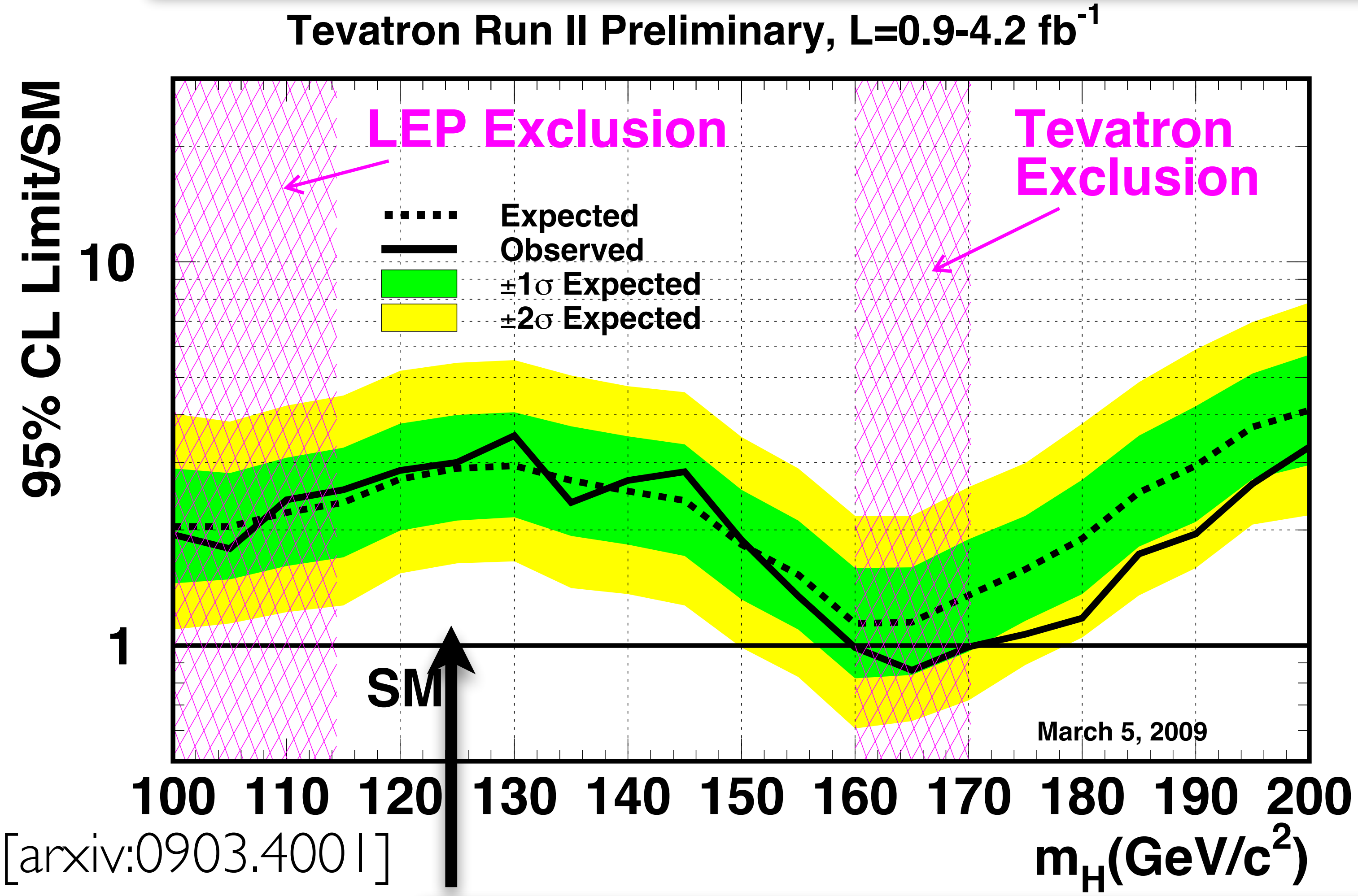
Problem of induction

- Our justification can be
    - ***deductive***: following by definition (logic/ mathematics)
    - ***inductive***: generalizing a universal based on limited data
- Induction is always susceptible possible "black swans".
- Later, 20th century positivism can largely be seen as a project staying true to the epistemological methods of science, but without the statistical confidence to make claims about the reality of their models (metaphysics).

David Hume (1711-1776)

A main goal of this talk

I want to facilitate an appreciation for statistical confidence intervals like below, and try to touch ground with how LHC physicists go from collecting and reducing data to performing a statistical test.



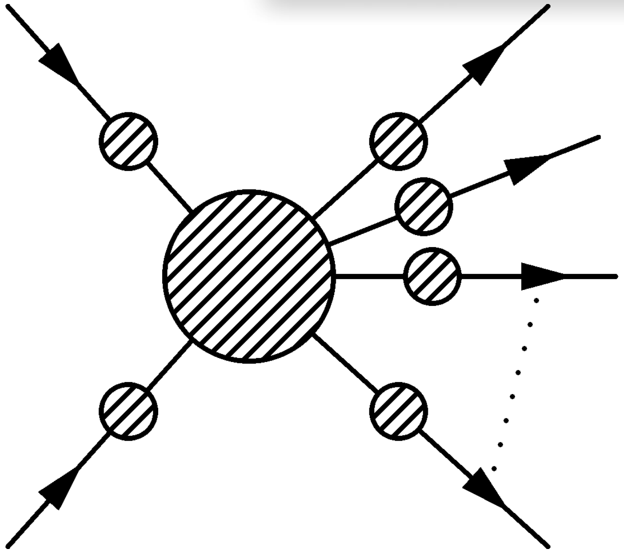
[arxiv:0903.4001]

Lot’s of important details glossed over. Great pedagogical documents on statistics:

- Cranmer, K. “Practical Statistics for the LHC”. [arxiv:1503.07622]
- Cowan, G. (1998). *Statistical Data Analysis.*
- ATLAS Higgs Combination [arxiv:1207.0319]
- Cranmer *et al.* [arxiv:1007.1727]

Higgs was later found in 2012 by ATLAS and CMS with $m_H \approx 125$ GeV

Scattering cross sections



At colliders, it can be shown that the differential rate of any given process factors as

$$dN = \varepsilon \, L \, dt \, d\sigma$$

= (efficiency) (luminosity) d(time) d(cross section)

QFT shows that the ***cross section*** can be calculated in terms of a ***matrix element***.

$$d\sigma = \prod_f \left(\frac{d^3 p_f}{(2\,\pi)^3 \, 2\, E_f} \right) \; \frac{|\mathcal{M}|^2}{4\, E_1 \, E_2 \, |v_1 - v_2|} \; (2\,\pi)^4 \, \delta^4\Big(p_1 + p_2 - \sum_f p_f\Big)$$

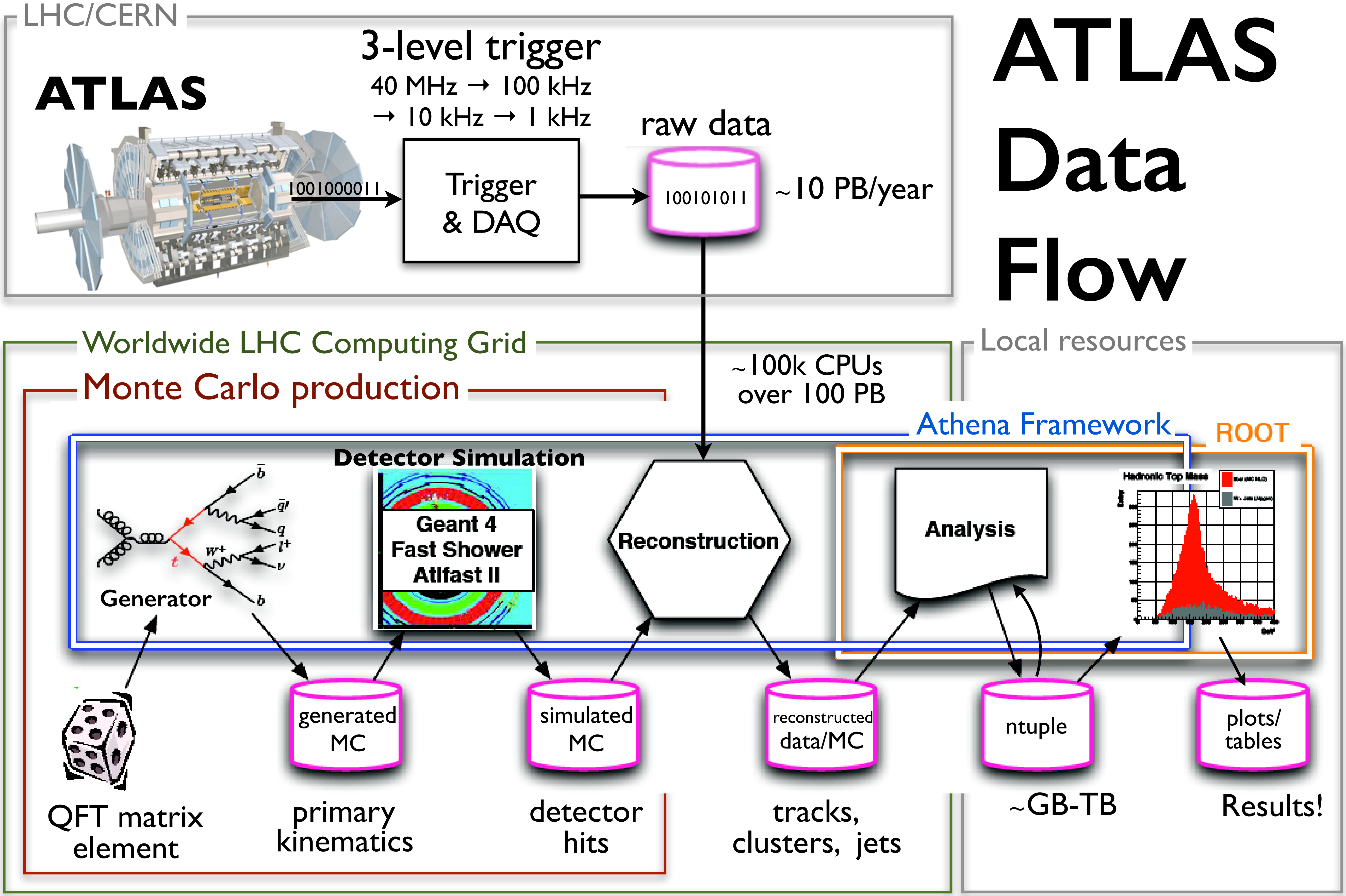
The number of expected events can be calculated by integrating the differential cross section over the running of the experiment.

$$N = \int dt \, L \int d\sigma \, \varepsilon$$

$$= \left(\int dt \, L \right) A \, C \, \sigma$$

= (integrated luminosity) (acceptance) (efficiency) (cross section)

ATLAS Data Flow



LHC/CERN

ATLAS

3-level trigger

40 MHz → 100 kHz
→ 10 kHz → 1 kHz

1001000011

Trigger
& DAQ

raw data

100101011

~10 PB/year

Worldwide LHC Computing Grid

Monte Carlo production

~100k CPUs
over 100 PB

Local resources

Athena Framework

ROOT

Detector Simulation

Geant 4
Fast Shower
Atlfast II

Generator

Reconstruction

Analysis

generated
MC

simulated
MC

reconstructed
data/MC

ntuple

plots/
tables

QFT matrix
element

primary
kinematics

detector
hits

tracks,
clusters, jets

~GB-TB

Results!

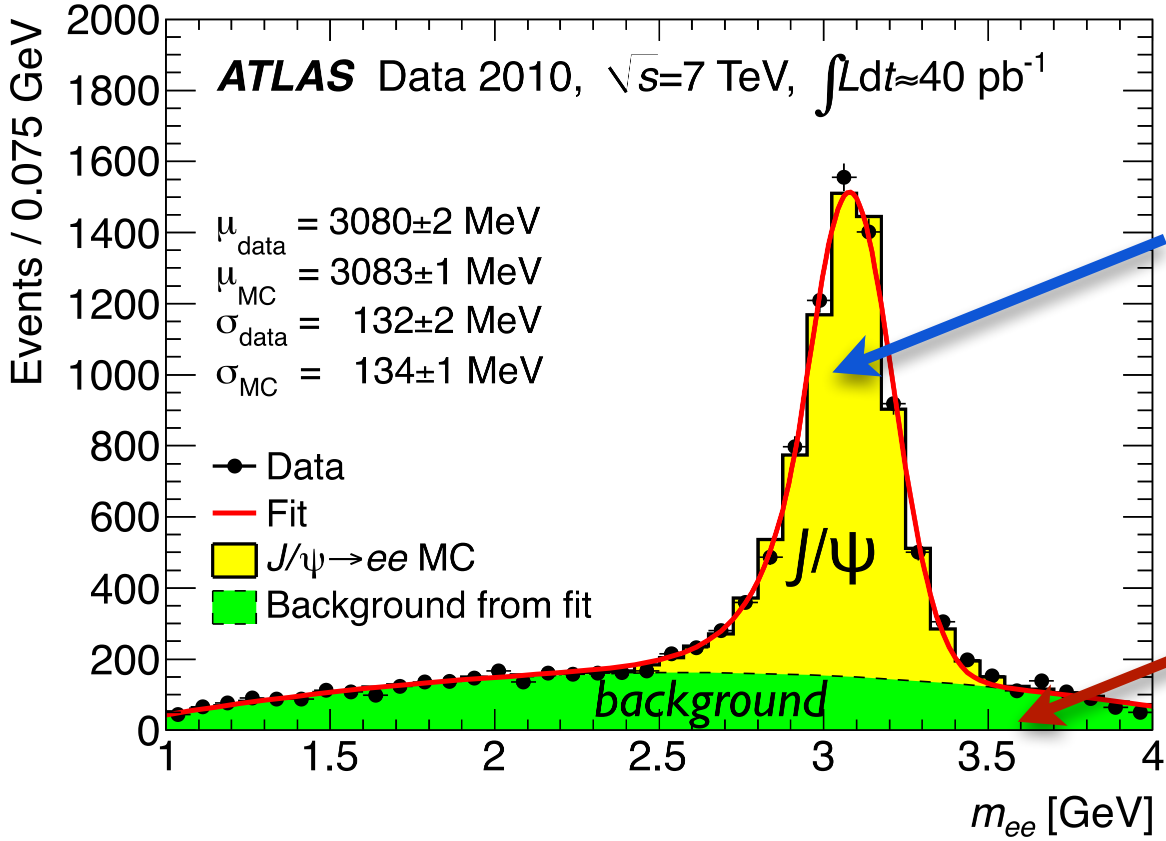
ab initio simulation

Building a model

$$N(expected) = \underbrace{N(correct\text{-}ID)} + \underbrace{N(fake)}$$

- ***Bottom-up***
- well-identified objects have scale factors from control regions
- estimated with detailed Monte Carlo simulation

- ***Top-down*** **, *"data-driven"***
- various magic with data depending on the analysis and your creativity
- side-band fit
- fake-factor method



Bottom-up
Monte Carlo

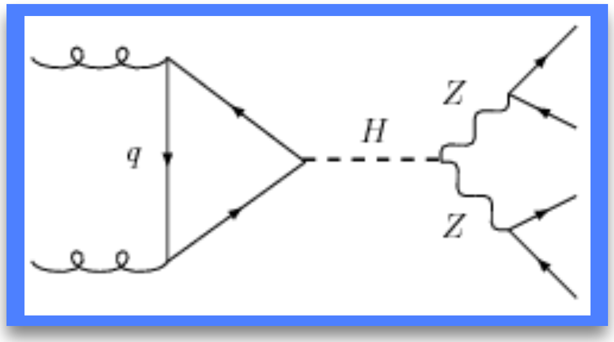
Data-driven
side-band fit

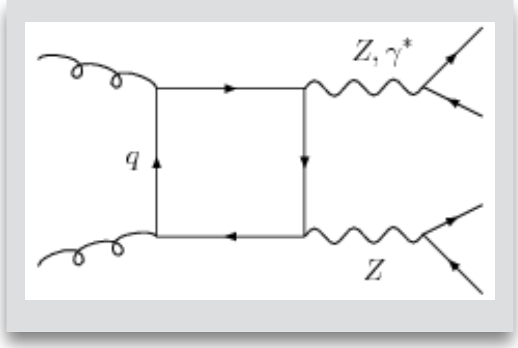
[arxiv:1110.3174]

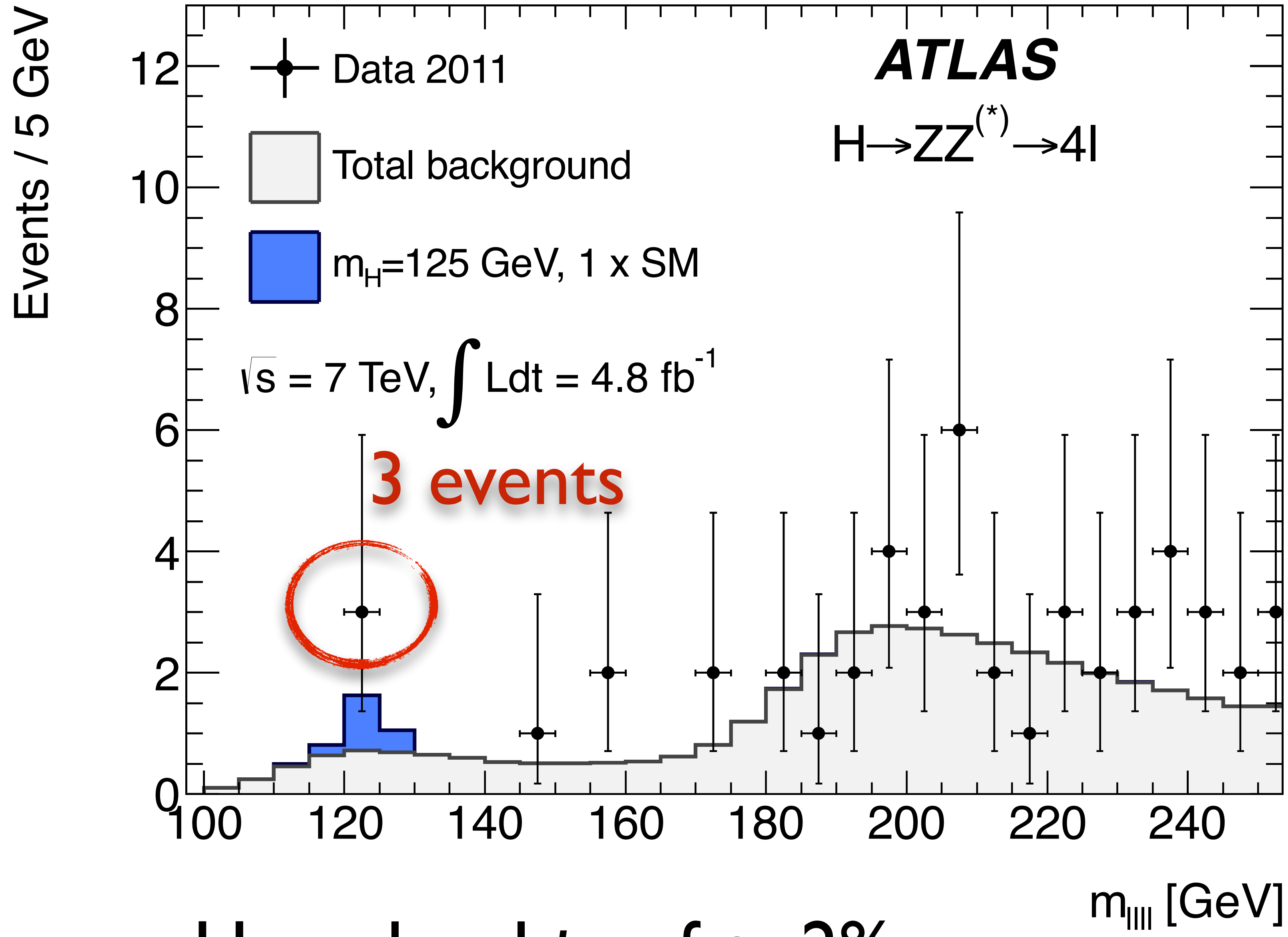
Is this significant?

Statistical (and philosophical) questions:

- How can we be precise and rigorous about how confident we are that a model is wrong?
  - *Hypothesis testing*
- How can we calculate the best-fit estimate of some parameter?
  - *Point estimation and confidence intervals*







3 events

Has a local p_0 of ≈ 2%

[arxiv:1207.0319]

Confidence Intervals

- A ***frequentist confidence interval*** is constructed such that, given the model, if the experiment were repeated, each time creating an interval, 95% (or other CL) of the intervals would contain the true population parameter (*i.e.* the interval has ≈95% *coverage*).
  - They can be one-sided exclusions, e.g. $m(Z') > 2.0$ TeV at 95% CL
  - Two-sided measurements, e.g. $m_{\mathrm{H}} = 125.1 \pm 0.2$ GeV at 68% CL
  - Contours in 2 or more parameters

- This ***is not the same*** as saying *"There is a 95% probability that the true parameter is in my interval"*. Any probability assigned to a parameter *strictly involves a Bayesian prior probability*.
- Bayes' theorem: P(Theory | Data) ∝ P(Data | Theory) P(Theory)

"likelihood" "prior"

Statistical model

“Marked Poisson” Probability model (PDF):

$\mathcal{D} = \{x_1, \ldots, x_n\}$

data #expected events observable

$$\mathbf{f}(\mathcal{D}|\nu, \alpha) = \mathrm{Pois}(n|\nu) \prod_{e=1}^{n} f(x_e|\alpha) = L(\alpha)$$

histograms

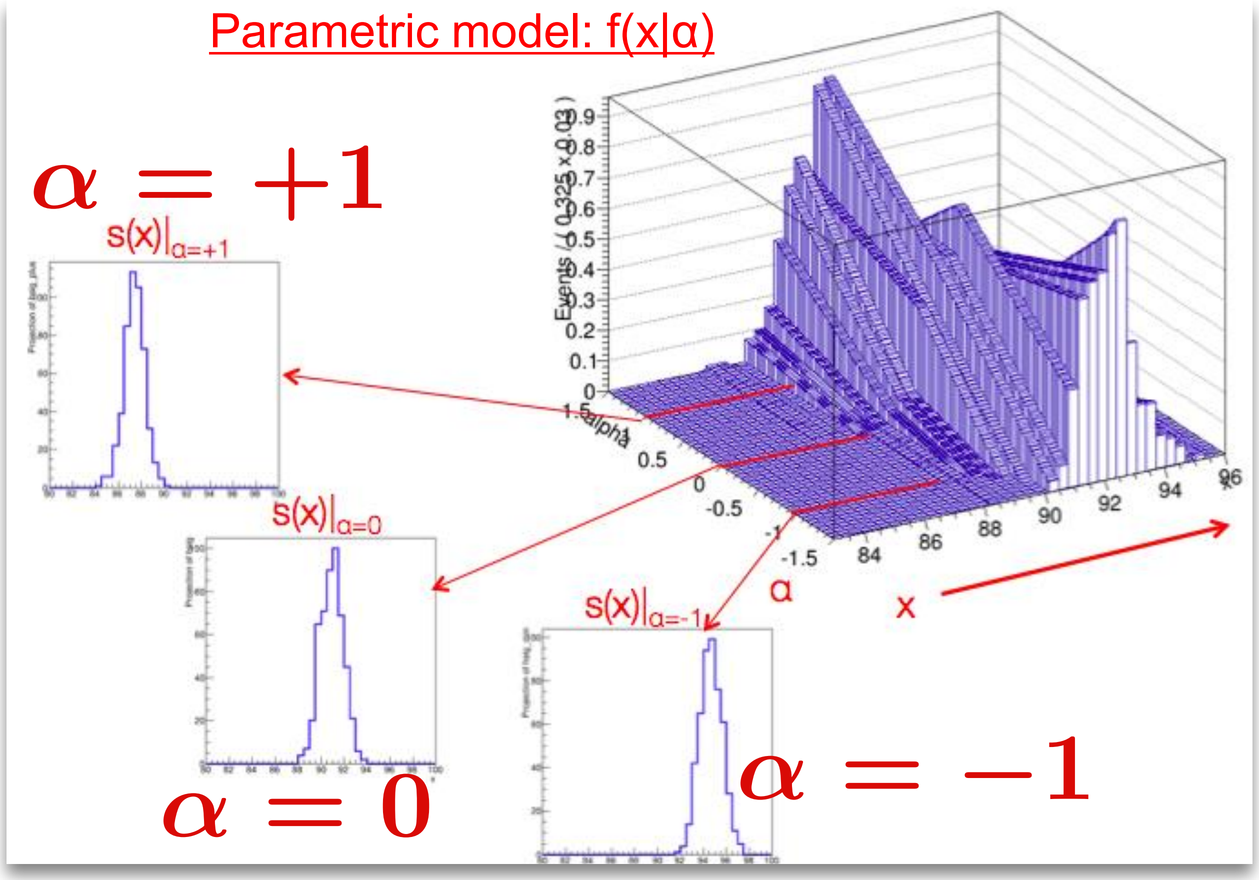
“*Likelihood*”

function of params with data fixed

$\{\alpha\}$ parameters include:

1. parameters of interest $\{\mu\}$:
    - e.g. Higgs mass (m_H) and signal strength (μ)
      μ=0 no signal, μ=1 nominal signal
2. nuisance parameters $\{\theta\}$:
   systematic uncertainties to be “profiled” away by maximizing *L* for a given μ.
    - e.g. luminosity uncert., jet-energy scale, electron energy scale, electron identification efficiency, etc.

template morphing

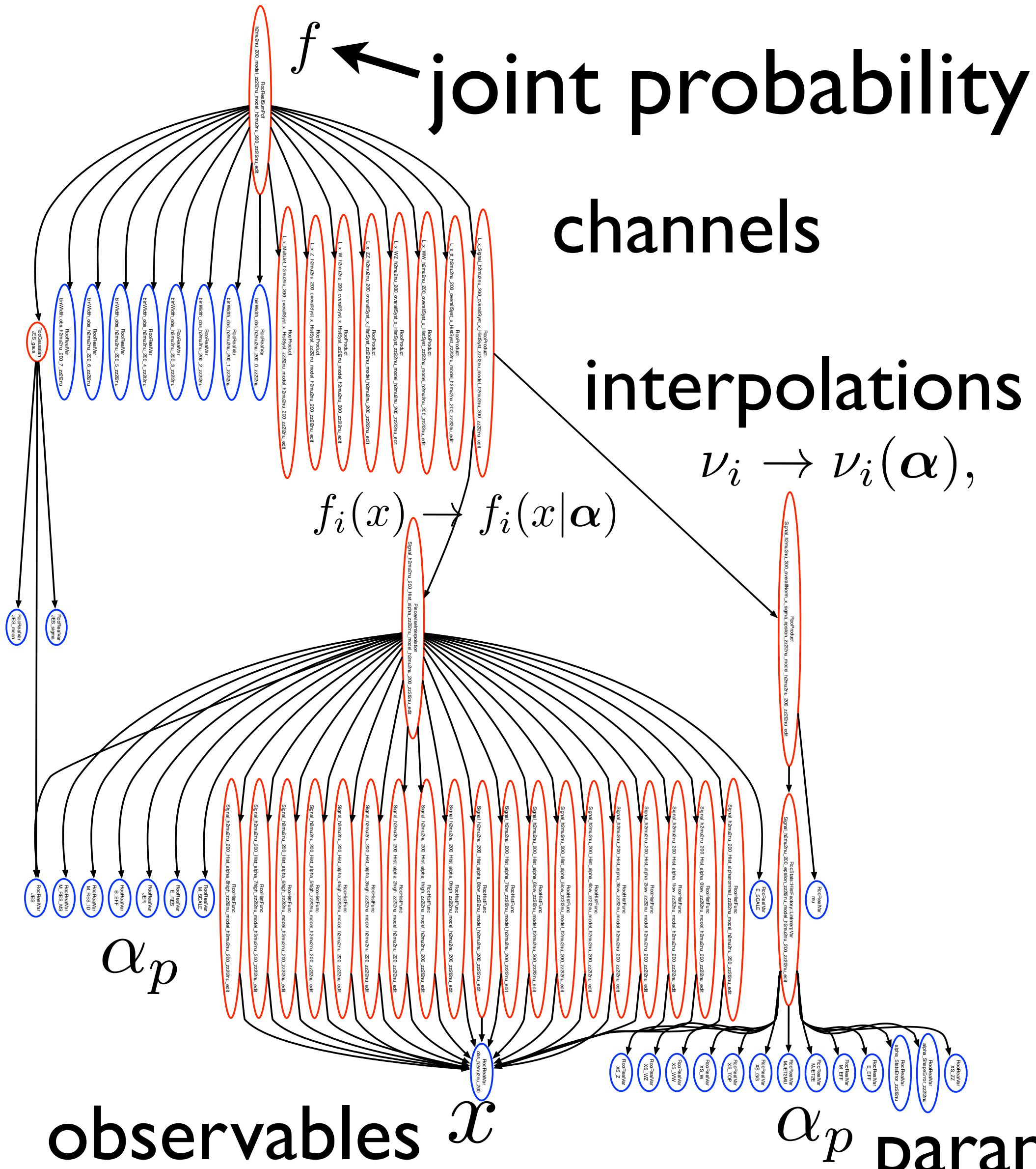


[arxiv:1503.07622, https://indico.cern.ch/event/243641/]

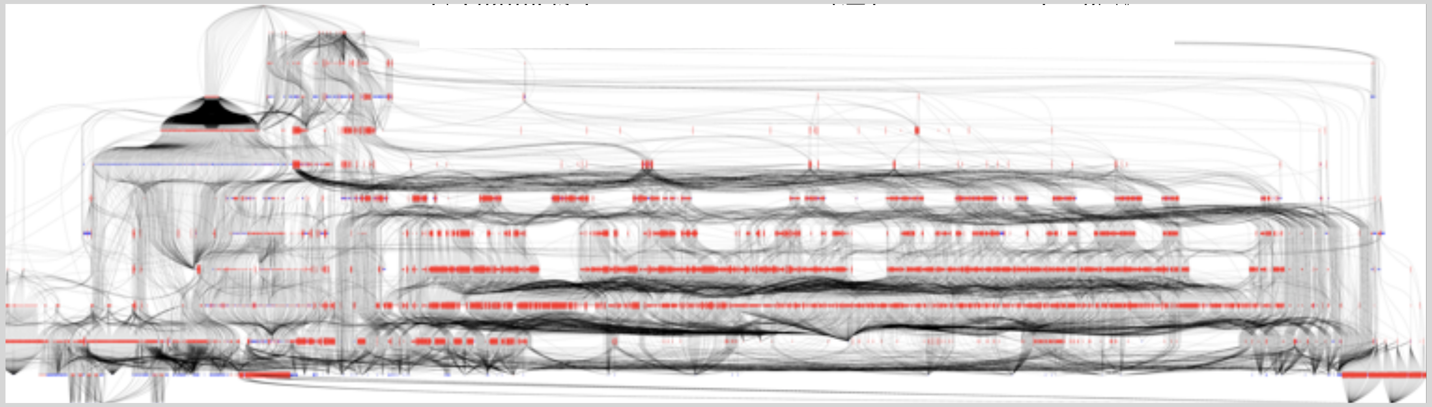
Statistical model

<u>General probability model with many channels and constraints:</u>

$$\mathbf{f}_{\mathrm{tot}}(\mathcal{D}_{\mathrm{sim}}, \mathcal{G}|\alpha) = \overset{\text{channels}}{\prod_{c\in\text{channels}}} \left[\mathrm{Pois}(n_c|\nu_c(\alpha)) \overset{\text{events}}{\prod_{e=1}^{n_c}} f_c(x_{ce}|\alpha)\right] \cdot \overset{\text{constraints}}{\prod_{p\in\mathbb{S}}} f_p(a_p|\alpha_p)$$

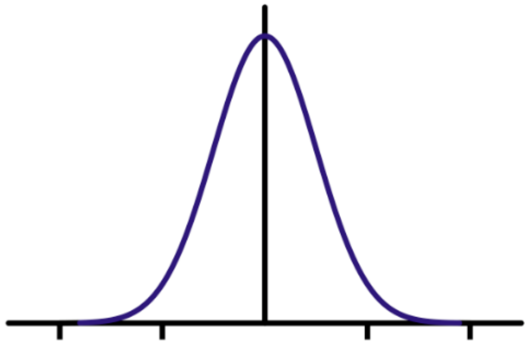


At the time of the Higgs discovery (2012) combined model had 100 channels & 500+ nuisance params.



[arxiv:1503.07622, https://indico.cern.ch/event/243641/]

Maximum Likelihood Estimate



Consider an experiment with N repeated measurements that are Gaussian distributed. The likelihood function is therefore

$$L = \prod_{i=1}^{N} \frac{1}{\sigma\sqrt{2\pi}} \, \exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2} \right)$$

The MLE for the mean, μ, can be found by maximizing the likelihood function, or equivalently, its natural logrhythm.

$$\ln L = -N \, \ln(\sigma\sqrt{2\pi}) - \sum_{i=1}^{N} \frac{(x_i - \mu)^2}{2\sigma^2}$$

$$0 = \frac{\partial \ln L}{\partial \mu} = \sum_{i=1}^{N} \frac{(x_i - \mu)}{\sigma^2}$$

$$\Rightarrow \qquad \hat{\mu} = \frac{1}{N} \sum_{i=1}^{N} x_i$$

Statistical point estimation

"Profiling"

Calculate the profile likelihood ratio:

$$\lambda(\mu) = \frac{L(\mu, \hat{\hat{\boldsymbol{\theta}}})}{L(\hat{\mu}, \hat{\boldsymbol{\theta}})}$$

Similar to the Likelihood, but does not depend on the nuisance parameters.

$\hat{}$ = Maximum Likelihood Estimates

$\hat{\hat{}}$ = MLE for given μ

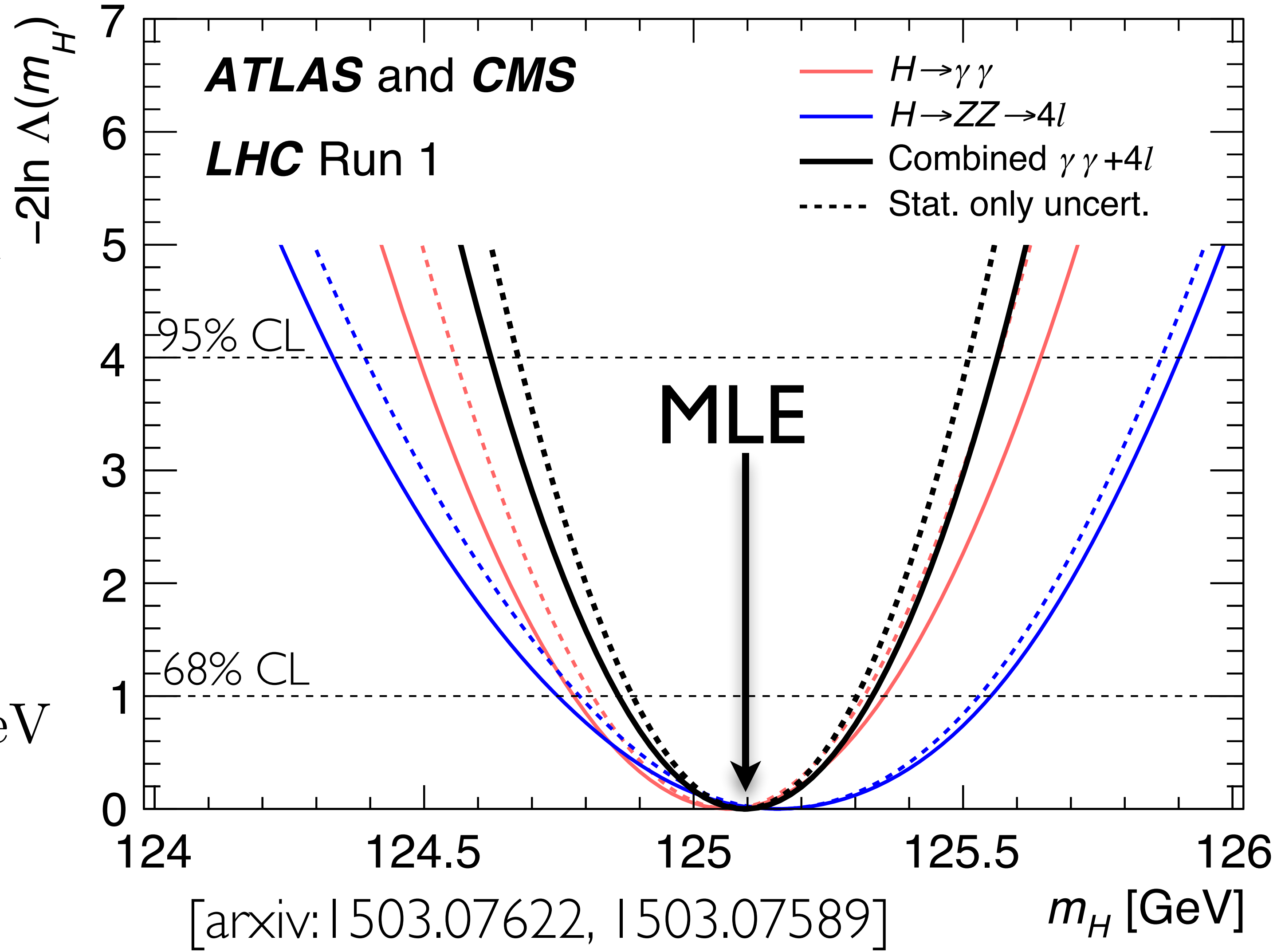
Maximum Likelihood Estimate

Maximize $\lambda(\mu)$

$\Rightarrow$ minimize $-\ln\lambda(\mu)$

e.g. Best-fit Higgs mass

$$m_H = 125.09 \pm 0.21\ (\text{stat.}) \pm 0.11\ (\text{syst.})\ \text{GeV}$$



[arxiv:1503.07622, 1503.07589]

Hypothesis testing

- Null hypothesis, H_0: the SM
- Alternative hypothesis, H_1: some new physics
- Type-I error: false positive rate (α)
- Type-II error: false negative rate (β)

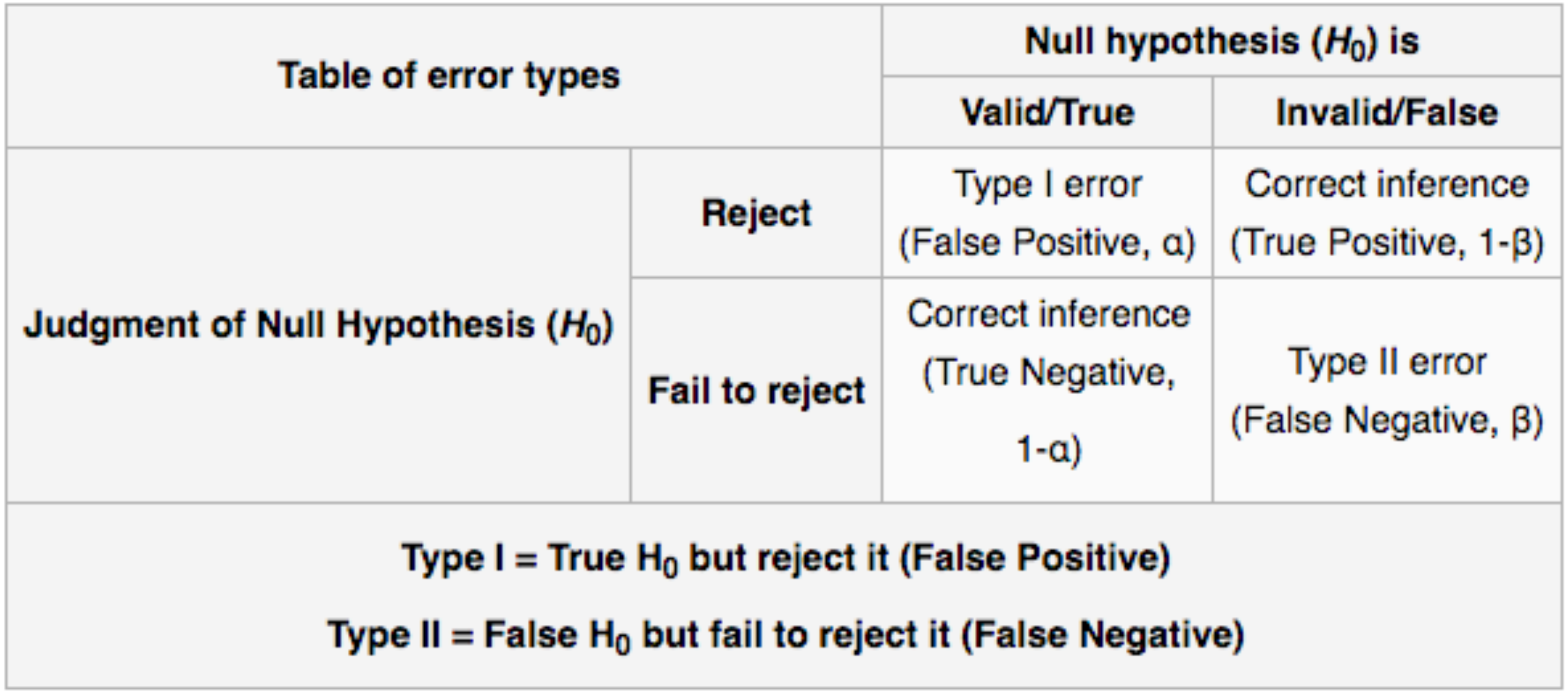


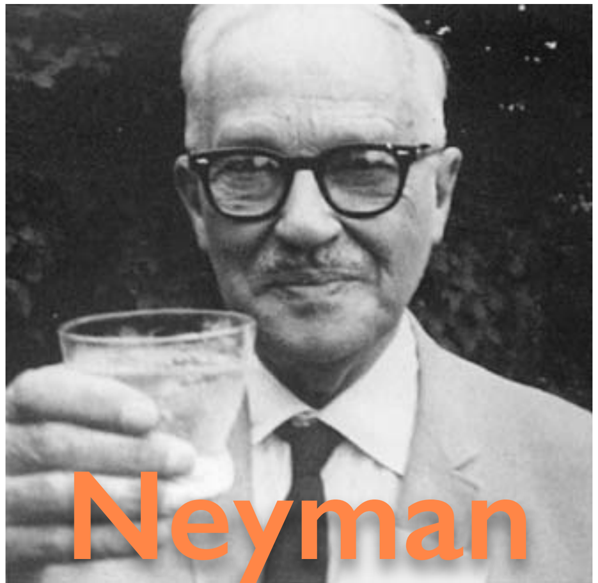
Table of error types		Null hypothesis (H_0) is	
		Valid/True	**Invalid/False**
Judgment of Null Hypothesis (H_0)	**Reject**	Type I error (False Positive, α)	Correct inference (True Positive, 1-β)
	Fail to reject	Correct inference (True Negative, 1-α)	Type II error (False Negative, β)

Type I = True H_0 but reject it (False Positive)

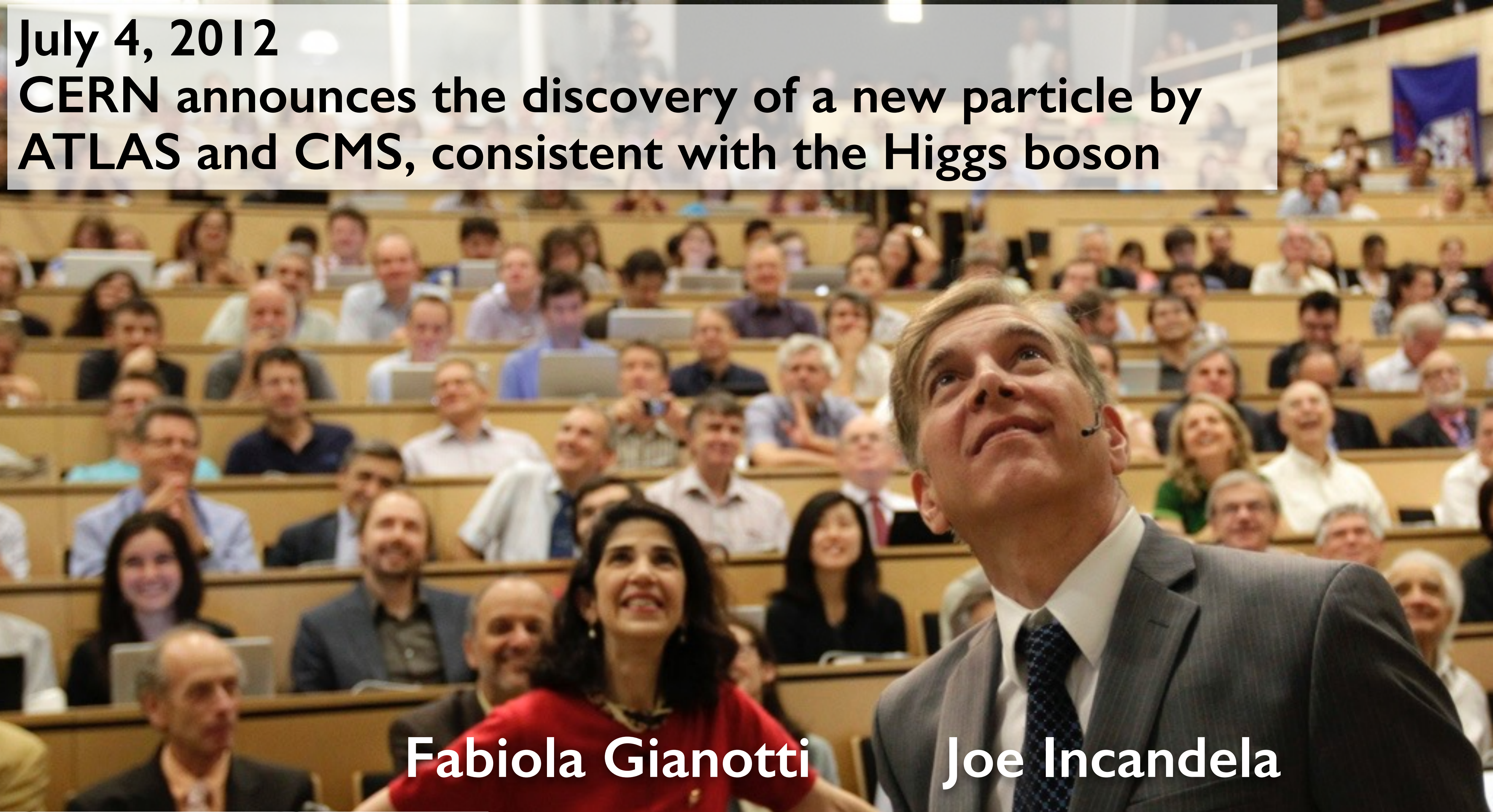
Type II = False H_0 but fail to reject it (False Negative)

- Power: 1-β
- Want to maximize power for a fixed false positive rate
- Particle physics has a tradition of claiming discovery at $5\sigma \Rightarrow p_0 = 2.9\times10^{-7}$ = 1 in 3.5 million, and presents exclusion with p_0 = 5%, (95% CL "coverage").
- Neyman-Pearson lemma (1933): the most powerful test for fixed α is the likelihood ratio:

$$\frac{L(x|H_0)}{L(x|H_1)} > k_\alpha$$



July 4, 2012
CERN announces the discovery of a new particle by ATLAS and CMS, consistent with the Higgs boson



Fabiola Gianotti **Joe Incandela**

ATLAS and CMS spokespersons

The New York Times

Late Edition

NEW YORK, THURSDAY, JULY 5, 2012

$2.50

ROMNEY NOW SAYS HEALTH MANDATE BY OBAMA IS A TAX

SHIFT RENEWS CRITICISM

Move Aligns Him With Conservative Voices Within His Party

Physicists Find Elusive Particle Seen as Key to Universe

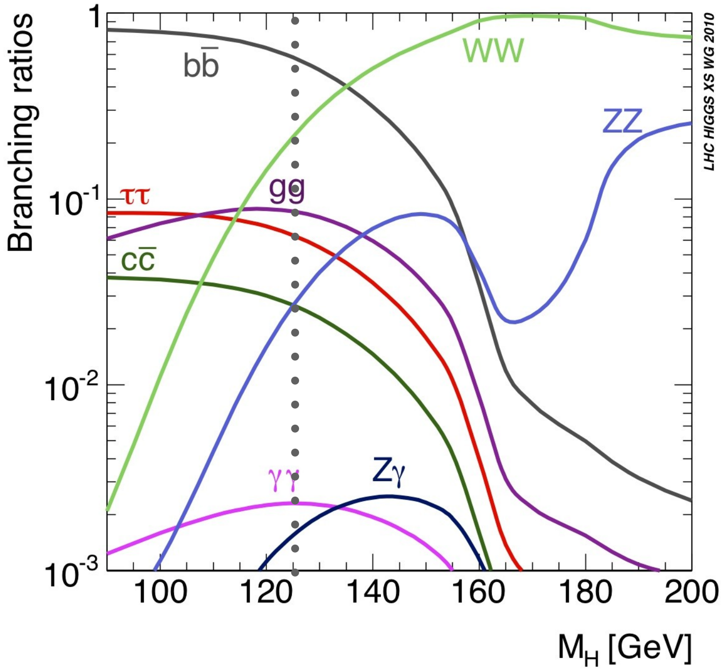
July 5 cover of the New York Times:

“Physicists Find Elusive Particle Seen as Key to the Universe”

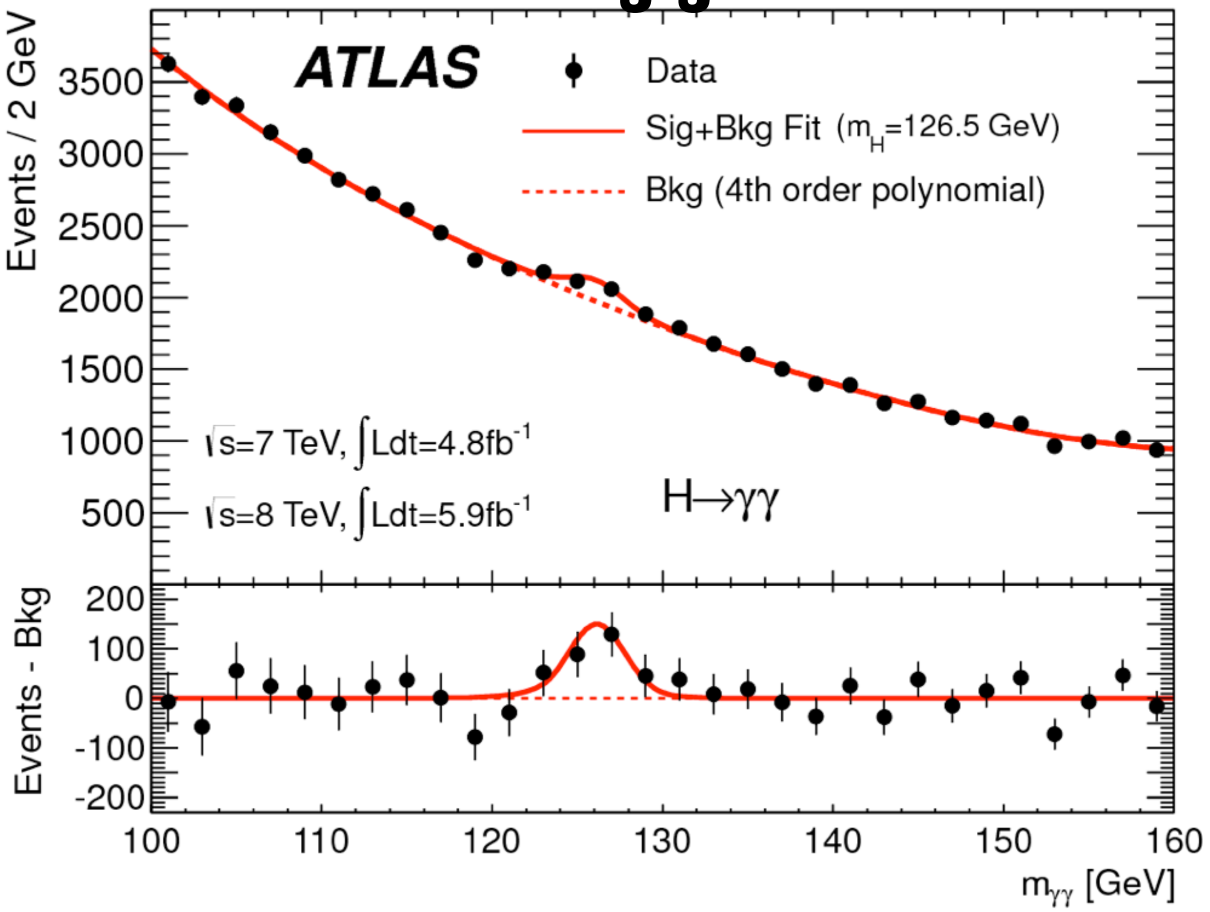
Higgs discovery

- Two channels with precise mass measurements: $H\to\gamma\gamma$ and $H\to ZZ\to 4l$.
- $H\to WW$ observes a broad but clear excess.

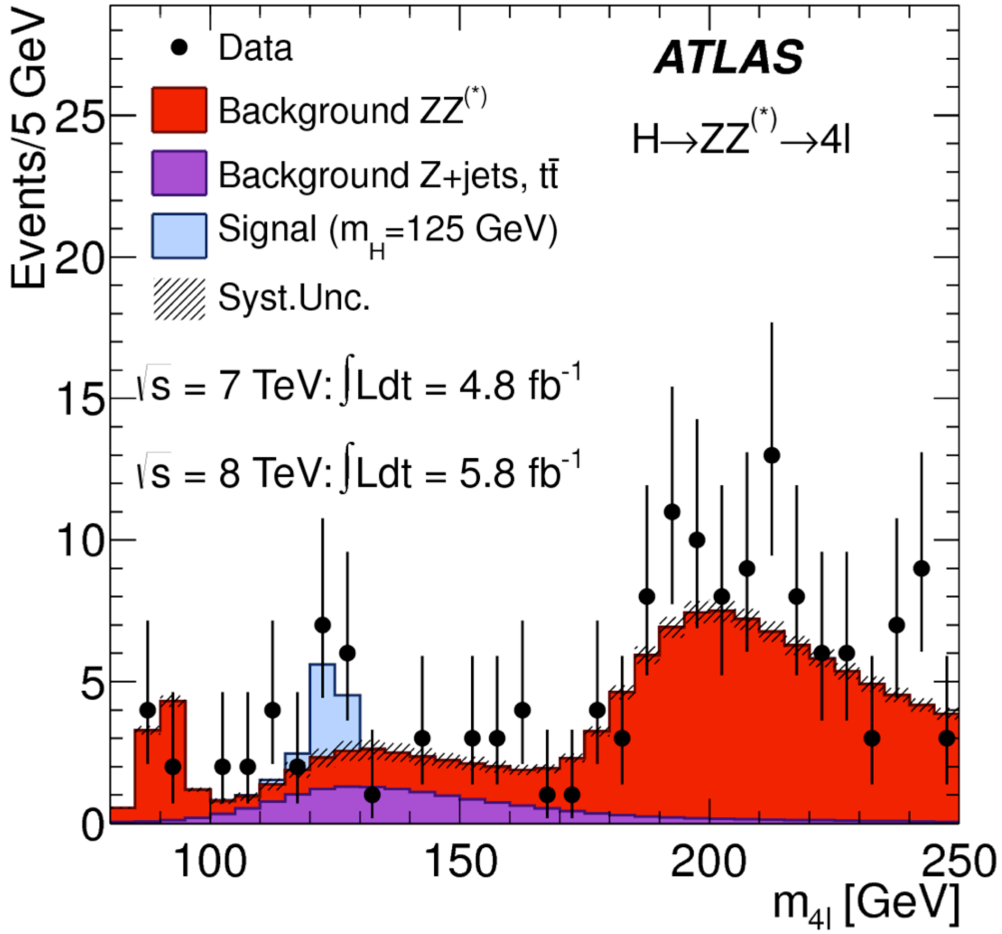
channel	bb	ττ	WW	ZZ	γγ
BR	58%	6%	22%	3%	0.2%



$H\to\gamma\gamma$

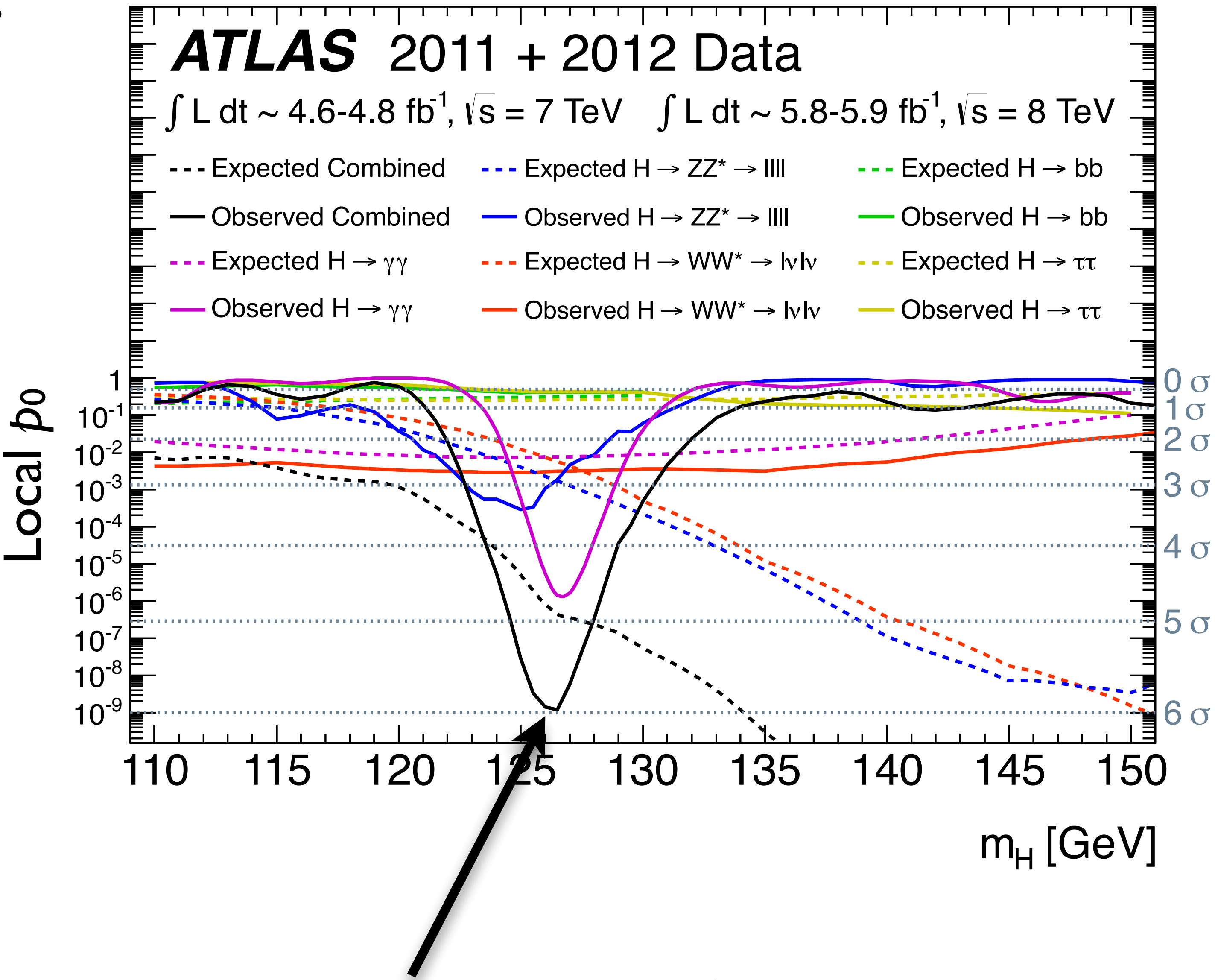


$H\to ZZ\to 4l$

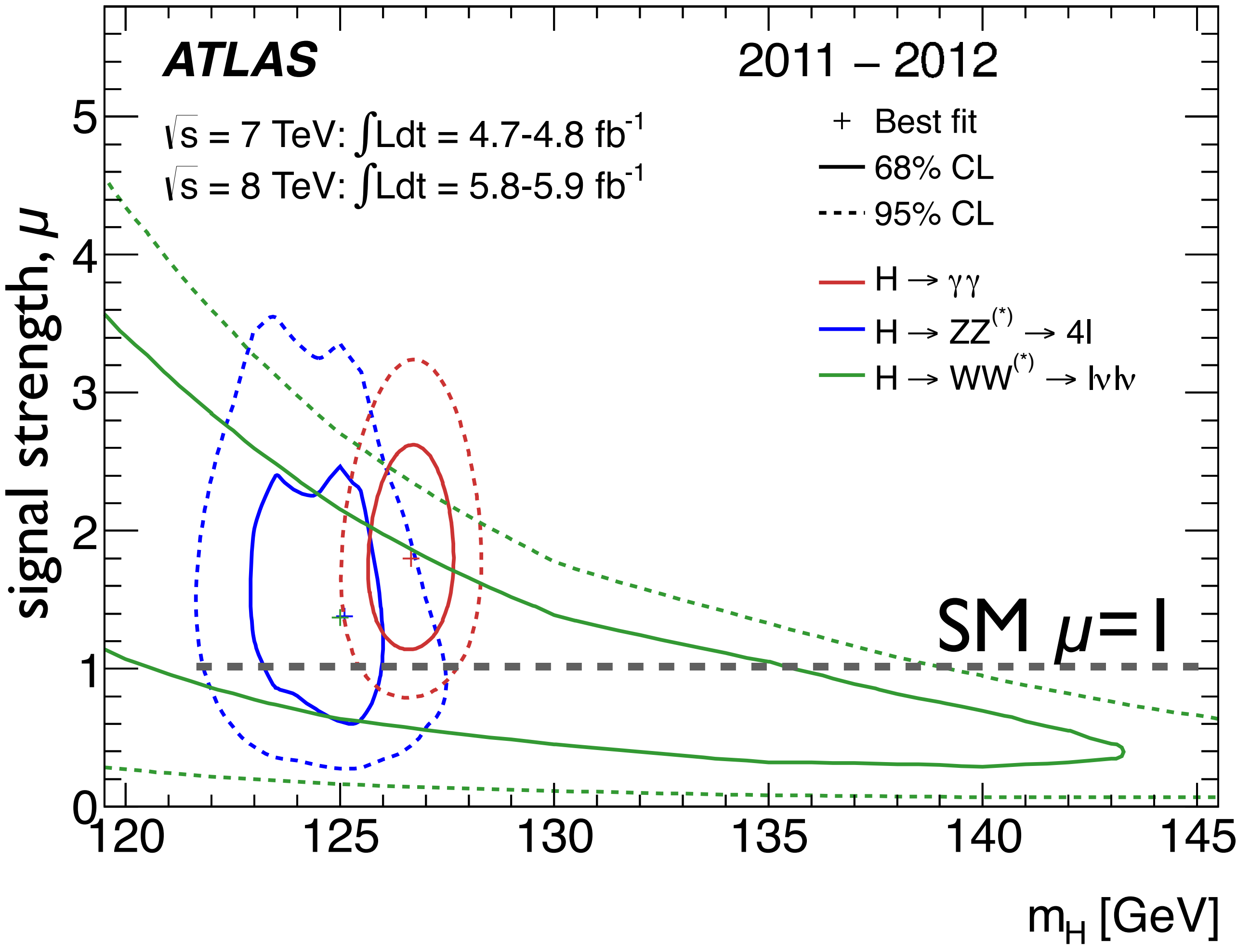


Higgs Confidence

Inconsistent with background only



Consistent with SM Higgs



- Local $p_0 = 1.7\times10^{-9}$, corresponding to 5.9

[arxiv: 1207.7214]

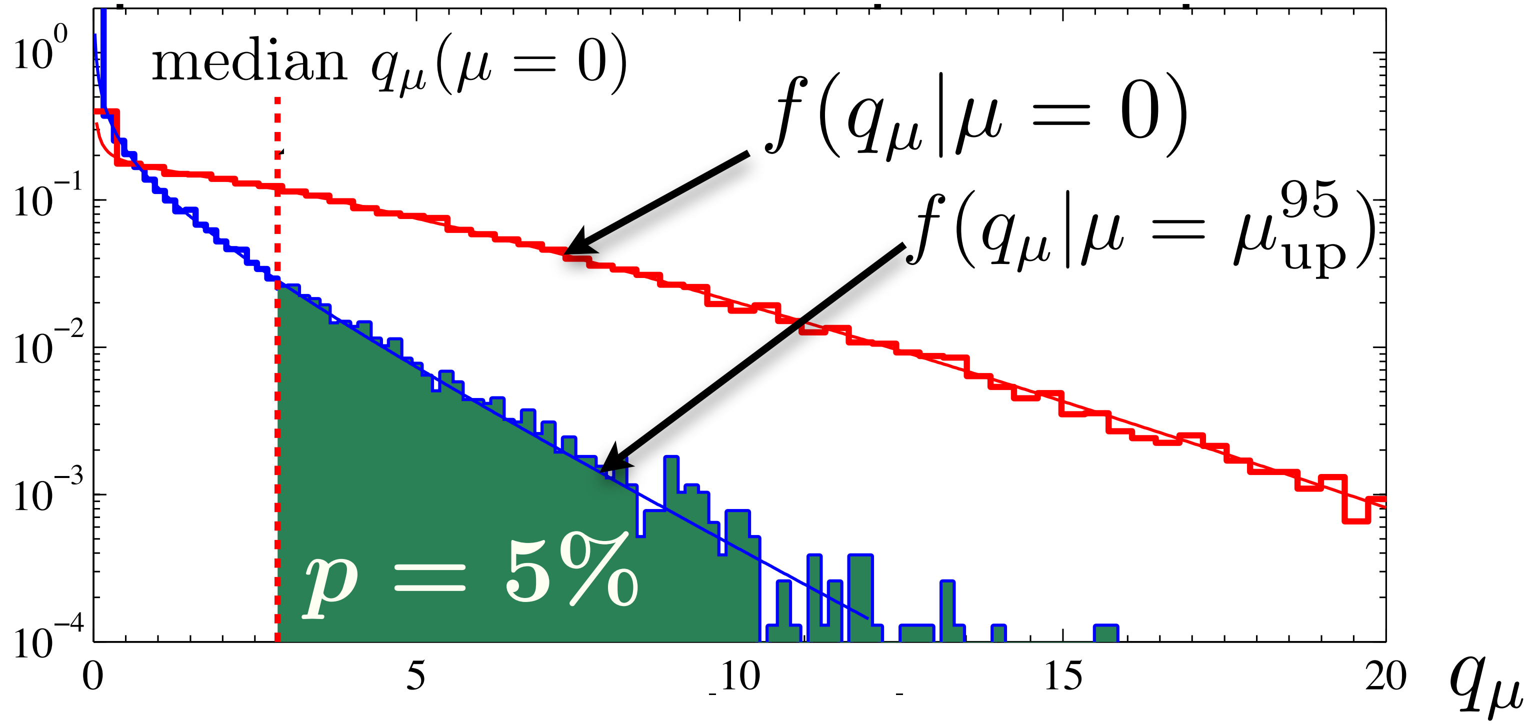
Excluding instead of discovering

- Construct test statistic:

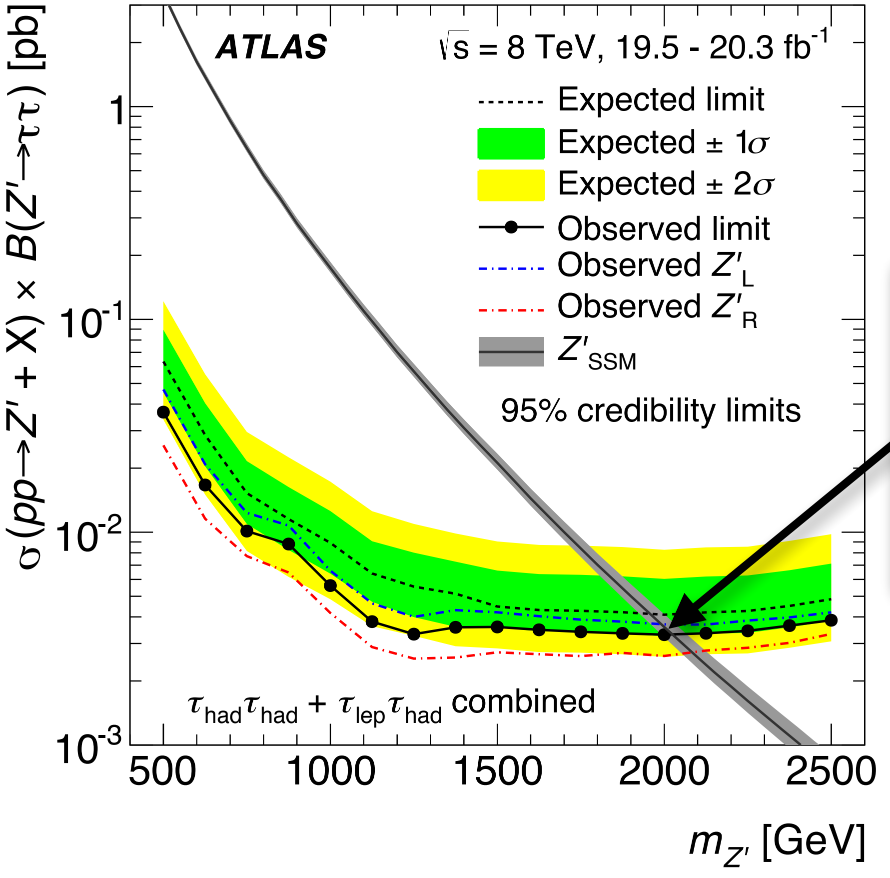
$$q_\mu = -2\ln(\lambda(\mu))$$

- Wilks' theorem: asymptotic to to a χ^2 distribution, larger values indicate greater incompatibility.
- Throw Monte Carlo pseudo experiments to find μ_{up}^{95} which has a p-value of 5%.
- If this signal strength were there, only 5% of experiments would have higher q_μ. ⇒ 95% CL or 2σ

Sample the test statistic with pseudo experiments



Example limit on new physics: Z'→ττ



(my PhD thesis)

excludes $m(Z'_{SSM}) < 2.02$ TeV @ 95% CL

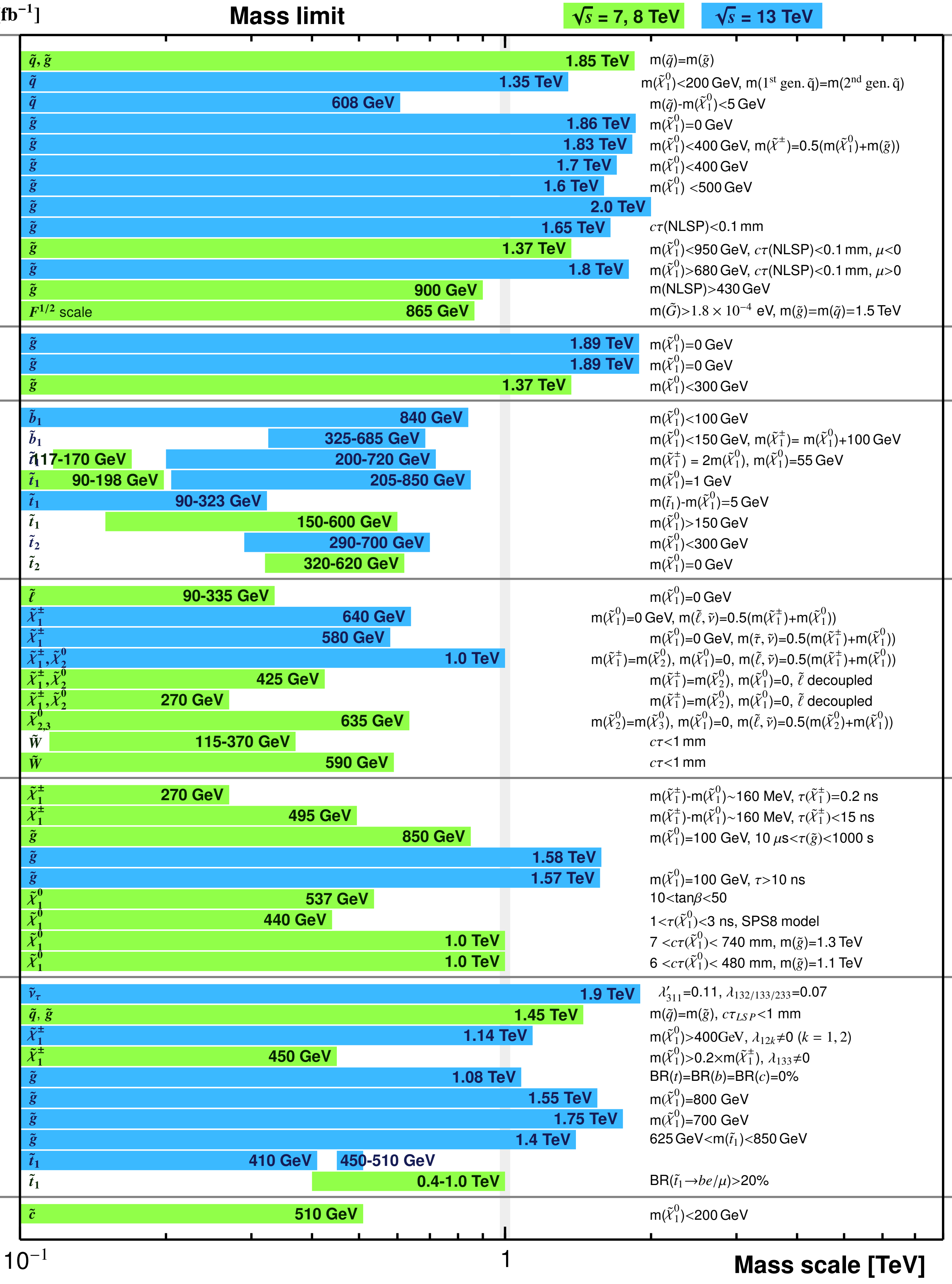
[arxiv:1007.1727, 1502.07177]

ATLAS SUSY Searches* - 95% CL Lower Limits

Status: August 2016

ATLAS Preliminary

$\sqrt{s}$ = 7, 8, 13 TeV

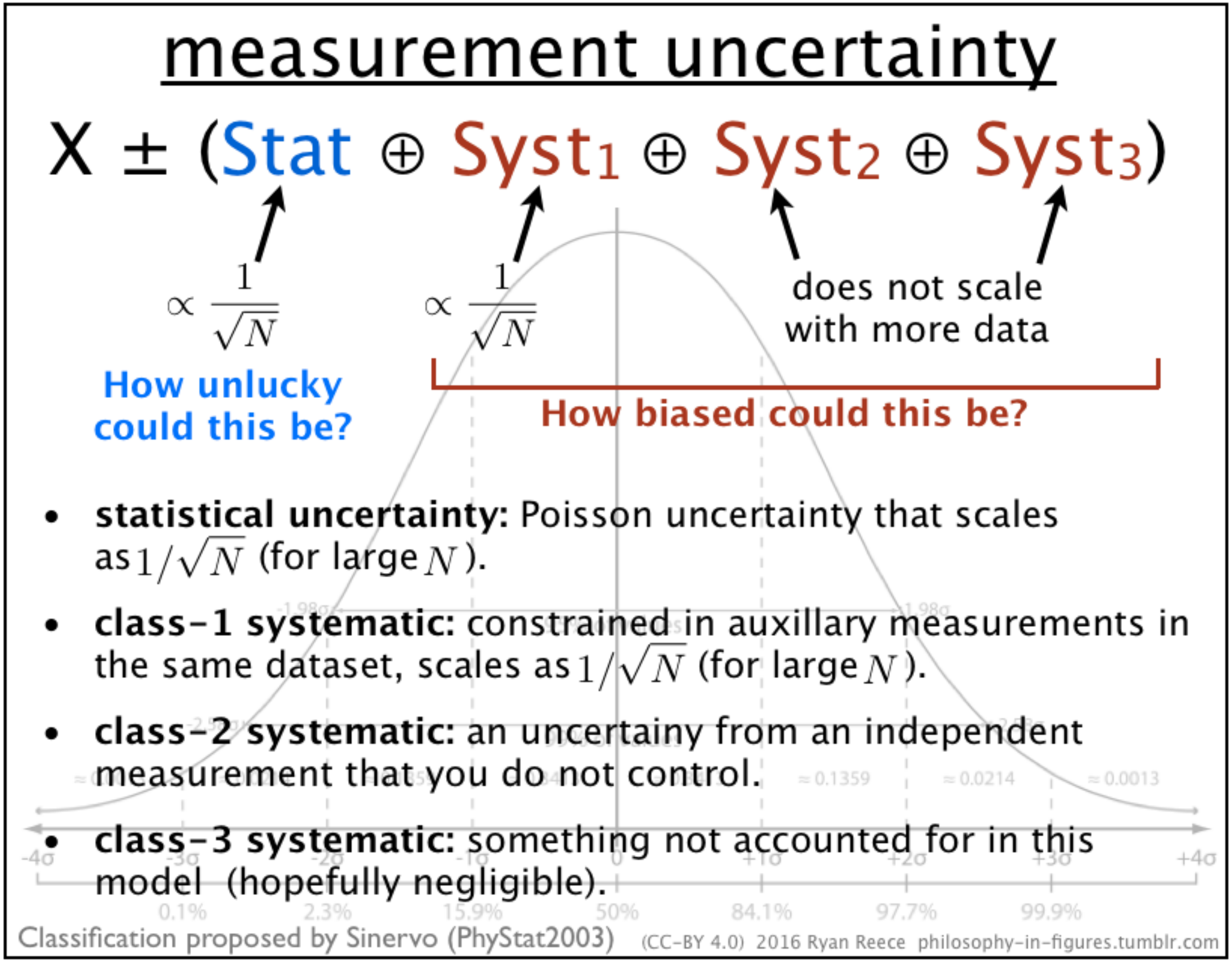


🟩 $\sqrt{s}$ = 7, 8 TeV 🟦 $\sqrt{s}$ = 13 TeV

	Model	e,μ,τ,γ	Jets	E_T^{miss}	$\int\mathcal{L}\,dt[\mathrm{fb}^{-1}]$	Mass limit			Reference
Inclusive Searches	MSUGRA/CMSSM	0-3 e,μ/1-2 τ	2-10 jets/3 b	Yes	20.3	$\tilde{q},\tilde{g}$	1.85 TeV 🟩	m($\tilde{q}$)=m($\tilde{g}$)	1507.05525
	$\tilde{q}\tilde{q}$, $\tilde{q}\rightarrow q\tilde{\chi}_1^0$	0	2-6 jets	Yes	13.3	$\tilde{q}$	1.35 TeV 🟦	m($\tilde{\chi}_1^0$)<200 GeV, m(1$^{\mathrm{st}}$ gen. q̃)=m(2$^{\mathrm{nd}}$ gen. q̃)	ATLAS-CONF-2016-078
	$\tilde{q}\tilde{q}$, $\tilde{q}\rightarrow q\tilde{\chi}_1^0$ (compressed)	mono-jet	1-3 jets	Yes	3.2	$\tilde{q}$	608 GeV 🟦	m($\tilde{q}$)-m($\tilde{\chi}_1^0$)<5 GeV	1604.07773
	$\tilde{g}\tilde{g}$, $\tilde{g}\rightarrow q\bar{q}\tilde{\chi}_1^0$	0	2-6 jets	Yes	13.3	$\tilde{g}$	1.86 TeV 🟦	m($\tilde{\chi}_1^0$)=0 GeV	ATLAS-CONF-2016-078
	$\tilde{g}\tilde{g}$, $\tilde{g}\rightarrow qq\tilde{\chi}_1^{\pm}\rightarrow qqW^{\pm}\tilde{\chi}_1^0$	0	2-6 jets	Yes	13.3	$\tilde{g}$	1.83 TeV 🟦	m($\tilde{\chi}_1^0$)<400 GeV, m($\tilde{\chi}^{\pm}$)=0.5(m($\tilde{\chi}_1^0$)+m($\tilde{g}$))	ATLAS-CONF-2016-078
	$\tilde{g}\tilde{g}$, $\tilde{g}\rightarrow qq(\ell\ell/\nu\nu)\tilde{\chi}_1^0$	3 e,μ	4 jets	-	13.2	$\tilde{g}$	1.7 TeV 🟦	m($\tilde{\chi}_1^0$)<400 GeV	ATLAS-CONF-2016-037
	$\tilde{g}\tilde{g}$, $\tilde{g}\rightarrow qqWZ\tilde{\chi}_1^0$	2 e,μ (SS)	0-3 jets	Yes	13.2	$\tilde{g}$	1.6 TeV 🟦	m($\tilde{\chi}_1^0$) <500 GeV	ATLAS-CONF-2016-037
	GMSB ($\tilde{\ell}$ NLSP)	1-2 τ + 0-1 ℓ	0-2 jets	Yes	3.2	$\tilde{g}$	2.0 TeV 🟦		1607.05979
	GGM (bino NLSP)	2 γ	-	Yes	3.2	$\tilde{g}$	1.65 TeV 🟦	$c\tau$(NLSP)<0.1 mm	1606.09150
	GGM (higgsino-bino NLSP)	γ	1 b	Yes	20.3	$\tilde{g}$	1.37 TeV 🟩	m($\tilde{\chi}_1^0$)<950 GeV, $c\tau$(NLSP)<0.1 mm, μ<0	1507.05493
	GGM (higgsino-bino NLSP)	γ	2 jets	Yes	13.3	$\tilde{g}$	1.8 TeV 🟦	m($\tilde{\chi}_1^0$)>680 GeV, $c\tau$(NLSP)<0.1 mm, μ>0	ATLAS-CONF-2016-066
	GGM (higgsino NLSP)	2 e,μ (Z)	2 jets	Yes	20.3	$\tilde{g}$	900 GeV 🟩	m(NLSP)>430 GeV	1503.03290
	Gravitino LSP	0	mono-jet	Yes	20.3	$F^{1/2}$ scale	865 GeV 🟩	m($\tilde{G}$)>1.8×10^{-4} eV, m($\tilde{g}$)=m($\tilde{q}$)=1.5 TeV	1502.01518
3^{rd} gen. $\tilde{g}$ med.	$\tilde{g}\tilde{g}$, $\tilde{g}\rightarrow b\bar{b}\tilde{\chi}_1^0$	0	3 b	Yes	14.8	$\tilde{g}$	1.89 TeV 🟦	m($\tilde{\chi}_1^0$)=0 GeV	ATLAS-CONF-2016-052
	$\tilde{g}\tilde{g}$, $\tilde{g}\rightarrow t\bar{t}\tilde{\chi}_1^0$	0-1 e,μ	3 b	Yes	14.8	$\tilde{g}$	1.89 TeV 🟦	m($\tilde{\chi}_1^0$)=0 GeV	ATLAS-CONF-2016-052
	$\tilde{g}\tilde{g}$, $\tilde{g}\rightarrow b\bar{t}\tilde{\chi}_1^+$	0-1 e,μ	3 b	Yes	20.1	$\tilde{g}$	1.37 TeV 🟩	m($\tilde{\chi}_1^0$)<300 GeV	1407.0600
3^{rd} gen. squarks direct production	$\tilde{b}_1\tilde{b}_1$, $\tilde{b}_1\rightarrow b\tilde{\chi}_1^0$	0	2 b	Yes	3.2	$\tilde{b}_1$	840 GeV 🟦	m($\tilde{\chi}_1^0$)<100 GeV	1606.08772
	$\tilde{b}_1\tilde{b}_1$, $\tilde{b}_1\rightarrow t\tilde{\chi}_1^{\pm}$	2 e,μ (SS)	1 b	Yes	13.2	$\tilde{b}_1$	325-685 GeV 🟦	m($\tilde{\chi}_1^0$)<150 GeV, m($\tilde{\chi}_1^{\pm}$)= m($\tilde{\chi}_1^0$)+100 GeV	ATLAS-CONF-2016-037
	$\tilde{t}_1\tilde{t}_1$, $\tilde{t}_1\rightarrow b\tilde{\chi}_1^{\pm}$	0-2 e,μ	1-2 b	Yes	4.7/13.3	$\tilde{t}_1$	117-170 GeV 🟩 200-720 GeV 🟦	m($\tilde{\chi}_1^{\pm}$) = 2m($\tilde{\chi}_1^0$), m($\tilde{\chi}_1^0$)=55 GeV	1209.2102, ATLAS-CONF-2016-077
	$\tilde{t}_1\tilde{t}_1$, $\tilde{t}_1\rightarrow Wb\tilde{\chi}_1^0$ or $t\tilde{\chi}_1^0$	0-2 e,μ	0-2 jets/1-2 b	Yes	4.7/13.3	$\tilde{t}_1$	90-198 GeV 🟩 205-850 GeV 🟦	m($\tilde{\chi}_1^0$)=1 GeV	1506.08616, ATLAS-CONF-2016-077
	$\tilde{t}_1\tilde{t}_1$, $\tilde{t}_1\rightarrow c\tilde{\chi}_1^0$	0	mono-jet	Yes	3.2	$\tilde{t}_1$	90-323 GeV 🟦	m($\tilde{t}_1$)-m($\tilde{\chi}_1^0$)=5 GeV	1604.07773
	$\tilde{t}_1\tilde{t}_1$(natural GMSB)	2 e,μ (Z)	1 b	Yes	20.3	$\tilde{t}_1$	150-600 GeV 🟩	m($\tilde{\chi}_1^0$)>150 GeV	1403.5222
	$\tilde{t}_2\tilde{t}_2$, $\tilde{t}_2\rightarrow\tilde{t}_1+Z$	3 e,μ (Z)	1 b	Yes	13.3	$\tilde{t}_2$	290-700 GeV 🟦	m($\tilde{\chi}_1^0$)<300 GeV	ATLAS-CONF-2016-038
	$\tilde{t}_2\tilde{t}_2$, $\tilde{t}_2\rightarrow\tilde{t}_1+h$	1 e,μ	6 jets + 2 b	Yes	20.3	$\tilde{t}_2$	320-620 GeV 🟩	m($\tilde{\chi}_1^0$)=0 GeV	1506.08616
EW direct	$\tilde{\ell}_{\mathrm{L,R}}\tilde{\ell}_{\mathrm{L,R}}$, $\tilde{\ell}\rightarrow\ell\tilde{\chi}_1^0$	2 e,μ	0	Yes	20.3	$\tilde{\ell}$	90-335 GeV 🟩	m($\tilde{\chi}_1^0$)=0 GeV	1403.5294
	$\tilde{\chi}_1^+\tilde{\chi}_1^-$, $\tilde{\chi}_1^+\rightarrow\tilde{\ell}\nu(\ell\tilde{\nu})$	2 e,μ	0	Yes	13.3	$\tilde{\chi}_1^{\pm}$	640 GeV 🟦	m($\tilde{\chi}_1^0$)=0 GeV, m($\tilde{\ell},\tilde{\nu}$)=0.5(m($\tilde{\chi}_1^{\pm}$)+m($\tilde{\chi}_1^0$))	ATLAS-CONF-2016-096
	$\tilde{\chi}_1^+\tilde{\chi}_1^-$, $\tilde{\chi}_1^+\rightarrow\tilde{\tau}\nu(\tau\tilde{\nu})$	2 τ	-	Yes	14.8	$\tilde{\chi}_1^{\pm}$	580 GeV 🟦	m($\tilde{\chi}_1^0$)=0 GeV, m($\tilde{\tau},\tilde{\nu}$)=0.5(m($\tilde{\chi}_1^{\pm}$)+m($\tilde{\chi}_1^0$))	ATLAS-CONF-2016-093
	$\tilde{\chi}_1^{\pm}\tilde{\chi}_2^0\rightarrow\tilde{\ell}_{\mathrm{L}}\nu\tilde{\ell}_{\mathrm{L}}\ell(\tilde{\nu}\nu)$, $\ell\tilde{\nu}\tilde{\ell}_{\mathrm{L}}\ell(\tilde{\nu}\nu)$	3 e,μ	0	Yes	13.3	$\tilde{\chi}_1^{\pm},\tilde{\chi}_2^0$	1.0 TeV 🟦	m($\tilde{\chi}_1^{\pm}$)=m($\tilde{\chi}_2^0$), m($\tilde{\chi}_1^0$)=0, m($\tilde{\ell},\tilde{\nu}$)=0.5(m($\tilde{\chi}_1^{\pm}$)+m($\tilde{\chi}_1^0$))	ATLAS-CONF-2016-096
	$\tilde{\chi}_1^{\pm}\tilde{\chi}_2^0\rightarrow W\tilde{\chi}_1^0Z\tilde{\chi}_1^0$	2-3 e,μ	0-2 jets	Yes	20.3	$\tilde{\chi}_1^{\pm},\tilde{\chi}_2^0$	425 GeV 🟩	m($\tilde{\chi}_1^{\pm}$)=m($\tilde{\chi}_2^0$), m($\tilde{\chi}_1^0$)=0, $\tilde{\ell}$ decoupled	1403.5294, 1402.7029
	$\tilde{\chi}_1^{\pm}\tilde{\chi}_2^0\rightarrow W\tilde{\chi}_1^0h\tilde{\chi}_1^0$, $h\rightarrow b\bar{b}/WW/\tau\tau/\gamma\gamma$	e,μ,γ	0-2 b	Yes	20.3	$\tilde{\chi}_1^{\pm},\tilde{\chi}_2^0$	270 GeV 🟩	m($\tilde{\chi}_1^{\pm}$)=m($\tilde{\chi}_2^0$), m($\tilde{\chi}_1^0$)=0, $\tilde{\ell}$ decoupled	1501.07110
	$\tilde{\chi}_2^0\tilde{\chi}_3^0$, $\tilde{\chi}_{2,3}^0\rightarrow\tilde{\ell}_{\mathrm{R}}\ell$	4 e,μ	0	Yes	20.3	$\tilde{\chi}_{2,3}^0$	635 GeV 🟩	m($\tilde{\chi}_2^0$)=m($\tilde{\chi}_3^0$), m($\tilde{\chi}_1^0$)=0, m($\tilde{\ell},\tilde{\nu}$)=0.5(m($\tilde{\chi}_2^0$)+m($\tilde{\chi}_1^0$))	1405.5086
	GGM (wino NLSP) weak prod.	1 $e,\mu+\gamma$	-	Yes	20.3	$\tilde{W}$	115-370 GeV 🟩	$c\tau$<1 mm	1507.05493
	GGM (bino NLSP) weak prod.	2 γ	-	Yes	20.3	$\tilde{W}$	590 GeV 🟩	$c\tau$<1 mm	1507.05493
Long-lived particles	Direct $\tilde{\chi}_1^+\tilde{\chi}_1^-$ prod., long-lived $\tilde{\chi}_1^{\pm}$	Disapp. trk	1 jet	Yes	20.3	$\tilde{\chi}_1^{\pm}$	270 GeV 🟩	m($\tilde{\chi}_1^{\pm}$)-m($\tilde{\chi}_1^0$)~160 MeV, τ($\tilde{\chi}_1^{\pm}$)=0.2 ns	1310.3675
	Direct $\tilde{\chi}_1^+\tilde{\chi}_1^-$ prod., long-lived $\tilde{\chi}_1^{\pm}$	dE/dx trk	-	Yes	18.4	$\tilde{\chi}_1^{\pm}$	495 GeV 🟩	m($\tilde{\chi}_1^{\pm}$)-m($\tilde{\chi}_1^0$)~160 MeV, τ($\tilde{\chi}_1^{\pm}$)<15 ns	1506.05332
	Stable, stopped $\tilde{g}$ R-hadron	0	1-5 jets	Yes	27.9	$\tilde{g}$	850 GeV 🟩	m($\tilde{\chi}_1^0$)=100 GeV, 10 μs<τ($\tilde{g}$)<1000 s	1310.6584
	Stable $\tilde{g}$ R-hadron	trk	-	-	3.2	$\tilde{g}$	1.58 TeV 🟦		1606.05129
	Metastable $\tilde{g}$ R-hadron	dE/dx trk	-	-	3.2	$\tilde{g}$	1.57 TeV 🟦	m($\tilde{\chi}_1^0$)=100 GeV, τ>10 ns	1604.04520
	GMSB, stable $\tilde{\tau}$, $\tilde{\chi}_1^0\rightarrow\tilde{\tau}(\tilde{e},\tilde{\mu})+\tau(e,\mu)$	1-2 μ	-	-	19.1	$\tilde{\chi}_1^0$	537 GeV 🟩	10<tanβ<50	1411.6795
	GMSB, $\tilde{\chi}_1^0\rightarrow\gamma\tilde{G}$, long-lived $\tilde{\chi}_1^0$	2 γ	-	Yes	20.3	$\tilde{\chi}_1^0$	440 GeV 🟩	1<τ($\tilde{\chi}_1^0$)<3 ns, SPS8 model	1409.5542
	$\tilde{g}\tilde{g}$, $\tilde{\chi}_1^0\rightarrow eev/e\mu\nu/\mu\mu\nu$	displ. $ee/e\mu/\mu\mu$	-	-	20.3	$\tilde{\chi}_1^0$	1.0 TeV 🟩	7 <$c\tau$($\tilde{\chi}_1^0$)< 740 mm, m($\tilde{g}$)=1.3 TeV	1504.05162
	GGM $\tilde{g}\tilde{g}$, $\tilde{\chi}_1^0\rightarrow Z\tilde{G}$	displ. vtx + jets	-	-	20.3	$\tilde{\chi}_1^0$	1.0 TeV 🟩	6 <$c\tau$($\tilde{\chi}_1^0$)< 480 mm, m($\tilde{g}$)=1.1 TeV	1504.05162
RPV	LFV $pp\rightarrow\tilde{\nu}_\tau+X$, $\tilde{\nu}_\tau\rightarrow e\mu/e\tau/\mu\tau$	$e\mu$,$e\tau$,$\mu\tau$	-	-	3.2	$\tilde{\nu}_\tau$	1.9 TeV 🟦	λ'_{311}=0.11, $\lambda_{132/133/233}$=0.07	1607.08079
	Bilinear RPV CMSSM	2 e,μ (SS)	0-3 b	Yes	20.3	$\tilde{q},\tilde{g}$	1.45 TeV 🟩	m($\tilde{q}$)=m($\tilde{g}$), $c\tau_{LSP}$<1 mm	1404.2500
	$\tilde{\chi}_1^+\tilde{\chi}_1^-$, $\tilde{\chi}_1^+\rightarrow W\tilde{\chi}_1^0$, $\tilde{\chi}_1^0\rightarrow ee\nu, e\mu\nu, \mu\mu\nu$	4 e,μ	-	Yes	13.3	$\tilde{\chi}_1^{\pm}$	1.14 TeV 🟦	m($\tilde{\chi}_1^0$)>400GeV, $\lambda_{12k}\neq0$ ($k=1,2$)	ATLAS-CONF-2016-075
	$\tilde{\chi}_1^+\tilde{\chi}_1^-$, $\tilde{\chi}_1^+\rightarrow W\tilde{\chi}_1^0$, $\tilde{\chi}_1^0\rightarrow\tau\tau\nu_e, e\tau\nu_\tau$	3 $e,\mu+\tau$	-	Yes	20.3	$\tilde{\chi}_1^{\pm}$	450 GeV 🟩	m($\tilde{\chi}_1^0$)>0.2×m($\tilde{\chi}_1^{\pm}$), $\lambda_{133}\neq0$	1405.5086
	$\tilde{g}\tilde{g}$, $\tilde{g}\rightarrow qqq$	0	4-5 large-R jets	-	14.8	$\tilde{g}$	1.08 TeV 🟦	BR(t)=BR(b)=BR(c)=0%	ATLAS-CONF-2016-057
	$\tilde{g}\tilde{g}$, $\tilde{g}\rightarrow qq\tilde{\chi}_1^0$, $\tilde{\chi}_1^0\rightarrow qqq$	0	4-5 large-R jets	-	14.8	$\tilde{g}$	1.55 TeV 🟦	m($\tilde{\chi}_1^0$)=800 GeV	ATLAS-CONF-2016-057
	$\tilde{g}\tilde{g}$, $\tilde{g}\rightarrow t\bar{t}\tilde{\chi}_1^0$, $\tilde{\chi}_1^0\rightarrow qqq$	1 e,μ	8-10 jets/0-4 b	-	14.8	$\tilde{g}$	1.75 TeV 🟦	m($\tilde{\chi}_1^0$)=700 GeV	ATLAS-CONF-2016-094
	$\tilde{g}\tilde{g}$, $\tilde{g}\rightarrow\tilde{t}_1t$, $\tilde{t}_1\rightarrow bs$	1 e,μ	8-10 jets/0-4 b	-	14.8	$\tilde{g}$	1.4 TeV 🟦	625 GeV<m($\tilde{t}_1$)<850 GeV	ATLAS-CONF-2016-094
	$\tilde{t}_1\tilde{t}_1$, $\tilde{t}_1\rightarrow bs$	0	2 jets + 2 b	-	15.4	$\tilde{t}_1$	410 GeV 🟦 450-510 GeV 🟦		ATLAS-CONF-2016-022, ATLAS-CONF-2016-084
	$\tilde{t}_1\tilde{t}_1$, $\tilde{t}_1\rightarrow b\ell$	2 e,μ	2 b	-	20.3	$\tilde{t}_1$	0.4-1.0 TeV 🟩	BR($\tilde{t}_1\rightarrow be/\mu$)>20%	ATLAS-CONF-2015-015
Other	Scalar charm, $\tilde{c}\rightarrow c\tilde{\chi}_1^0$	0	2 c	Yes	20.3	$\tilde{c}$	510 GeV 🟩	m($\tilde{\chi}_1^0$)<200 GeV	1501.01325

**Only a selection of the available mass limits on new states or phenomena is shown.*

Systematics



measurement uncertainty

$$X \pm (\text{Stat} \oplus \text{Syst}_1 \oplus \text{Syst}_2 \oplus \text{Syst}_3)$$

$\propto \frac{1}{\sqrt{N}}$

How unlucky could this be?

$\propto \frac{1}{\sqrt{N}}$

does not scale with more data

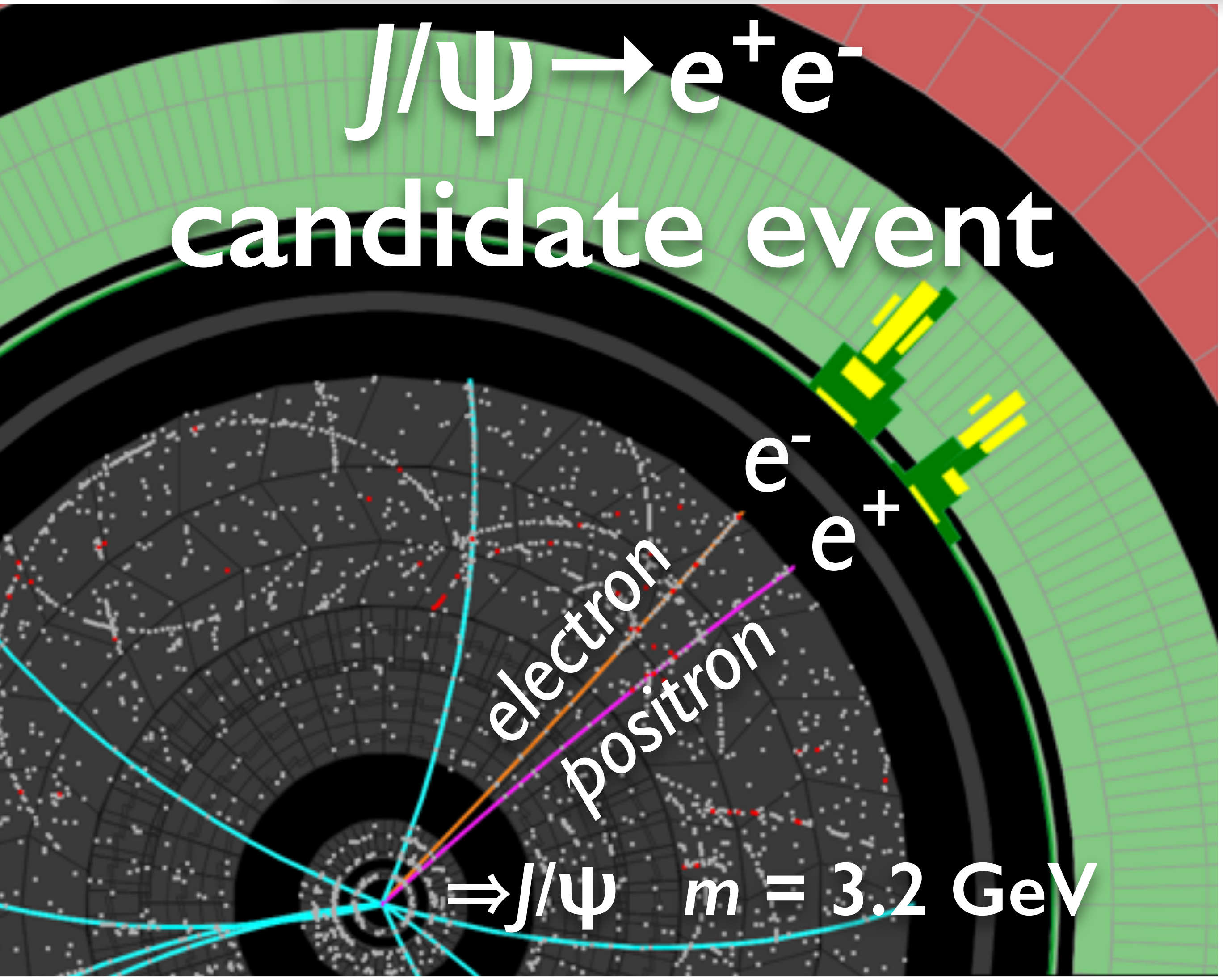
How biased could this be?

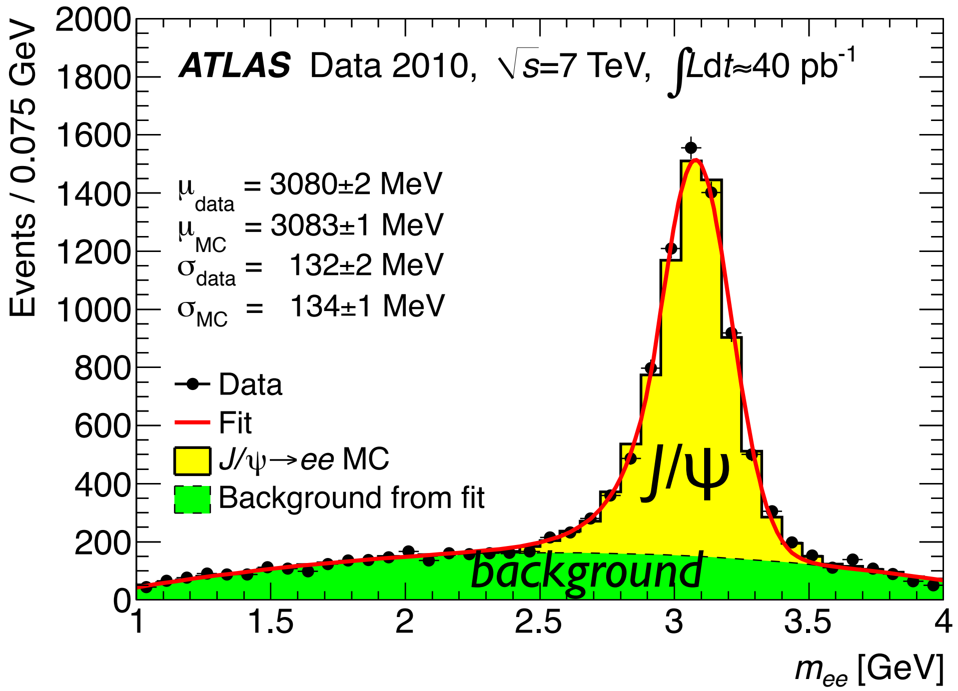
- **statistical uncertainty:** Poisson uncertainty that scales as $1/\sqrt{N}$ (for large N).
- **class-1 systematic:** constrained in auxillary measurements in the same dataset, scales as $1/\sqrt{N}$ (for large N).
- **class-2 systematic:** an uncertainy from an independent measurement that you do not control.
- **class-3 systematic:** something not accounted for in this model (hopefully negligible).

Classification proposed by Sinervo (PhyStat2003)

from: http://philosophy-in-figures.tumblr.com/

Real Patterns





What is an electron?

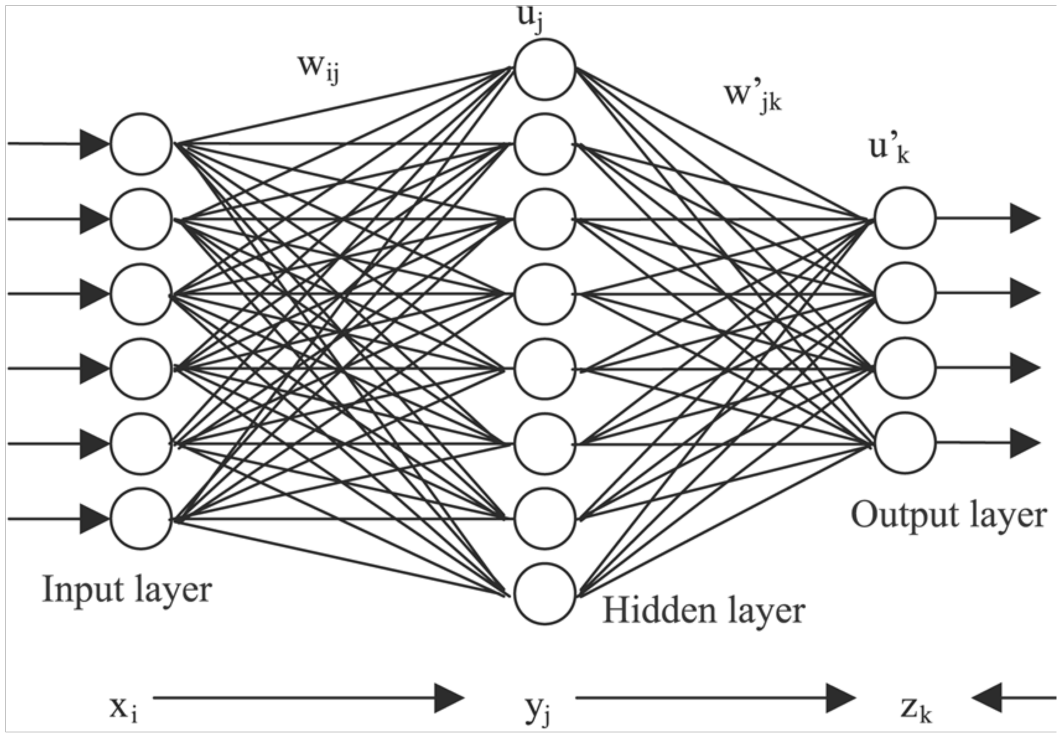
- An excitation in a ***Dirac spinor field*** representation of SU(2)xU(1), the *"Platonic electron"*.
- ***A software object*** with a reconstructed track and calorimeter deposit, passing some selection cuts, the *"pragmatist electron"*.
- ***A set of voltages*** and timings read-out from the detector, the *"Ramsified electron"*.

➡ Reality has a hierarchy of onion layers, but it has ***real patterns*** (Dennett 1991).

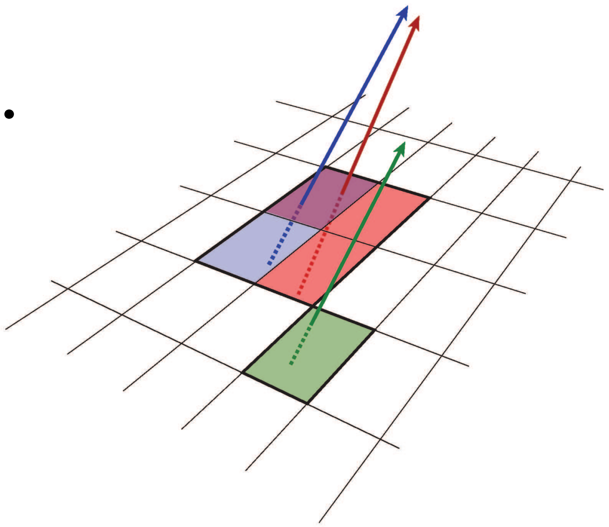
Machine Learning

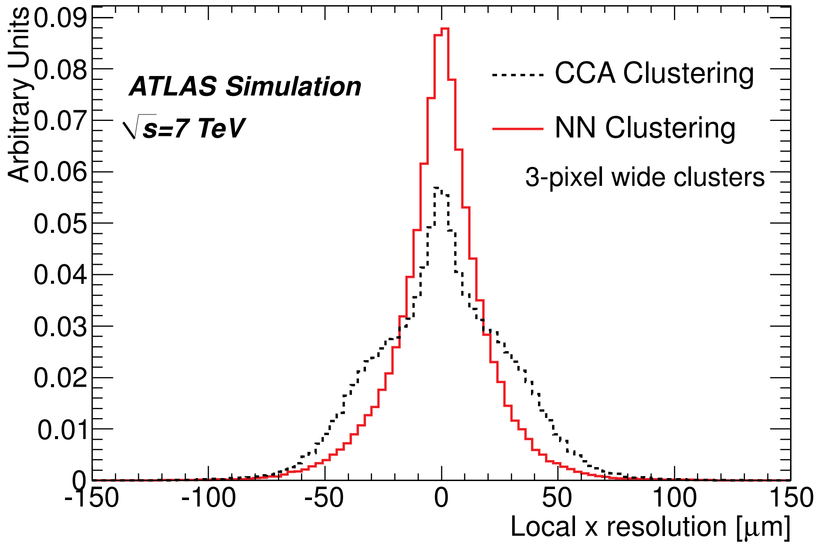
Neural Networks

- Inspired by the biological cortex
- Can be used for *classification* or *regression* with many input variables.
- Using NNs and other MVAs has been common in HEP for years, for pattern recognition, particle ID, event selection...
- In the past, always used *shallow* NNs.
- ATLAS uses NNs in many places, e.g. pixel clustering.
- Jet tagging for taus and b-quarks has used NNs in many iterations.



ATLAS pixel clustering with NNs



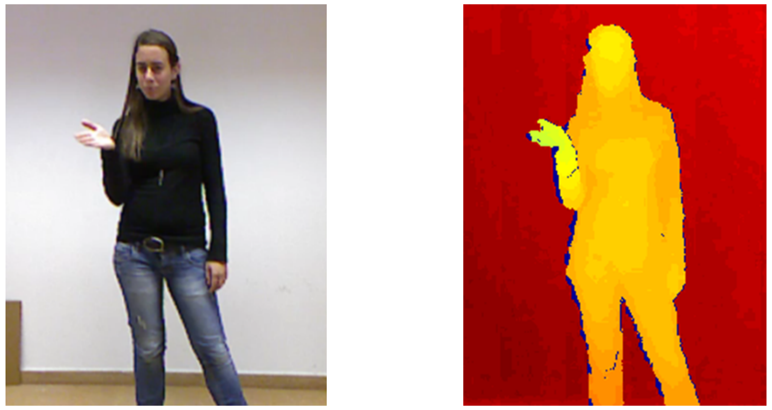


[1406.7690]

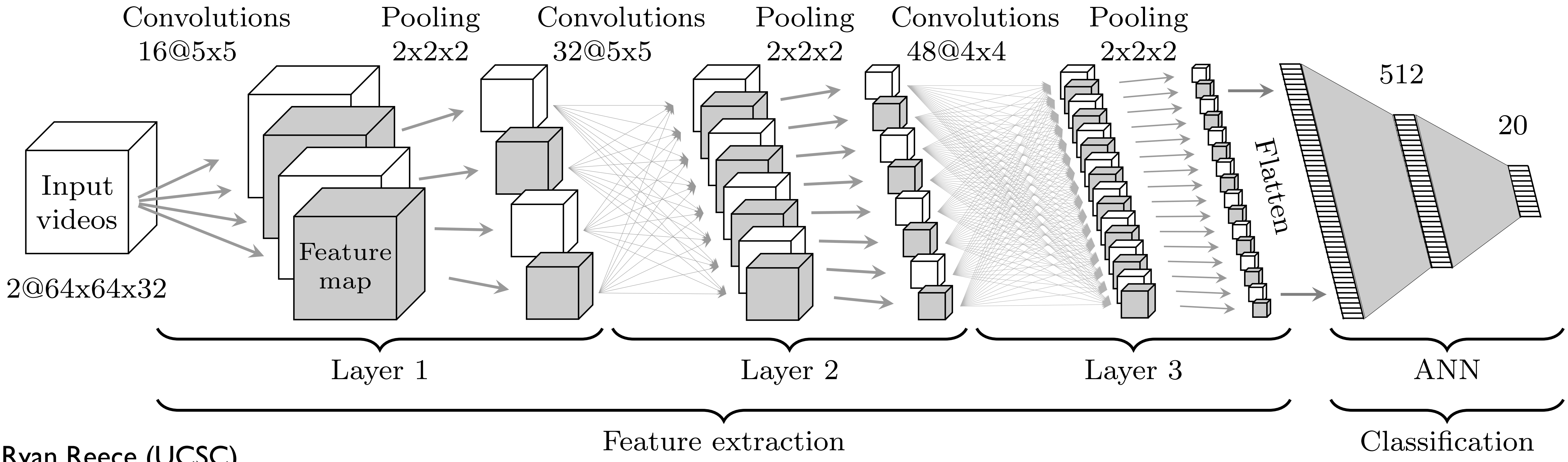
Examples of CNNs

- In 1990s, Yann LeCun pioneered Convolutional Neural Nets (CNN) and used them for Optical Character Recognition.
- Now it is standard in image recognition and captioning, NLP, computer vision, etc.

Pigou et al. (2014). Sign Language Recognition using Convolutional Neural Networks.

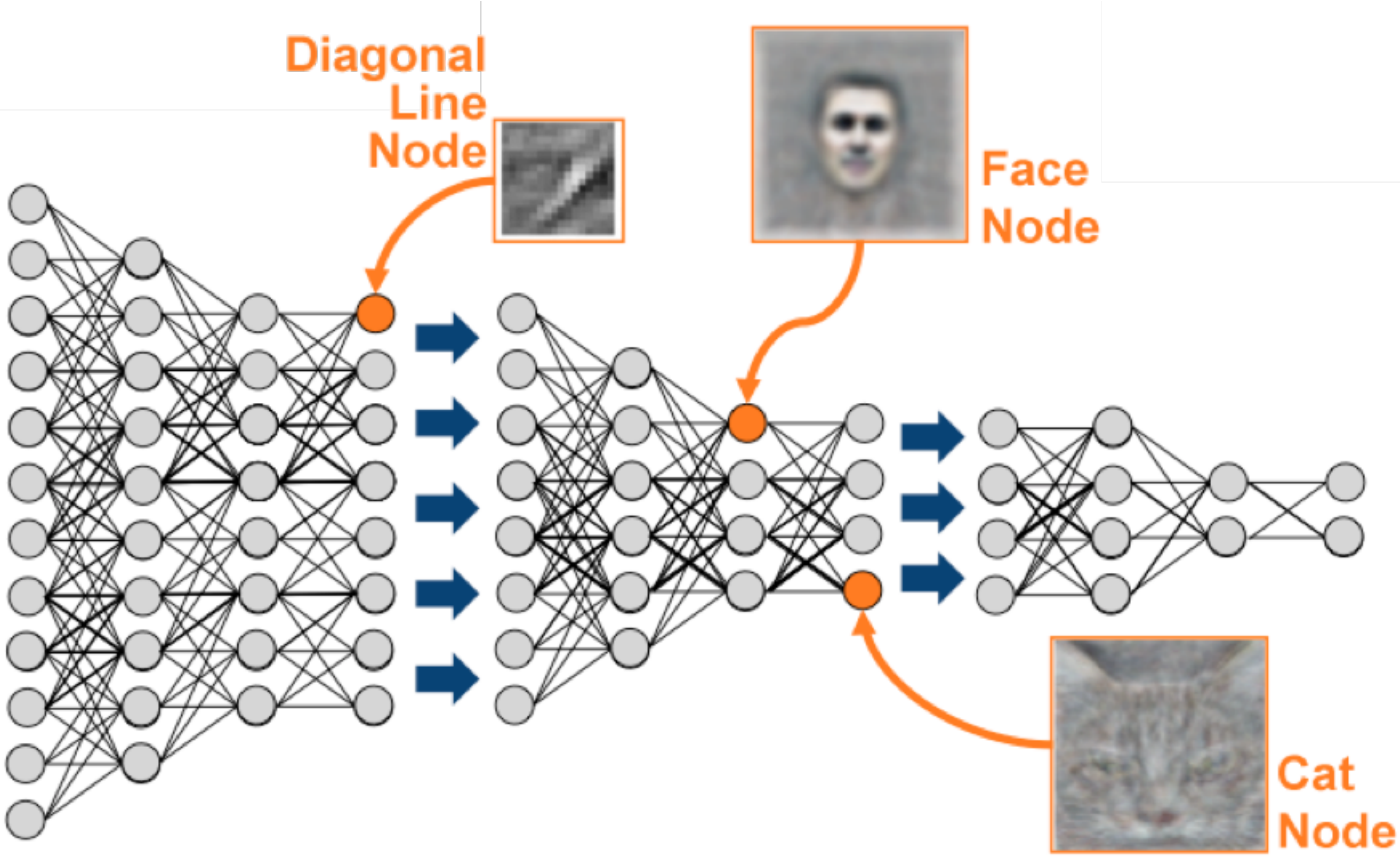


(a) RGB (b) Depth map



Why go *deep*?

- Multiple layers allow for *feature extraction.*
- "Vanishing gradient problem" → hard to train many layers.
- ***Now in "Deep Learning Renaissance"***

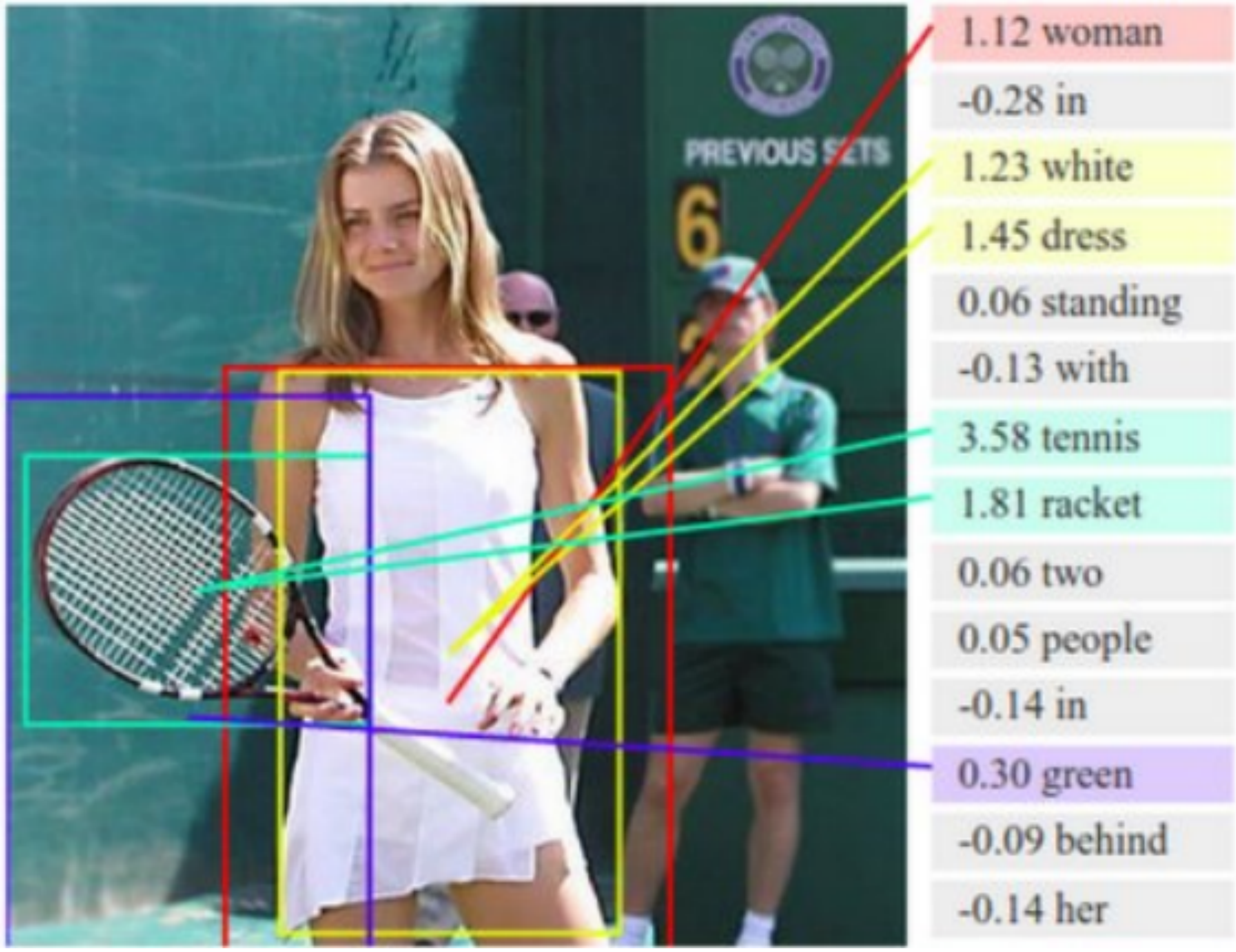


[Forbes/Google]

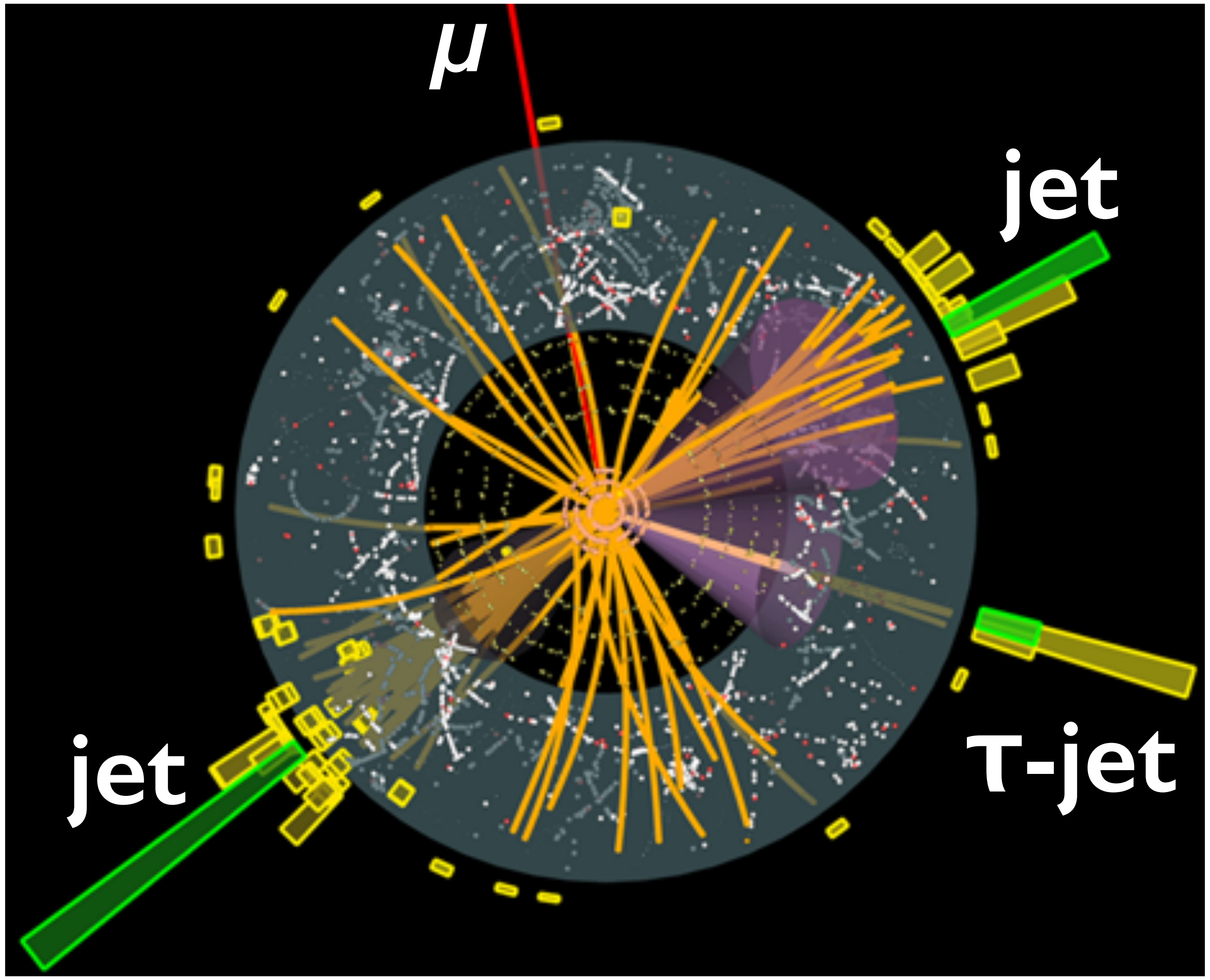
1. *Better training*: techniques and tools (e.g. smarter NN structures).
2. *Better hardware*: multicore, GPUs, bigger data centers, cloud computing, coming: neuromorphic computing.
3. *More training*: bigger datasets, search, the internet, open science.

Deep learning future?

Google ImageNet competition example

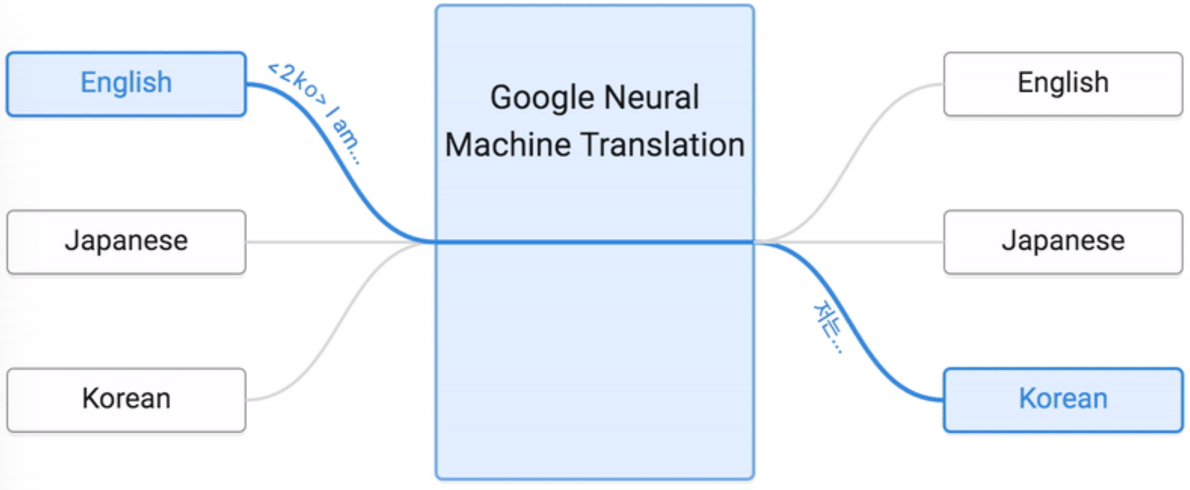


DNN future of ATLAS?



Natural Kinds?

- Seems like the possible uniqueness of the latent space representations (the features discovered by DNNs) says something interesting about the issue of ***natural kinds***, how to carve nature at its joints.
    - *Opposite sentiments shared by*:
      Bensusan, H. (Sussex) (2014). What can inductive machines suggest about the realism debate?
      Hennig, C. (UCL) (2015). What are the true clusters?
- What do results in machine translation say about arguments for the ***inscrutability of reference?***
    - "Zero-Shot Translation with Google's Multilingual Neural Machine Translation System" https://research.googleblog.com/2016/11/zero-shot-translation-with-googles.html

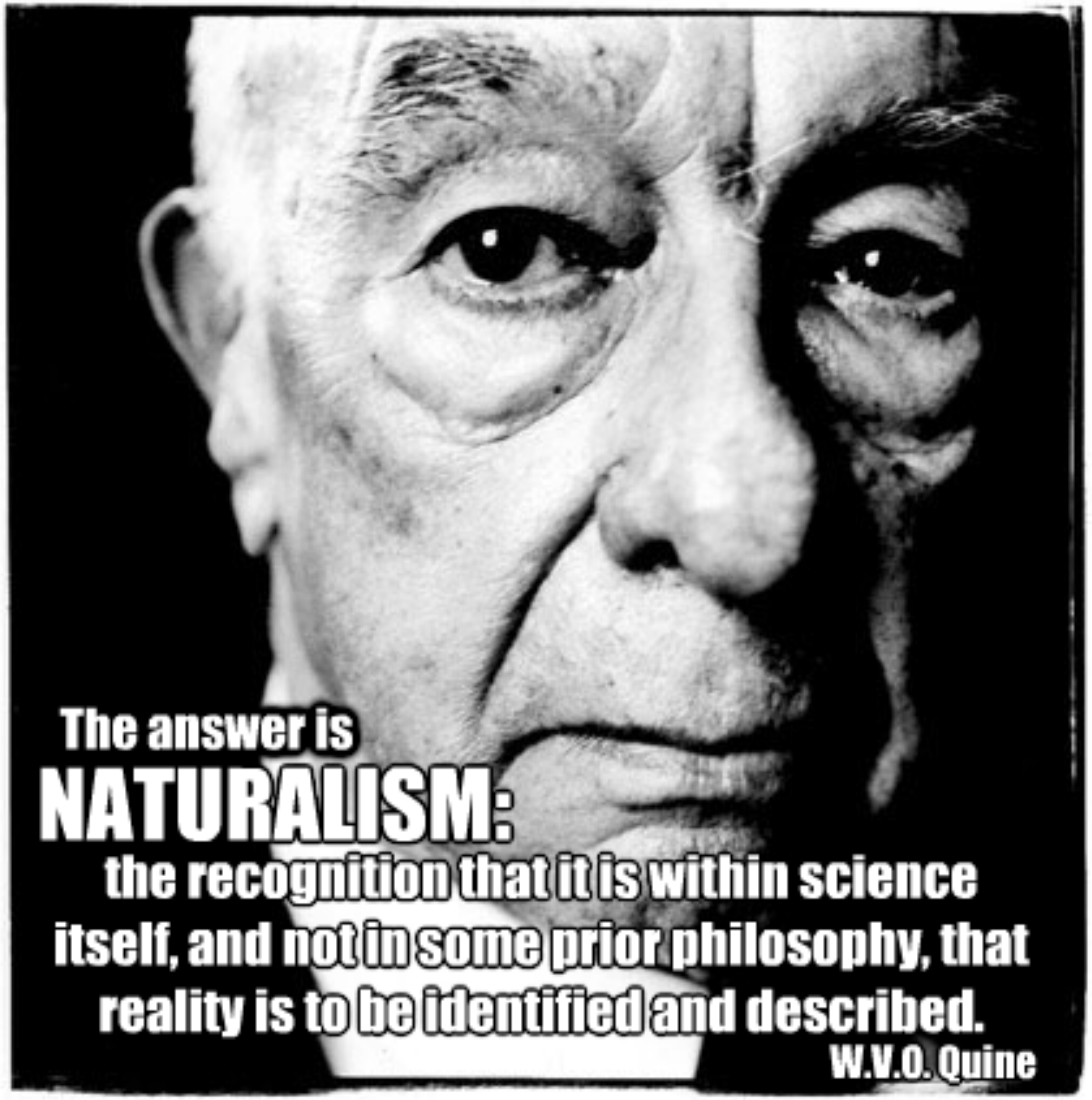


- I doubt one could rename-away the Higgs field, for example, being the only scalar field in the Standard Model.
    - My thoughts after reading: Button and Walsh. (2015). "Ideas and Results in Model Theory: Reference, Realism, Structure and Categoricity". arxiv:1501.00472.

Summary

- Particle physics probes very deep questions about what the world is made of, how it works, and how it got here.
- QFT is arguably the most impressive reductionist framework in science.
- Unlike previous eras of parts of physics seeming "near complete", QFT should be viewed as an *Effective Field Theory*.
- Experimental particle physics has consistently pushed the bounds of computing, and lead the big-data explosion until the 2000s.
- Physicists have learned to statistically justify their claims, and have often lead in developing statistical theory and methods.
- There are arguably ***Natural Kind*** characterizations of the degrees of freedom in nature, non-arbitrary choices in modeling the data.
- Realist cases can be made for the Standard Model, atoms, genes etc. despite what theory changes come in other regimes (structural realism, rainforest realism, Ladyman & Ross (2007) *Every Thing Must Go*).
- Discoveries in particle physics have the potential to explain the existence of dark matter and reveal details about the early universe.
- Machine learning is revolutionizing how induction can be automized. What does ML say about the realism debate?

Back up slides



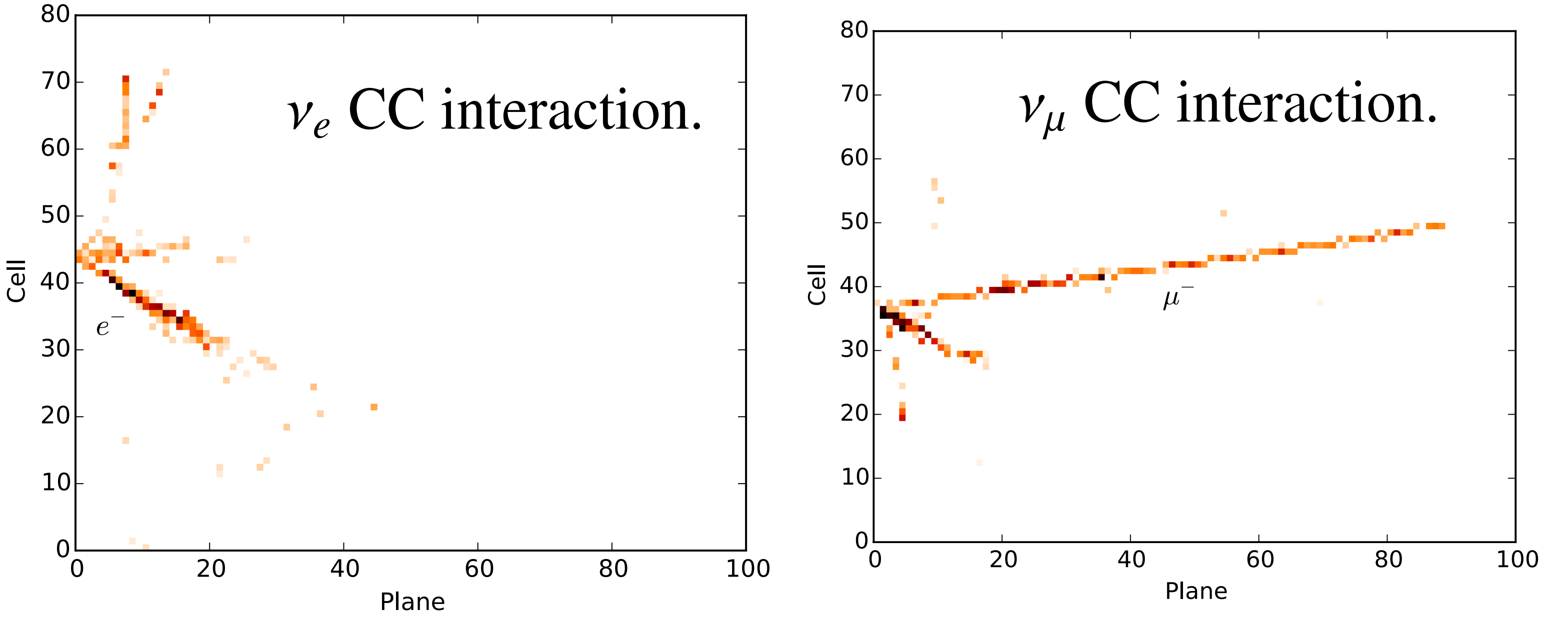
Deep Learning in HEP

- Deep learning does best with raw data and when there are unexploited features.
- raw channels→*tagging*
- basic kinematics→*features*

- Baldi *et al.* (2014). Searching for Exotic Particles in High-Energy Physics with Deep Learning. [1402.4735]
- 

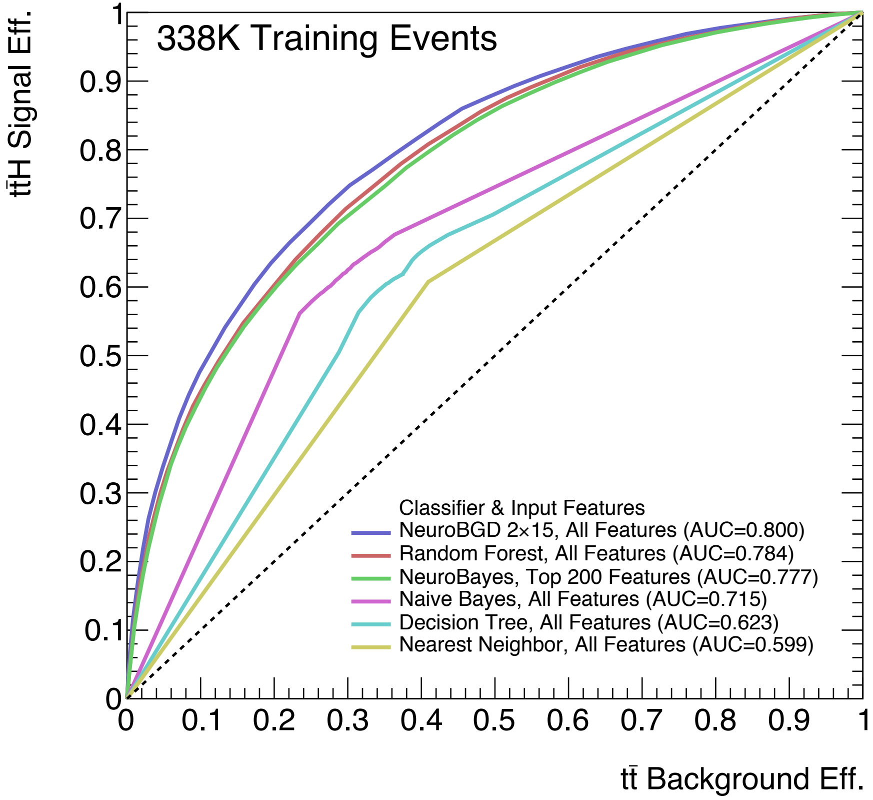
Baldi *et al.* (2015). Enhanced Higgs to $\tau^+\tau^-$ Search with Deep Learning. [1410.3469]

- Aurisano *et al.* (2016). A Convolutional Neural Network Neutrino Event Classifier. [1604.01444]



out performs NOvA's conventional reconstruction

- Santos *et al.* (2016). Machine learning techniques in searches for tth in the h→bb decay channel. [1610.03088]

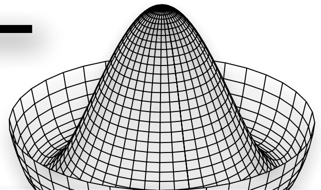


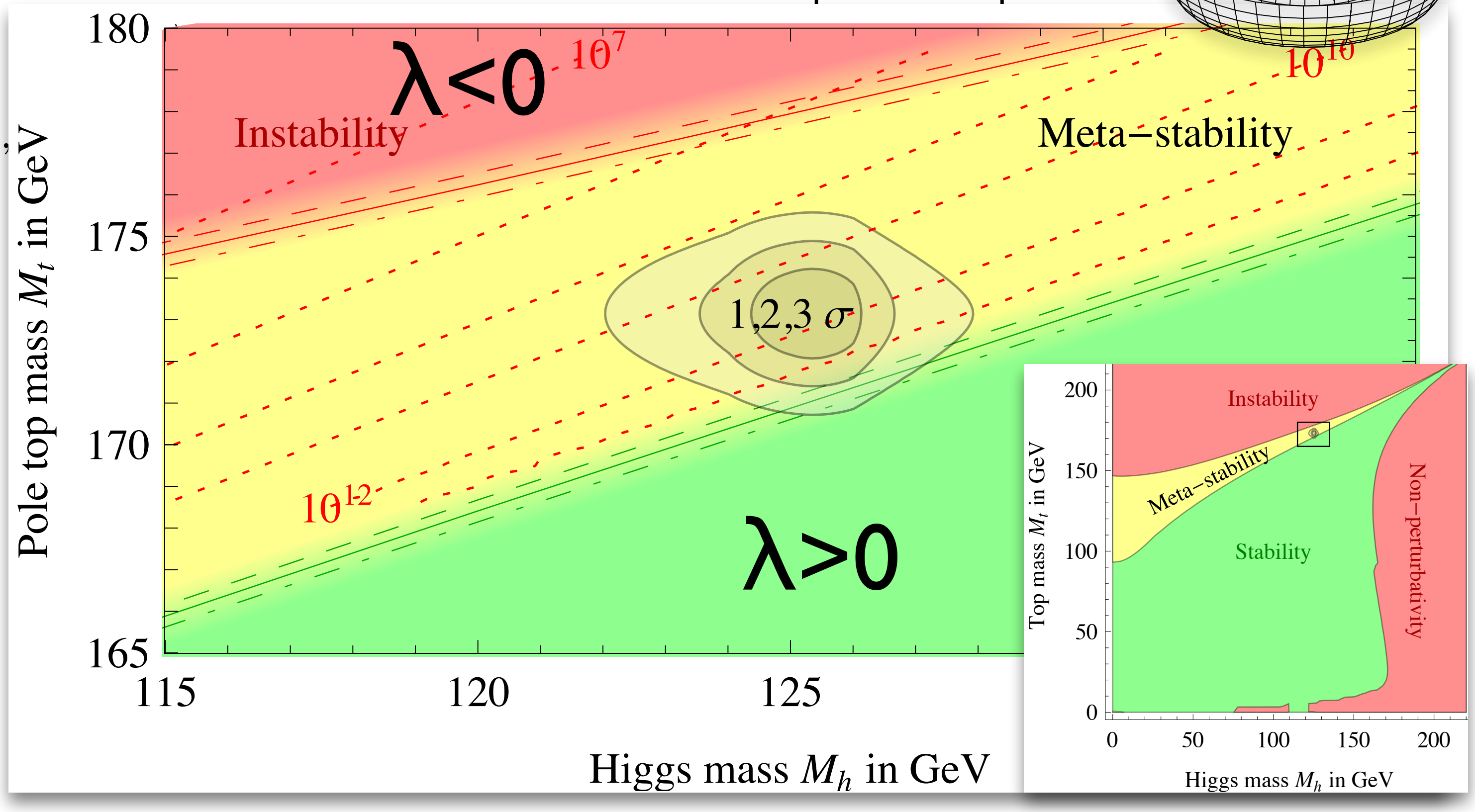
Naturalness or multiverse?

Higgs mass and vacuum stability in the Standard Model at NNLO

Giuseppe Degrassi[a], Stefano Di Vita[a], Joan Elias-Miró[b], José R. Espinosa[b,c], Gian F. Giudice[d], Gino Isidori[d,e], Alessandro Strumia[g,h]

$$V(\Phi) = \mu^2\, \Phi^\dagger\, \Phi + \lambda\, \left|\Phi^\dagger\, \Phi\right|^2$$



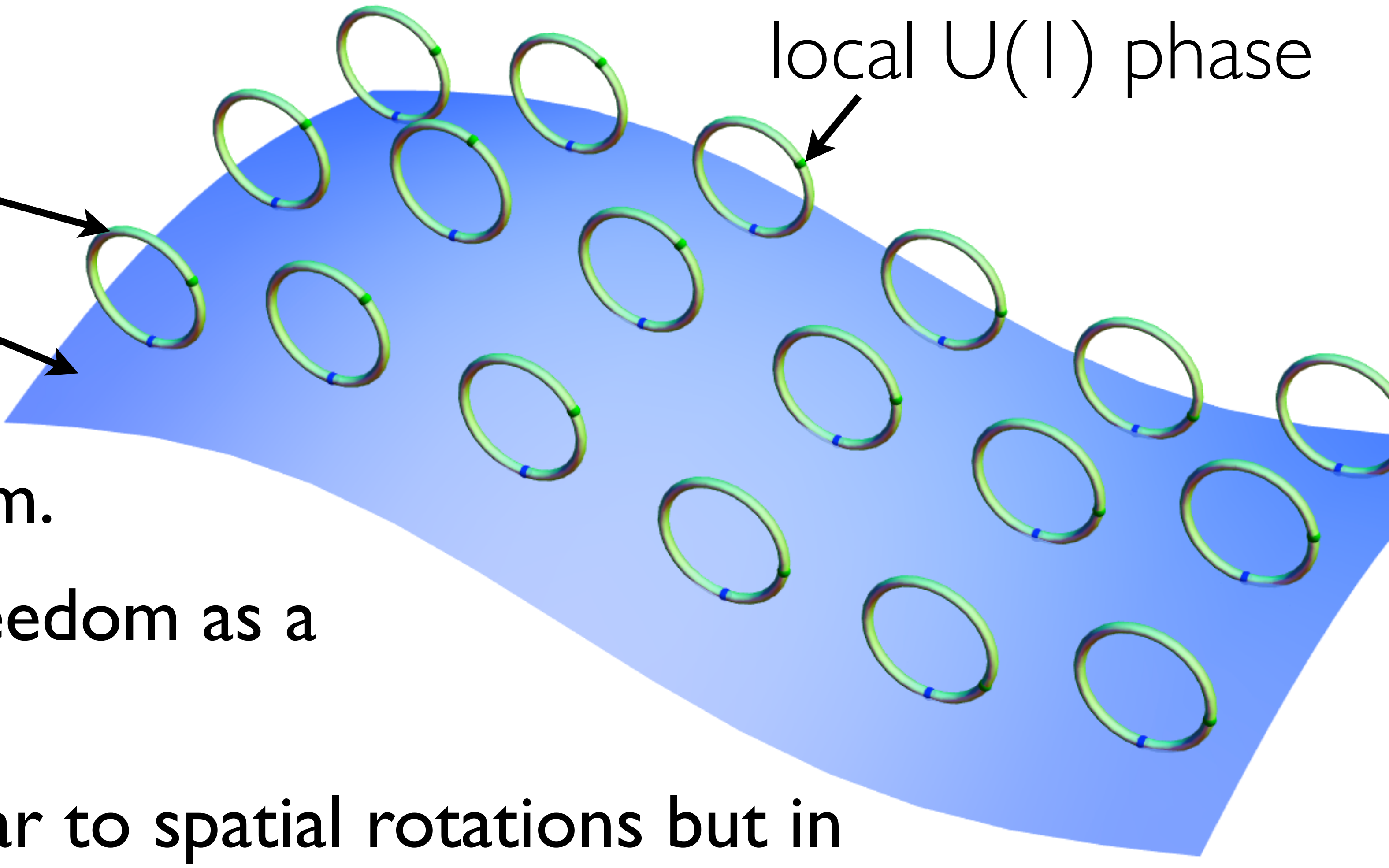


"If the LHC finds Higgs couplings deviating from the SM prediction and new degrees of freedom at the TeV scale, then the most important question will be to see if a consistent and natural (in the technical sense) explanation of EW breaking emerges from experimental data. But if the LHC discovers that the Higgs boson is not accompanied by any new physics, then it will be much harder for theorists to unveil the underlying organizing principles of nature. The multiverse, although being a stimulating physical concept, is discouragingly difficult to test from an empirical point of view. The measurement of the Higgs mass may provide a precious handle to gather some indirect information."

[arxiv: 1205.6497]

Gauge invariance is deep!

Why do gauge theories work?



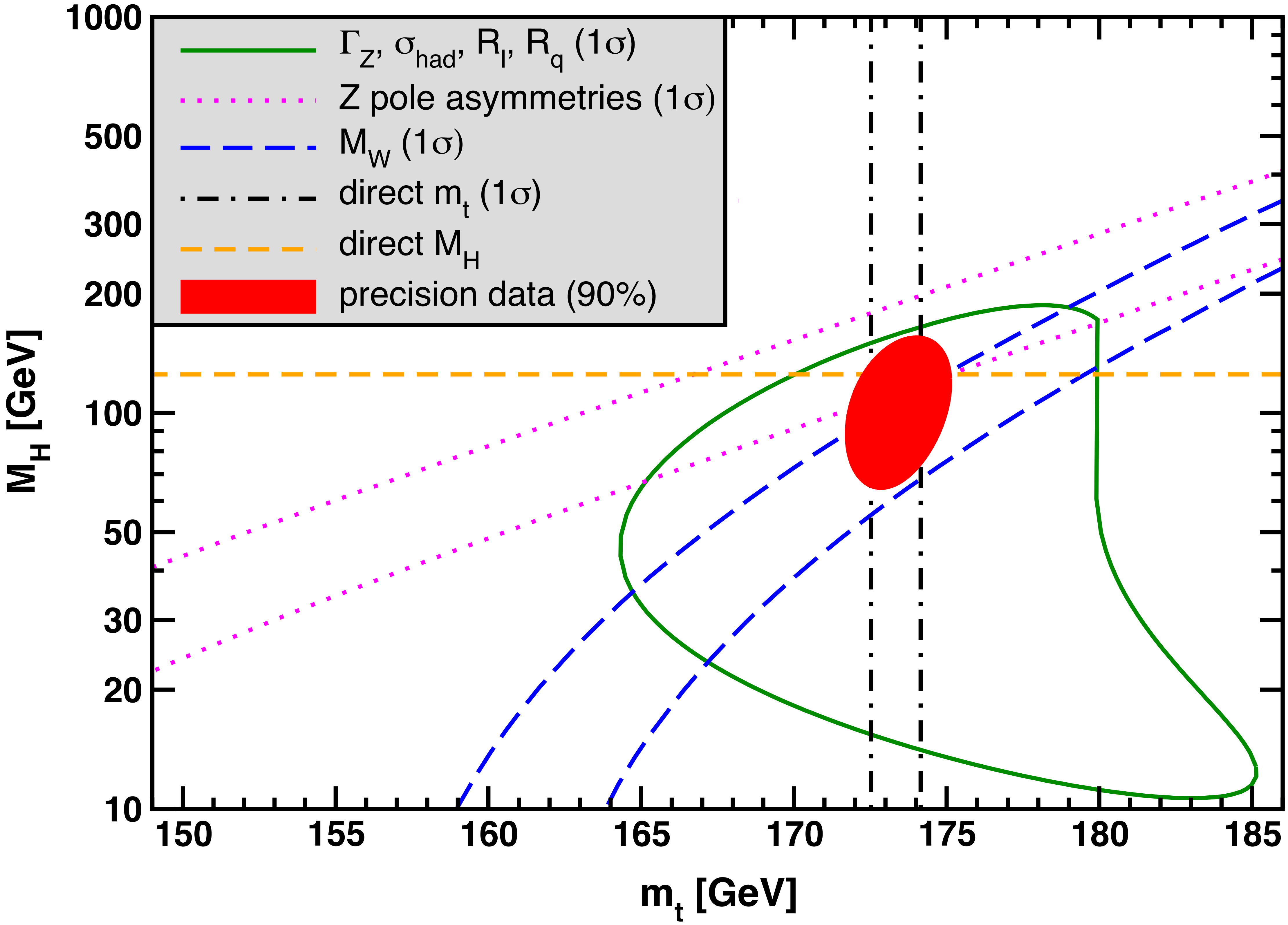
local U(1) phase

Internal gauge space

Spacetime

- Loyalty to the gauge principle motivated the Higgs mechanism.
- Some have described gauge freedom as a “redundancy of description”.
- But it is also a symmetry, similar to spatial rotations but in the *internal space of the field*.
- Can be rotated *locally*, independently at every spacetime point.
- What does it mean for the laws of nature to be describable by the continuous symmetries of Lie groups?
- What does it mean that the state of the universe can be represented as an element of a complex vector space, a Hilbert space?

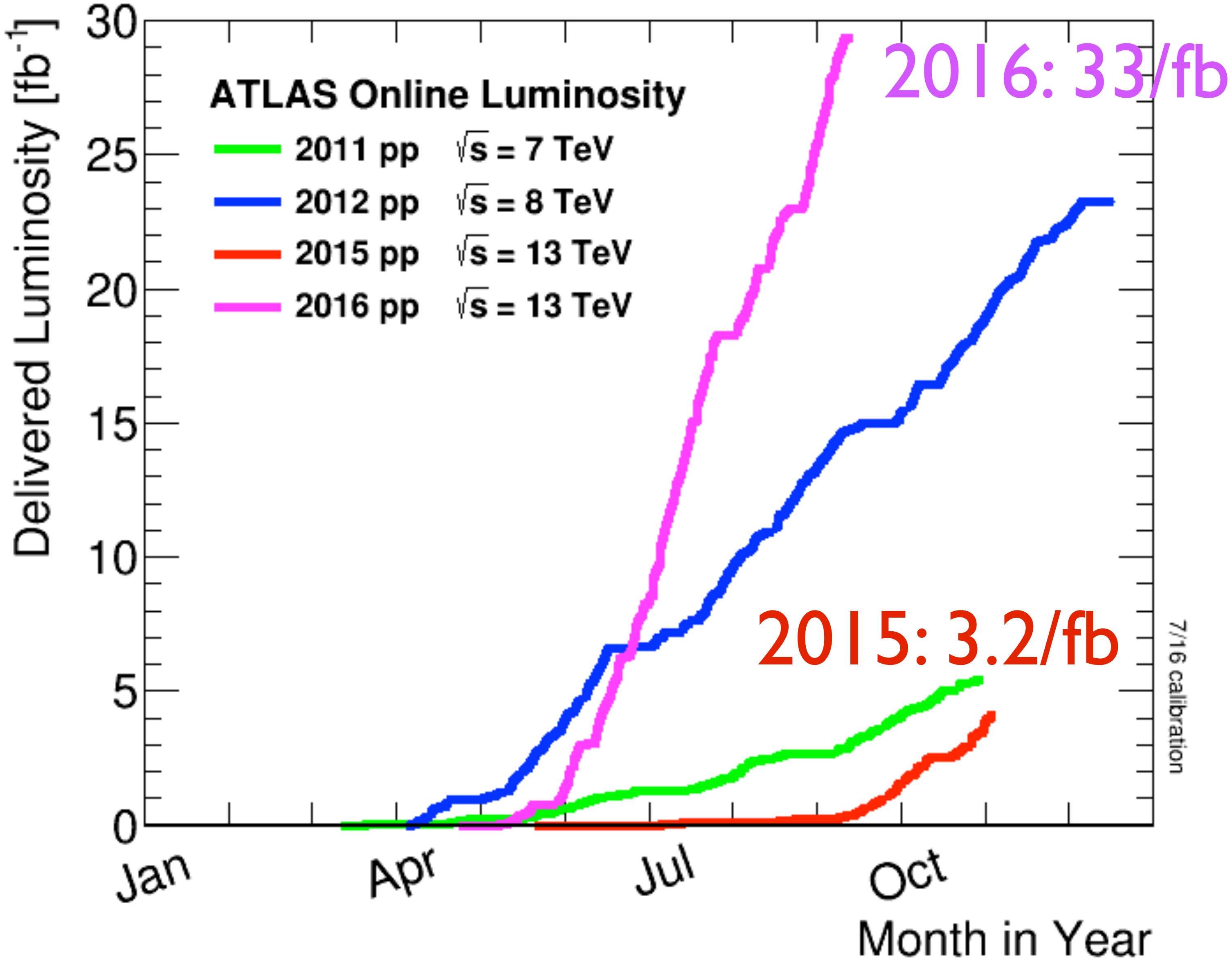
The SM is over constrained



[PDG 2016]

Datasets

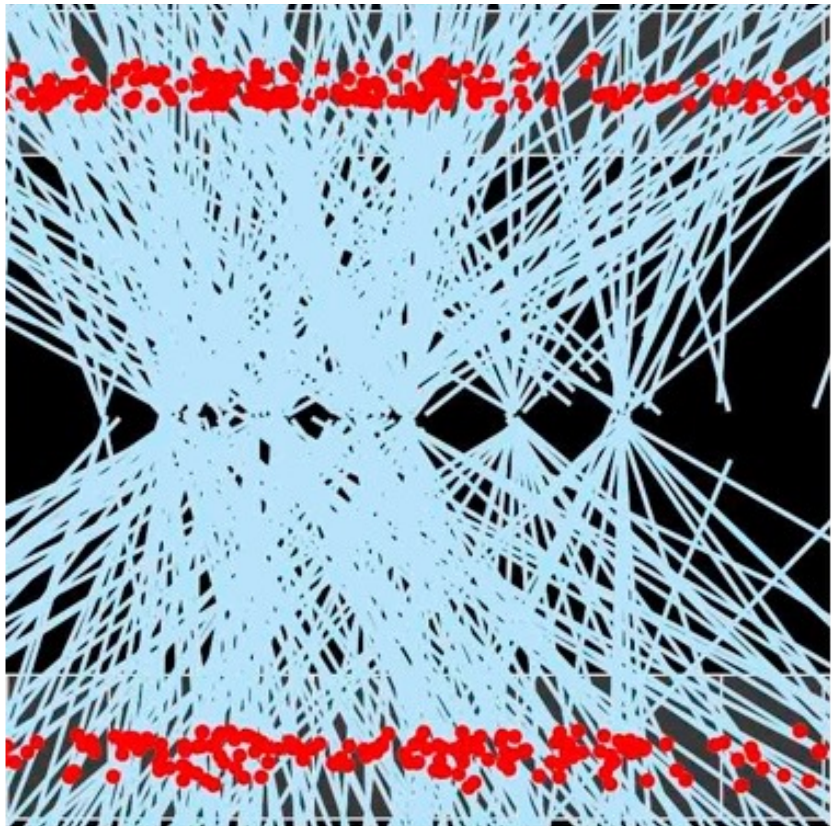
The LHC has performed extremely well!!



2016: 33/fb

2015: 3.2/fb

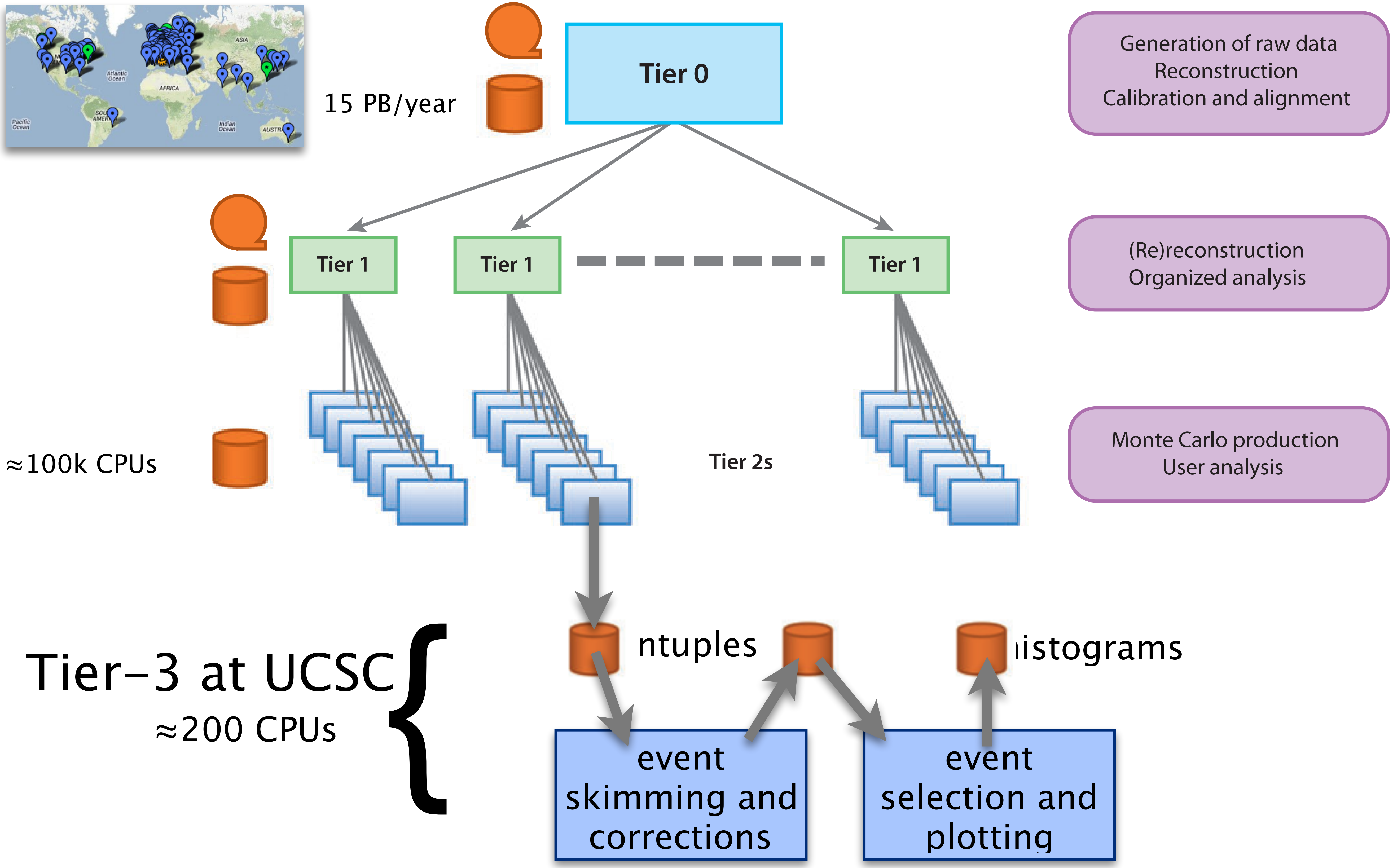
Recently broke inst. lumi. records > 10^{34} $\text{cm}^{-2}\text{s}^{-1}$



Typically 20-40 verticies per bunch crossing

Latest analyses combine collision data at $\sqrt{s}$=13TeV collected in the years 2015 and 2016, giving a total integrated lumi ≈ 13-15 fb^{-1}.

Computing infrastructure



Generation of raw data
Reconstruction
Calibration and alignment

(Re)reconstruction
Organized analysis

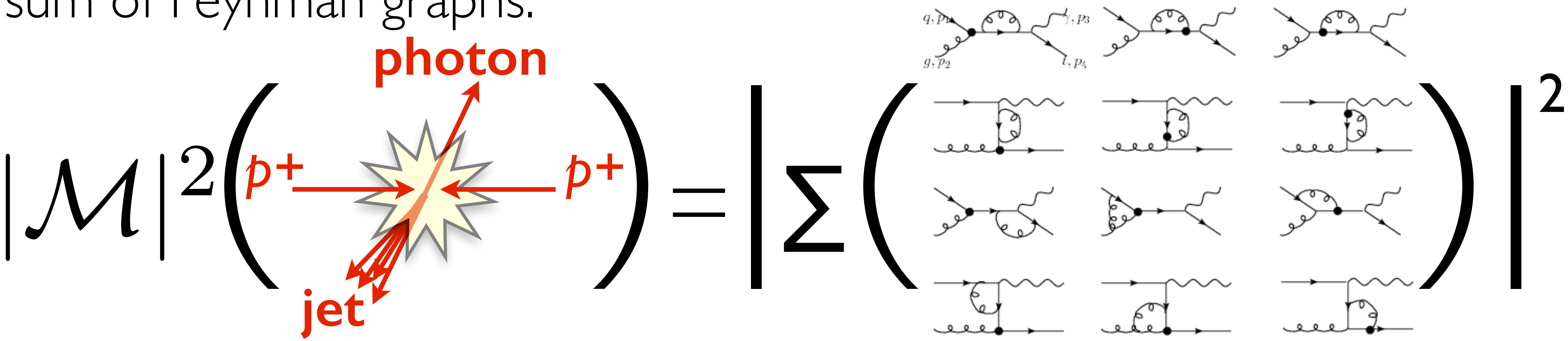
Monte Carlo production
User analysis

Tier-3 at UCSC

≈200 CPUs

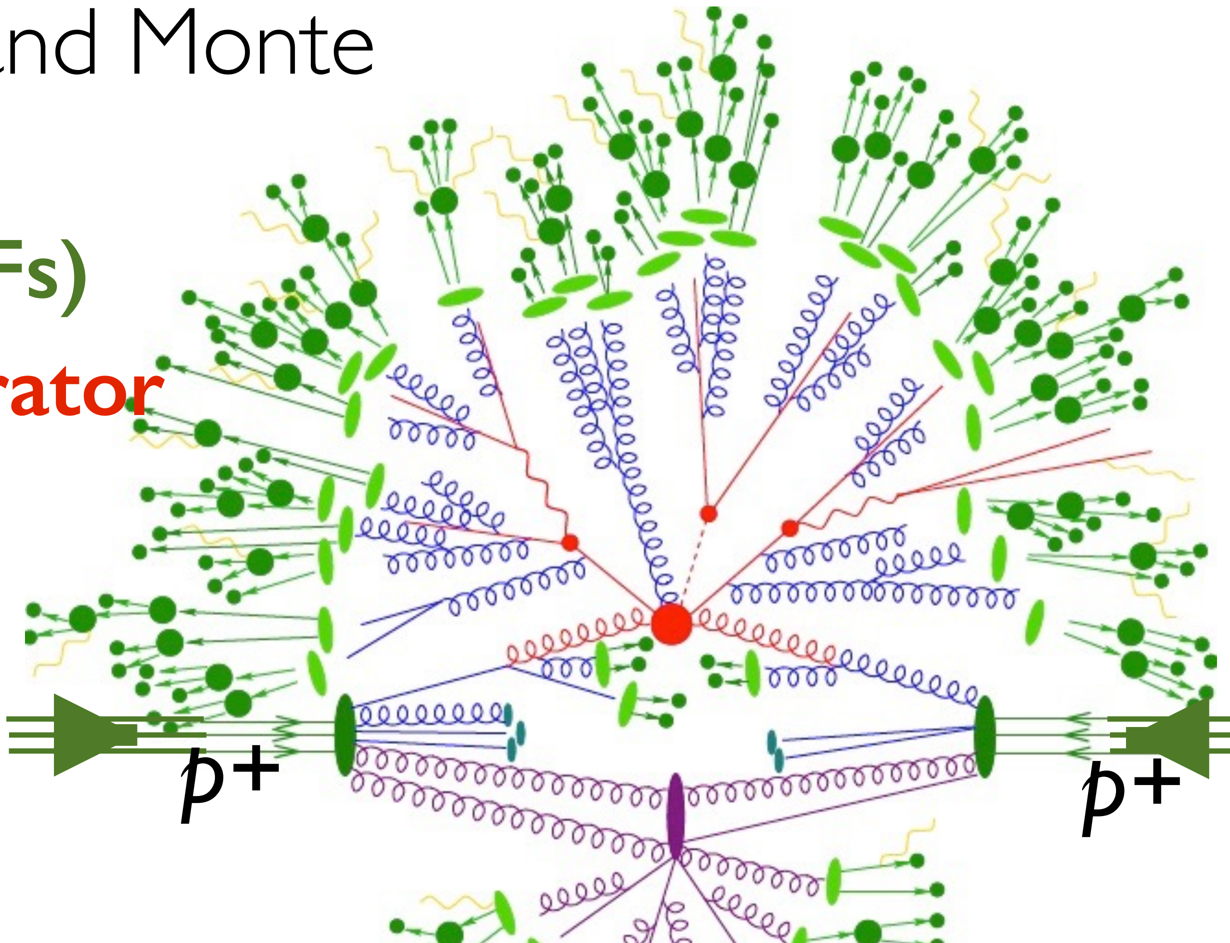
Matrix element

Hard-scatter matrix elements are calculated from a perturbative sum of Feynman graphs.

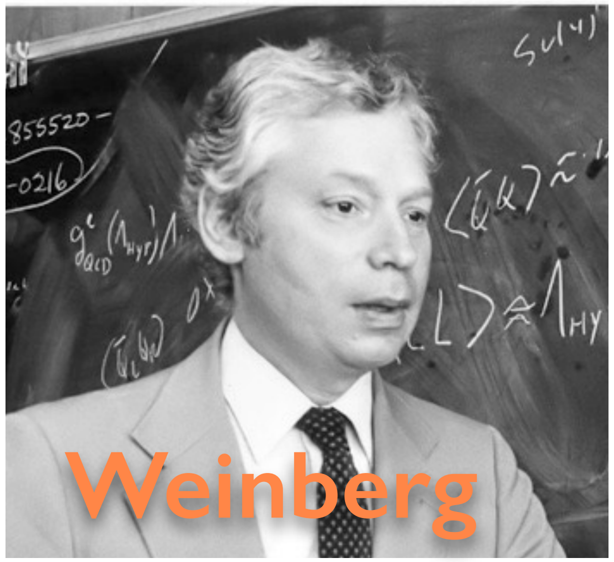


The strong force further complicates things by confining quarks in hadrons. Theorists and Monte Carlo simulations factor the problem:

- **“Parton Distribution Functions” (PDFs)**
- **“Hard-scatter” matrix element generator**
- **“Parton shower”, Bremsstrahlung, Initial/final-state radiation**
- **“Hadronization”**



Symmetry-first physics



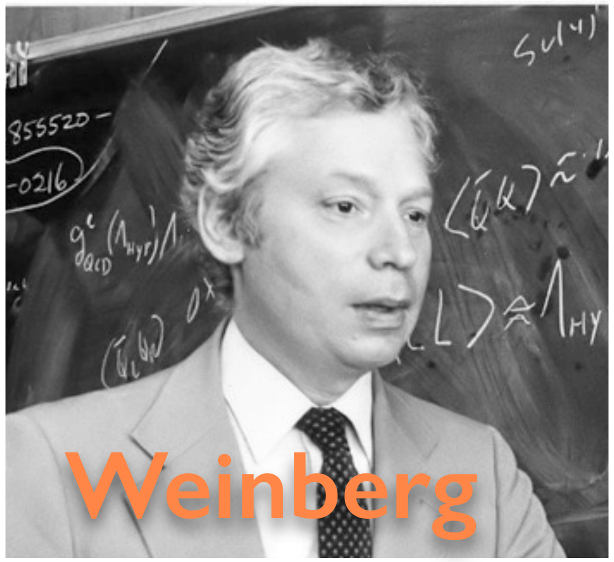
“Why do we enumerate possible theories by giving their Lagrangians rather than by writing down Hamiltonians? ... that **symmetries** imply the existence of Lie algebras of suitable quantum operators, and you need these Lie algebras to make sensible quantum theories. ... **if you start with a Lorentz invariant Lagrangian density then because of Noether’s theorem the Lorentz invariance of the S-matrix is automatic.”**

--

Weinberg, S. (1996). What is quantum field theory, and what did we think it is?

⇒ QFT is naturally relativistic if one requires that the Poincaré algebra be satisfied.

Effective Field Theories



"it is very likely that ***any quantum theory*** that at sufficiently low energy and large distances looks Lorentz invariant and satisfies the cluster decomposition principle will also at sufficiently low energy ***look like a quantum field theory.*** ...

This leads us to the idea of **effective field theories.** When you use quantum field theory to study low-energy phenomena, then according to the folk theorem you're not really making any assumption that could be wrong, unless of course Lorentz invariance or quantum mechanics or cluster decomposition is wrong, provided you don't say specifically what the Lagrangian is. As long as you let it be the most general possible Lagrangian consistent with the symmetries of the theory, you're simply writing down the most general theory you could possibly write down."

--

Weinberg, S. (1996). What is quantum field theory, and what did we think it is?

⇒ QFT is a way of parametrizing effective, local degrees of freedom.

Data science workflow

Data science done well looks easy *after your data is clean*.

■ **HEP/ATLAS equivalent**

1. Define the question of interest
   SM and BSM physics
2. Get the data
   dq2/rucio, Globus GridFTP
3. ***Clean and correct the data***
   GRLs, CP tools, RootCore, SUSYTools, QuickAna
4. Explore the data
   ROOT, event loops, histograms
5. Fit statistical models
   RooFit, RooStats, HistFitter, CLs/Bayesian methods
6. Communicate the results
   talks, notes, publications, axiv
7. Make your analysis reproducible
   CDS, SVN, HEPData, RECAST

Data cleaning can be a significant part of the analysis effort!

taken from: http://simplystatistics.org/2015/03/17/data-science-done-well-looks-easy-and-that-is-a-big-problem-for-data-scientists/