

UCX BOF Update @ SC'18



ASTRA

"Per aspera ad astra"

VANGUARD

Astra – the First Petscale Arm based Supercomputer



HPE Apollo 70 Chassis: 4 nodes



HPE Apollo 70 Rack



18 chassis/rack

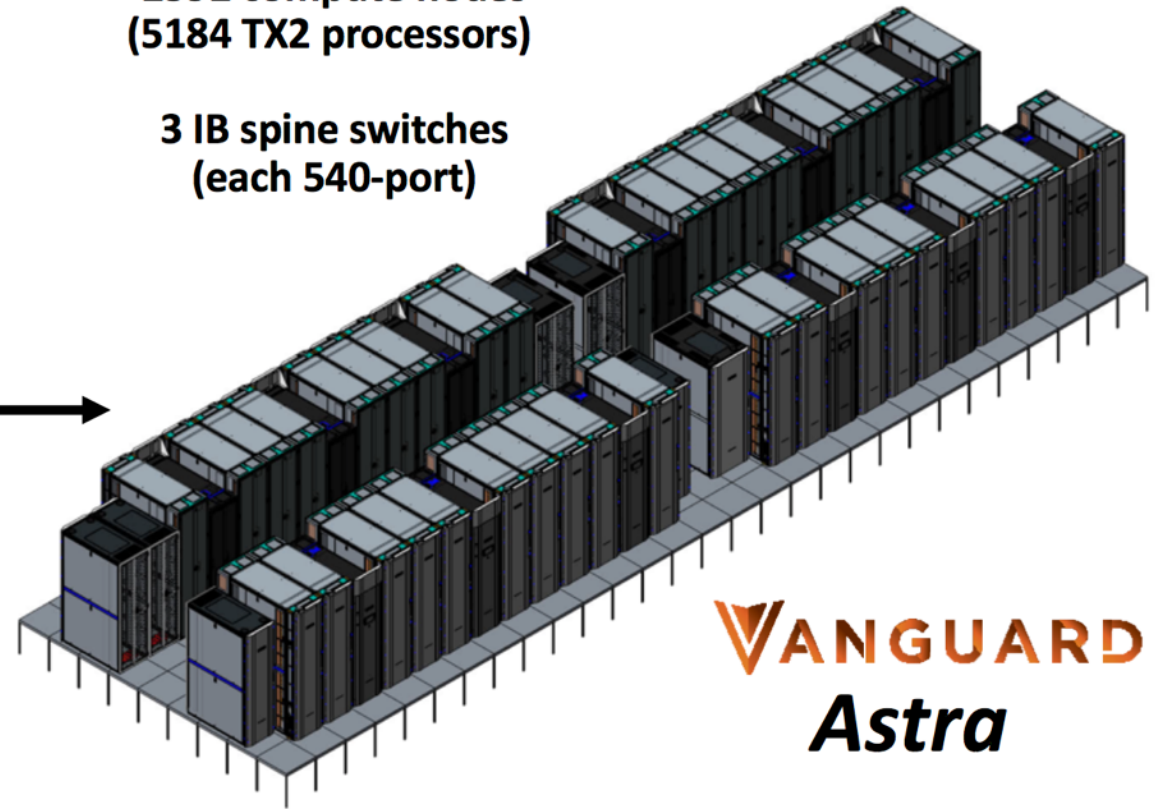
72 nodes/rack

**3 IB switches/rack
(one 36-port switch
per 6 chassis)**

**36 compute racks
(9 scalable units, each 4 racks)**

**2592 compute nodes
(5184 TX2 processors)**

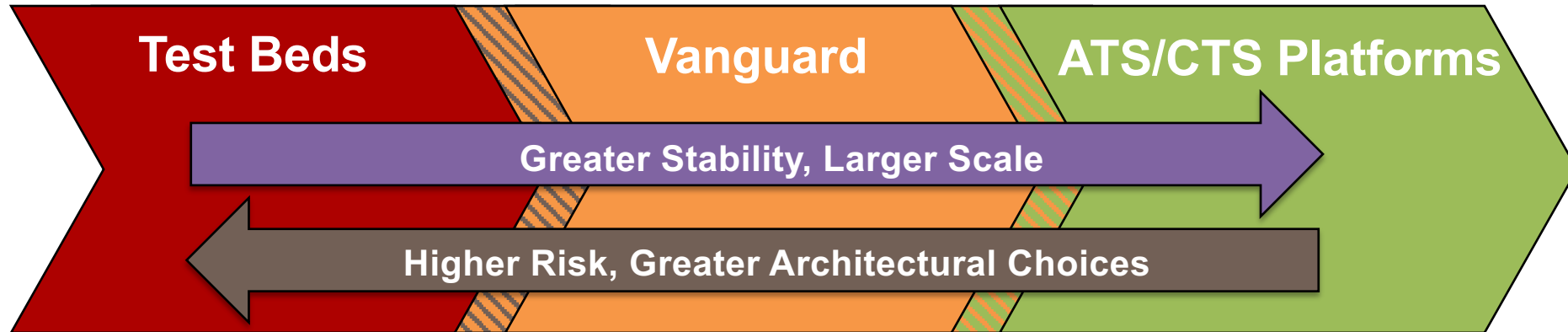
**3 IB spine switches
(each 540-port)**



**VANGUARD
Astra**



Where Vanguard Fits



Test Beds

- Small testbeds (~10-100 nodes)
- Breadth of architectures Key
- Brave users

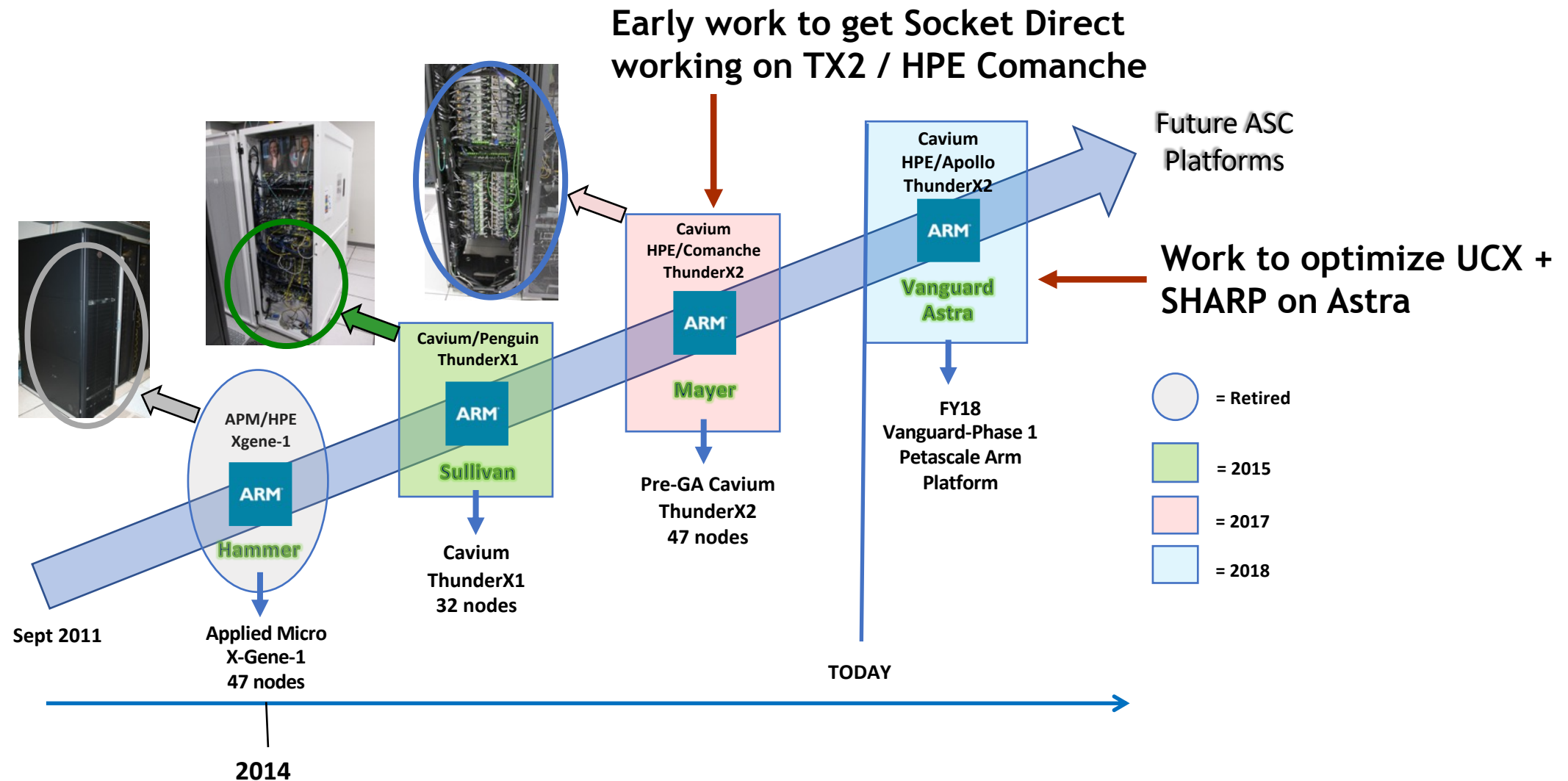
Vanguard

- Larger-scale experimental systems
- Focused efforts to mature new technologies
- Broader user-base
- Not Production
- **Tri-lab resource but not for ATCC runs**

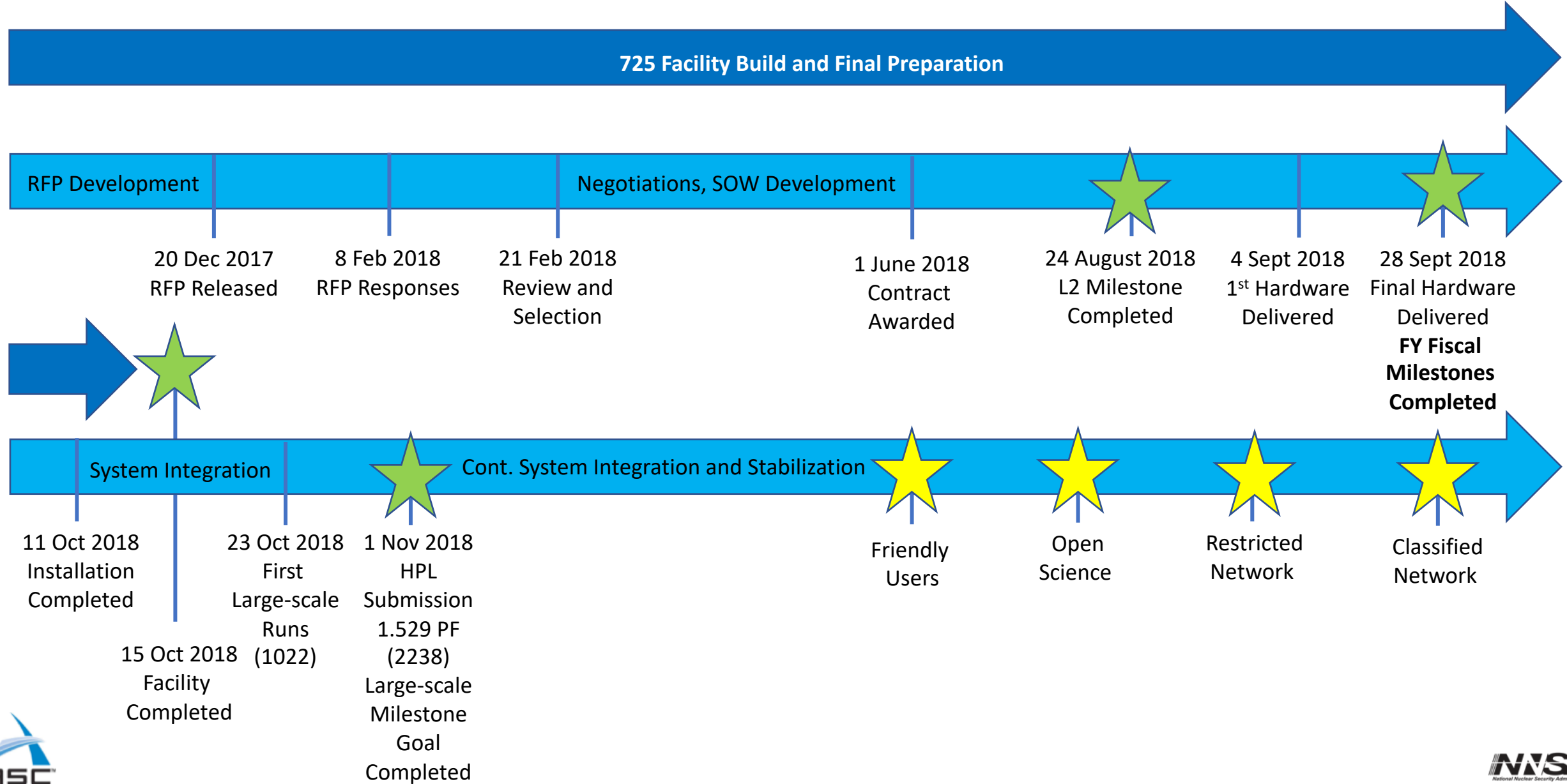
ATS/CTS Platforms

- Leadership-class systems (Petascale, Exascale, ...)
- Advanced technologies, sometimes first-of-kind
- Broad user-base
- PRODUCTION USE

Sandia's NNSA/ASC ARM Platform Evolution



Vanguard-Astra: Timeline



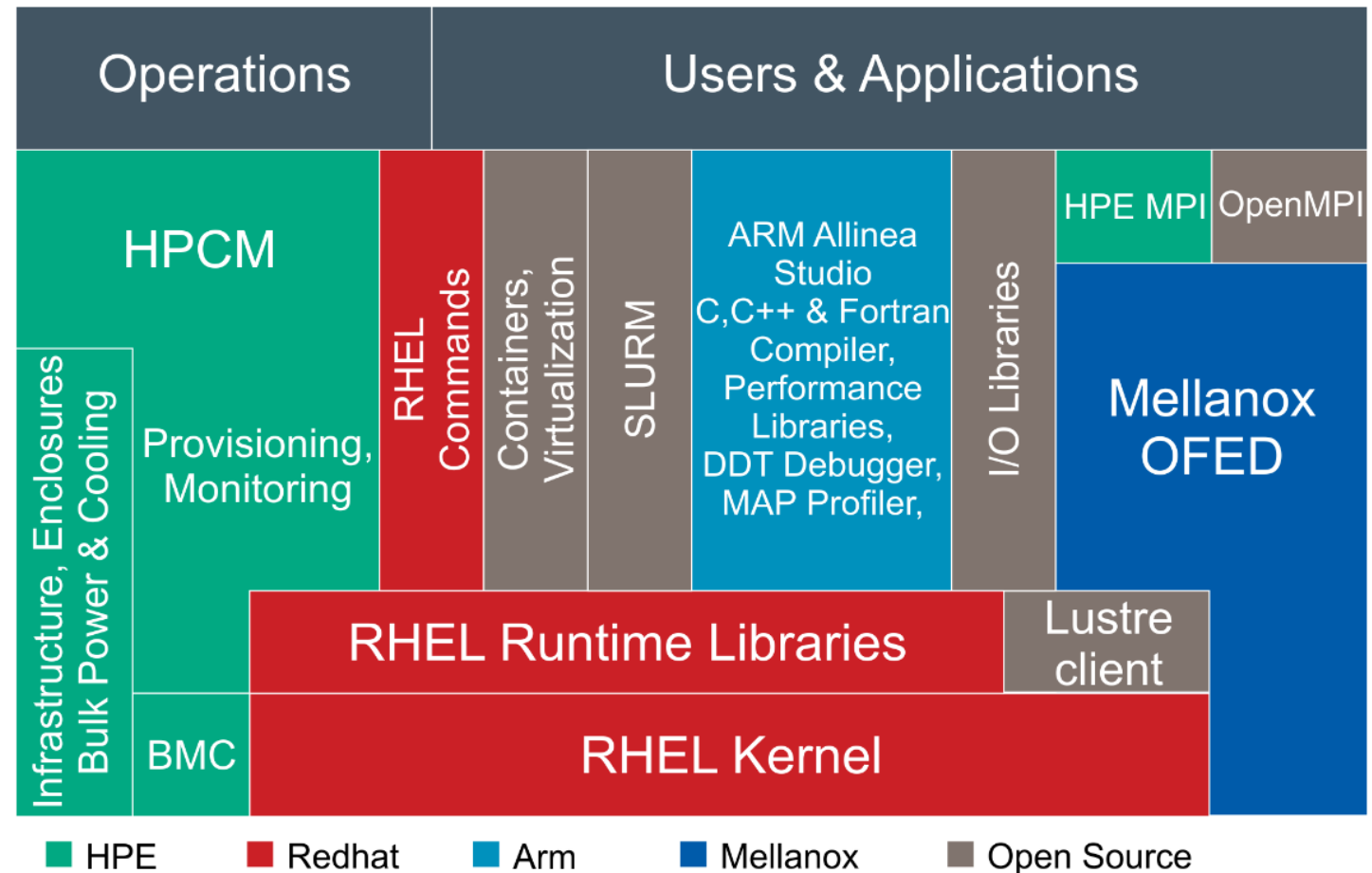
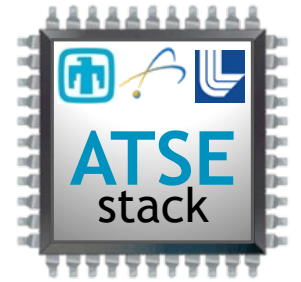


HPE's HPC Software Stack

HPE:

- HPE MPI (+ XPMEM)
- HPE Cluster Manager
- Arm:
 - Arm HPC Compilers
 - Arm Math Libraries
 - Allinea Tools
- Mellanox-OFED & HPC-X
- RedHat 7.x for aarch64


Hewlett Packard
Enterprise



Network Stack Used for HPL + HPCG Runs



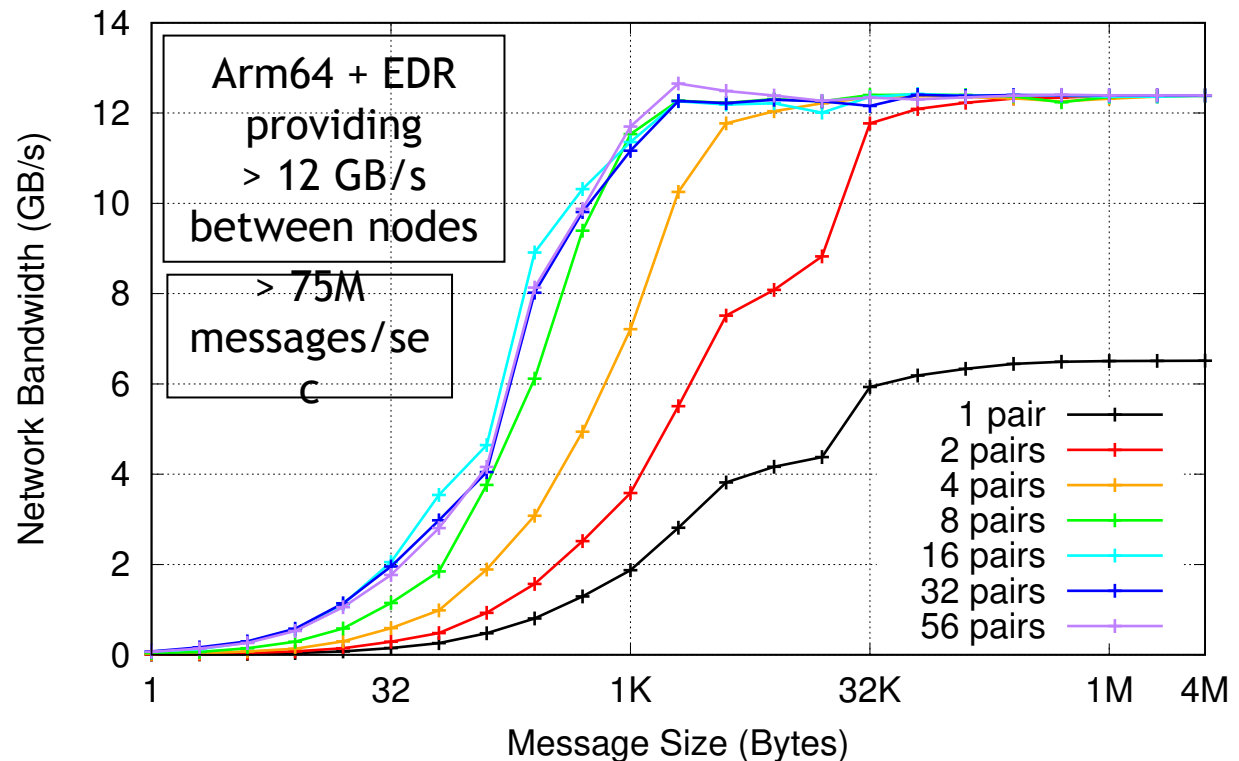
- OpenMPI 3.1.2
- UCX checkout from 20181020

Early Performance Results from Mayer InfiniBand

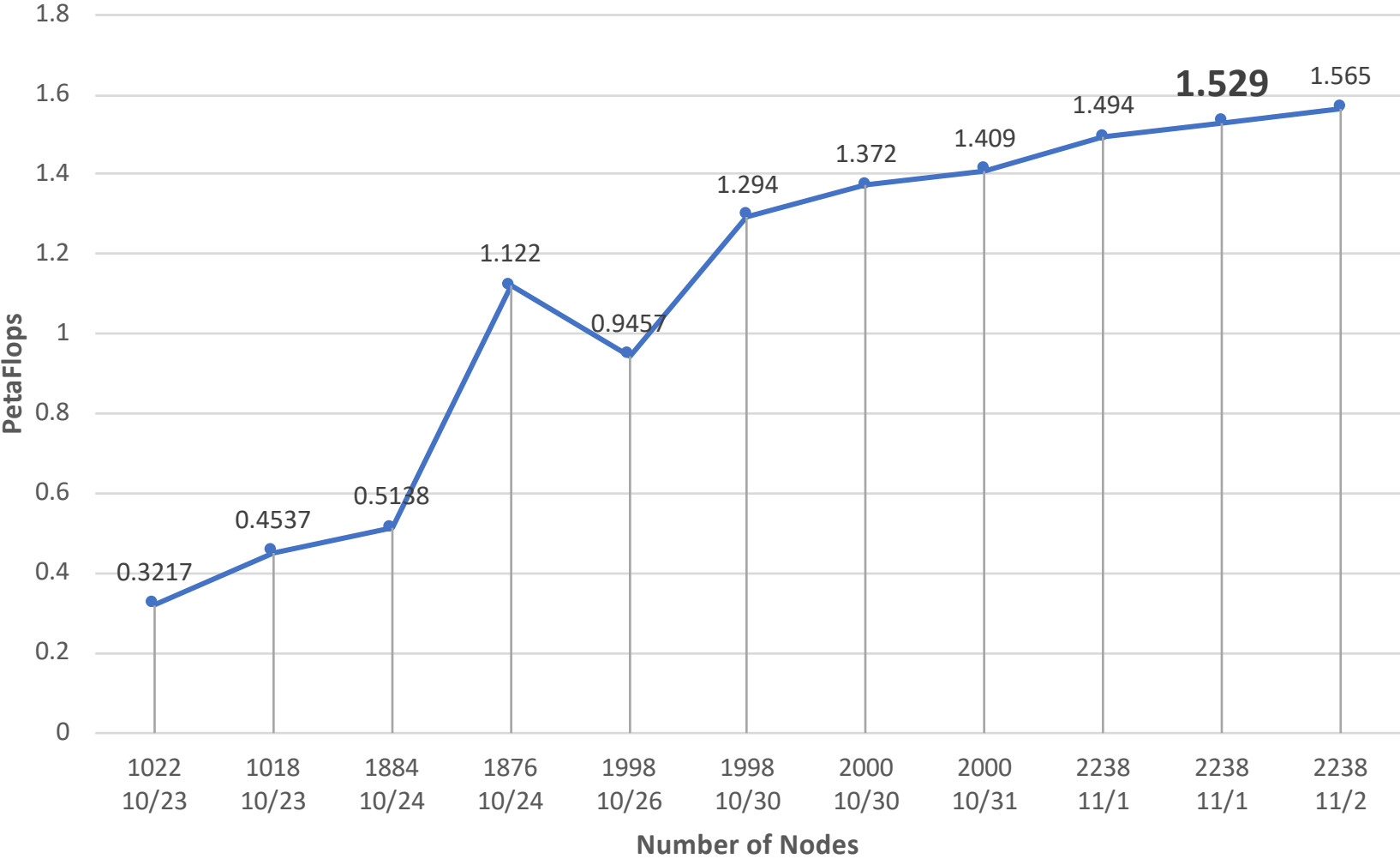
Socket Direct feature enables a single NIC to be shared by multiple host processor sockets

- Share a single physical link to reduce cabling complexity and costs
- NIC arbitrates between host processors to ensure a fair level of service
- Required some complex O/S patches early on in test systems

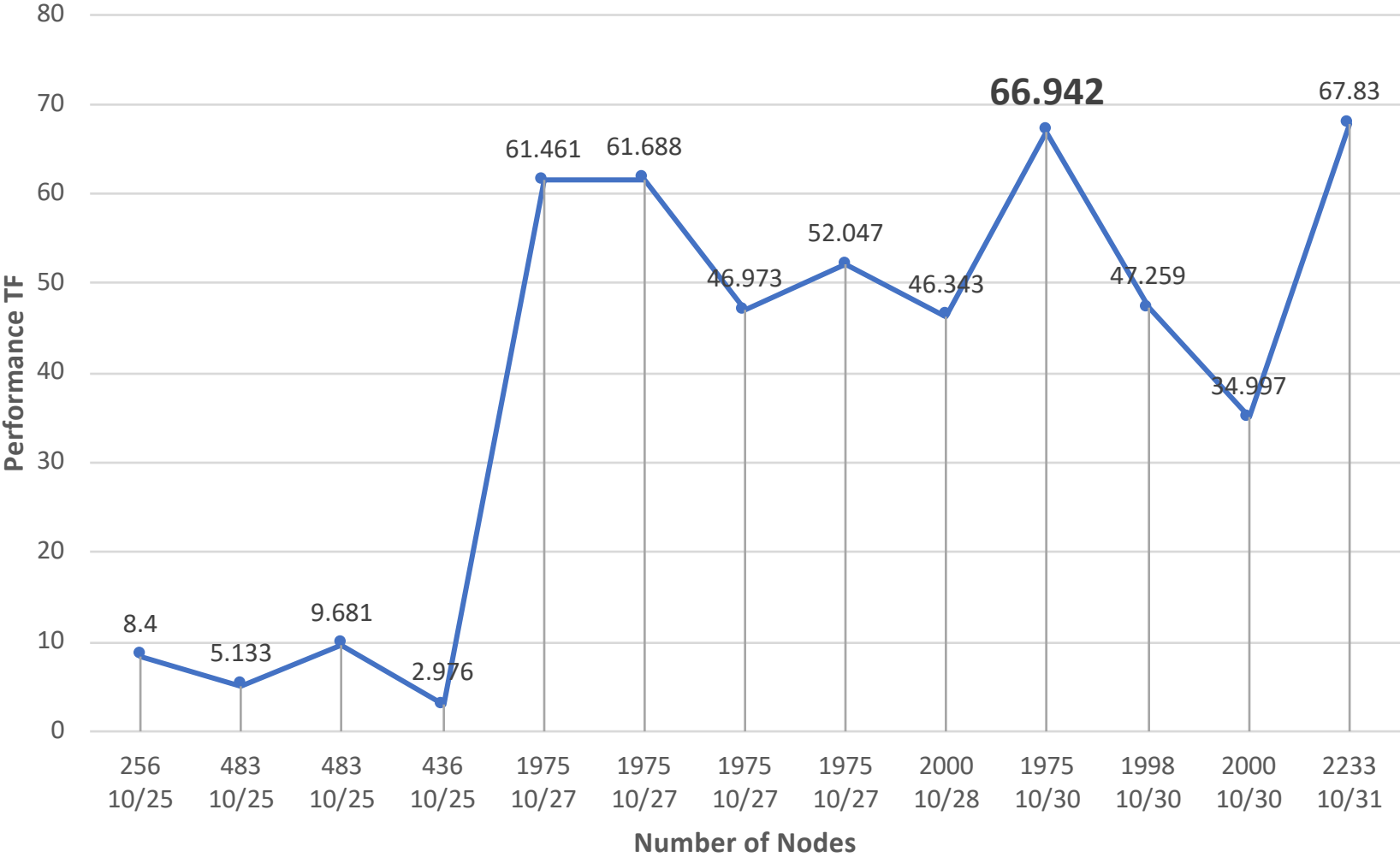
OSU MPI Multi-Network Bandwidth



HPL Performance (in PetaFlops)



HPCG Performance (TeraFlops)





Exceptional Service in the National Interest