



The Modular Supercomputer Architecture and its application in HPC and HPDA

Damian Alvarez

Jülich Supercomputing Centre, JSC (Germany)



The DEEP Projects

Research & innovation projects co-funded by the European Union

- DEEP (2012–2015): Cluster/Booster concept
- DEEP-ER (2013–2017): Scalable I/O and resiliency
- DEEP-EST (2017–2020): Generalized modular supercomputing architecture, support for data analytics and ML

Co-design leading to novel, highly efficient, heterogeneous HPC architectures

One of only two Exascale project series funded

Total EC funding: 30 M€



Motivation – One Size does not fit All

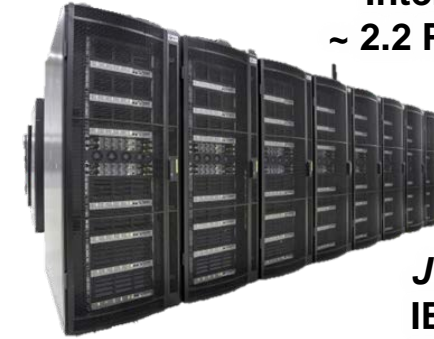
HPC systems come in two very different flavours

- General purpose Clusters with
 - *High flexibility & reliable performance*
 - *Preferred by many applications since “good enough” performance is easy to achieve*
 - *Relatively high power consumption*
- Dedicated, highly scalable HPC systems (MPPs)
 - *Highest degree of parallelism, specialized fabrics*
 - *Few (highly parallelizable) codes can fully exploit them*
 - *Highly energy efficient*

Can one combine the best of these two worlds into a single system?

Yes! → Exploit system-level heterogeneity!

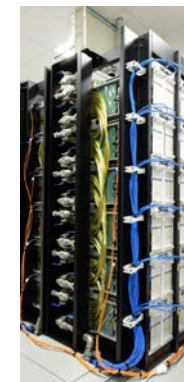
JSC JURECA Cluster
Intel® Xeon® Haswell
~ 2.2 PFlop/s in ~1 MW



JUQUEEN
IBM Blue Gene/Q
5.9 PFlop/s in 2.3 MW

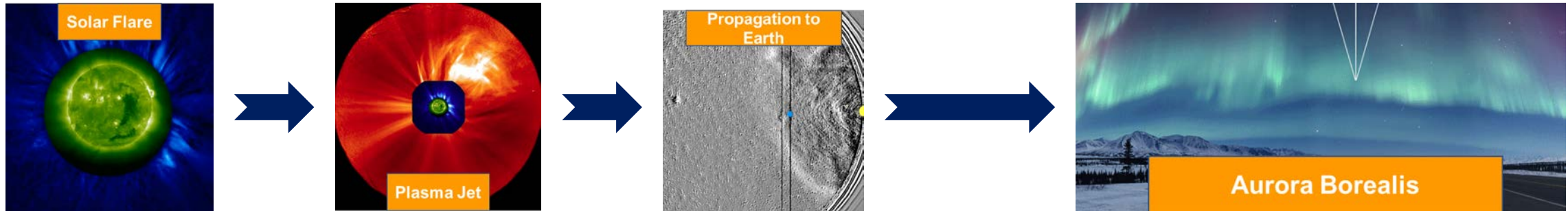


DEEP Prototype
Intel Xeon + Intel Xeon Phi
0.6 PFlop/s in 0.15 MW



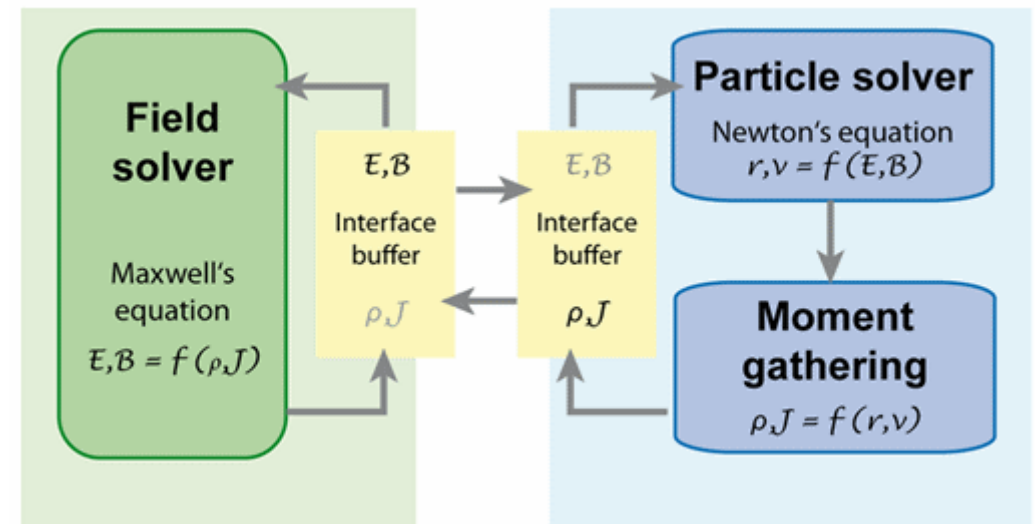
Motivation – Application View

Space Weather simulation (xPiC) from KU Leuven

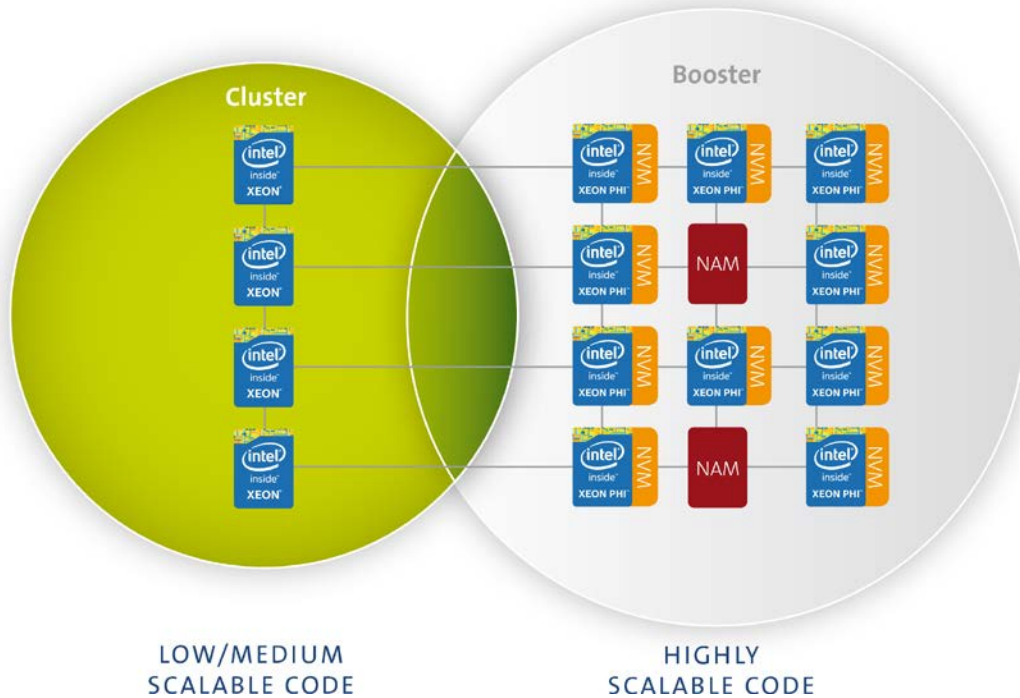


Algorithm iterates between computing magnetic field and particle movement

- Field equations run best on Intel® Xeon® CPUs
- Particle solver & moment computation runs best on Intel® Xeon Phi™

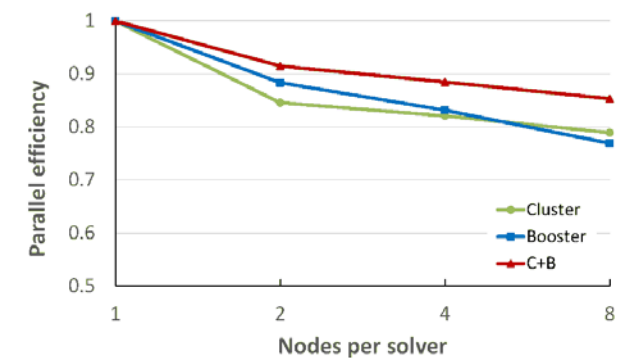
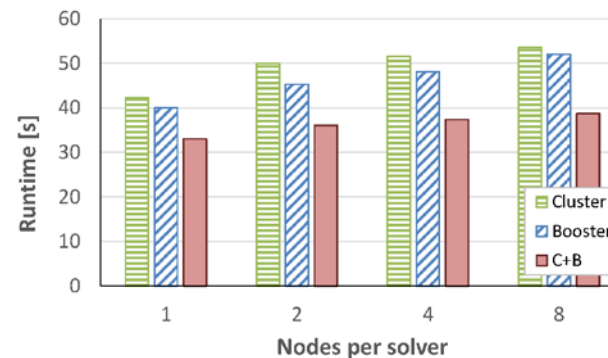


Cluster Booster Architecture & Results

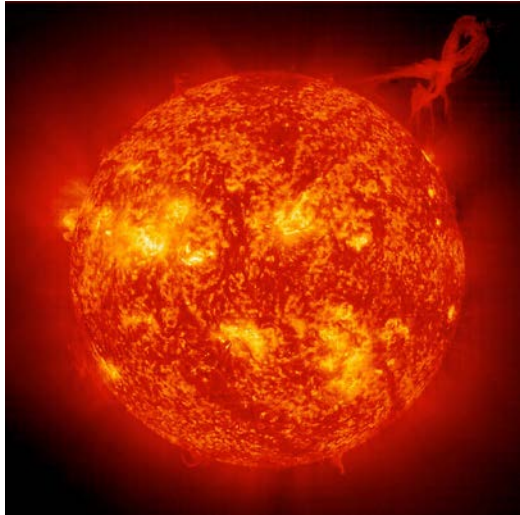


Implemented & validated in DEEP and DEEP-ER

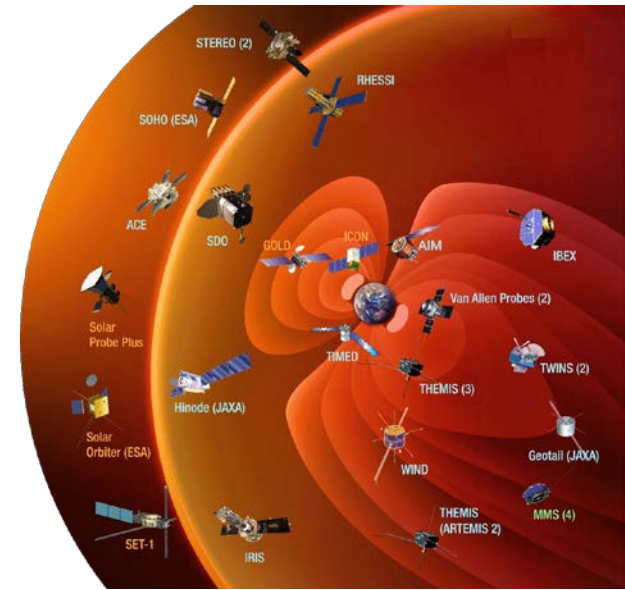
For xPiC, see approx. 35% gain in performance for combined Cluster and Booster system compared to same size homogenous system



Combining HPC and ML Steps in a Workflow



Observation satellites orbiting L1 point



Data Analytics

Forecast solar wind conditions at L1 from remote image analysis of the Sun

DL analysis triggers full simulation

Space Weather Prediction

Detailed physics simulations of the Earth environment given the solar wind conditions at L1



Modular Supercomputing Architecture

DEEP-EST prototype includes 3 compute modules

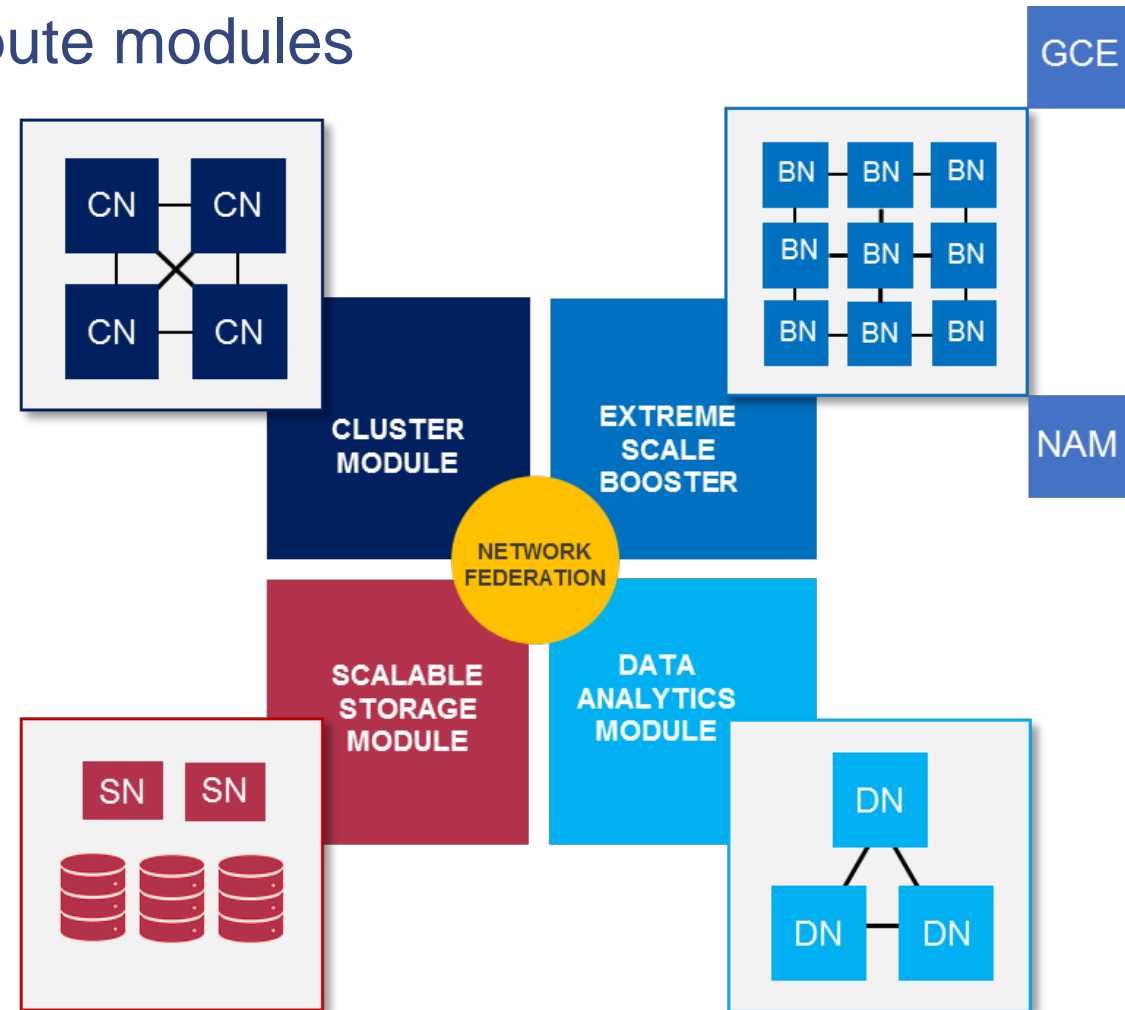
- Cluster Module
- Extreme Scale Booster
- Data Analytics Module

Storage module handles warm workflow data

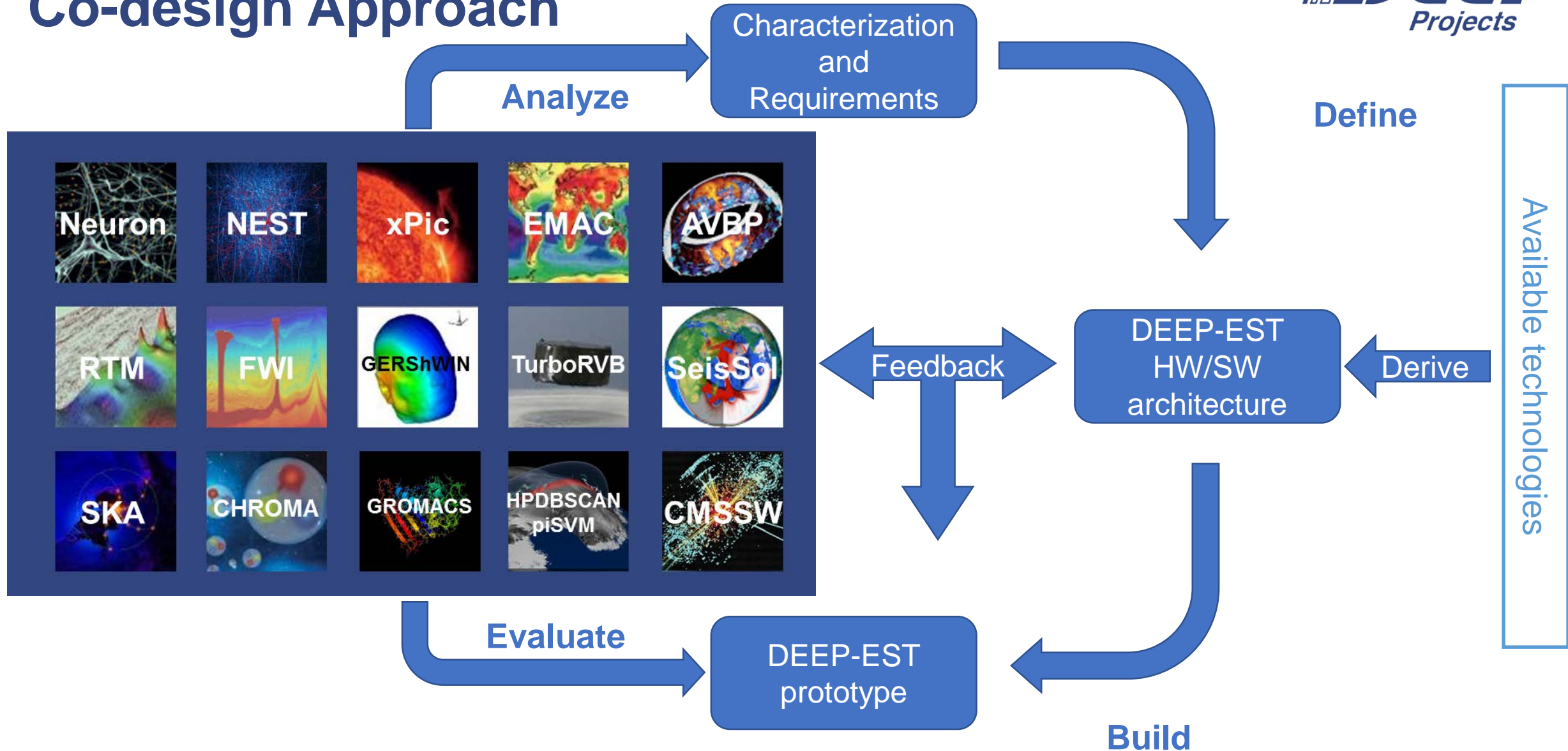
Local storage and the network attached memory (NAM) for hot application data

Global communication engine (GCE) accelerates MPI collective operations

Network federation glues all parts together



Co-design Approach

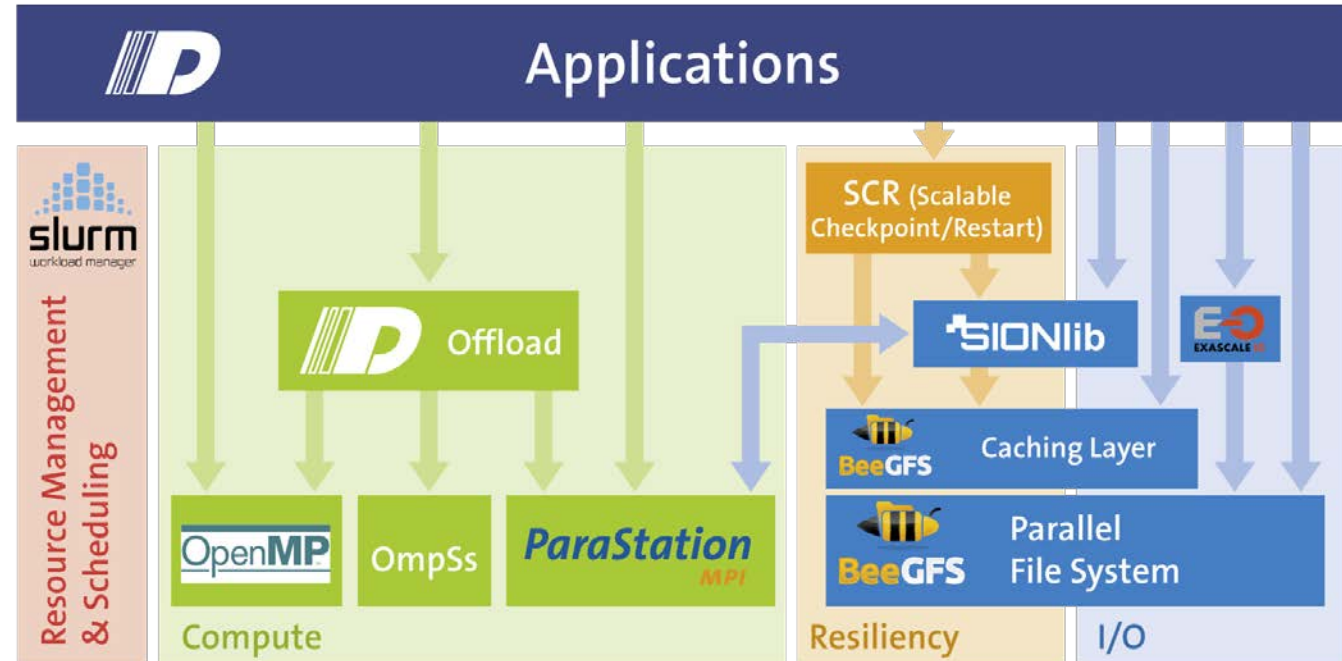


DEEP-EST Software Stack

Provides standard APIs to applications

- MPI, OpenMP, OmpSs (tasking)
- Scalable Posix I/O and MPI-I/O
- SCR for checkpointing
- Deep Learning frameworks
- SLURM as resource manager & scheduler

SLURM extensions will support workloads spread across modules



European System Technology in DEEP-EST

EXTOLL TOURMALET network and Fabri³ nexus

MEGWARE packaging, cooling and power distribution

ParaStation Cluster Middleware



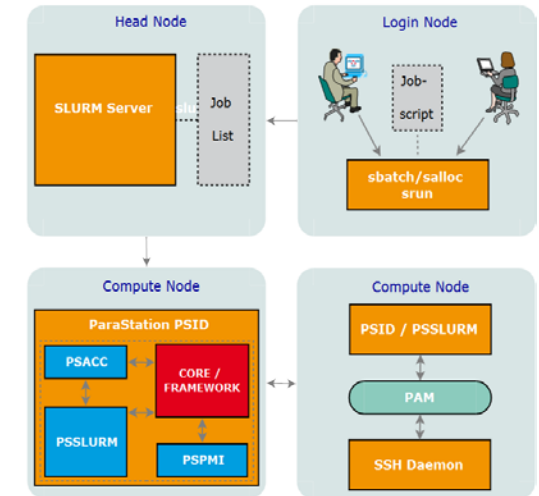
EXTOLL TOURMALET ASIC



MEGWARE SlideSX-LC[®]

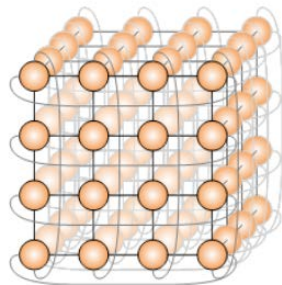


MEGWARE ColdCon[®]



ParaStation MPI and Resource Management

ParaStation
CLUSTERTOOLS



EXTOLL TOURMALET 3D-Torus



EXTOLL Fabri³



MEGWARE ClustSafe[®]



Summary

The DEEP journey

- Starts with the Cluster-Booster architecture
 - *Enables flexible association and use of heterogeneous resources*
 - *Accelerates heterogeneous applications*
 - *Extension of standard API*
- Continued with the DEEP-ER system
 - *Integrate local and network attached storage*
 - *Provide scalable I/O and check pointing*
- Next destination: Modular Supercomputer Architecture
 - *Generalization of the Cluster-Booster concept*
 - *Supports HPC, data analytics and machine learning workloads*
 - *First implementation to be ready mid 2019*

First production use of Cluster/Booster concept by JSC

- JURECA Cluster and JURECA Booster in 2018