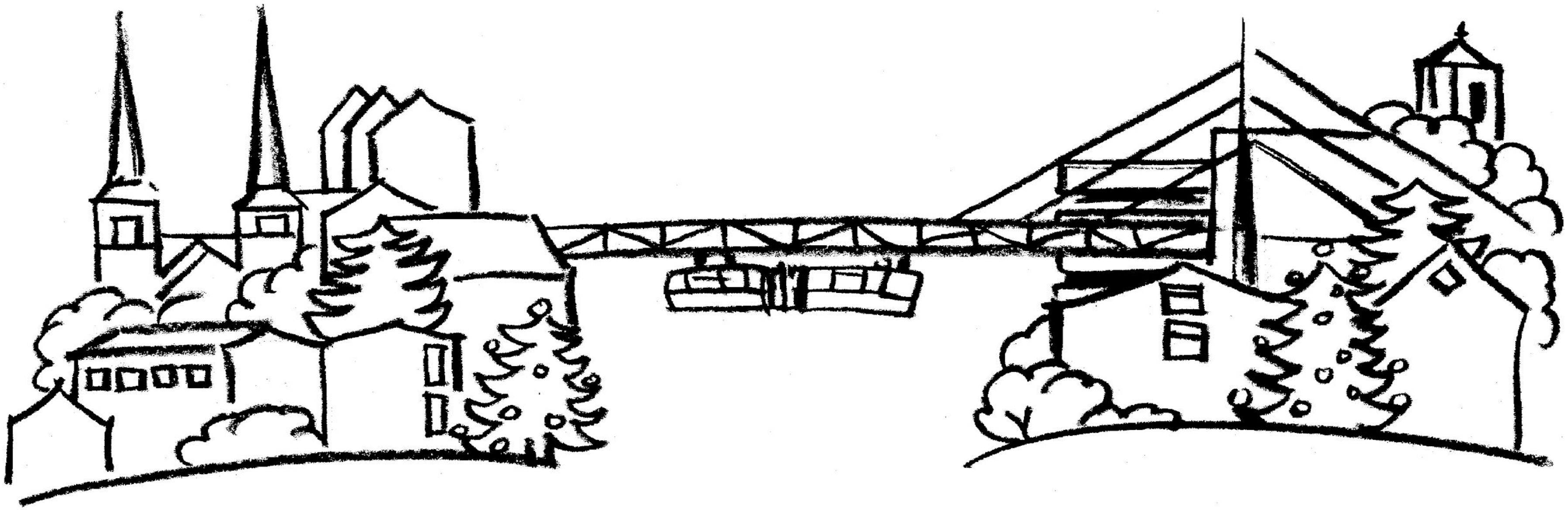
Status Wuppertal



Anna Vorländer
Marisa Sandhoff
Torsten Harenberg

Bergische Universität Wuppertal

Wuppertal ATLAS **Computing** Team

Management

Wolfgang Wagner, Christian Zeitnitz

ATLAS / HPC Computing

permanent:

Marisa Sandhoff – ATLAS, (local) HPC, all Tier-2 services, Strategy, ATLAS DE Cloud
Torsten Harenberg – ATLAS, (local) HPC, all Tier-2 services, Strategy, ATLAS DE Cloud co-administration, TAB
Martin Errenst – ATLAS, main HPC.NRW contact, local HPC (esp. Slurm and BeeGFS)
Frank Ellinghaus – JEM/PAVER
Ralf Häusling – Hardware support

non-permanent:

Shayma Wahdan – ATLAS, HPC.NRW
Raphael Kleinemühl – FIDIUM
Anna Vorländer – JEM/PAVER
Johanna Kraus – JEM/PAVER
Mustafa Schmidt – Detector Simulations (GEANT) (M&O-A)

ATLAS Tier-2 - some key facts -

Fully integrated into local HPC center “PLEIADES”.

- 17152 AMD EPYC 7452 CPU Cores, full HDR100 Infiniband
  - 4 GB RAM / Core (this is “high mem” for ATLAS)
  - 25 GB / Core local, mixed-mode SSD for fast, local scratch space
- ~1 TB BeeGFS Cluster File System
- ~3.5 PB dCache (for ATLAS)
- GPU Cluster with 40 NVidia A100 GPUs
- own 10 GBit/s WAN, all local LAN connections >= 10GBit/s
- Other than ATLAS Grid usage: ~220 users (local and associated institutes)
  - Shibboleth account creation for outside users (dev. in cooperation with our “ZIM”)
- integrated into the HPC.NRW competence network

Grid Services are all containerized (or are currently migrated) and orchestrated through Ansible, most other services (like LDAP) as well. All solutions developed on-site.

Home of JEM / PAVER (see Anna’s slides later)

HEPSPEC23 tests – machine design **essential** for HEP performance (e.g. fast local hard drive)

AMD EPYC 7452 32-Core Processor	IFIC	12.496	35
	MPPMU	12.515	46
	OU_OSCER_ATLAS	13.189	12
	OU_OSCER_ATLAS_TEST	13.346	12
	RAL	16.609	82
	RAL-LCG2_MCORE	16.214	20
	UKI-SCOTGRID-GLASGOW_CEPH	15.970	114
	UKI-SOUTHGRID-OX-HEP	14.975	48
	UKI-SOUTHGRID-RALPP	16.403	12
	ifae	13.454	58
	pic	12.988	67
	wuppertal	20.494	66

same CPU but different "HEP performance"

Credit: Michael Böhler, GridKa Cloud monthly meeting, 05-07-2023

substantial over-the-pledge contributions by DE Unis (and DESY)

ATLAS comp. ressources mostly from Unis (in contrast to CMS)

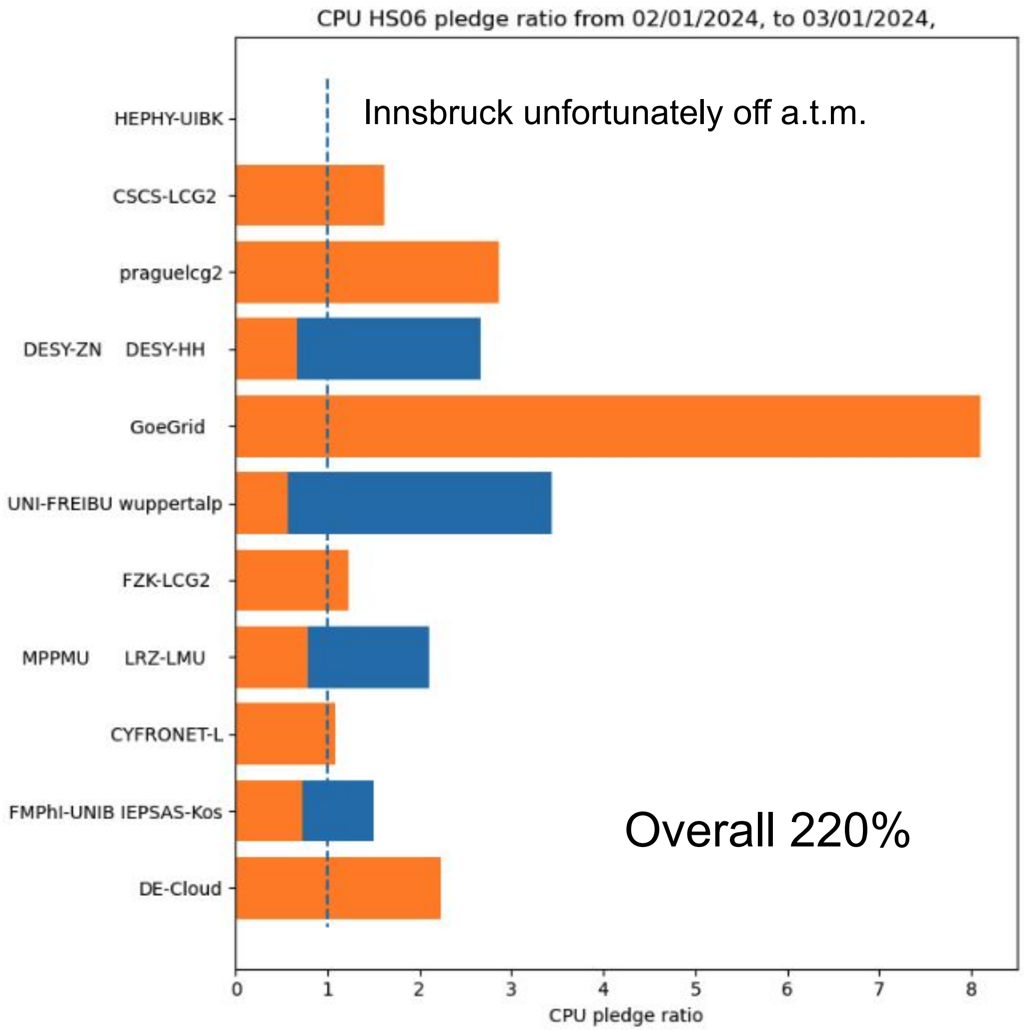
GridKa : 25.10%
DESY/MPI : 31.91%
DE Unis : 42.99%

} w.r.t. wall clock time

Jan-Mar 2024, source: M. Boehler, T.H., TAB Report March 2024

Wall clock time. All jobs (HS23 seconds) Oct 23

	Value	Percent
FZK-LCG2	578 Bil	18%
CSCS-LCG2	419 Bil	13%
wuppertalprod	401 Bil	12%



source: G.Duckeck, GridKa Cloud monthly meeting 13-MAR-2024

Computing: some current developments (ATLAS related)

- XCache + Cobald/Tardis + AUDITOR (R. Kleinemühl for FIDIUM)

  Setup of an XCache instance (req. full Kubernetes K8s setup with SLATE CI)
  Setup of a Cobald/Tardis queue for Wuppertal – accounting via AUDITOR
  bandwidth
  Kubernetes plugin for AUDITOR

- optimizing PLEIADES usage for ATLAS

  constant SLURM adjustments to run special high-mem ATLAS jobs in Wuppertal
  (in collaboration with Rod Walker)

Computing: some current developments (ATLAS related)

- “federated dCache” (M.S./T.H. with Thomas Hartmann and Christian Voss (DESY) for FIDIUM)

  development of local, cache pool nodes within a large dCache instance

  Idea:
  - have the central dCache head node @ DESY, satellite cache pool nodes @ site(s)
  - Developing secure intra-dCache communication channels and measuring performance

  Status: Test system is running @ Wuppertal, currently working on ad-hoc VPN structures

- containerization and semi-automatic rollout of a Grid Site (M.S./T.H.)
  – as an alternative to Cobald/Tardis –

  - developed/ing standardized Grid-enabled (incl IPv4/6 dual-stack) Alma9 containers with Ansible support
  - developed Ansible playbooks for all Grid services (ARC, Squid, BDII, WNs, dCache SNs), should be easy to adopt to other sites

The PAVER framework

- MC validation in ATLAS is done using the PAVER framework
- **Developed**, **maintained** and constantly **improved** by ATLAS group @ **Bergische Universität Wuppertal**
- Can be accessed via **nice webpage**: `jem.cern.ch` (CERN SSO necessary)

BERGISCHE UNIVERSITÄT WUPPERTAL

PAVER

PMG **A**rchitecture for **V**alidating **E**vgen with **R**ivet

Physics **M**odeling **G**roup

Event **gen**eration = MC

Johanna Kraus Anna Vorländer DPG spring meeting 2024 4 / 11

Why validation is important

official MC production
80 billion events in 2023

- MC production campaigns for the collaboration (e.g. for LHC Run 3) are **huge**

⇒ Important to **find issues/bugs/features etc. beforehand**

Latest "MC23d" re-reconstrucion campaign produced ∼**11Mt CO_2**
≈ 1 mio. × the yearly production of one person living in Germany (11t)
≈ 3 mio. scientists flying from Karlsruhe to LA (USA) and back (3.5t)

⇒ We should make sure we trust our MCs!

Validation samples
0.5 billion events in 2023

- Generate **small MC samples** at first

⇒ Test and **validate** them **extensively**

- Save a lot of time, money and resources by careful validations

⇒ Important aspect towards **sustainability** in ATLAS!

Validation workflow using PAVER

1. **Production of MC samples**
   - **Different physics processes** to cover all scenarios (∼7 samples)
     - SM and BSM: e.g. $t\bar{t}$, W+jets, SUSY, ...
   - New samples for **each new version** (with otherwise identical setup)

2. **Fill histograms**
   - Run automatically using RIVET
   - Select set of ∼**10-20** RIVET **analyses** for each sample
   - ⇒ Get ∼**200-400 histograms** with just a few clicks

3. **Validation**
   - Each new sample is compared to a **reference file**
   - ⇒ Usually the **last validated version** of this generator/process
   - Show validation **results** on the PAVER **webpage**
   - ⇒ Large set of **checks and comparisons** between reference and new samples
   - ⇒ Can be shared directly with e.g. generator **experts**

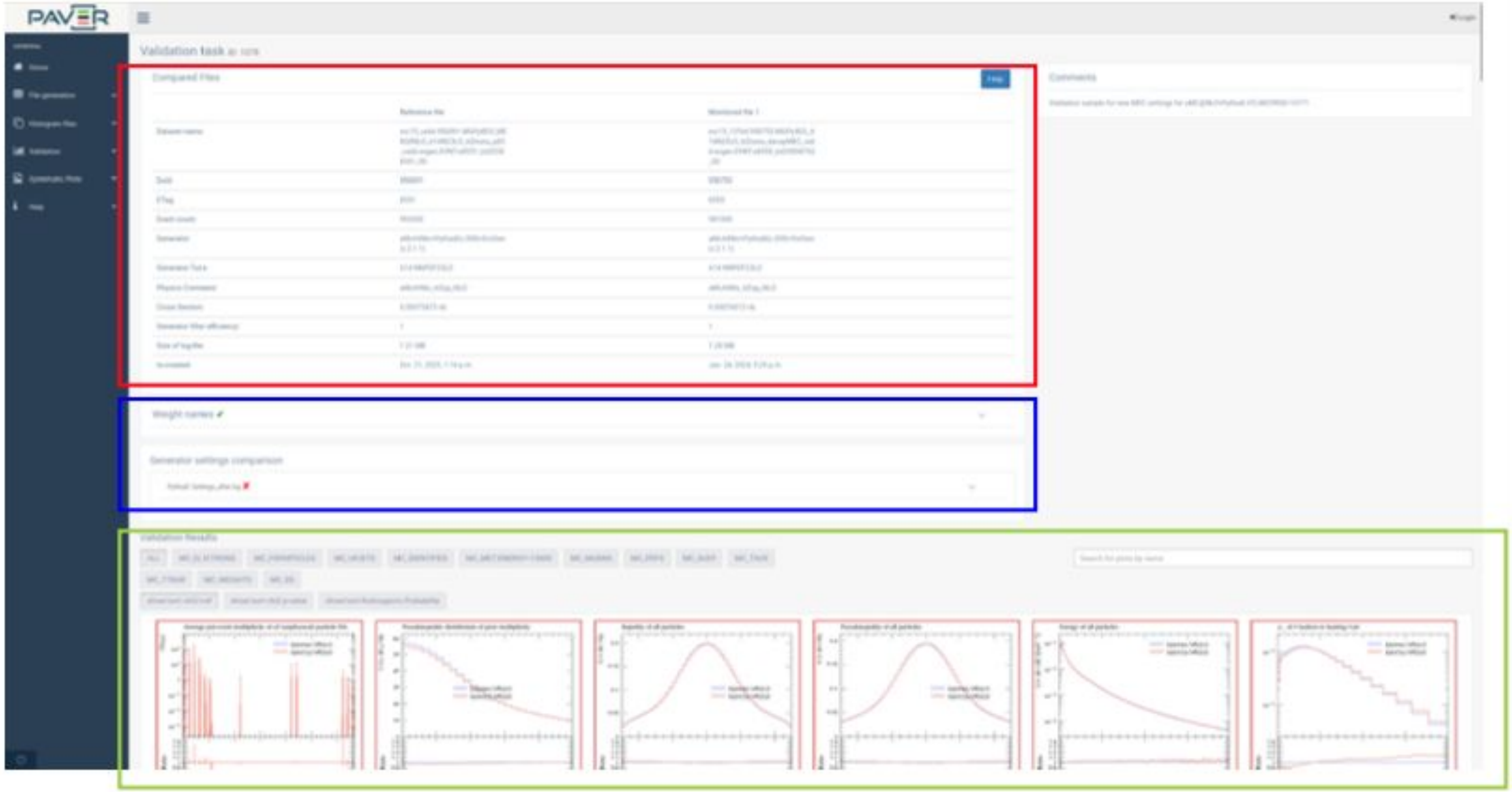
PAVER

PAVER Johanna Kraus Anna Vorländer DPG spring meeting 2024 6 / 11

The PAVER webpage - exemplary results page

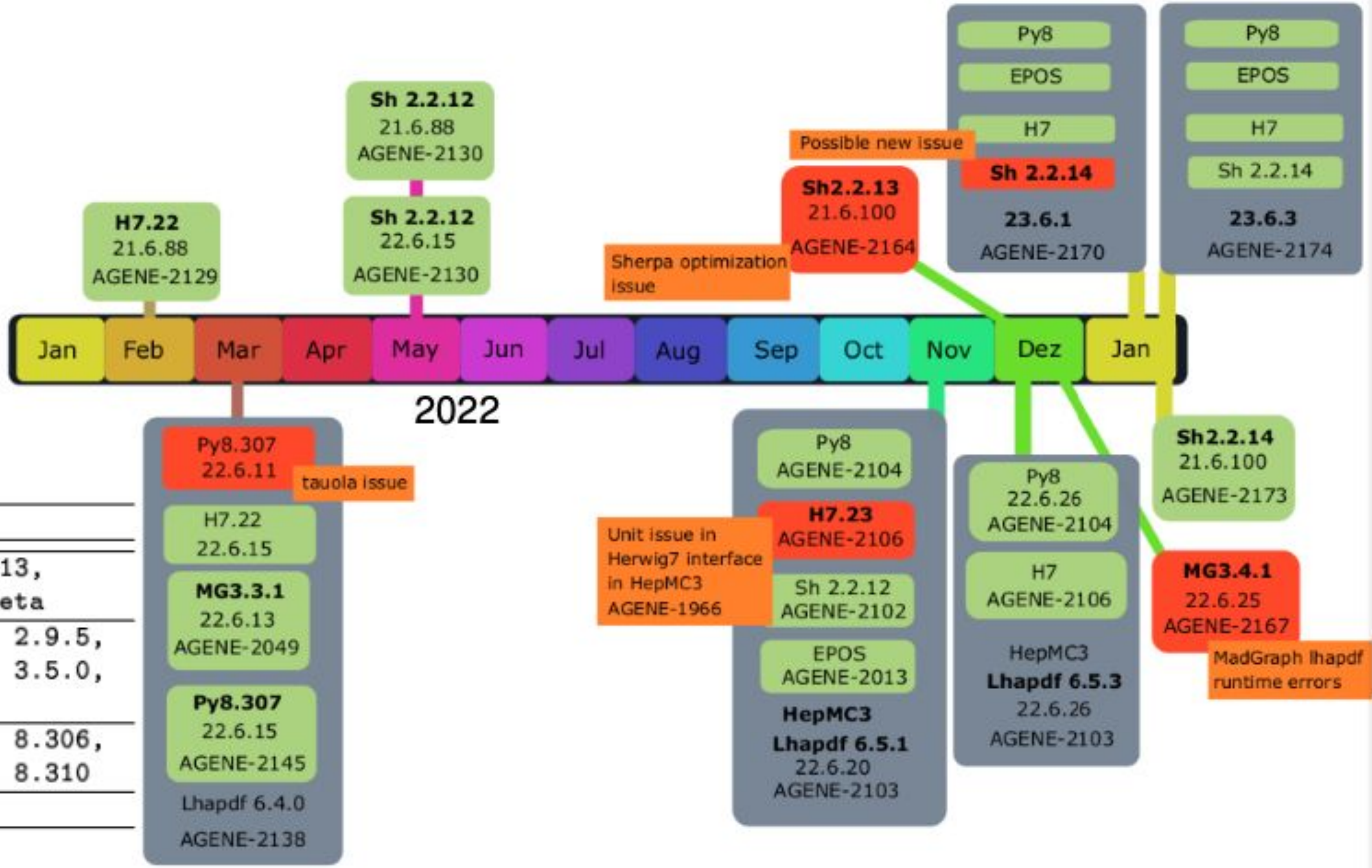
- Table with **meta info**
- **Automated comparison** of
  - generator **weights**
  - generator input **settings**
- **Hundreds of histograms** with
  - Many **sorting** and **filtering** features
  - **Statistical tests**: χ^2 and Kolmogorov-Smirnov
  - **Color code** based on p-value

  $\Rightarrow$ **Identify histograms with disagreement** at one glance



Validation successes

- Massive **validation program** over the last years
- Many **successfully validated** generator (or software) updates



Generator	Reference	Monitored versions
SHERPA	2.2.10	2.2.11, 2.2.12, 2.2.13, 2.2.14, 3.0.0alpha/beta
MG_AMC@NLO	2.7.3	2.8.1, 2.9.2, 2.9.3, 2.9.5, 2.9.9, 3.3.1, 3.4.1, 3.5.0, 3.5.1, 3.5.3
PYTHIA	8.244	8.245, 8.303, 8.305, 8.306, 8.307, 8.308, 8.309, 8.310
HERWIG	7.2.1	7.2.1, 7.2.3, 7.3.0

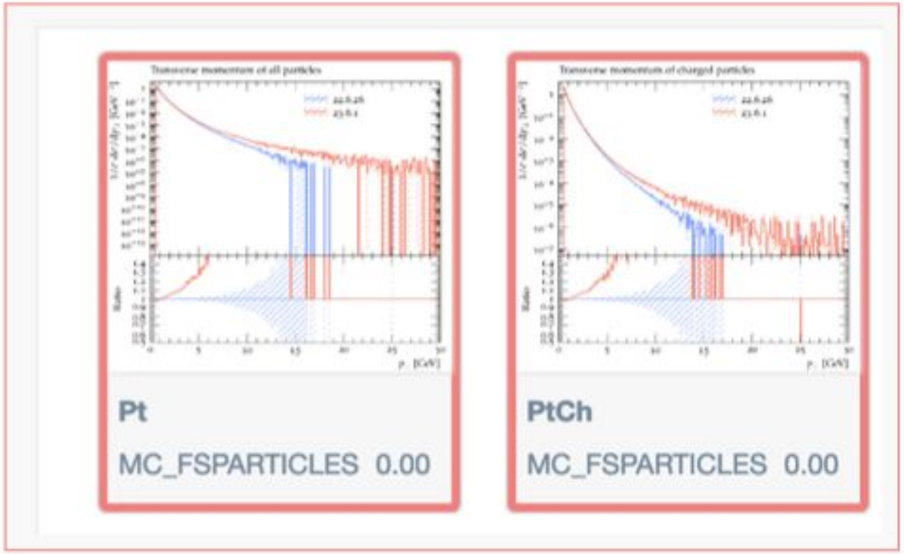
Later validations: reference = most recent successfully validated version

Validation successes

- **Found** important **issues** and bugs with the help of `PAVER` ✓
- Automated production and sorting of hundreds of plots:

⇒ Find issues **affecting** only **certain processes, variables or regions** (that might be unexpected)

- Example 1:



- AthGeneration `23.6.1`, `Pythia 8` validation
- Differences only in jetjet_JZ0 filtered sample

⇒ **Filters** were **not working** in this version → Fixed in newer version

PAVER Johanna Kraus Anna Vorländer DPG spring meeting 2024 10 / 11

Conclusions

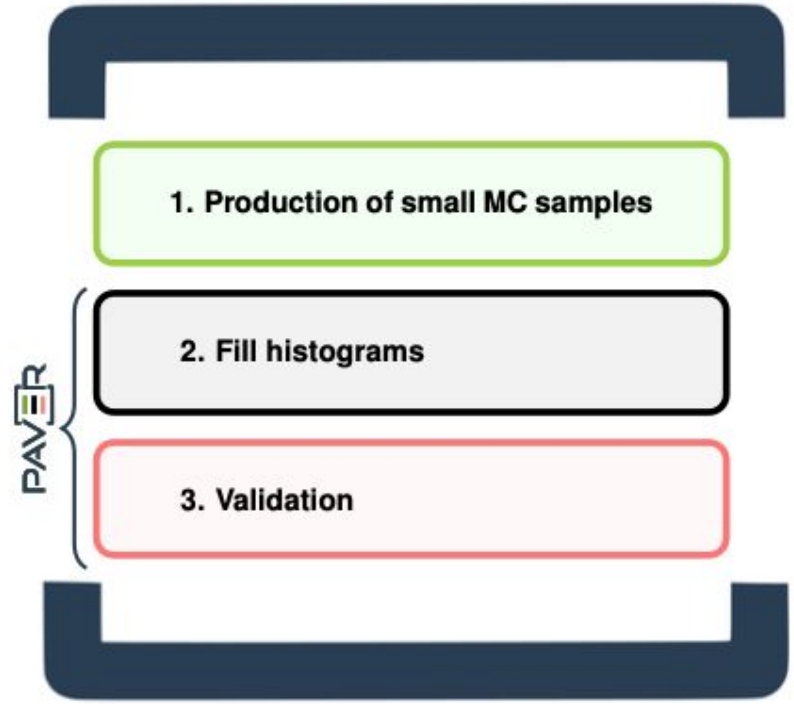
- **Monte-Carlo** (MC) simulations play a **key role** in high energy physics
- **Continuous validation** of new versions is immensely **important**

⇒ Make sure we only produce **huge MC sets** for the ATLAS collaboration if we are sure they are **correct** ⇒ **Sustainability**!

- **3-step validation procedure** using the PAVER framework
  Available via website: `jem.cern.ch`

⇒ **Automated**, **robust**, **fast**, **intuitive** and easily **accessible** MC validation setup

✓ **Hundreds of histograms** are produced with just **a couple of clicks**
  - \+ including **statistical tests** and many **sorting** & **filtering** features
  - \+ checks of **meta info**, **weights** and **settings**



- Plan for the **future**: open PAVER to a **large community** (within ATLAS)

PAVER Johanna Kraus Anna Vorländer DPG spring meeting 2024 11 / 11

BERGISCHE UNIVERSITÄT WUPPERTAL

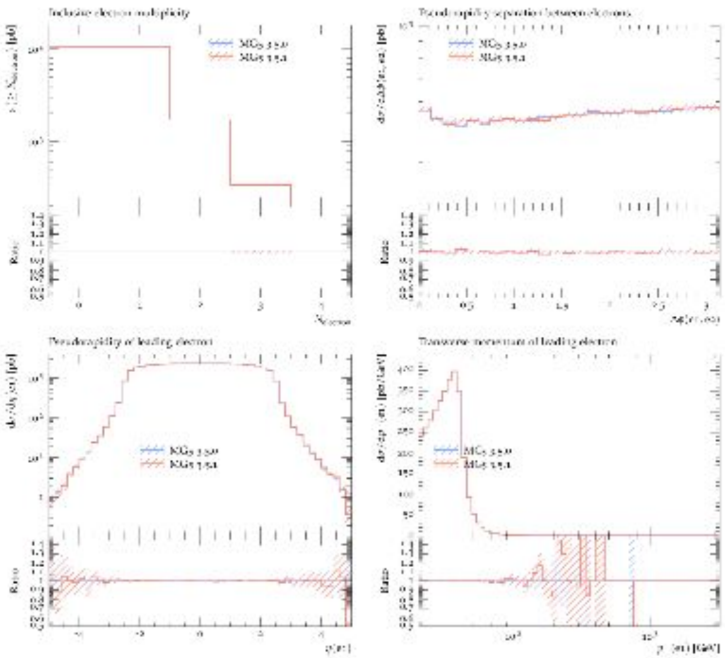
Backup

RIVET

- RIVET: **R**obust **I**ndependent **V**alidation of **E**xperiment and **T**heory

- **Used widely** across **HEP community**
- Is updated and maintained regularly
- **Preserve analysis** code from HEP experiments (e.g. ATLAS, CMS)

exemplarily 4/20 histograms of
MC_ELECTRONS



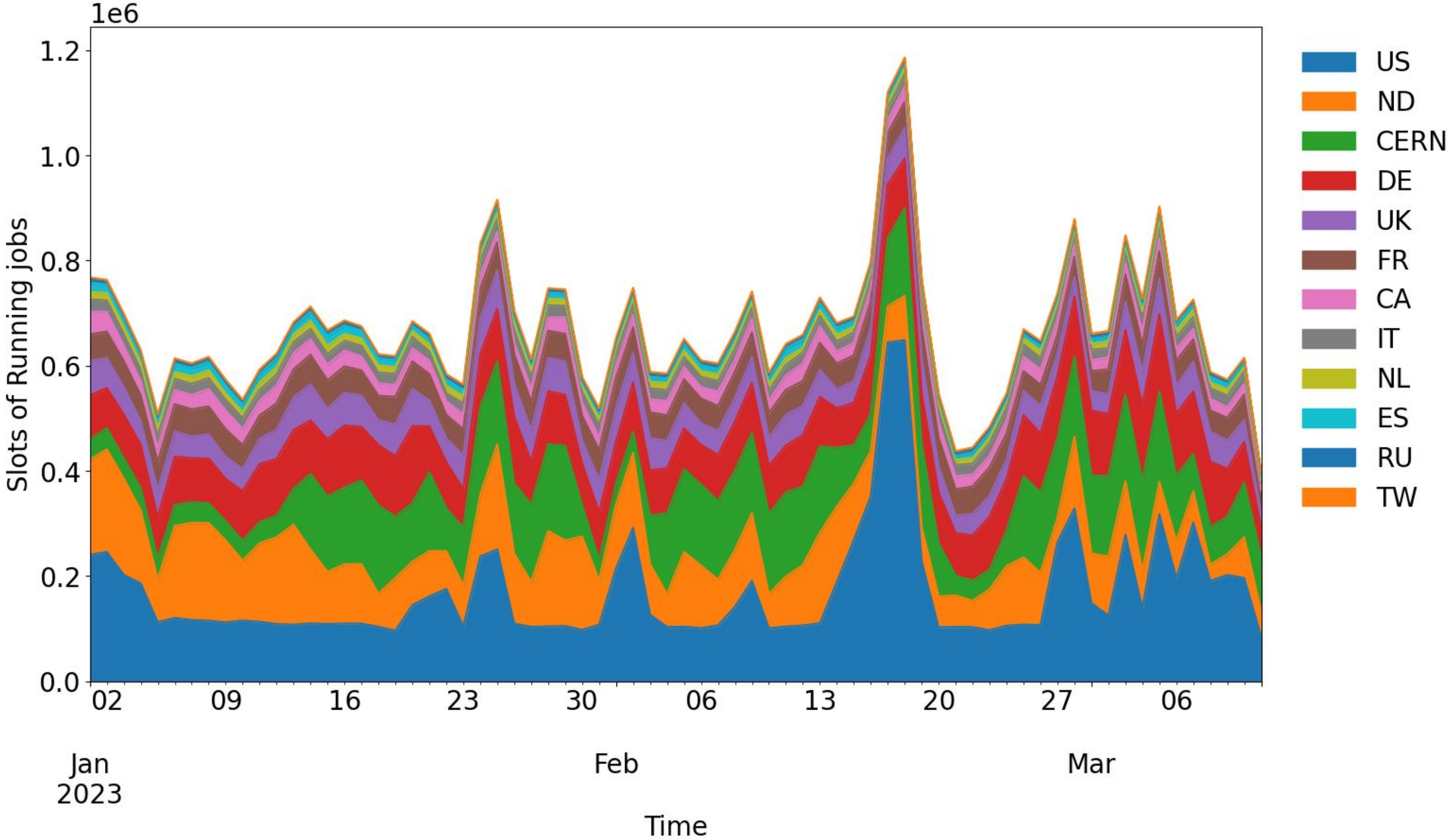
- Ever-increasing **collection of pre-defined analysis** based on
  - Final states → e.g. "MC_ELECTRONS"
  - Processes → e.g. "MC_TTBAR"
  - Measured data → e.g. "ATLAS_2018_I1656578"
- Defines a **fixed set of histograms** for each analysis

- Output files are called YODA files (**Y**et More **O**bjects for **D**ata **A**nalysis)

PAVER **Johanna Kraus** Anna Vorländer **DPG spring meeting 2024** **5 / 11**

ATLAS Clouds Jan - Mar 2024

- stable contribution of the DE sites: 14.7 % #4 of all world-wide resources

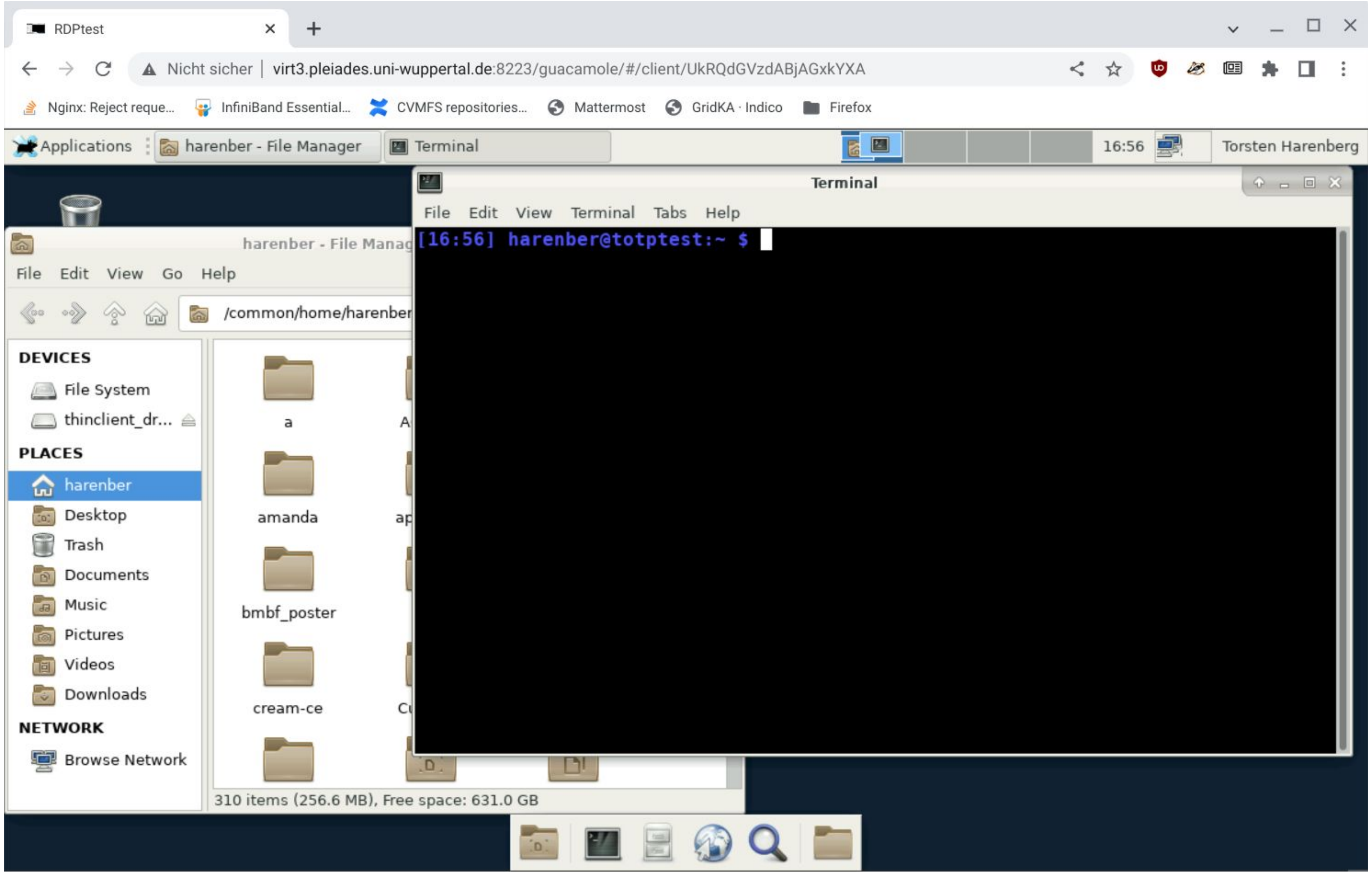


25.31 %
16.52 %
16.24 %
14.66 %
7.80 %
7.33 %
3.92 %
3.17 %
1.86 %
1.76 %
0.90 %
0.53 %

Jan-Mar 2024, source: M. Boehler, T.H., TAB Report March 2024

BERGISCHE UNIVERSITÄT WUPPERTAL

other Wuppertal Computing projects: Shibboleth-based account creation + Web-based login to HPC resources



other Wuppertal Computing projects:
Accessing HPC resources via Jupyterhub (web-based)