Simulation in Particle Physics

Rob Roser and Tom LeCompte

Outline

Geant4 and its importance to High Energy Physics

The Process of Transforming Geant4 to take advantage of the emerging parallel architectures

Projecting our Needs

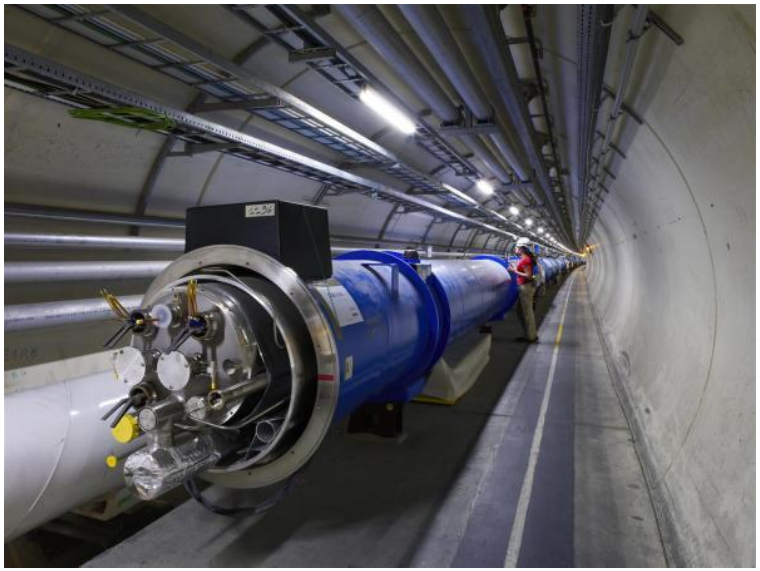
The Energy Frontier

The Large Hadron Collider is the world's largest particle accelerator.

- It collides beams of protons at 8 TeV.
- Possibly the largest machine of any kind.

There are four large experiments there:

- ATLAS & CMS (general purpose)
- ALICE & LHCb (specialized)

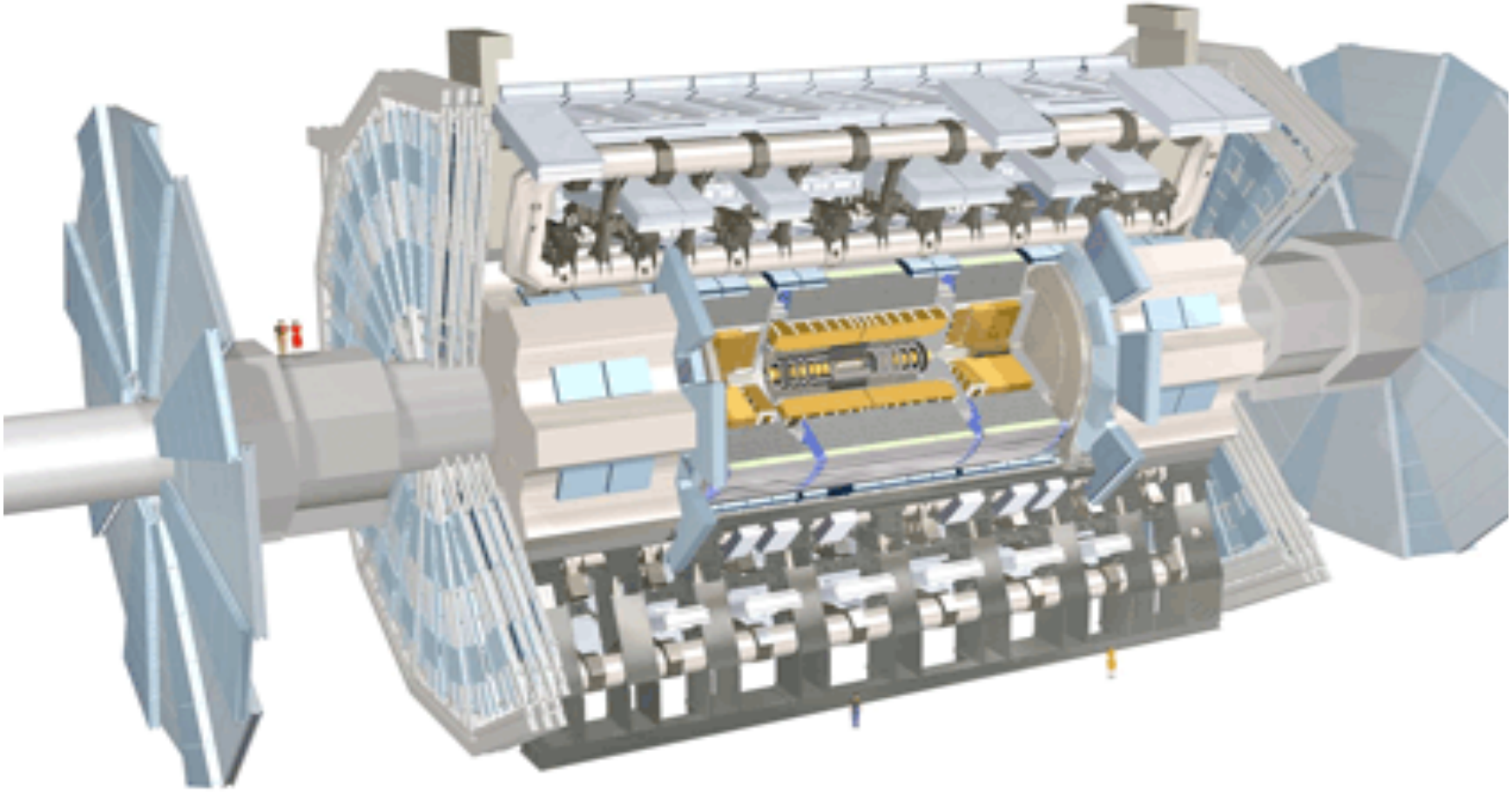


The experiments intend to shed light on several fundamental questions:

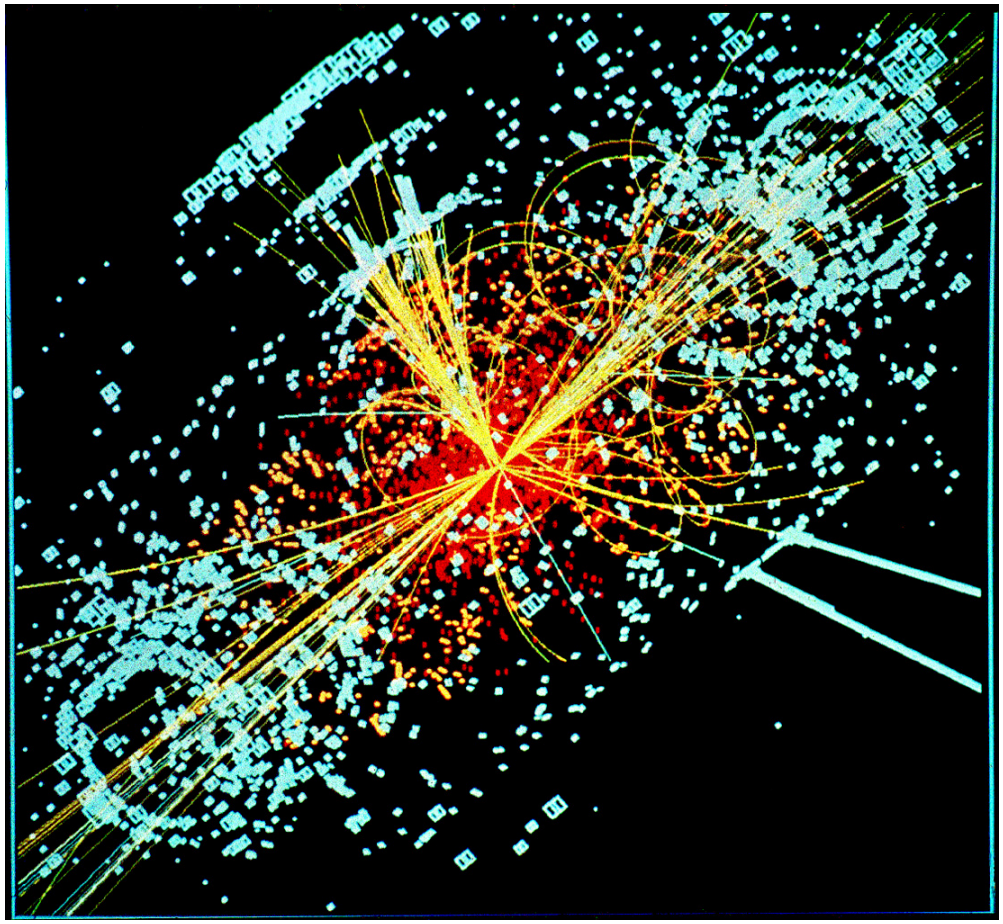
- Why is the weak nuclear force so weak?
  The "Higgs mechanism" is one answer to this.
- What is the Dark Matter (80% of the universe?)
- Why is the universe made out of matter, not antimatter?
- Why is gravity so much weaker than the other forces?

The ATLAS Detector

It takes a lot of “stuff” to detect particles!



Simulated Higgs Event in CMS



What is Geant4?

Geant4 is a C++ tool kit that tracks particles through matter, breaking the particle motion into small segments, applying appropriate physical processes and probabilities at each segment.

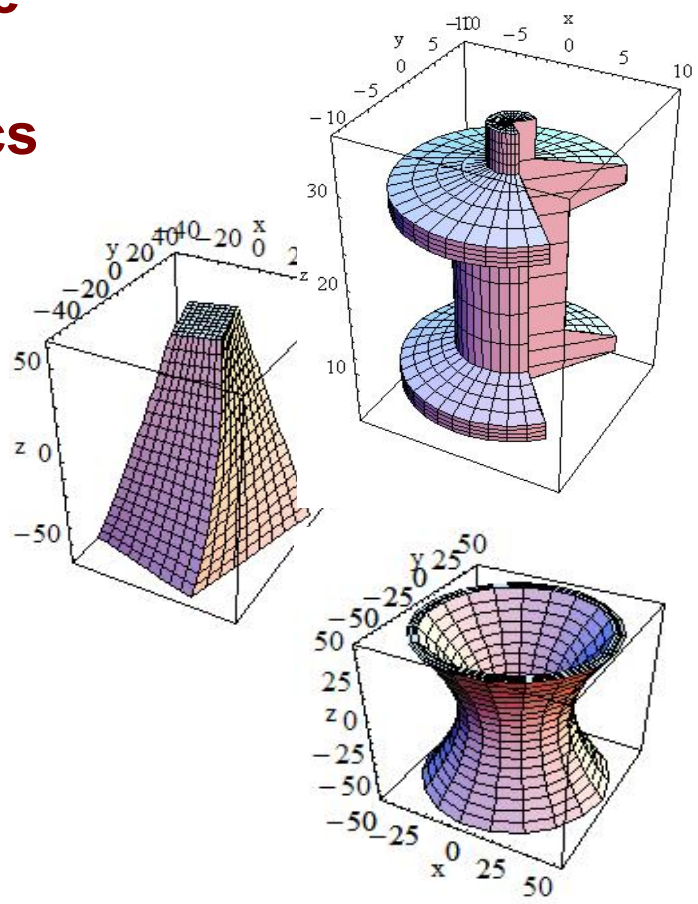
- **These processes can destroy old particles, modify state or create new ones**
- **Processes include atomic processes like ionization and excitation, decay processes, photonic transitions, secondary emission, etc.**
- **The wide coverage of physical processes comes from mixture of theory-driven, parameterization, and empirical formulae.**
- **Successor to Geant3, the Geant4 Project began in 1994 with the first public release in 1998**

Geant4 is Unique

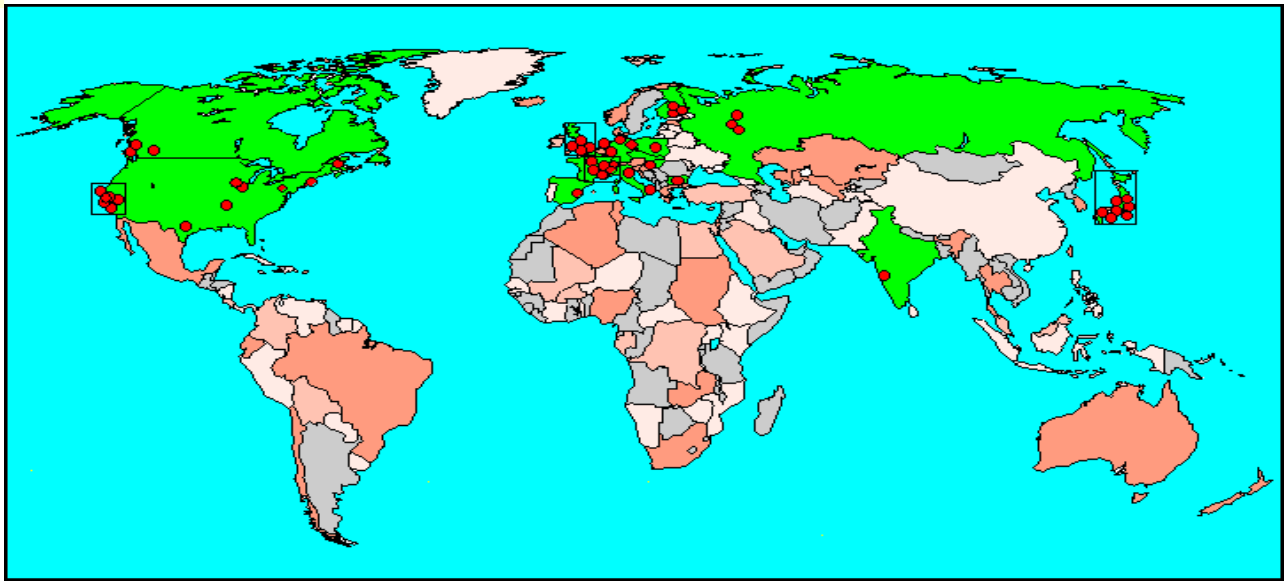
Geant4 is distinguished from other Monte Carlo Particle Transport codes by

- **The comprehensive suite of physics processes and particle types**
- **The complexity of geometrical descriptions leads to realistic representations.**
- **A collaborative open source model leveraging international expertise**

Enables the user to select physics processes/models and choice of GUI, visualization, persistency, and histogramming technologies



Geant4 is an International Collaboration



TRIUMF

Lebedev

J.W.Goethe Universität

Collaborators also from non-member institutions, including

Budker Inst. of Physics

IHEP Protvino

MEPHI Moscow

Pittsburg University

Geant4 is Big

Big Computation

- Large ensembles of sequential jobs
- Runs on a worldwide Grid of processors
- Significant Computing Hardware investment worldwide in order to satisfy demand
- Not clear that computing capacity will continue to scale with luminosity
  - (it is the luminosity (total number of collisions) that sets the scale for how many simulated events we require.

Big Data

- HEP has more simulated than collected data
- Approaching O(100) Pbytes by the end of this year

Geant4 Supports the Full Experiment Lifecycle

1. Detector design
2. Software development
3. Commissioning and calibration
4. Data analysis

Data Analysis is by far the primary use case for simulation

Any useful data curation must also curate the experiment's Geant4 program

What do we use Simulation For

- Model physics processes and design a set of analysis “cuts” that optimally find what we are looking for
- Understand how often we should see a certain type of event in our collision
- Train neural networks and other advanced analysis techniques on specific signatures
- Simulate a new physics model invented by a theorist to see whether we can detect it
- Etc…

Challenges for Geant4's Future

Geant4 is a sequential C++ toolkit

- MC runs are ensembles dispatched to the Grid
- It can take months to simulate a billion particles
- Code as it stands now is extremely "serial"

CPU capability has plateaued

- Dennard scaling has ended

Potential to constrain progress in HEP if we don't react

Joint HEP/ASCR Workrshop held this Spring

Bring people from both communitites together to look at what could be done to parallelize Geant4

About 50 attendee's equally split between HEP and ASCR communities

Summary of Charge from Dan and Jim

Bob Lucas and Rob Roser to co-chair workshop
Ceren Suset and Lali Chatterjee are DOE contacts

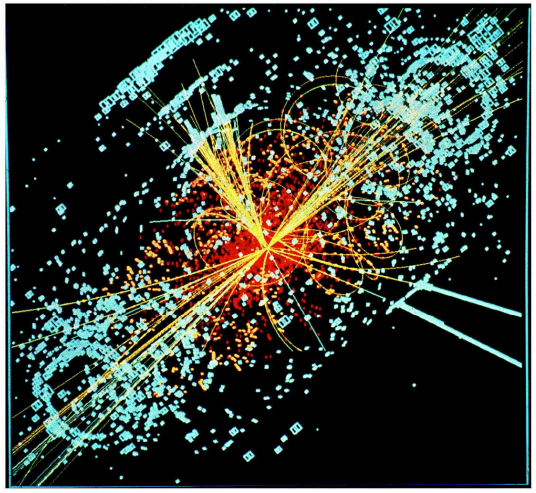
Goals:

- **Review status, successes and limits of Geant4**
- **Determine challenges posed by new architectures**
- **Consider opportunities in algorithms and optimization**
- **Ascertain research for robust, sustainable code**
- **Create foundation among ASCR and HEP investigators**
- **Understand and not duplicate international efforts**
- **Explore transformative advances via HEP-ASCR collaboration**

Report

Transforming Geant4 for the Future

Report from the Workshop on
Transforming Geant4 for the Future
September 2012



Event Generation

The first step in the simulation process – and something that often goes un-noticed

These software packages perform the detailed feynman diagram calculations in order to probabilistically determine the characteristics of a specific event from a specific physics process

- **The sophisticated generators don't just do the "tree level" process but also worry about radiation,bremsstrahlung and loop corrections**
- **There are a wide range of generators used by the HEP community.**
- **For Hadronic events – PYTHIA, HERWIG and SHERPA and MADGRAPH are most common.**
- **There are more specialized ones like ALPGEN, MC@NLO, and JIMMY which are also widely used**

Event Generation(2)

We could actually (and do) decouple the event generation from the detector simulation effort.

Both are quite CPU intensive

MADGRAPH is very CPU intensive and could benefit from running on NERSC style machines

A Snapshot Now

Currently, CMS and ATLAS has each produce between 7-10 billion events over its lifetime.

These events are produced WORLD-WIDE via the GRID and then stored centrally

An average size of a simulated event is ~500kb

Full simulation of one event on one core today ~2-5 min depending on complexity of physics process

Demand for simulation has peaks as the collaboration prepares for certain prestigious conferences

Looking into the Future

As luminosity of the machine improves, the demand for simulated events will grow. 200M Simulated events/fb is a reasonable approximation

Expect we need to produce 10's of Billions of simulated events in 2017 and more when the high luminosity machine upgrades come online in 2020.

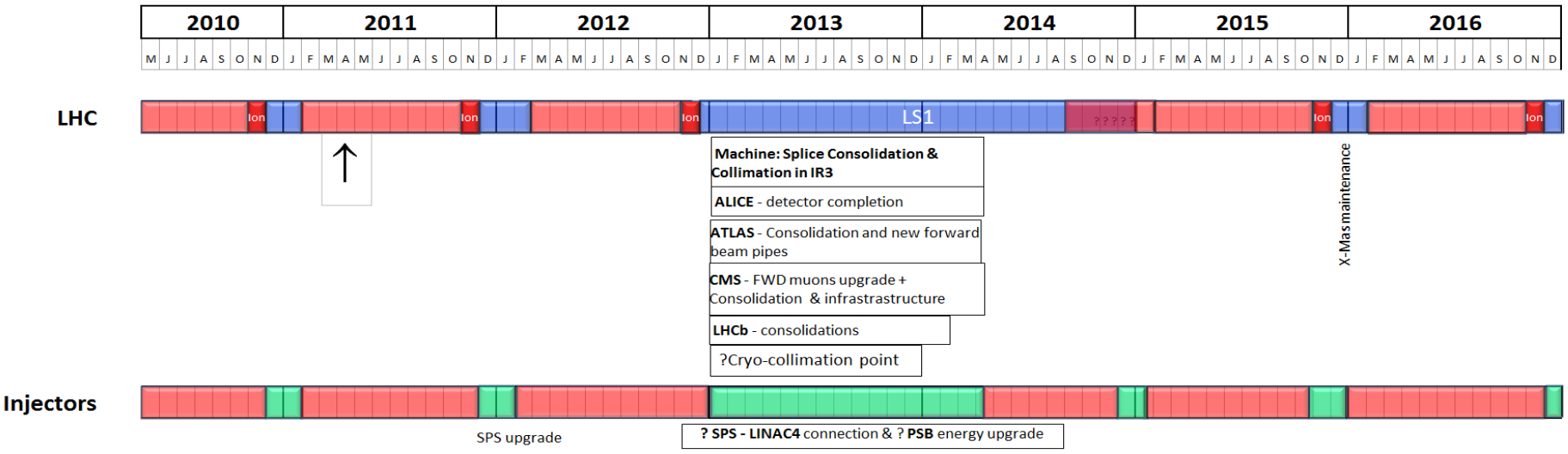
The detectors are not changing dramatically and thus the average simulated event will remain at ~500kb

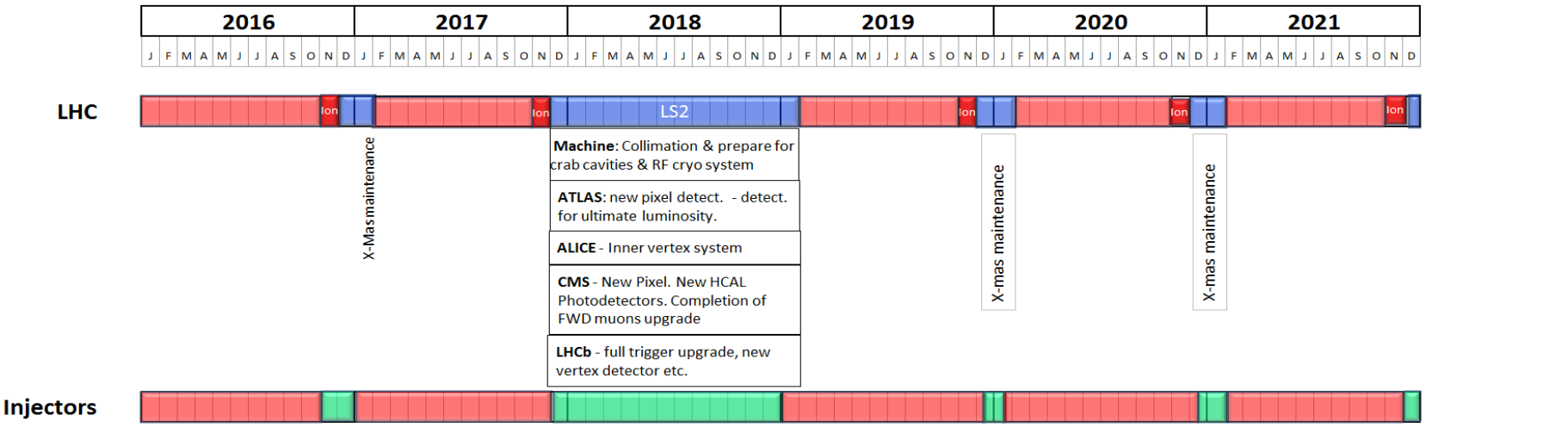
Work is underway to improve the parallelization of Geant4 and thus reduce the time to produce a single event

A facility like NERSC would be tremendously useful to HEP – even to just help with peak demand periods

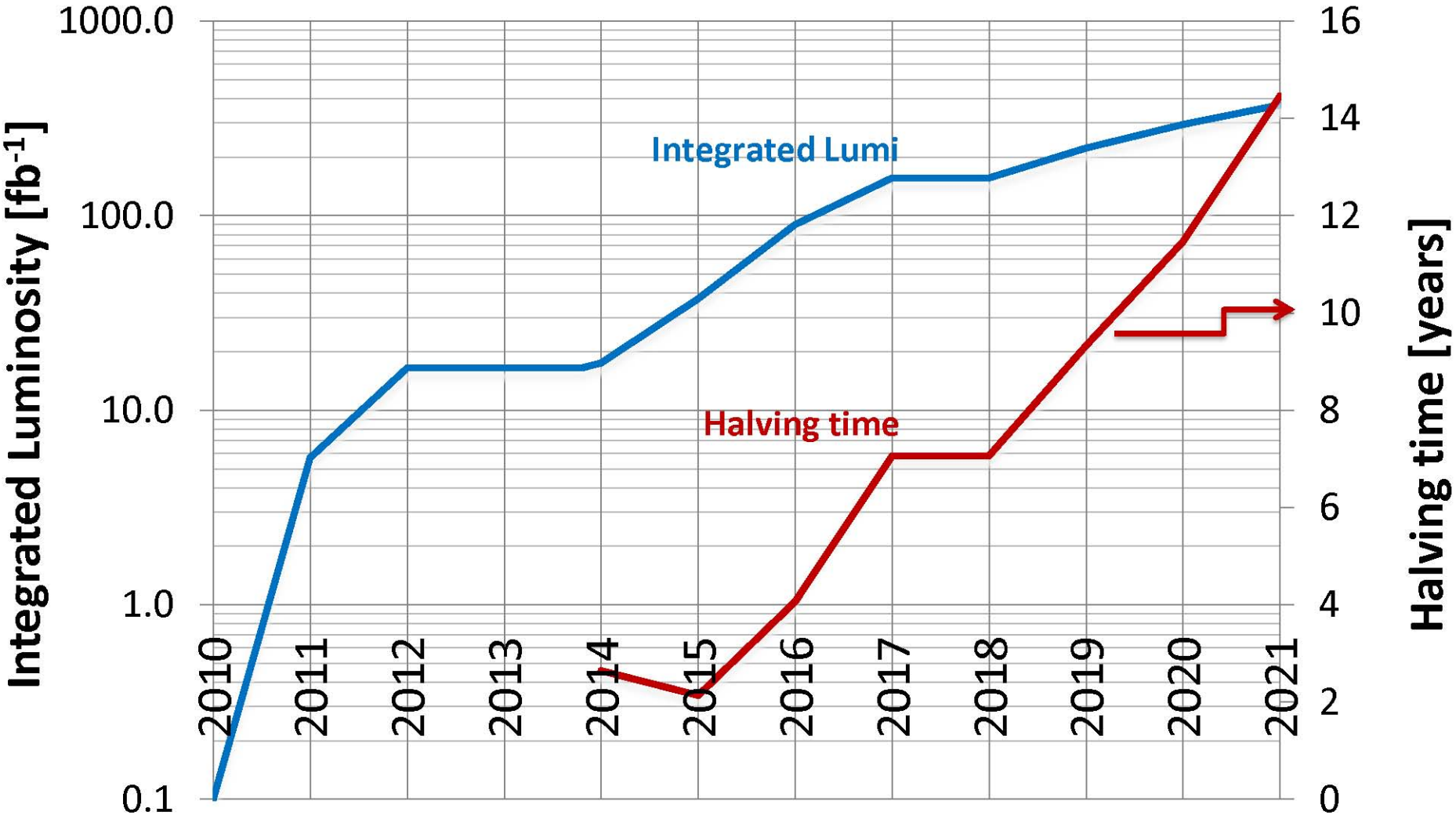
10 YEAR PLAN 2011-2021

New rough draft 10 year plan





10-year luminosity forecast



Take Away Message

HEP Simulation is critical to the success of the field.

As luminosity grows, the demand for simulation will also grow.

HEP needs to either find new resources or do more with current resources in order to not negatively impact the science

We are working now to "parallelize" our codes to take advantage of multi-core architectures

However – our computing requirements fall more into the category of high throughput computing rather than high performance computing – our simulation problems are different than say that of galaxies and we don't need the fast interconnects (however certain applications could benefit!)

The Request

NERSC has HPC's, but also more traditional clusters like PDSF. We would like to exploit both types of technologies.

We don't just need HPC, we need a way to get jobs in and out – ideally with something like OSG that we are used to

We would like to get started sooner than later in utilizing these facilities and see how well we can do.

By 2017, tens of millions of CPU hours at NERSC should be very beneficial to this science.