Compute Testing and ARM Provision at Glasgow's Tier2

Dr Dwayne Spiteri, Dr Emanuele Simili

Prof David Britton
Dr Gordon Stewart
Dr Samuel Skipsey
Dr Bruno Borbely

HEPiX Spring Workshop - 16/04/2024

Introduction

- Who we are as a site

- Tested Equipment
    - Equipment in production
    - In-house test boxes
    - Equipment Tested Remotely

- Our latest benchmark results
    - Brief discussion on power metrics

- Upscaling benchmark tests - ARM Farm
    - Ease of setup
    - Results from LHC Experimental tests

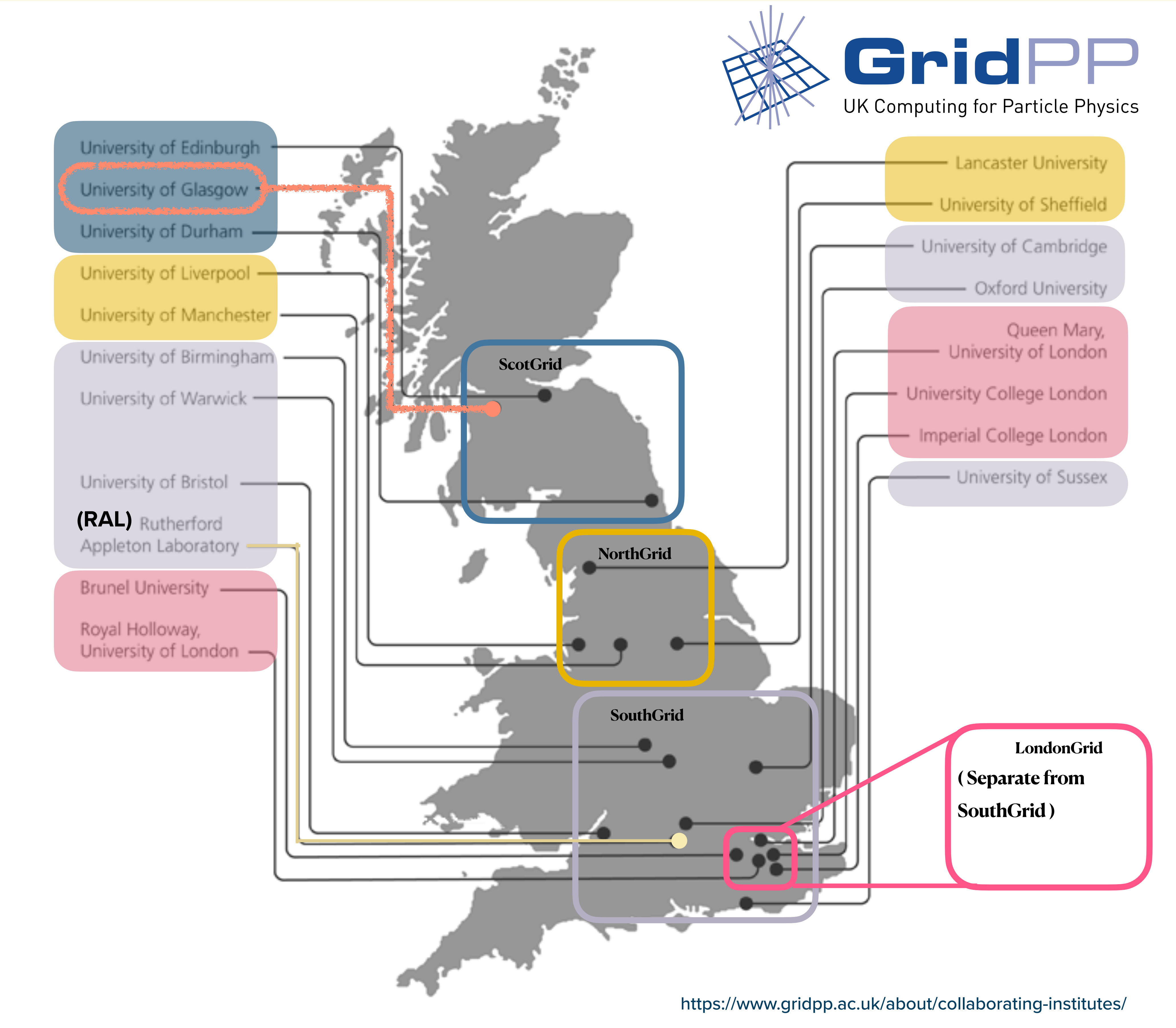
GridPP and Glasgow

- GridPP is the UK's branch of the WLCG. Its 18 Tier-2 sites and 1 Tier-1 (RAL) are split into 4 groups (ScotGrid, NorthGrid, SouthGrid and LondonGrid).

- Glasgow is the largest Tier-2 site in ScotGrid.

Glasgow Data-centre Highlights

- **Compute**
  ~15.5k x86 cores
  ~2k ARM cores
- **Storage**
  12PB raw storage
- **Networking**
  10 GbE to worker nodes
  10/40 GbE to storage
  40 GbE to the main campus
  2x10GbE to Janet (UK's inter-university system)



Glasgow Production Compute

- **Compute**
  ~15.5k x86 cores
  ~2k ARM cores

2xIntel40ht: Dual Socket Intel XEON 10-Core CPU E5-2630 v4 (HP)

CPU: 2 * Intel(R) Xeon(R) 10-core CPU E5-2630 v4 @ 2.2GHz (TDP 85W)
RAM: 160GB (4*32GiB + 4*8GiB) DDR4 2400 MHz → 4 GB/core
HDD: 2TB disk SATA 6Gb/s @ 7200 RPM (SEAGATE)
OS: CentOS 7.9 →

2xAMD64ht: Dual Socket AMD EPYC 7513 32-Core Processor (DELL)

CPU: 2 * x86 AMD EPYC 7513, 32C/64HT @ 2.6GHz (TDP 200W)
RAM: 512GB (16 x 32GB) DDR4 3200MT/s → 4 GB/core
HDD: 3.84TB SSD SATA Read Intensive
OS: CentOS 7.9 → Alma 9

2xAMD64ht: Dual Socket AMD EPYC 7452 32-Core Processor (DELL)

CPU: 2 * x86 AMD EPYC 7452, 32C/64HT @ 2.32GHz (TDP 200W)
RAM: 512GB (16 x 32GB) DDR4 3200MT/s → 4 GB/core
HDD: 3.84TB SSD SATA Read Intensive
OS: CentOS 7.9 → Alma 9

2*ARM80c: Dual Socket Ampere Altra Q80-30 80-Core Processor (Ampere)

CPU: 2 * ARM Q80-30 80C @ 3GHz (TDP 210W)
RAM: 512GB (32 x 16GB or 16 x 32GB) DDR4 3200MT/s → 3.2 GB/core
HDD: 2 * 1Tb NVMe (INTEL + SAMSUNG)
OS: Rocky 9.2

Generous Donation from AMPERE

Glasgow in-House Tested Compute

AMD96ht: Single AMD EPYC 7643 48-Core Processor (GIGABYTE)

CPU: x86 AMD EPYC 7643 48C/96HT @ 2.3GHz (TDP 225W)
RAM: 256GB (16 x 16GB) DDR4 3200MHz → 2.7 GB/core
HDD: 3.84TB SSD SATA (SAMSUNG)
OS: Alma 9

2xAMD96ht: Single AMD EPYC 7643 48-Core Processor (DELL)

CPU: 2* AMD EPYC 7643 24-Core Processor @ 4GHz @ 2.3GHz (TDP 200W)
GPU: 2* NVidia A100 PCIe 80GB (TDP 300W)
RAM: 256GB (16 x 16GB) DDR4 3200MHz → 2.7 GB/core
HDD: 480GB SSD SATA + 5TB SSD SCSI (DELL)
OS: Rocky 8

ARM80c: Single socket Ampere Altra Q80-30 80-Core Processor (GIGABYTE)

CPU: ARM Q80-30 80C @ 3GHz (TDP 210W)
RAM: 256GB (16 x 16GB) DDR4 3200MHz → 3.2 GB/core
HDD: 3.84TB SSD SATA (SAMSUNG)
OS: Alma 9

Grace144c: Dual Socket NVidia Grace 144-Core Processor (SuperMicro)

CPU: NVidia Grace 144-Core 480GB DDR5 @ 3.4GHz (TDP 500W)
RAM: 480GB (on chip) DDR5 4237MHz → 3.3 GB/core
HDD: 1TB NVMe + 4TB NVMe (SAMSUNG)
OS: Alma 9

Glasgow Remotely Tested Compute

2*AMD256ht: Dual Socket AMD EPYC 9754 128-Core Processor (SuperMicro)

CPU: 2 * x86 AMD EPYC 9754, 128C/256HT @ 3.1GHz (TDP 360W)
RAM: 1.536TB (24 x 64GB) DDR4 3200MHz → 3 GB/core
HDD: 512GB NVMe + 3.84TB SSD
OS: Rocky 9.2

Bergamo

AMD128c: Single Socket AMD EPYC 8534P 64-Core Processor (SuperMicro)

CPU: AMD EPYC 8534P @ 3.1GHz (TDP 200W)
RAM: 576GB (6 x 96GB) DDR5 3200MT/s → 4.5 GB/core
HDD: 1Tb NVMe Storage
OS: Rocky 8.8 Green Obsidian

Siena

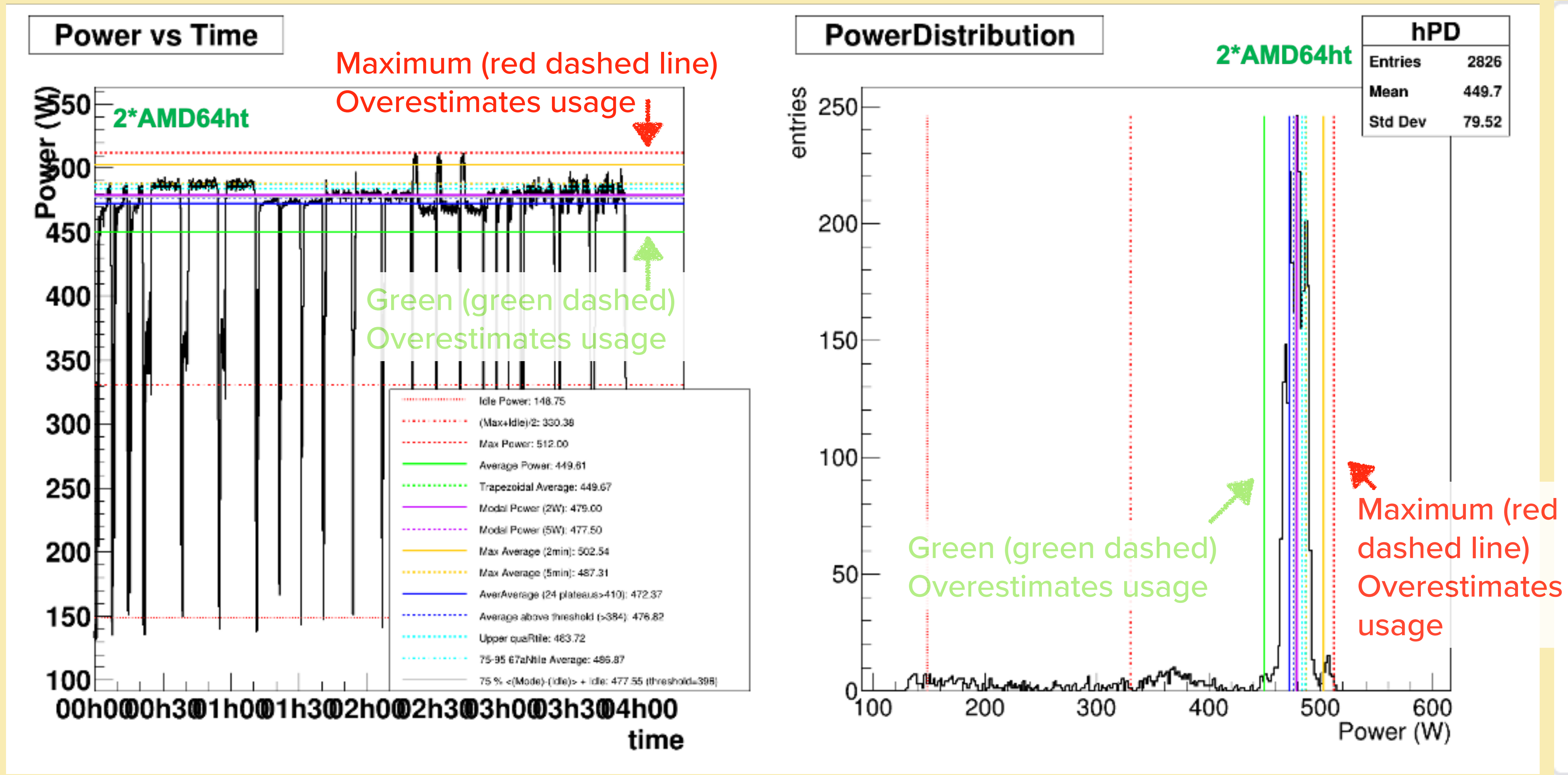
ARM128c: Single Socket Ampere Altra Max M128-28 128-Core Processor (XMA)

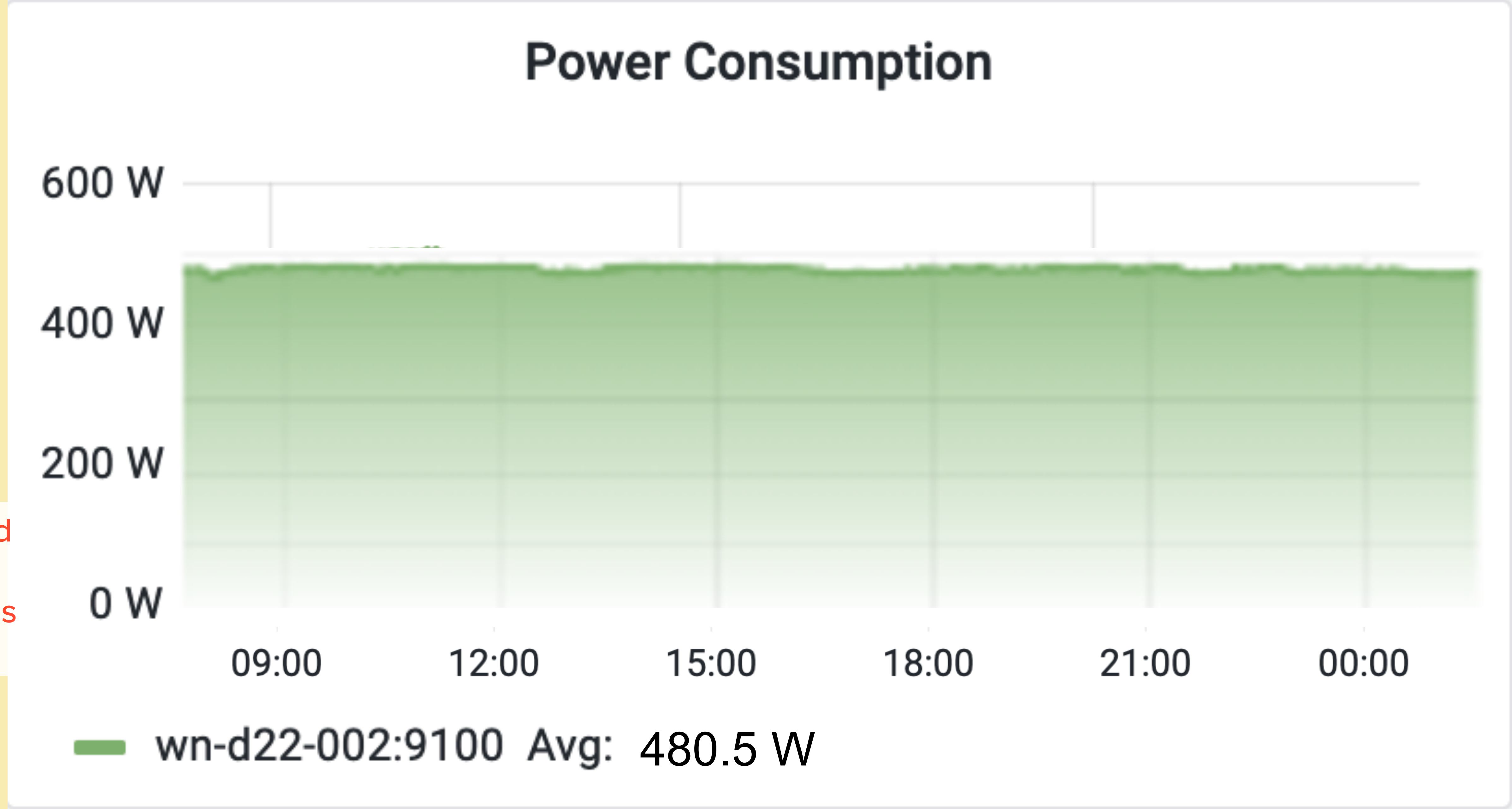
CPU: ARM M128-28 @ 2.8GHz (TDP 250W)
RAM: 512GB (64 x 8GB) DDR4 3200MHz → 4 GB/core
HDD: 1Tb NVMe Storage
OS: Rocky 8.8 Green Obsidian

- More on the way. Vendors have been notified of our interest in; AmpereOne, Blackwell, Grace Hopper (CPU + GPU),

What do we do with these machines

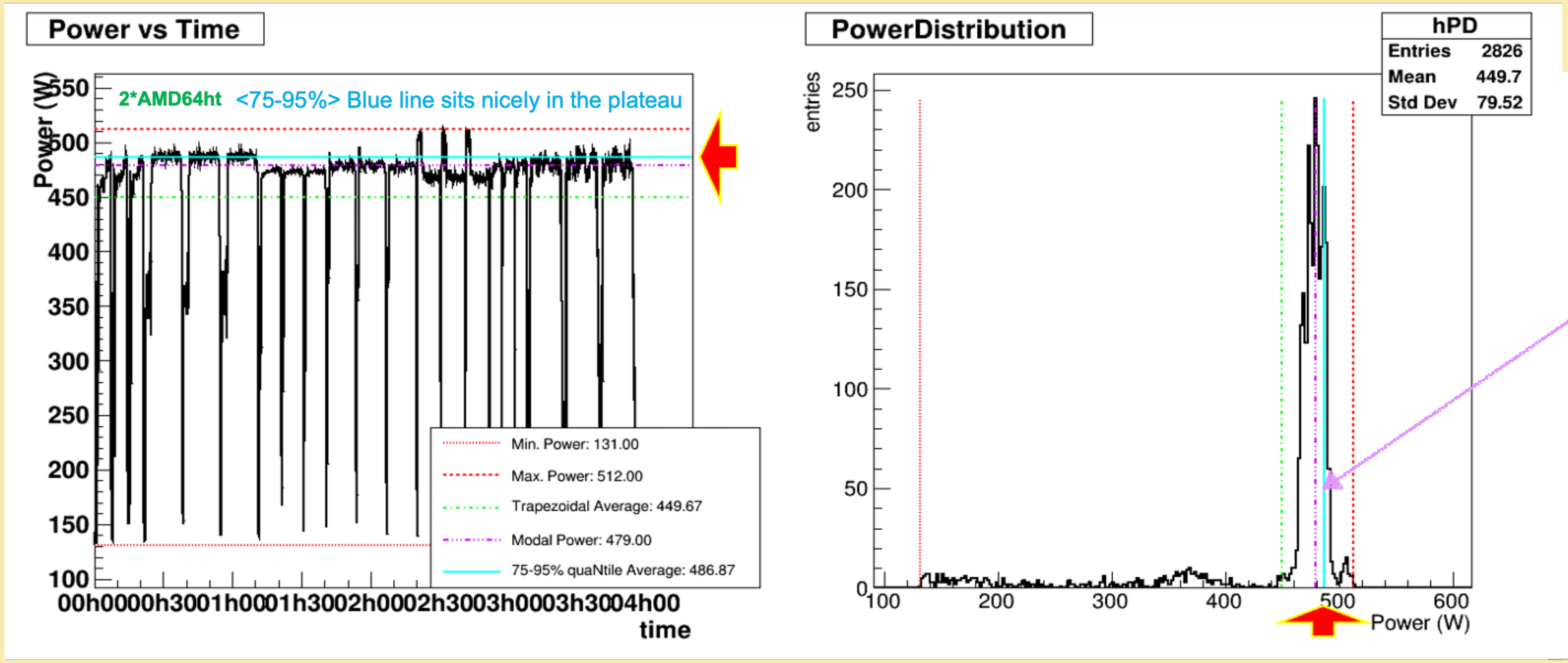
- The nodes we test are benchmarked using the **HEP-Score suite** (i.e., 3 x 7 workloads of *10-30 min.* each ~3hr including start-up phase). We created an add-on to record power used during the benchmarking.
- Typically the power profile of these runs (left) don't represent the standard running conditions at grid sites (right) where typical jobs last longer and multiple jobs run simultaneously, keeping the CPU at ~100%.
- The current community standards for power reporting; Maximum and Average power, over and underestimate usage of benchmark runs respectively.

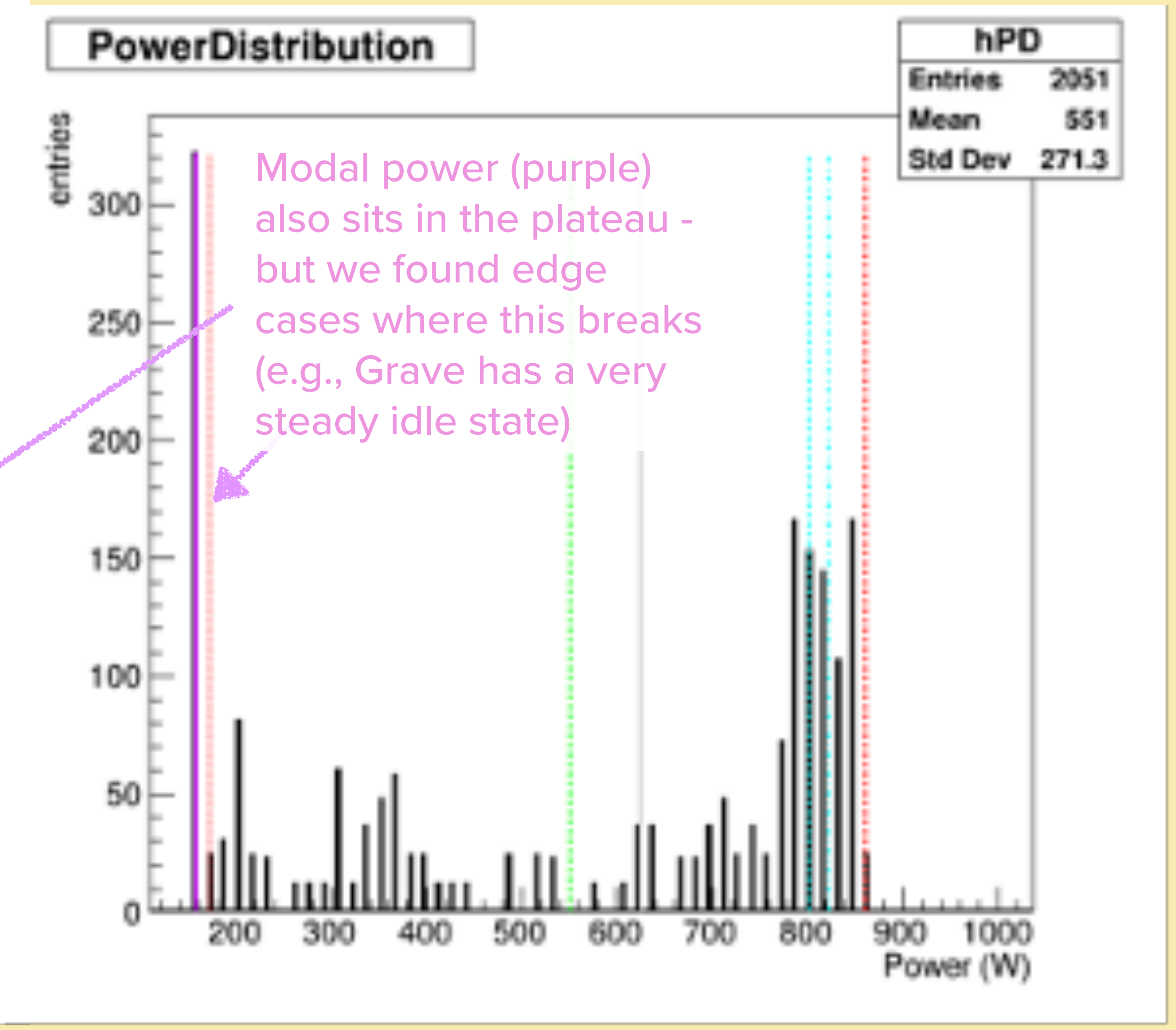




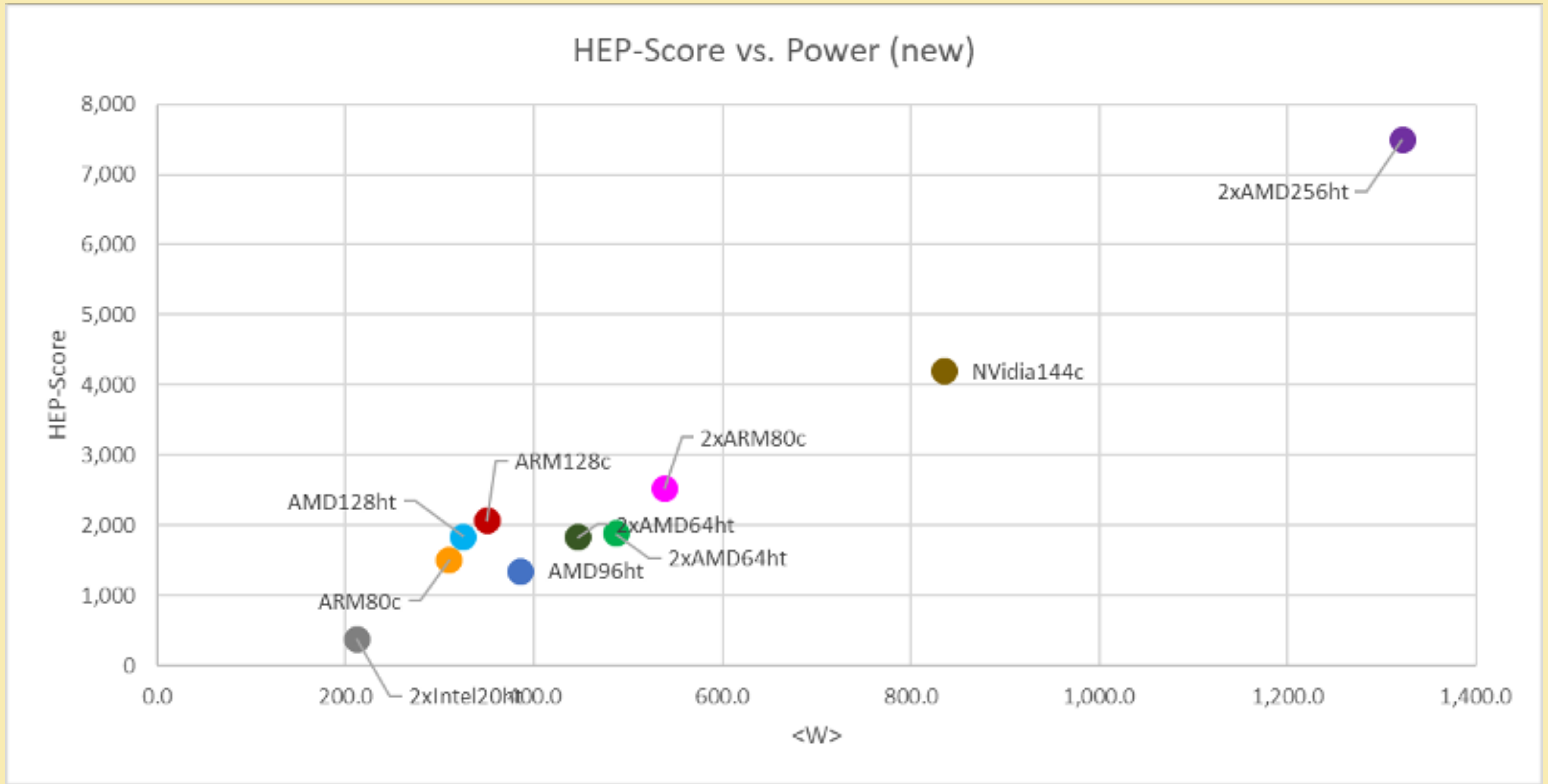
Watt are we talking about

- Want a new figure of merit (FoM) for power reporting that lies in the plateau of the runs. Don’t want to be swayed by short periods in idle (Average), or odd spikes above the plateau (Maximum).
- FoM should be easy to understand and implement. Arrange data in power order and perform an upper quaRtile average but remove the top 5% of data (75-95% quaNtile average <75-95%>).

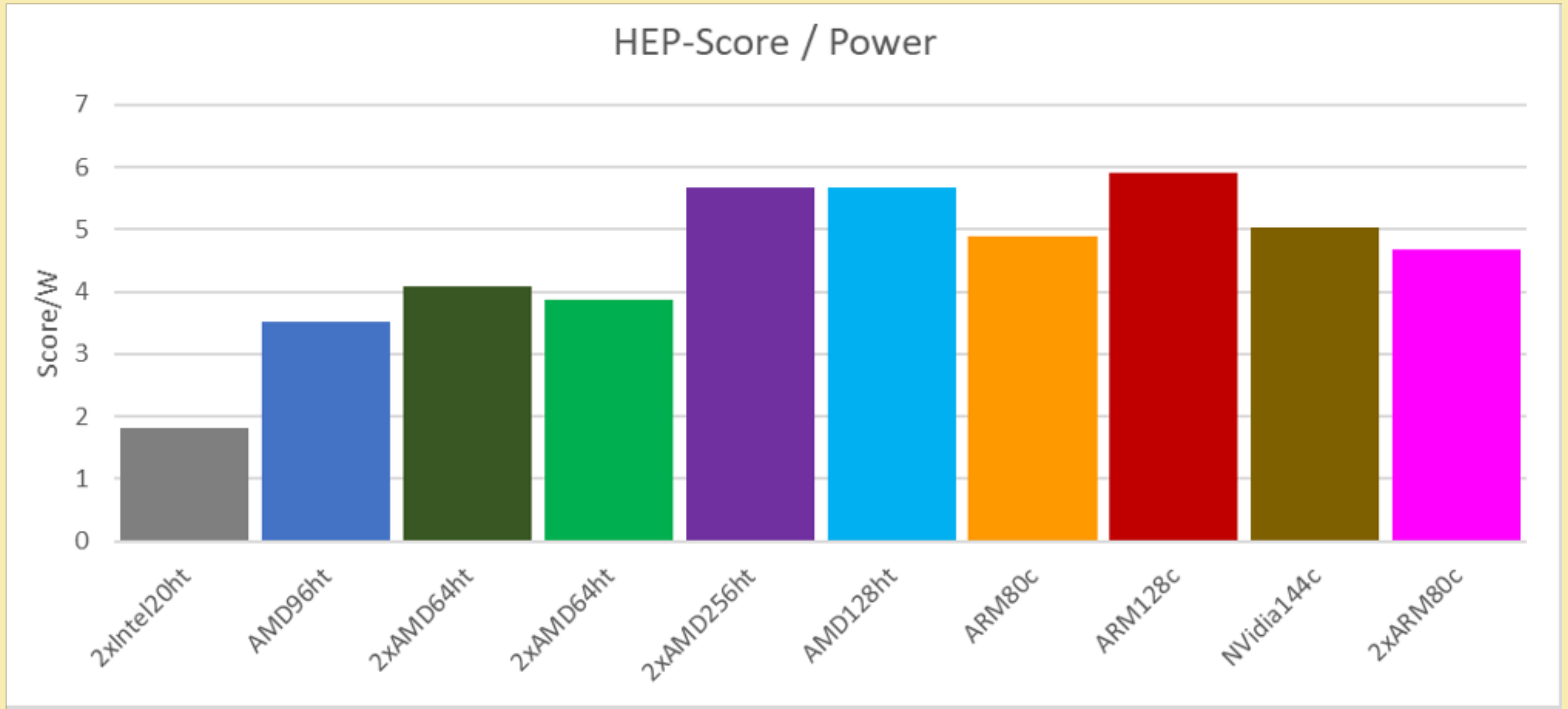




HEPScore/Watt

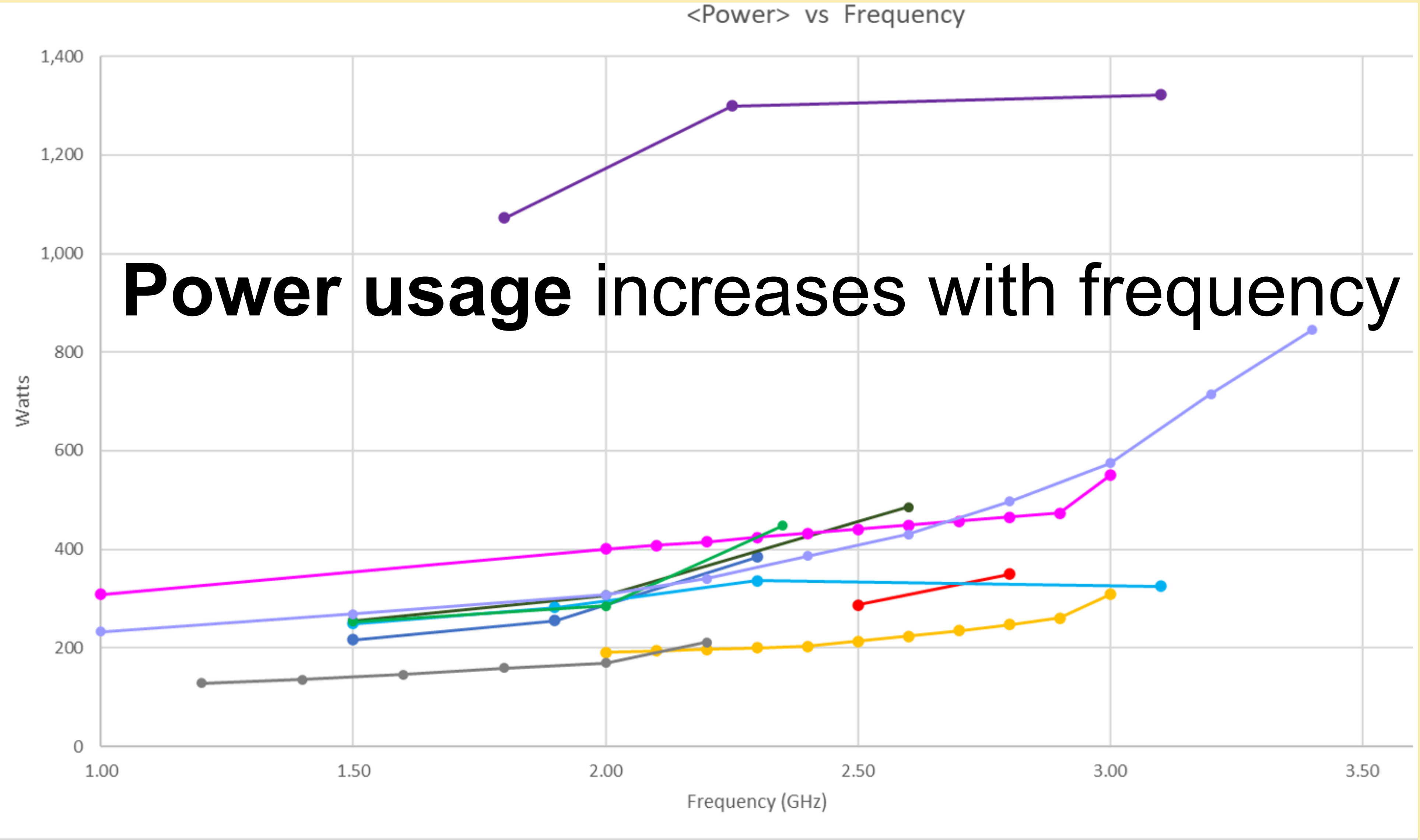


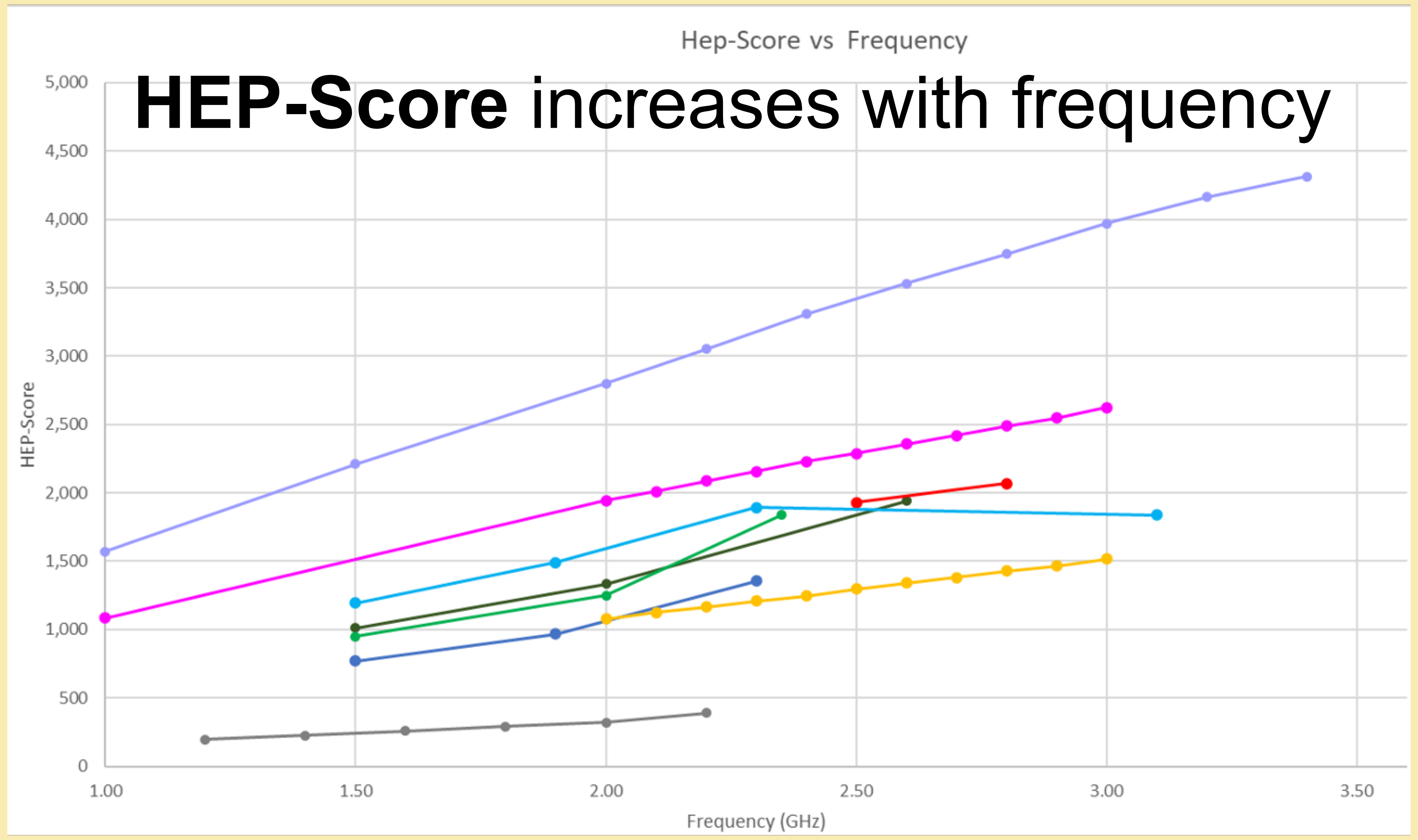
- Using <75-95%> as our new power FoM, we retested our machines and replotted data.
- The rankings of the machines in terms of HEPScore/Power did not change.



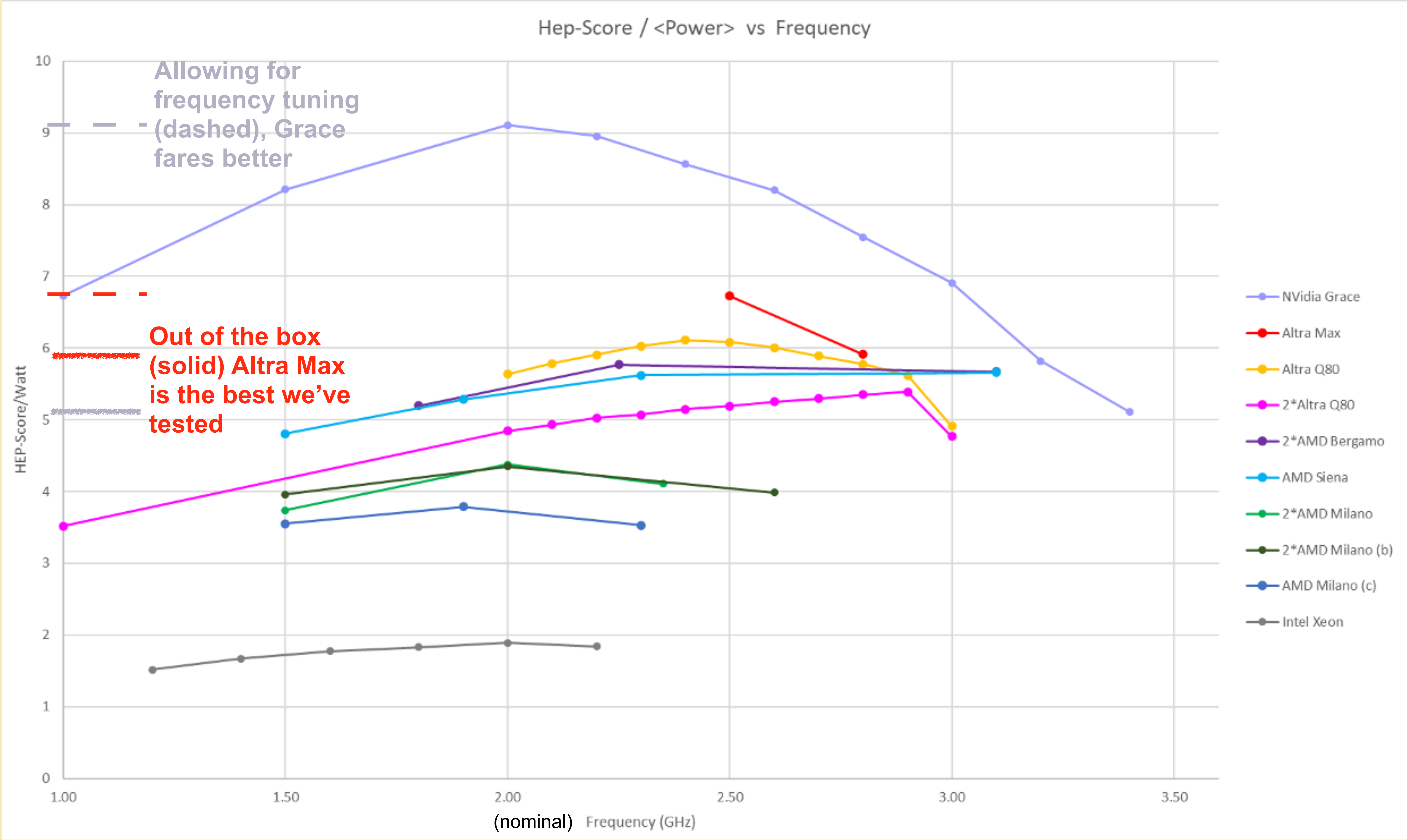
- For example Grace box (brown) less efficient than ARM128c (red)
- **Not the whole story though**

Frequency Scaling

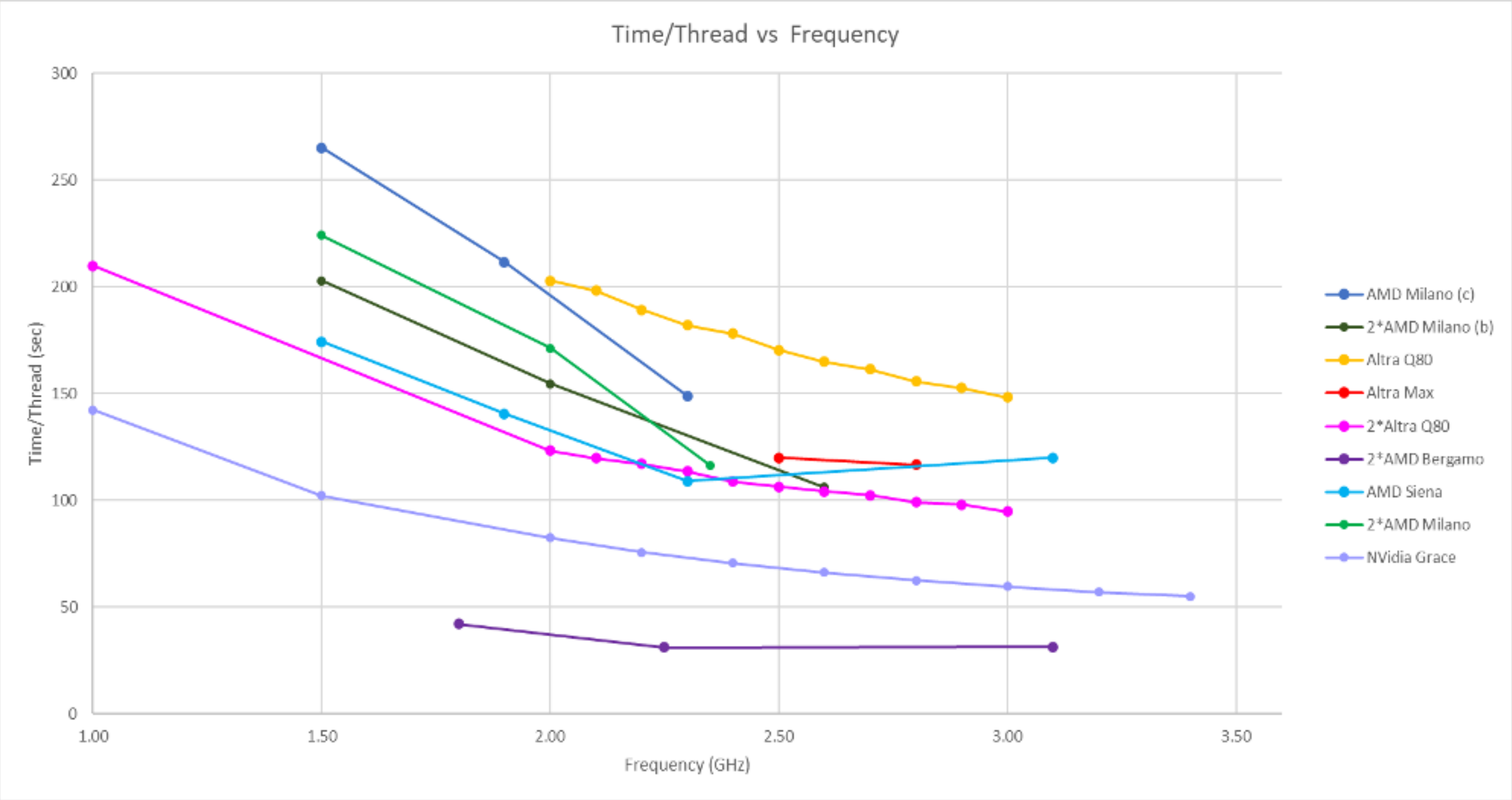




- **HEP-Score/Watt** vs. **CPU Frequency** gives a better picture of hardware potentials, and also shows optimal performances per watt at mid range.
- No machine tested was maximally performant at their highest nominal frequency.

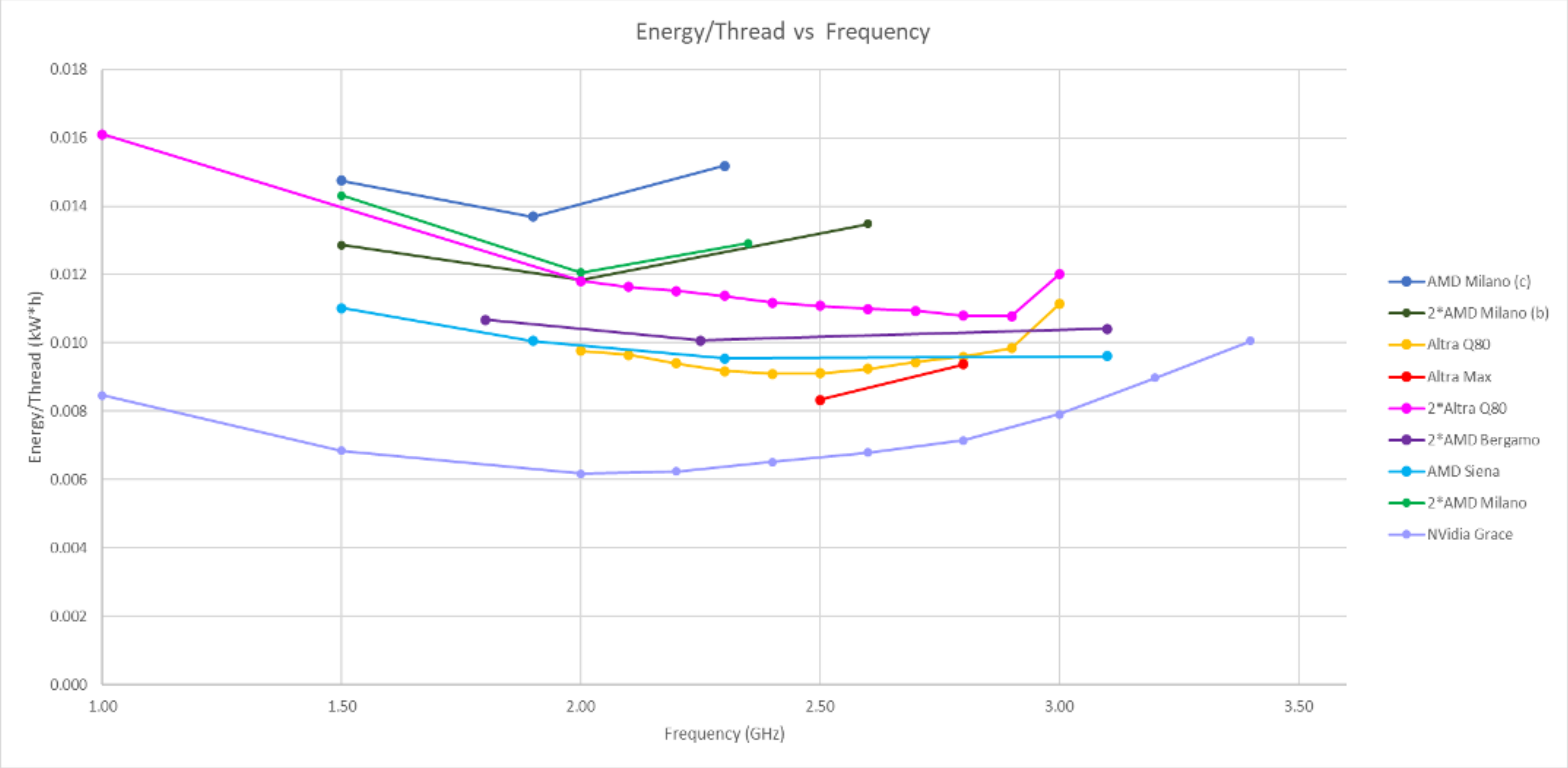


Frequency Scaling (2)



- Execution time decreases almost linearly with frequency

- While the total energy per job has a minimum below max. frequency (on all hardware)



- What is clear is that tuning the frequency down a step can save quite some energy at the cost of an increase in time, but by how much time and the energy you save greatly depend on the hardware!

- In the end, it is up to each sites to find a compromise …

The ARM Farm

- From these tests, ARM compute have shown indications that it could outperform x86 in terms of energy efficiency for HEP-style workloads

- Following our work in this field, Ampere donated some cores which we used to create an 'ARM farm

2*ARM80c: Dual Socket Ampere Altra Q80-30 80-Core Processor (Ampere)

CPU: 2 * ARM Q80-30 80C @ 3GHz (TDP 210W)
RAM: 512GB (32 x 16GB or 16 x 32GB) DDR4 3200MT/s → 3.2 GB/core
HDD: 2 * 1Tb NVMe (INTEL + SAMSUNG)
OS: Rocky 9.2

Generous Donation from AMPERE

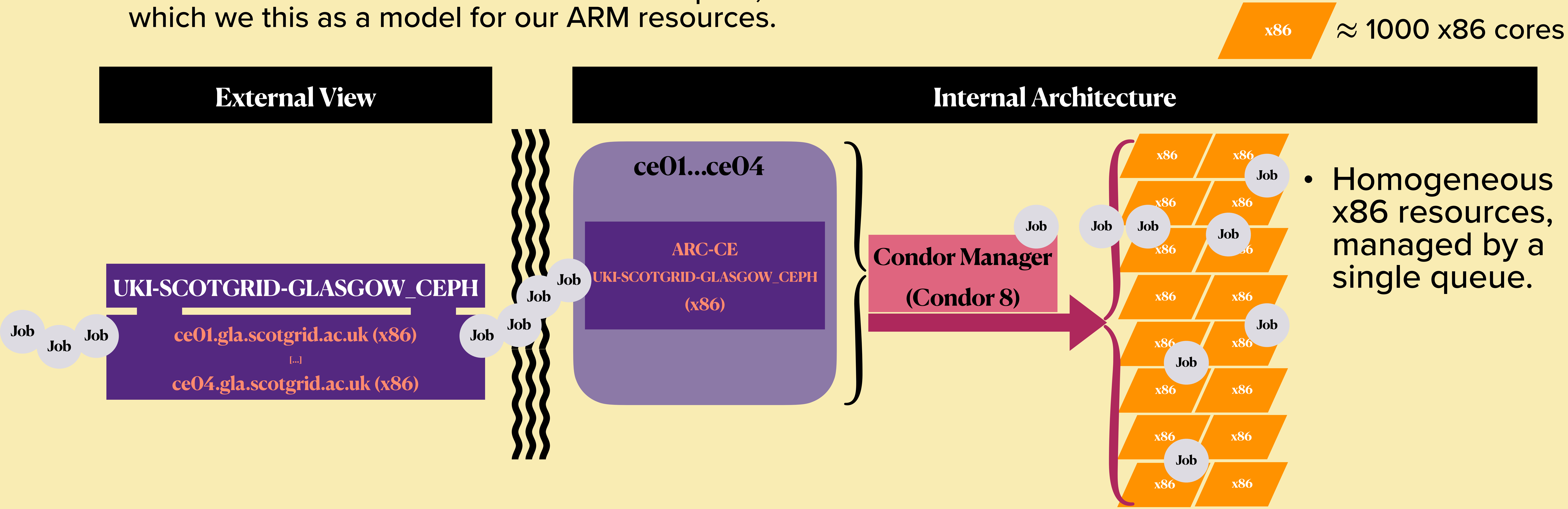
- We wanted to test to see how easy it would be to advertise ARM resource at a Tier-2 grid site

- Initially offered out for testing the to 4 main LHC experiments, with the idea being that once they had run and validated their physics outputs, we could move to fully integrating said ARM resource at our grid site

- Will take you through our journey with ATLAS who were the first to take us up on this offer and then summarise the other experiments findings

How do we get VO's to see and use our ARM resource?

- This is a small visualisation of our x86 compute, which we this as a model for our ARM resources.



- Homogeneous x86 resources, managed by a single queue.

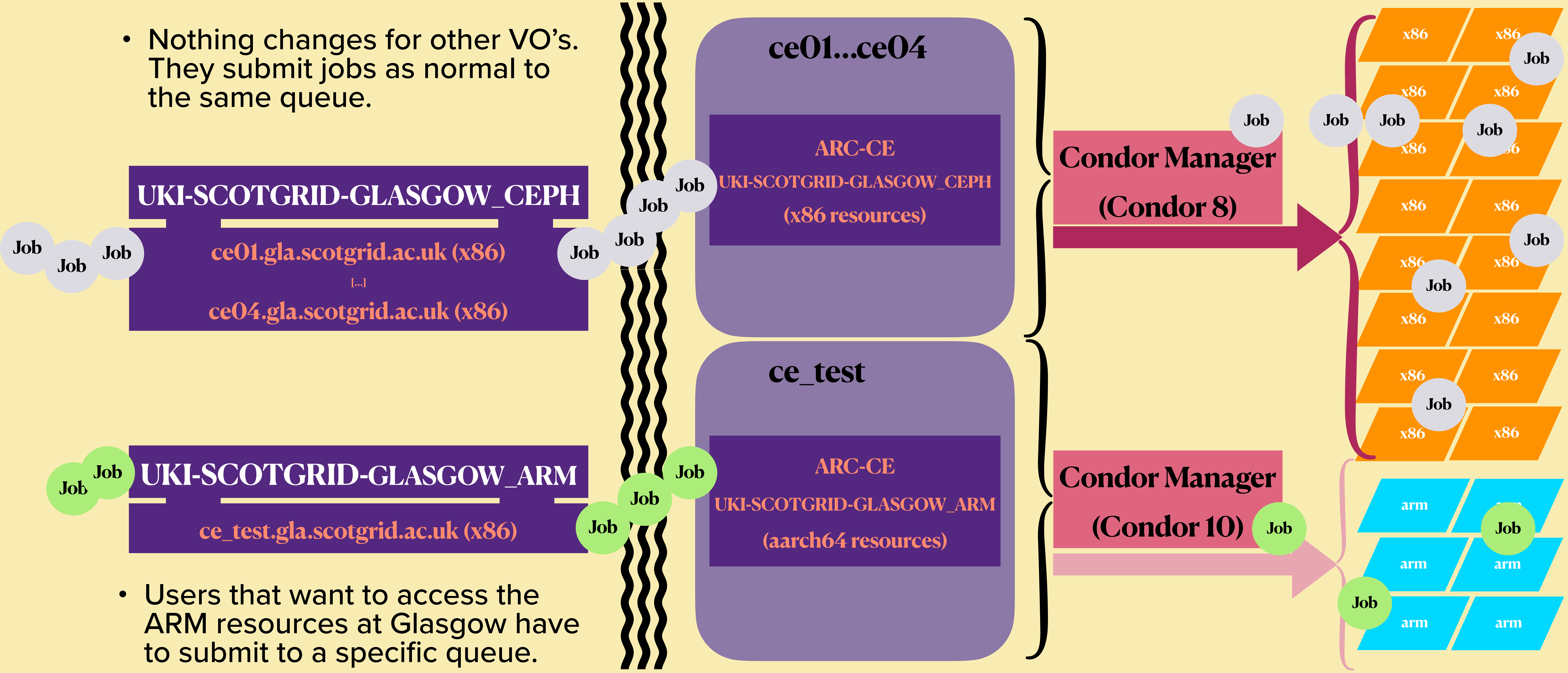
How do we get VO's to see and use our ARM resource?

arm ≈ 2x160 arm cores
x86 ≈ 1000 x86 cores

External View

- Nothing changes for other VO's. They submit jobs as normal to the same queue.

Internal Architecture



- Users that want to access the ARM resources at Glasgow have to submit to a specific queue.

- Want to move to condor 10 on all worker nodes so we will have two types of architecture in one condor pool.
- Have to ensure that jobs run on nodes that have the right architecture!

Technical setup differences between ARM and x86

- Our ARM nodes are on Rocky9 / HTCondor 10, attached to their own CE and Condor manager, while the rest of the site remains at Centos7 / HTC8 (this will change in the near future).

- While there are configuration changes between x86 to ARM, some changes are due the move from EL7 to EL9. By and large the ARM nodes are treated the same as x86 ones.

- For Provisioning, we currently use a bespoke PXE- / kickstart-based system (think Cobbler but targeted a little more closely at our needs) and that works across all our systems.

- Configuration management is via Ansible, and works across all hosts.

- Automatic updates are run across the site, with a couple of exceptions, and this is also true of the ARM nodes - patching normally involves checking applied updates rather than pushing them out. All possible through Ansible (SSH and Python underneath).

Setup: ATLAS-side

PanDA setup stores data locally (now)

Points to our temporary CE and our condor_arm queue in the ARC

aarch architecture only

- Now we just have to run some proper jobs on them

PanDA Queue: UKI-SCOTGRID-GLASGOW_ARM

Object details

ID	930
Name	UKI-SCOTGRID-GLASGOW_ARM
PanDA Site	UKI-SCOTGRID-GLASGOW-CEPH
Description	
Default object	UKI-SCOTGRID-GLASGOW_VIRTUAL
Corecount	8 overwritten
Corepower	8.74 inherited.rcsite
Coreenergy	10.0 inherited.rcsite

State

Object state	ACTIVE
State comment	Object was cloned from UKI-SCOTGRID-GLASGOW_CEPH via WebUI
Last modification date	2023-08-11 13:26:51.509877

Associated DDMEndpoints

Search:

DDMEndpoint	Type	Experiment Site	Activities
UKI-SCOTGRID-GLASGOW-CEPH_DATADISK	DATADISK	UKI-SCOTGRID-GLASGOW	pl/0, read_lan/0, write_lan/0
UKI-SCOTGRID-GLASGOW-CEPH_SCRATCHDISK	SCRATCHDISK	UKI-SCOTGRID-GLASGOW	write_lan_analysis/0, read_lan/1, write_lan/1

Showing 1 to 2 of 2 entries

Associated Queues

Search:

CE	Queue	flavour	version	qstatus	site	cputime	wclock	ETF	Ops
ce-test.gla.scotgrid.ac.uk	condor_arm	ARC-CE	None	production	UKI-SCOTGRID-GLASGOW	0	0	✖	

Showing 1 to 1 of 1 entries

Associated PandaQueue architectures

Type	Architectures	Vendor	Instructions	Model	Ops
cpu	aarch64 excl				

ATLAS Testing - ARM Work

- The ARM nodes ran jobs starting with 50M events and with simulation, reconstruction and derivation tasks

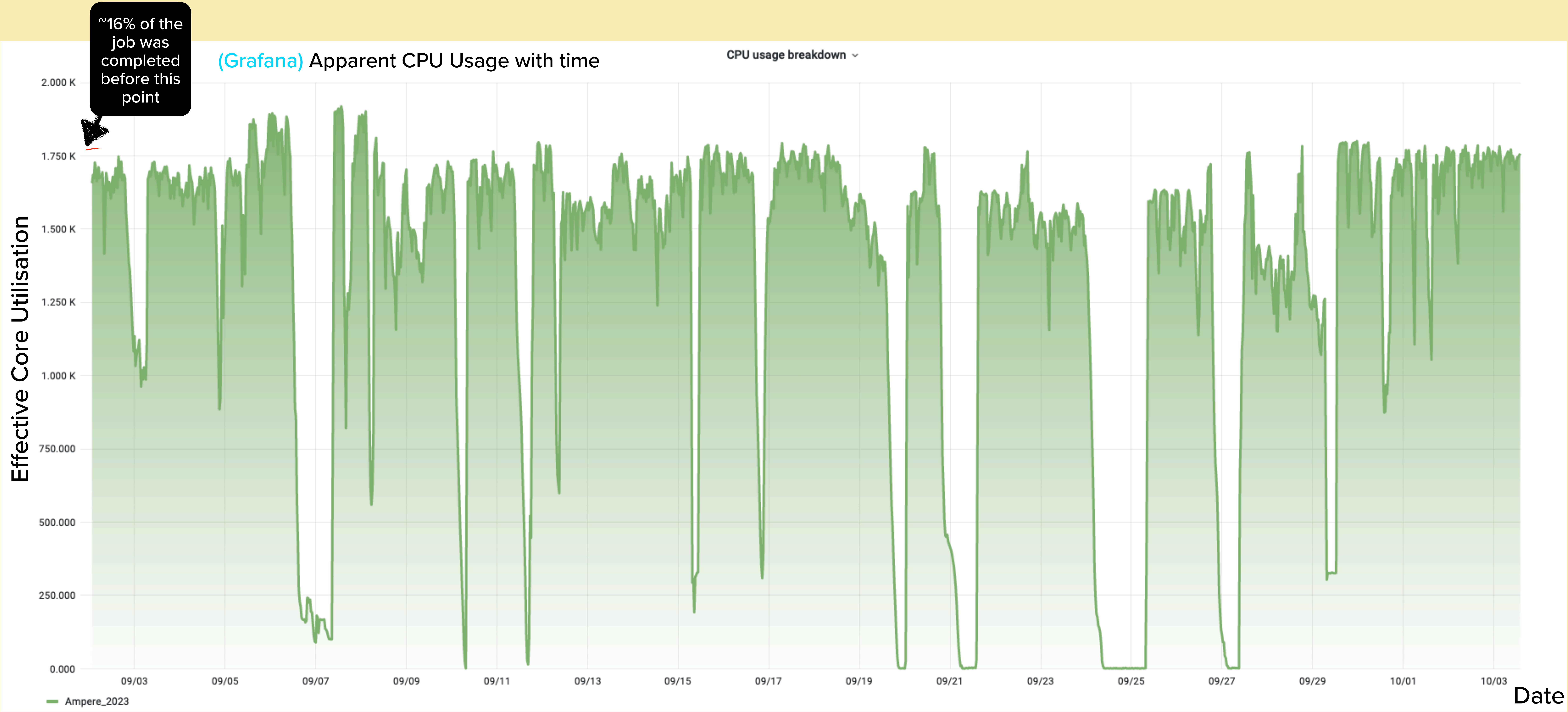
- Meant to run for about ~21 days, ended up taking 58 days. (Mid August to Early October

Estimated CO_2 emission in grams, based on the weighted average of regional electricity sources for ATLAS Grid computing

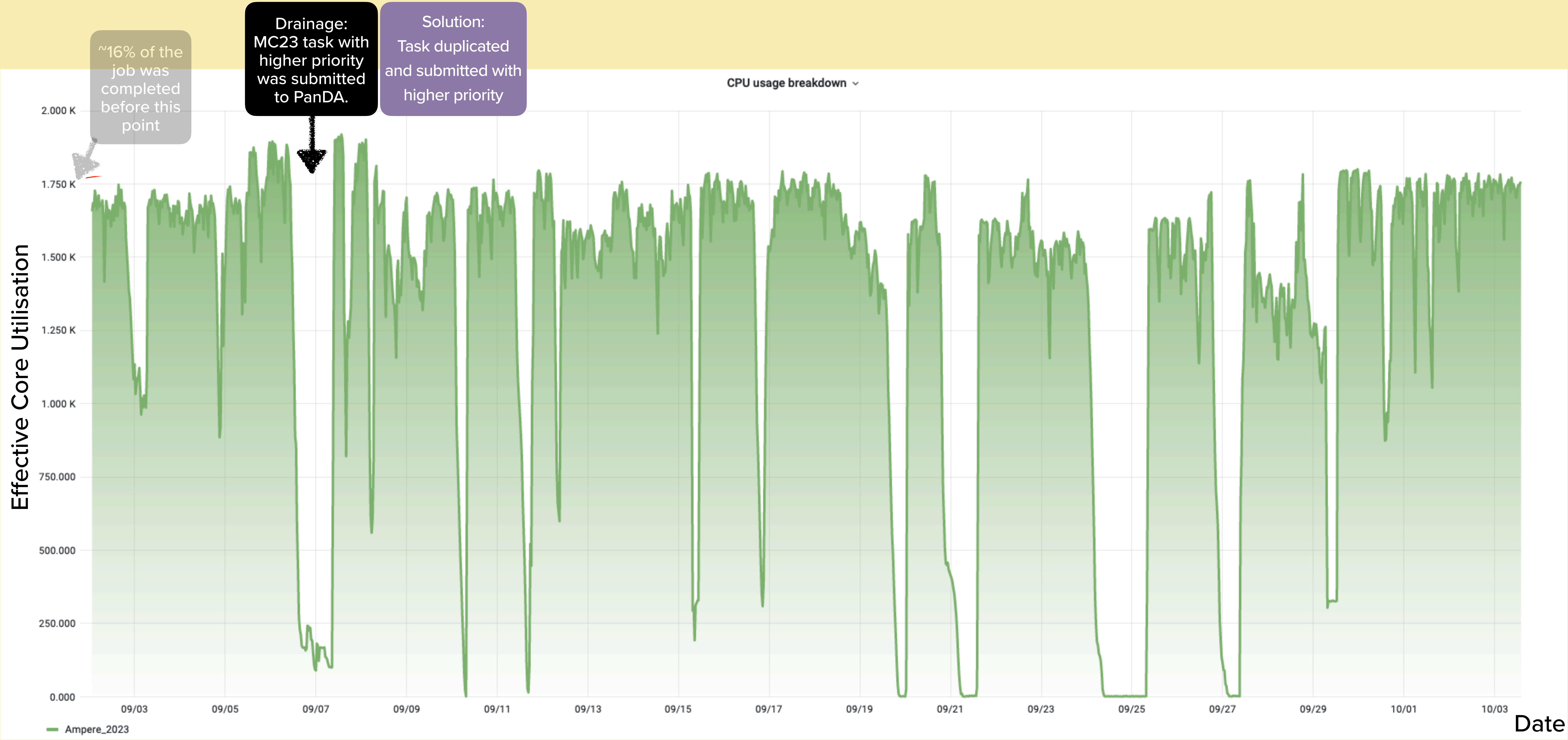
34427583 task

Task ID	Campaign	Request	Type	Processing type	Working Group	User	Nucleus	Status	N input files finished	N input events used	N output events	HS23s Expected Total done failed	Time stamps: created last modified	Cores	Priority: original current	Attempt	Parent	Tracker	CO_2 total done failed
34427583	MC23c	50572	prod	merge	AP_MCGN	janders	RAL-LCG2	finished	24,989 (100%) 24,989	49,978,000 (100%) 49,978,000	49,978,000	187.0M 179.3M 166.6M 12.6M	2023-08-21 07:10:11 2023-10-07 15:49:31	1	885 900	0	34412357	JIRA	15kg 14kg 1kg

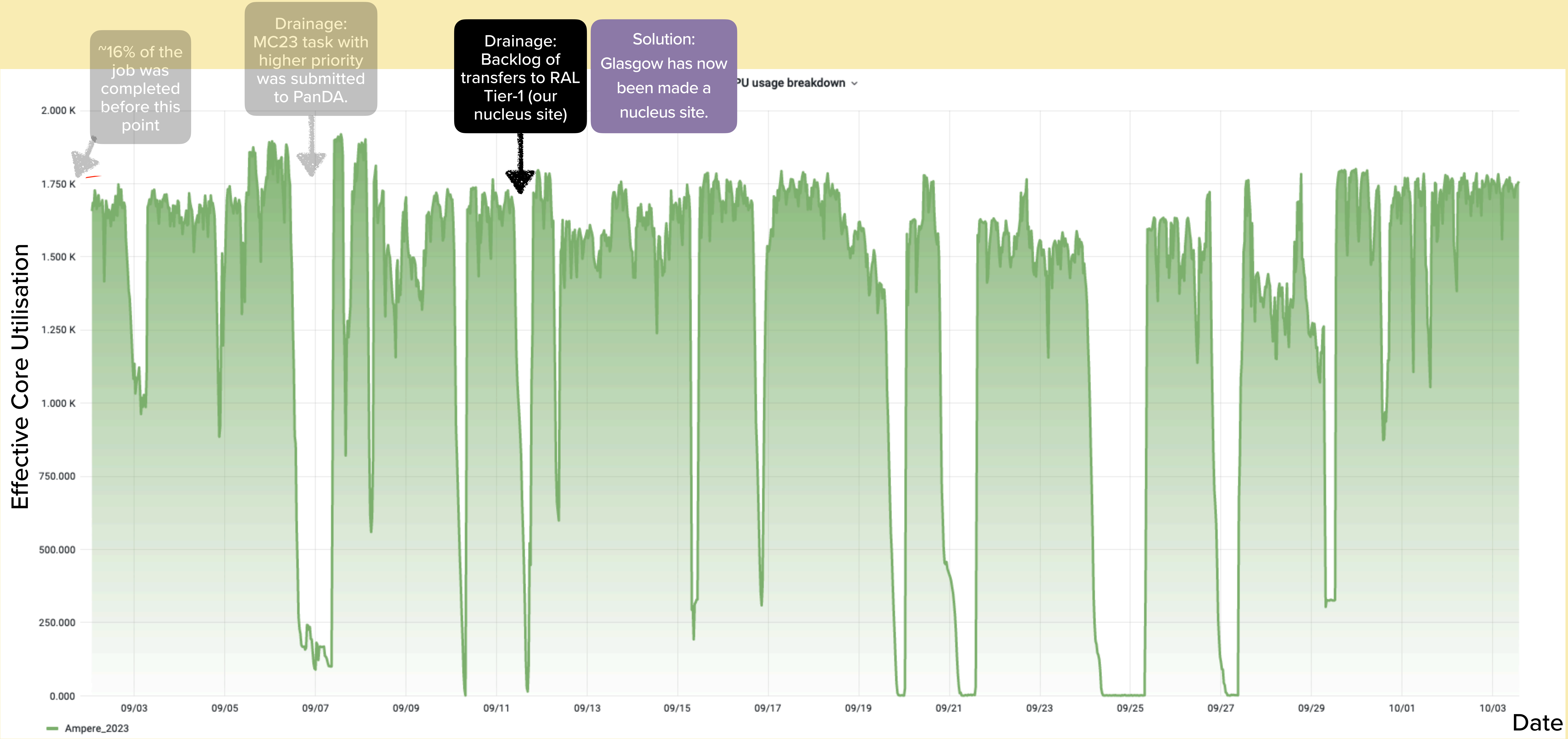
ATLAS ARM Tests



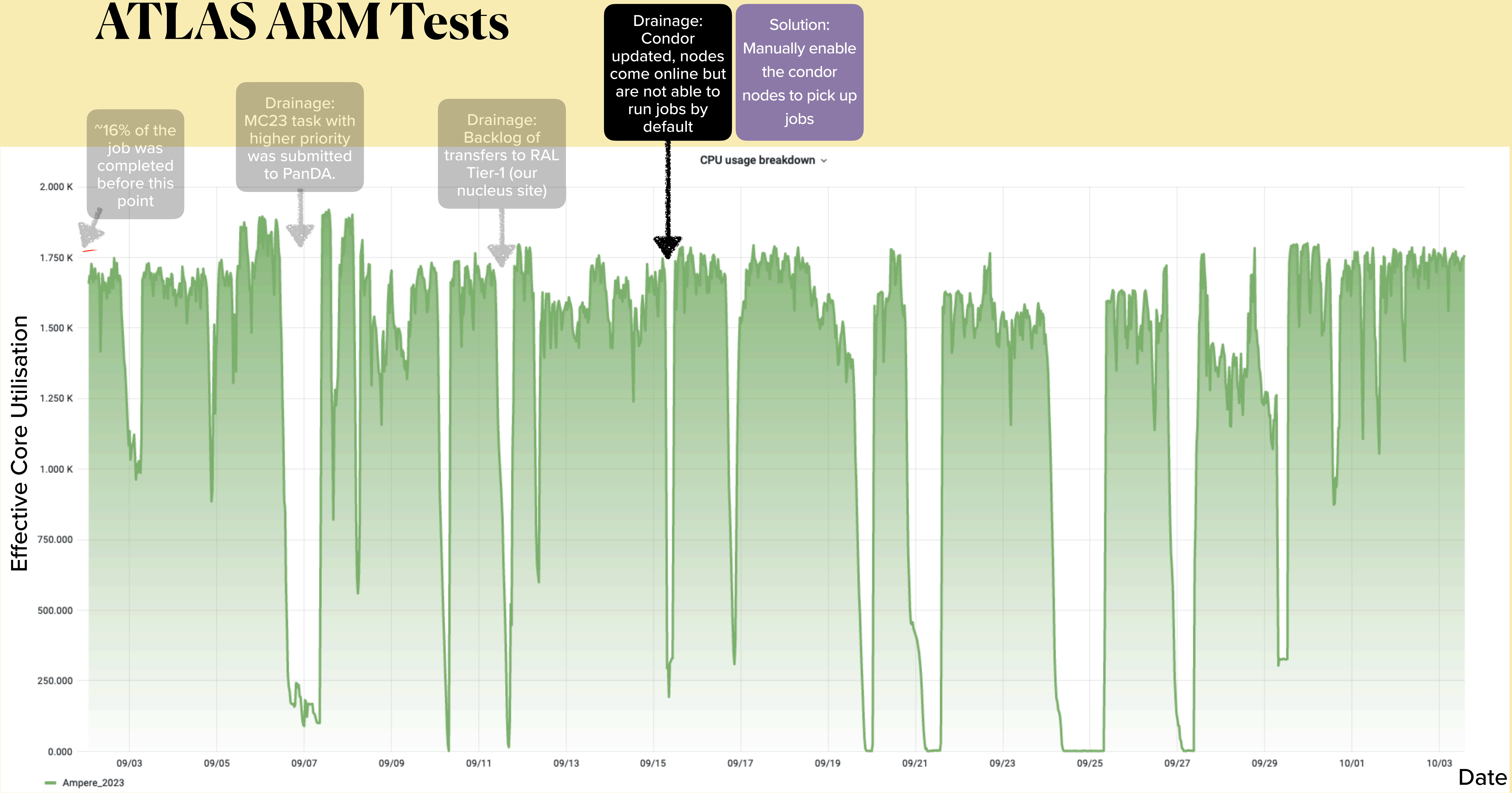
ATLAS ARM Tests



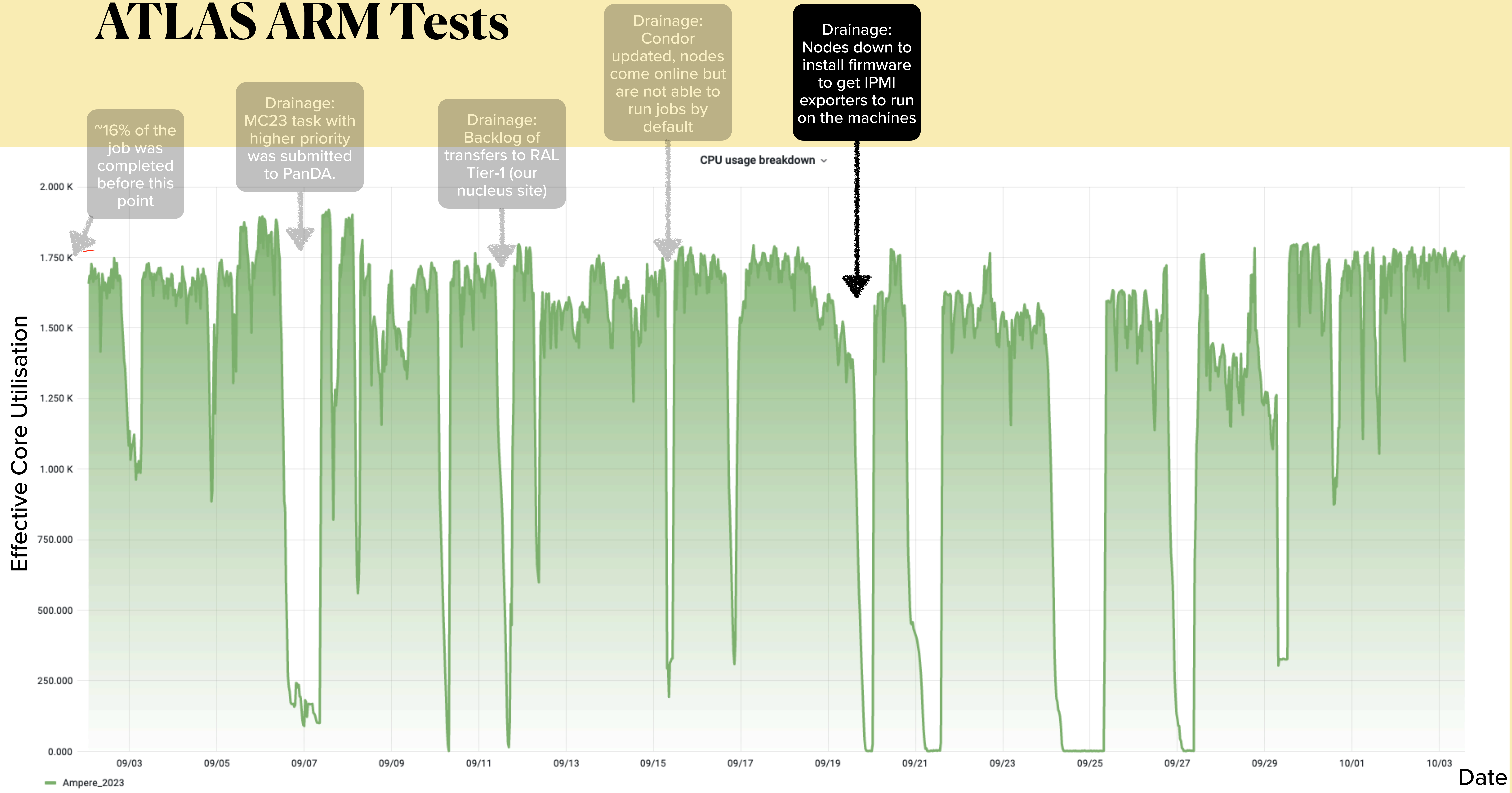
ATLAS ARM Tests



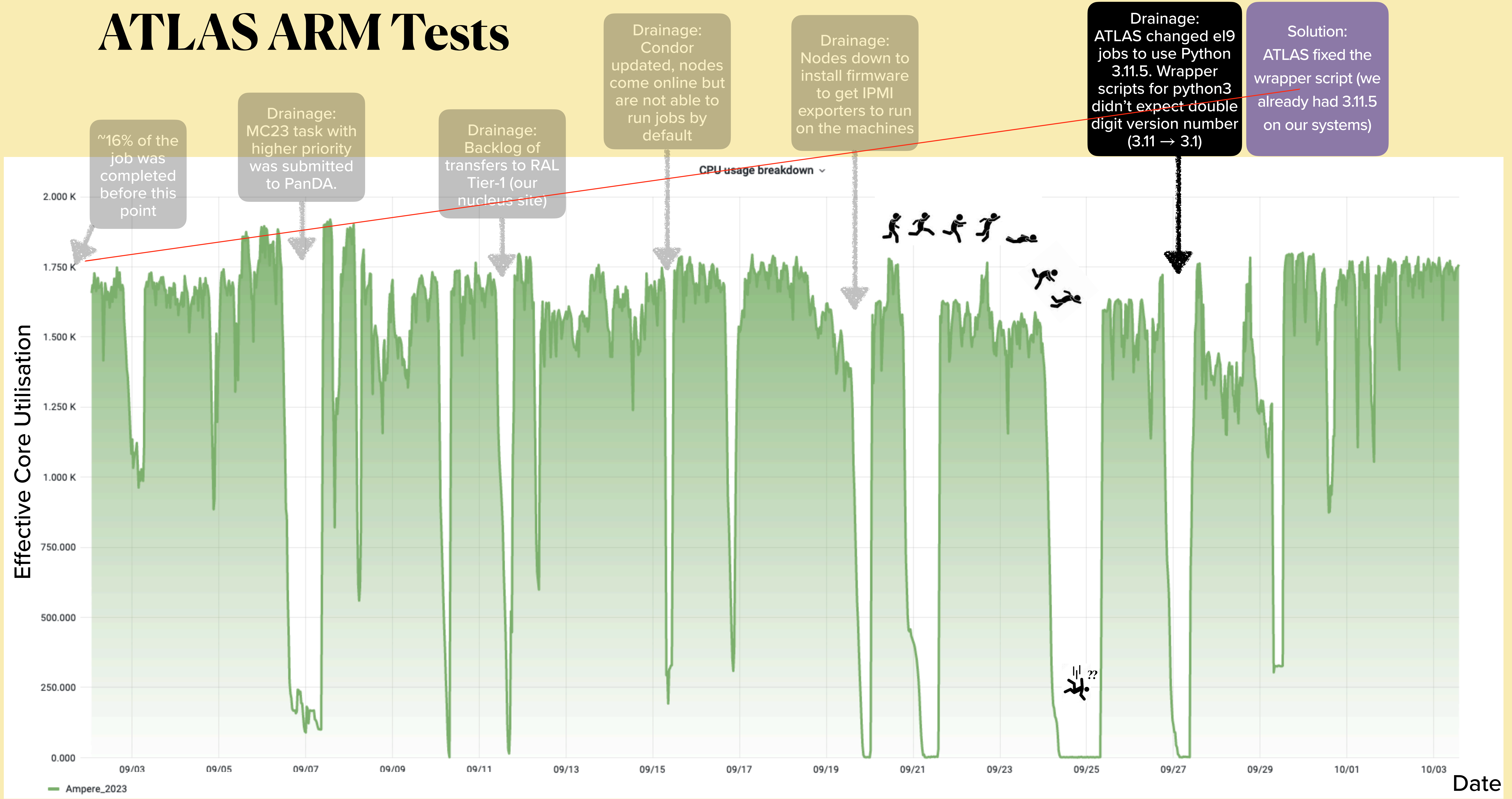
ATLAS ARM Tests



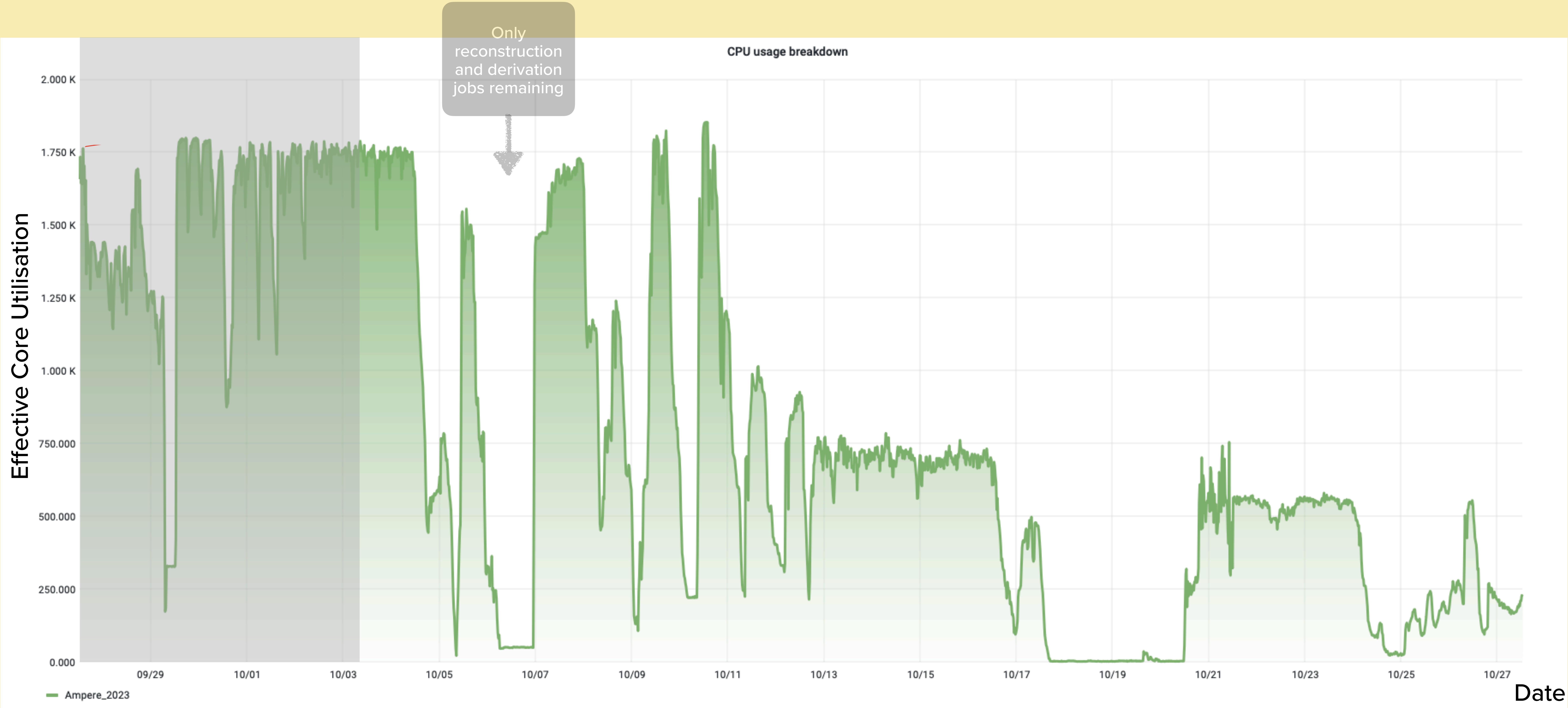
ATLAS ARM Tests



ATLAS ARM Tests



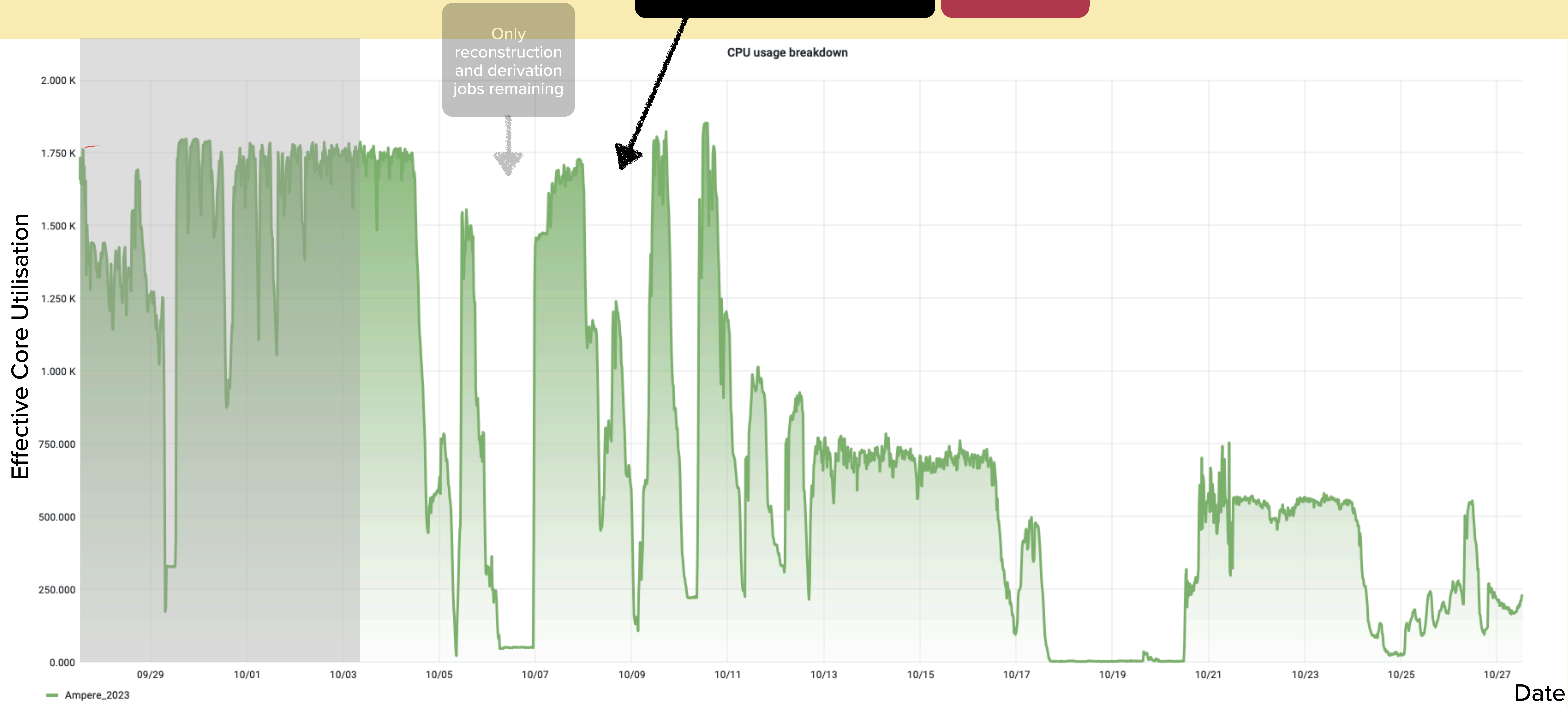
ATLAS ARM Tests



ATLAS ARM Tests

Inconsistent running : Due to the fact that each node has about 1.7TB of scratch space = ~10.5GB per core. For a multicore (8) job there should be 85GB of scratch space to play with, but reco ATLAS jobs apparently require 100GB.

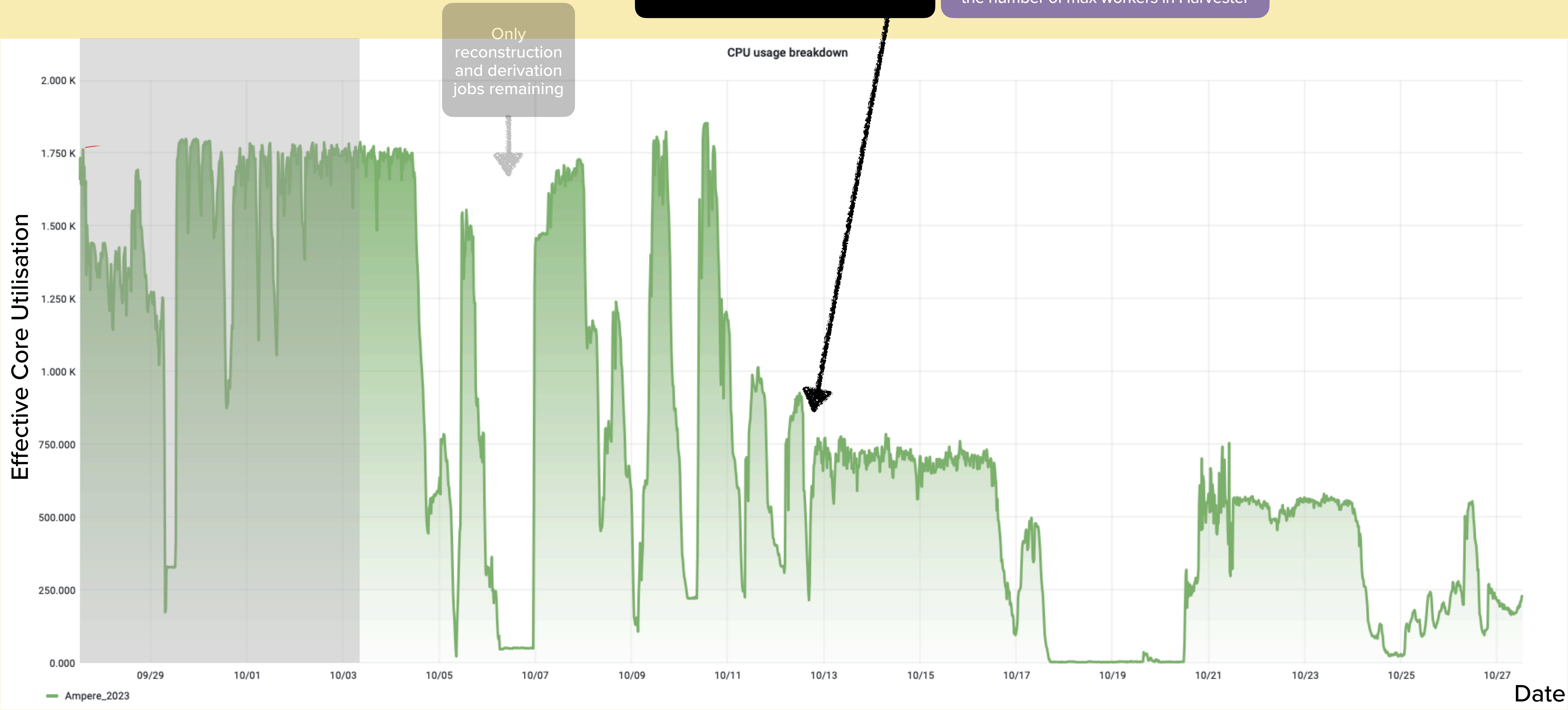
Note:
When we can we might have to add more storage to our nodes



ATLAS ARM Tests

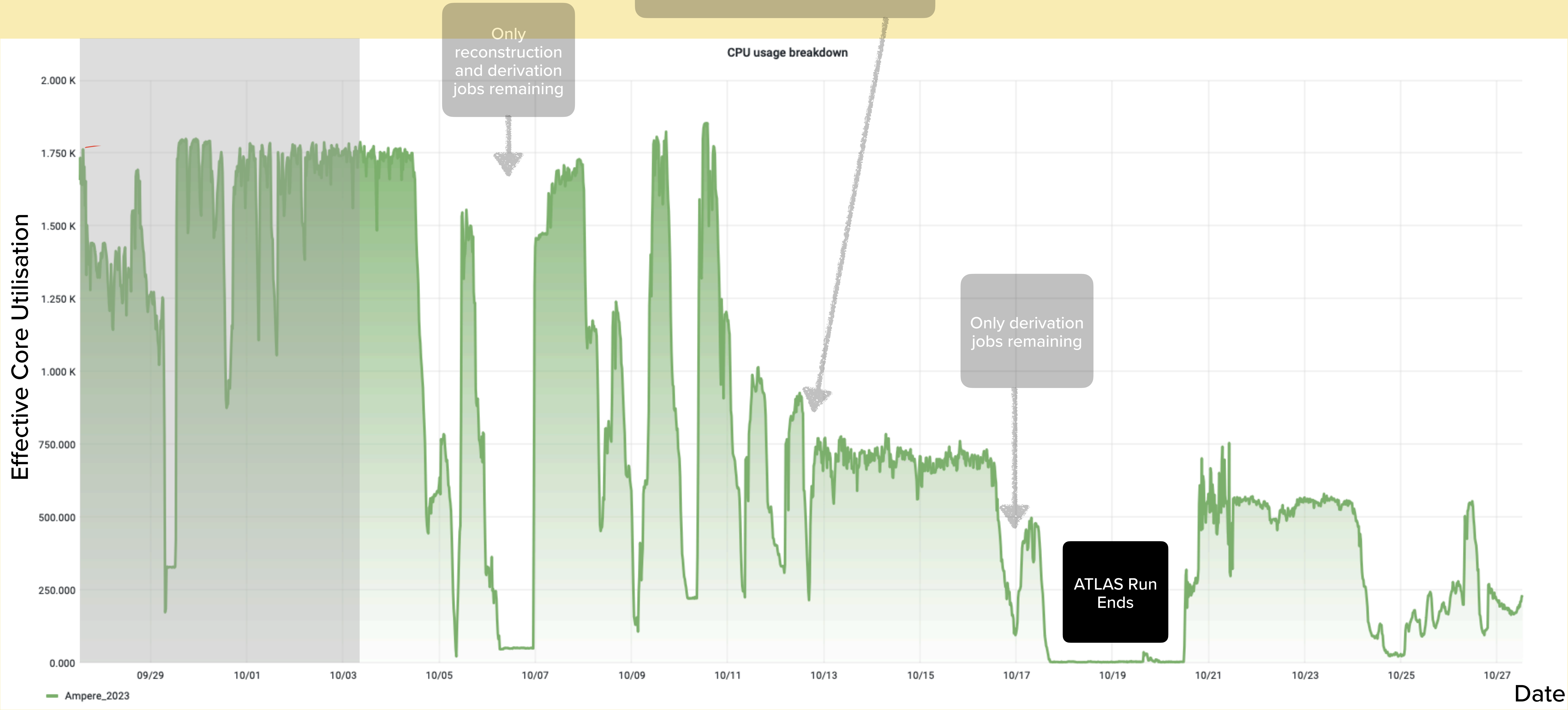
Inconsistent running : Due to the fact that each node has about 1.7TB of scratch space = ~10.5GB per core. For a multicore (8) job there should be 85GB of scratch space to play with, but reco ATLAS jobs apparently require 100GB.

Solution:
1) Setting added on ATLAS-side to delete intermediate files when running multi-step Reco transforms.
2) Nodes were half-filled by using a cap on the number of max workers in Harvester



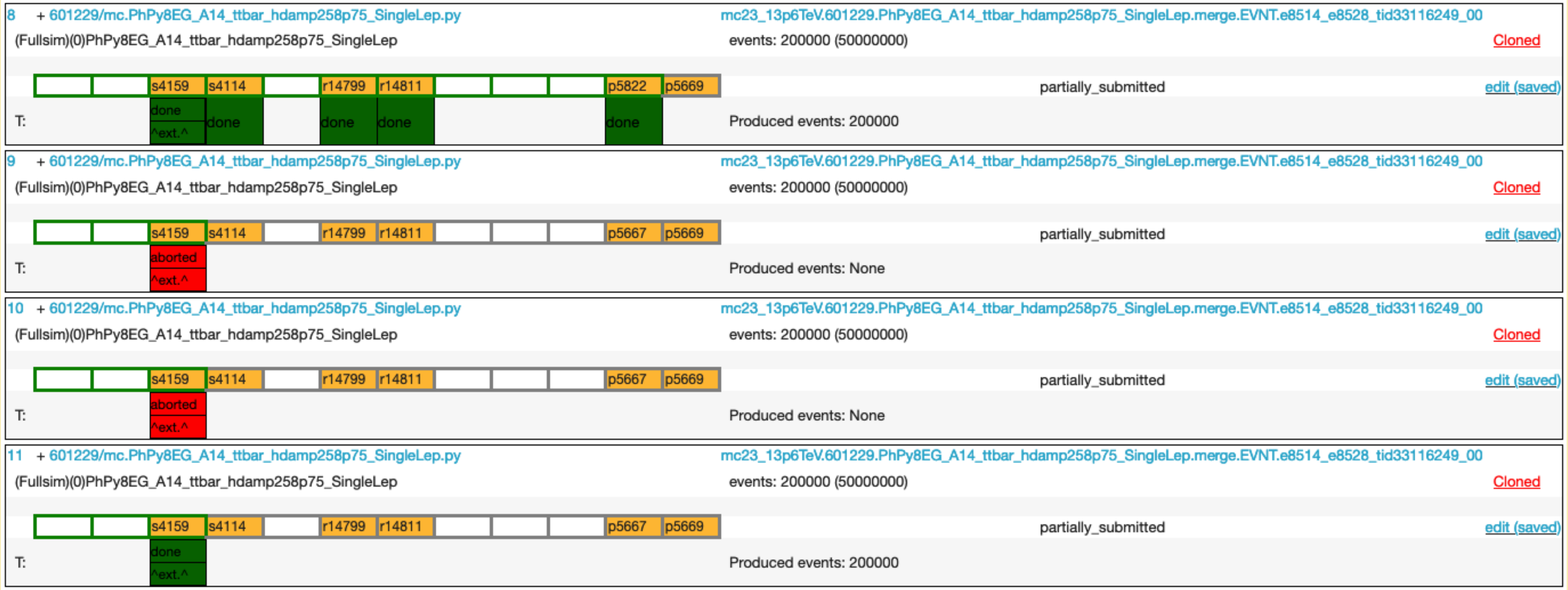
ATLAS ARM Tests

Inconsistent running : Due to the fact that each node has about 1.7TB of scratch space = ~10.5GB per core. For a multicore (8) job there should be 85GB of scratch space to play with, but reco ATLAS jobs apparently require 100GB.



ATLAS Testing - x86 Work

- Using the same s-, r- and p-tags as the ARM work, jobs were also sent our x86 nodes to get some metrics for direct comparison.
- Only 200k events were sent in this way, site still live and taking our usual mix of work from other VO's
- Took ~26 days to complete the full chain.



34714849 task

Task ID	Campaign	Request	Type	Processing type	Working Group	User	Nucleus	Status	N input files finished	N input events used	N output events	HS06*sec Expected Total done failed	Time stamps: created last modified	Cores	Priority: original current	Attempt	Tracker	gCO_2 total done failed
34714849	MC23c	50572	prod	simul	AP_MCGN	janders	SWT2_CPB	done	100 (100%) 100	200,000 (100%) 200,000	200,000	223,400,000 175,231,218 174,063,834 1,167,384	2023-09-07 07:45:09 2023-09-08 11:05:42	8	220 900	0	JIRA	15,717 15,611 105

Comparison Methodology

- Use PanDA data taken on the ATLAS-side to estimate usage metrics.
- Estimate how many cores were used for how long to calculate the average CPU-hours taken for each type of ATLAS job (area of orange to the right)
- Then use a combination of reported and measured data to calculate HEPScore/Watt

States of jobs in this task [drop mode]

	pending	defined	waiting	assigned	throttled	activated	sent	starting	running	holding	transferring	merging	finished	failed	cancelled	closed
Run													100			

Job list. Sort by PandaID, time since last state change, ascending creation time, descending creation time, ascending mod time, descending mod time, priority, attemptnr, ascending duration, descending duration

PanDA ID Attempt#	Owner Group	Request Task ID	Transformation	Status	Created	Time to start d:h:m:s	Duration d:h:m:s	Mod	Cloud Site	Priority	N input events (N input files)	Max PSS/core, GB	Job info
5957143510 Attempt 2	janders AP_MCGN	50572 34714849	Sim_tf.py	finished	2023-09-08 02:36:02	0:0:04:33	0:4:46:37	2023-09-08 10:27:09	WORLD UKI-SCOTGRID-GLASGOW_CEPH online no active blacklisting rules defined	900	2000 (1)	0.34	

Job name: mc23_valid.601229.PhPy8EG_A14_ttbar_hdamp258p75_SingleLep.simul.e8514_e8528_s4159.5956895702 #2

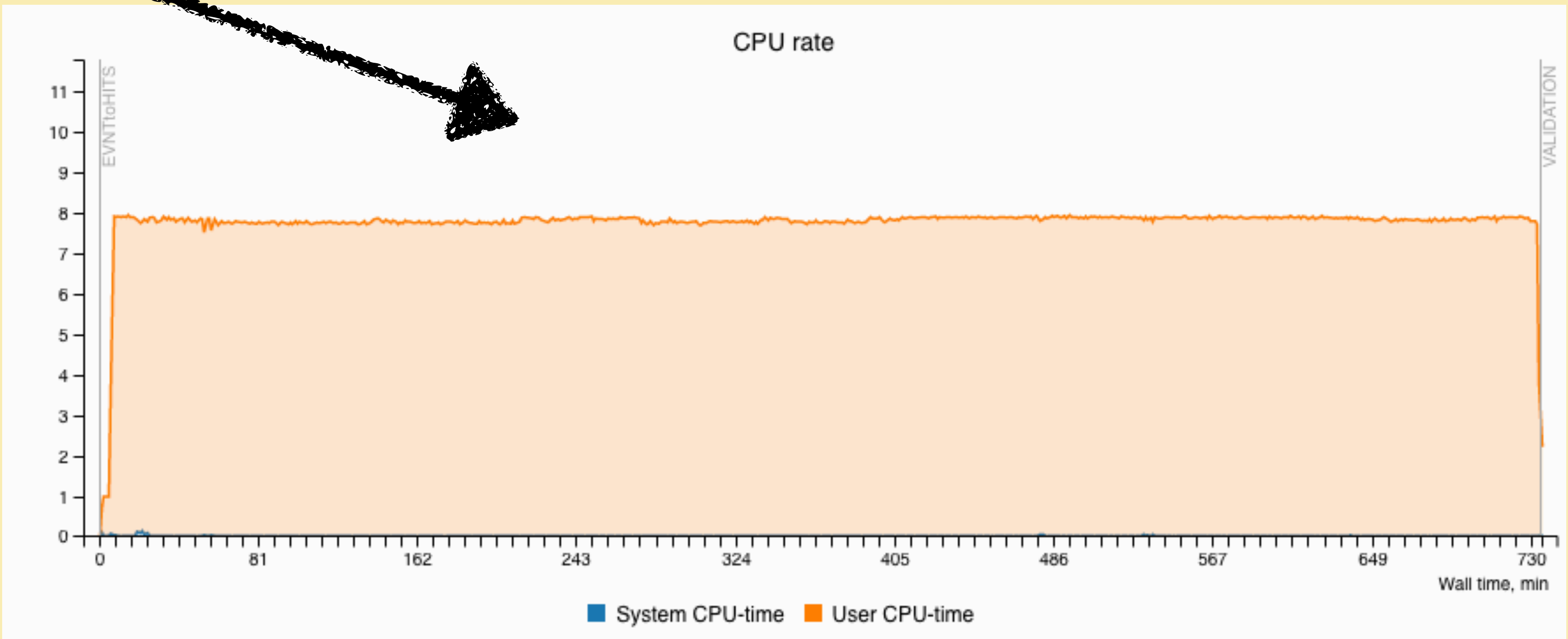
Datasets: **In:** mc23_13p6TeV:mc23_13p6TeV.601229.PhPy8EG_A14_ttbar_hdamp258p75_SingleLep.merge.EVNT.e8514_e8528_tid33116249_00 || Rucio link
Out: mc23_valid.601229.PhPy8EG_A14_ttbar_hdamp258p75_SingleLep.simul.log.e8514_e8528_s4159_tid34714849_00

Job name: mc23_valid.601229.PhPy8EG_A14_ttbar_hdamp258p75_SingleLep.simul.e8514_e8528_s4159.5956895785

PanDA ID	Owner	WG	Request Task ID	Status	Type	Transformation	Created Last modified	Time to start Duration [d:h:m:s]	Site	Harvester instance Worker ID	Cores	Priority	Attempt
5956895785	janders	AP_MCGN	50572 34714849	finished	simul	Sim_tf.py	2023-09-07 20:40:37 2023-09-08 11:05:27	0:1:35:29 0:12:18:23	UKI-SCOTGRID-GLASGOW_CEPH	CERN_central_B 468057624	8	900	1

Datasets:	**In:** mc23_13p6TeV:mc23_13p6TeV.601229.PhPy8EG_A14_ttbar_hdamp258p75_SingleLep.merge.EVNT.e8514_e8528_tid33116249_00 **Out:** mc23_valid.601229.PhPy8EG_A14_ttbar_hdamp258p75_SingleLep.simul.log.e8514_e8528_s4159_tid34714849_00
Files summary:	input: 1, size: 579.80 (MB); log: 1; output: 1; pseudo_input: 1

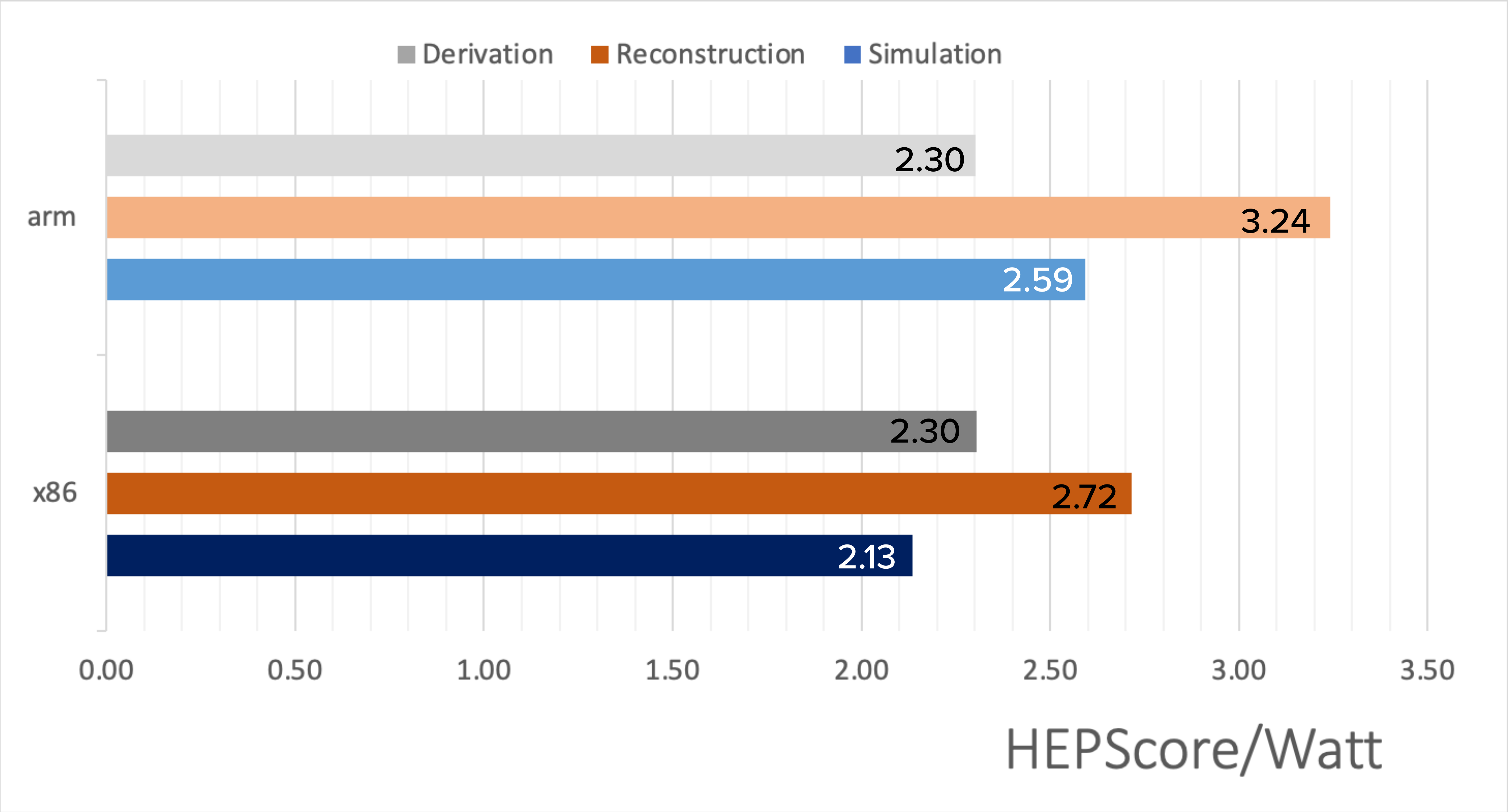
Logs ▾ | Go to ▾ | Show ▾ | Jump to ▾ | Memory and IO plots ▾

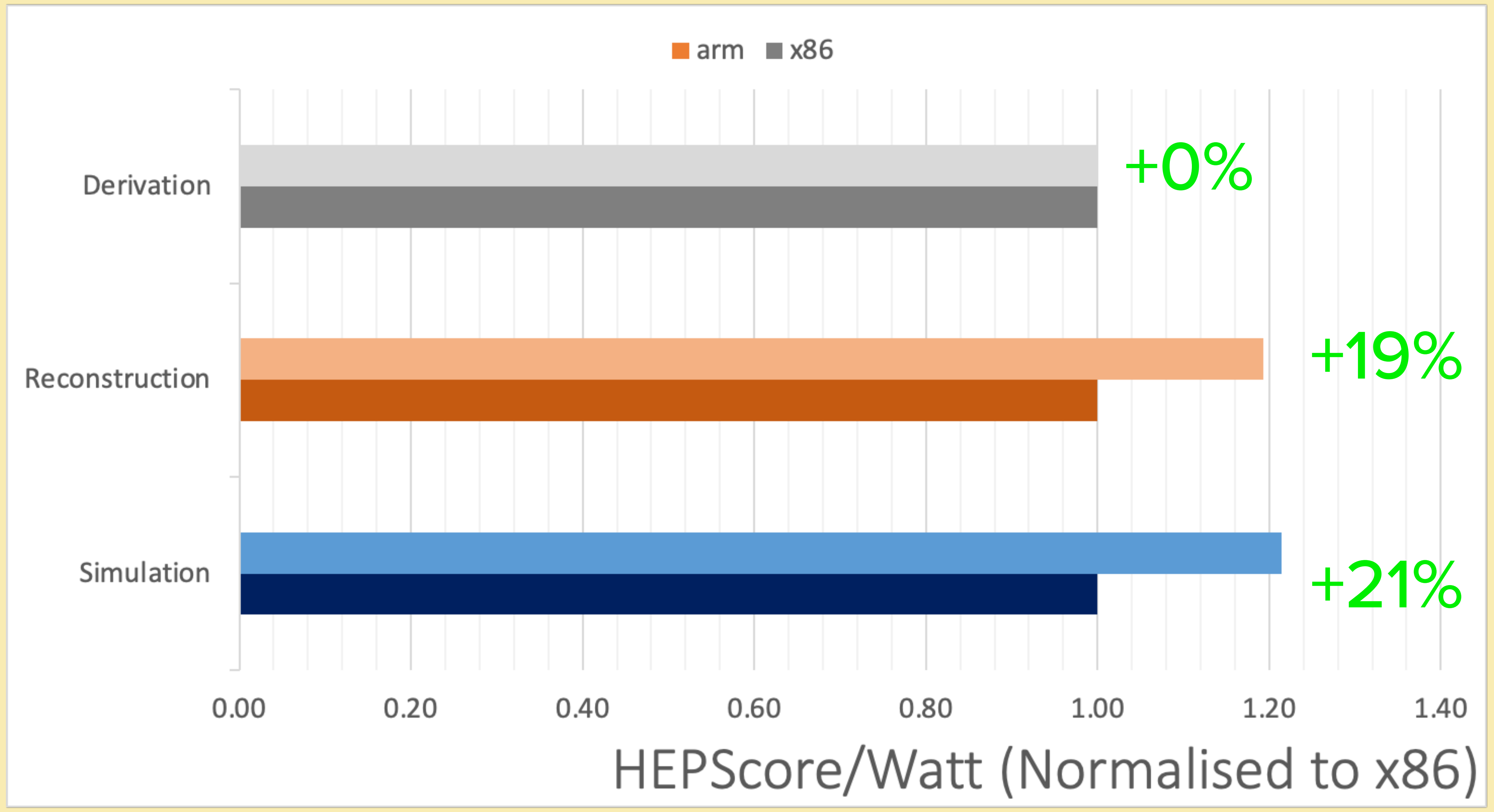


Estimates of HEPScore/Watt

$$\frac{\text{HEPScore}}{\text{Watt}} = \frac{\text{HEPScore}}{\text{Job}} \div \left(\frac{\text{Cores/Hour}}{\text{Job}} \times \frac{\text{Watt*Hours}}{\text{Core}}\right)$$

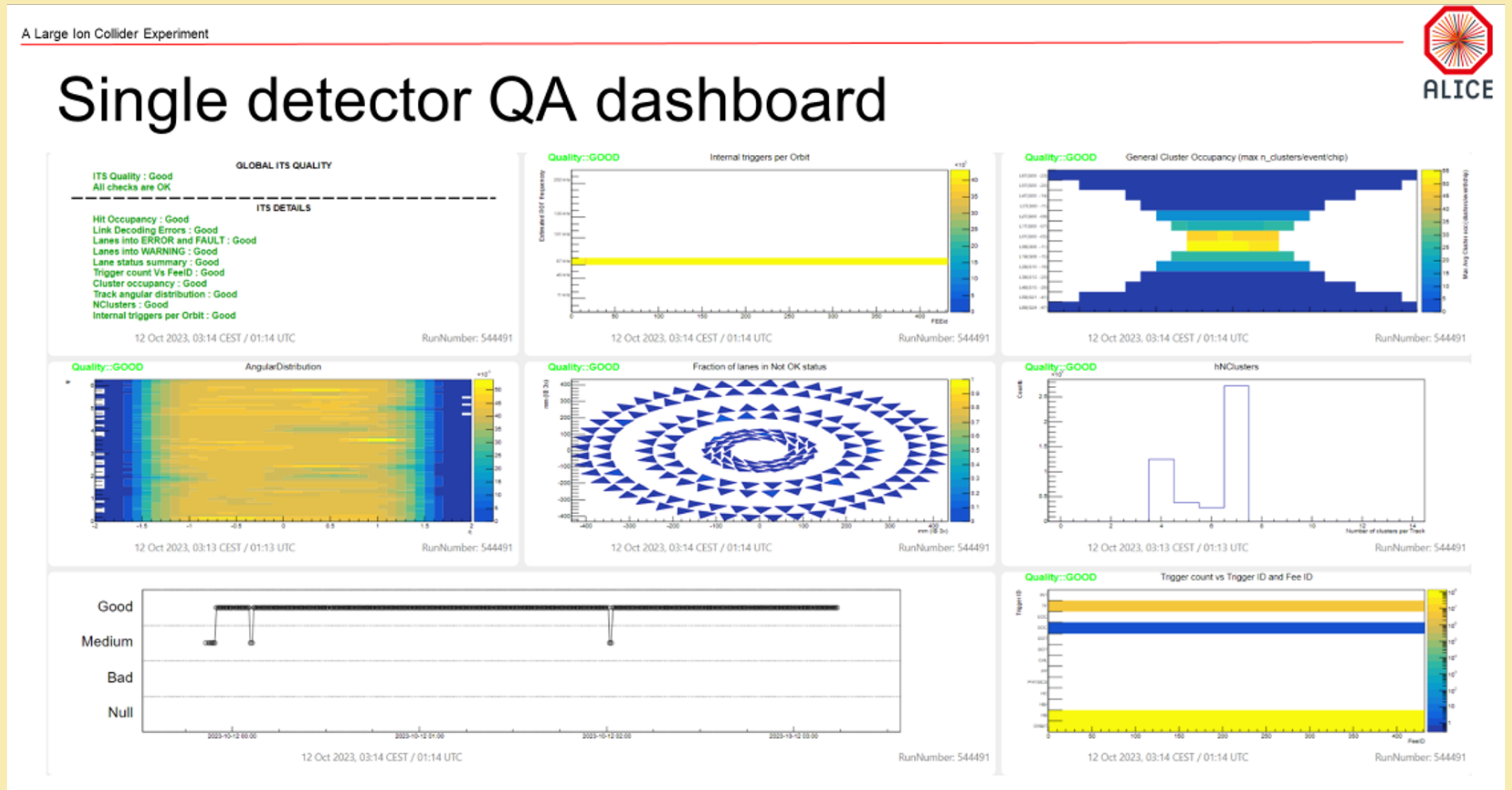
- HEPScore/Job: Obtained from ATLAS PanDA
- Cores/Hour / Job: Obtained using methodology from slide 29
- Watt*Hours / Core: Measured on site





- The x86 bars here are weighted averages from the different x86 machines we have at our cluster.
- HEPScore/Watt better for ARM for reconstruction and simulation work
- HEPScore/Watt worse for derivation work - *ARM load order of magnitude larger, expect merge steps (I/O bound processes) of jobs to more CPU intensive relative to the jobs for x86?*

LHC ARM Validation Campaign Summary



Currently **ATLAS**, **CMS**, and **ALICE**, have finished running extensive campaigns against our ARM cluster.

- **ATLAS**: Successfully ran at our site and physics validated
  ATLAS have already fully validated ARM work
  https://its.cern.ch/jira/browse/ATLPHYSVAL-919

- **ALICE**: Successfully ran work and it is physics validated (see image)
  - Green means statistical match with x86.

- **CMS**: ARM work was run for physics validation purposes …
  Ran into problems with AAA – not a CMS site so data not local
  Physics Validation unable to be statistically validated -
  CMS mostly successful with Monte Carlo data, but ran into problems with detector data that are not yet understood - more testing at CNAF

- **LHCb**: Have successfully ran test jobs, and are about to start their large-scale submission.

Conclusions and Future Work

- Glasgow has run benchmarking tests on a variety of different architectures, and will continue to do so
- Future HEPScore/Watt will hopefully be more accurate with the **<75-95%> quantile average** power, and this will be standard measurement for power we report going forward
- Due to increased power efficiency, frequency throttling should be investigated by sites if they want to reduce power/carbon (more on this after the break)

- Benchmarking shows that ARM has better performance/Watt, and that this performance translates to site-level comparisons (albeit not at the same degree)
- Experiments are willing to engage with sites to validate ARM workloads and increasing amounts of architecture on their side are being built on ARM.
- From this, we are roughly doubling the capacity of our ARM farm provision - M128-30's are coming
- We hope that it's possible from 2025 to start pledging ARM resources to experiments - Potentially at 20% to start with.