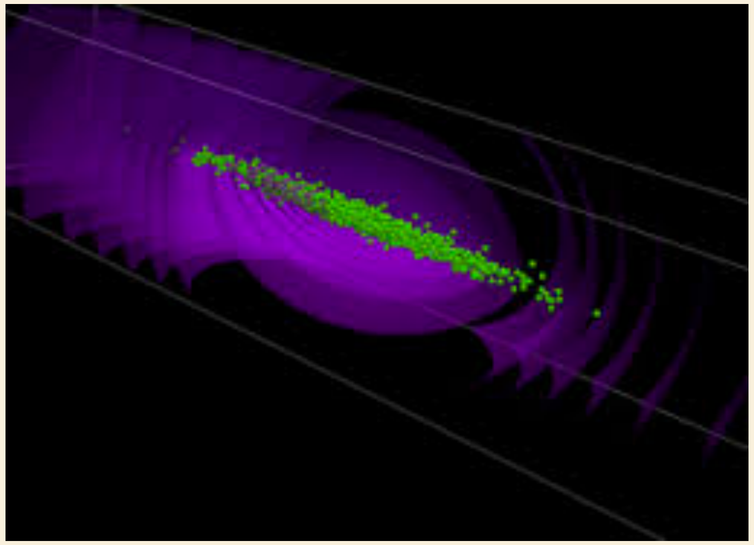
Simulations for Discoveries at Fermilab

Adam Lyon / Fermilab SCD / Muon g-2

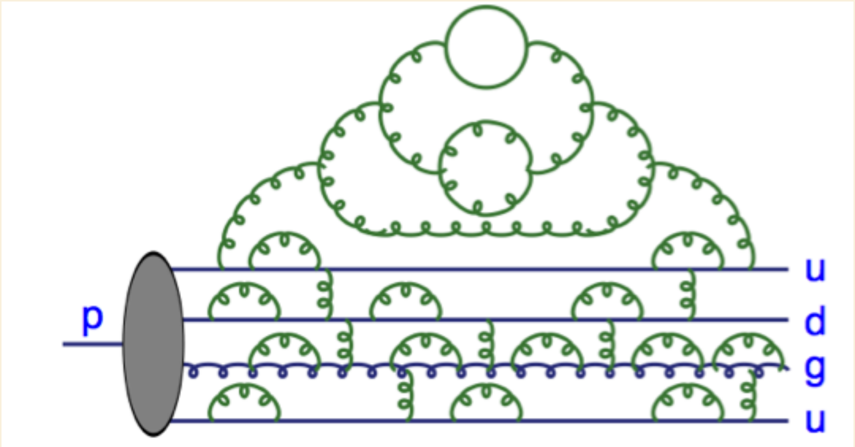
IIT Applied Math Colloquia 2 March 2015

I'll be talking about…

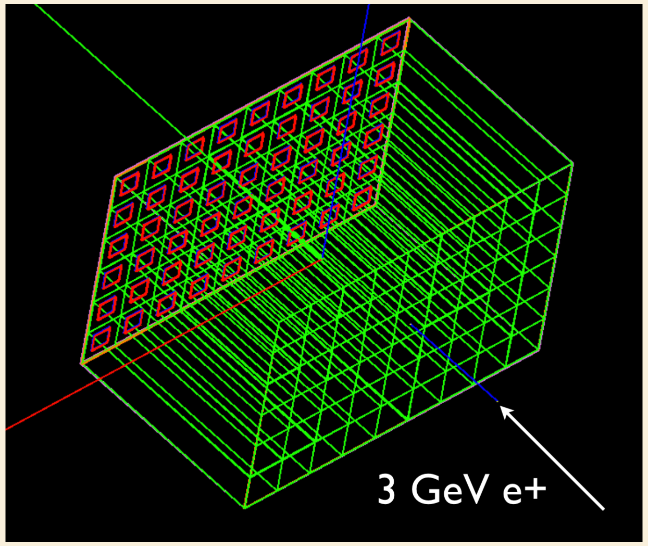
Introduction



I: Accelerator Simulations

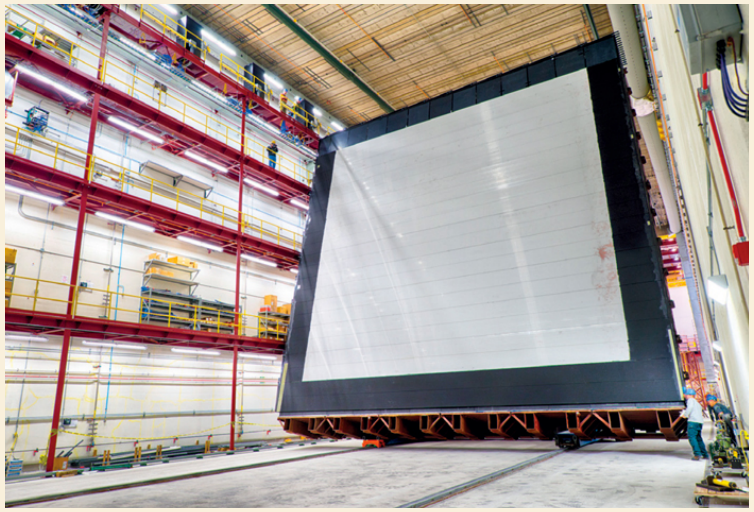


II: Collider Experiment Simulations



III: Detector Simulations

IV: Running Simulations

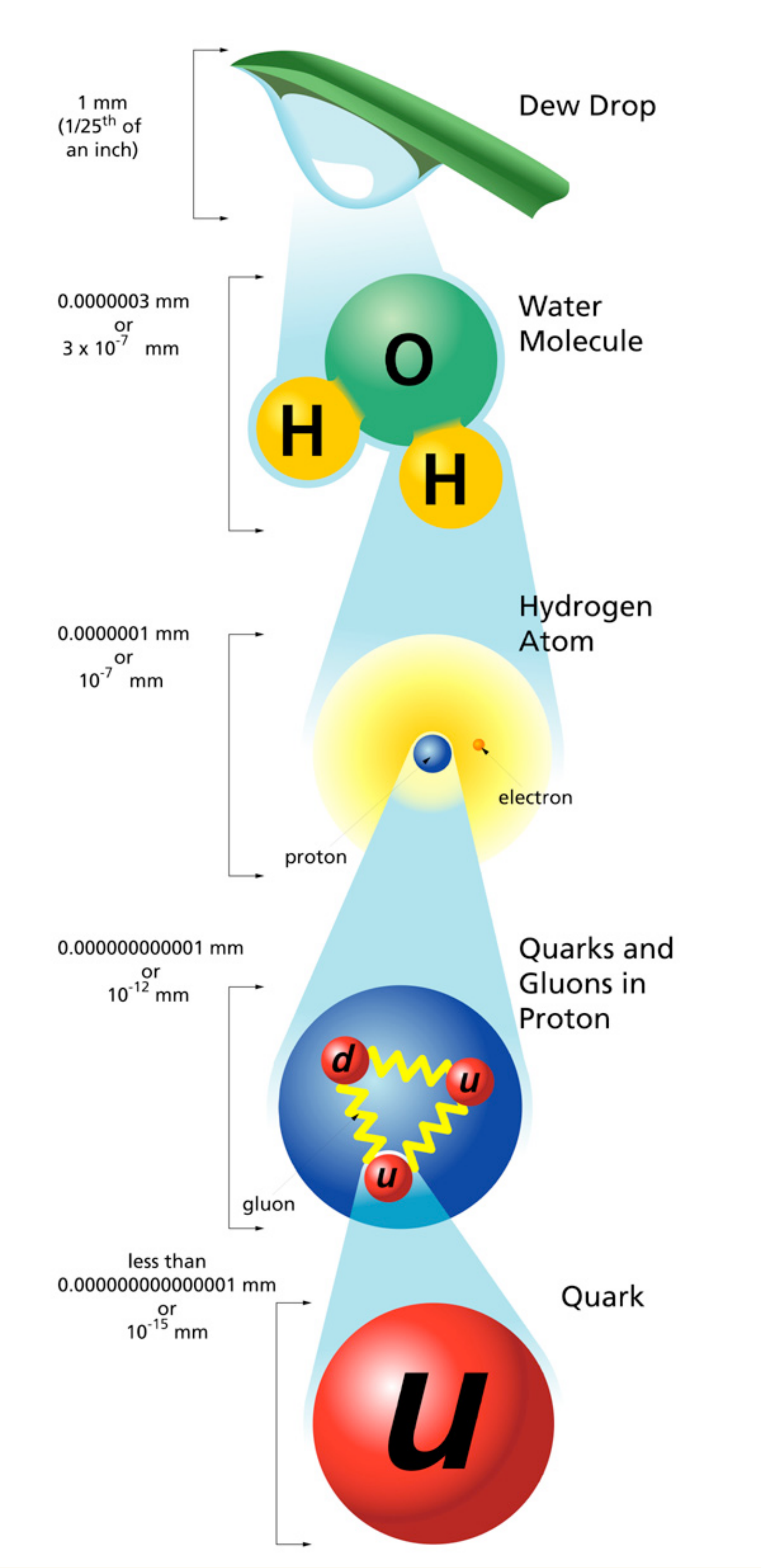


V: Simulations and "Big Data"

What is particle physics?

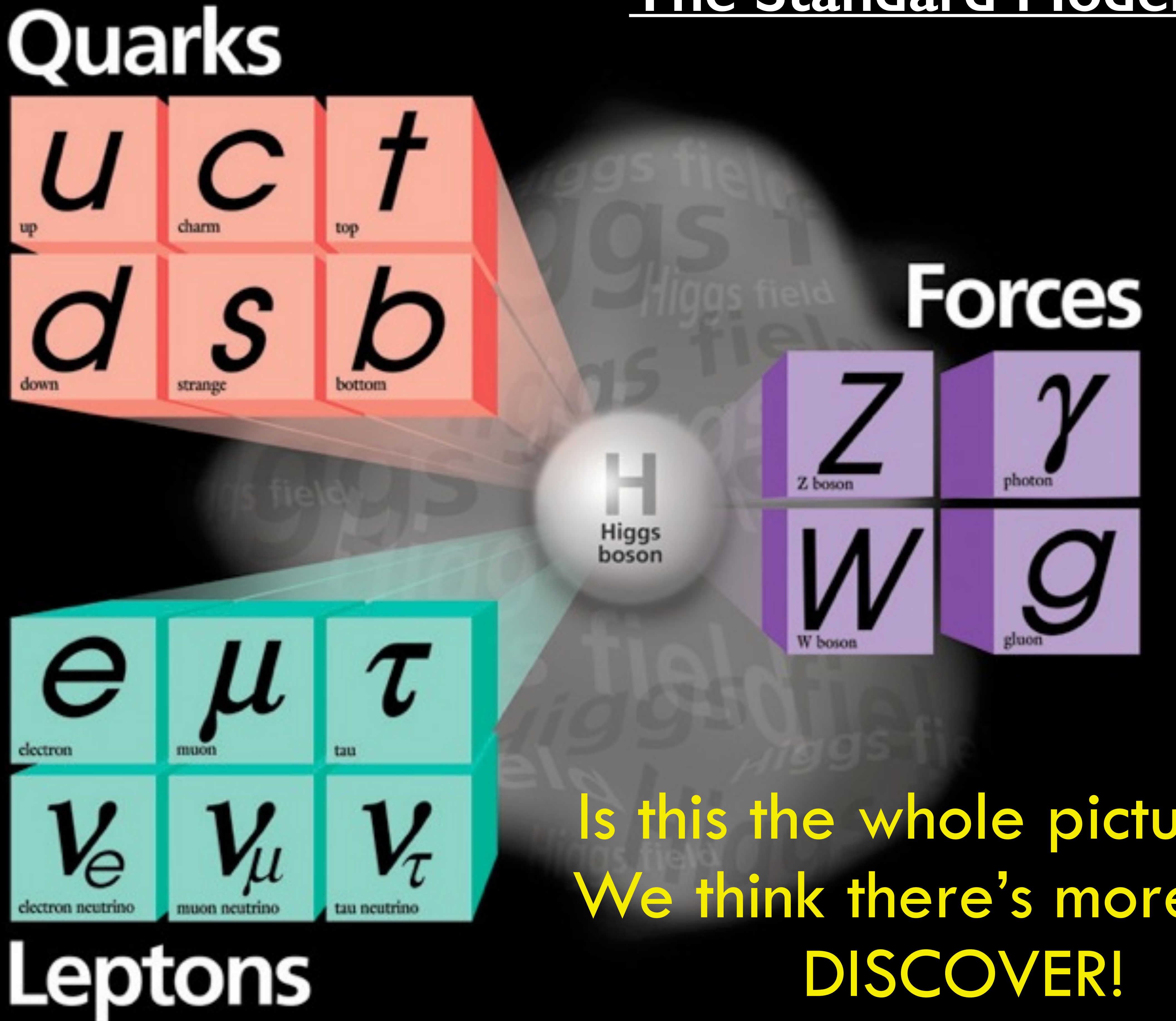
Q: What is matter?
How does it all work?

A: We've learned a lot!



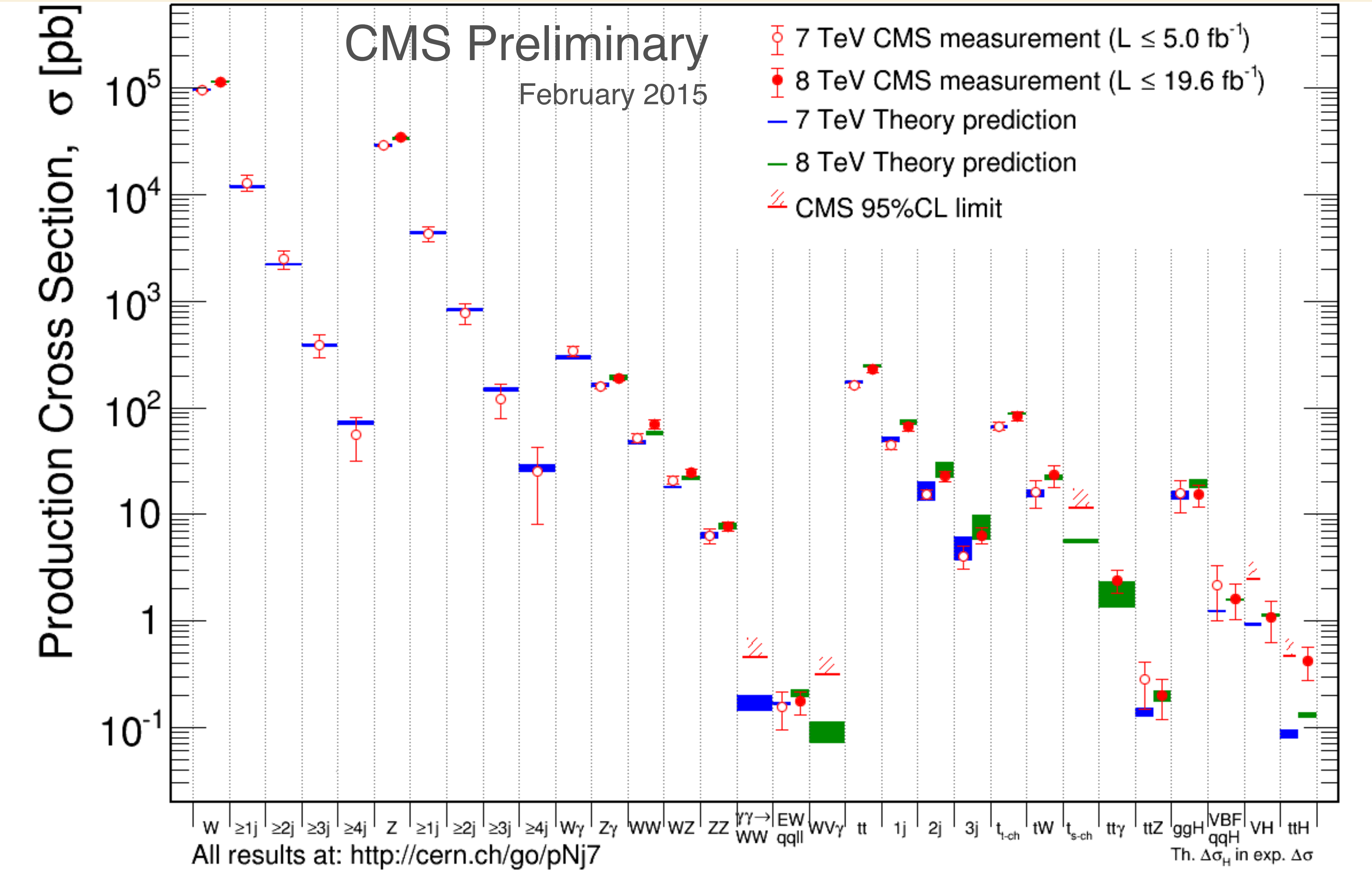
Fermilab

The Standard Model



Is this the whole picture?
We think there's more to
DISCOVER!

While the Standard Model is incredibly successful,



it gives an **incomplete** picture of nature:

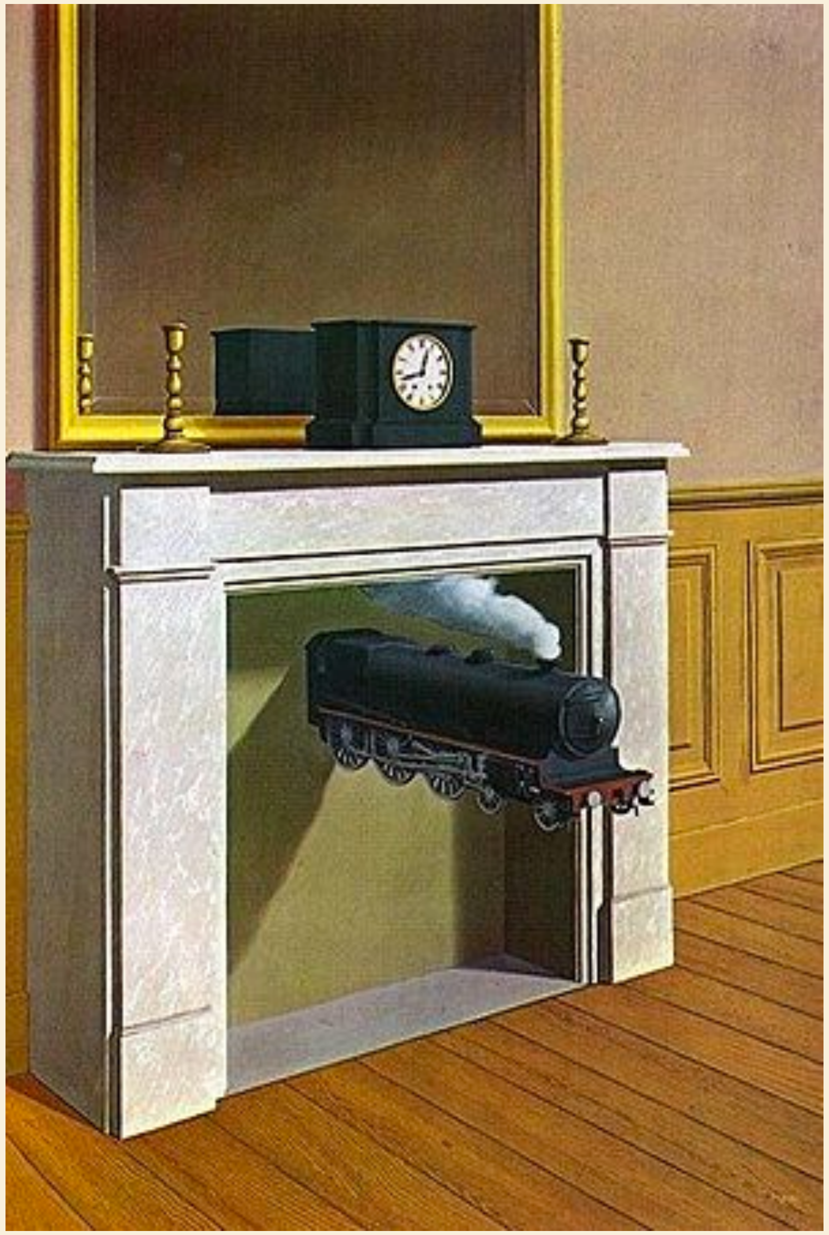
Gravity? Matter/antimatter asymmetry? Dark Matter? SUSY?

Search for new phenomena **beyond** the Standard Model

Some important concepts

The universe we live in is a far stranger place than our everyday intuition leads us to expect.

(Chad Orzel in Teach Relativity to Your Dog)



Magritte, Art Institute of Chicago

Particles live in the quantum world and they are ruled by probabilities, not certainty

Particles don't have personalities
e.g. All electrons follow the same rules, no exceptions

Mass and energy are interchangeable
$E = mc^2$

Some particles change with time
Neutrinos oscillate into different types

Empty space is not empty
Proton is uud quarks, but in reality more complicated

These concepts are all important to experiments

How particle collisions work

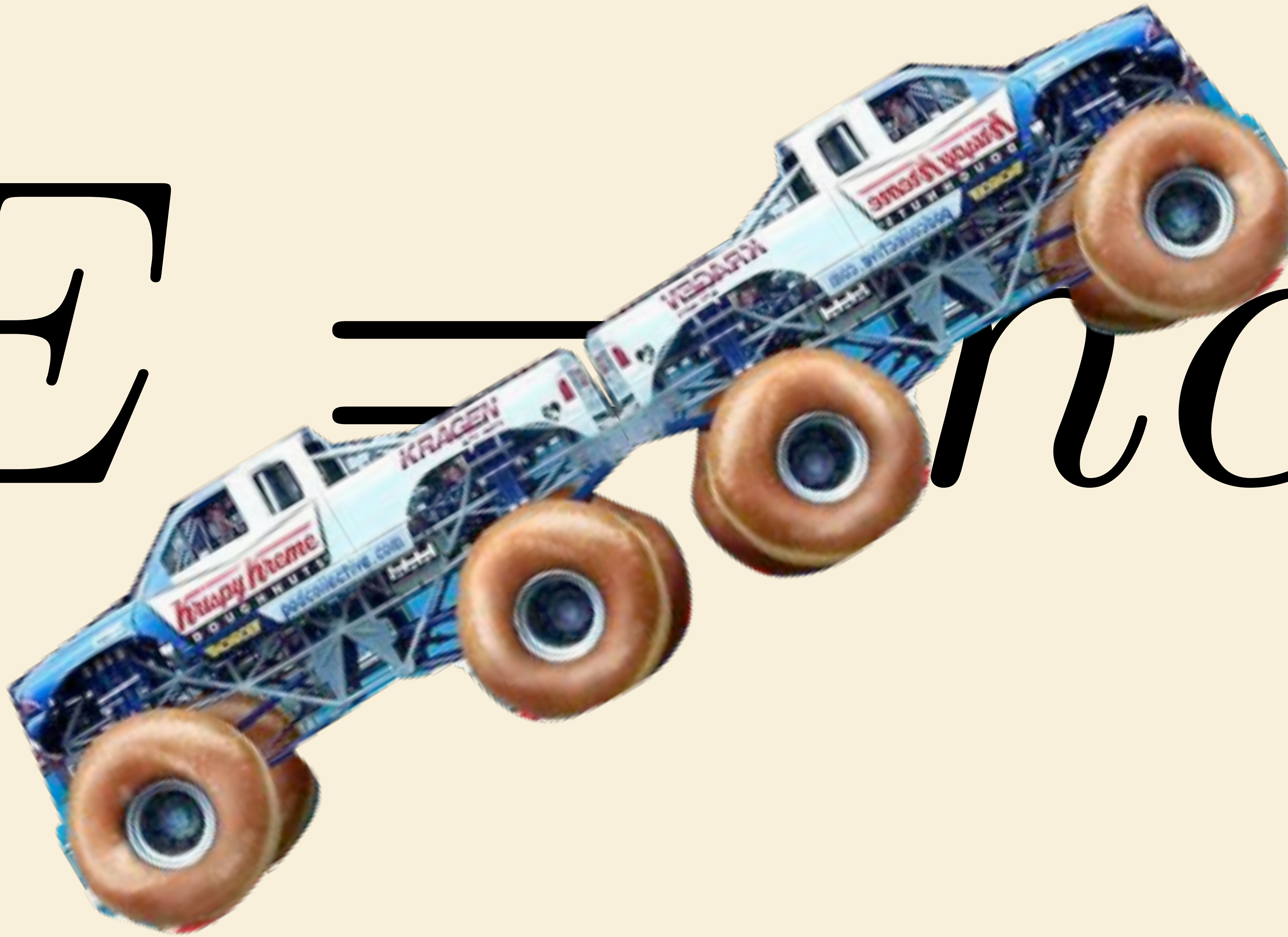
$$E = mc^2$$

How particle collisions work

$$E = mc^2$$

How particle collisions work

$$E = mc^2$$



How particle collisions work

Collisions into fixed targets

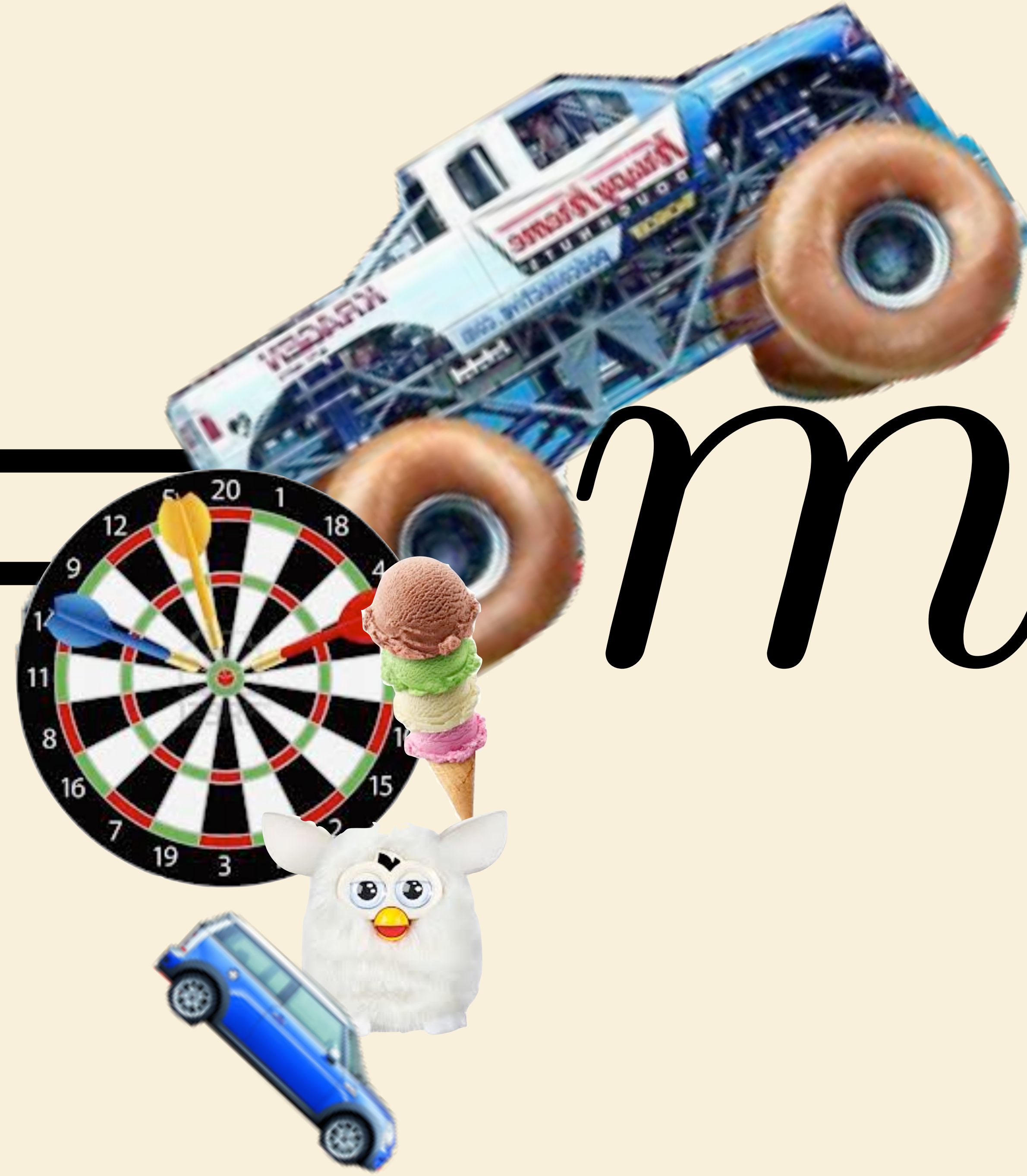
$$E = mc^2$$

Collisions into fixed targets

$$E = mc^2$$

Collisions into fixed targets

$$E = mc^2$$



Collisions into fixed targets

$$E = mc^2$$

Quantum Physics
PROBABILITIES

Quantum Physics
PROBABILITIES

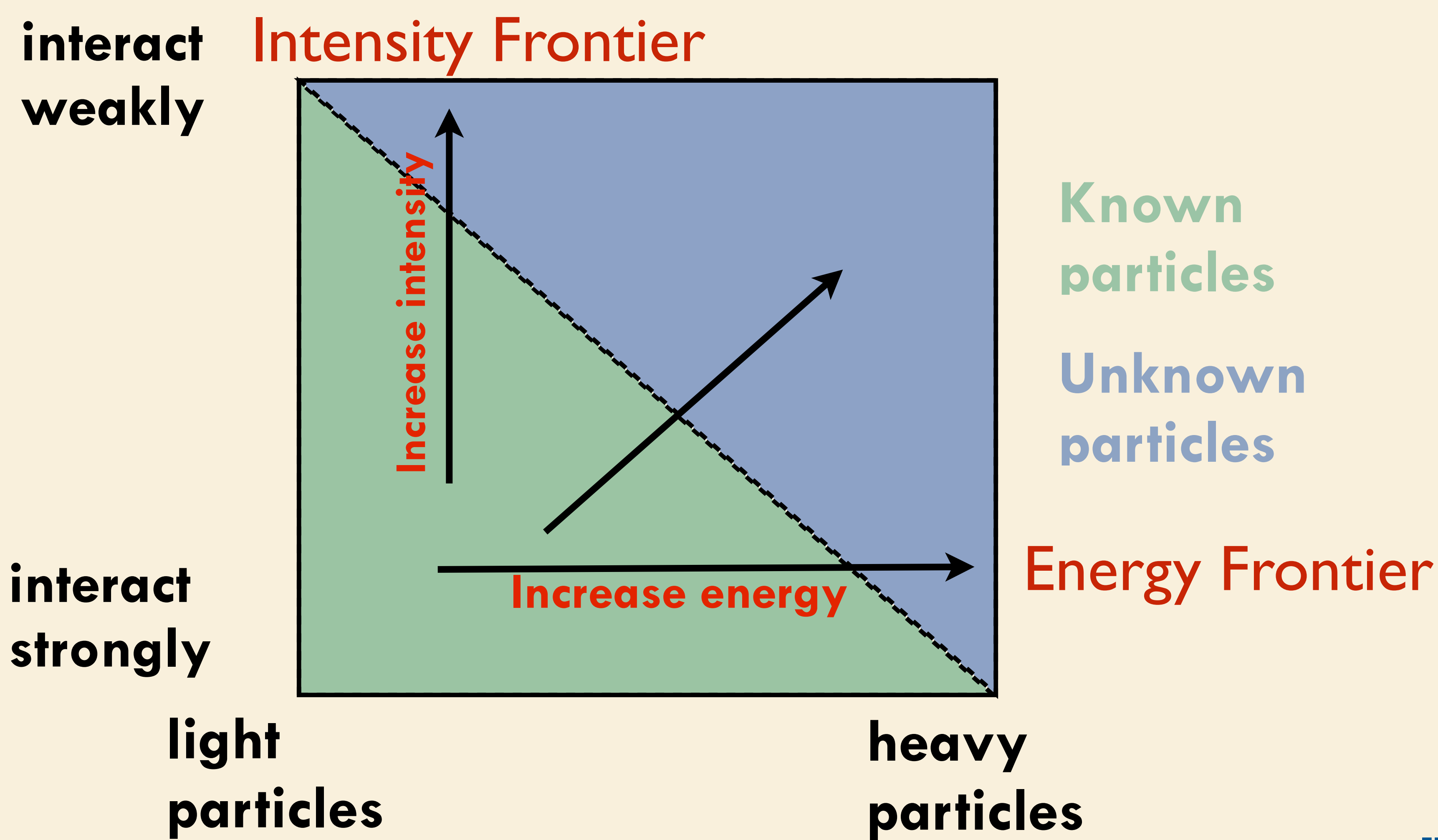
Quantum Physics
PROBABILITIES

Quantum Physics
PROBABILITIES

Ways to look for new physics

Find new heavy particles
- **Need to collide at high energy to make them**

Find particles that weakly interact (low probabilities)
- **Need to collide LOTS of particles to catch rare interactions**



What is Fermilab?

America's Premier Particle Physics Laboratory

~1700 employees

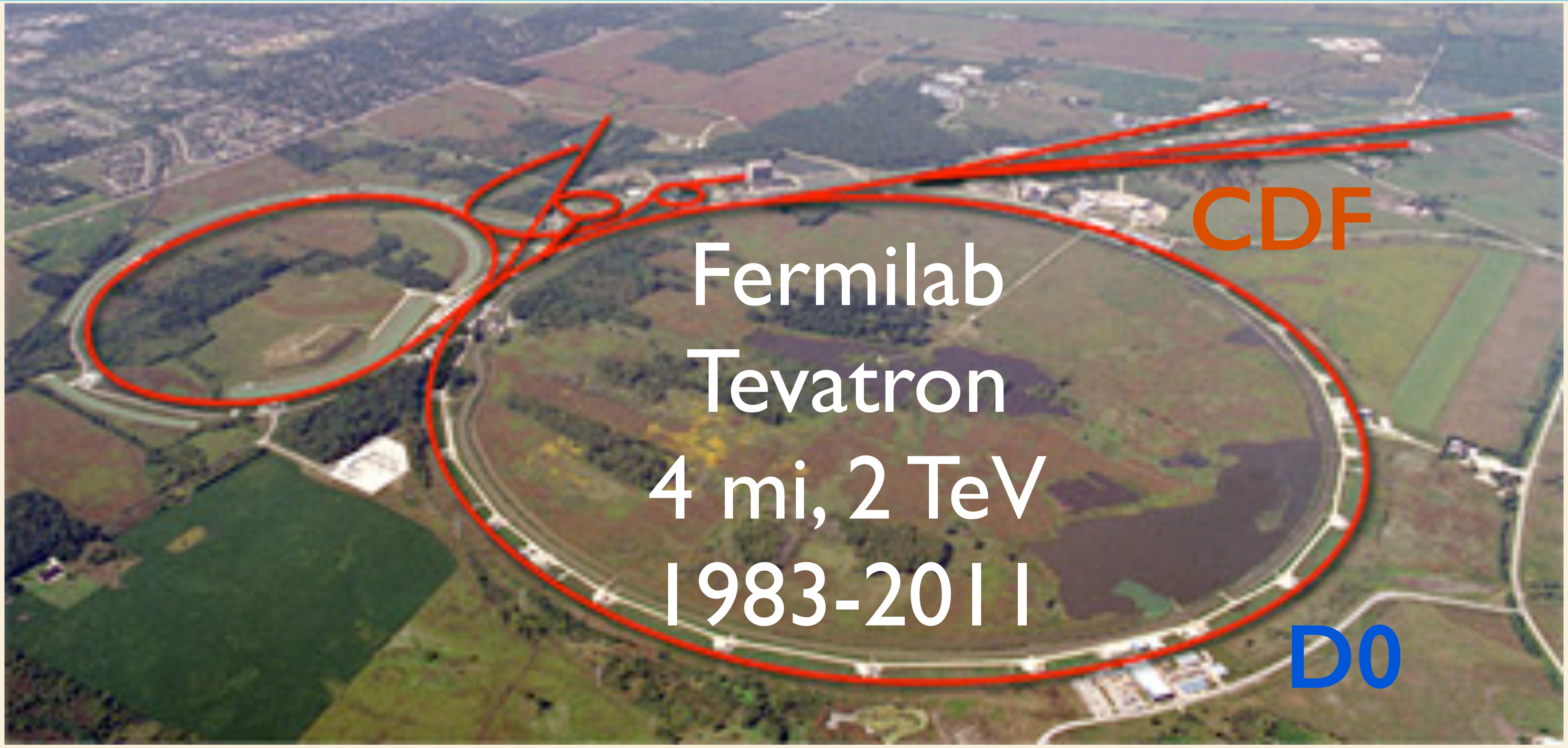
A US Department of Energy Office of Science National Laboratory (10 Office of Science labs, 17 total)

All basic, non-classified research

Open to visitors - come and visit!

www.fnal.gov

Tools of the Energy Frontier

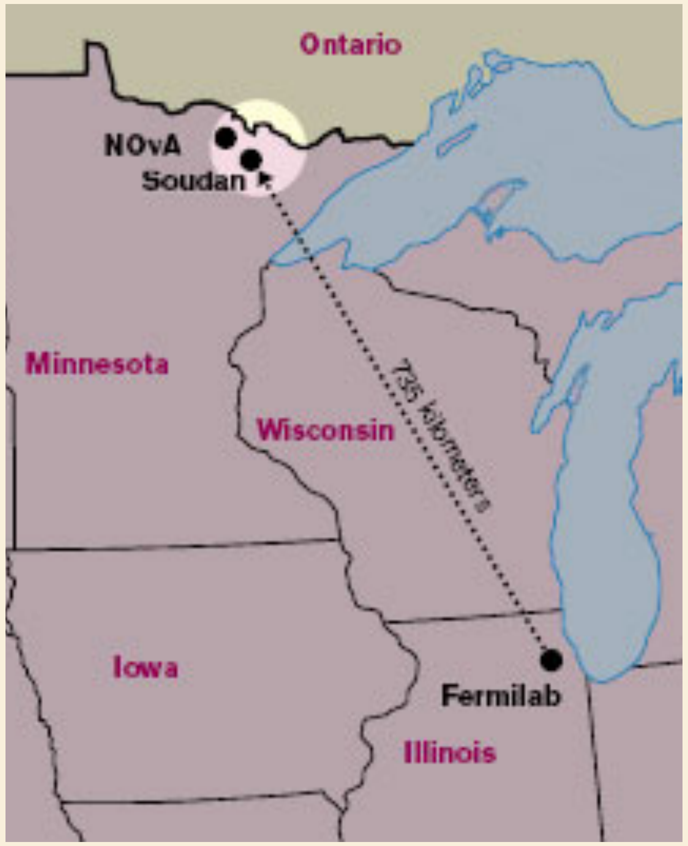


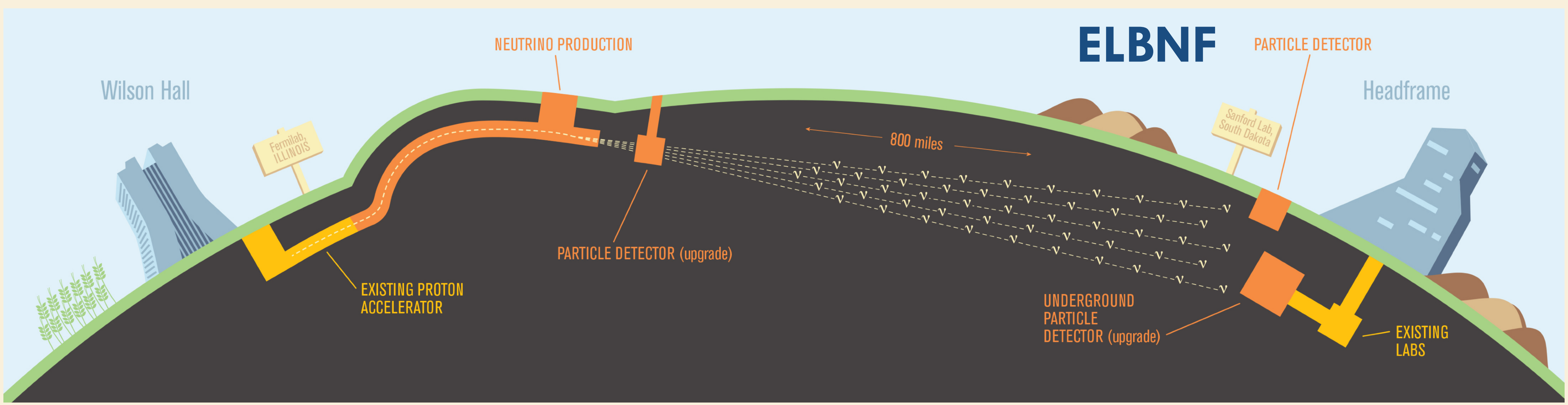
Intensity Frontier (Neutrino Oscillations)

Intensity Frontier (Neutrino Oscillations)

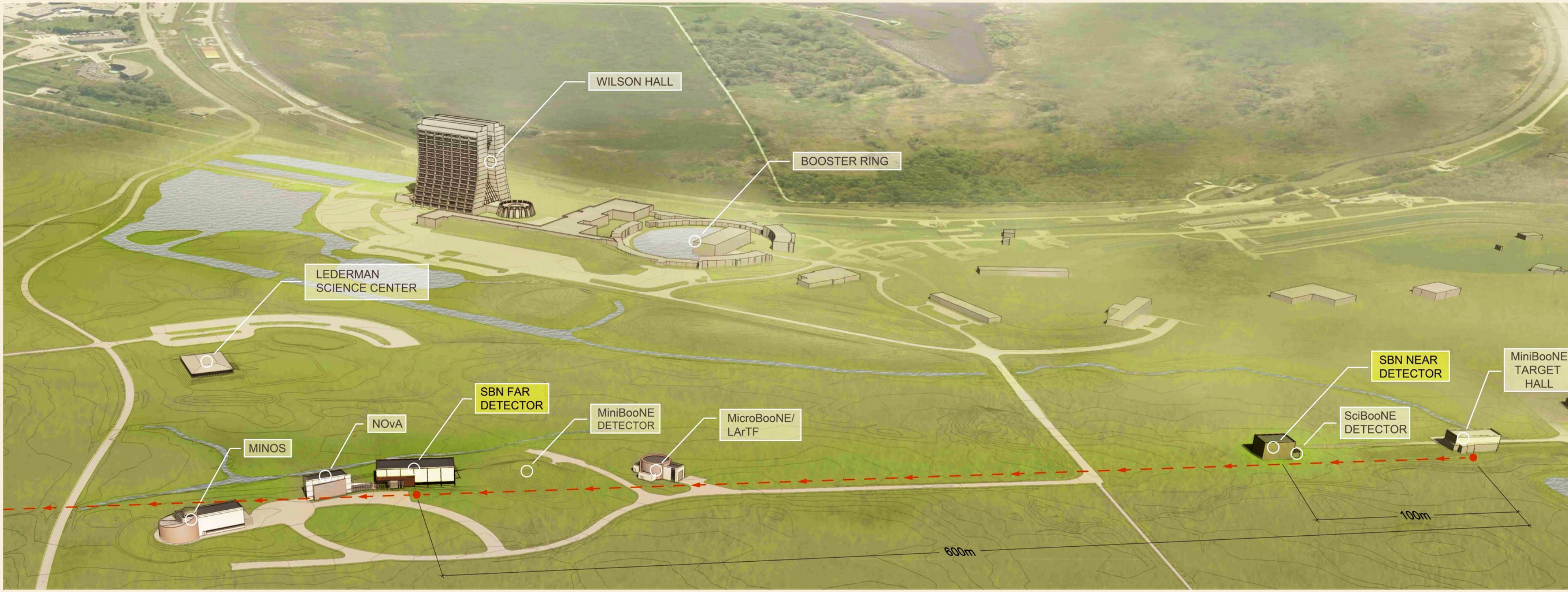
Intensity Frontier (Neutrino Oscillations)

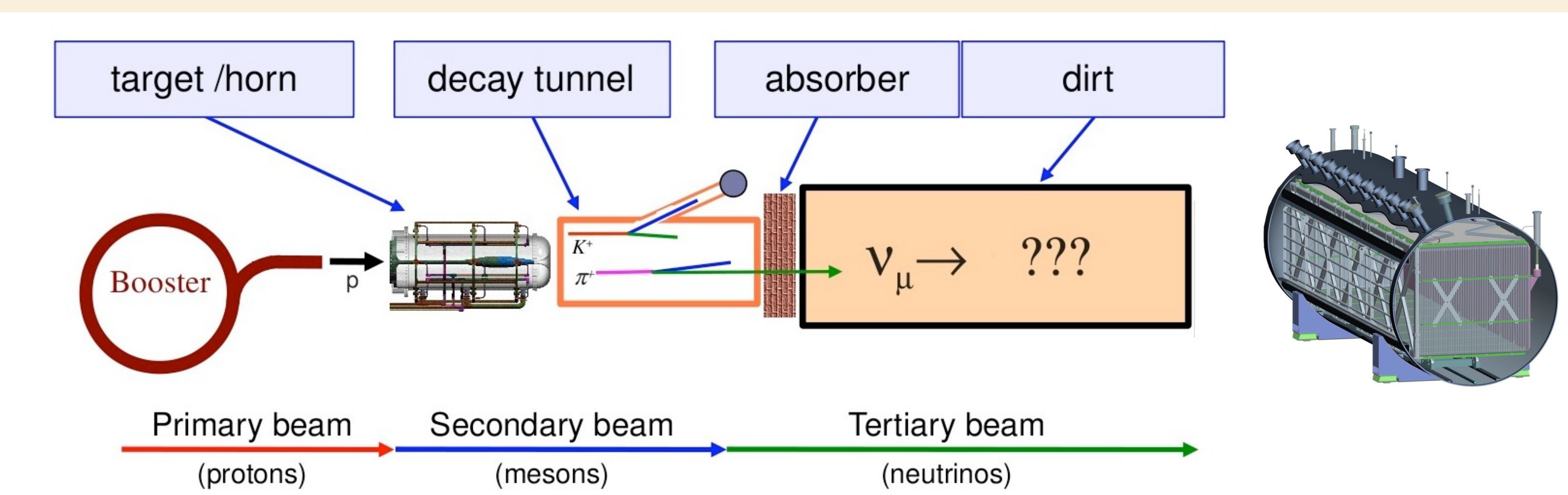
Long baseline neutrino experiments

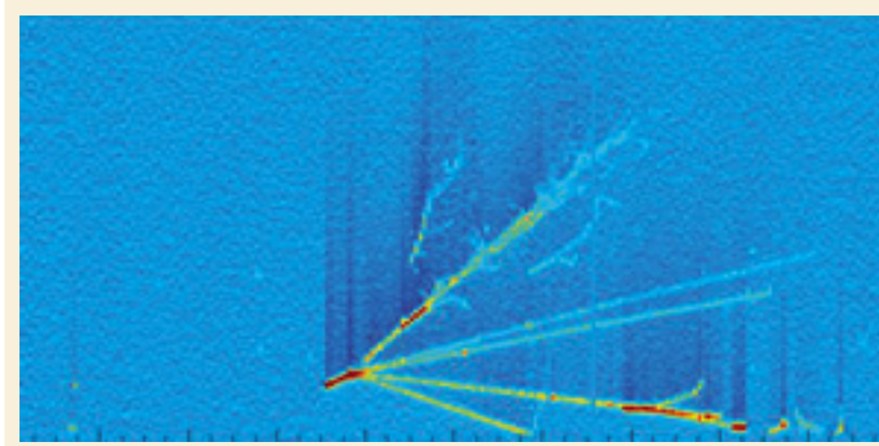




Short baseline neutrino experiments

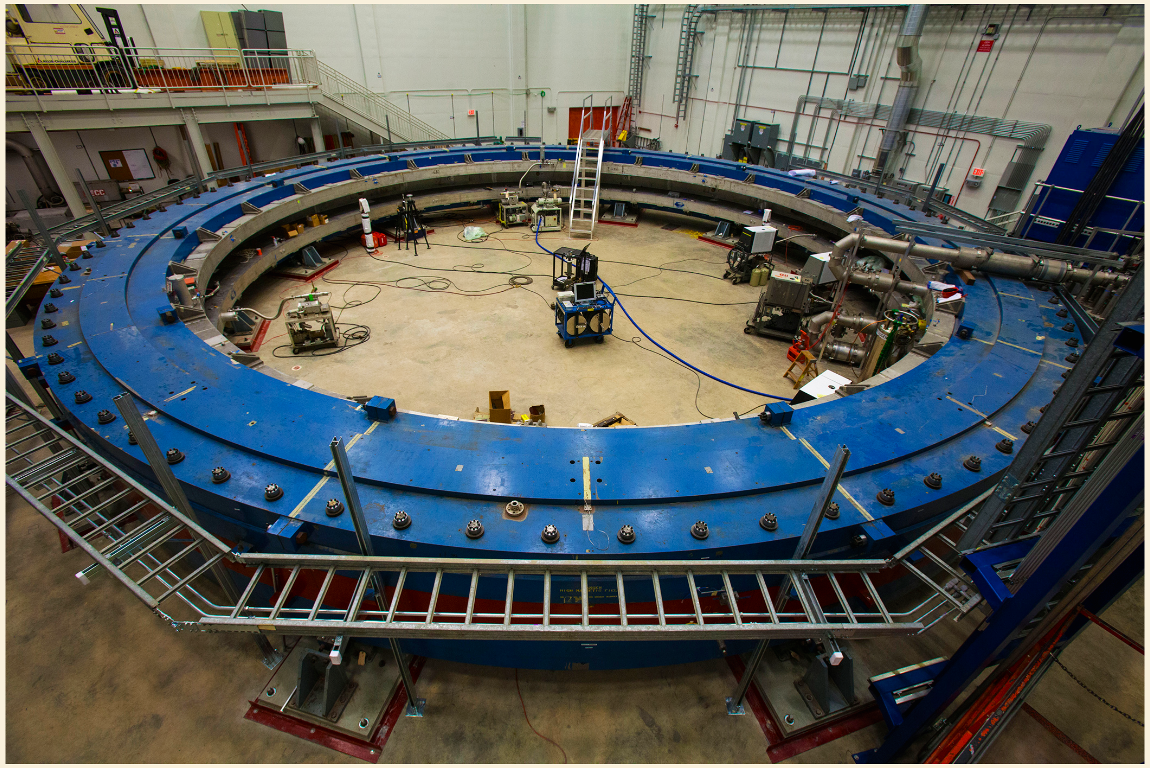


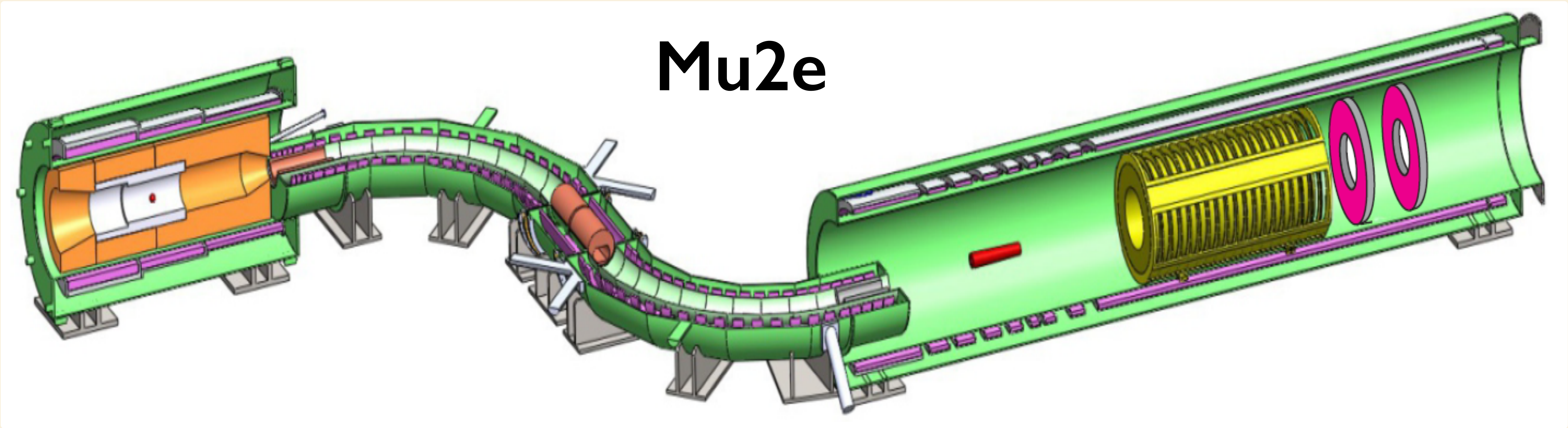




Fermilab

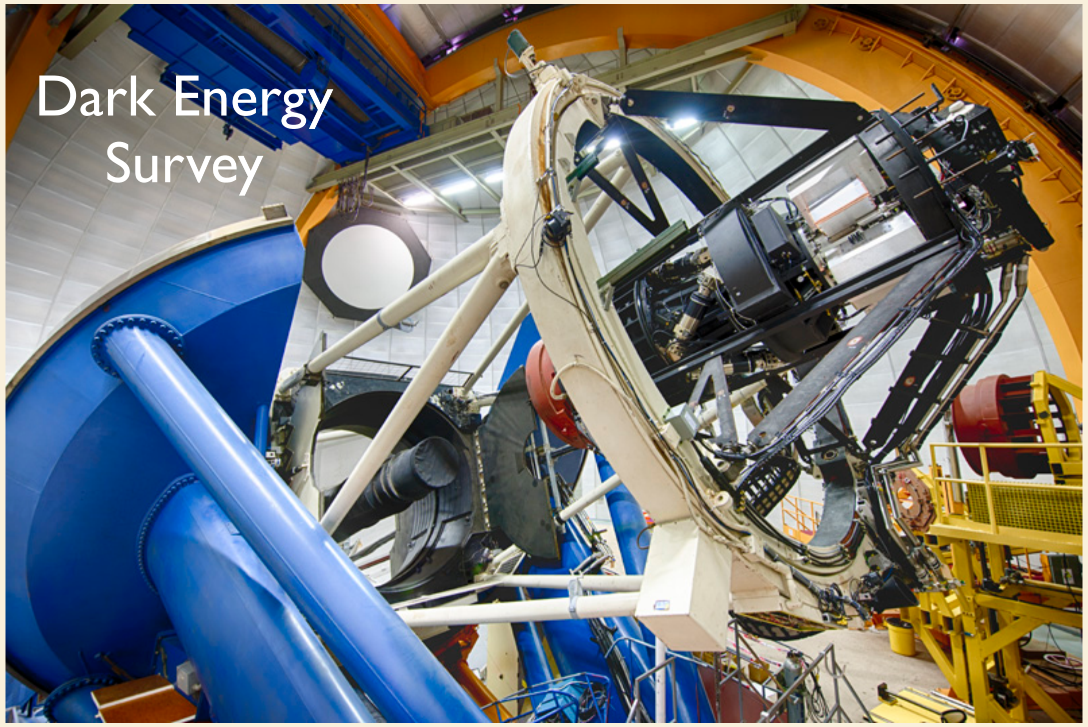
Precision Muon Physics

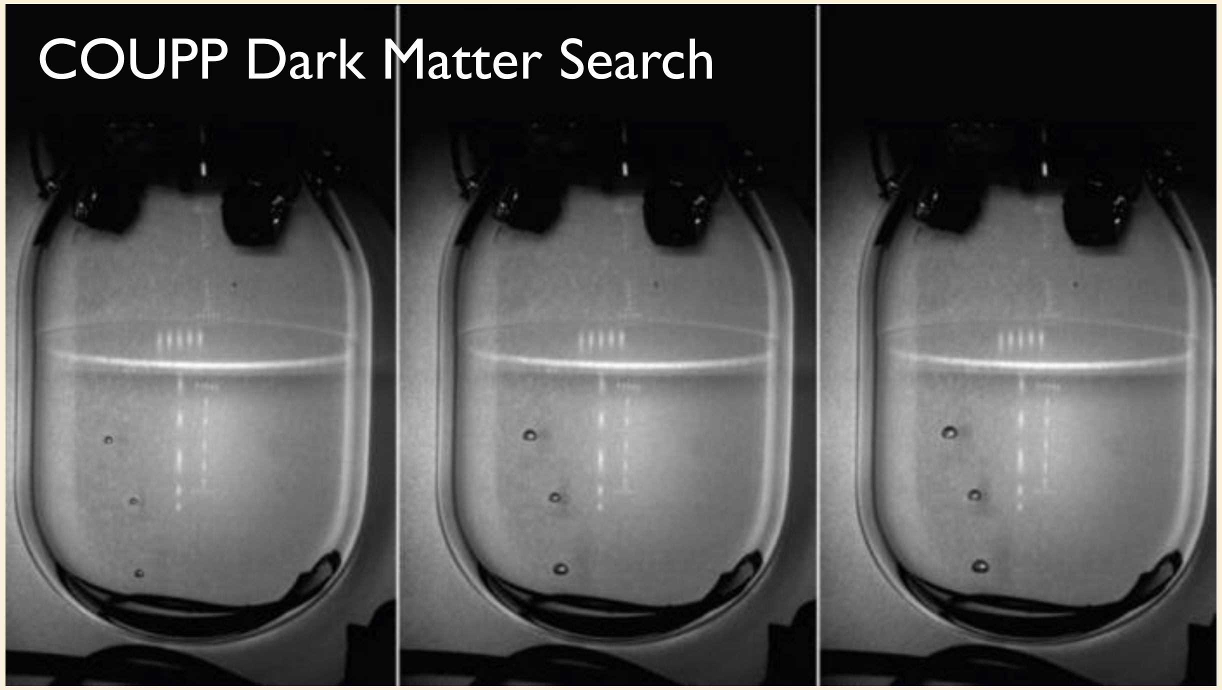




Fermilab at the Cosmic Frontier

Understanding Dark Energy and Dark Matter





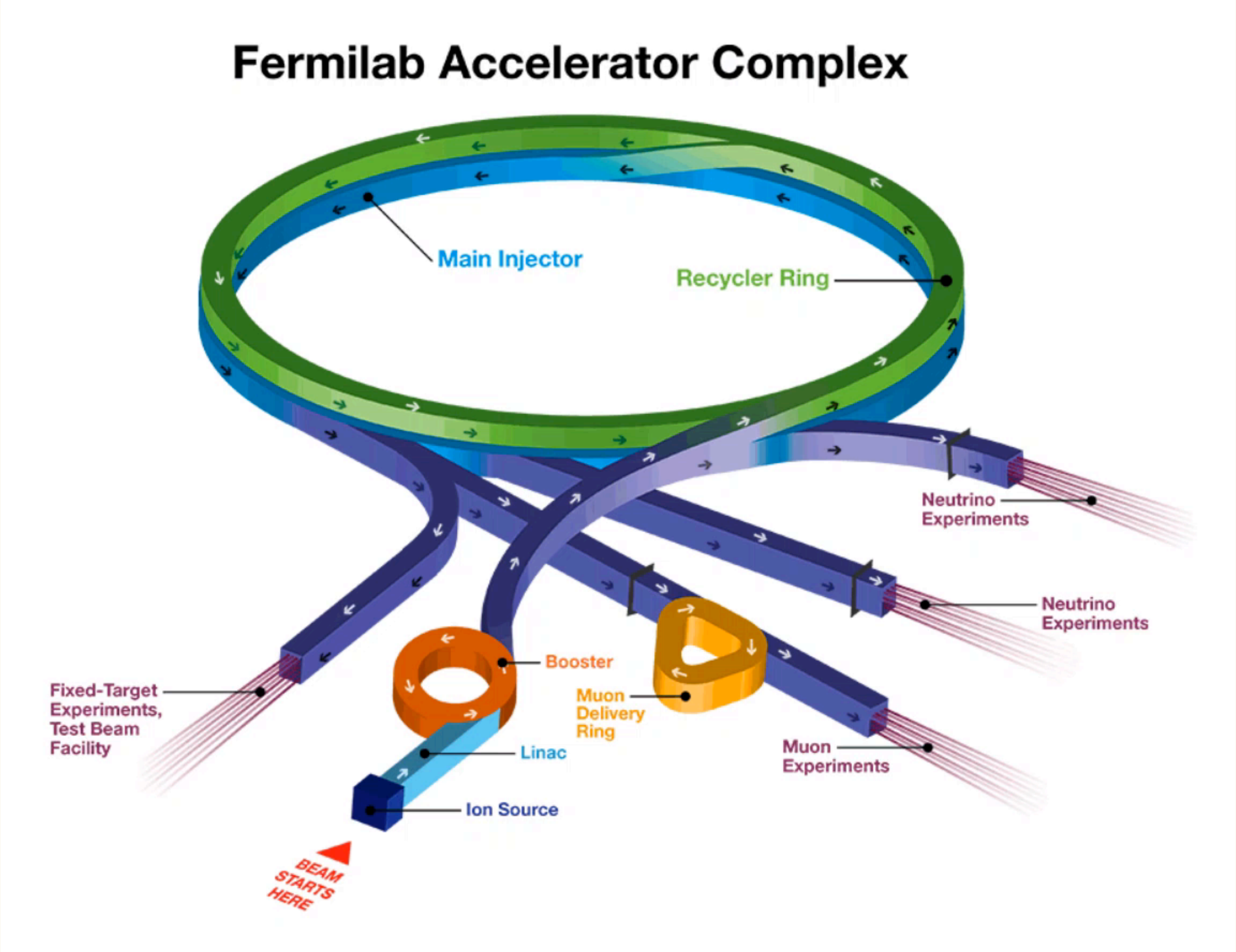
ILLINOIS INSTITUTE OF TECHNOLOGY @ Fermilab

The Center for Accelerator and Particle Physics

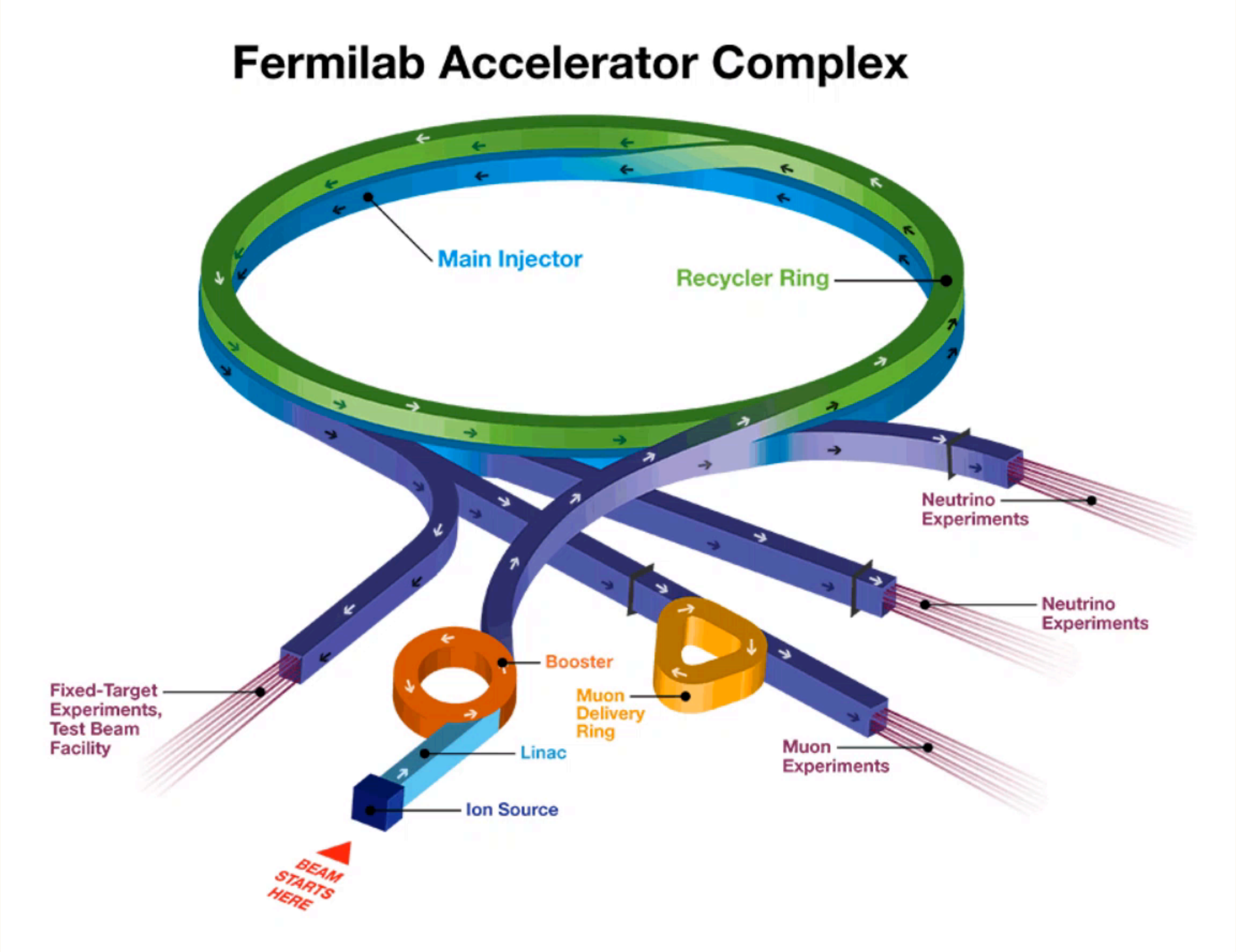
Illinois Consortium for Accelerator Research

- **Long involvement on Fermilab experiments**
- **MINOS – Long baseline neutrino experiment**
- **MuCool / MICE – new particle beam technology for muon accelerators**
- **Accelerator simulation**

Fermilab's Current Accelerators

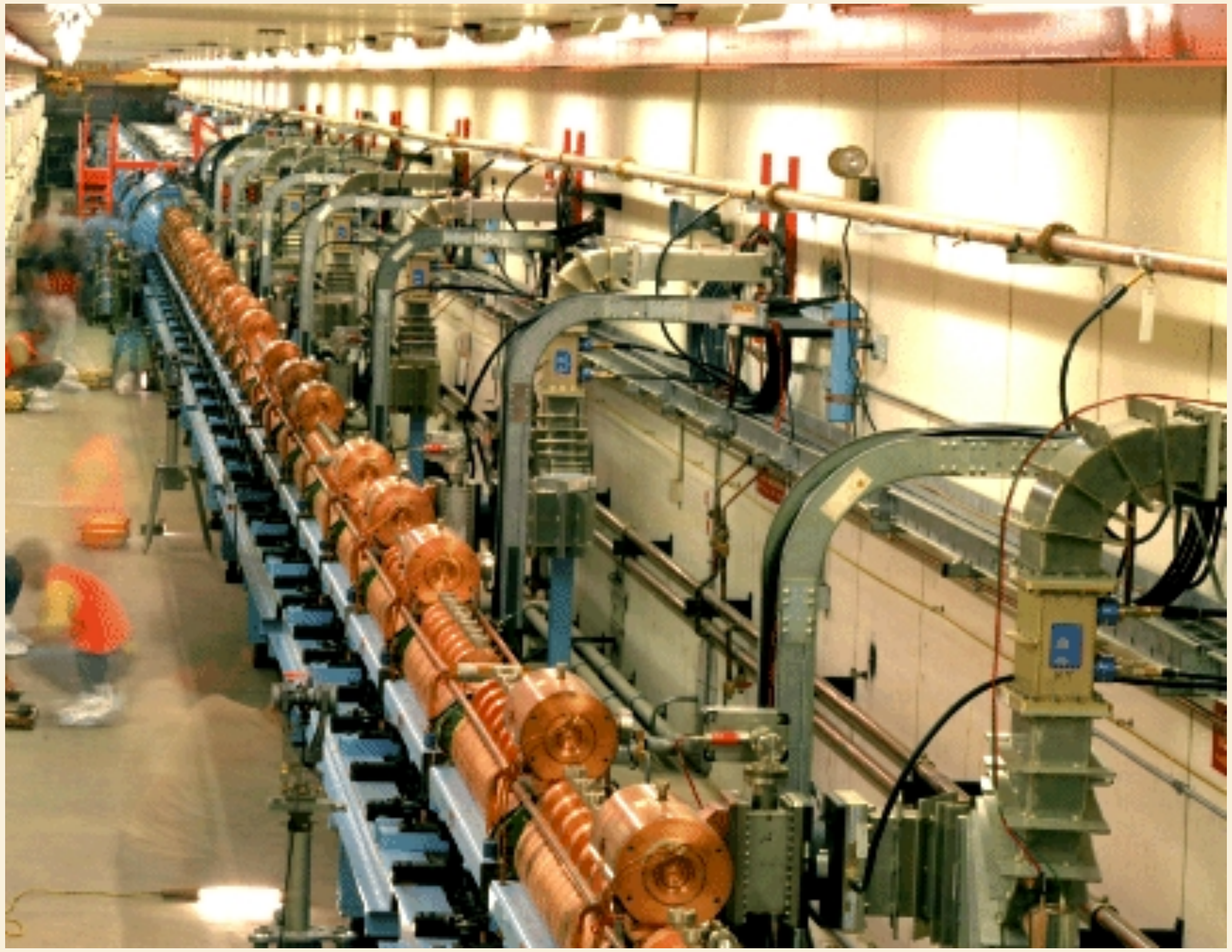


Fermilab's Current Accelerators



Accelerating particles

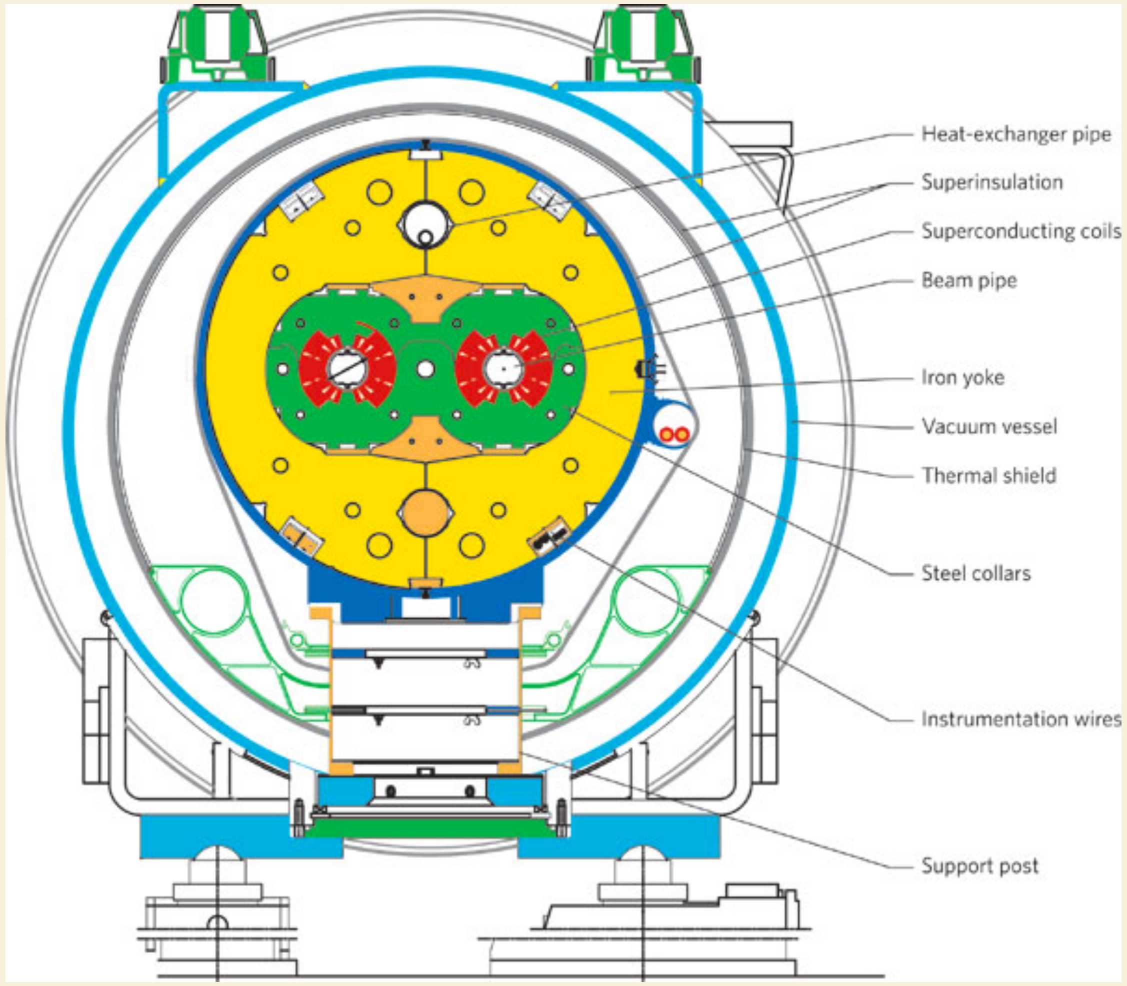
Electromagnetic waves from RF cavities accelerate charged particles



Accelerating particles

Circulate beam to kick repeatedly with same cavity

Magnets bend beam to circle and focus



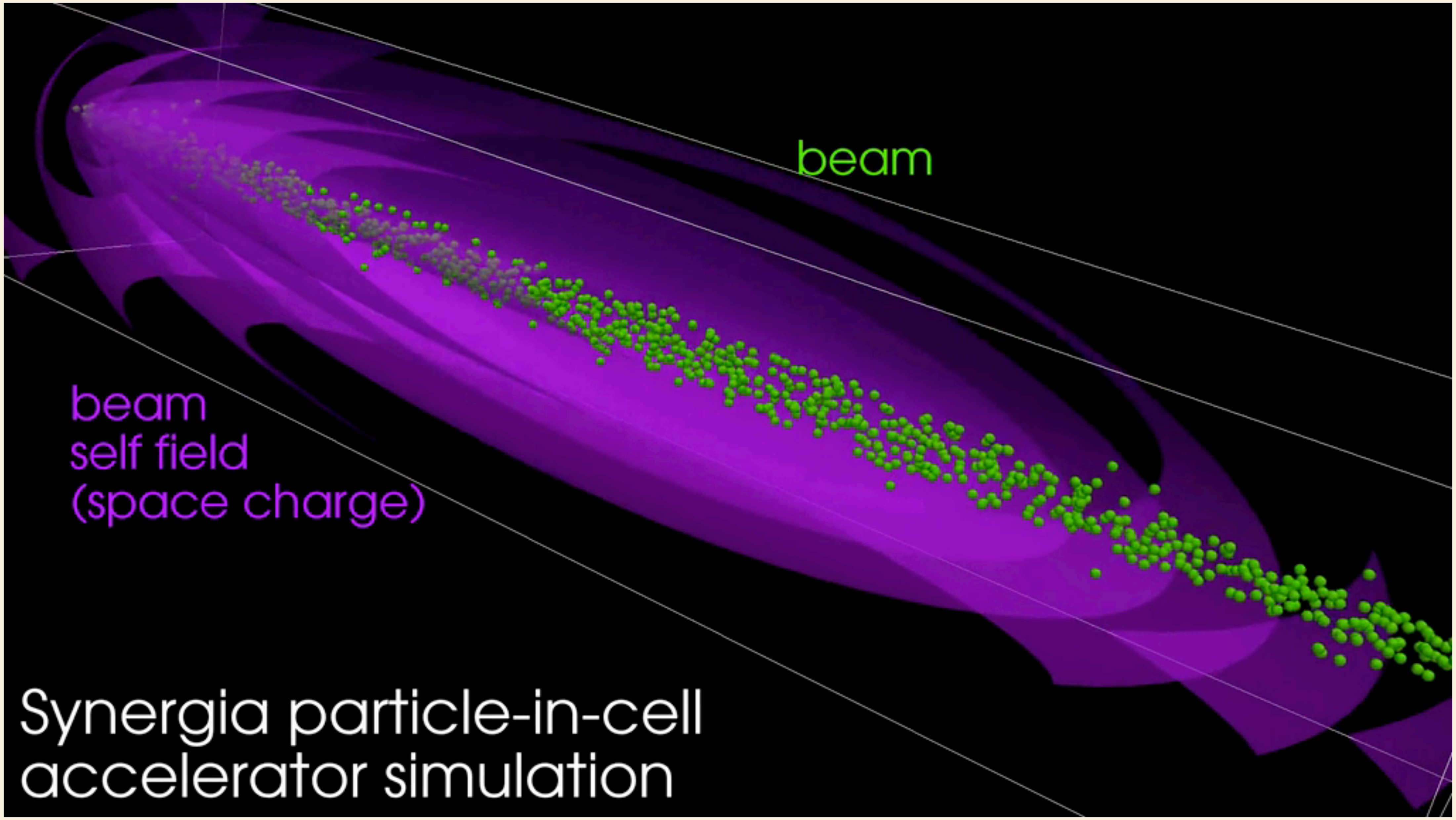
Case Study I

Simulating Particle Accelerators

Accelerators deal with lots of effects

We want super-intense beams (hard to do)

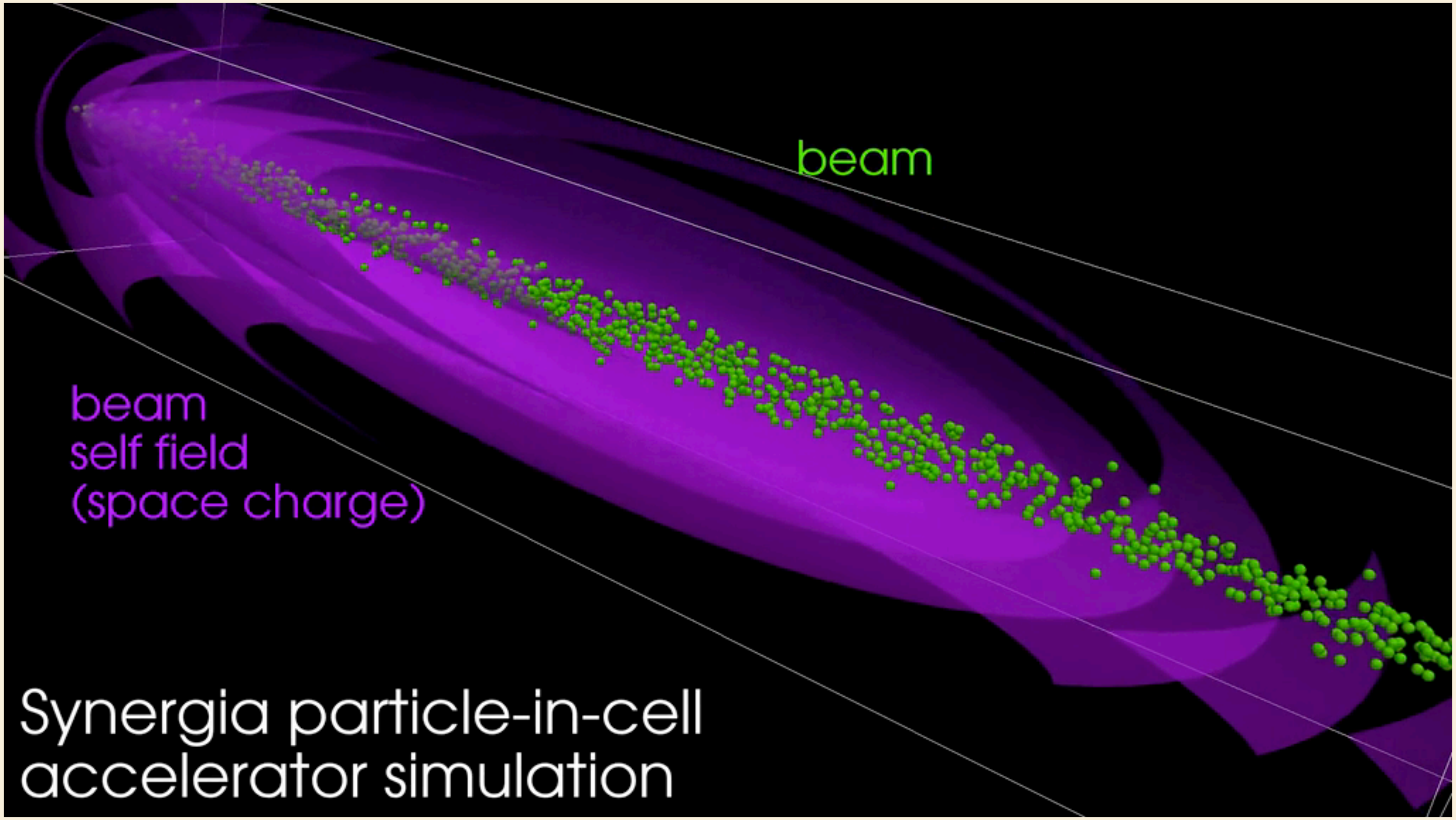
Beam instabilities must be understood



Accelerators deal with lots of effects

We want super-intense beams (hard to do)

Beam instabilities must be understood



Synergia

Fermilab
Accelerator Simulation Group

Computational beam dynamics for existing and planned accelerators

Essentially solving the Lorentz force equation for charged particles accelerating through a lattice of cavities and magnets (elements)

$$\frac{d\vec{p}}{dt} = q\vec{E} + q\vec{v} \times \vec{B}$$

1000s of elements (10s of types), thousands to millions of revolutions, 1-1000s of bunches with O(10^{12}) particles, 50-1000 steps/revolution

Synergia & ComPASS Funded by DOE SciDAC

IIT Physics is a collaborator

Simulation techniques

Internal (beam generated) fields are hard to simulate

Particle in Cell (PIC) methods:

- o Track individual motions of particles (embarrassingly parallel)
- o Collect particle density on grid (cells) and solve PDEs
  Tightly coupled - requires work to parallelize

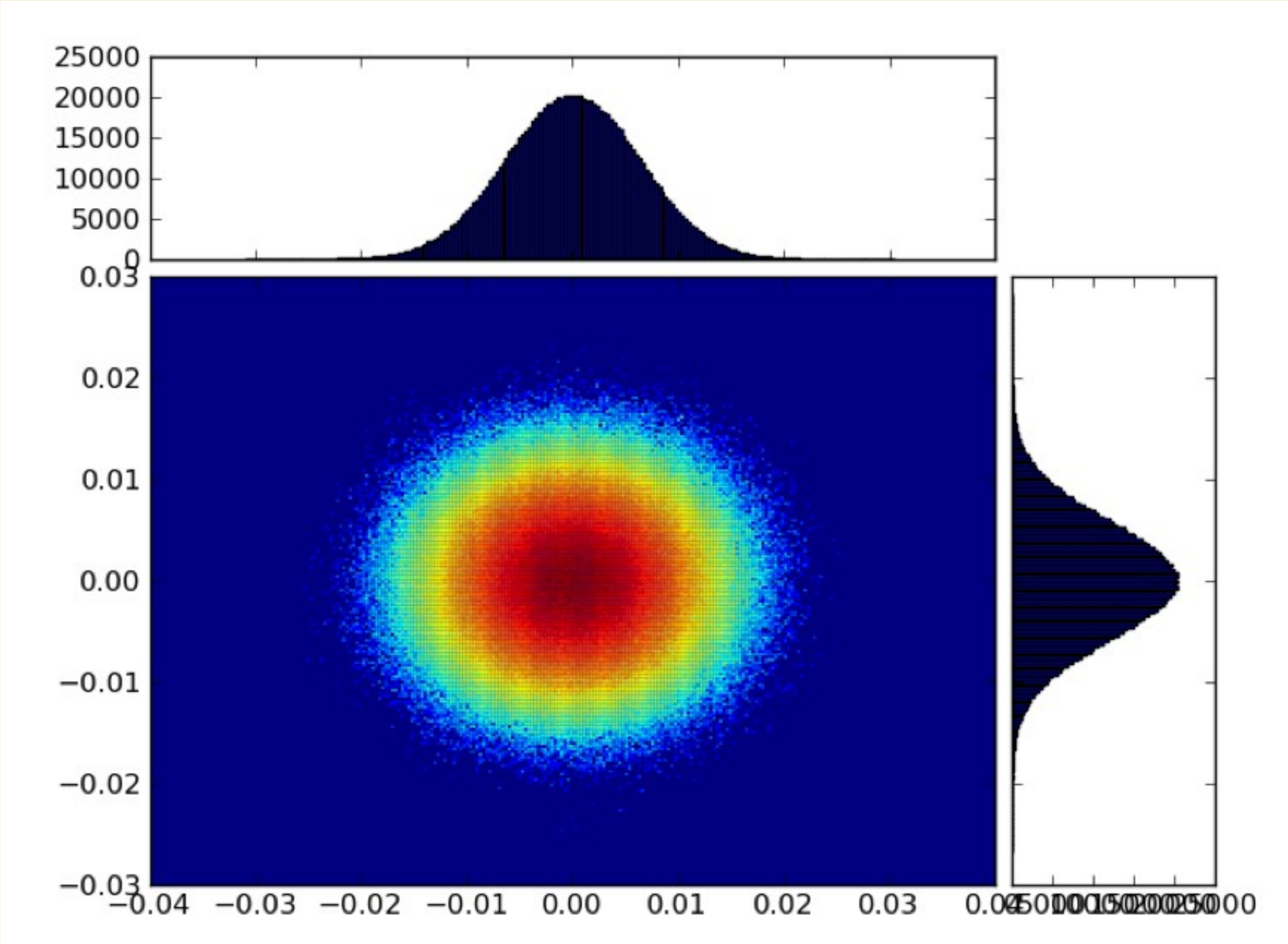
Runs on laptops to Supercomputers (good for tight coupling)

Difficulties:
Typically 64^3 - 128^3 grid (2×10^5 – 2×10^6 degrees of freedom)
10^5 – 10^8 time steps

External fields can cause particles to move over many grid cells in a single step - complicated point-to-point communication

What do we measure?

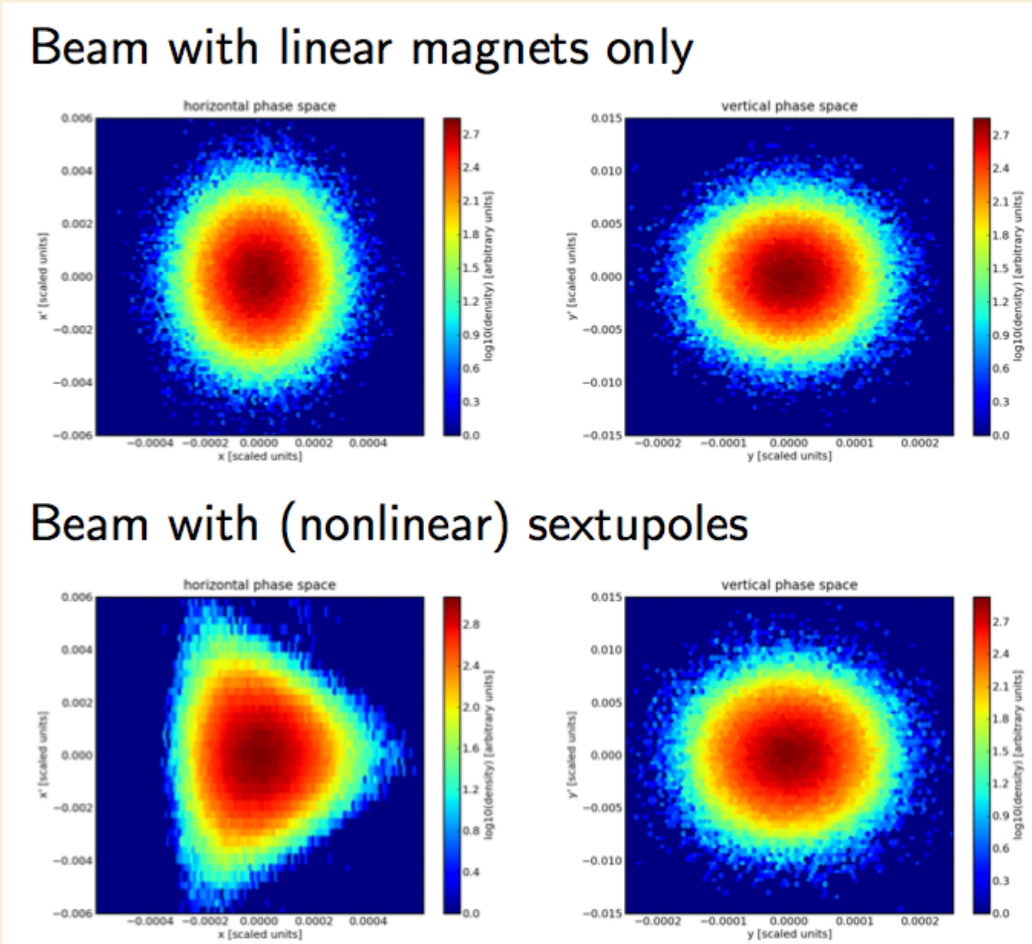
Transverse Momentum



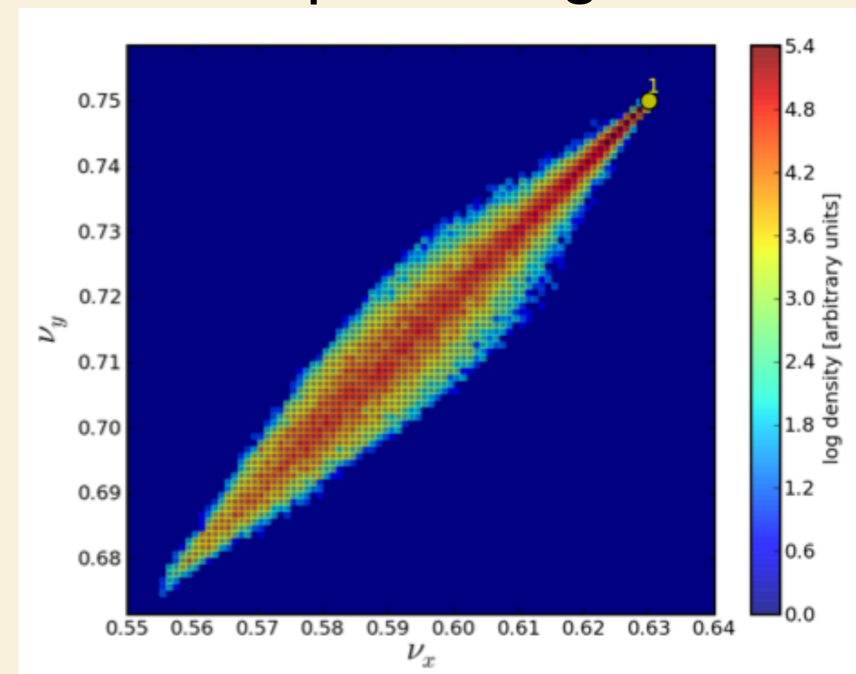
Transverse Position

The "Phase Space" - want the beam focused to the center

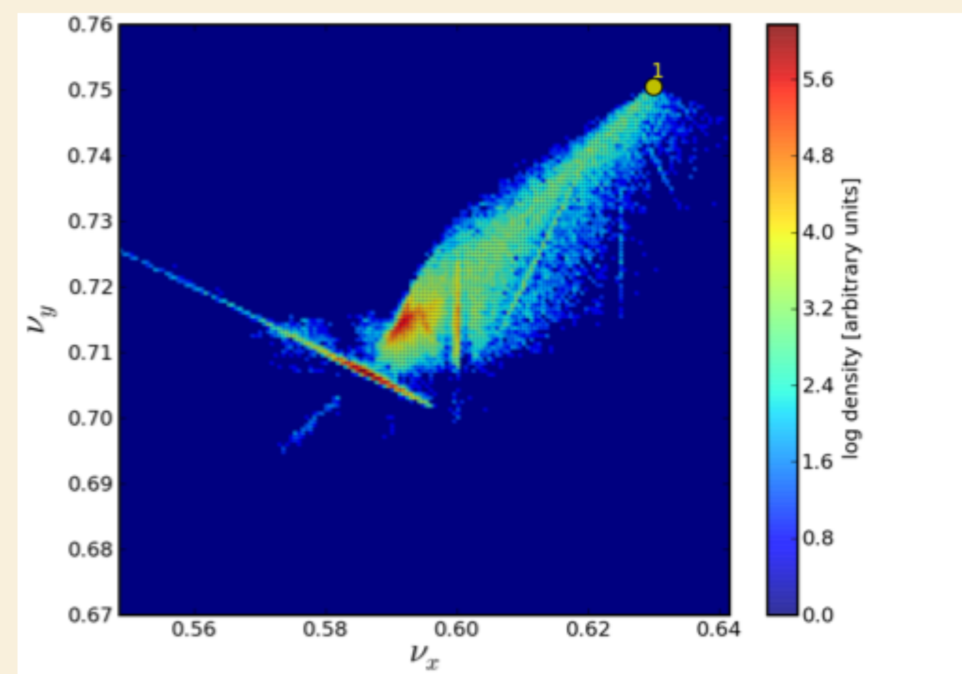
Accelerator design



Space charge



Space charge & sextuple magnets

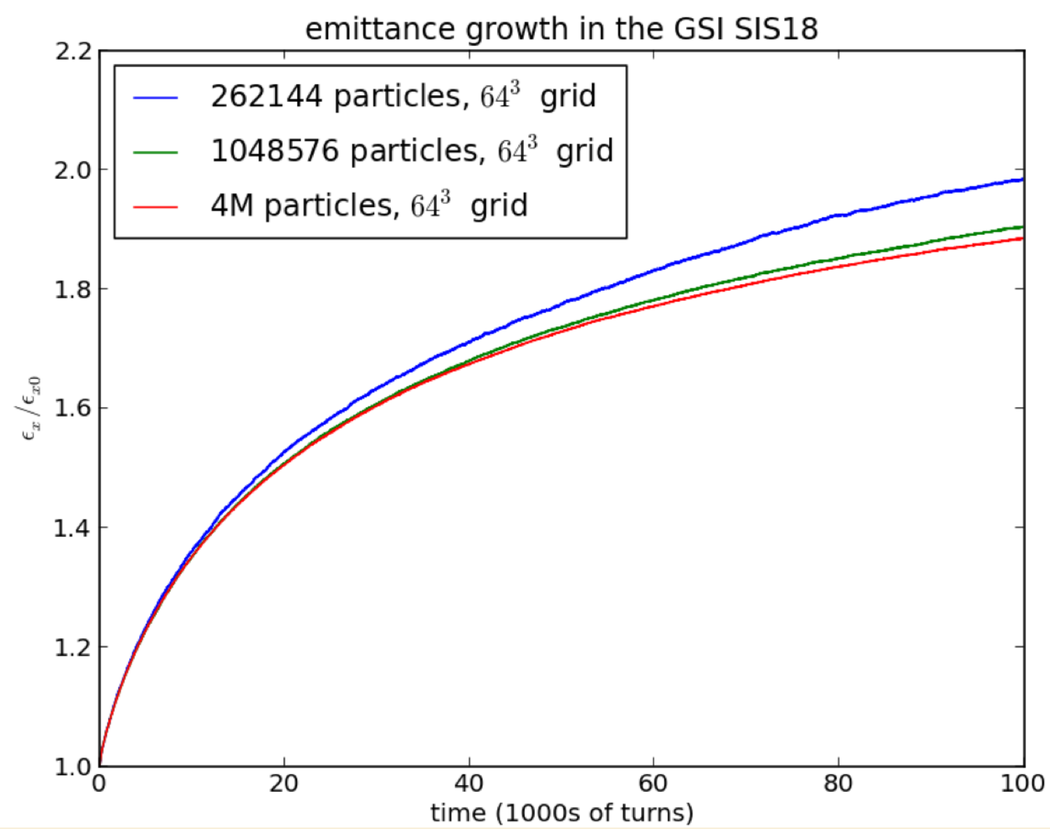


Simulations show phase space in different scenarios

See resonances on lower right plot – unacceptable losses result

“Big” simulations

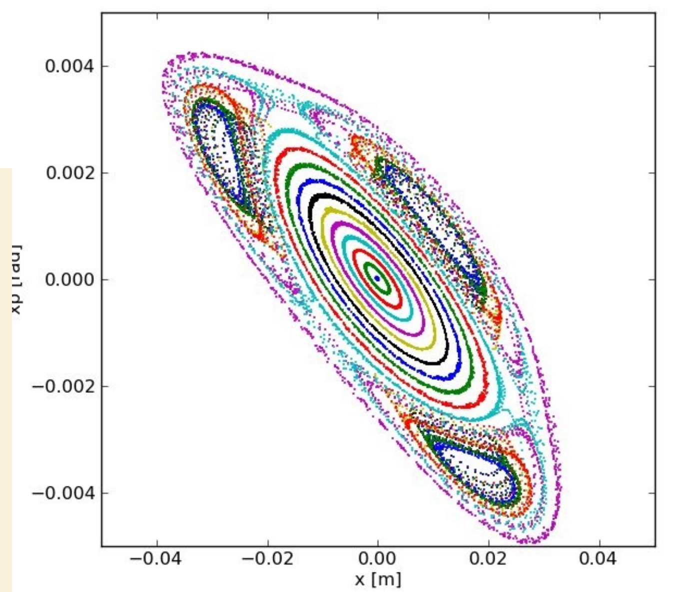
Longest accelerator simulation ever performed (Two weeks on 120 Intel cores)



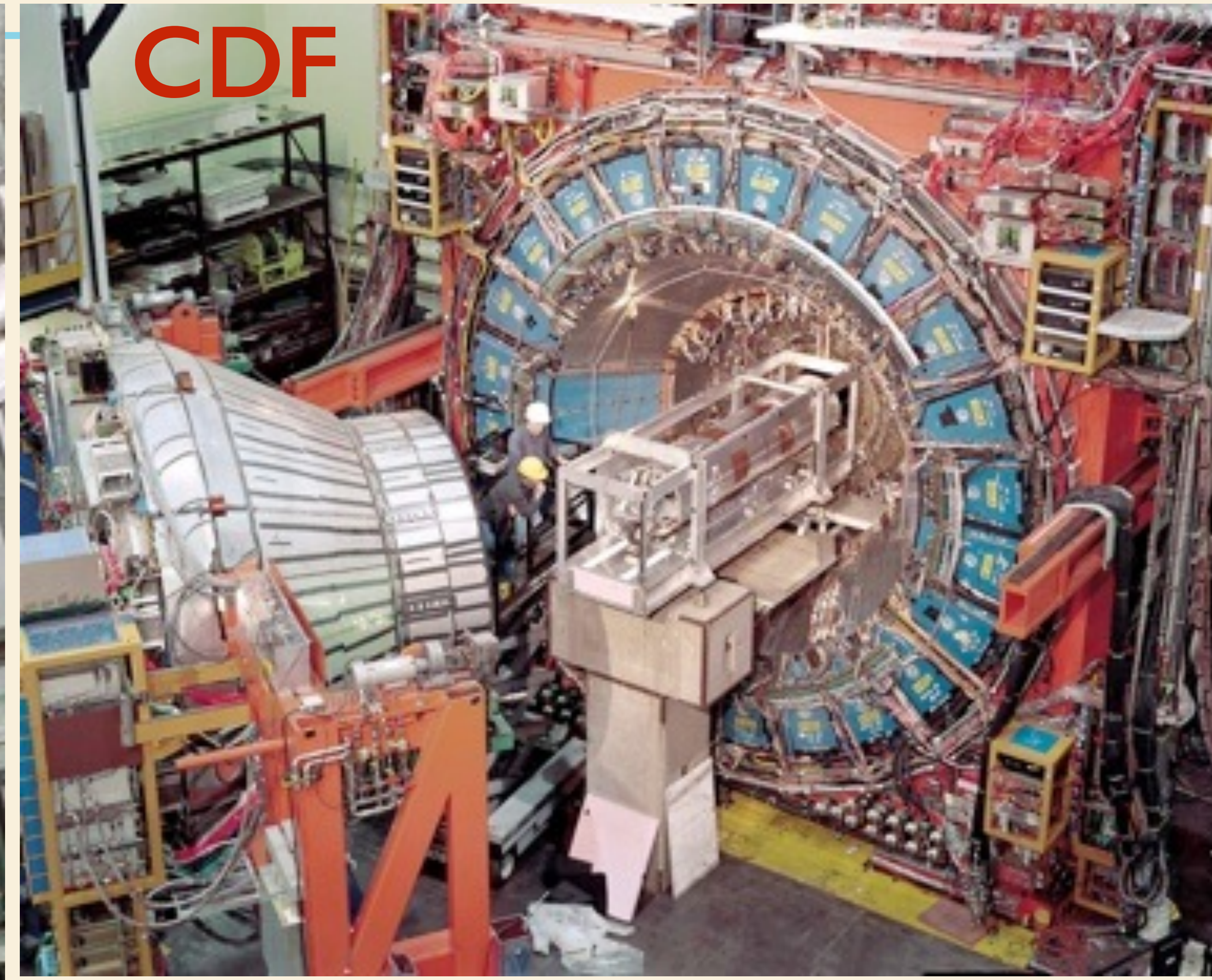
71 steps/turn
7,100,000 steps
4,194,304 particles
29,779,558,400,000 particle-steps
1,238,158,540,800,000 calls to “drift”

Yes, that's over a quadrillion

Can reproduce phase-space islands of actual accelerator



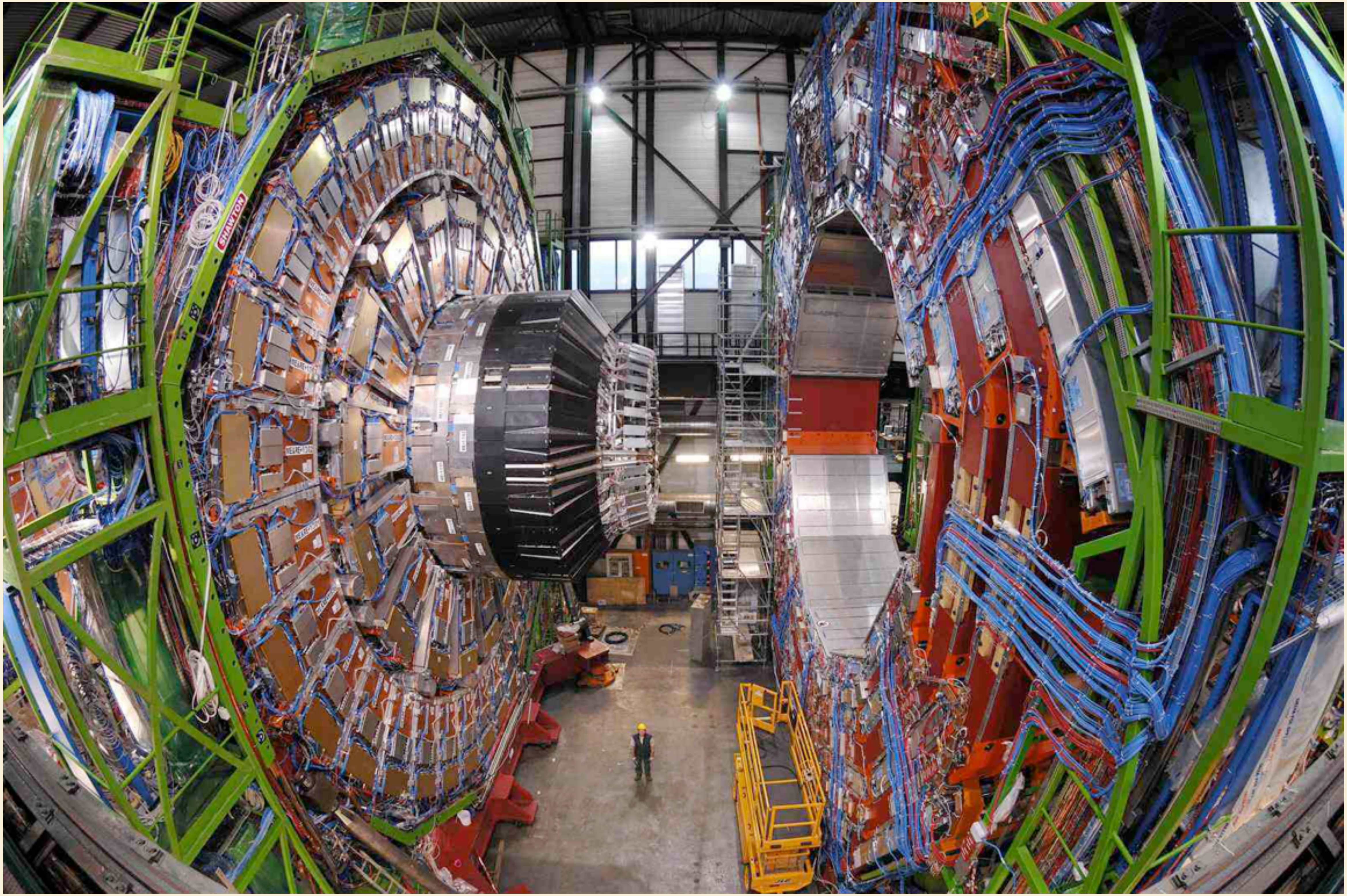
On To Detectors…



<u>Detectors</u>

We use the very big to learn about the very small!

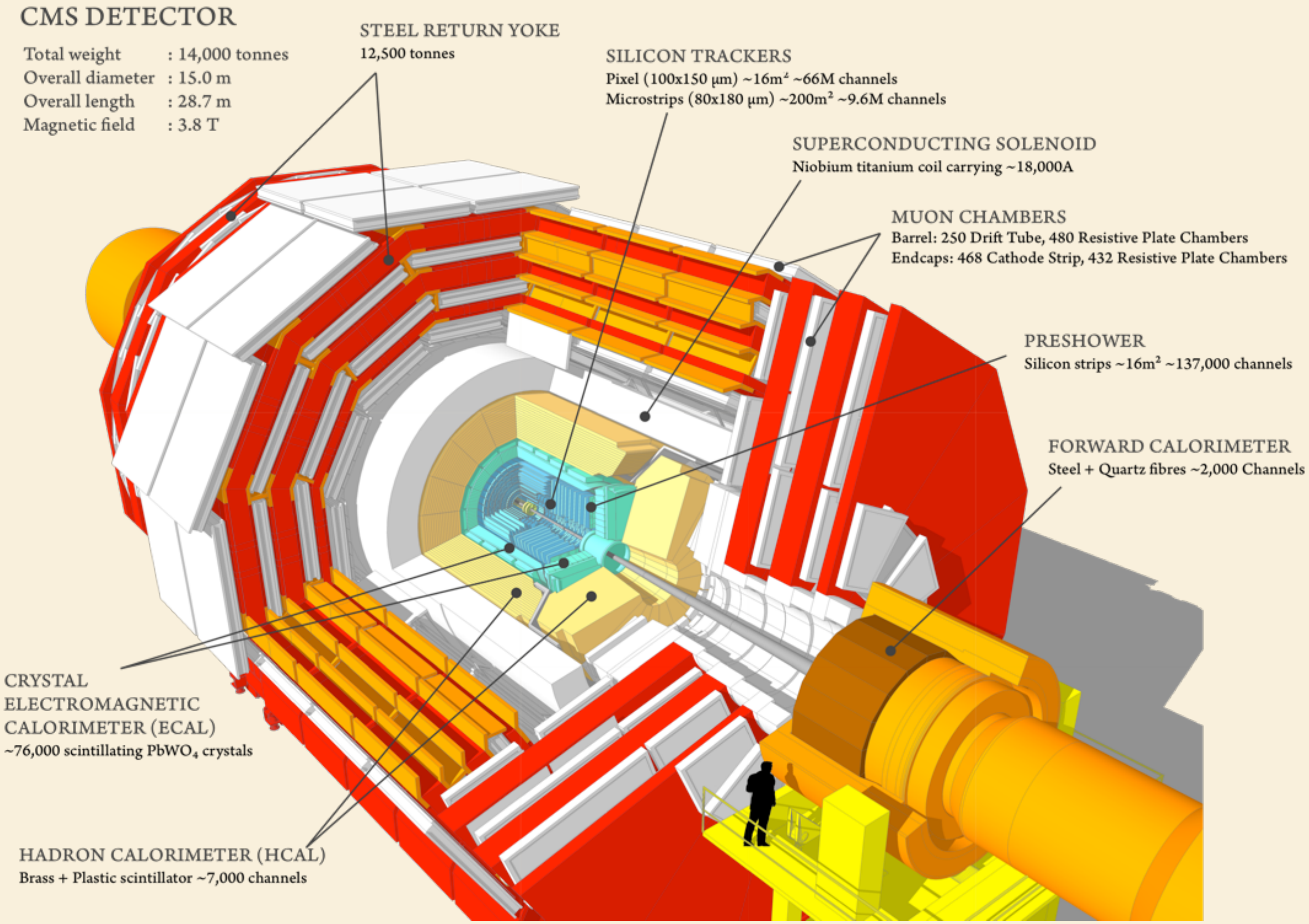
The CMS Detector at CERN



Finding the Higgs @ CMS

Large Fermilab involvement on CMS

Detector is built around collision point



Records flight path of every particle leaving collision

100M individual measurements (channels)

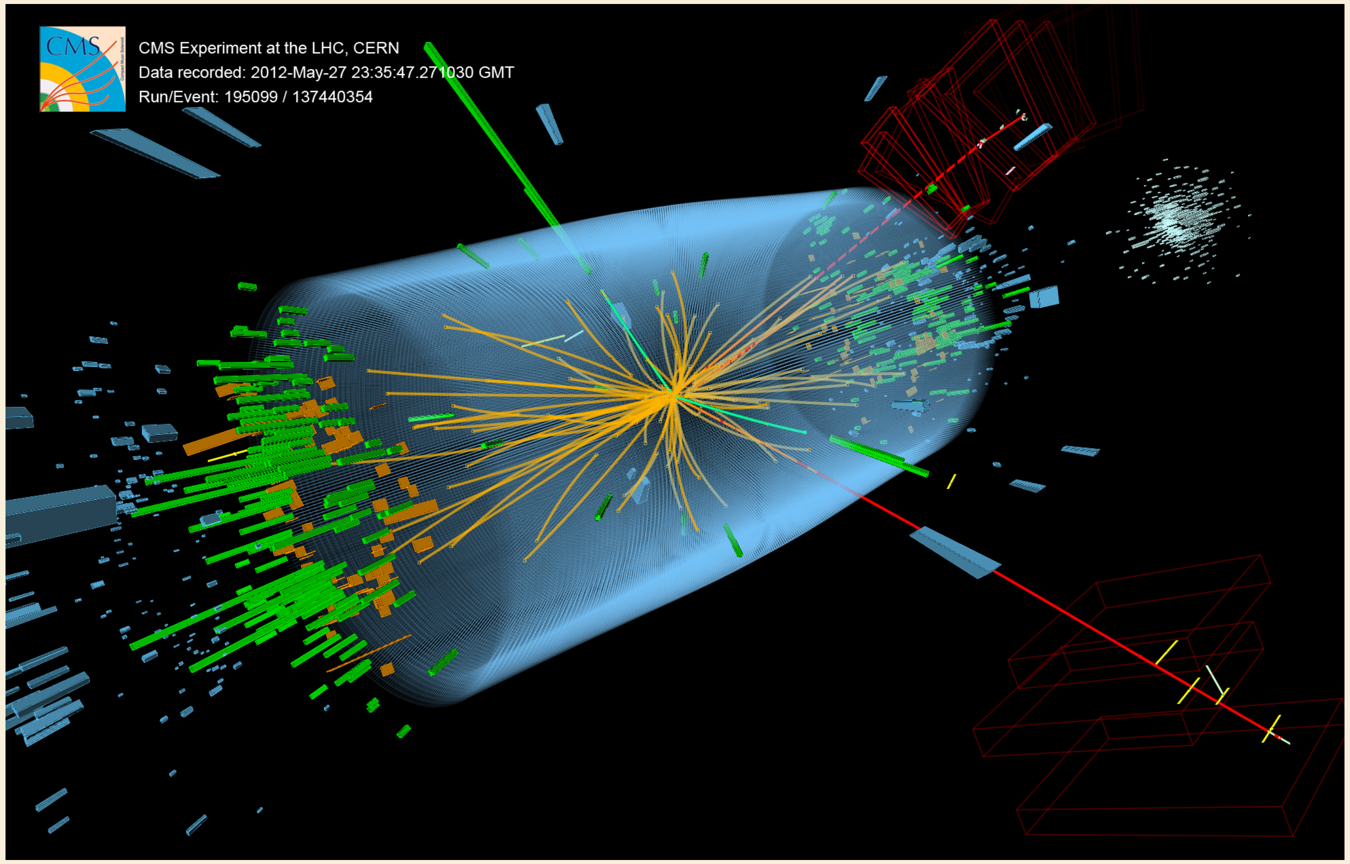
All measurements for a collision is an "event"

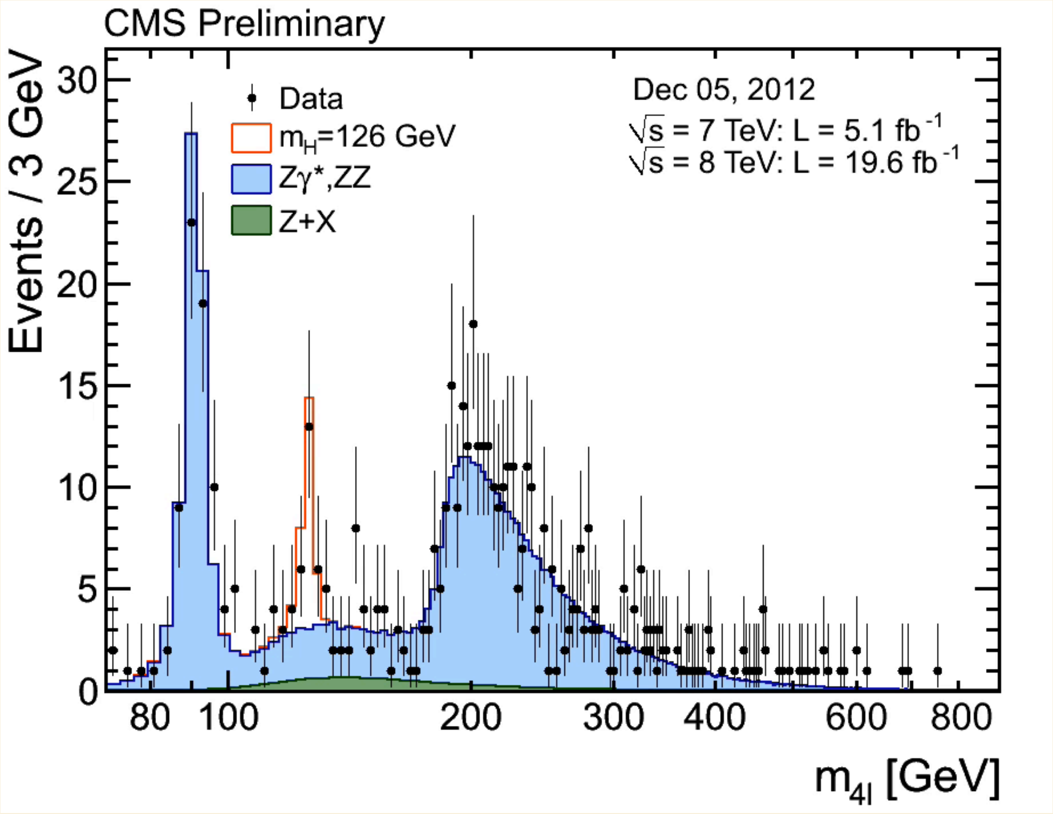
Events are independent

Discovering the Higgs

Need lots of energy to "knock a Higgs particle out of space"

It almost immediately decays into other particles





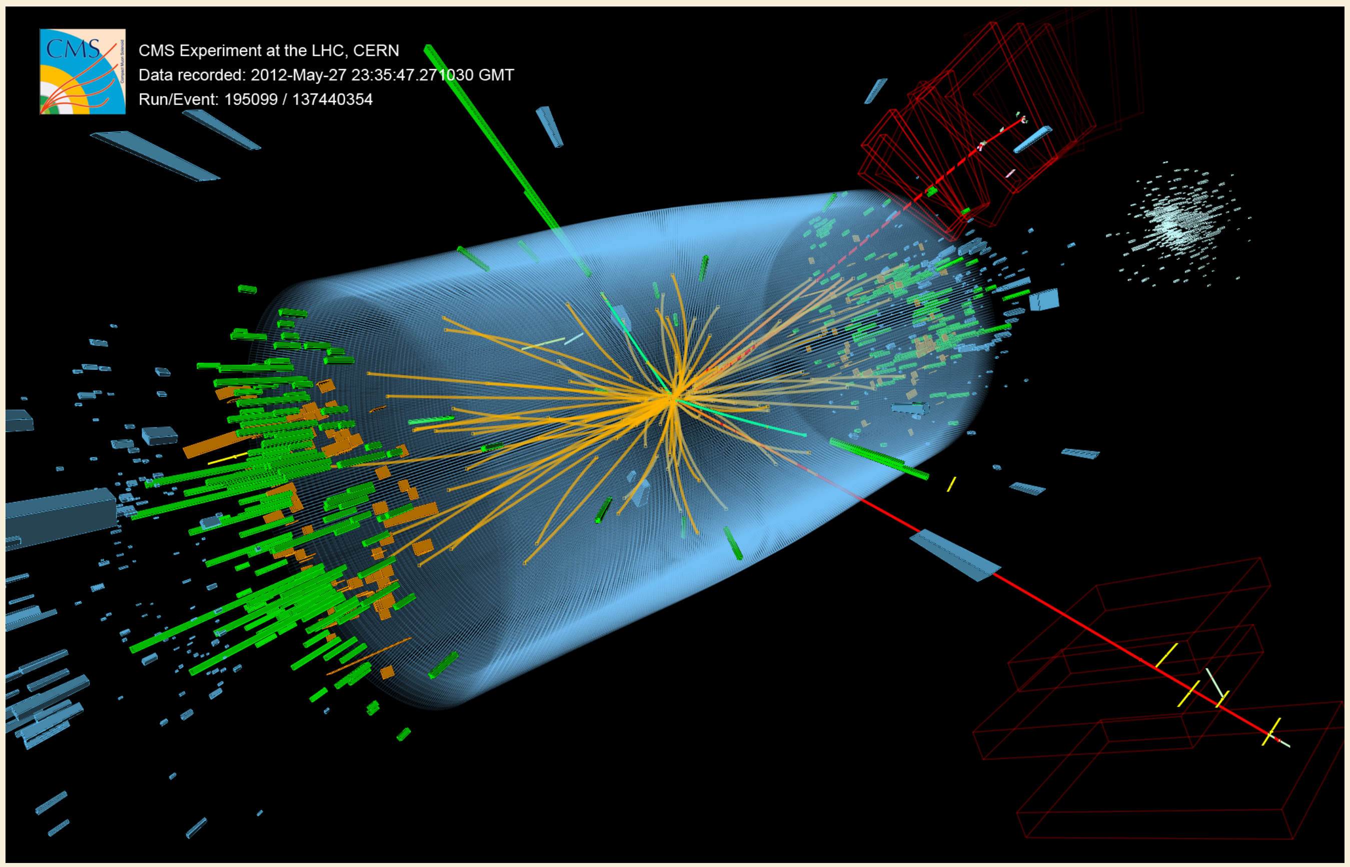
Find ~100 events from billions [600M events/s, Keep 600/s]

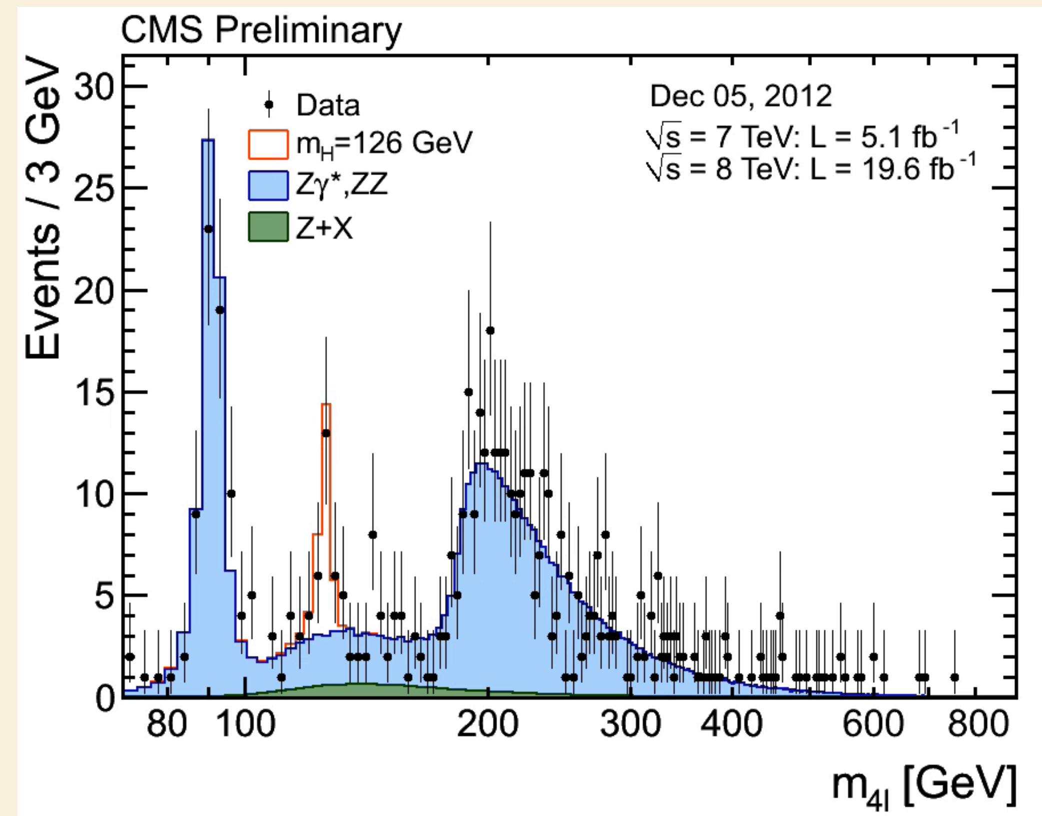
Want to end up with a clear signal Note simulations above

Discovering the Higgs

Need lots of energy to “knock a Higgs particle out of space”

It almost immediately decays into other particles





Find ~100 events from billions [600M events/s, Keep 600/s]

Want to end up with a clear signal   Note simulations above

Fermilab

Case Study II

Simulations in collider experiments

Simulations are key

Simulations are crucial for achieving and interpreting a result

Simulate signal and background

Understand efficiency and biases of algorithms

Understand the acceptance of the detector

Understand fluctuations

Understand detector response

How well the simulations mimic reality are important for systematic uncertainties

Flow of simulations

Event generator → **Geant Simulation** → **Detector response** → **Recon-struction**

Written by Phenomenologists

Pythia, Herwig, Sherpa, MadGraph, PhoJet, Dipsy, CompHep, MC@NLO, MCFM, POWHEG, TrueNoir, ResBos, …

Simulates particles passing through the detector

Determines energy deposits

May be replaced by Fast MC

Simulates the detector response to those energy deposits (what does the detector do?)

Output is meant to look like raw data

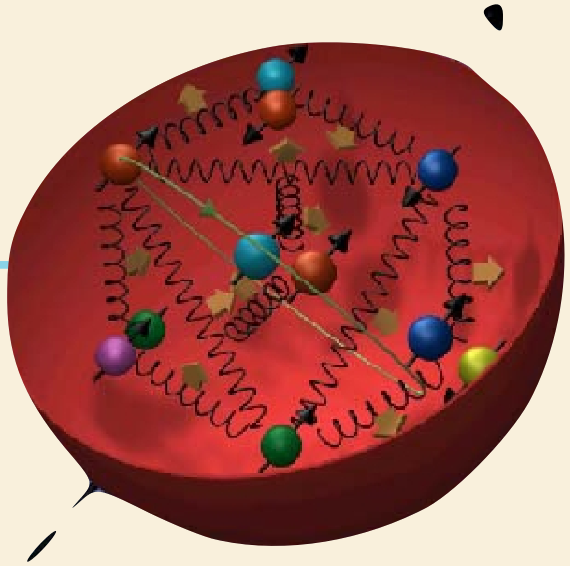
Converts raw data (real or simulated) into physics objects for analyses

Written by:

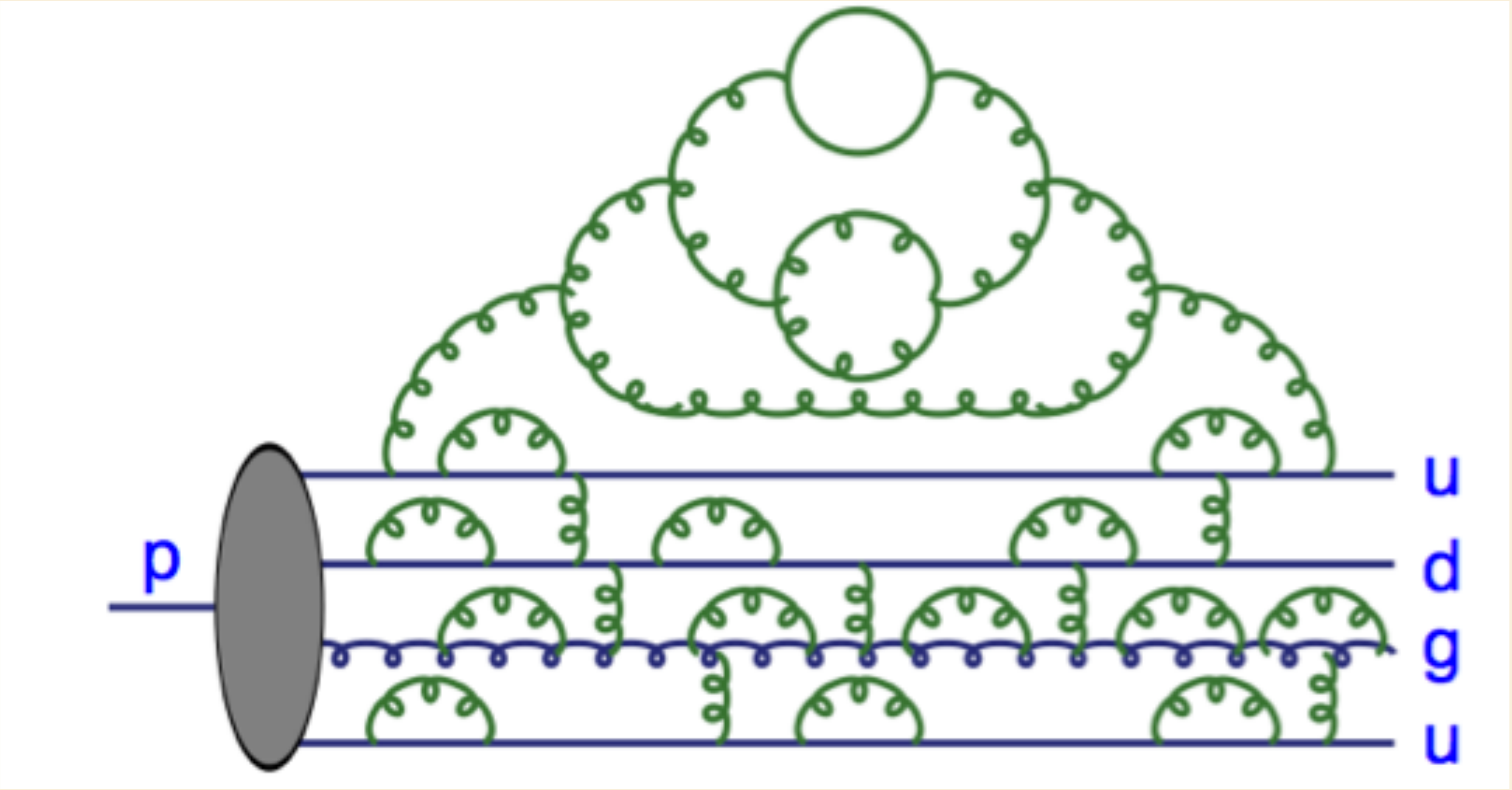
- Many theorists (blue)
- Geant4 collab (orange)
- The experiment (green)

Fermilab

Event generators

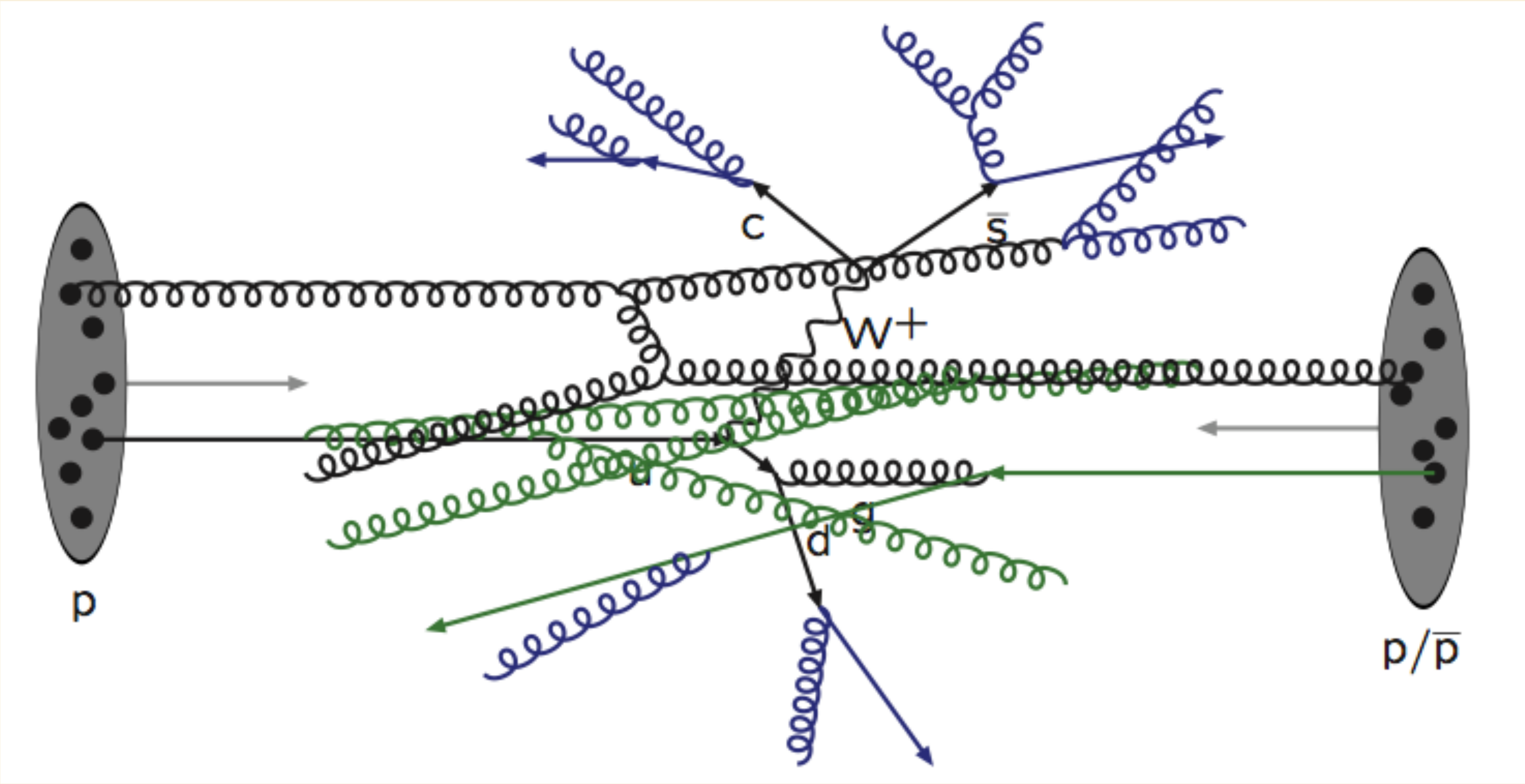


The proton is a complicated beast



From Torbjorn Sjostrand

The collision is super-complicated



From Torbjorn Sjostrand

Hard scattering and decays

Initial and final state radiation

Soft interactions with their radiation

Generally trying to solve…

$$\mathcal{P}(\mathbf{p}_i^{\text{vis}}|\alpha) = \frac{1}{\sigma(\alpha)} \sum_{k,l} \int dx_1 dx_2 \frac{f_k(x_1) f_l(x_2)}{2 s x_1 x_2} \times \left[\prod_{j \in \text{inv.}} \int \frac{d^3 p_j}{(2\pi)^3 2E_j} \right] |\mathcal{M}_{kl}(p_i^{\text{vis}}, p_j; \alpha)|^2$$

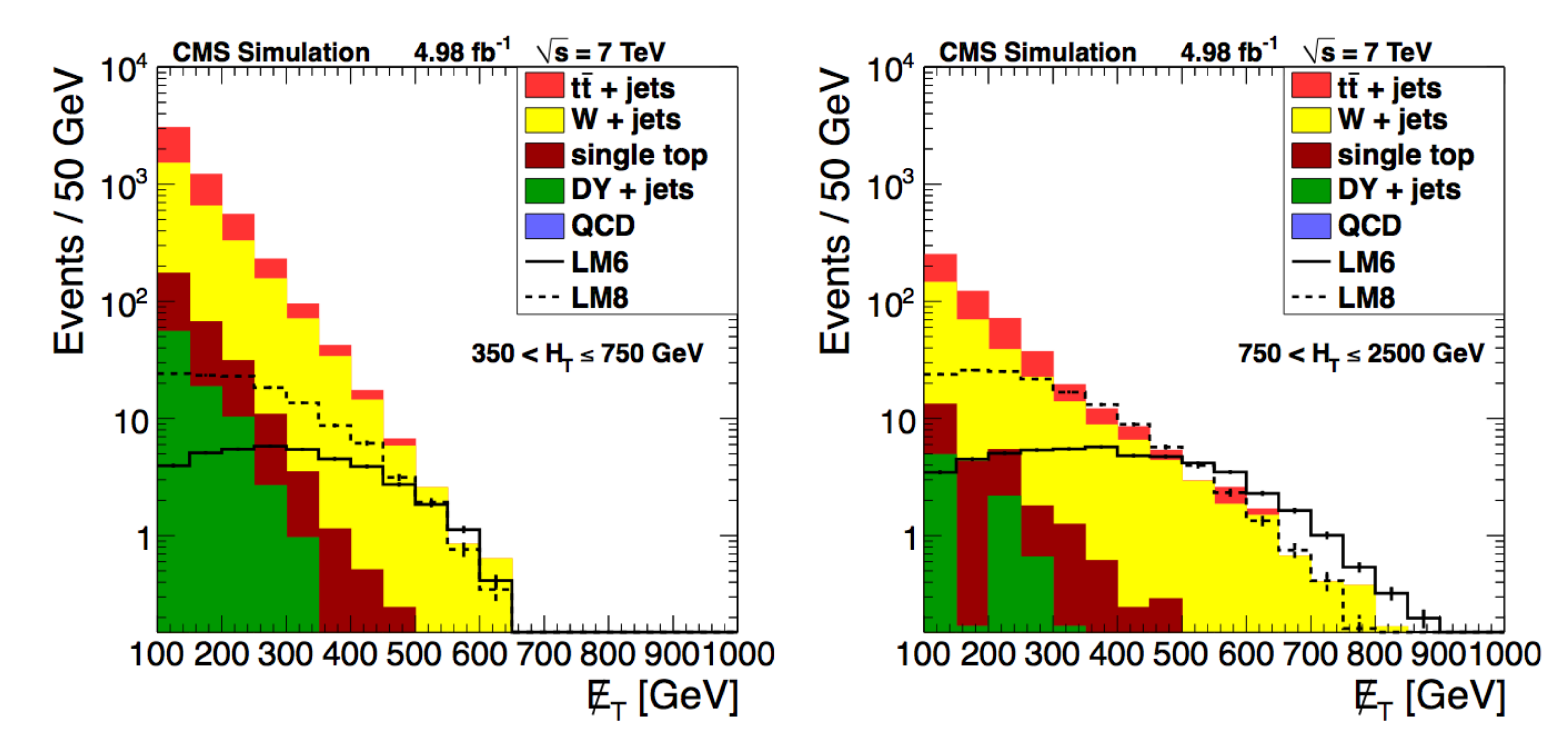
Likelihood for a given event with measured momentum with underlying model parameters

The Matrix element contains the interesting physics

Use Monte Carlo methods to solve (written by experts)

An example from SUSY

Exploring signal to background

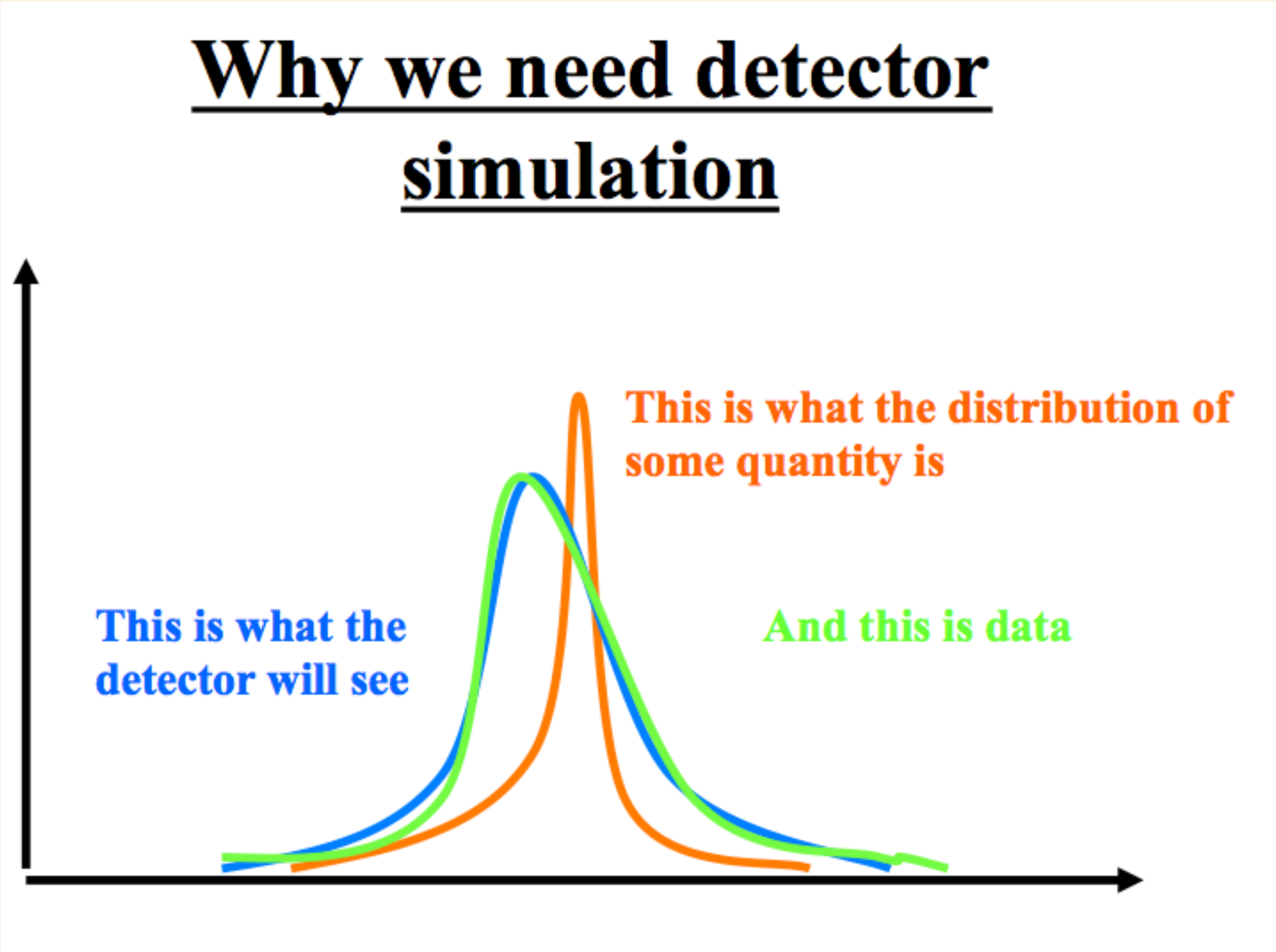


From Phys. Rev. D 87 (2013) 052006

Case Study III

Detector simulation

Detectors sculpt distributions

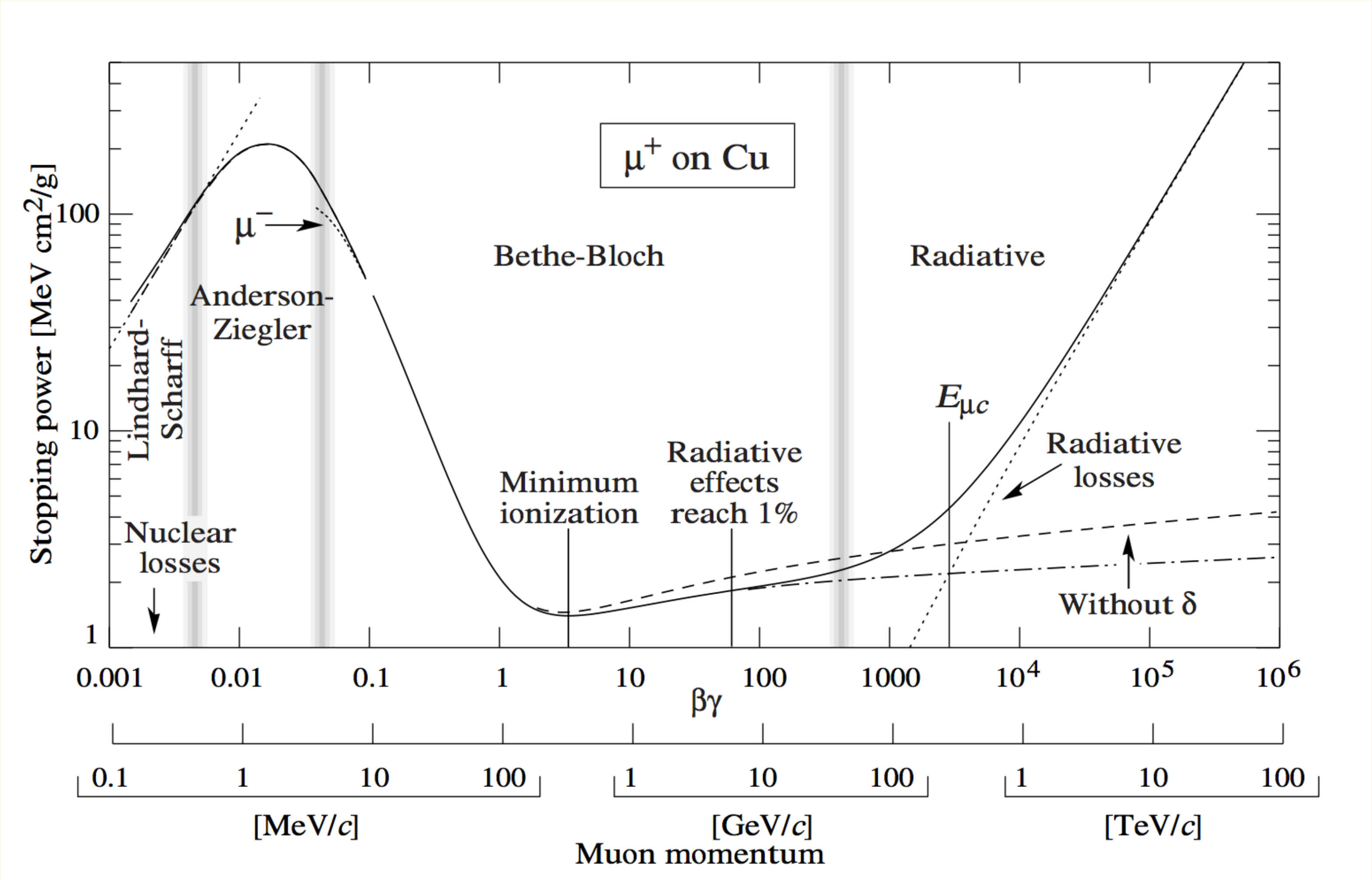


From Tomasz Wlodek

Understand particles in a detector

Particles can loose energy

$$-\left\langle \frac{dE}{dx} \right\rangle = Kz^2 \frac{Z}{A} \frac{1}{\beta^2} \left[\frac{1}{2} \ln \frac{2m_e c^2 \beta^2 \gamma^2 T_{\max}}{I^2} - \beta^2 - \frac{\delta(\beta\gamma)}{2} \right]$$

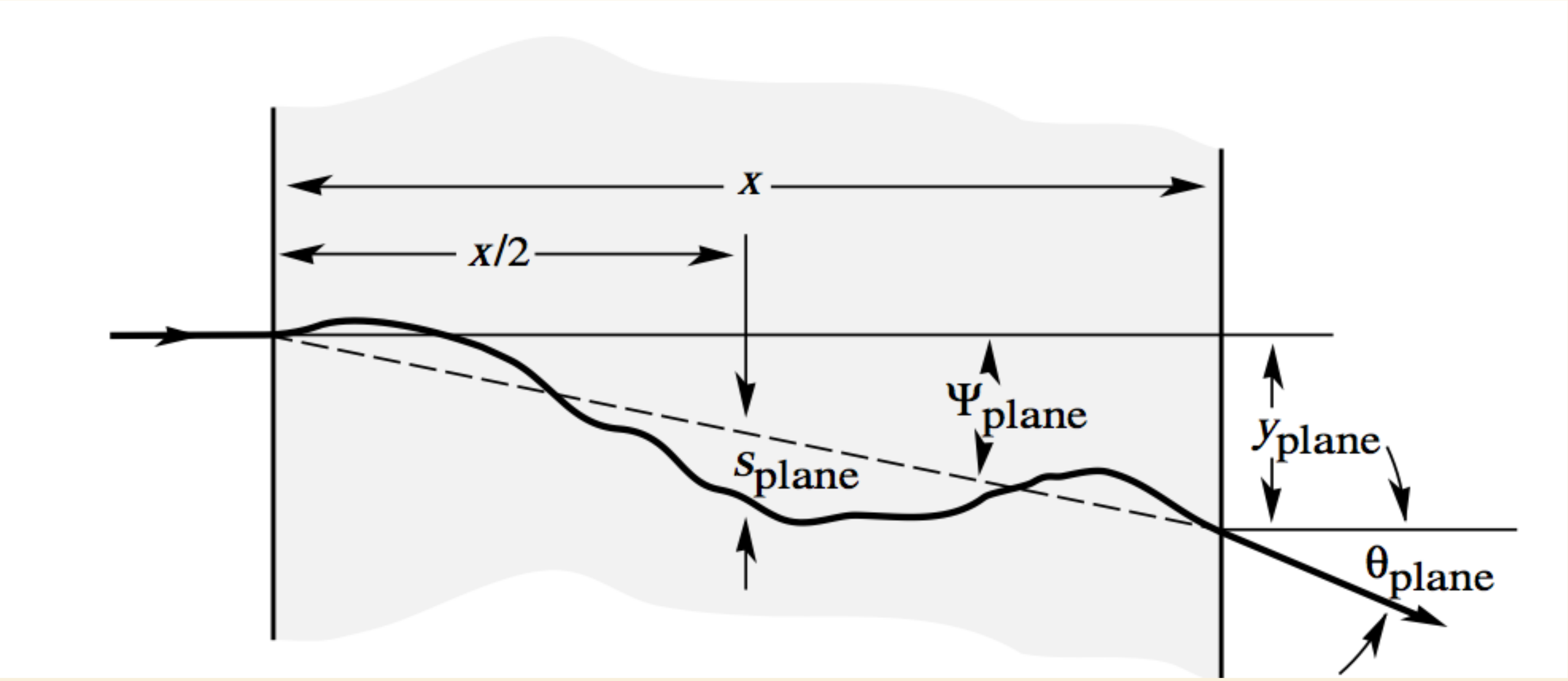


From Particle Data Group 2010

Understand particles in a detector

And can multiple scatter

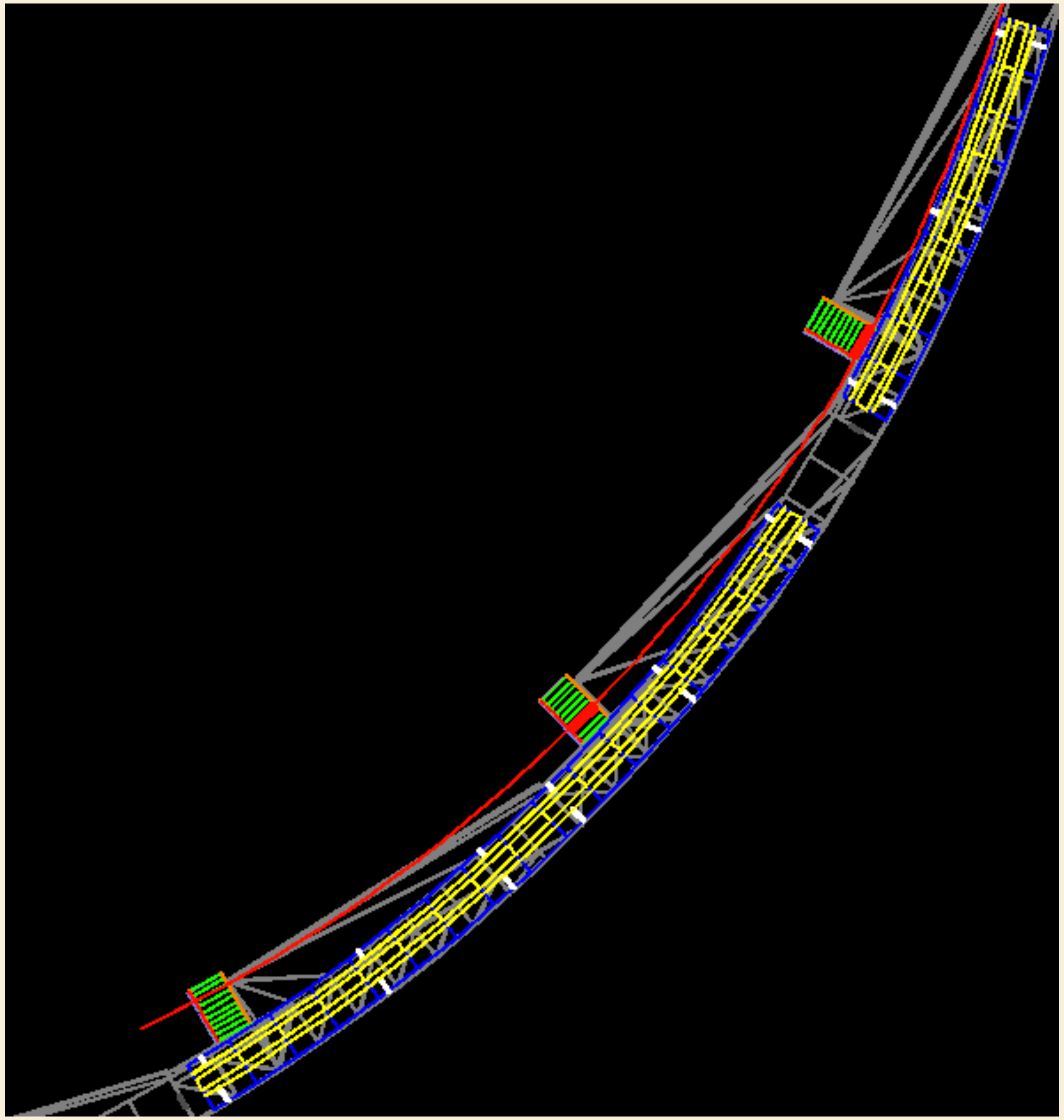
$$\theta_0 = \frac{13.6\ \text{MeV}}{\beta c p}\, z \sqrt{x/X_0}\Big[1 + 0.038 \ln(x/X_0)\Big]$$

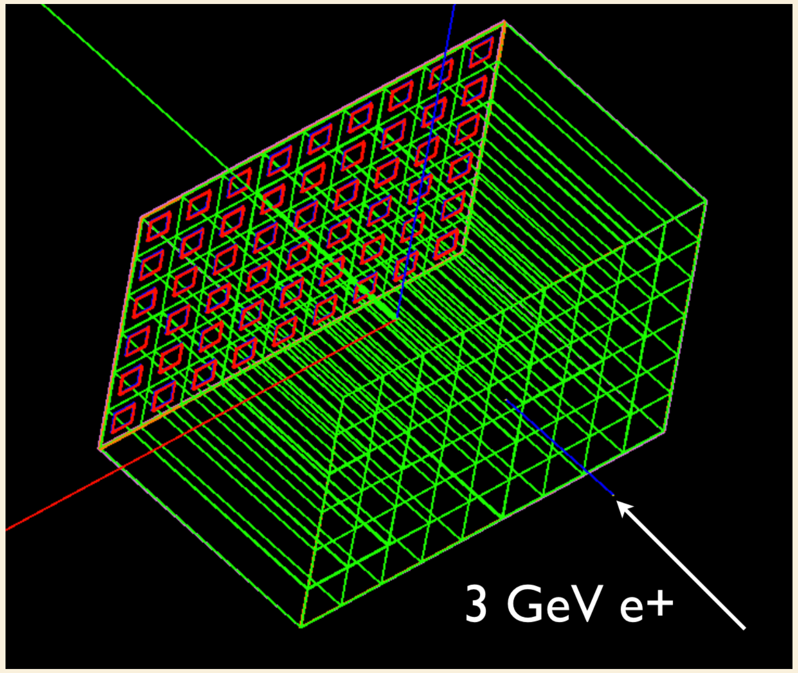


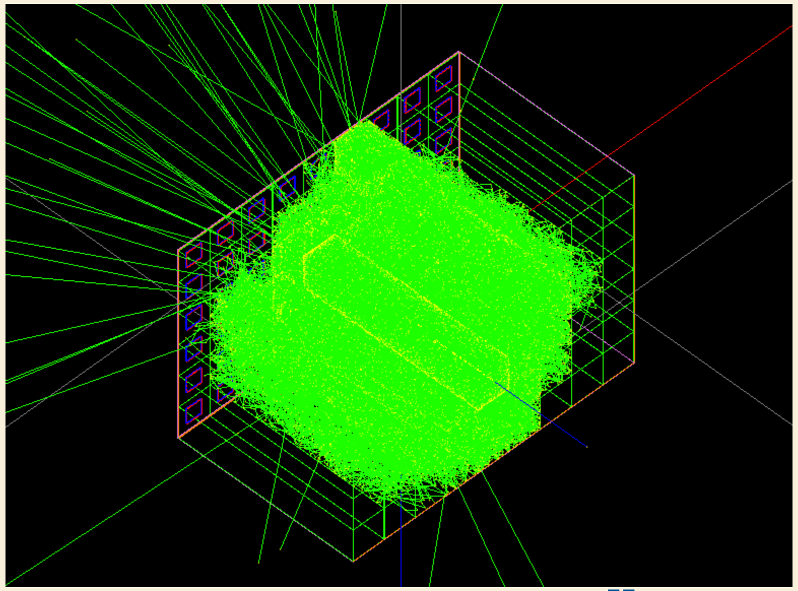
From Particle Data Group 2010

Geant 4

Started 20 years ago in C++
Knows about passage through matter and E&M fields

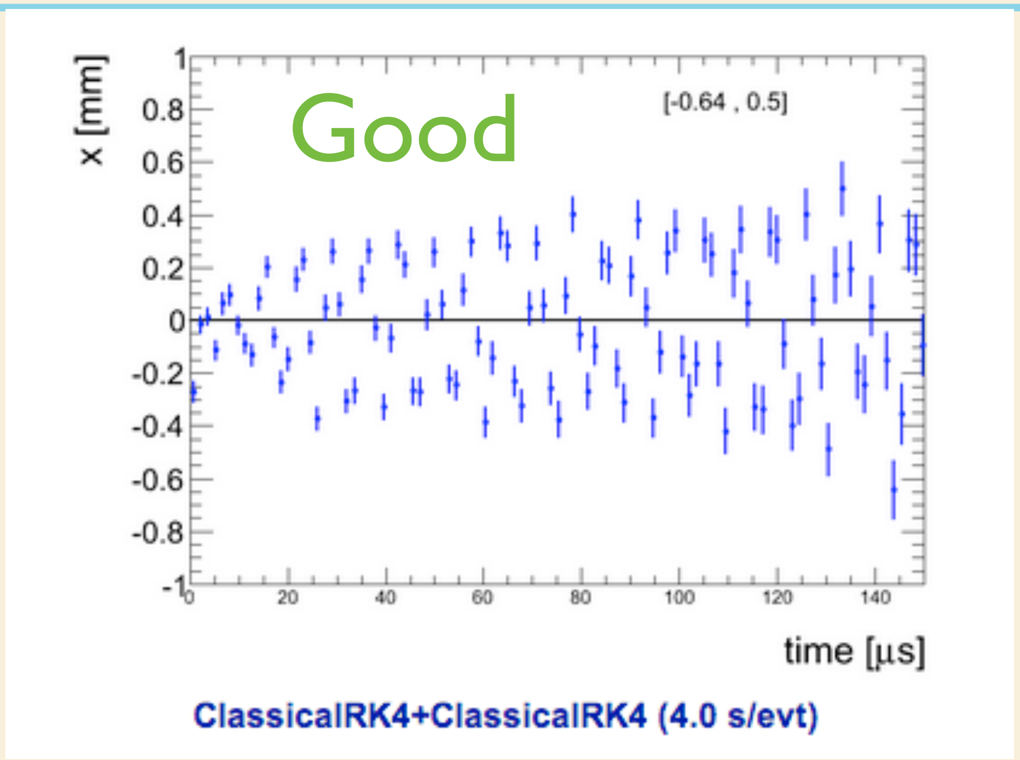


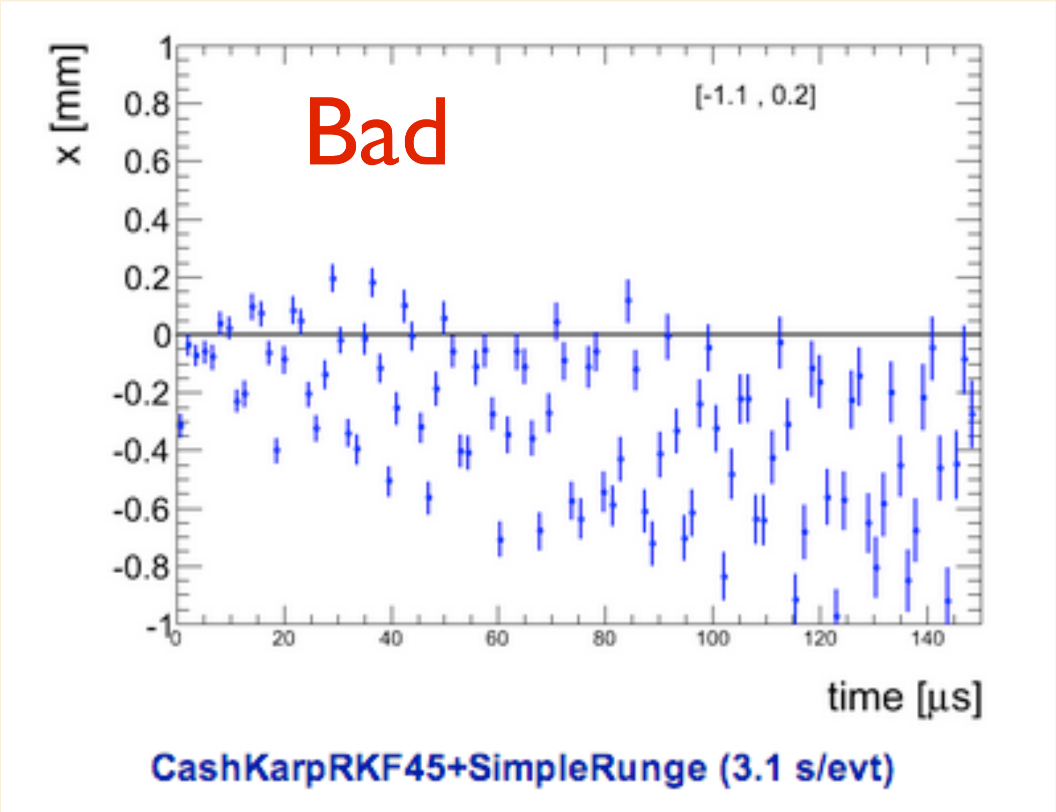


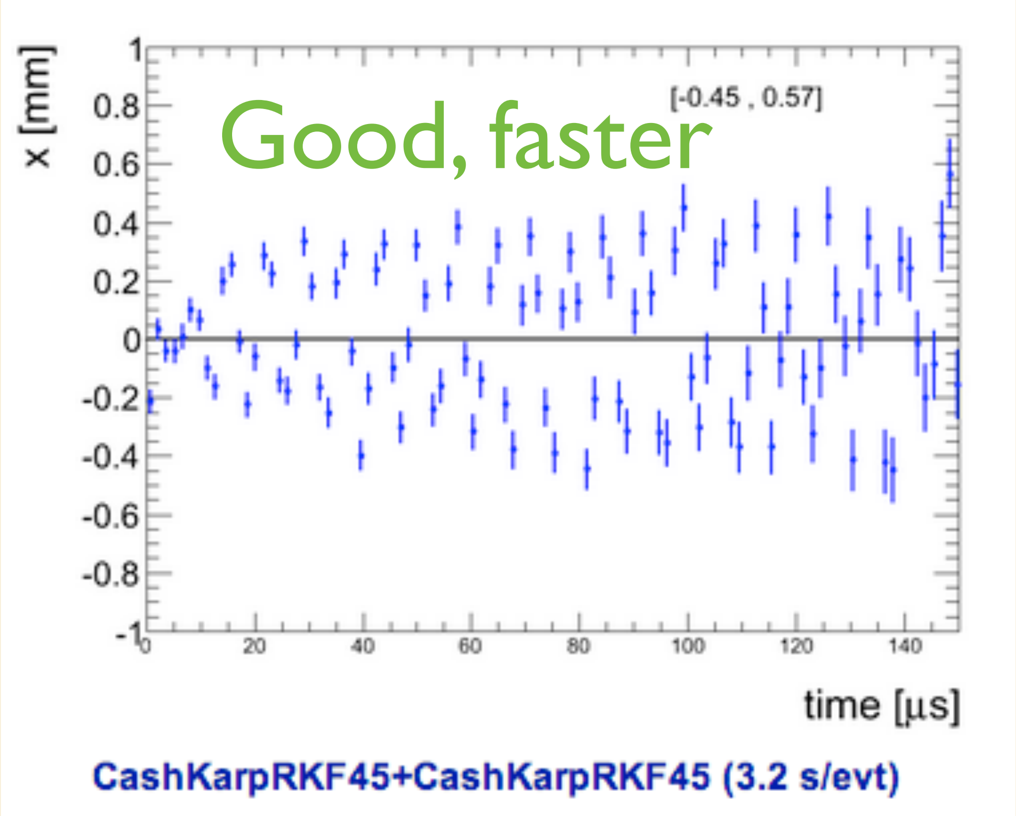


Must be careful

Must choose a robust integrator







Case Study IV

Mechanics of Running Simulations

How do we run these?

Compute Farm of worker nodes

Tape Robot

Disk Storage

Events are independent, embarrassingly parallel at the event level; no inter-processor communication required

So High Performance Computers (HPC; Supercomputers) are not necessary. Instead rely on *High Throughput* Computing

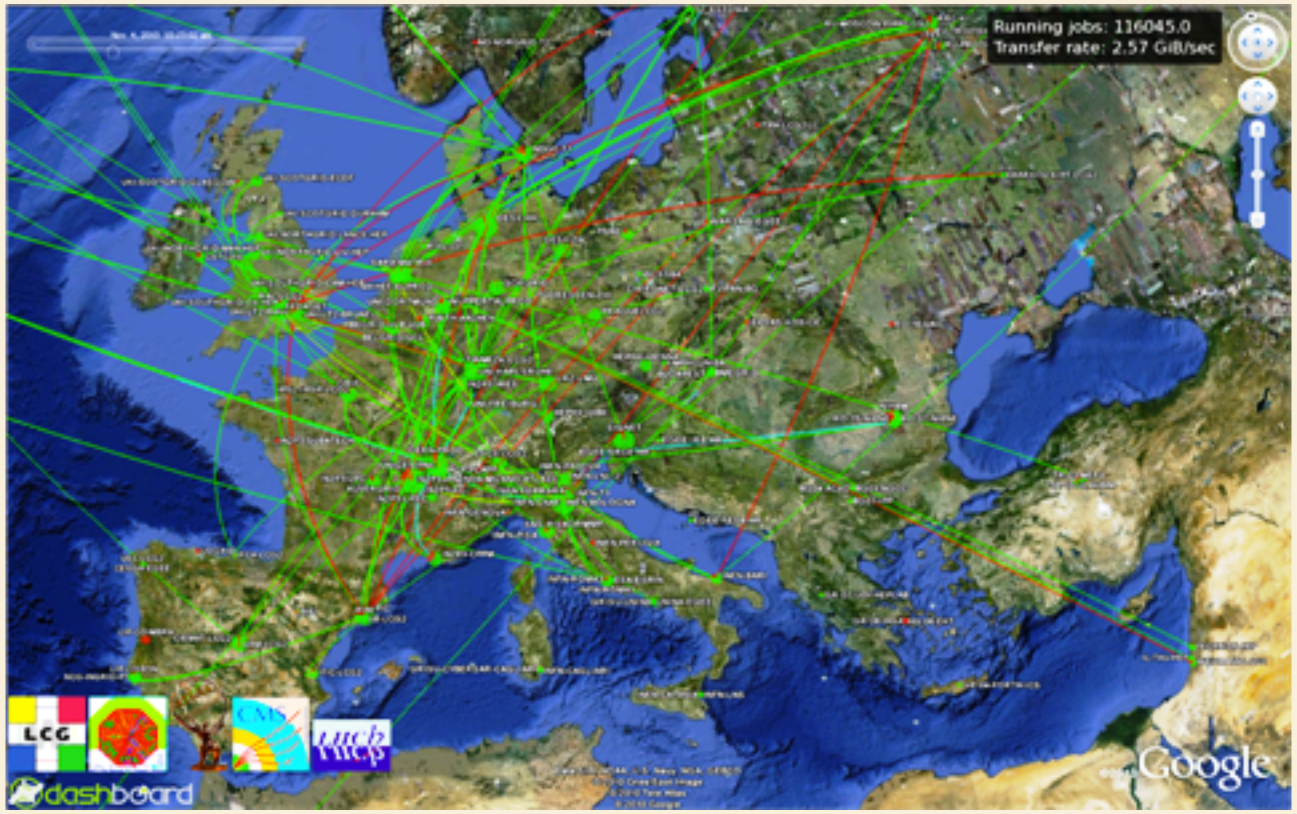
Run on farms of commodity linux nodes; typically 64 cores with 2 GB/core

Resources

Fermilab (for CMS and IF) has 27K batch slots (one core/slot), 84 PB on tape, 18 PB disk cache
- Sized for mainly average, not peak

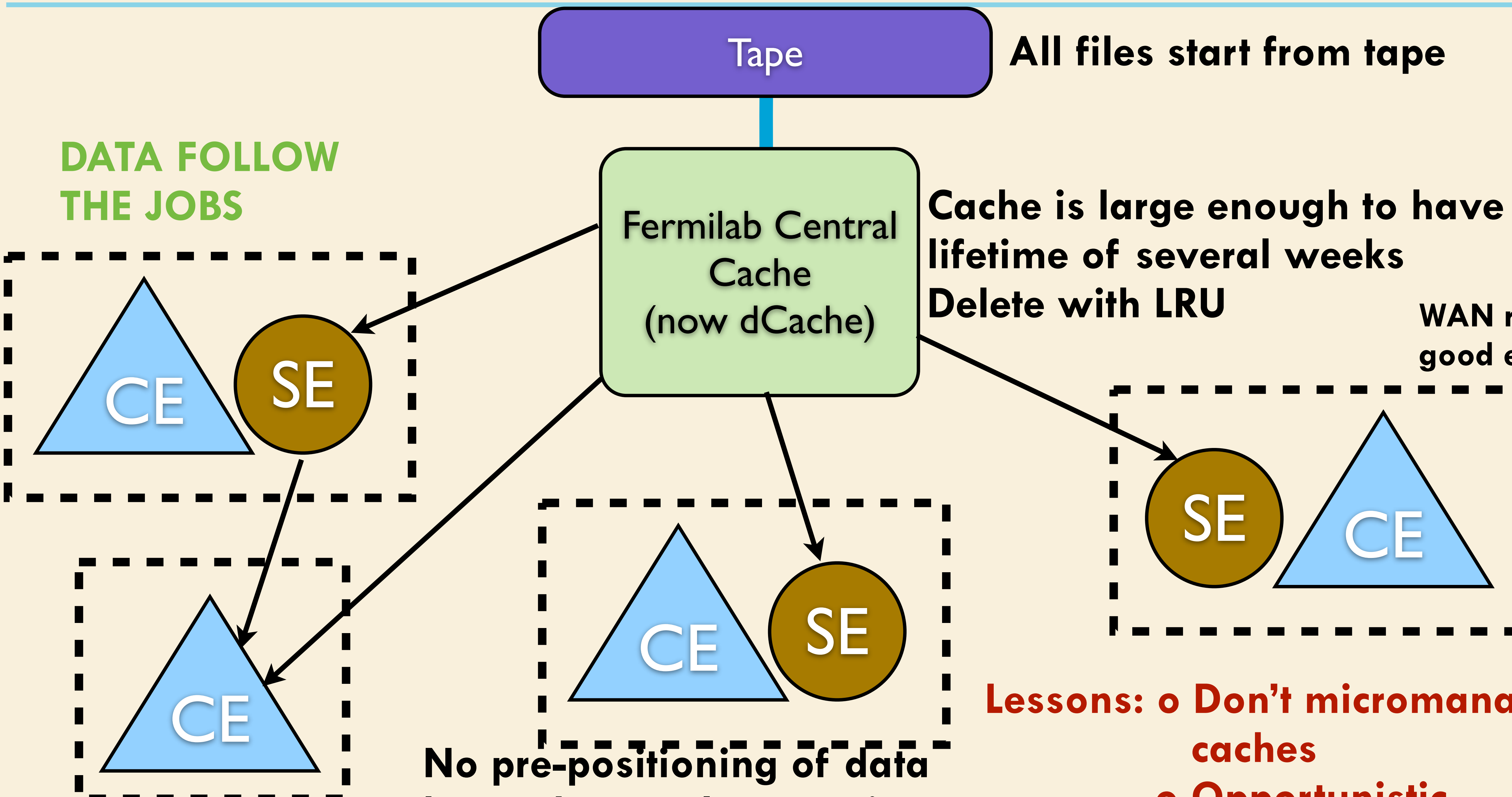
Open Science Grid - a consortium of computing farms that we can use opportunistically

LHC Computing Grid for CMS



Commercial clouds – Resources on Amazon (and Microsoft Azure) for peak overflow

IF Data Management Philosophy



All files start from tape

DATA FOLLOW THE JOBS

Cache is large enough to have lifetime of several weeks
Delete with LRU

WAN rates good enough

No pre-positioning of data beyond normal prestaging

Lessons:
- Don't micromanage caches
- Opportunistic running requires agility

Favor scalability over efficiency

Run as many sites as we can, even on ones less productive

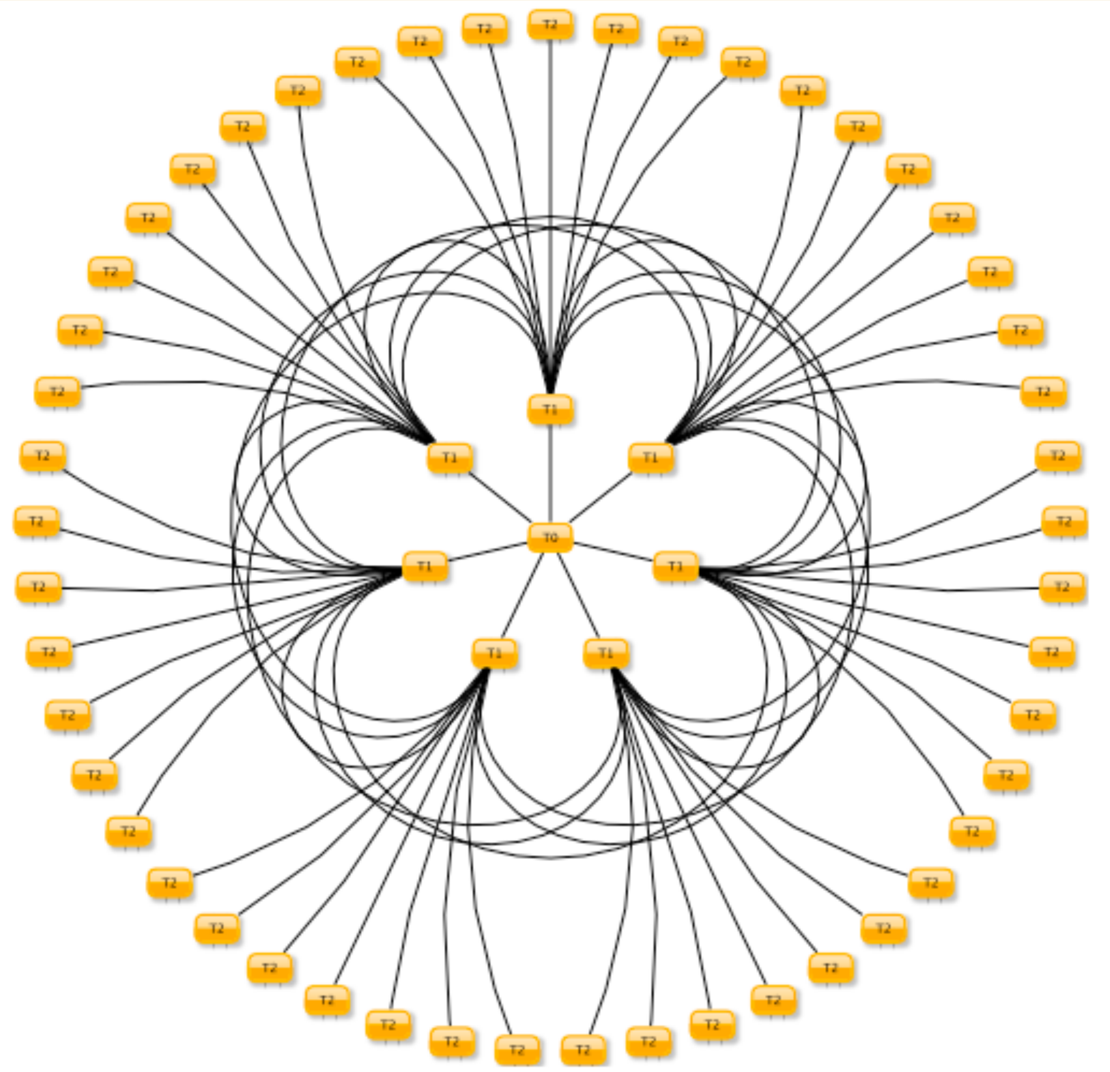
LHC Data Handling (CMS/ATLAS)

- Similar functionality to Tevatron Data Handling. The basic pieces are there, but implemented differently (PhEDEx, DBS3, PANDA, RUCIO, WMAgent/CRAB3, DIRAC)

- LHC Run 1: Hierarchical structure
  - Tier 0 @ CERN: Prompt reco
  - Tier 1 (regional): Re-reco, skimming calibration, make physics format
  - Tier 2 (local): analysis & MC

- Assign datasets to T1's and send jobs to the right place (JOBS FOLLOW THE DATA)

- Isolation of tape, limiting data movement allows for slow/non-robust networks, but no opportunistic running

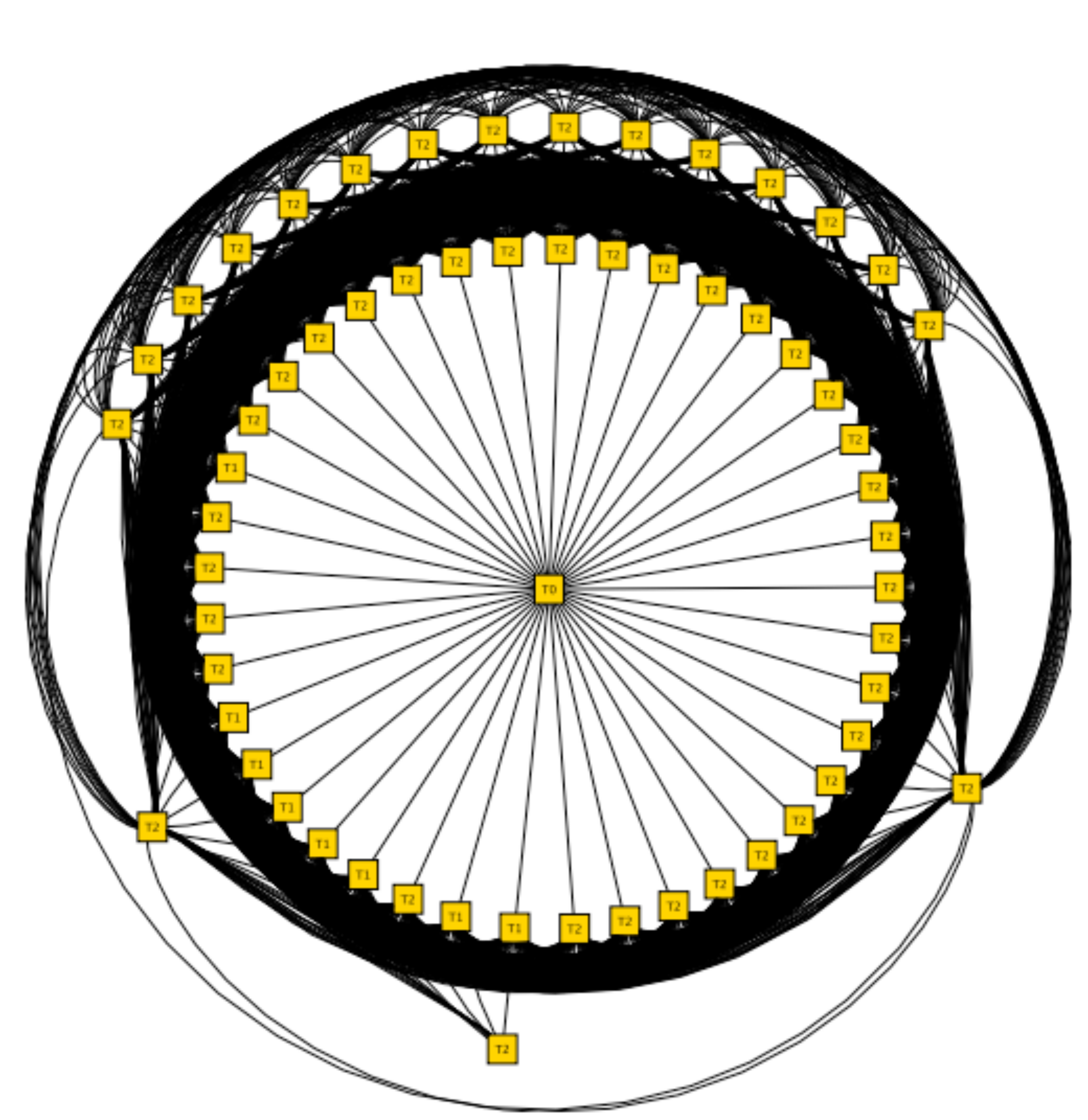


Favor efficiency over scalability

We control a set of sites, so maximize their productivity

Future LHC Data Handling

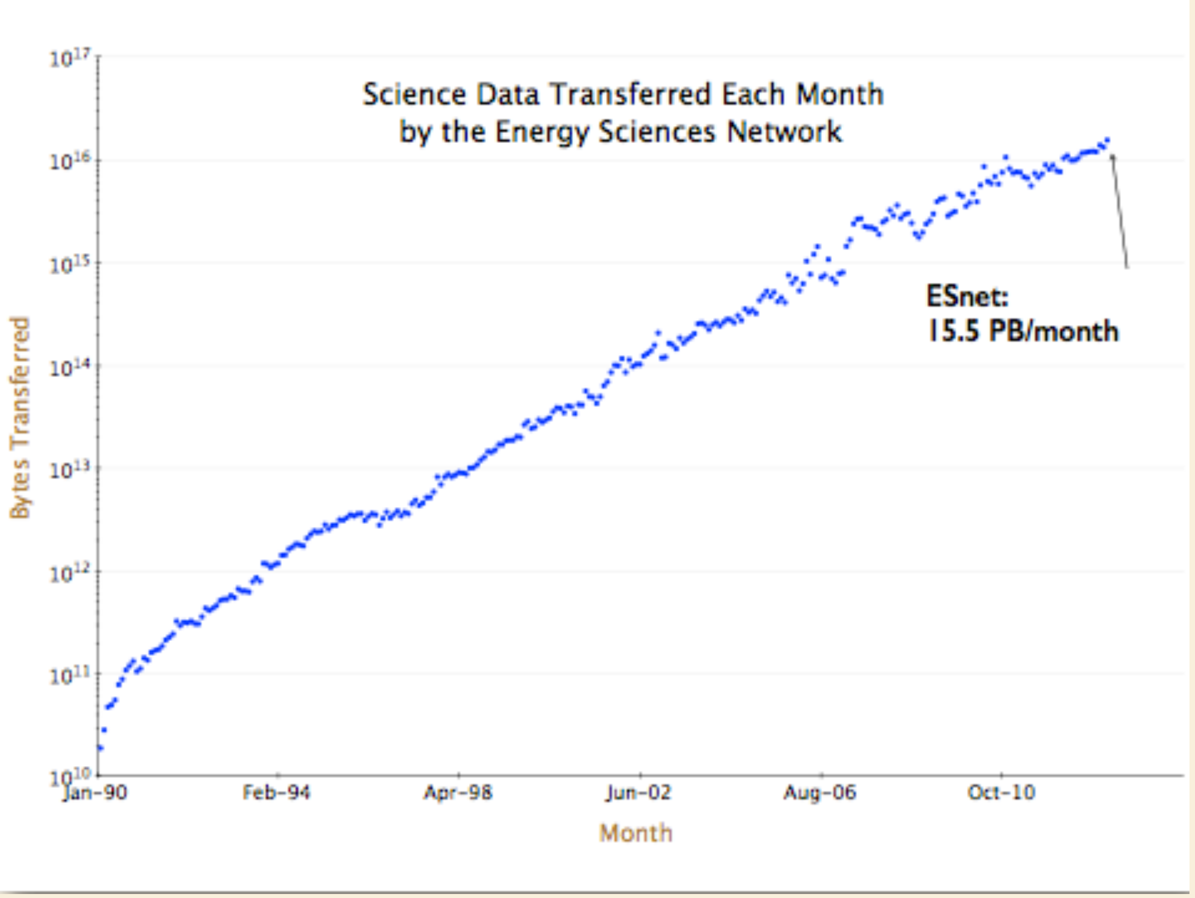
- Preparing for Run 2, the next level
- Networks are more robust and faster than predicted
- Connect everything together
- Still pre-position data on T1's but use federated XRootD as a fallback (Any data, Anytime, Anywhere)
- Makes opportunistic running possible (crucial for LHC Run 2)



Maximize reliability, scalability, and efficiency

Wow - the best of all worlds

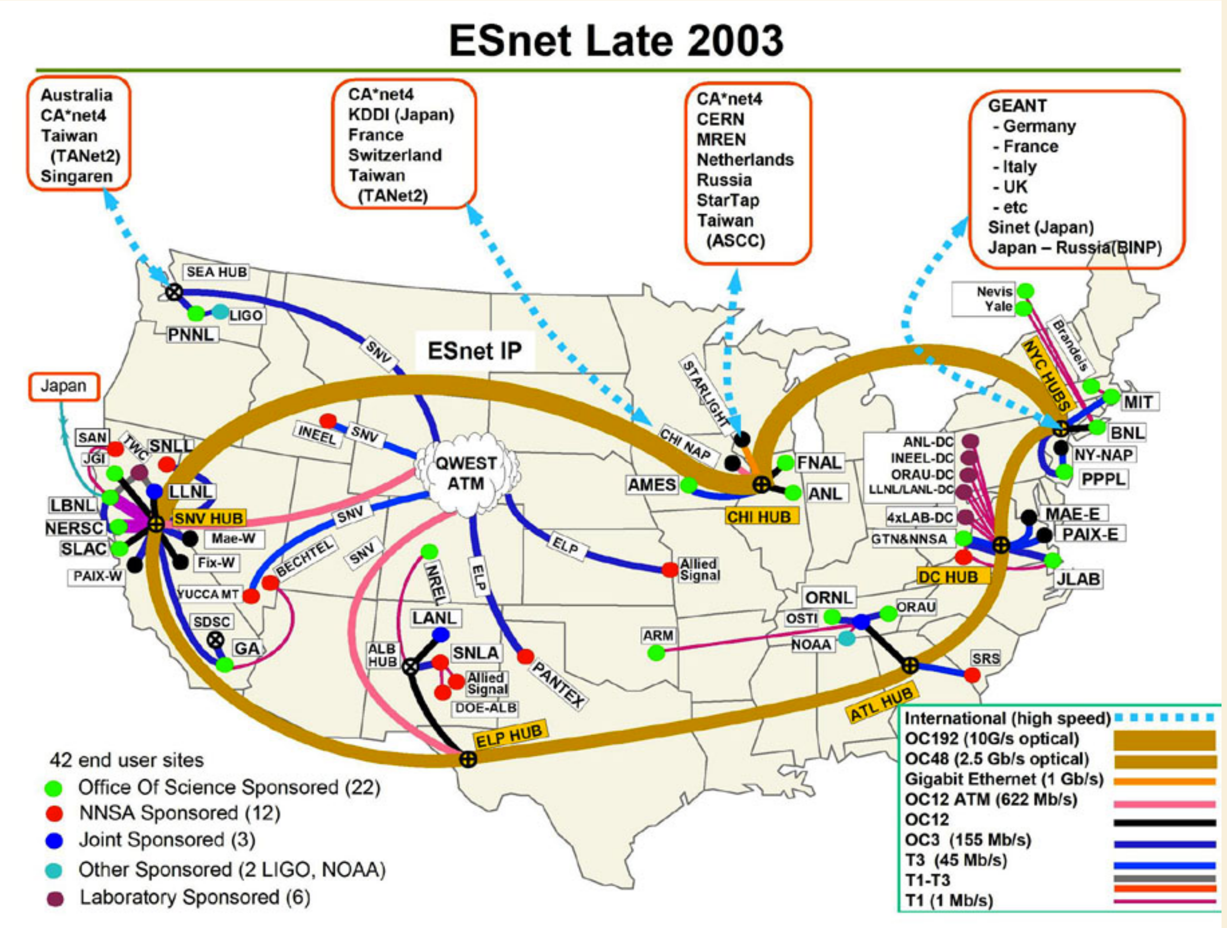
Robust Networks

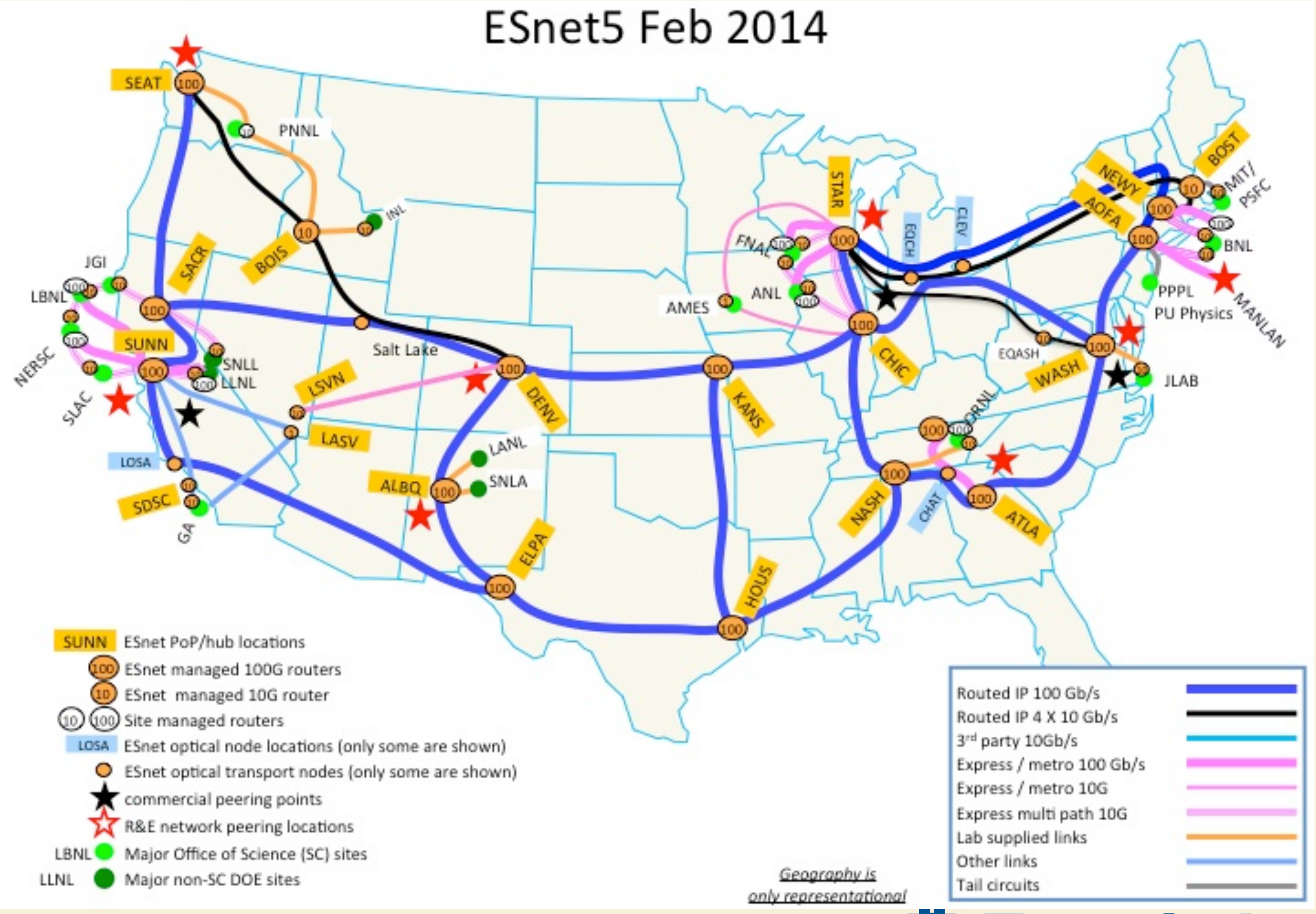


2003: 10 GB/s trunk lines
2014: 100 GB/s trunk lines

Large transcontinental pipes

DOE ESNet From GB to PB





Fermilab

The Future

- The computing landscape is changing
- CPUs can't go faster (no 10 GHz?)
  - Limited by power density
  - So integrate more of them together
  - Multicore CPUs (parallel computing)
  - GPGPUs (with attached ARMs)
  - Many integrated cores (MICs) [Intel PHI]
- Using these technologies is challenging, especially for non-experts
- Much R&D to try these architectures e.g. Physics codes, Geant, Synergia
- Just starting to explore Quantum Computing

The Free Lunch is Over (http://www.gotw.ca/publications/concurrency-ddj.htm)

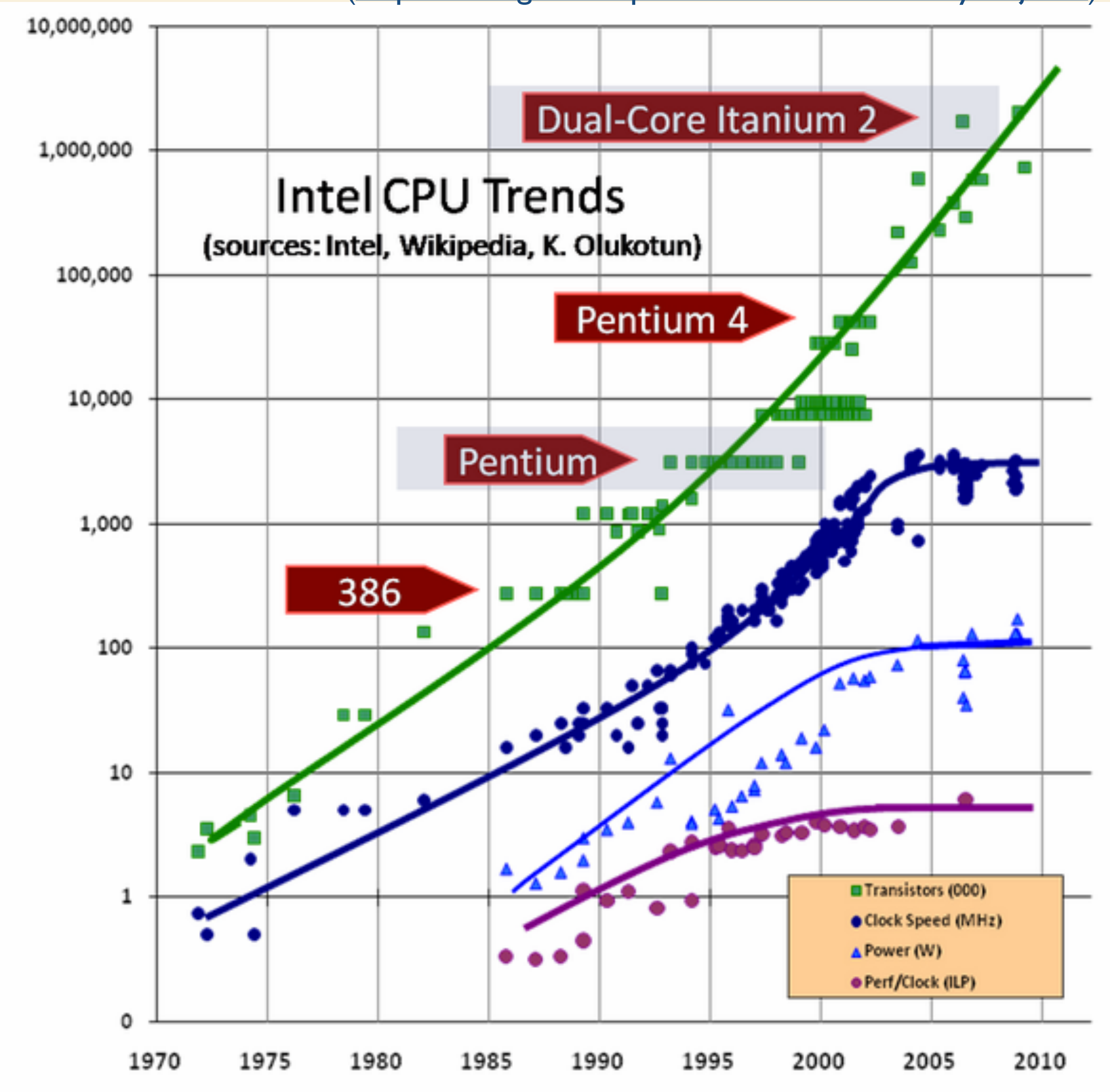
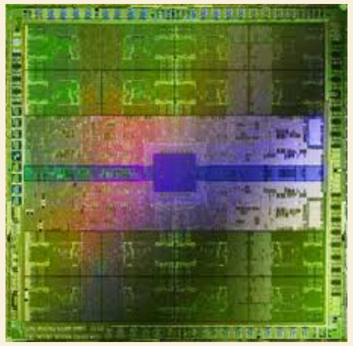
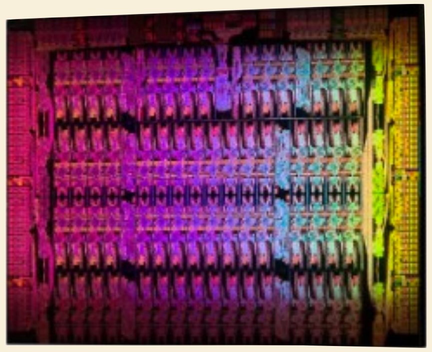
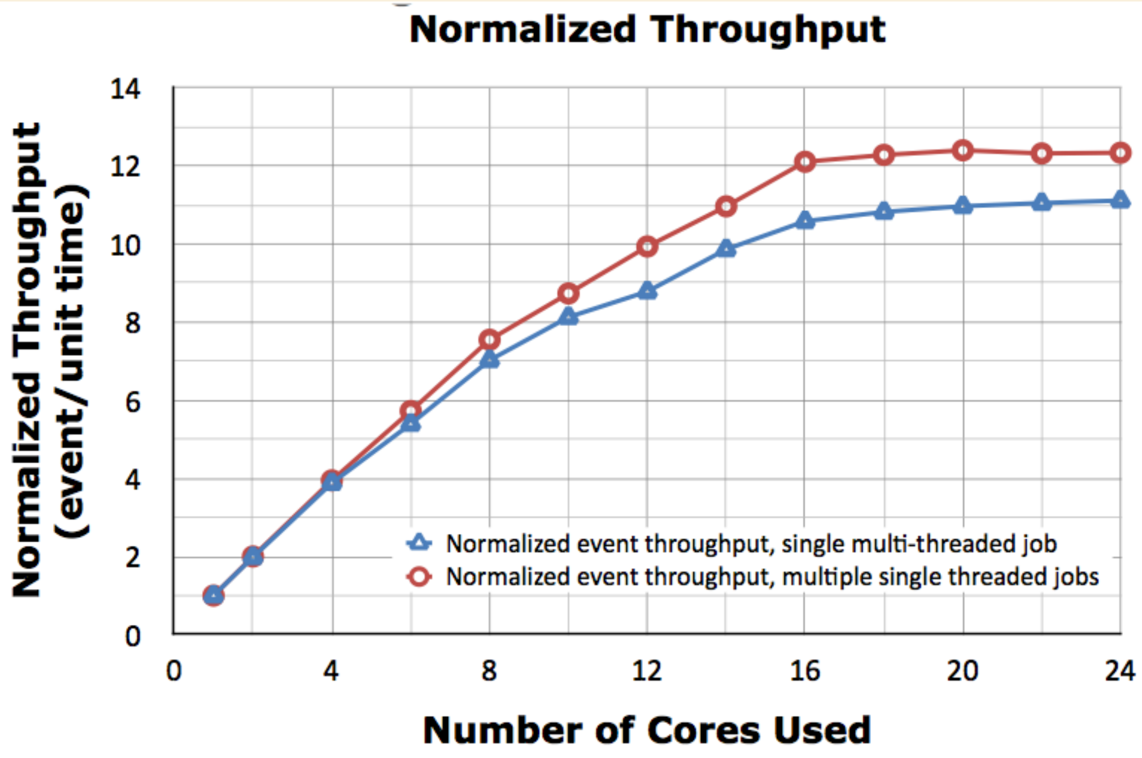


Figure 1: Intel CPU Introductions (graph updated August 2009; article text original from December 2004)

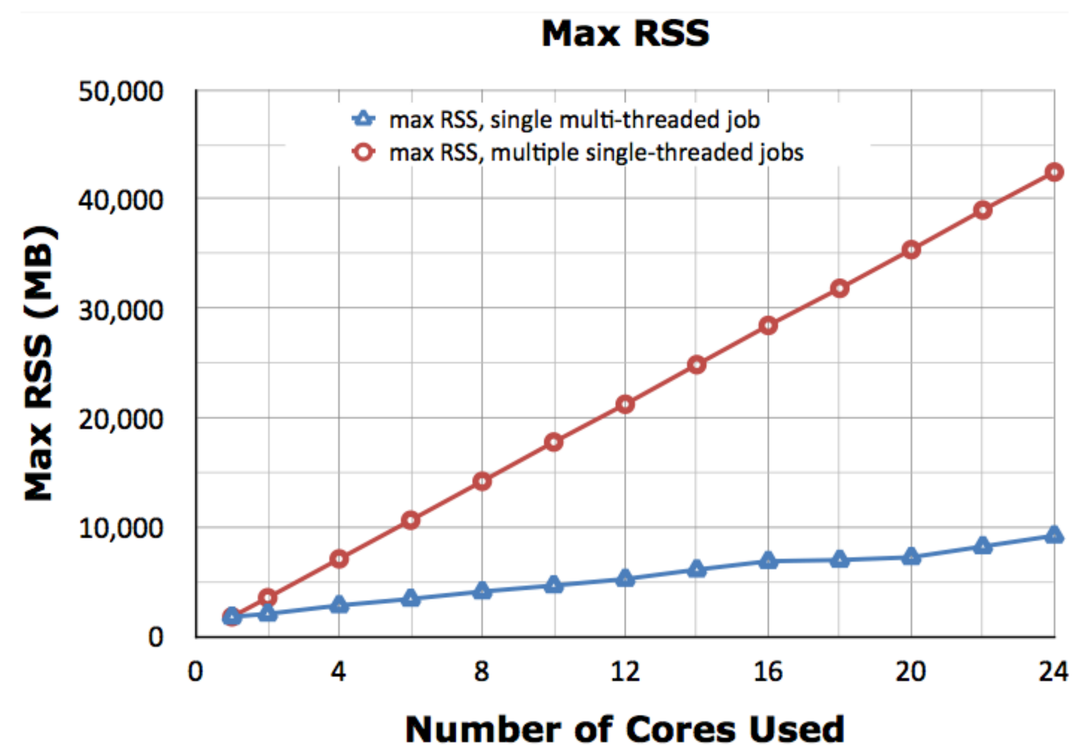


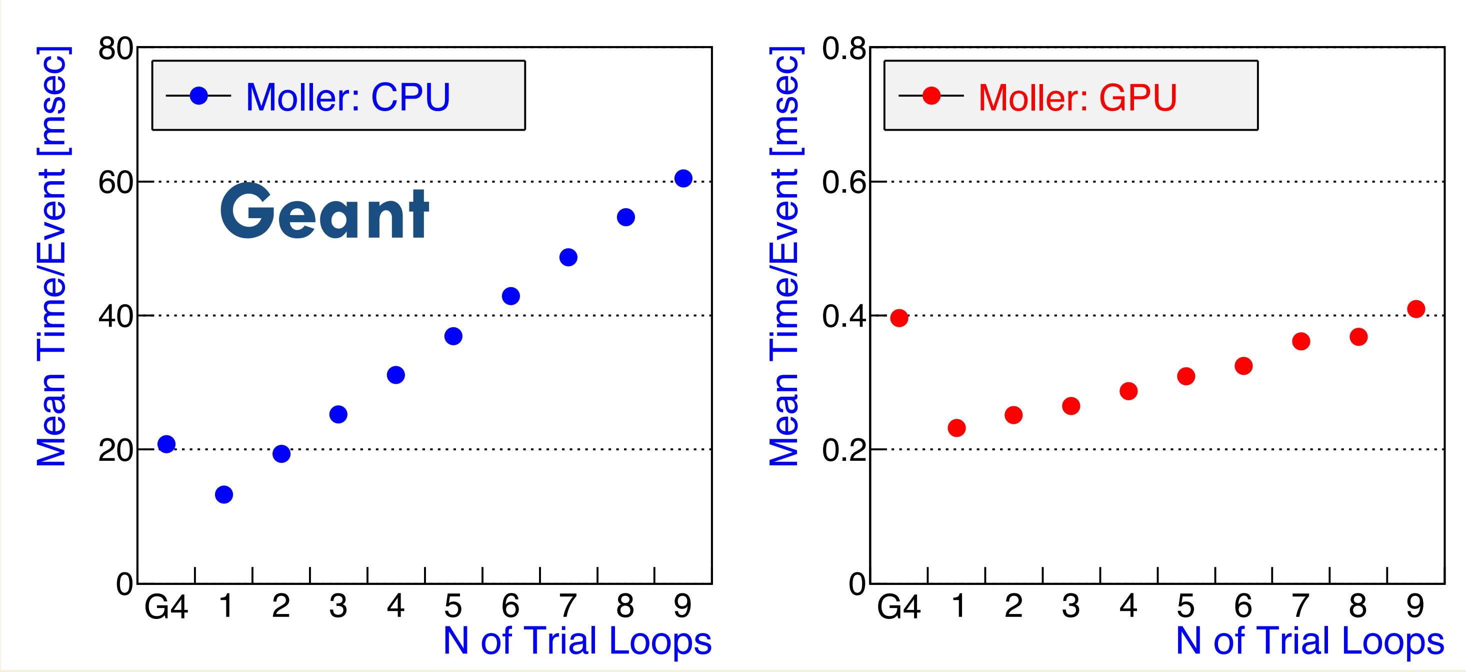


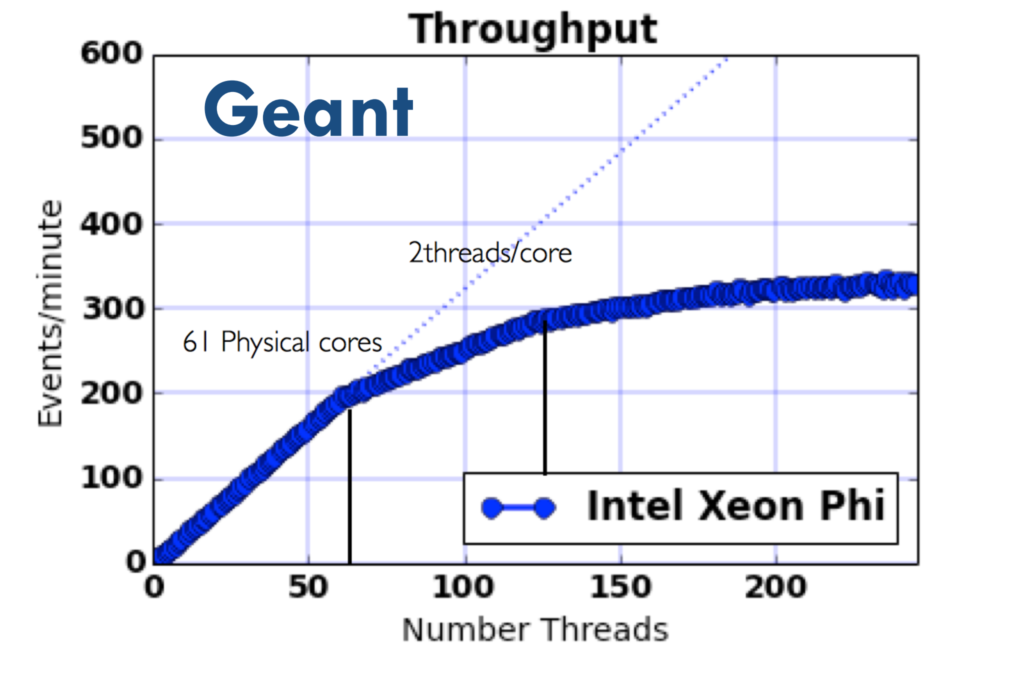
Effect of multiprocessing



95% efficiency compared to single-threaded jobs at significant memory savings







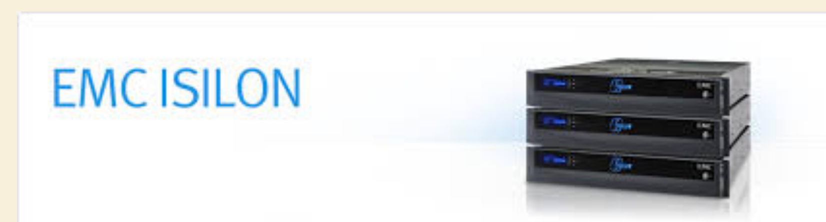
Is "Big Data" Technology Useful?

- Nothing small about our data! (CMS produces 20 PB/yr)

- Industry has commoditized the Map-Reduce paradigm (Hadoop, Spark, No-SQL DBs) - Processing occurs at the data

- Does the notion of a file survive? E.g. No-SQL databases for events?

- Can we use <u>Data Appliances</u>?
  e.g. EMC Isilon - Very high speed i/o with
  fast disk caching - can perform map-reduce &
  DB functions in the cache! Disks become compute
  nodes! Wow!



  e.g. YARC - "purpose built" data appliance from Cray
  Does a "graph analysis" to find connections in data

YarcData
Getting to Eureka! faster

  - Must pre-load these appliances.
    Is opportunistic use possible?

Case Study V

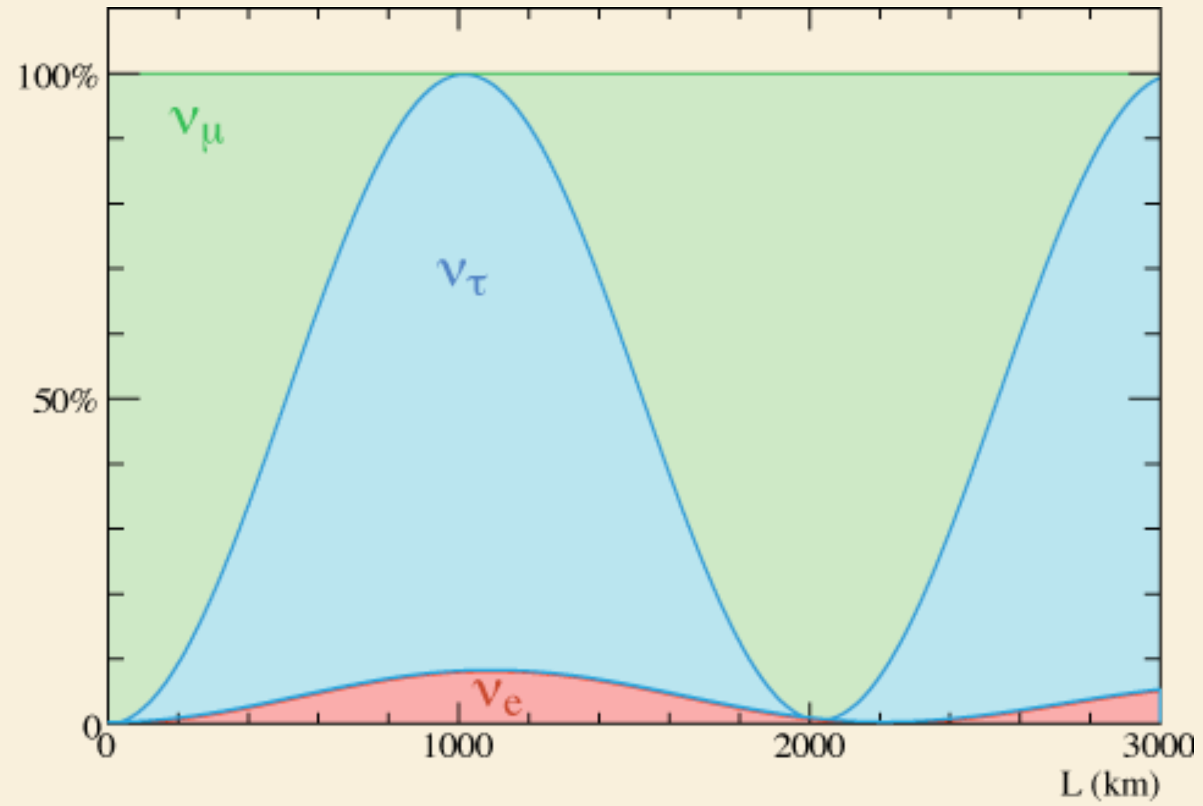
NOvA Template Matching

NOvA Reminder

Off-axis neutrino experiment in northern MN

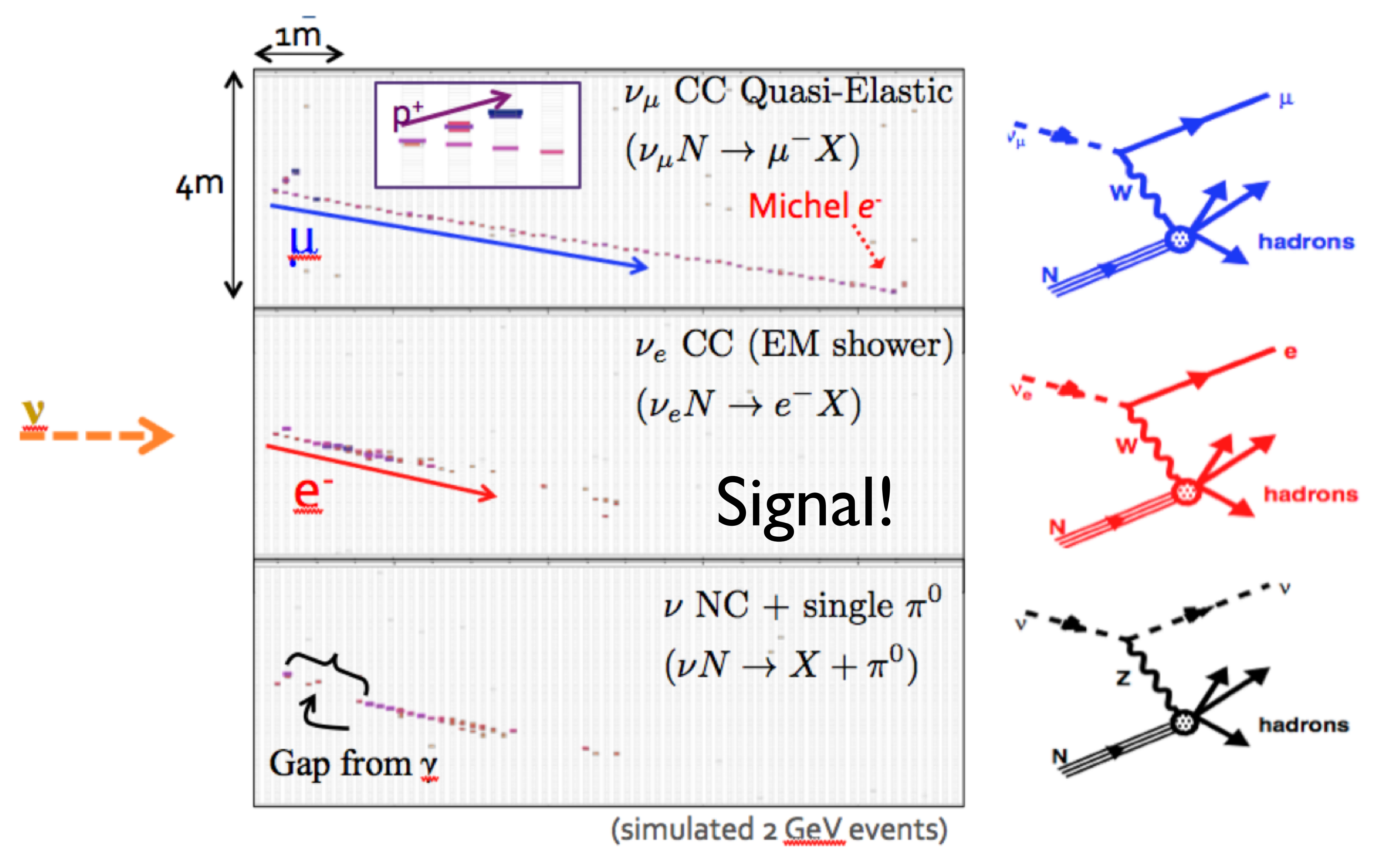
896 planes; 16m x 16m; 344,064 cells filled with liquid scintillator

Look for $\nu_\mu \to \nu_e$



NOvA Simulations

Simulated signals

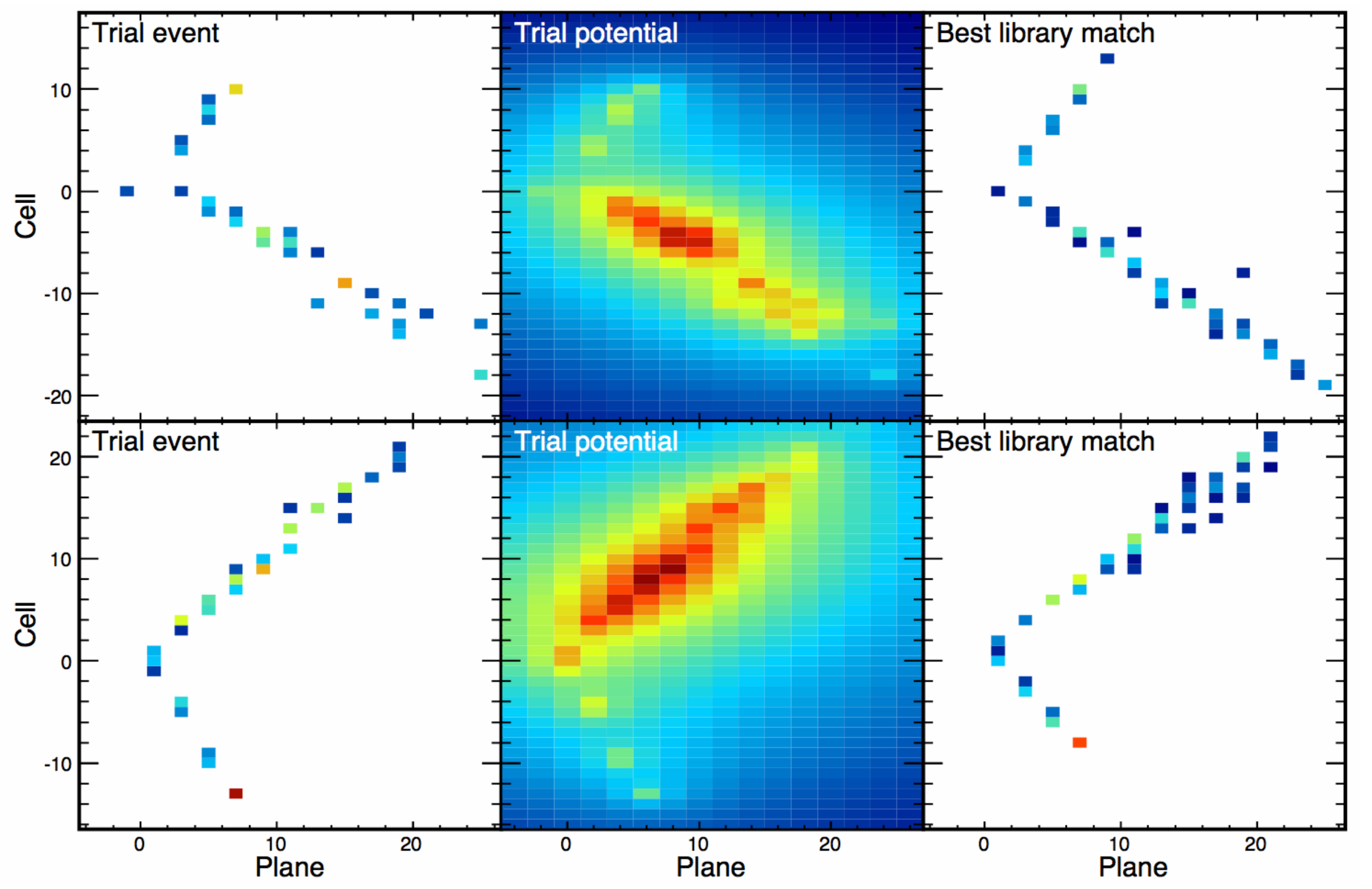


Fermilab

Library Event Matching

Problem: Identify and classify a data event

Solution: Compare to a large library of simulated events of known identity and find best match



Use "energy potential" to match trial event to library

77M simulated events

131 GB of memory

Use a dedicated large memory machine to run (bottleneck)

Improve performance

Signal efficiency is 55%, Background misid is 2%

Improvements come from a larger more varied library

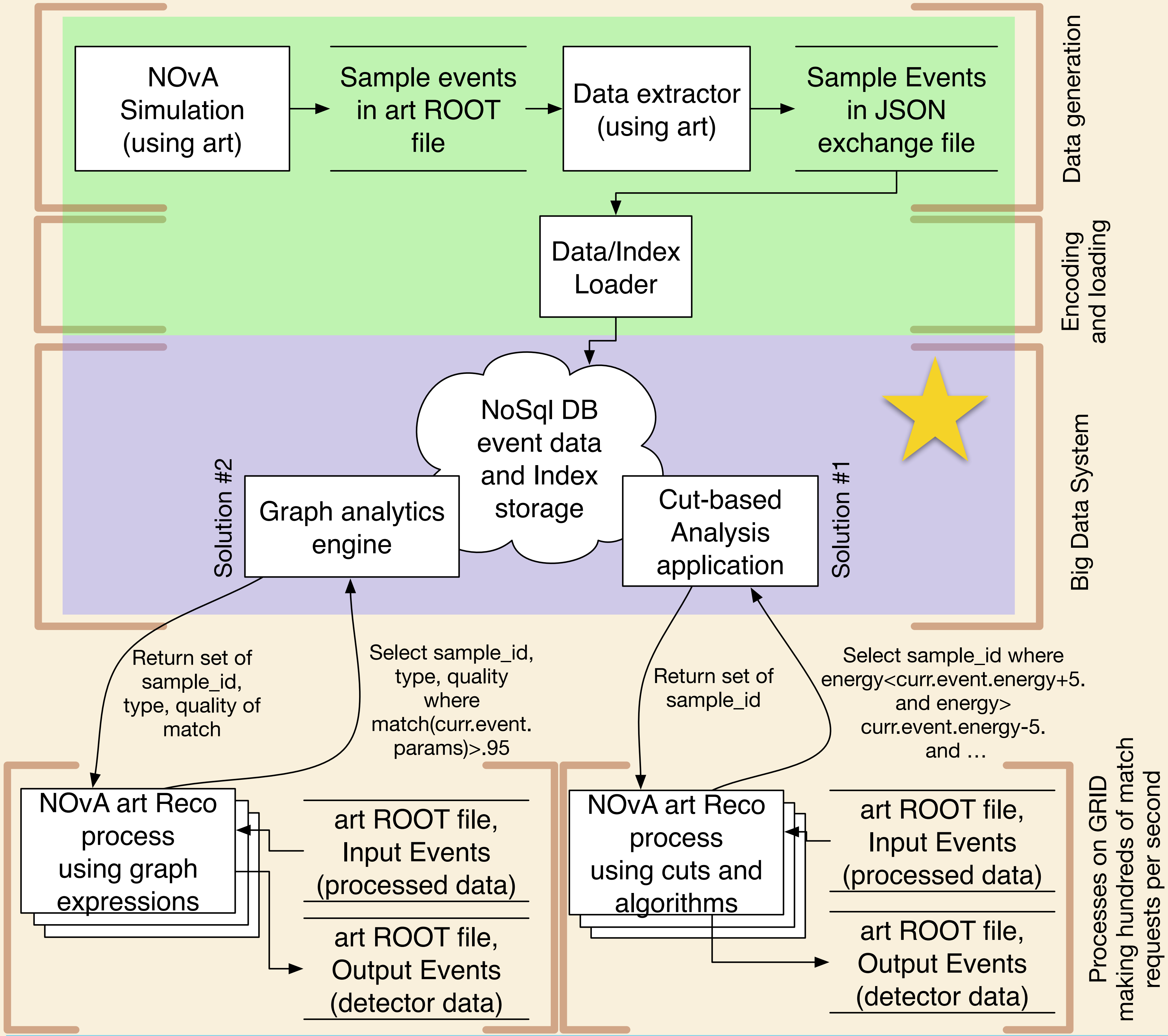
e.g. 10B template library would require ~ 10 TB

No longer possible on a dedicated machine

Traditional High Throughput Grid paradigm is very inefficient for this purpose (library would need to be local to every node for speed)

Are commodity Big Data techniques an answer?

Explore Big Data



NoSQL databases

- very fast
- calc in server near data
- but limitations

Graph analytics engine

- (entities and their relationships)
- Next big thing?

Starting to explore these ideas and tools - New for us

Wrapping up

Simulations are absolutely crucial for understanding,

- o physics
- o accelerators
- o detectors
- o data

Running these simulations requires major computing horsepower

The computing landscape is changing:
New architectures, new techniques, new tools

We have to keep up - and it's not easy!

We're very open to collaborating!!

An exciting time for particle physics: e.g. LHC Run 2, Short & long baseline neutrino experiments running, Muon program takes data 2017