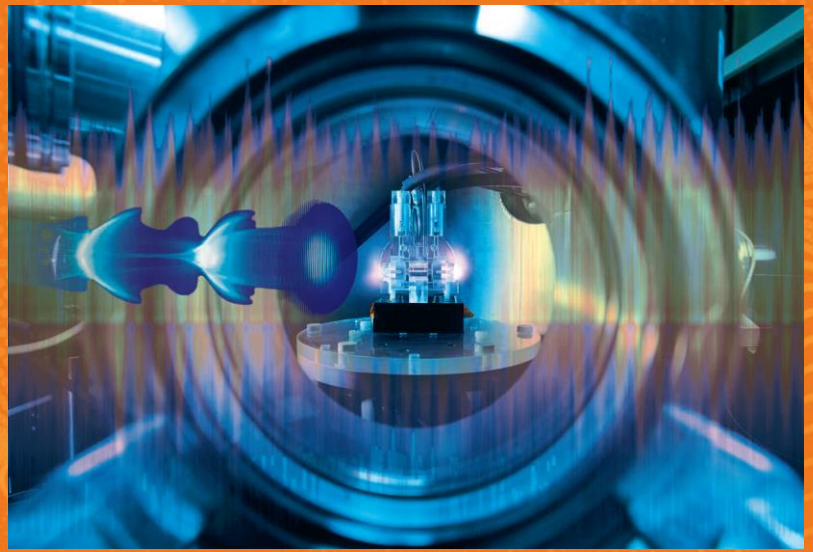
HELMHOLTZ

RESEARCH FOR
GRAND CHALLENGES

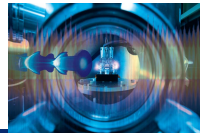
DMA ST 2 – The Digital Scientific Method

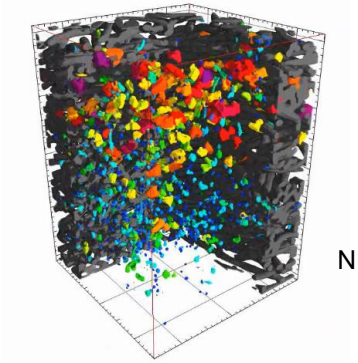
M. Al-Turany (GSI)
G. Juckeland. (HZDR)



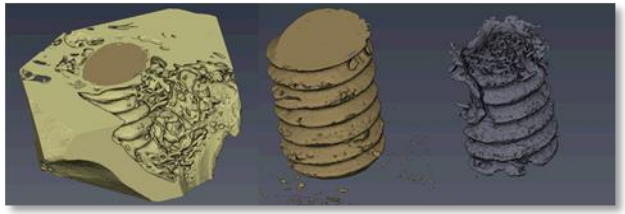
DESY. JÜLICH Forschungszentrum GSI HI JENA HELMHOLTZ Helmholtz Institute Jena HIM HELMHOLTZ Helmholtz-Institut Mainz HZB Helmholtz Zentrum Berlin HZDR HELMHOLTZ ZENTRUM DRESDEN ROSSENDORF Helmholtz-Zentrum Geesthacht Zentrum für Material- und Küstenforschung IPP KIT Karlsruhe Institute of Technology

I. Focus – ST 2 / The Digital Scientific Method

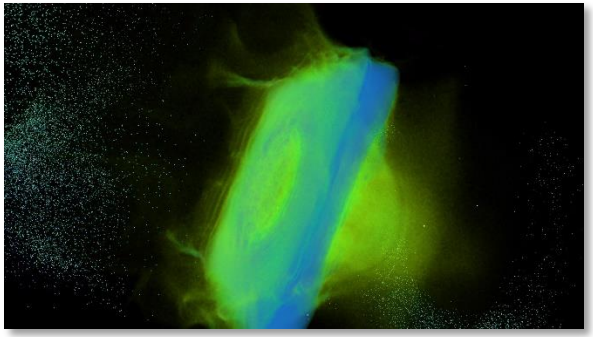




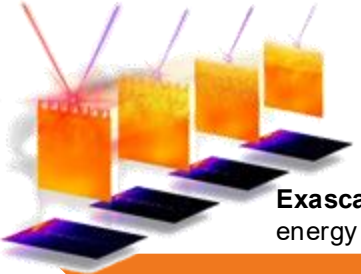
Near real-time **segmentation** of battery electrode data by **AI**



Near real-time **segmentation** of bone implant data by **AI**



Exascale simulations of laser-driven ion acceleration



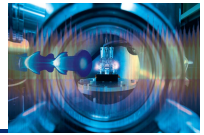
Exascale simulations of high energy density plasmas

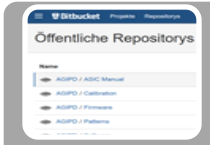
Develop, apply and share **cutting edge digital methods and frontier technologies** for research in Matter.

- Artificial Intelligence
- Exascale Computing
- High Throughput Computing
- Quantum Computing
- Near real-time analysis

2023: DMA Open Solutions Toolbox → 2025: Near real-time capabilities → 2027: Surrogate modelling

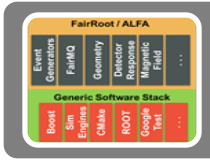
Milestones:





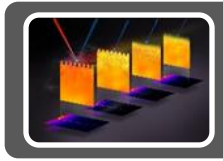
2023

- DMA repository of interconnectable, modular software in full operation



2025

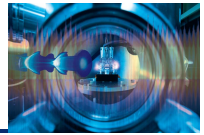
- Toolbox for near-realtime data analysis at extreme scales available



2027

- Surrogate models of multi-source, multi-modal experiments

Cooperation with EOSC



2023

- DMA repository of interconnectable, modular software in full operation

02/2019 - 08/2022

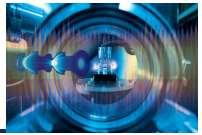
ESCAPE

European Science Cluster of Astronomy & Particle physics ESFRI research infrastructures

ESCAPE OSSR | Open-source Scientific Software and Service Repository

A sustainable open-access repository to share scientific software and services to the science community and enable open science

The DMA Repository Implementation



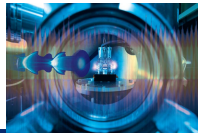
- **ESCAPE-OSSR** provides a strong curation workflow
  (https://zenodo.org/communities/escape2020/)

- **HIFIS Research Software Directory** offers a flexible platform
  (https://helmholtz.software)

- Current strategy: Combining workflows from both platforms to maximize effectiveness

- RSD + Matter content is "not yet rich" but serves as foundation for Milestone DMA

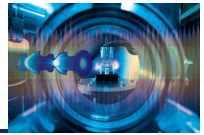
Current Status (2023 Milestone Achieved)



- The DMA repository of interconnectable, modular software is now in full operation
- The repository provides a directory of software packages with examples covering the whole simulation and experiment life cycle
- Basic requirements established for software inclusion:
  - Mature state following good software engineering practices
  - Publicly available with Open-source license
  - Introductory documentation
  - Proper attribution in AUTHORS and CONTRIBUTORS files

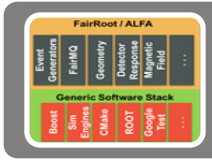
See Onboarding DMA Software to ESCAPE-OSSR Catalogue
https://indico.desy.de/event/33132/contributions/129472/

Milestones:



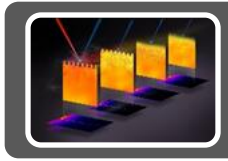
2023

- DMA repository of interconnectable, modular software in full operation



2025

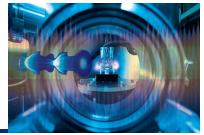
- Toolbox for near-realtime data analysis at extreme scales available

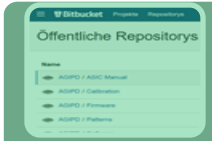


2027

- Surrogate models of multi-source, multi-modal experiments

Milestones:

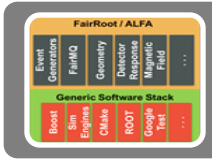




2023

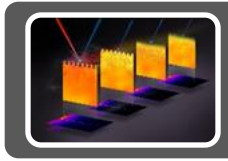
- DMA repository of interconnectable, modular software in full operation

Done



2025

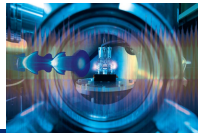
- Toolbox for near-realtime data analysis at extreme scales available



2027

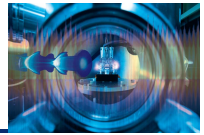
- Surrogate models of multi-source, multi-modal experiments

Challenges

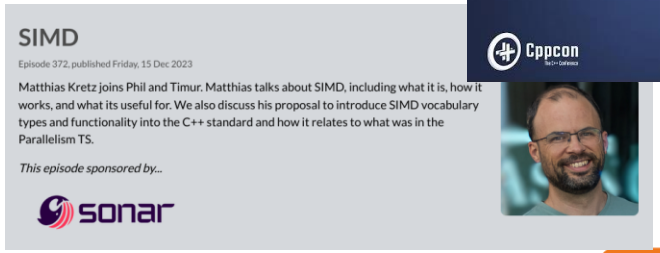


- Content for the repository is "not yet rich" and requires more onboarding
  - Need to identify and approach relevant groups for software contributions
  - Requires regular meetings for effective onboarding and collaboration

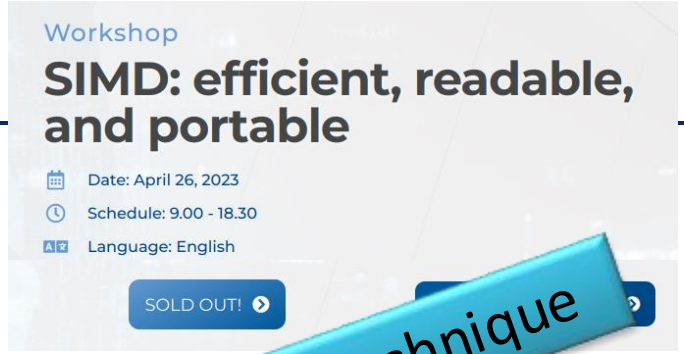
Data-parallel Types for C++

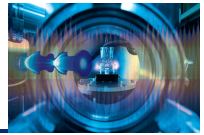


- GCC implementation fully usable and documented on cppreference.com
- Tutorial on std::experimental::simd at “using std::cpp 2023”
- Design approval of std::simd for C++26
- Collaboration with Intel on adding more features to simd in C++26
- CppCon 2023 presentation
- CppCast interview



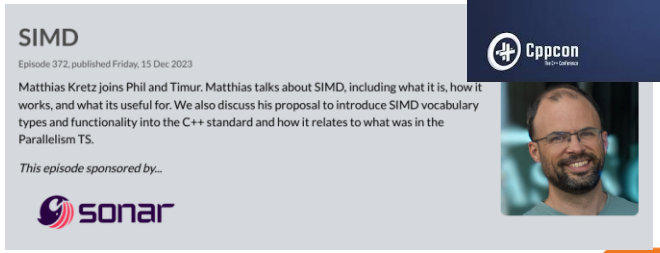
Data-parallel Types for C++



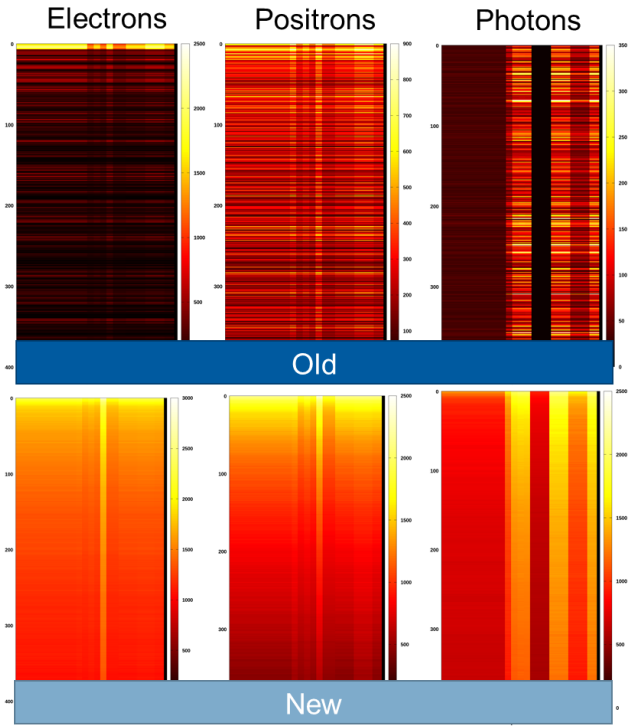


- GCC implementation fully usable and documented on cppreference.com
- Tutorial on std::experimental::simd at "…pp 2023"
- Design approval of std::simd for C++26
- Collaboration with Intel on …atures to simd in C++26
- CppCon 2023 …
- CppCast interview

hardware-level vectorization technique



Common solutions for experiments & simulations



alpaka

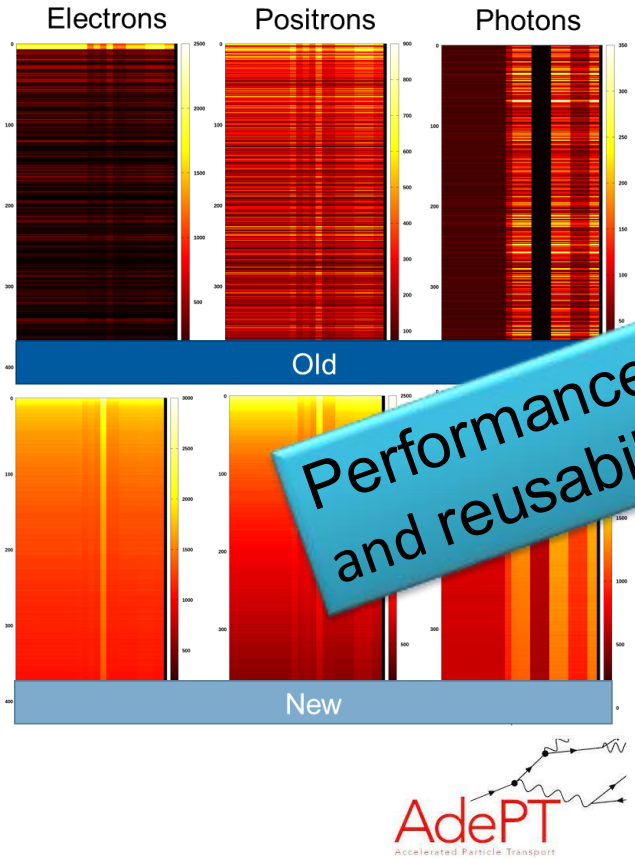
AMA

open PMD

Exascale software stack for scalable data analytics & high performance computing

- Heterogeneous computing with CPUs, GPUs, FPGAs
- F.A.I.R. I/O, data streaming and open data formats
- True cross-community use: HEP, Photon Science, Neutron Science, Astrophysics, Plasma Science, Life science

Common solutions for experiments & simulations



alpaka

Performance portability, data consistency, and reusability of the software

openPMD

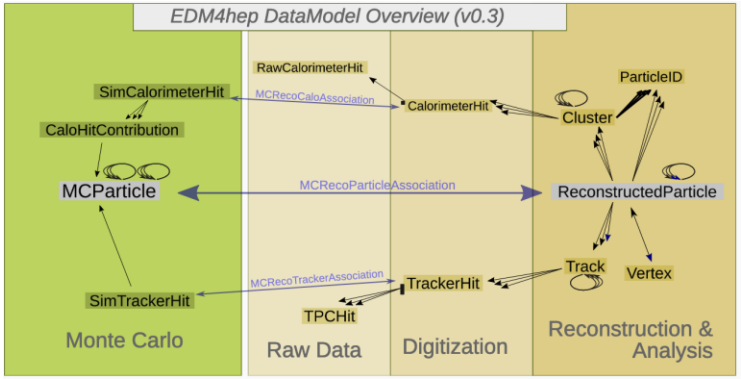
…are stack for …nalytics & high …ance computing

- Heterogeneous computing with CPUs, GPUs, FPGAs
- F.A.I.R. I/O, data streaming and open data formats
- True cross-community use: HEP, Photon Science, Neutron Science, Astrophysics, Plasma Science, Life science

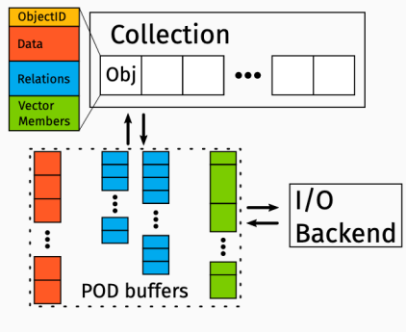
EDM4HEP: AN EVENT DATA MODEL FOR FUTURE COLLIDERS

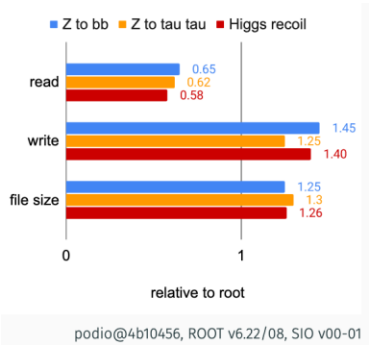
- event data models (EDM) are at the core of every HEP event processing framework
  - they define the interface for any algorithm and analysis
- **EDM4hep** aims at providing the EDM for all future collider projects in the context of the **Key4hep** project
  - used by FCC-ee, FCC-hh, ILC, CLICdp, CEPC,...
- developed at DESY (**DMA-ST2**) and CERN
  - also supported by AIDAinnova

For more details see talk at vCHEP 2021
T.Madlener: **EDM4hep and podio - The event data model of the Key4hep project and its implementation**
https://indico.cern.ch/event/948465/contributions/4323705



- **EDM4hep** uses the **PODIO** event data model toolkit
- define data entities in yaml files
- autogenerated C++ and Python code (using *Jinja*)
- data stored as **POD**s (**plain old data objects)**
- several backends exist: ROOT, SIO, HDF5 (WIP)
- -> choose what's best for your use case





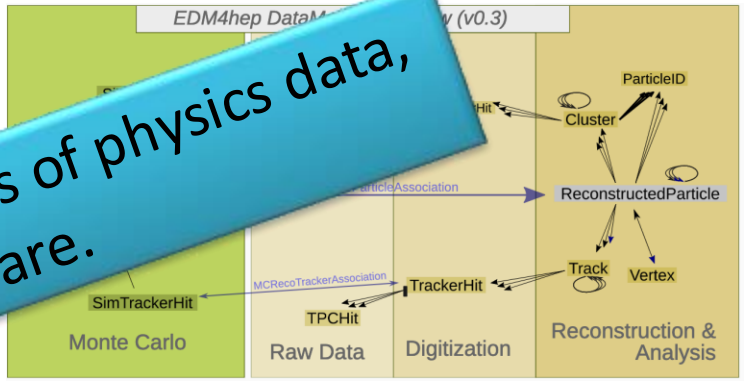
- **EDM4hep** event data model heavily inspired by LCIO
- EDM used in the linear collider community for almost 20 years
- modernize the syntax
- improve the performance

EDM4HEP: AN EVENT DATA MODEL FOR FUTURE COLLIDERS

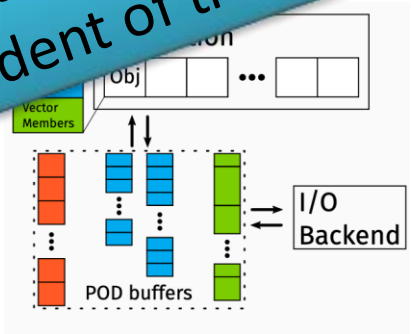
- event data models (EDM) are at the core of every HEP event processing framework
  - they define the interface for any algorithm and analysis
- **EDM4hep** aims at providing the EDM for all future collider projects in the context of the **Key4hep** project
  - used by FCC-ee, FCC-hh, ILC, CLICdp, CEPC,...
- developed at DESY (**DMA-ST2**) and CERN
  - also supported by AIDAinnova

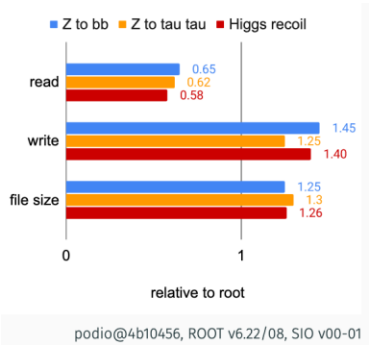
For more details see talk at vCHEP 2021
T.Madlener: **EDM4hep and podio - The ev… and its implementation**
https://indico.cern.ch/event/…

Defining the structure and semantics of physics data, independent of transport or hardware.



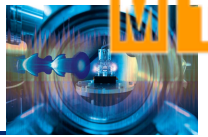
- **EDM4hep** us… event data m…
- define data en…
- autogenerated … code (using *Jinja*…
- data stored as **PODs** (**plain old data objects)**
- several backends exist: ROOT, SIO, HDF5 (WIP)
- -> choose what's best for your use case

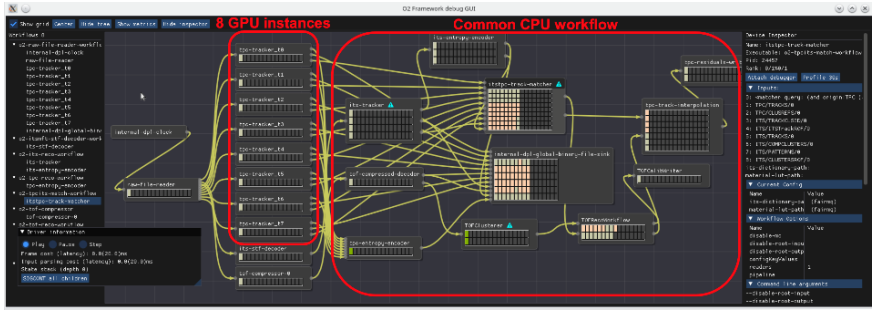


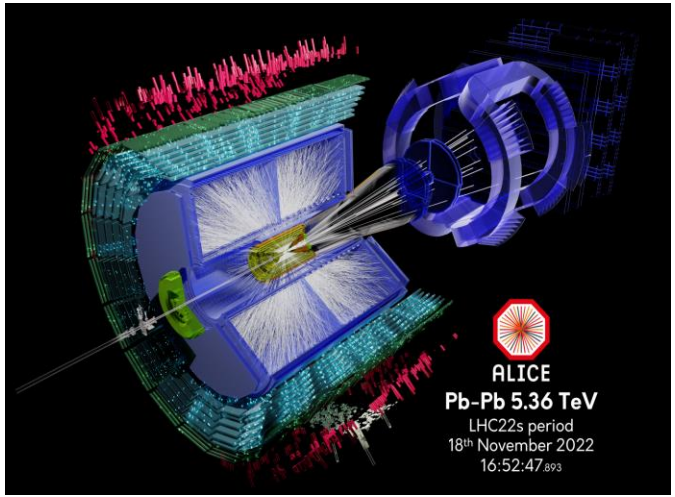


- **EDM4hep** event data model heavily inspired by LCIO
- EDM used in the linear collider community for almost 20 years
- modernize the syntax
- improve the performance

Integrating HPC into online experimental workflows







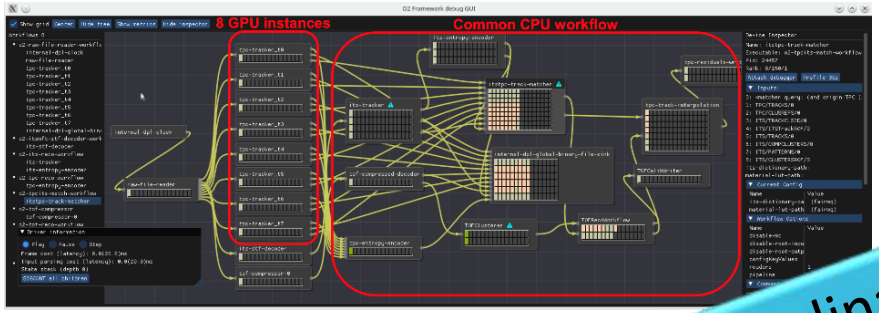
Alice O² with FAIRMQ + DDS + ODC

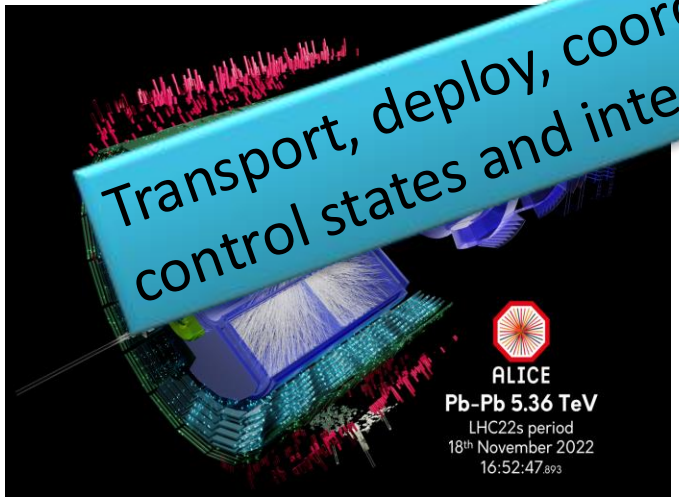
Modern heterogeneous compute systems are complex to use and provide dedicated resources for a variety of analysis and simulation workflows.

Optimum mapping of compute resources to the needs of experiments now requires to make use of large-scale high performance compute resources.

Breaking down complex workflows and dynamically distributing them on heterogeneous HPC systems is a key prerequisite to optimally use these systems.

Integrating HPC into online experimental workflows





Alice O² with FAIRMQ

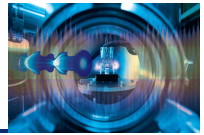
Modern [illegible] ems are [illegible] dedicated resources for [illegible]nd simulation workflows.

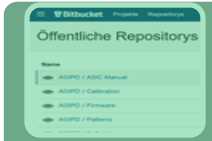
[illegible]am mapping of compute resources to the needs of experiments now requires to make use of large-scale high performance compute resources.

Breaking down complex workflows and dynamically distributing them on heterogeneous HPC systems is a key prerequisite to optimally use these systems.

Transport, deploy, coordinate data and processes, control states and integration with the experiment.

Milestones:

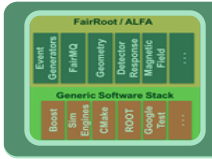




2023

- DMA repository of interconnectable, modular software in full operation

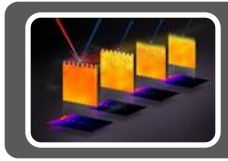
Done



2025

- Toolbox for near-realtime data analysis at extreme scales available

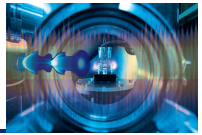
Done



2027

- Surrogate models of multi-source, multi-modal experiments

Next Steps (POF IV)



- Launch the bi-weekly meeting schedule immediately
- Finalize the current software list for DMA repository
- Expand collaboration with other relevant initiatives (NFDI, EOSC, ..)
- Continue developing the toolbox for near-realtime data analysis with specific use cases
- Prepare for 2027 milestone planning: "Surrogate models of multi-source, multi-modal experiments"