

ErUM Data Pilot Project

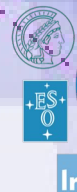
Innovative Digital Technologies for Research on Universe and Matter



Bundesministerium
für Bildung
und Forschung

ErUM-Data
Community Meeting
18.01.2021

Thomas Kuhr
LMU München

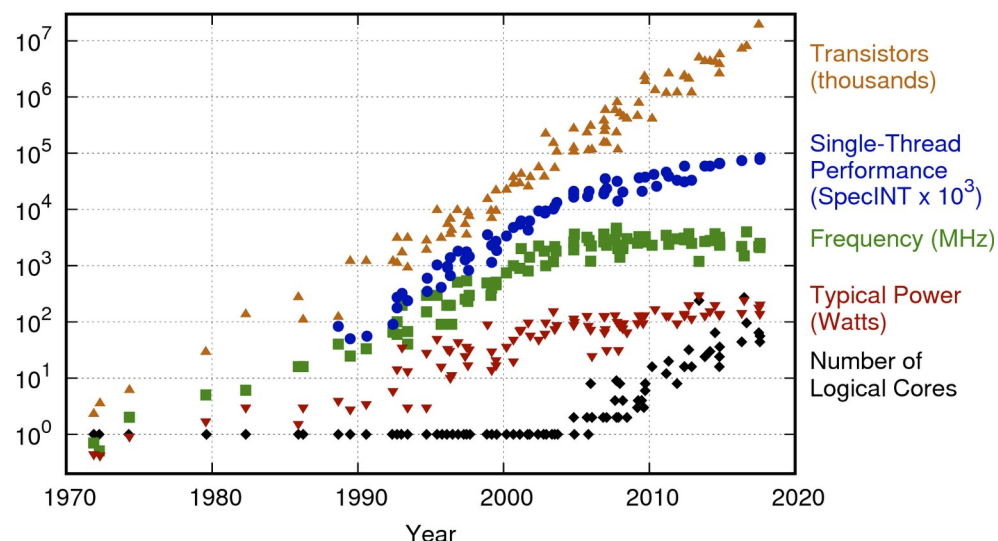


Common Challenges and Opportunities

Challenges and opportunities:

- Data volumes/rate
- New technologies (e.g. AI, hardware architectures)
- Data management part of BMBF call for particle/nuclear physics in 2017
- ➔ Experiment overarching "Verbundantrag" on "Querschnittsthema" computing and software

42 Years of Microprocessor Trend Data

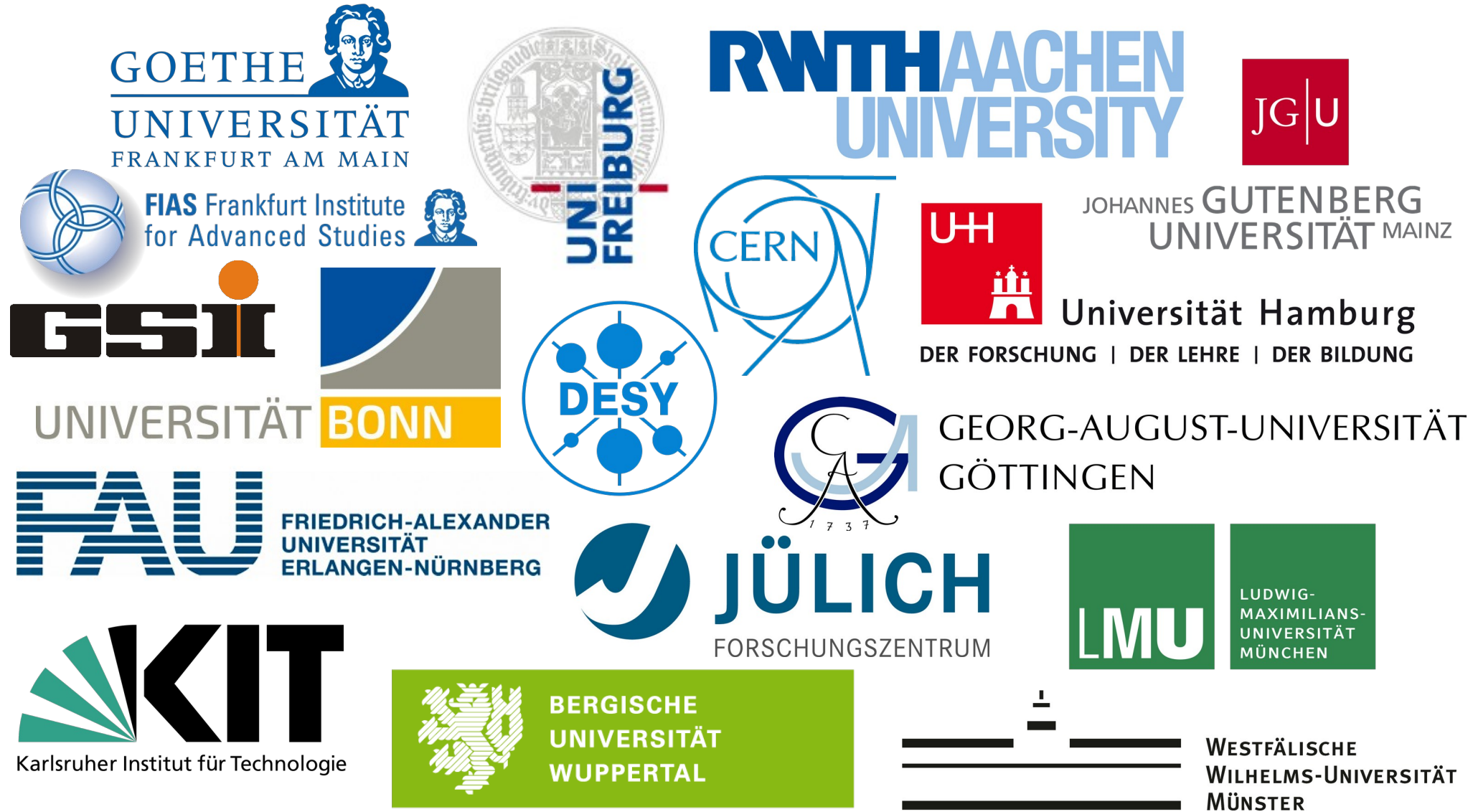


Original data up to the year 2010 collected and plotted by M. Horowitz, F. Labonte, O. Shacham, K. Olukotun, L. Hammond, and C. Batten
New plot and data collected for 2010-2017 by K. Rupp



- **Innovative Digital Technologies for Research on Universe and Matter**
- Application of partners from
 - ✓ Particle Physics (ATLAS, Belle II, CMS)
 - ✓ Hadron and Nuclear Physics (ALICE, CBM, PANDA)
 - ✓ Astroparticle Physics (Auger, CTA, IceCube)
- to develop experiment overarching solutions
- Evaluated by panel including computer scientists
- Got 3.6 M€ for 3 years, started October 2018

Project Partners



- Developments for the provision of technologies for the use of heterogeneous computing resources

A1: Tools for integration <ul style="list-style-type: none">• Scheduling of cloud jobs• Container technologies• Database access	A2: Efficient Use <ul style="list-style-type: none">• Transient data caches• Transparent access to distributed data
A3: Workflow Control <ul style="list-style-type: none">• Optimization with data mining	

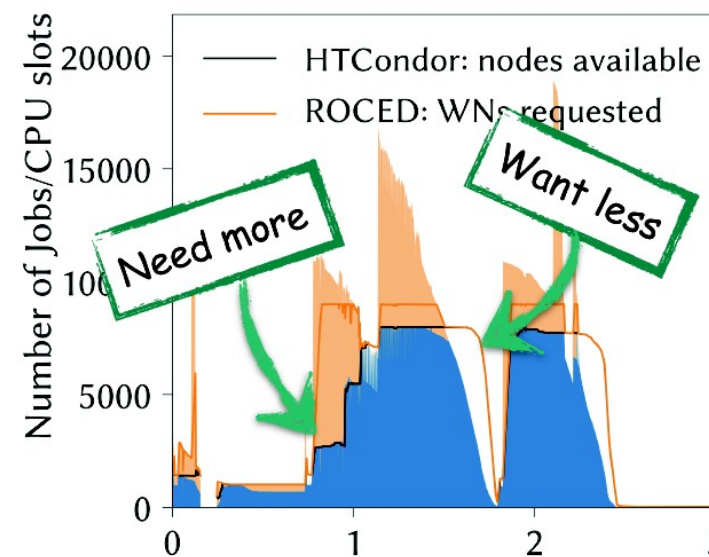
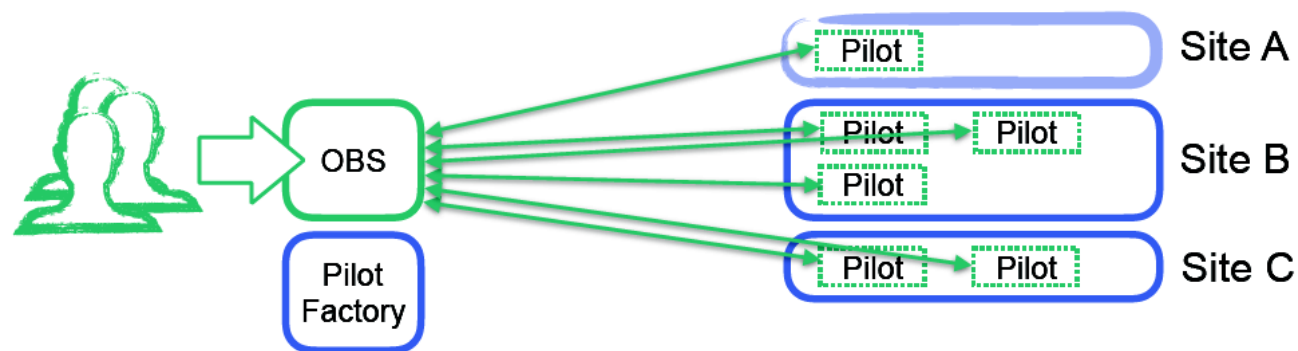
- Application and test of virtualized software components in the environment of heterogeneous computing resources

<p>B1: Tests of Components Implementation and test on different platforms</p> <ul style="list-style-type: none">• Storage and caching solutions• Virtualized services (databases, monitoring, accounting)	<p>B2: Job and Resource Management Job distribution and monitoring in a heterogeneous computing resource environment using container technologies</p>
<p>B3: Virtualization of User Jobs</p> <ul style="list-style-type: none">• Requirement capture• Determination and creation of run time environment• Creation of container and meta data	<p>B4: Combined Tests Test of complete system on different platforms regarding</p> <ul style="list-style-type: none">• Installation and maintenance• Performance• Scalability• Robustness

Example of a Common Solution

COBaID / TARDIS

- Overlay Batch System (OBS)
- Adjustment of allocated resources to demands



COBaID/TARDIS helps to fight COVID-19

April 14, 2020

The project [COBaID/TARDIS](#) developed at the Karlsruhe Institute of Technology (KIT) and funded within the scope of ErUM-Data IDT is supporting the fight against COVID-19. COBaID/TARDIS, otherwise used to dynamically provision and transparently integrate opportunistic resources for High Energy Physics, is currently acting as a job factory and submission infrastructure for work-units of the [Folding@Home](#) and [Rosetta@Home](#) projects at the GridKa Tier 1 centre at KIT. [[SCC News](#)]

- **Deep Learning, Gain of knowledge by substantiated data-driven methods**

C1: Processing of Sensor Data <ul style="list-style-type: none">• Signal filter, noise suppression• Processing of time dependent data	C2: Object Reconstruction <ul style="list-style-type: none">• Track and cluster reconstruction, jet forming, event reconstruction• Questions of placement, order, assignment of data• Extraction of small signals in case of large backgrounds
C3: Network Accelerated Simulations <ul style="list-style-type: none">• Generative adversarial networks, adjustment of simulation to data• Methods for the evaluation of the quality of network simulations	C4: Quality of Network Predictions <ul style="list-style-type: none">• Reduction of experimental systematic uncertainties• Special learning strategies• Prediction relevant information• Uncertainty of predictions

Example of a Common Solution

- Python module that makes a variety of datasets from the involved science fields easily available for the development of machine learning algorithms

```
import erum_data_data as edd

# loading training data into RAM (downloads dataset first time)
X_train, y_train = edd.load('top', dataset='train', cache_dir = './', cache_subdir = 'datasets')
```

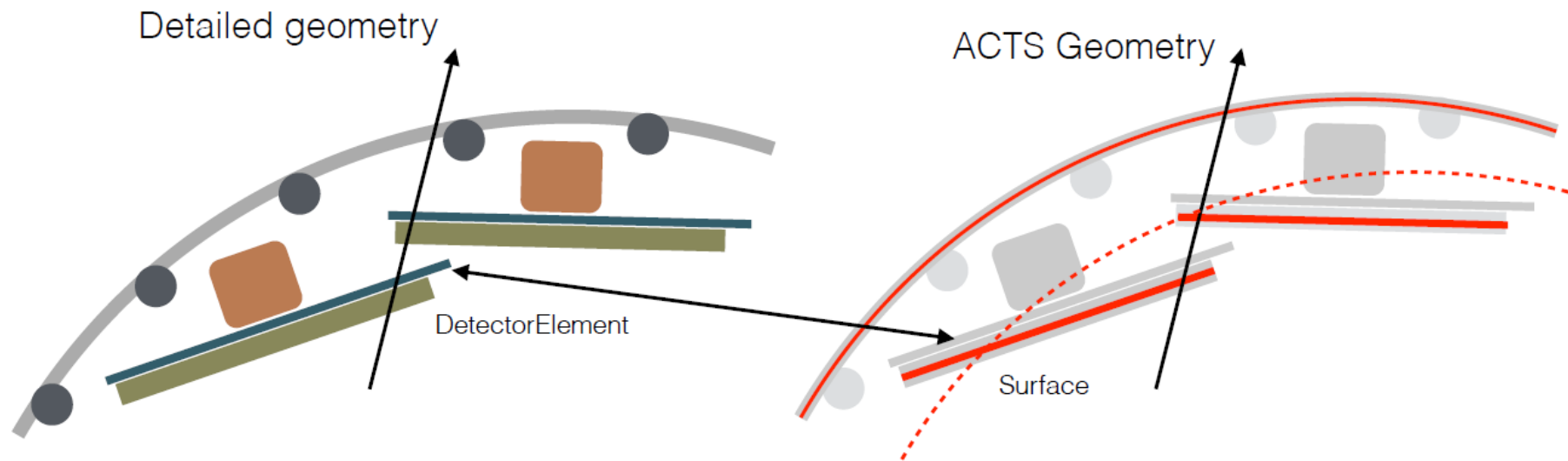
1. Top Tagging at the LHC [link](#), Publication: 1902.09914
2. Spinodal or not? [link](#), Publication: 1906:06562
3. EOSL or EOSQ [link](#), Publication: 1910.11530
4. Cosmic Airshower [link](#)
5. LHC Olympics 2020 (Unsupervised anomaly detection) [link](#)
6. SmartBKG dataset (Belle II - generated events passing downstream selection) [link](#)

- Event reconstruction: Cost- and energy-efficient use of computing resources

D1: Track Finding <ul style="list-style-type: none">• Alternative algorithms, e.g. cellular automata• Alternative architectures, e.g. GPUs	D2: Parameter Determination <ul style="list-style-type: none">• Connection of GenFit2-ACTS

Example of a Common Solution

- ▶ A Common Tracking Software
- ACTS core, framework, simulation, data



IDT-UM Further Information



Home Partners Research Areas Talks Logo

Innovative Digital Technologies for Research on Universe and Matter

Progress in fundamental research on universe and matter (ErUM) is made by studying structures at smaller and smaller scales. The high resolution of modern instruments in particle, hadron and nuclear, and astroparticle physics results in huge amounts of research data, at the order of millions of terabytes. And the next generation of experiments will increase the dataset sizes even more, exceeding the growth expected from advances in storage technologies.

- Web page: <https://www.erum-data-idt.de/>
- Mailing list: computing-verbund@lists.lrz.de
- [erum-data-idt organization on github](#)
- **Two collaboration meetings per year**

Computing Strategy Workshop

- **Workshop on Computing Strategy in the HL-LHC Era, May 2020**
- <https://indico.physik.uni-muenchen.de/event/33/>
- ➔ **Document:**
<https://indico.physik.uni-muenchen.de/event/33/attachments/142/242/Abschlussdokument.pdf>

Zusammenfassung der Strategiediskussion zum Computing in der HL-LHC-Ära

Präambel

Am 6. und 7. Mai 2020 wurden strategische Fragen zum Computing in Deutschland für die kommenden 10 Jahre auf einem offenen Workshop von Wissenschaftlern aus den Gebieten Teilchen-, Hadronen- und Kern-, Astroteilchenphysik und Astronomie sowie den beteiligten Rechenzentren diskutiert. Die Agenda mit den Vorträgen und Fragen ist auf der Indico-Seite <https://indico.physik.uni-muenchen.de/event/33/> zu finden. Dieses Dokument fasst die Ergebnisse der Diskussion zusammen.