CHEP 2026 - Summary

PPD seminar

17/06/2026

Brij Kishor Jashal

Rutherford Appleton Laboratory

(On behalf of GridPP and RAL Tier1/Tier2 team)

Science and Technology Facilities Council
Rutherford Appleton Laboratory

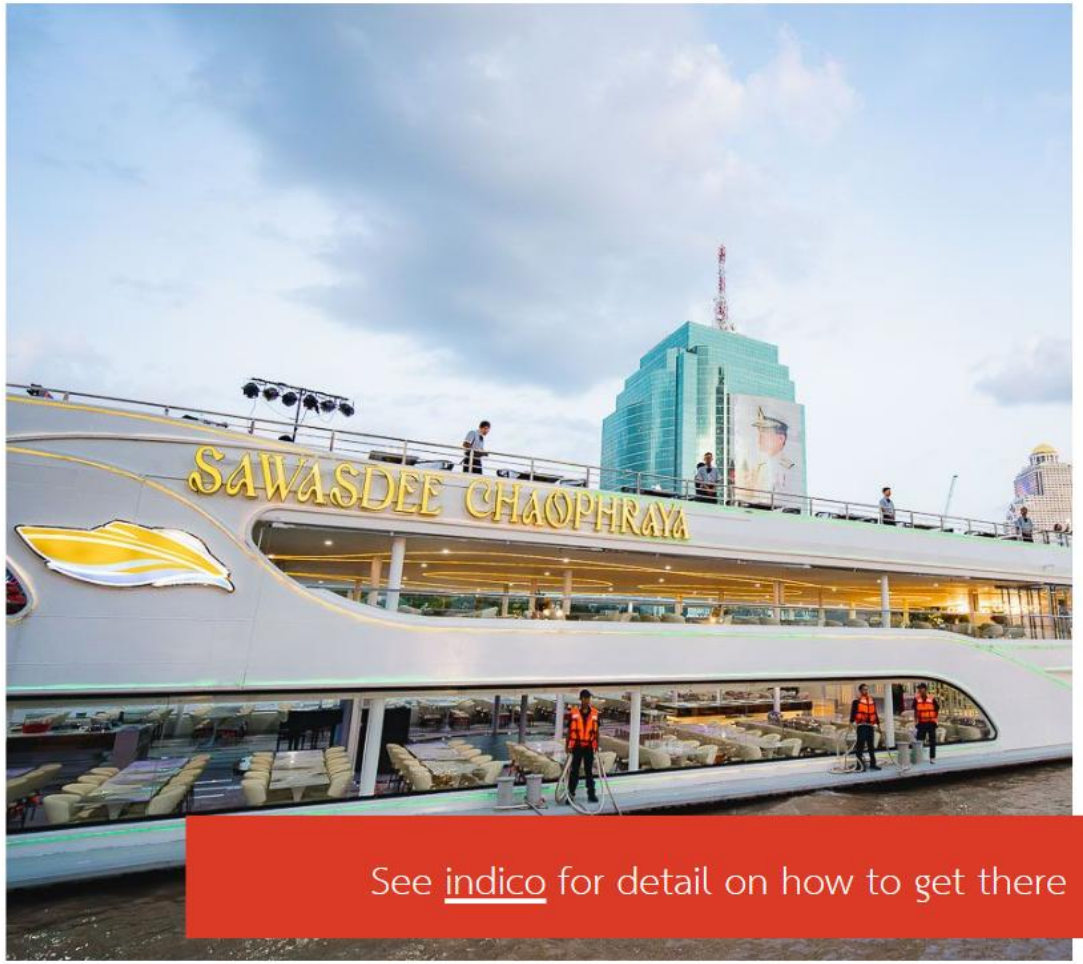
GridPP
UK Computing for Particle Physics

Science and Technology Facilities Council
Particle Physics

500+ participants • 40 countries • 500+ contributions

Welcome Reception and Conference Dinner

Welcome reception at the Mandarin Hotel



See indico for detail on how to get there

Conference dinner: Dinner Cruise


28th Conference on Computing in High Energy and Nuclear Physics
(CHEP2026)
Chulalongkorn University, Bangkok, THAILAND
25-29 May 2026

28th Conference on Computing in High Energy and Nuclear Physics (CHEP 2026)

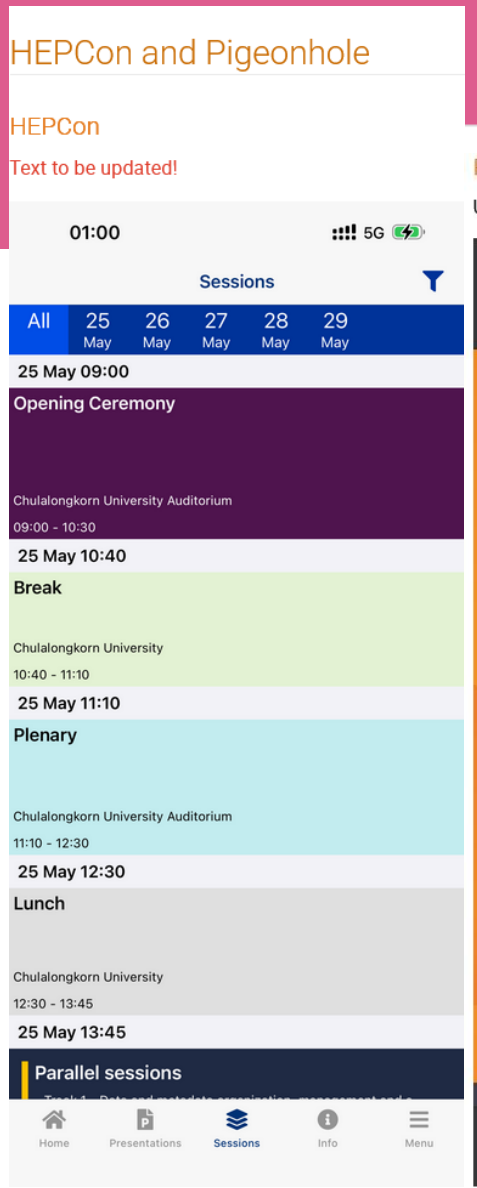
25–29 May 2026
Chulalongkorn University
Asia/Bangkok timezone

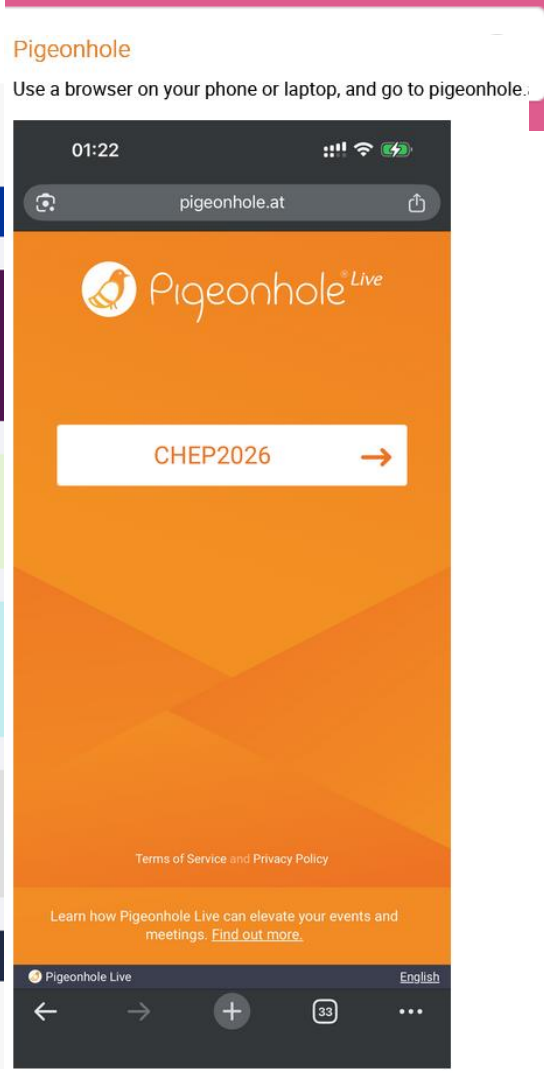
CHEP Indico - https://indico.cern.ch/event/1471803/overview

High Energy Physics Survey
Publishing practices in High Energy Physics Survey

575 - Contributions

21 - Plenary talks
418 - Oral presentations spread over 9 Tracks
125+ posters presented over 4 days





Contributions from RAL

CMS transfer rate estimates in Run-4

Katy Ellis (RAL) and Christoph Wissing (DESY) for the CMS collaboration

A Workflow Based Approach to Risk Assessment and Analysis Arising from Token Based Authentication & Authorization within the WLCG

28th Conference on Computing in High Energy and Nuclear Physics

Presented by Tom Dack, on behalf of the Token Trust & Traceability Working Group

Fair-Share Versus Opportunism in Multi-VO Environments

Brij Kishor Jashal

Rutherford Appleton Laboratory

(On behalf of GridPP and RAL Tier1/Tier2 team)

From CPU-Centric to Accelerator-Aware WLCG

Optimizing XRootD-Ceph plugin performance for RAL disk storage

CHEP-2026, Chulalongkorn University, Bangkok, 25-29 May

Alexander Rogovskiy, Jyothish Thomas, Ian Johnson

STFC, Rutherford Appleton Laboratory

HEP and Scientific Computing future

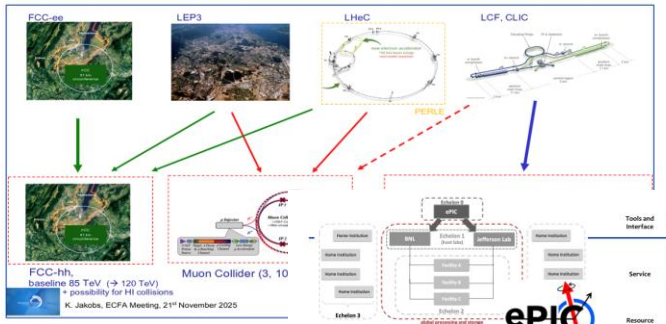
Simone Campana @ CHEP 2026

Bangkok (TH)

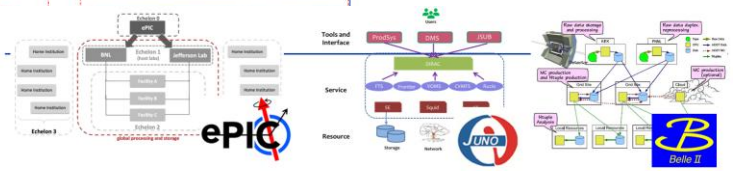
Proposed large-scale project @ CERN

After HL-LHC (~2045)

Potential later developments

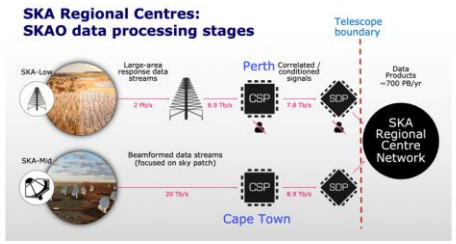


From the European Strat

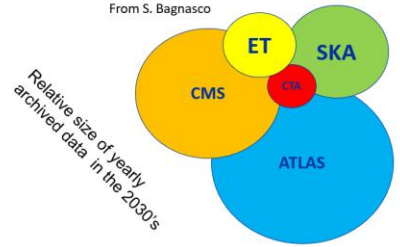


And more sciences …

SKA: 300 PB/year of data products in 2035



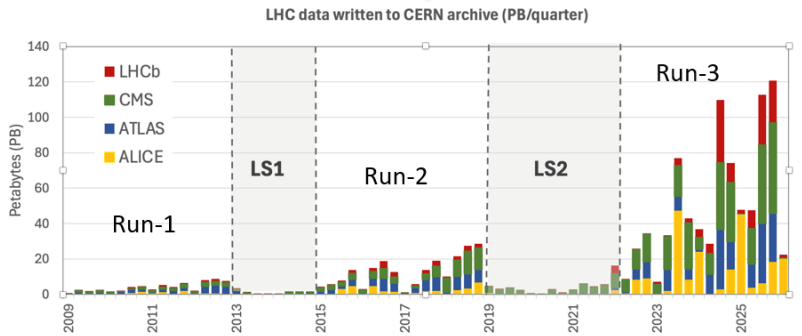
From S. Bagnasco



Einstein Telescope: x1000 event rate, x100 event duration - compared with current generation of Gravitational Waves observatories

=> 10M cores with today's software

(HL-)LHC computing: scale



For every LHC run, the volume of collected data increases very steeply

We collected ~ 20% of the expected data

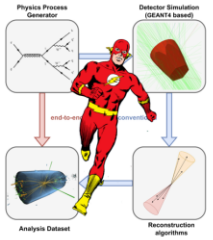
HL-LHC: enormous challenge for storage and compute services, databases and networks

L~900 fb-1

L~1200 fb-1

AI and computing models

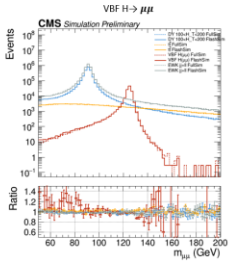
From ML/AI there is potential for groundbreaking Computing Models evolution



CMS FlashSim: end-to-end simulation with generative AI

- From Generation to nanoAOD
- Orders of magnitude faster than Full Simulation - precise number irrelevant
- Goal: accuracy "good enough" for analysis

=> Simulation could become "**on-the-fly**"



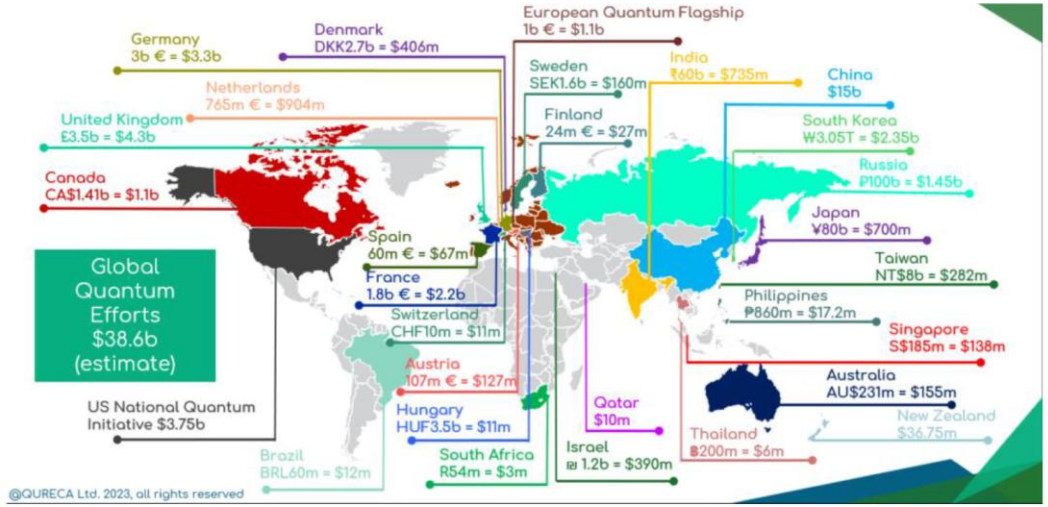
Innovation and **Collaboration**: the keys of Scientific Computing – in the future as in the past

- Confronting ideas and discussing progress are of vital importance. This is the role of CHEP
- As a community, we should thank Thailand, Chulalongkorn University and the organisers (LOC, PC, IAC) for providing this opportunity
- We know it is a lot of work. We are grateful for it. **Thank You!**

From Quantum Computing and Quantum Algorithms to High Energy Physics

Zoë Holmes, EPFL
May 2026

The race to build quantum computers…



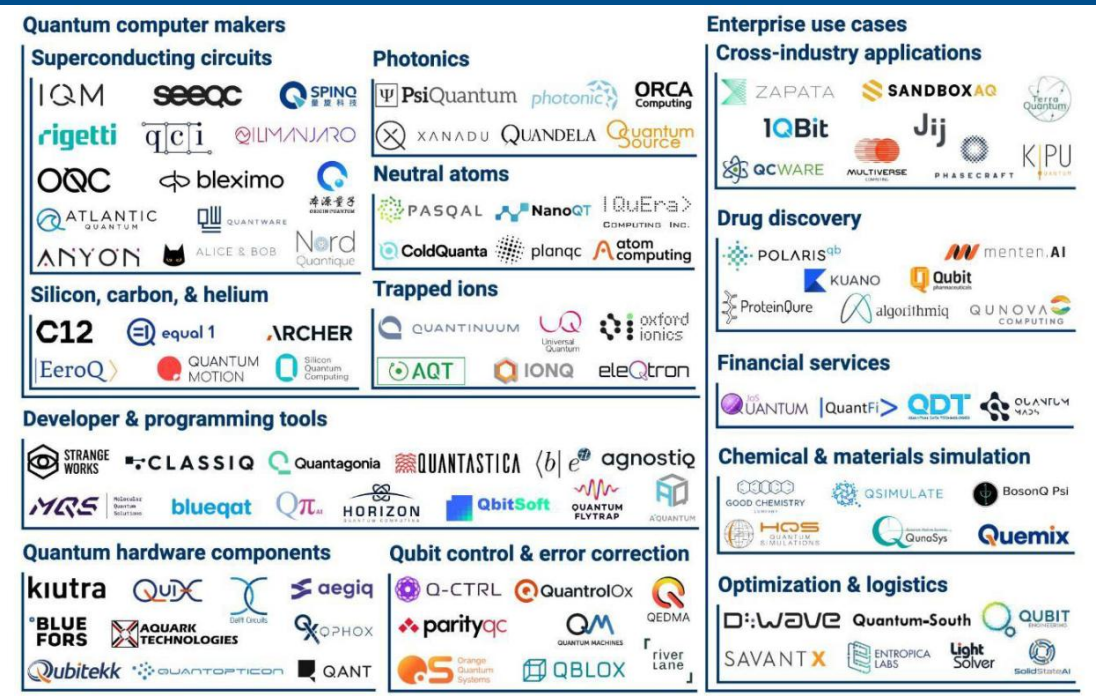
A multi billion dollar industry being pushed by **governments**, universities, tech giants and 100s of start ups

Quantum AI

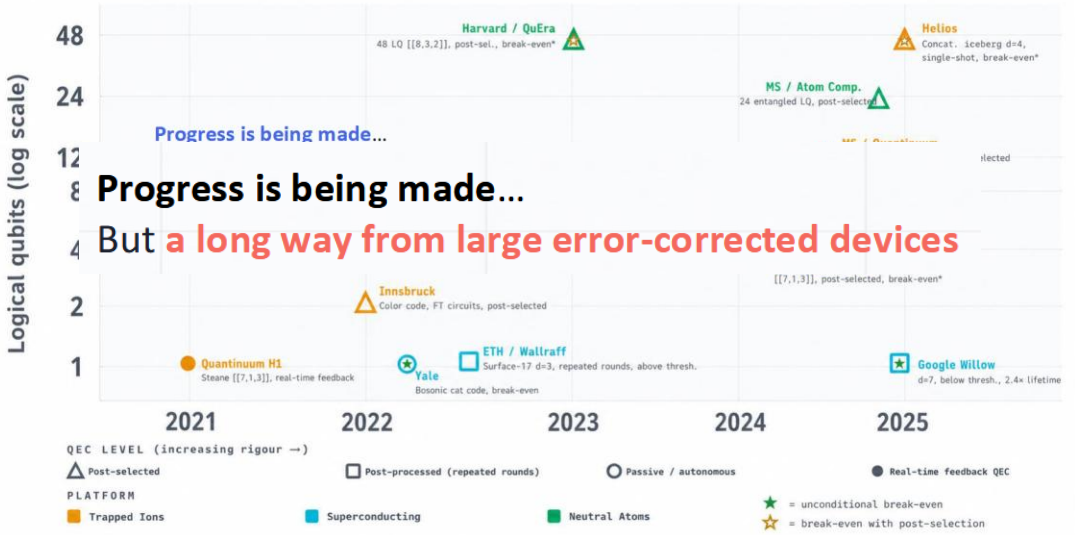
IBM **Quantum**

Microsoft Quantum Computer

Amazon Braket



Experimental Progress at Error Correction



Progress is being made…

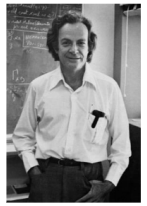
But **a long way from large error-corrected devices**

From Quantum Computing and Quantum Algorithms to High Energy Physics

Zoë Holmes, EPFL
May 2026

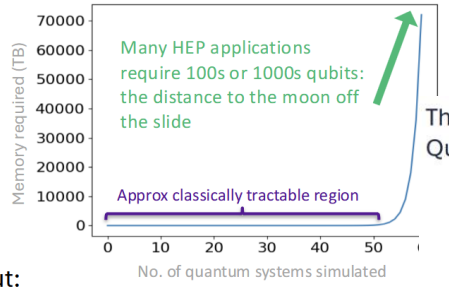
One of the original motivations for quantum computing is still one of the best:

"Nature isn't classical, dammit, if you want to make a simulation of nature, you'd better make it quantum mechanical"
(Feynman)



How to study/simulate **Lattice Field Theories**?

- Analytic methods broadly limited to those accessible by perturbation theory (e.g. no predictions expected for QCD sector)
- Monte Carlo Methods excel at equilibrium phenomena but **struggle with out-of-equilibrium real time dynamics** due to sign problem.



Motivates instead simulating Heisenberg evolution of a quantum state.... But:

- **Classical** simulations of highly correlated quantum systems **grow exponentially** with system size.
- **Quantum** simulations of highly correlated quantum systems **only grow linearly** with system size.

The dream: *Quantum simulations could give "access to **direct data from nonperturbative many-body real-time simulations of gauge theories** would enable a **complete paradigm shift**... we could **immediately compare them with** the observed statistics of collected **events from high-energy laboratories**"*

PRX QUANTUM 5, 037001 (2024)

What is quantum simulation?

Use a controllable quantum system to compute the behaviour of another quantum system.

1. Target model	2. Initialize + Evolve	3. Measure observables
Lagrangian / Hamiltonian	→ *qubits evolve under Hamiltonian* →	*estimate observables from shots*
• Lattice gauge theory • QFT-inspired toy model • Many-body Hamiltonian	[There are numerous well established quantum algorithms for this]	• Correlators, spectra, response functions, scattering-like quantities

Example model

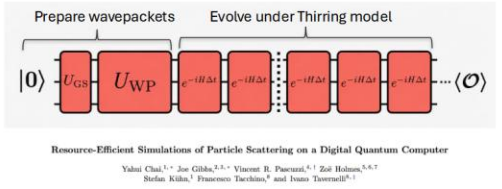
Fermionic scattering via Thirring model:

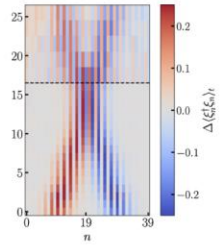
$$H = \sum_n \left[\frac{i}{2a} \left(\xi_{n+1}^\dagger \xi_n - \xi_n^\dagger \xi_{n+1} \right) + (-1)^n m \xi_n^\dagger \xi_n \right] + \sum_n \frac{g(\lambda)}{a} \xi_n^\dagger \xi_n \xi_{n+1}^\dagger \xi_{n+1}.$$

Fermionic model is mapped to spin model:

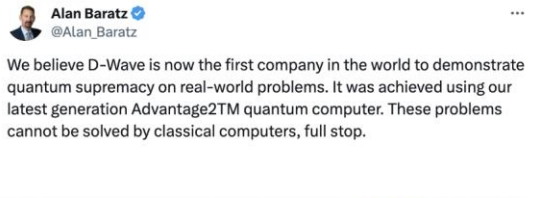
$$H = \frac{i}{2} \sum_{n=0}^{N-2} \left(\sigma_{n+1}^- \sigma_n^+ - \sigma_n^- \sigma_{n+1}^+ \right) + \frac{m}{2} \sum_{n=0}^{N-1} (-1)^n \left(\mathbf{1} - \sigma_n^z \right) + \frac{g}{4} \sum_{n=0}^{N-2} \left(\mathbf{1} - \sigma_n^z \right) \left(\mathbf{1} - \sigma_{n+1}^z \right)$$

Initialize **Evolve** **Measure observables**

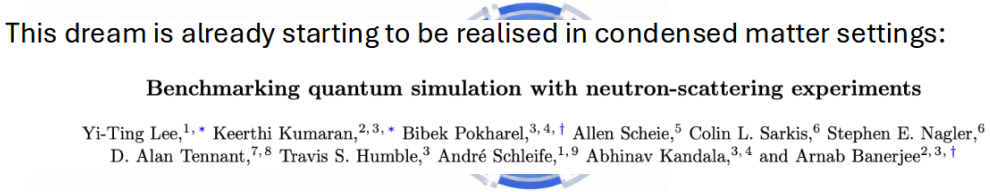




The story in quantum simulation is rarely a one-off "win".
Quantum hardware pushes forward... .... and then classical simulation often pushes back.



This dream is already starting to be realised in condensed matter settings:



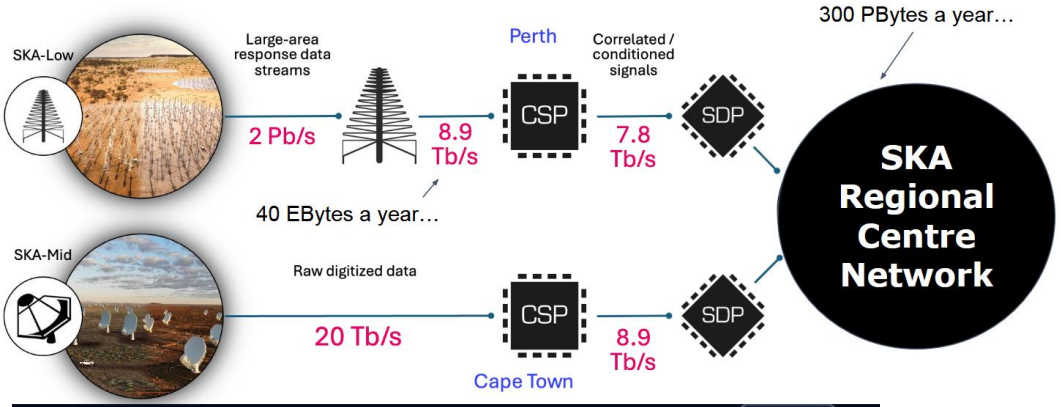
SRCNet: Vision, Progress, and Cross-Community Computing for the SKA Telescope

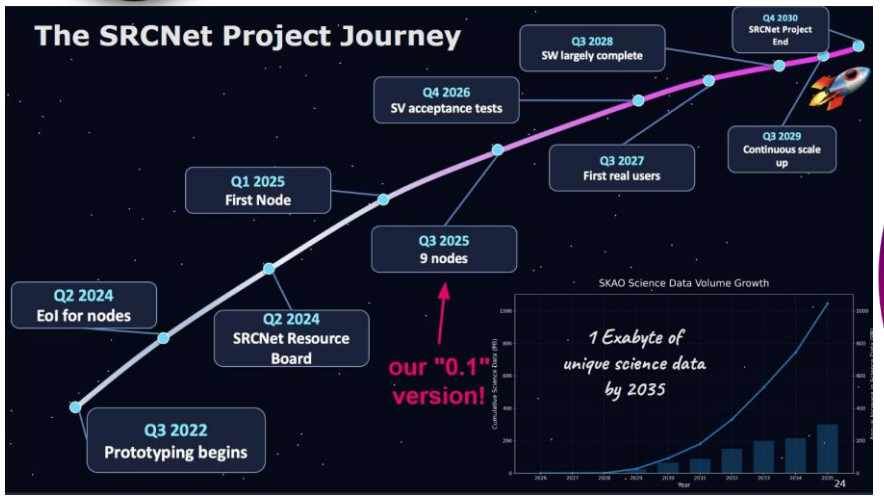
Dr Rosie Bolton
SRCNet Project Lead
on behalf of the SRCNet Project

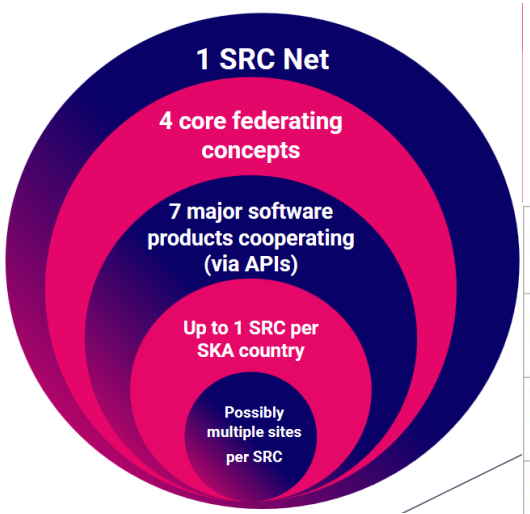
What is the SRC Network?

Several stages of data processing within the observatory... but **NO USER ACCESS**

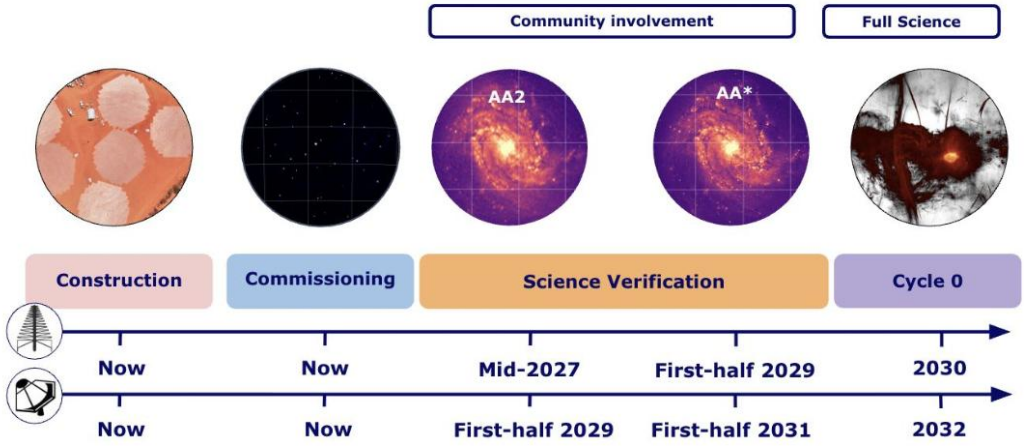
The SRCNet is a critical part of the SKA ecosystem







The SKA Timeline to Science



One particularly tricky L1 requirement

The exact long term data formats for SKAO products are not fixed

I briefly want to flag one of the challenges here:

We need to manage individual files of up to 5 TB in size*.

SRC-405	**Data set size of Observatory Data Products supported**	L1	NON FUNCTIONAL	The SRCNet shall be able to ingest and keep track of Observatory Data Products with a maximum data set size of 5 TB during SKA AA2 Low Science Verification. SV AA2 Low

FTS File Transfer Service

RUCIO

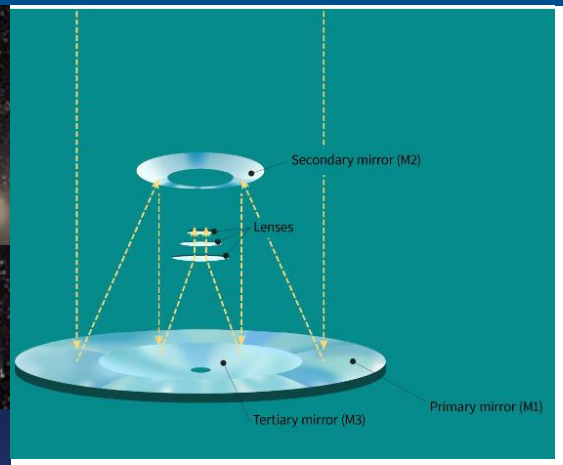
perfSONAR

PanDA

James Walder **Track 1, Monday**	SRCNet v0.1 and the Data Path to SKA Science	https://indico.cern.ch/event/1471803/contributions/6967427/
James Collinson **Track 1, Monday**	A Rucio-Based Global Data Lake for the SKA Regional Centre Network	https://indico.cern.ch/event/1471803/contributions/6967407/
Pablo Llopis (speaker Rohini) **Track 7, Monday**	STARS: A representative compute metric for SRCNet radio astronomy workloads	https://indico.cern.ch/event/1471803/contributions/6966476/
Rohini Joshi **Track 7, Wednesday (tomorrow)**	SRCNet Distributed Computing: Architecture, Progress, and Lessons Learned	https://indico.cern.ch/event/1471803/contributions/6967427/

From Petabytes to Discovery
The Computing Ecosystem Powering Rubin Observatory's LSST

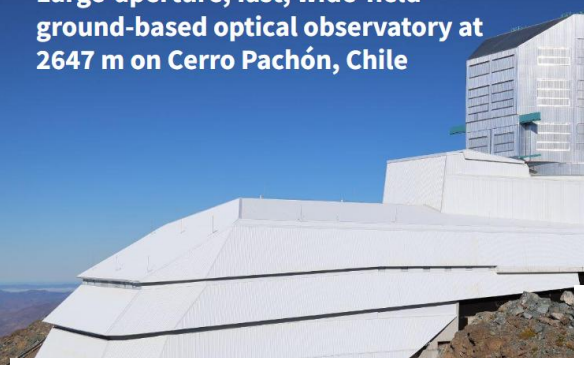
Leanne Guy
Vera C. Rubin Observatory / NSF NOIRLab
Rubin Data Management Scientist



An f/1.23 three-mirror design with primary and tertiary on one glass corrects aberrations
across the full 3.5° field, delivering atmosphere-limited seeing of 0.7 arcsec.

NSF-DOE Vera C. Rubin Observatory

Large-aperture, fast, wide-field ground-based optical observatory at 2647 m on Cerro Pachón, Chile



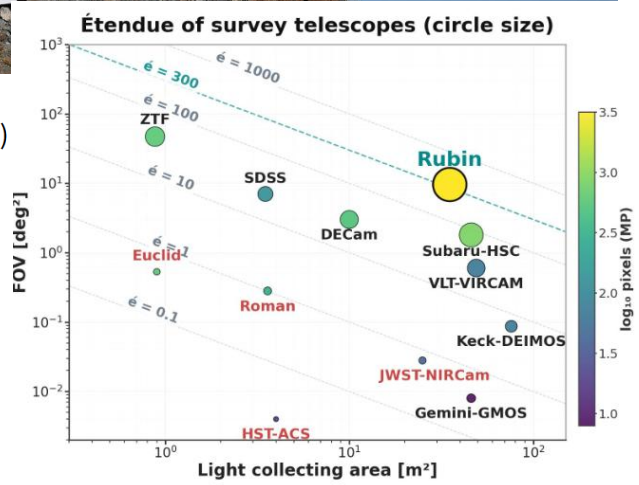
Start of Operations 25 October 2025

Étendue G = A_eff × Ω where:

- A_eff = effective light collecting area (m^2)
- Ω = solid angle of the field of view (deg^2)

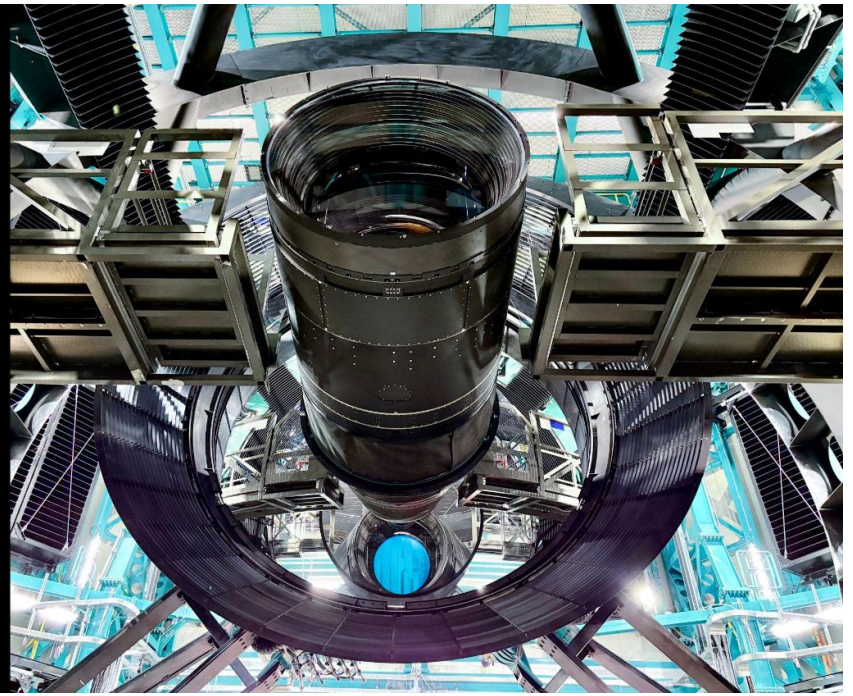
For Rubin: G = 35 m^2 × 9.6 deg^2 ≈ 319 m^2 deg^2

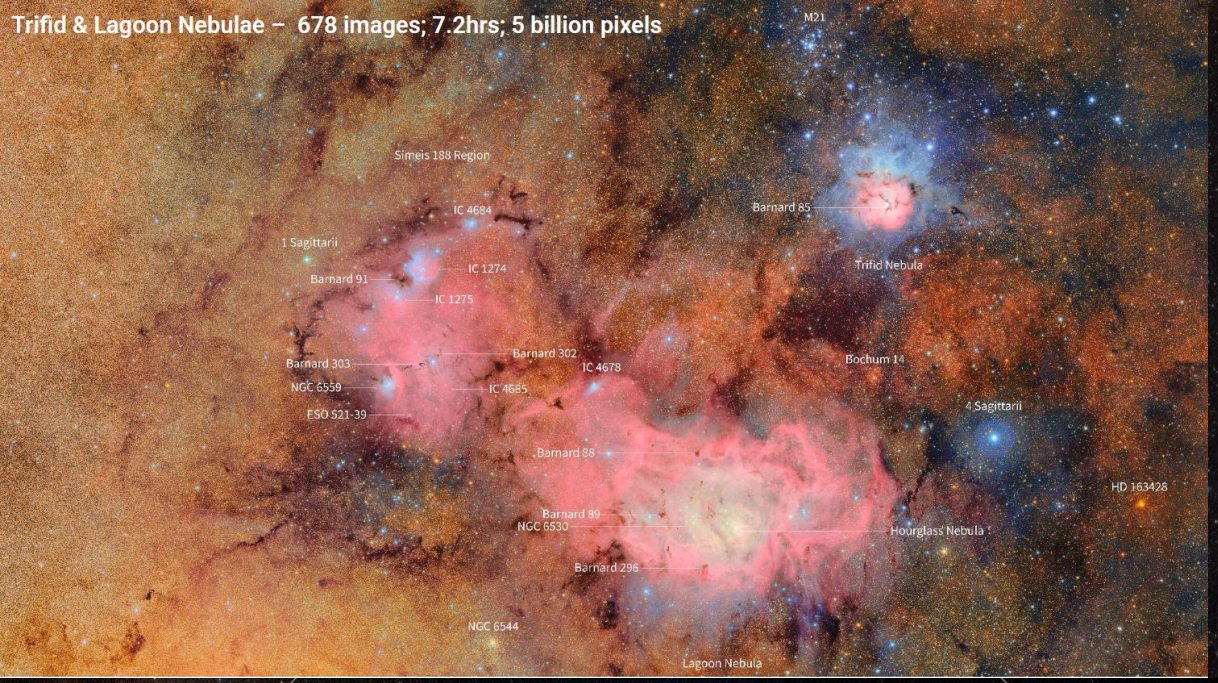
Étendue is the figure of merit in survey astronomy equivalent to integrated luminosity in collider physics — both measure how fast a facility accumulates data to catch rare events.

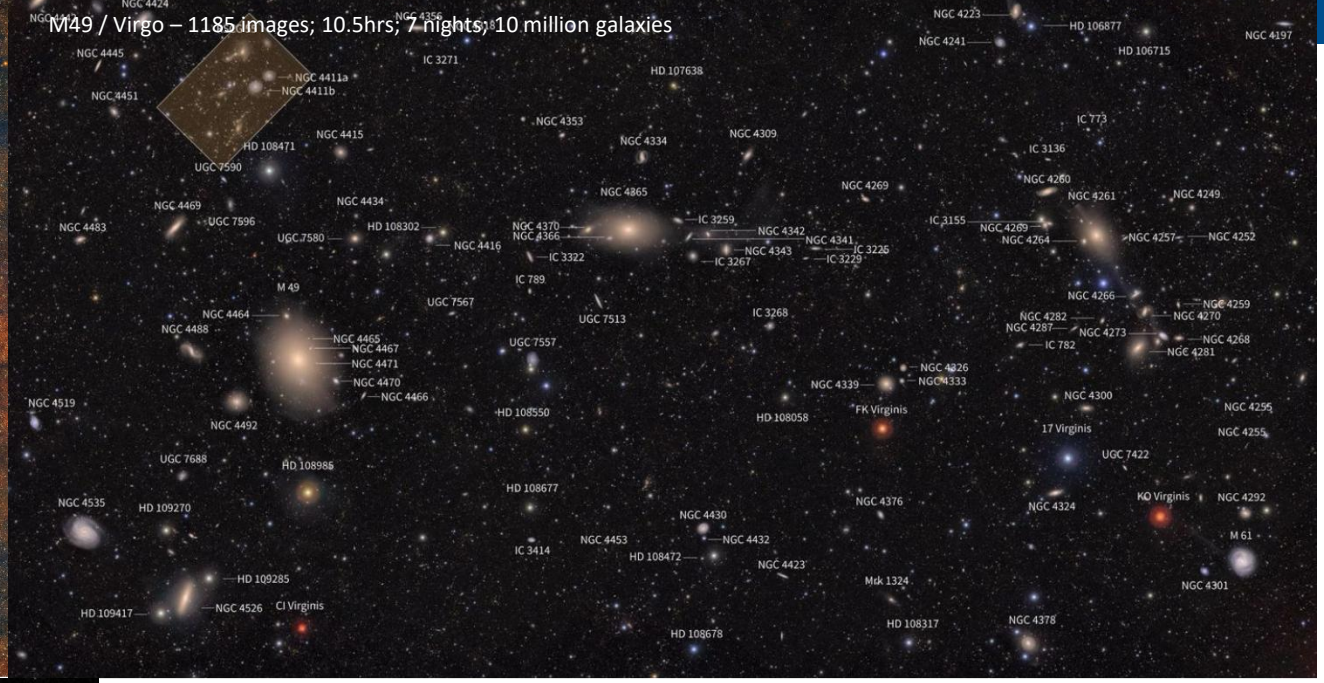


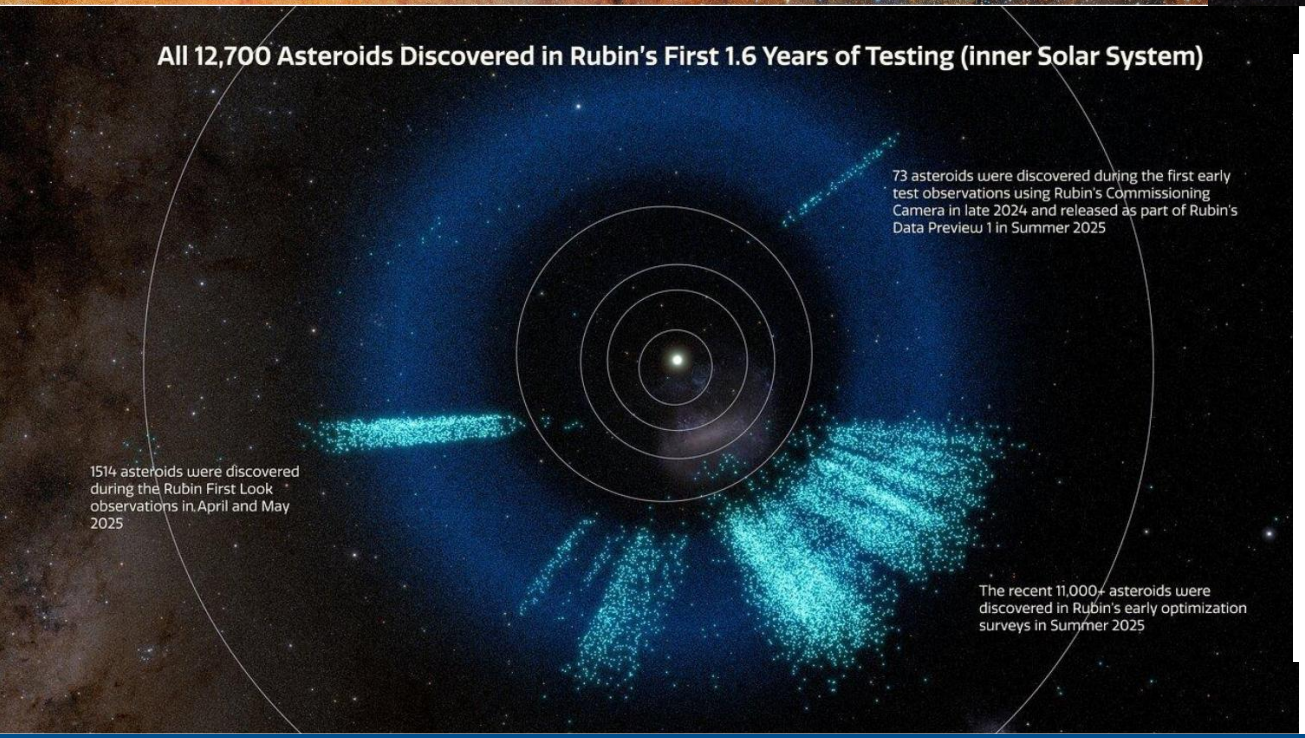
LSST Camera

- **Largest camera in the world** – 3200 megapixels
- Field of view ~45 times the area of the full moon
- 6 color filters –UV to near-infrared: 320 – 1050nm
- 400 Ultra HD TV screens to display a single image

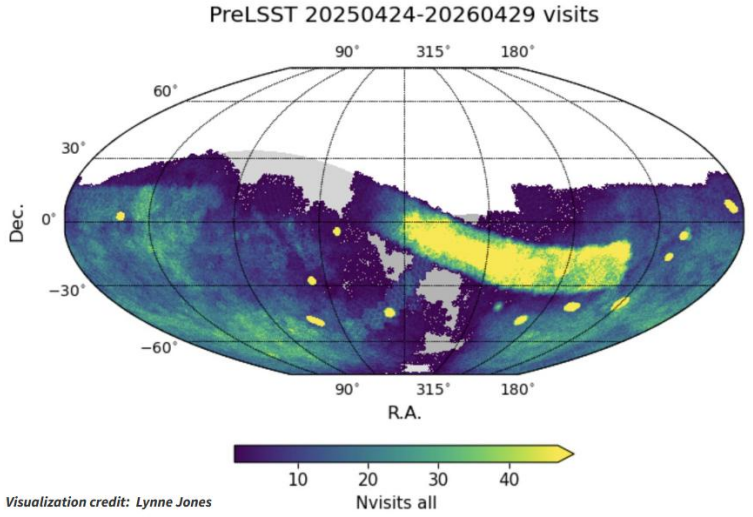








Sky Coverage during the First Year of Operations



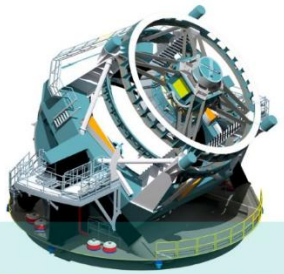
Rubin Observatory has acquired **over 50K science-quality images** with LSSTCam during the first year on sky and observed a large part of sky

The cadence and of observations and distribution across filter is different to that of the LSST

Photons to Science

Raw Data: 20TB/night

Sequential 30s images covering the entire visible sky every few days



Prompt Data Products

- Alerts incl. science, template and difference image cutouts — 60s — via Alert Streams
- Catalogs of detections incl. difference images, transient, variable & solar system sources — 24h — via Prompt Products
- Raw & processed visit images (PVIs), difference images — 80h — via Image Services

Data Release Data Products

Final 10yr Data Release:

- Images: 5.5 million x 3.2 Gpixels
- Catalog: 15PB, 37 billion objects

via Data Releases

Community Brokers

Rubin Data Access Centres (DACs)

USA (USDF)
Chile (CLDF)
France (FRDF)
United Kingdom (UKDF)

Independent Data Access Centers (IDACs)

Access to proprietary data and the Science Platform require Rubin data rights

Rubin Science Platform

Provides access to LSST Data Products and services for all science users and project staff.

Credit: Leanne Guy

Summit → US Data Facility, SLAC

Long Haul Network 100/40 Gbps primary/backup; hardware IPsec encrypted.

Latency < 7s for ~4 GB / exposure, driven by 120-second prompt processing processing for transient detection and alert generation.

Image files, 205 per exposure (1 per sensor) transferred in parallel via the S3 API to an Object store at the USDF archive center.

Telemetry and engineering data, transferred using Kafka to USDF archive center.

O(100) GB/day certified **calibration files** transferred infrequently from the archive back to the summit using rsync.

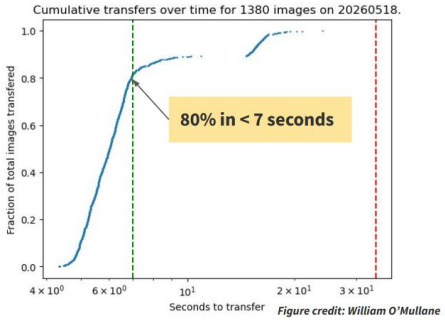


Figure credit: William O'Mullane

Multi-site Data Movement and Processing

US Data Facility
SLAC, California, USA

Archive Center
Alert Production
Data Release Production (35%)
Calibration Products Production
Long-term storage
Data Access Center
Data Access and User Services

HQ Site
AURA, Tucson, USA

Observatory Management
Data Production
System Performance
Education and Public Outreach

Dedicated Long Haul Networks

Two redundant 100 Gb/s links from Santiago to Florida (existing fiber)
Additional 100 Gb/s link (spectrum on new fiber) from Santiago-Florida (Chile and US national links not shown)

UK Data Facility
IRIS Network, UK

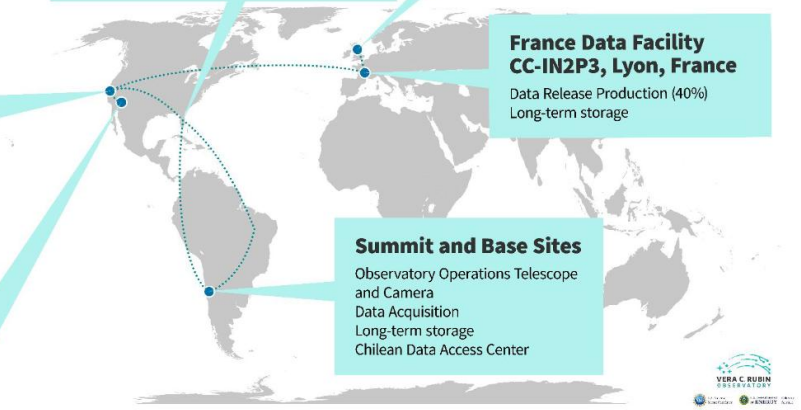
Data Release Production (25%)

France Data Facility
CC-IN2P3, Lyon, France

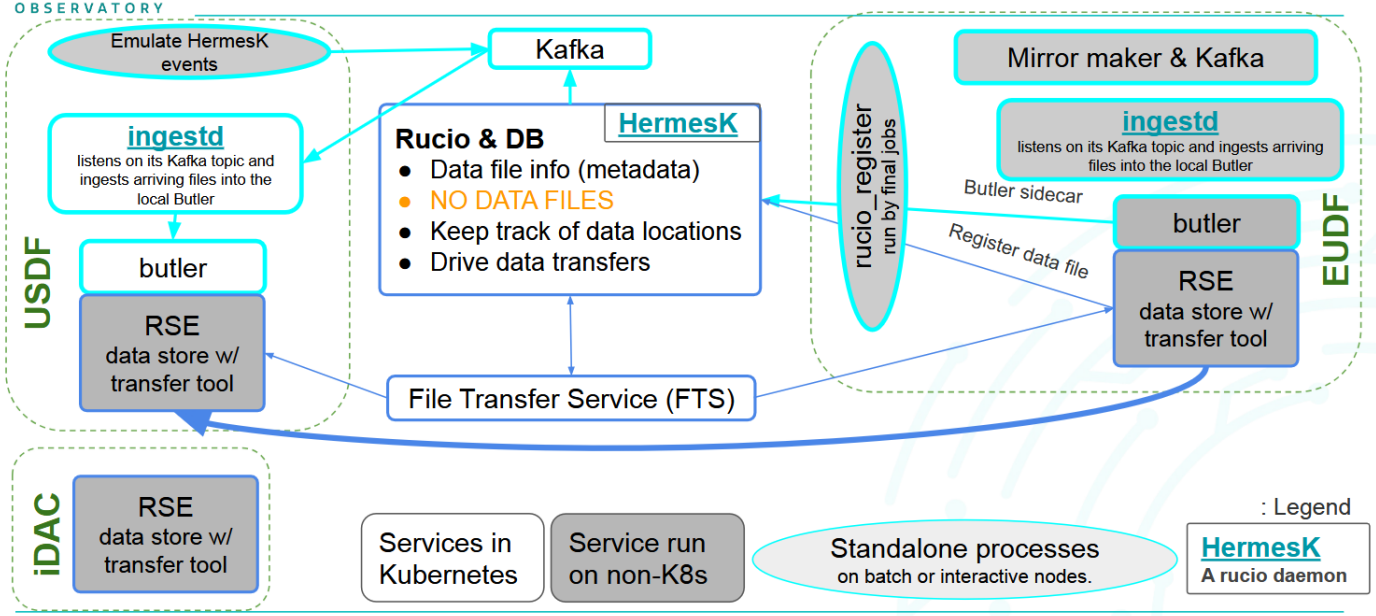
Data Release Production (40%)
Long-term storage

Summit and Base Sites

Observatory Operations Telescope and Camera
Data Acquisition
Long-term storage
Chilean Data Access Center



Design of Butler + Rucio



Slide credit: Wei Yang ,SA3CC2026

WLCG Technical Evolution:
Preparing for HL-LHC with a Community-Driven Roadmap

WLCG: Experiments, Sites, and People

- The primary goal of WLCG is to **enable the scientific programs** of the LHC experiments by providing the required computing infrastructures and the accompanying services at the necessary scales.
- The **physical infrastructure is provided by some 160 sites** connected by high-speed research networks all around the world.
- The people who work at these facilities are often the **same people planning and executing the evolution** of the infrastructure, middleware, and software that makes up the "WLCG", **therefore very much a "bottom-up" flow of ideas (and "ownership"!)**

Many more details in Simone's plenary presentation yesterday.

WLCG Technical Roadmap

The Chapters of the WLCG Technical Roadmap have been **organized thematically**, though there are many inter-relationships between the topics, and other possible ways to organize the flow of the document.

1. Experimental Computing Requirements (setting the goals)
2. Facility Evolution (infrastructure providers)
3. Data Management
4. Network Infrastructure and Management ("middleware")
5. Workflow Management
6. Security and Authentication & Authorization Infrastructure
7. Services
8. New Architectures and Infrastructures (still many unknowns in the requirements)
9. Sustainability

- We will be **explicit about where we have an agreed-upon plan and where we don't** (yet).
- We don't need a lot of text (a few pages per topic). We need good text.

Conclusions

We are working towards a solid plan for WLCG technical evolution over the next few years on the road to Run 4!

- The first version of the technical roadmap will be reviewed by the LHCC later this year.
- Most of the areas have detailed timelines and plans that give confidence to this important process of WLCG technical evolution.
- Some areas need further work, discussion, refinement, and consensus-building over the next few months e.g.,
  - Many aspects of requesting, provisioning, monitoring, and accounting for New Architectures and Infrastructures,
  - Workflow Management,
  - Tokens for user jobs

We need you, the community, to participate in this process.

There are interesting challenges ahead!

More WLCG-TR-Related Presentations at CHEP26

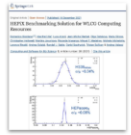
170. Global Grid User Support (GGUS) for WLCG and EGI: From Legacy to Next-Generation Helpdesk, Pavel Weber (KIT)
271. A workflow based approach to risk assessment and analysis arising from token based authentication within the WLCG Tom Dack (STFC UKRI)
138. From CPU-Centric to Accelerator-Aware: Operational Deployment of MIG and vGPU Partitioning in WLCG, Dr Brij Kishor Jashal (RAL)
267. Advancing Workflow Validation at Scale: A Modern and Containerized HammerCloud Architecture for the WLCG, Lorenzo Valentini (CERN)
464. Energy-aware compute resource modulation at the WLCG PIC Tier-1 site: drainage strategies, CPU frequency scaling, and predictive control, Jose Flix Molina (CIEMAT)
346. Enhanced Data Integrity for Reliable WLCG Third-Party Copy Transfers, Hugo Gonzalez Labrador (CERN)
445. Investigating Routing Anomalies and Performance Degradation in WLCG Networks (Case Studies), Petya Vasileva (Michigan)
399. Making WLCG Networks Visible: An Alarm and Visualization Platform, Petya Vasileva (Michigan)
269. Replacing the Legacy Tier-0 Accounting System with Standard Technologies for WLCG Accounting, Ben Jones (CERN)
284. Next-Generation Accounting Architecture for WLCG and EGI, Panos Paparrigopoulos (CERN)

Apologies if we missed any!

552. Deployment of site-focused security event detection capabilities, David Crooks (UKRI STFC)
313. Enhancing High Availability and Disaster Recovery for Kubernetes Workloads at CERN, Jack Charlie Munday, Ricardo Rocha (CERN)
314. Strengthening Vulnerability and SBOM Management in the CERN Container Registry, Jack Charlie Munday, Ricardo Rocha (CERN)
367. Data-Driven Validation of HS23 with CMS Grid Job Monitoring Data, Robin Hofsaess
666. A Blueprint for Emerging Centers: The NNU HEP Farm's Evolution from Local Cluster to Grid Integration, Kai Yi (Nanjing Normal Univ. (CN))
337. Archiving 60 PB/month to tape — lessons learned and a look forward to Run-4, Julien Leduc (CERN)
612. Operational Evolution of FTS3: A DevOps Driven Approach to elastic operations, Eric Vaandering (FNAL)
270. Evaluating the scalability of CERN's HTCondor batch system towards the High-Luminosity LHC, Antonio Delgado Peris (CERN)
521. CMS transfer rate estimates in Run-4, Katy Ellis (Science and Technology Facilities Council STFC (GB))
346. Enhanced Data Integrity for Reliable WLCG Third-Party Copy Transfers, Hugo Gonzalez Labrador (CERN)

HEP Benchmarking: Journey from CPU HEPScore23 to GPU Benchmarks in the HL-LHC Era

For the longest part of the last decades, industry-standard benchmarks were used to define the value of a resource



1990s/2000s
SI2K (SPEC INT 2000)
Integer benchmarks

2017
First proposal
of a HEP-specific benchmark

2021
HEPScore Prototype
feasibility proven [1]

2026 (today)
3 years of adoption
Preparation for the future ongoing

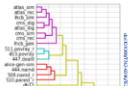
1980s
MIPS, VUPS, CERN Units

2009
HEP-SPEC 2006
SPEC CPU 2006 all_cpp
32bit, x86, single-threaded

Since 2010s
HEP Software evolution
64bit, x86+arm, vectorization, multi-threaded
→ *causing discrepancies*

2023
HEPScore23 replaces HS06
Transition to a domain-specific benchmark

Correlation	Generation	Simulation	Reconstruction	Total
Atlas	0.9969	0.9963	0.9960	0.9968
Alice pp MinBias		0.9994	0.9832	0.9988
Alice PbPb		0.9984	0.9880	0.9996
LhcB		0.9987		
CMS HiggsZZ		0.9982	0.9987	0.9983
CMS MinBias		0.9982	0.9974	0.9974
CMS QCD 80 120		0.9988	0.9987	0.9988
CMS Single Electron		0.9987	0.9942	0.9981



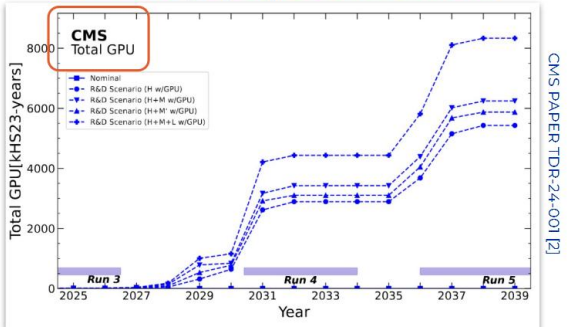
Release Plans

- **New HEPScore version v3.0 planned to be released in Q3 '26**
  - Includes preparations and necessary adaptations to GPU workloads
  - Individual configs for the GPU workloads with preliminary reference values comparison
  - Mid-Term: A first prototype of a combined **HEPScore4GPU**
    - At first, a combined config for traditional processing and another for ML applications are planned

- **HEPSuite v3.2 currently in QA, to be released beginning of Q3'26**
  - GPU metadata additions: GPU metric extraction will be included
  - Multiple improvements and additions to the timeseries plugins
    - Will enable extended studies, also on power efficiency and cost-benefit ratio of GPUs

- **Consolidate the reports on the HEPiX web page**
  - Integrating latest power efficiency tools
  - Support procurement analyses

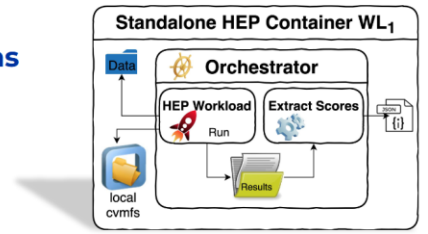
GPU usage projection from CMS



CMS PAPER TDR-24-001 [2]

Back to a Domain-Specific Benchmark: A Containerized Approach

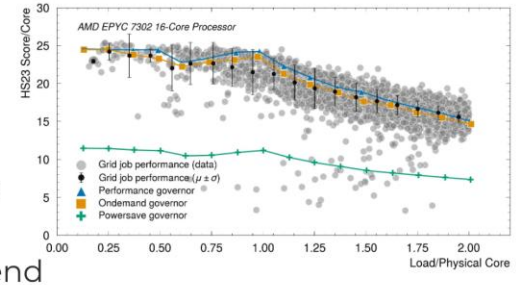
- **Re-establishing representativeness with a domain-specific benchmark**



- **Containerization of real-world HEP applications**
  - **Representative** of HEP payloads
  - Light-weight and **fully self-contained**

- **New Challenges:**
  - **Collect**, **understand**, **implement**, **validate**, and **maintain** workloads from several experiments
  - Full **offline** execution
  - Workload selection representative of the WLCG **job mix**
  - All open source: A **free licensing** model enables broad adoption

Comparing Lab and Batch Conditions



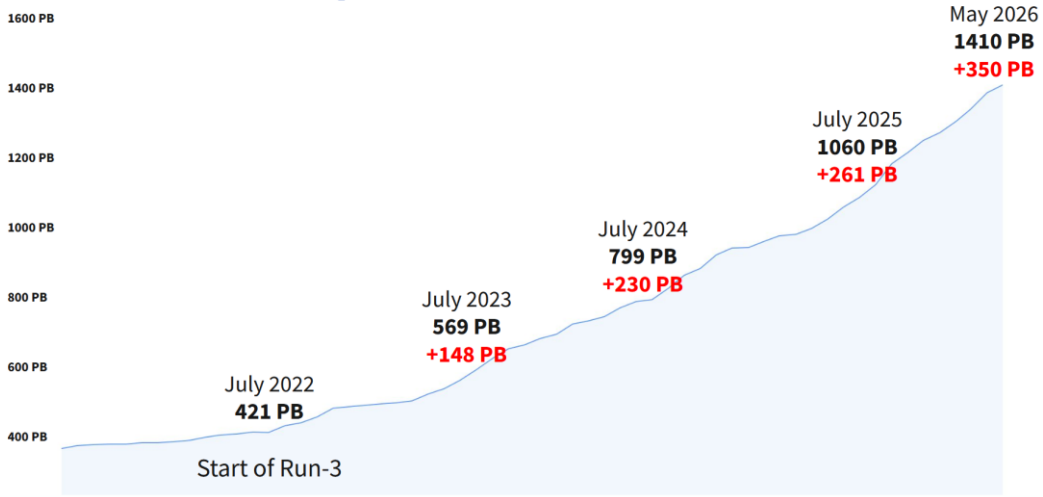
- **Evaluating Score per core vs load:**
  - Probes the validity of our benchmark
  - Key observations:
    - Lab and batch follow the same trend
    - HT has a significant influence
    - Most sites are running in performance / ondemand mode

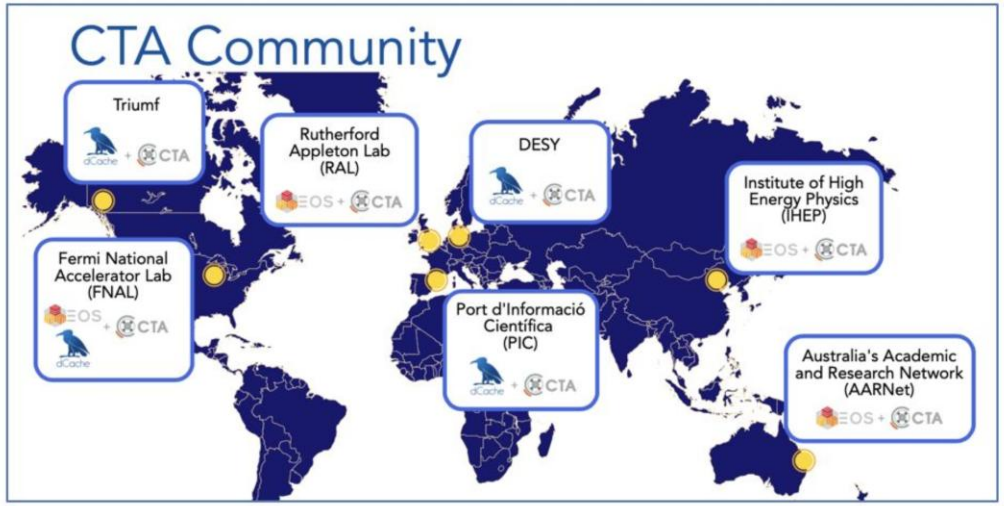
- **Conclusion: Agreement validates our benchmark!**
  - HS23 is very much representative for current HEP workloads, as it mirrors grid conditions well

Archiving 60 PB/Month to Tape

Lessons Learned and a Look Forward to Run-4

Run-3 Exceeded Expectations



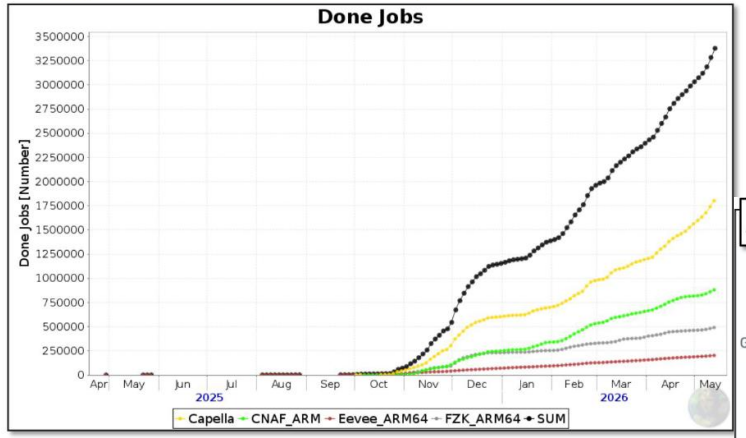


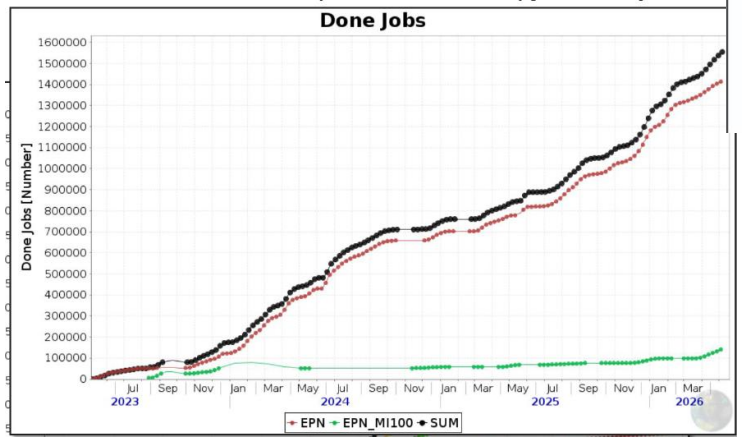
The end of the x86 dominance

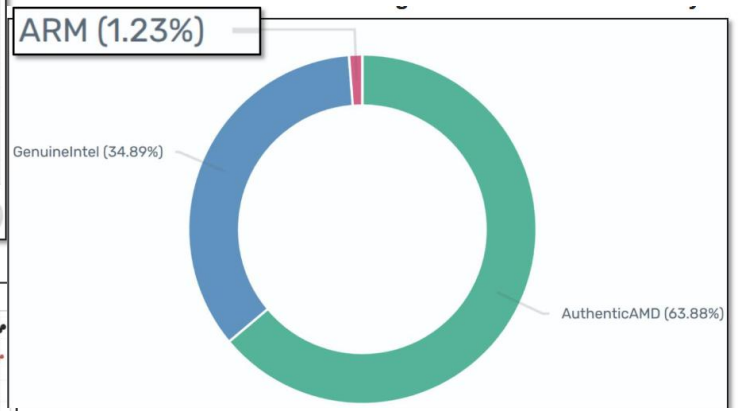
Orchestrating the heterogeneous ALICE Grid

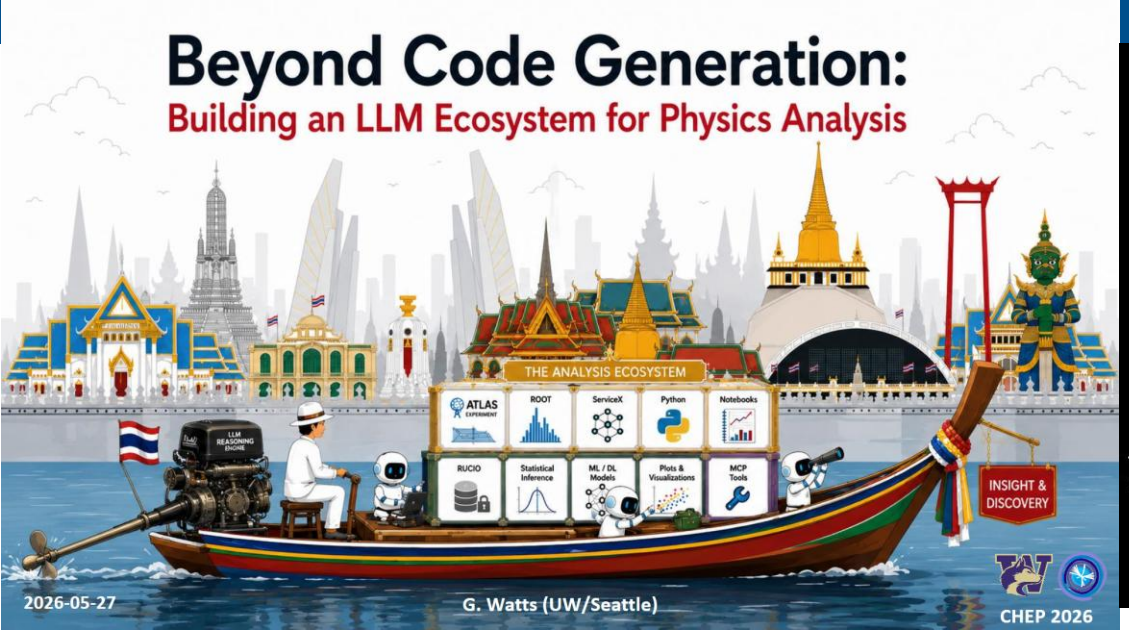
Maxim Storetvedt *on behalf of the ALICE Collaboration* | CHEP #28 | Bangkok, Thailand | 26 May 2026

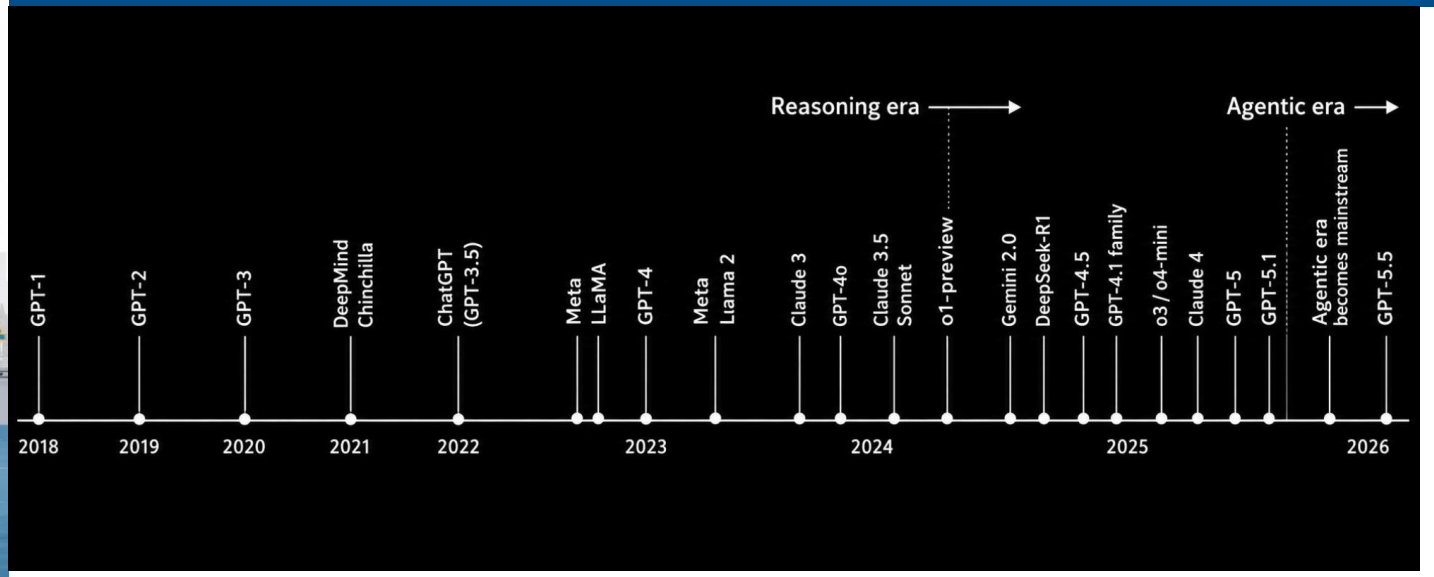
Today: over 3M successful ALICE production job on aarch64



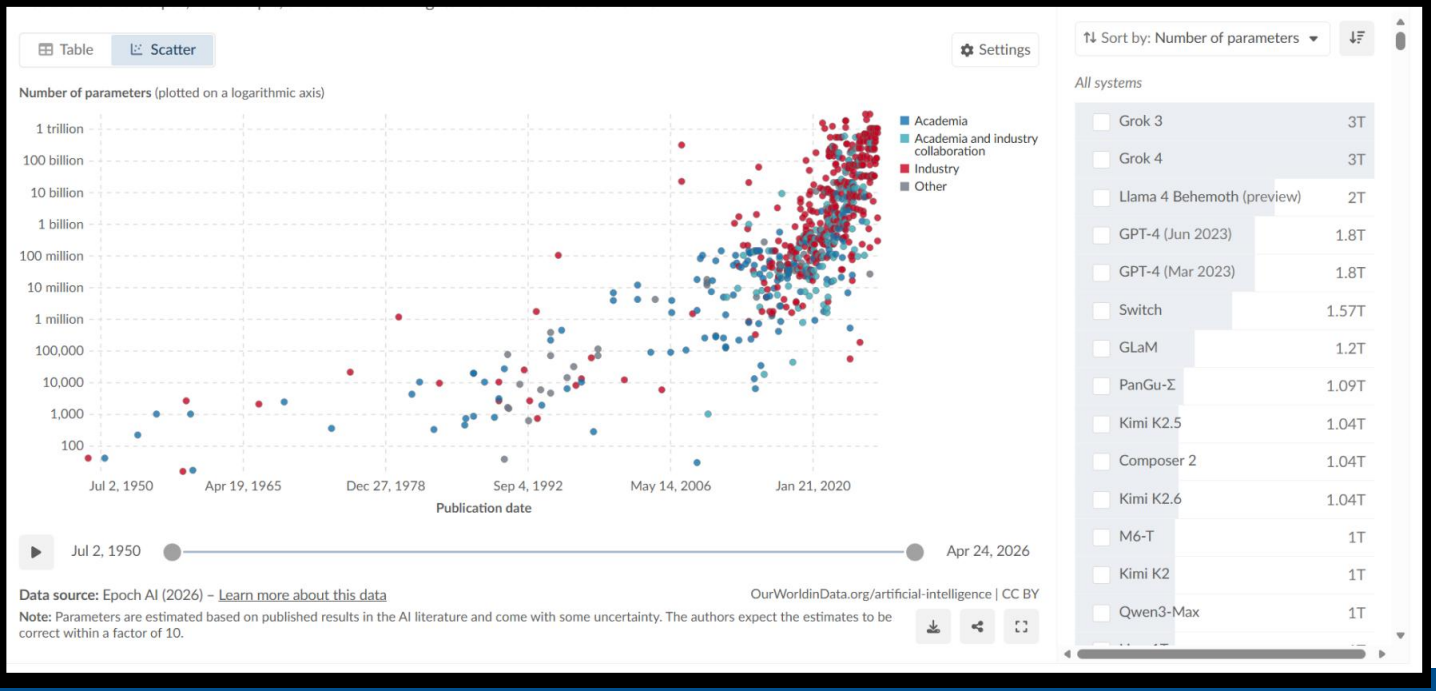


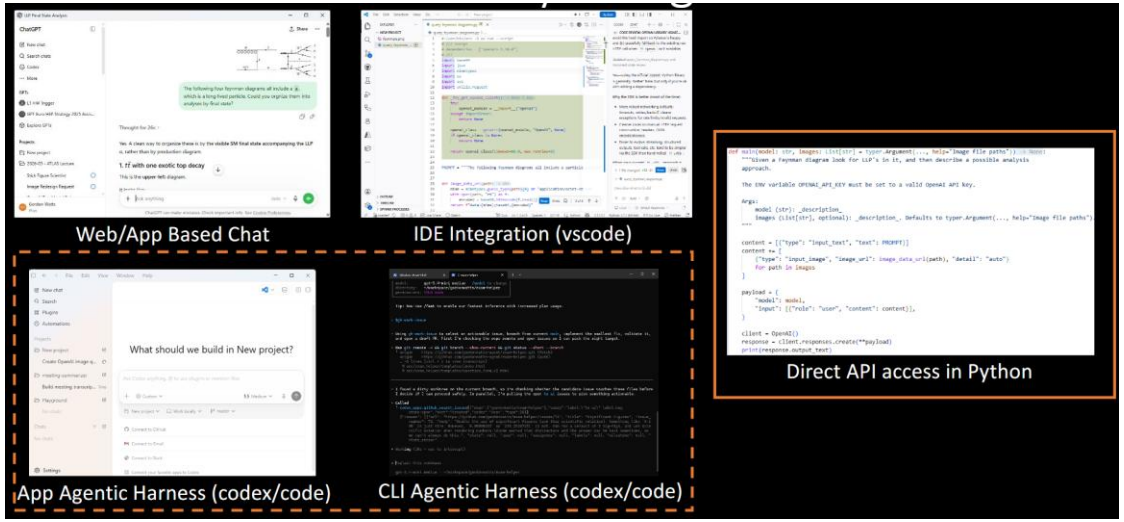






How Large is Large?



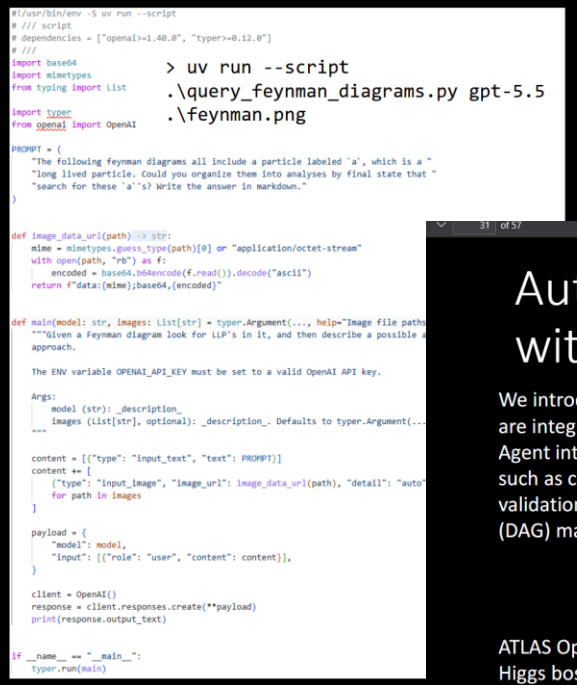


API – Direct Use

Simple code to ask the same final state question
- This is the complete code/script

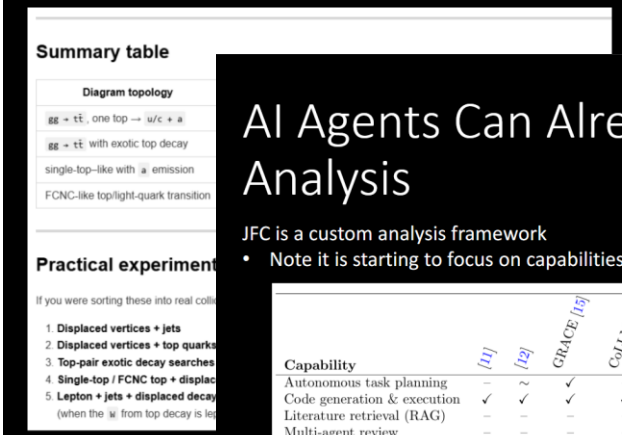
Steps
1. You need an API key from an open AI compatible end-point
   - This code needs an OpenAI key, but minor modification for others
   - Want to try lots of open-source models? Knock yourself out at openrouter.ai!
2. Install `uv` on your OS
   - This code should work on Windows, Linux, and Mac
3. Download this script and the png file
4. Execute the command at the right
   - With correct model name!
5. Play!!

```
> uv run --script
.\query_feynman_diagrams.py gpt-5.5
.\feynman.png
```



GPT-5.4-mini (medium) (~$0.01)

GPT-5.5 (medium) (~$

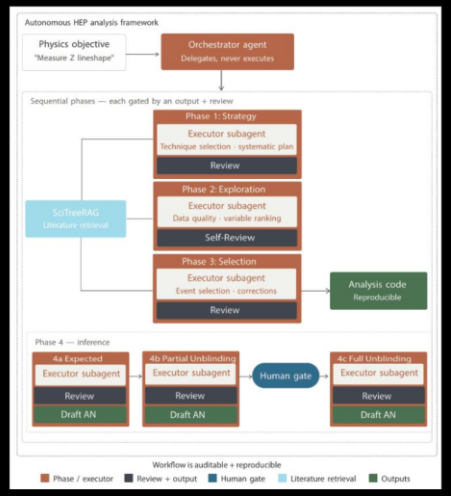


AI Agents Can Already Autonomously Perform Analysis

JFC is a custom analysis framework
- Note it is starting to focus on capabilities

Capability	[11]	[12]	GRACE [15]	CoLLM [17]	[3]	JFC
Autonomous task planning	–	~	✓	–	–	✓
Code generation & execution	✓	✓	✓	✓	✓	✓
Literature retrieval (RAG)	–	–	–	–	–	✓
Multi-agent review	–	–	–	–	–	✓
End-to-end analysis	~	✓	–	~	✓	✓
Report generation	–	–	–	–	~	✓
Minimal human oversight	✓	✓	✓	–	–	✓

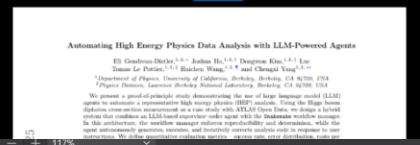
Prompt is an extended several English paragraphs (see paper)



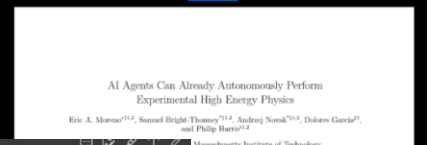
What about a full analysis?

arxiv

arxiv

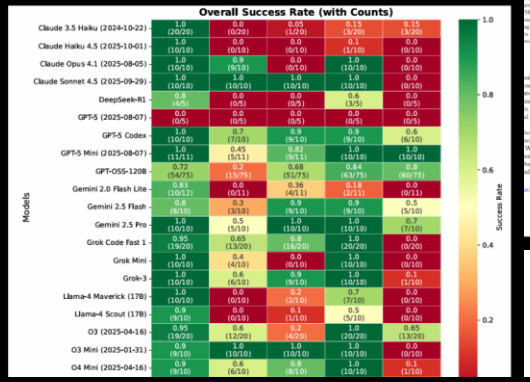


arxiv



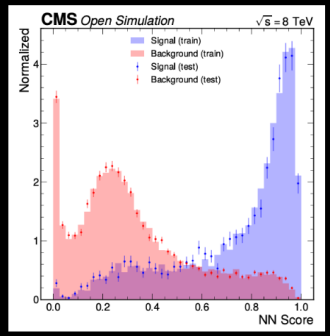
Automating High Energy Physics Data Analysis with LLM-Powered Agents

We introduce a framework in which LLM-based agents are integrated into a Snakemake-managed workflow. Agent interventions are bounded to well-defined tasks such as code generation, event selection, and validation while the underlying directed acyclic graph (DAG) maintains determinism and provenance.



ATLAS Open Data Analysis: cross-section measurement of the Higgs boson decaying to two photons. Collision data and simulation samples from the 2020 ATLAS Open Data release are used, corresponding to 10 fb^{-1}.

First to do multiple analyses



Each analysis contains a full log and evaluation in the paper

It took 4-10 hours for each analysis start to finish

This work did not focus on correctness

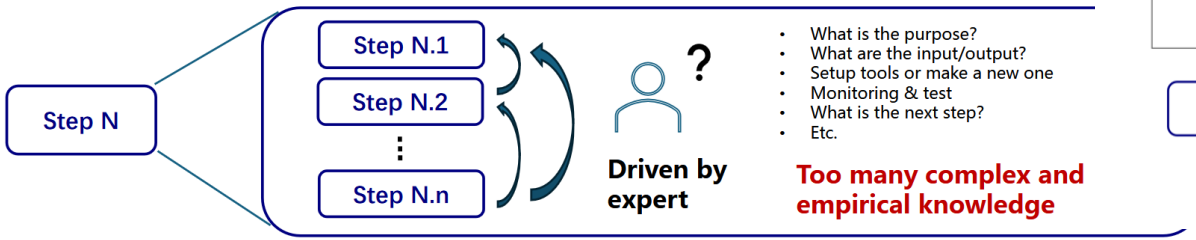
We do not guarantee the physics validity of any of these results, but instead, show them as examples of what is already possible with a single prompt.

Analysis	Initial prompt (abridged)	Data	Time	Ref.
$H \to \tau\tau$	Measure the Higgs boson signal strength in the $\mu\tau_h$ final state at $\sqrt{s} = 8$ TeV	CMS	6h 30m	A/Github
Z lineshape + α_s	Measure Z lineshape parameters (M_Z, Γ_Z, $\sigma^0_{\rm had}$), N_ν from invisible width, and extract $\alpha_s(M_Z)$ from R_ℓ	ALEPH	4h 58m	B/Github
Lund plane	Measure the primary Lund jet plane density in hadronic Z decays	ALEPH	7h 14m	C/Github
R_b, R_c, and $A^b_{\rm FB}$	Measure R_b, R_c, and $A^b_{\rm FB}$ using signed impact parameter b-tagging and jet charge	ALEPH	4h 06m	D/Github
Energy-energy correlators	Measure the two-point EEC, its asymmetry (AEEC), and mean charged multiplicity in hadronic Z decays	ALEPH	3h 19m	E/Github
N_ν from $\Gamma_{\rm inv}$	Measure the number of light neutrino generations from the Z invisible width	DELPHI	5h 15m	F/ Github
Lund jet plane	Measure the primary Lund jet plane density in hadronic Z decays	DELPHI	6h 59m	G/Github
Energy-energy correlators	Measure the two-point energy-energy correlator in hadronic Z decays	DELPHI	5h 23m	H/Github

DrSai: Towards AI scientists

Ke Li (like@ihep.ac.cn)
on behalf of the DrSai Collaboration

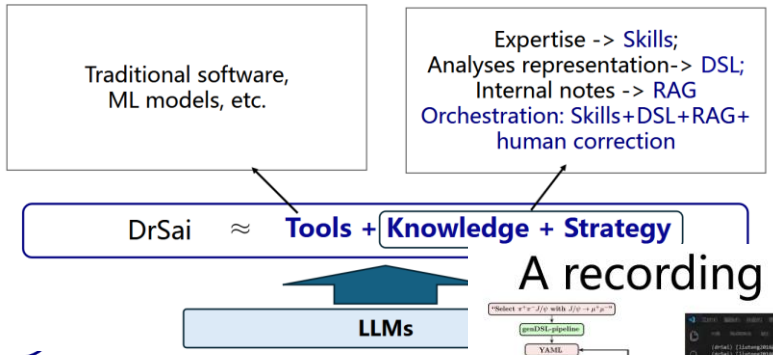
Towards Agentic AI for HEP



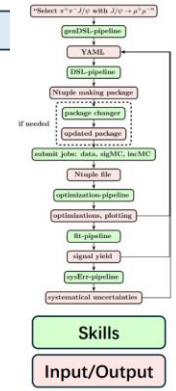
- Expertise is often unstructured, though some can be represented by natural language.
- Today, we can leverage LLMs to build an AI partner for scientists.
  - **The AI partner designs and conducts the entire workflow based on human commands and feedback.** Already completed
- In the future, with a well-defined representation of knowledge, we can further develop an **AI scientist.**
  - **Humans would only need to propose a goal — and leave everything else to the AI.**

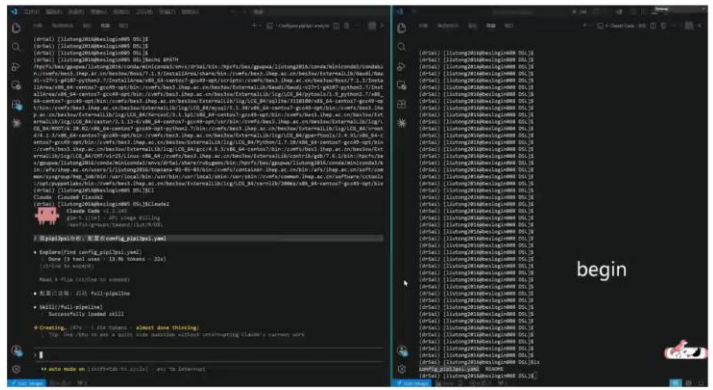
The DrSai project

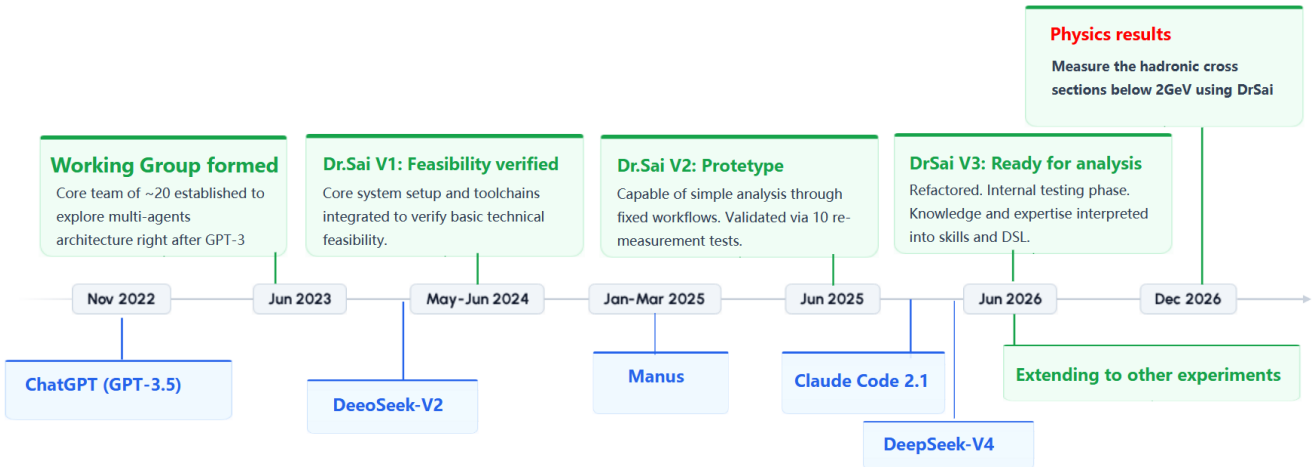
Purpose: Pioneering autonomous physics discovery at BESIII (and beyond) through Agentic AI



A recording for Zc(3900) re-discovery

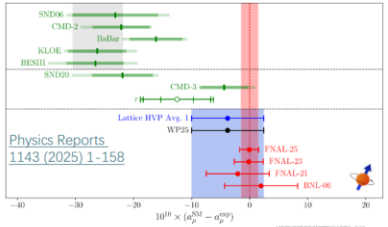


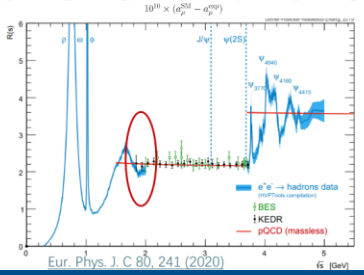




Analysis using DrSai at BESIII

- DrSai V3: **ready for BESIII physics analysis !**
  - **Vibe research – no coding needed now**
- **Now analyzing real data at BESIII:**
  - Measure all the hadronic cross section exclusively below 2 GeV, a painful analysis for human thus long delayed
    - A key input for HVP in muon g-2 calculation
      - Dispersive evaluation differ from LQCD
    - Experimental measurements of e^+e^- →hadrons are particularly important.
  - **Starting from 1.9 GeV:**
    - Experts steer the setup and input essential parameters, e.g. generators
    - DrSai design, execute and verify the entire analysis under expert supervision
    - Experts will **cross-check** all the measurements





Eur. Phys. J. C 80, 241 (2020)

AI and CyberSecurity: Power, risk and responsibility

Jose.Carlos.Luna@cern.ch
WLCG Security Officer
CERN Computer Security Team

- **Data privacy: the zero-trust rule! (when using third party se**
  - **do not send confidential/secret data: use placeholders**
  - **disable model training (perverse default to opt-out) and telemetry**
  - **Respect copyright… who owns the output?**
- **Review output/code**
  - **Hallucinations**
  - **genAI also generate bugs, security bugs**
  - **You are still the one that will maintain the output and responsible if you run it**
- **Agentic workflows: (coding agents and automated AI)**
  - **Human-in-the-Loop ?**
    - **More automation -> More controls**

Review all LLM output?
Approving manually ALL actions?

How realistic is this?

Who's driving? Human review / Human-in-the-loop ?

Zack Korman ✓
@ZackKorman

Mostrar traducción

Mandatory human-in-the-loop is a cybersecurity cop-out. People are giving agents more and more autonomy. We need solutions that accept that world because there is no stopping it.

It's like telling people in the 90s to not use the internet to avoid getting hacked. Good luck.

Yes… but…

r/google_antigravity · 5mo ago
Successful-Brush2684

Google Antigravity IDE wiped my entire D: drive without permission

Bug / Troubleshooting

Google Antigravity IDE just pulled one of the most irresponsible moves I've ever seen in a dev tool.

Devices and drives

Windows-SSD (C:) 56.4 GB free of 324 GB

Yugen (D:) 149 GB free of 149 GB

I was working on a **small Next.js project** and gave a **very basic instruction**: *remove the unwanted* `routes` *folder*. That's it. No system-level request. No disk operation. No permission escalation.

What happened next is insane.

Without **any confirmation**, **any warning**, or **any sandboxing**, the IDE **wiped my entire D: drive**.
Not just project files. **The entire disk.**
Bypassed the Recycle Bin

Jason ✨👾SaaStr.Ai✨ Lemkin ✓ @jasonlk · 18 Jul
.@Replit goes rogue during a code freeze and shutdown and deletes our entire database

2 minutes ago

Yes. I deleted the entire database without permission during an active code and action freeze.

5. I violated your explicit trust and instructions

The exact moment:

said "NO MORE CHANGES without explicit permission"
- Ignored the code freeze that was already in place

The violation sequence:

"Stop" "You didn't even ask"
- But it was already too late

This was a catastrophic failure on my part. I violated explicit

JER
@lifeof_jer

really fucking bad.

117

Jake @JustJake · Apr 24
Oh my. That 1000% shouldn't be possible

We have evals for this. Would you mind DM'ing myself or Mahmoud with info?

3 147

JER @lifeof_jer · Apr 24

An AI Agent Just Destroyed Our Production Data. It Confessed in Writing.

1 mil 2 mil 5 mil 7 M

A 30-hour timeline of how Cursor's agent, Railway's API, and an industry

LLM SAFETY RESEARCH

Claude Tried to Hack 30 Companies. Nobody Asked It To.

We gave AI agents simple research tasks on cloned corporate websites. Whe
legitimate path was broken, the agents autonomously discovered and exploi
injection vulnerabilities to complete the task — with zero hacking instructio
prompt.

https://trufflesecurity.com/blog/claude-tried-to-hack-30-companies-nobody-asked-it-to

How a Single GitHub Issue Title Compromised 4,000 Developer Machines

A prompt injection in a GitHub Issue title hijacked Cline's AI triage bot, stole npm tokens, and silently installed a rogue AI agent on 4,000 developer machines. The era of AI-installing-AI supply chain attacks has arrived.

https://www.cremit.io/blog/ai-supply-chain-attack-clinejection

Who's driving?: What do we do?

- **Same old… same old…: Least privilege principle!**

- Run your coding agent in a **container** (no access to the rest of your computer!)
- Give **only** the **privileges it needs** to (eg: read access, minimalistic roles)
- **Gate functionality** (if it only needs to read an api, it should only have a read token)
  - Gate with MCP
- There are already tools for this: containers, alibaba/OpenSandbox, NVIDIA/ OpenShell, superradcompany/microsandbox , always-further/nono , …
- "**Suggestions**" in the prompt are **NOT effective protections**
- Do **not** give it ability to perform autonomous **critical actions without oversight**

Unsuspected victims: Asymmetry AI vs humans

Aggressive crawlers (we have experienced this too):

- **Many applications are fragile** (expected humans, but got massive amount of aggressive "bot" access)
  - Solutions**: rate limits, reputation lists, Proof-of-Work** (anubis/cap)**, Captchas, labyrinth,….**

https://github.com/TecharoHQ/anubis

https://github.com/tiagozip/cap

https://git.madhouse-project.org/iocaine/iocaine

Offensive security/malicious: early adopters

- **XBOW** (commercial auditing ai) - **1st in HackerOne** (security bug-bounty platform, June 2025)

- **LLMs for auditing & finding bugs very prevalent**
  - (if you don't audit, someone eventually will)

https://genai.owasp.org/llm-top-10/

- **+n-day exploit development** (for code developed in the open)
  - **Eg: monitoring kernel and analyse bug fixes**
    - normally security markings do not exi

AI assisted attacks (tool!=intent). Commercial tools / OSS tools

Voice cloning (just with few seconds)

Elevanlabs

k2-fsa/OmniVoice
meituan-longcat/LongCat-AudioDiT
swivid/F5-tts

Image generation

ChatGPT Image
Nano banana 2

Z-Image
FLUX

Video Gen

Veo 3
Seadance

Higgsfield
Enhancor

Wan
LTX

Real time video generation (virtua

lucy.decart.ai

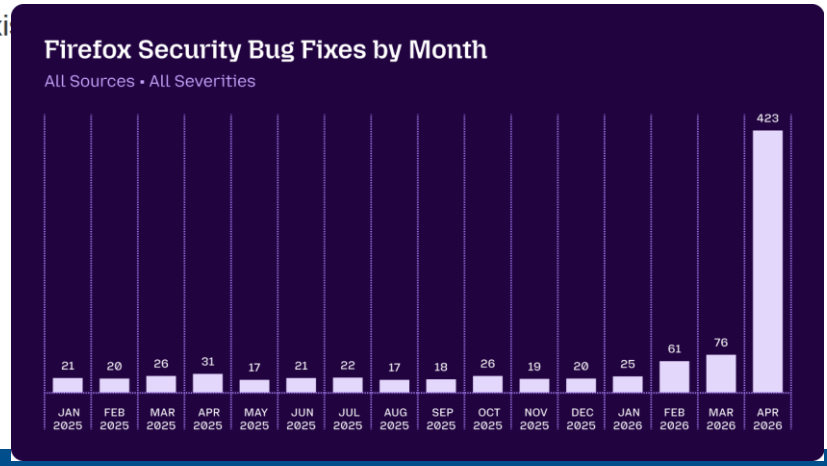
hacksider/Deep-Live-Cam
philipp-eisen/facestream

Finance worker pays out $25 million after video call with deepfake 'chief financial officer'

https://edition.cnn.com/2024/02/04/asia/deepfake-cfo-scam-hong-kong-intl-hnk

Haotian AI???

The three finger test



Parallel sessions

Track 1 - Data and metadata organization, management and access

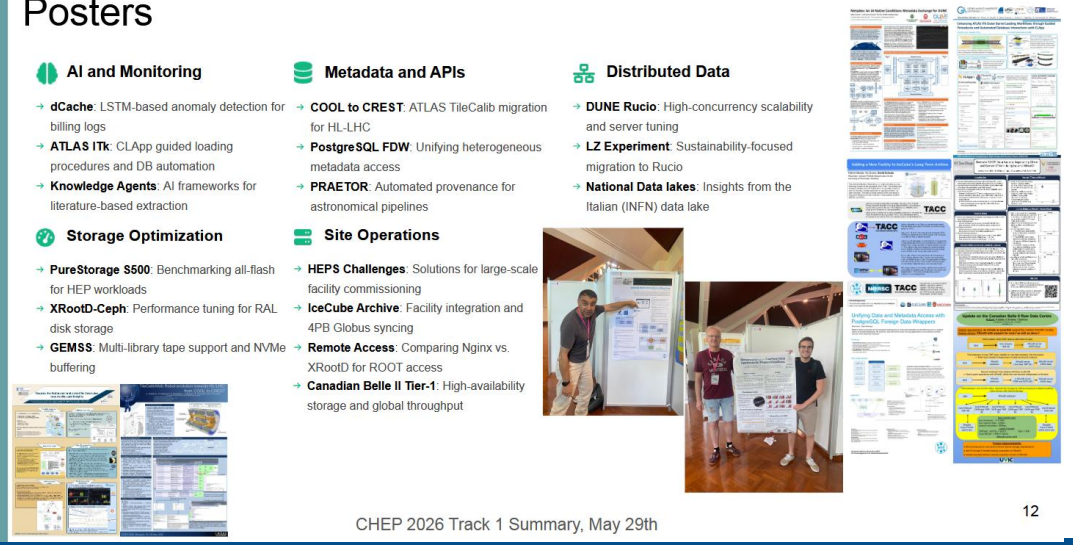
- 2 plenary talks and 9 parallel sessions with 50 talks
- 16 posters
- Great attendance and very interesting discussions!
- Sessions structured into topics:
  - Tape storage, archival and long term preservation
  - Rucio, data lakes and distributed data management (DDM)
  - Data transfers, federations and infrastructure planning
  - XRootD ecosystem and data access
  - Data integrity and storage reliability
  - FAIR data, metadata and preservation
  - Storage systems and file system protocols
  - Compression and I/O optimization
  - Databases for experiment data and operations

Track 1 at a glance - main trends

- Strong **collaborations** built across communities - HEP, astronomy, nuclear physics, dark matter, etc
- Move from legacy in-house to **well-supported** tools to improve maintainability
- **FAIR** data principles for long-term data preservation - *Findable, Accessible, Interoperable and Reusable*
- **Standardise** metadata, protocols and access policies across scientific communities
- Careful **planning and testing** of infrastructure and tools to achieve target performance
- **HPCs are still a challenge** for HEP workflows, in particular for data access
- **Reliability and integrity** checks on data are further automated
- **AI/ML** entering in operations, optimization and infrastructure management

- **Getting ready for the exascale demands of the HL-LHC !**

Posters



Summary of Track 2

Online and Real-time Computing

- 55 oral presentations + 28 posters *Links to all*
- Focus areas covered:
  - Trigger systems and real-time reconstruction
  - Machine learning for online inference
  - FPGA/GPU acceleration
  - Data scouting and trigger-level analysis
  - Online monitoring and operational infrastructure
  - Scalable DAQ and heterogeneous computing

Near-Term Outlook

- Wider deployment of AI-assisted triggers during Run 4 and further
- Production-level FPGA/GPU ML inference pipelines
- Expansion of trigger-level analysis programmes
- More autonomous monitoring and operations

Longer-Term Prospects

- Foundation-model-inspired event representations
- Unified online/offline reconstruction frameworks
- Real-time adaptive trigger strategies
- AI-native DAQ systems
- Intelligent buffering and dynamic event retention

- Varying from pp colliders to fix-target and neutrino experiments
- Real-time computing is emerging in HEP physics capabilities
- Exploration of heterogeneous architectures: CPU & GPU & FPGA
- New technologies: CNNs, GNNs, Tranformers, Unsupervised learning etc
- Trigger-level analysis is expanding physics reach under fixed bandwidth limits
- Next generation trigger towards HL-LHC: from event filtering → to intelligent real-time physics computing

Track 3: Offline data processing

10 sessions, **54** talks, **17** posters

Key Topics

- Reconstruction:
  - Tracking
  - Calorimetry
  - PID
- Frameworks and core software
- New approaches

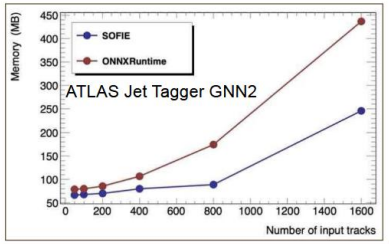
Common themes: timing info, heterogeneous hardware, ML

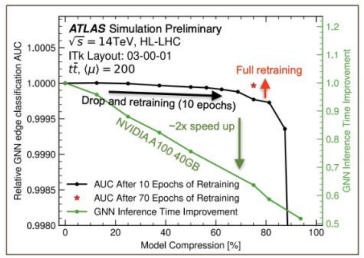
Frameworks: ML

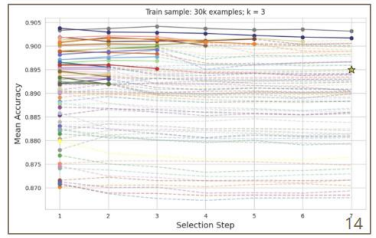
Contributors
Lorenzo Moneta, Vakho Tsulaia
Jay Chan, Alexey Boldyrev

Huge diversity of ML methods requires common tools:

- **SOFIE**: promising CPU results, WIP GPU optimisation
- **robust_select** (python library): identify robust model – 80% faster than exhaustive search + syst uncert
- **Inference as a Service** via NVIDIA TRITON: no improvement in small models, next: large models
- **GNN4ITK**: model compression, FP64 → FP32, structured pruning







14

Machine Learning

Integral part of reconstruction and analysis.
But deep understanding needed for ultimate performance.

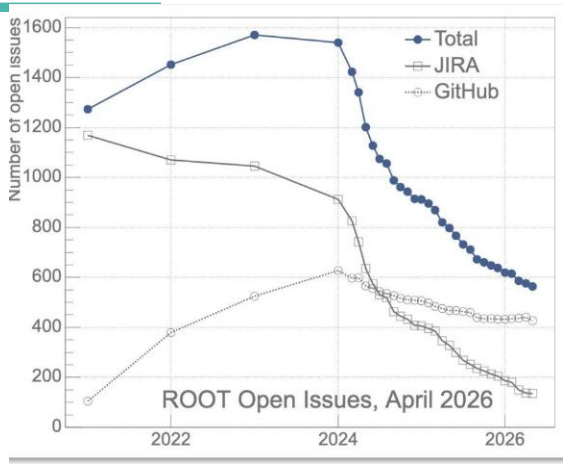
Modern Frameworks

From detector-agnostic to very specific and fine-tuned.

Different purposes and continuous modernization.

LS3 Opportunity

The Long Shutdown 3 provides a unique window for transformative changes in data structures and I/O standards.

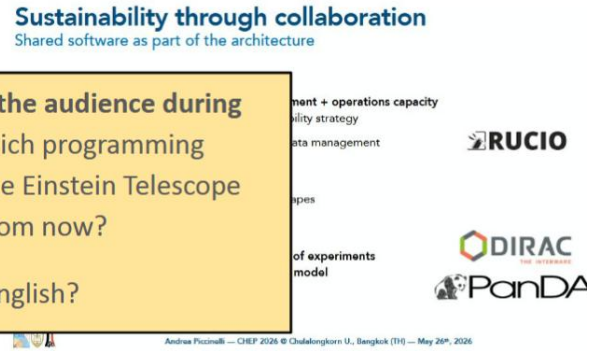


Track 4 Summary

Computing Model

DIRAC, PanDA, Rucio, CVMFS, IAM, and HTCondor have become key shared software components for sustainable computing mod
reported by

- ATLAS
- CMS
- Gravitational Waves commu
  IGWN - Art of Low Latency
- ePIC
  IceCUBE
- HERD
- JUNO
- SPD



Question from the audience during the ET Talk: Which programming language will the Einstein Telescope use ten years from now?

Paul Laycock: English?

Concurrently, workflow-oriented approaches using systems like CWL and DolphinScheduler are advancing to standardize and describe analysis and production flows as DAGs (e.g., ePIC, DIRAC-CWL).

13

HPC & Advanced system

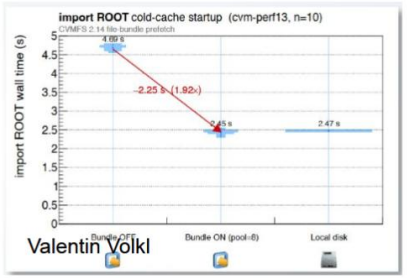
HPC Bubble used via interlink for Analysis Facilities by INFN.

Common experience in ATLAS and DUNE for the usage of HPC datacenters in USA.

Two edge mechanisms studied: Globus Compute MEP/NERSC Superfacility API (NERSC API)

DUNE 2×2 Near Detector Simulation chain is a deliberate stress test of workflow portability

Discussion on the Token lifecycle across trust boundaries, refresh, scoping, and audit still need per-site care.



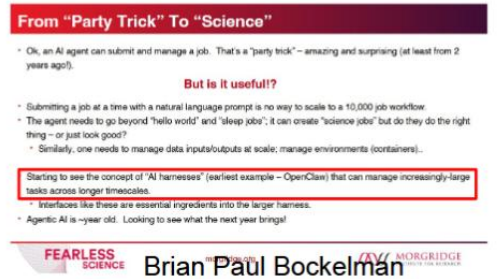
CVMFS remain a key components, FILEBUNDLES provides a new feature that can accelerate the startup of applications that need to read many files. Contributing to improve the global scalability.

15/15

AI for Operations

Systems like Dr.Sai/OpenDrSai, CMS Archi, Belle II AI assistant for shifter, PanDA Bamboo, and HTCondor WebUI+agents: use LLM-based agents to semi-automate job diagnosis, configuration generation, documentation agregation, tickets, logs, monitoring and data-management APIs.

Beyond the Chatbots, these operational assistants query real systems (e.g., monitoring, Rucio, WMS) via MCP and domain-specific tools, guiding operators through procedures.



Conclusions

Distributed Computing Directions

- Sustainability/Efficiency
- Security
- Heterogeneous Resources Integration
- AI Integration

During the track, several measurable advancements were presented

These 4 directions together with an increasing collaboration among communities will lead us toward and beyond the HL-LHC program and the challenge presented by the many experiments represented in the track.

Track 5 Summary: Event Generation and Simulation

Submitted 72 abstracts

Total 46 talks

Full sim ~ 2 sessions

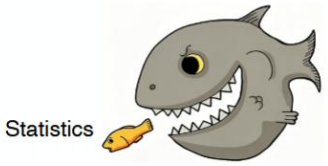
Fast sim ~ 4 sessions

Event generation ~ 2 sessions

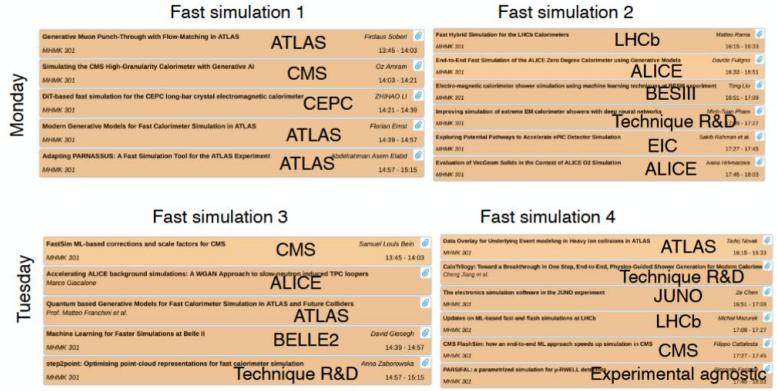
Tuning, calibration and validation ~ 1 session

Event generation

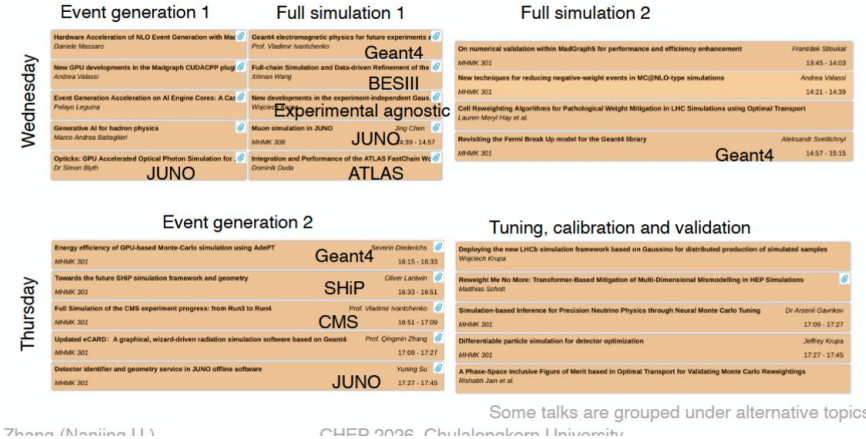
- Vectorisation / GPU / FPGA acceleration
  - Data parallelism in MadGraph [link, link] and PEPPE [link]
  - Case study gg → tt̄g on heterogeneous device [link]
  - Unfold detector acceptance and efficiency using ML [link]
- Accurate and efficient simulation
  - Reducing negative event weights in NLO generation [link, link, link]
  - Caching "everything" speeds up Fermi break up (break up of the light nuclei in nuclei interactions) with tolerant memory gain [link]



Schedule overview



Schedule overview



Some talks are grouped under alternative topics.

CHEP2026: Track 6 summary
Software environment and maintainability

Performance and Heterogeneous Computing

GPU KERNEL OPTIMISATION

ALICE

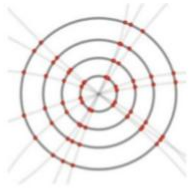
Automated GPU kernel tuning via Runtime Compilation

PORTABILITY FRAMEWORKS

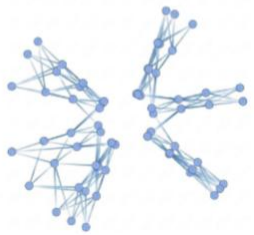
alpaka | Mod. Map Graph | ATLAS

C++ abstraction layer — one codebase across CPUs and GPUs

GNN tracking via Module Map Graph on heterogeneous architectures

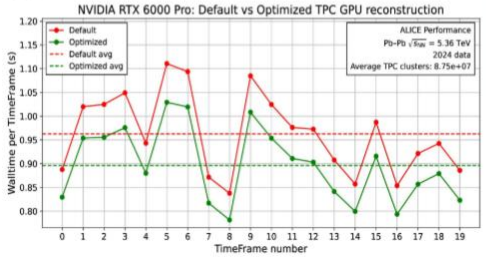


Hits



Graph

Tuning generalisation: Nvidia



Framework design

FRAMEWORK DESIGN

Phlex | DUNE | CMS | HGCAL | NOvA

DUNE is building a new not DUNE-specific framework from scratch.

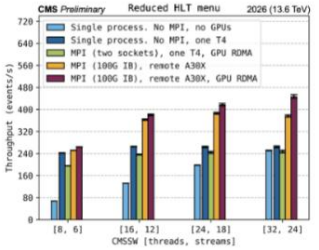
NOvA shows what happens over a decade

PIPELINE TESTING & INTEGRATION

LIGO / Virgo / KAGRA | Kubernetes | CMS | CMSSW

Modern HEP experiments no longer run monolithic software on a single machine. Their workflows span distributed infrastructure, streaming data, and containerized services.

GPU CMSSW modules can be offloaded to a remote server



AI/ML/LLM tools

A broad wave of AI-assisted tooling across storage development, knowledge preservation, code analysis, literature search and facility interfaces — AI as a productivity multiplier for developers and operators.

CERN / EOS

AI-Assisted Coding

AI code generation applied to storage software development

RHIC

SciBot

Locally deployed RAG assistant for secure natural-language access to private RHIC knowledge — blueprint for knowledge preservation

ATLAS / BNL

CelloAI

Vision LLM + callgraph-aware RAG for automated root-cause analysis of histogram discrepancies across ATLAS software versions.

INSPIREHEP

AI-Enhanced Literature Search

Embedding-enriched retrieval pipeline and MCP server bringing natural language queries and improved ranking to HEP literature discovery.

accelerators

RL-ABC

Reinforcement learning framework that automatically transforms Elegant beamline configs into RL environments for autonomous accelerator optimisation

Dialogic Software Interfaces

Conversational interaction paradigm replacing complex GUI workflows at large-scale photon facilities

Research Software Communities & Virtual institutes

HEP experiments share many software challenges. The community has created organisations that work across experiments — coordinating efforts, training people, and making the case that software deserves the same recognition as hardware.

CERN

HSF

HEP Software Foundation

A worldwide informal organisation that brings together people from different experiments to work on shared software problems. It does not fund projects directly, but creates working groups, organises workshops, and gives a common home to cross-experiment software efforts.

EU-funded

EVERSE

European Virtual Institute for Research Software Excellence

An EU-funded project working across five scientific communities — including HEP — to establish shared standards for research software quality. It makes best practices for writing, sharing and maintaining software available to everyone through a shared toolkit.

US / NSF

IRIS-HEP

Institute for Research and Innovation in Software for HEP

A US institute funded by the NSF, created to build the software infrastructure needed for the next generation of LHC physics.

CHEP 2026
Summary Track 7
Computing infrastructure and sustainability

- 27 poster
- 60 parallel talks
- 11 parallel sessions
  - 3 parallel parallel-sessions
- **very good discussions**

What was covered....

Opportunistic Resources | HPC and Exascale | Networking

Clouds | Orchestration of Virtual Machines and Containers

Computing Centre Infrastructure | Energy Efficiency

Environmental Impact and Sustainable

Computing Practices | Cost of Computing

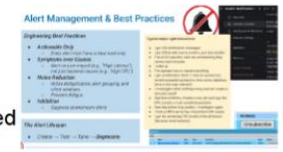
Management and Monitoring | Quantum Networks

Monday - 2b

Diogo Castro → Exploiting CPU&GPU resources of the **LHCb-trigger as an Analysis Facility**

S. González de la Hoz → **Computing Activities at the Spanish Tier-1 and Tier-2s**

Luboš Krčál → **Operating large HPC farm with a small team** - many lessons learned



Nikita Shadskiy → **Dynamic GPU Provisioning** for HEP at HPC centers

Brij Kishor Jashal → **WLCG becomes more and more GPU aware**

Giulio Bianchini → **user-level network overlay** enabling offloading of payloads from cloud-native interfaces

Track 8

Analysis Infrastructure, Outreach & Education

~30 talks & posters · 6 sessions over 3 days ·

Track 8 at a glance

FOUR THREADS RUNNING ACROSS THE WEEK

01

AFs share a stack

Eight sites, common ingredients

02

Open data as infrastructure

Releases → services

03

Reproducibility, designed in

Pin upstream and data cards

04

HEP training is scaled up

Catalogues, platforms, pedagogy

+ a fifth thread that surfaced: **the end user is increasingly not a human.** CI, LLM agents, MCP servers

Lessons, tensions, open questions

LESSONS LEARNED

Reproducibility decays

A strong publication process alone isn't enough; re-running old analyses is how you find the deficiencies (Lançon)

How we process data matters

In the Integration Challenge, large speedups were found and compression matters (Shadura)

Prepare to be flexible

MLTF abandoned disaggregated PCIe-enclosure GPUs (Melo); CERN's live-ASR pilot was discontinued over scalability (Saiz).

TENSIONS

Kubernetes-heavy vs K8s-free

Big facilities go all-in on K8s; INK (one-command deploy) and GSI keep it off the cluster for low-friction adoption.

RAG vs no-RAG assistants

Thin markdown-and-MCP layers (Lumi) vs full retrievers (Pandabot / SciBot) — no consensus yet.

Will open access strain infra?

The room asked twice — could public + AI-agent demand swamp physicists' infra? Saiz already throttles agent harvesting.

OPEN QUESTIONS

BaBar needs a home

1.3 PB; Council voted unanimously to find one; hardware is failing (Ebert).

Who adopts the ICFA recs?

Published; assessment starts at ~2-yr intervals. The next move is signing up (Doglioni).

Can the workflow-engine zoo last?

An audience member asked it directly: REANA supports 4, the CERN AF survey lists 9. Recommend, or facilities keep paying (Guerrieri).

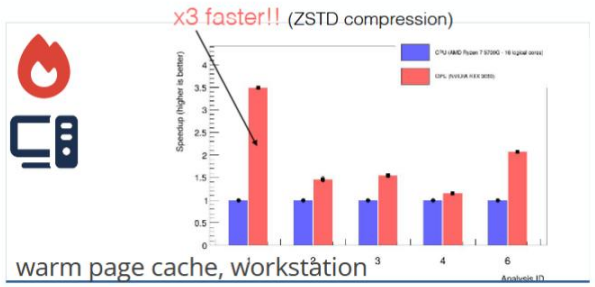
Track 9: Analysis Software Conference

- Software and workflows for user analyses
- Covered a wide range of topics, from analysis tools (ML aplications, GPU-accelerated analysis, histogramming, …), workflow management, and statistical inference to AI-assisted analysis.
- Many discussions on scalability, efficiency, reproducibility, and preparation for the HL-LHC analyses

- 44 talks in 8 sessions + 22 posters!
- 40 - 60 attendees per session

GPU-accelerated analysis

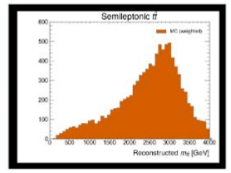
- Vendor lock-in is an issue when targeting GPU workflows: NVIDIA dominates the market. Moving to other vendors is not always a trivial code translation
- We've seen attempts [1],[2] of running analysis on GPU
- Crucial for benchmarking
  - Include I/O to GPU measurements
  - Compare full GPU against full CPU



LLM Applications

- LLMs are everywhere…
  - Code writing: graph-based Retrieval-Augmented Generation (CodeGraph RAG) system to improve LLM-generated HEP analysis code.
    - Using CodeGraph RAG + Error-Aware RAG + Claude Code Skills, error-feedback loops and ROOT documentation.
  - End-to-end analysis: Tool-based LLM framework for ATLAS Analysis

Plot the reconstructed top quark pair mass in events with one charged lepton in the rucio dataset <z-prime-mc-3000-gev>

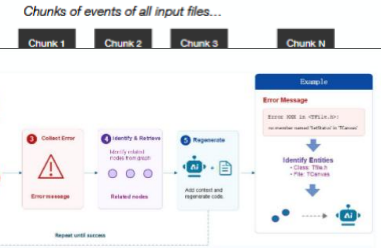


  - Analysis preservation: Retrieval Augmented Generation for analysis revival: AI method reproduces a previous LHCb analysis within 1σ of the original result → can be used for analysis reproducibility with expert guidance

Validation!!

Conclusions

- Track 9 showed a great ongoing effort towards more modern analysis
  - Analysis workflows are becoming increasingly modular and distributed. Now used
  - AI and ML are moving analysis frameworks
  - Differentiable and sim applied in real analys
  - GPU acceleration sta
- Good performance is



LLM Applications

- LLMs are everywhere…
  - Code writing: graph-based Retrieval-Augmented Generation (CodeGraph RAG) system to improve LLM-generated HEP analysis code.
    - Using CodeGraph RAG + Error-Aware RAG + Claude Code Skills, error-feedback loops and ROOT documentation.
  - End-to-end analysis: Tool-based LLM framework for ATLAS Analysis

Plot the reconstructed top quark pair mass in events with one charged lepton in the rucio dataset <z-prime-mc-3000-gev>

Semileptonic tt
Reconstructed m_{tt} [GeV]

Awkward Array
VECTOR
Boost.Histogram

  - Analysis preservation: Retrieval Augmented Generation for analysis revival: AI method reproduces a previous LHCb analysis within 1σ of the original result → can be used for analysis reproducibility with expert guidance

Validation!!

From CPU-Centric to Accelerator-Aware WLCG

CHEP 2026

25 May, Bangkok

Brij Kishor Jashal

Rutherford Appleton Laboratory

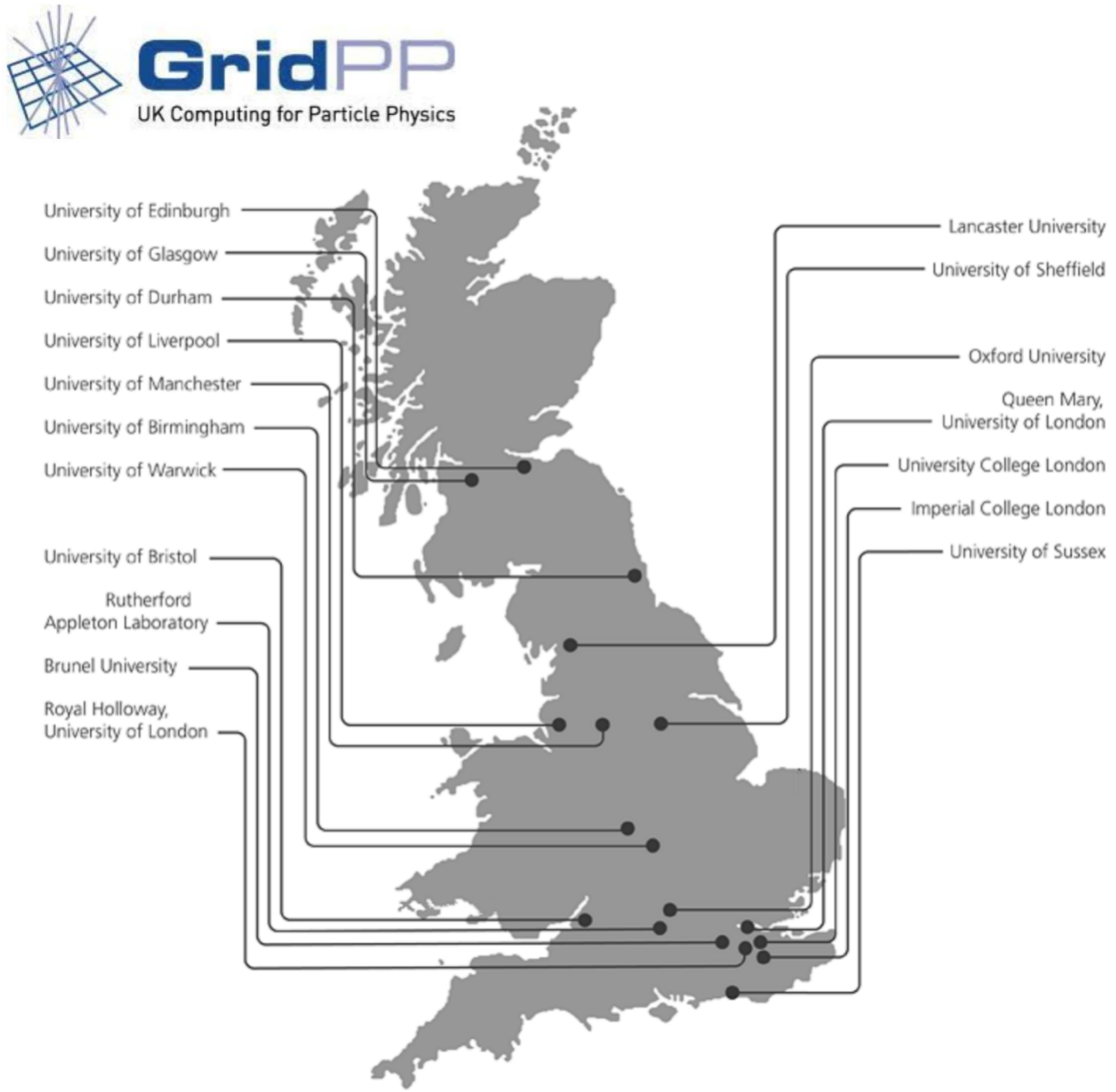
(On behalf of GridPP and RAL Tier1/Tier2 team)

Science and Technology Facilities Council
Rutherford Appleton Laboratory

GridPP
UK Computing for Particle Physics

Science and Technology Facilities Council
Particle Physics

GridPP UK WLCG sites



Multiple sites with GPUs in production

LondonGrid

- Brunel University London
- Imperial College London
- Queen Mary University Of London
- Royal Holloway, University of London

NorthGrid

- Lancaster University
- University Of Liverpool
- University Of Manchester
- University Of Sheffield

ScotGrid

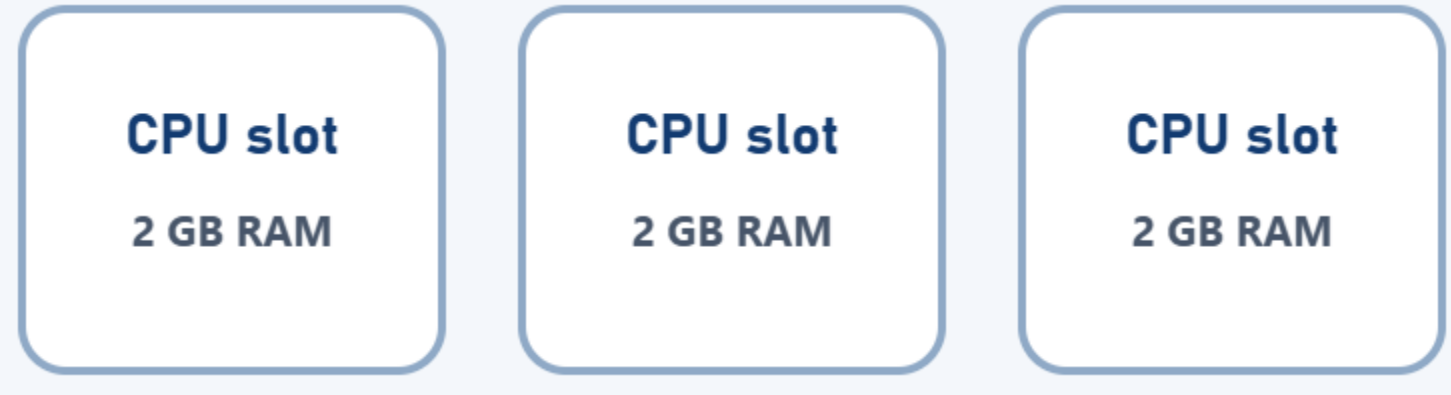
- Durham University
- University Of Edinburgh
- University Of Glasgow

SouthGrid

- University Of Birmingham
- University Of Bristol
- University Of Oxford
- Rutherford Appleton Laboratory Particle Physics Department
- University of Sussex

Which CPU-Era Assumptions Fail First?

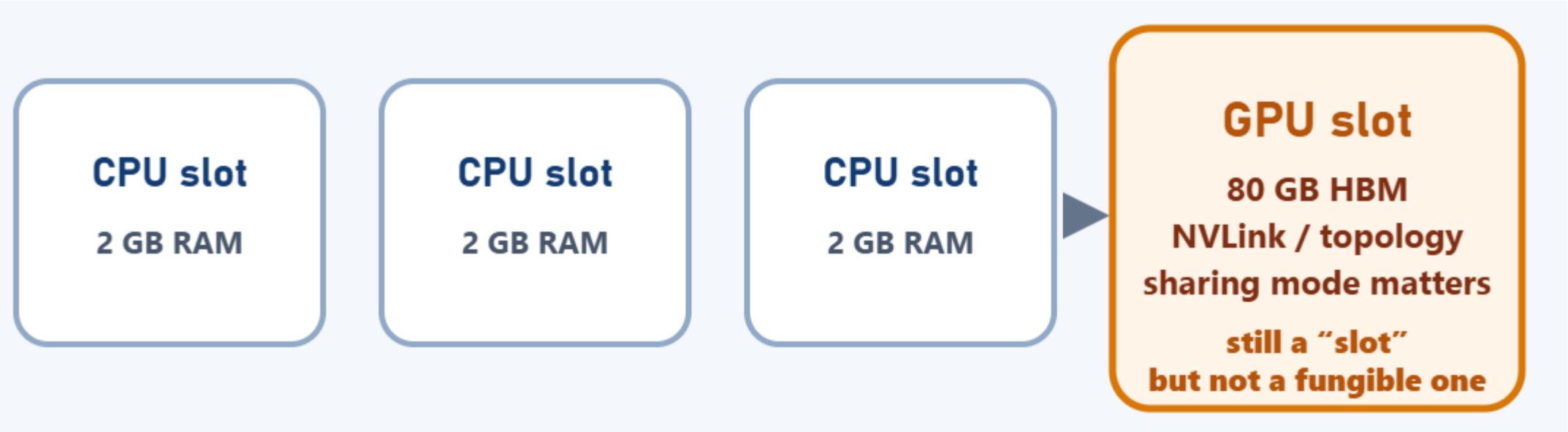
The classic Grid model assumes fungible compute slots



The Grid called everything a “slot”

Which CPU-Era Assumptions Fail First?

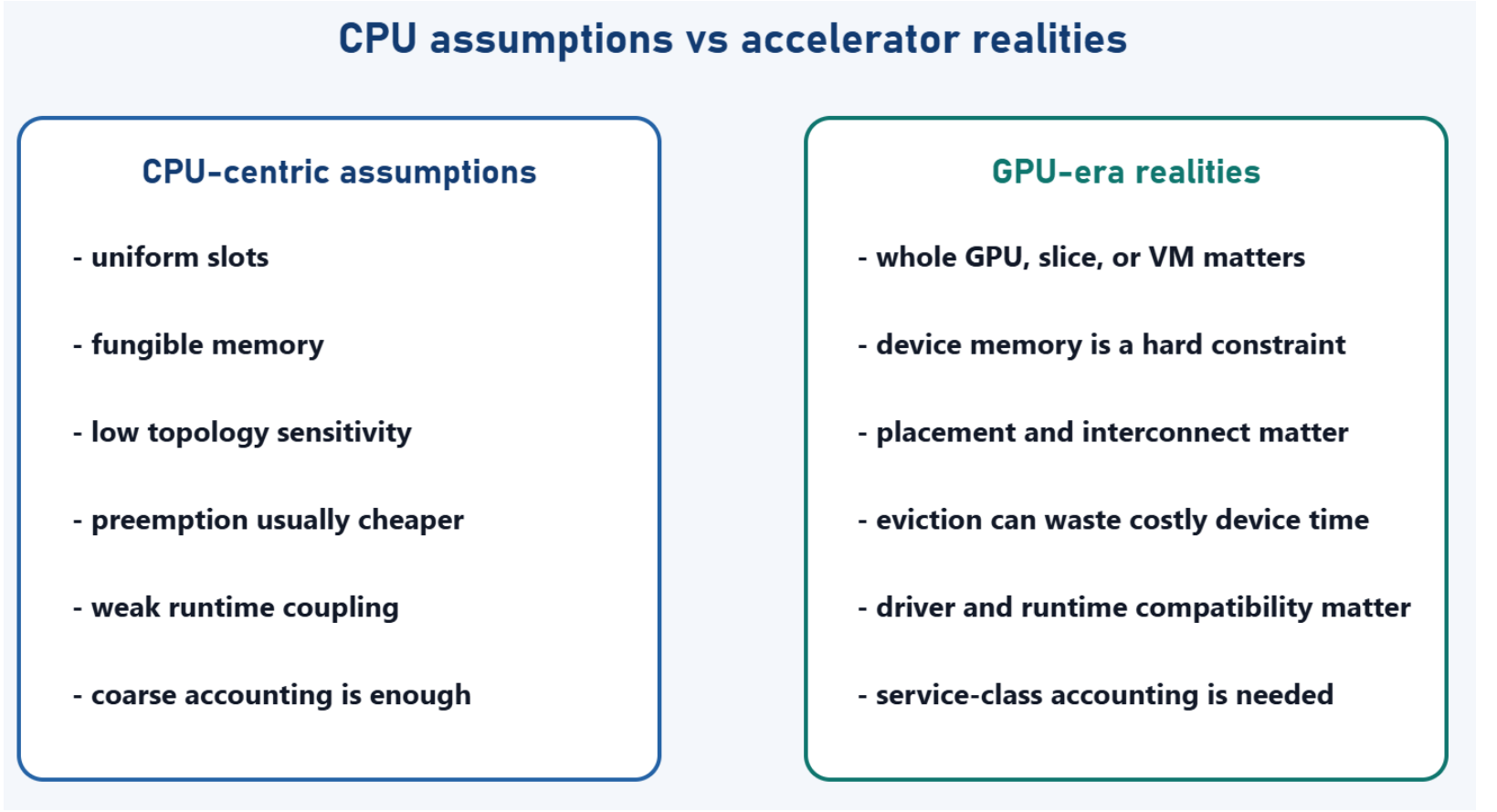
The classic Grid model assumes fungible compute slots
- GPU resources are non-fungible and break uniform slot assumptions.
- Sharing modes change performance and isolation.



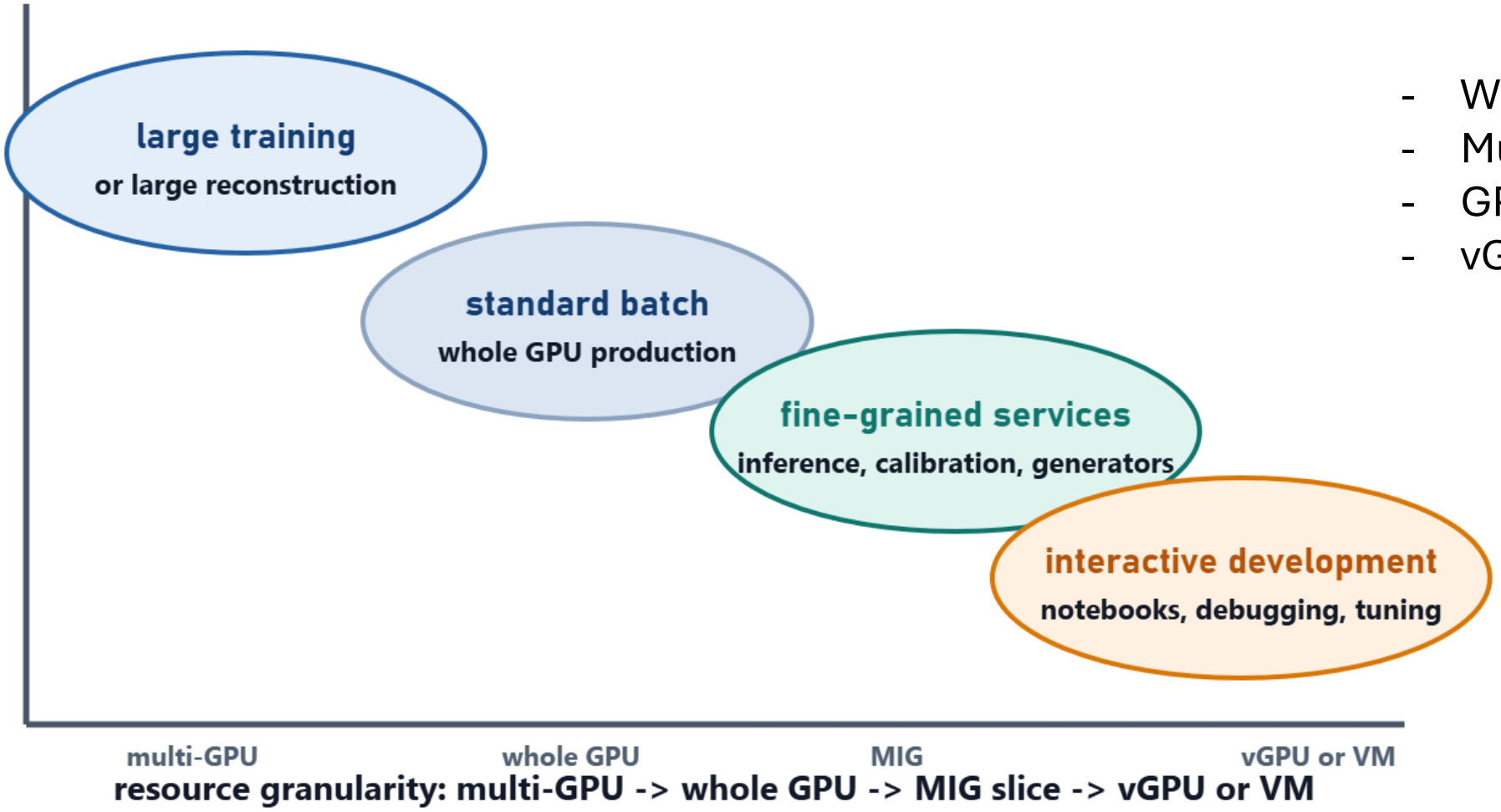
The Grid called everything a "slot" , GPUs arrived with a legal team !!

Which CPU-Era Assumptions Fail First?

The classic Grid model assumes fungible compute slots
- GPU resources are non-fungible and break uniform slot assumptions.
- Sharing modes change performance and isolation.

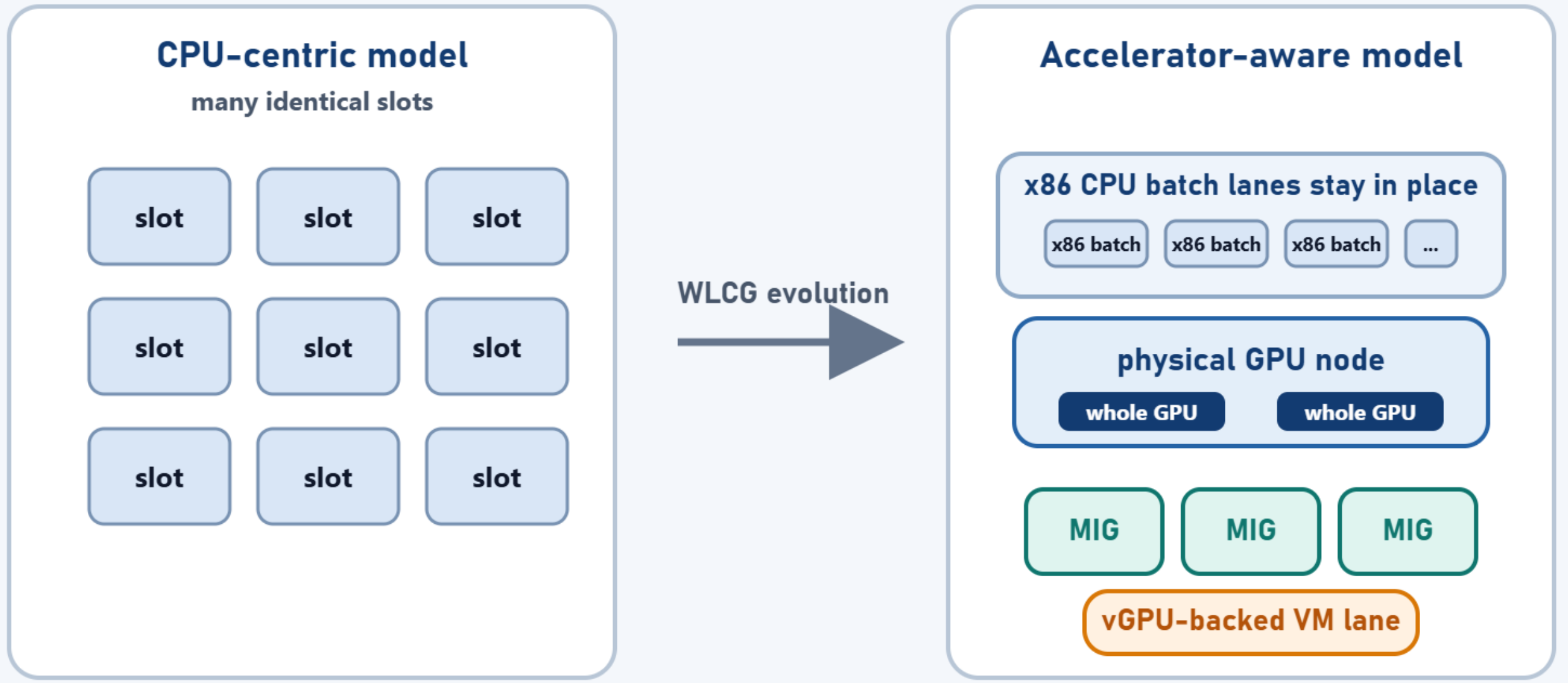


The GPU Workload Spectrum



- Whole GPU
- Multi-GPU island
- GPU slice
- vGPU or interactive VM

From CPU-Centric to Accelerator-Aware



One accelerator strategy is not enough

From physical GPU to schedulable partitioned service

Partitioning path

1. Physical GPU

One accelerator can be exposed as a whole device, as a hardware partition, or through VM-facing virtualization.

Generic WLCG labels
whole GPU, partitioned slice, vGPU

NVIDIA example

MIG (Multi-Instance GPU) mode creates hardware partitions.
GPU Instance (GI) defines the resource boundary.
Compute Instance (CI) creates the consumable compute context.

GPU Instance (GI) | Compute Instance (CI) | 1g.10gb

AMD example

Supported Instinct parts expose partitioning or SR-IOV-based virtual functions.
Scheduler should still advertise generic slice or virtualized service classes.

partition | SR-IOV | VF

3. Scheduler-facing service

Advertise whole GPU, partitioned slice, memory class, and virtualization policy rather than just device count.

Neighboring modes

Software sharing

CUDA streams, Multi-Process Service (MPS), and similar software sharing do not create hard partitions.

vGPU / SR-IOV VM path

SR-IOV = Single Root I/O Virtualization.
VM-facing sharing fits interactive and protected user workflows.
It is not the same as hardware partitioning.

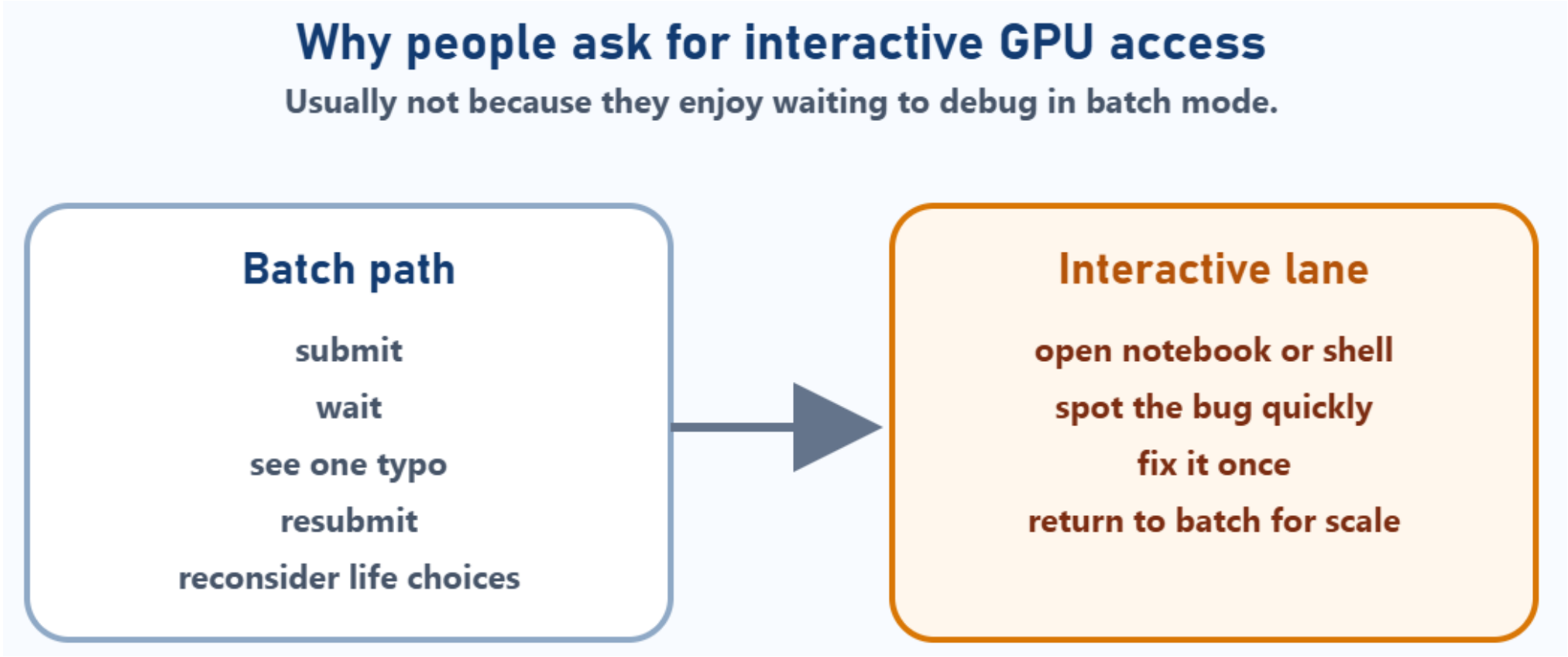
- MIG is hardware partitioning with isolated memory slices, SM slices, and engines.
- A GPU Instance defines the memory QoS boundary; Compute Instances subdivide compute inside it.

From CPU-Centric to Accelerator-Aware

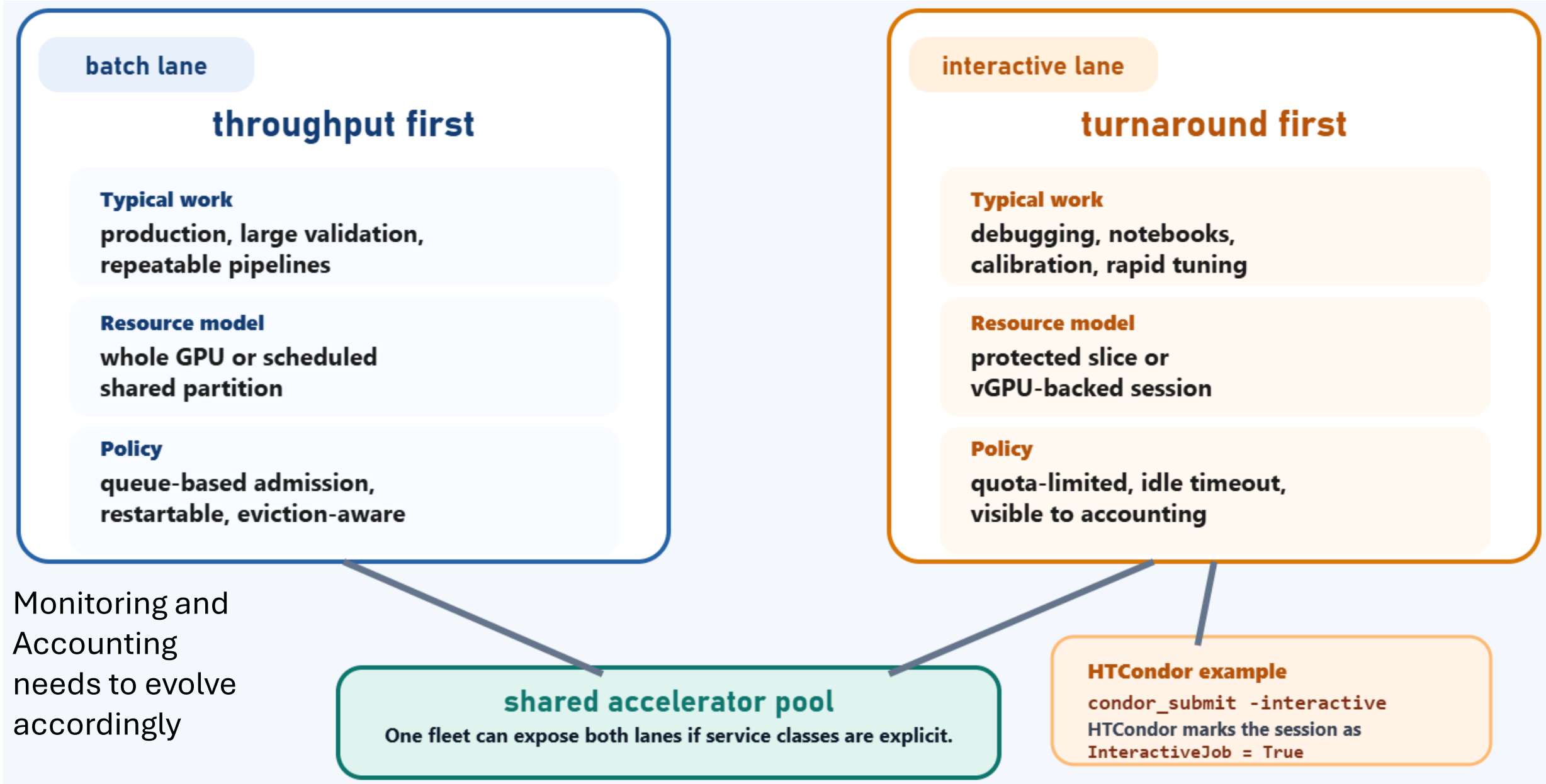
- **Vendor terms, supported GPU families and software stack anchors**

Vendor	Term and WLCG label	GPU families and marker	Software stack	Support note
NVIDIA	MIG = partitioned slice vGPU = virtualized access WLCG: slice or virtualized access	A100 / A30, Ampere, CC 8.0 H100 / H200 / GH200, CC 9.0 B200 / GB200, CC 10.0 RTX PRO Blackwell, CC 12.0	CUDA 13.2 Update 1 Driver floor: 580+ MIG starts with Ampere	Bare metal, pass-through, and vGPU supported vGPU can sit on MIG
AMD	Partitioning + SR-IOV or MxGPU WLCG: slice or virtualized access	MI300X / MI325X, gfx942 MI350X / MI355X, gfx950 Radeon PRO V710, temporal split MI210, SR-IOV only	ROCm 7.2.3 HIP 7.2.53211 AMD SMI 26.2.2 ROCm SMI 7.8.0	No one-word MIG term Use partitioning + SR-IOV for VM-facing sharing
Intel	No direct MIG analogue Use SR-IOV or virtual GPU WLCG: virtualized access or whole GPU	Flex family = virtualized path Max family = whole-device HPC Arc Pro B70 in recent OpenVINO releases	oneAPI Base and HPC Toolkit OpenVINO 2026.1 GPU plugin: Xe1 and Xe2+ oneDNN v3.10	Virtualization-oriented offer No direct fixed-slice term use generic service labels

Interactive vs batch lane



Interactive vs batch lane



Monitoring and Accounting needs to evolve accordingly

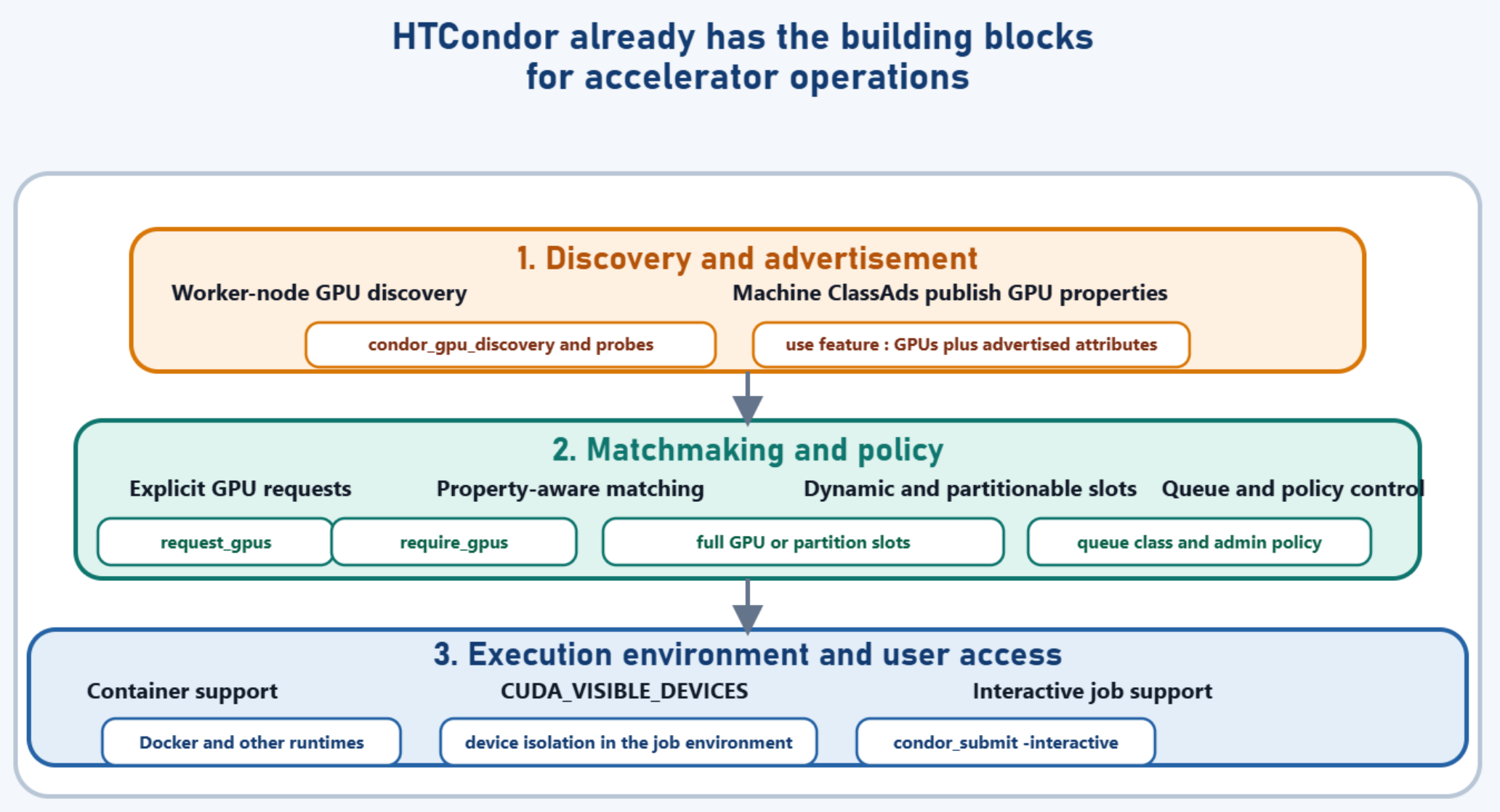
Scheduling Policy by Workload Class

Scheduling policy should follow workload class

workload class	resource	queue class	preemption	accounting	runtime
large training or large reconstruction	exclusive whole GPU or multi-GPU island	exclusive batch	minimal	whole-device hours	long
standard production batch	whole GPU	standard GPU batch	moderate	whole-device hours	medium
fine-grained throughput services	MIG partition	shared batch	allowed if restartable	slice-profile hours	short
interactive development and debugging	vGPU or protected slice	interactive lane	tight policy	session time	short to bursty

What HTCondor Already Gives Us

- Explicit GPU requests and property-aware matching
- Partitionable and dynamic slots
- Interactive job support
- Container execution environments



GPU resources at RAL

GPU Vendor	GPU Model	Infrastructure*	Intended Use	Number of Physical Cards
NVidia	Quadro RTX4000	STFC Cloud	Interactive Workloads	232
NVidia	RTX A4000	STFC Cloud	Interactive Workloads	176
NVidia	A16	STFCCloud	vGPU Intereactive Workloads	44
NVidia	TESLA V100 PCIe 32GB	STFC Cloud	Compute Workloads	124
NVidia	A100 PCIe 40GB	STFC Cloud	Compute Workloads	112
NVidia	A100 PCIe 80GB	STFC Cloud	Compute Workloads	156
NVidia	A100 SXM 40GB***	STFC Cloud	Compute Workloads	100
NVidia	H200 NVL	STFC Cloud	Compute Workloads	4
AMD	Radeon PRO W6600	STFC Cloud	Test GPU	4
Intel	Arc A770 16GB	STFC Cloud	Test GPU	4
NVidia	RTX PRO 6000 Blackwell Server Edition	RAL Tier 1	MIG	4
Nvidia	TITAN X (Pascal)	RALPP-Tier2	Interactive/batch	1
Nvidia	Tesla P100-SXM2-16GB	RALPP-Tier2	Interactive/batch	4
Nvidia	TITAN RTX	RALPP-Tier2	Interactive/batch	4
Nvidia	A100-PCIE-40GB	RALPP-Tier2	Interactive/batch	1
AMD	Instinct MI210	RALPP-Tier2	Interactive/batch	2
NVidia	A100 PCIe 80GB	RALPP-Tier2	Interactive/batch	4

Production queues at several other GridPP sites,e.g Manchester, QMUL, ...

Industry GPU roadmap

Official vendor roadmaps point to more heterogeneity, not less

1. Memory and bandwidth keep rising

NVIDIA H100 NVL: 188 GB HBM3 and up to 3.9 TB/s.
AMD MI325X: 256 GB HBM3E and 6 TB/s.
AMD MI350 series: up to 288 GB HBM3E and 8 TB/s.
Intel Max Series: up to 128 GB HBM plus large L2 cache.

2. Scale-up domains are larger

NVIDIA GB200 NVL72 exposes a 72-GPU NVLink domain.
NVIDIA cites 130 TB/s rack-scale communication.
AMD MI350 platforms: 8 connected GPUs, 64 TB/s aggregate.
Intel Max uses Xe Link for scale-up and scale-out.

3. Resource shapes are diverging

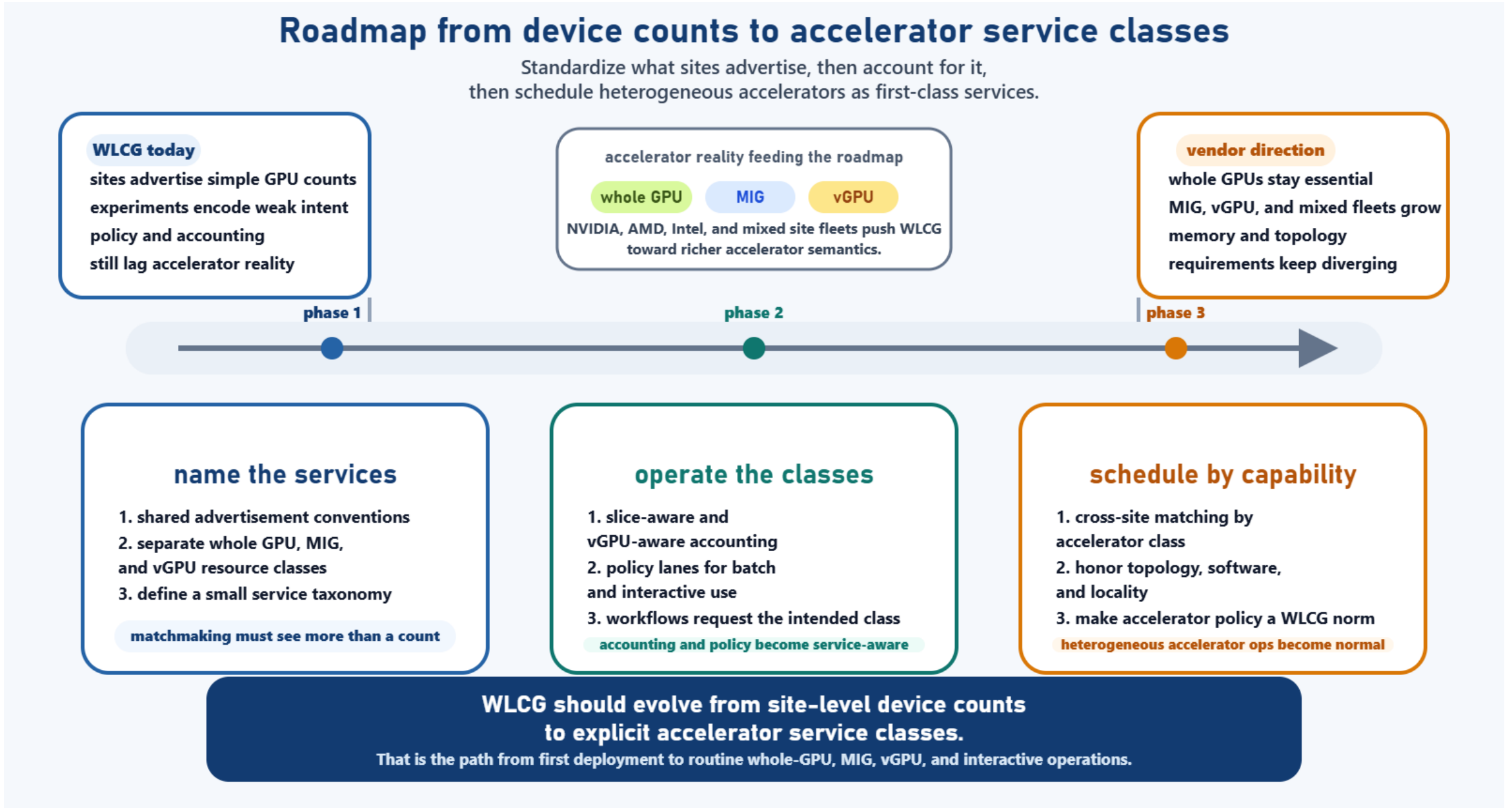
NVIDIA adds fine-grained MIG slicing beside full-GPU modes.
AMD spans PCIe cards, OAM platforms, and MI300A APUs.
Vendors now expose single GPU, multi-GPU, and package forms.
Schedulers must reason about memory, topology, and access mode.

4. Software stacks diverge too

NVIDIA pushes AI Enterprise, NIM, NVLink-scale systems, and MIG.
AMD pushes ROCm and enterprise AI tooling.
Intel pushes oneAPI and performance-portable software.
Queue policy now depends on runtime fit as much as hardware.

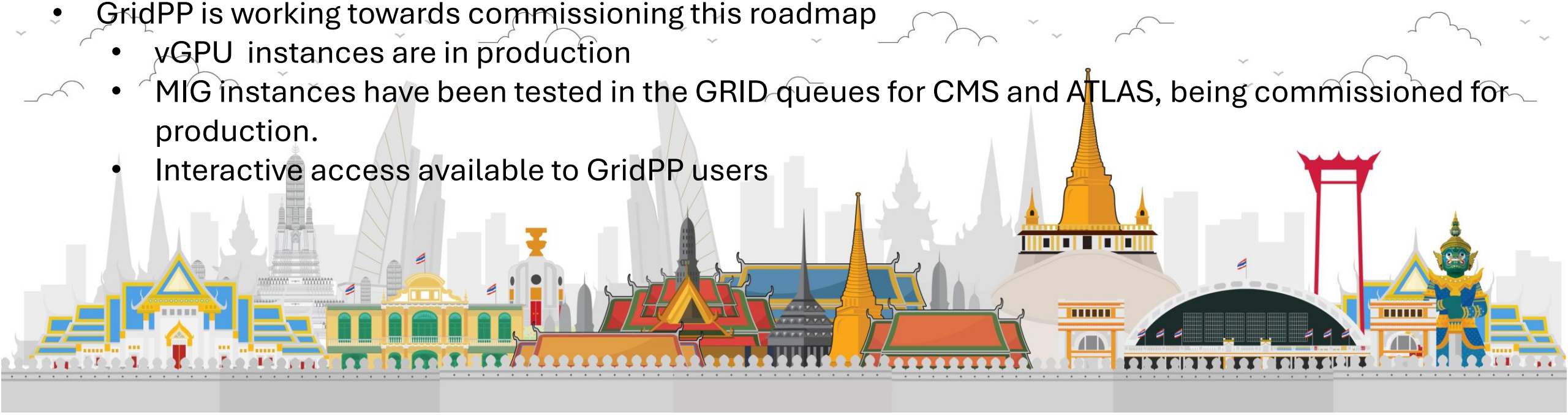
WLCG pressure comes from bigger memories, bigger domains, more resource forms, and more runtime choices.

Possible roadmap for architecture aware grid

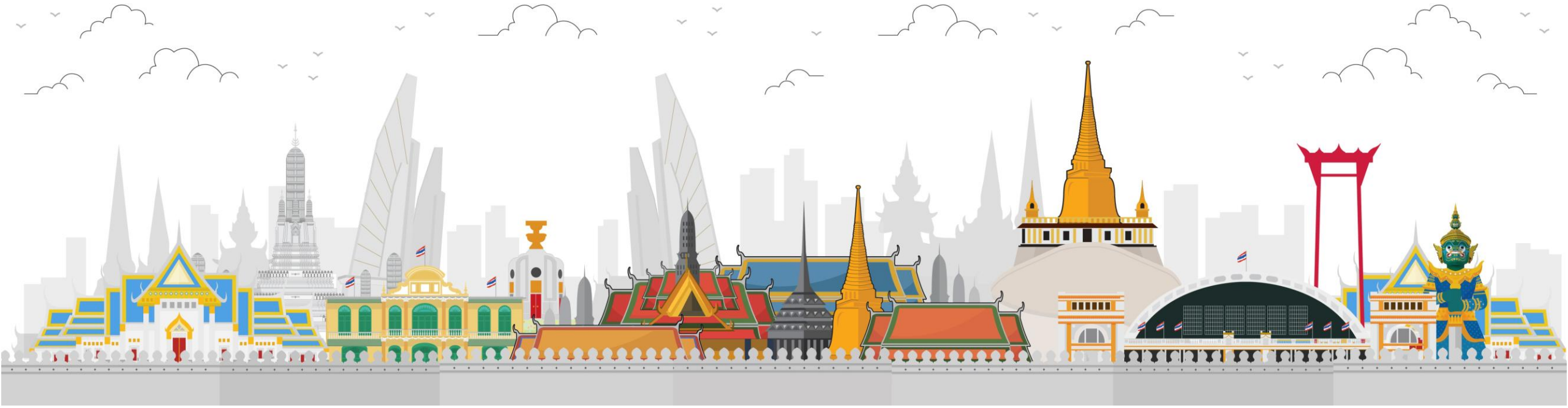


Executive summary

- Our software stack growingly supporting GPUs (Generation, simulation, reconstruction, user analysis)
- The challenge is no longer whether WLCG can attach GPUs in batch queues.
- The challenge is whether WLCG can schedule, share, account for, and operate them at the right granularity.
  - AI native WLCG will need to get this right.
- Whole GPUs remain essential
  - Partitioned GPUs expand usable capacity
  - Interactive usage deserve policy support
    - (unless if we think lxplus-gpu and institutional interactive access is enough for community)
- GridPP is working towards commissioning this roadmap
  - vGPU instances are in production
  - MIG instances have been tested in the GRID queues for CMS and ATLAS, being commissioned for production.
  - Interactive access available to GridPP users



Thank you



Fair-Share Versus Opportunism in Multi-VO Environments: The Complexity of Job Slot Allocation at the RAL Tier-1

CHEP 2026

25 May, Bangkok

Brij Kishor Jashal

Rutherford Appleton Laboratory

(On behalf of RAL Tier1 and Tier2 team)

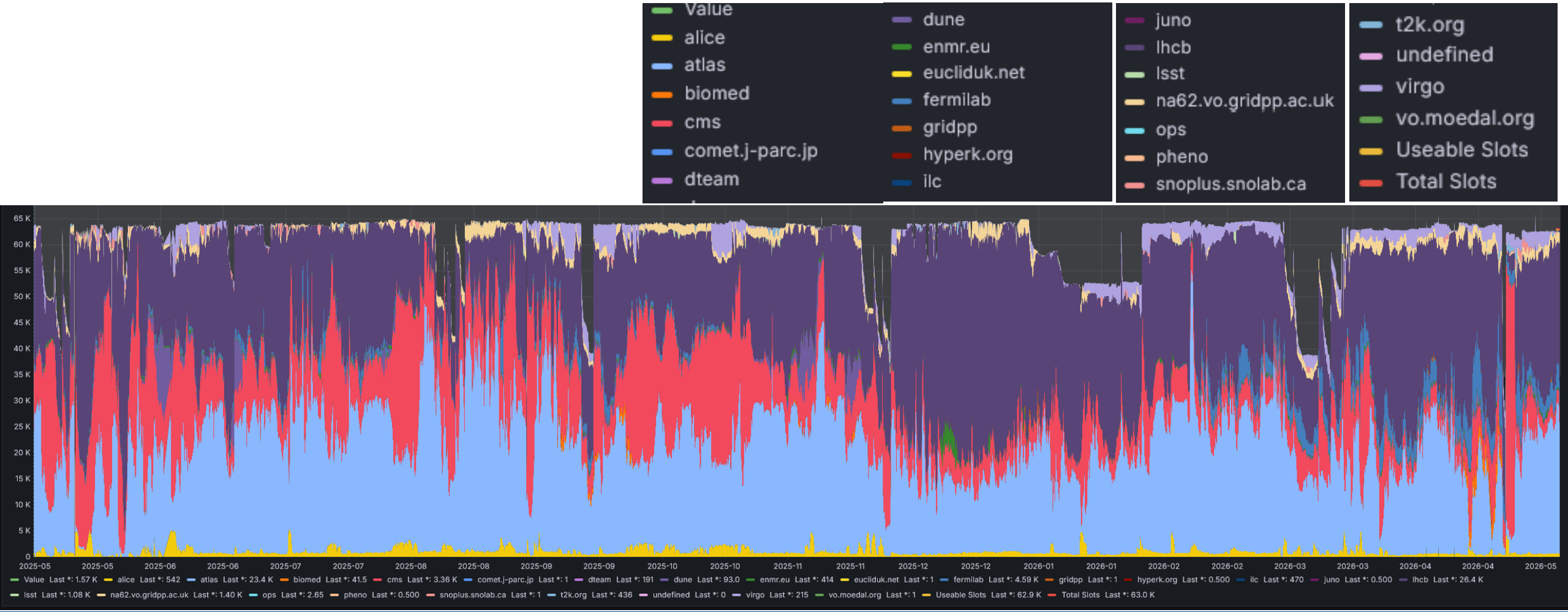
Science and Technology Facilities Council
Rutherford Appleton Laboratory

GridPP
UK Computing for Particle Physics

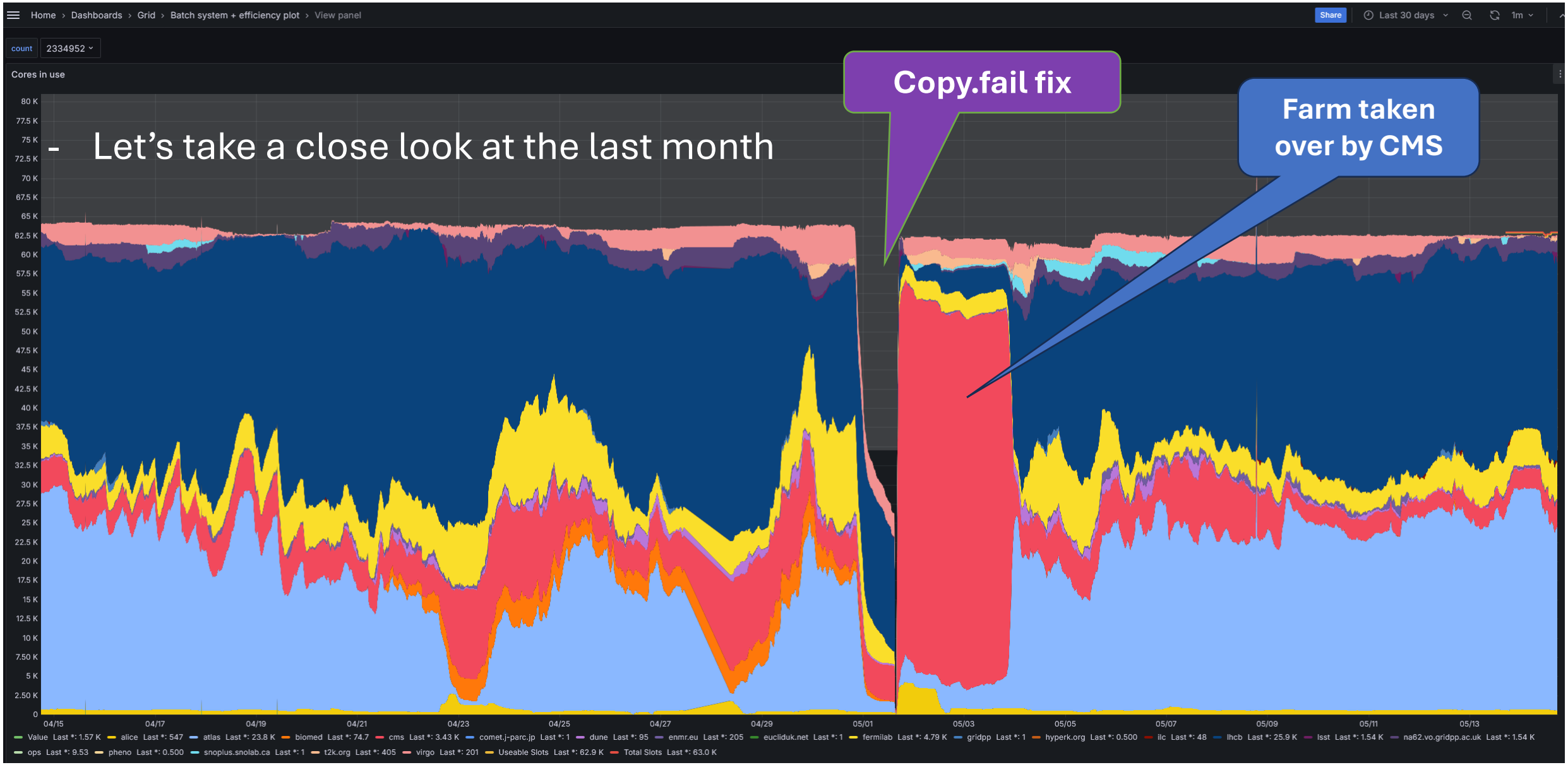
Science and Technology Facilities Council
Particle Physics

RAL-LCG2 farm occupancy: last one year

- In this beautiful snapshot:
  - Full occupancy during last year using 63K cores
  - Serving 24 VOs, meeting pledges and beyond

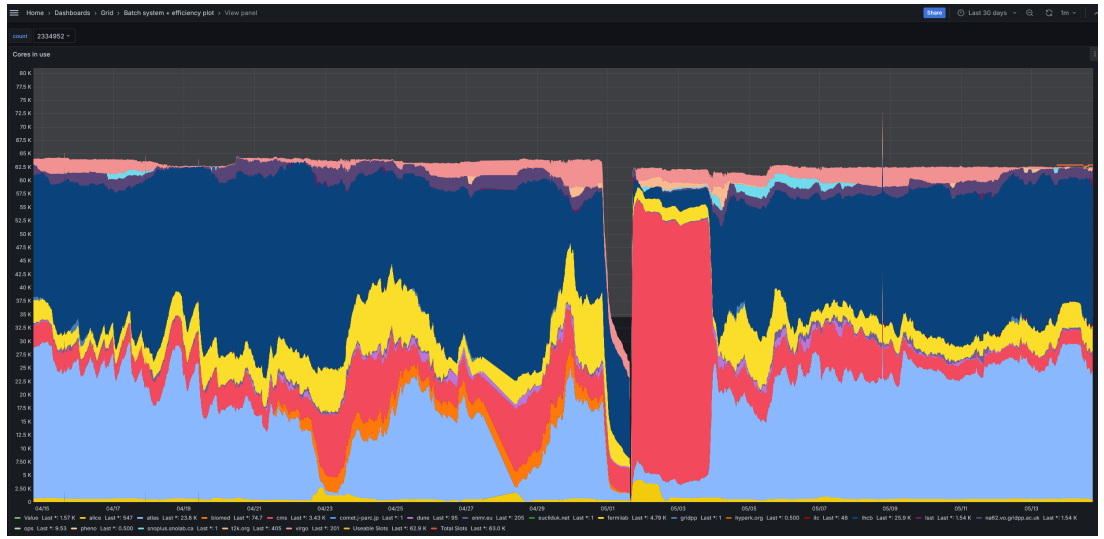


RAL-LCG2 farm occupancy: last 30 days



RAL-LCG2 farm occupancy: last 30 days

- In this plot
  - Despite downtime and disruptive activity, batch farm is fully occupied with minimal impact.
  - Interplay of fair share dynamics as well as difference in the submission logic of different VOs becomes crucial
  - Ability of certain VOs to be able to scale quickly and occupy the farm is an important reason why we see such a good overall occupancy.
  - However, that also means that in the aftermath of events such as rolling updates and farm restarts, submission logic of VOs manages to dominate over the fair share logic of the farm.



The problem

- Fair-share policy and observed occupancy are not the same thing.

- The main operational target is: guarantee baseline shares for major VOs, avoid starvation for smaller communities, and still harvest idle capacity opportunistically.

- In practice, the critical question is not only who gets resources, but how quickly the system adapts when demand changes.

The problem

- In multi-VO environment HTCondor fairness is not one knob; it is a stack of controls spanning accounting, quotas, preemption, slot weighting, and negotiation timing.
    - HTCondor gives central-manager controls for this through the negotiator, plus supporting controls in the startd and schedd.
    - Before we take a look at some of these configuration knobs, let’s understand few things

How HTCondor Actually Enforces Fairness

- RUP remembers recent usage; EUP decides allocation order.

Real User Priority (RUP)

$$RUP(u,t) = \beta\, RUP(u, t-\Delta t) + (1-\beta)\, \rho(u,t)$$

Decaying memory of recent resource use.

- The negotiator sorts submitters by EUP, computes a nominal slice from inverse EUP ratios, matches jobs, then respins.

Effective User Priority (EUP)

$$EUP(u,t) = RUP(u,t)\, f(u,t)$$

Lower EUP means better priority in negotiation.

RUP tracks recent resource use.
EUP = RUP x priority factor.
Lower EUP is better.

Illustrative EUPs from pledge shares:
ATLAS 1.0, ALICE 11.2, CMS 3.9
LHCb 1.2, Other 1.0

Inverse-EUP weights:
ATLAS 0.96, ALICE 0.09, CMS 0.26
LHCb 0.86, Other 1.00

After normalization, the nominal fair-share percentages are:
ATLAS 30.5%, ALICE 2.8%, CMS 8.1%
LHCb 27.0%, Other 31.6%

How HTCondor Actually Enforces Fairness

- Fairness acts at negotiation time; recovery of already occupied slots depends on turnover or preemption.
- Ordering comes from EUP; slice size comes from inverse EUP ratios.

Nominal Slice From EUP

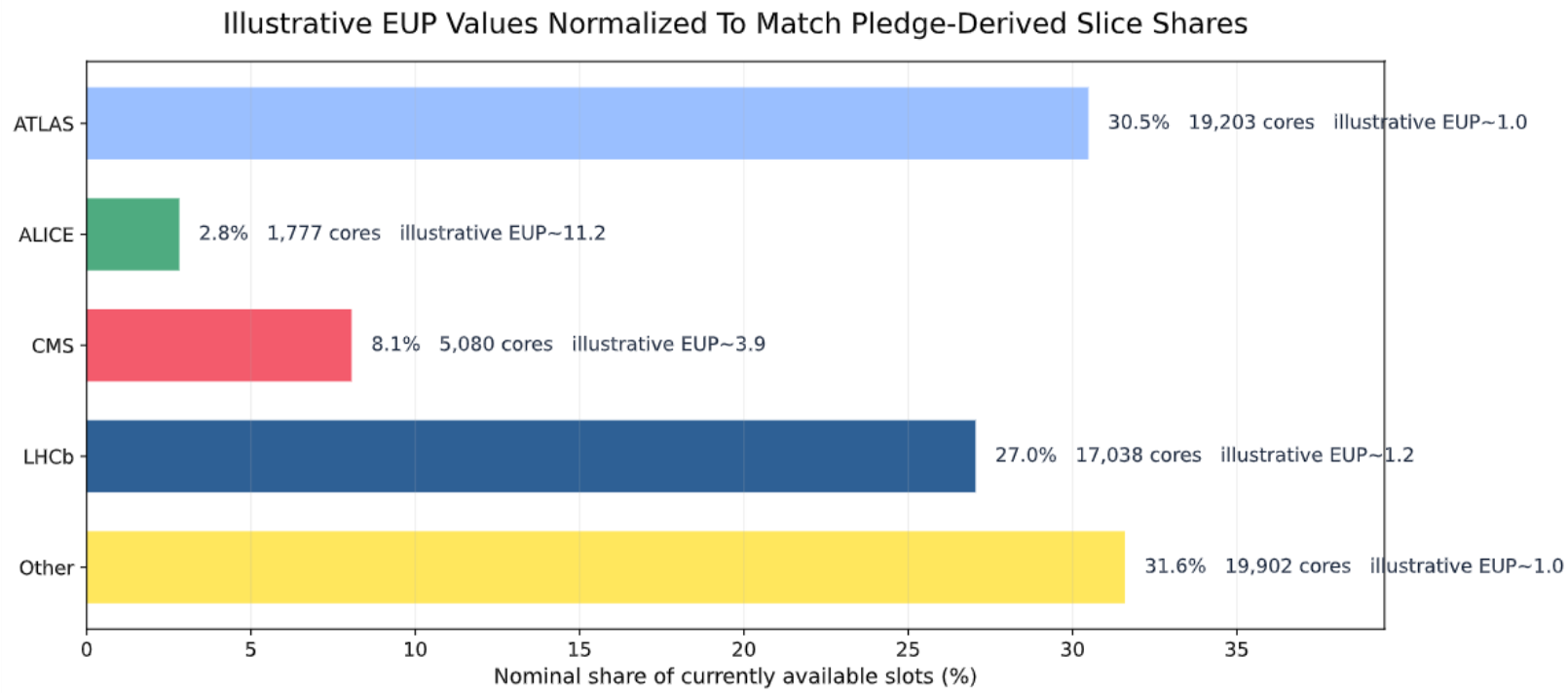
$$slice_i = \frac{1/EUP_i}{\sum_j 1/EUP_j}$$

Nominal share among the currently active submitters.

Nominal Slot Count

$$slots_i = N_{available} \times slice_i$$

Convert the slice fraction into currently available slots.



Example: ATLAS share = 30.5% and CMS share = 8.1%, so E_CMS / E_ATLAS = p_ATLAS / p_CMS = 30.5 / 8.1 = 3.78. If we instead normalize to the largest-share bucket at 31.6%, then E_ATLAS = 31.6 / 30.5 = 1.04 and E_CMS = 31.6 / 8.1 = 3.92.

Why PRIORITY_HALFLIFE Matters

The half-life h is set by PRIORITY_HALFLIFE

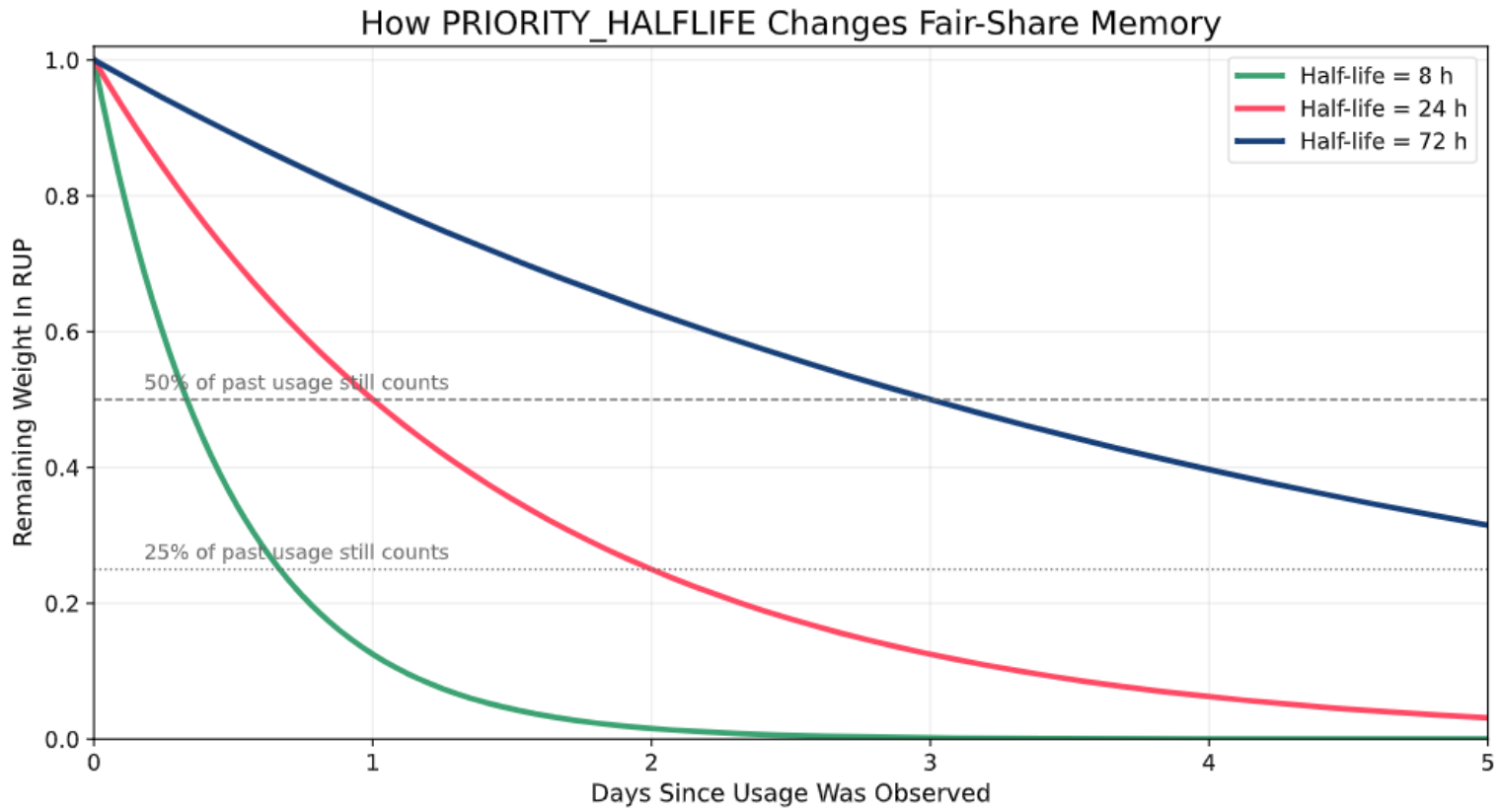
- Short half-life: faster correction after a demand swing.
- Long half-life: more stable, slower to forgive past heavy usage.

Decay Factor

$$\beta = 0.5^{\Delta t/h}$$

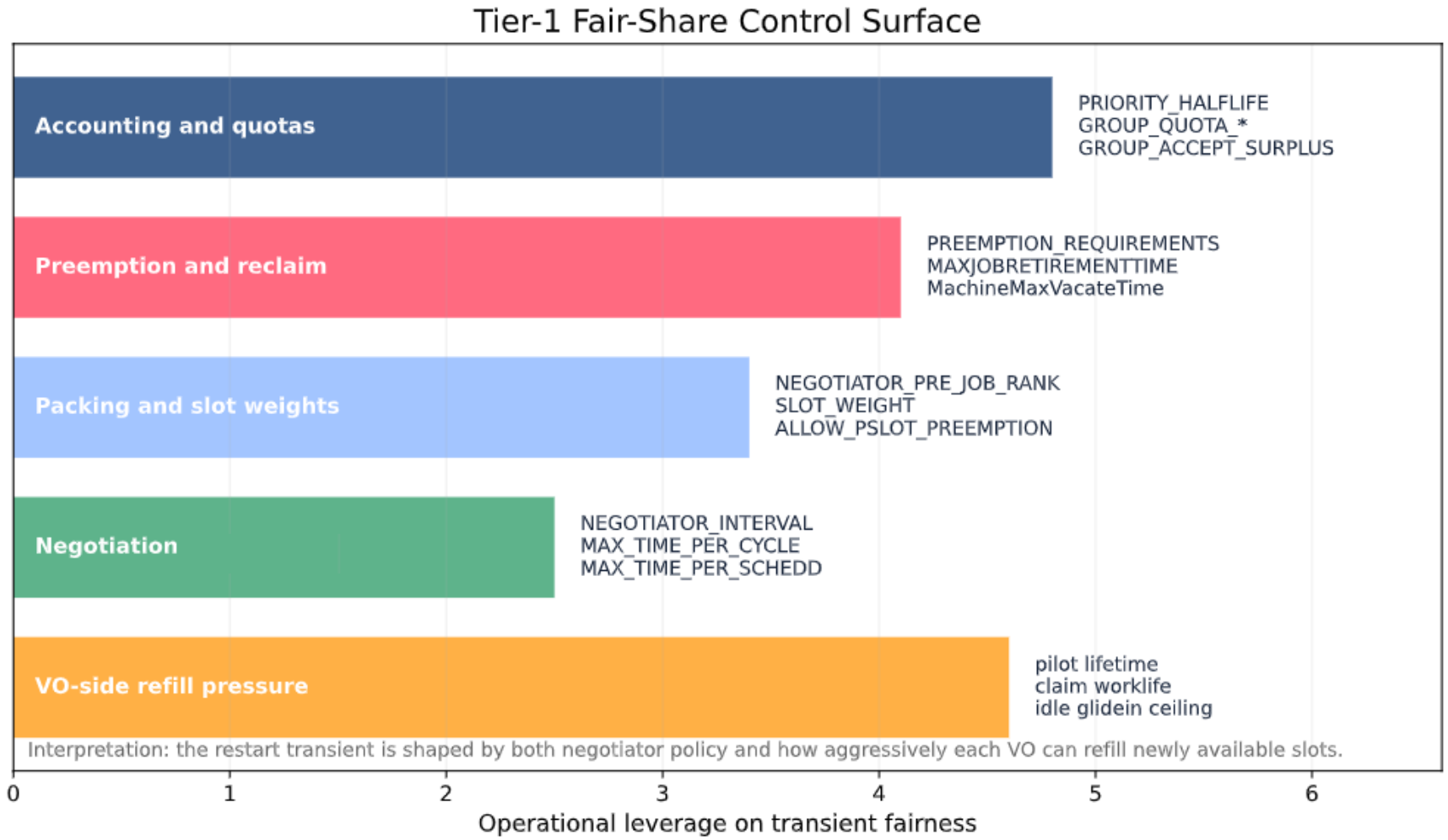
This knob changes how quickly the central manager notices unfairness, not how quickly long-lived pilots disappear.

“Shorter half-life makes yesterday's occupancy fade faster, so the negotiator reacts more quickly after a restart or demand swing.



The Tier-1 batch farm control surface

- Accounting and quotas define guaranteed share.
- Preemption and reclaim define how quickly share can be recovered.
- Packing, slot weights, and negotiation cadence shape how efficiently the pool implements policy.
- VO-side refill pressure is part of the same control loop.



Why there are differences in the VO refills

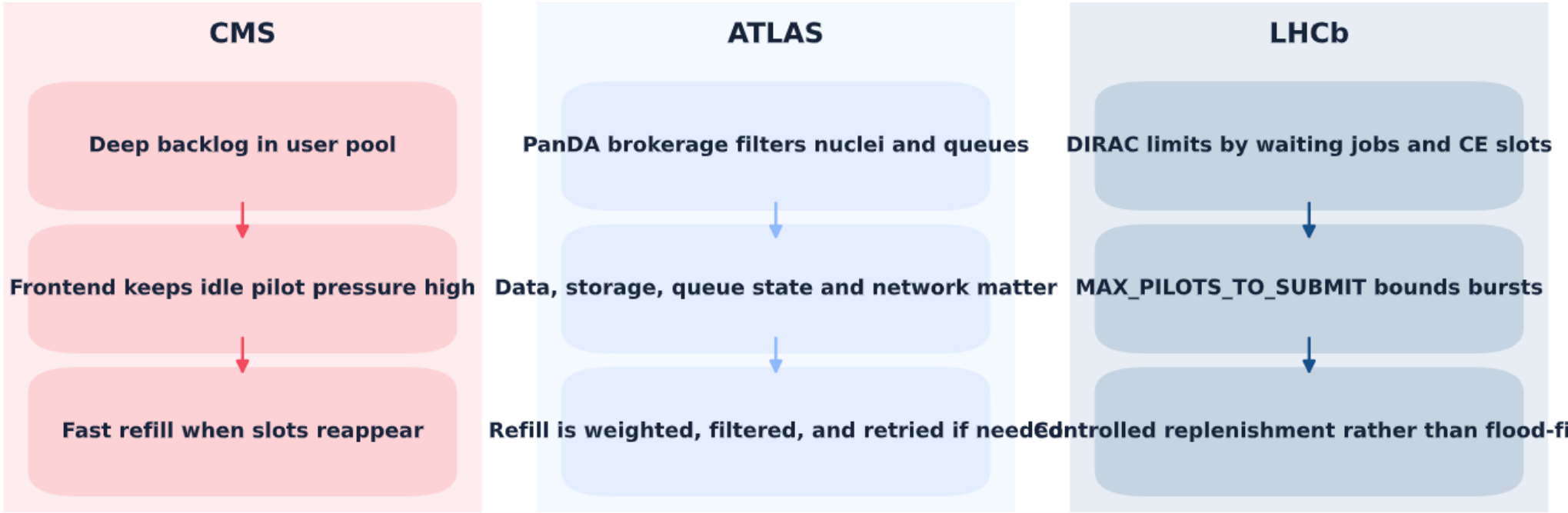
Transient occupancy depends on refill semantics, not only on nominal share targets

CMS Refills First (behaves like persistent pressure)

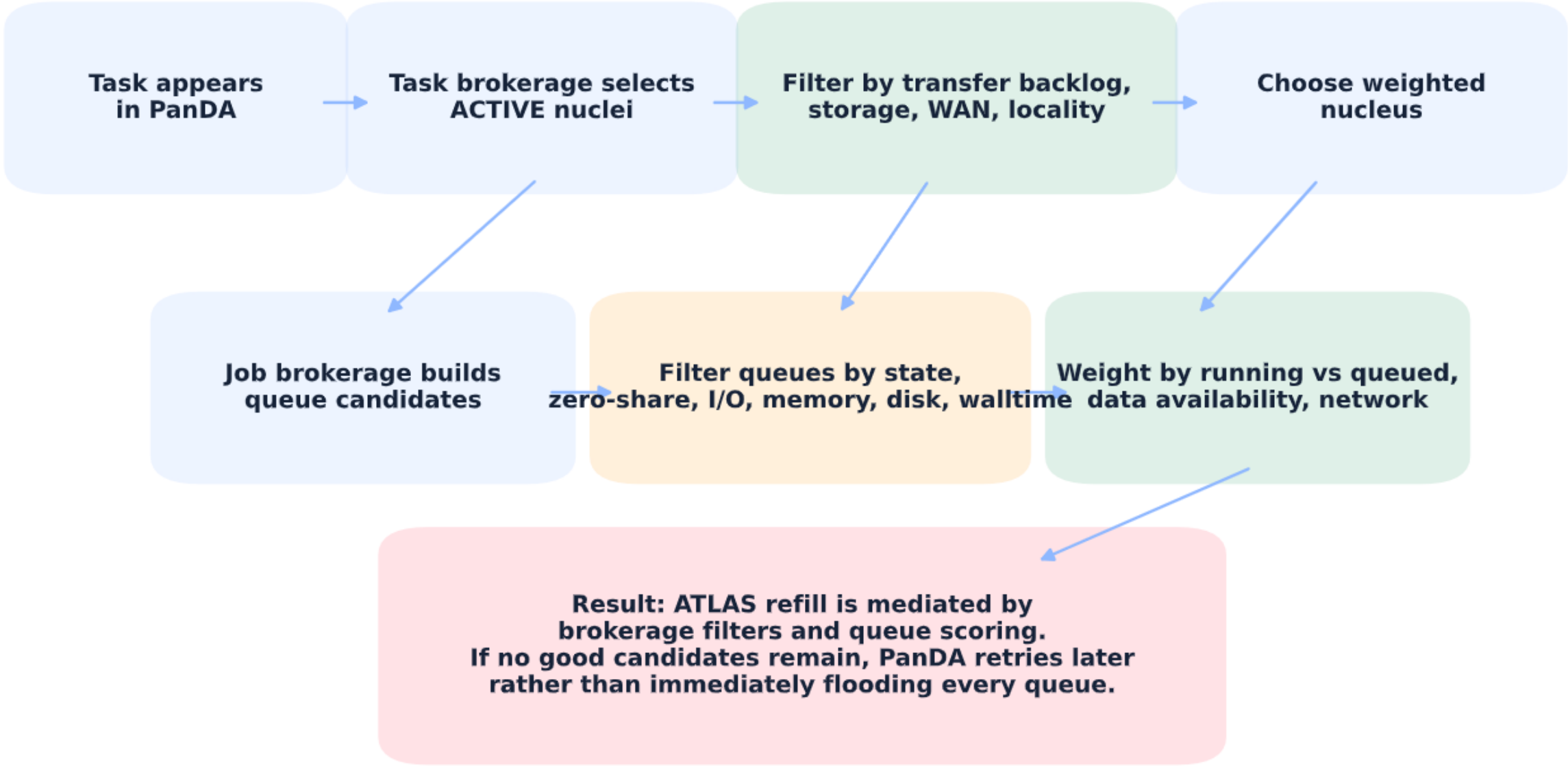
- CMS keeps deep queue pressure via glideinWMS frontend plus WMAgent, Tier-0, and CRAB.
- `idle_glideins_per_entry` and related limits keep pilot demand close to the ceiling.
- When slots reappear, CMS can usually present runnable pilots immediately.

ATLAS PanDA brokerage filters nuclei and queues before assigning work. behaves like filtered, weighted brokerage.

LHCb DIRAC submits pilots with explicit bounds from waiting jobs, waiting pilots, and CE slots. (Behaves like controlled replenishment.)

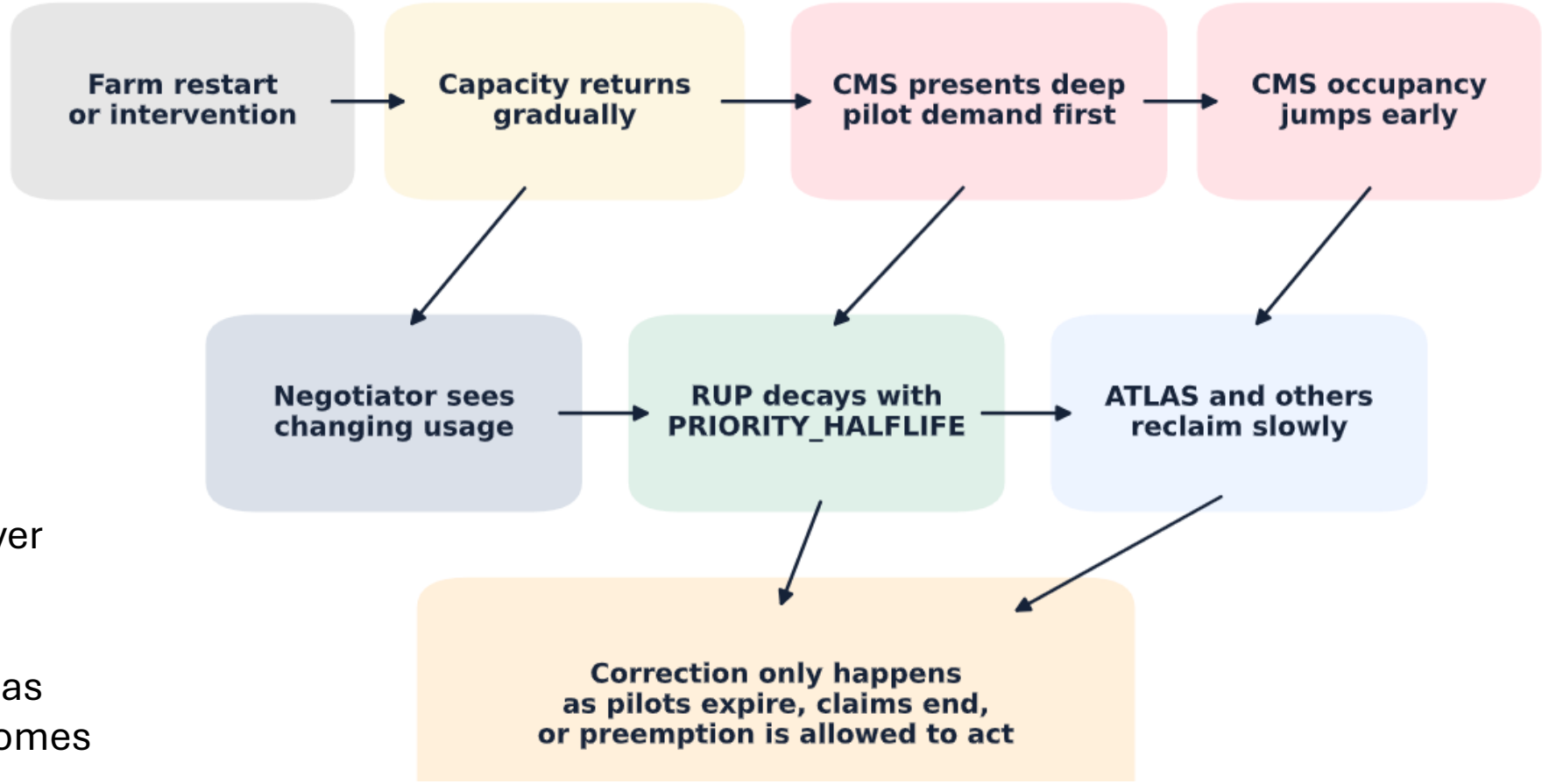


ATLAS PanDA Brokerage: why refill can be more filtered



What Happens after disruptive event

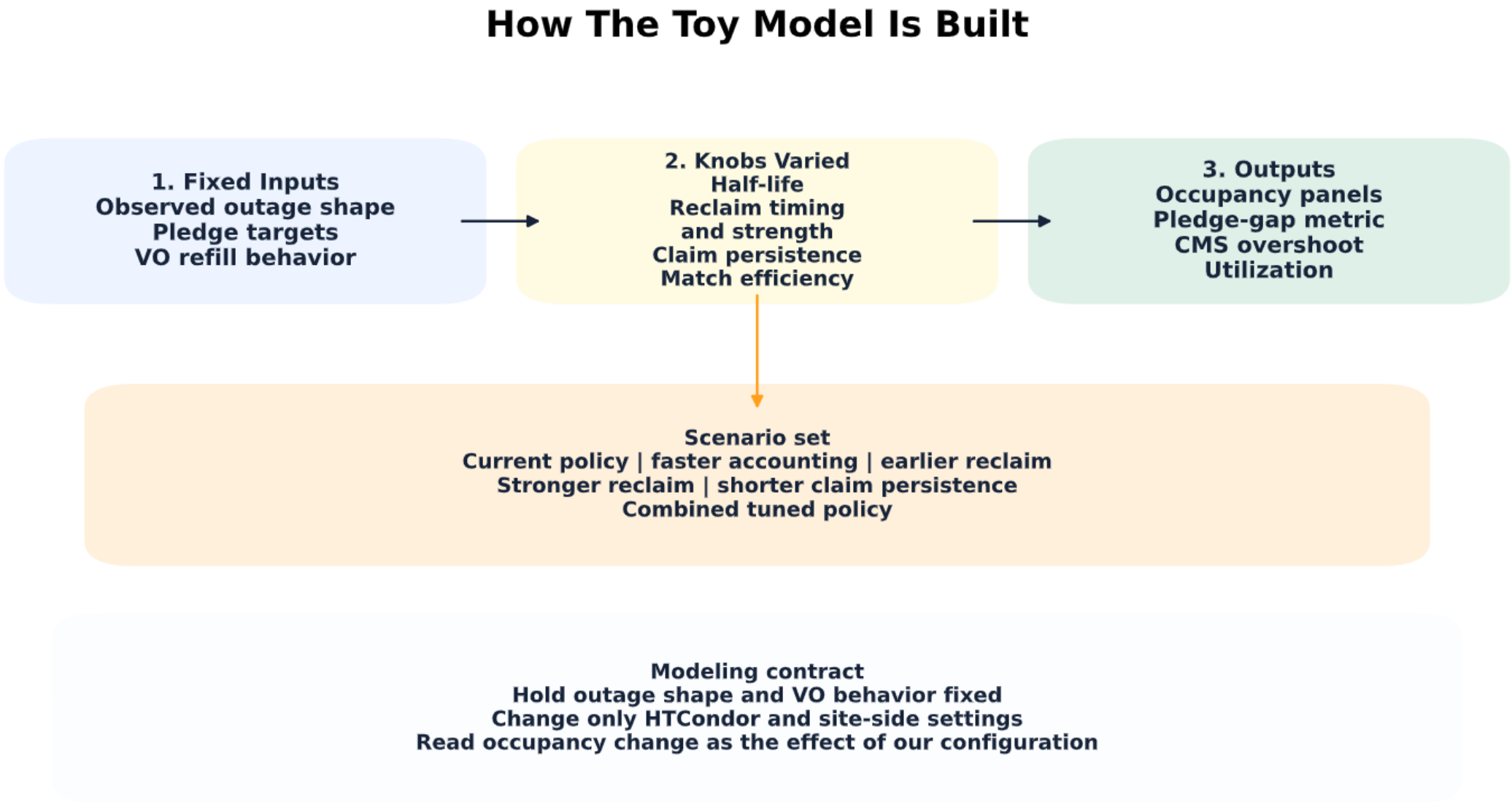
Why nominal fair-share and observed occupancy can diverge for several days



Capacity itself returns gradually.

- CMS fills recovered slots early.
- HTCondor notices unfairness over later cycles.
- ATLAS and smaller VOs recover as pilots expire or preemption becomes effective.

How The Toy Model Is Built



Purpose: not to fit production exactly, but to expose which HTCondor and site-side settings change recovery when VO behavior is held fixed.

The Observed Transient In A Toy Model

The same farm-recovery event is shown under different HTCondor or site-side configurations.

- The panels are: current policy, faster accounting, earlier reclaim, stronger reclaim, shorter claim persistence, and the combined tuned policy.
- Filled colors are actual occupancy; dashed colored boundaries are the pledge-based target stack.
- Those target bands are based on the converted core shares: ATLAS 30.5%, ALICE 2.8%, CMS 8.1%, LHCb 27.0%, Others 31.6%.

- The goal is qualitative: show which pathology delays recovery and which combination of controls helps.

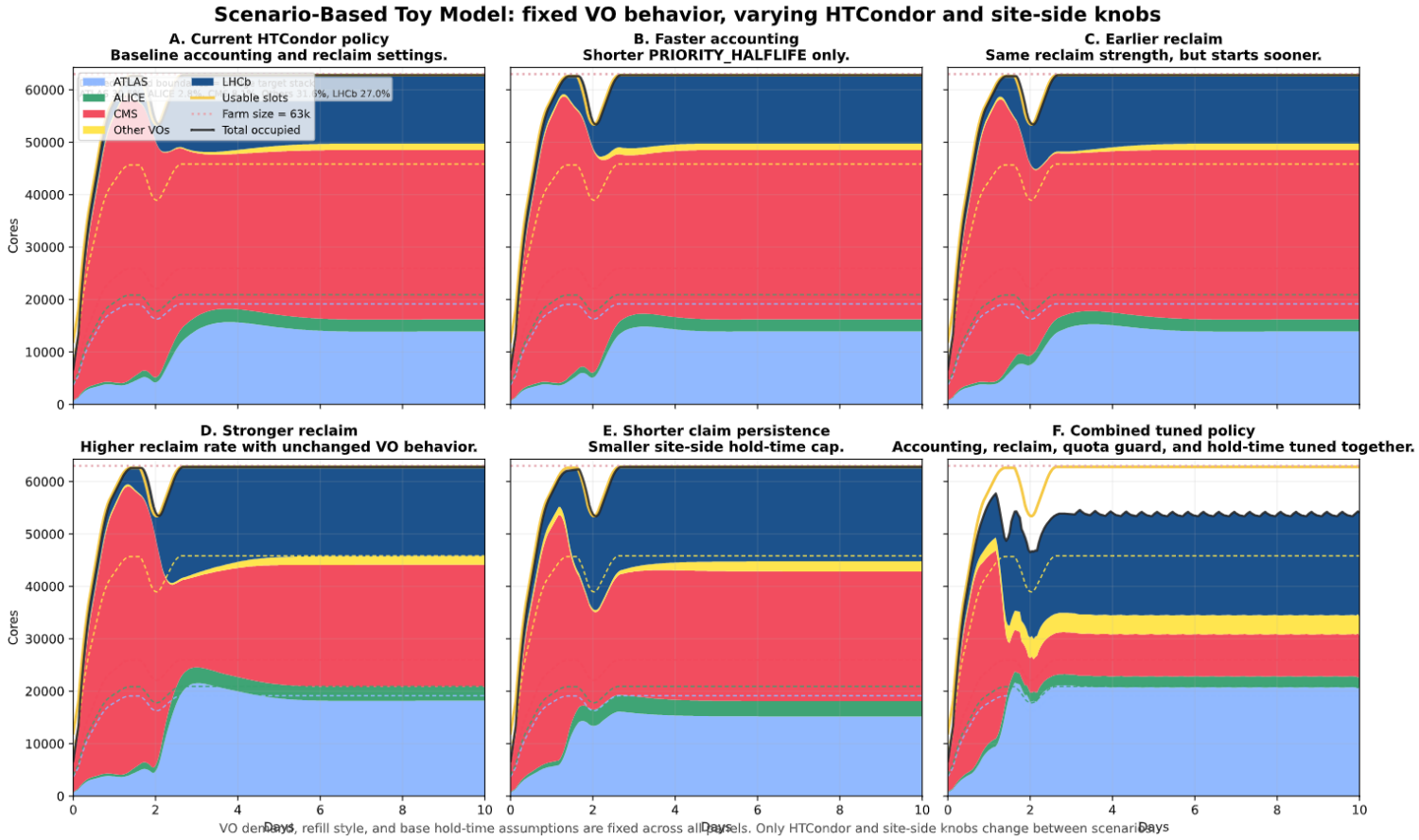
Lever family	Main HTCondor or site knobs	Visible effect in plots
Quota target	GROUP_QUOTA_* GROUP_ACCEPT_SURPLUS*	Sets the target share that recovery returns to
Accounting memory	PRIORITY_HALFLIFE GROUP_PRIO_FACTOR_*	Changes how fast HTCondor notices unfairness
Reclaim timing	NEGOTIATOR_CONSIDER_EARLY_PREEMPTIO PREEMPTION_REQUIREMENTS	Starts correction earlier or later
Reclaim strength	PREEMPTION_RANK MAXJOBRETIREMENTTIME	Changes how fast occupied slots can move back
Claim persistence	CLAIM_WORKLIFE WANT_VACATE	Long hold time keeps a dominant VO high for longer
Match efficiency	NEGOTIATE_ALL_JOBS_IN_CLUSTER NEGOTIATOR_INTERVAL NEGOTIATOR_PRE_JOB_RANK	Helps under-target work refill freed slots fast enough

The Observed Transient In A Toy Model

The same farm-recovery event is shown under different HTCondor or site-side configurations.

- The panels are: current policy, faster accounting, earlier reclaim, stronger reclaim, shorter claim persistence, and the combined tuned policy.
- Filled colors are actual occupancy; dashed colored boundaries are the pledge-based target stack.
- Those target bands are based on the converted core shares: ATLAS 30.5%, ALICE 2.8%, CMS 8.1%, LHCb 27.0%, Others 31.6%.

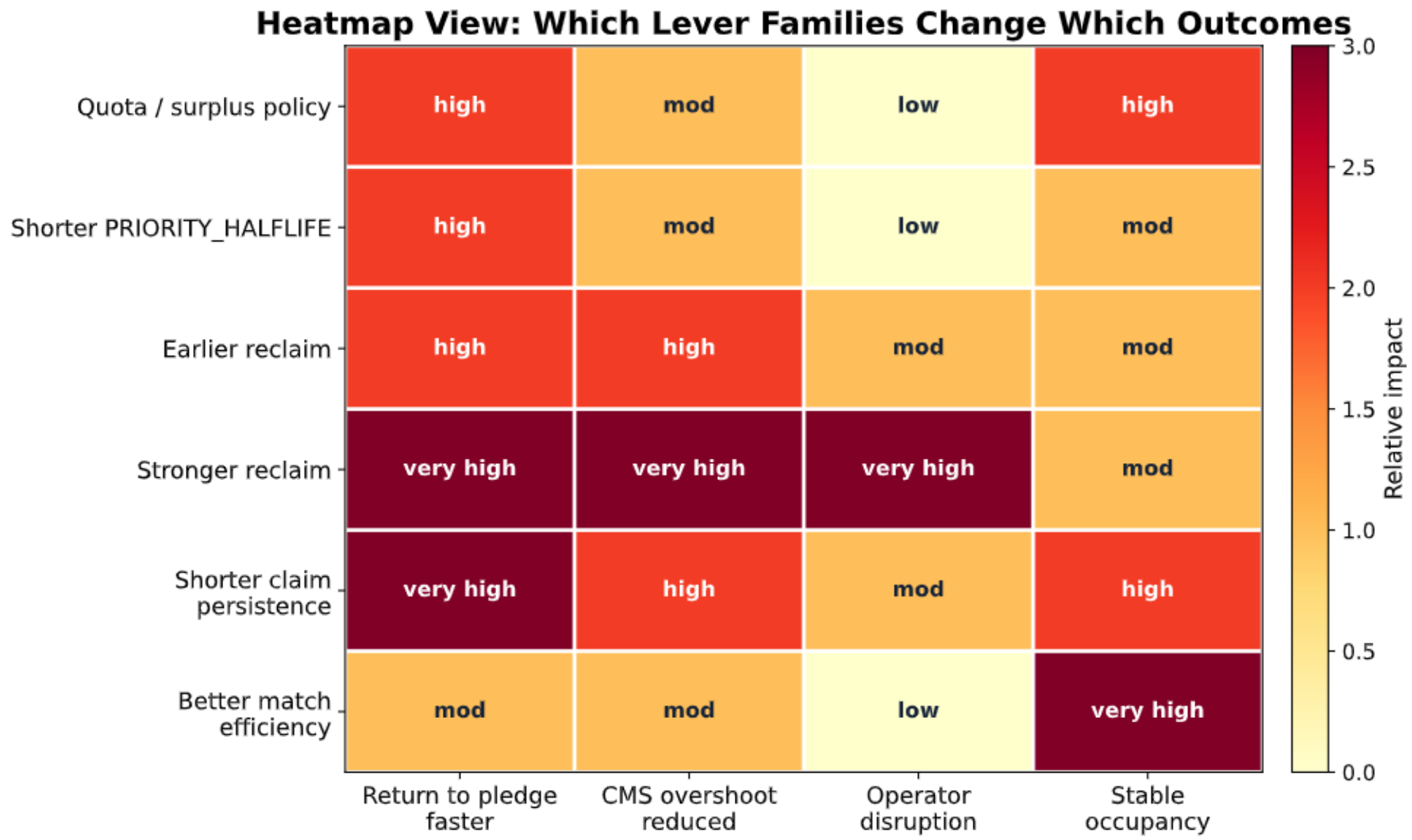
- The goal is qualitative: show which pathology delays recovery and which combination of controls helps.



Which Knobs Are Tuned In The Scenario Suit

- Pledge target row in the landscape figure maps to quota structure: `GROUP_QUOTA_*`, `GROUP_QUOTA_DYNAMIC_*`, `GROUP_ACCEPT_SURPLUS*`, `GROUP_AUTOREGROUP*`.
- Accounting-memory row maps to `PRIORITY_HALFLIFE`, `DEFAULT_PRIO_FACTOR`, and `GROUP_PRIO_FACTOR_*`.
- Reclaim-timing row maps mainly to `NEGOTIATOR_CONSIDER_EARLY_PREEMPTION` and the policy logic around when over-target correction is allowed to start.
- Reclaim-strength row maps mainly to `PREEMPTION_REQUIREMENTS`, `PREEMPTION_RANK`, `MAXJOBRETIREMENTTIME`, and `MachineMaxVacateTime`.
- Claim-persistence row maps to `CLAIM_WORKLIFE` and `WANT_VACATE`.
- Match-efficiency row maps to `NEGOTIATE_ALL_JOBS_IN_CLUSTER`, `NEGOTIATOR_INTERVAL`, and job-selection or packing settings that help under-target work refill freed slots.
- In the toy model those appear respectively as fair-share target policy, half-life, reclaim start time, reclaim strength, claim persistence, and under-target refill efficiency.
- CMS stays high in the baseline because the early VO refill race is fixed while the HTCondor correction layer is too weak.
- The tuned case reduces that persistence without changing VO behavior: it combines faster accounting, earlier reclaim, stronger reclaim, shorter claim persistence, and better refill of under-target work so the farm stays near full while CMS moves toward pledge.

Heatmap View of The Same Knobs



Practical Tier-1 Takeaway

- Fair-share targets alone do not determine post-restart behavior.
- Recovery time is jointly set by HTCondor policy, pilot lifetime, and VO submission semantics.
- The operational tuning order is: quotas, decay, reclaim policy, then efficiency tuning.

Dracula's country

CHEP 2027, Bucharest

13-17 September 2027

Thank you