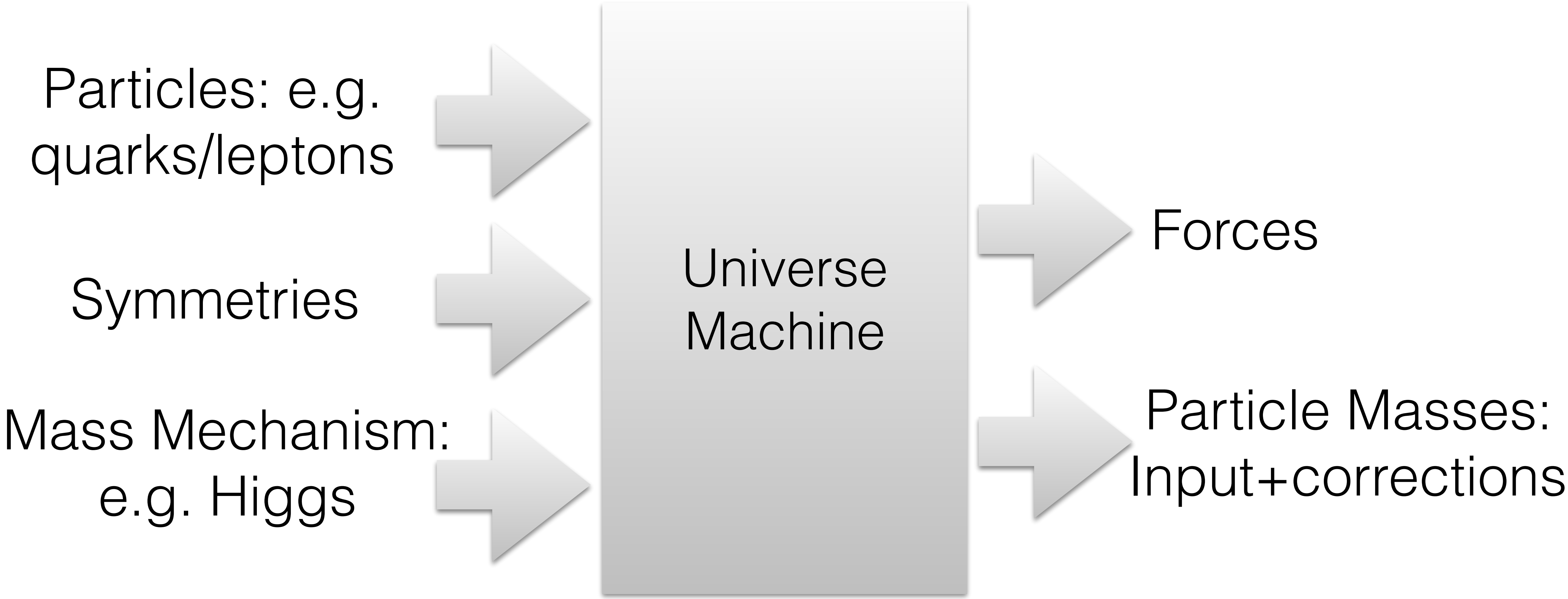
Deep Learning in HEP

How AI will tell us if the Universe was an accident.

Amir Farbin
University of Texas at Arlington

Building a Universe



Particles: e.g. quarks/leptons

Symmetries

Mass Mechanism: e.g. Higgs

Universe Machine

Forces

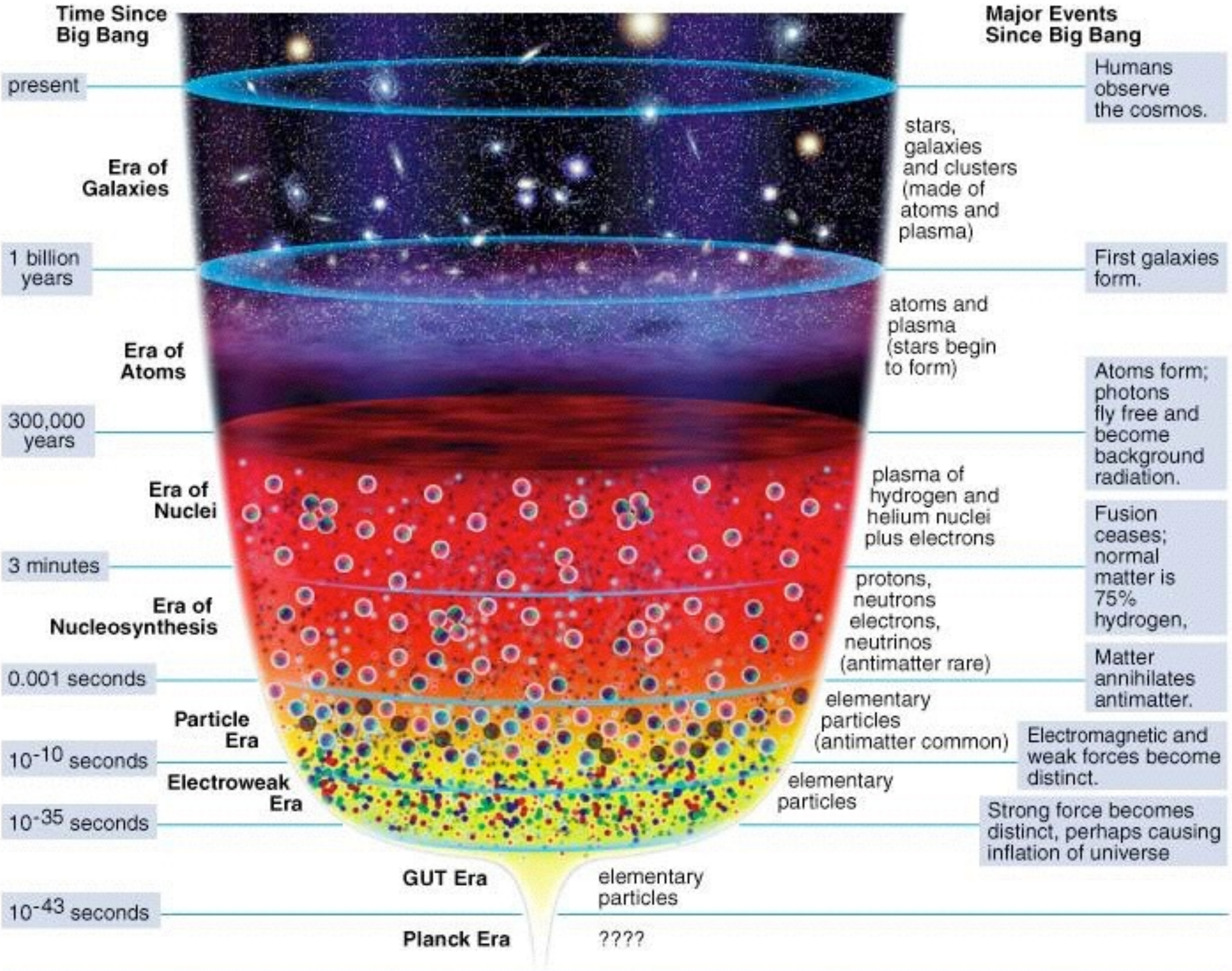
Particle Masses: Input+corrections

Emergence of Structure

Today

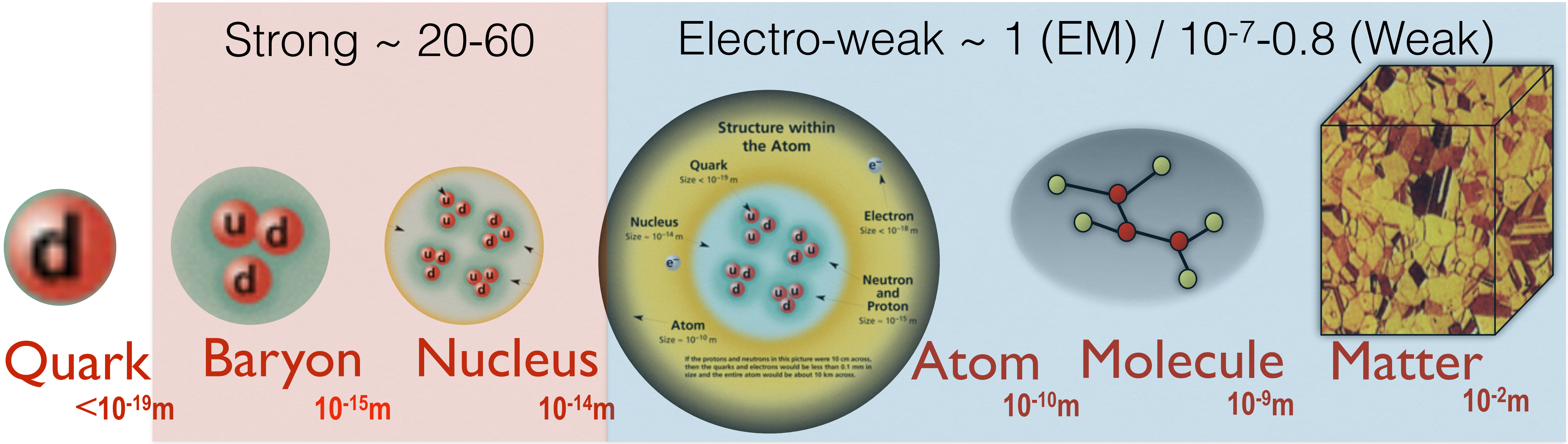
13.7 Billion Years

"Beginning"



Building a Universe…

- Each “structure” is due to some fundamental force.
    - The stronger the force the smaller the structure.
    - The weaker the force the larger the structure.



Strong ~ 20-60

Electro-weak ~ 1 (EM) / 10^{-7}-0.8 (Weak)

Quark $<10^{-19}$m

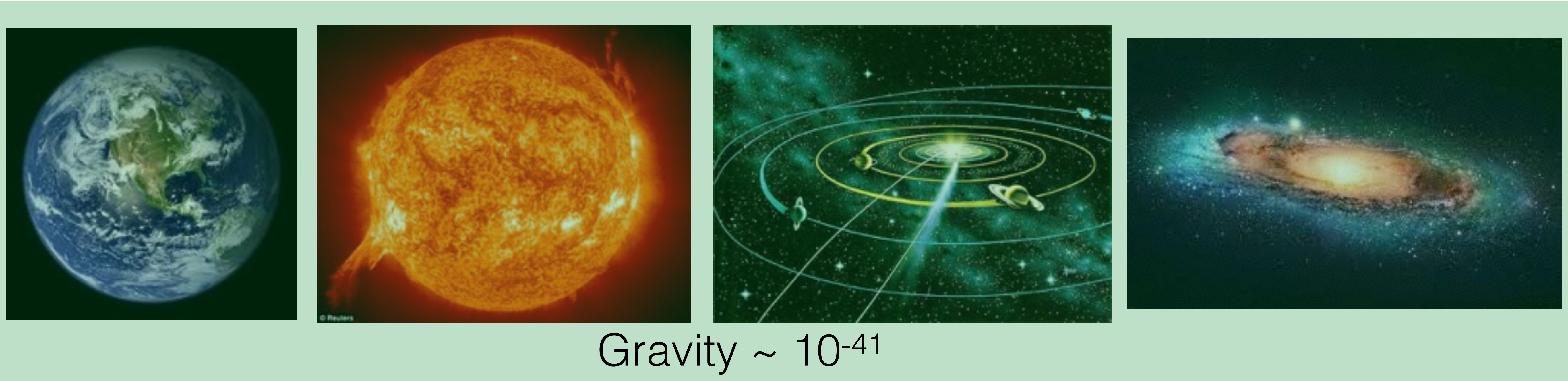
Baryon 10^{-15}m

Nucleus 10^{-14}m

Atom 10^{-10}m

Molecule 10^{-9}m

Matter 10^{-2}m



Gravity ~ 10^{-41}

Higgs Fine-tuning

Measured = Bare + Correction

$$m_H^2 = m_0^2 + \delta m_0^2$$

Measured 125 GeV

We input Bare

$\delta m_H^2 \sim \Lambda^2$ (ie large)

Need in part in 10^{16} cancellation
to get m_H correct.

Alternative: New Physics at energy Λ fixes the problem.

Value of Λ depends on how much fine-tuning.

Why is the Higgs light?

- ***Chance (Fine-tuned)—*** very very unlikely to get these parameters…
  - perhaps:
    - *multiverse*- there are lots of Universes.
    - *anthropic principle*- we are in a Universe in which we can exist.
- ***Naturalness***- Small numbers don't in nature.
  - There is some symmetry, force, structure that control the constants…
    - Add new particles / symmetries
  - A aesthetic principle that constants should be of order 1.
    - Therefore any observed small/fine-tuned number is due to some phenomena.
  - For example for the Higgs mass, it can be Supersymmetry, extra-dimensions, additional sub-structure.
  - This is LHC's primary mission. Basically look for something new.
- ***Design?***

Deep Learning

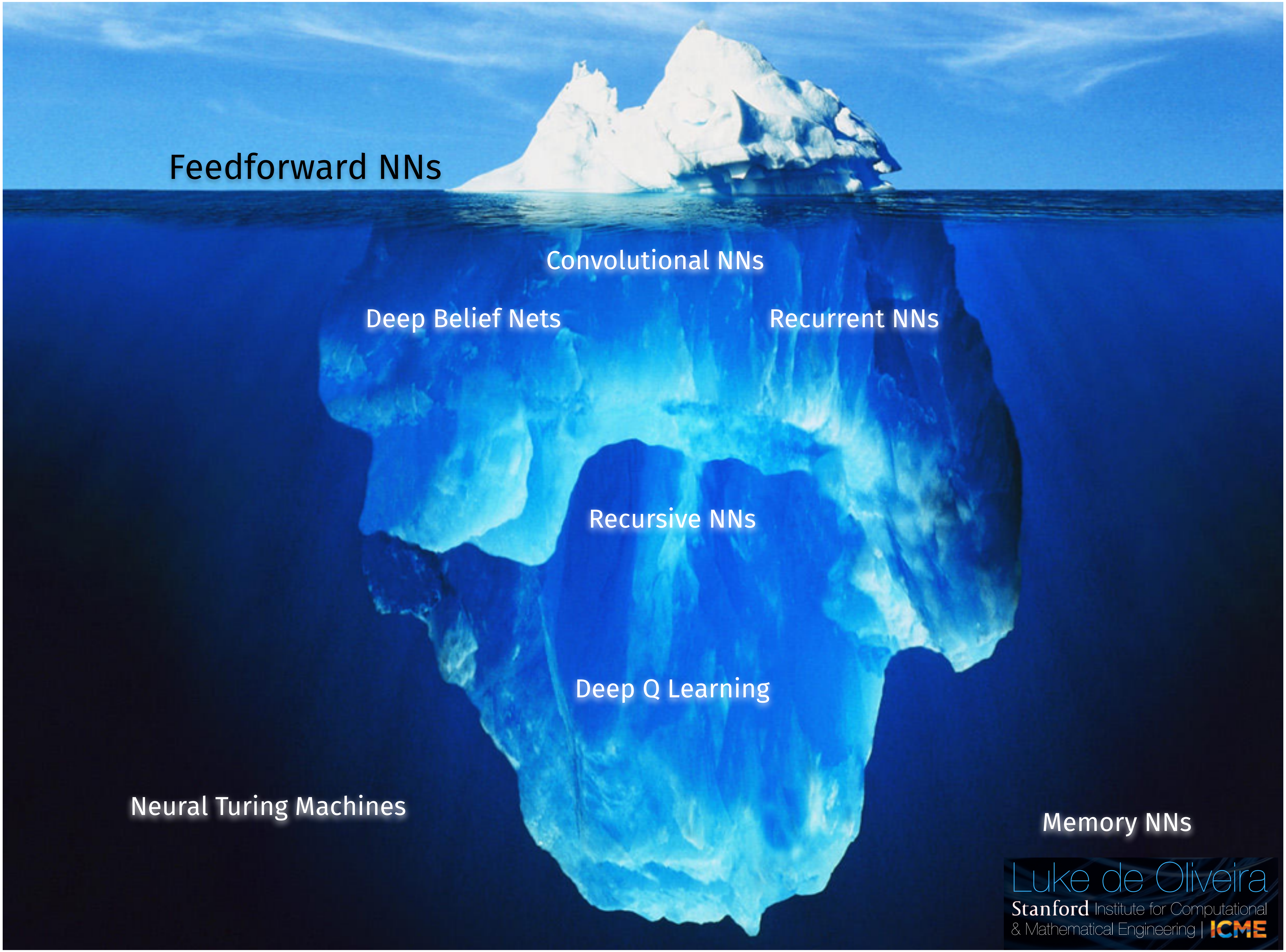
Deep Learning

- What is it?
  - Many layer Neural Networks with large number of parameters.
- Why now? Difficulty training such big networks in the past… now:
  - Solutions to difficulties in training (vanishing gradient problem)
    - Better activation. Longer training with bigger Data sets. Unsupervised Learning.
  - Big Data provides the necessary large datasets for training
  - GPUs

Recent History

- Deep Learning feats that sparked broad interest:
    - 2012, Google 1B DNN learns to identify cats (and 20000 other types of objects) (Wired Article, paper)
        - *Raw input*: trained with 200x200 pixel images from YouTube
        - *Unsupervised*: the pictures were unlabeled.
        - Google cluster 16000 cores ~ $1M. Redone with $20k system with GPUs.
    - 2013: Deep Mind builds AI that plays ATARI (Blogpost, Nature,YouTube,YouTube)

Examples



Feedforward NNs

Convolutional NNs

Deep Belief Nets

Recurrent NNs

Recursive NNs

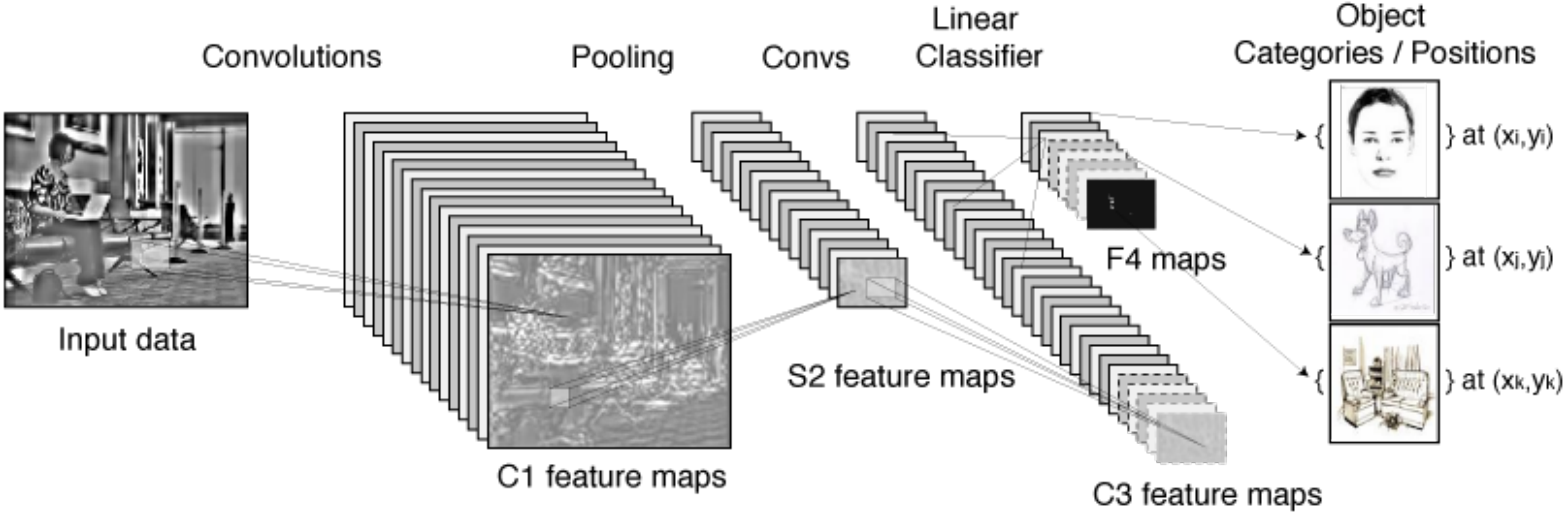
Deep Q Learning

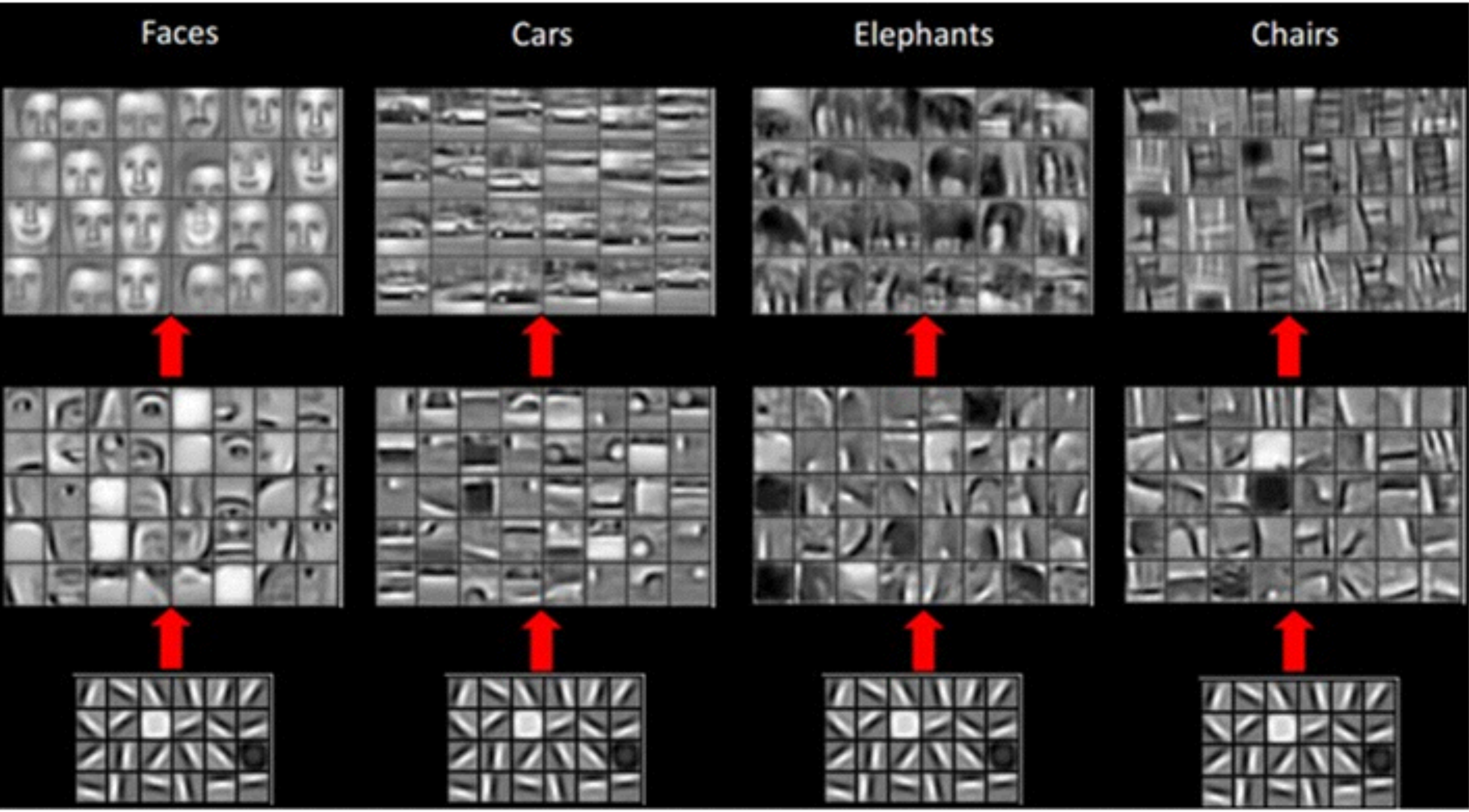
Neural Turing Machines

Memory NNs

Convolutional NN

- 1D: Time series, 2D: images, 3D: video

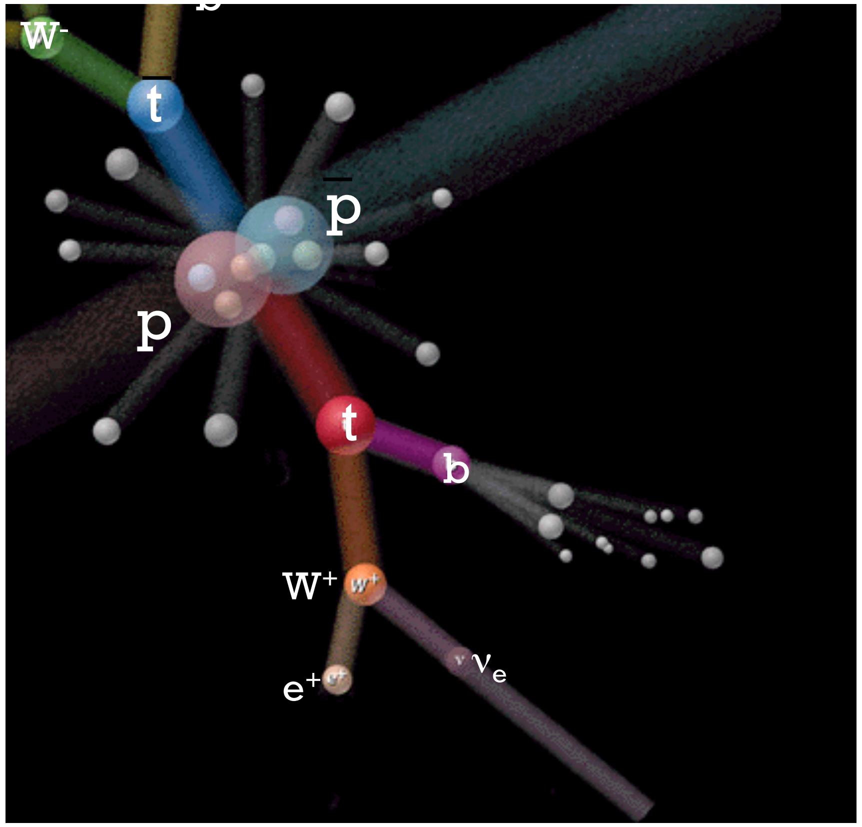



Faces
Cars
Elephants
Chairs

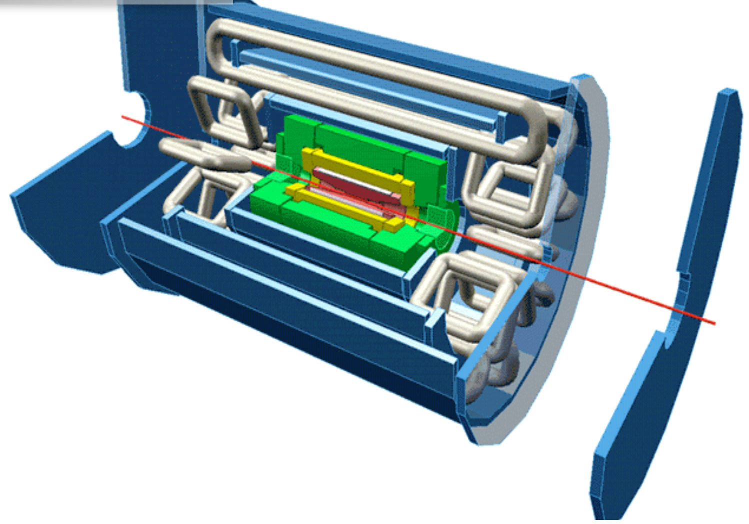
Context

HEP Experiments

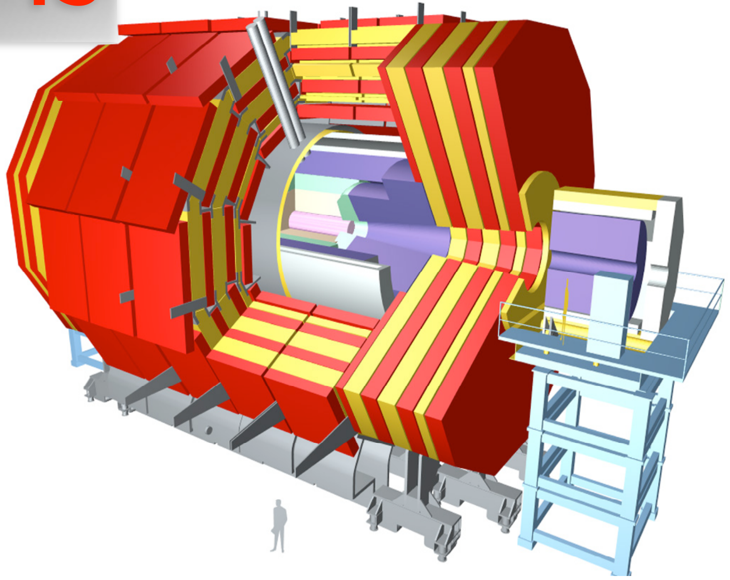
- 2 parts to HEP experiment:
  - *source*: e.g. LHC collisions creating quickly decaying heavy particles
  - *detector*: a big camera
    - pictures of long-lived decay products of short lived heavy/interesting particles.
  - Detectors parts: Tracking, Calorimeters, Muon system, Particle ID (e.g. Cherenkov, Time of Flight)

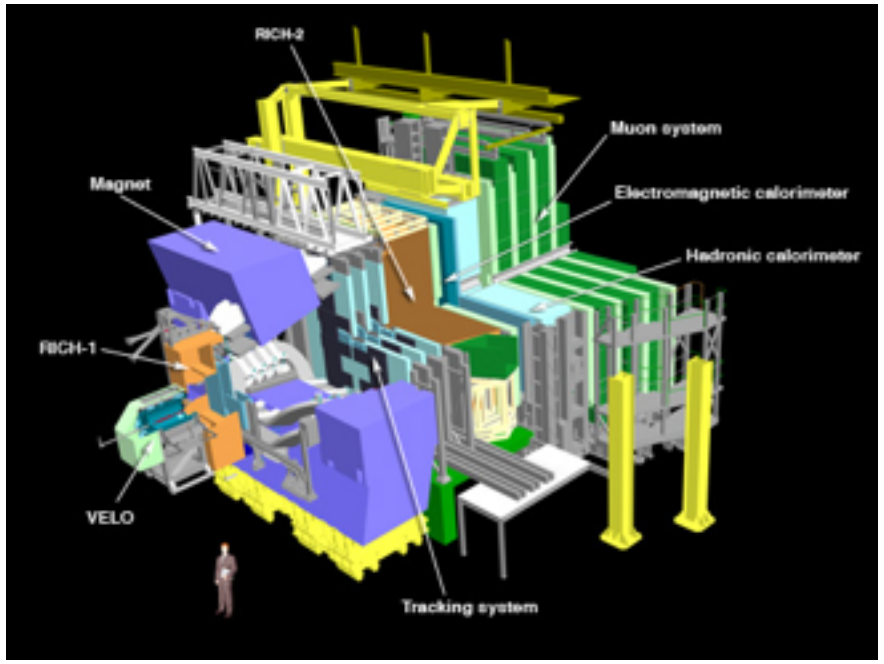


ATLAS



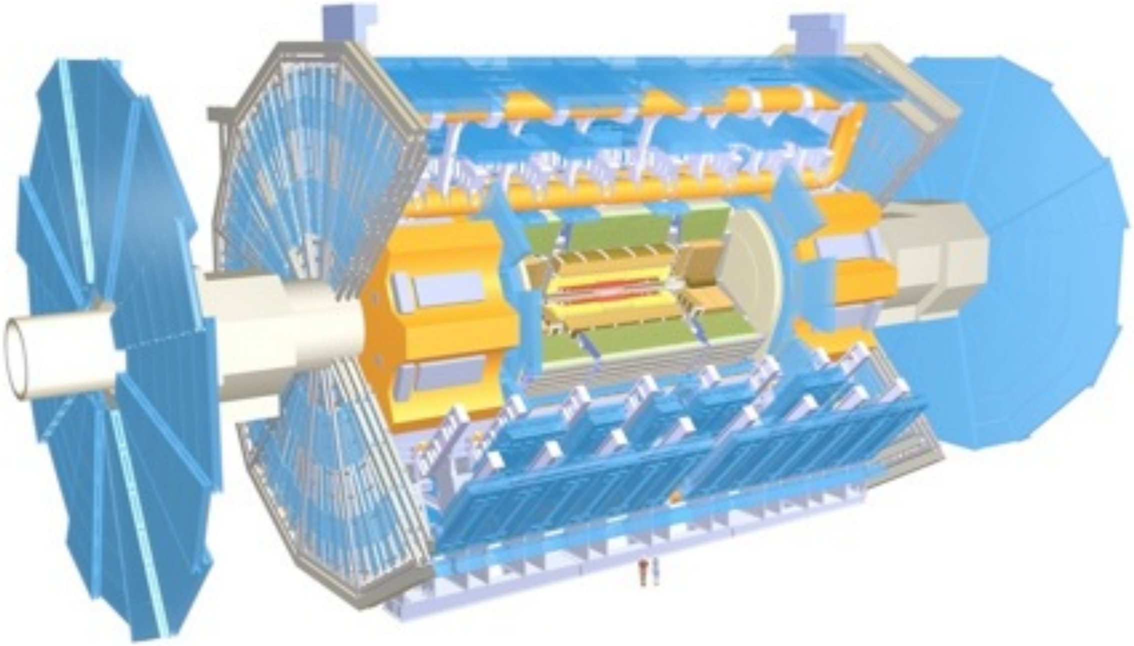
CMS





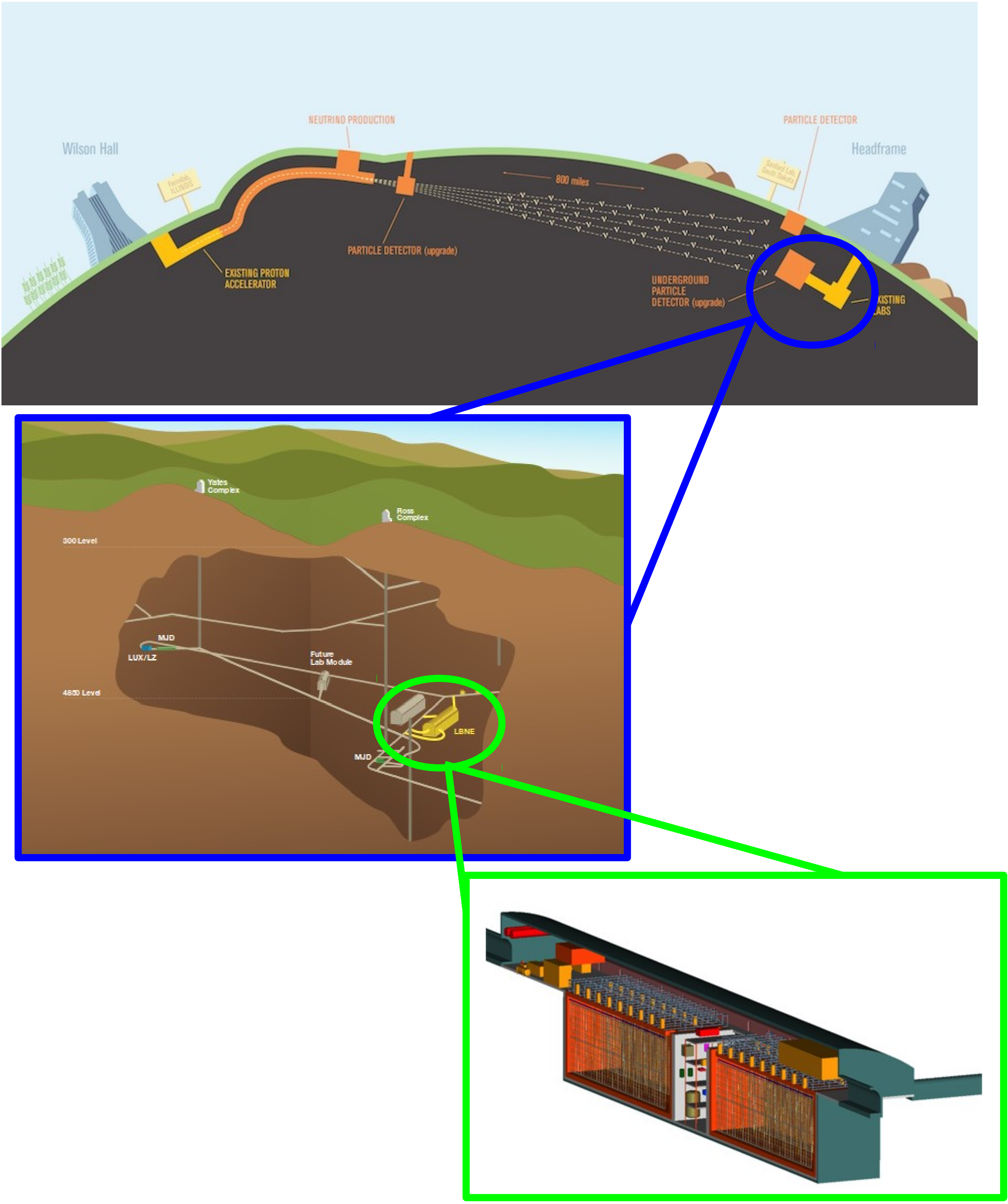
Europe

- **Europe:** *LHC at Energy Frontier*: World's most energetic proton-proton machine.
  - Found the Higgs in Run 1…
  - Next goals:
    - Test naturalness (Was the Universe and accident?) by searching for New Physics like Supersymmetry.
    - Find Dark Matter (reasons to think related to 1)
    - Study the SM Higgs find new Higgses
  - Run 2 at higher energy now.
  - Run 3 at higher luminosity by end of decade.
  - High Luminosity- LHC by 2025.
  - 100 TeV Machine later in the century? (In China?)



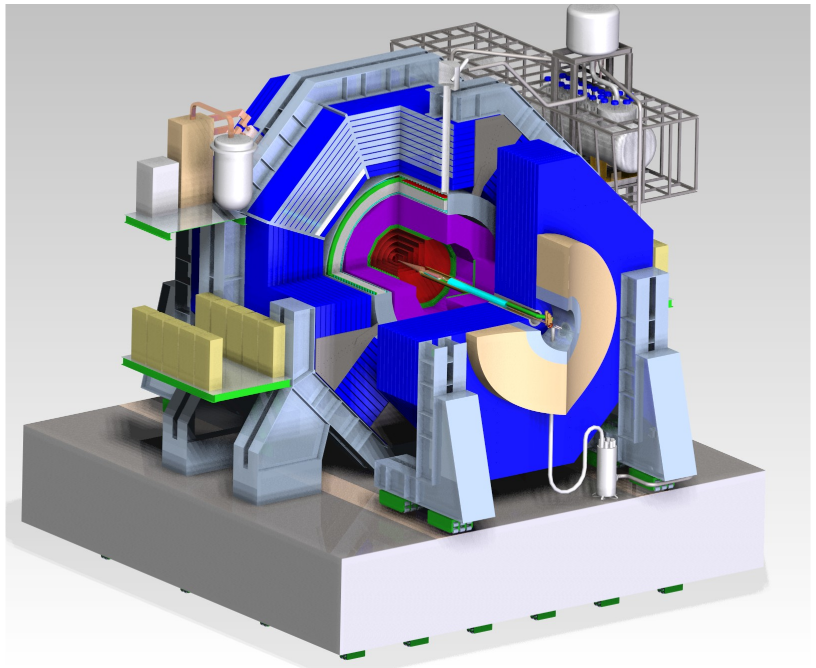
US

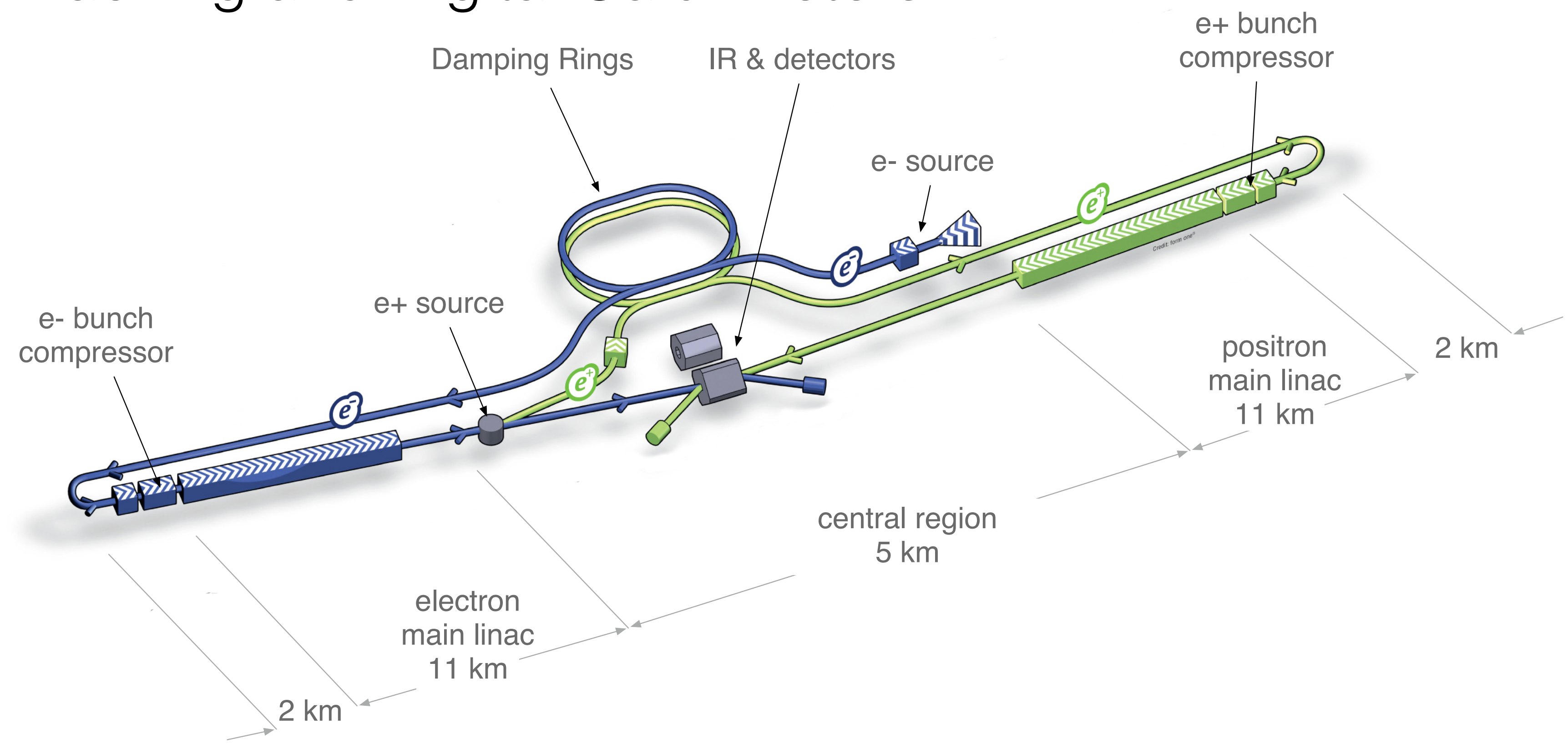
- **US:** *Long Baseline Neutrino Facility (LBNF)/Deep Underground Neutrino Experiment (DUNE) at Intensity Frontier*
  - Shoot intense neutrino beam through earth at a Near and Far (1300km) detector.
  - Physics Goals:
    - Study Neutrinos, especially Charge Matter Violation (Why is there Matter in the Universe?)
    - Supernova
    - Proton Decay
    - Dark Matter
  - *Liquid Argon Time Projection Chambers* (LArTPC) detector technology.
  - Short Base Line program and LArTPC R&D until ~2020. (Many experiments ~ 100 Ton)
  - Beam to 10 kiloton DUNE in 2025…
  - Gradually expand to 40 kilotons and run for 30 years.

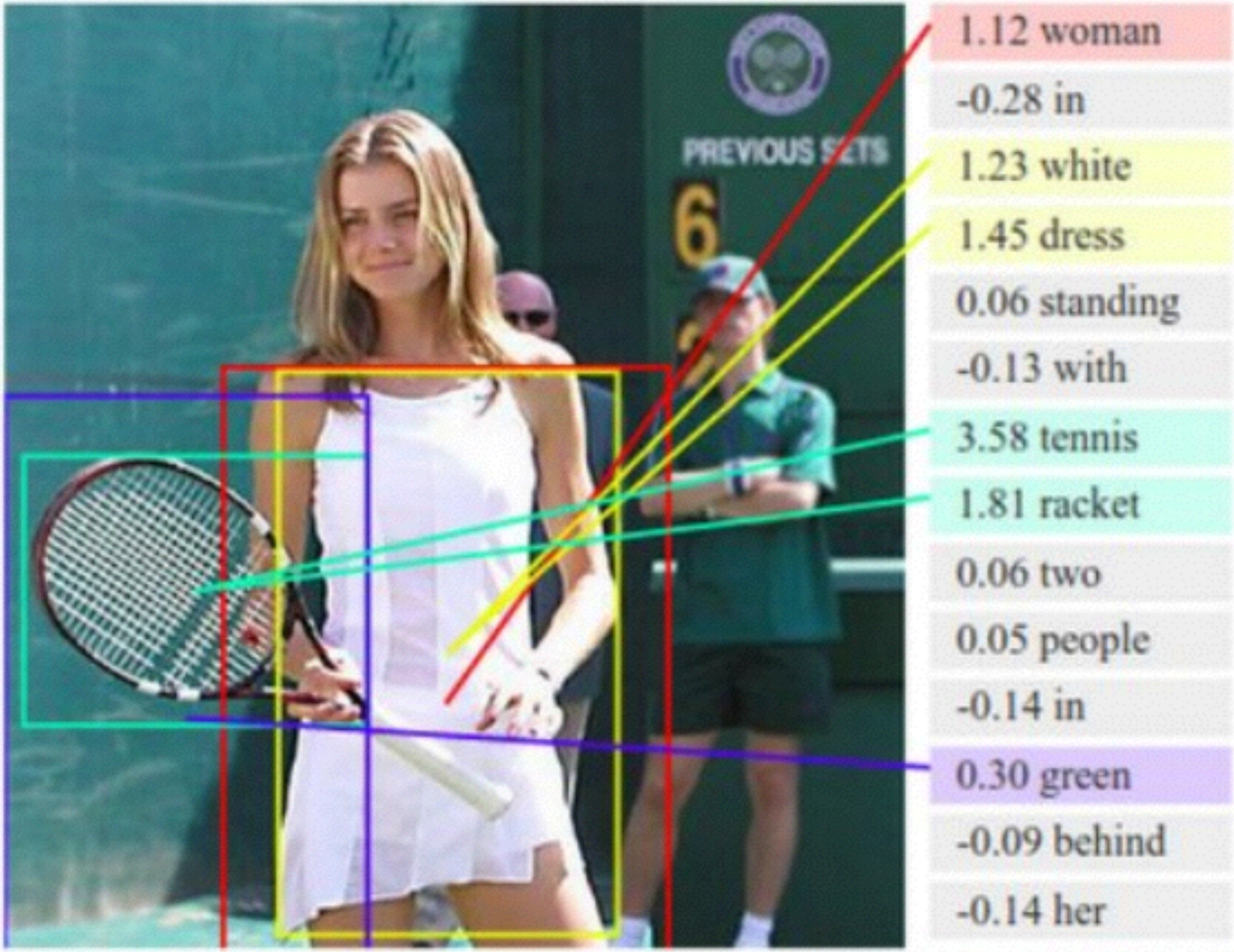


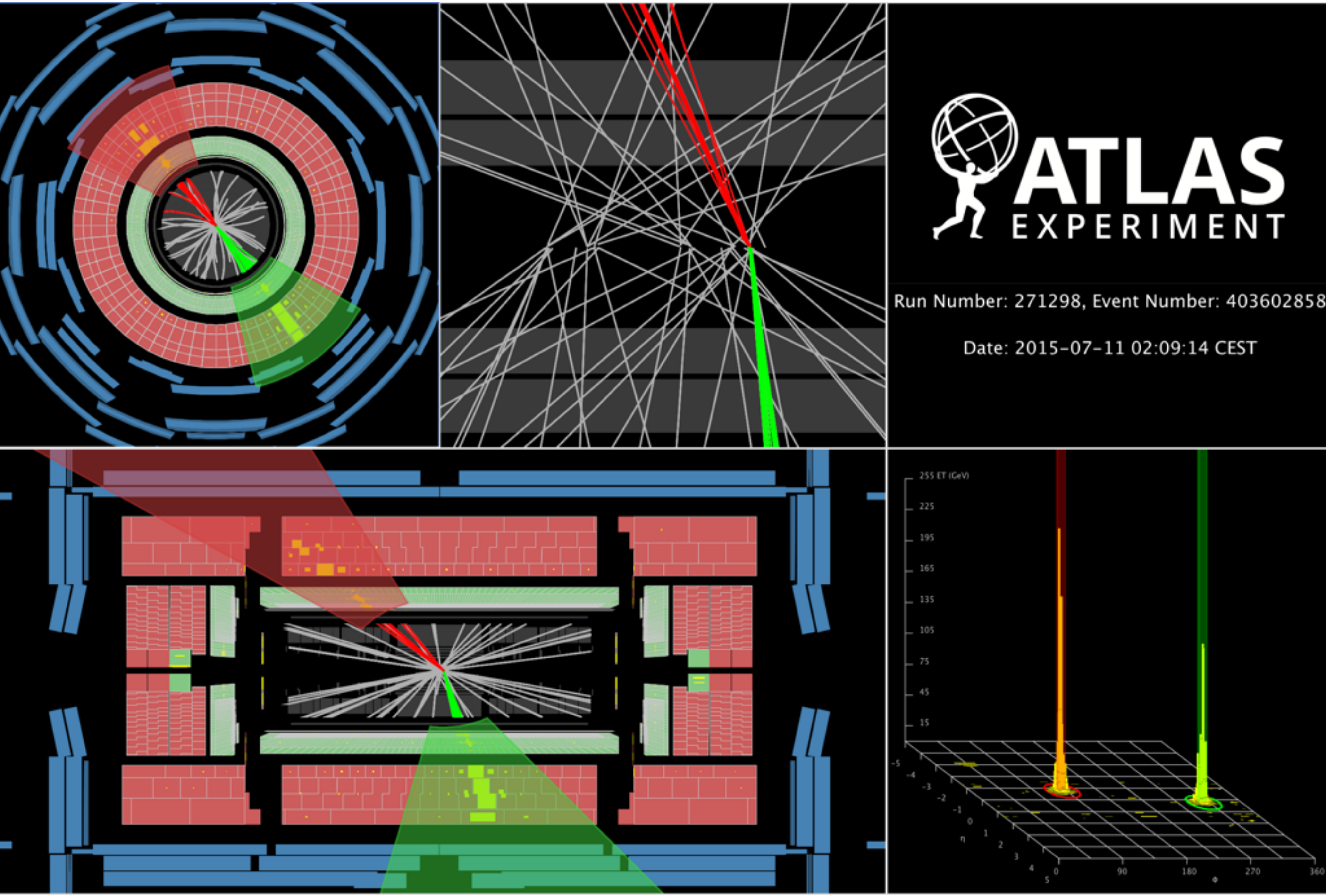
Japan

- **Europe:** *LHC at Energy Frontier*
- **US:** *LBNF/DUNE at Intensity Frontier*
- **Japan:** *International Linear Collider (ILC):* Most energetic e^+e^- machine.
  - Japanese will hopefully build this in 2020s.
  - Precision studies of Higgs and hopefully new particles found at LHC.
  - High granularity Silicon Tracking and Digital Calorimeters.






PREVIOUS SETS
6
1.12 woman
-0.28 in
1.23 white
1.45 dress
0.06 standing
-0.13 with
3.58 tennis
1.81 racket
0.06 two
0.05 people
-0.14 in
0.30 green
-0.09 behind
-0.14 her



Run Number: 271298, Event Number: 403602858

Date: 2015-07-11 02:09:14 CEST

Why go Deep?

- ***Better*** Algorithms
  - Hopefully DNN-based classification/regression *out performs* hand crafted algorithms.
    - For LArTPC, it may be able to do something we cannot do well algorithmically.
  - *Unsupervised learning*: DNNs classify without being told what are the classes.
    - The hope is that DNNs could make sense of complicated data that we don't understand or expect (e.g. anomaly detection).
- ***Faster*** Algorithms
  - After training, DNN inference is sometimes *faster* than algorithmic approach. e.g. Playing go.
  - Already *parallelized* and optimized for GPUs/HPCs. First broadly applicable and low threshold use of GPUs.
    - Industry building highly optimized software, chips, systems (HPCs), and cloud services.
  - DNN can *encapsulate expensive computations*, e.g. Matrix Element Method or simulation.
- ***Easier*** Algorithm Development: *Feature Learning* instead of *Feature Engineering*
  - Reduce time physicists spend writing developing algorithms that process raw data into the inputs features (e.g. Reconstruction) to traditional analysis or Machine Learning.
  - Save on development time and costs.

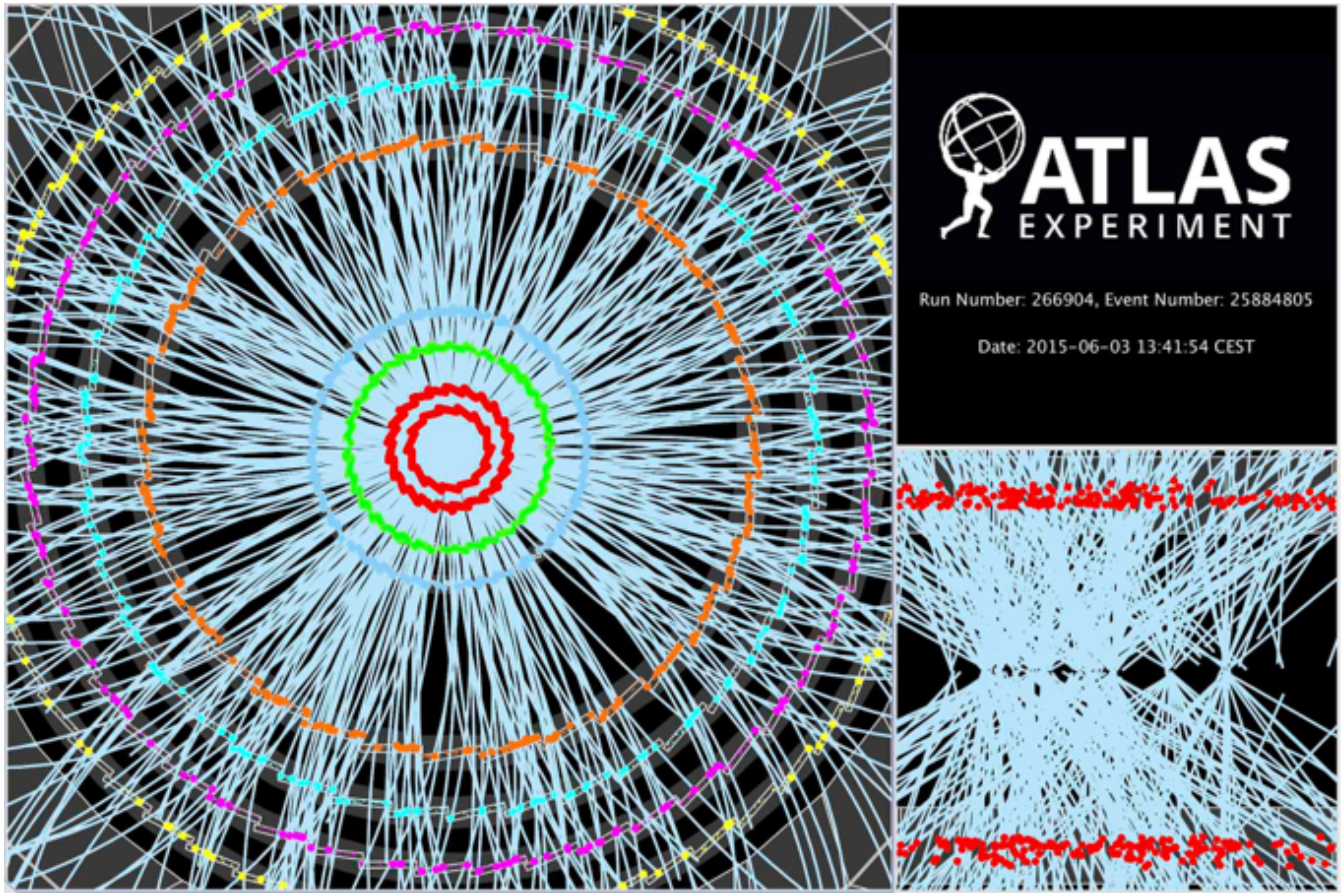
Moore’s law?

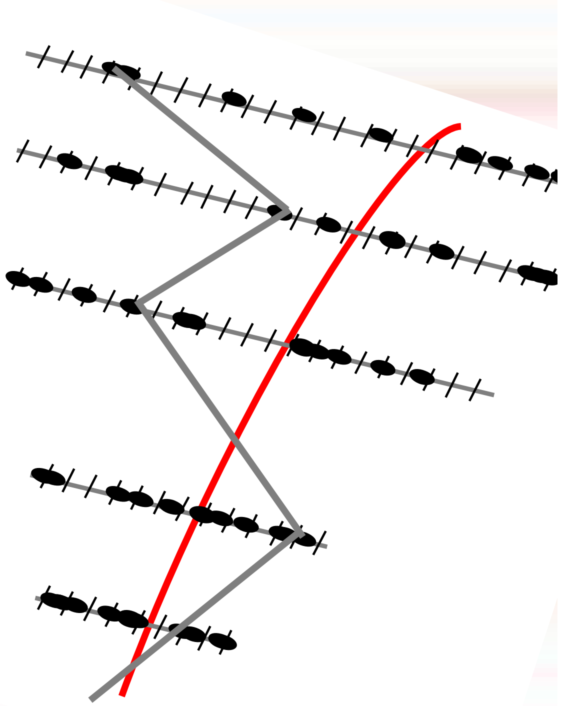
- For the first time, the cost of adding more transistors/silicon area has increased recently.
- HL-LHC computing requirements will outpace Moore’s Law.
    - We cannot assume that we will easily get 10x the computing power for same price in 10 years.
    - First estimates of cost of HL-LHC computing is several times LHC, even assuming Moore Law.
- Solutions:
    - Quantum computers are no good for us…
    - Highly parallel processors (e.g. GPUs) are already > 10x CPUs for certain computations.
        - Unfortunately parallelization (i.e. Multi-core/GPU) has been difficult.
    - Trend is away from x86 towards custom hardware (e.g. GPUs, Mics, FPGAs, Custom DL Chips)
    - Deep Learning and Neuromorphic chips are a possible solution.
        - Think of the DL “seeing” tracks in silicon detectors like how DeepMinds’s AI sees moves on the go board.
        - Neuromorphic chips are incredibly power efficient.

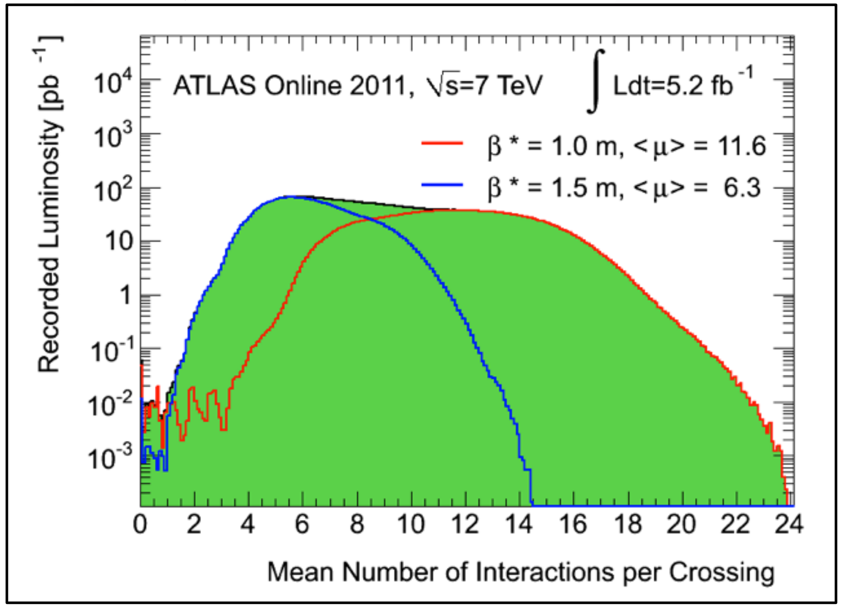
Particle Detectors

Tracking

- Measure Charged particle trajectories. If B-field, then measure momentum.

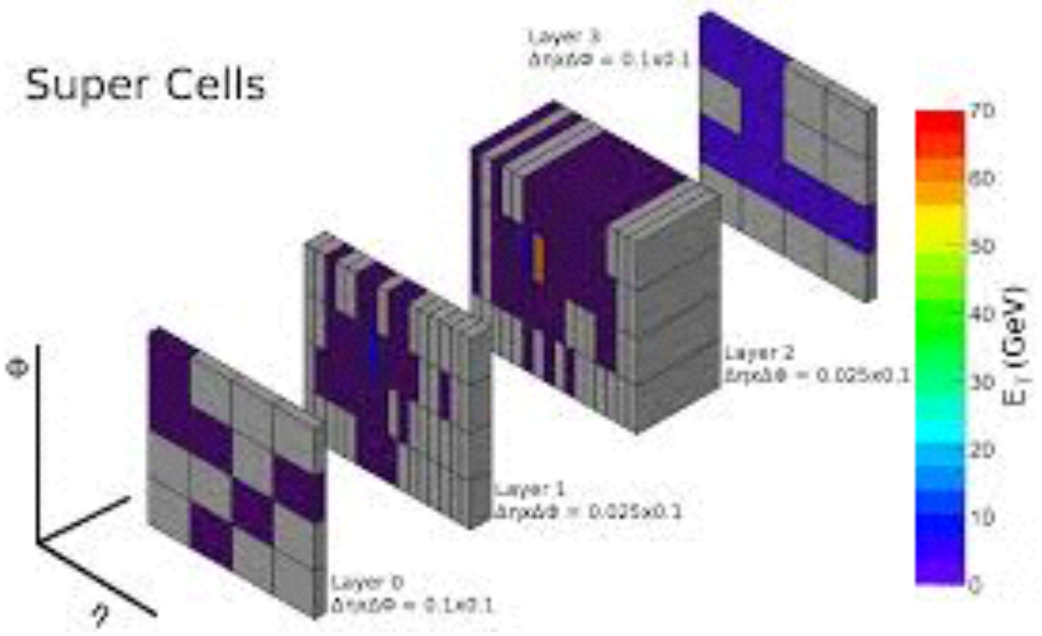


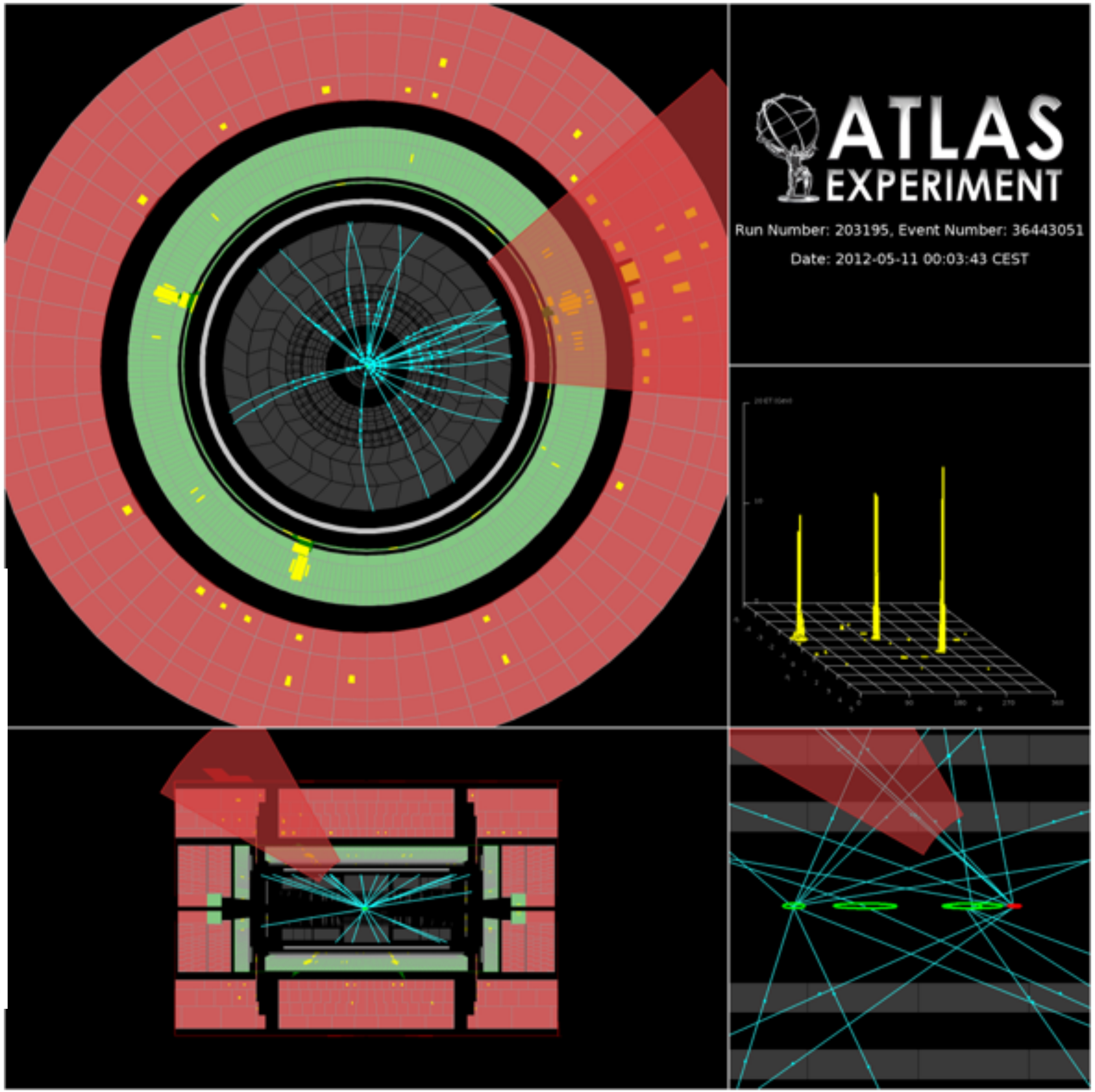




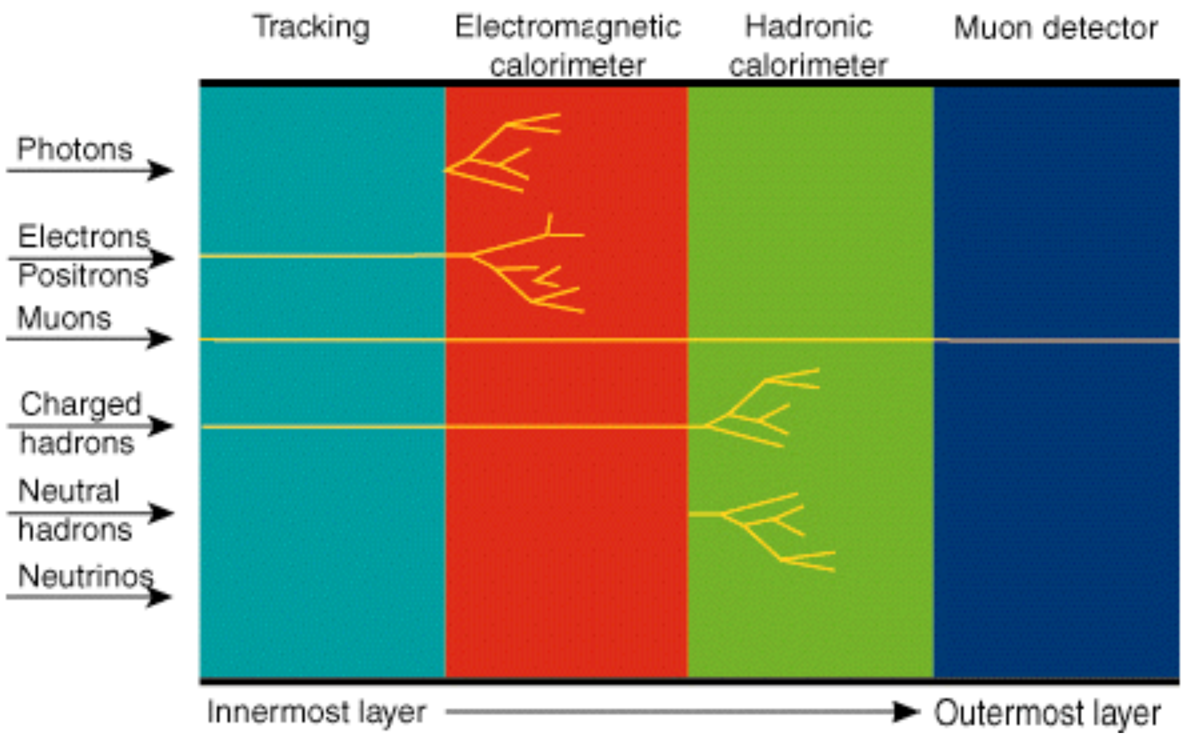
Calorimetry

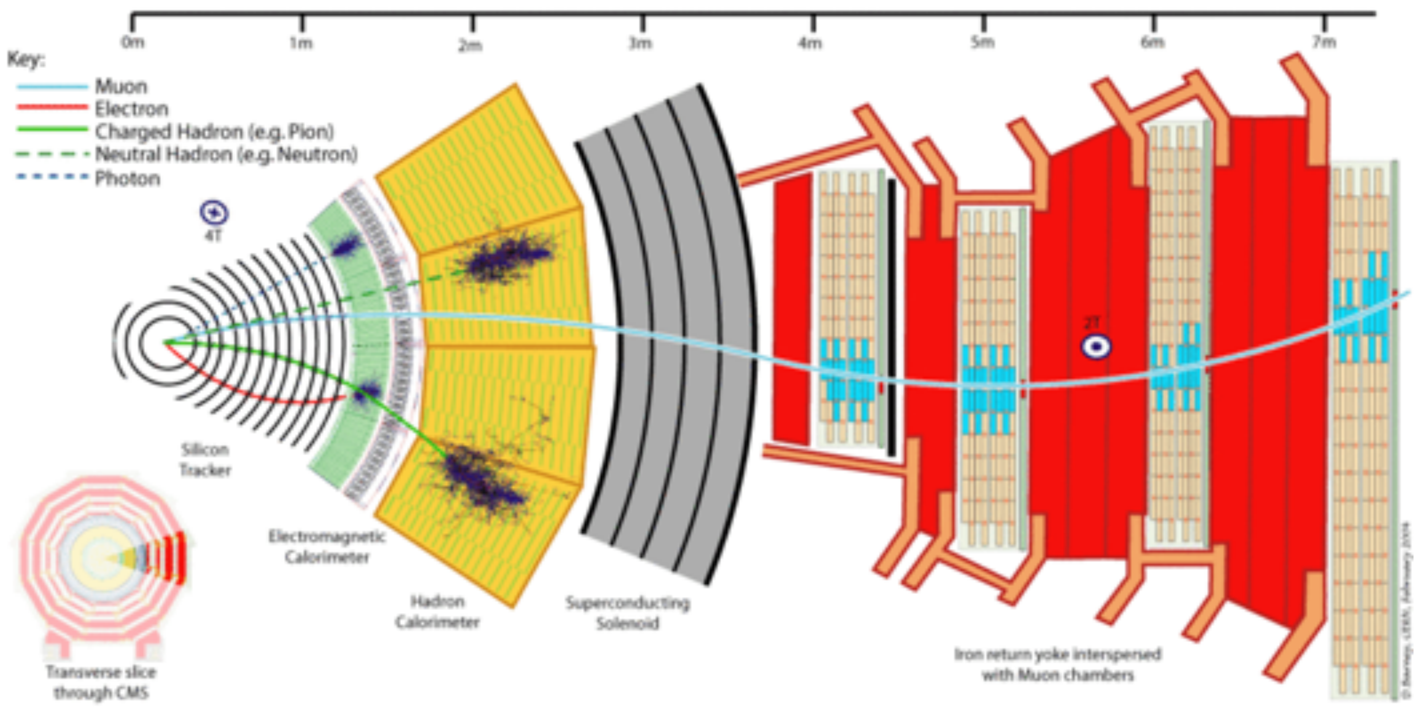
- Make particle interact and loose all energy, which we measure. 2 types:
  - Electromagnetic: e.g. crystals in CMS, Liquid Argon in ATLAS.
  - Hadronic: e.g. steel + scintillators
  - e.g ATLAS:
    - 200K Calorimeter cells measure energy deposits.
    - 64 x 36 x 7 3D Image

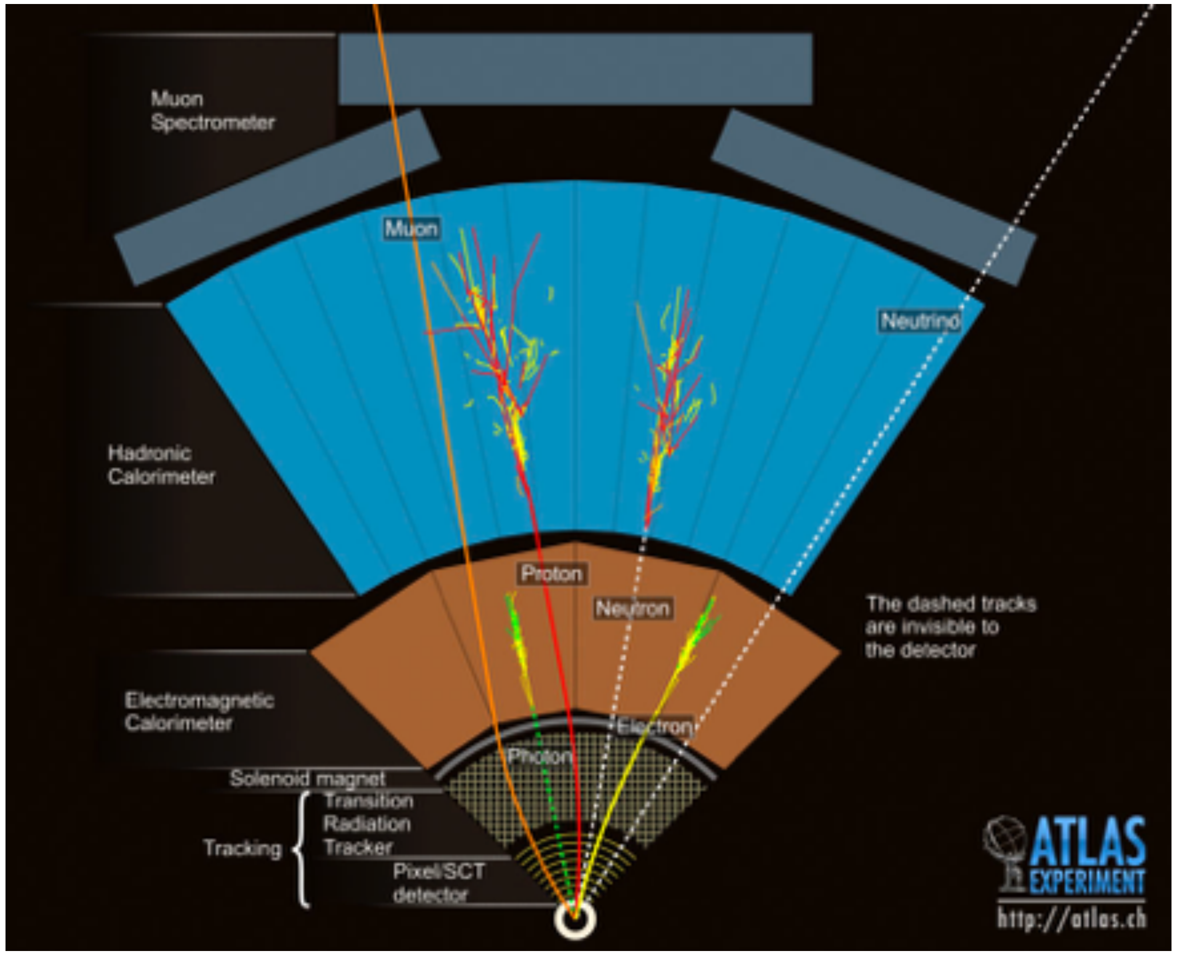




LHC detectors

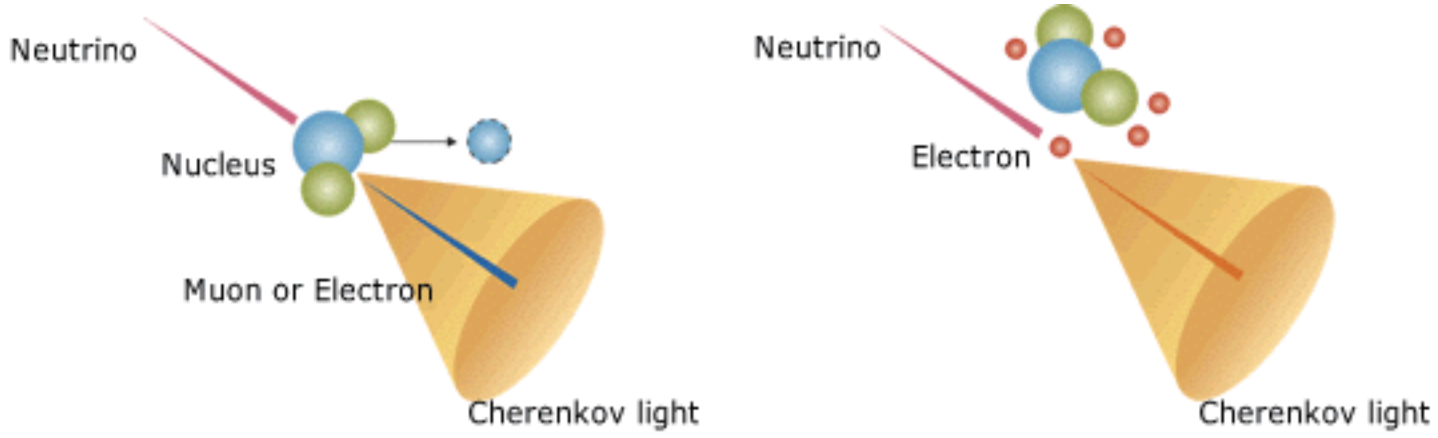




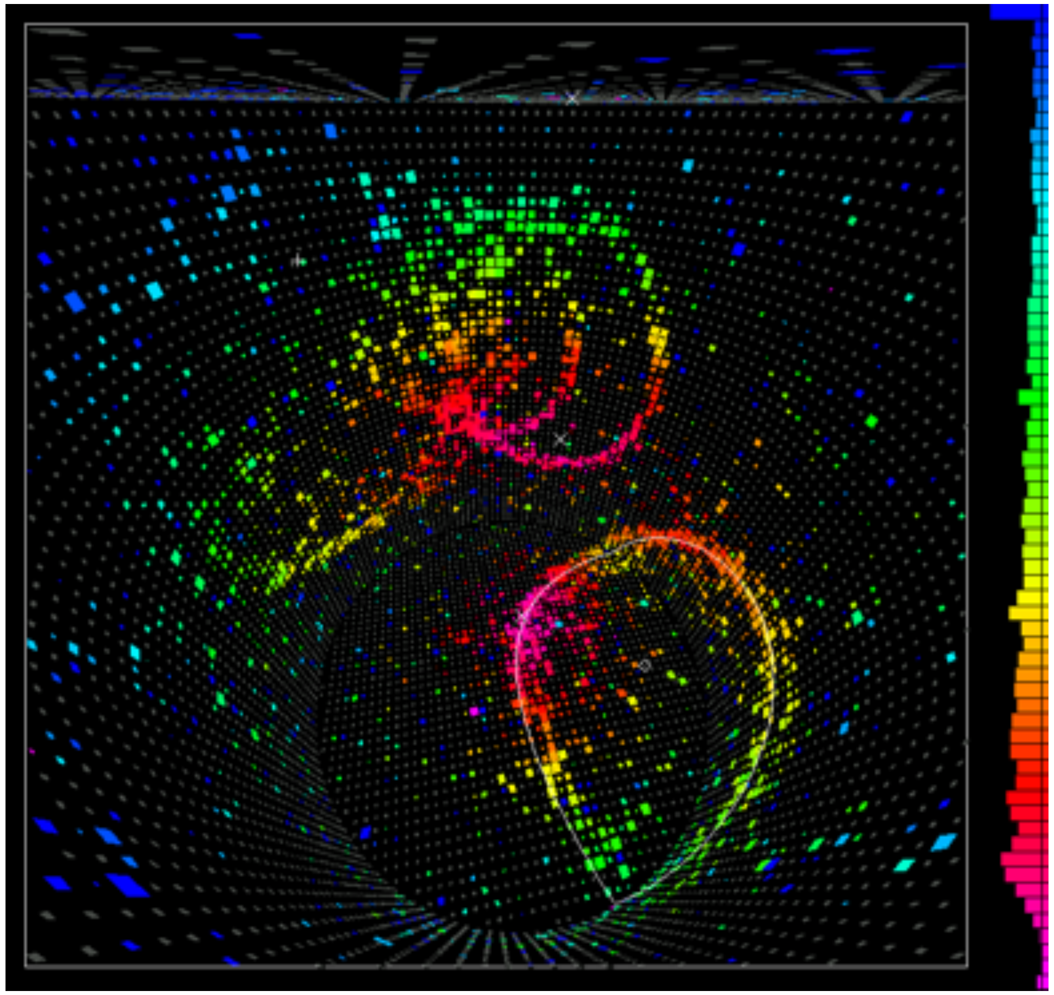


How do we “see” particles?

- Charged Particles traveling faster than speed of light in medium emit Cherenkov light (analogous to sonic boom).
  - Light emitted in cone, with angle function of speed and mass.
  - Depending on context, allow for particle identification and/or speed measurement.



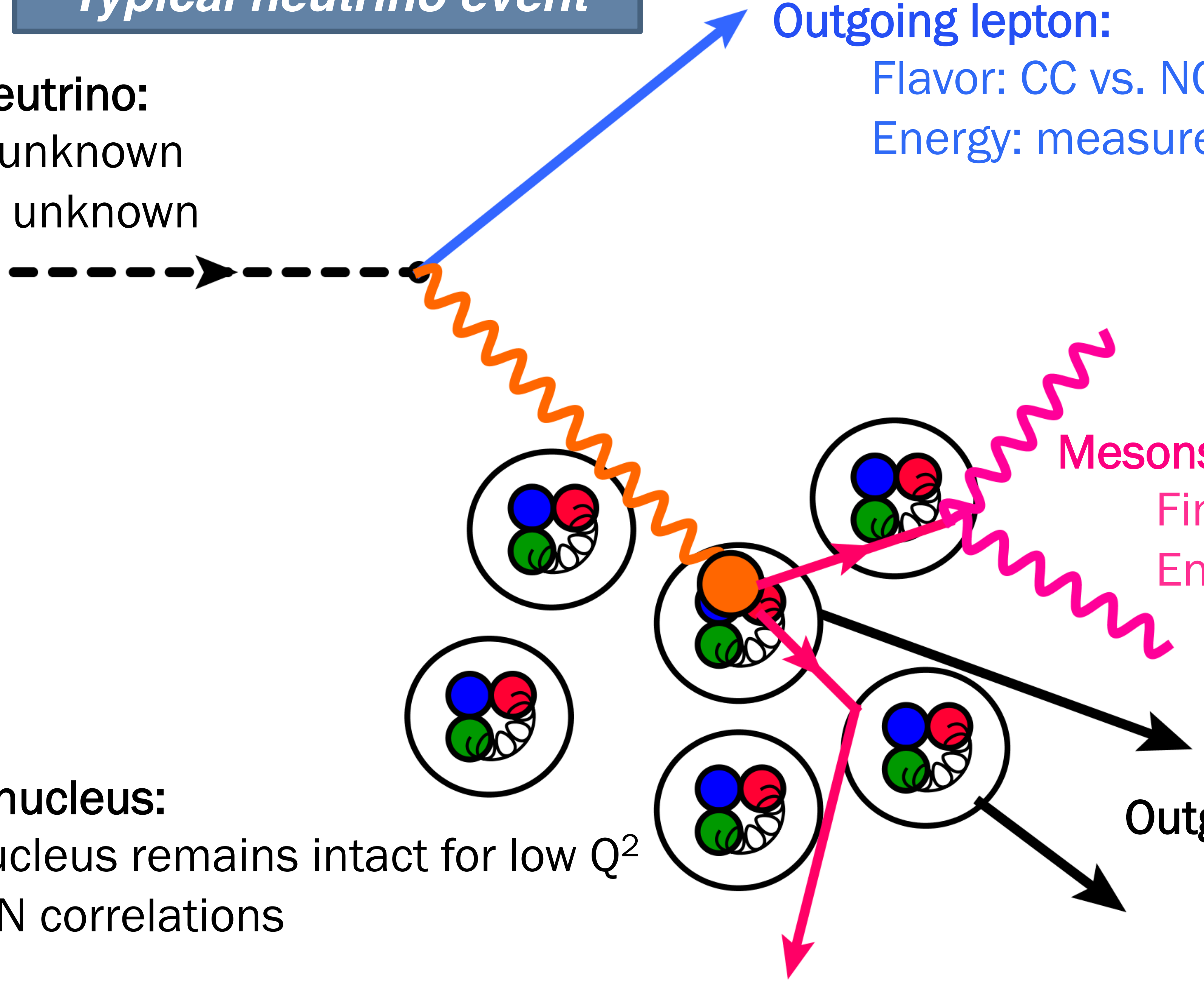
The generated charged particle emits the Cherenkov light.



Neutrino Detection

In neutrino experiments, try to determine flavor and estimate energy of incoming neutrino by looking at outgoing products of the interaction.

Typical neutrino event



Outgoing lepton:

Flavor: CC vs. NC, μ^+ vs. μ^-, e vs. γ

Energy: measure

Incoming neutrino:

Flavor unknown

Energy unknown

Mesons:

Final State Interactions

Energy? Identity?

Target nucleus:

Nucleus remains intact for low Q^2

N-N correlations

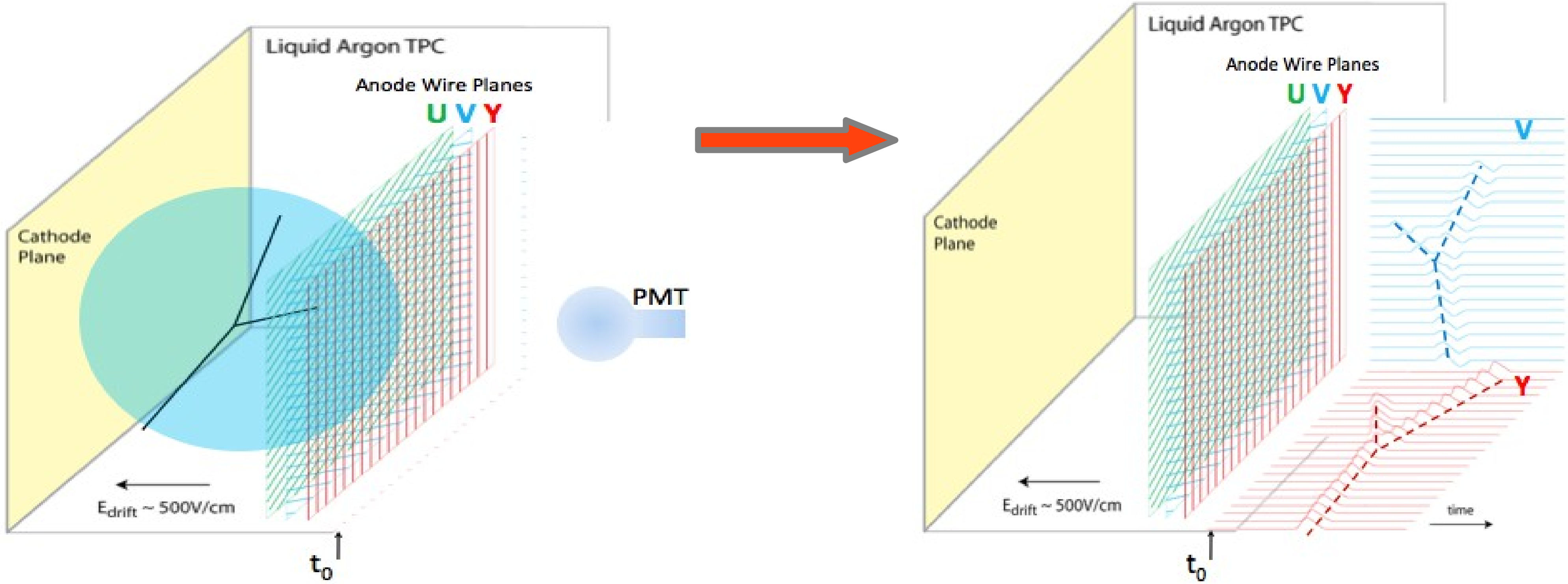
Outgoing nucleons:

Visible? Energy?

Jen Raaf

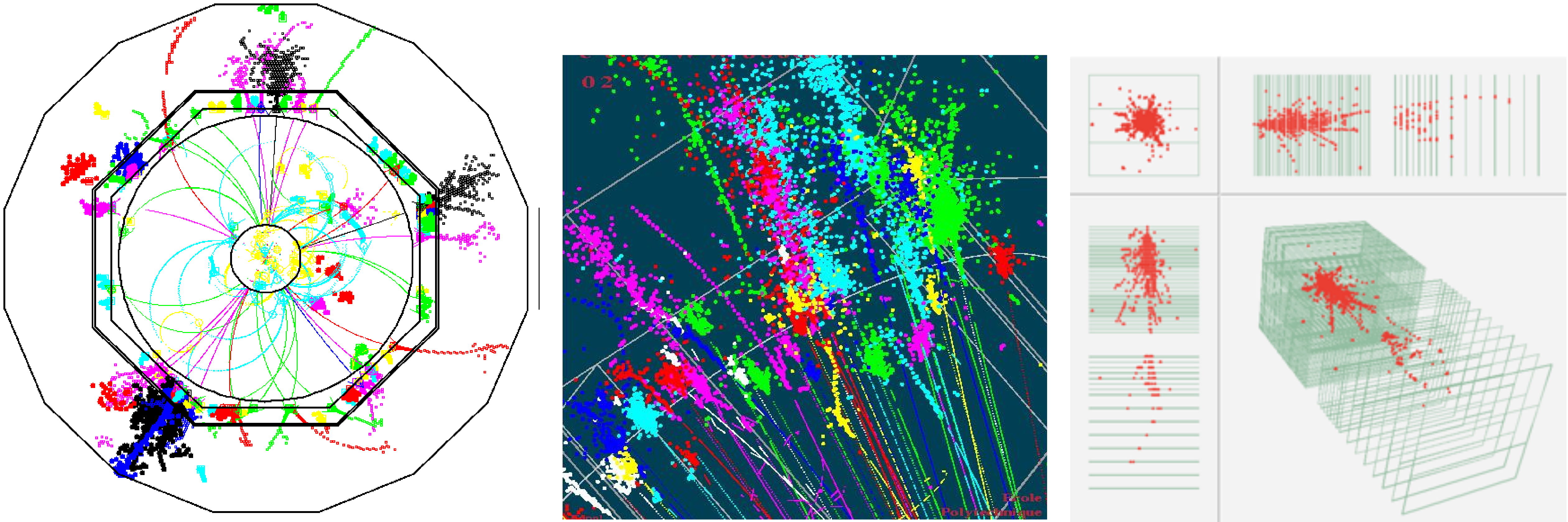
Neutrino Detectors

- Need large mass/volume to maximize chance of neutrino interaction.
- Technologies:
    - Water/Oil Cherenkov
    - Segmented Scintillators
    - Liquid Argon Time Projection Chamber: promises ~ 2x detection efficiency.
        - Provides tracking, calorimetry, and ID all in same detector.
        - Usually 2D read-out… 3D inferred.
    - Gas TPC: full 3D



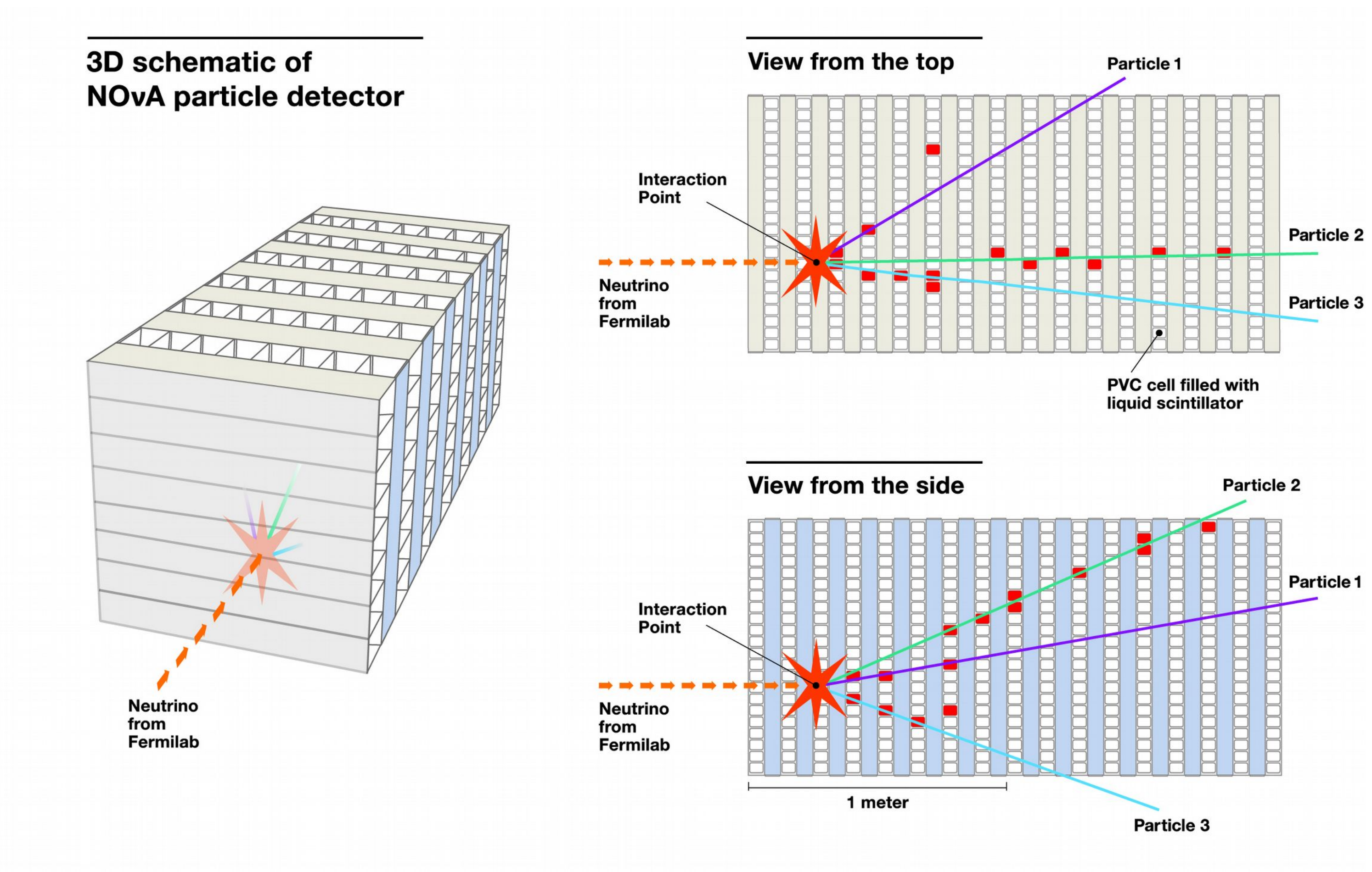
ILC Detectors

- Precision measurements require excellent calorimetry
  - Aim for jet energy resolution giving di-jet mass resolution similar to Gauge boson widths
  - Various concepts ~ digital/high granularity calorimetry + particle flow.
  - Similarities to upgrade LHC forward detectors



Examples

Nova

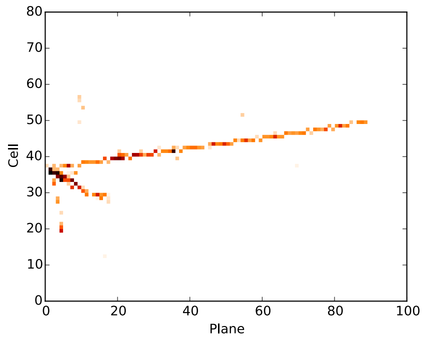


	CVN Selection Value	ν_e sig	Tot bkg	NC	ν_μ CC	Beam ν_e	Signal Efficiency	Purity
Contained Events	–	88.4	509.0	344.8	132.1	32.1	–	14.8%
$s/\sqrt{b}$ opt	0.94	43.4	6.7	2.1	0.4	4.3	49.1%	86.6%
$s/\sqrt{s+b}$ opt	0.72	58.8	18.6	10.3	2.1	6.1	66.4%	76.0%

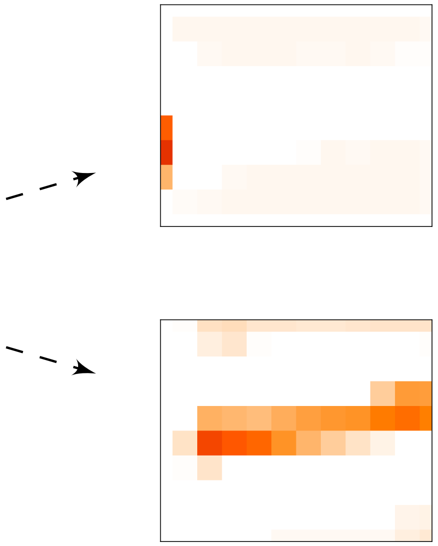
	CVN Selection Value	ν_μ sig	Tot bkg	NC	Appeared ν_e	Beam ν_e	Signal Efficiency	Purity
Contained Events	–	355.5	1269.8	1099.7	135.7	34.4	–	21.9%
$s/\sqrt{b}$ opt	0.99	61.8	0.1	0.1	0.0	0.0	17.4%	99.9%
$s/\sqrt{s+b}$ opt	0.45	206.8	7.6	6.8	0.7	0.1	58.2%	96.4%

- 40% Better Electron Efficiency for same background.

http://arxiv.org/pdf/1604.01444.pdf

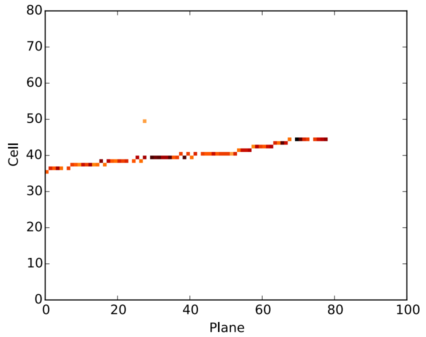


Muon Neutrino DIS CC

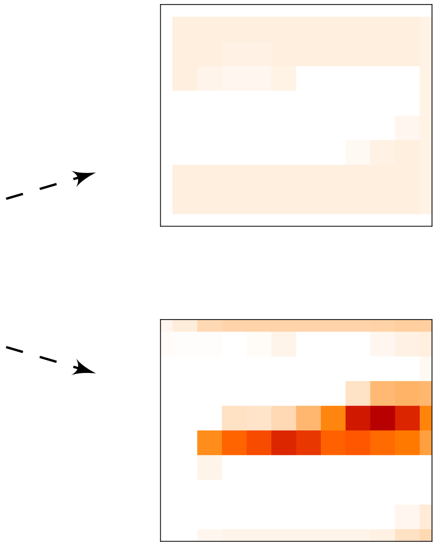


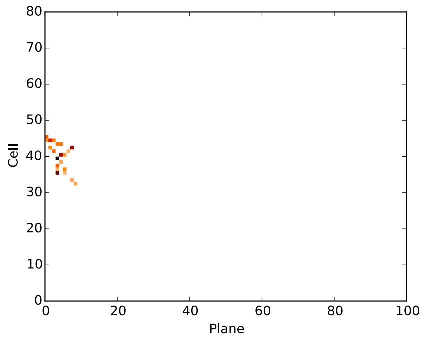
Hadronic Feature Map

Muon Feature Map

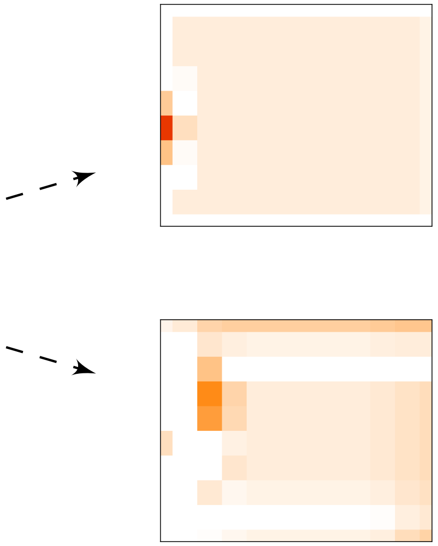


Muon Neutrino QE CC



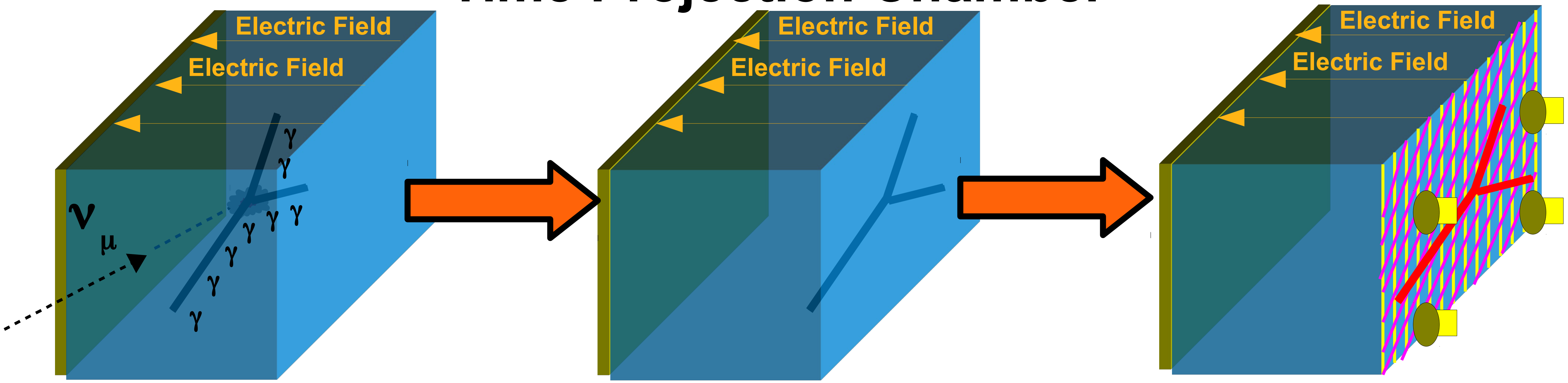


Muon Neutrino NC



LArTPC

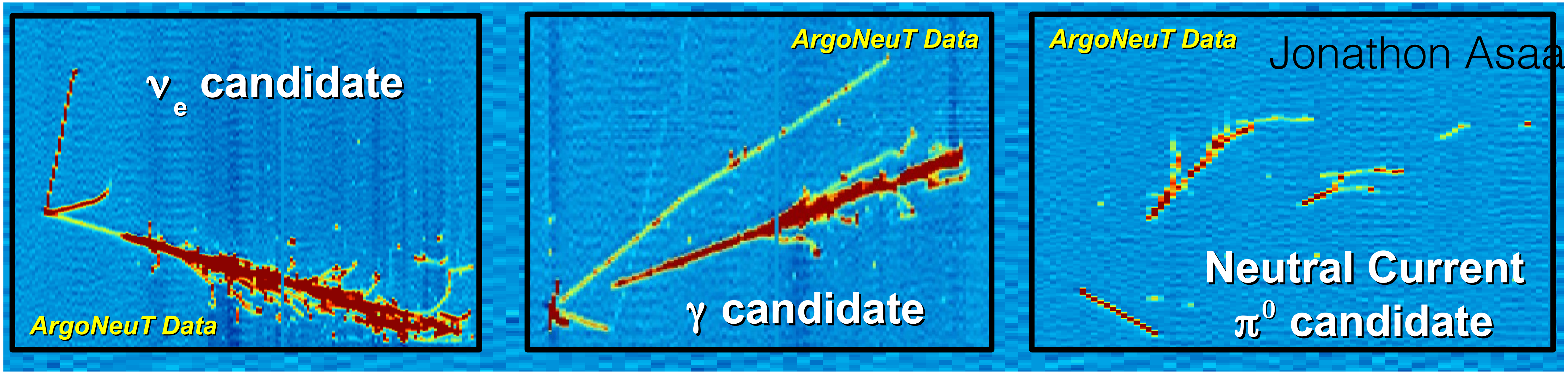
Time Projection Chamber



Neutrino interaction in LAr produces ionization and scintillation light

Drift the ionization charge in a uniform electric field

Read out charge and light produced using precision wires and PMT's

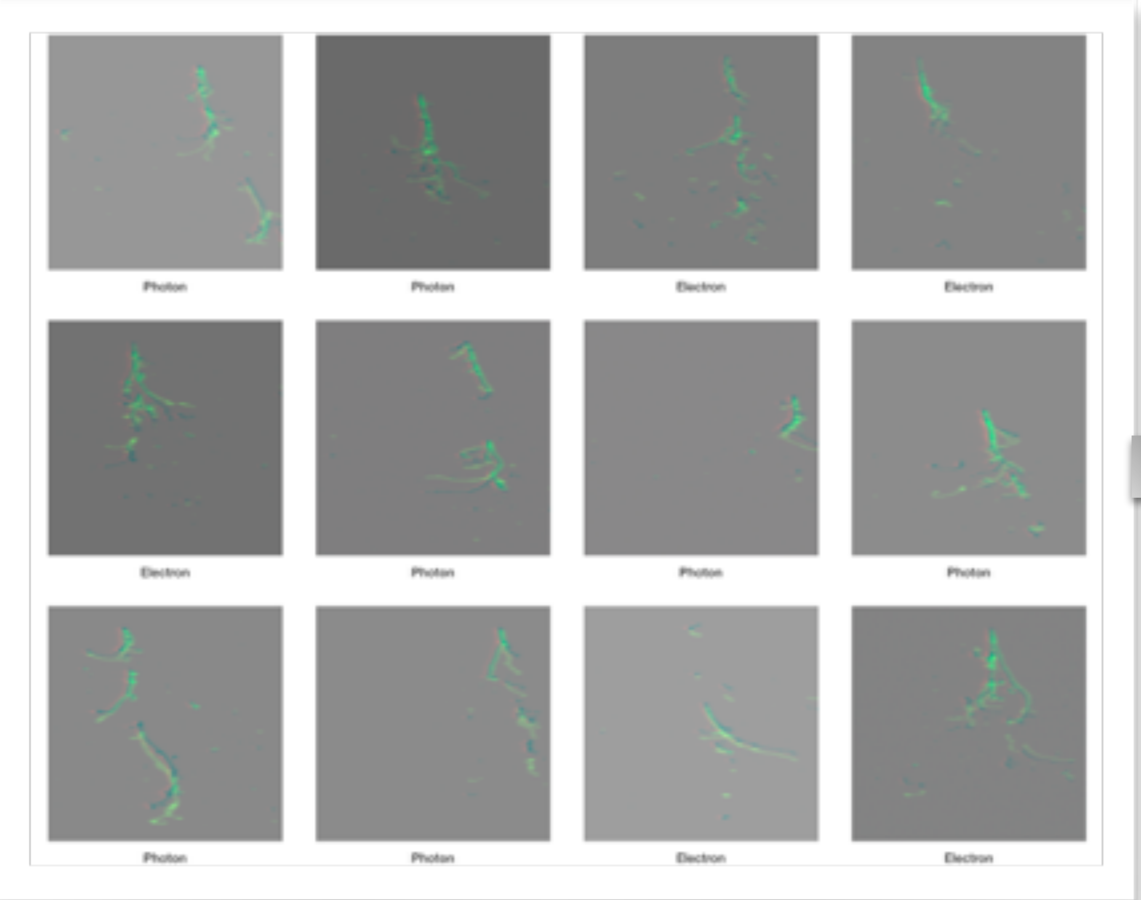


Jonathon Asaadi

Tracking, Calorimetry, and Particle ID in same detector.
Goal ~80% Neutrino Efficiency.
All you need for Physics is neutrino flavor and energy.

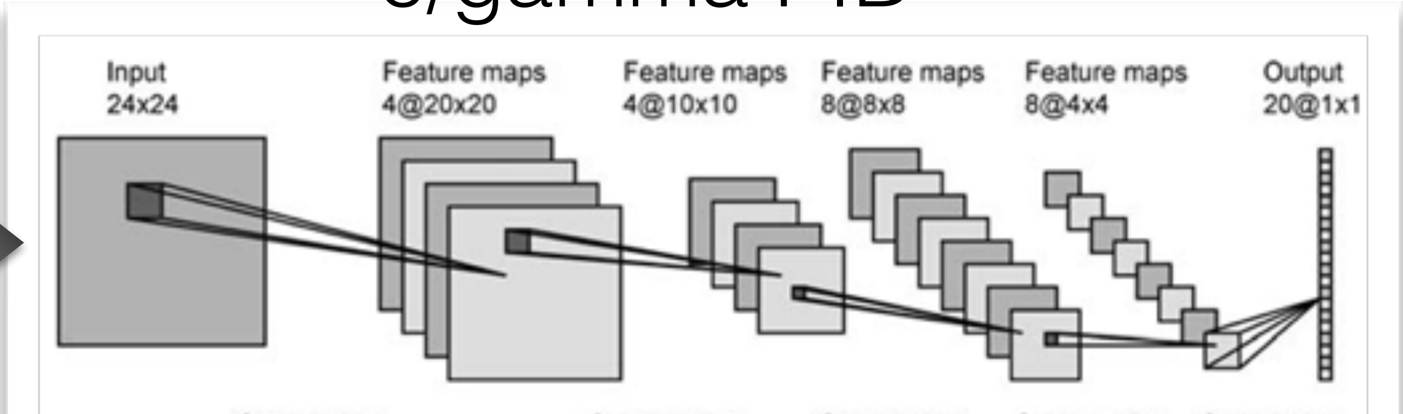
LArIAT

e/gamma PID



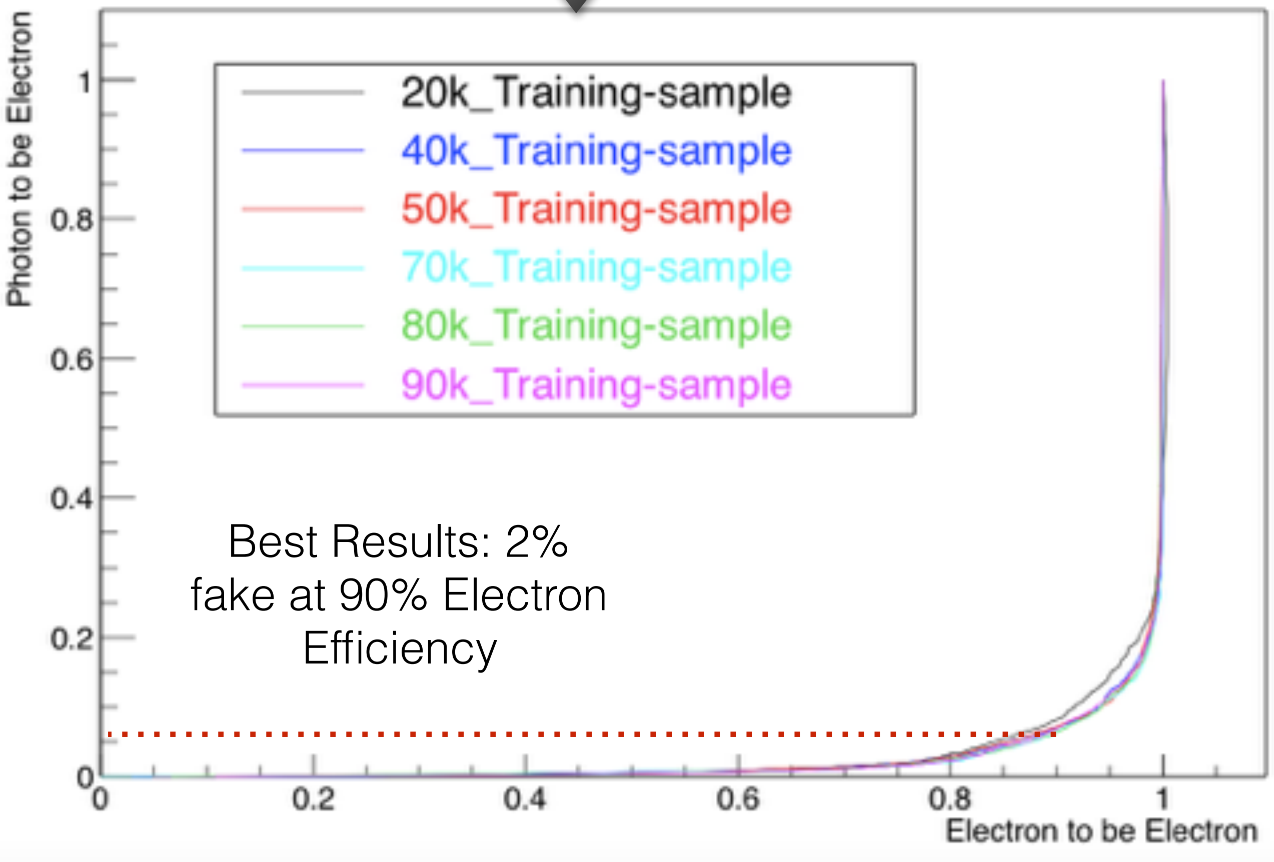
Raw Data: Wire ADC vs Time x Planes
(LArIAT Simulation)

Train



Deep Convolutional Neutral Network
(GoogLeNet)

Out of the box *Feasibility Study* with No attempt at optimization.

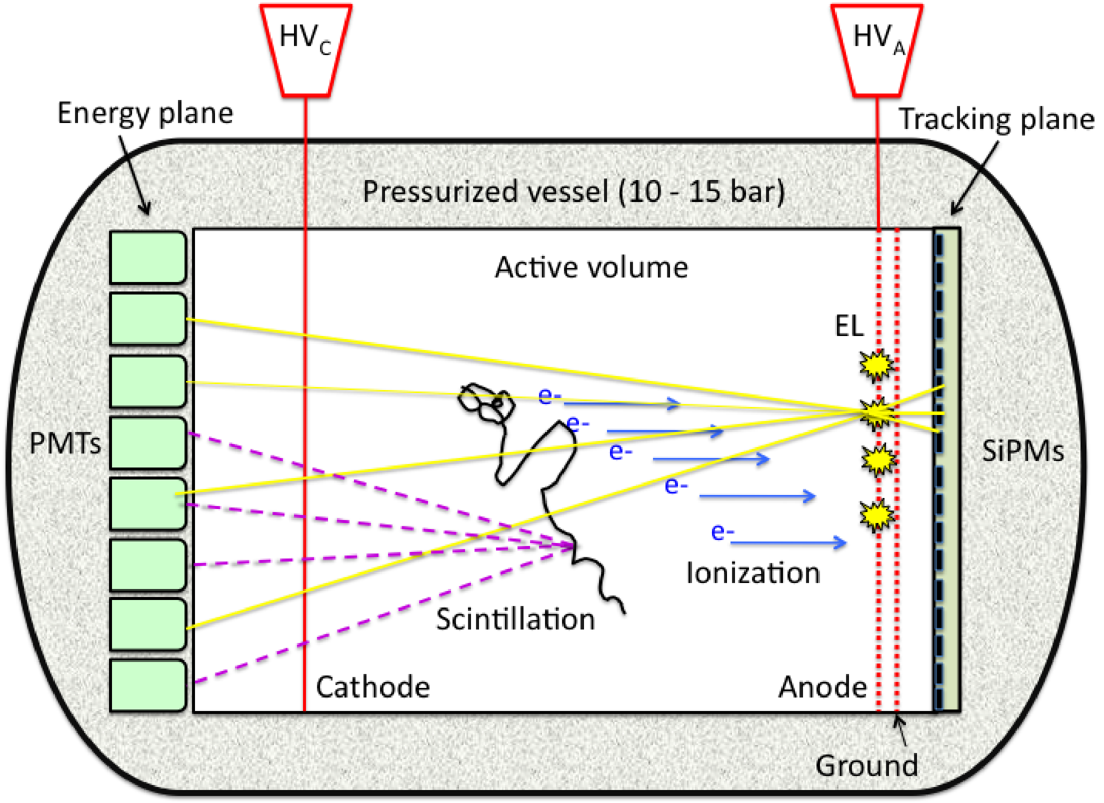


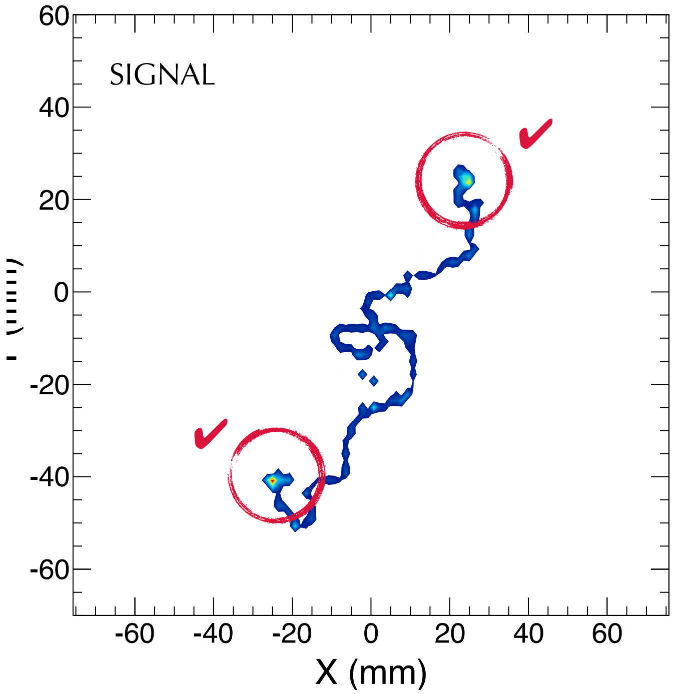
Best Results: 2% fake at 90% Electron Efficiency

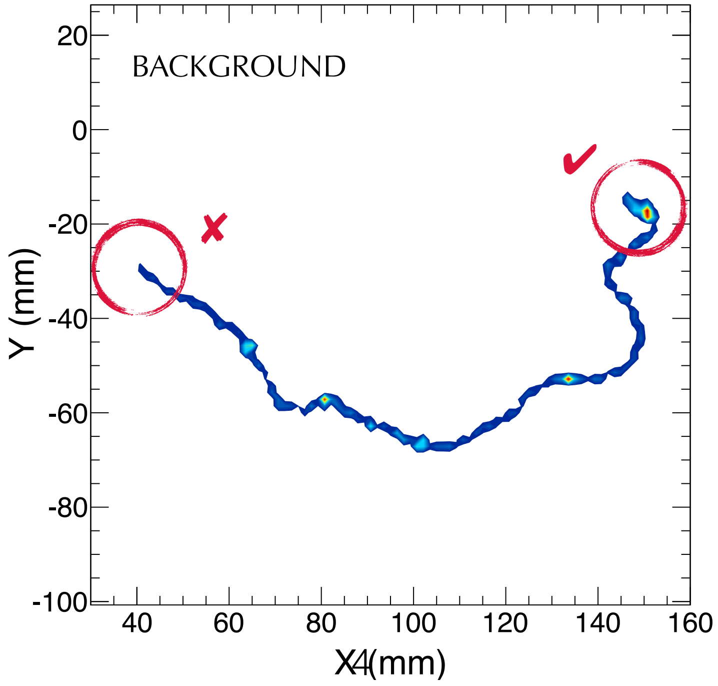
- First results with neutrinos:
  - 5% NC at 80% CC
  - 15% Muon CC at 80% Electron CC
- Regression working on Neutrino Energy
- DL efforts present also in other LArTPC experiments (not yet public).
- May be easy and ideal tool for Detector Optimization.

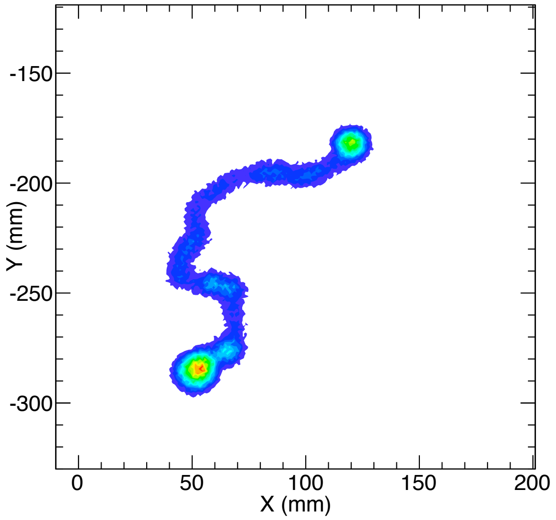
NEXT Experiment

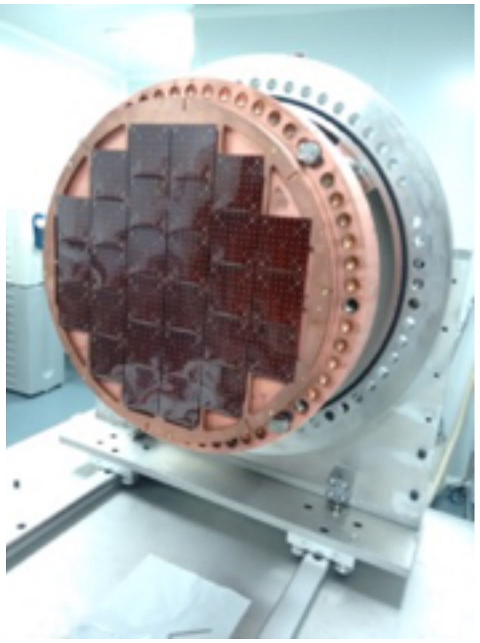
- Neutrinoless Double Beta Decay using Gas TPC/SiPMs
- Signal: 2 Electrons. Bkg: 1 Electron.
- 3D readout… candidate for 3D Conv Nets.
- Just a handful of signal events will lead to noble prize
  - Can we trust a DNN at this level?









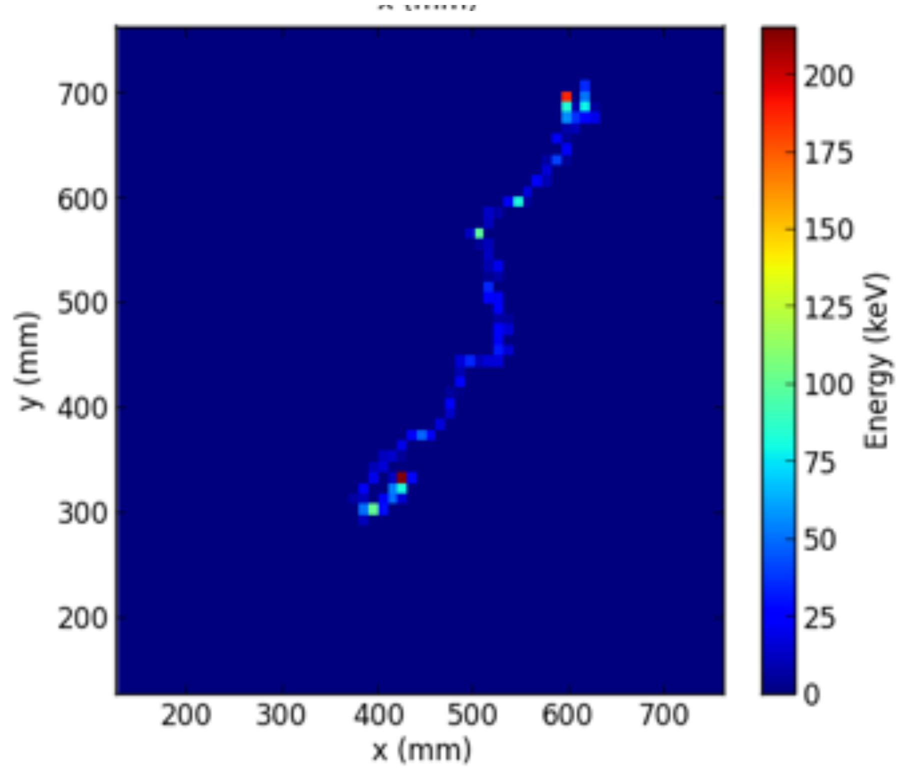


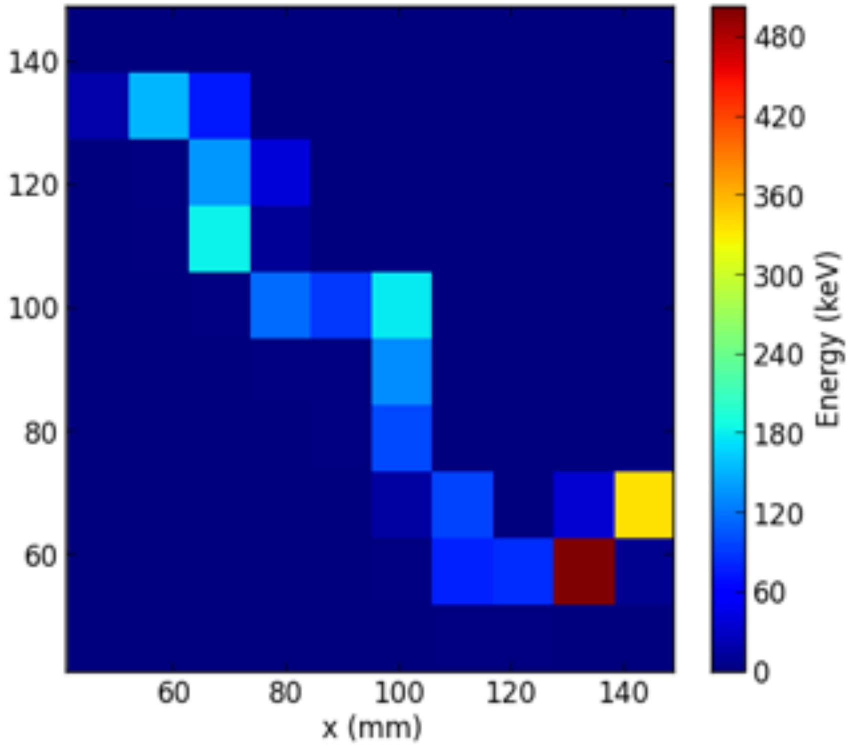
NEXT Detector Optimization

- Idea 1: use DNNs to optimize detector.
  - Simulate data at different resolutions
  - Use DNN to quickly/easily assess best performance for given resolution.
- Idea 2: understand the relative importance of various physics/ detector effects.
  - Start with simplified simulation. Use DNN to assess performance.
  - Turn on effects one-by-one.

Run (2x2x2 voxels, unless otherwise noted)	**Accuracy (%)**
toy MC, ideal	99.8
toy MC, realistic 0vbb E distribution	98.9
MAGBOX, no deltas, no E-fluctuations	**98.3**
MAGBOX, no deltas, no E-fluctuations, no brem	98.3
toy MC, realistic 0vbb E distribution, double MS	97.8
MAGBOX, no deltas	94.6
NEXT-100 fast analysis	**93.1**
MAGBOX, no E-fluctuations	93.0
MAGBOX, no brem	92.4
MAGBOX, all physics	**92.1**
10x10x5 NEXT-100 fast analysis	**86.5**

(Preliminary results)





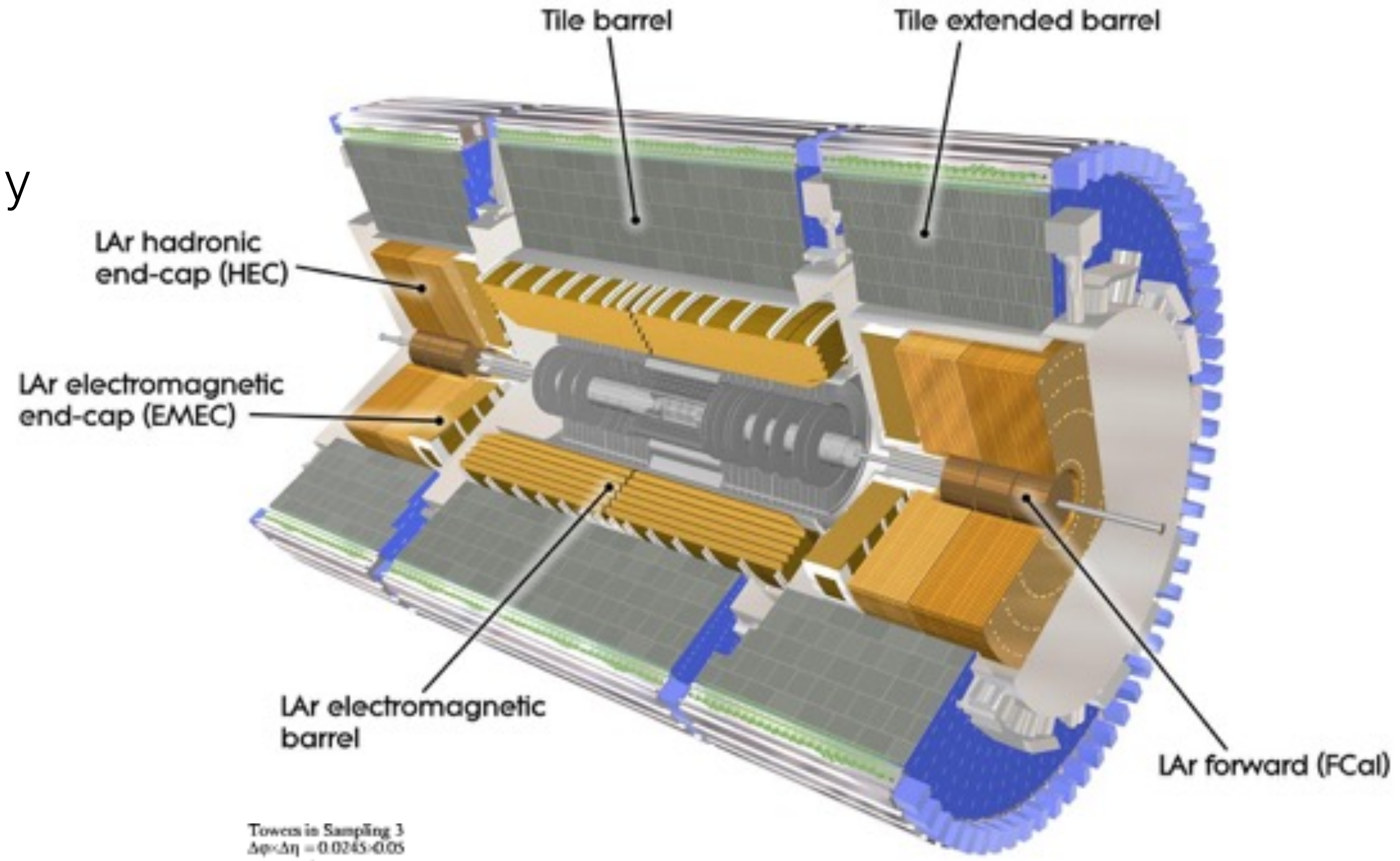
Done!

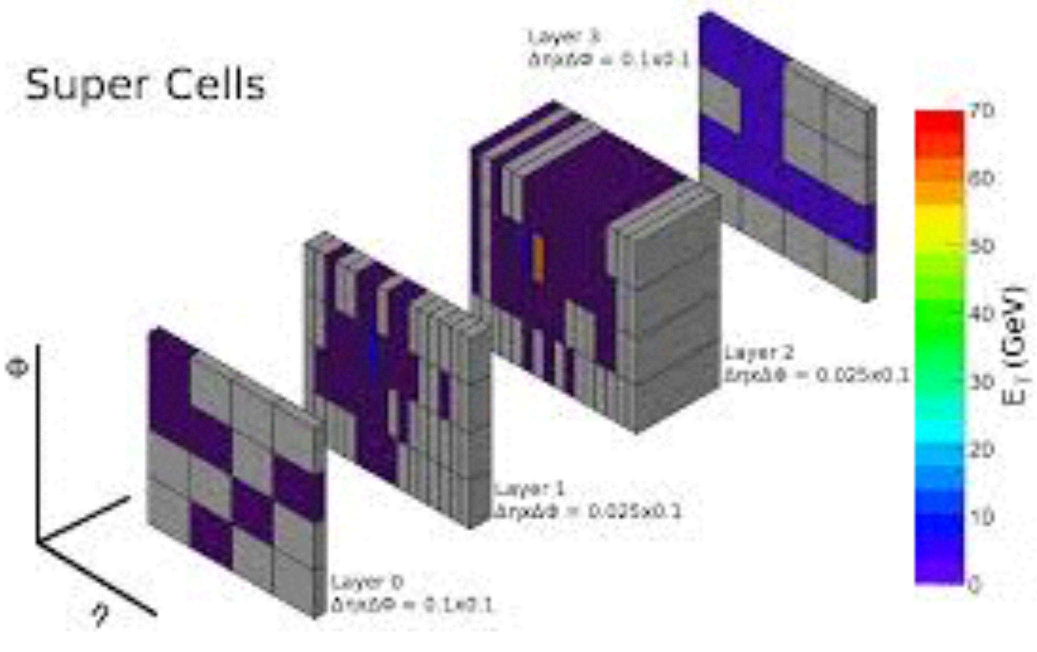
Fine-tuning

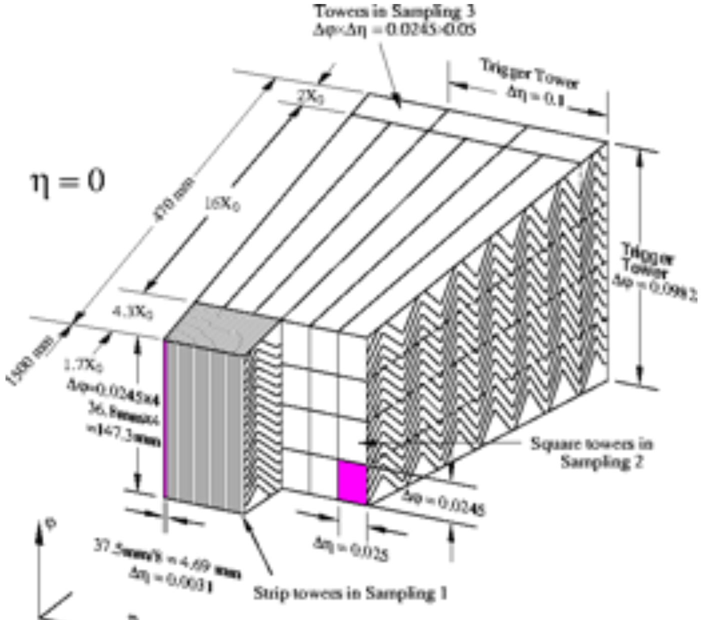
- Our existence depends on physical constants being very precisely tuned.
  - **Force of Gravity**… must be within 1 part in 10^{60}.
  - *and* **Cosmological Constant** (dark energy)… must be within 1 part in 10^{120}.
    - Or the Universe would either blow itself apart or collapse.
- **Distribution of mass energy in early Universe** must be smoothly distributed by 1 part in $(10^{10})^{123}$.
  - Or we wouldn't get structures we see today.
- The observed **Higgs mass** (observed by LHC in 2012) is naively due to a fine-tuning of 1 part in 10^{16}.
  - Or Forces and masses would be very different.
  - Only one that we have a clue on how to investigate.

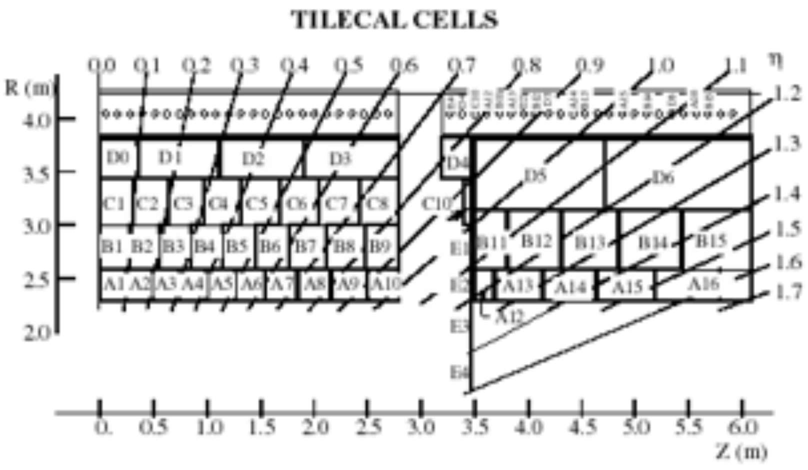
ATLAS Calorimeter

- Ideally suited for “imaging”
  - Electromagnetic- Highly transverse and longitudinal segmented.
  - Hadronic- Longitudinal sampling
  - 200K Calorimeter cells measure energy deposits.
  - ~ 64 x 36 x 7 3D Image
  - Interesting Challenges: non-uniform granularity, cylindrical geometry.



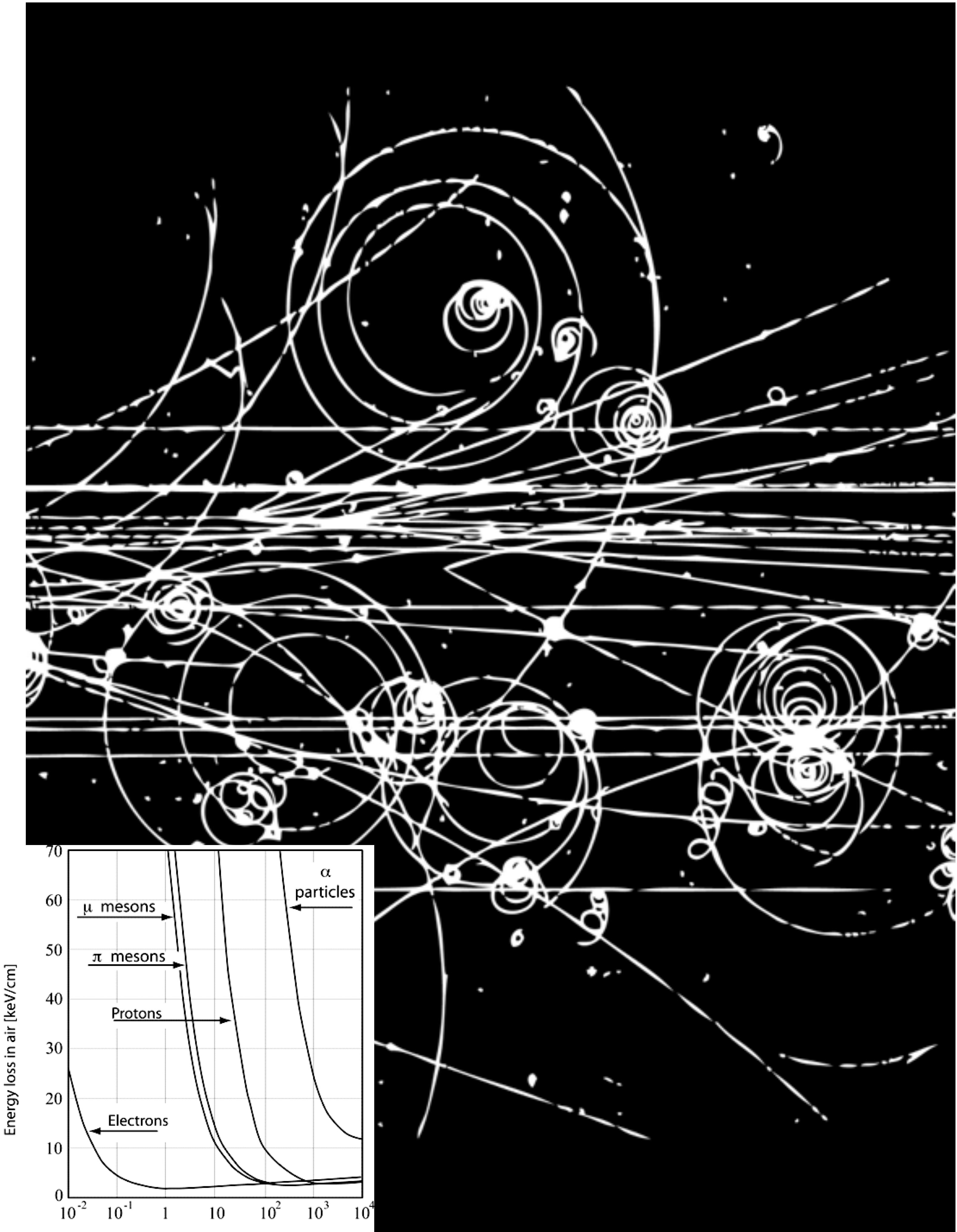






How do we "see" particles?

- Charged particles ionize media
    - Image the ions.
        - In Magnetic Field the curvature of trajectory measures momentum.
        - Momentum resolution degrades as less curvature: $\sigma(p) \sim c\,p \oplus d$.
            - d due to multiple scattering.
    - Measure Energy Loss (~ # ions)
        - dE/dx = Energy Loss / Unit Length = f(m, v) = Bethe-Block Function
            - Identify the particle type
        - Stochastic process (Laudau)
    - Loose all energy → range out.
        - Range characteristic of particle type.



How do we “see” particles?

- Particles deposit their energy in a stochastic process know as “showering”, secondary particles, that in turn also shower.
    - Number of secondary particles ~ Energy of initial particle.
    - Energy resolution improves with energy: $\sigma(E) / E = a/\sqrt{E} \oplus b/E \oplus c$.
        - a = sampling, b = noise, c = leakage.
    - Density and Shape of shower characteristic of type of particle.
- Electromagnetic calorimeter: Low Z medium
    - Light particles: electrons, photons, $\pi^0 \rightarrow \gamma\gamma$ interact with electrons in medium
- Hadronic calorimeters: High Z medium
    - Heavy particles: Hadrons (particles with quarks, e.g. charged pions/protons, neutrons, or jets of such particles)
        - Punch through low Z.
        - Produce secondaries through strong interactions with the nucleus in medium.
        - Unlike EM interactions, not all energy is observed.