Particle Physics: The Big Picture

How we compute our way to discovery

Gavin S. Davies University of Mississippi

EPIC Summer School · Monday, July 13, 2026

Part 1

Computing *is* physics

Picture a particle physicist at work

What do you actually see? A lab coat? A blackboard covered in equations? A giant machine underground?

Picture a particle physicist at work

What do you actually see? A lab coat? A blackboard covered in equations? A giant machine underground?

Most days it's a laptop, a terminal, and code, closer to what you'll be doing this week than you'd think.

Picture a particle physicist at work

What do you actually see? A lab coat? A blackboard covered in equations? A giant machine underground?

Most days it's a laptop, a terminal, and code, closer to what you'll be doing this week than you'd think.

There's still real hardware, though. This week you can see our undergrads building a cloud chamber, next year Dr. Quinn's lab starts testing TPC electronics, and I spend some of my own time doing QA on components for the DUNE Near Detector's TMS.

The modern scientific method

1. Idea
What if neutrinos oscillate between flavors?

→

2. Detector
Build something that can "see" it

→

3. Data
Record what happens, millions of times

→

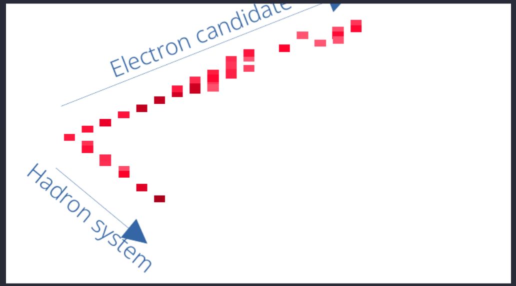
4. Computing
Turn raw data into an answer

Step 4 used to be an afterthought. Today it's most of the experiment.

Detectors are cameras

A particle detector is, at its core, a strange digital camera:

- Instead of light hitting pixels, particles hit sensors: wires, tubes, crystals
- Instead of one photo, you get millions of "photos" per second
- Instead of a person choosing when to press the shutter, a computer decides, in real time, which "photos" are worth keeping

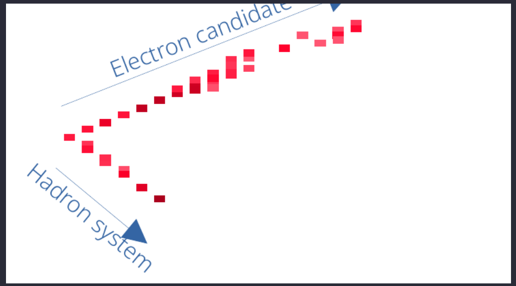


A real NOvA interaction: the "photo" a camera like this actually takes

Detectors are cameras

A particle detector is, at its core, a strange digital camera:

- Instead of light hitting pixels, particles hit sensors: wires, tubes, crystals
- Instead of one photo, you get millions of "photos" per second
- Instead of a person choosing when to press the shutter, a computer decides, in real time, which "photos" are worth keeping



A real NOvA interaction: the "photo" a camera like this actually takes

You do this too: you keep the good shots, not every frame. Particle physics does it at 40 million shots per second, in millionths of a second.

The trigger: deciding what to keep

At the Large Hadron Collider, particle collisions happen 40 million times per second. Almost none of them are interesting.

- **Level-1 trigger**: custom hardware, decides in microseconds whether to keep it or drop it
- **High-level trigger**: a computing farm of thousands of machines, double-checks in milliseconds
- Out of 40,000,000 collisions a second, only a few thousand get written to disk

Every experiment you'll hear about this week makes this same keep-or-drop decision, just tuned to its own physics.

Quick poll: keep it, or drop it?

You're the Level-1 trigger. One second to decide, out loud, for each:

- A cosmic ray passing straight through, the most common thing your detector sees
- A pattern matching a known particle decay nobody's ever measured this precisely
- Random electronic noise in one sensor
- Something that doesn't match any pattern in your simulation library

Quick poll: keep it, or drop it?

You're the Level-1 trigger. One second to decide, out loud, for each:

- A cosmic ray passing straight through, the most common thing your detector sees
- A pattern matching a known particle decay nobody's ever measured this precisely
- Random electronic noise in one sensor
- Something that doesn't match any pattern in your simulation library

Real triggers make this exact call, millions of times a second: familiar noise gets dropped, and anything rare or unrecognized is kept. That last one, the pattern nobody's simulated, is exactly how you'd stumble onto new physics you weren't looking for.

How much data is that?

~1 PB/s
raw data hitting LHC sensors, before any trigger

~90 PB
stored at CERN per year, after triggering

~30 PB
DUNE alone, expected per year once running

1 petabyte (PB) = 1,000,000 gigabytes. Your phone probably holds ~0.0001 PB.

A phone photo is about 4 MB, so 1 PB is roughly 250 million photos: fill your phone completely, then do that another 7,800 times. CERN stores that much 90 times over every year. DUNE alone is expected to add 30 phones' worth, every year, on its own.

Where did the Web come from?

In 1989, a physicist named Tim Berners-Lee, working at CERN, wrote a proposal called "Information Management." His boss's only comment in the margin: *"Vague, but exciting."*

The problem was ordinary: thousands of physicists, in dozens of countries, on incompatible computers, needed to share documents and data. His fix (hyperlinks, HTML, HTTP, a URL) is what we now call the Web.

Remember where the internet came from

The actual server room at CERN: yes, that rack really is labeled "World Wide Web Servers." Photo: Hugovanmeijeren, CC BY-SA 3.0.

The Web was step one. The Grid was step two.

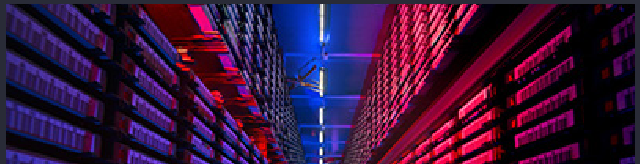
The Web let physicists **share documents.** It didn't solve the much bigger problem: sharing **petabytes of data and millions of hours of computing power.**

The answer, built through the 2000s: the **Worldwide LHC Computing Grid (WLCG)**

170+ computing centers	**40+** countries	**1M+** CPU cores, pooled together

Same instinct as the Web, just applied to computers instead of pages.

Fermilab Feynman Computing Center: one of 170+ Grid sites



Big data, before "Big Data" was a buzzword

Particle physics has been doing this for decades:

- Splitting enormous datasets across many computers (**not one giant supercomputer**)
- Running the same analysis job **thousands of times in parallel**
- Building databases and file catalogs that track petabytes across sites
- Automating pipelines so a physicist doesn't hand-manage every job

Big data, before "Big Data" was a buzzword

Particle physics has been doing this for decades:

- Splitting enormous datasets across many computers (**not one giant supercomputer**)
- Running the same analysis job **thousands of times in parallel**
- Building databases and file catalogs that track petabytes across sites
- Automating pipelines so a physicist doesn't hand-manage every job

Silicon Valley started calling this "Big Data" and "the Cloud" in the mid-2000s. Particle physics was already doing it, at the Web's own home lab, a decade earlier, just without the marketing department.

From raw signals to physics

the pipeline every experiment shares

What computing actually *does*

Every experiment you'll hear about this week runs data through the same four jobs, wearing different clothes:

1. Trigger
Decide what's worth keeping, in real time

2. Reconstruction
Turn raw sensor readings into physics objects (a track, a shower, an energy deposit)

3. Simulation
Predict what the detector should see, so you know what "normal" looks like

4. Analysis
Statistics: does the data support your idea, or not?

Simulation: rolling the dice, billions of times

You can't get a "background-free" photo of a neutrino interaction to compare against. So physicists **build a fake universe inside a computer** and run it billions of times.

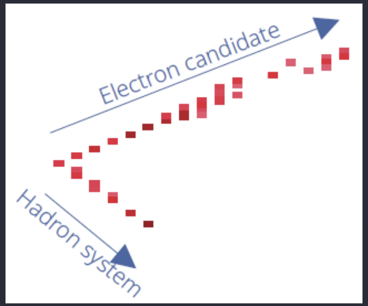
- Software like **GEANT4** tracks fake particles through a virtual copy of the real detector, obeying the known laws of physics
- Run it enough times and you learn: *"if my theory is right, here's what my detector should record"*
- Compare that prediction to the real data: differences are where new physics might hide

> This is a Monte Carlo method, named after the casino. Same core idea as counting how many times a die lands on 6: do it enough times, randomness turns into a reliable answer.

Reconstruction: turning static into a picture

Raw detector data is not a picture. It's a long list of numbers: which wire fired, at what time, with how much charge.

Reconstruction software's job is to turn that list of numbers into something a human, or another algorithm, can reason about: a track, a particle, an energy.



A real NOvA event: raw hits (small squares) reconstructed into a labeled electron track and hadron system

Same idea as turning a spreadsheet of pixel values back into a photo, except the "photo" is a 3D particle interaction, and getting it right can decide a whole result.

Software collaboration at the same scale as the physics

These experiments don't buy commercial software. They **build their own**, together, the same way open-source projects work:

- **ROOT**: the data-analysis framework nearly every HEP experiment builds on
- **art / LArSoft**: the framework NOvA and DUNE build their reconstruction on
- Thousands of contributors, across institutions, coordinated with **git and GitHub**

This afternoon's tools session (git) and Wednesday's Python/Jupyter work are the entry-level version of what runs these experiments.

Where this leads: careers

The skills this pipeline demands (**large-scale data handling, statistics, simulation, distributed computing, machine learning**) are exactly the skills industry pays for.

Physics-trained computing people go on to tech companies, finance, national labs, biotech, climate modeling, and sometimes back into physics.

You don't have to become a particle physicist for this week to be useful. The computing habits are the transferable part.

Part 2

NOvA & DUNE: a snapshot

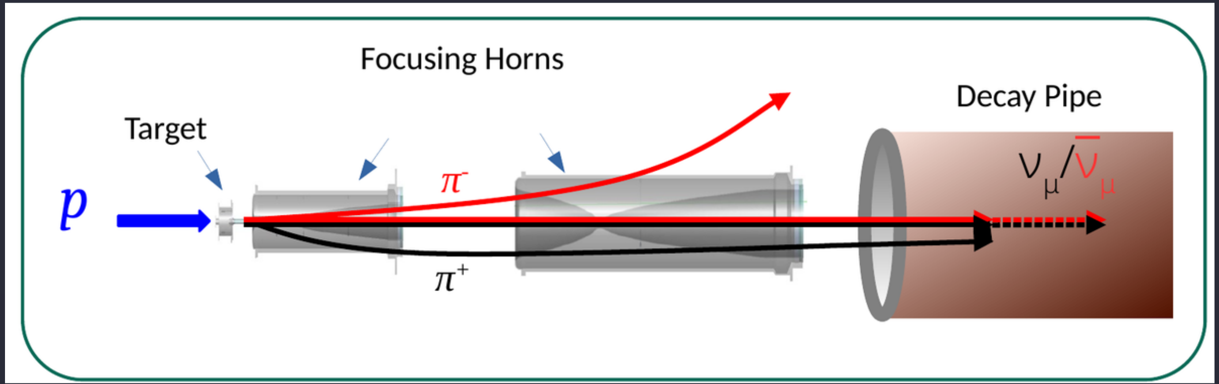
60-second neutrino refresher

- Neutrinos are the lightest, most abundant, and most antisocial known particles: trillions pass through you every second, unnoticed
- They come in **three flavors**: electron, muon, tau
- They **oscillate**: a neutrino created as one flavor can be measured as another, later, having traveled some distance
- That oscillation depends on their (tiny, nonzero) mass, one of the biggest open cracks in the Standard Model

Wednesday's neutrino oscillation tutorial goes deep. Today: just enough to understand what NOvA and DUNE are chasing.

How do you make a neutrino beam?

Gotta make the neutrinos before you can send them 800 km and catch them.



Protons from Fermilab's accelerator complex slam into a target. The resulting pions and kaons get squeezed by magnetic focusing horns into a decay pipe, where they decay in flight into a beam of (mostly) muon neutrinos.

- Fermilab's **NuMI** beamline feeds NOvA; a newer, more powerful **LBNF** beamline feeds DUNE, same idea, longer and stronger
- Flip the horns' polarity and the beam flips from neutrinos to antineutrinos, the trick both experiments use to compare the two

NOvA: running today

- **NuMI beam** at Fermilab, Illinois, fires a beam of muon neutrinos
- **Near Detector** at Fermilab measures the beam before it oscillates
- **Far Detector**: 14 kilotons of PVC and liquid scintillator, sits 810 km away in Ash River, Minnesota
- 344,064 individual cells: the actual pixel count of this morning's "camera"
- Taking data since 2014, still running today

Top of NOvA Far Detector, Ash River, Minnesota

The real Far Detector, built almost entirely by hand on the Canadian border, to catch a particle that barely interacts with anything. Photo: Fermilab.

NOvA, to scale

The 810 km baseline, with the actual buildings and detector structure behind the numbers.

Fermilab (Wilson Hall, bottom) to the Far Detector in Ash River, Minnesota (top): the same PVC-and-scintillator "camera" technology at both ends. Photo compilation: Fermilab/NOvA collaboration.

5 milliseconds of real NOvA data

nova-nusoft.fnal.gov/public/neutrinos/5ms.png

That link is a live snapshot straight from the Far Detector: every pixel is one PVC cell lighting up, color-coded by charge, refreshed continuously as real cosmic rays and beam events land.

Most of what you'll see is cosmic-ray clutter. Buried in it, occasionally, is a genuine NuMI beam neutrino, which is exactly why the trigger from this morning has to be so good at telling the two apart.

NOvA: the computing snapshot

- Every real event needs **~thousands of simulated events** to compare against, generated and reconstructed the same way as real data
- Analysis runs on **CAFAna**, NOvA's shared analysis framework, built on ROOT, used collaboration-wide so results are reproducible across sites
- Jobs run on the **Open Science Grid / FermiGrid**, the same WLCG idea from this morning, applied to Fermilab's neutrino program instead of the LHC
- **Ole Miss / MESH contributes directly** to NOvA software and analysis
- The collaboration itself: **203 scientists and engineers, 53 institutions, 9 countries**

10+ yrs
running continuously since 2014

~2 PB+
on tape, about 400,000 movies' worth, growing every day

~25M
core-hours of computing used per year

NOvA taught a computer to "see" neutrinos

Sorting real neutrino events from look-alike background used to mean comparing each event against a library of **77 million** past examples, slow, and it topped out around 2015.

NOvA taught a computer to "see" neutrinos

Sorting real neutrino events from look-alike background used to mean comparing each event against a library of **77 million** past examples, slow, and it topped out around 2015.

In 2016, NOvA swapped that step for a convolutional neural network, the same kind of AI that recognizes cats in photos, trained on 4.7 million simulated events to recognize neutrino interactions instead.

The result: 30% more usable data, instantly, from the exact same detector, just by getting better at recognizing what it already saw. It was one of the first times a neural network was used in a published particle physics result, anywhere.

Your turn: human vs. neural network

NOvA runs a **live, public event display** of real detector data, and a companion page that challenges you to identify real neutrino events by eye, the same task that convolutional neural network learned in 2016.

nova-nusoft.fnal.gov/public/nova-events/nova-events/

Let's pause here and go through a few together.

When the Grid isn't fast enough

In 2018, NOvA needed a huge statistics calculation finished in time for a conference deadline.

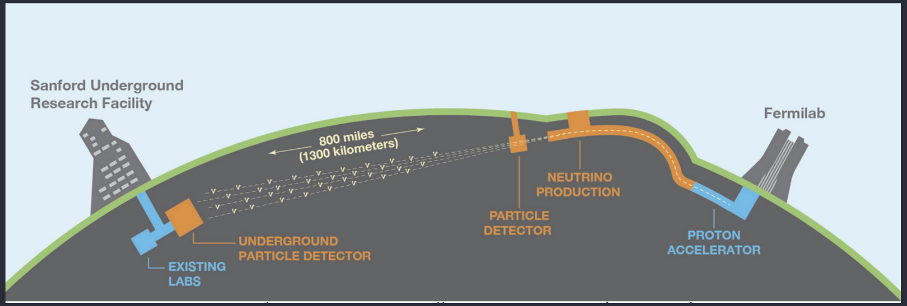
On the usual Grid
Estimated ~5 months

On a national supercomputer (NERSC)
Done in ~36 hours, using over 1 million CPU cores at once

Same idea as this morning's Grid story, one level up: sometimes 170 shared computing centers still aren't enough, and physicists rent time on one of the world's biggest supercomputers instead.

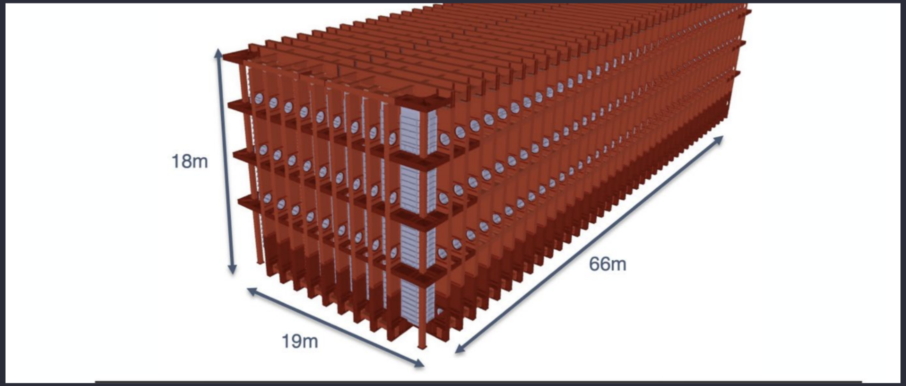
DUNE: the next generation

- **LBNF beam** at Fermilab fires toward South Dakota
- **Near Detector** at Fermilab, **Far Detector** at **SURF**, the Sanford Underground Research Facility (a former gold mine), 1,300 km away and about 1.5 km underground
- Far Detector uses liquid argon time projection chambers (LArTPCs), a more precise "camera" than NOvA's
- Currently under construction, first beam data expected around 2029

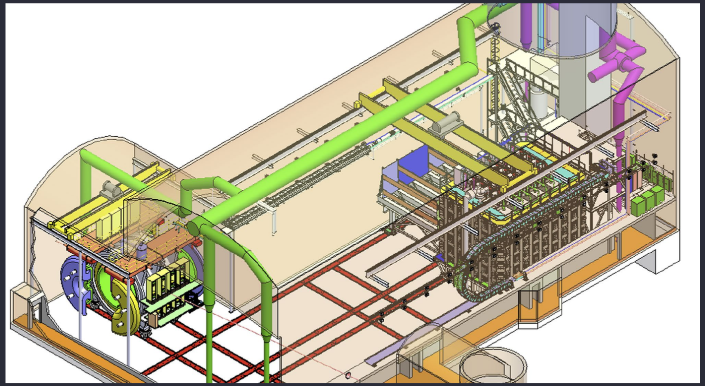


Same idea as NuMI, longer and stronger: LBNF sends the beam 1,300 km through the Earth to SURF. Source: dunescience.org.

DUNE, to scale



One Far Detector cryostat: 66 m × 19 m × 18 m. Four of these, a mile underground.



The Near Detector cavern at Fermilab: three sub-detectors, including the Muon Spectrometer (TMS) Gavin QA's components for.

DUNE: the computing snapshot

~30 PB
expected raw data per year, about 6 million movies' worth

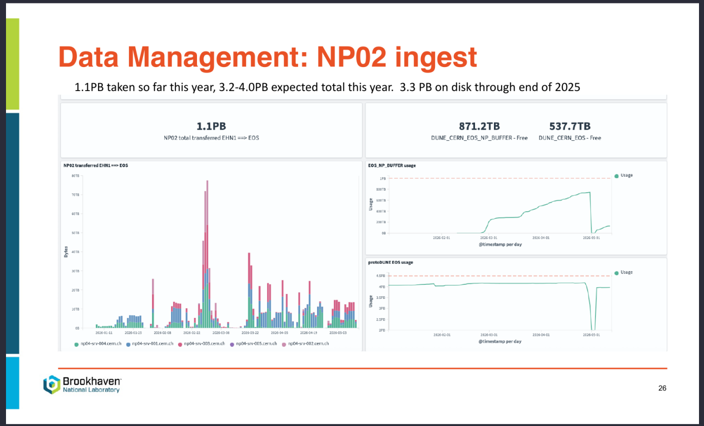
10s of millions
readout channels across the far detector modules

1,500+
collaborators, 228 institutions, 38 countries, coordinating on shared software

- LArTPCs produce detailed 3D images, not just simple pulses, so reconstruction is a harder computer-vision problem, leaning increasingly on machine learning
- DUNE also has to reconstruct in near real time for one rare, precious case: catching a supernova neutrino burst, which gives only a few seconds of warning
- Software is being built now: this is a live engineering problem, not settled infrastructure

Gavin's own role in DUNE: Software Lead, coordinating exactly this kind of collaboration-wide computing effort.

DUNE's scale, this month



DUNE Core Software & Computing plenary, May 21, 2026: real protoDUNE (NP02) data-transfer and storage-usage monitoring, not a projection.

1.1 PB moved so far this year, on track for 3.2–4.0 PB total by year's end, and that's from a prototype, not the real DUNE far detector.

Quick quiz: NOvA or DUNE?

Shout out an answer for each one:

- Running since 2014, still taking data today
- Detector technology is a liquid argon time projection chamber
- Far detector sits 1.5 km underground in a former gold mine
- Baseline is 810 km, from Fermilab to Minnesota
- ~30 PB/year expected, once it's fully running

Quick quiz: NOvA or DUNE?

Shout out an answer for each one:

- Running since 2014, still taking data today
- Detector technology is a liquid argon time projection chamber
- Far detector sits 1.5 km underground in a former gold mine
- Baseline is 810 km, from Fermilab to Minnesota
- ~30 PB/year expected, once it's fully running

NOvA, DUNE, DUNE, NOvA, DUNE. If you got 4 or 5, you were paying attention this afternoon.

Ole Miss is part of this Grid

This isn't just something happening at Fermilab and CERN. It's happening in this building.

- The Ole Miss group runs an active Open Science Grid (OSG) site for Belle II's distributed computing
- Ole Miss physicists and students already process real DUNE, NOvA, and Belle II data day to day
- A new local cluster is being installed right now: 144 CPU cores, 1 TB RAM, 864 TB storage, built to serve both DUNE and Belle II
- A May 2026 workshop here at UM brought DUNE, NOvA, and Belle II computing people together to set it up, some of the same people running your week

The "Grid" from this morning isn't a textbook diagram. There's a node of it down the hall, right now, for the exact two experiments you're about to meet.

The common thread

Both experiments (one running for over a decade, one still under construction) live or die by the same four-step pipeline from this morning: trigger → reconstruction → simulation → analysis.

Neither NOvA nor DUNE is only a detector story. It's a computing story with a detector attached, and that's true of nearly every modern particle physics experiment. Which is exactly why this week starts with tools, not equations.

Where you fit in this week

Today
Unix, git, GitHub: the actual tools these collaborations run on

→

Tue
Belle II hadronic physics, team project launch

→

Wed
Neutrino oscillations, hands-on tutorial

→

Thu–Fri
Analyze real(-ish) data, present it like a physicist would

You are not doing toy exercises this week. You're using the same tools, the same culture, and pieces of the same pipeline that run NOvA and DUNE right now.

This afternoon: Python and Jupyter

Everything so far has been *watching* the pipeline. Now you drive.

- One notebook, start to finish: Jupyter basics, core Python, your first physics histograms, and a real data cut
- Work at your own pace, pair up if it helps, but be able to explain your own code
- Errors are expected. Reading the error is the skill, not avoiding it

Repo: github.com/mesh-collab/epic-curriculum
Notebook: `notebooks/00_python_and_jupyter_intro.ipynb`

Get set up

```
# new to uv? install it first (skip if you already have it):
curl -LsSf https://astral.sh/uv/install.sh | sh     # Mac/Linux
irm https://astral.sh/uv/install.ps1 | iex          # Windows

git clone https://github.com/mesh-collab/epic-curriculum.git
cd epic-curriculum
uv sync
uv run jupyter lab
```

Open `notebooks/00_python_and_jupyter_intro.ipynb` and check the kernel in the top-right says **EPIC Curriculum**.

Run the first cell (`Shift+Enter`). If you see `Hello, particle physics!` you're ready to go.

Questions?

Gavin S. Davies · gsdavies@olemiss.edu