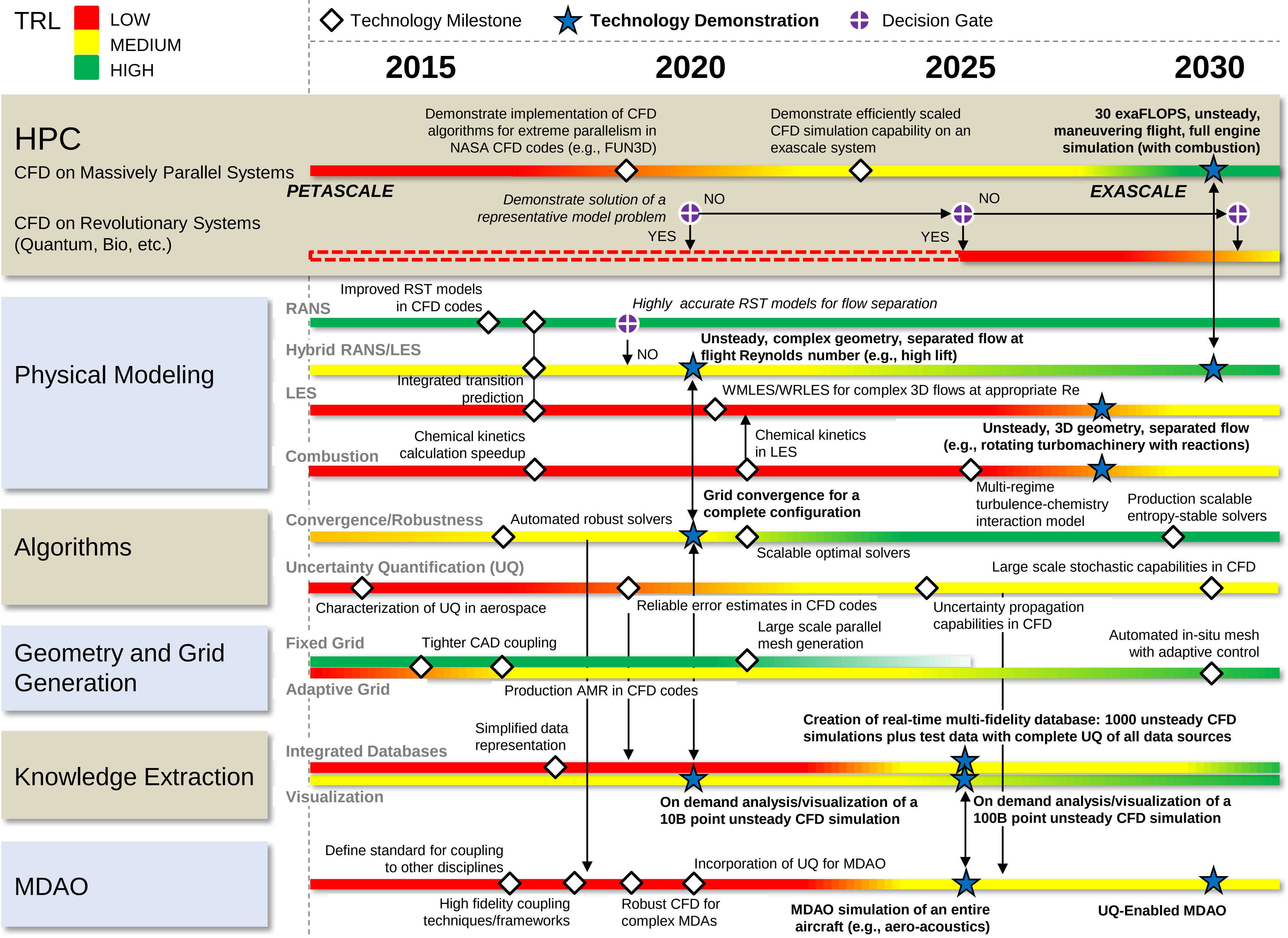


TRL 🟥 LOW 🟨 MEDIUM 🟩 HIGH

◇ Technology Milestone ★ **Technology Demonstration** ⊕ Decision Gate

2015 **2020** **2025** **2030**

HPC

CFD on Massively Parallel Systems

Demonstrate implementation of CFD algorithms for extreme parallelism in NASA CFD codes (e.g., FUN3D)

Demonstrate efficiently scaled CFD simulation capability on an exascale system

30 exaFLOPS, unsteady, maneuvering flight, full engine simulation (with combustion)

PETASCALE

EXASCALE

CFD on Revolutionary Systems (Quantum, Bio, etc.)

Demonstrate solution of a representative model problem

NO

YES

NO

YES

Physical Modeling

RANS

Improved RST models in CFD codes

Highly accurate RST models for flow separation

Hybrid RANS/LES

NO

Unsteady, complex geometry, separated flow at flight Reynolds number (e.g., high lift)

LES

Integrated transition prediction

WMLES/WRLES for complex 3D flows at appropriate Re

Combustion

Chemical kinetics calculation speedup

Chemical kinetics in LES

Unsteady, 3D geometry, separated flow (e.g., rotating turbomachinery with reactions)

Multi-regime turbulence-chemistry interaction model

Algorithms

Convergence/Robustness

Automated robust solvers

Grid convergence for a complete configuration

Scalable optimal solvers

Production scalable entropy-stable solvers

Uncertainty Quantification (UQ)

Characterization of UQ in aerospace

Reliable error estimates in CFD codes

Uncertainty propagation capabilities in CFD

Large scale stochastic capabilities in CFD

Geometry and Grid Generation

Fixed Grid

Tighter CAD coupling

Large scale parallel mesh generation

Adaptive Grid

Production AMR in CFD codes

Automated in-situ mesh with adaptive control

Knowledge Extraction

Integrated Databases

Simplified data representation

Creation of real-time multi-fidelity database: 1000 unsteady CFD simulations plus test data with complete UQ of all data sources

Visualization

On demand analysis/visualization of a 10B point unsteady CFD simulation

On demand analysis/visualization of a 100B point unsteady CFD simulation

MDAO

Define standard for coupling to other disciplines

High fidelity coupling techniques/frameworks

Robust CFD for complex MDAs

Incorporation of UQ for MDAO

MDAO simulation of an entire aircraft (e.g., aero-acoustics)

UQ-Enabled MDAO

The Future of High-End Computing

Dr. Roy Campbell
9 January 2020

Distribution A

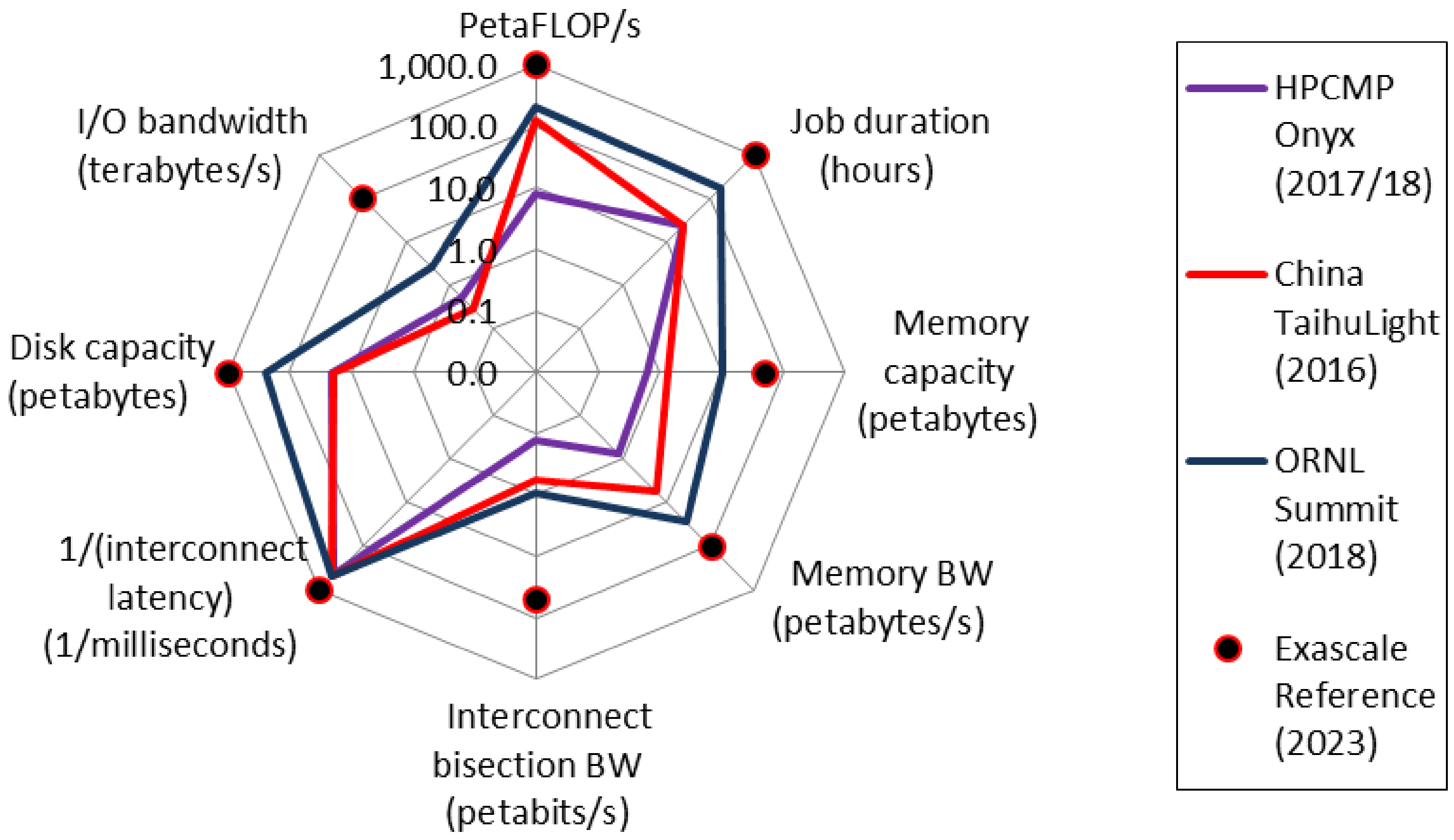
High-End Computing

Most Difficult Defense Use Cases

- **Hypersonic vehicles**
  - Aerodynamic structure
  - Advanced combustion
  - Durable/resilient materials
  - Precise maneuverability
  - Numerous payload options
- **Trade space analysis, technical issue resolution, and life-cycle management**
  - Airplanes
  - Ships
  - Rotorcraft
  - Ground vehicles
  - Submarines
  - Missiles
- **Autonomous swarms**
  - Air, land, and sea
  - Coordinate thousands of devices for complex missions
  - Account for physics of each device
  - Little to no human intervention
  - Adapt to current situation and environment
- **Complete situational awareness**
  - Global scale
  - Fine-grained local refinement
  - Ingestion and curation of large volumes of disparate information in real-time
  - Uncover true threats
  - Overcome deception

High-End System Design

Headwind: Difficulty Maintaining System Balance



High-End System Design

Headwind: Difficulty Maintaining System Balance *(cont.)*

	HPCMP Onyx (2017/18)	China TaihuLight (2016)	ORNL Summit (2018)	Exascale Reference (2023)
Capacity per FLOP/s				
Memory *(bytes)*	0.08	0.01	0.05	0.05
	163%	**21%**	**101%**	**100%**
I/O *(bytes)*	2.8	0.2	1.2	1.0
	275%	**16%**	**125%**	**100%**
Throughput per FLOP				
Memory *(bytes)*	0.10	0.04	0.14	0.10
	100%	**45%**	**135%**	**100%**
Interconnect *(bytes)*	0.002	0.001	0.001	0.006
	34%	**9%**	**9%**	**100%**
I/O *(bytes)*	7.E-05	2.E-06	1.E-05	1.E-04
	67%	**2%**	**12%**	**100%**

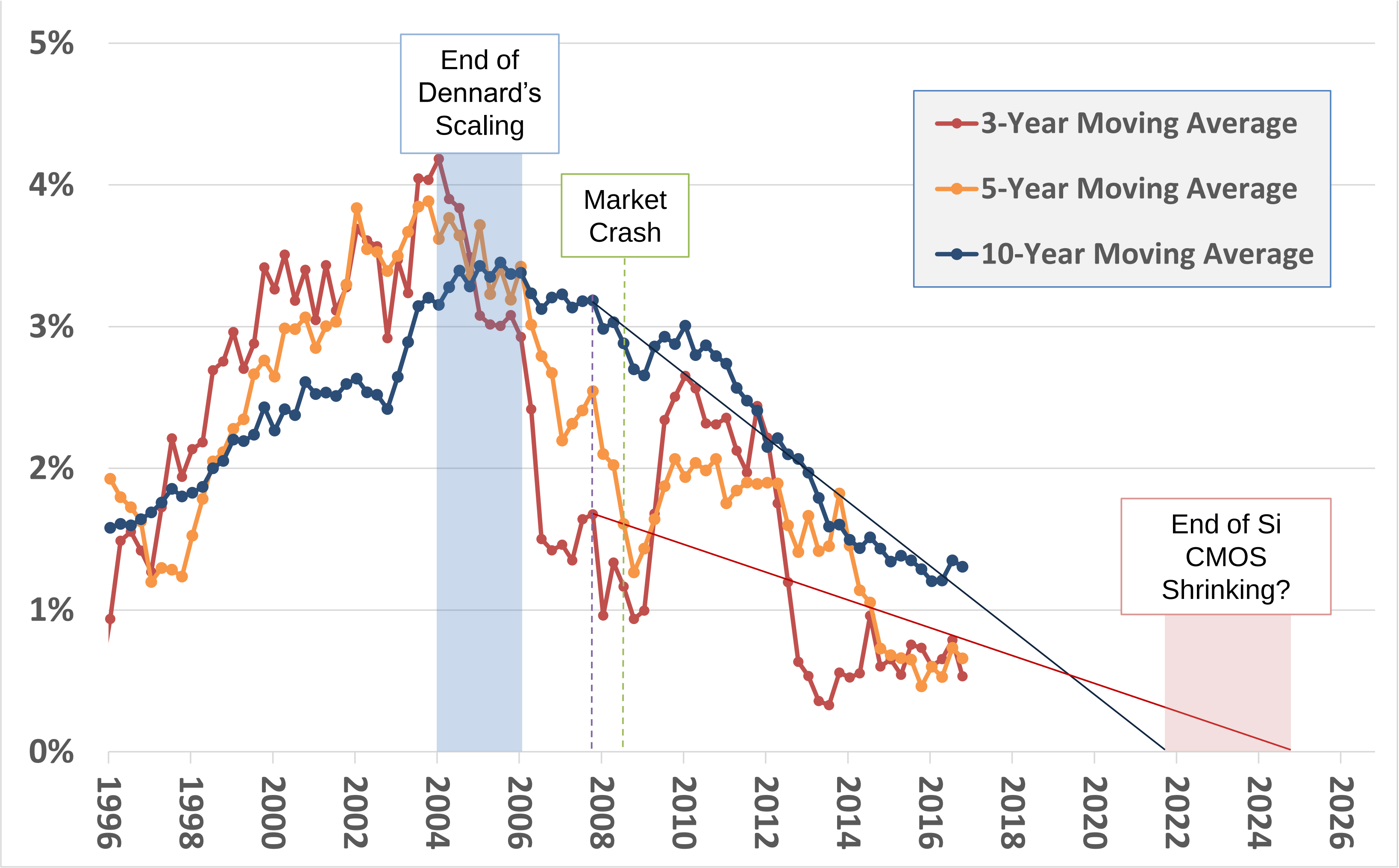
decimal
target attribute divided by peak FLOP/s

percentage
target decimal divided by exascale reference decimal

red
low value relative to exascale reference

U.S. Productivity Growth★

Headwind: End of Silicon CMOS Shrinking



★ *Bureau of Labor Statistics (8 Mar 2017): Change in Non-Farm Output Per Labor Hour*

Chip Fabrication Challenges

Headwind: End of General Purpose / Von Neumann Computing

Major foundries ***(late 2019 status)***
Global Foundries *(14nm)*, Intel *(10nm)*, TSMC *(7nm)*, Samsung *(7nm)*

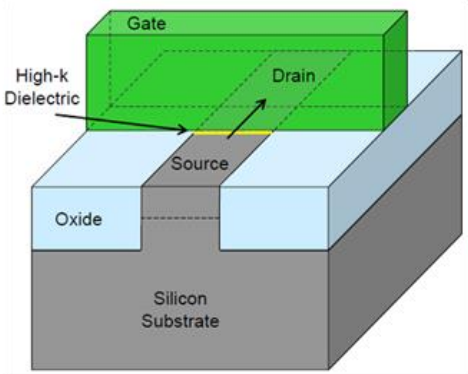
Feature size

- **32nm – Intel / Planar *(TI-11/12)***
- **22nm – Intel / Tri-Gate *(TI-13/14/15)***
- **14nm – Intel / Tri-Gate *(TI-16/17)***
- **7nm – TSMC / FinFET *(TI-18)***
- **Silicon – 0.2nm covalent diameter**

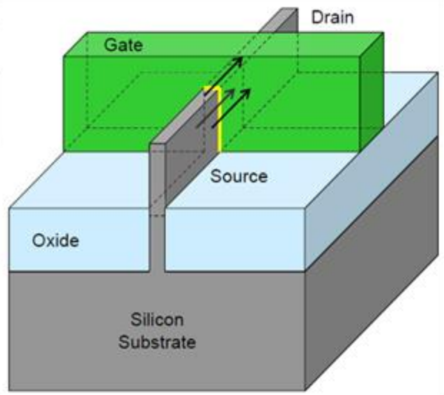
New lithography method
13.5nm λ extreme UV (EUV)
(5nm TSMC & Samsung 2020)

New fabrication methods
gate-all-around (3nm Samsung 2021), 3D stacking

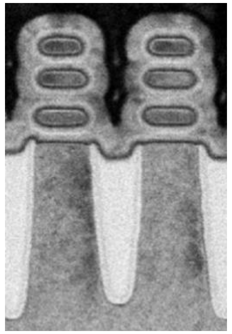
New materials
indium gallium arsenide (InGaAs) + indium phosphide (InP)



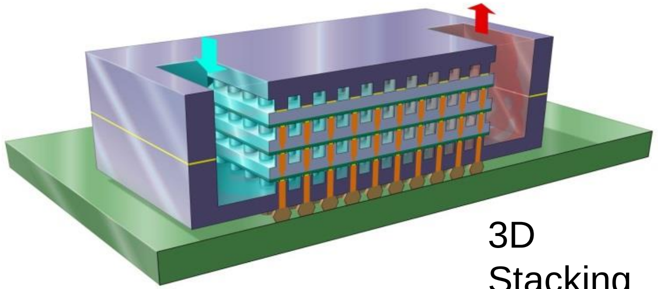
Planar



Tri-Gate



Gate All Around

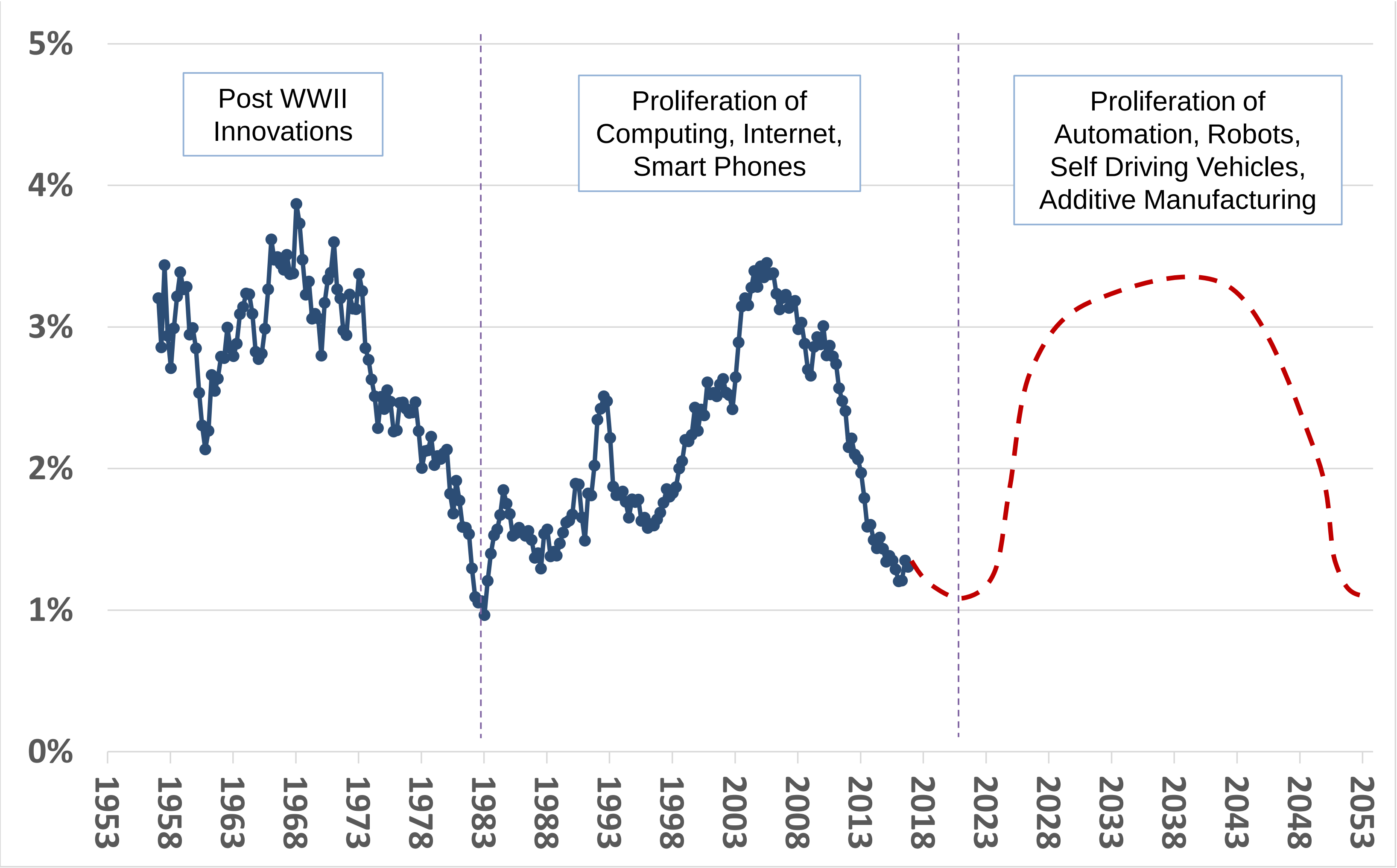


3D Stacking

FY22	FY23	FY24	FY25	FY26-FY30	FY31-FY35	FY36-FY40
End of Si CMOS Shrinking			3D Stacking		Age of Special Purpose / Non Von Neumann Computing	

U.S. Productivity Growth *(10Y-MA)*★

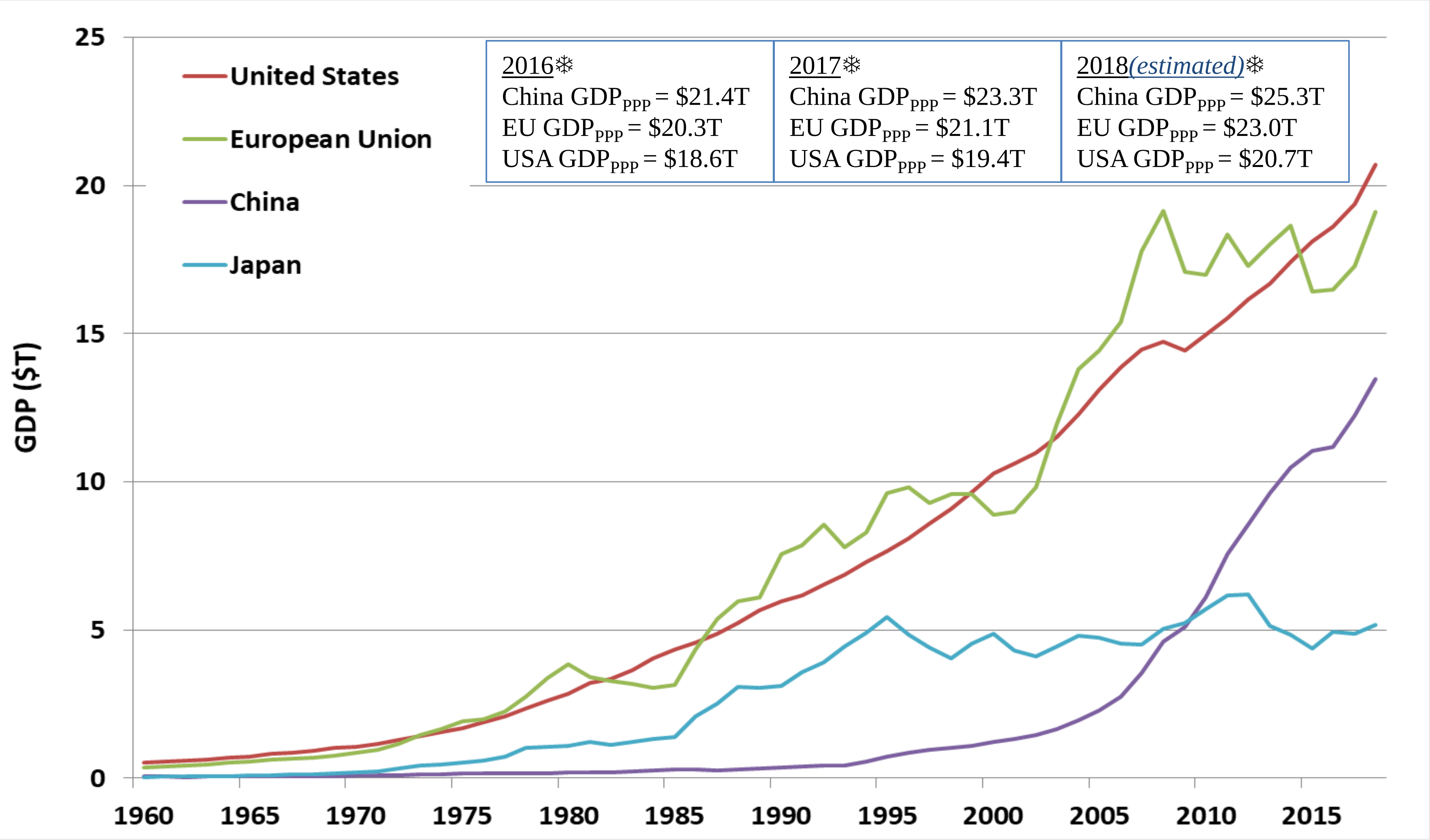
Headwind: Transition to New Productivity Driver



★ *Bureau of Labor Statistics (8 Mar 2017): Change in Non-Farm Output Per Labor Hour*

Gross Domestic Product *(GDP)*★

Headwind: Challenge to U.S. Economic Leadership



2016❄	2017❄	2018*(estimated)*❄
China $\text{GDP}_{\text{PPP}} = \$21.4\text{T}$	China $\text{GDP}_{\text{PPP}} = \$23.3\text{T}$	China $\text{GDP}_{\text{PPP}} = \$25.3\text{T}$
EU $\text{GDP}_{\text{PPP}} = \$20.3\text{T}$	EU $\text{GDP}_{\text{PPP}} = \$21.1\text{T}$	EU $\text{GDP}_{\text{PPP}} = \$23.0\text{T}$
USA $\text{GDP}_{\text{PPP}} = \$18.6\text{T}$	USA $\text{GDP}_{\text{PPP}} = \$19.4\text{T}$	USA $\text{GDP}_{\text{PPP}} = \$20.7\text{T}$

★ *World Bank (2018)*

❄ *PPP = purchasing power parity*

Supercomputing Top 10 *(June 2019)*

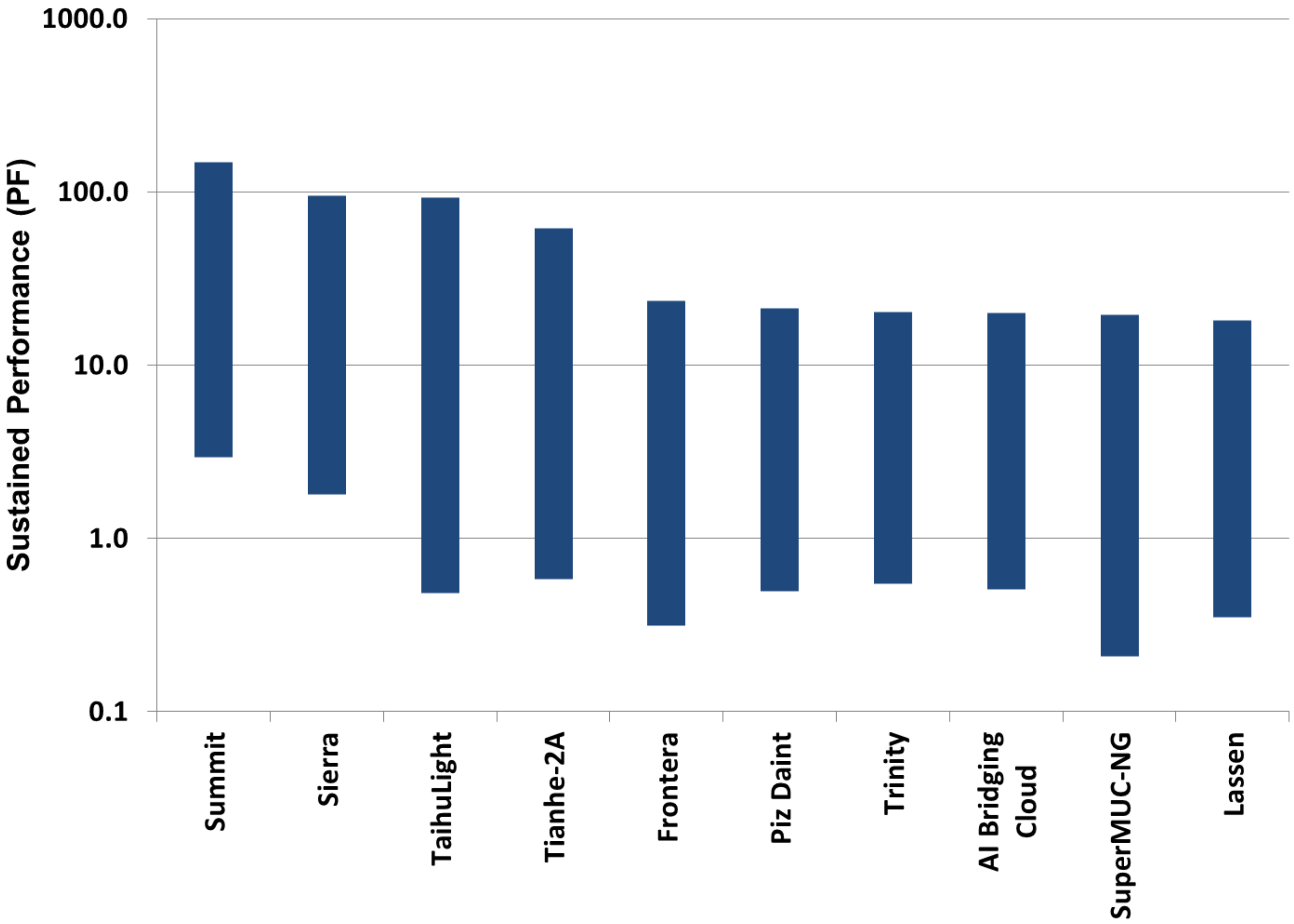
Headwind: Challenge to U.S. Technological Leadership

HPL Rank	System Name	System Architecture	Site	Cores (M)	HPL R_{max} (PF)	R_{peak} (PF)	HPL R_{max} / R_{peak}	Power (MW)	GF/W	HPCG R_{max} (PF)	HPCG R_{max} / R_{peak}	HPCG Rank
1	Summit	IBM AC922: IBM P9 (22 cores) + NVIDIA Volta + dual-rail IB-4X EDR interconnect	USA (DOE/ORNL)	2.4	148.6	200.8	74%	10	14.7	2.93	1.5%	1
2	Sierra	IBM S922LC: IBM P9 (22 cores) + NVIDIA Volta + dual-rail IB-4X EDR interconnect	USA (DOE/LLNL)	1.6	94.6	125.7	75%	7	12.7	1.80	1.4%	2
3	TaihuLight	Sunway (260 cores) + Chinese interconnect	China (Wuxi)	10.6	93.0	125.4	74%	15	6.1	0.48	0.4%	8
4	Tianhe-2A	Intel IvyBridge (12 cores) + NUDT Matrix-2000 + Chinese interconnect	China (Guangzhou)	5.0	61.4	100.7	61%	18	3.3	0.58	0.6%	4
5	Frontera	Dell: Intel Cascade Lake (28 cores) + IB-4X HDR	USA (NSF/TACC)	0.4	23.5	38.7	61%	6	3.9	0.31	0.8%	
6	Piz Daint	Cray XC50: Intel Haswell (12 cores) + NVIDIA Pascal + Cray Aries interconnect	Switzerland	0.4	21.2	27.2	78%	2	8.9	0.50	1.8%	7
7	Trinity	Cray XC40: Intel Haswell (16 cores) + Intel KNL (68 cores) + Cray Aries interconnect	USA (LANL/SNL)	1.0	20.2	41.5	49%	8	2.7	0.55	1.3%	5
8	AI Bridging Cloud	Fujitsu: Intel Skylake (20 cores) + NVIDIA Volta + IB-4X EDR interconnect	Japan	0.4	19.9	32.6	61%	2	12.1	0.51	1.6%	6
9	SuperMUC-NG	Lenovo: Intel Skylake (24 cores) + Intel Omni-Path interconnect	Germany	0.3	19.5	26.9	72%	4	5.2	0.21	0.8%	
10	Lassen	IBM S922LC: IBM P9 (22 cores) + NVIDIA Volta + dual-rail IB-4X EDR interconnect	USA (DOE/LLNL)	0.3	18.2	23.0	79%	1	15.1	0.35	1.5%	

Sustained Performance

Headwind: Challenge to U.S. Technological Leadership

- **HPL: dense linear algebra *(easy)***
- **HPCG: sparse linear algebra *(hard)***
- **USA the first to exceed 100PF for HPL *(easy)***
- **USA the first to exceed 1PF for HPCG *(hard)***
- **Application performance often similar to HPCG**



High-End Computing

Strategic Challenges

- **Difficulty maintaining system balance**
  - Traditional Von Neumann data flows become challenging
  - Traditional parallel approaches become challenging
- **End of general purpose / Von Neumann computing**
  - Quickly reaching the end of Silicon CMOS shrinking
  - Other approaches not viable (yet), provide short-term relief, or are not broadly applicable
  - Only long-term options: (a) special purpose computing, (b) non Von Neumann computing
- **Transition to new productivity driver**
  - Compute-centric advents no longer benefit from economic interest; pace slows
  - Data-centric advents benefit from economic interest; pace increases
  - Compute-centric forced to converge with (or overcome by) data-centric
- **Challenge to U.S. economic and technological leadership**
  - Increasing difficulty for U.S. to drive industry roadmaps
  - Influence through U.S. Government funding both temporary and waning

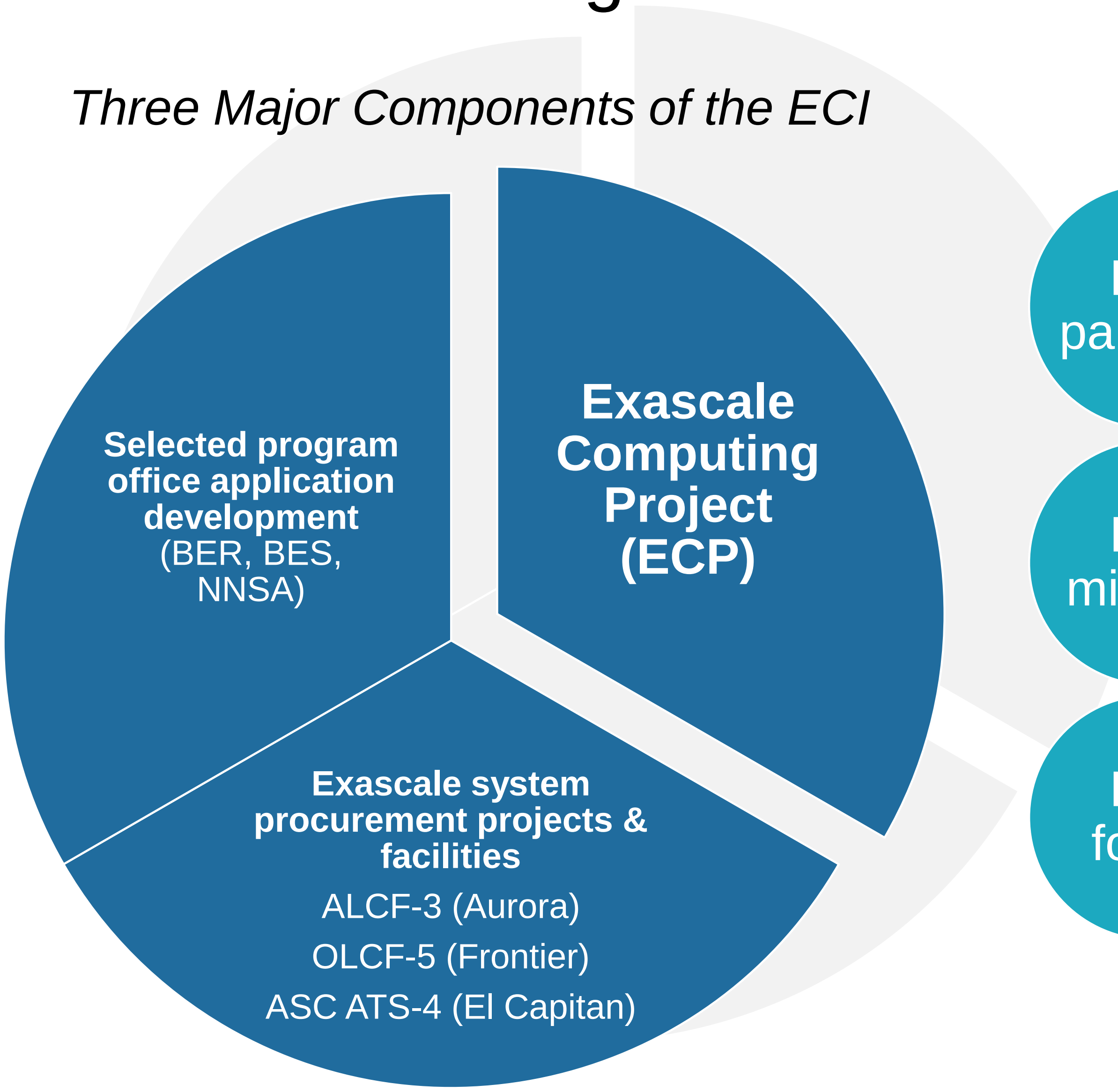
Can DOE's Exascale Computing Project Help the AIAA Community?

Douglas B. Kothe, Oak Ridge National Laboratory
Director, Exascale Computing Project

AIAA SciTech Forum 360
Orlando, FL
January 9, 2019

DOE Exascale Program: The Exascale Computing Initiative (ECI)

Three Major Components of the ECI



Exascale Computing Project (ECP)

Selected program office application development (BER, BES, NNSA)

Exascale system procurement projects & facilities

ALCF-3 (Aurora)

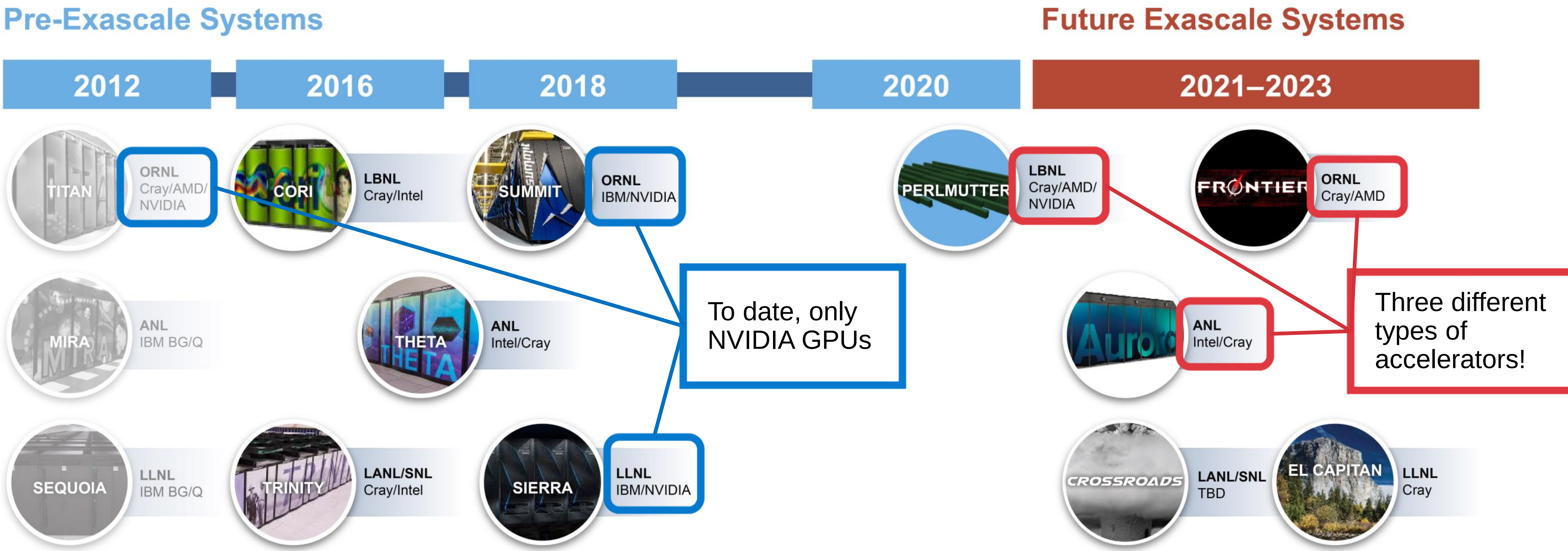
OLCF-5 (Frontier)

ASC ATS-4 (El Capitan)

ECI partners	US DOE Office of Science (SC) and National Nuclear Security Administration (NNSA)
ECI mission	Accelerate R&D, acquisition, and deployment to deliver exascale computing capability to DOE national labs by the early- to mid-2020s
ECI focus	Delivery of an *enduring and capable exascale computing capability for use by a wide range of applications* of importance to DOE and the US

Department of Energy (DOE) Roadmap to Exascale Systems

An impressive, productive lineup of *accelerated node* systems supporting DOE's mission



Pre-Exascale Systems

Future Exascale Systems

2012

TITAN
ORNL
Cray/AMD/NVIDIA

MIRA
ANL
IBM BG/Q

SEQUOIA
LLNL
IBM BG/Q

2016

CORI
LBNL
Cray/Intel

TRINITY
LANL/SNL
Cray/Intel

2018

SUMMIT
ORNL
IBM/NVIDIA

THETA
ANL
Intel/Cray

SIERRA
LLNL
IBM/NVIDIA

To date, only NVIDIA GPUs

2020

PERLMUTTER
LBNL
Cray/AMD/NVIDIA

2021–2023

FRONTIER
ORNL
Cray/AMD

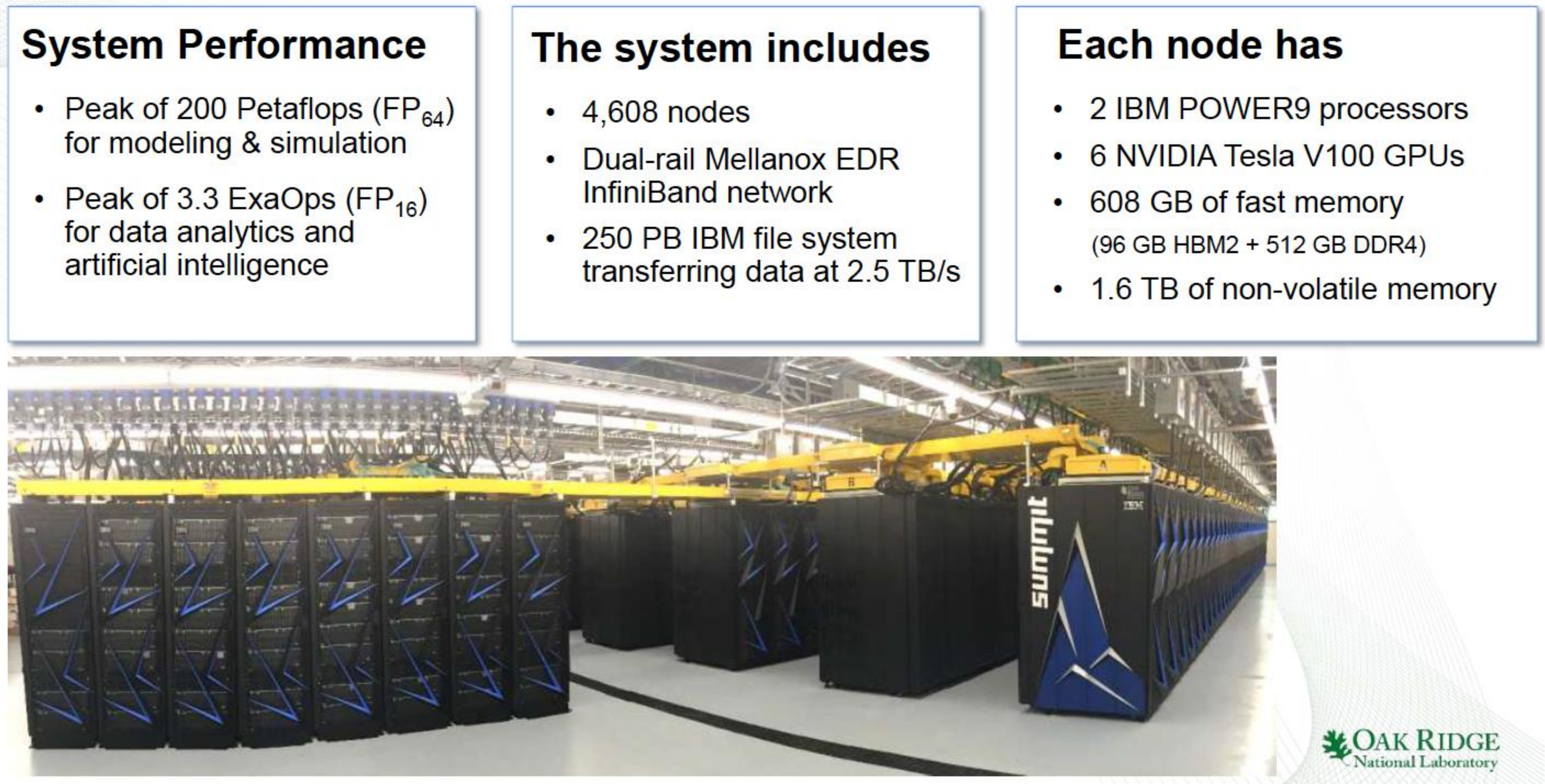
Aurora
ANL
Intel/Cray

CROSSROADS
LANL/SNL
TBD

EL CAPITAN
LLNL
Cray

Three different types of accelerators!

ORNL Summit: IBM System Overview



System Performance

- Peak of 200 Petaflops (FP_{64}) for modeling & simulation
- Peak of 3.3 ExaOps (FP_{16}) for data analytics and artificial intelligence

The system includes

- 4,608 nodes
- Dual-rail Mellanox EDR InfiniBand network
- 250 PB IBM file system transferring data at 2.5 TB/s

Each node has

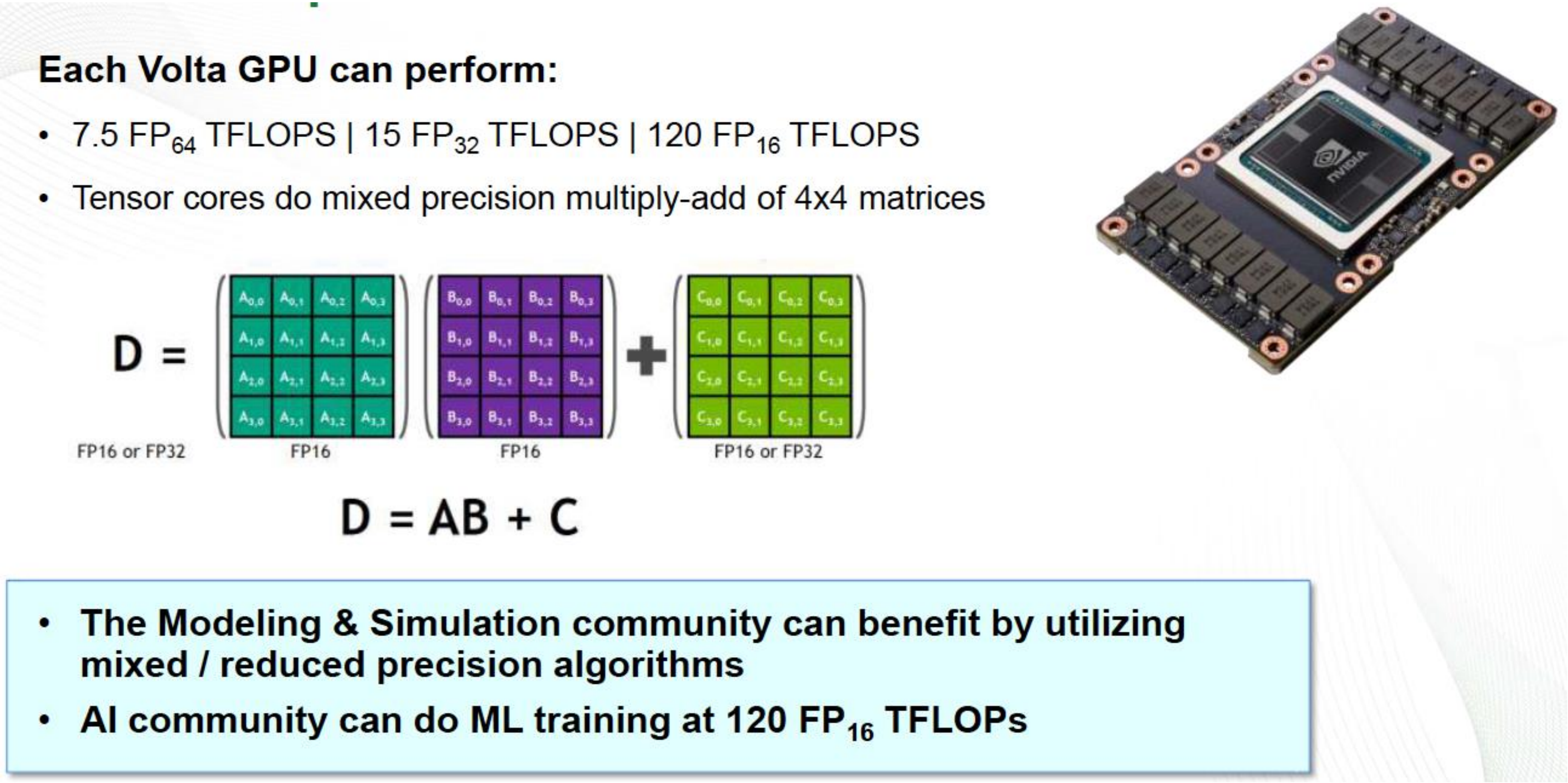
- 2 IBM POWER9 processors
- 6 NVIDIA Tesla V100 GPUs
- 608 GB of fast memory
  (96 GB HBM2 + 512 GB DDR4)
- 1.6 TB of non-volatile memory

OAK RIDGE National Laboratory

EXASCALE COMPUTING PROJECT

Summit has 27,648 NVIDIA Volta GPUs

Each with optimized AI performance



Each Volta GPU can perform:

- 7.5 FP_{64} TFLOPS | 15 FP_{32} TFLOPS | 120 FP_{16} TFLOPS
- Tensor cores do mixed precision multiply-add of 4x4 matrices

$$D = \begin{pmatrix} A_{0,0} & A_{0,1} & A_{0,2} & A_{0,3} \\ A_{1,0} & A_{1,1} & A_{1,2} & A_{1,3} \\ A_{2,0} & A_{2,1} & A_{2,2} & A_{2,3} \\ A_{3,0} & A_{3,1} & A_{3,2} & A_{3,3} \end{pmatrix} \begin{pmatrix} B_{0,0} & B_{0,1} & B_{0,2} & B_{0,3} \\ B_{1,0} & B_{1,1} & B_{1,2} & B_{1,3} \\ B_{2,0} & B_{2,1} & B_{2,2} & B_{2,3} \\ B_{3,0} & B_{3,1} & B_{3,2} & B_{3,3} \end{pmatrix} + \begin{pmatrix} C_{0,0} & C_{0,1} & C_{0,2} & C_{0,3} \\ C_{1,0} & C_{1,1} & C_{1,2} & C_{1,3} \\ C_{2,0} & C_{2,1} & C_{2,2} & C_{2,3} \\ C_{3,0} & C_{3,1} & C_{3,2} & C_{3,3} \end{pmatrix}$$

FP16 or FP32 FP16 FP16 FP16 or FP32

$$D = AB + C$$

- **The Modeling & Simulation community can benefit by utilizing mixed / reduced precision algorithms**
- **AI community can do ML training at 120 FP_{16} TFLOPs**

Summit Node

- **3 GPU per CPU**
- Coherent memory across entire node
- 1.6 TB of on-node NVM

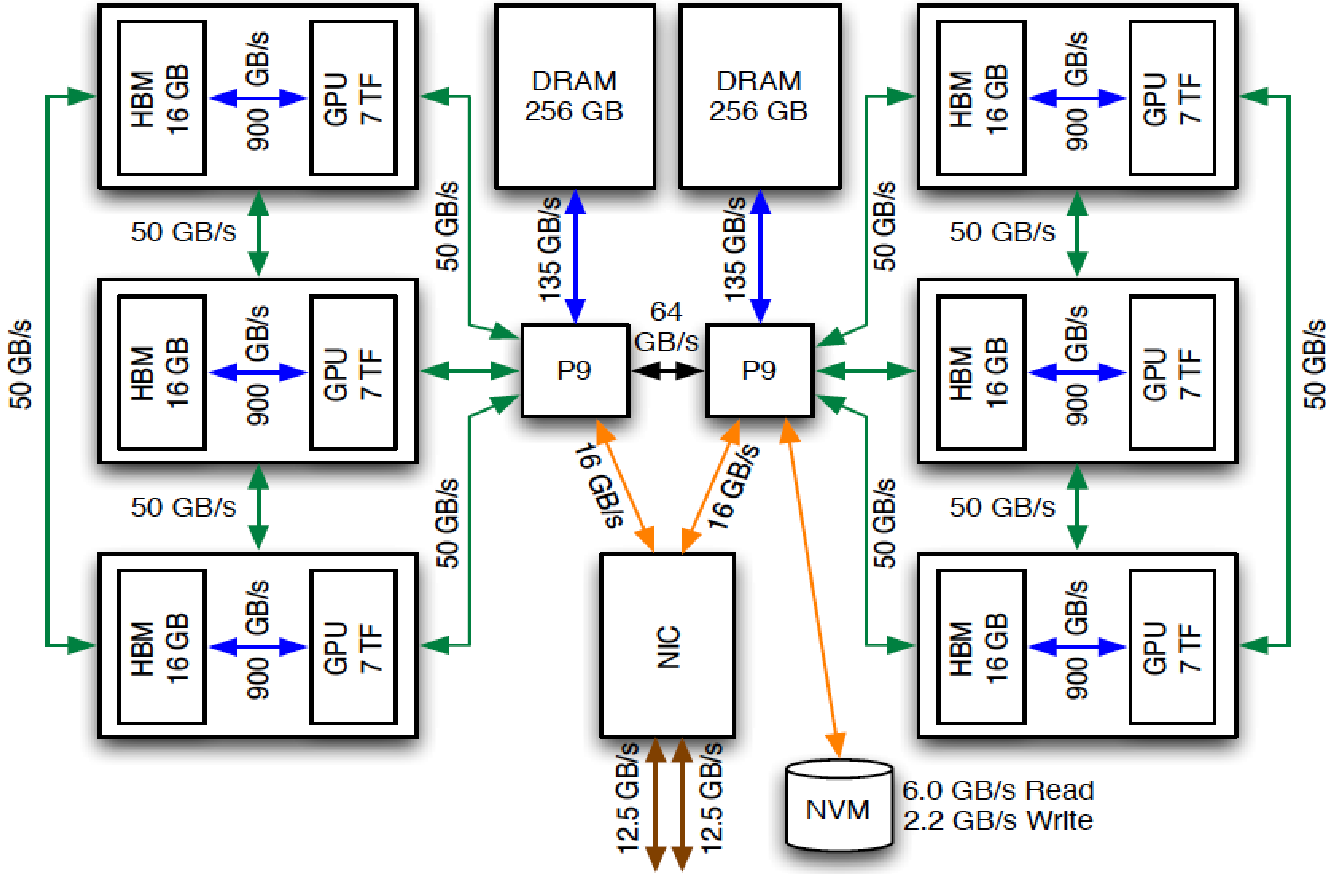
Sierra Node

- **2 GPU per CPU**
- 75 GB/s GPU-CPU links

Titan Node

- **1 GPU per CPU**
  AMD CPU / Nvidia GPU

6 GPU, 2 CPU Summit Node



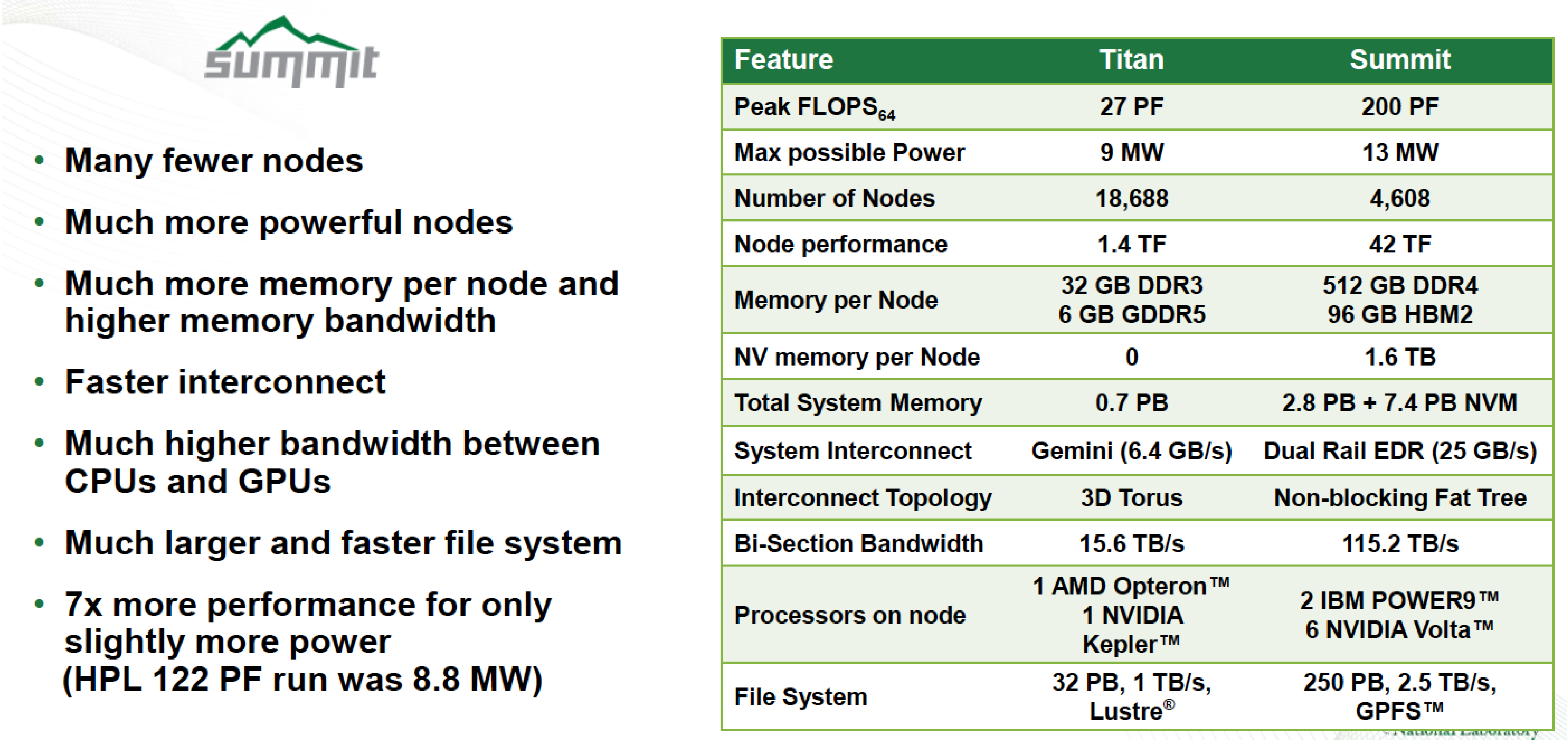
TF	42 TF (6x7 TF)
HBM	96 GB (6x16 GB)
DRAM	512 GB (2x16x16 GB)
NET	25 GB/s (2x12.5 GB/s)
MMsg/s	83

↔ HBM/DRAM Bus (aggregate B/W)
↔ NVLINK
↔ X-Bus (SMP)
↔ PCIe Gen4
↔ EDR IB

HBM & DRAM speeds are aggregate (Read+Write).
All other speeds (X-Bus, NVLink, PCIe, IB) are bi-directional.

What Makes Summit's Architecture Better Than Titan?

(these same concepts are being carried over into Frontier)



- **Many fewer nodes**
- **Much more powerful nodes**
- **Much more memory per node and higher memory bandwidth**
- **Faster interconnect**
- **Much higher bandwidth between CPUs and GPUs**
- **Much larger and faster file system**
- **7x more performance for only slightly more power (HPL 122 PF run was 8.8 MW)**

Feature	Titan	Summit
Peak FLOPS$_{64}$	27 PF	200 PF
Max possible Power	9 MW	13 MW
Number of Nodes	18,688	4,608
Node performance	1.4 TF	42 TF
Memory per Node	32 GB DDR3 6 GB GDDR5	512 GB DDR4 96 GB HBM2
NV memory per Node	0	1.6 TB
Total System Memory	0.7 PB	2.8 PB + 7.4 PB NVM
System Interconnect	Gemini (6.4 GB/s)	Dual Rail EDR (25 GB/s)
Interconnect Topology	3D Torus	Non-blocking Fat Tree
Bi-Section Bandwidth	15.6 TB/s	115.2 TB/s
Processors on node	1 AMD Opteron™ 1 NVIDIA Kepler™	2 IBM POWER9™ 6 NVIDIA Volta™
File System	32 PB, 1 TB/s, Lustre®	250 PB, 2.5 TB/s, GPFS™

Frontier Continues the Accelerated Node Design

Begun at ORNL with Titan and continued with Summit

Frontier Node -- 4 GPUs per CPU:

- One purpose-built AMD EPYC™ processor
- Four HPC and AI optimized AMD Radeon Instinct™ GPUs
- Fully connected with high speed AMD Infinity Fabric links
- Coherent memory across the node
- 100 GB/s node injection bandwidth
- On-node NVM storage

Partnership between ORNL, Cray, and AMD

The Frontier system will be delivered in 2021

Peak FP_{64} Performance greater than 1.5 EF

Max Power Consumption 29 MW

Cray Shasta cabinets Connected by Slingshot™ interconnect

- with adaptive routing, congestion control, and quality of service

Comparison of Titan, Summit, and Frontier Systems

System Specs	Titan	Summit	Frontier
Peak	27 PF	200 PF	~1.5 EF
# cabinets	200	256	> 100
Node	1 AMD Opteron CPU 1 NVIDIA Kepler GPU	2 IBM POWER9™ CPUs 6 NVIDIA Volta GPUs	1 AMD EPYC CPU 4 AMD Radeon Instinct GPUs
On-node interconnect	PCI Gen2 No coherence across the node	NVIDIA NVLINK Coherent memory across the node	AMD Infinity Fabric Coherent memory across the node
System Interconnect	Cray Gemini network 6.4 GB/s	Mellanox Dual-port EDR IB network 25 GB/s	Cray four-port Slingshot network 100 GB/s
Topology	3D Torus	Non-blocking Fat Tree	Dragonfly
Storage	32 PB, 1 TB/s, Lustre Filesystem	250 PB, 2.5 TB/s, IBM Spectrum Scale™ with GPFS™	2-4x performance and capacity of Summit's I/O subsystem.
Near-node NVM	No	Yes	Yes
Power	9 MW	13 MW	29 MW

ECP by the Numbers

7
YEARS

$1.7B

A seven-year, $1.8B R&D effort that launched in 2016

6
CORE DOE
LABS

Six core DOE National Laboratories: Argonne, Lawrence Berkeley, Lawrence Livermore, Oak Ridge, Sandia, Los Alamos

- Staff from most of the 17 DOE national laboratories take part in the project

3
FOCUS
AREAS

Three technical focus areas: Hardware and Integration, Software Technology, Application Development supported by a Project Management Office

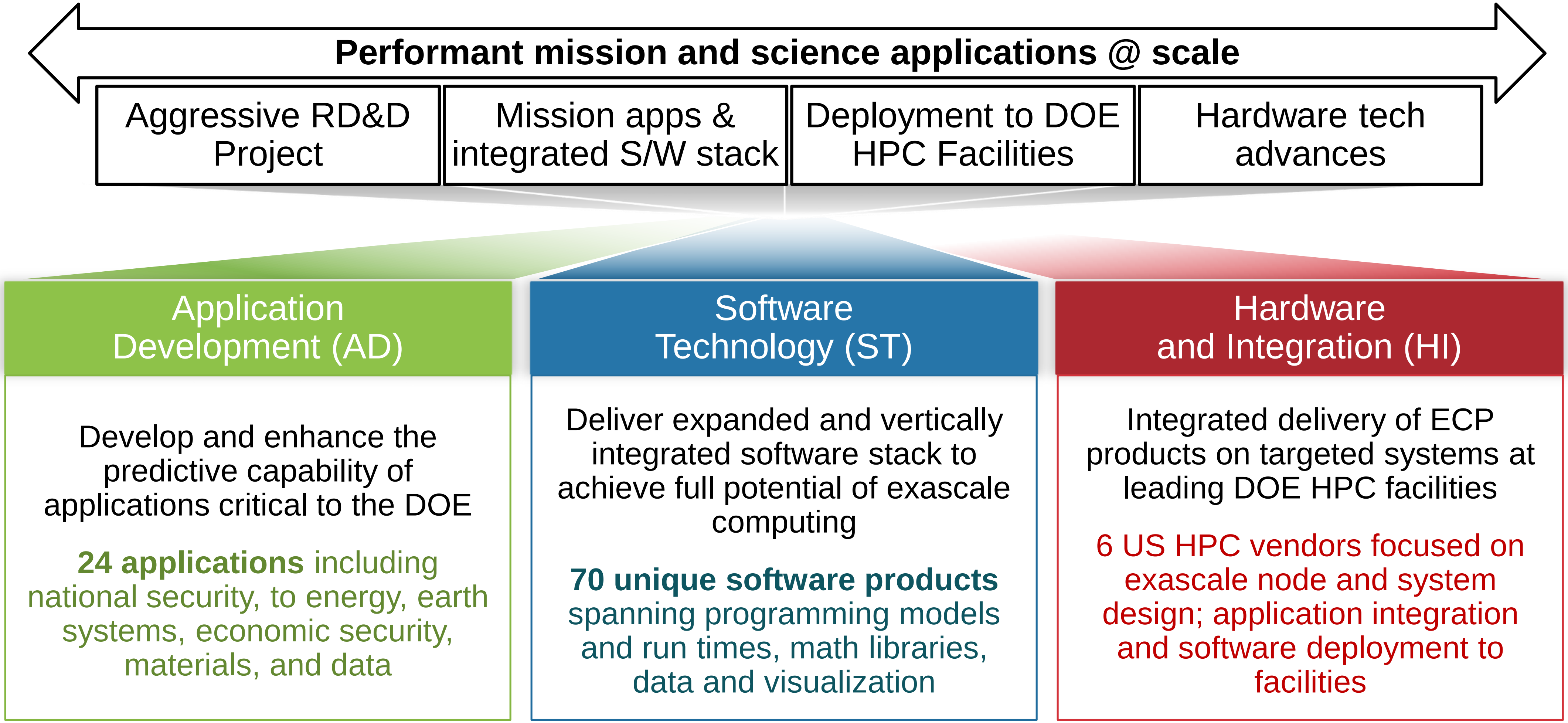
80+
R&D TEAMS

1000
RESEARCHERS

More than 80 top-notch R&D teams

Hundreds of consequential milestones delivered on schedule and within budget since project inception

ECP’s three technical areas have the necessary components to meet national goals



Performant mission and science applications @ scale

Aggressive RD&D Project

Mission apps & integrated S/W stack

Deployment to DOE HPC Facilities

Hardware tech advances

Application Development (AD)

Develop and enhance the predictive capability of applications critical to the DOE

24 applications including national security, to energy, earth systems, economic security, materials, and data

Software Technology (ST)

Deliver expanded and vertically integrated software stack to achieve full potential of exascale computing

70 unique software products spanning programming models and run times, math libraries, data and visualization

Hardware and Integration (HI)

Integrated delivery of ECP products on targeted systems at leading DOE HPC facilities

6 US HPC vendors focused on exascale node and system design; application integration and software deployment to facilities

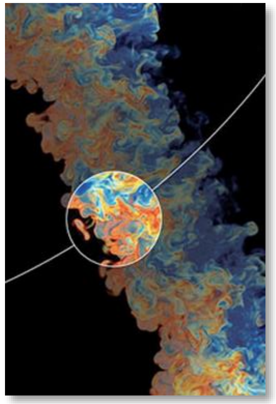
ECP applications target national problems in DOE mission areas

National security

Next-generation, **stockpile stewardship** codes

Reentry-vehicle-environment simulation

Multi-physics science simulations **of high-energy density physics** conditions



Energy security

Turbine **wind plant** efficiency

Design and commercialization of **SMR**s

Nuclear fission and fusion reactor **materials design**

Subsurface use for **carbon capture**, petroleum extraction, waste disposal

High-efficiency, low-emission **combustion engine** and gas turbine design

Scale up of **clean fossil fuel** combustion

Biofuel catalyst design

Economic security

Additive manufacturing of qualifiable metal parts

Reliable and efficient planning of the **power grid**

Seismic hazard risk assessment

Scientific discovery

Cosmological probe of the standard model of particle physics

Validate fundamental laws of nature

Plasma wakefield accelerator design

Light source-enabled **analysis of protein and molecular structure** and design

Find, predict, and control materials and properties

Predict and control **magnetically confined fusion plasmas**

Demystify **origin of chemical elements**

Earth system

Accurate regional impact assessments in **Earth system models**

Stress-resistant crop analysis and catalytic conversion of **biomass-derived alcohols**

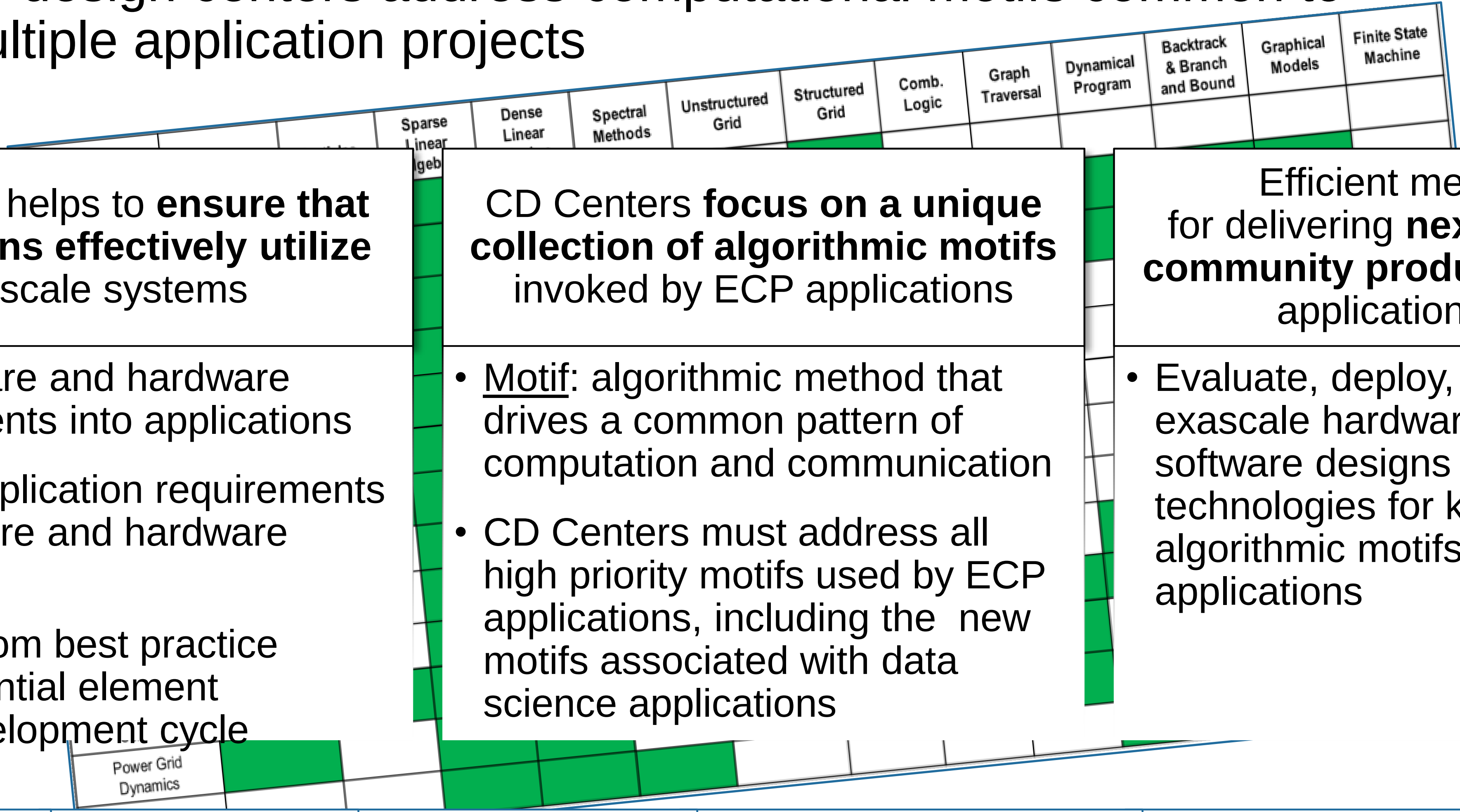
Metagenomics for analysis of biogeochemical cycles, climate change, environmental remediation

Health care

Accelerate and translate **cancer research** (partnership with NIH)

Co-design Projects

- Co-design centers address computational motifs common to multiple application projects



Co-design helps to **ensure that applications effectively utilize** exascale systems

- Pull software and hardware developments into applications
- Pushes application requirements into software and hardware RD&D
- Evolved from best practice to an essential element of the development cycle

CD Centers **focus on a unique collection of algorithmic motifs** invoked by ECP applications

- Motif: algorithmic method that drives a common pattern of computation and communication
- CD Centers must address all high priority motifs used by ECP applications, including the new motifs associated with data science applications

Efficient mechanism for delivering **next-generation community products** with broad application impact

- Evaluate, deploy, and integrate exascale hardware-aware software designs and technologies for key crosscutting algorithmic motifs into applications

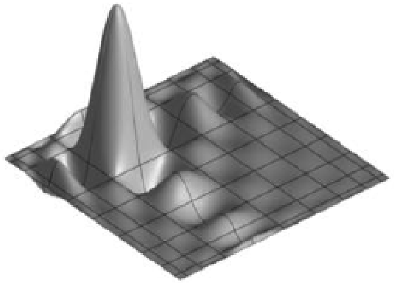
CODAR	**COPA**	**AMReX**	**CEED**	**ExaGraph**	**ExaLearn**
Data and workflows	*Particles/mesh methods*	*Block structured AMR*	*Finite element discretization*	*Graph-based algorithms*	*Machine Learning*

Center for Efficient Exascale Discretizations (CEED)

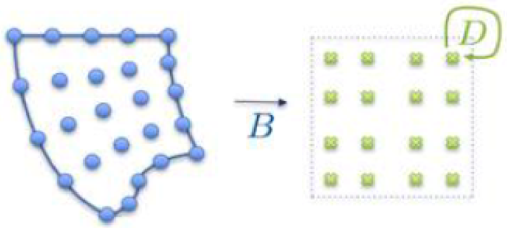
Co-Design of unstructured mesh, FE-based PDE discretizations

Goal

- Develop algorithms and software to enable more efficient HPC simulations in a wide range of PDE-based science applications.
- Focus on next-generation discretization methods: high-order finite elements on general unstructured grids.
- Target high performance on a variety of hardware: CPU, GPU, A21 in a flexible and user-friendly way.

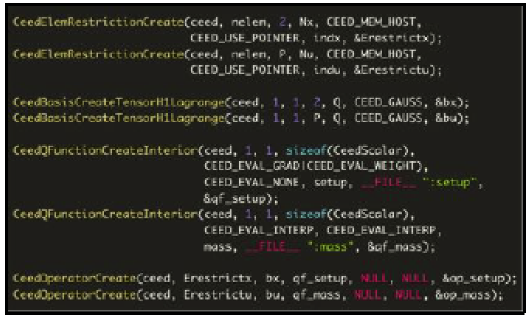


$$A = P^T G^T B^T D B G P$$



Approach

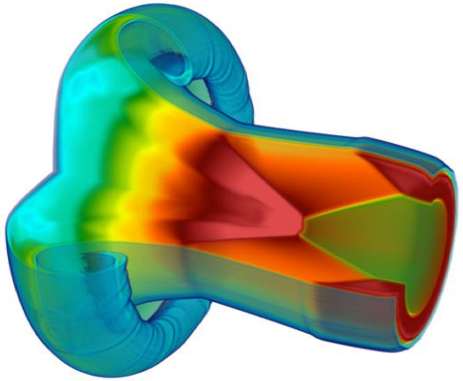
- Performance-enabling math foundation: high-order operator decomposition
- Fast kernels: CEED benchmarks, combine expertise, engage community
- Library integration: high-level API (MFEM, Nek5000), low-level API (libCEED)
- Application engagement: liaisons, CEED miniapps (Nekbone, Laghos)
- Collaborate with ECP/ST, broader community (SciDAC, xSDK, deal.ii, …)
- High-order software ecosystem: operator format, FMS, matrix-free solvers



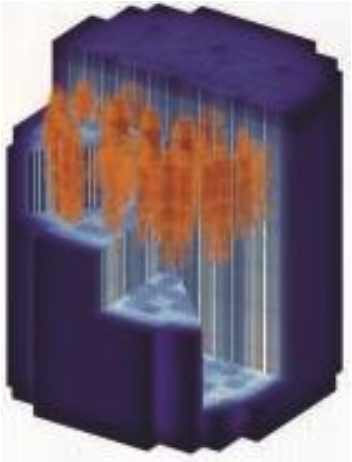
```
CeedElemRestrictionCreate(ceed, nelem, 2, Nx, CEED_MEM_HOST,
                          CEED_USE_POINTER, indx, &Erestrictx);
CeedElemRestrictionCreate(ceed, nelem, P, Nu, CEED_MEM_HOST,
                          CEED_USE_POINTER, indu, &Erestrictu);

CeedBasisCreateTensorH1Lagrange(ceed, 1, 1, 2, Q, CEED_GAUSS, &bx);
CeedBasisCreateTensorH1Lagrange(ceed, 1, 1, P, Q, CEED_GAUSS, &bu);

CeedQFunctionCreateInterior(ceed, 1, 1, sizeof(CeedScalar),
                            CEED_EVAL_GRAD|CEED_EVAL_WEIGHT),
                            CEED_EVAL_NONE, setup, __FILE__ ":setup",
                            &qf_setup);
CeedQFunctionCreateInterior(ceed, 1, 1, sizeof(CeedScalar),
                            CEED_EVAL_INTERP, CEED_EVAL_INTERP,
                            mass, __FILE__ ":mass", &qf_mass);

CeedOperatorCreate(ceed, Erestrictx, bx, qf_setup, NULL, NULL, &op_setup);
CeedOperatorCreate(ceed, Erestrictu, bu, qf_mass, NULL, NULL, &op_mass);
```

PI: Tzanio Kolev (LLNL)

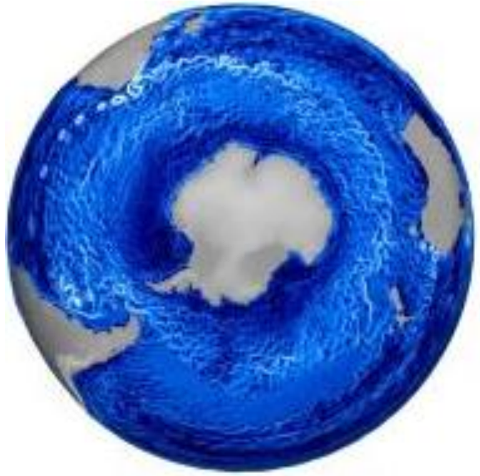
CEED is targeting several ECP applications



Compressible flow (MARBL)

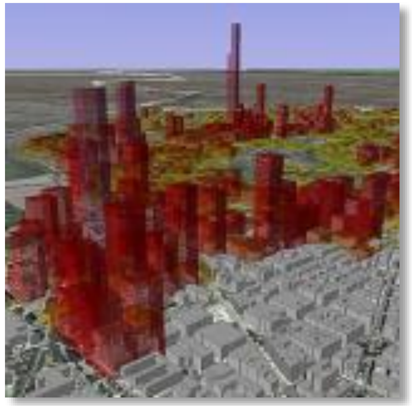


Modular Nuclear Reactors (ExaSMR)



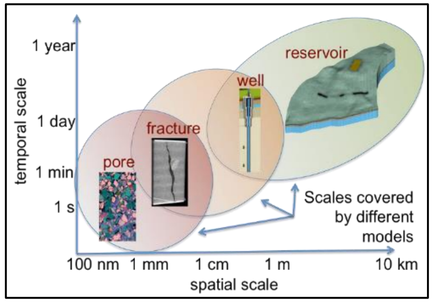
Climate (E3SM)

Wind Energy (ExaWind)

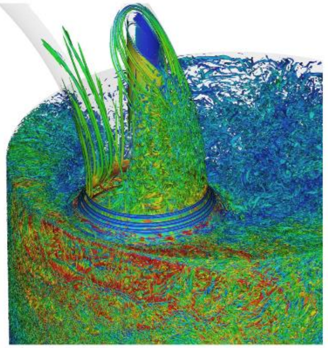


Urban systems (Urban)

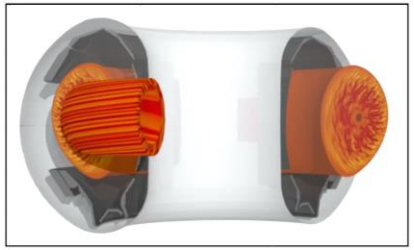
Additive Manufacturing (ExaAM)



Subsurface (GEOS)



Combustion (Nek5000)

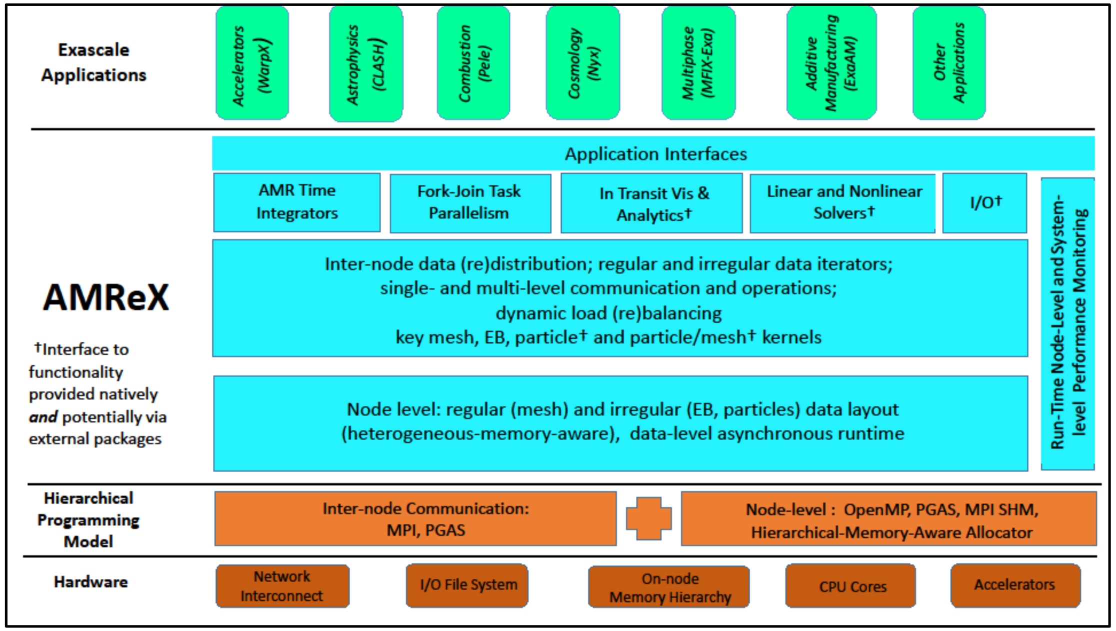


Magnetic Fusion (WDMApp)

PI: Tzanio Kolev (LLNL)

ECP’s Adaptive Mesh Refinement Co-Design Center: AMReX

- Develop and deploy software to support block-structured adaptive mesh refinement on exascale architectures
  - Core AMR functionality
  - Particles coupled to AMR meshes
  - Embedded boundary (EB) representation of complex geometry
  - Linear solvers
  - Supports two modalities of use
    - Library support for AMR
    - Framework for constructing AMR applications
- Provide direct support to ECP applications that need AMR for their application
- Evaluate software technologies and integrate with AMReXwhen appropriate
- Interact with hardware technologies / vendors



Exascale Applications

Accelerators (WarpX)
Astrophysics (CLASH)
Combustion (Pele)
Cosmology (Nyx)
Multiphase (MFIX-Exa)
Additive Manufacturing (ExaAM)
Other Applications

AMReX

†Interface to functionality provided natively ***and*** potentially via external packages

Application Interfaces

AMR Time Integrators
Fork-Join Task Parallelism
In Transit Vis & Analytics†
Linear and Nonlinear Solvers†
I/O†

Inter-node data (re)distribution; regular and irregular data iterators;
single- and multi-level communication and operations;
dynamic load (re)balancing
key mesh, EB, particle† and particle/mesh† kernels

Node level: regular (mesh) and irregular (EB, particles) data layout
(heterogeneous-memory-aware), data-level asynchronous runtime

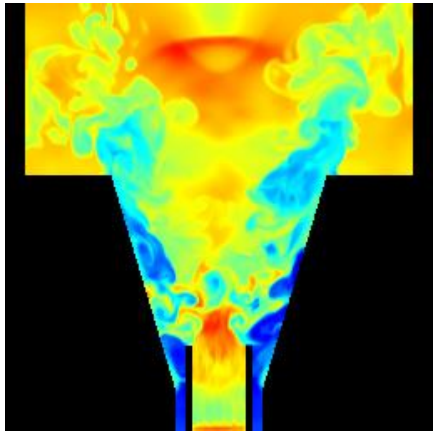
Run-Time Node-Level and System-level Performance Monitoring

Hierarchical Programming Model

Inter-node Communication: MPI, PGAS
Node-level : OpenMP, PGAS, MPI SHM, Hierarchical-Memory-Aware Allocator

Hardware

Network Interconnect
I/O File System
On-node Memory Hierarchy
CPU Cores
Accelerators



Application	Particles	ODEs	Linear Solvers	EB
Combustion	X	X	X	X
Multiphase	X		X	X
Cosmology	X	X	X	
Astrophysics	X	X	X	
Accelerators	X			

PI: John Bell (LBNL)

ECP's Co-Design Center for Online Data Analysis and Reduction

CODAR

Traditional approach:
Compute…output…analyze [offline]

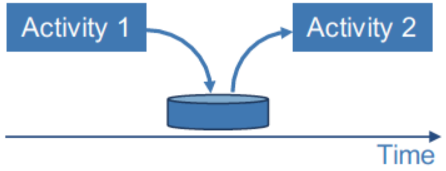
Write simulation output to secondary storage; read back for analysis

Decimate in time when simulation output rate exceeds output rate of computer

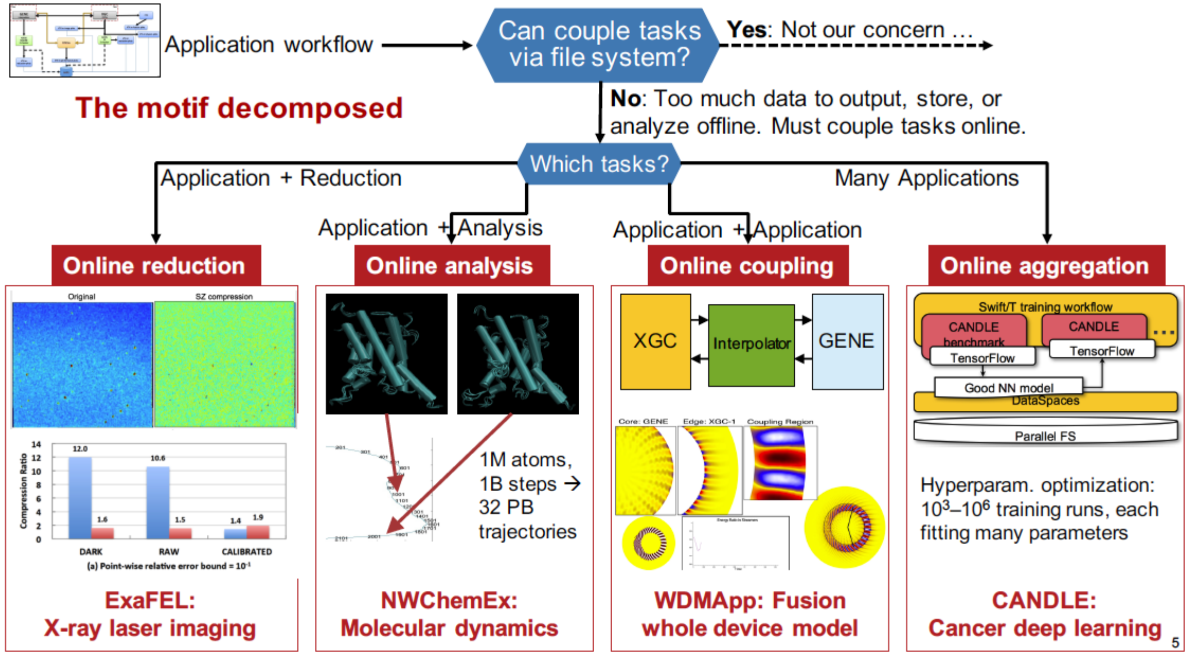
New approach:
Online data analysis and reduction

Co-optimize simulation, analysis, reduction for performance and information output

Substitute CPU cycles for I/O, via online data (de)compression and/or online data analysis



Provide the right information at the right time and place to accelerate discovery!



Goal: Replace the activities in HPC workflow that have been mediated through file I/O with in-situ methods / workflows. data reduction, analysis, code coupling, aggregation (e.g. parameter studies).

Cross-cutting tools:

- Workflow setup, manager (Cheetah, Savanna); Data coupler (ADIOS-SST); Compression methods (MGARD, FTK, SZ), compression checker (Z-checker)
- Performance tools (TAU, Chimbuco, SOSFlow)

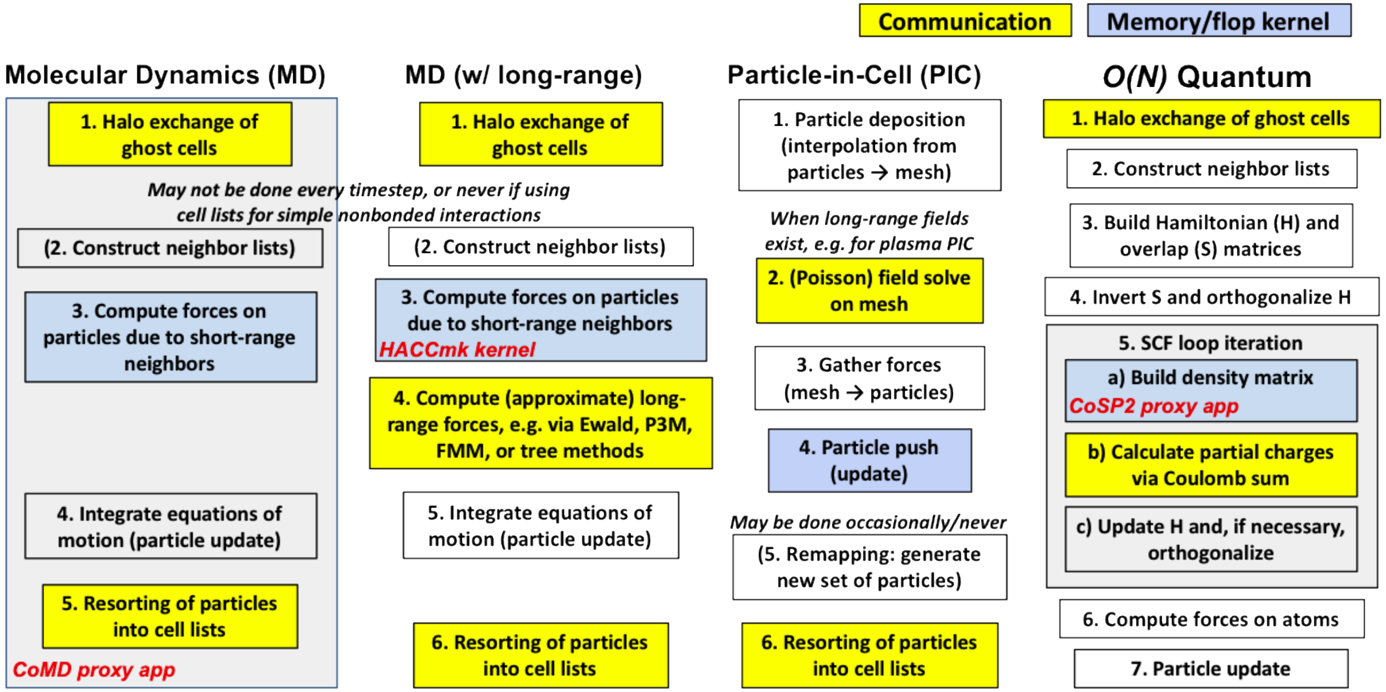
PI: Ian Foster (ANL)

ECP’s Co-Design Center for Particle Applications: CoPA

Goal: Develop algorithms and software for particle methods,

Cross-cutting capabilities:

- Specialized solvers for quantum molecular dynamics (Progress / BML).
- Performance-portable libraries for classical particle methods in MD, PDE (Cabana).
- FFT-based Poisson solvers for long-range forces.



Communication

Memory/flop kernel

Molecular Dynamics (MD)

1. Halo exchange of ghost cells

May not be done every timestep, or never if using cell lists for simple nonbonded interactions

(2. Construct neighbor lists)

3. Compute forces on particles due to short-range neighbors

4. Integrate equations of motion (particle update)

5. Resorting of particles into cell lists

CoMD proxy app

MD (w/ long-range)

1. Halo exchange of ghost cells

(2. Construct neighbor lists)

3. Compute forces on particles due to short-range neighbors *HACCmk kernel*

4. Compute (approximate) long-range forces, e.g. via Ewald, P3M, FMM, or tree methods

5. Integrate equations of motion (particle update)

6. Resorting of particles into cell lists

Particle-in-Cell (PIC)

1. Particle deposition (interpolation from particles → mesh)

When long-range fields exist, e.g. for plasma PIC

2. (Poisson) field solve on mesh

3. Gather forces (mesh → particles)

4. Particle push (update)

May be done occasionally/never

(5. Remapping: generate new set of particles)

6. Resorting of particles into cell lists

***O(N)* Quantum**

1. Halo exchange of ghost cells

2. Construct neighbor lists

3. Build Hamiltonian (H) and overlap (S) matrices

4. Invert S and orthogonalize H

5. SCF loop iteration

a) Build density matrix *CoSP2 proxy app*

b) Calculate partial charges via Coulomb sum

c) Update H and, if necessary, orthogonalize

6. Compute forces on atoms

7. Particle update

Technical approach:

- High-level C++ APIs, plus a Fortran interface (Cabana).
- Leverage existing / planned FFT software.
- Extensive use of miniapps / proxy apps as part of the development process.

PI: Sue Mniszewski (LANL) recently replacing Tim Germann (LANL), who is taking on a larger role in ECP

ECP’s Co-Design Center for Machine Learning: ExaLearn

Bringing together experts from 8 DOE Laboratories

- AI has the potential to accelerate scientific discovery or enable prediction in areas currently too complex for direct simulation (ML for HPC and HPC for ML)
- AI use cases of interest to ECP:
  - *Classification and regression*, including but not limited to image classification and analysis, e.g. scientific data output from DOE experimental facilities or from national security programs.
  - *Surrogate models* in high-fidelity and multiscale simulations, including uncertainty quantification and error estimation.
  - *Structure-to-function relationships*, including genome-to-phenome, the prediction of materials performance based on atomistic structures, or the prediction of performance margins based on manufacturing data.
  - *Control systems*, e.g., for wind plants, nuclear power plants, experimental steering and autonomous vehicles.
  - *Inverse problems* and optimization. This area would include, for example, inverse imaging and materials design.
- Areas in need of research
  - Data quality and statistics
  - Learning algorithms
  - Physics-Informed AI
  - Verification and Validation
  - Performance and scalability
  - Workflow and deployment

Expected Work Product: A Toolset That . . .

- Has a line-of-sight to exascale computing, e.g. through using exascale platforms directly, or providing essential components to an exascale workflow
- Does not replicate capabilities easily obtainable from existing, widely-available packages
- Builds in domain knowledge where possible “Physics”-based ML and AI
- Quantifies uncertainty in predictive capacity
- Is interpretable
- Is reproducible
- Tracks provenance

PI: Frank Alexander (BNL)

ECP Apps: Delivering on Challenge Problems

Requires Overcoming Computational Hurdles

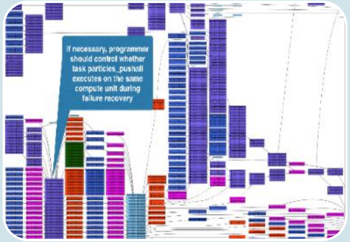
Domain	Challenge Problem	Computational Hurdles
Wind Energy	Optimize 50-100 turbine wind farms	Linear solvers; structured / unstructured overset meshes
Nuclear Energy	Virtualize small & micro reactors	Coupled CFD + Monte Carlo neutronics; MC on GPUs
Fossil Energy	Burn fossil fuels cleanly with CLRs	AMR + EB + DEM + multiphase incompressible CFD
Combustion	Reactivity controlled compression ignition	AMR + EB + CFD + LES/DNS + reactive chemistry
Accelerator Design	TeV-class 100-1000X cheaper & smaller	AMR on Maxwell’s equations + FFT linear solvers + PIC
Magnetic Fusion	Coupled gyrokinetics for ITER in H-mode	Coupled continuum delta-F + stochastic full-F gyrokinetics
Nuclear Physics: Lattice QCD	Use correct light quark masses for first principle light nuclei properties	Critical slowing down; strong scaling performance of MG-preconditioned Krylov solvers
Chemistry	Heterogeneous catalysis: MSN reactions	HF + DFT + coupled cluster (CC) + fragmentation methods
Chemistry	Catalytic conversion of biomass	Hybrid DFT + CC; CC energy gradients
Extreme Materials	Microstructure evolution in nuclear matls	AMD via replica dynamics; OTF quantum-based potentials
Additive Manufacturing	Born-qualified 3D printed metal alloys	Coupled micro + meso + continuum; linear solvers
Quantum Materials	Predict & control matls @ quantum level	Parallel on-node performance of Markov-chain Monte Carlo
Astrophysics	Supernovae explosions & neutron star mergers	AMR + nucleosynthesis + GR + neutrino transport

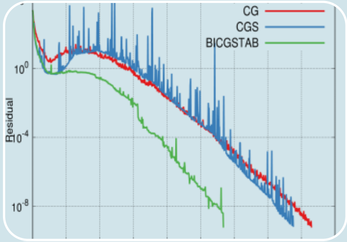
ECP Apps: Delivering on Challenge Problems

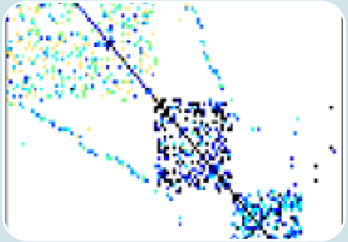
Requires Overcoming Computational Hurdles

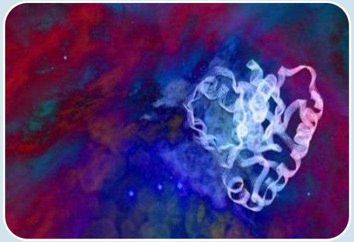
Domain	Challenge Problem	Computational Hurdles
Cosmology	Extract “dark sector” physics from upcoming cosmological surveys	AMR or particles (PIC & SPH); subgrid model accuracy; insitu data analytics
Earthquakes	Regional hazard and risk assessment	Seismic wave propagation coupled to structural mechanics
Geoscience	Geomechanical and geochemical evolution of a wellbore system at near-reservoir scale	Coupled AMR flow + transport + reactions to Lagrangian mechanics and fracture
Earth System	Assess regional impacts of climate change on the water cycle @ 5 SYPD	Viability of Multiscale Modeling Framework (MMF) approach for cloud-resolving model; GPU port of radiation and ocean
Power Grid	Efficient planning; underfrequency response	Parallel performance of nonlinear optimization based on discrete algebraic equations and MIP
Cancer Research	Predictive preclinical models and accelerate diagnostic and targeted therapy	Increasing accelerator utilization for model search; exploiting reduced/mixed precision; preparing for any data management or communication bottlenecks
Metagenomics	Discover, understand (find genes) and control species in microbial communities	Efficient and performant implementation of UPC, UPC++, GASNet; graph algorithms; SpGEMM performance
FEL Light Source	Light source-enabled analysis of protein and molecular structure and design	Strong scaling (one event processed over many cores) of compute-intensive algorithms (ray tracing, M-TIP) on accelerators

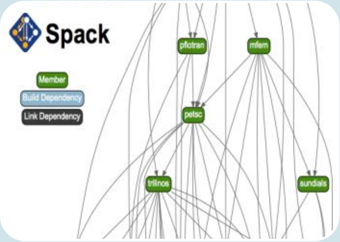
ECP software technologies are a fundamental underpinning in delivering on DOE’s exascale mission











Programming Models & Runtimes

- Enhance & prepare OpenMP and MPI programming models (hybrid programming models, deep memory copies) for exascale
- Development of performance portability tools (e.g., Kokkos and Raja)
- Support alternate models for potential benefits and risk mitigation: PGAS (UPC++/GASNet) ,task-based models (Legion, PaRSEC)
- Libraries for deep memory hierarchy & power management

Development Tools

- Continued, multifaceted capabilities in portable, open-source LLVM compiler ecosystem to support expected ECP architectures, including support for F18
- Performance analysis tools that accommodate new architectures, programming models, e.g., PAPI, Tau

Math Libraries

- Linear algebra, iterative linear solvers, direct linear solvers, integrators and nonlinear solvers, optimization, FFTs, etc.
- Performance on new node architectures; extreme strong scalability
- Advanced algorithms for multi-physics, multiscale simulation and outer-loop analysis
- Increasing quality, interoperability, complementarity of math libraries

Data and Visualization

- I/O libraries: HDF5, ADIOS, PnetCDF,
- I/O via the HDF5 API
- Insightful, memory-efficient in-situ visualization and analysis – Data reduction via scientific data compression
- Checkpoint restart
- Filesystem support for emerging solid state technologies

Software Ecosystem

- Develop features in Spack necessary to support all ST products in E4S, and the AD projects that adopt it
- Development of Spack stacks for reproducible turnkey deployment of large collections of software
- Optimization and interoperability of containers on HPC systems
- Regular E4S releases of the ST software stack and SDKs with regular integration of new ST products

NNSA ST

- Projects that have both mission role and open science role
- Major technical areas: New programming abstractions, math libraries, data and viz libraries
- Cover most ST technology areas
- Open source NNSA Software projects
- Subject to the same planning, reporting and review processes

ST Ecosystem: From products to SDKs to an integrated stack

ECP ST Individual Products

Levels of Integration	Product	Source and Delivery
• Standard workflow • Existed before ECP	ST Products	• **Source:** ECP L4 teams; Non-ECP Developers; Standards Groups • **Delivery:** Apps directly; spack; vendor stack; facility stack
• Group similar products • Make interoperable • Assure policy compliant • Include external products	SDKs	• **Source:** ECP SDK teams; Non-ECP Products (policy compliant, spackified) • **Delivery:** Apps directly; spack install sdk; future: vendor/facility
• Build all SDKs • Build complete stack • Containerize binaries	E4S	•**Source:** ECP E4S team; Non-ECP Products (all dependencies) •**Delivery:** spack install e4s; containers; CI Testing

ECP ST Open Product Integration Architecture

ECP is formalizing software stack deployment mechanisms to deliver high quality software to DOE Facilities

ECP must provide robust, reliable and supported ECP software products for DOE Facilities to adopt them

To ensure this ECP invests in:

- Software Development Toolkits (SDKs):
  - improves interoperability among tools with similar functionality
  - improves ecosystem robustness through development of community policies for how software interacts
  - Improves quality of software through adoption of standard best practices
- Extreme-scale scientific software stack (E4S):
  - A Spack-based distribution of ECP ST products
  - Related and dependent software extensively tested for interoperability and portability to multiple architectures
  - Version 1.0 release in November 2019: 50 full-release and 6 partial-release products

http://e4s.io

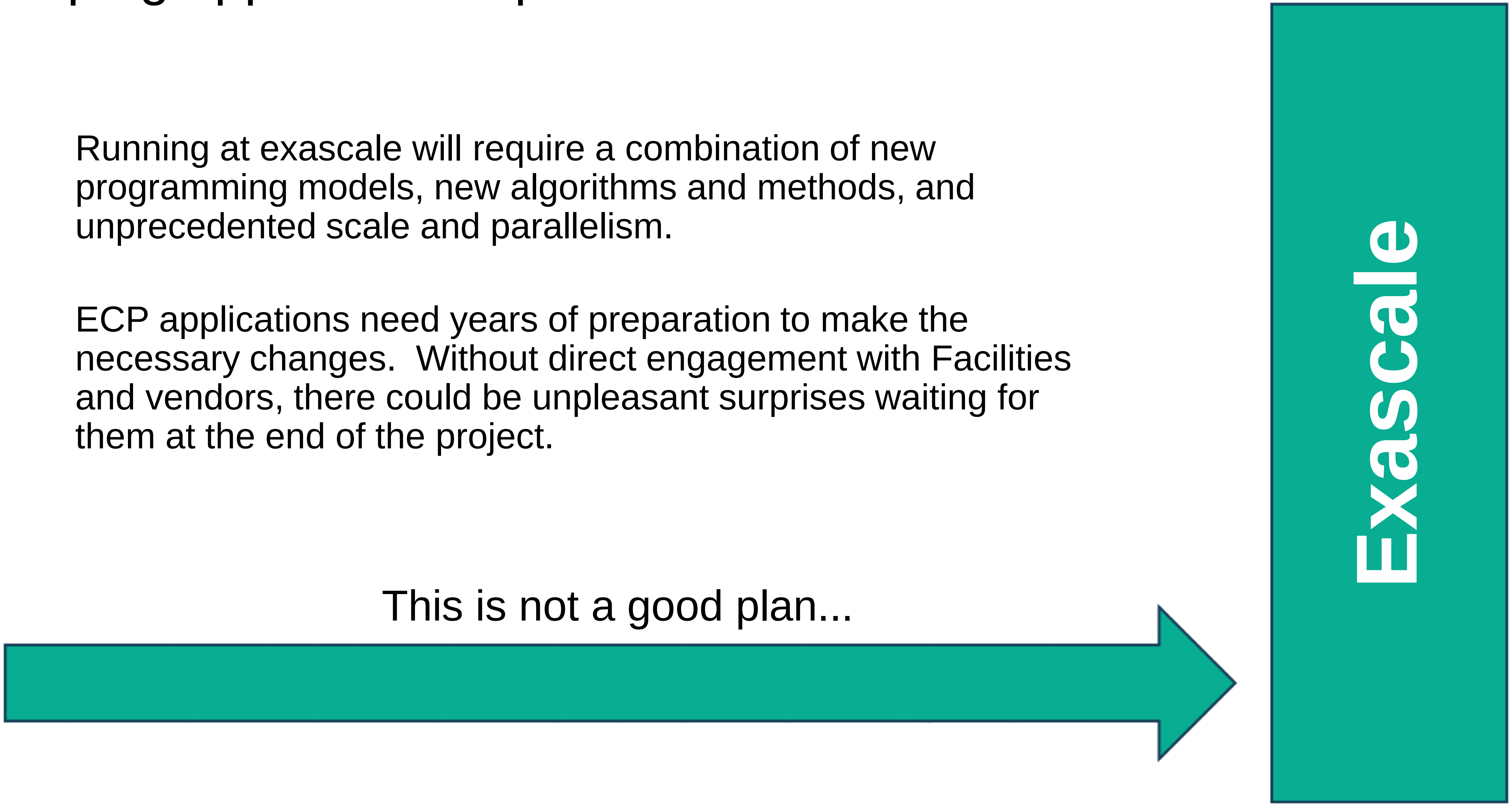
Pre-exascale GPU machines have been critical in preparing for exascale

- ECP teams have had early access to large NVIDIA GPU supercomputers.
- For many applications, refactoring the code to run well on a heterogeneous machine has required fundamental changes to data structures, data motion and algorithms that could be made independently of specific accelerator features.
- Code teams are refactoring with the knowledge that CUDA will not be available on Aurora or Frontier (or many other machines they may want to run on). Many are adopting portable programming models (e.g. Kokkos) or isolating machine-specific kernels.

Ramping applications up to exascale

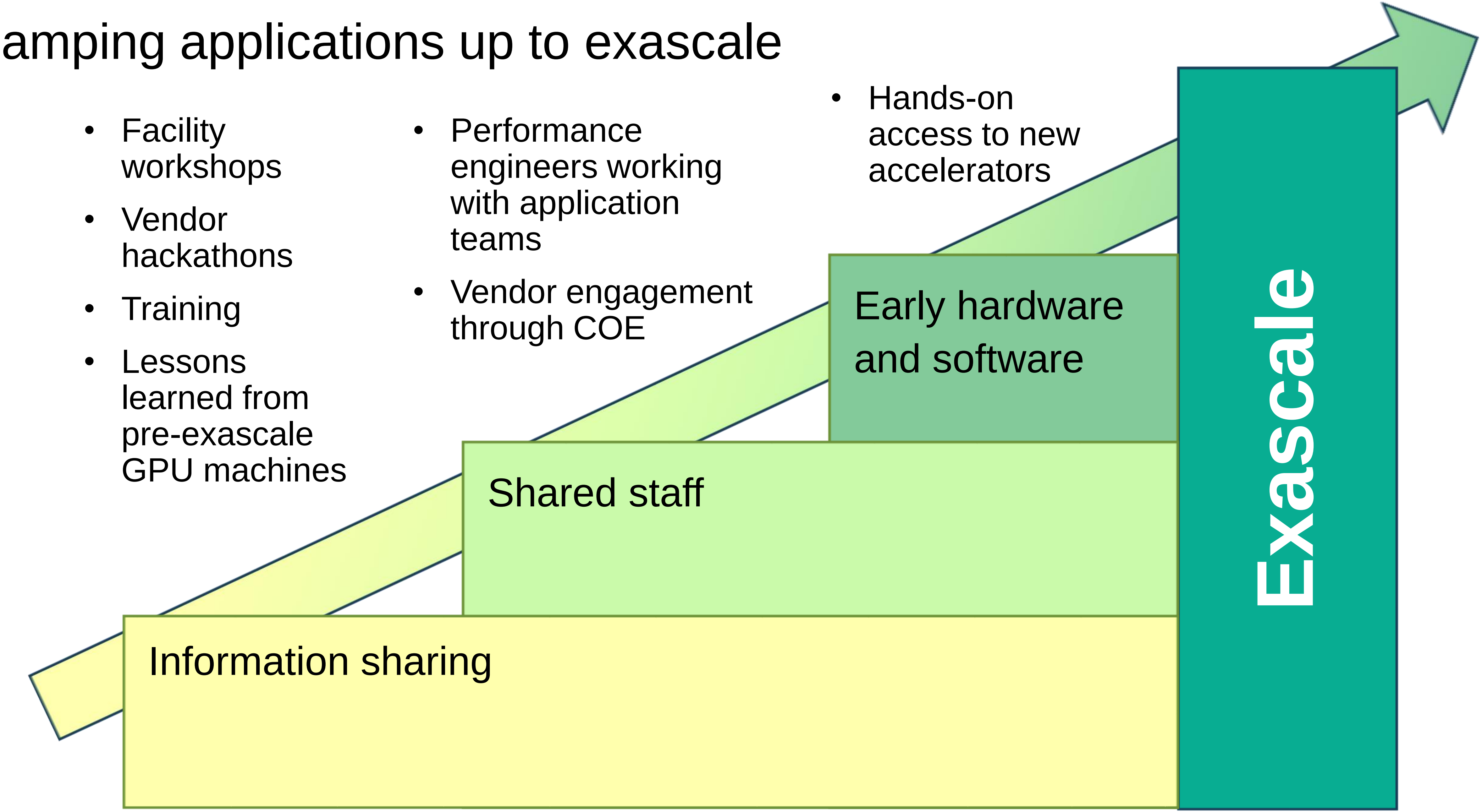
Running at exascale will require a combination of new programming models, new algorithms and methods, and unprecedented scale and parallelism.

ECP applications need years of preparation to make the necessary changes. Without direct engagement with Facilities and vendors, there could be unpleasant surprises waiting for them at the end of the project.



This is not a good plan...

Ramping applications up to exascale



- Facility workshops
- Vendor hackathons
- Training
- Lessons learned from pre-exascale GPU machines

- Performance engineers working with application teams
- Vendor engagement through COE

- Hands-on access to new accelerators

Information sharing

Shared staff

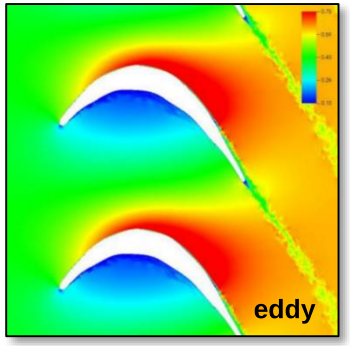
Early hardware and software

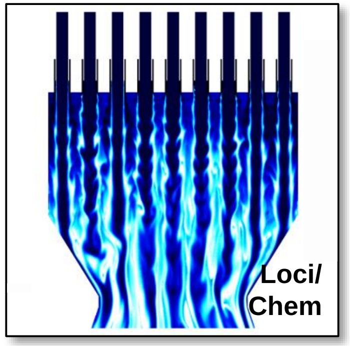
Exascale

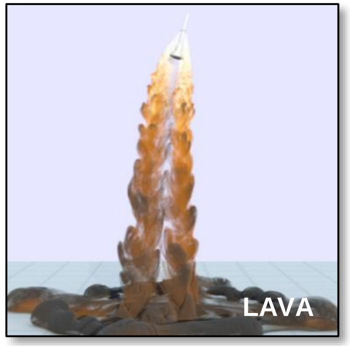
High Performance Computing's Impact on Aerospace Prediction

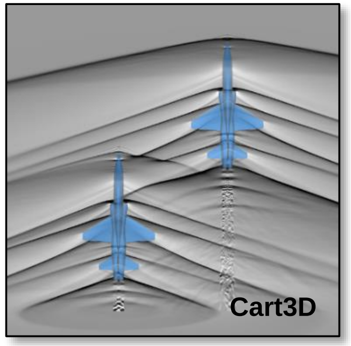
Eric Nielsen
Computational AeroSciences Branch
NASA Langley Research Center

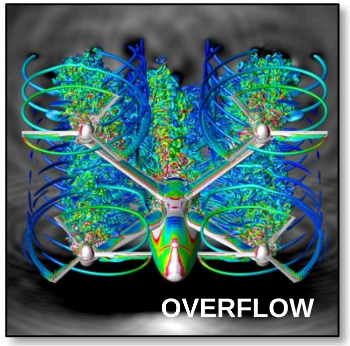
AIAA SciTech Forum 360
January 9, 2019



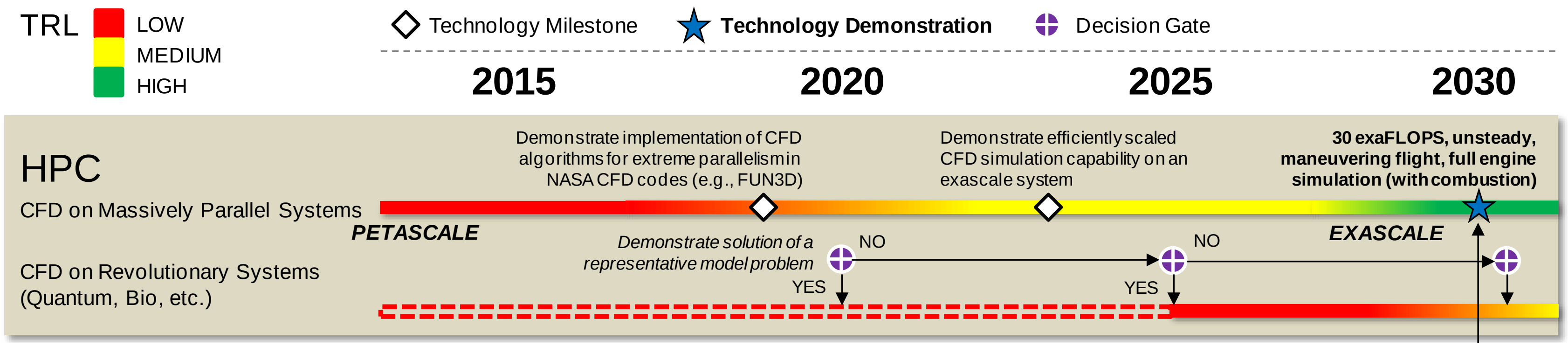








HPC in the CFD Vision 2030 Roadmap



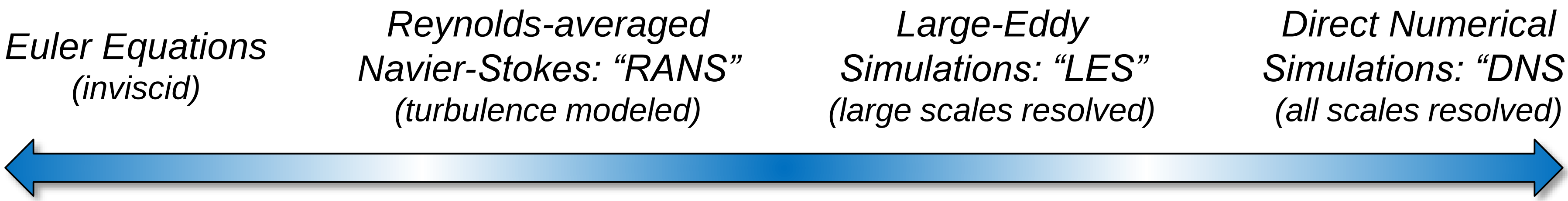
TRL
🟥 LOW
🟨 MEDIUM
🟩 HIGH

◇ Technology Milestone
★ **Technology Demonstration**
⊕ Decision Gate

1. *"Current simulation software must be ported to evolving and emerging HPC architectures with a view toward efficiency and software maintainability."*
2. *"Investments must be made in the development of new algorithms, discretizations, and solvers that are well suited for the massive levels of parallelism and deep memory architectures anticipated in future HPC architectures."*
3. *"Increased access to the latest large-scale computer hardware must be provided and maintained, not only for production runs, but also for algorithmic research and software development projects…"*

"The aerospace CFD community is notoriously insular, publishing in AIAA or similar venues, with scant presence in computational science meetings hosted by SIAM, IEEE, and ACM…at a minimum, NASA should establish a presence at these meetings to keep abreast of developments in these areas…"

Capacity vs Capability Computing



Euler Equations
(inviscid)

Reynolds-averaged
Navier-Stokes: “RANS”
(turbulence modeled)

Large-Eddy
Simulations: “LES”
(large scales resolved)

Direct Numerical
Simulations: “DNS”
(all scales resolved)

Increasing physics, increasing cost →

- Today’s analysis and design approaches typically rely on Euler and RANS simulations, each requiring O(10^2)-O(10^4) CPU hours on moderate HPC resources
- CFD Vision 2030 calls for prediction of unsteady separated flows
- Current projection is full aircraft WMLES as a grand challenge problem in the 2060 timeframe using an entire leadership-class machine, with DNS following at infinity*
- For now, hybrid RANS-LES and canonical wall-modeled LES simulations are state of the art and may require O(10^8) CPU hours on large HPC resources

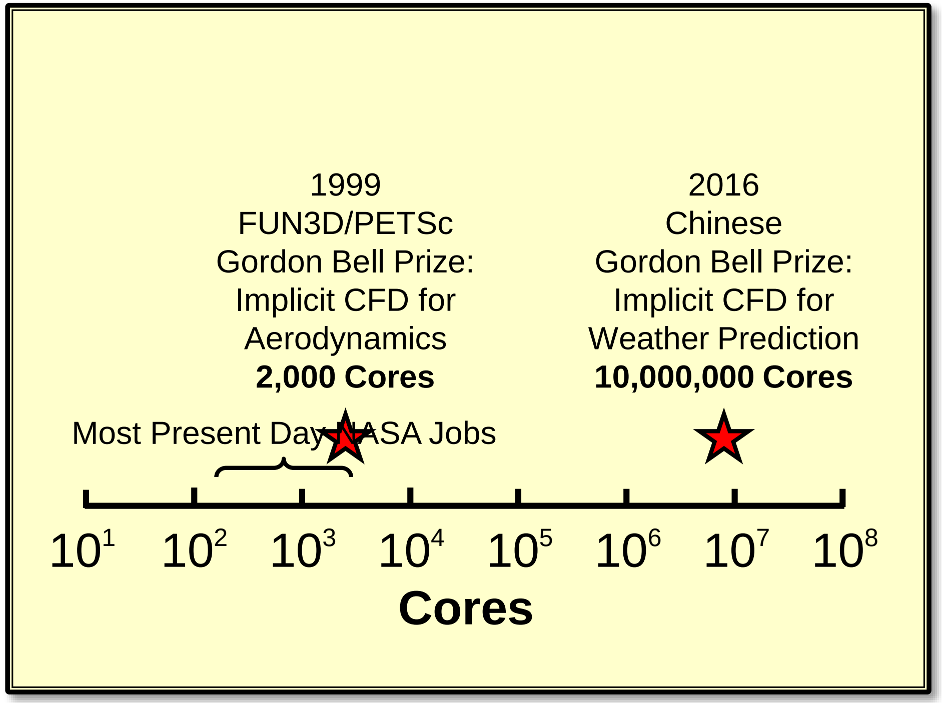
Space Launch System Database Generation

Database	**Code**	**Solutions (Grid Size)**	**Wall Time**
Ascent	FUN3D	1,380 (60M)	2-4 weeks
Ascent	OVERFLOW	1,000 (500M)	2-3 months
F & M Wind Tunnel	FUN3D	600 (40M)	1 week
Booster Separation	FUN3D	13,780	3 months
Booster Separation	Cart3D	25,000	3 months

* Spalart, P., Private communication.

Derek Dalle, NASA ARC

An Historical Anecdote on Capacity vs Capability



1999
FUN3D/PETSc
Gordon Bell Prize:
Implicit CFD for
Aerodynamics
2,000 Cores

2016
Chinese
Gordon Bell Prize:
Implicit CFD for
Weather Prediction
10,000,000 Cores

Most Present Day NASA Jobs

Software and Algorithmic Challenges

Software

- Strategic, long-term software development programs
- Project-driven environment leaves little time for skills development and exploring new paradigms
  - Heterogeneity is expected to grow dramatically over the next decade
  - Codes may run *slower* on new architectures if we stand pat
  - “I want the performance but I don’t want to change my code”
- Proliferation of programming models
- Compilers can leave a bit to be desired

Algorithms

- Exposing vastly more concurrency
- Asynchronous, communication hiding, dynamic task-based schedulers
- Strong scaling, particularly for simulations with long time durations (HRLES, LES, …)
- V&V in the face of asynchronous execution
- Mixed-precision approaches leveraging specialized hardware driven by AI community

Hardware and Other Challenges

Hardware

- Job scheduling for capability development and applications
- Chicken and the egg, and the diversity of hardware in the data center
- Leveraging DOE hardware is very helpful, but not a silver bullet
  - Extremely high bar to qualify: equivalent of ~1 million cores now; ~10 million cores soon
  - Not enough compute available
  - Dependence on highly competitive, proposal-based systems not conducive to planning; high overhead for researcher
  - Sensitive data restrictions

Other

- Complex engineering workflows with items that have not been ported / scaled: Amdahl is a killer!
- Are we leveraging the potential of ML / AI / Big Data?
- Valuing computational science expertise on equal footing with traditional core competencies
  - Then attracting and hiring such workforce

Some Encouraging Recent Activities

Use of FUN3D on Summit for Mars Entry

- FUN3D shows 35x node-level speedup (6 NVIDIA Tesla V100s vs IBM POWER9)
- Scaled to 1,024 Summit nodes – performance equivalent of 1.1M Xeon Skylake cores
- Up to 200 TB of flowfield data stored per run
- Ensembles requiring years on capacity-managed Xeon systems are done in a workweek

NASA Langley HPC Incubator

- 3-year activity aimed at workforce development, HPC infusion into conventional projects
- Over 1700 participants in training courses; numerous guest speakers and collaborators
- Travel to HPC / computational science forums seldom supported by projects

FY20 New Start: "New CFD Algorithms Tailored to Emerging Computer Hardware Technology"

- 5-year project spanning 4 centers, funded by NASA Office of Chief Engineer
- Aimed at exploring workhorse aerosciences codes on new HPC architectures
- Workforce development through training, hackathons, boot camps, travel support
- Strengthening ties with OGAs, vendors, academia
- Procurement of testbed hardware
- Close coordination with NASA Advanced Supercomputing Division

DoD High Performance Computing Trends and Requirements for Meeting Milestones of the CFD2030 Vision

DOD HPC MODERNIZATION PROGRAM

Scott A. Morton, **CREATE-AV Project Manager**

January 9, 2020

Distribution A: Approved for Public Release (PA# 19-68)

DoD HPCMP at a Glance

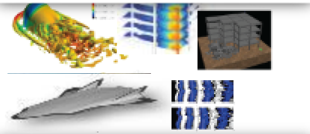
A national asset providing high performance computing capabilities and expertise to solve our most critical DoD mission challenges

DoD Services & Agencies

- Army
- Navy
- Air Force
- DTRA
- MDA
- DARPA

Major Functions

Science and Technology



Test and Evaluation

Acquisition Engineering

Hardware + Network + Software

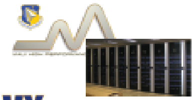
DoD Supercomputing Resource Centers (DSRCs)

U.S. Air Force Research Laboratory DSRC

U.S. Army Research Laboratory DSRC

U.S. Army Engineer Research and Development Center DSRC

Maui High Performance Computing Center DSRC



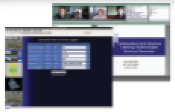
U.S. Navy DSRC

Networking and Security

Defense Research & Engineering Network (DREN)

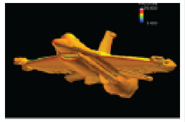
Computer Network Defense, Security R&D, and Security Integration

Software Applications



Core Software

Computational Environments

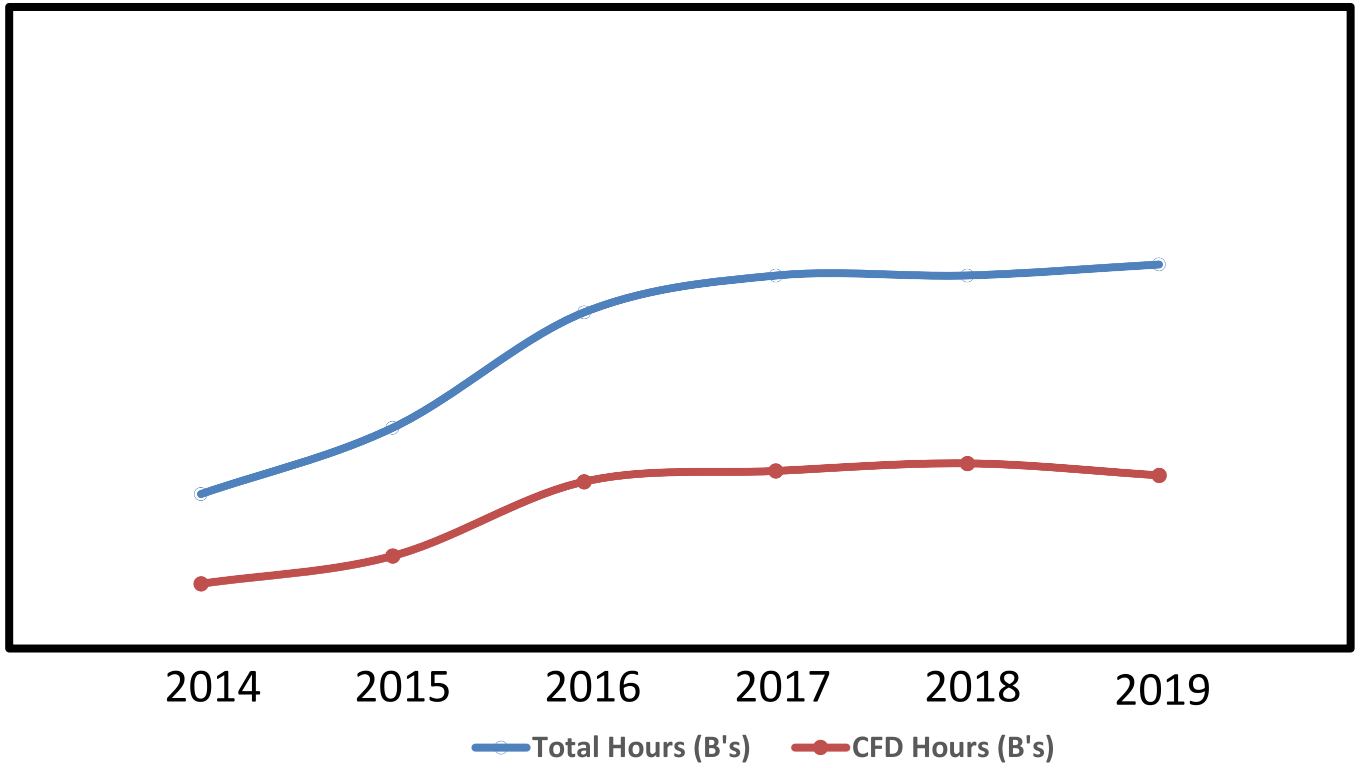


Education and Training

HPC User Support

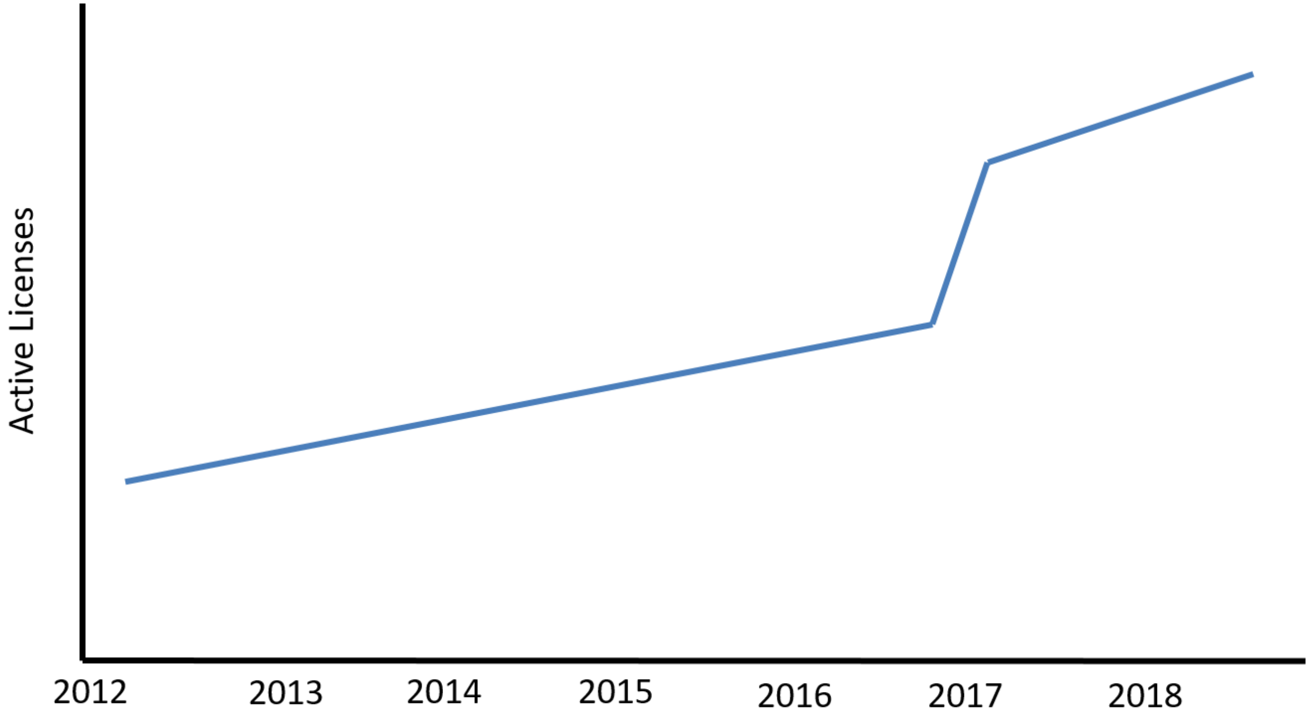
Partners DoD/Federal Agencies + PEO's/PM's + Academia + Industry

Hardware Utilization



- Hardware Use Trends for CFD and Total
  - Machine utilization typically constant above 80%
  - Typical use is for S&T and T&E
  - Future growth expected for Acquisition programs' Digital Engineering (DE) support

Software Utilization for DoD Acquisition (Multidisciplinary CFD)



- Software Use Trends for Air Vehicles
  - Jump in uptake corresponds roughly to OEMs use
  - Sustained 14% growth prior and 17% growth after
  - Future software requirements are in direct support of Acquisition Program’s DE paradigm shift

The Paradigm to Change

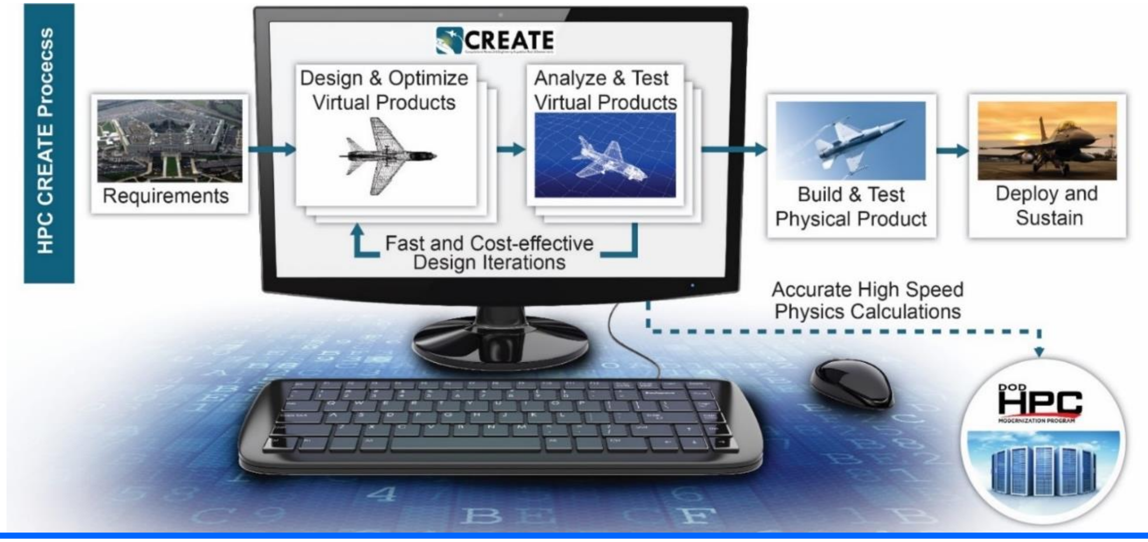
From

Reliance on physical test as the driver for design iteration and primary source for “actionable engineering data”, e.g. support warrant holder requirements, system certifications, etc.

To

Use physics-informed analysis and **virtual test to drive design iterations**, and as a source of actionable engineering data.

Use virtual test to drive design iterations.



Use physical test to validate the design.

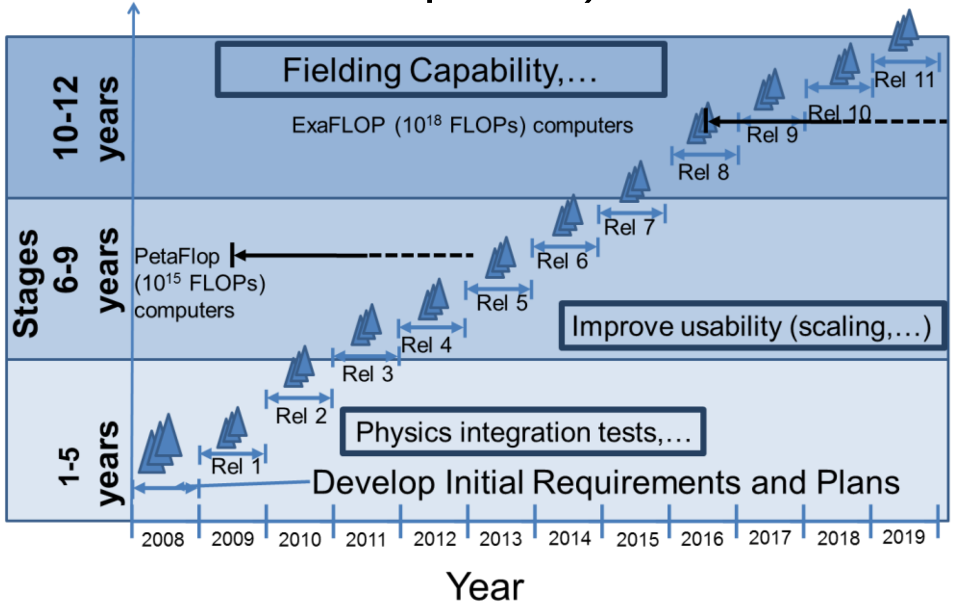
CREATE Software Engineering

Production quality software, designed for a service life measured in decades.

A technology bridge – transitioning maturing research into production technology.

Annual Release Cadence (with intermediate updates)

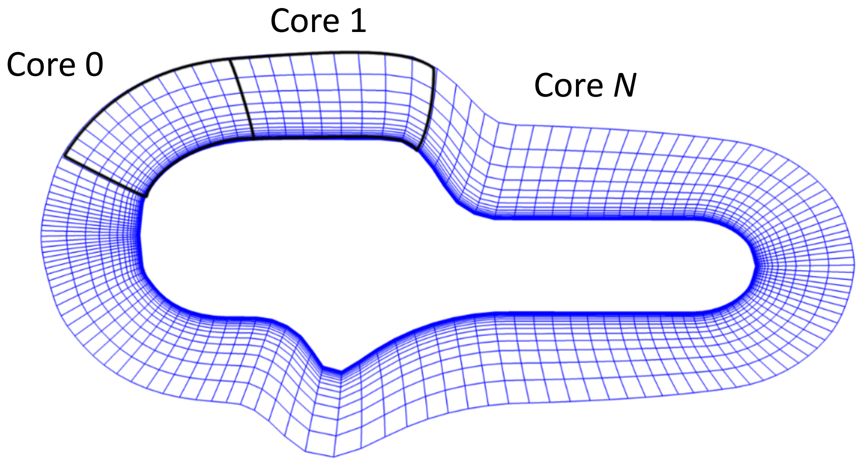
- Agile software development practices
- Rigorous unit, system, and integration testing
- Automated verification and validation testing
- Constantly growing multi-disciplinary capabilities
- Difficult to follow fast changes in machine architectures



12 Physics-based Simulation Software Products

High Performance Computing

- **Supercomputers with many CPU cores are used to accelerate computation**
  - Meshes are divided into partitions
  - Compute clusters offer millions of CPU cores.



- **The fastest, modern compute cluster is Summit**
  - Summit (#1) has fewer CPU cores than Sunway TaihuLight (#3)

#	Name and Location		CPU Cores	GPUs
1	Summit	USA	2,282,544	Yes
2	Sierra	USA	1,572,480	Yes
3	Sunway TaihuLight	China	10,646,600	No
4	Tianhe-2A	China	4,981,760	No
	…			

Top supercomputers in the world by performance

https://www.top500.org/lists/2018/11/

Next-generation CFD codes should also use new compute architectures to achieve better performance.

CFD on GPUs

- **Primary challenge of GPU computing:**
  - All code should be re-written with NVIDIA CUDA programming language.
  - Memory bandwidth from CPU to GPU is low; sharing computation is often inefficient.

- **Many other research groups have successfully used GPUs for Computational Fluid Dynamics:**
  - Crabill, Witherden, Jameson: “A Parallel Direct Cut Algorithm for High-Order Overset Methods with Application to a Spinning Golf Ball.”
  - K. Soni, D. D. Chandar, and J. Sitaraman: “Development of an overset grid computational fluid dynamics solver on graphical processing units”
  - B. P. Pickering, C. W. Jackson, T. R. Scogland, W.-C. Feng, and C. J. Roy: “Directive-based gpu programming for computational fluid dynamics”
  - E. Nielsen and A. Walden, “Preparing the FUN3D CFD solver for the exascale era.”

6 × speedup compared to a CPU

- **GPUs have yet to fit into a production-level framework:**
  - A framework using multiple solvers, with some using GPUs has not been demonstrated.

CFD2030 Vision Hardware Needs

- CFD2030 Vision is “in sync” with DoD plans for use in supporting Acquisition Program DE
  - Growing use of computational methods to support acquisition decisions
  - Paradigm shift becoming a reality in the DoD
- DoD Acquisition Support software represents millions of lines of codes primarily suited to cpus
- Hardware necessary to support DoD plans requires growth of support for the HPC infrastructure ($$)
  - Need more computational resources to support acquisition
  - Moore’s Law drop-off means additional funding required to accomplish needed HPC growth

HPC Considerations for Industrial (OEM) CFD

- Primarily **capacity-based** (not capability-based)
    - Large number (10s-100s) of cases using full configuration geometry is now routine for steady-state CFD analysis in all phases (concept development, detailed design, certification, and product support)
    - Larger numbers (1000s) for aerodynamic optimization
- Increasing use of **unsteady, time-accurate CFD** including **turbulence-resolving methods** (hybrid RANS/LES, WMLES) and exploring new technologies (e.g. UQ)
- Critical need to **increase CFD computational efficiency**
    - Simulations to help build aero databases, and expand further into flight envelope – **balancing accuracy with throughput**
    - Recognizing the move to **next-generation HPC architectures** (e.g. GPUs) and the need to **re-factor codes (software)** to exploit hardware (e.g DoE Exascale Project)
- Assessment of evolving CFD technology on **leadership-class systems** would be useful in helping industry size future HPC systems to maximize ROI