Enabling Discoveries at the LHC through Advanced Computation and Machine Learning

Mark Neubauer

University of Illinois at Urbana-Champaign

Blue Waters Symposium
Sunriver, OR
May 16-19, 2017

The Pursuit of Particle Physics

To understand the the Universe at its most fundamental level

Primary questions: What are the

- elementary constituents of matter?
- forces that dictate their behavior?
- the nature of space and time?

The Standard Model*

(a.k.a. our best theory of Nature)

Ordinary Matter

Quarks

u (up)	c (charm)	t (top)
d (down)	s (strange)	b (bottom)

Leptons

e (electron)	μ (muon)	τ (tau)
ν_e (electron neutrino)	ν_μ (muon neutrino)	ν_τ (tau neutrino)

Mediate Matter Interactions

Forces

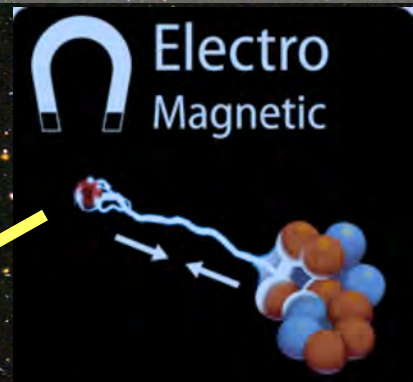
Z (Z boson)	γ (photon)
W (W boson)	g (gluon)

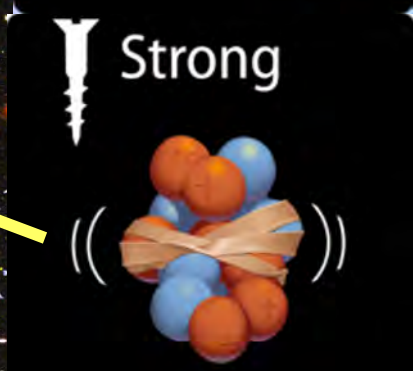
H Higgs boson

Heavy!

m=0

Before July 4, 2012,
never directly observed!



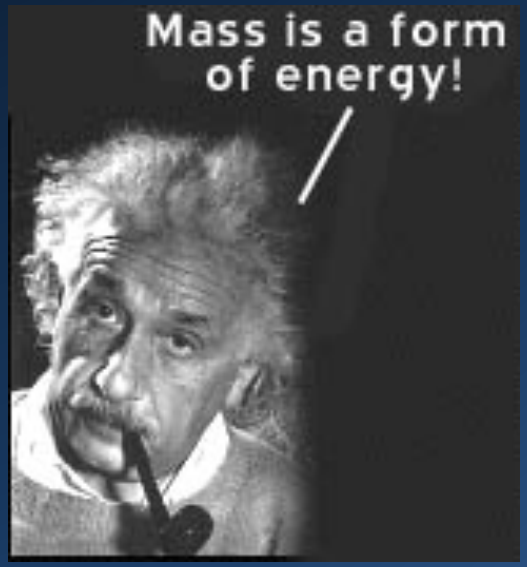


*Some assembly required. Gravity not included

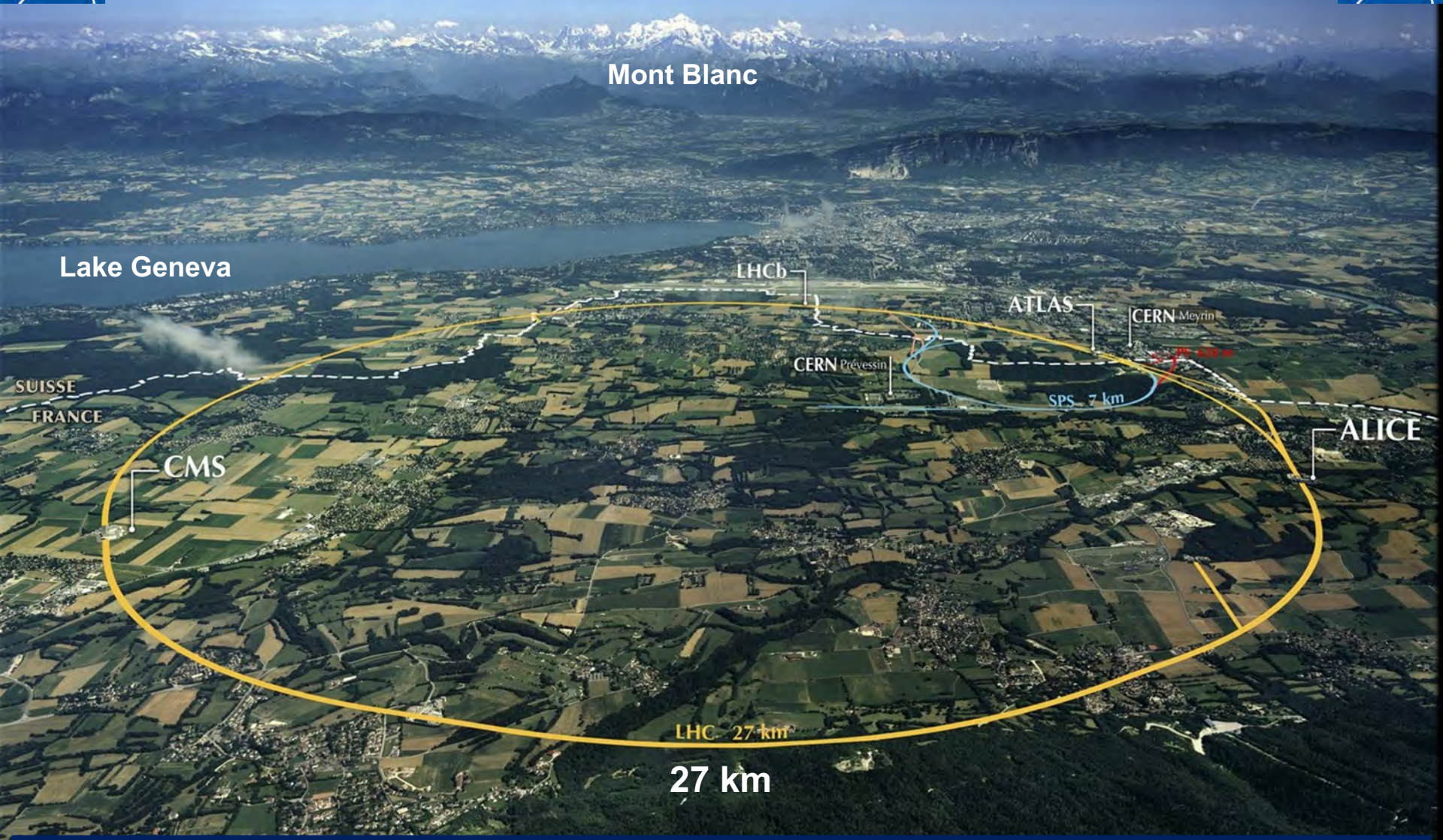
How do we look for new particles?

At Hadron Collide

Challenge: Massive particles are unstable and must be detected by signatures in stable decay products they leave behind

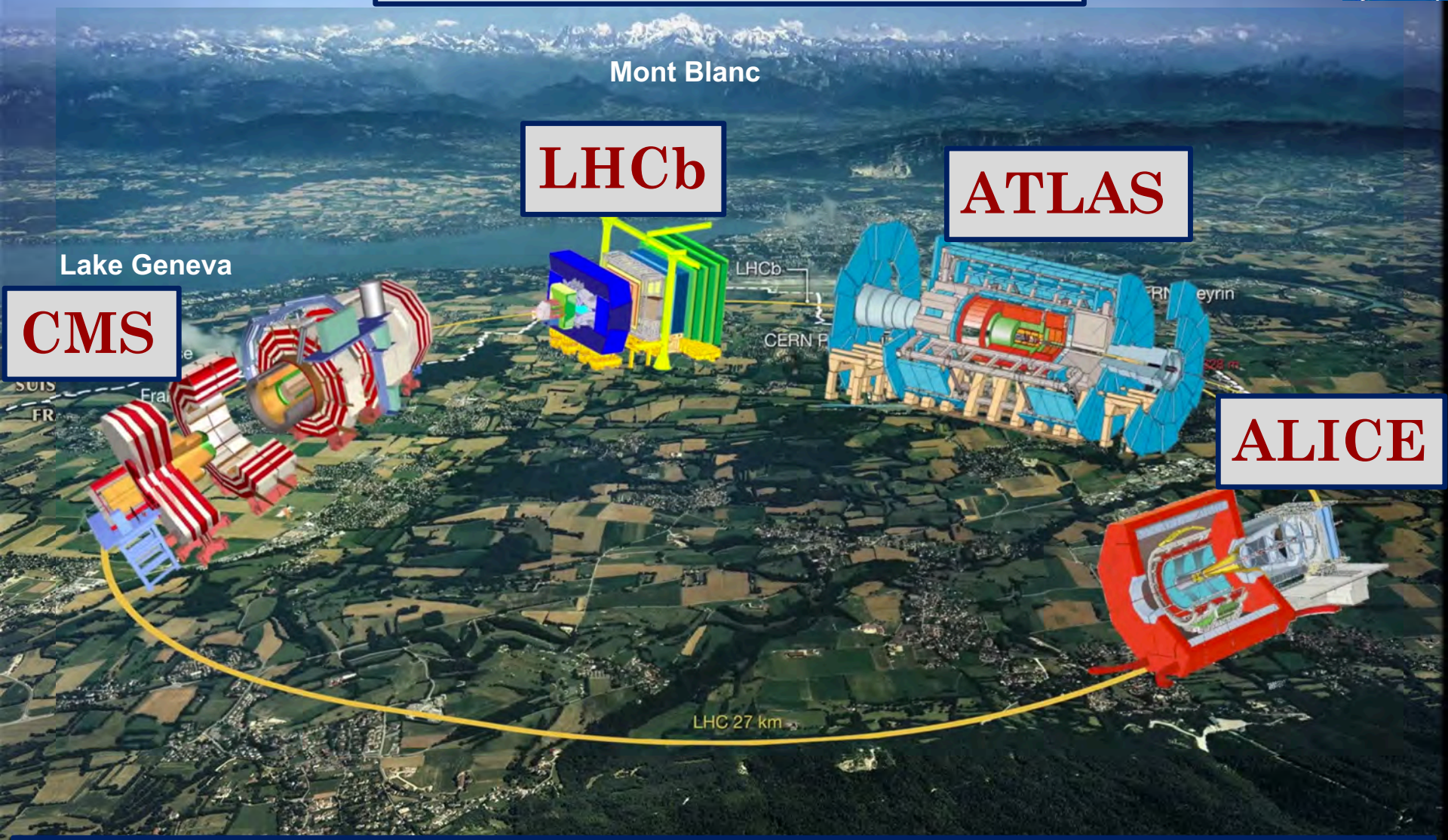


The Large Hadron Collider (LHC)



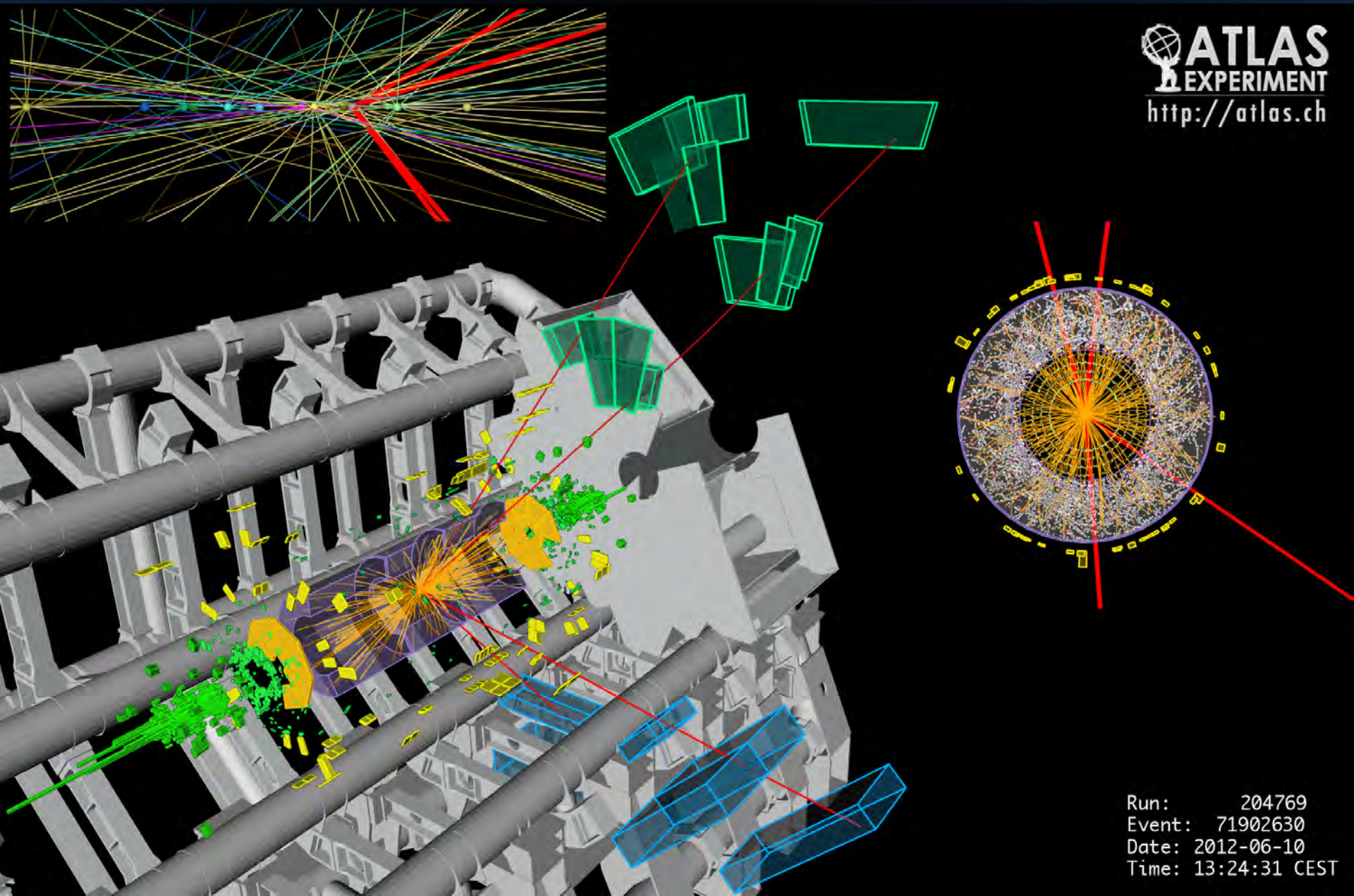
pp, pPb and PbPb collisions at highest energies

LHC Experiments

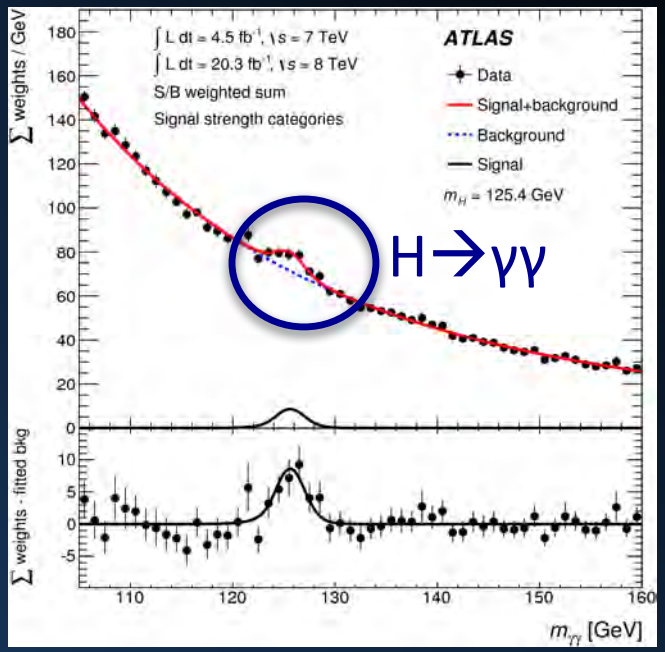


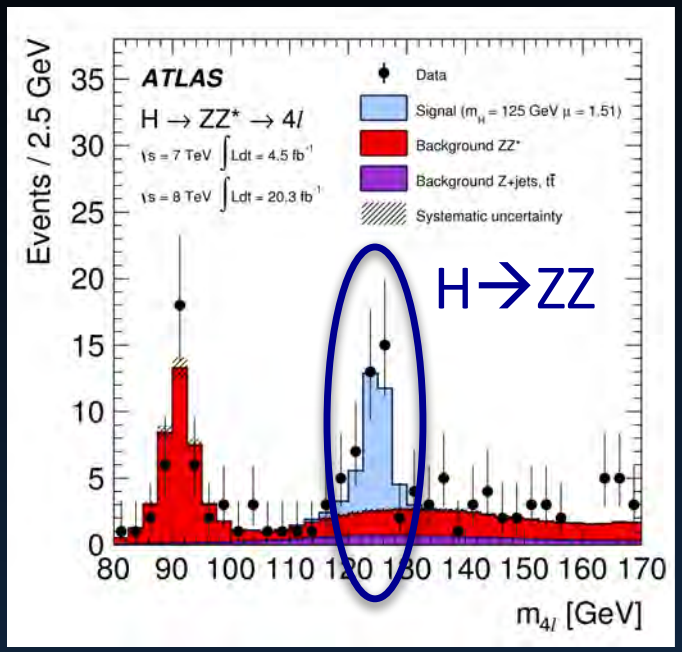
pp, pPb and PbPb collisions at highest energies

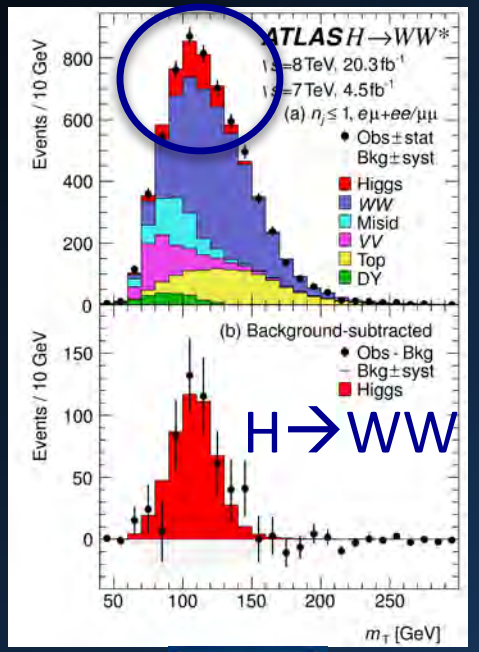
Higgs Boson Candidate: H→ZZ*→μμμμ



Higgs Boson Discovery! (2012)







2013 Nobel prize in Physics to Peter Higgs and Francois Englert

A new era in particle physics. The discovery of a Higgs boson with mass 125 GeV opens up a new window to search for beyond-the-SM physics

Higgs: Needle in a Haystack of Needles

- The Higgs(-like) boson discovery based on analysis of 1 quadrillion (10^{15}) proton collisions!
  - 2 million Higgs bosons produced ($7000/Higgs)
- The vast majority look virtually the same as less interesting processes
  - Only a few really stand out (e.g. H→ZZ→μμμμ)

Big Data. Bon Appétit!

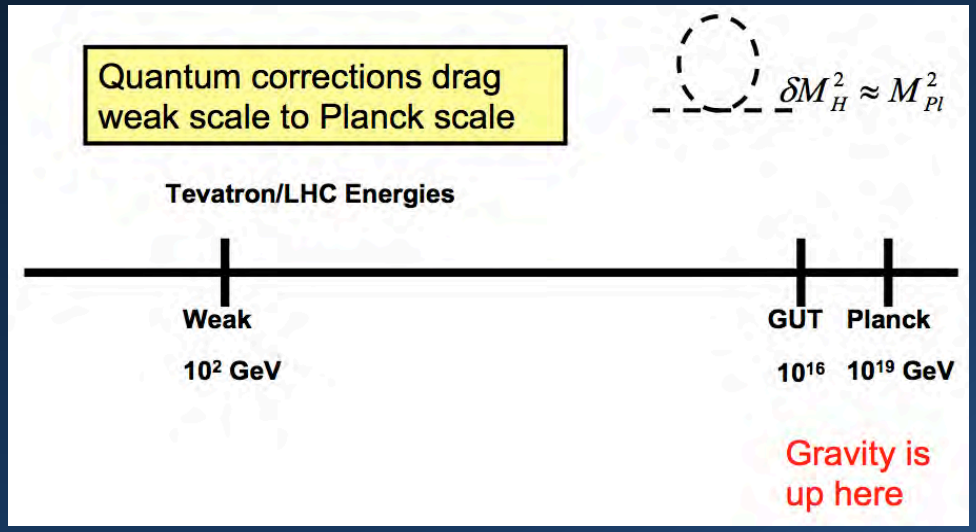
- LHC annually generates 15 Petabytes of data (300,000 Blue-Ray discs)
- Data analyzed by 1000s of physicists located all over the world
- Higgs boson discovery:

A success story of Big Science with Big Data!

Open Questions after Higgs Discovery

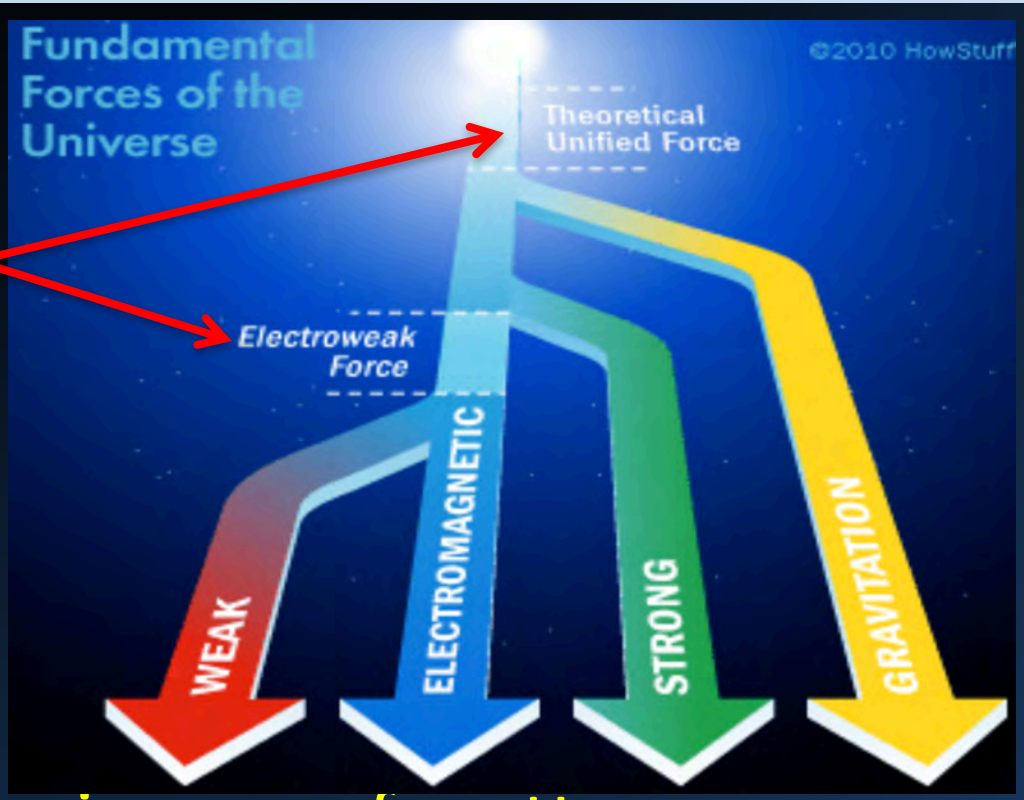
- Is the newly discovered particle THE Higgs Boson in the SM or something else?
  - Does the Higgs boson have properties predicted by SM?
  - Is the Higgs boson fundamental or composite?
  - Are there more Higgs bosons?
  - How does the Higgs get its mass?

- What protects against large quantum corrections that connect weak and Planck scales?



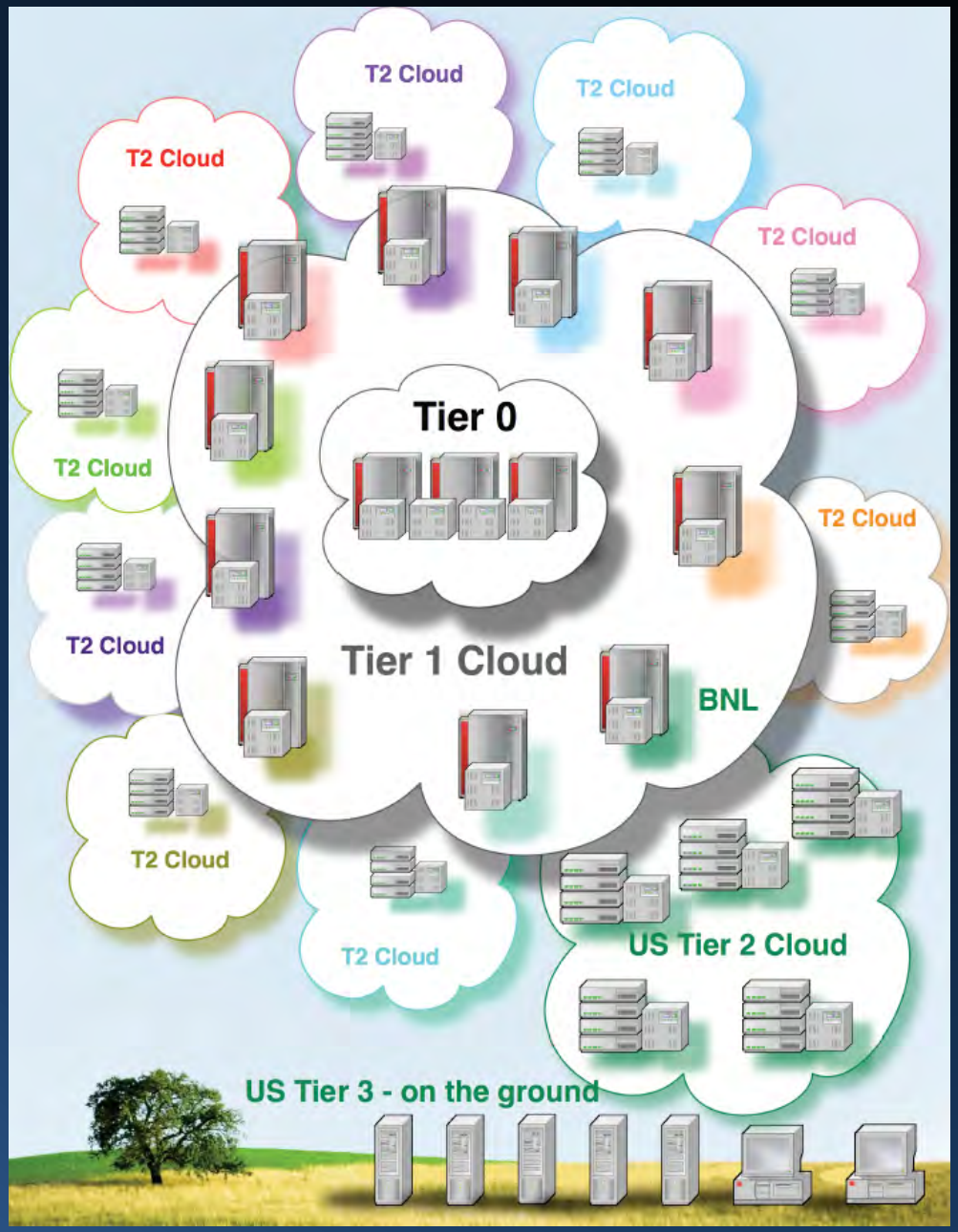
Open Questions after Higgs Discovery

- Do forces unify at some energy? What GUT describes that?
- What drove cosmic inflation? Role for Higgs boson in this?
- Quantum Gravity?



- Why is there a vast dominance of matter over antimatter in the Universe? Role for Higgs Boson?
- Are neutrinos their own anti-particle? Sterile v's? How do neutrinos get their mass? Why so small?
- What is the nature of Dark Matter? Dark Energy?

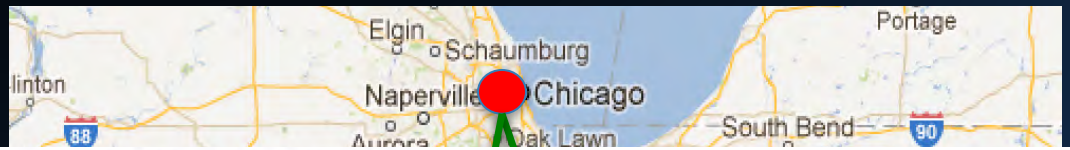
LHC Computing Challenges

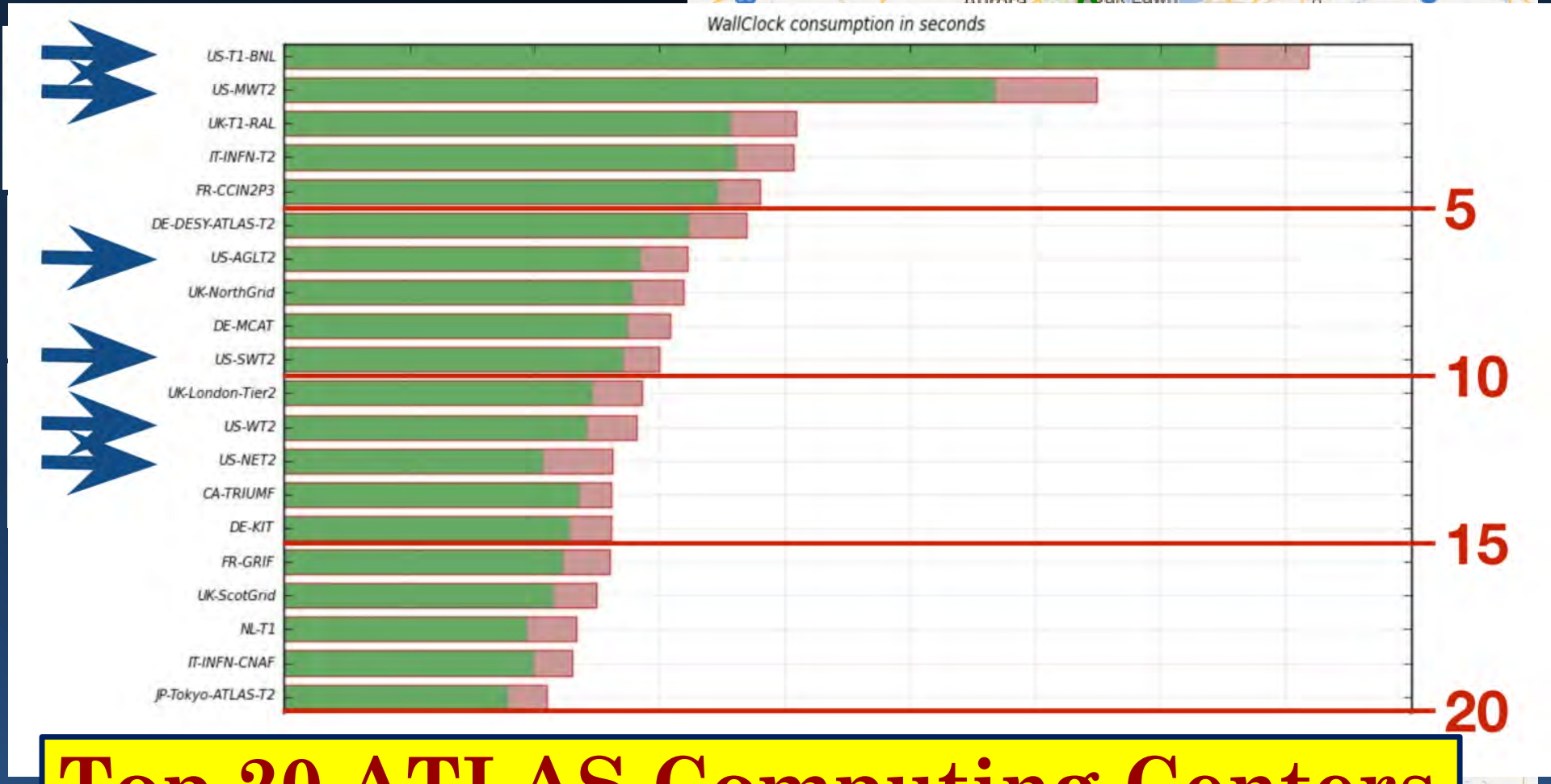


- PBs of data must be processed to generate high-level objects
- 100s of Billions of simulated events must be generated to interpret the data
- Tier-2 Centers are where the vast majority of simulation, data processing and data analysis occurs

The Midwest Tier-2 (MWT2) Center

- MWT2: A three-site Tier-2 Consortia





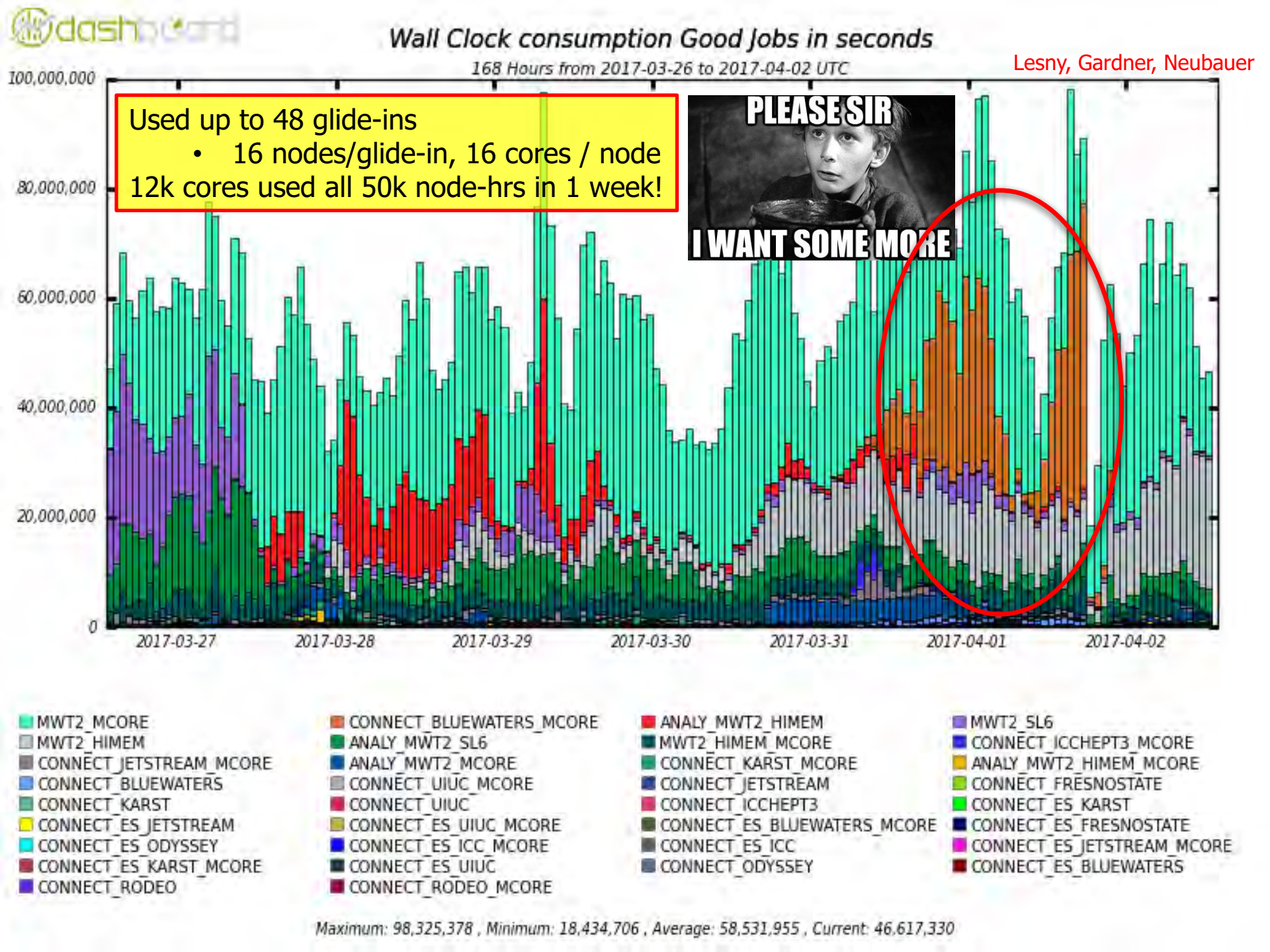
Top 20 ATLAS Computing Centers

LHC Science on Blue Waters

1) Large-scale simulation, data processing and data analysis of proton-proton collisions at the LHC
   - ***Approach***: Leverage MWT2/OSG Services and the Shifter/Docker container technologies available on BW
   - BW platform is ideal for this application due to its large CPU, storage and network resources
2) Machine Learning applications to improve the LHC discovery potential
   - ***Approach***: Utilize industry-standard ML tools (e.g. Keras, TensorFlow, scikit-learn, ..) available on BW
   - BW platform is ideal for this due to its large GPU resources for training of large and complex networks

NOTE: 1) greatly enhances the capabilities of 2) !

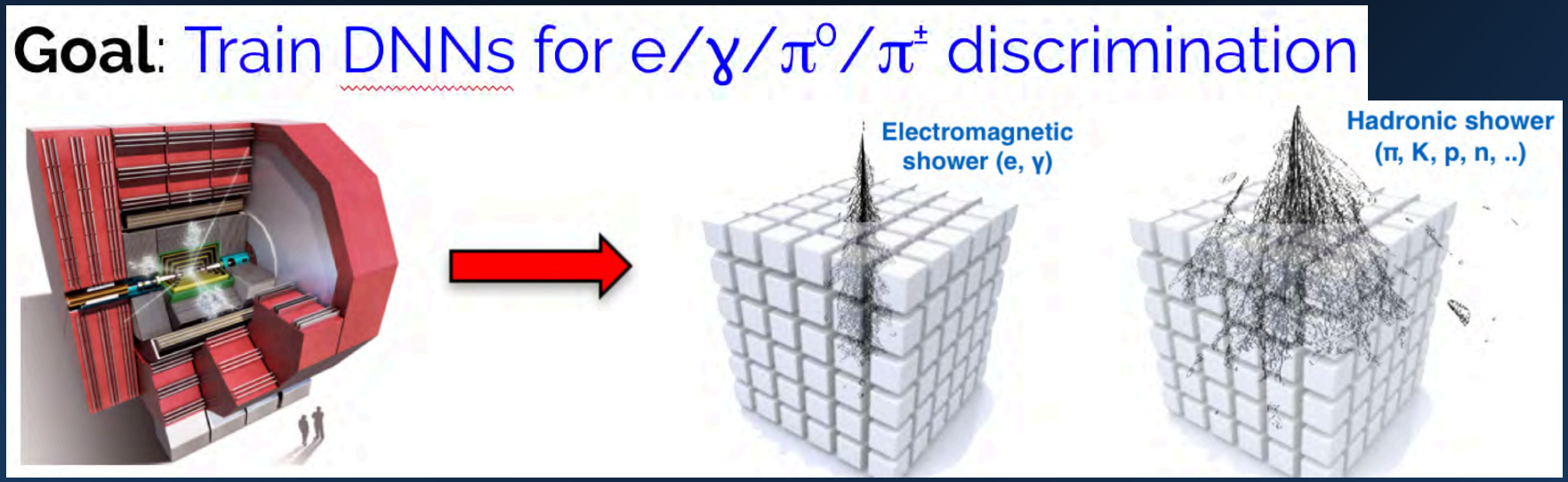
Lesny, Gardner, Neubauer



Machine Learning on BW

Several Machine Leaning applications have been pursued with our initial Blue Waters allocation

1) Stable Particle Identification (Farbin, Hooberman, *et. al.*)



2) Decaying particle ID (e.g. boosted Higgs taggers) (Neubauer, Zhong, *et. al.*)

3) Sustainable Matrix Element Calculations (Neubauer, Zhong, *et. al.*)

The Matrix Element Method

Probability density („weight“) for event **x** given hypothesis α?

ACAT2016 - Sébastien Wertz

Possible uses:

Sample likelihood
→ M.L. parameter fit
$\prod_{i \in events} P(\mathbf{x}_i|\alpha)$

Neyman-Pearson discriminant [4]
→ Hypothesis testing/search for rare process
$P(\mathbf{x}|S)/\sum_i r_i P(\mathbf{x}|B_i)$

… Can be computed!

$$P(\mathbf{x}|\alpha) = \frac{1}{A_\alpha \sigma_\alpha} \int \mathrm{d}\Phi(y) \frac{\mathrm{d}x_1 \mathrm{d}x_2}{x_1 x_2 s} f(x_1) f(x_2) |\mathcal{M}_\alpha(y, x_1, x_2)|^2 W(\mathbf{x}|y) \epsilon_\alpha(y)$$

Theoretical hypothesis (Matrix Element) + Parton shower + Detector (transfer functions, efficiencies) + Experimental information (whole event **x**)

- Multi-dim (~8D) integration of strongly-peaked fcns → very hard!
- Investigating application of DNN to encode integrals → fast, re-usable
  → Building collaboration between ATLAS, CMS, theorists and CS

Summary and Outlook

- The LHC program is in early stages of a multi-decade exploration into physics at the high-energy frontier
- Blue Waters provides a unique resource to extend the discovery reach of the LHC experiments
  - We have successfully deployed our "application" on BW – simulation, data processing & analysis, Machine Learning
  - This brings a large-scale ***Distributed High-Throughput Computing*** (DHTC) capability to the BW ecosystem
    - DHTC on HPCs is critical to realize the promise of deep learning in HEP, in addition to facilitating data-enabled discovery
- On the horizon – a possible NSF Scientific Software Innovation Institute for HEP (in conceptualization phase)