

Into the DEEP

A Modular Supercomputing Architecture towards ExaScale

SC Asia 2019, Singapore
13 March 2019

HERBERT CORNELIUS
AXEL AUWETER

MEGWARE COMPUTER GMBH

www.megware.com

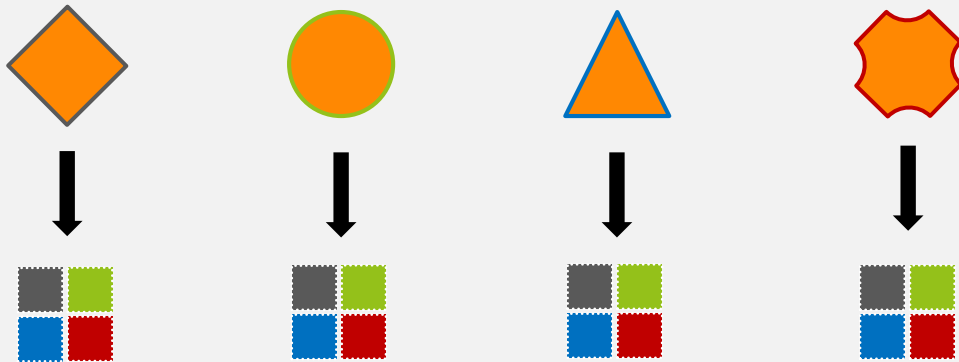
Modular Supercomputing Architecture

General Concept

The **Modular Supercomputing Architecture** (MSA) concept was invented by **Jülich Supercomputing Centre** (JSC) as a generalization of the previously implemented **Cluster-Booster*** concept.

Traditional Monolithic Supercomputing Architecture

APPLICATION & WORKLOAD MODULES/CHARACTERISTICS

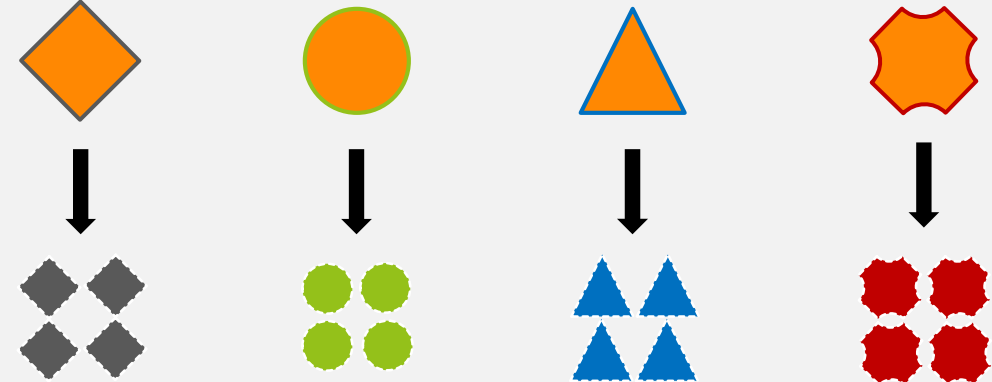


MONOLITHIC HARDWARE

SINGLE MODULE - WITH ALL NODES THE SAME

Modern Modular Supercomputing Architecture

APPLICATION & WORKLOAD MODULES/CHARACTERISTICS



MODULAR HARDWARE

MULTIPLE MODULES - EACH WITH TARGETED (DIFFERENT) NODES

*Eicker, N., Lippert, T., Moschny, T., Suarez, E., *The DEEP Project An alternative approach to heterogeneous cluster-computing in the many-core era*. Concurrency and Computation: Practice and Experience Vol. 28, Issue 8, 2394–2411 (2016).

Modular Architectures

Performance & Functionality

Macro-Level

MSA
(MODULES)

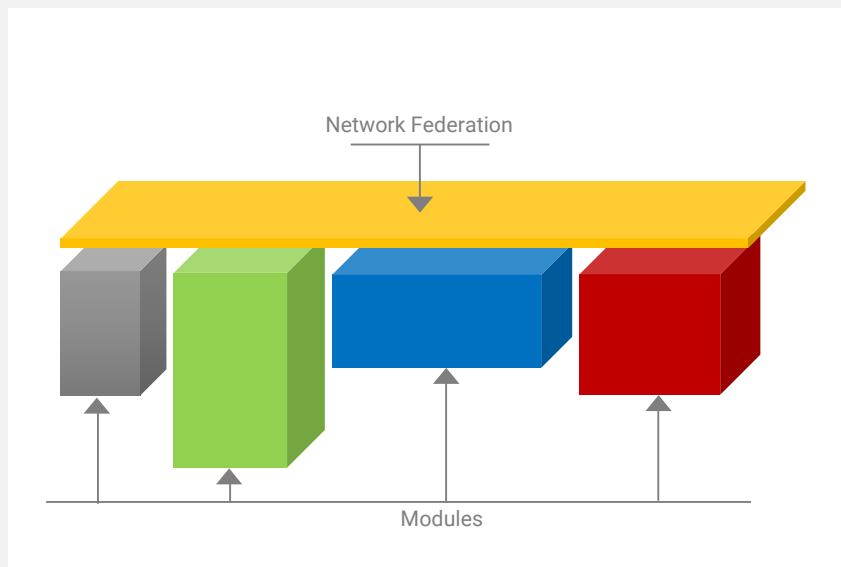


Image: Megware

Micro-Level

Hybrid
(CHIPLETS)

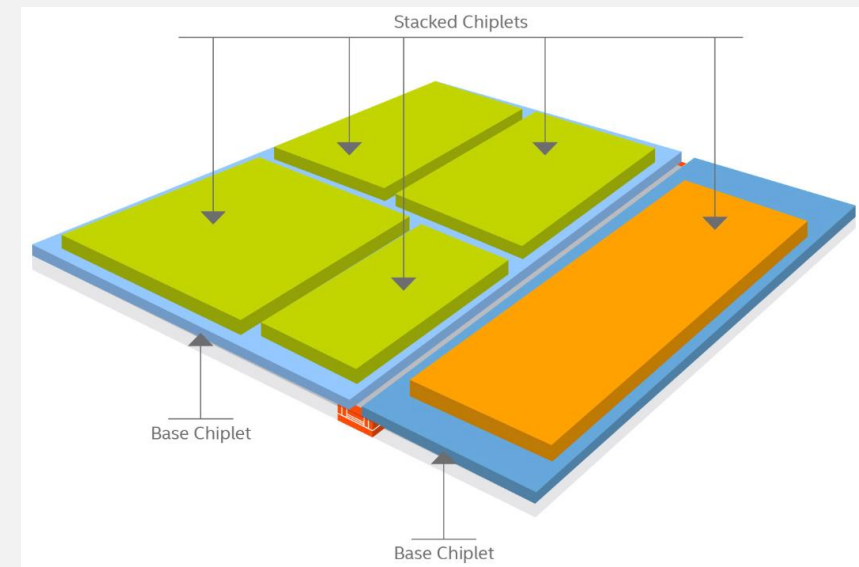
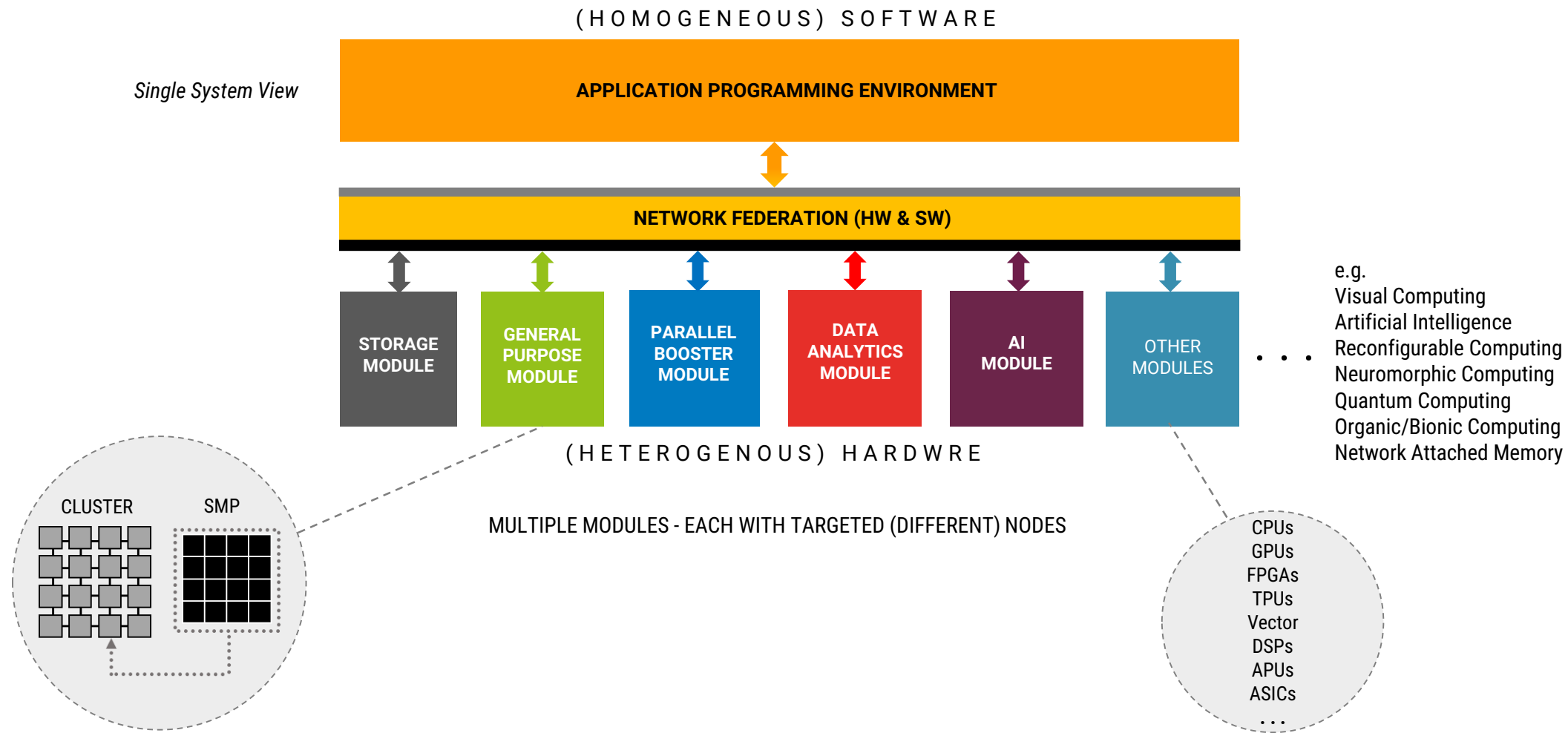


Image: Intel

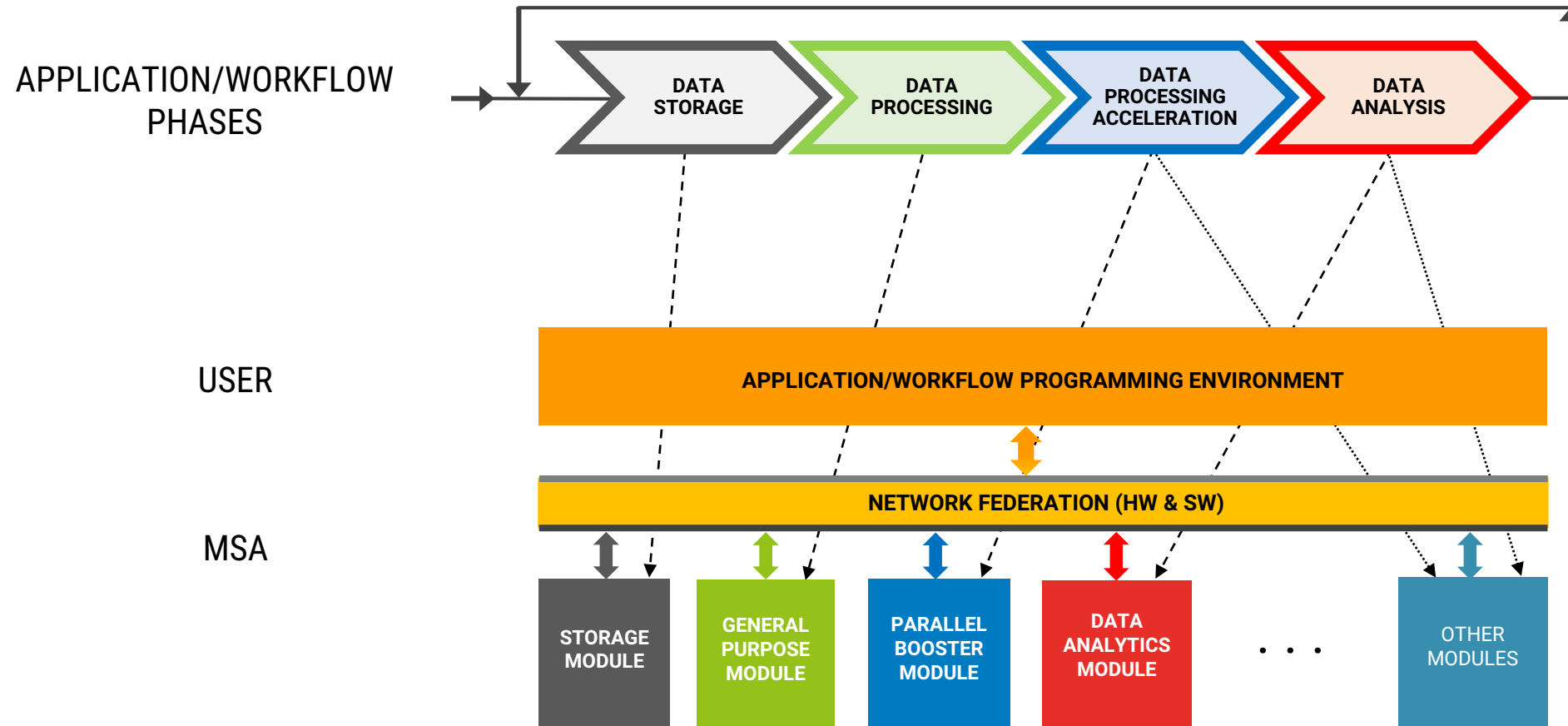
Modular Supercomputing Architecture (MSA)

High Level View



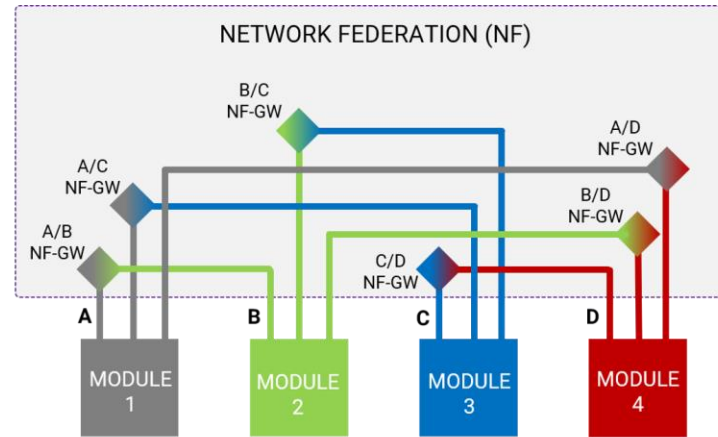
Modular Supercomputing Architecture (MSA)

Application & Workflow Mapping

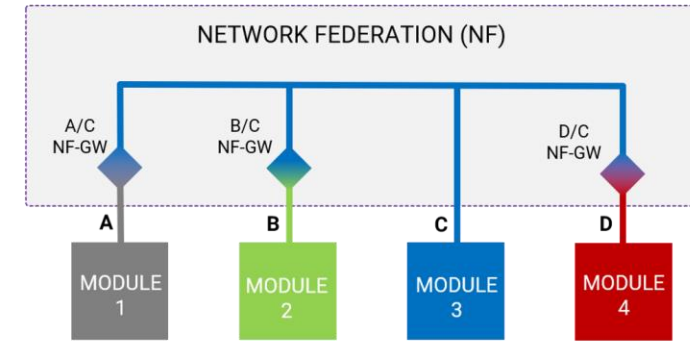


Network Federation (NF) Hardware – High Level

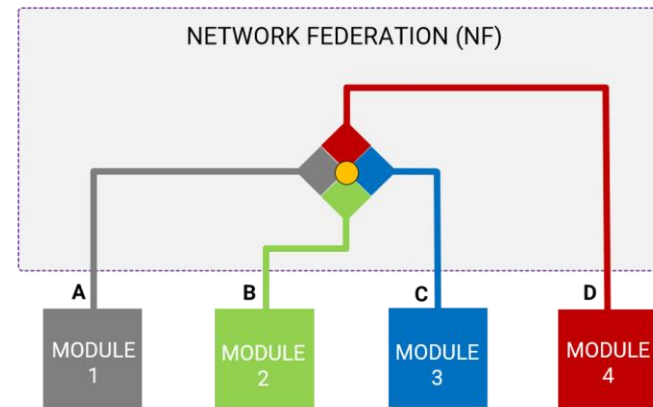
Some Examples



a. complex NF



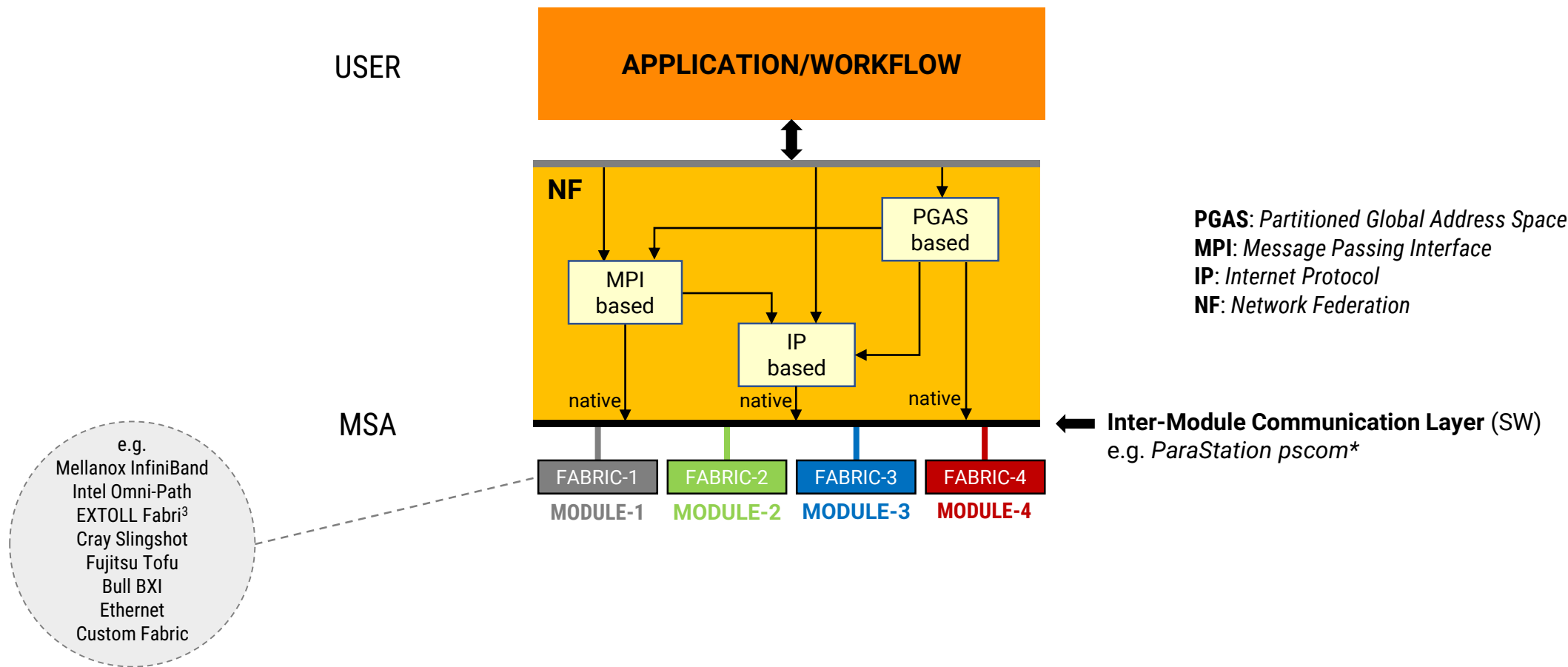
b. simplified NF



c. merged NF

Network Federation (NF) Software – High Level

Communication/Messaging

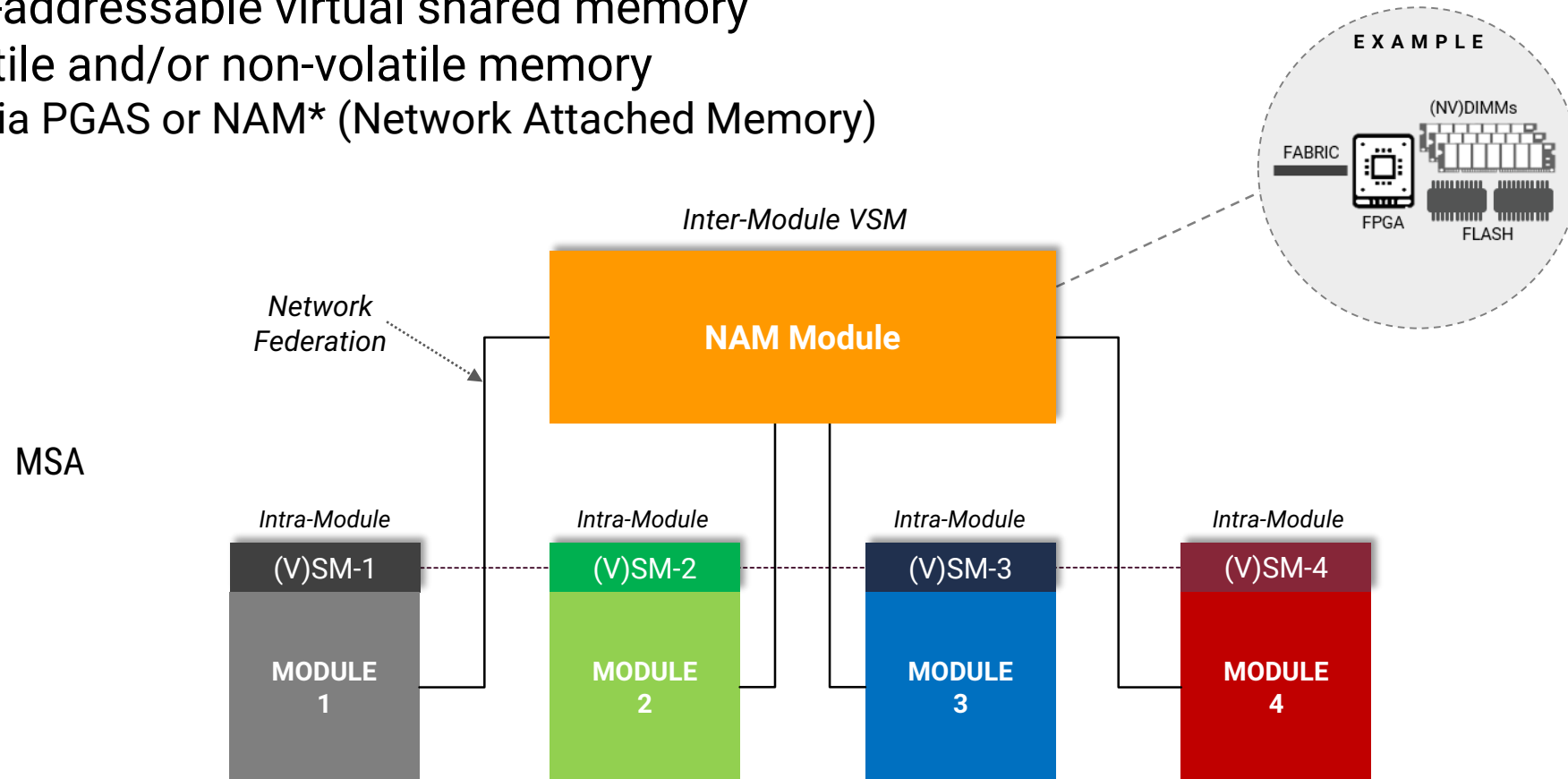


*Eicker, N., "ParaStation MPI", MPICH BoF SC16, www.deep-projects.eu/images/161115_ParaStationMPI.pdf

NF Virtual Shared Memory (VSM) Concept

Data-Centric Computing

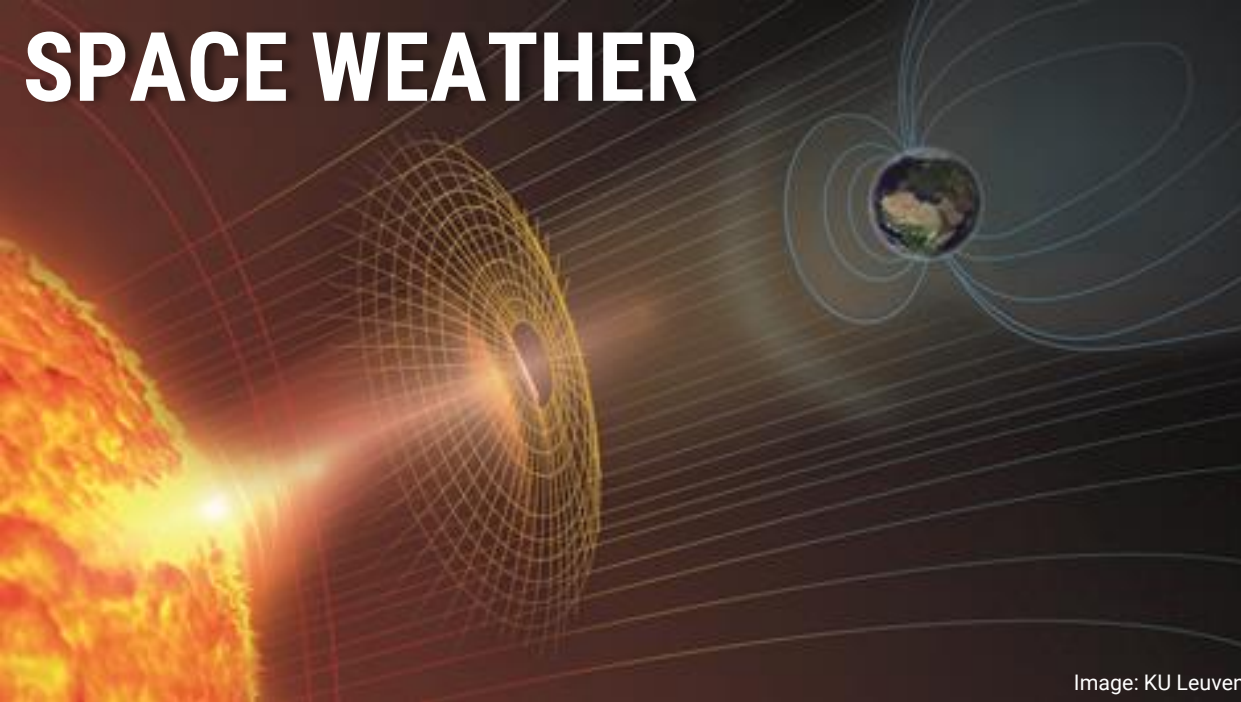
- Byte-addressable virtual shared memory
- Volatile and/or non-volatile memory
- e.g. via PGAS or NAM* (Network Attached Memory)



*Schmidt, J., "Accelerating Checkpoint/Restart Application Performance in Large-Scale Systems with Network Attached Memory", https://archiv.ub.uni-heidelberg.de/volltextserver/23800/1/dissertation_juri_schmidt_publish.pdf.

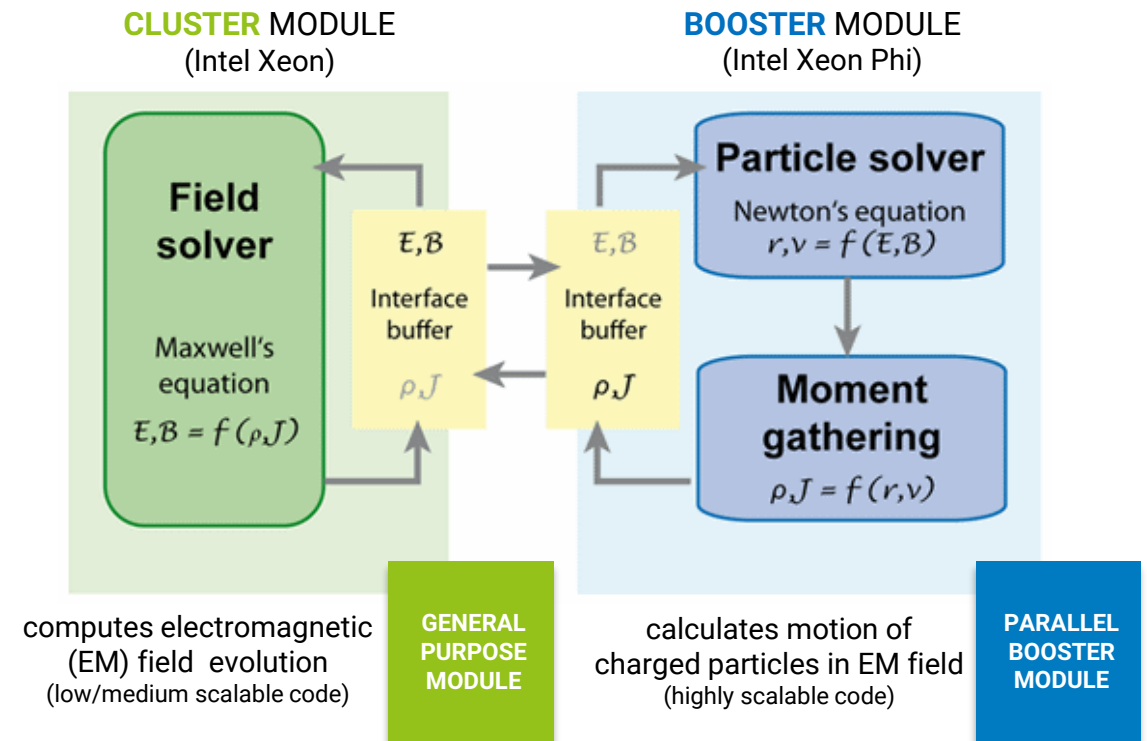
SPACE WEATHER

- Simulates plasma produced in solar eruptions and its interaction with the Earth magnetosphere
- Particle-in-Cell code: **xPIC**
- Author: **KU Leuven**, Belgium

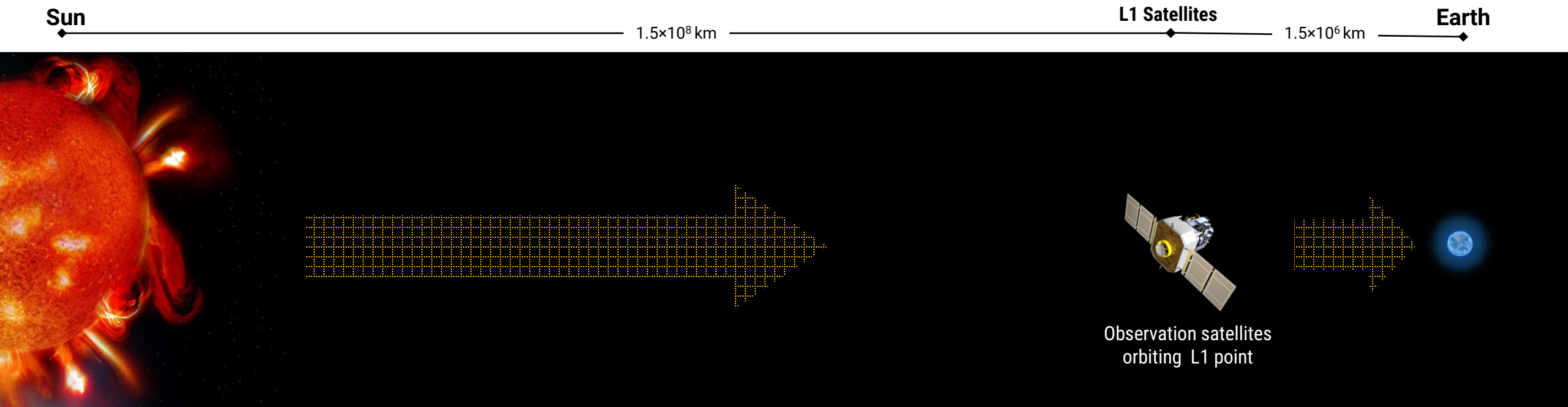


35% gain in performance* for
combined Cluster and Booster system
compared to same size homogenous system

*Kreuzer, A.; Eicker, N.; Amaya, J.; Suarez, E., "Application Performance on a Cluster-Booster System", 2018 IEEE IPDPS Workshops (IPDPSW), Vancouver, Canada, pp. 69 - 78 (2018) [10.1109/IPDPSW.2018.00019]



SPACE WEATHER on MSA (Jorge Amaya)



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



DL analysis triggers full simulation



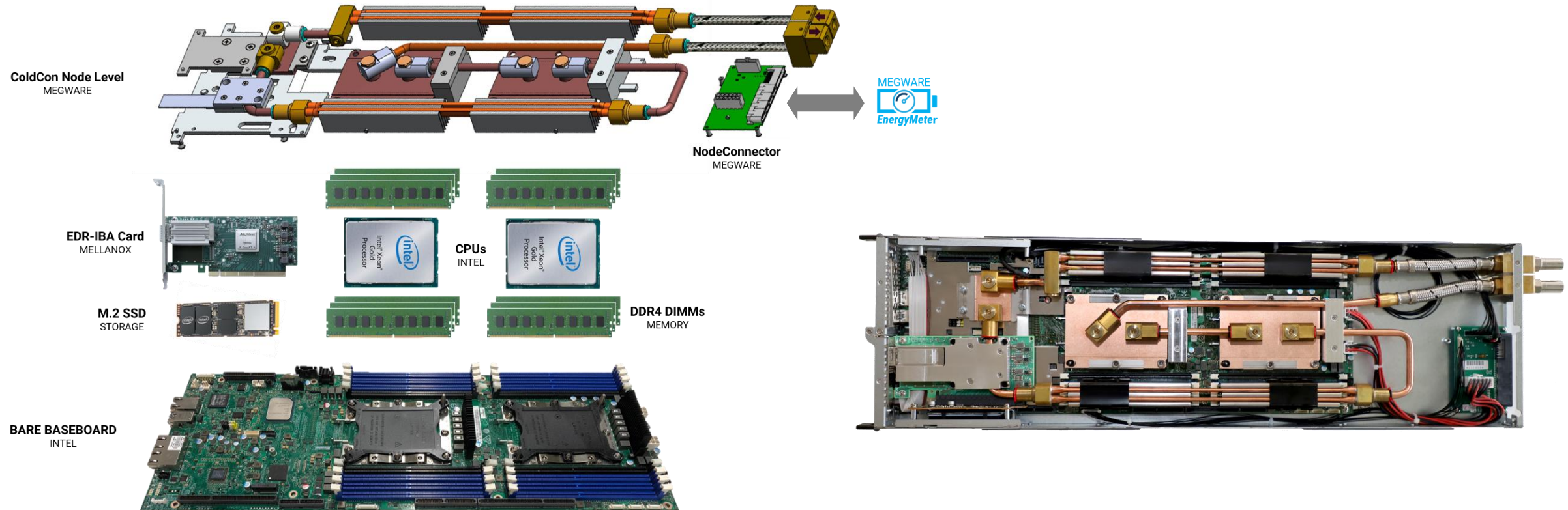
Space Weather Prediction
Detailed physics simulations of the Earth environment given the solar wind conditions at L1



Images: KU Leuven

SlideSX-LC® ColdCon® Direct Liquid Cooling

Example: Intel® Xeon® Scalable Processor 2S-Node



www.megware.com/en/products/slidesx.html

ColdCon® Direct Liquid Cooling Technology

CLUSTER NODES



FABRIC SWITCH



POWER SUPPLY



- 4th generation of direct liquid (warm-water) cooling
- Intel® Xeon® and Xeon® Phi™ Processor based cluster nodes
- Dual 100 Gbps links to Intel® Omni-Path fabric
- MEGWARE ColdCon® for nodes, power supplies and Intel® Omni-Path switches
- **Energy Efficient: 100% fanless**
- Thermal insulation with **less than 3% heat loss** to computer room air (full system)
- **Scalable** and **sustainable**

CoolMUC-3 Supercomputer

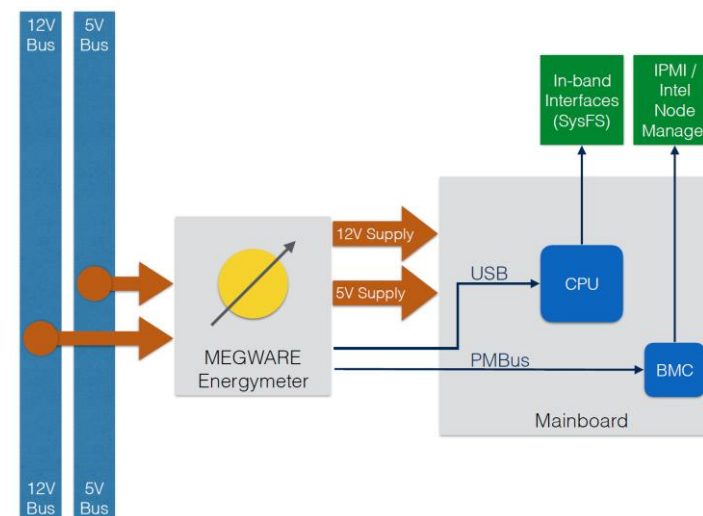
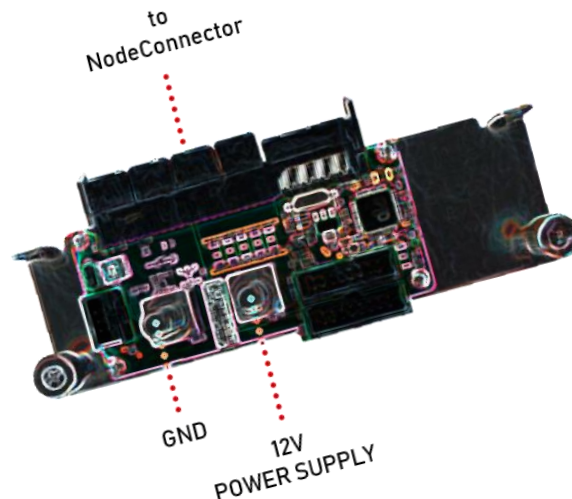


Leibniz Supercomputing Centre
of the Bavarian Academy of Sciences and Humanities



www.megware.com/en/products/coldcon.html

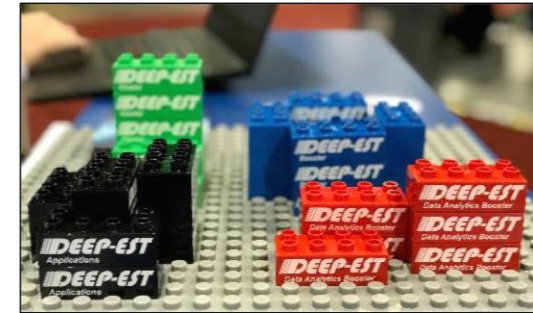
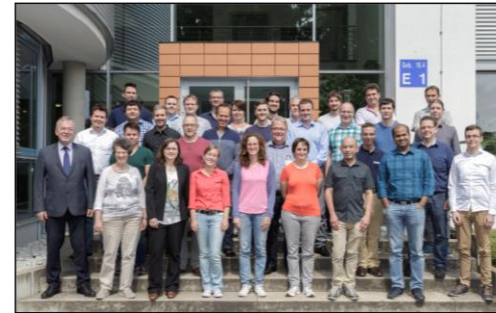
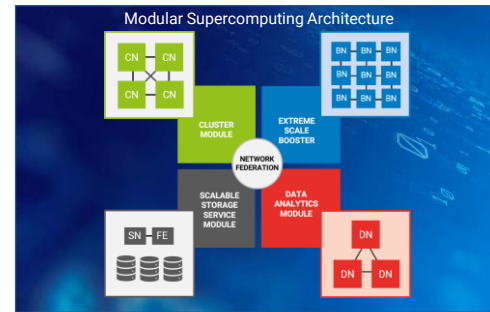
EnergyMeter



- Calibrated **high-accuracy, high-frequency** measurements for voltage, current, and 12V/5V power rails
- Fully **integrated power & energy metering** hardware device with no impact on application performance
- Data available **in-band** via CLI and Linux SysFS and **out-of-band** via IPMI
- Seamless integration into the NodeManager platform

www.deep-projects.eu

Dynamical Exascale Entry Platform



The DEEP projects DEEP, DEEP-ER and DEEP-EST have received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no ICT-610476 and no ICT-287530 as well as the Horizon2020 funding framework under grand agreement no. 754304.

Thank You.