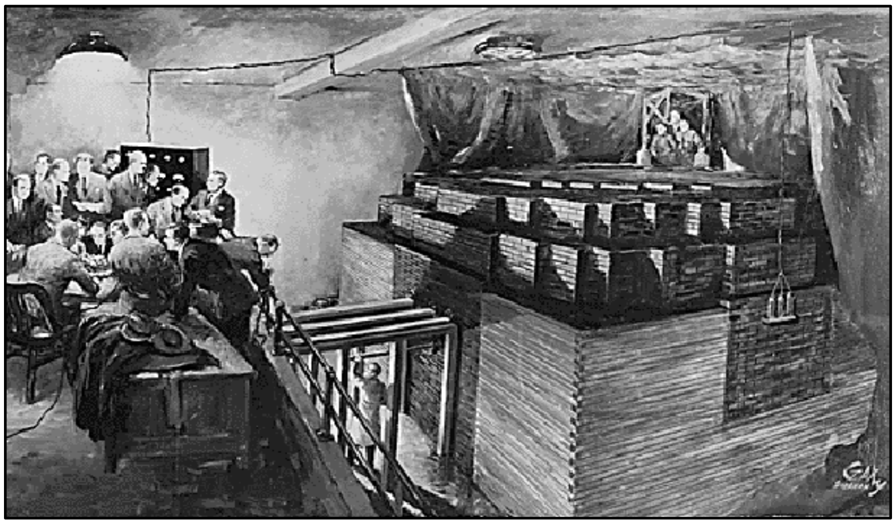
Supercomputers – Why are they different?



Katherine Riley
Argonne Leadership Computing Facility

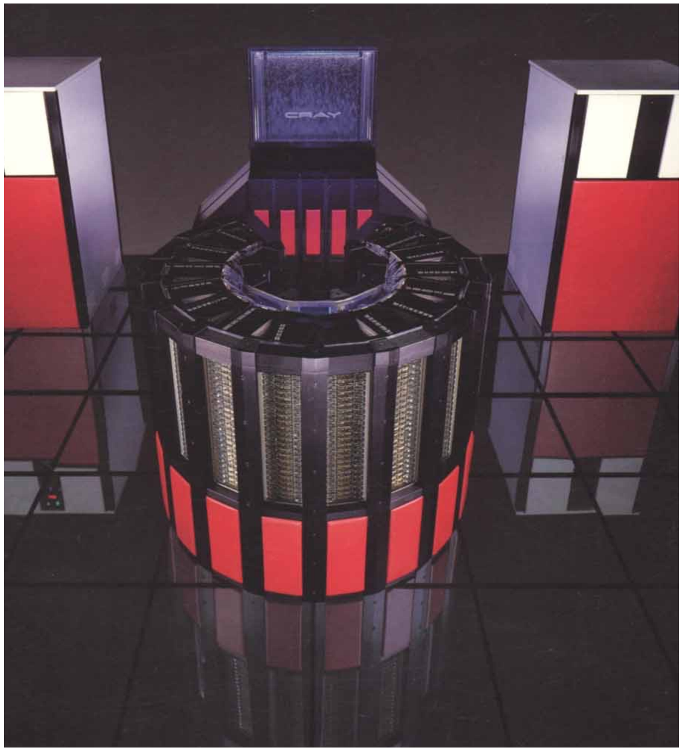
Argonne National Laboratory
July 2016

OUTLINE

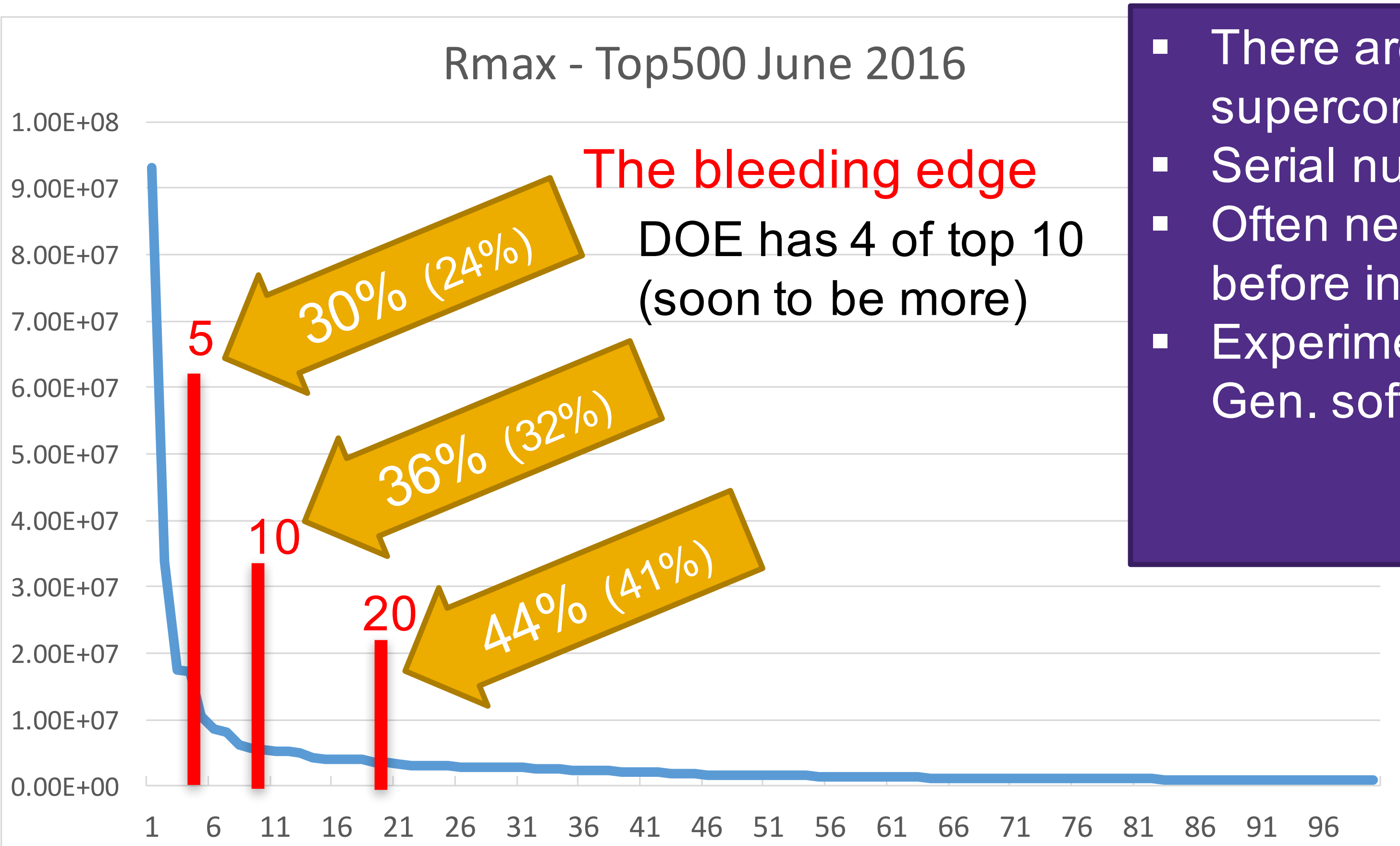
- Unique resources
- Science driven design
- Ripples
- Continuous improvement

IPAD VS SUPERCOMPUTER

- IPad 2 ~= Cray-2 from 1985
  - 1.9, 1.5-1.65 GFLOP/s
  - 2GB vs 64GB memory
- Is IPad 2 a supercomputer?
  - Just an old one?
  - Forefront of tech

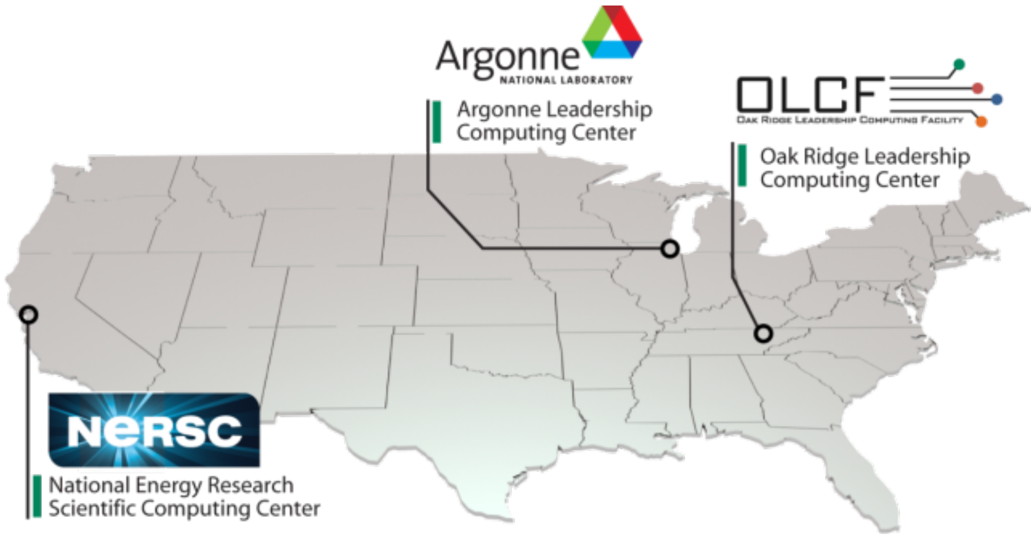


TOP 1% HAS A LOT OF THE COMPUTE POWER



- There are not many huge supercomputers
- Serial number .9, 1, 2
- Often never constructed before install
- Experimental/Rare/First Gen. software and hardware

DOE'S OFFICE OF SCIENCE COMPUTATION USER FACILITIES



NERSC
Edison is 2.57 PF

ALCF
Mira is 10 PF

OLCF
Titan is 27 PF

- DOE is leader in open High-Performance Computing
- Provide the world's most powerful computational tools for open science
- Access is free to researchers who publish
- Boost US competitiveness
- Attract the best and brightest researchers

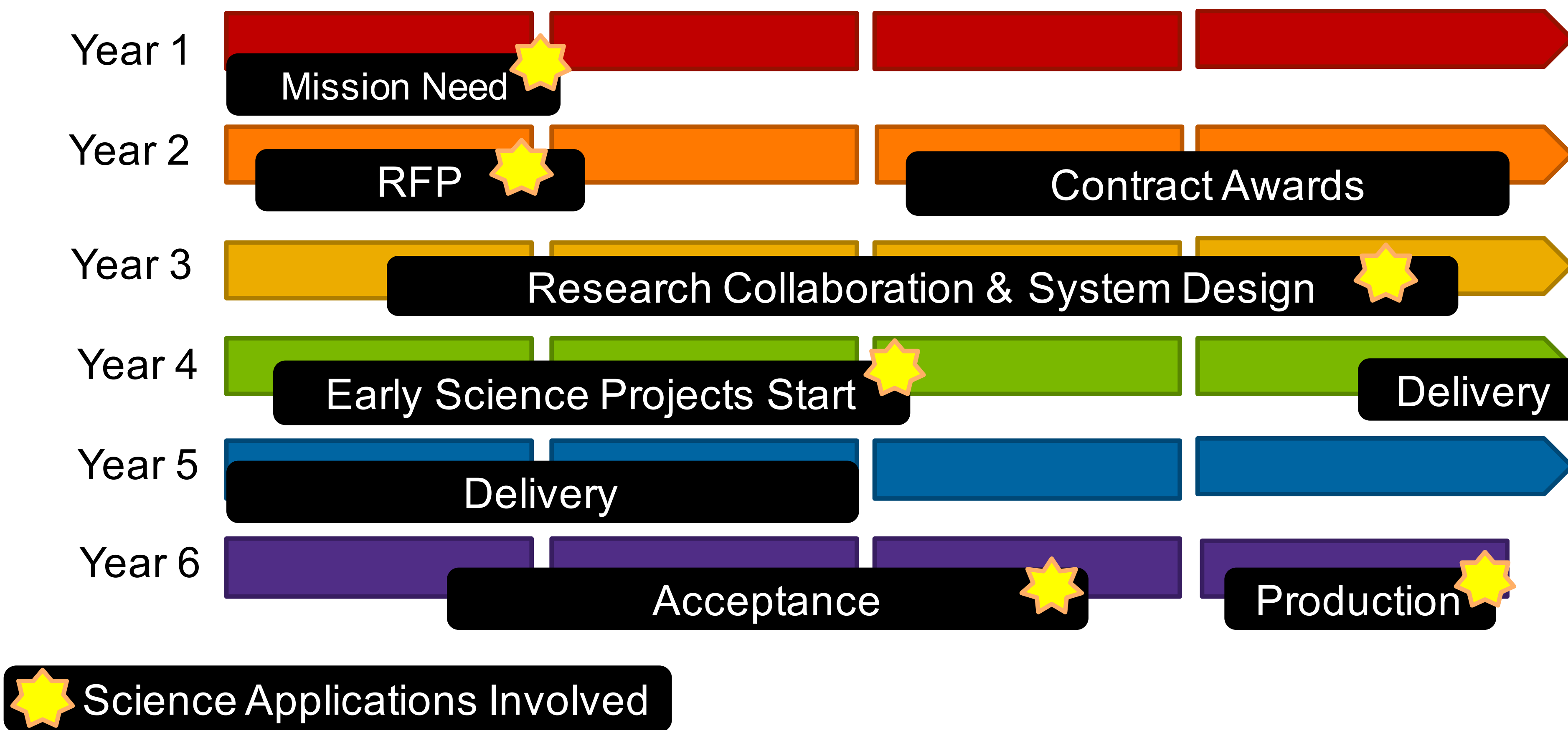
SCIENCE DRIVEN DESIGN

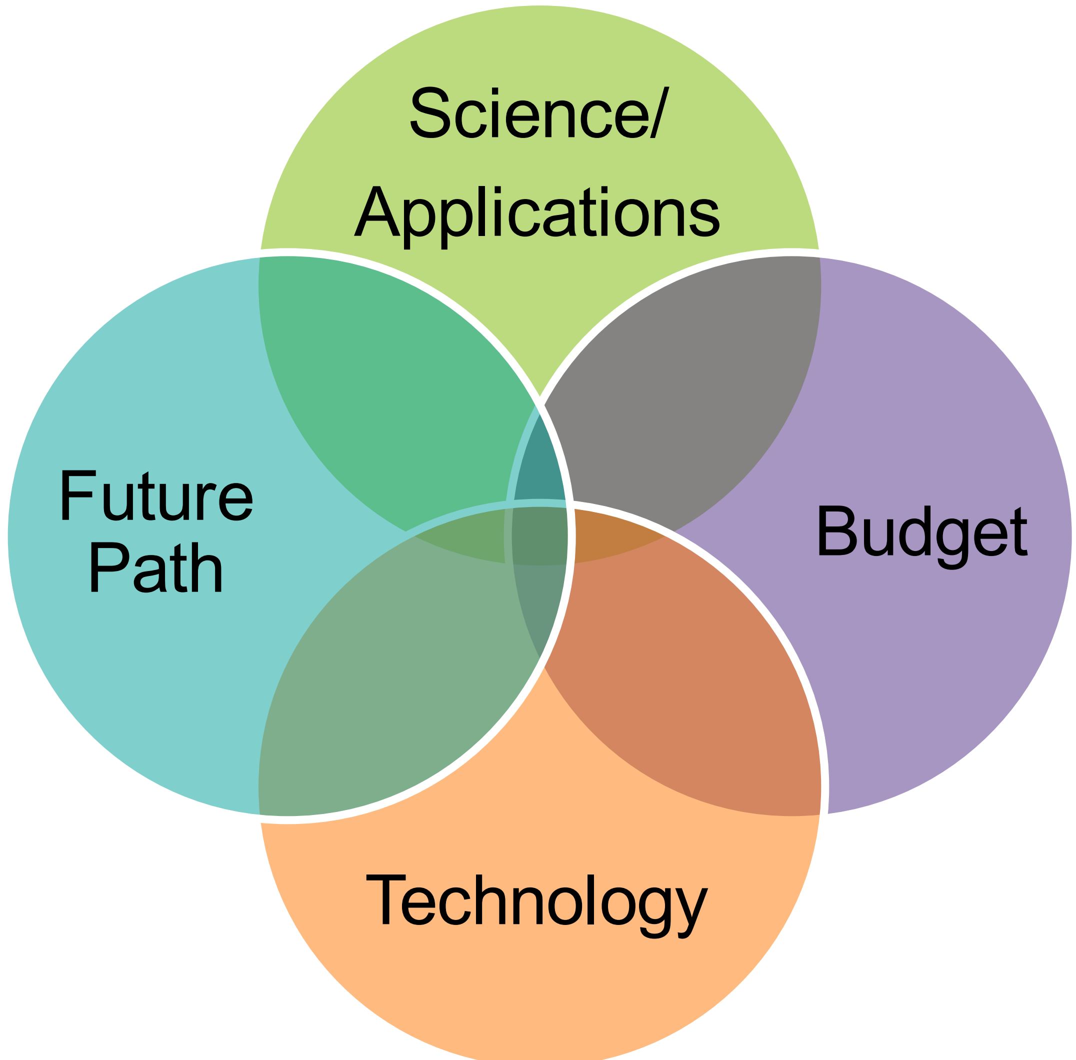
- Mission **need** from high impact science drivers
- Machine **design** proposals include application tests
- Science based project **performance** parameters
- **Collaboration** with vendors driven by science applications
- Early science programs to **test and evaluate**

CORAL
COLLABORATION
OAK RIDGE • ARGONNE • LIVERMORE

https://asc.llnl.gov/CORAL-benchmarks/

SUPERCOMPUTER TIMELINE




Science/
Applications
Future
Path
Budget
Technology

REAL IMPACTS

- Memory BW & footprint
- Compute node capability
- IO
- Software stack

CANNOT AFFORD WEAK LINKS

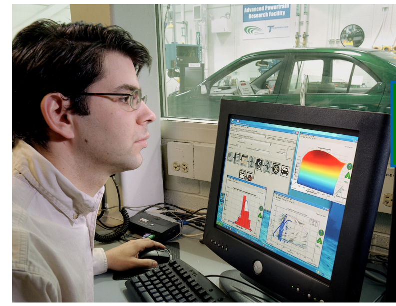
- Performance *does* matter
- Machine is only as fast as the slowest component
- Reducing overall capability reduces ability to delivery mission
- Reduce interconnect?
  - Less scaling
  - More contention
- Reduce memory footprint?
  - Limits science
- Reduce memory bandwidth?
  - Less performance

RIPPLES

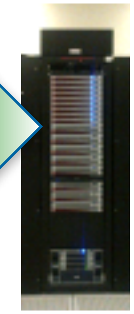
- Cost Limits on Topology
- New Topology
- New Algorithms
- New Technology Chips
- New Node Layout

- More OPS, memory, BW
- New Power
- New Heat
- Better Cooling
- Higher Density Racks
- Better Floors

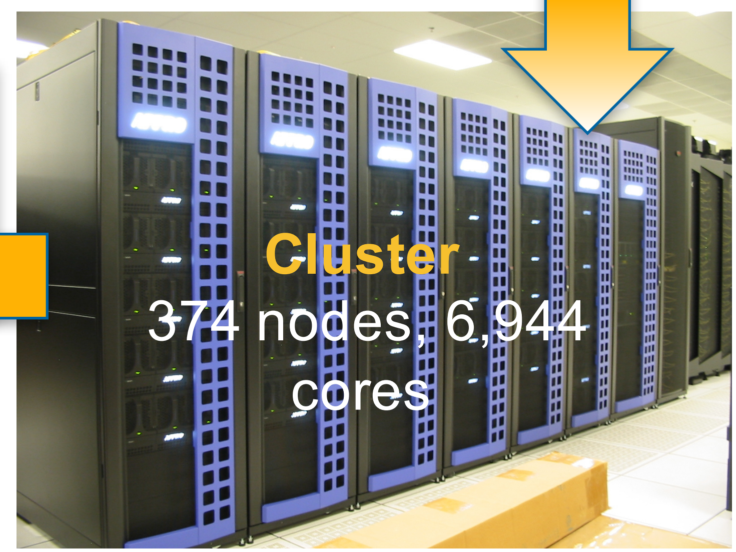
SCALE JUMPS TOUGH: COMPLEXITY



Desktop

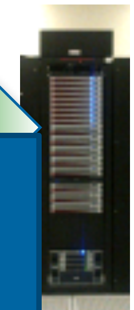


Work Group



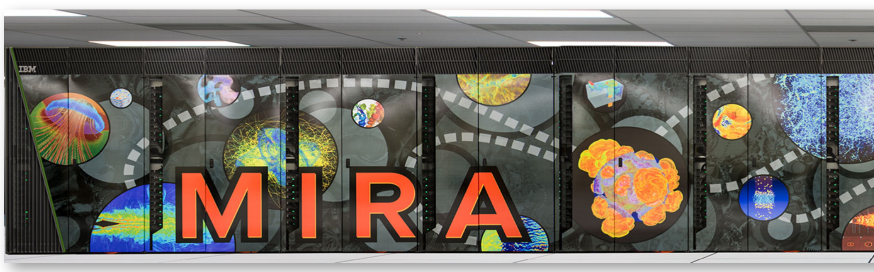
Supercomputer
Mira: 49,152 nodes, 786,432 cores

SCALE JUMPS TOUGH: COMPLEXITY



Group

- Science driven design has constraints
- Chain or ripple, there are consequences

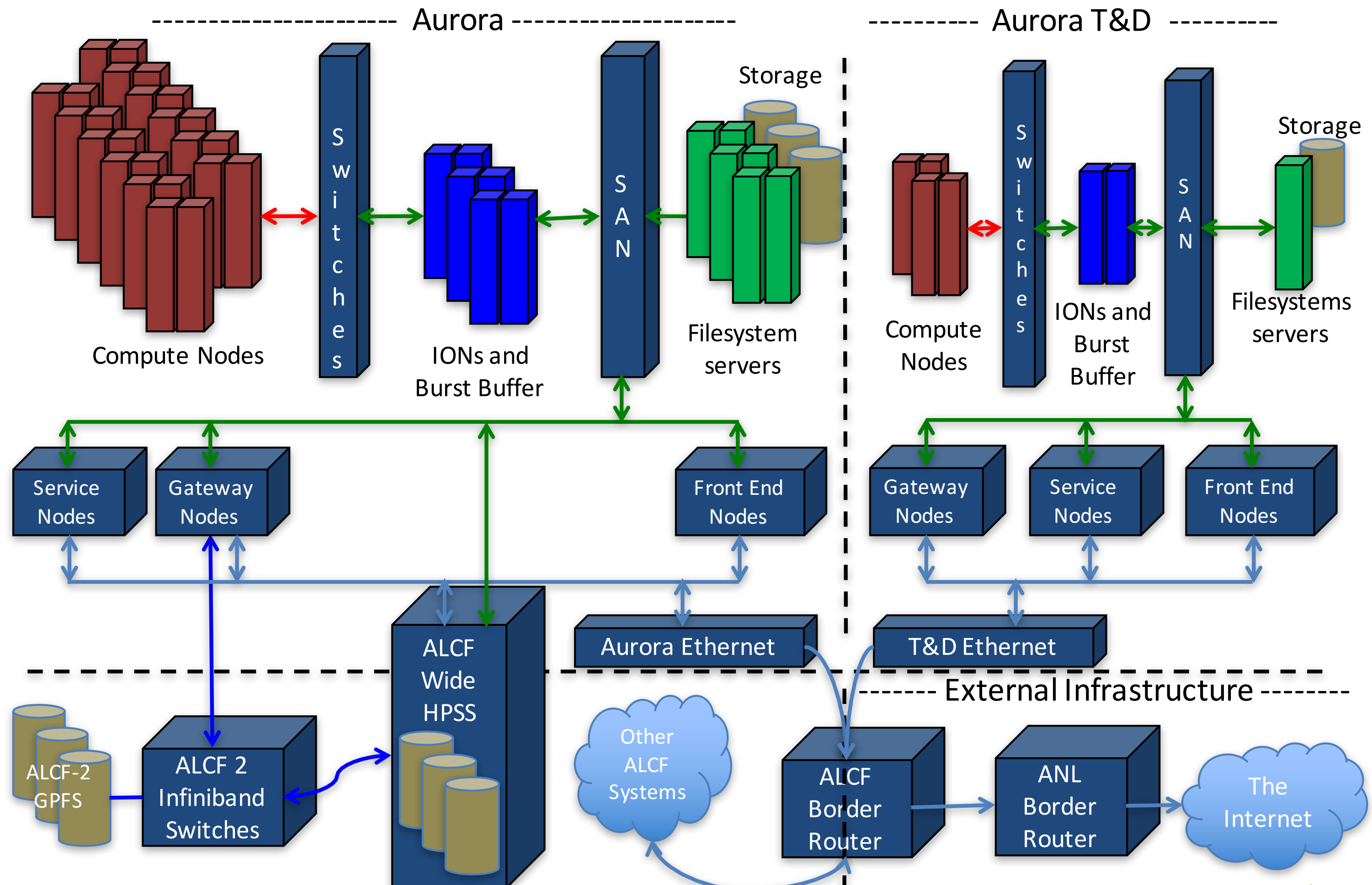


Supercomputer
Mira: 49,152 nodes, 786,432 cores

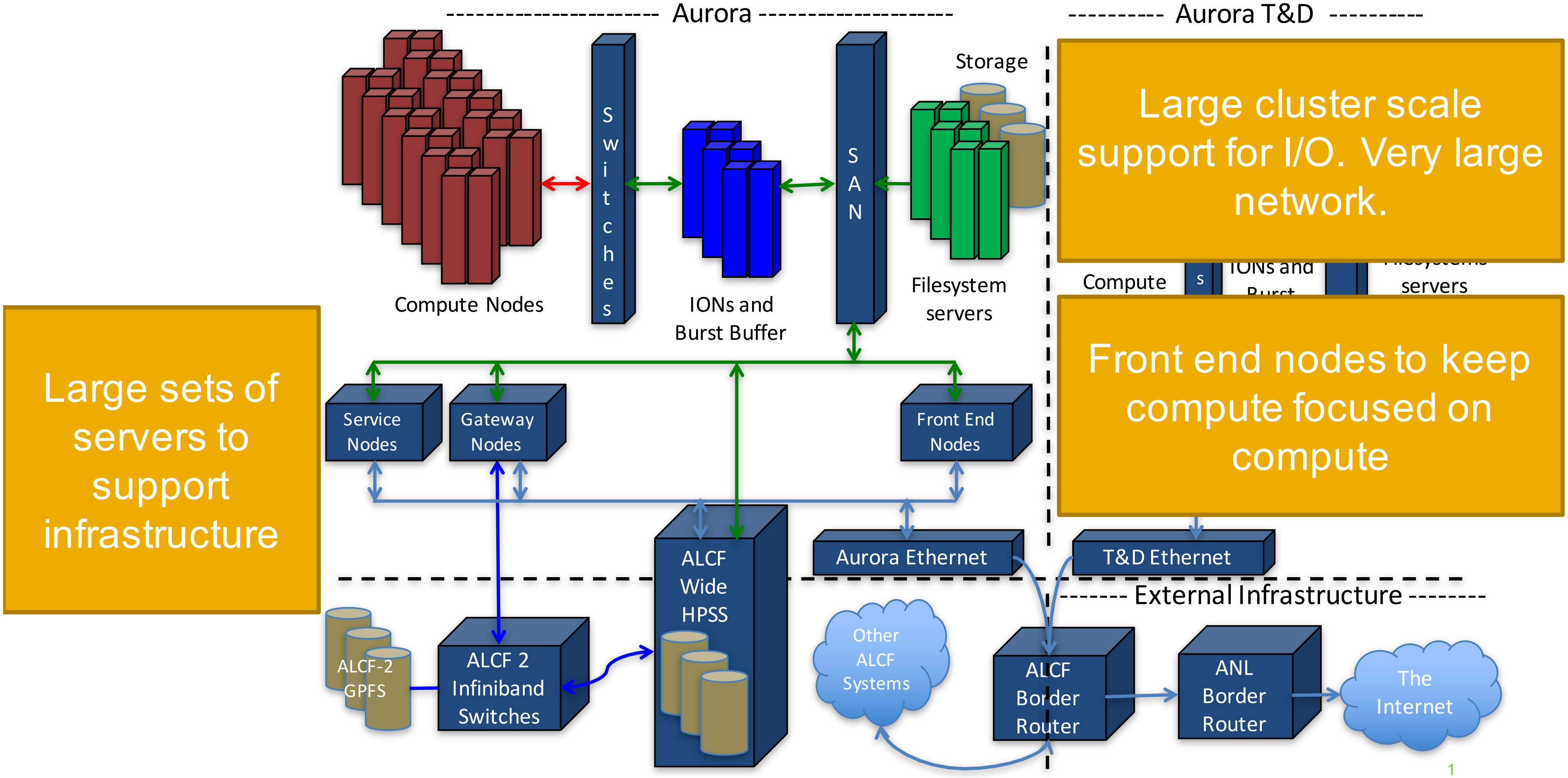
COMMON CHALLENGES

- Memory per task
- Scaling
- Managing data
- Portability
- Challenge for OPS
- Authentication
- Programming language/model availability
- Batch queuing
- File transfer
- Scaling
- Compiler capabilities
- Libraries at scale
- System calls/compute node capabilities
- Contention

HIGH LEVEL FACILITY DIAGRAM



HIGH LEVEL FACILITY DIAGRAM



HARDWARE RIPPLES

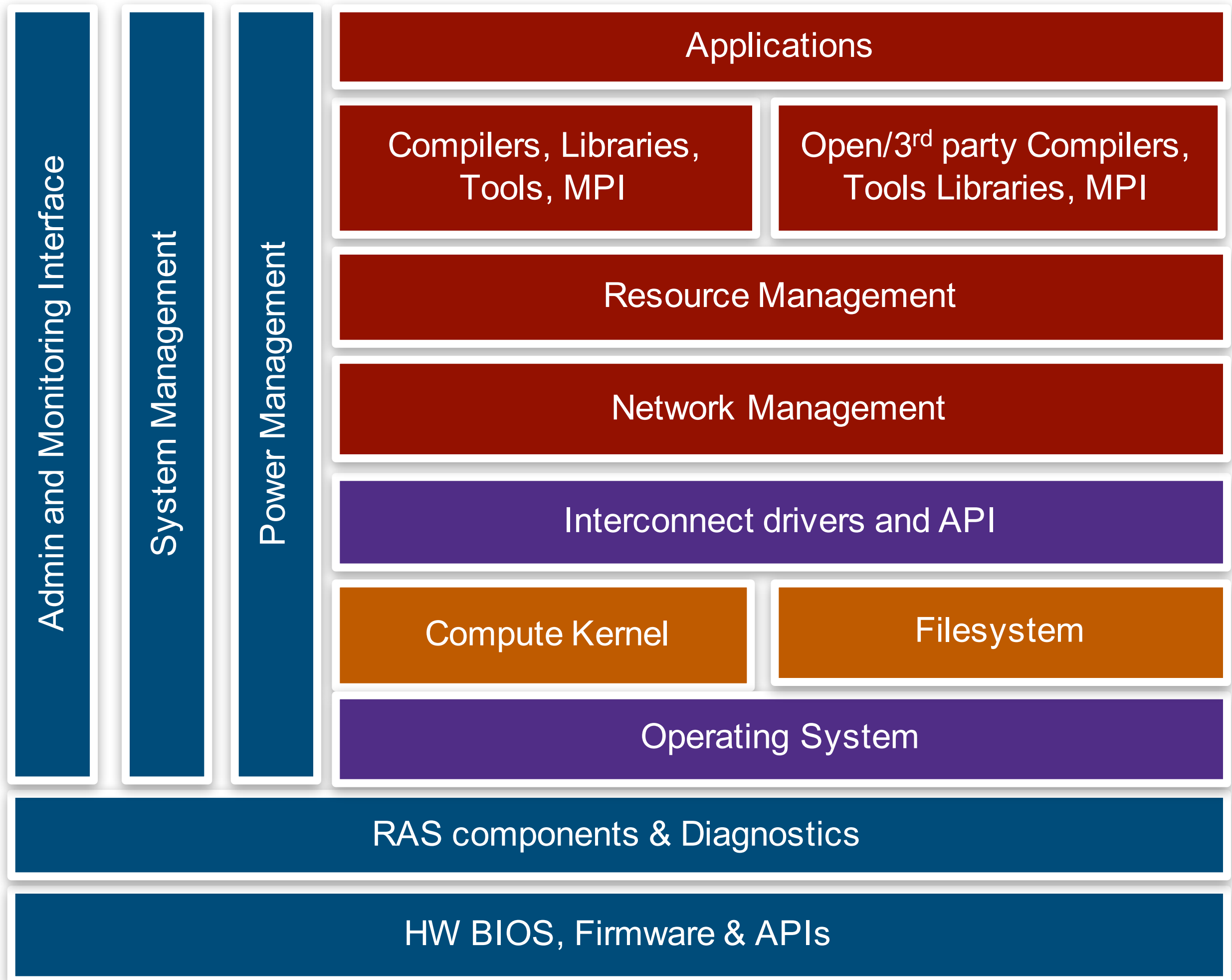
- Scaling for O(1K) is not the same for O(10K) or O(100K)
  - Network technology and software
  - Jitter (software)
  - Application capability
- Exposing OPS
  - Parallelism
  - Software stack
- I/O
  - Portability
  - Managing

SOFTWARE RIPPLES

- Bleeding edge HW requires clever SW solutions
- Third party tools need to scale to be functional
- Compute OS functionality
  - Jitter
  - Performance
- Behavior of common libraries varies, e.g. MPI

SOFTWARE STACK

- All crucial and specialized for a system
- Immediate high impact touch by application
- Application impact
- Some apps will notice



DESIGN POINT FOR PROGRAMMING MODELS

Program Model Target	Past	PMs are here → Now & Future
Cost	FLOPS	Data movement
Memory Scaling	BW, byte/FLOP	Tiered BW speeds, OPS/byte
Locality	MPI (+X) Off-node costs ok	MPI+X+Y+… Off-node cost growing
Concurrency	Node count growth, Node concurrency slow	Node count shrinking, Node concurrency growing
Uniformity	Homogenous	Heterogeneous
Reliability	Hardware Issue	Whose Issue is It? (System Software)
Power	Operational/Facility Plan	+System software

APPLICATION RIPPLES

- Science applications are about performance
  - Time to solution, not speed of opening PPT
- Development environment
- New capabilities invite new science
- Programming models and languages

CONTAINERS STILL STARTING

Promising but need to investigate efficacy

- Jitter and overhead for performance
- Exposure to optimized libraries
- Scalability

IDEAS FOR PATHS FORWARD

Computational Environment

- New use models: in-time, real-time, complex workflows, streaming, co-scheduling, more in-situ
- Portability and Performance
- Diverse environments and requirements for access
- Programming models and runtime systems

Data

- Improved tools, algorithms, approaches
- Data discovery, deep/machine learning, etc
- Data management, provenance, & storage
- Data where you need it when you need it – all tiers of data

Community Involvement

- Education and training
- Standards committees
- Training materials, best practices, examples, etc.

SUMMARY

- Biggest supercomputers are rare beasts
- Science plays a large role in design
- Everything counts in design
- There is an application challenge
- Facilities and DOE working on improvements

QUESTIONS