



Modular Supercomputing

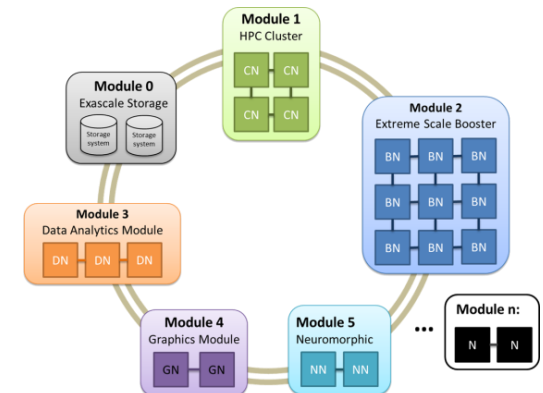
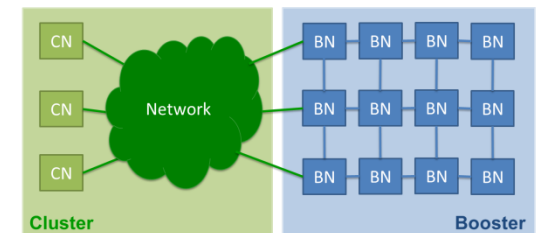
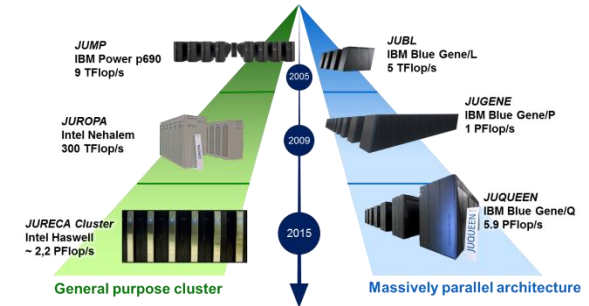
An overview of
computing architecture
evolution at JSC

Estela Suarez

Jülich Supercomputing Centre (JSC), Forschungszentrum Jülich

Outline

- Supercomputing evolution
 - Cluster computing
- Architectures at JSC
 - **Dual** architecture approach
 - **Cluster-Booster** architecture
 - *The DEEP projects*
 - **Modular Supercomputing** architecture
- JSC future vision

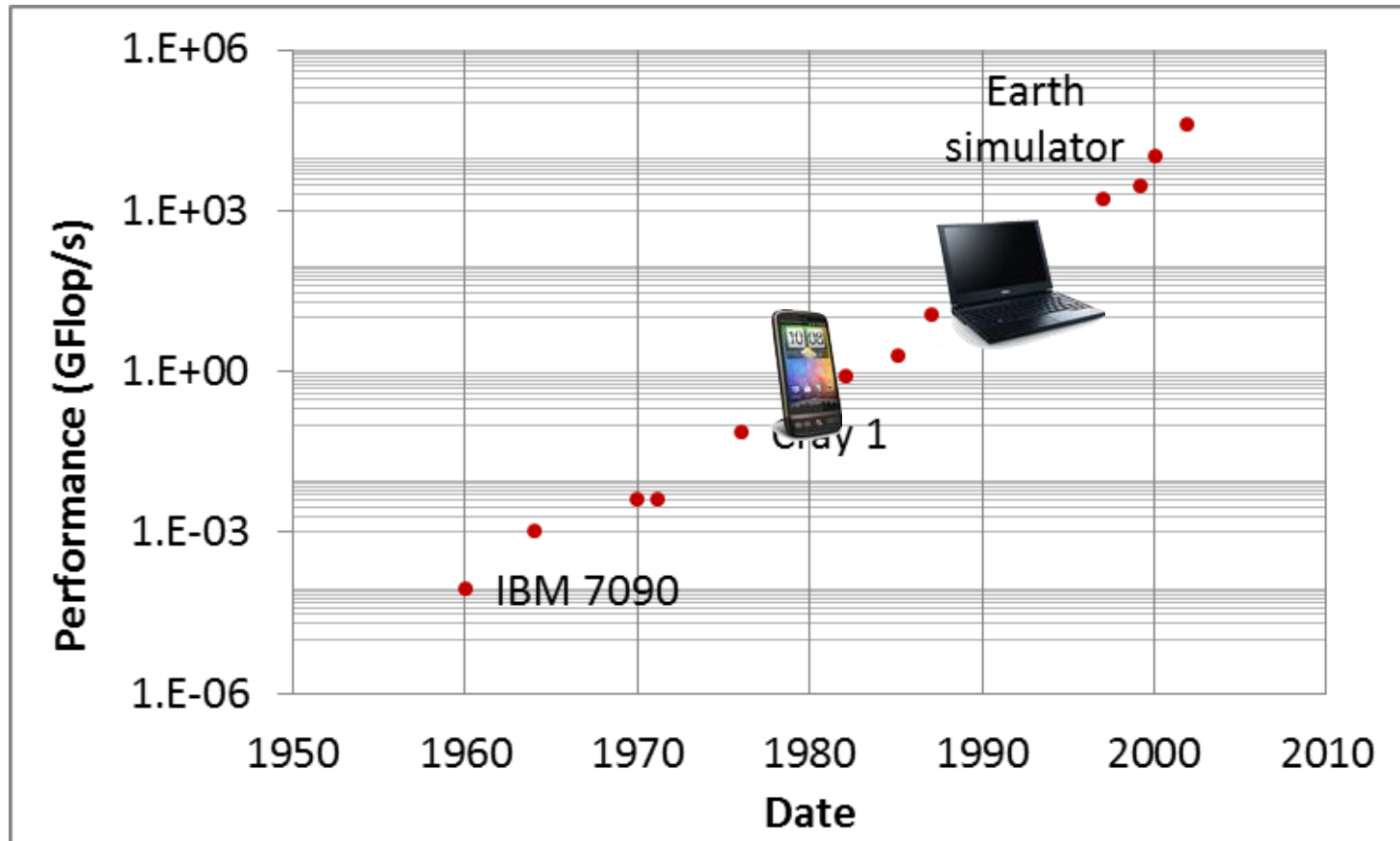




Architecture changes in time

SUPERCOMPUTING EVOLUTION

Supercomputing evolution – Performance



Moore's law

- Every 18-24 months the # transistors in microprocessors doubles

SC Evolution – Architecture paradigms

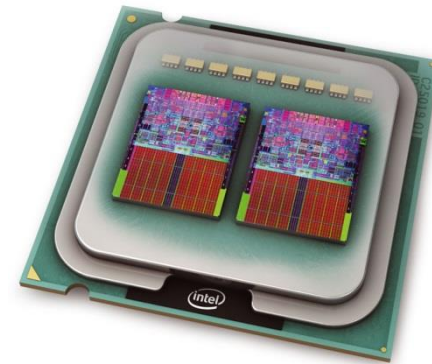
- **1940 – 1950:** first computers are Supercomputers
 - Specialized, very expensive
- **1960 – 1980:** general purpose computers appear
 - Still special machines needed to solve very complex problems
→ Supercomputing (High Performance Computing - HPC)
- 
 - Focus: floating operations (linear Algebra)
 - Special purpose technologies (fast vector processors, parallel architectures)
 - Only few machines produced
- **1990 – 2000:** integrate standard processors
 - Many „computers“ connected through fast network
 - Distributed memory → MPI
 - Both in proprietary Massively Parallel (MPP) and Cluster Computing
- **2010 – :** heterogeneous cluster systems
 - Use accelerator technologies (GPU, Xeon Phi)



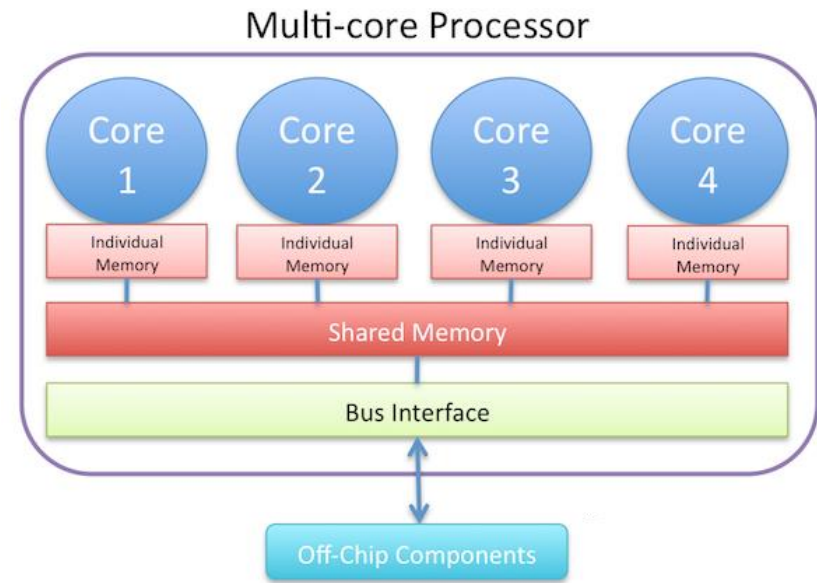
General purpose processors

Characteristics

- Broad range of capabilities
- Today always multi-core
 - Up to about 20 cores
- High single thread performance
 - High frequency
 - Out of order processing
- High memory per core
- Runs standard programming environment (MPI, OpenMP, etc.)
- Disadvantages
 - Limited energy efficiency
 - Larger \$/FLOP



e.g.



Accelerators

Many core (Intel Xeon Phi)

- 60 to 80 cores, 4 threads / core
- Rather low single thread performance
 - *low frequency*
- Energy efficient (\$/FLOP)
- x86 architecture → standard programming with MPI, OpenMP, etc.
- Can run autonomously (without host)

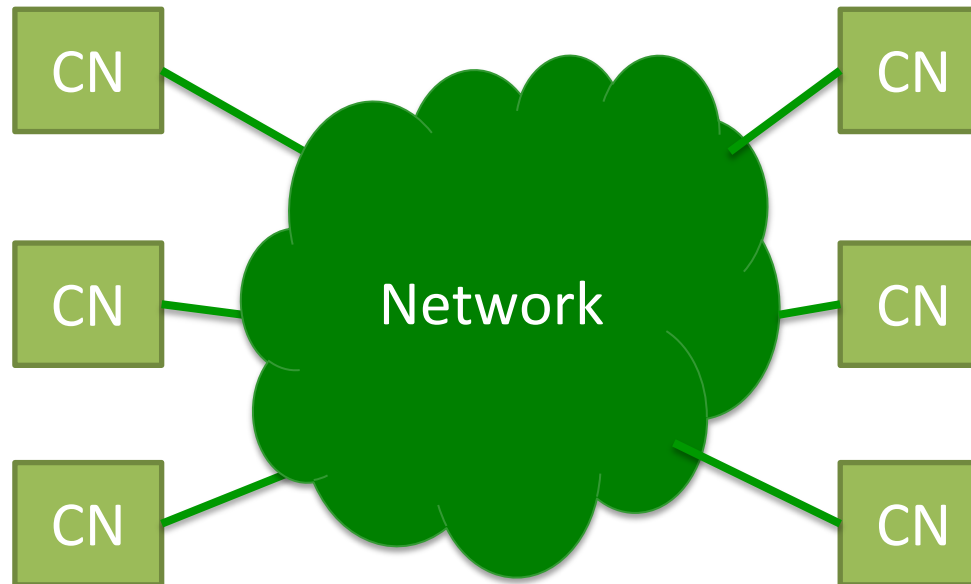


GPGPU (Graphic cards)

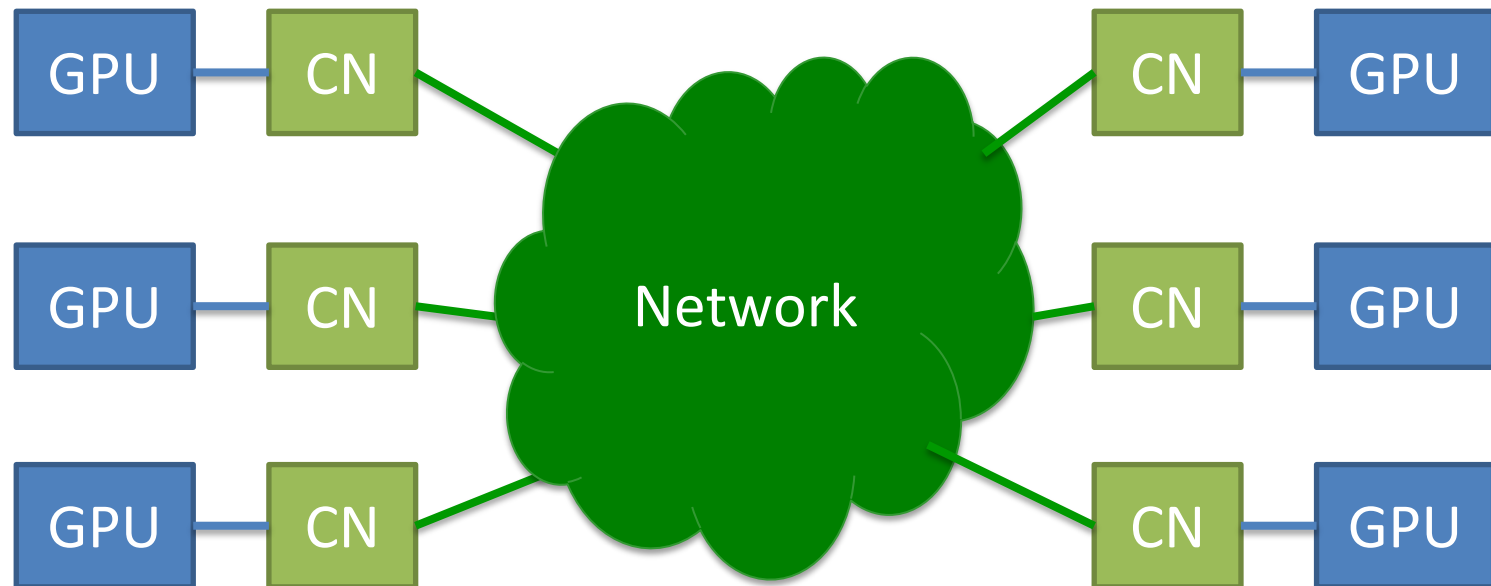
- Designed for graphics but evolved into general purpose
- Hundreds of (weak) computing cores
- Very energy efficient (\$/FLOP)
- Require specific programming models (CUDA, OpenCL)
- Require a host CPU



Homogeneous Cluster



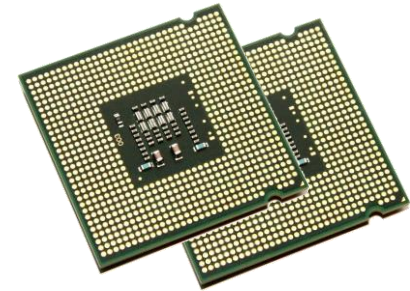
Heterogeneous Cluster



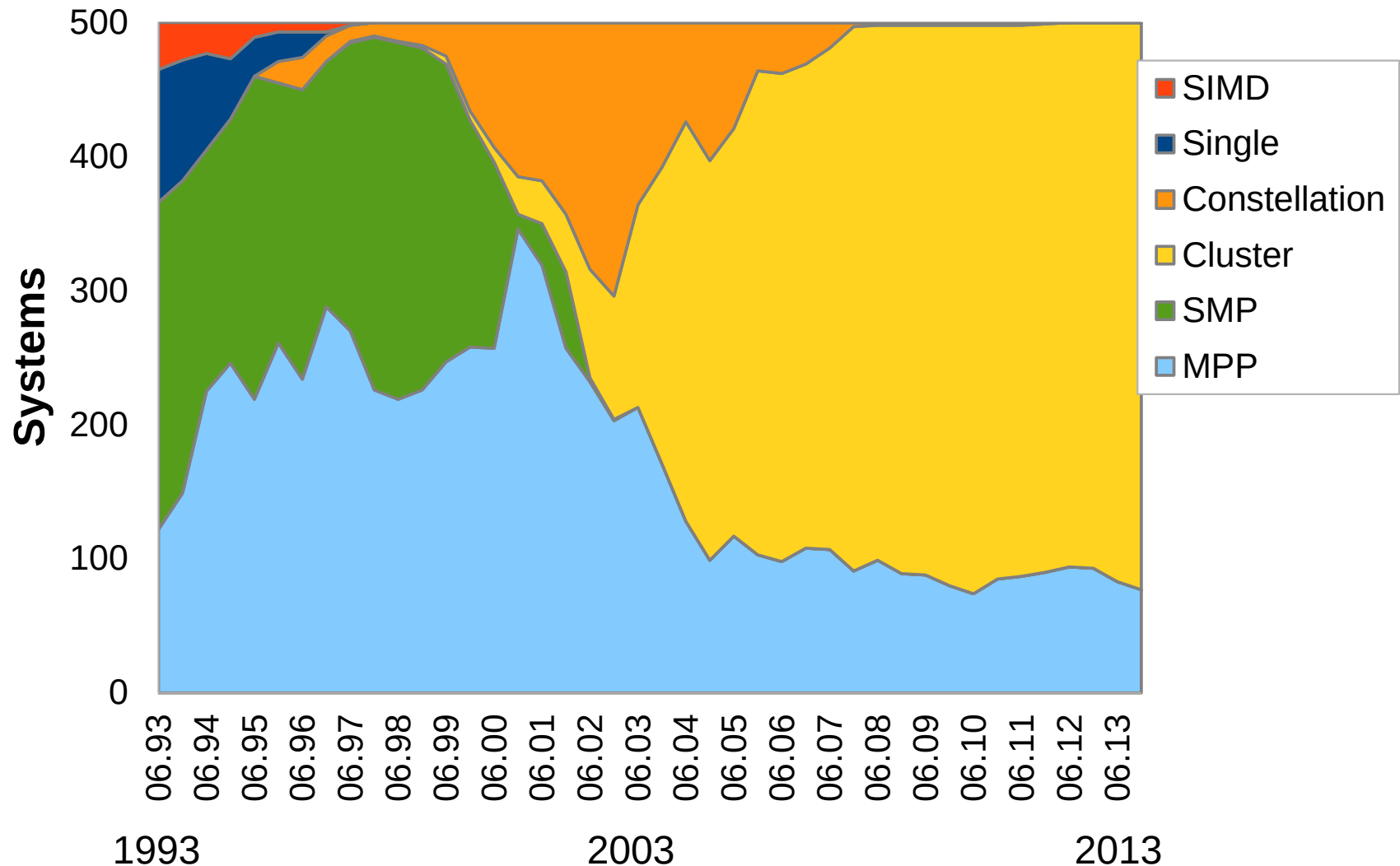
Cluster „anatomy“

Components selected/tuned to fit user needs

- **Heart:** Processor providing compute performance
- **Brain:** Memory / Storage
 - With many layers (Caches, DRAM, SSD, HD, Tape)
- **Nerve system:** Network
 - Often more than one (MPI, Administration, I/O...)
- **Consciousness:** Software
 - Including various middleware layers (node management, process management, MPI, etc.)
 - OpenSource
- **Balance** is more important than performance of individual components
 - Challenge: „slow“ memory and network evolution vs. computing performance



SC Evolution – Top500 architectures

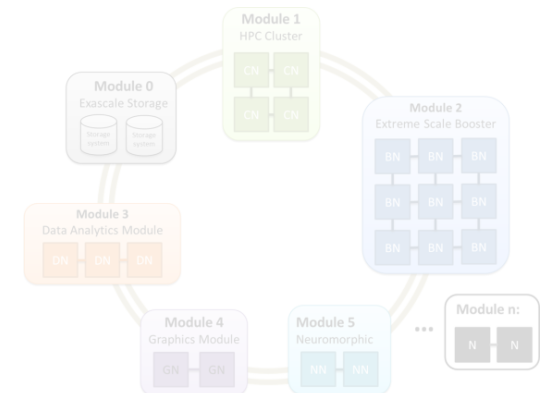
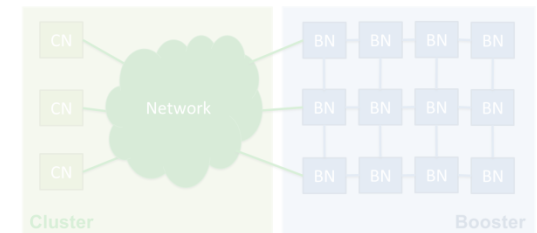
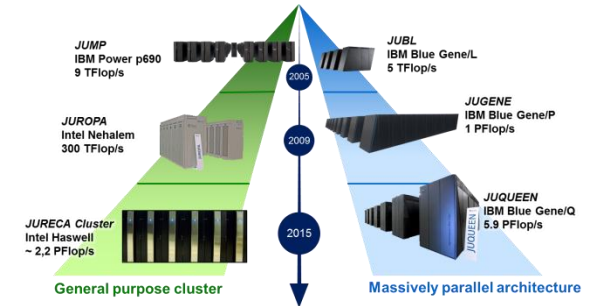




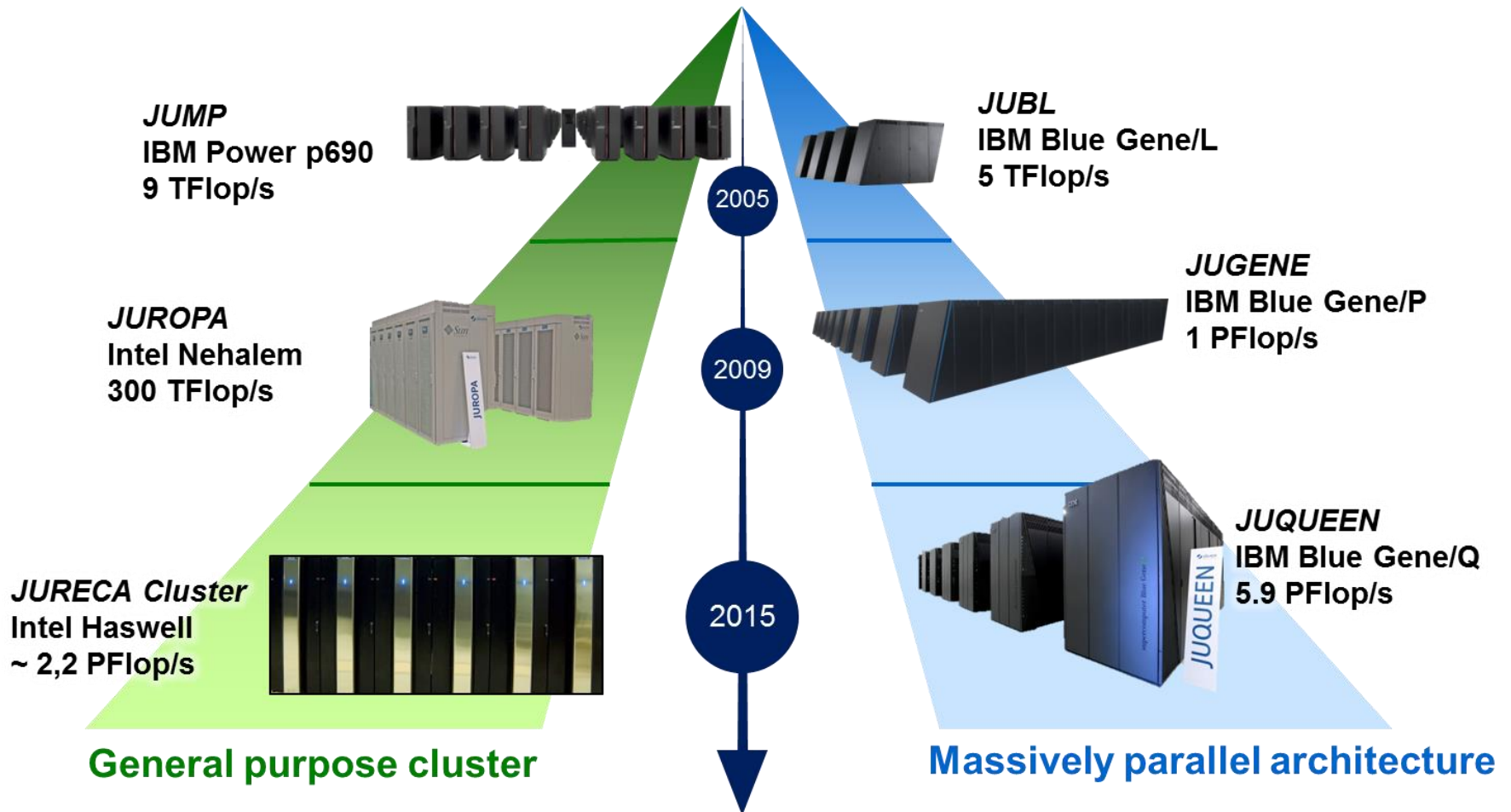
ARCHITECTURES AT JSC

Outline

- Supercomputing evolution
 - Cluster computing
- Architectures at JSC
 - **Dual** architecture approach
 - **Cluster-Booster** architecture
 - *The DEEP projects*
 - **Modular Supercomputing** architecture
- JSC future vision

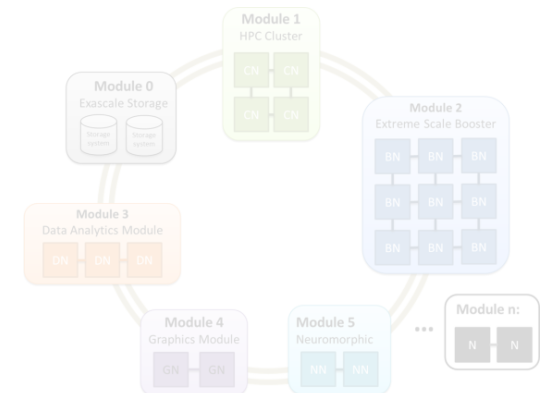
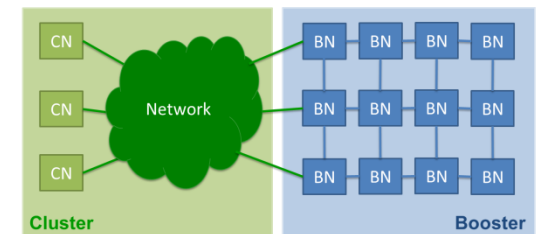
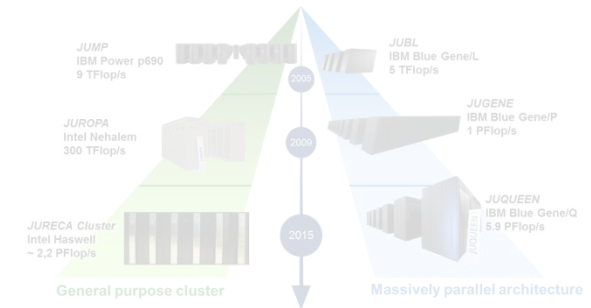


Dual Architecture at JSC

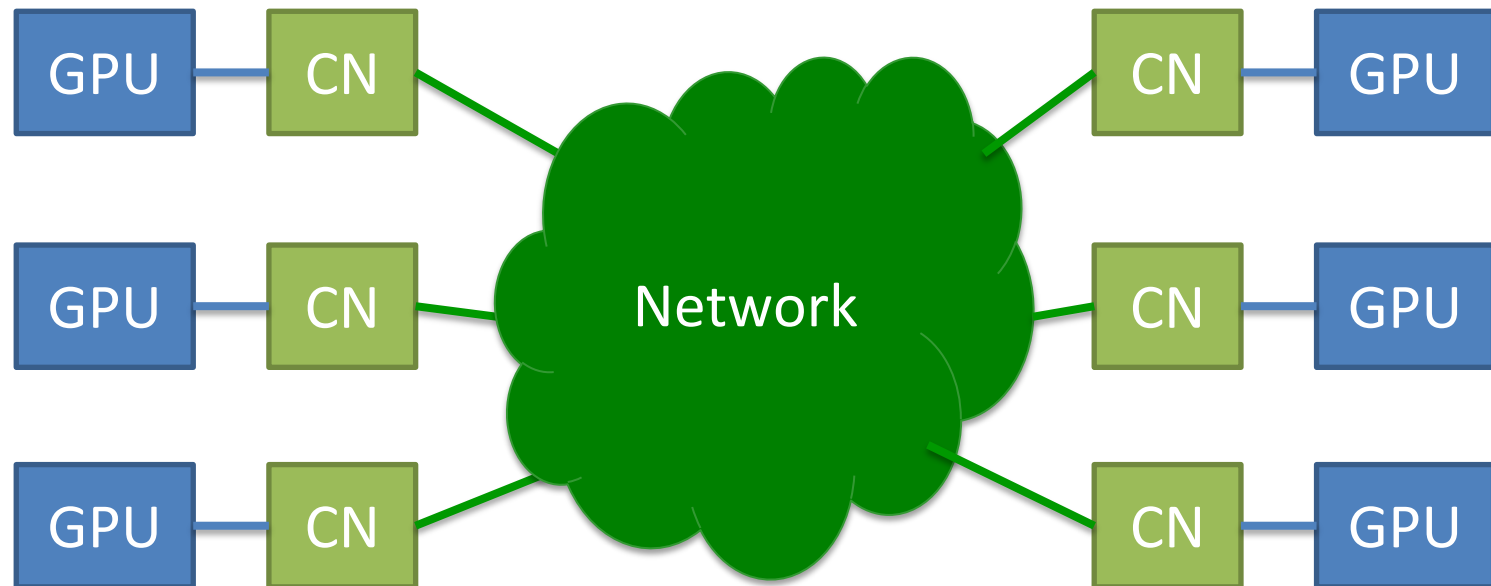


Outline

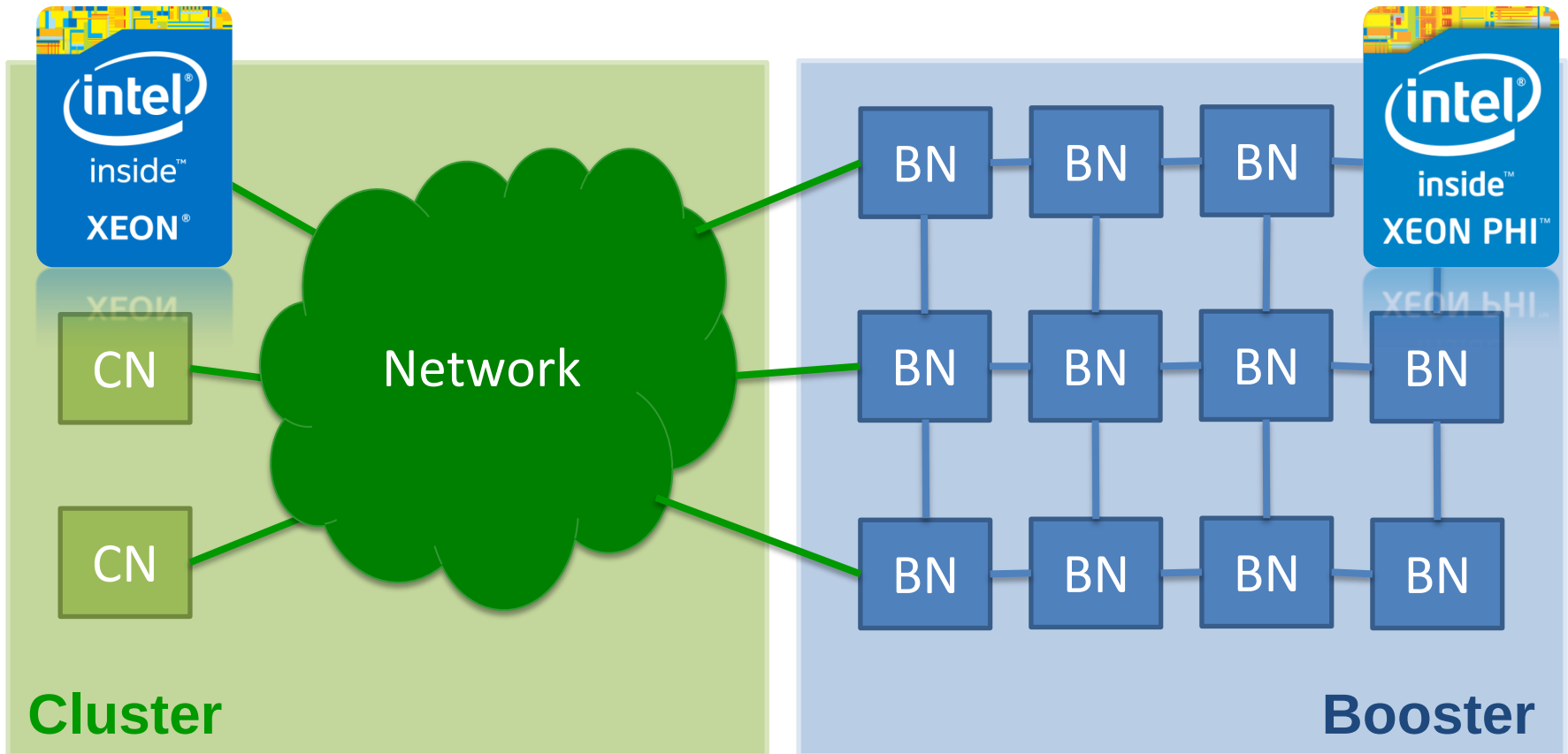
- Supercomputing evolution
 - Cluster computing
- Architectures at JSC
 - **Dual** architecture approach
 - **Cluster-Booster** architecture
 - *The DEEP projects*
 - Modular Supercomputing architecture
- JSC future vision



Heterogeneous Cluster



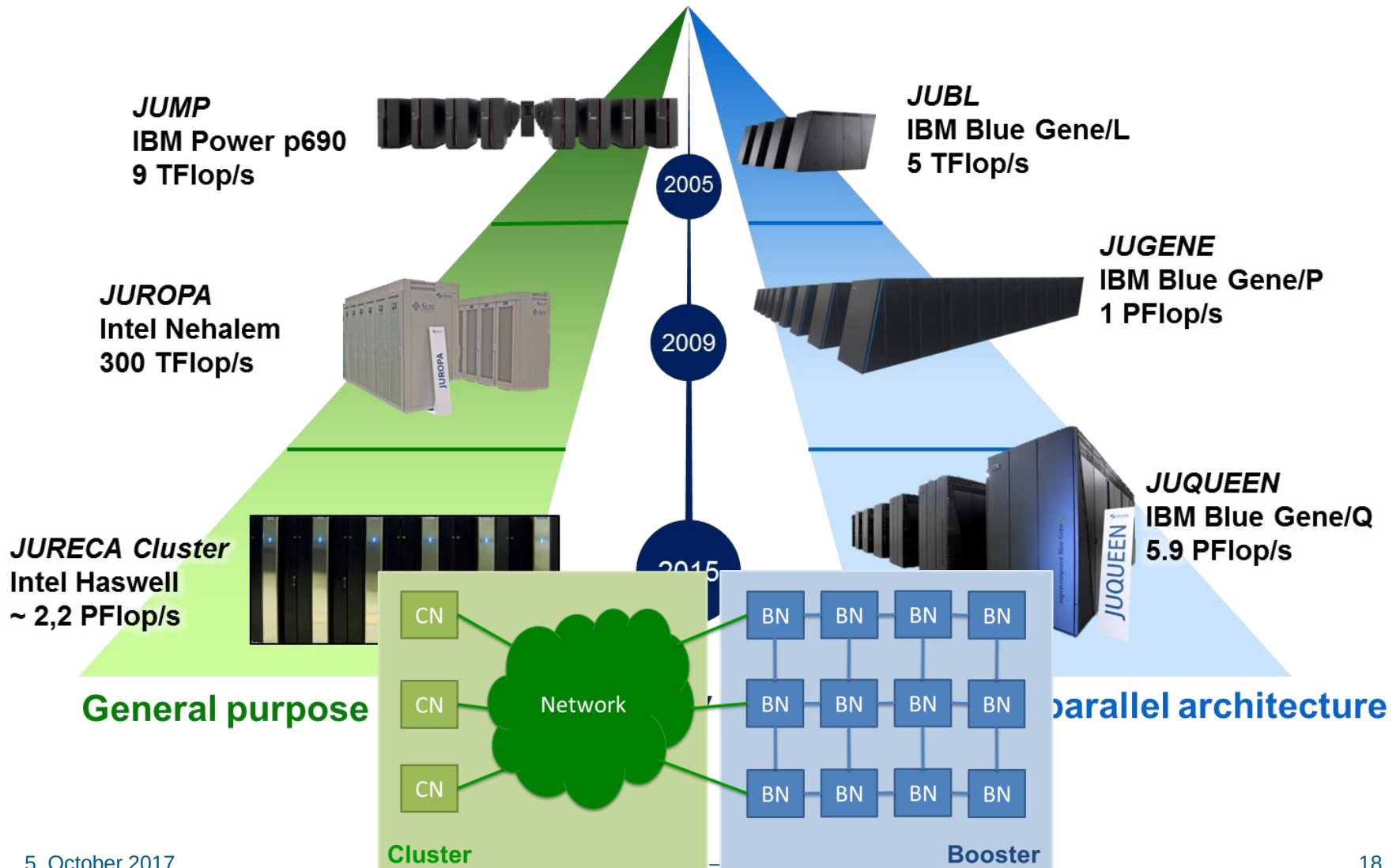
Cluster-Booster architecture



Low/Medium scalable code parts

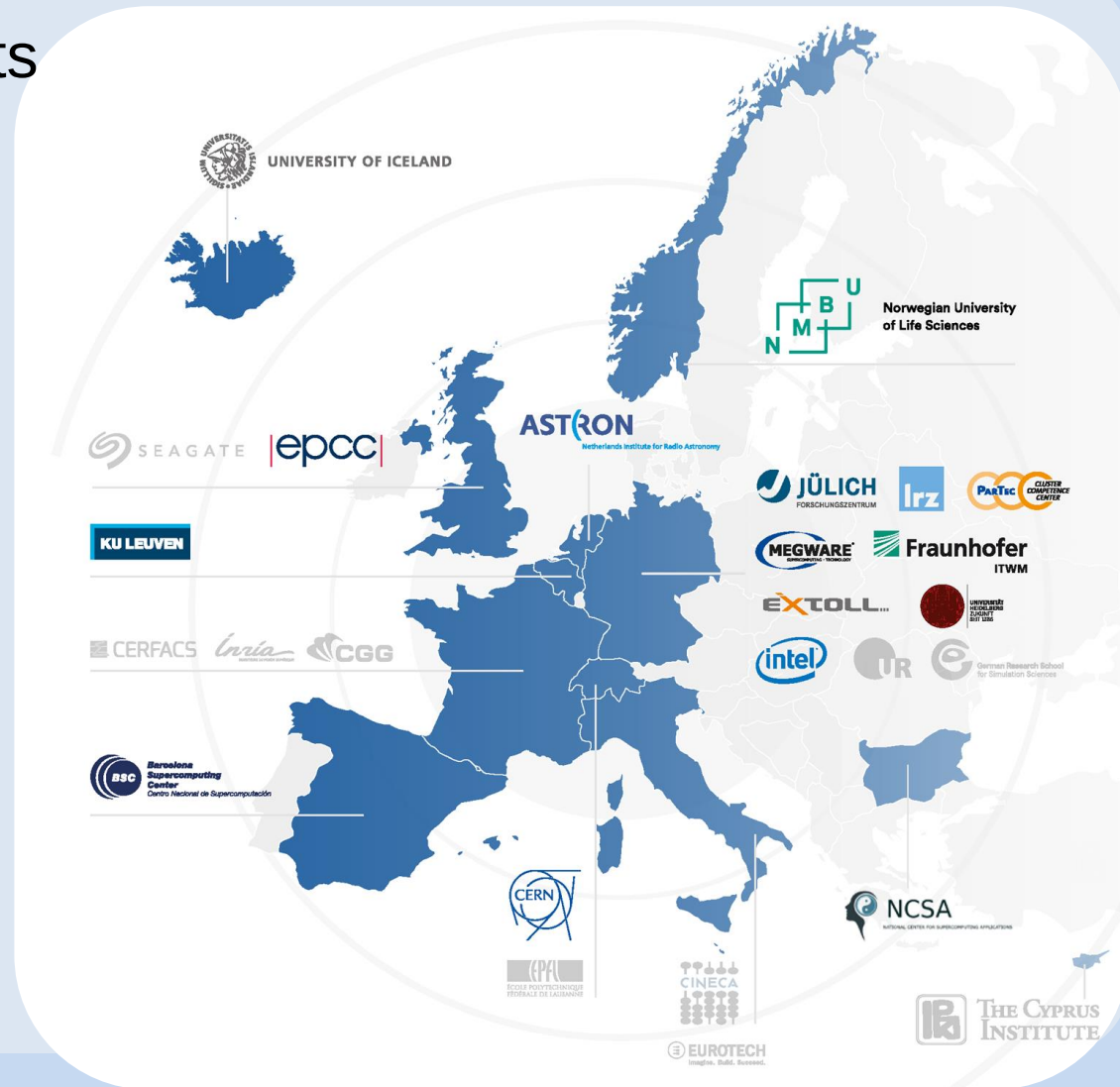
Highly scalable code parts

Dual Architecture at JSC



- 3x EU Exascale projects
DEEP
DEEP-ER
DEEP-EST
- 27 partners
Coordinated by JSC
- EU-funding: 30 M€
- Nov 2011 – Jun 2020

Co-design:
Hardware
Software
Applications

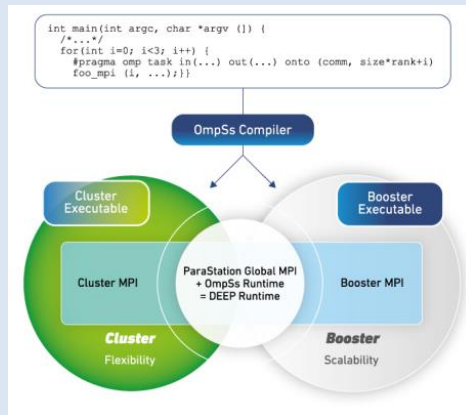
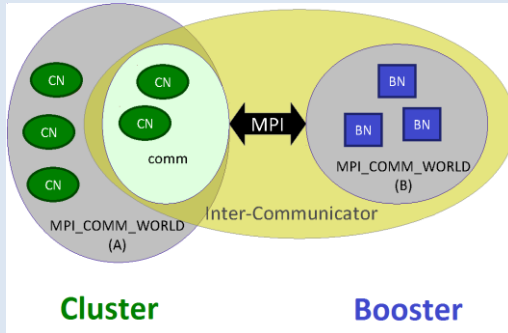


- Installed at JSC
- 1,5 racks
- 500 TFlop/s
(peak perf.)
- 3.5 GFlop/s/W
- Water cooled

Cluster
(128 Xeon)



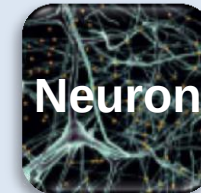
Booster
(384 Xeon Phi
KNC)



- **Low-level SW:** Cluster-Booster protocol
- **Scheduler:** Torque/Maui → SLURM
- **Filesystem:** BeeGFS
- **Compilers:** Intel, gcc, PGI
- **Debuggers:** Intel Inspector, TotalView
- **Programming:** ParaStation MPI (mpich), OpenMP, OmpSs
- **Performance analysis tools:** Scalasca, Extrae/Paraver, Intel Advisor, VTune...
- **Benchmarking tools:** JUBE
- **Libraries:** SIONlib, SCR, E10, HDF5...

DEEP projects applications (15):

- Brain simulation (EPFL + NMBU)
- Space weather simulation (KULeuven)
- Climate simulation (Cyprus Institute)
- Computational fluid engineering (CERFACS)
- High temperature superconductivity (CINECA)
- Seismic imaging (CGG + BSC)
- Human exposure to electromagnetic fields (INRIA)
- Geoscience (LRZ)
- Radio astronomy (Astron)
- Lattice QCD (University of Regensburg)
- Molecular dynamics (NCSA)
- Data analytics in Earth Science (UoI)
- High Energy Physics (CERN)



Neuron



NEST



iPiC3D



EMAC



AVBP



TurboRVB



RTM



FW



GERSHWIN



SeisSol



LOFAR
SKA



CHROMA



GROMACS

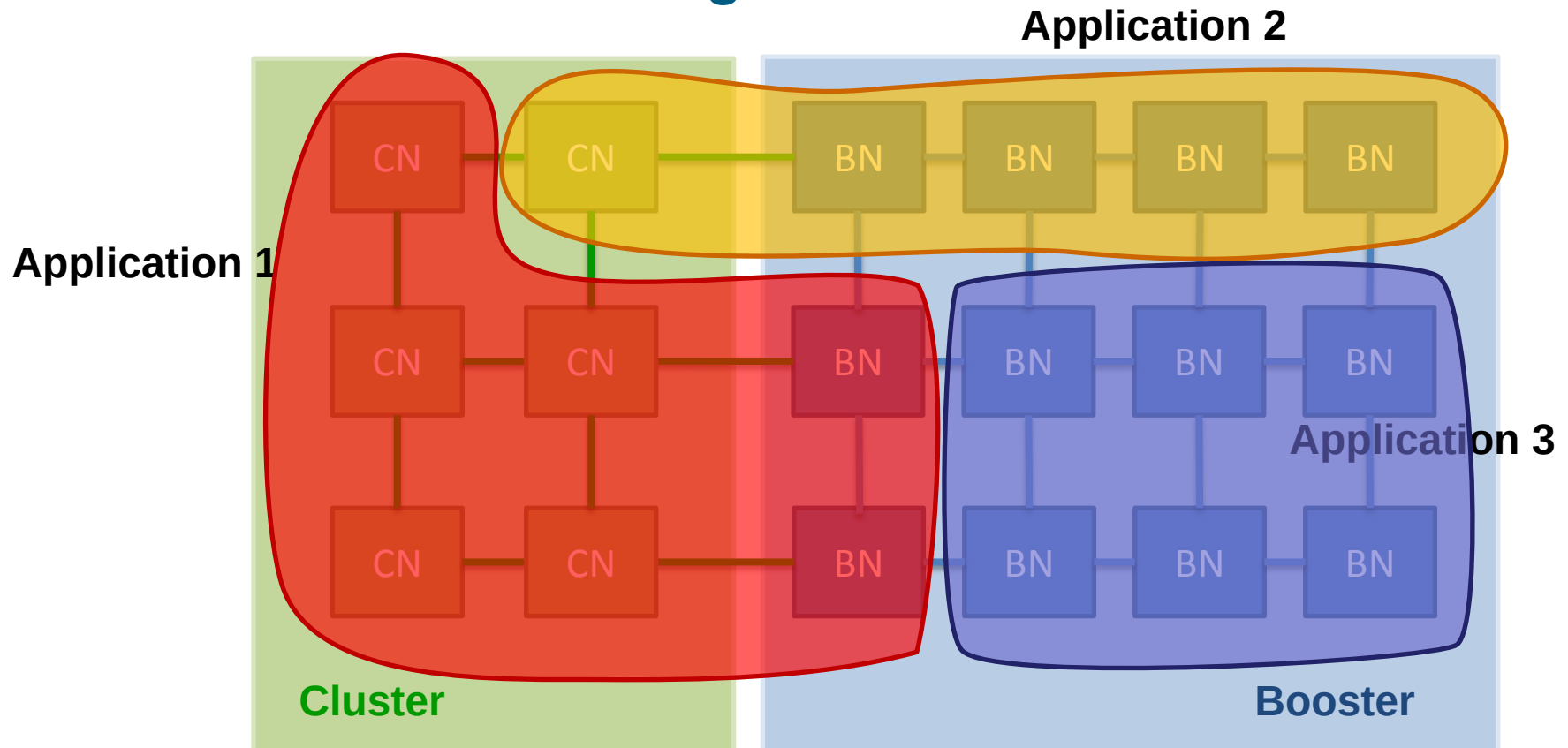


HPDBSCAN
piSVM



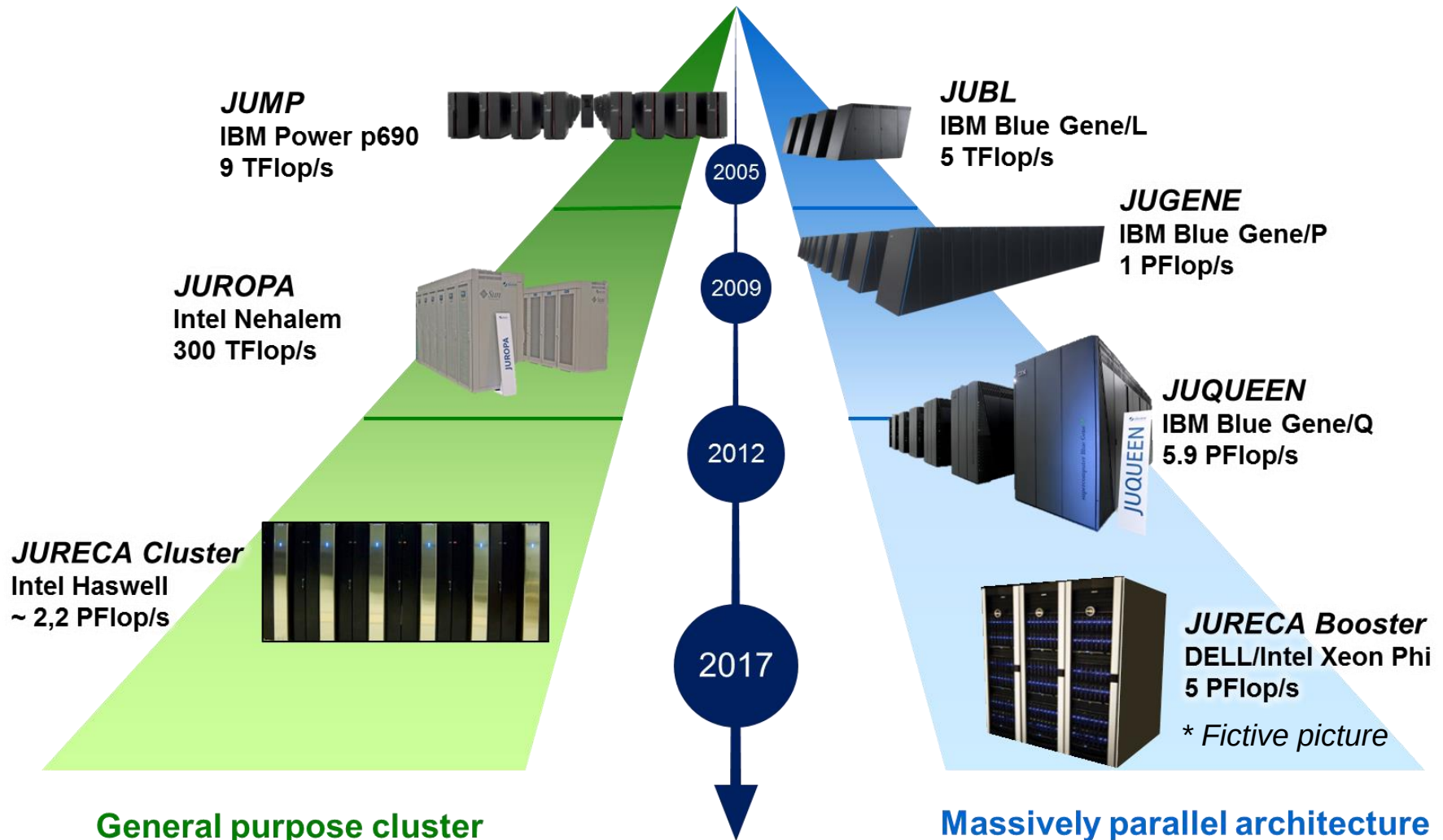
CMSSW

Cluster-Booster advantages

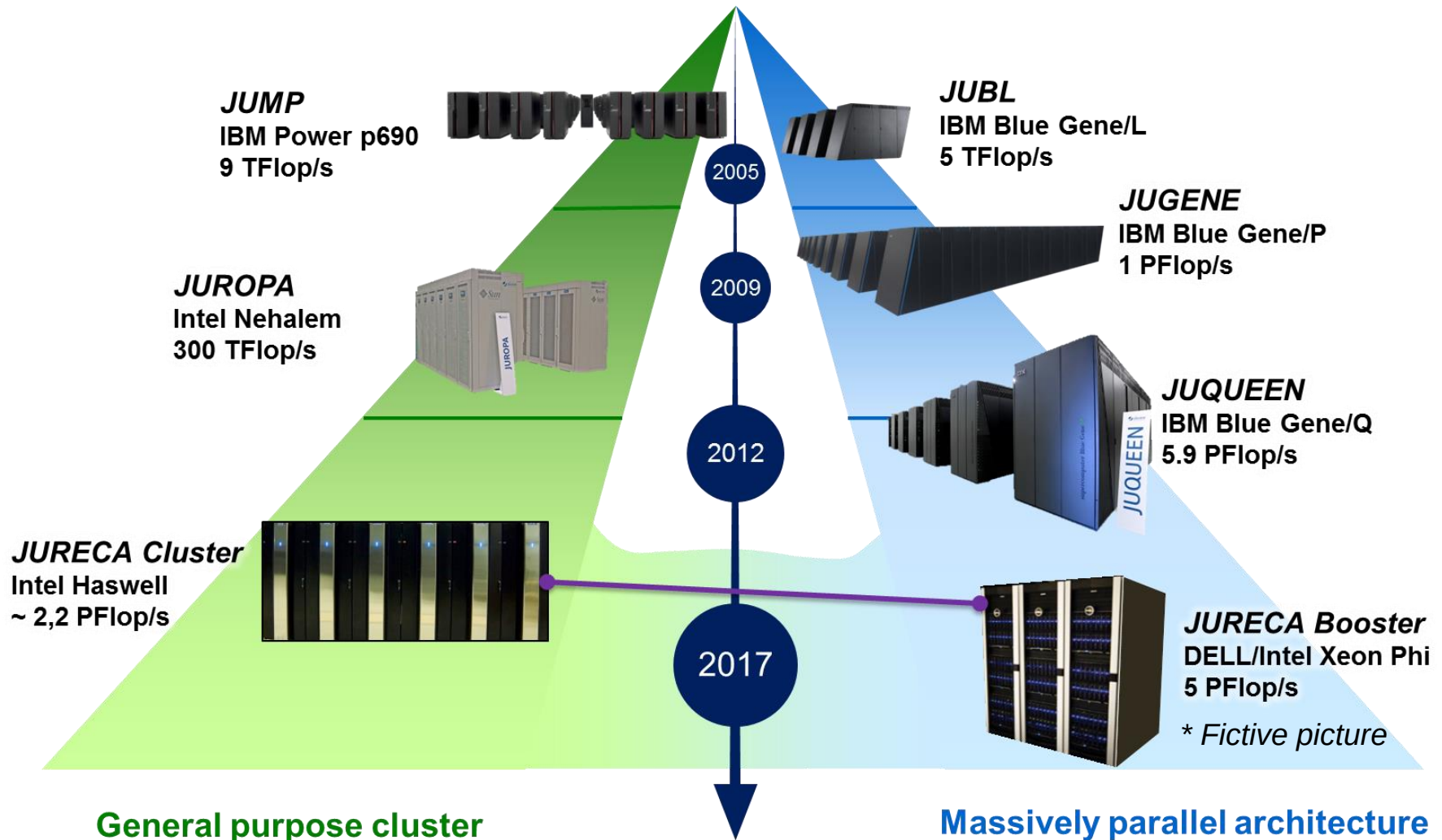


- **Full user flexibility** – many possible use modes
- **Efficient resources use** – only used nodes are blocked

Dual → Cluster Booster

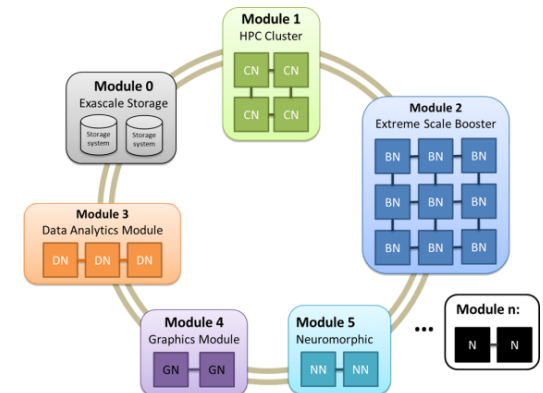
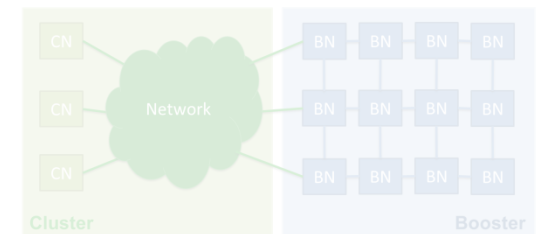
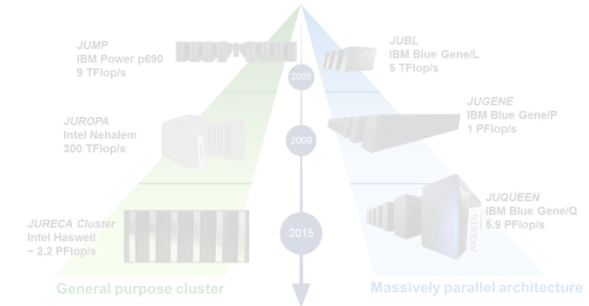


Dual → Cluster Booster

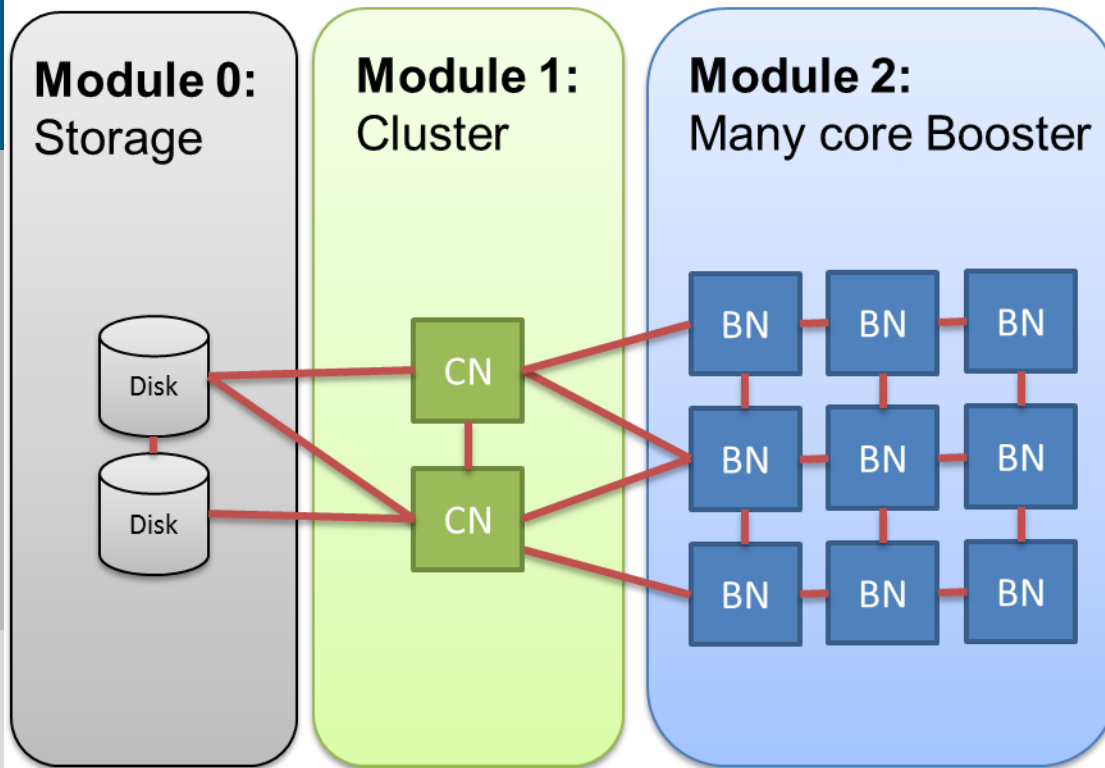


Outline

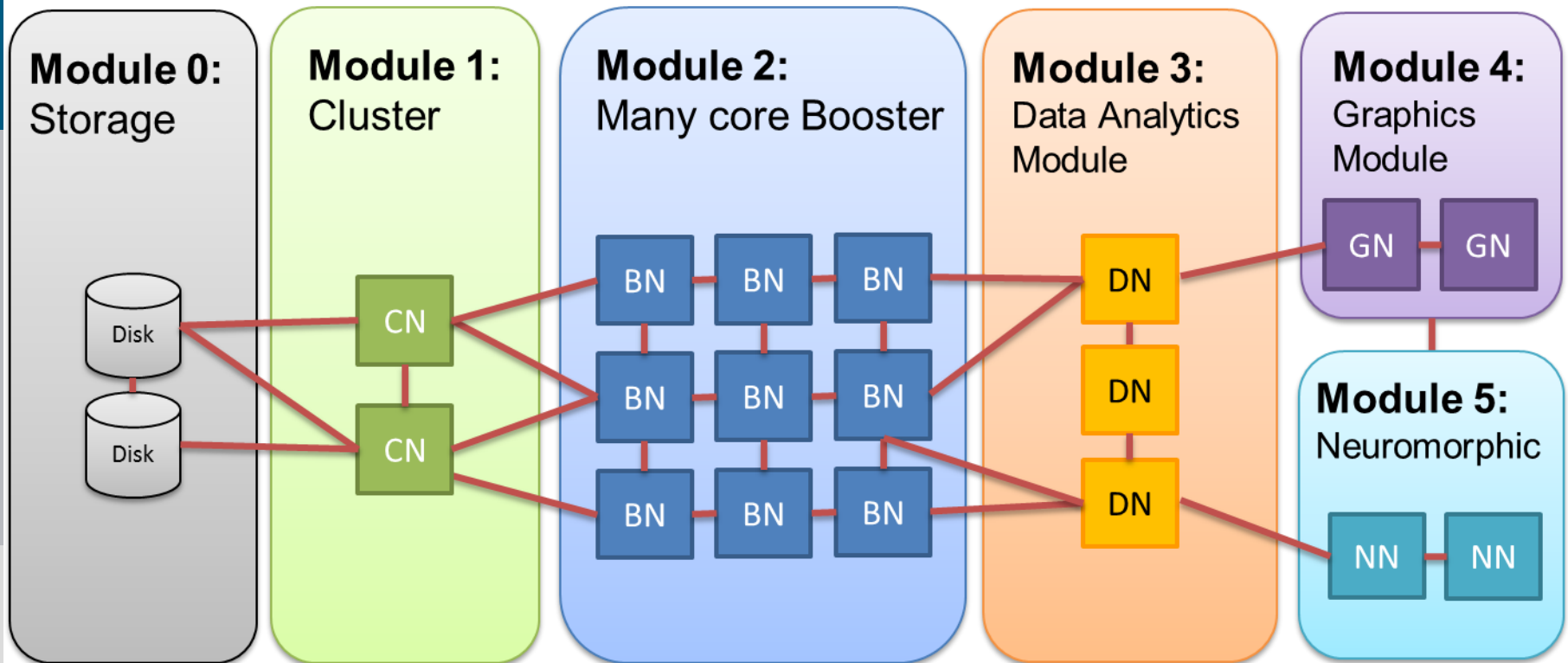
- Supercomputing evolution
 - Cluster computing
- Architectures at JSC
 - **Dual** architecture approach
 - **Cluster-Booster** architecture
 - *The DEEP projects*
 - **Modular Supercomputing** architecture
- JSC future vision



Cluster – Booster architecture

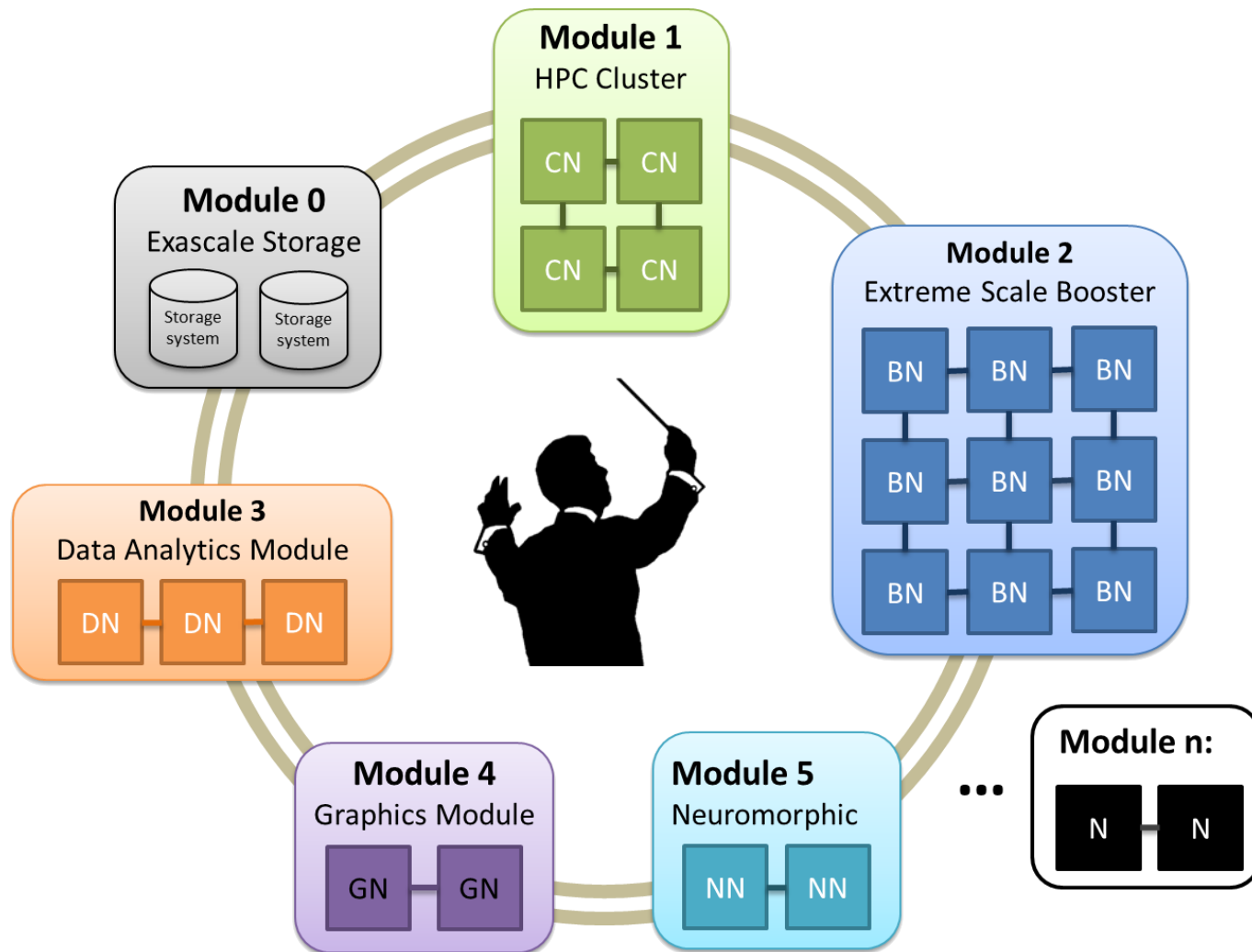


Modular Supercomputing

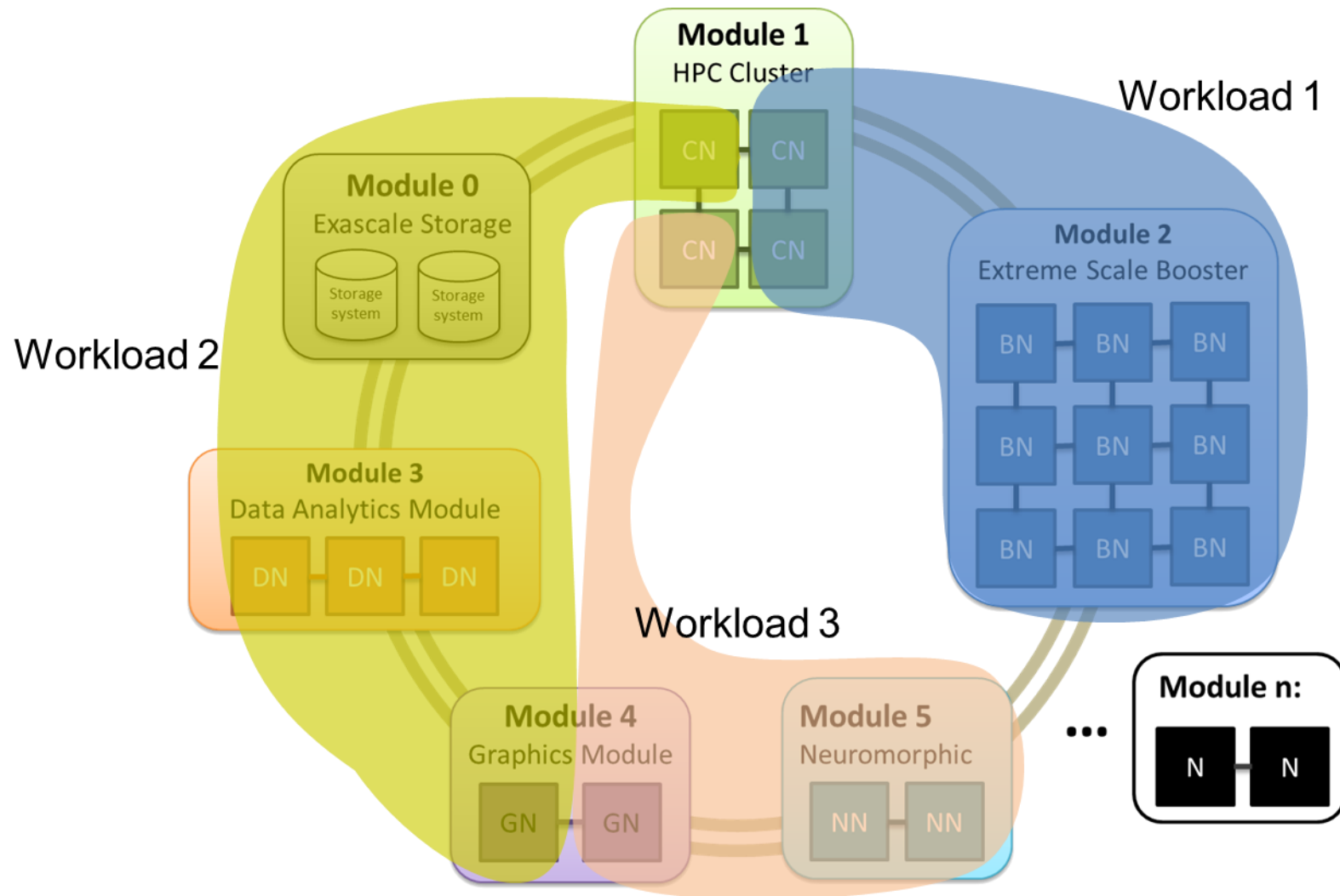


Generalization of the Cluster-Booster concept

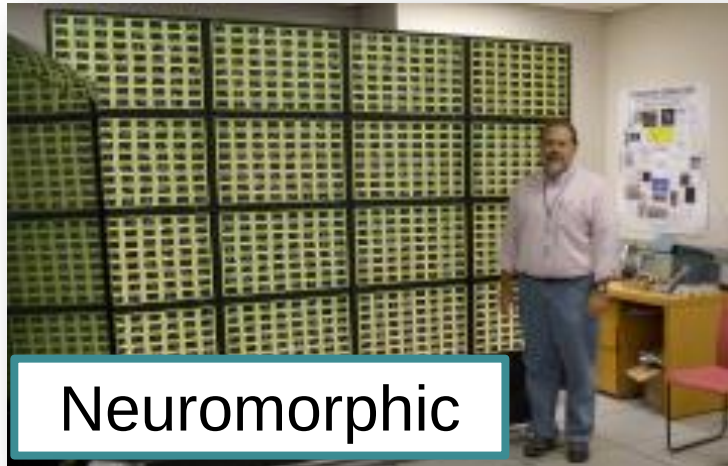
Modular Supercomputing



Modular Supercomputing



JSC future vision



Neuromorphic



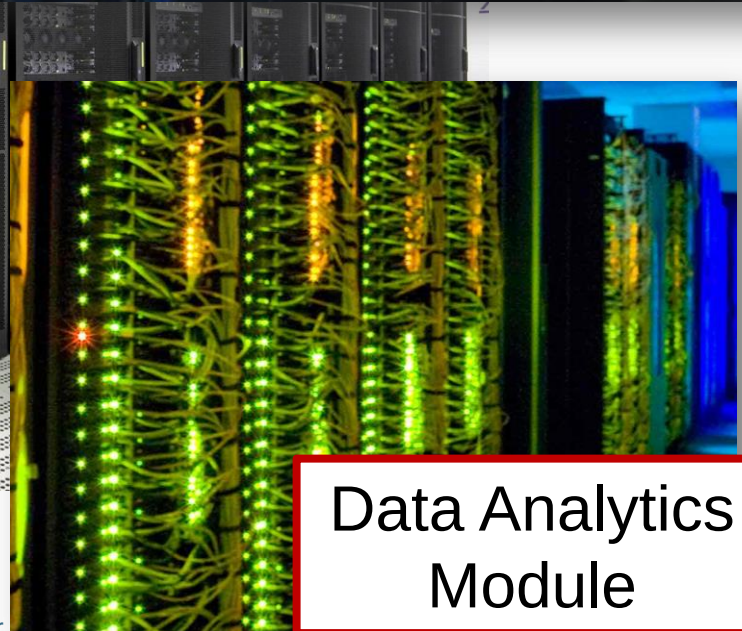
Booster



Quantum Computer



Cluster



Data Analytics
Module