Au service de la recherche scientifique, l'innovation et la compétitivité des entreprises

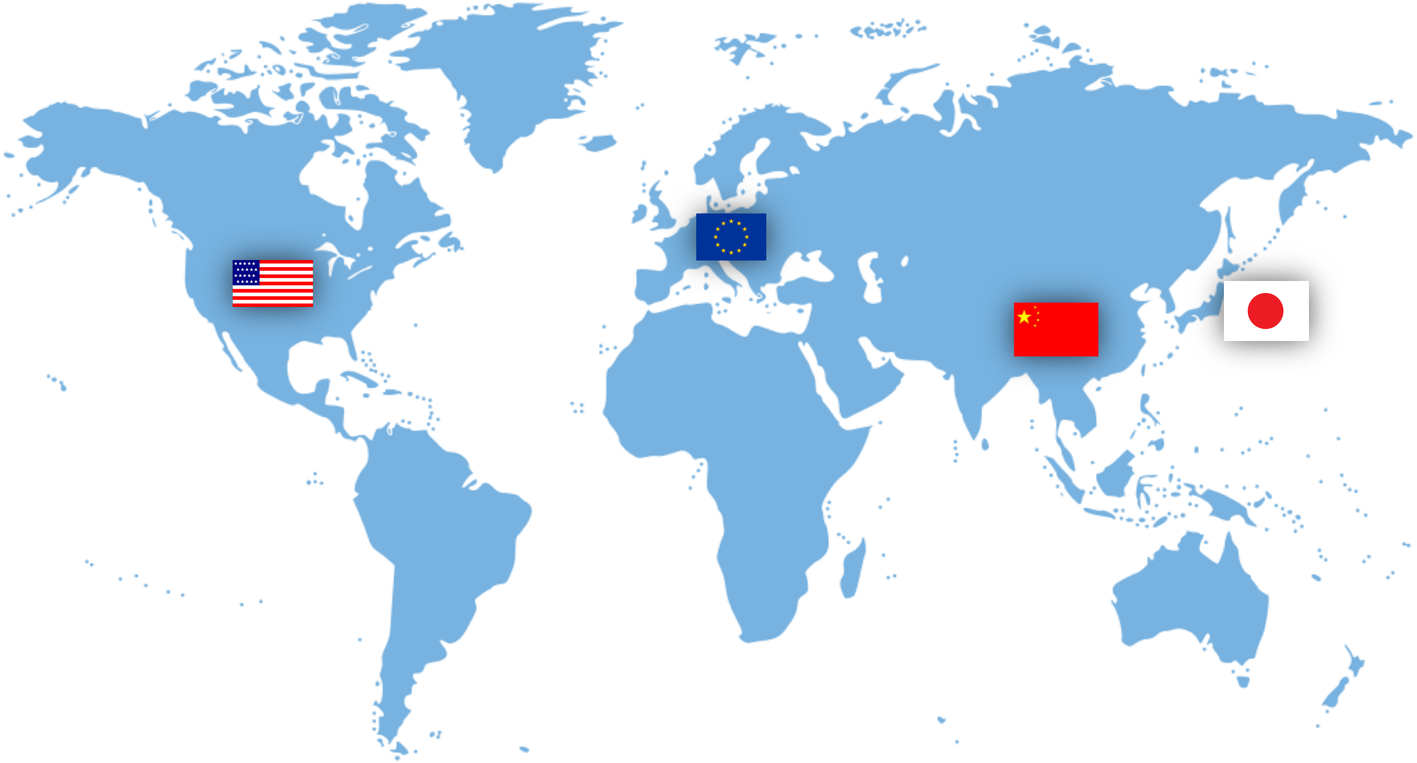
À la croisée des révolutions numériques

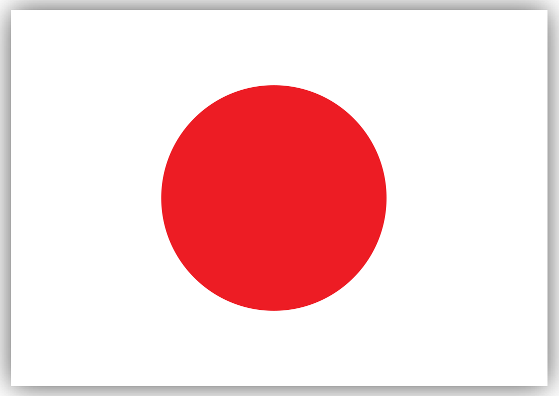
Vision des architectures pour l'Exascale, panorama international

S. Réquéna – GENCI
C. Calvin – ARISTOTE
F. Boillod-Cerneux - CEA

Séminaire Aristote « En route vers l'Exascale » | 23 mai 2019

AGENDA



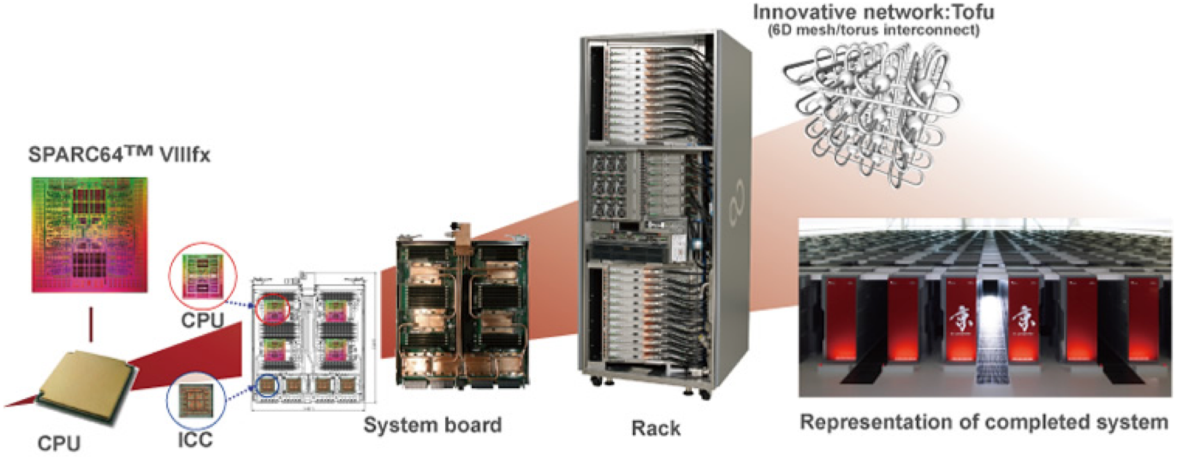


L'EXASCALE AU JAPON

Post-K au RIKEN

❑Successeur du K-Computer (2011-2019) 京, *kei*, qui signifie 10^{16}

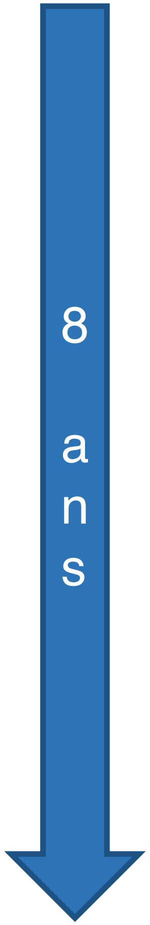
- Système Fujistsu de 11.3 PFlop/s crête installé au Riken (Kobe) en 2011
- Architecture MPP
  - 864 racks avec chacun 24 cartes de 4 nœuds mono socket Sparc VIIIfx 8-core 2.0 GHz (45nm, 128 GF), 16 GB de mémoire
  - 88 128 processeurs **soit 705 024 cœurs et 1.26 PB de mémoire dist.** au total
  - Interconnect propriétaire **Tofu 6D torique** avec liens 5GB/s bi directionnels
  - Refroidissement liquide - consommation totale = **12.6MW**
  - Stack logicielle optimisée : micro kernel, OpenMPI, MPI, auto // Fortran 90, Lustre, …
  - Noeud de calcul très équilibré : **0.5 bytes/flops** (bw mémoire, réseau, …)
- #1 du top500 de juin 2011 à juin 2012 (now #18), #1 graph500 (BFS) depuis 2014 et #1 ou #2 du HPCG depuis 2016



L’EXASCALE AU JAPON

Post-K au RIKEN

❑Comme pour tout Tier0 japonais → Méthodologie de co design


8 ans

- 2012-2013 : études de faisabilité : 3 équipes architecture en compétition (NEC, Hitachi et **Fujitsu**) et 1 équipe applications
  - 9 applications retenues : médical/pharma, environnement/prévention risques, énergie, manufacturing, …
  - Utilisation de simulateurs/estimateurs performances sur kernels, mini et full apps
  - Impact vectorisation, #core, #NUMA node, interconnect, HBM, …
- Avril 2014 : démarrage projet Post K avec budget proche de 1Md$ et **objectif 100x soutenus / K computer sur applications concrètes**
- Aout 2017 : Hotchips conférence -> annonce support processeur ARM
- Aout 2018 : annonces processeur ARM64fx et réseau Tofu2 par Fujitsu
- 1H 2019 : démarrage fabrication Post-K
- Aout 2019 : fin des activités du K-Computer
- 4Q2019-1Q2020 : installation Post-K
- 1H2020 : pré production Post-K
- 2020-2021 : pleine production Post-K

L'EXASCALE AU JAPON

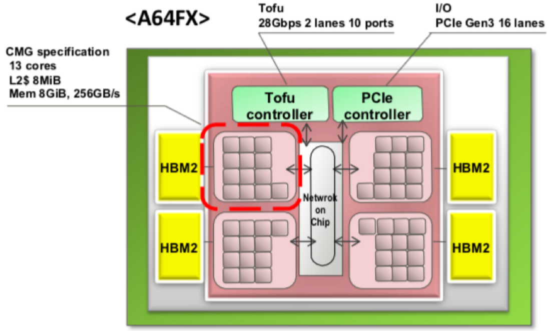
Post-K au RIKEN

❑ Post K une architecture MPP manycore **convergée** (HPC+IA)

- Nœud mono ARM v8-A A64FX (**48 compute+4 OS core**), SVE 512, 7nm
- 4 core memory group (4 espaces NUMA) de 12 + 1 cœurs
- **>2.7 TF peak** (FP64), >5.4 TF (FP32), >10.8 (INT16 DP) et > 21.6 (INT8 DP)
- **32 GB HBM3 (**1 Go/s débit, >80%@Stream TRIAD) → 0.66 GB/core
- Bon équilibre nœud avec **0.4 Bytes/Flop** et >90% @DGEMM

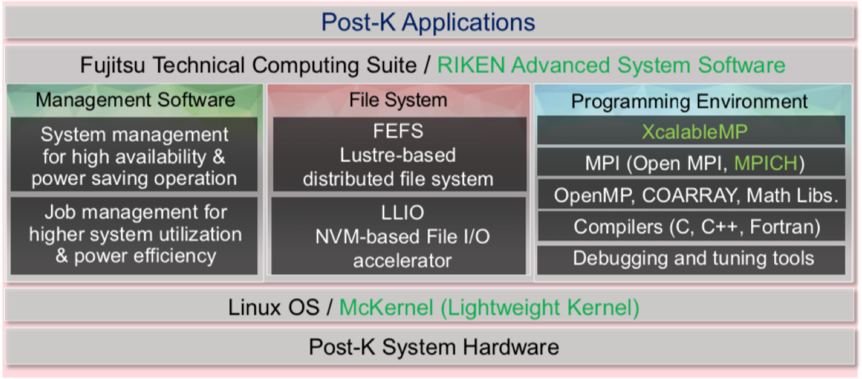
❑Fédérée par un interconnect TOFU-D next gen

- Topologie 6D Torique (comme K computer)
- *0,49 us de latence et 38.1 Go/s débit soutenu per nœud*



❑Specs machine complète

- *>150 000 nodes, 8 millions de cœurs de calcul -> environ 0.5 EF peak*
- 3 niveaux stockage (burst buffers, Lustre, Cloud)
- 400 racks, refroidissement DLC
- 40MW consommation totale
- Stack logicielle optimisée



Post-K Applications

Fujitsu Technical Computing Suite / RIKEN Advanced System Software

Management Software

System management for high availability & power saving operation

Job management for higher system utilization & power efficiency

File System

FEFS Lustre-based distributed file system

LLIO NVM-based File I/O accelerator

Programming Environment

XcalableMP

MPI (Open MPI, MPICH)

OpenMP, COARRAY, Math Libs.

Compilers (C, C++, Fortran)

Debugging and tuning tools

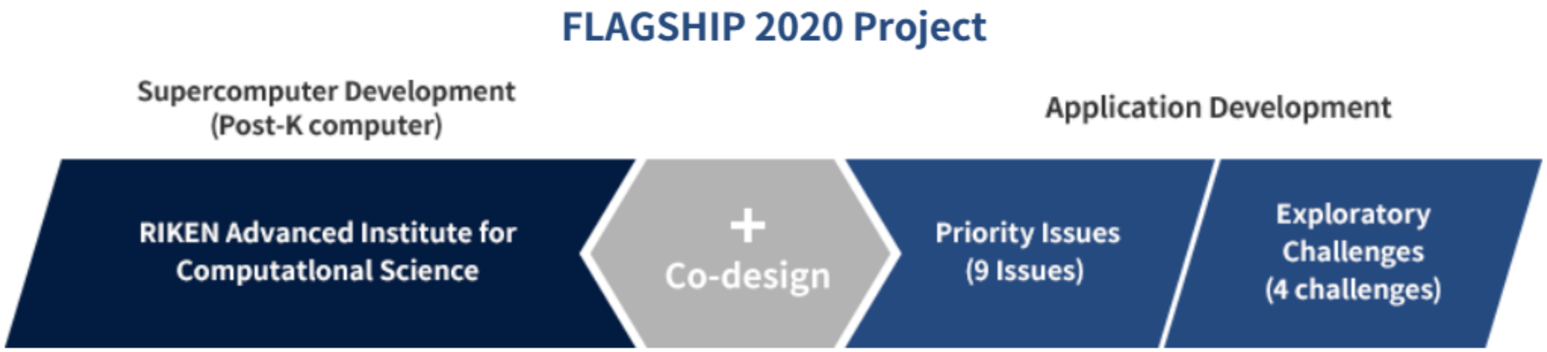
Linux OS / McKernel (Lightweight Kernel)

Post-K System Hardware

Post K sera un produit commercial de Fujistu

L'EXASCALE AU JAPON

Post-K au RIKEN



FLAGSHIP 2020 Project

Supercomputer Development (Post-K computer)

Application Development

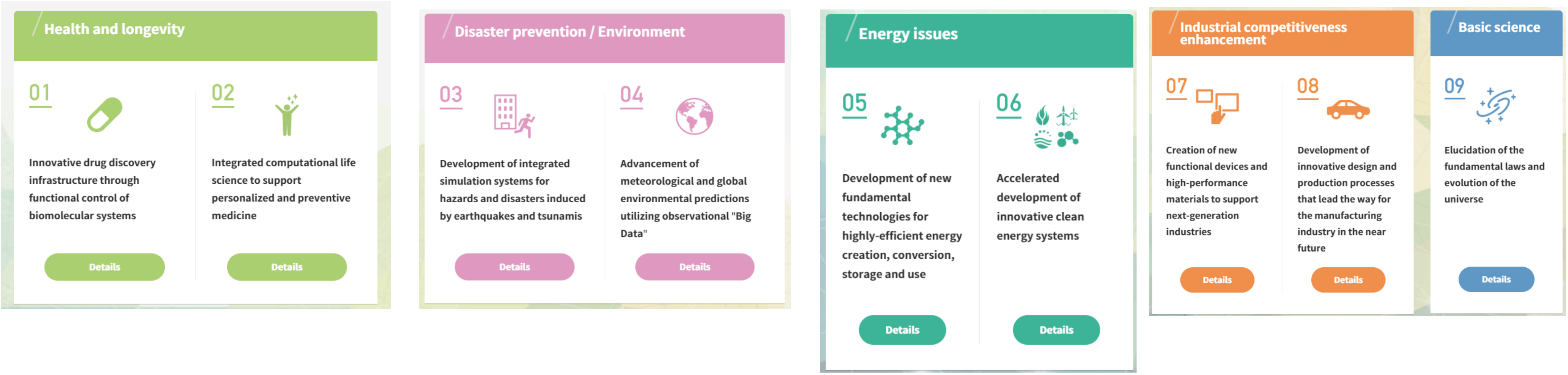
RIKEN Advanced Institute for Computatlonal Science

+ Co-design

Priority Issues (9 Issues)

Exploratory Challenges (4 challenges)

The project is backed by Japan's Ministry of Education, Culture, Sports, Science and Technology.



Health and longevity

01 Innovative drug discovery infrastructure through functional control of biomolecular systems

Details

02 Integrated computational life science to support personalized and preventive medicine

Details

Disaster prevention / Environment

03 Development of integrated simulation systems for hazards and disasters induced by earthquakes and tsunamis

Details

04 Advancement of meteorological and global environmental predictions utilizing observational "Big Data"

Details

Energy issues

05 Development of new fundamental technologies for highly-efficient energy creation, conversion, storage and use

Details

06 Accelerated development of innovative clean energy systems

Details

Industrial competitiveness enhancement

07 Creation of new functional devices and high-performance materials to support next-generation industries

Details

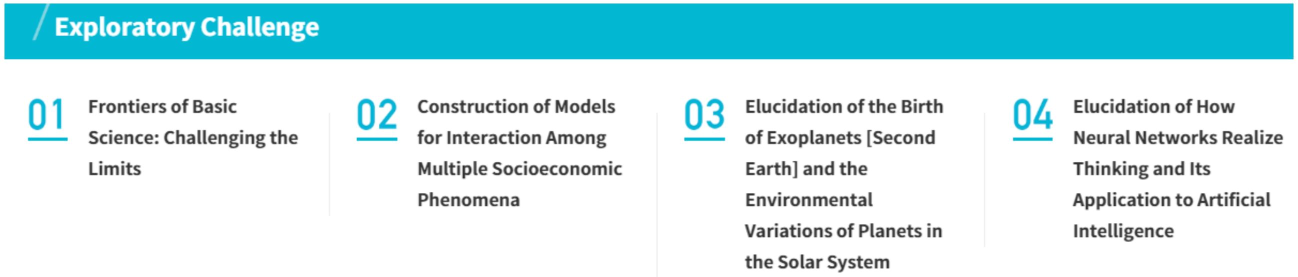
08 Development of innovative design and production processes that lead the way for the manufacturing industry in the near future

Details

Basic science

09 Elucidation of the fundamental laws and evolution of the universe

Details



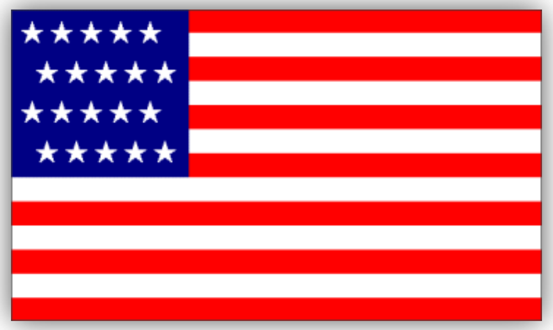
Exploratory Challenge

01 Frontiers of Basic Science: Challenging the Limits

02 Construction of Models for Interaction Among Multiple Socioeconomic Phenomena

03 Elucidation of the Birth of Exoplanets [Second Earth] and the Environmental Variations of Planets in the Solar System

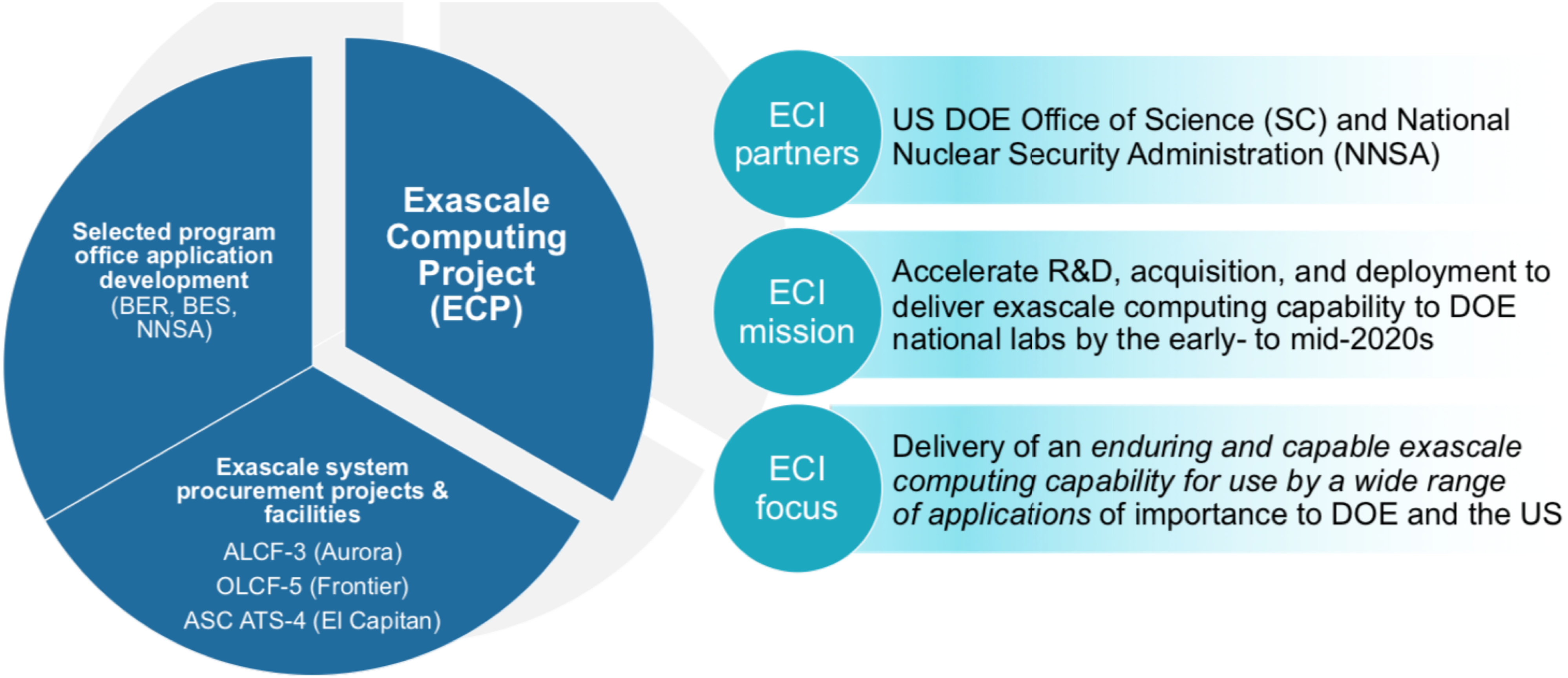
04 Elucidation of How Neural Networks Realize Thinking and Its Application to Artificial Intelligence



L'EXASCALE AUX ETATS-UNIS

ECI : Exascale Computing Initiative

- Juillet 2015 : signature de l'executive order Exascale par B. Obama et mise en place de la National Strategic Computing Initiative (NSCI)
- 2017 : Exascale Computing Initiative pour DoE et NNSA



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

L'EXASCALE AUX ETATS-UNIS

Ne pas oublier aussi !

Décret de D. Trump sur l'IA US

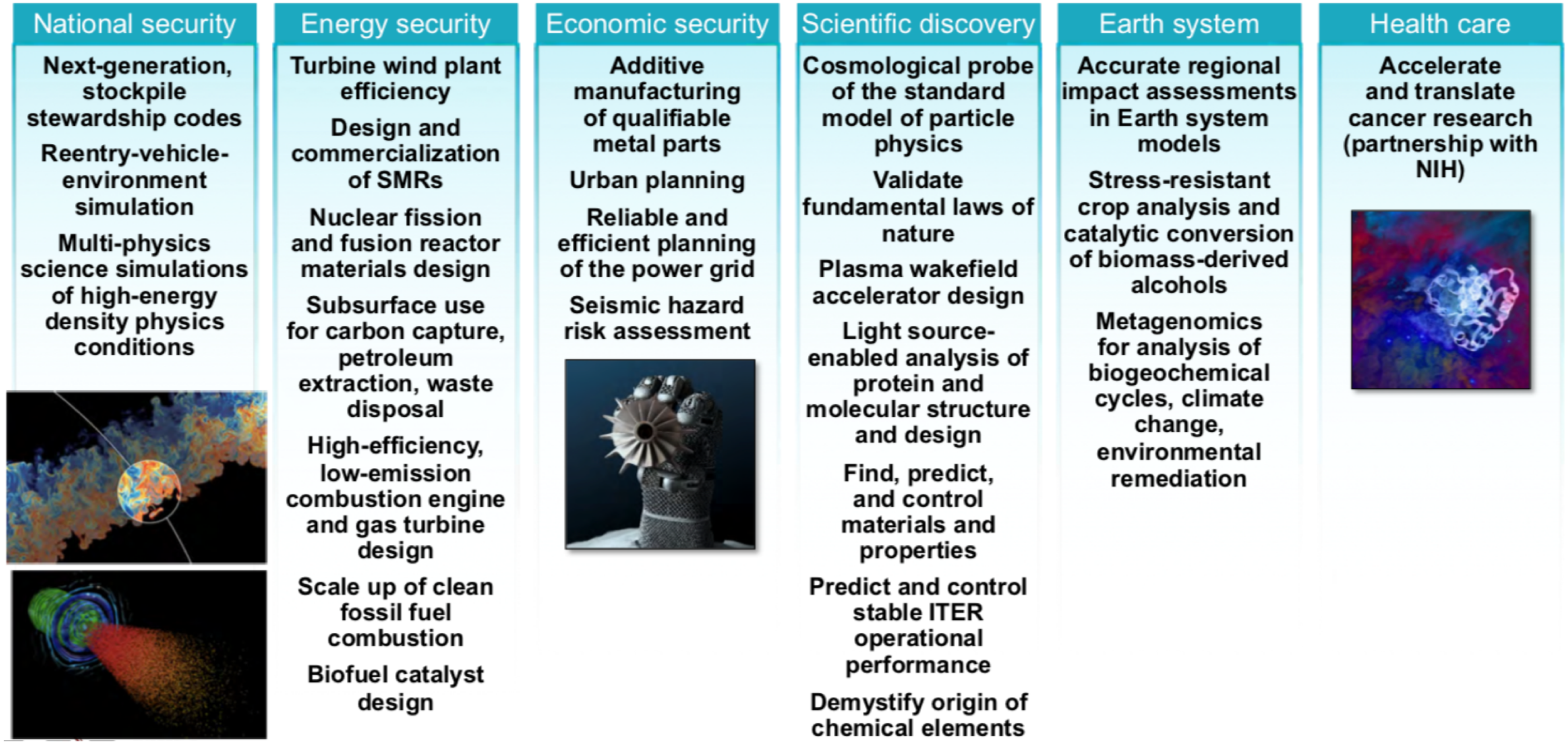
- ❑ Executive Order en date du 11 février
- ❑U.S Artificial Intelligence Initiative « *Maintaining/Accelerating American Leadership in Artificial Intelligence »*
- ❑ 6 objectifs dont notamment
  - ▪ Investissements publics dans R&D en IA
    - Priorisation des investissements agences publiques dans IA
  - ▪ Mise à disposition de ressources HPC pour IA
    - Agences fédérales devront ouvrir leurs moyens HPC à R&D IA
  - ▪ Gouvernance pour guider développement IA
    - Fiabilité, sécurité et interopérabilité (standards)
  - ▪ Formation
    - Tous cycles & filières apprentissage pour aide au changement travailleurs impactés
- ❑Pas de budget annoncé

https://www.whitehouse.gov/articles/accelerating-americas-leadership-in-artificial-intelligence/

L’EXASCALE AUX ETATS-UNIS

ECP : Exascale Computing Project

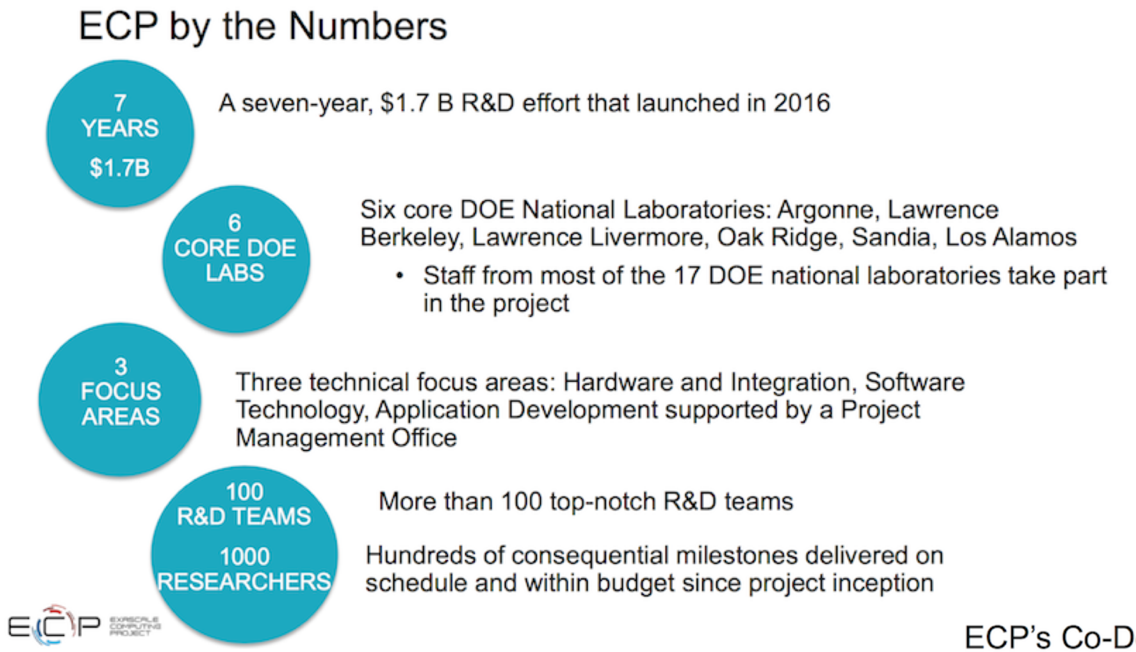
- ❑ Co design et développement de nouvelles approches (HPC+IA)
  - 25 applications sélectionnées
  - > 80 outils et environnements : MPI, OpenMP, OpenACC, LLVM, PAPI/TAU, BLAS, ADIOS, HDF5, Paraview … mais aussi Raja, Kokkos, spack, UnifyCR, …
  - Extreme Scale Scientific Software Stack (e4s.io)
  - Forte contribution dans standards



National security	Energy security	Economic security	Scientific discovery	Earth system	Health care
Next-generation, stockpile stewardship codes Reentry-vehicle-environment simulation Multi-physics science simulations of high-energy density physics conditions	Turbine wind plant efficiency Design and commercialization of SMRs 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	Additive manufacturing of qualifiable metal parts Urban planning Reliable and efficient planning of the power grid Seismic hazard risk assessment	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 stable ITER operational performance Demystify origin of chemical elements	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	Accelerate and translate cancer research (partnership with NIH)

L'EXASCALE AUX ETATS-UNIS

ECP : Exascale Computing Project



ECP by the Numbers

7 YEARS $1.7B — A seven-year, $1.7 B 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

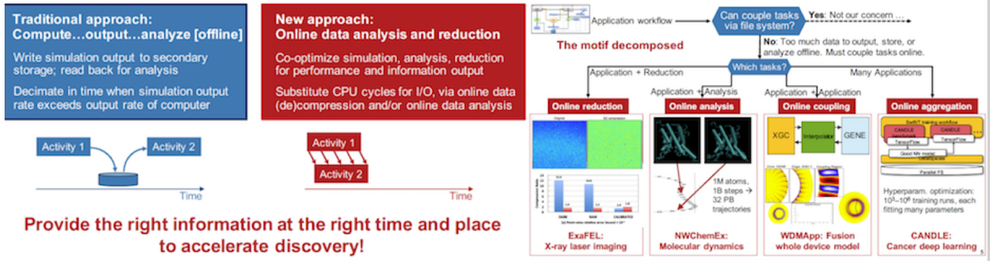
- Staff from most of the 17 DOE national laboratories take part in the project

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

100 R&D TEAMS — More than 100 top-notch R&D teams

1000 RESEARCHERS — Hundreds of consequential milestones delivered on schedule and within budget since project inception

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



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

Activity 1
Activity 2
Time
Activity 1
Activity 2
Time
Application workflow
The motif decomposed
Can couple tasks via file system?
Yes: Not our concern …
No: Too much data to output, store, or analyze offline. Must couple tasks online.
Which tasks?
Application + Reduction
Application + Analysis
Application + Application
Many Applications
Online reduction
Online analysis
Online coupling
Online aggregation
1M atoms, 1B steps → 32 PB trajectories
XGC
Interpolator
GENE
Hyperparam. optimization: 10³–10⁶ training runs, each fitting many parameters
ExaFEL: X-ray laser imaging
NWChemEx: Molecular dynamics
WDMApp: Fusion whole device model
CANDLE: Cancer deep learning

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).

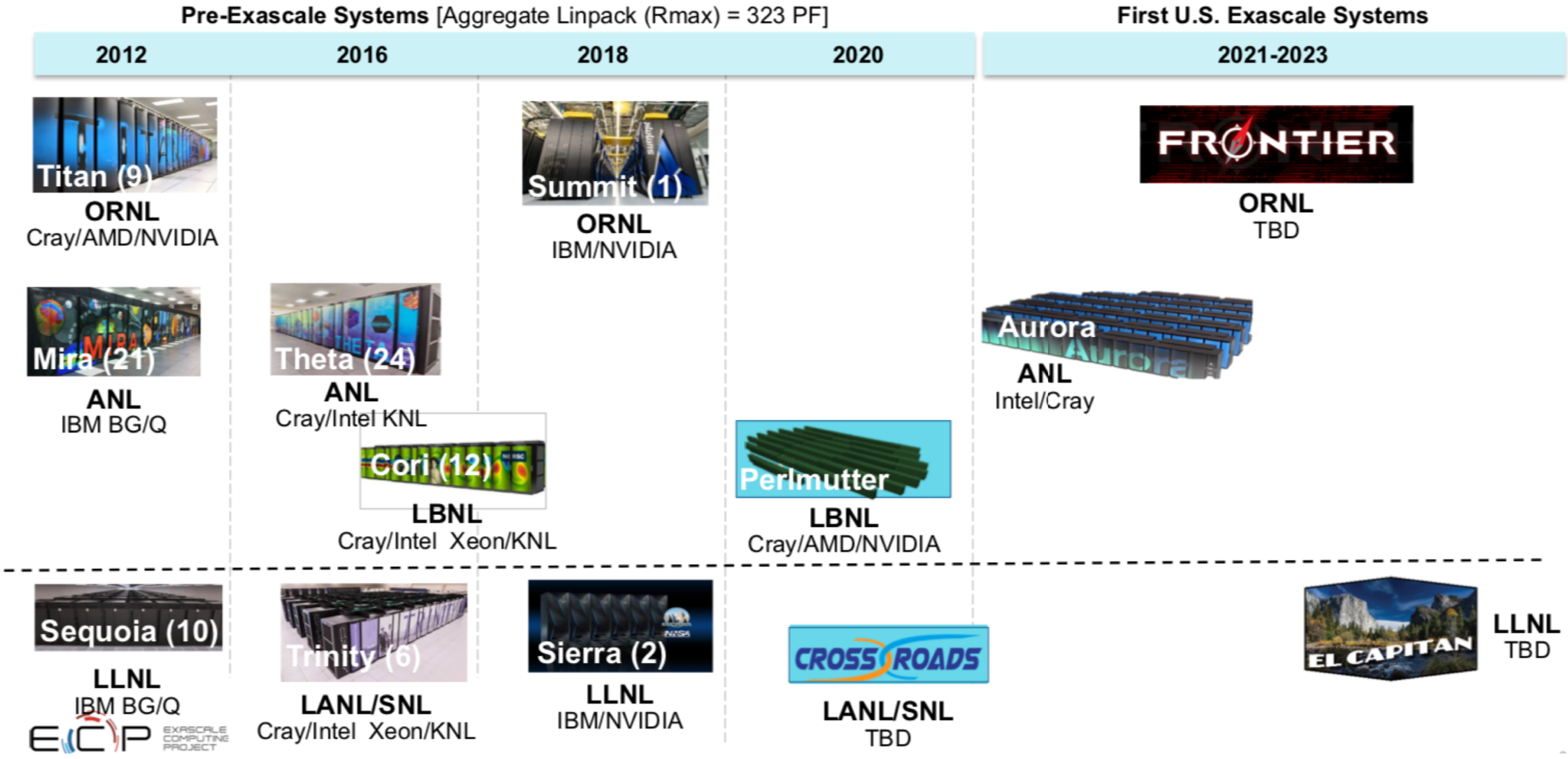
Cross-cutting tools:

- 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)

L'EXASCALE AUX ETATS-UNIS

Roadmap DoE sur 1 slide



Pre-Exascale Systems [Aggregate Linpack (Rmax) = 323 PF]

First U.S. Exascale Systems

2012 | 2016 | 2018 | 2020 | 2021-2023

Titan (9)
ORNL
Cray/AMD/NVIDIA

Summit (1)
ORNL
IBM/NVIDIA

FRONTIER
ORNL
TBD

Mira (21)
ANL
IBM BG/Q

Theta (24)
ANL
Cray/Intel KNL

Aurora
ANL
Intel/Cray

Cori (12)
LBNL
Cray/Intel Xeon/KNL

Perlmutter
LBNL
Cray/AMD/NVIDIA

Sequoia (10)
LLNL
IBM BG/Q

Trinity (6)
LANL/SNL
Cray/Intel Xeon/KNL

Sierra (2)
LLNL
IBM/NVIDIA

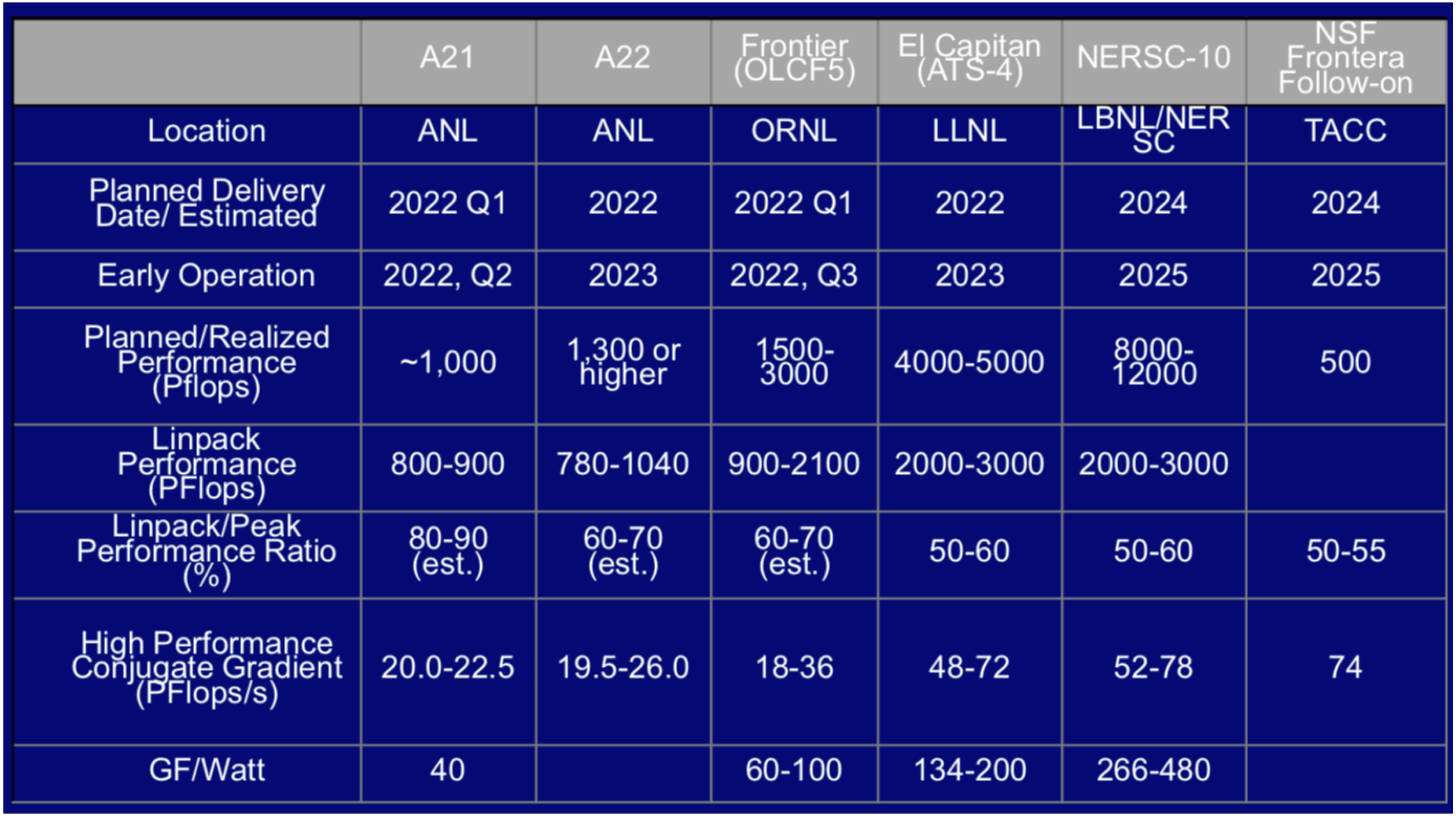
CROSSROADS
LANL/SNL
TBD

EL CAPITAN
LLNL
TBD

ECP EXASCALE COMPUTING PROJECT

L'EXASCALE AUX ETATS-UNIS

Ou deux avec en plus NSF (TACC)



	A21	A22	Frontier (OLCF5)	El Capitan (ATS-4)	NERSC-10	NSF Frontera Follow-on
Location	ANL	ANL	ORNL	LLNL	LBNL/NERSC	TACC
Planned Delivery Date/ Estimated	2022 Q1	2022	2022 Q1	2022	2024	2024
Early Operation	2022, Q2	2023	2022, Q3	2023	2025	2025
Planned/Realized Performance (Pflops)	~1,000	1,300 or higher	1500-3000	4000-5000	8000-12000	500
Linpack Performance (PFlops)	800-900	780-1040	900-2100	2000-3000	2000-3000	
Linpack/Peak Performance Ratio (%)	80-90 (est.)	60-70 (est.)	60-70 (est.)	50-60	50-60	50-55
High Performance Conjugate Gradient (PFlops/s)	20.0-22.5	19.5-26.0	18-36	48-72	52-78	74
GF/Watt	40		60-100	134-200	266-480	

Source : Hyperion HPC User Forum, mars 2019

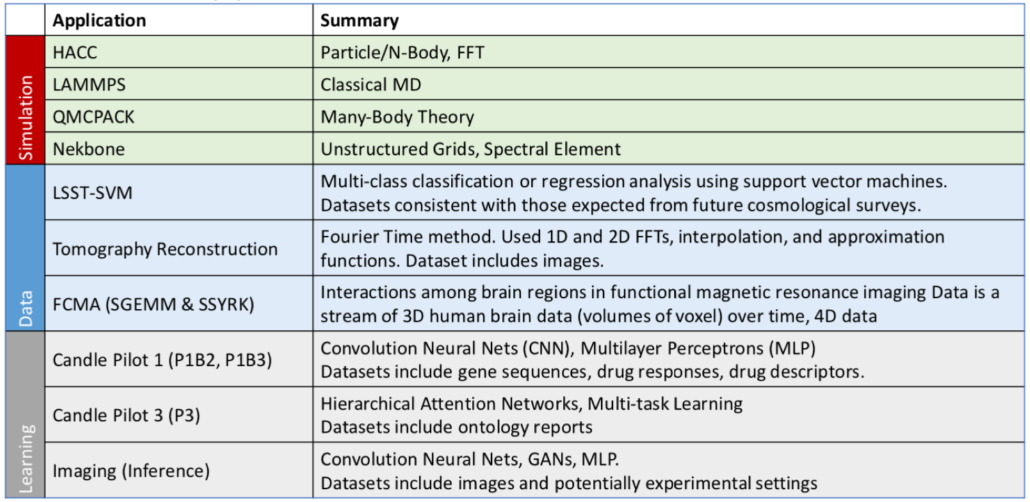
L'EXASCALE AUX ETATS-UNIS

DoE : A21 à Argonne Labs (ANL)

❑A21 (ex Aurora) un système MPP hybride en 2021

- **1^er^ système Exacale US**, fourni par Intel et Cray, report du 1^er^ deal CORAL
- Axé convergence HPC / HPDA / AI
- Annonce 18 mars 2019 du DoE, budget 500M$ (CAPEX)
  - Au moins 200 racks Shasta Cray
  - next gen Intel Xeon CPU 10 nm (3D Foveros) couplés à Intel X^e discrete GPU
  - mémoire persistente Optane DC
  - Interconnect Cray Slingshot Ethernet low latency, topologie Dragonfly, 200 Gbs
- Stack logicielle Intel OneAPI
  - Couche abstraction pour utilisation CPU, GPU, CSA, AI chips et FPGA

A21 FOM Applications



	Application	Summary
Simulation	HACC	Particle/N-Body, FFT
	LAMMPS	Classical MD
	QMCPACK	Many-Body Theory
	Nekbone	Unstructured Grids, Spectral Element
Data	LSST-SVM	Multi-class classification or regression analysis using support vector machines. Datasets consistent with those expected from future cosmological surveys.
	Tomography Reconstruction	Fourier Time method. Used 1D and 2D FFTs, interpolation, and approximation functions. Dataset includes images.
	FCMA (SGEMM & SSYRK)	Interactions among brain regions in functional magnetic resonance imaging Data is a stream of 3D human brain data (volumes of voxel) over time, 4D data
Learning	Candle Pilot 1 (P1B2, P1B3)	Convolution Neural Nets (CNN), Multilayer Perceptrons (MLP) Datasets include gene sequences, drug responses, drug descriptors.
	Candle Pilot 3 (P3)	Hierarchical Attention Networks, Multi-task Learning Datasets include ontology reports
	Imaging (Inference)	Convolution Neural Nets, GANs, MLP. Datasets include images and potentially experimental settings

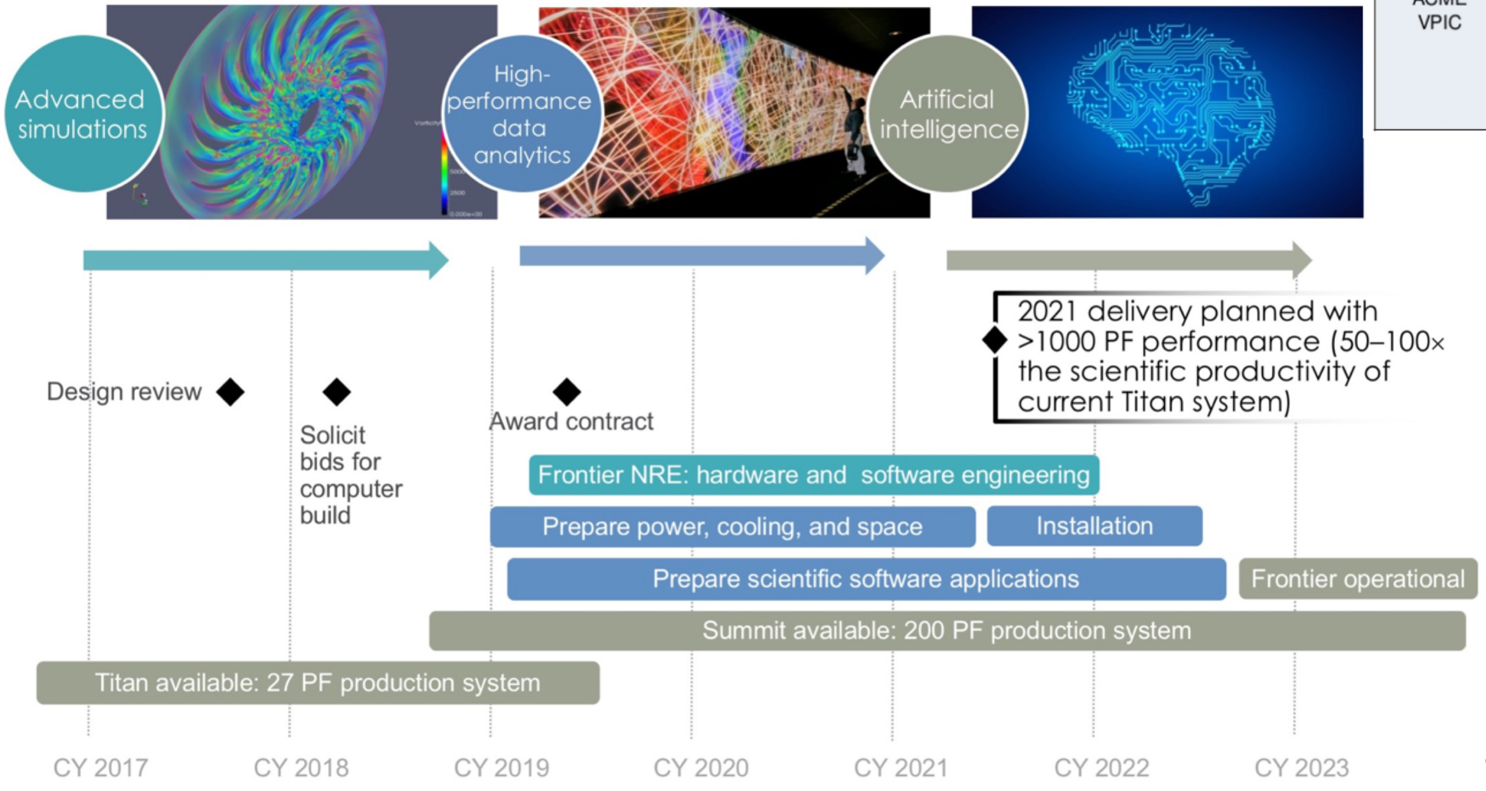
L'EXASCALE AUX ETATS-UNIS

DoE : Frontier à Oak Ridge National Laboratory (ORNL)

❑Appel d'offres CORAL2 du DoE (ORNL + LLNL)

- 1.8 Md$ pour systèmes ORNL et LLNL + update système ANL
- Au moins 1300 PF (FP64) peak performance, 1 GB per MPI task et au moins 8 PB de mémoire totale, max 40MW, 20-30MW de préférence
- Au moins 50x / (Titan, Sequoia) sur apps publiques ->
- CRAY (AMD – CPU+GPU)

Scalable Science	Throughput	Data Science Deep Learning	Skeleton
QMCPACK	Quicksilver	Integer sort	MPI suite
HACC	Kripke	Havoq	Memory suite
NEKbone	LAMMPS	Big Data Suite	GEMM
ACME	AMG	Deep Learning Suite	Pynamic
VPIC	PENNANT		CLOMP
	Laghos		ML suite
			I/O suite
			RAJA suite
			Hash

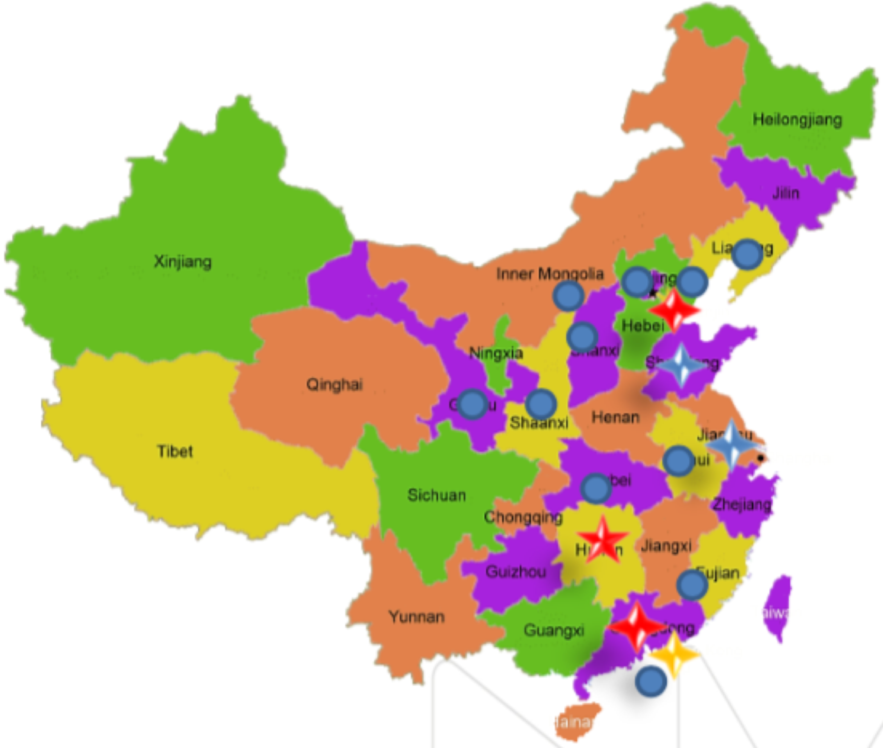

Advanced simulations
High-performance data analytics
Artificial intelligence
Design review
Solicit bids for computer build
Award contract
2021 delivery planned with >1000 PF performance (50–100× the scientific productivity of current Titan system)
Frontier NRE: hardware and software engineering
Prepare power, cooling, and space
Installation
Prepare scientific software applications
Frontier operational
Summit available: 200 PF production system
Titan available: 27 PF production system
CY 2017
CY 2018
CY 2019
CY 2020
CY 2021
CY 2022
CY 2023

L’EXASCALE EN CHINE

Contexte

Status of Chinese SC

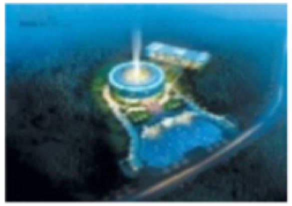
Supercomputing Centers in China



CNGrid

NSCC-Guangzhou,2013
Tianhe-2

NSCC-Wuxi,2016
Shenwei-Taihu Light



NSCC-Changsha,2012
Tianhe-1A

NSCC-Jinan,2012
Shenwei-Bluelight

NSCC-Tianjin,2010
Tianhe-1A

NSCC-Shenzhen,2011
Dawning-6000

Courtesy : **Yutong Lu**

L'EXASCALE EN CHINE

Contexte

- ❑HPC, AI, Quantum = technos stratégiques pour la Chine (13ème plan)
- ❑Exascale targets :
  - ▪ Exaflops in peak performance with HPL efficiency >60%
  - ▪ Node performance > 10 TF, >30 GF/W energy efficiency
  - ▪ 10 PB memory and EB storage
  - ▪ >400 Gbs interconnect b/w
  - ▪ Large scale system management and resource, system monitoring and fault tolerance
  - ▪ Support for large scale applications
    - Numerical nuclear reactor
    - Numerical aircraft (CFD, structure, MDO) and engines (4-stage unsteady LES simulation)
    - Astrophysics and earth system
    - Drug discovery and life sciences
    - Seismic and oil exploration
    - Material sciences
    - Complex engineering (ex: 3 Gorges full dam modeling)
- ❑3 projets en compétition ;

1. NUDT : CPU scalaire + accélérateur (DSP)
2. NRCPC = manycore CPU
3. Dawning (Sugon) : processeur x86 (AMD) + accélérateurs (DCU)

L'EXASCALE EN CHINE

3 projets concurrents, systèmes convergés HPC/HPDA/AI

NUDT (Tianhe-3)

- Reconfigurable flexible heterogeneous arch.
- High-speed interconnect
- based on Chinese-designed Arm technology, likely some version of Phytium's Xiaomi
- ARM processor (>64 cores, > 2TF) + Matrix 3000 DSP accelerator (>96 cores, 10 TF, HBM2, support FP16)
- final version will be operational by 2020
- 200 times faster and 100 times more storage capacity than Tianhe-1
- Water cooling PUE<1.1

Sunway

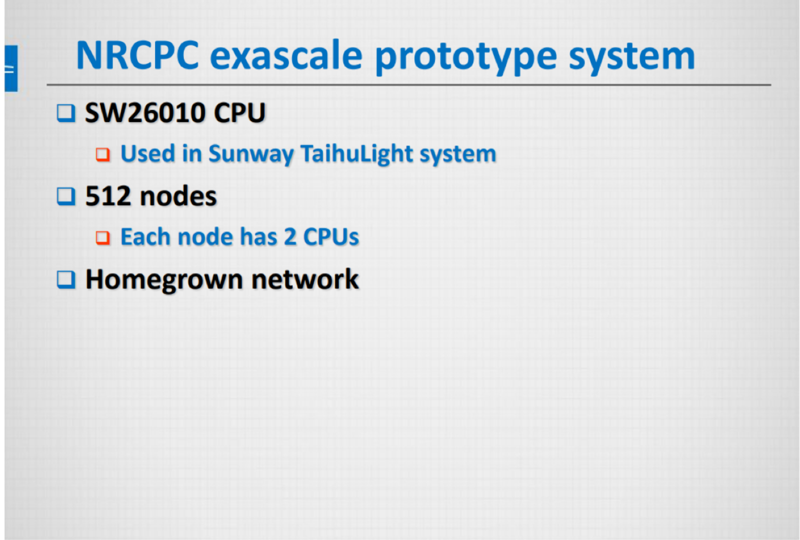
- developed by the National Research Center of Parallel Computer Engineering and Technology (NRCPC)
- New manycore processor (evolution of the current 260-core ShenWei 26010)
- new interconnect
- expect to build the exascale computer in the second half of 2020 or the first half of 2021

Sugon

- home grown Hygon x86 (>1TF) + DCU accelerators (>15 TF) / node
- 400 Gbs 6D torus, 384 ports routers
- Less than 32768 nodes, with 32 TF/node
- China has a licensing agreement between Hygon and AMD (Zen 1 architecture for the moment)
- Immersive cooling technology (Imm058)

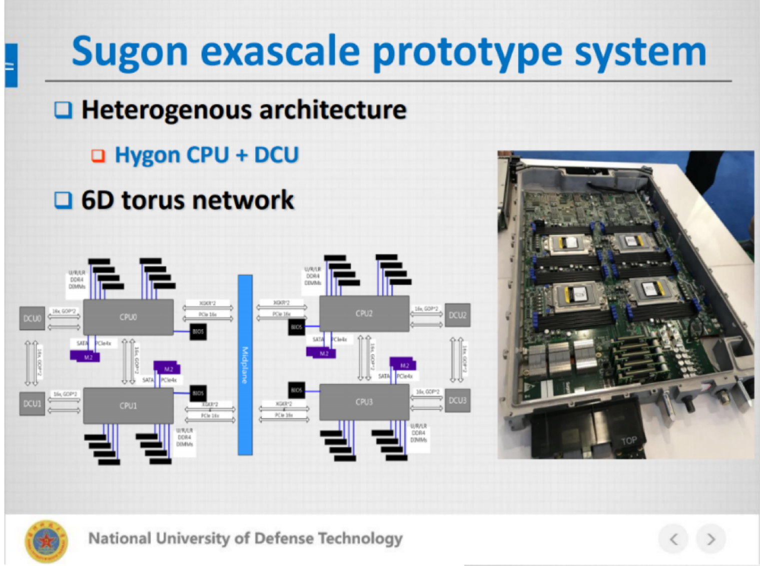
L'EXASCALE EN CHINE

3 projets concurrents, systèmes convergés HPC/HPDA/AI



NRCPC exascale prototype system

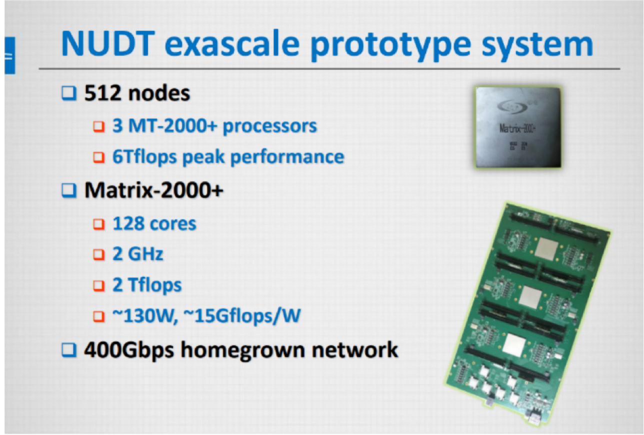
- SW26010 CPU
  - Used in Sunway TaihuLight system
- 512 nodes
  - Each node has 2 CPUs
- Homegrown network



Sugon exascale prototype system

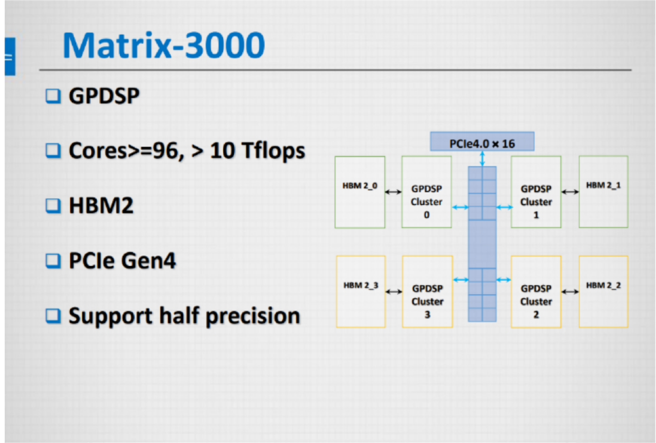
- Heterogenous architecture
  - Hygon CPU + DCU
- 6D torus network

National University of Defense Technology



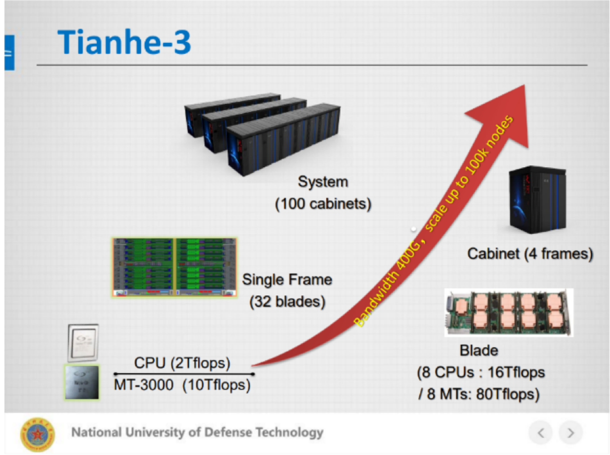
NUDT exascale prototype system

- 512 nodes
  - 3 MT-2000+ processors
  - 6Tflops peak performance
- Matrix-2000+
  - 128 cores
  - 2 GHz
  - 2 Tflops
  - ~130W, ~15Gflops/W
- 400Gbps homegrown network



Matrix-3000

- GPDSP
- Cores>=96, > 10 Tflops
- HBM2
- PCIe Gen4
- Support half precision

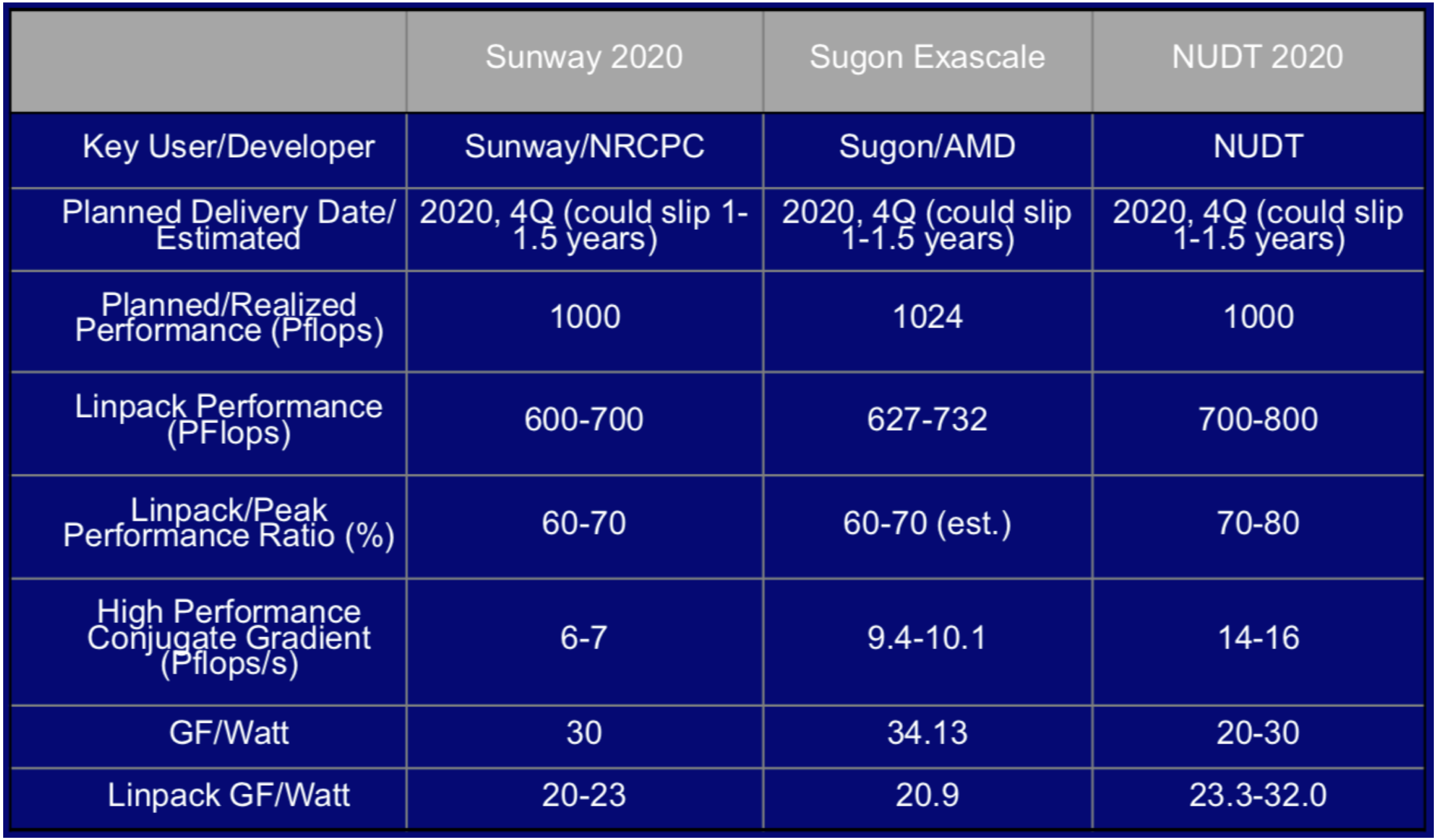


Tianhe-3

National University of Defense Technology

L'EXASCALE EN CHINE

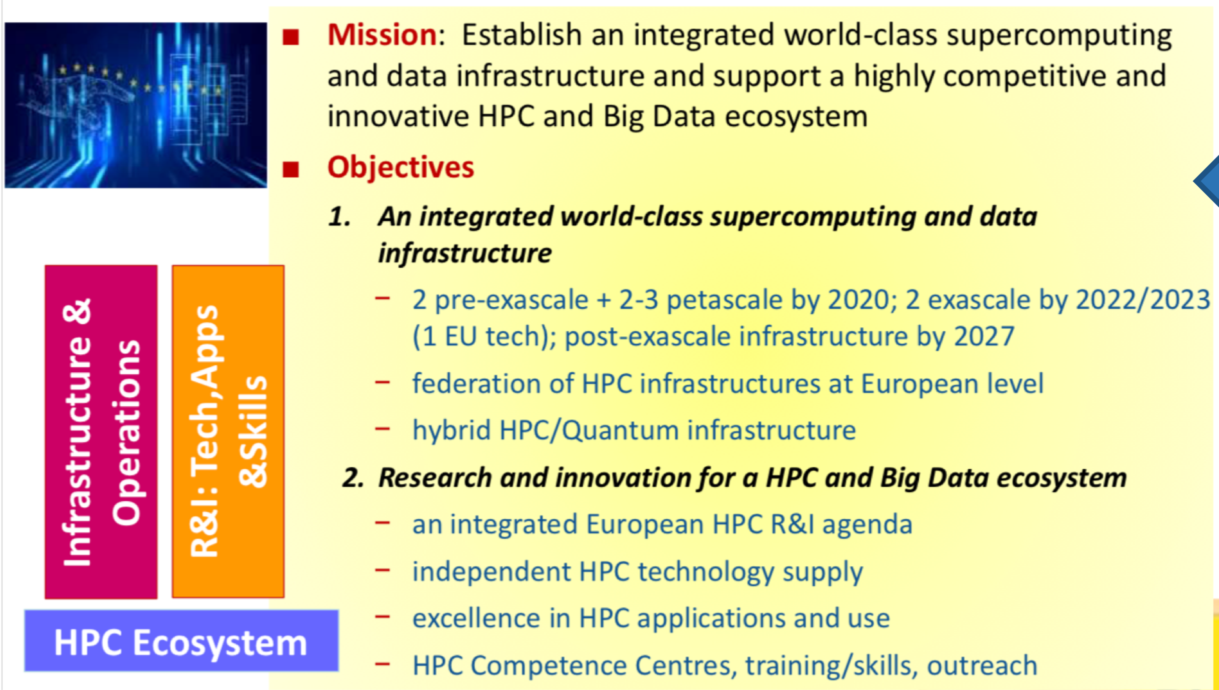
Sur 1 slide les 3 projets en lice



	Sunway 2020	Sugon Exascale	NUDT 2020
Key User/Developer	Sunway/NRCPC	Sugon/AMD	NUDT
Planned Delivery Date/ Estimated	2020, 4Q (could slip 1-1.5 years)	2020, 4Q (could slip 1-1.5 years)	2020, 4Q (could slip 1-1.5 years)
Planned/Realized Performance (Pflops)	1000	1024	1000
Linpack Performance (PFlops)	600-700	627-732	700-800
Linpack/Peak Performance Ratio (%)	60-70	60-70 (est.)	70-80
High Performance Conjugate Gradient (Pflops/s)	6-7	9.4-10.1	14-16
GF/Watt	30	34.13	20-30
Linpack GF/Watt	20-23	20.9	23.3-32.0

EURO-HPC

Mutualiser pour aller vers l'Exascale



- **Mission**: Establish an integrated world-class supercomputing and data infrastructure and support a highly competitive and innovative HPC and Big Data ecosystem
- **Objectives**
  1. ***An integrated world-class supercomputing and data infrastructure***
     - 2 pre-exascale + 2-3 petascale by 2020; 2 exascale by 2022/2023 (1 EU tech); post-exascale infrastructure by 2027
     - federation of HPC infrastructures at European level
     - hybrid HPC/Quantum infrastructure
  2. ***Research and innovation for a HPC and Big Data ecosystem***
     - an integrated European HPC R&I agenda
     - independent HPC technology supply
     - excellence in HPC applications and use
     - HPC Competence Centres, training/skills, outreach

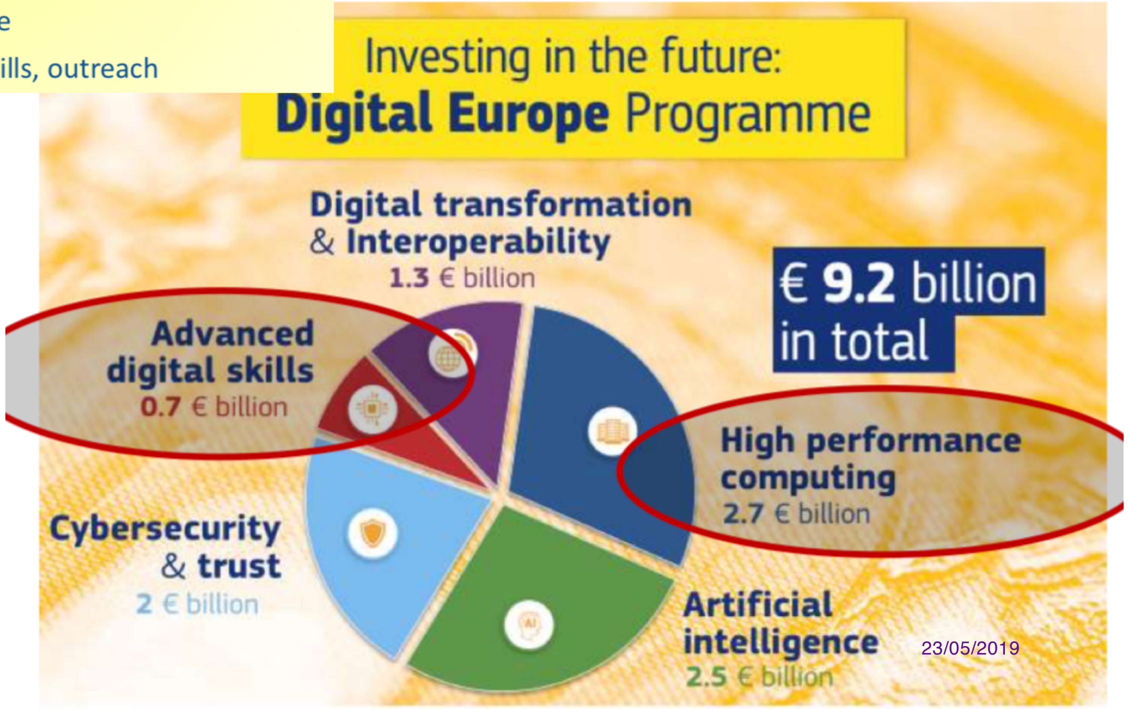
>1 Md€ alloués pour la 1ere phase (Pré Exascale)

Plus de 3Md€ visés dans le FP9 pour la 2ème phase (Exascale)

"Our goal is for Europe to become one of the top 3 world leaders in high-performance computing by 2020."

Jean-Claude Juncker, 27 October 2015

Fin 2018 : 25 pays partenaires !



EXASCALE EN EUROPE

Le projet EPI : European Processor Initiative

One of the lighthouse R&I project of EuroHPC Joint Undertaking

epi European Processor Initiative

BMW GROUP · MINI · Rolls-Royce Motor Cars Limited · Atos · Infineon · BSC Barcelona Supercomputing Center Centro Nacional de Supercomputación

KALRAY · JÜLICH Forschungszentrum · semidynamics silicon design and verification services · TÉCNICO LISBOA · Fraunhofer ITWM

ALMA MATER STUDIORUM UNIVERSITÀ DI BOLOGNA · CHALMERS · UNIVERSITÀ DI PISA · FER · E4 COMPUTER ENGINEERING · CINECA · cea

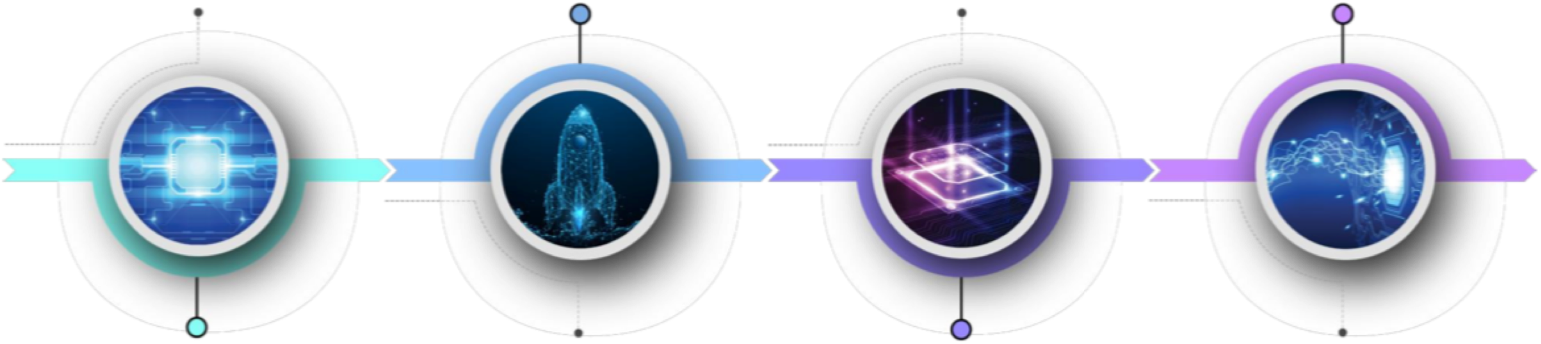
SURF SARA · GENCI · FORTH INSTITUTE OF COMPUTER SCIENCE · ST · EXTOLL latency matters.

ETH zürich · Elektrobit

EXASCALE EN EUROPE

Le projet EPI : European Processor Initiative

Current roadmap



2021

Rhea Family - Gen1 GPP

EPI Common Platform
ARM & RISC-V
External IPs
•
HPC System PreExascale
Automotive PoC

EPI IP's Launch Pad
&
Pan European Research Platform for HPC and AI

2021-2022

2022-2023

Cronos Family - Gen2 GPP

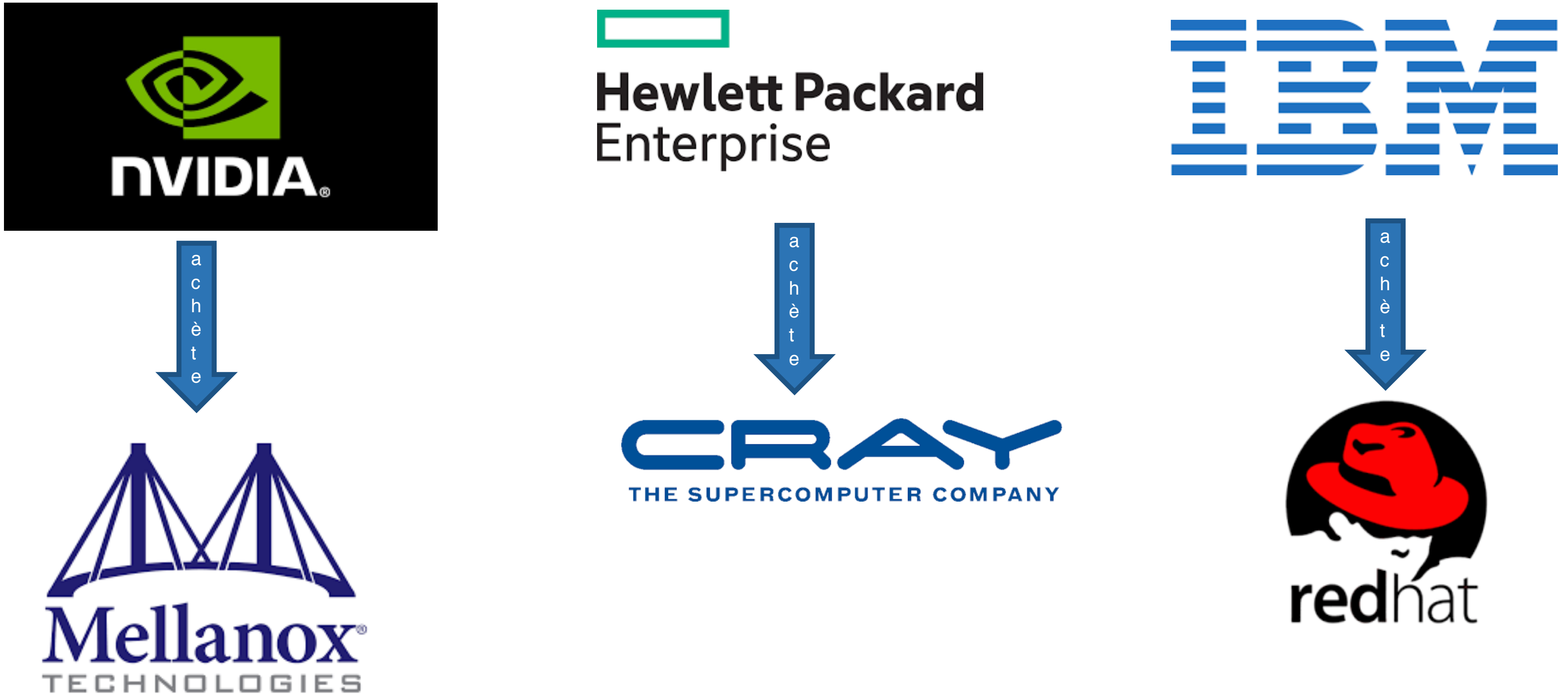
EPI Common Platform
ARM & RISC-V
•
HPC System Exascale
Automotive CPU

Gen3 GPP Family

2024-...

LA CONTRACTION DES ACTEURS

- Mouvements importants sur le marché du HPC sur les derniers 6 mois



NVIDIA

achète

Mellanox TECHNOLOGIES

Hewlett Packard Enterprise

achète

CRAY THE SUPERCOMPUTER COMPANY

IBM

achète

redhat

- D'autres mouvements à venir?

CONCLUSION

❑L’exascale est un projet cohérent à part entière :

- Pas du business-as-usual
- Nécessite de se définir des cibles applicatives et de travailler sur un couple {cibles applicatives ; architecture}
  - Sachant que les architectures doivent satisfaire à des contraintes fortes (comme la consommation énergétique par exemple)
  - L’adaptation (refonte) des codes est donc nécessaire – MAIS ne peut se faire indépendamment de l’archi HW
  - Comme l’archi HW ne peut se définir indépendamment des cibles applicatives

❑Nécessite la mobilisation de compétences complémentaires au service de l’objectif

- Ce n’est pas que du financement pour acheter et concevoir du matériel

❑Projet dans la durée

- Si le design d’une architecture exascale prend du temps, la conception d’une application pour l’exascale prend encore plus de temps !