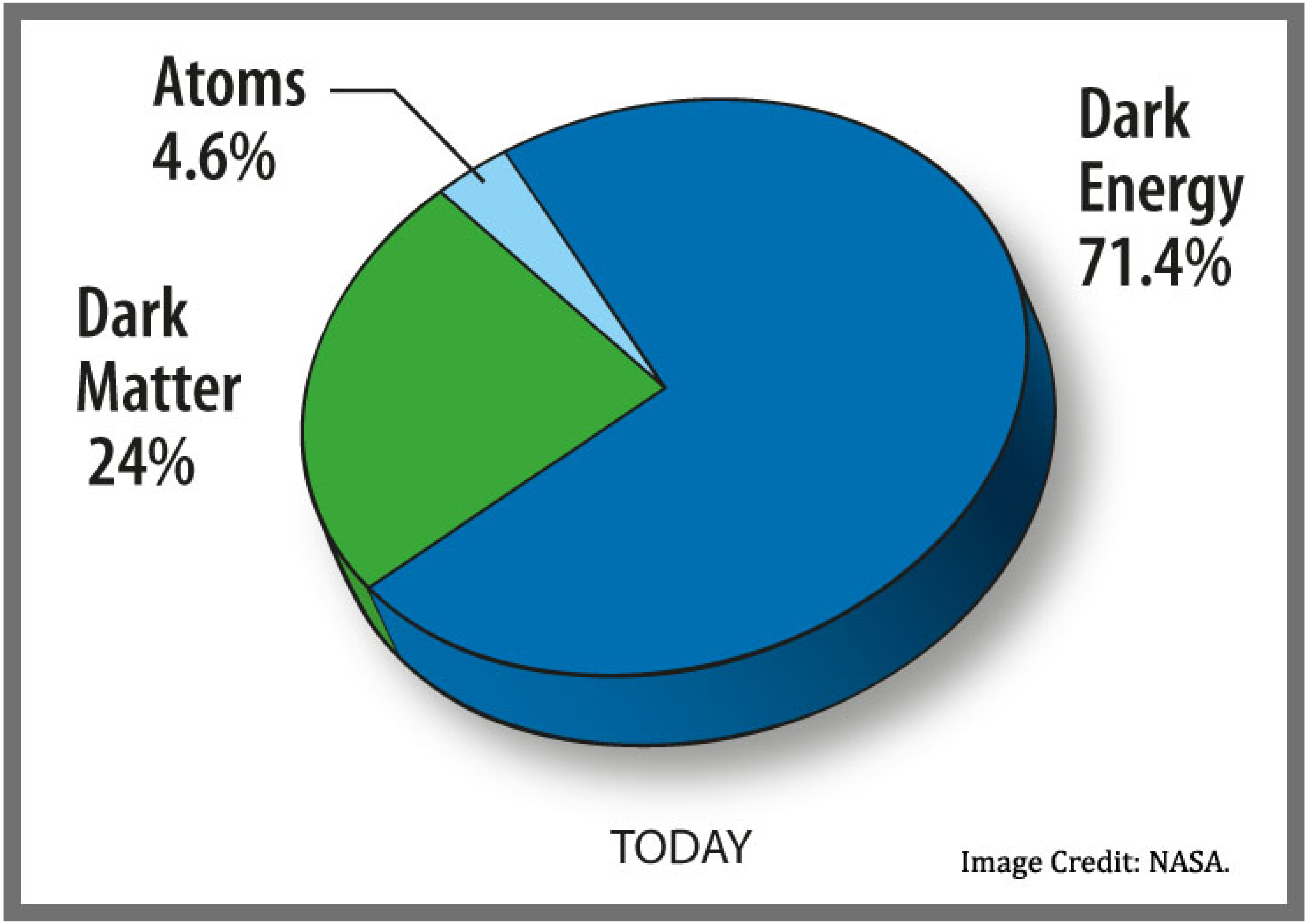
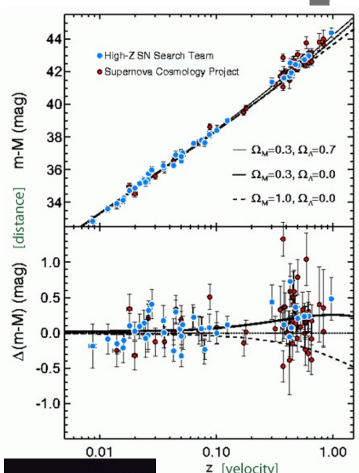
Prelude

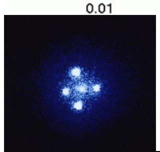
The Matter and Energy Content of The Universe

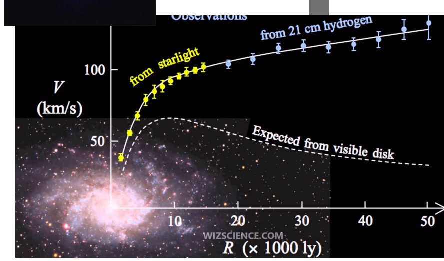


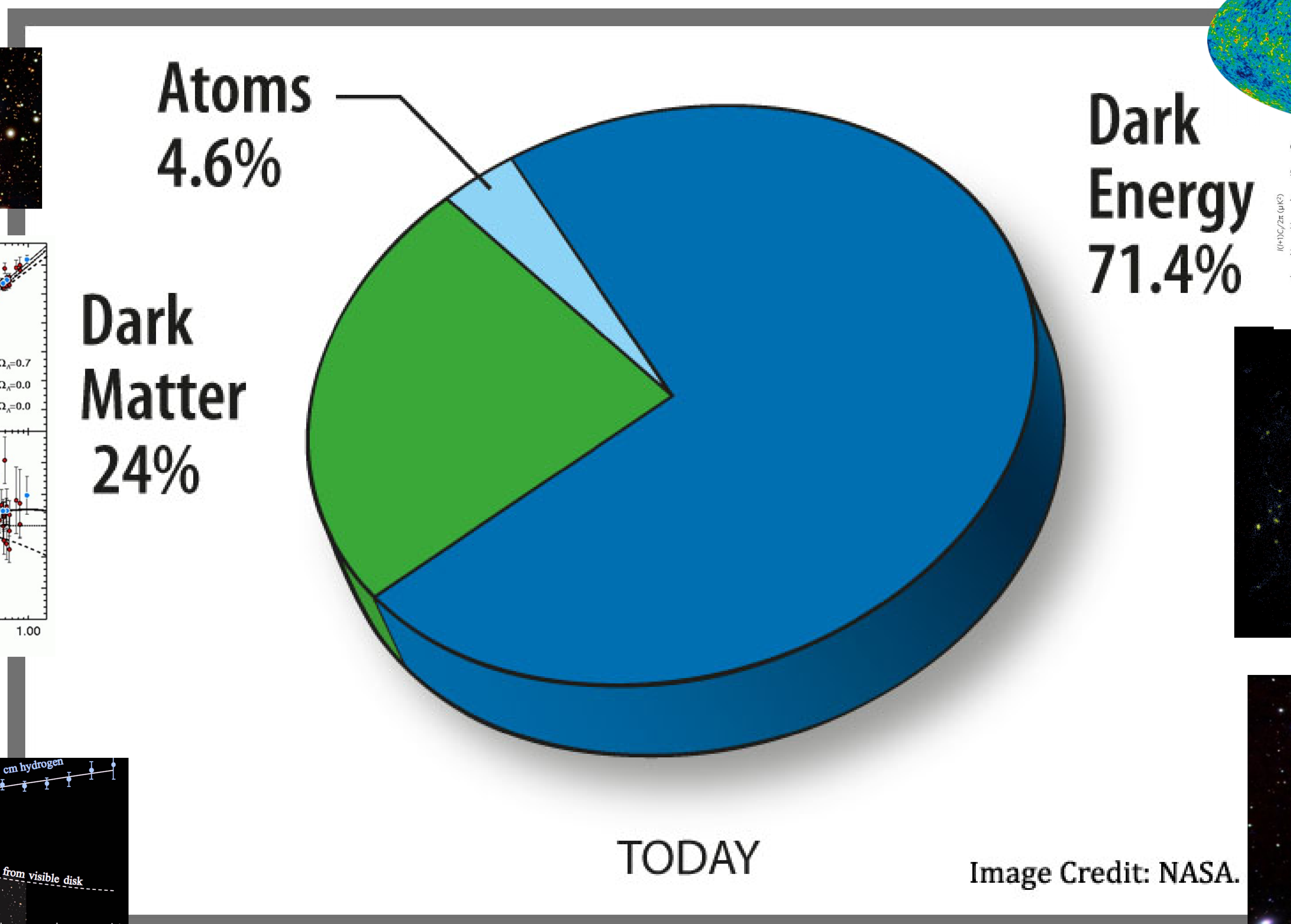
We know that the fundamental nature of > 95% of the matter+energy content of the universe eludes us: we don’t know what it is. But we know it is there!

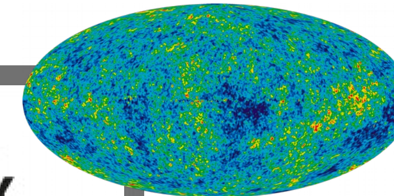
The Matter and Energy Content of The Universe

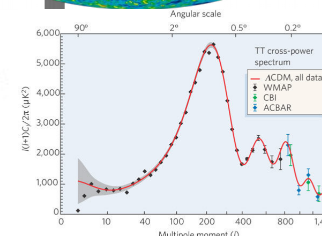


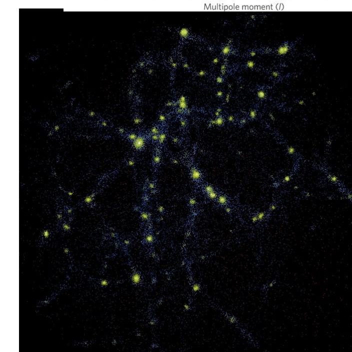






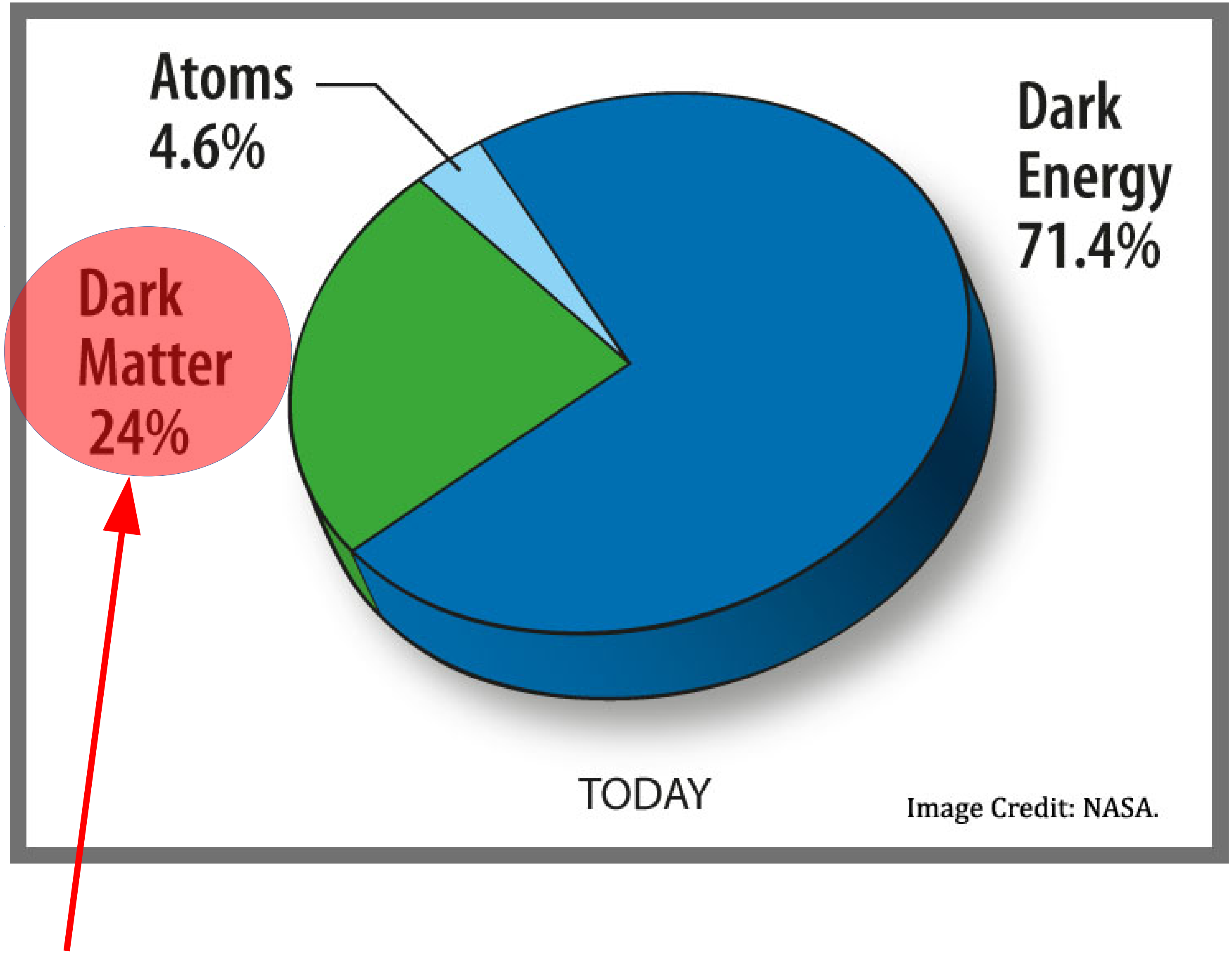






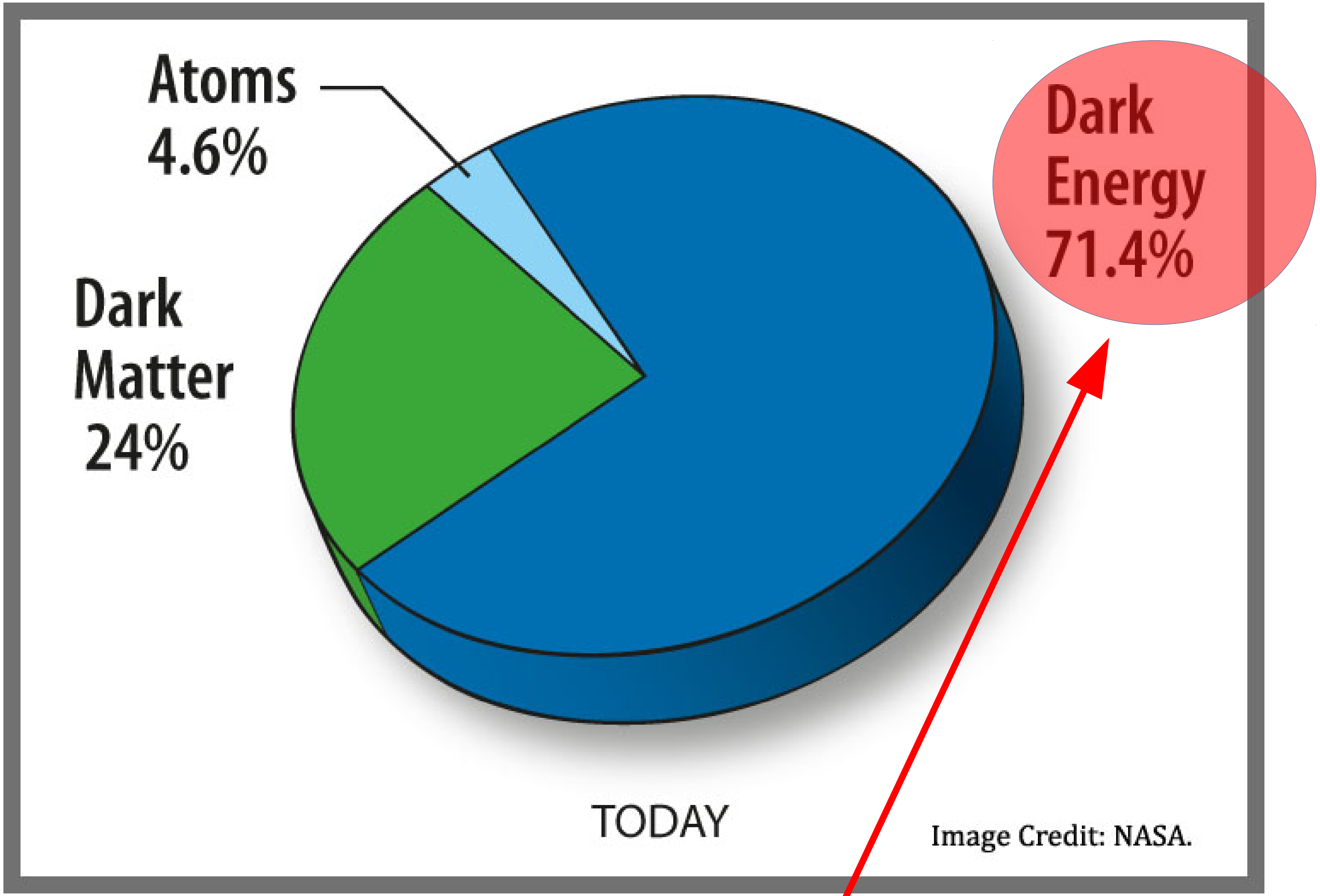
We know that the fundamental nature of > 95% of the matter+energy content of the universe eludes us: we don't know what it is. But we know it is there!

The Matter and Energy Content of The Universe



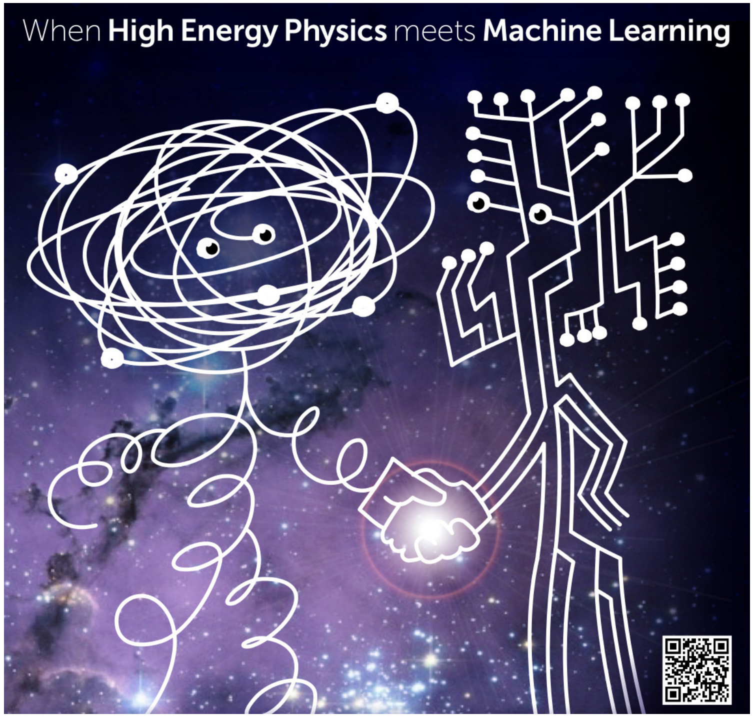
In our lifetime, we have a fair chance of shedding light [sic!] on this guy, possibly at CERN.

The Matter and Energy Content of The Universe



This guy seems to be a tougher cookie (and it pushes the universe apart)

On Particle Physics, Information, and Machines That Learn



Wolfgang Waltenberger

Institute of High Energy Physics, Austrian Academy of Sciences

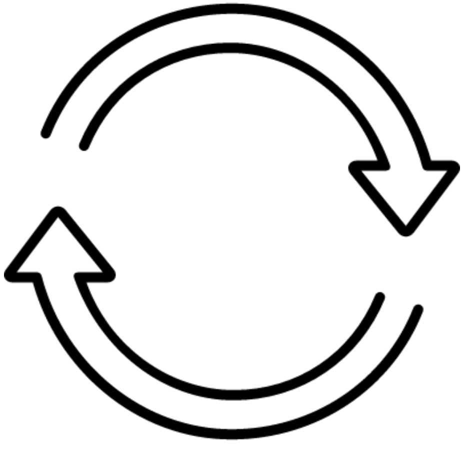
Linz, Jan 25, 2018

HEPHY
INSTITUT FÜR HOCHENERGIEPHYSIK

Outlook

In this talk I want to convey two central messages:

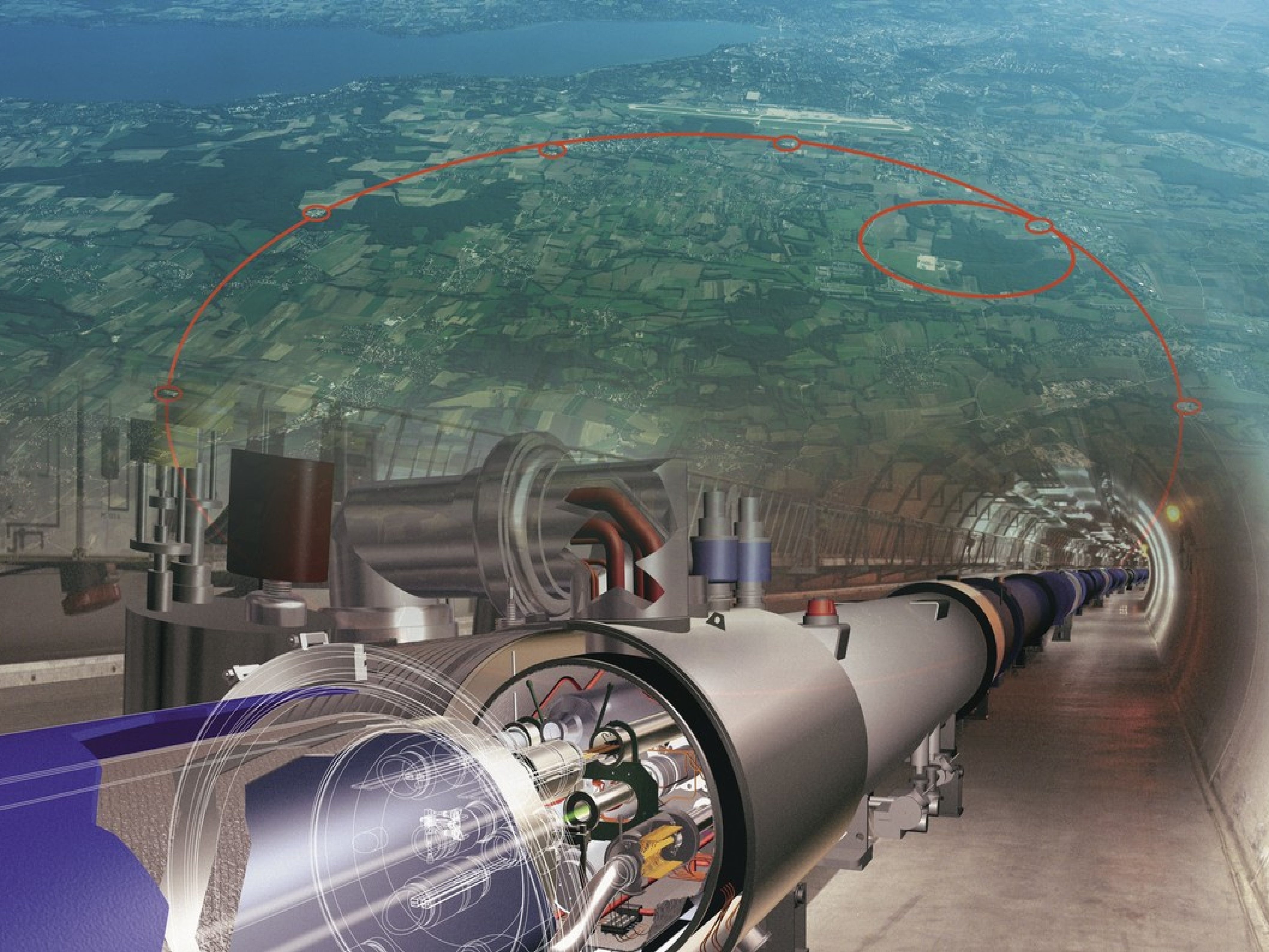
- Machine learning pushes Particle Physics

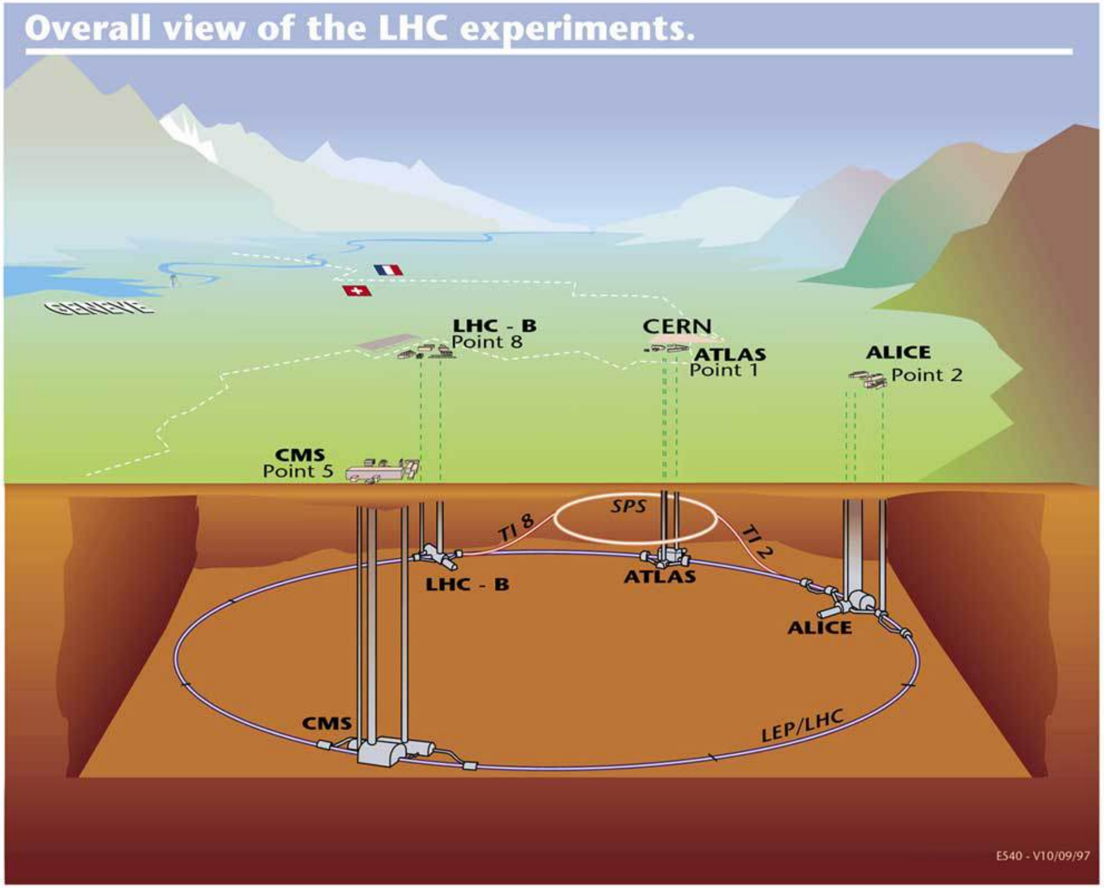


- Physics can inspire and help advance Machine Learning

Part I:

Machine Learning pushes Particle Physics





The Large Hadron Collider:

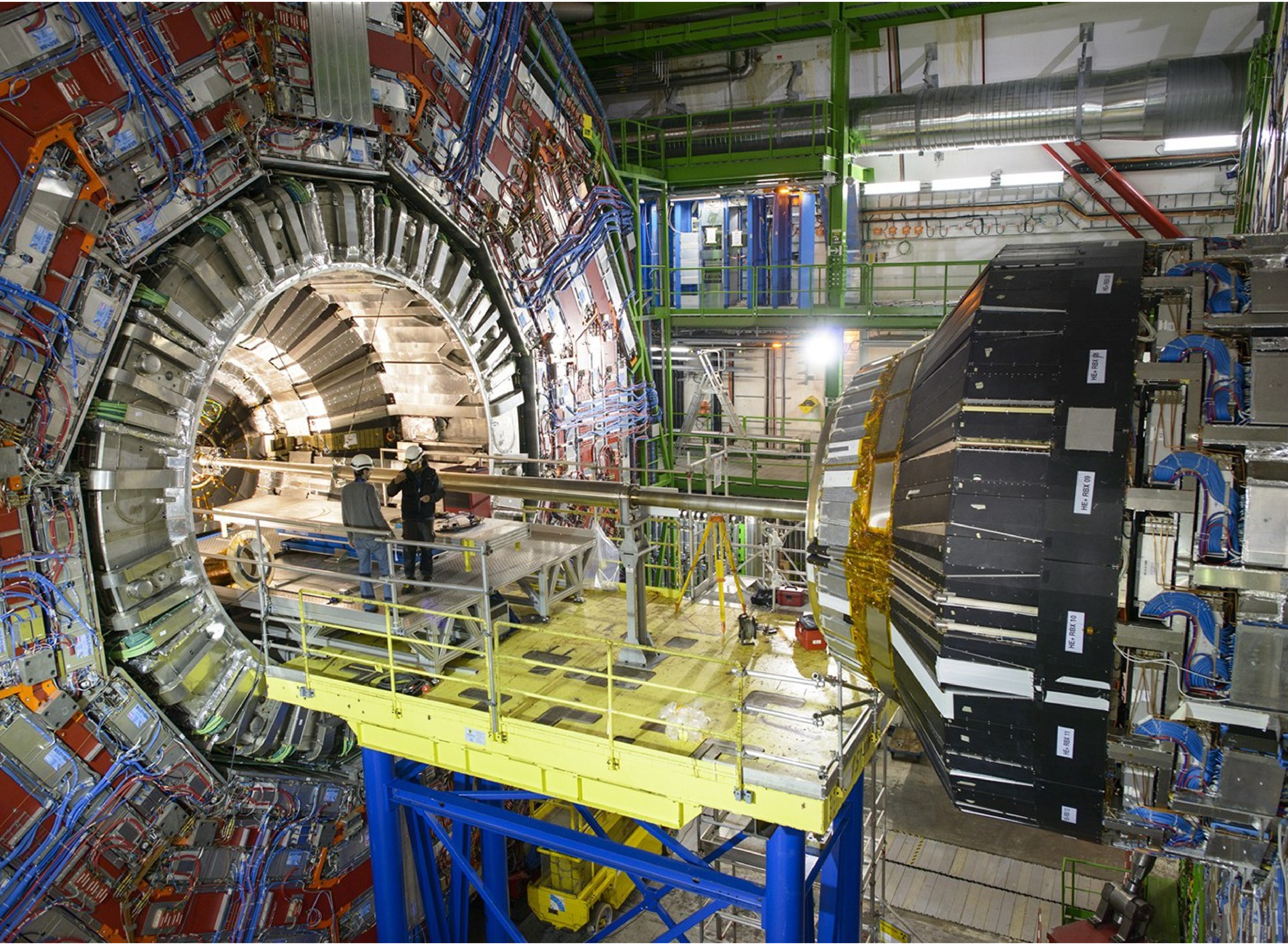
– the world's largest microscope

– 27 km circumference

– accelerating protons to 99.999999% the speed of light

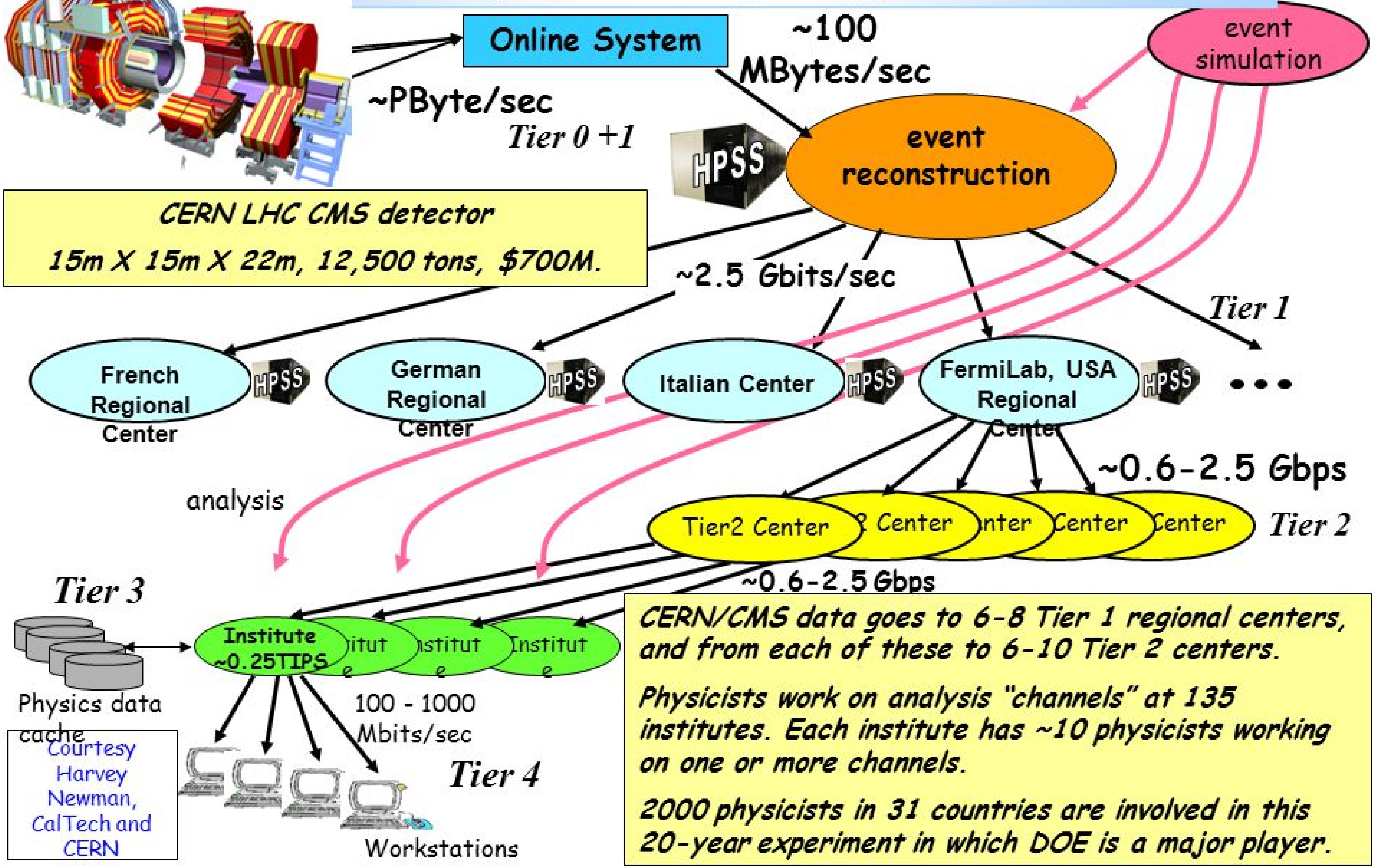
– superconducting magnets cooled to 1.9 K (-271 °C)

– operating at 40 MHz!


HE+ RBX 09
HE+ RBX 10
HE+ RBX 11

Example: High Energy Physics Data Management

CERN / LHC Data: One of Science's most challenging data management problems



CERN/CMS data goes to 6-8 Tier 1 regional centers, and from each of these to 6-10 Tier 2 centers.

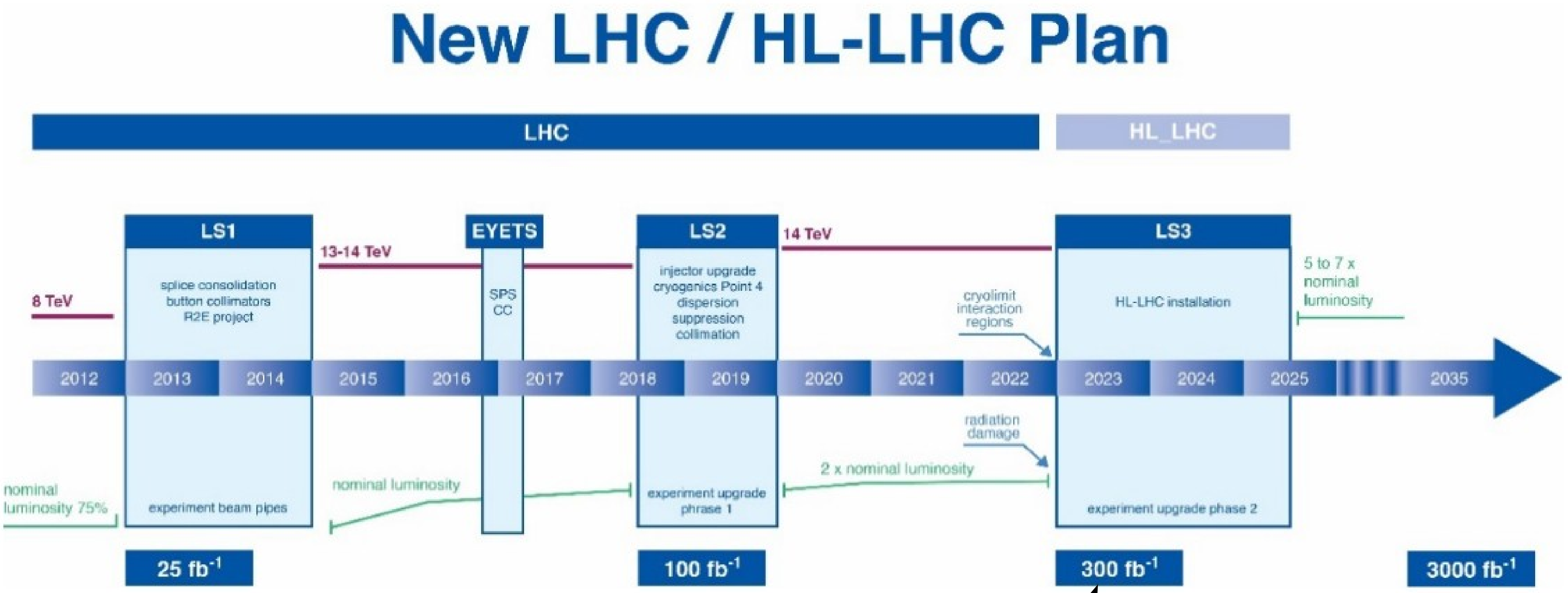
Physicists work on analysis "channels" at 135 institutes. Each institute has ~10 physicists working on one or more channels.

2000 physicists in 31 countries are involved in this 20-year experiment in which DOE is a major player.


MonALISA
File
View
Discovery
Groups
Security
Help
Normal view
OnTop view
Nodes:
No. of users
0
22
Links:
Reset
Action:
Zoom
Rotate
Scale nodes
Speed
3D Map
Groups
GMap
TabPan
Load
Clients
IO Traffic
Multi-view
CALTECH
node(s) under cursor:

Prepare for the Future

But we also start preparing for ~ 10 x more data!

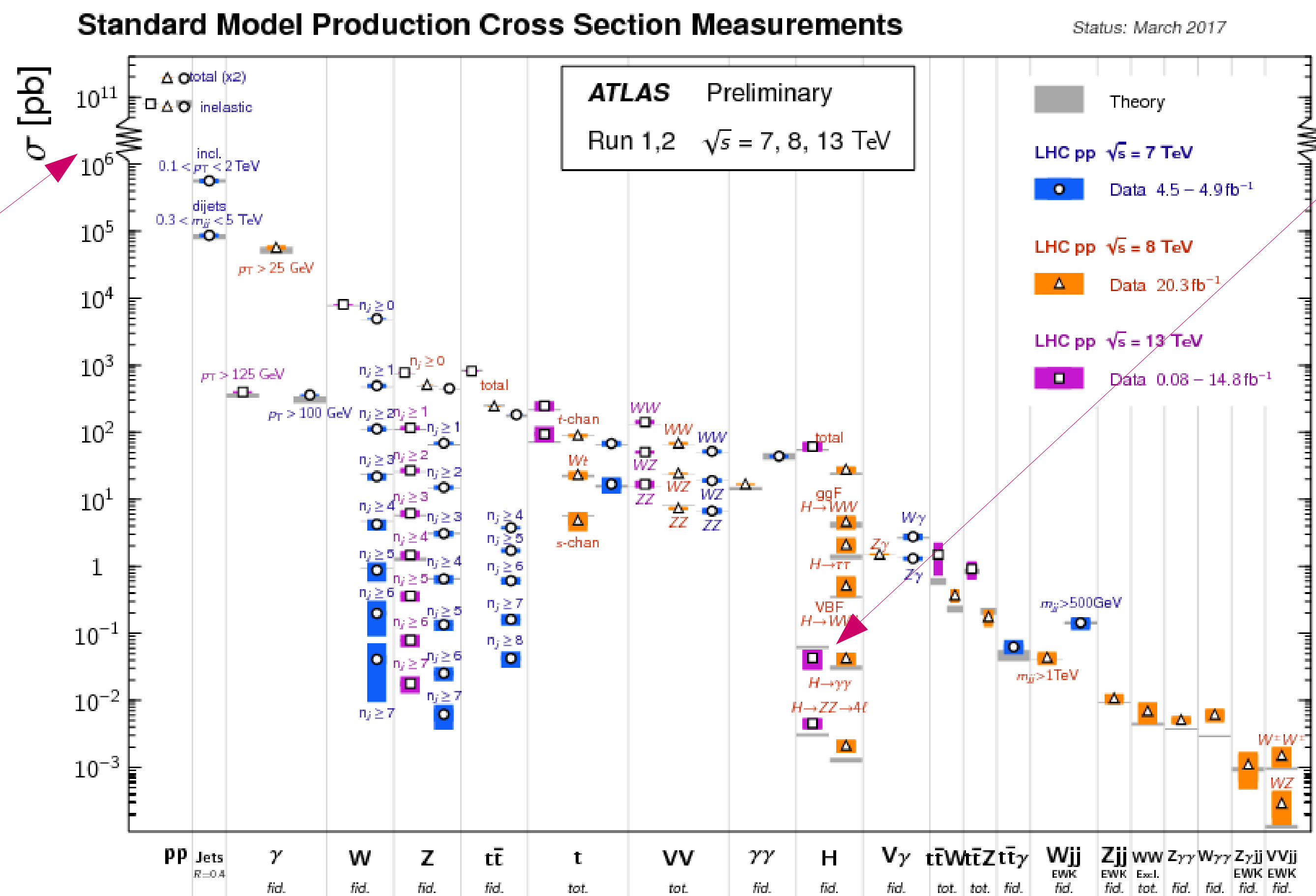


data now
(in funny units)

soon: ~ 3
times more
data

long term
future: ~ 10
more data

Why so much data? Because what we are interested in, happens only in a very few collisions. All that happens more often, we understand very well: boring!



y-axis: Measure for how often processes occur (logarithmic, with a huge gap!)

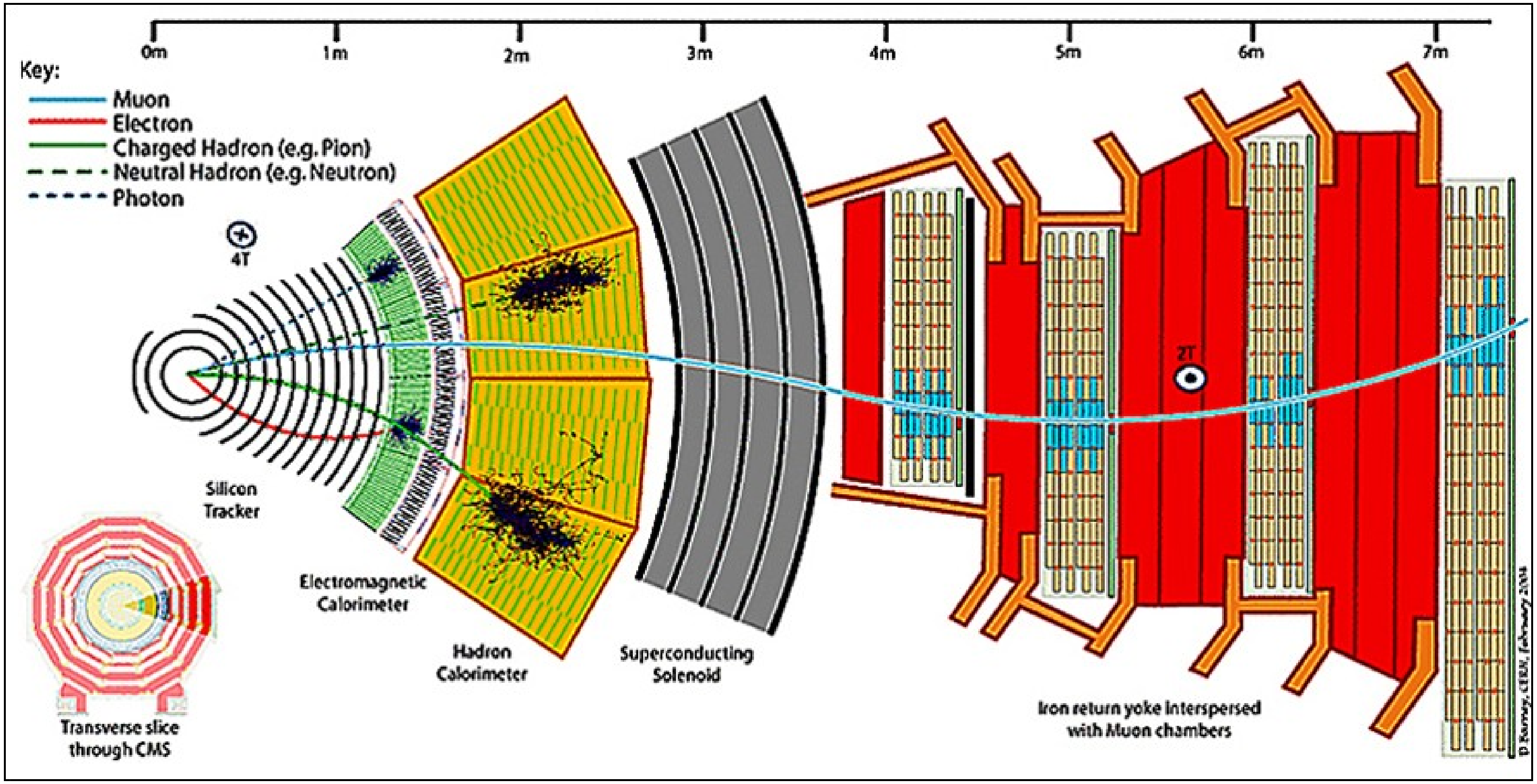
Higgs, Nobel Prize 2013

x-axis: Type of Physics Process

Pattern Recognition

From the raw data to the final statement on our fundamental physical laws, we encounter a great number of pattern recognition and statistics problems. Let me just highlight one:

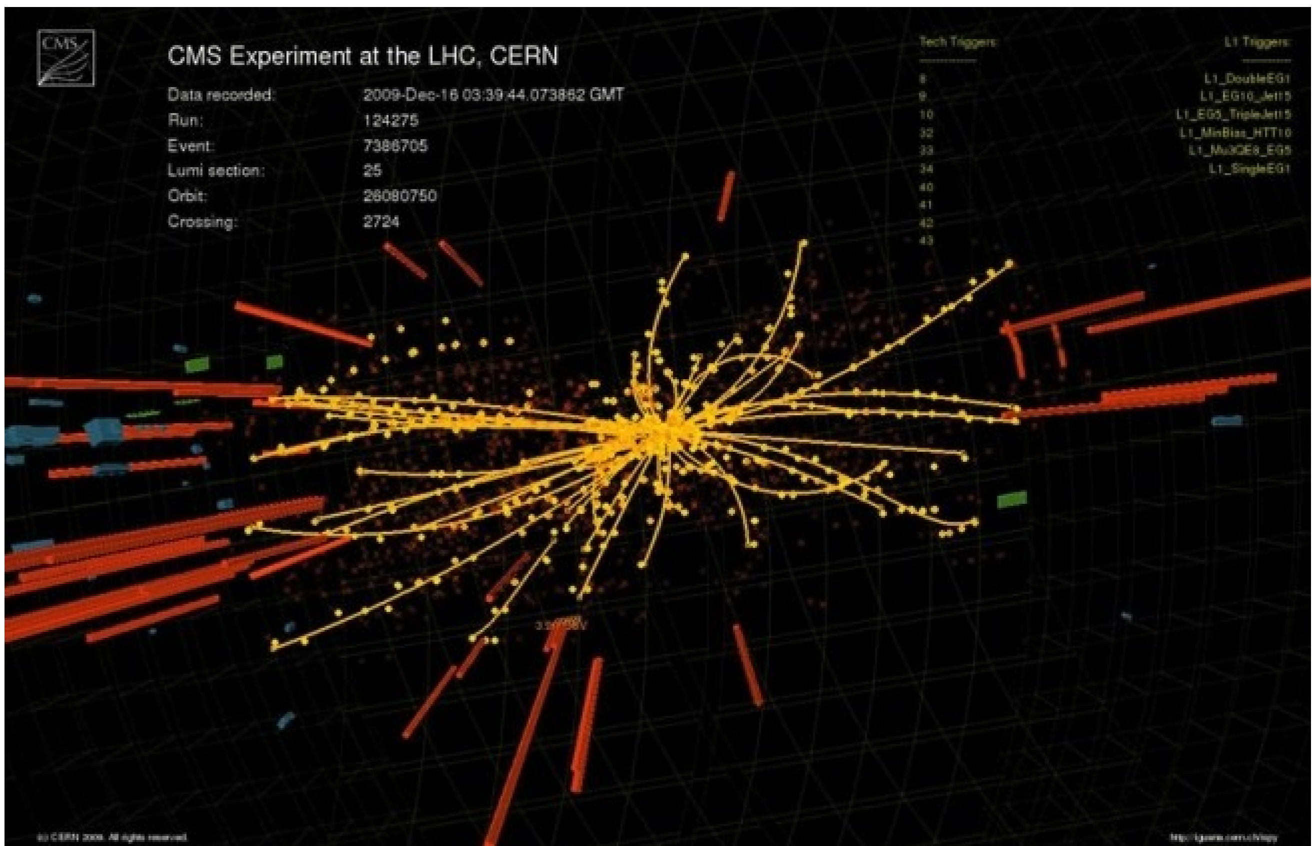
reconstruction of particle tracks



Pattern Recognition

From the raw data to the final statement on our fundamental physical laws, we encounter a great number of pattern recognition and statistics problems. Let me just highlight one:

reconstruction of particle tracks



In the past: using Kalman filters to find and fit the particle trajectories!

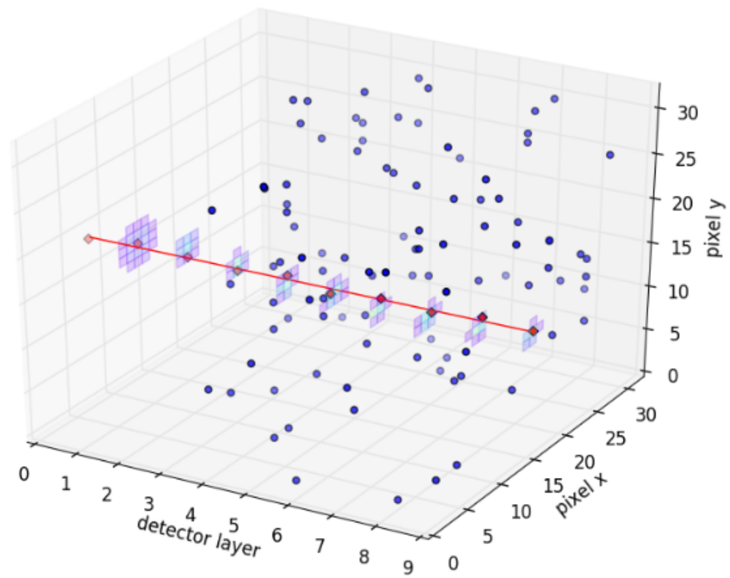
Pattern Recognition

From the raw data to the final statement on our fundamental physical laws, we encounter a great number of pattern recognition and statistics problems. Let me just highlight one:

reconstruction of particle tracks

But we are also learning lots and lots from you:

LSTM model prediction: colored squares



$$\boxed{\vec{r}_0, \vec{r}_1, \ldots, \vec{r}_N} \rightarrow \text{bi-GRU} \rightarrow \text{bi-GRU} \rightarrow \text{bi-GRU} \rightarrow \text{FC} \rightarrow \boxed{\vec{p}_0, \vec{p}_1, \ldots, \vec{p}_N}$$

$\vec{r} = (r, \phi, z)$

$\vec{p} = (p_{\mathrm{trk1}}, p_{\mathrm{trk2}}, \ldots, p_{\mathrm{trkMax}})$

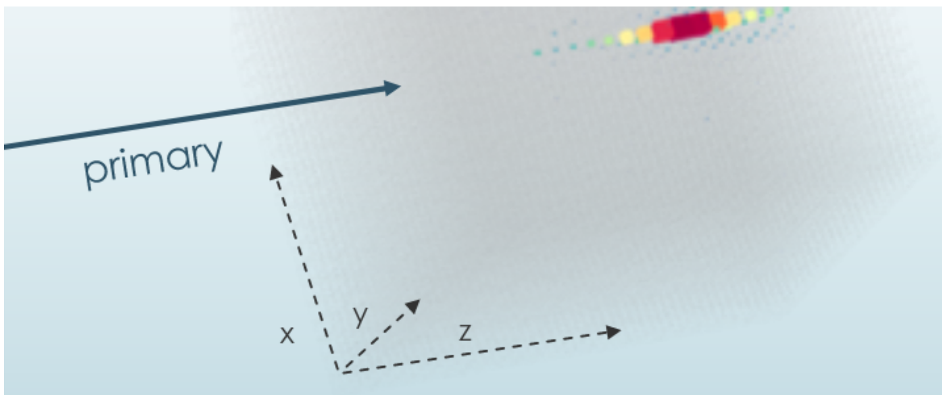
Performance not yet good enough for HL-LHC, so they intend to tackle graph-network learning next.

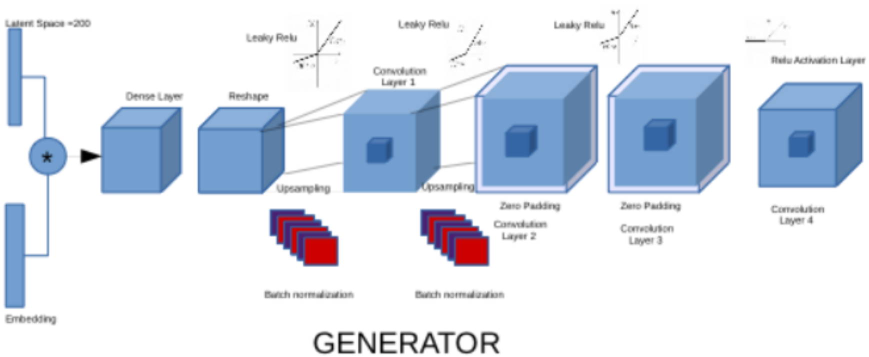
Deep Learning for Physical Sciences, NIPS 2017

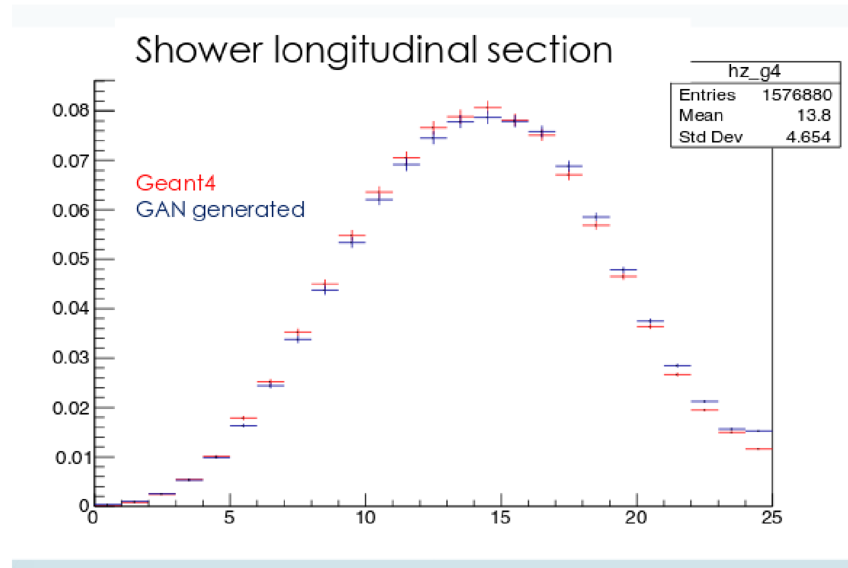
Fast simulation of “events”

Another adaption of a novel machine learning technique by particle physcists:

Using GANs for a fast simulation of proton-proton collision “events”:







“Fully simulated events” (red) and much faster “GAN-generated events” (blue) are similar

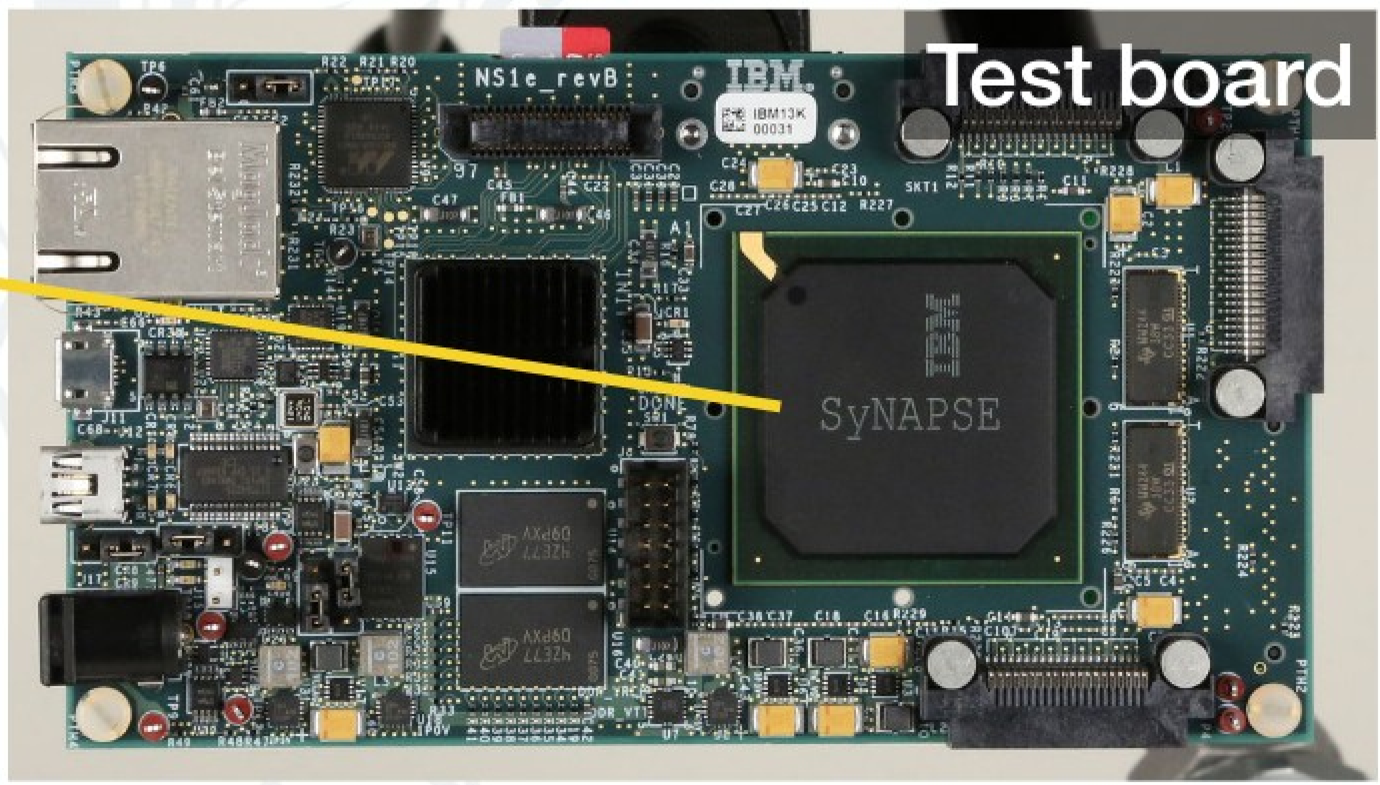
ACAT 2017, “Generative models for fast simulation”

Neuromorphic Hardware

Neuromorphic Chips: experimental, post-von-Neumann architecture, loosely modeled after human brain (analog – “spiking” signals).

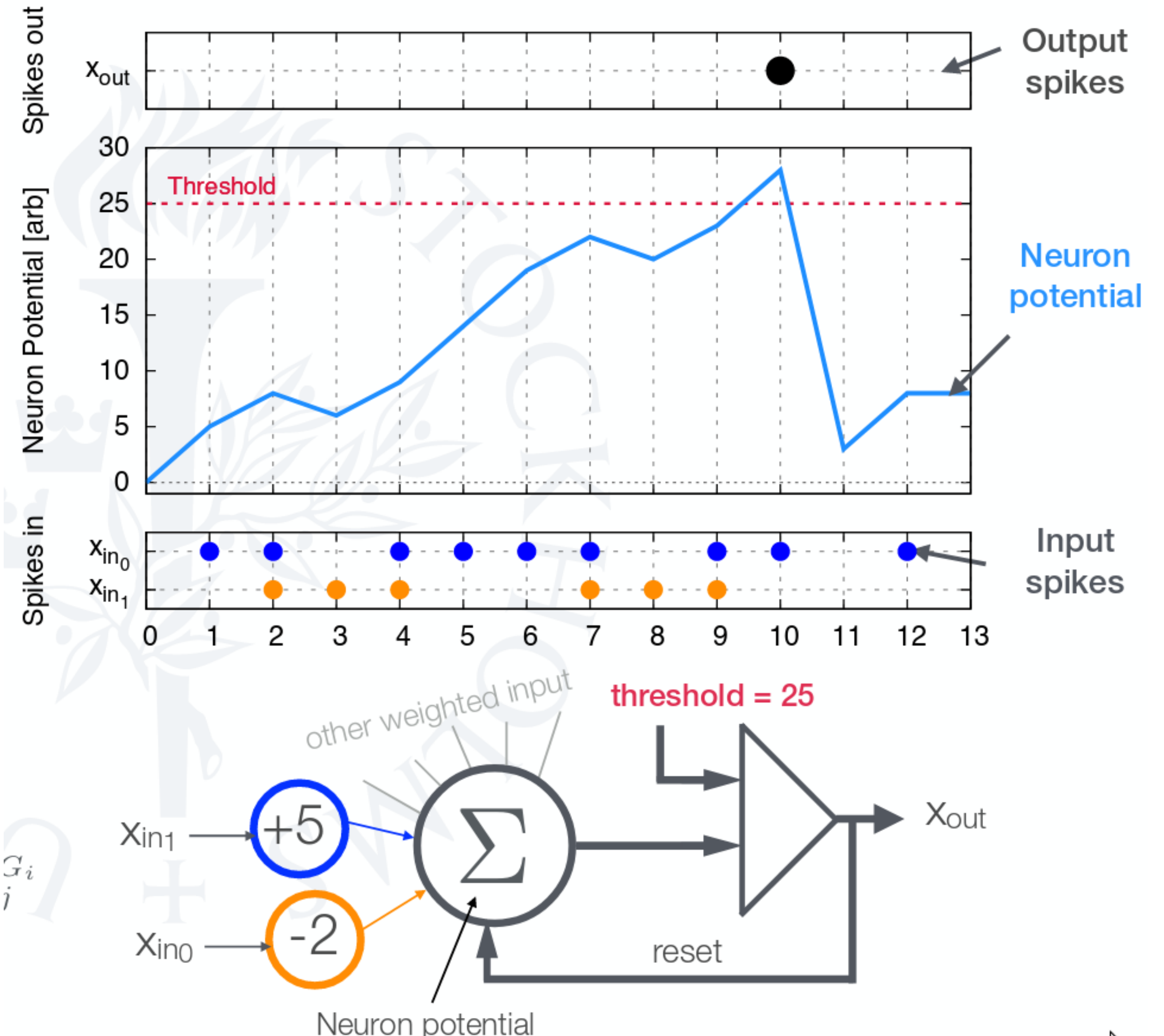
“True North” architecture: experimental program conducted by DARPA and IBM

Power consumption tens of Milliwatts!!!!



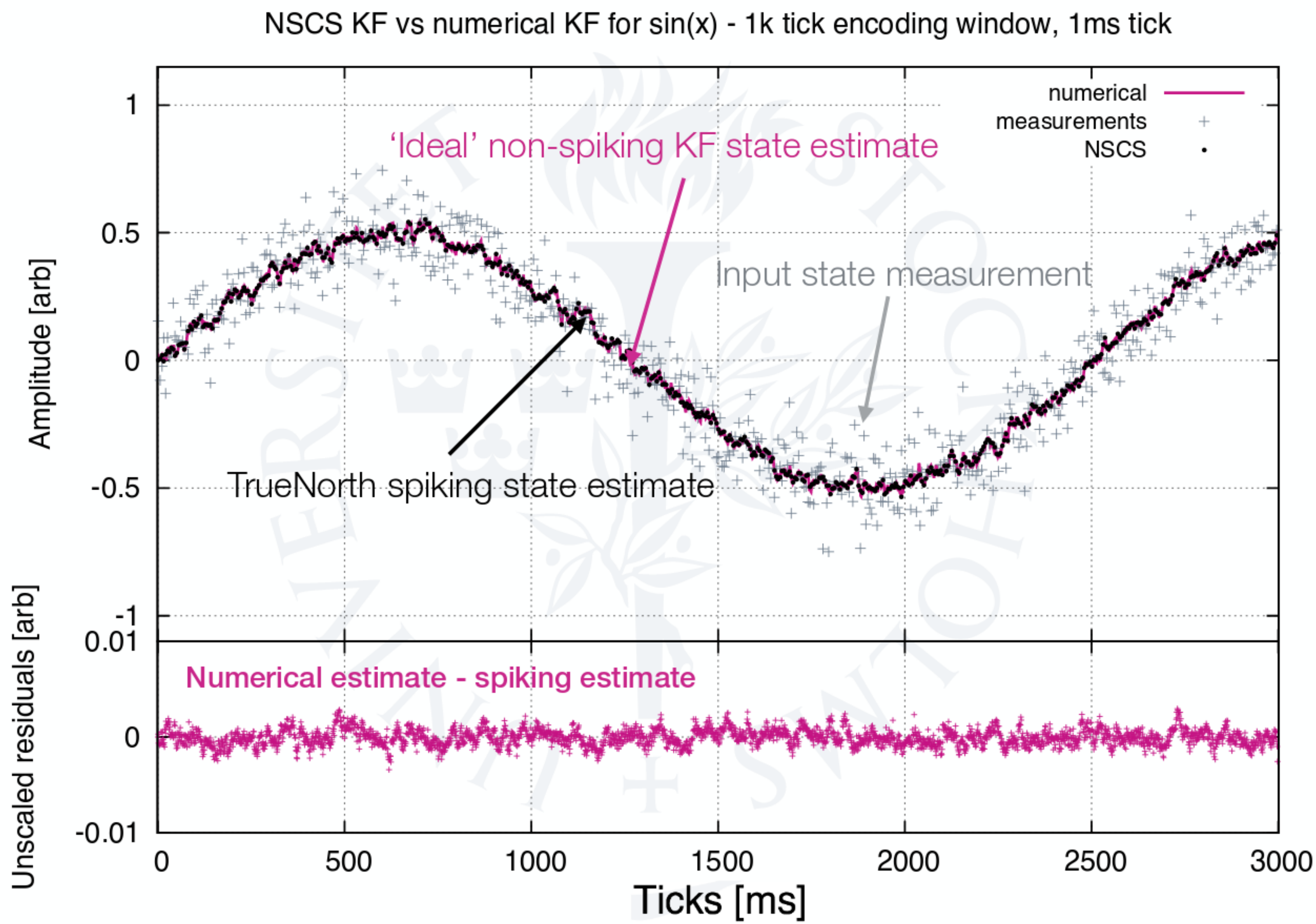
Neuromorphic Hardware

Neuromorphic neurons in TrueNorth



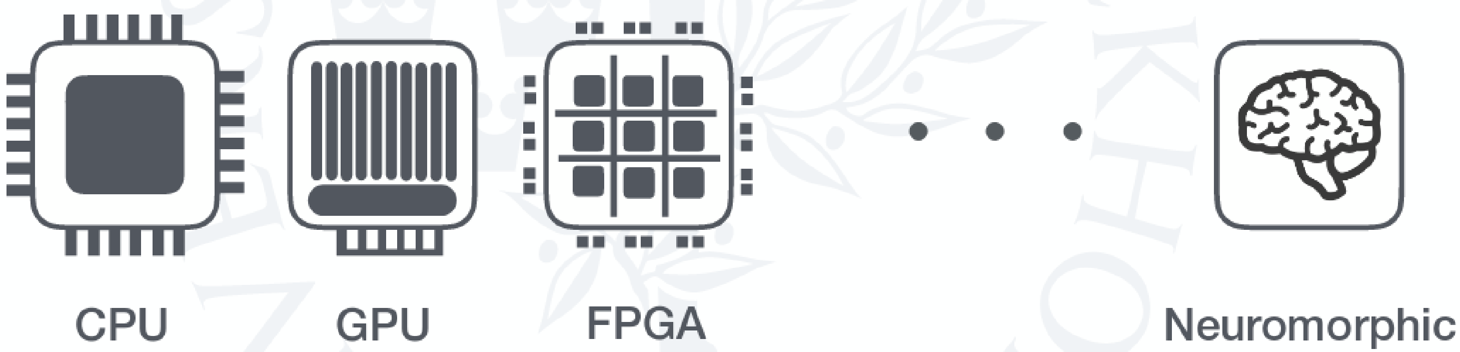
Rebecca Carney, CTD2017 Talk

Neuromorphic Hardware



Rebecca Carney, CTD2017 Talk

Neuromorphic Hardware



- We have adapted CPU's, GPU's and FPGA's for our needs. Could neuromorphic chips feature in our toolkit in 10 years time? Perhaps but it seems more likely we would use an ASIC ANN.

Rebecca Carney, CTD2017 Talk

Quantum Machine Learning

- Q: And what about the intersection between quantum computing and machine learning?

- A: Not here. Not today. Not me.

  http://www.quantummachinelearning.org/

Summary: Part I

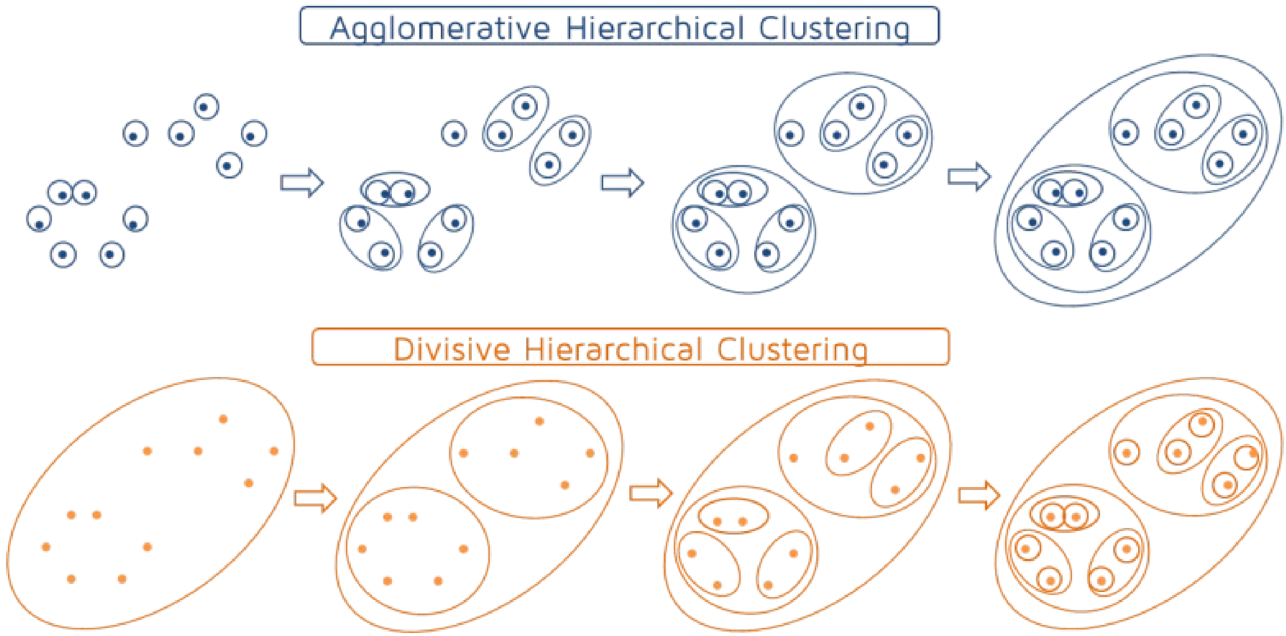
- We have $\rightarrow \infty$ amount of data
- We have $\rightarrow \infty$ computing ressources
- We have a lot of exciting machine learning problems
- For the “online” data processing we have
  $t \rightarrow 0$ time. So we need ML also in FPGAs, ASICs
- We need to prepare for the future “High-Lumi” LHC: 10x more data!!
  Investigating in novel hardware and algorithms
- Contributions also from purely data-scientific
  institutions (e.g Yandex)

Part II:
Physics can inspire and elucidate Machine Learning

Dynamic Quantum Clustering

- Task: to find natural clusters of points in abstract “feature” spaces. Find an algorithm that does well with many dimensions and big data

- Millions of algorithms exist, e.g. very simple hierarchical clusterers:



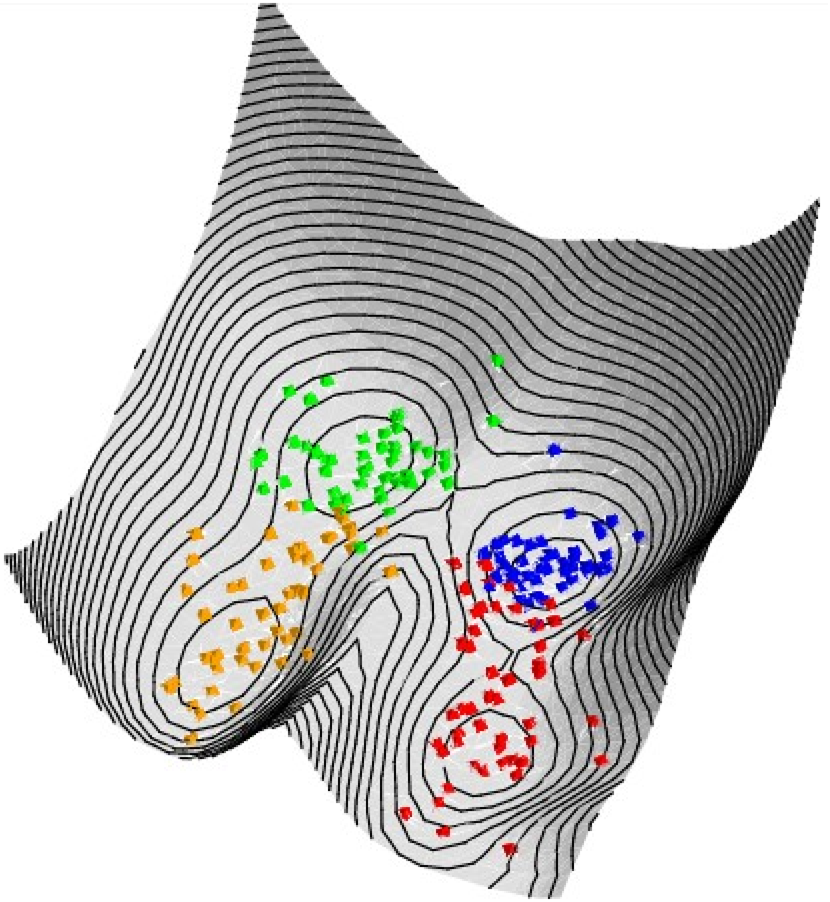
Dynamic Quantum Clustering

- But can we exploit the notion of quantum mechanical superposition to identify an algorithm with low computational complexity?

  Idea:

- Treat data points as if they were electrons (in their ground states).
- Superimpose them (=add up their

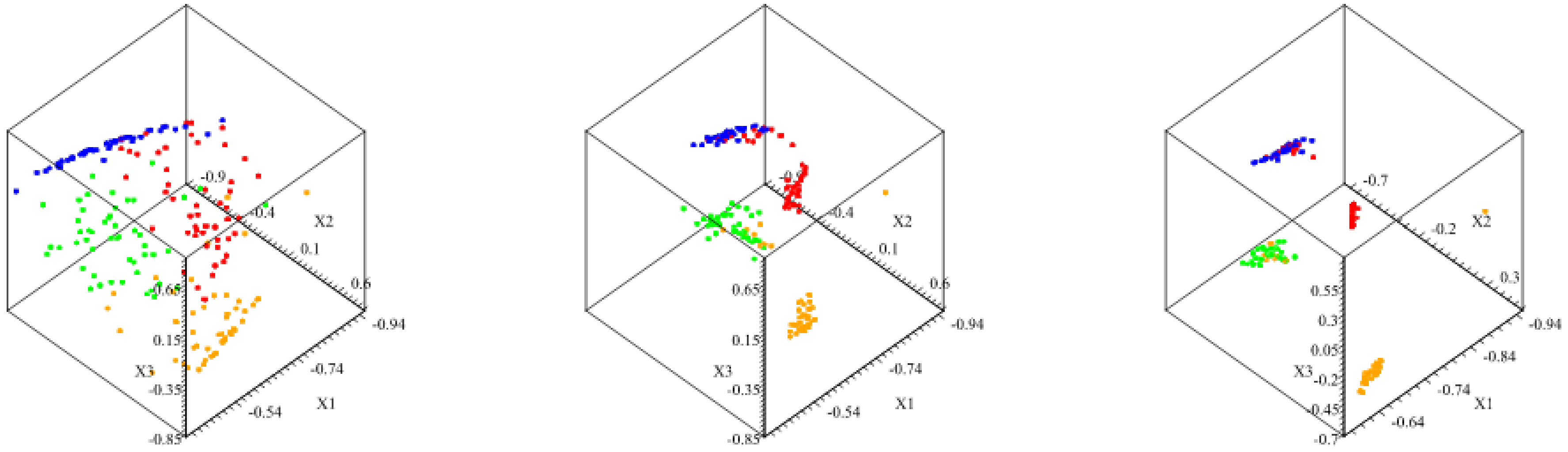
  wave functions)
- Use Schrödinger equation to find the according potential (easy!!).
- The minima of the potential are the cluster centers.



Dynamic Quantum Clustering

Use time-dependent Schrödinger equation (pretend electrons attract each other) to move data points into cluster centers.

→ Clustering and visualisation algorithm with low computational complexity!



https://arxiv.org/pdf/0908.2644.pdf

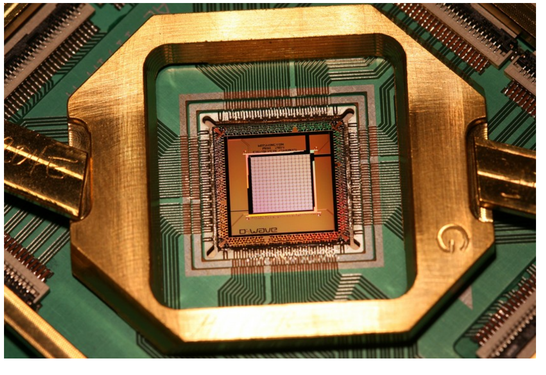
https://arxiv.org/pdf/1310.2700.pdf

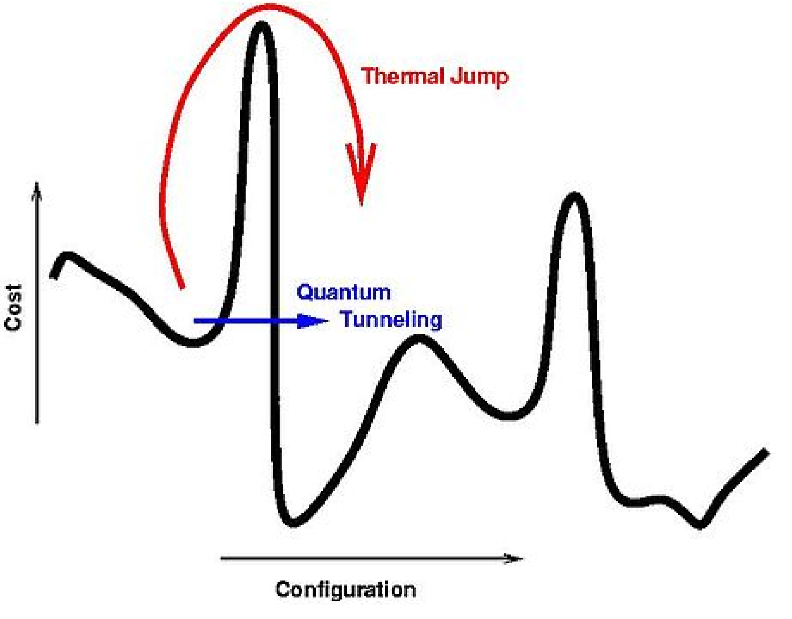
open source GPU implementation
https://github.com/peterwittek/dqc-gpu

Quantum Annealing

- Task: find global minimum of function; avoid local minima
- Classical strategies: e.g. simulated annealing.
- Quantum annealing: exploit the tunneling effect of quantum mechanics to overcome local minima. Can be done on a d-

  “quantum” computer





Article

Information Geometry

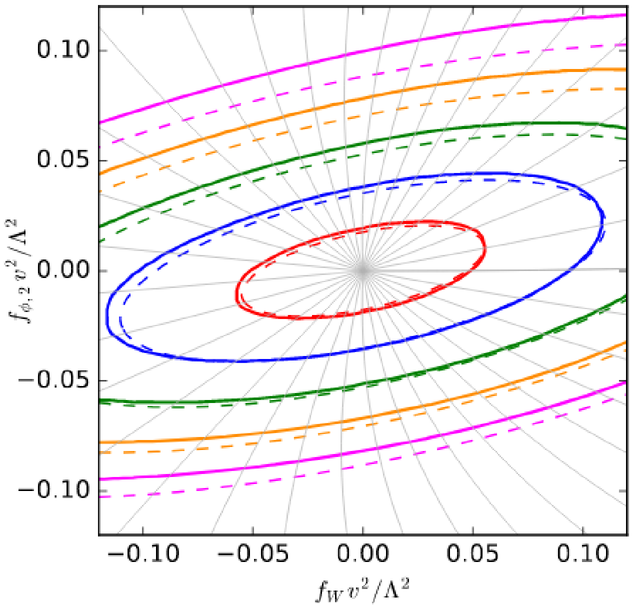
- Idea: given an abstract feature space, interpret Fisher information matrix as the “curvature” of the information space – similar to curved spacetimes (general relativity) in modern physics

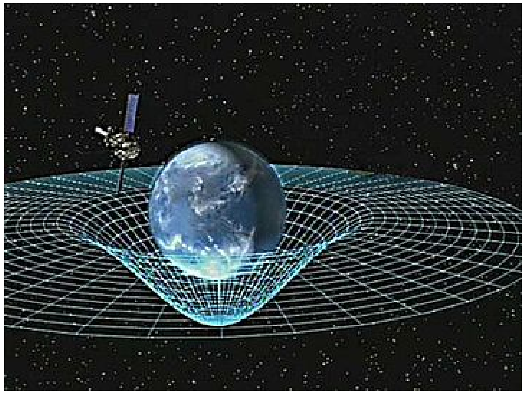
- Distance between points is a direct statistical measure for how more (un)unlikely it is to measure a value.

$$d(\mathbf{g}_b, \mathbf{g}_a) = \min_{\mathbf{g}(s)} \int_{s_a}^{s_b} ds \sqrt{\frac{dg_i(s)}{ds} I_{ij}(\mathbf{g}(s)) \frac{dg_j(s)}{ds}} ,$$

https://arxiv.org/pdf/1612.05261.pdf

Information Geometry





Information geometry showing what we know about two specific attributes of the Higgs boson. Contours are lines with same "distance" (compatibility) from center.

May help optimize our data analyses.

https://arxiv.org/pdf/1612.05261.pdf

Why is deep learning so cheap?

- E.g. image recognition: the space of all possible inputs is huge! In comparison, the training samples are tiny. How can this work at all?

- Ref arXiv:1608.08225 tackles the problem with an analogy with physics: why can we understand the universe in terms of fundamental Hamiltonians?

  (Hamiltonians are operators that “encode” the law of conservation of total energy)

Why is deep learning so cheap?

Physics	Machine learning
Hamiltonian	Surprisal $-\ln p$
Simple H	Cheap learning
Quadratic H	Gaussian p
Locality	Sparsity
Translationally symmetric H	Convnet
Computing p from H	Softmaxing
Spin	Bit
Free energy difference	KL-divergence
Effective theory	Nearly lossless data distillation
Irrelevant operator	Noise
Relevant operator	Feature

TABLE I: Physics-ML dictionary.

Example. If a Hamiltonian is something similar to a negative log-likelihood, then softmaxing might be understandable as an application of Bayes Theorem and have a lot in common with the Boltzmann distribution of thermodynamics:

$$\tilde{\boldsymbol{\sigma}}(\mathbf{x}) \equiv \frac{e^{\mathbf{x}}}{\sum_i e^{y_i}}.$$

$$p(y|\mathbf{x}) = \frac{p(\mathbf{x}|y)p(y)}{\sum_{y'} p(\mathbf{x}|y')(y')},$$

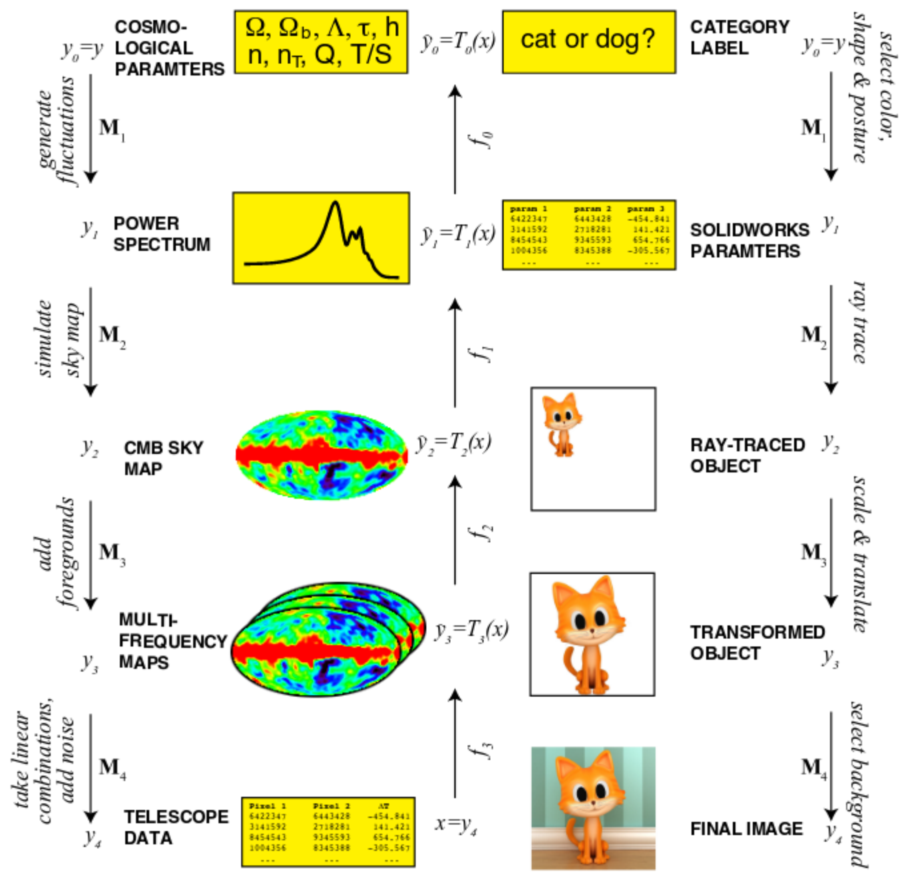
$$p(y|\mathbf{x}) = \frac{1}{N(\mathbf{x})} e^{-[H_y(\mathbf{x})+\mu_x]}, \tag{3}$$

where

$$N(\mathbf{x}) \equiv \sum_y e^{-[H_y(\mathbf{x})+\mu_y]}. \tag{4}$$

Why is deep learning so cheap?

The authors identify a few (intuitive) factors: locality, symmetry, log-polynomial NLLs, and compositeness:



→ **No-Flattening Theorems** as practical consequences (when can you or can you not “flatten” a deep network?)

Summary part II: physics-inspirations

- Metaphors from quantum mechanics and other fields in physics may be exploited for the purposes of statistics and machine learning

- The question of physical / thermodynamic foundations of (machine) learning arises. Can we hope to improve learning from such analogies?

CERN openlab (computing, not physics!) takes applications for their summer student program:

http://cern.ch/go/SLd9

Google Summer of Code

Introduction

Google Summer of Code is a program that allows students to contribute to development of open-source projects, mentored by participating organizations.

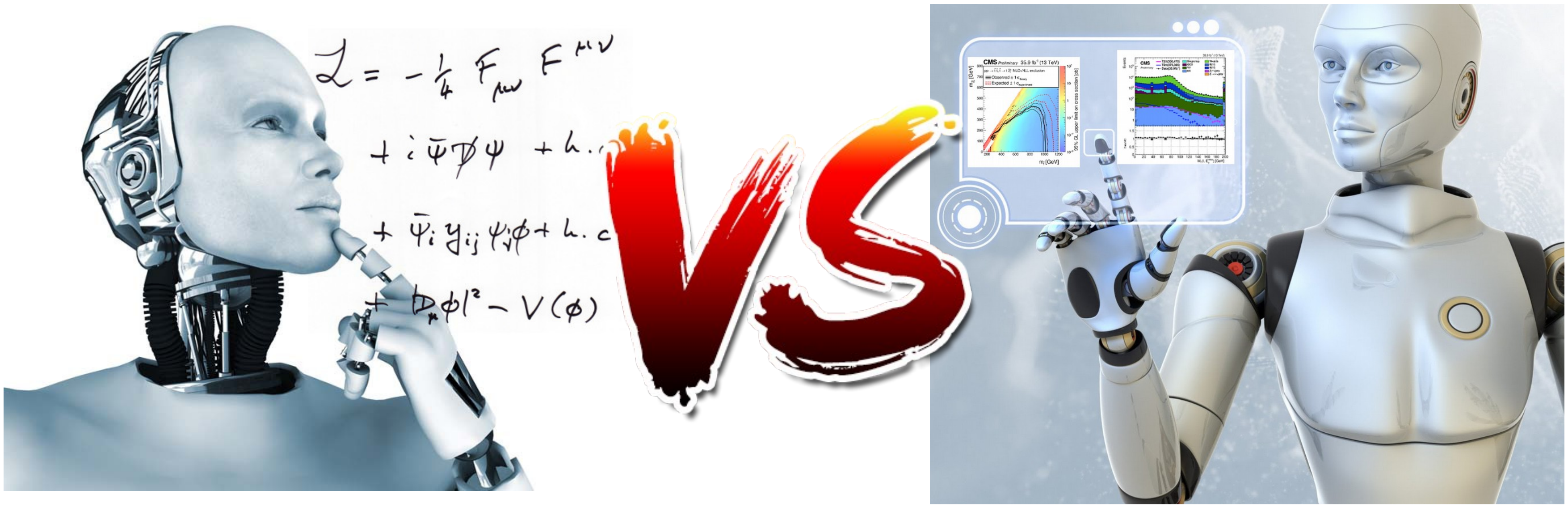
Particle physics is an exciting field where large collaborations of scientists collect and analyze petabytes data from high-energy physics experiments, such as the Large Hadron Collider, hosted at the CERN laboratory in Geneva, Switzerland. Some of the questions that we collectively ask are:

- what are the fundamental blocks that make up our Universe?
- what is the nature of dark matter and dark energy?
- what is the nature of the asymmetry between matter and antimatter?
- what was early Universe like?

http://hepsoftwarefoundation.org/activities/gsoc.html

And now let me briefly flash one of my own future projects (with ~ 10 colleagues + students) for the next few years ….

Unsupervised Learning of the Next Standard Model: a roadmap

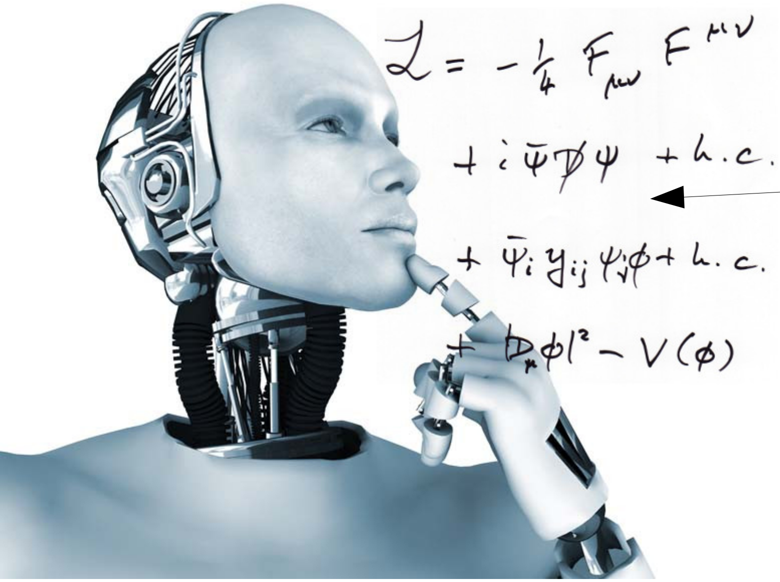


Wolfgang Waltenberger
HEPHY Vienna

with the SModelS group

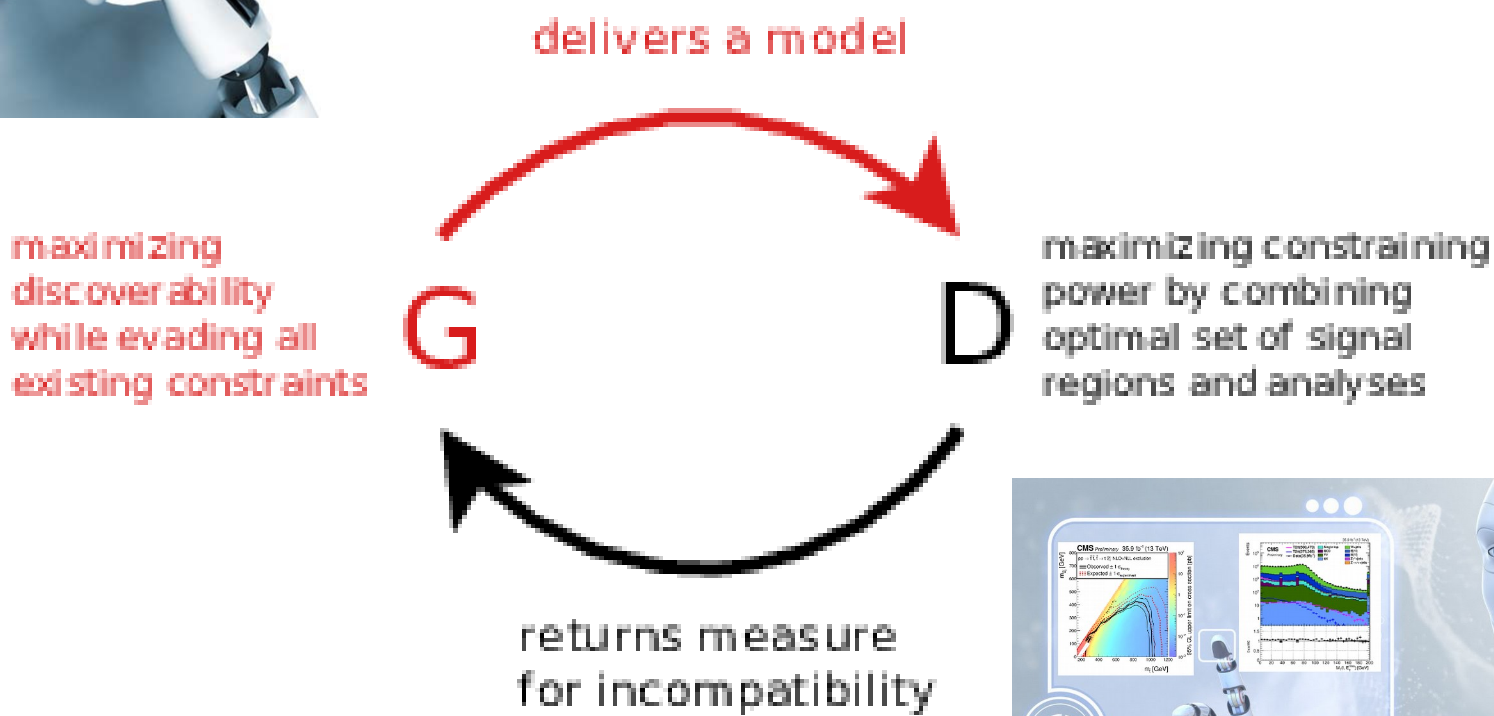
Workshop: LHC Chapter II
Natal, Brazil, 6 – 17 November 2017

Slides



Similar to Generative Adversarial Networks:

A “theoretical physicist machine” **G** invents new fundamental theories



A “data analyst machine” tries to statistically falsify the theories based on LHC data

Summary

- Particle Physics is a field with lots of data, and lots of interesting applications of machine learning
- Physics (like the neuro sciences, or biology) may help elucidate on the effectiveness of machine learning
- Lots and lots of activity in machine learning, in particle physics, and in machine learning in particle physics
- Must machines become eligible for Physics Nobel Prizes? ;)

Thank you!