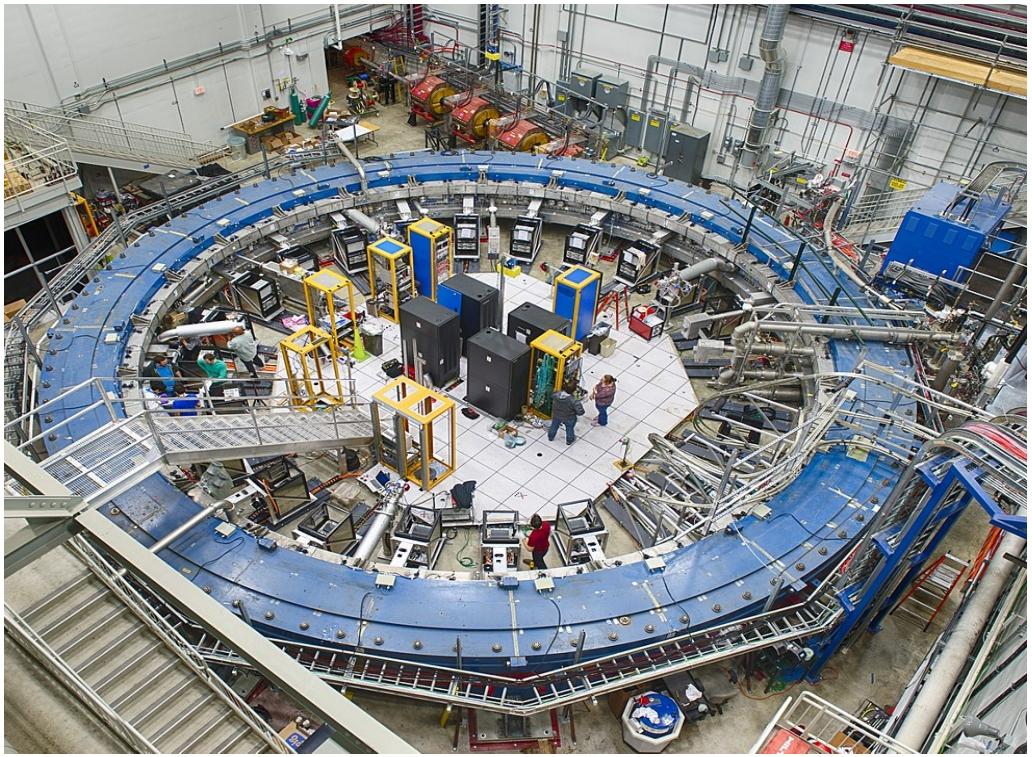


Small experiment, Big Data:
the data production of the Muon g-2 Experiment

41st **I**nternational **C**onference
on **H**igh **E**nergy **P**hysics
8 July 2022

Paolo Girotti (INFN Pisa)
on behalf of the Muon g-2 collaboration

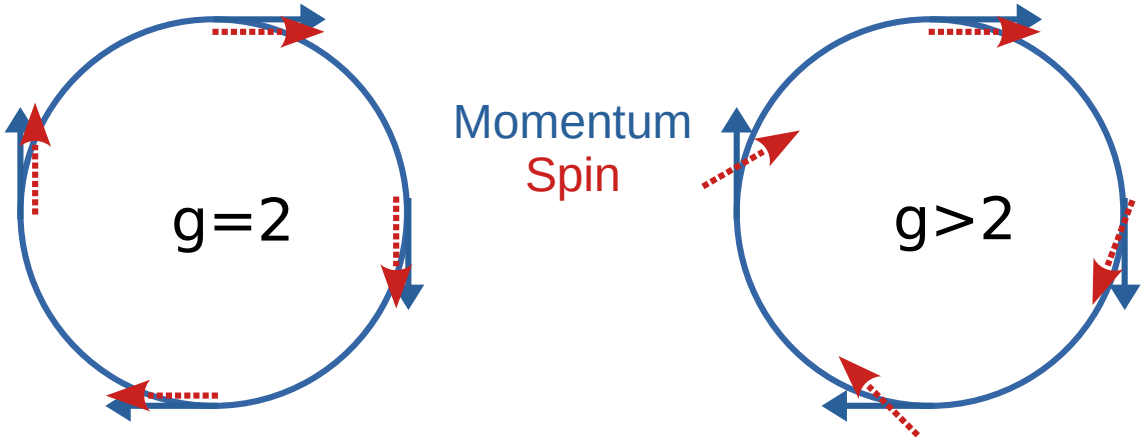
Outline

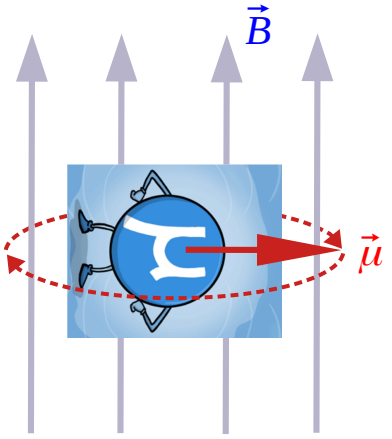
- Muon g-2 in a nutshell
- Small experiment, big data
- Reconstructing the data
- Technical tools
- Production challenges

Muon g-2 primer

- The Muon g-2 Experiment (E989) at Fermilab measures the muon anomalous magnetic moment very precisely
  - Goal: repeat and improve BNL (2001) measurement with 4x precision
- A beam of polarized muons circulates inside a storage ring
- The magnetic field strength is measured by NMR probes
- Decay positrons carry the precession signal and are detected by calorimeters

$$\vec{\omega}_a = \vec{\omega}_s - \vec{\omega}_c = -\left(\frac{g-2}{2}\right)\frac{e\vec{B}}{m} \equiv -a_\mu \frac{e\vec{B}}{m}$$





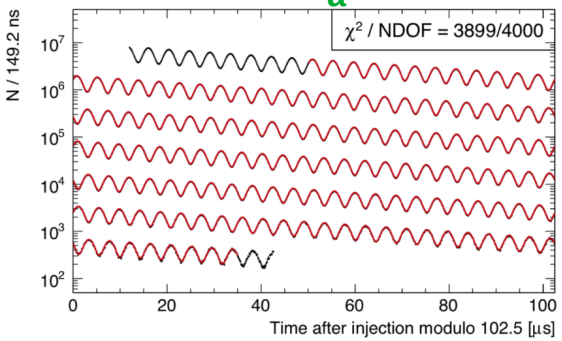
The muon precession frequency is influenced by the interactions with **all** possible virtual particles

Master formula

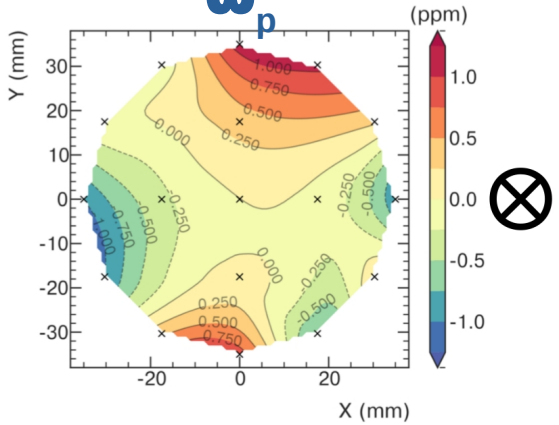
$$a_\mu = \frac{\omega_a}{\tilde{\omega}'_p(T_r)} \frac{\mu'_p(T_r)}{\mu_e(H)} \frac{\mu_e(H)}{\mu_e} \frac{m_\mu}{m_e} \frac{g_e}{2}$$

Constants known from other experiments with high precision

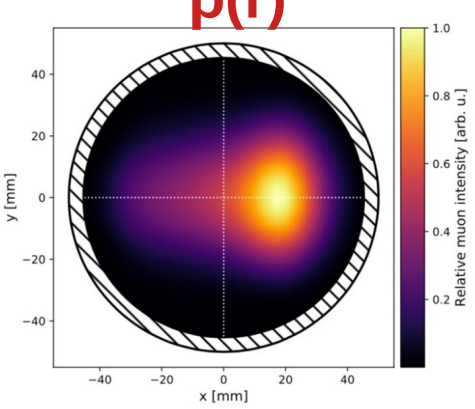
ω_a



ω_p



$\rho(r)$

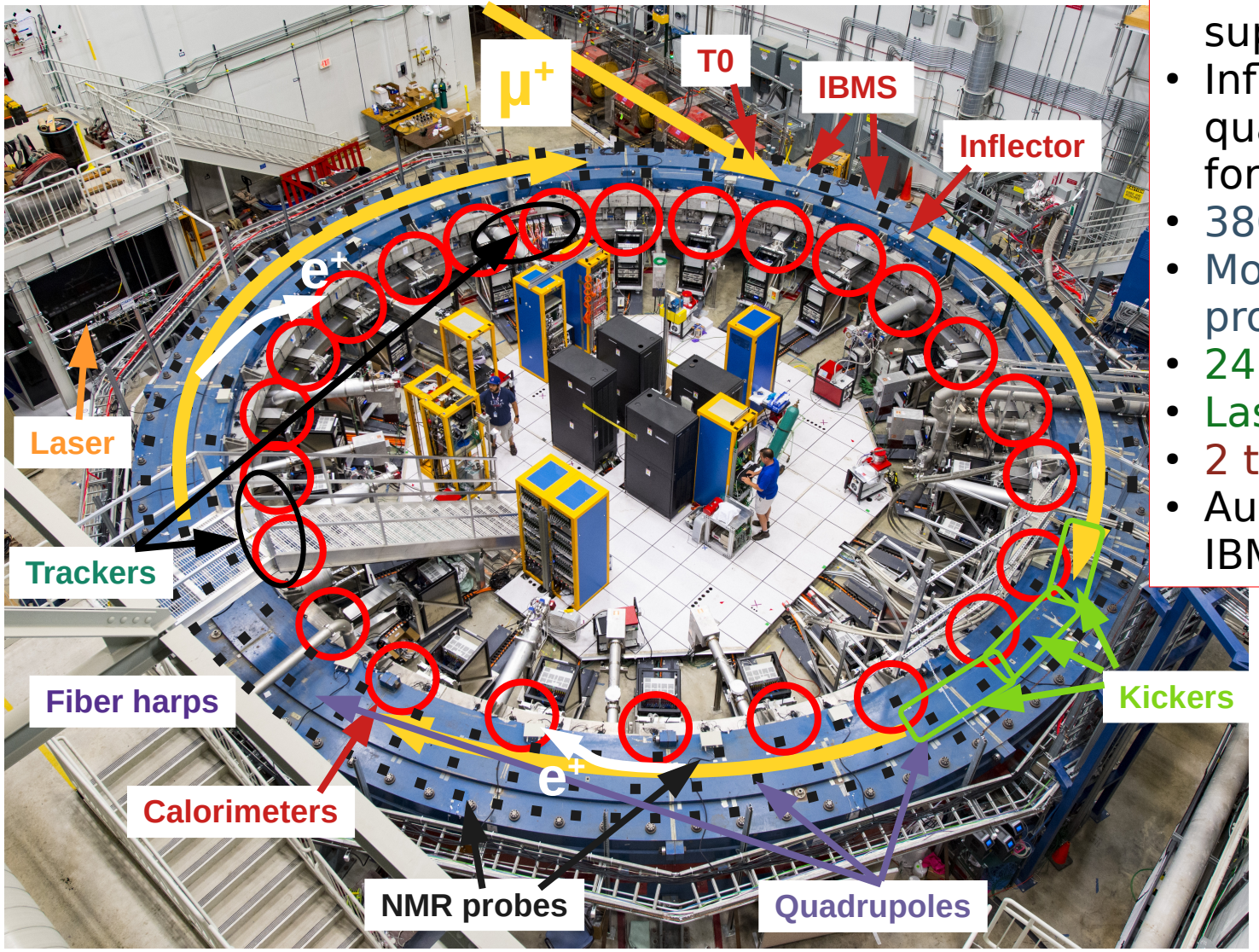


Three measured quantities:

- ω_a: Muon anomalous precession frequency
- ω_p: Larmor precession frequency of protons in water (mapping B)
- ρ_r: Muon distribution in the storage ring

Goal: measure a_μ with 140 ppb accuracy (100 stat + 100 syst)

The Muon g-2 Experiment

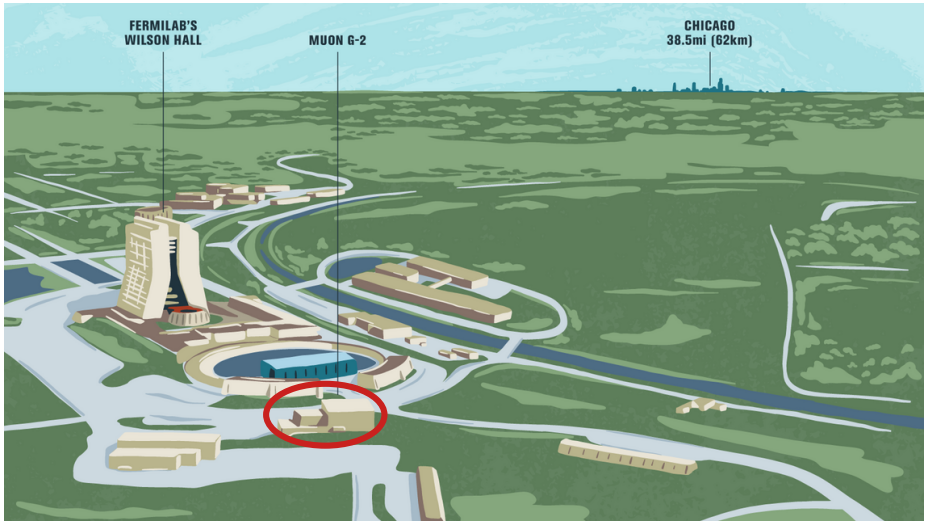


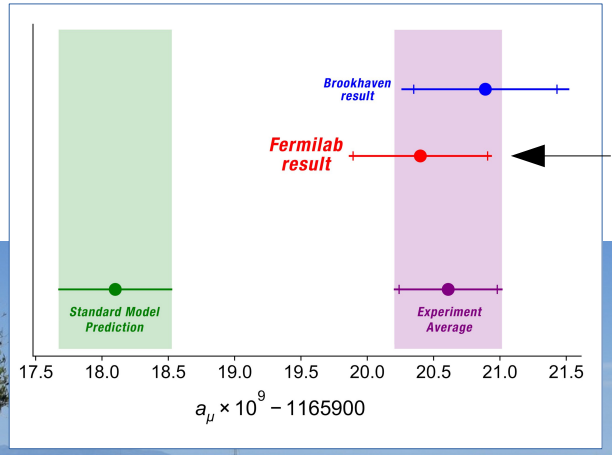
- 15 meter wide dipole superconducting magnet
- Inflector, kickers, quadrupoles, collimators for beam insertion
- 386 NMR probes
- Moving trolley with 17 ω_p probes
- 24 calorimeters ω_a
- Laser calibration system
- 2 tracker stations $\rho(r)$
- Auxiliary detectors: T0, IBMs, Fiber harps

A lot of components and detectors
→ a lot of **data**!

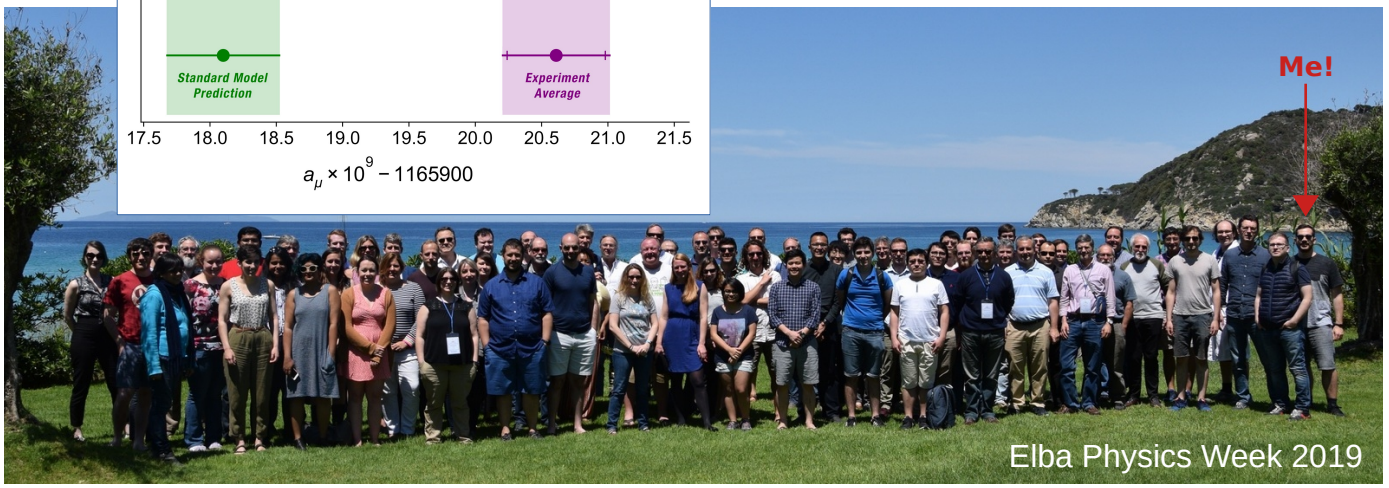
“Small” experiment

- Muon g-2 is a relatively small HEP experiment
- Part of Fermilab Muon Campus
- ~200 collaborators from 7 countries
- Very young community (40% under 35 yo)





First physics result
7 April 2021

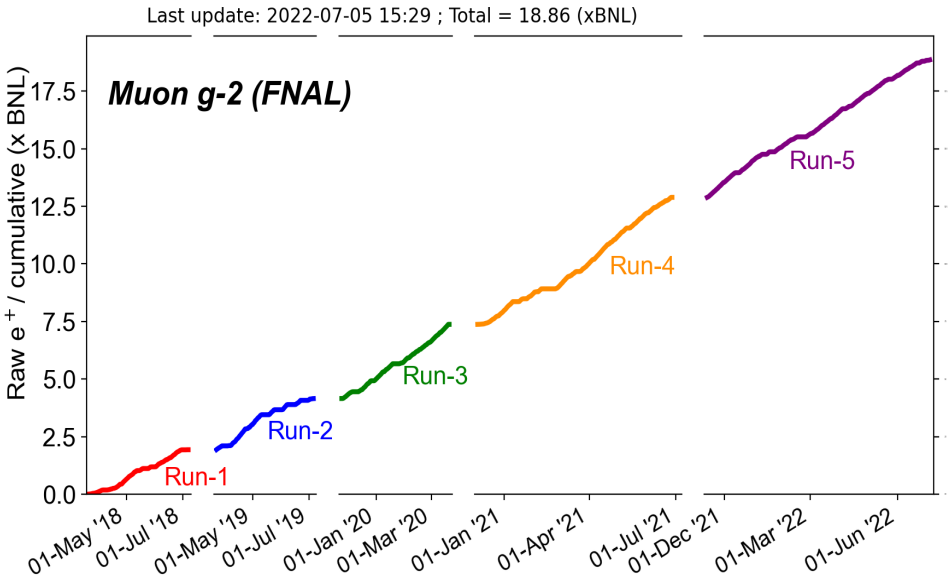


“Big” data

- 16 bunches of muons every 1.4 s
- Muons circulate for 700 μs, generating ~2000 positron hits on calorimeters
- 1296 calorimeter channels @800 MHz
- 2048 tracker channels @400 MHz
- One event is a stored bunch
- One subrun (2 GB) every ~8 seconds
- 24/7 running for 4-7 months each year
- Currently running 5th year of production

  → Collected **~6.5 PB** of raw data so far

  → Plus simulation ~2 PB

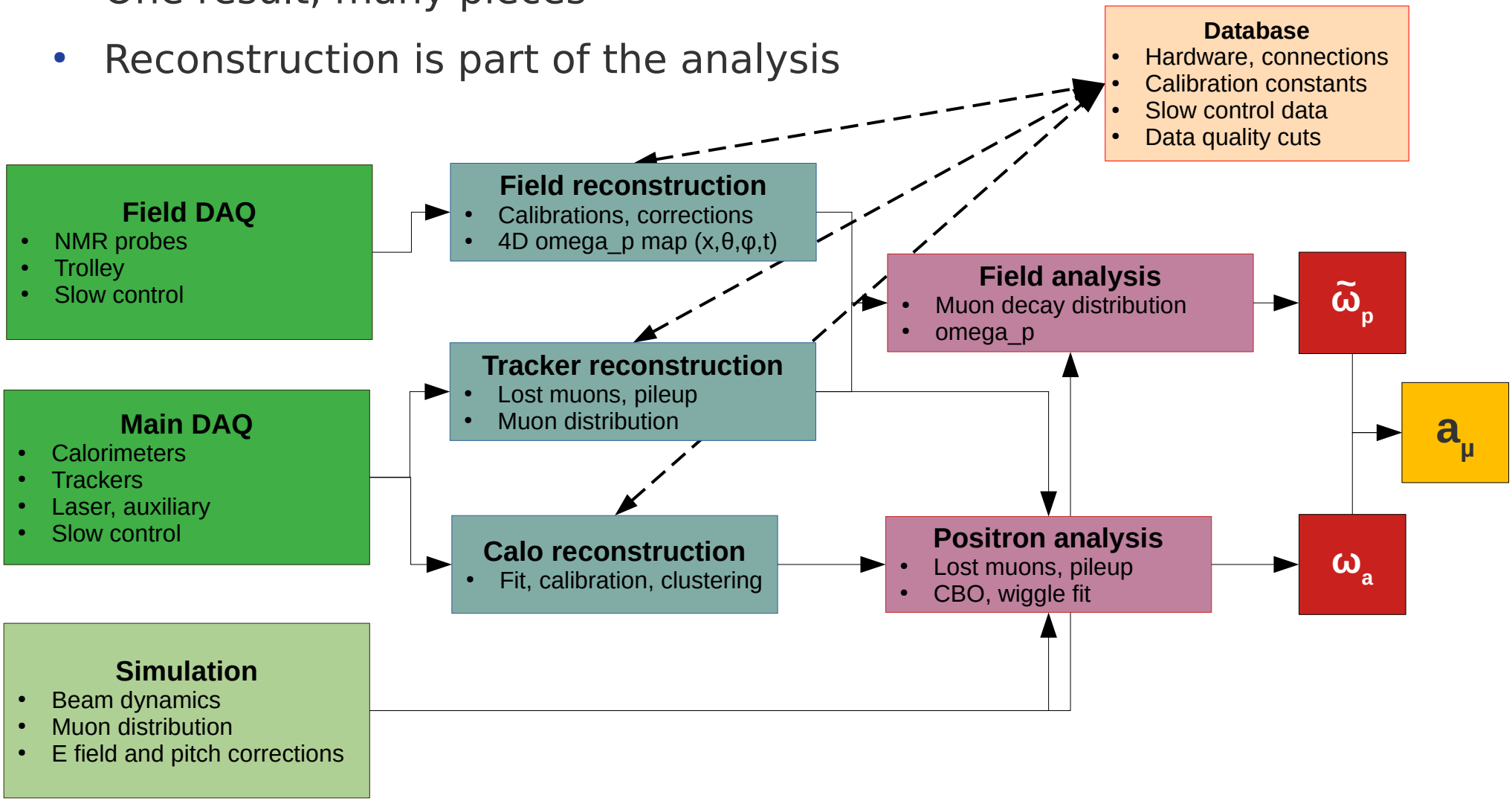


	Raw	Offline	Subruns
Run1	0.4 PB	0.3 PB	258108
Run2	0.5 PB	0.5 PB	294207
Run3	1.0 PB	1.2 PB	559377
Run4	2.0 PB	1.7 PB*	1121703
Run5	2.5 PB*	-	1423000*

*Under production

From raw to a_μ

- One result, many pieces
- Reconstruction is part of the analysis

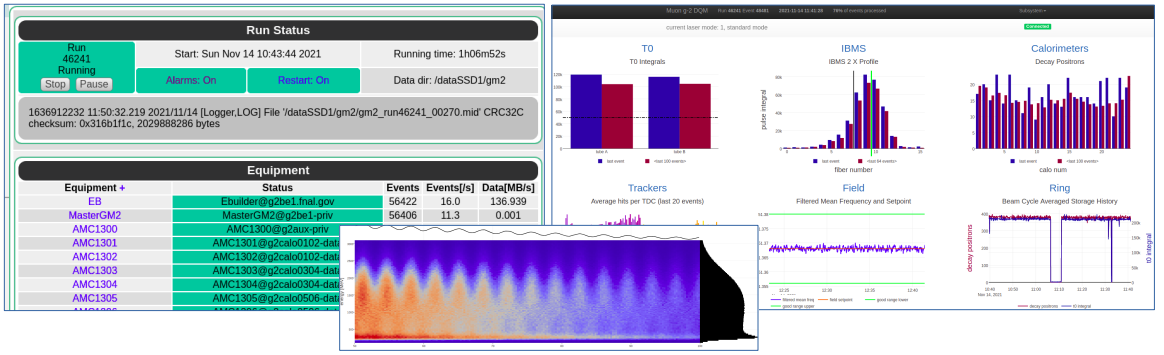


Data reconstruction

Raw data

Online

- Fraction of data
- Instantaneous plots
- DAQ and DQM

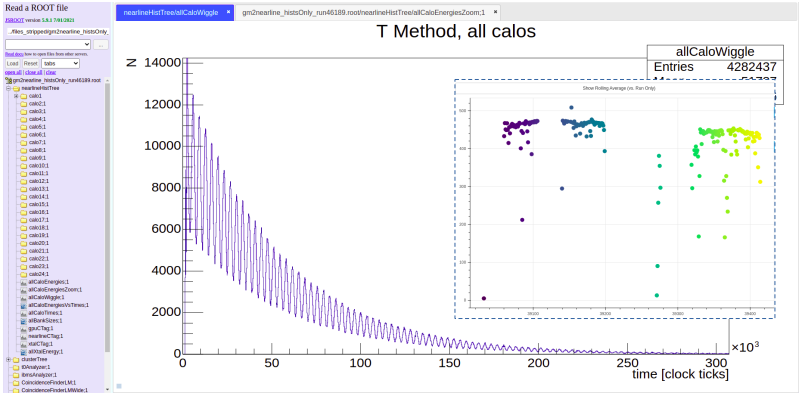


MIDAS

plotly

Nearline

- All data analyzed and plotted
- Standard calibration
- Useful for quick physics analysis
- Results in ~30’



ROOT Data Analysis Framework

bokeh

Offline

- Complete and fully calibrated reconstruction
- Involves pre- and full- productions
- Multiple reconstruction techniques
- Version-controlled, produced with POMS on grid
- Final results in >1 year

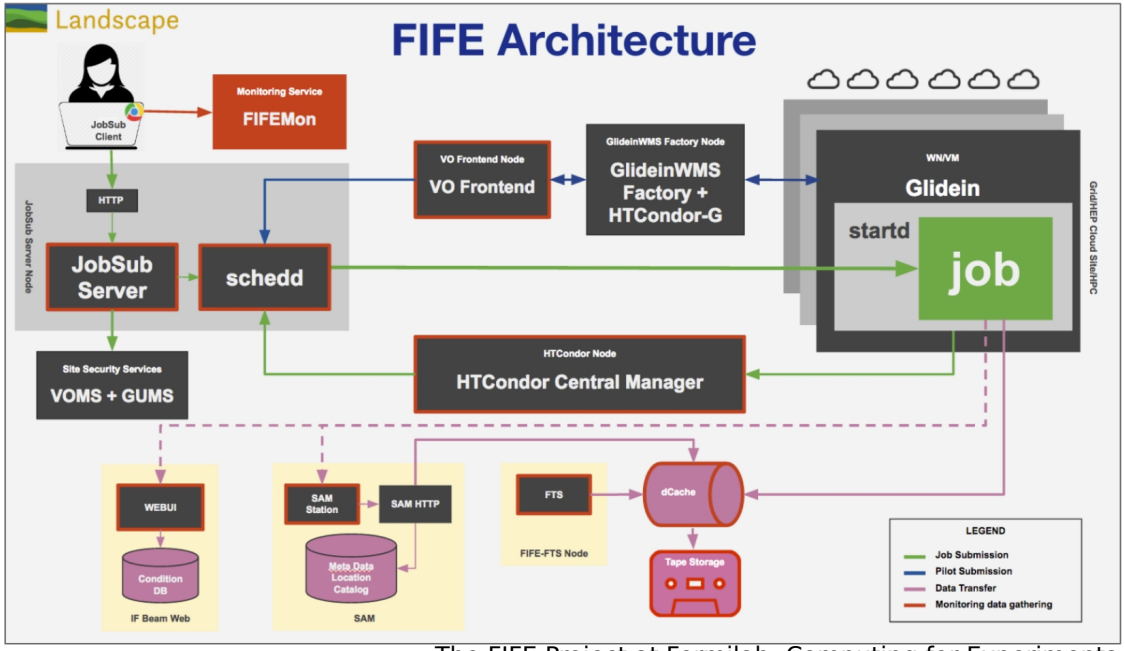
Pre-production → Calibration → Full-production → DQC

art

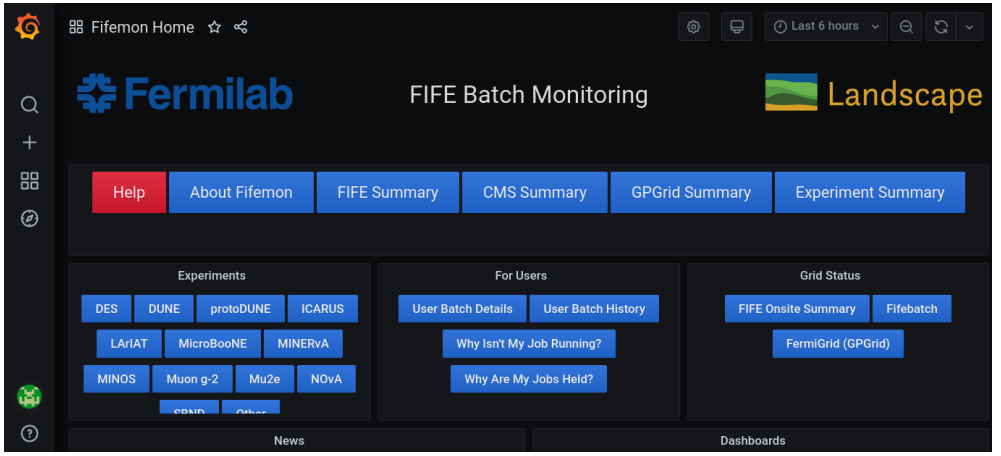
Production Operations Management System

FIFE environment

- **F**abr**I**c for **F**rontier **E**xperiments toolkit for data management and job submission
- SAM for management of file metadata
- Samweb for job visualization
- FTS-IFDH system for high speed transfer between tape, cache, and grid nodes
- Jobsub scheduler via HTCondor
- Monitoring tools on Grafana
- Designed for Fermilab infrastructure
- Well suited for Muon g-2 needs



The FIFE Project at Fermilab: Computing for Experiments
https://doi.org/10.22323/1.282.0176

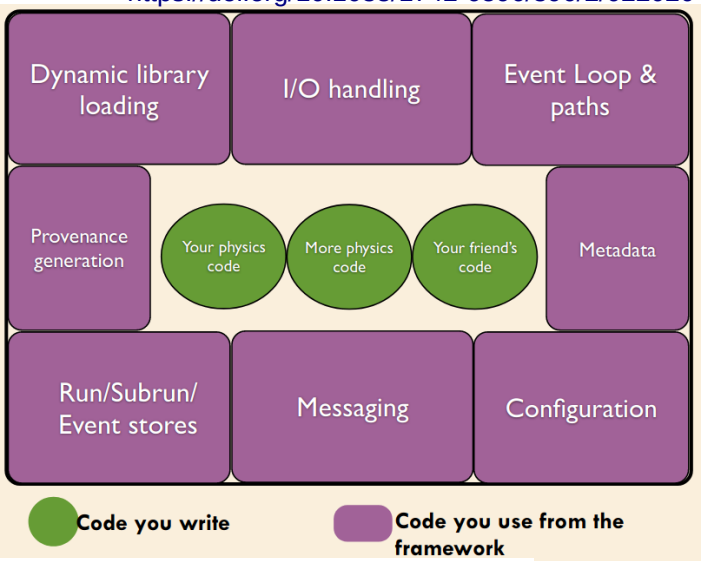


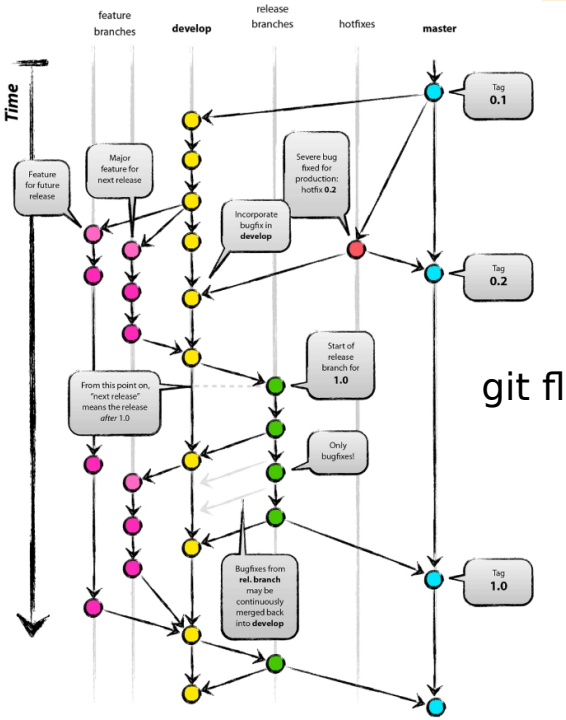
Offline reconstruction

- We use ***art*** framework
  - Highly modular
  - Made for HEP physics
  - Seamless transition between simulation and real data
- **ROOT/C++** based software
- Code repository with version control
  - One of the first Fermilab experiments to use **git**
  - Hosted by Redmine

https://art.fnal.gov/

https://doi.org/10.1088/1742-6596/396/2/022020

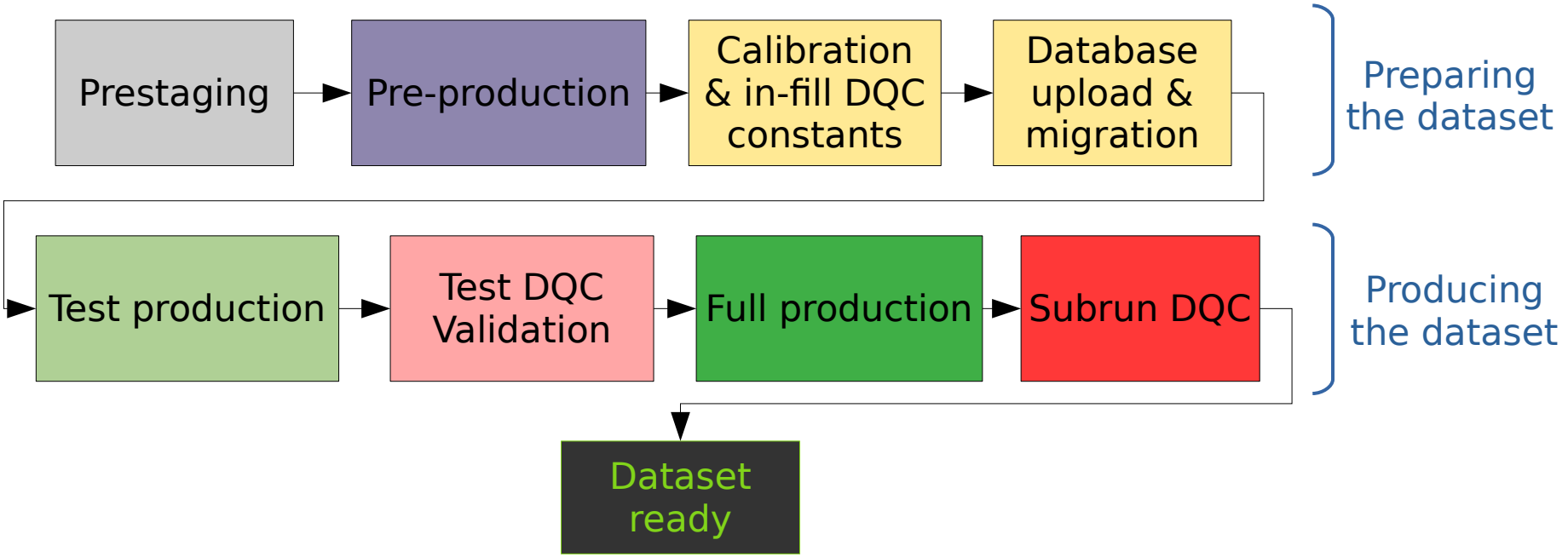




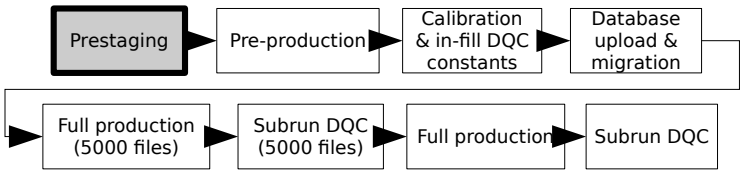
git flow

Production workflow

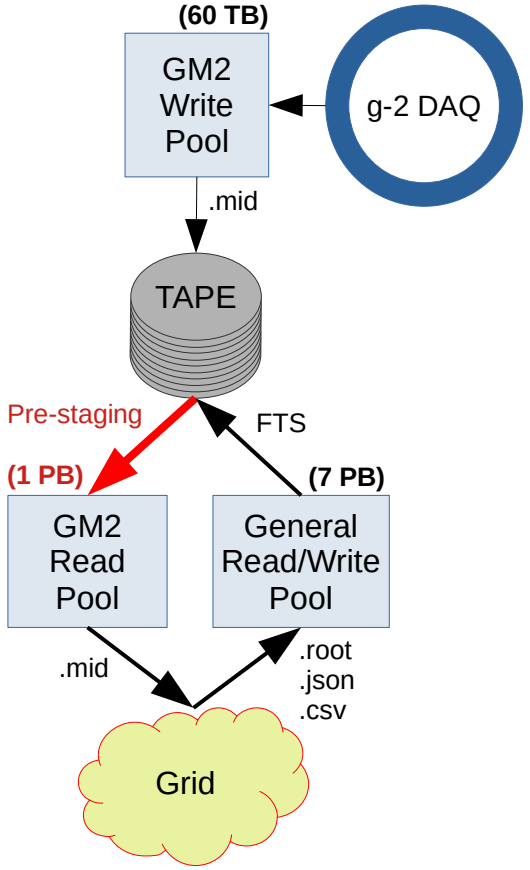
- Production of datasets happens in multiple steps
  - Multiple datasets proceeding in parallel
- **Pre-production** is a light processing of the data with the aim to obtain calibration constants and *per-bunch* DataQualityCuts (DQC)
- **Full-production** is the complete production of calibrated data and *subrun* DQC



Data staging

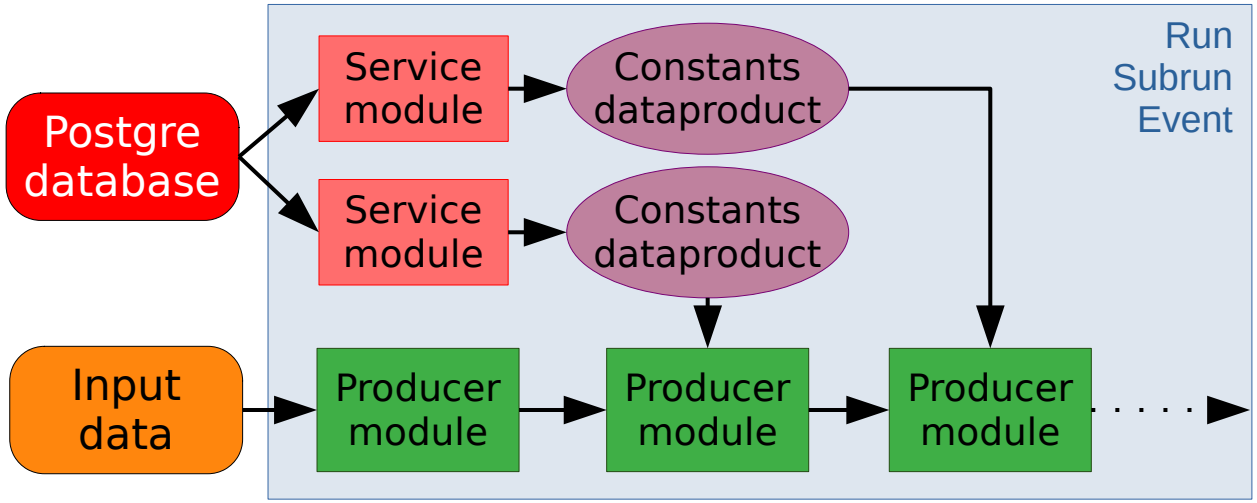


- First step is to pre-stage the dataset
  - Typical dataset size: ~100k files, ~200 TiB
- Dedicated 1 PB GM2 Read pool for prestaging raw files
- Careful calendar management of datasets by offline managers
- At any given time, typically:
  - 3 datasets are staged and under reconstruction
  - 3 datasets are being prestaged
  - 3 datasets are leaving the pool
- Depending on infrastructure load, 3-6 days to prestage 200 TB of data from tape to disk
- New files from grid get stored on a public ~7 PB pool and get transferred on tape via FTS



Database

- Database extensively used for both online DAQ configuration and offline data reconstruction
- Recent transition for having critical reconstuction parameters and calibrations to a condition database supported by Fermilab SCD
- **PostgreSQL** interface
- **I**nterval **O**f **V**alidity based on run, subrun, event numbers

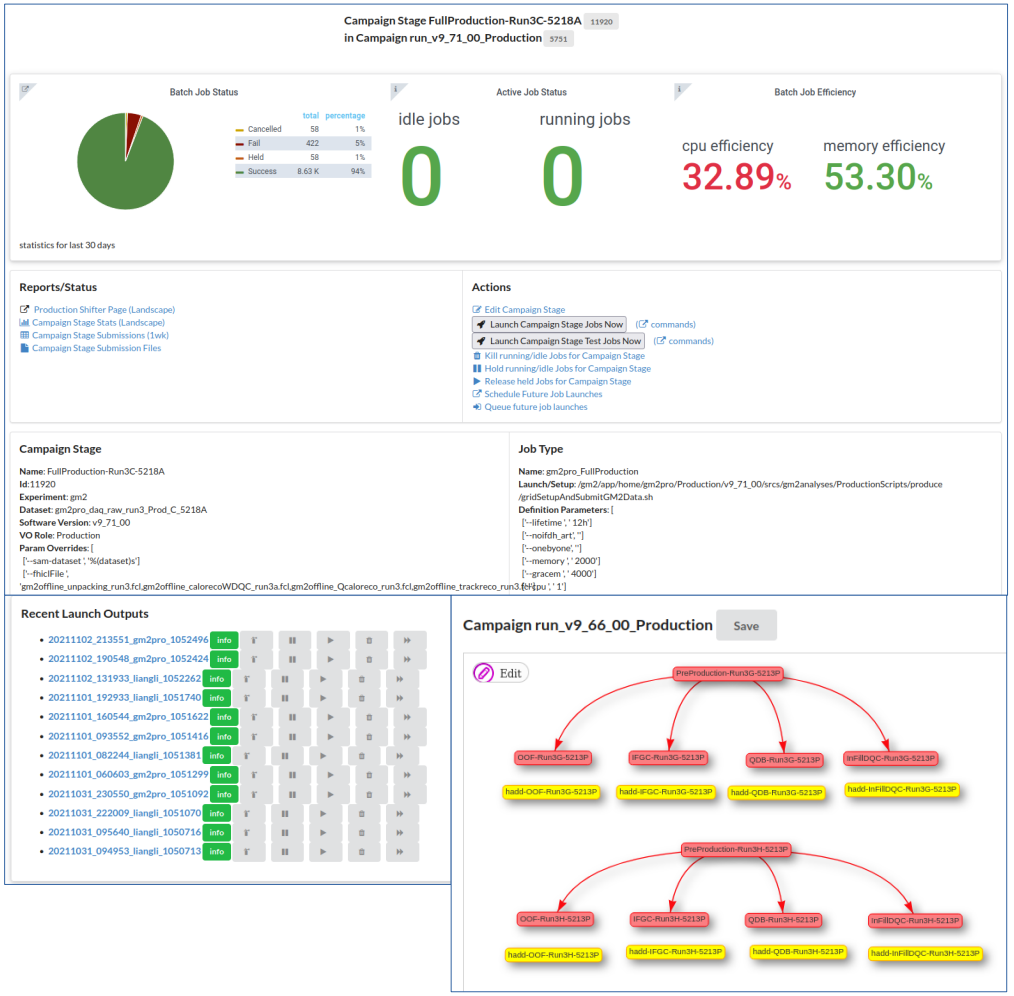


Terminal interface with **psql**

```
gm2_conditions_dev=> select calonum,xtalnum,oofcorrec
tion from oof_correction_data,oof_correction_iovs whe
re begin_time = 243370196 and calonum=1 and iov_id=10
 limit 10;
 calonum | xtalnum | oofcorrection
---------+---------+---------------
       1 |       0 |      0.932184
       1 |       1 |       0.94373
       1 |       2 |      0.933066
       1 |       3 |      0.928901
       1 |       4 |      0.930891
       1 |       5 |      0.918331
       1 |       6 |      0.940355
       1 |       7 |      0.901765
       1 |       8 |       0.93363
       1 |       9 |      0.939446
(10 rows)
```

Running on GRID

- Production jobs run on **Grid**
  - Both onsite and offsite
  - 6000 reserved slots for g-2
  - Data I/O is handled by SAM, IFDH and FTS
- **POMS** interface
  - Useful tool to keep track of which version a dataset has been reconstructed with
  - Automatic slicing of dataset, recovery of failed jobs and grid submissions with crontabs
  - Frequent monitoring together with FIFE tools is useful to ensure job success

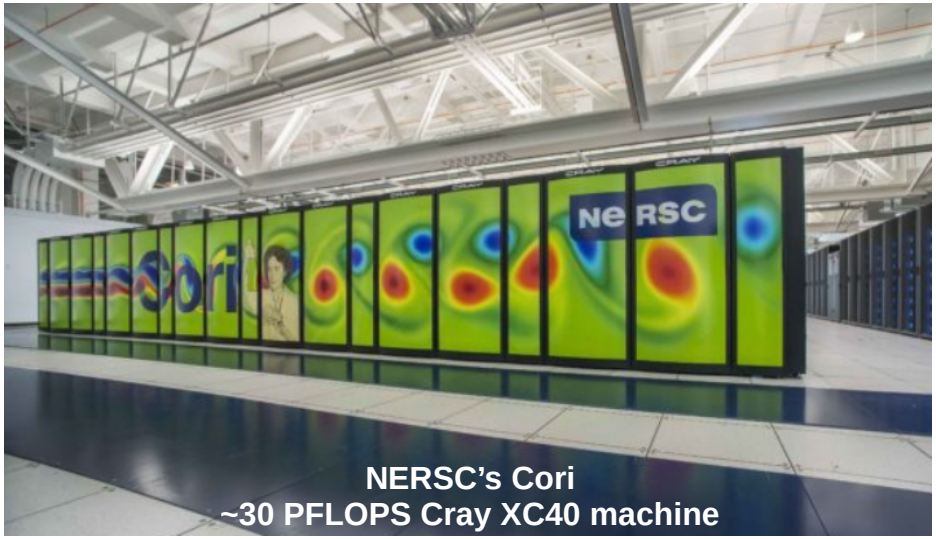


POMS roles
- Gm2pro for managers
- Gm2shifter for shifters

Simulation

- Important piece for g-2 physics, parallel to data taking
- Many simulation packages for the various parts of the beamline and the storage ring
  - MARS (Proton target)
  - BMAD (beamlines & g-2)
  - G4beamline (beamlines)
  - Gm2ringsim (injection & g-2)
  - COSY (g-2 storage ring)

CPU intensive



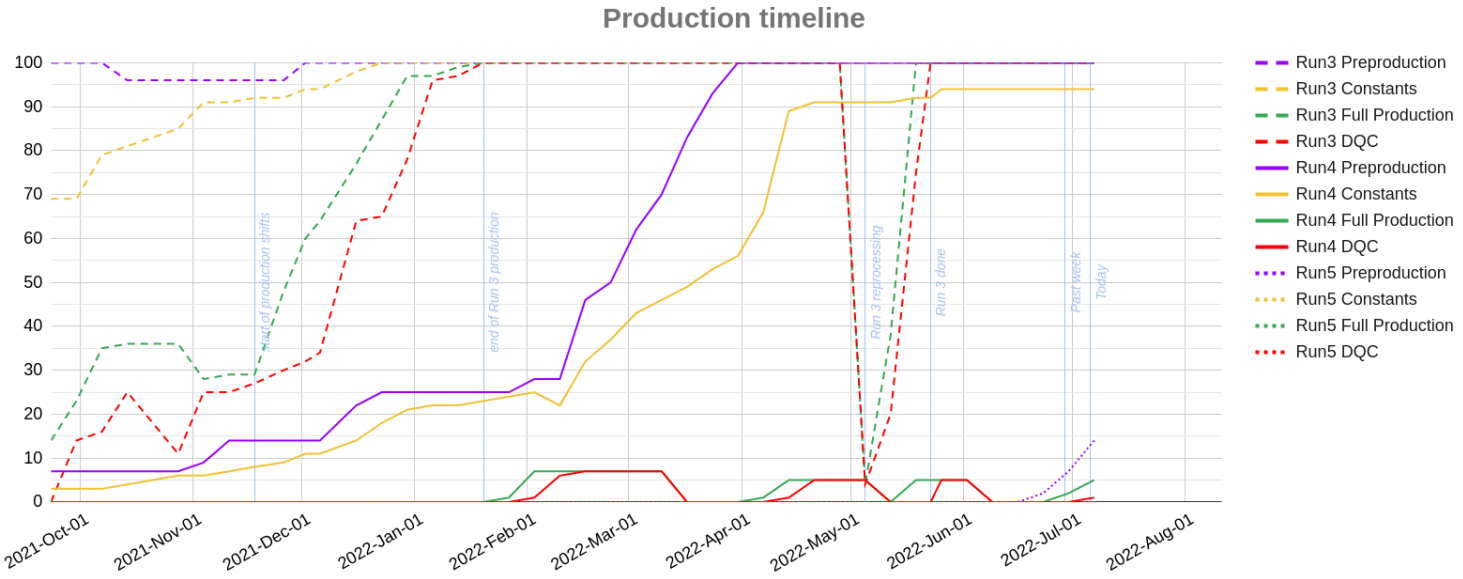
NERSC's Cori
~30 PFLOPS Cray XC40 machine

- Complete simulation of fields, beam dynamics, muon decays, and detector interactions in the storage ring
- Precision tracking over 200 km (up to 5000 turns in g-2 ring)
- Making use of HPC computing **@NERSC**
  - First production completed with 10 B events

HEPCloud

Production shifts

- Notable innovation in the Muon g-2 workflow
- All the steps of data production are now pushed and monitored by production shifters, with institutional quota
  - Procedures leveraged from online shift operations
- Proved to be fundamental for keeping production on time for the experiment publications
- Number of weekly shifters lowered from 8 to 3 during the past year, thanks to improvements in the workflow efficiency



Challenges

A list of **challenges** that we faced and how we **solved** them

- Many large datasets to process
  - Separate pre- from full- production, enstablish rolling production scheme
  - Prestaging schedule with increased dedicated GM2 disk pool
  - Multiple datasets now processed merged together
- A lot of calibration constants, hard bookkeeping
  - Move everything to psql database
- Lack of workforce to produce the data
  - Introduce production shifters → expand to collaboration (thanks to the new Production-shifter role on POMS too)
- Memory usage of full-production
  - Managed to keep most jobs under 2 GB of memory, which is the “unit” of grid slots. More memory would require more (wasted) cpu

Conclusions

- The Muon g-2 Experiment is a relatively small-sized experiment
- Nonetheless, being a precision experiment, the final result requires reconstructing and analysing multiple PetaBytes of raw data
- Reconstruction and analysis are not really independent – multiple reconstructions are sometimes required (especially for the first Runs)
- Production is now a collaboration-wide effort with shifts
- New results expected early next year!

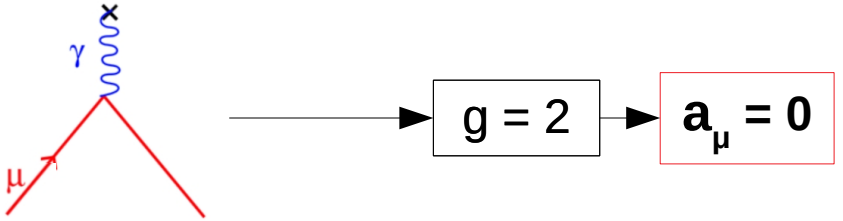
Thank you for the attention

Paolo Girotti | pgirotti@fnal.gov

Backup

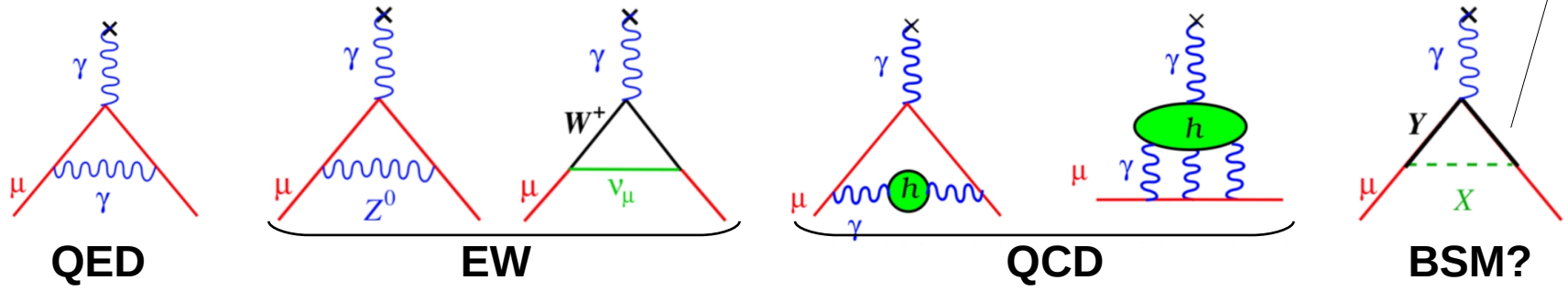
The muon anomaly

- Deviation from 2 arise from quantum loop corrections
- The simple muon-photon interaction gives g=2



In the vacuum, the muon interacts with all the possible virtual particles in all the possible ways
Undiscovered particles could affect the value of g-2!

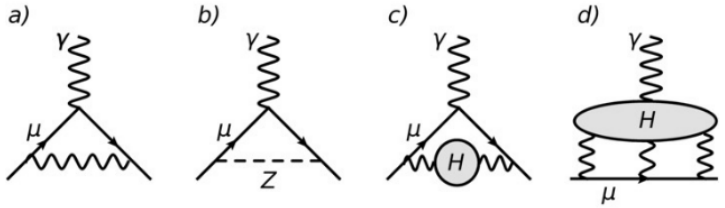
- But considering corrections:



QED EW QCD BSM?

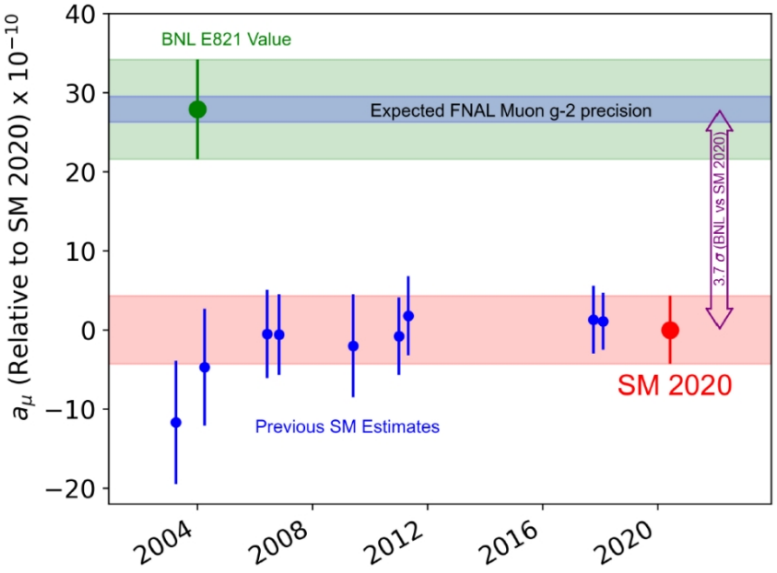
a_μ = 0.0011658… +0.000000001536... +0.000000069383… +?

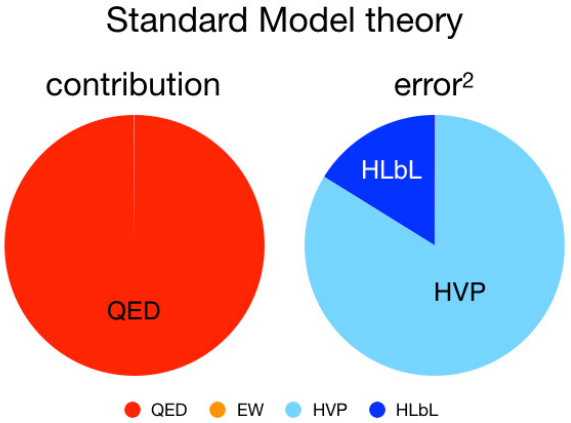
Status before 2021



Source	Value (a_μ x 10^{-11})	Error
a) QED	116 584 718.9	0.1
b) EW	154	1
c) HVP	6845	40
d) HLBL	92	18

Muon g-2 Theory Initiative arXiv:2006.04822

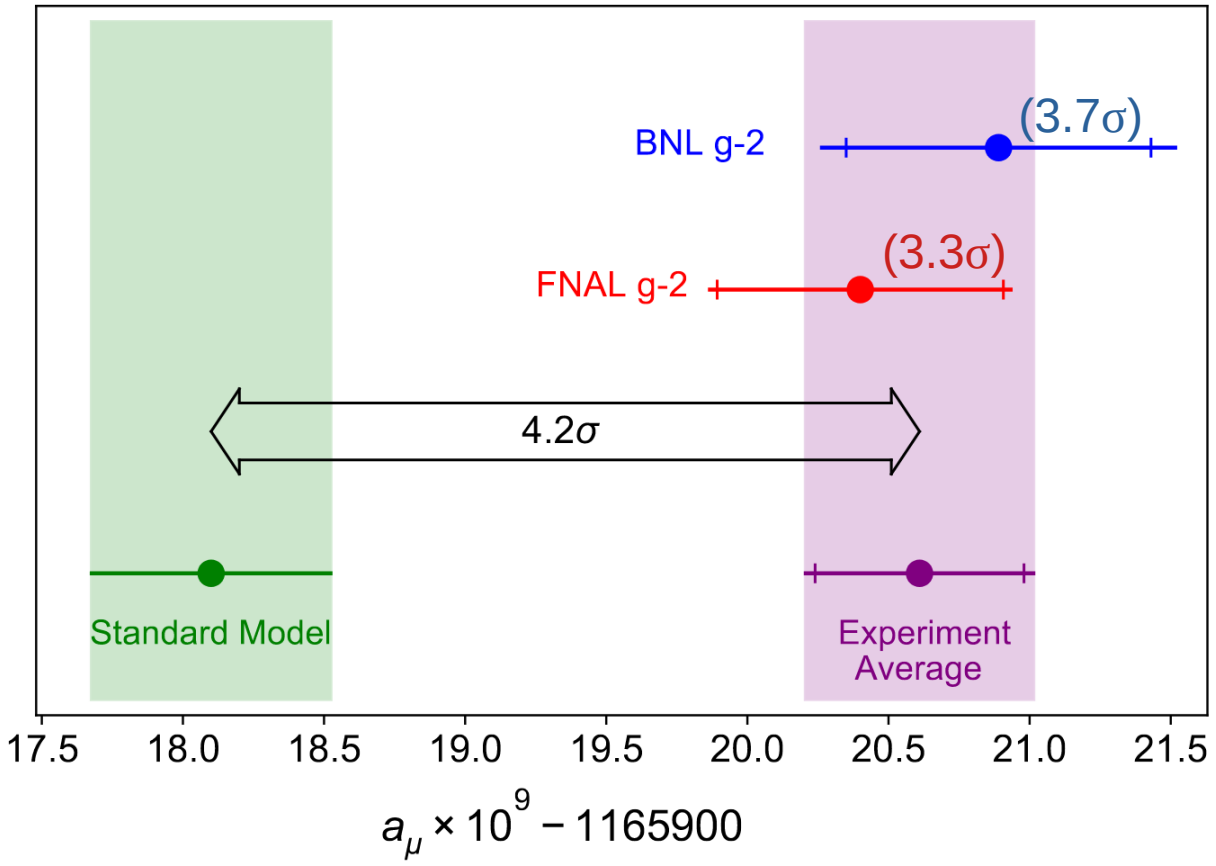




QCD loops account for:
- **0.006%** of the contribution
- **99.95%** of the uncertainty

- a_μ measured at Brookhaven National Lab (BNL, 2006), and the result differs by **3.7σ** with respect to SM prediction
- Bringing the magnet from BNL to Fermilab's powerful accelerator beam
  - Goal: reduce the error by a factor of 4 to 140 ppb

Run 1 result



Quantity	Correction Terms (ppb)	Uncertainty (ppb)
ω_a^m (statistical)	–	434
ω_a^m (systematic)	–	56
C_e	489	53
C_p	180	13
C_{ml}	-11	5
C_{pa}	-158	75
$f_{\rm calib}\langle\omega_p(x,y,\phi)\times M(x,y,\phi)\rangle$	–	56
B_k	-27	37
B_q	-17	92
$\mu_p'(34.7^\circ)/\mu_e$	–	10
m_μ/m_e	–	22
$g_e/2$	–	0
Total systematic	–	157
Total fundamental factors	–	25
Totals	544	462

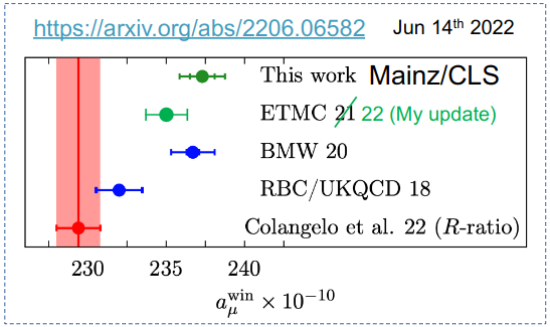
a_μ (FNAL) = 116 592 040(54) × 10^{-11} (0.46 ppm)

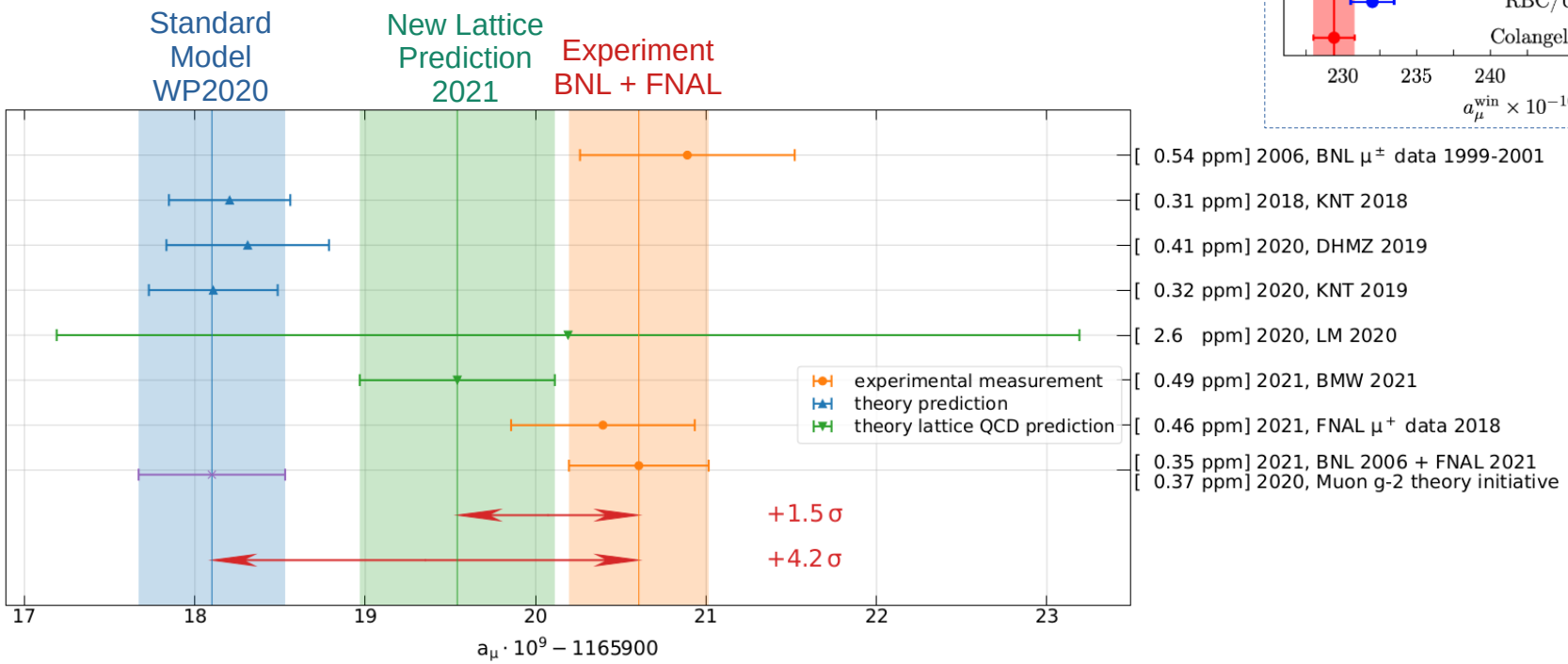
a_μ (Exp) = 116 592 061(41) × 10^{-11} (0.35 ppm)

a_μ (Th) = 116 591 810(43) × 10^{-11} (0.37 ppm)

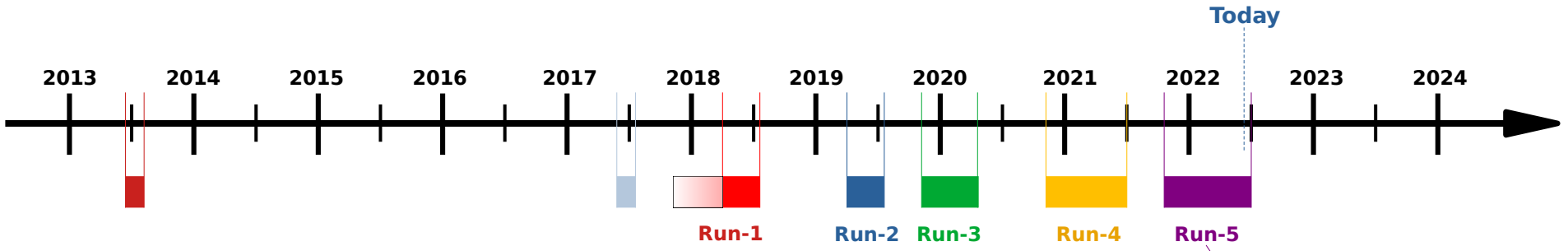
Lattice results

- New **Lattice QCD** estimation of the hadronic contribution
  - Ab-initio calculation, does not rely on experimental data
  - First lattice result to provide an estimate with the error comparable to the dispersive evaluations (< 1%)
  - Currently <u>confirmed</u> by other lattice groups
  - "*New g-2 puzzle*" in the theory side

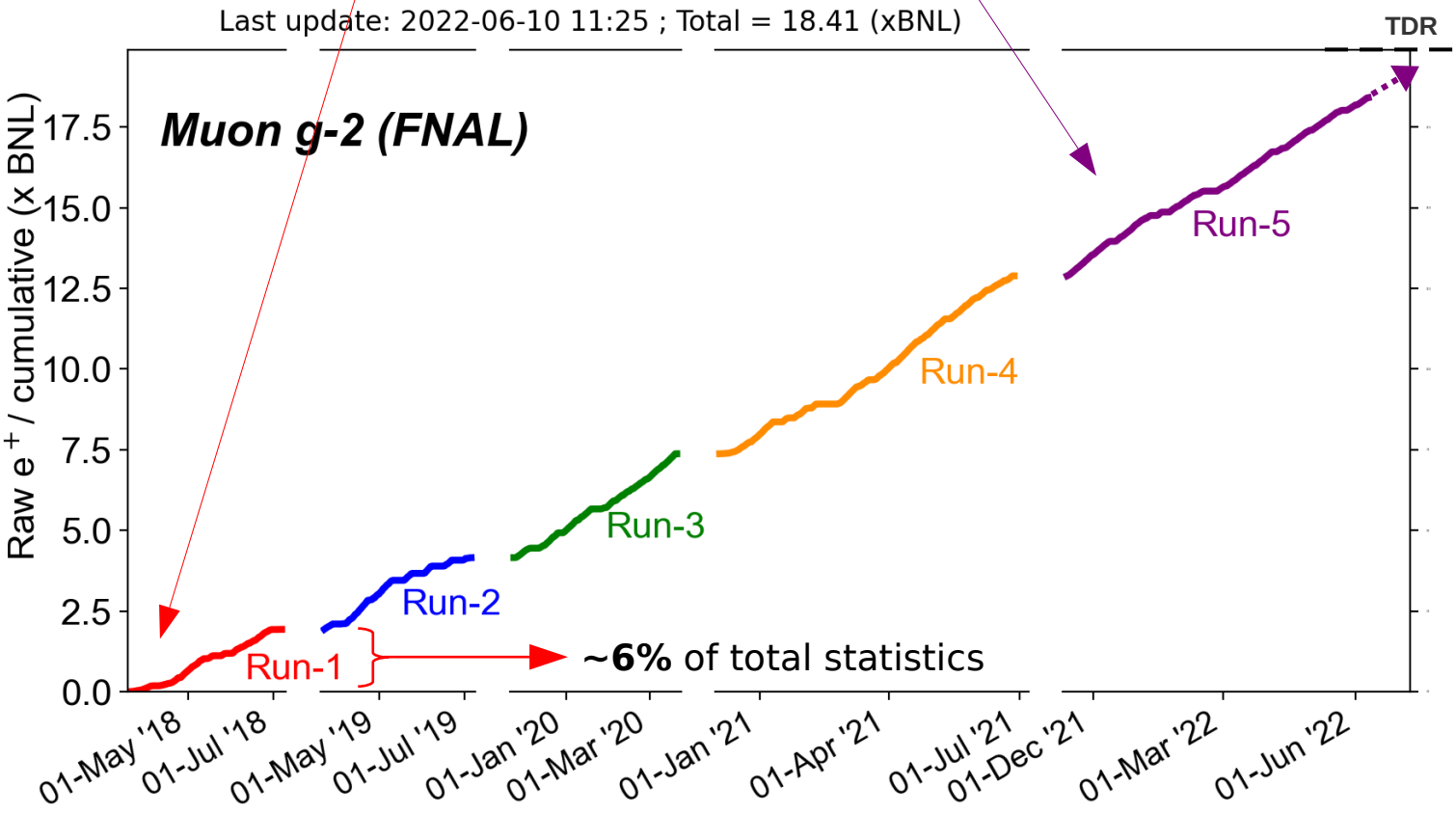




Timeline



- Many improvements from one Run to the next
- Managed to push through Covid isolations with fully remote shifts
- On track to reach design statistics
- A lot of data to analyze



Production time

- Because of the multiple reprocessings, the overall processing time of Runs varied by a lot
  - Run1 took ~1.6 years (258k files)
  - Run2 took ~2.3 years (294k files)
  - Run3 took ~1.6 years (559k files)
- If no reprocessing needed, production now takes ~1 month per 200k files → Run4 projection of 4/5 months