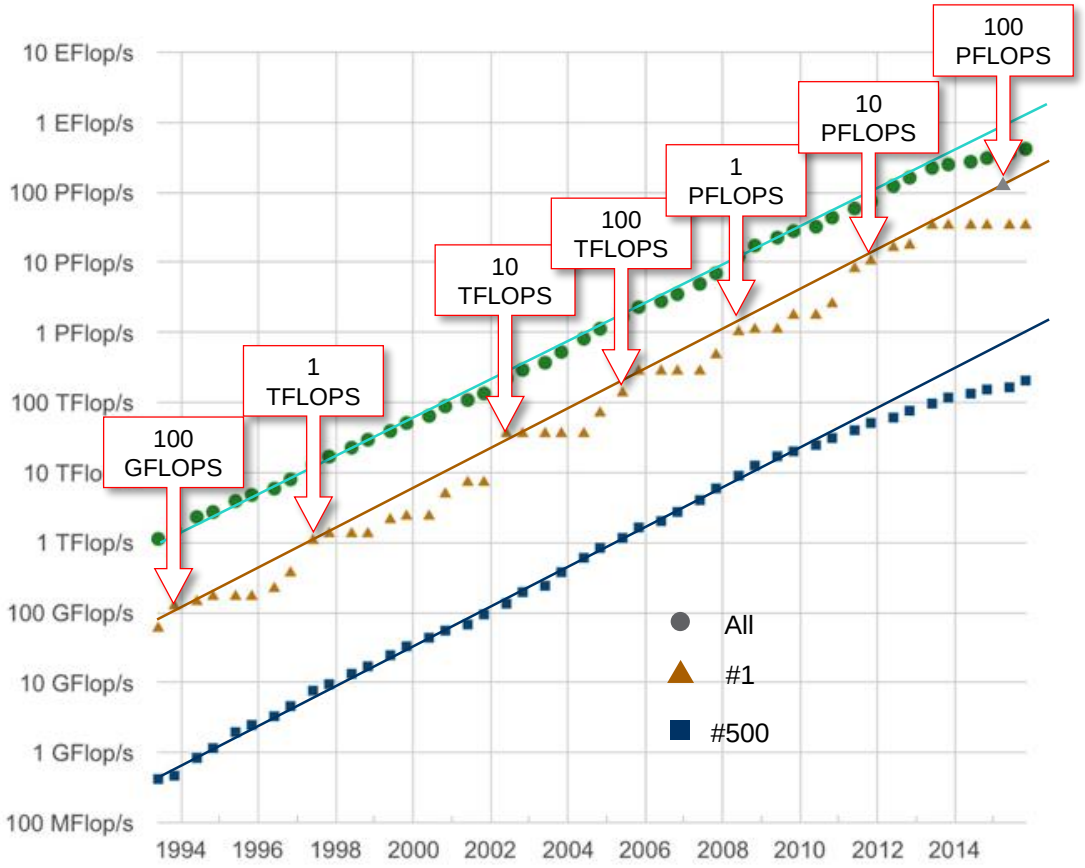
Head to the Exascale …

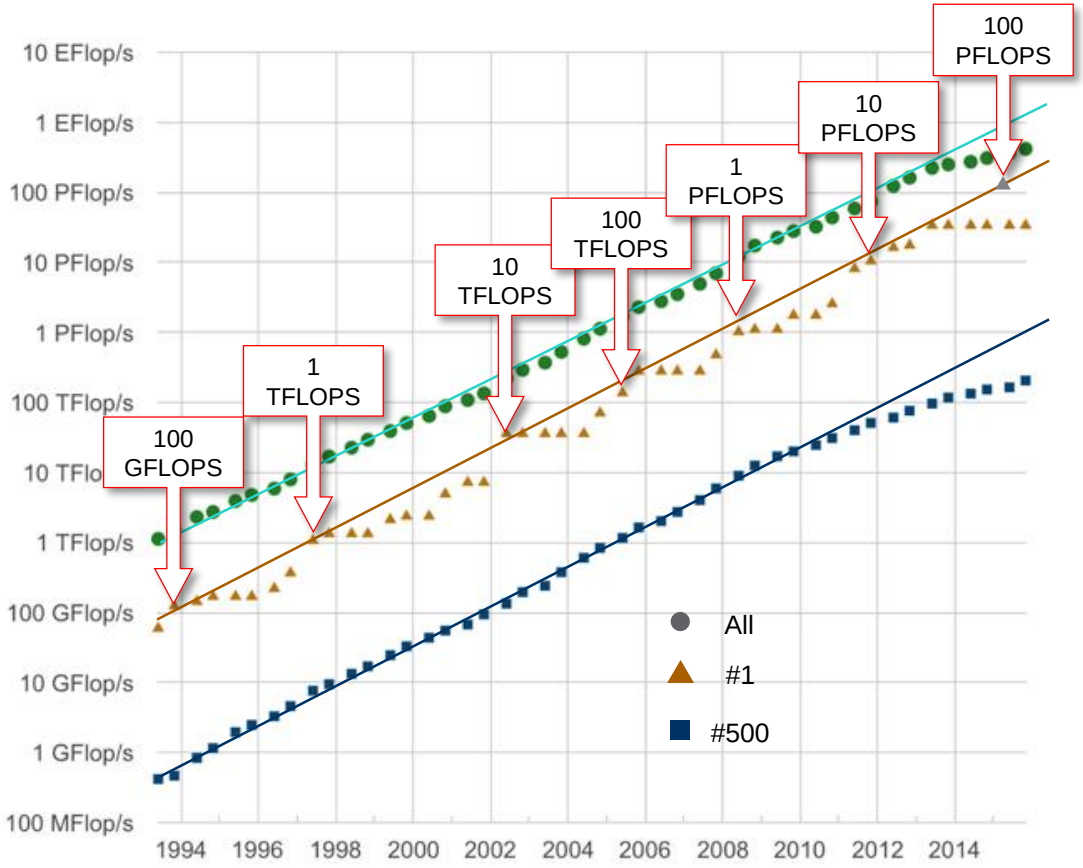
Tibor Temesi
Silicon Computers

July 11/2019

Top500 effects



Top500 effects



What we are aiming at

the top10 now

Rank	System	Cores	Rmax (TFlop/s)	Rpeak (TFlop/s)	Power (kW)
1	**Summit** - IBM Power System AC922, IBM POWER9 22C 3.07GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband , **IBM** DOE/SC/Oak Ridge National Laboratory **United States**	2,414,592	148,600.0	200,794.9	10,096
2	**Sierra** - IBM Power System S922LC, IBM POWER9 22C 3.1GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband , **IBM / NVIDIA / Mellanox** DOE/NNSA/LLNL **United States**	1,572,480	94,640.0	125,712.0	7,438
3	**Sunway TaihuLight** - Sunway MPP, Sunway SW26010 260C 1.45GHz, Sunway , **NRCPC** National Supercomputing Center in Wuxi **China**	10,649,600	93,014.6	125,435.9	15,371
4	**Tianhe-2A** - TH-IVB-FEP Cluster, Intel Xeon E5-2692v2 12C 2.2GHz, TH Express-2, Matrix-2000 , **NUDT** National Super Computer Center in Guangzhou **China**	4,981,760	61,444.5	100,678.7	18,482
5	**Frontera** - Dell C6420, Xeon Platinum 8280 28C 2.7GHz, Mellanox InfiniBand HDR , **Dell EMC** Texas Advanced Computing Center/Univ. of Texas **United States**	448,448	23,516.4	38,745.9	
6	**Piz Daint** - Cray XC50, Xeon E5-2690v3 12C 2.6GHz, Aries interconnect , NVIDIA Tesla P100 , **Cray Inc.** Swiss National Supercomputing Centre (CSCS) **Switzerland**	387,872	21,230.0	27,154.3	2,384
7	**Trinity** - Cray XC40, Xeon E5-2698v3 16C 2.3GHz, Intel Xeon Phi 7250 68C 1.4GHz, Aries interconnect , **Cray Inc.** DOE/NNSA/LANL/SNL **United States**	979,072	20,158.7	41,461.2	7,578
8	**AI Bridging Cloud Infrastructure (ABCI)** - PRIMERGY CX2570 M4, Xeon Gold 6148 20C 2.4GHz, NVIDIA Tesla V100 SXM2, Infiniband EDR , **Fujitsu** National Institute of Advanced Industrial Science and Technology (AIST) **Japan**	391,680	19,880.0	32,576.6	1,649
9	**SuperMUC-NG** - ThinkSystem SD650, Xeon Platinum 8174 24C 3.1GHz, Intel Omni-Path , **Lenovo** Leibniz Rechenzentrum **Germany**	305,856	19,476.6	26,873.9	
10	**Lassen** - IBM Power System S922LC, IBM POWER9 22C 3.1GHz, Dual-rail Mellanox EDR Infiniband, NVIDIA Tesla V100 , **IBM / NVIDIA / Mellanox** DOE/NNSA/LLNL **United States**	288,288	18,200.0	23,047.2	

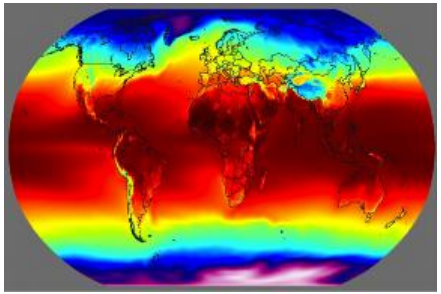
10x

Exascale

100x

HPC and AI can dramatically improve the way we live, work, and innovate

Weather



- Timely and more precise weather forecasting
- Improved understanding of climate change

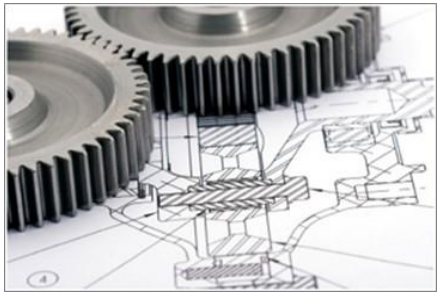
Energy

- Wind energy optimization
- Better photovoltaic efficiency

Life Sciences

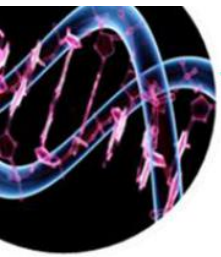
- Accelerated drug discovery
- More personalized healthcare

Manufacturing

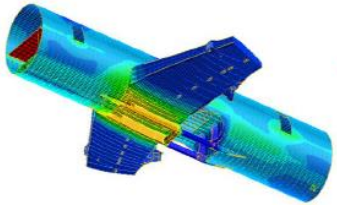


- Predictive and prescriptive maintenance
- Automating product lifecycle management
- Short design cycles

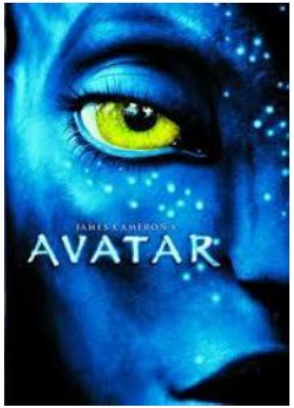
– **Better quality, competitiveness, and time to market across every sector**

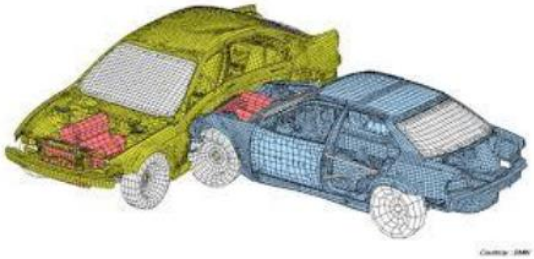


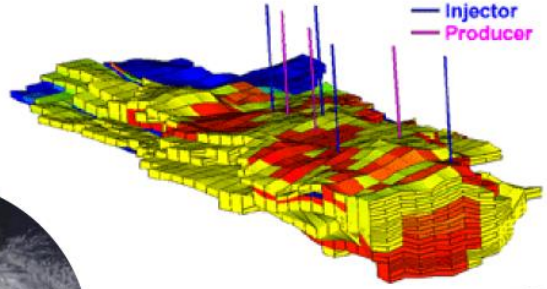
Where is HPC ?

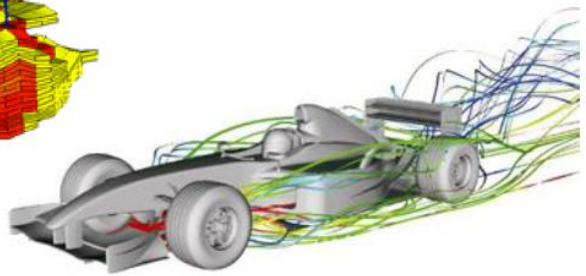


- Human genome sequencing
- **Nuclear Stockpile Simulation**
- **Airplane/car manufacturers**
- **Military systems**
- Rendering farms
- **Oil & Gas, reservoir simulation, seismic…**
- Chemistry
- Banks
- Formula 1
- **Weather forecast**
- Universities









The main trends

- After a boring period, increasing demand for the higher computing performance, the reasons:
  - HPC delivers compelling financial returns
  - New HPC areas: AI, ML, and DL

- the HPC technologies are constantly evolving – how to increase the computing performance?
  - New (Exascale) architecture
  - Computing devices (CPU, GPU, FPGA, …)
  - Bandwidth between the computing devices and the memory (HBM)
  - (Optical) Interconnect: more I/O and less latency

HPC delivers compelling financial returns

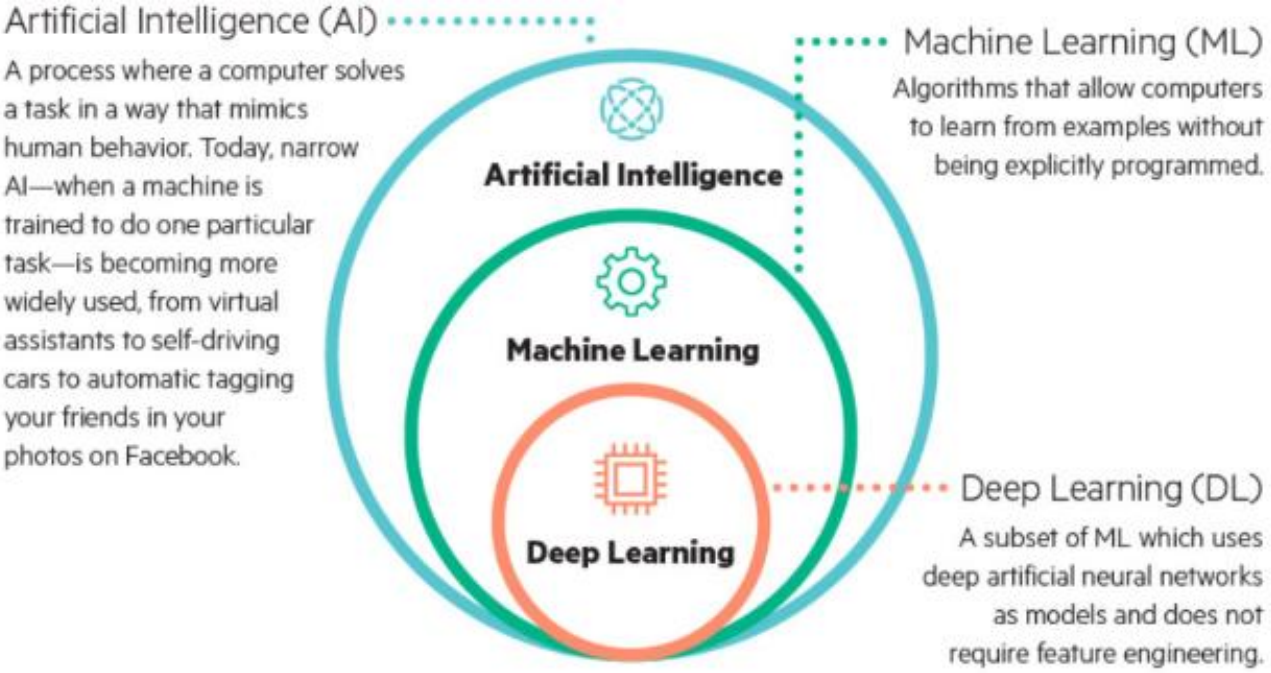
Up to $463 dollars on average in revenue and up to $44 on average of profits (or cost savings) per dollar invested in HPC[1]

	Financial Services	Oil and Gas	Life Sciences	Manufacturing
				
	Accelerating trading & portfolio management	Transforming exploration and production	Enhancing patient care and predictive medicine	Automating product lifecycle management
Avg. of Revenue $ per HPC $	**$834**	**$416**	**$160**	**$83**
Avg. of Profit or Cost Savings $ per HPC $	**$61**	**$54**	**$41**	**$20**

[1] Hyperion Research Study of HPC ROI, June 2018

New fields of use: diving into Deep Learning

"We need to go deeper"



The main trends

- After a boring period increasing demand for the higher computing performance, the reasons:
  - HPC delivers compelling financial returns
  - New HPC areas: AI, ML, and DL

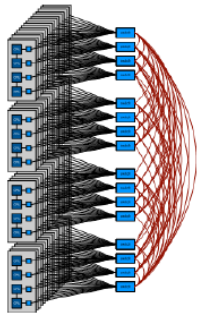
- the HPC technologies are constantly evolving – how to increase the computing performance?
  - New (Exascale) architecture
  - Computing devices (CPU, **GPU**, FPGA, …)
  - Bandwidth between the computing devices and the memory (HBM)
  - (Optical) Interconnect: more I/O and less latency

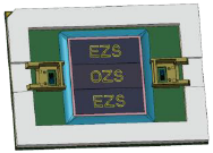
PathForward - Exascale R&D program

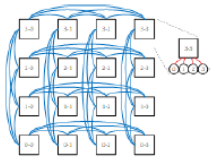
Central element of DOE's Exascale Compute Program (ECP)

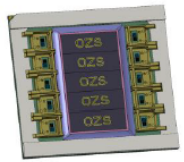
U.S. DEPARTMENT OF ENERGY

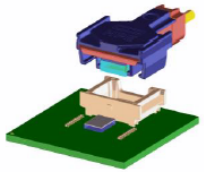
System architecture	Silicon design	Fabrics	Optical interconnects	I/O Subsystem

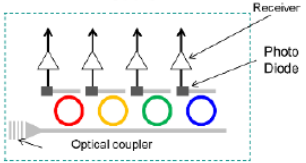


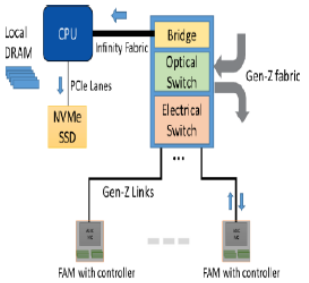












Technology Development

- Exascale architecture
- Application co-design
- System monitoring
- Power and cooling

- Gen-Z Chipset Development and Integration

GEN Z

- Network modeling
- Router architecture
- Topology (OZS ASIC)
- Gen-Z software (MPI)

- Photonic Volume Fab
- Packaging of photonics, electronics, fibers

- Media controller
- File and data access library

Explosion of processing elements

Despite recent turmoil, Xeon will remain strong and will keep going

US exascale remains a strong technology driver

Dedicated to built a very “capable” CPU

Lots of HBM2, rich I/O subsystem

Laid the “golden egg”, keep iterating (MMA, etc). CUDA was developed with almost perfect timing

HPC used to be able to scoop up the good consumer parts for a premium, but now has to battle with the growing AI market

Google TPUs and a many many silicon valley startups

Embraced the MCM route first: allows for more die periphery, easier thermals, enhanced connectivity

GPU roadmap/software catching up

Vector Engine now available to any system vendor - with good HBM2

First ”credible” HPC ARM CPU

8 channels, good flops, flexibility to adopt new technologies

European Processor Initiative

Risc-V or ARM?

WW availability?

Others… Some more or less ”relevant”…

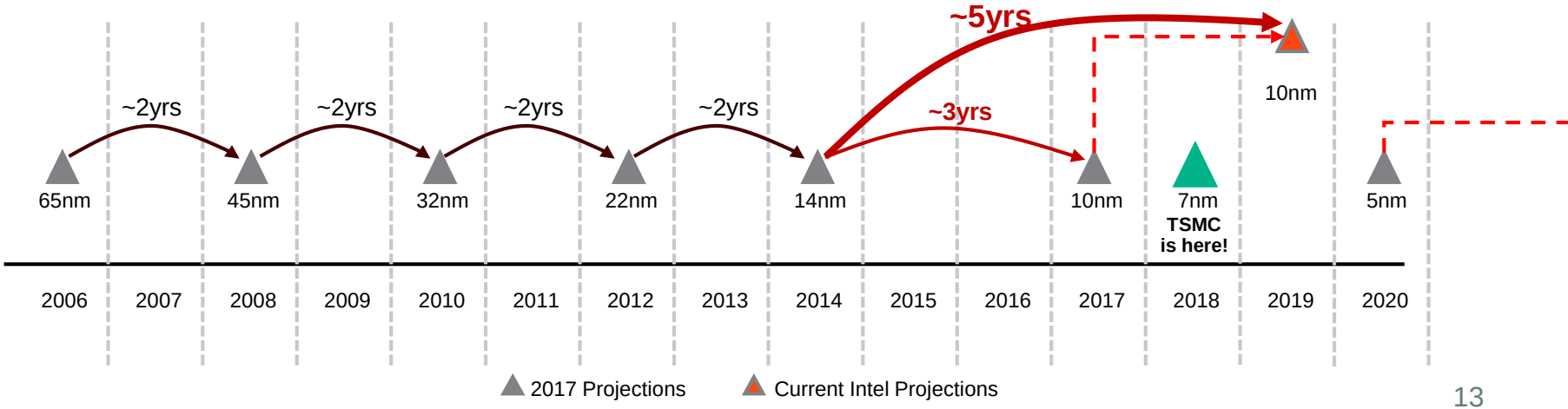
Committed to HPC through its integrators

Stepping up on software enablement, next gen features, go to market

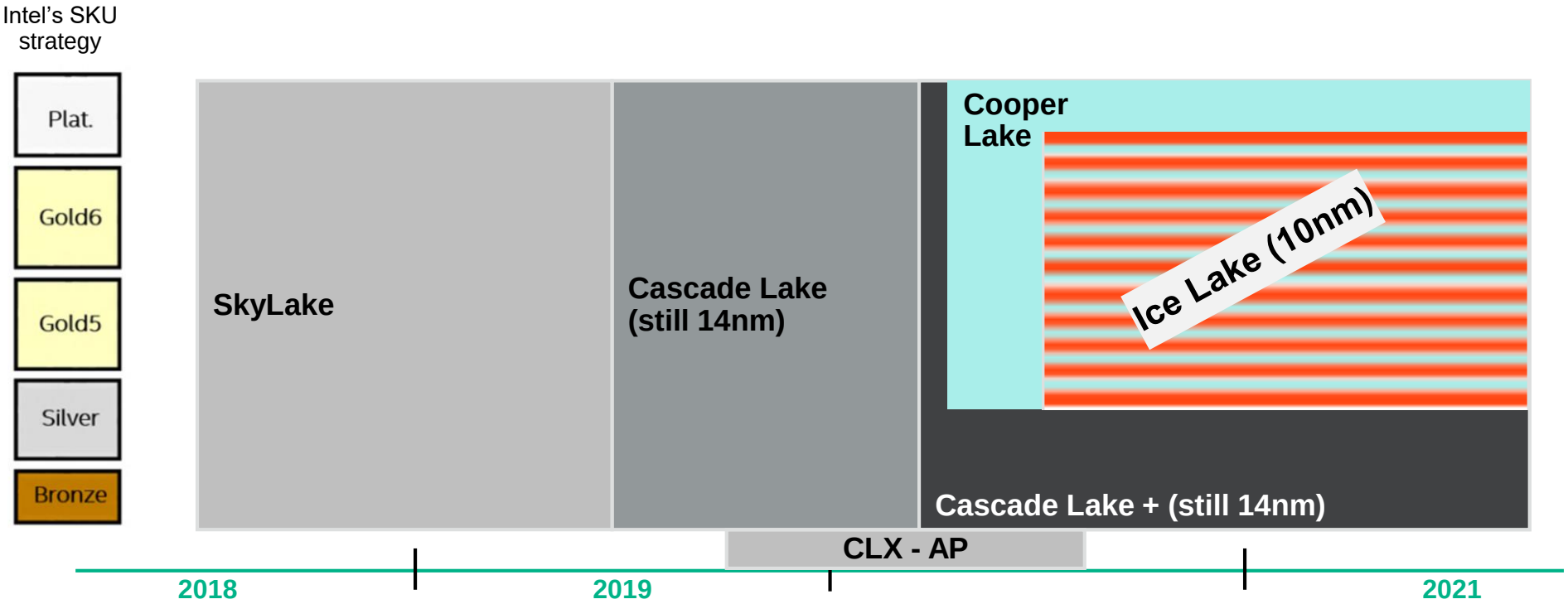
The Compute Paradigm is undergoing Transformation

- Moore's Law has been struggling for some time.
  - Single-threaded performance hasn't increased in ~10 years
  - Paradigms are shifting to increase performance and lower licensing costs
  - Accelerators gaining acceptance

- Intel stumbling in fab lithography transitions
  - Slipped 2 years on an already protracted transition
  - TSMC is yielding well at 7nm; Intel yet to reach 10nm
  - Opportunities for AMD, ARM, and others

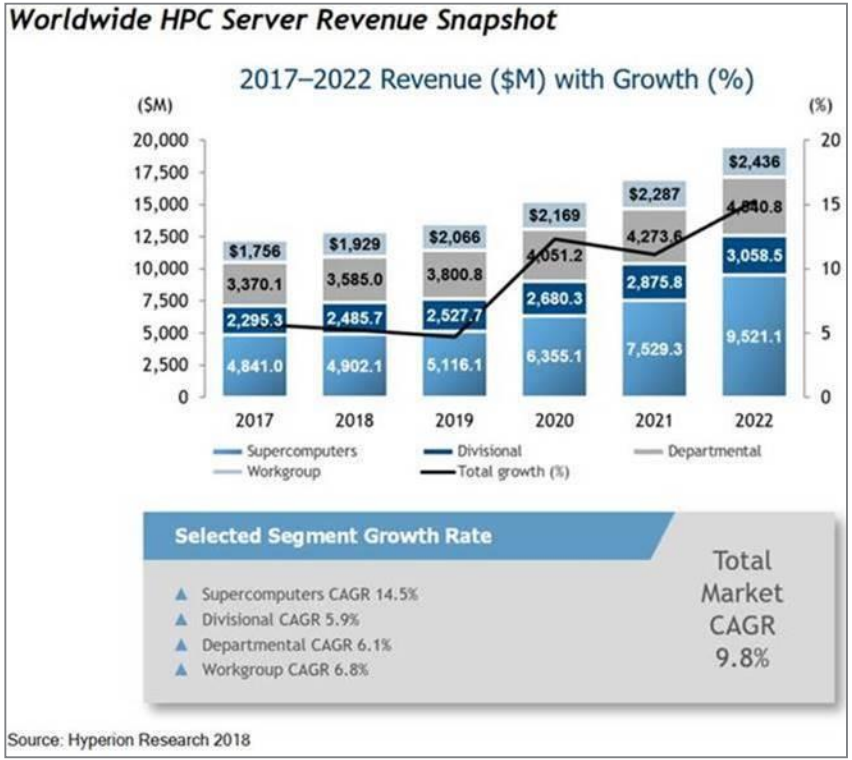


Intel’s General Xeon Roadmap



The good news: the HPC industry is growing fast again!!!

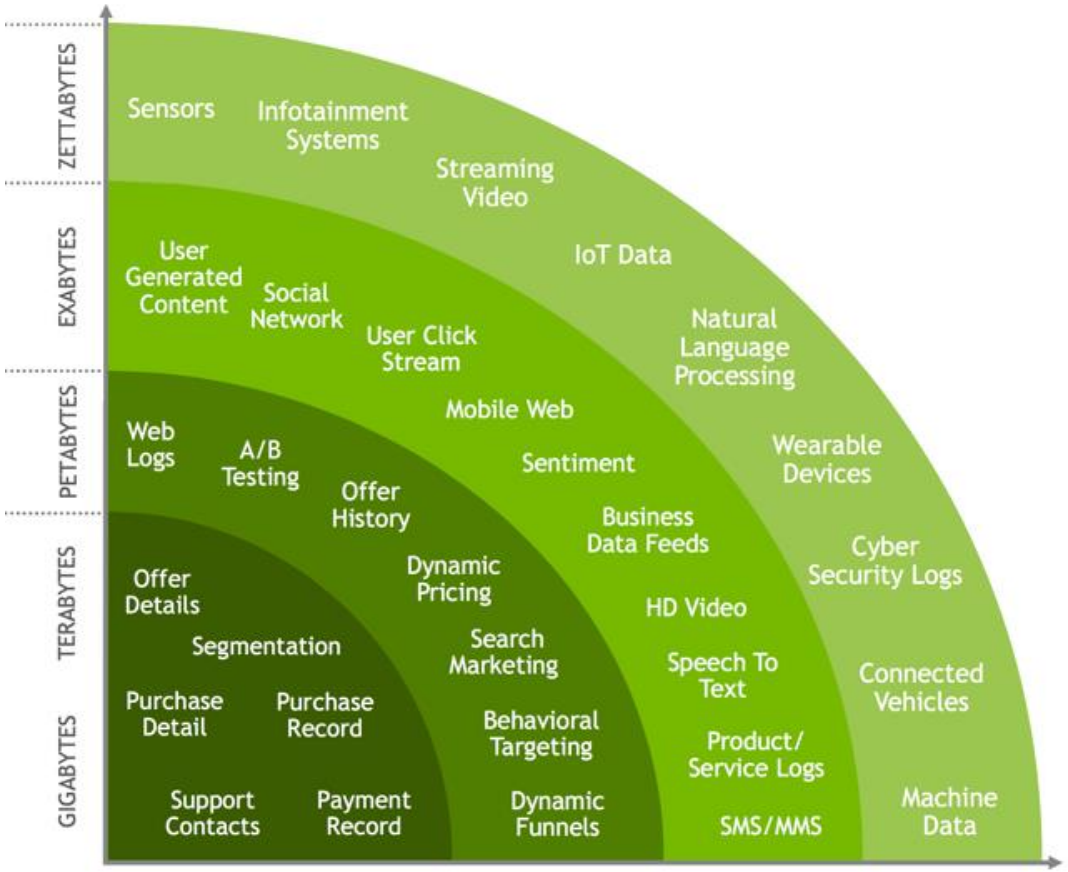
And the high-end is the fastest growing segment



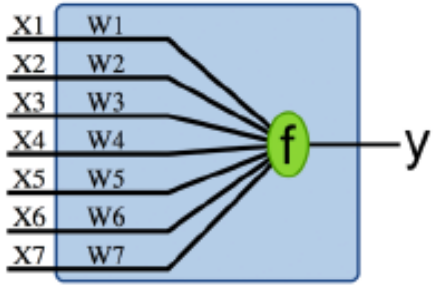
- Exascale initiatives
- New HPC areas and workflows
  HPDA, AI, AI in the loop, data driven simulation, muti-scale, multi-physics
- Operational HPC
  Weather, emergency response, smart cities
- Dramatic increase in Average System Price
  Less systems but a lot bigger!

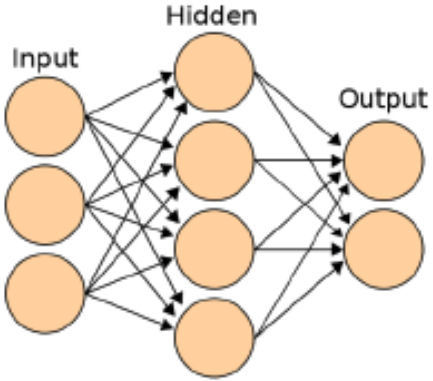
AI + HPC, the new drivers of the REVOLUTION

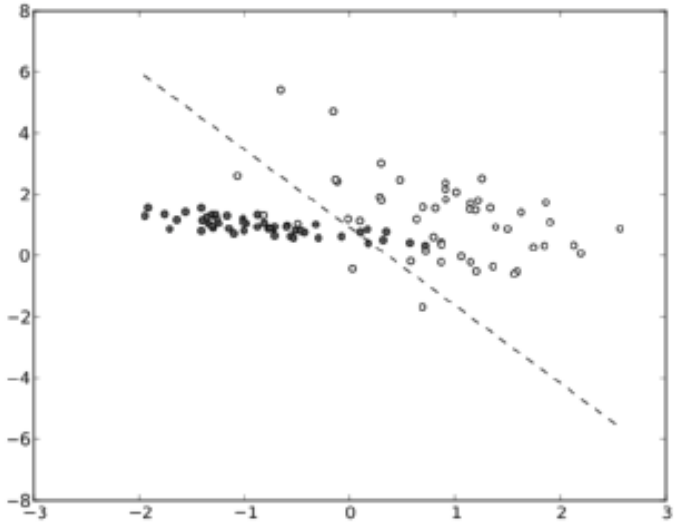
Ingredients: Big Data

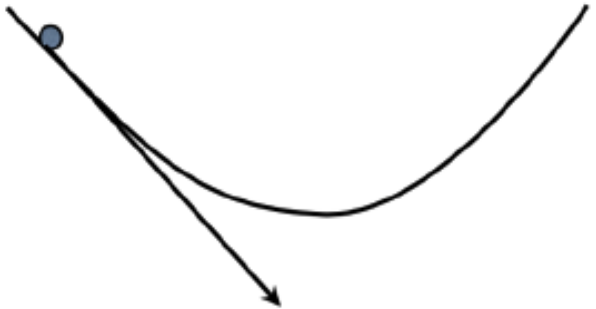


Ingredients: AI algorithms

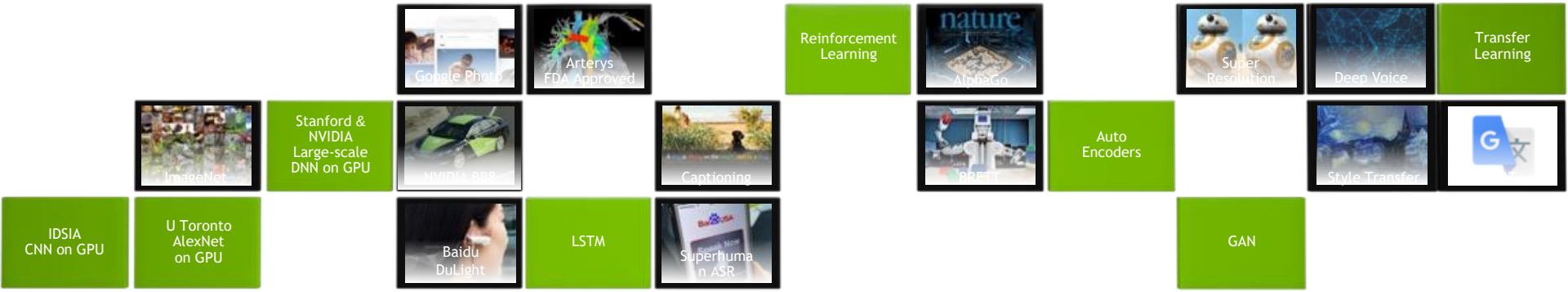








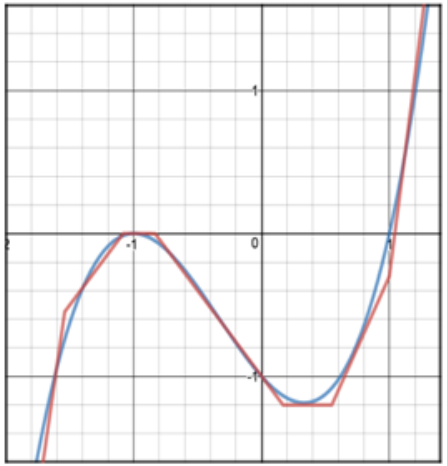
Algorithms + Big Data + GPUs = The Big Bang of modern AI



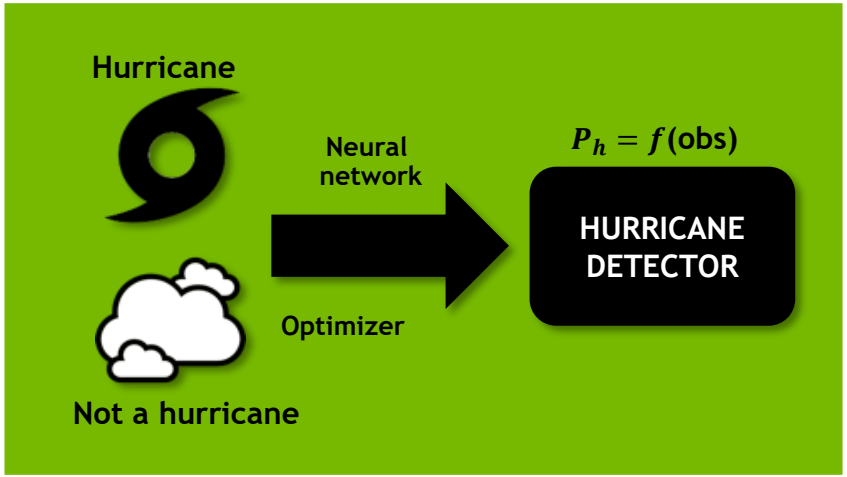
recognition/classification -> recursion/time series -> generative

Software, by example

Deep Learning builds functions from examples of desired behavior



Functions are the building blocks of software. DL can approximate any function.

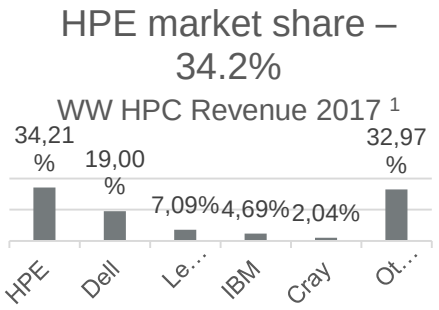


Some functions are too complex to code by hand. Generate complex functions by example.

Mix freely with conventional software and algorithms

HPE & HPC

HPE is the market leader in HPC



HPE's #1 share is 15 pts > #2 Dell

HYPERION RESEARCH

"HPE's next-generation and new Apollo systems will facilitate that adoption by providing easier integration and management while delivering extreme density to reduce data center footprint and extend the range of HPC and AI use cases."

Steve Conway,
SVP of Research

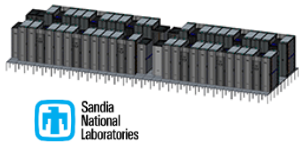
Creating value for customers

ROLLS ROYCE · Zentralanstalt für Meteorologie und Geodynamik ZAMG · PSC Pittsburgh Supercomputing Center · NREL National Renewable Energy Laboratory · Tokyo Institute of Technology · Notre Dame · NCAR National Center for Atmospheric Research · Sandia National Laboratories

HPE helps EPFL Blue Brain Project unlock the secrets of the brain

Blue Brain Project

World's largest Arm Supercomputer for U.S. Department of Energy[2]



Best HPC Server – 2017[3]
HPE Apollo 6000 Gen10

Best HPC Server Product or Technology
HPC wire
Readers' Choice Awards 2017

Fastest parallel processing performance[4]

HPE SGI 8600

HPE GreenLake – best-in-class[5] consumption model

[1-5] See speakers notes for substantiation.

HPE continues to be the market leader in HPC

MAY 17, 2019 • PRESS RELEASE:

HPE to acquire supercomputing leader Cray

https://www.hpe.com/us/en/newsroom/press-release/2019/05/hpe-to-acquire-supercomputing-leader-cray.html

CRAY RESEARCH
A Silicon Graphics Company

sgi

CRAY

Hewlett Packard Enterprise

Top100: 15 HPE + 26 Cray = 41% (2019)

HPE purpose-built portfolio for HPC

Advisory, Professional, Operational Services | HPE GreenLake Flex Capacity | Hybrid HPC

Supercomputing / Enterprise / Commercial HPC

HPE SGI 8600

Liquid cooled, Petaflop scale for HPC and AI

HPE Apollo 6000 Gen10

Air-cooled, HPC at rack scale

HPE Apollo 2000 Gen10

The bridge to enterprise scale-out architecture

HPE Apollo sx40

Max GPU capacity and performance with lower TCO

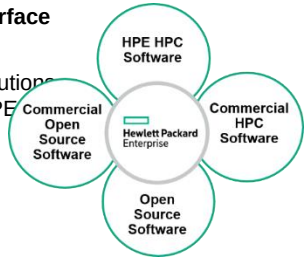
Adaptive Cooling

HPE Adaptive Rack Cooling System

Higher power density with less datacenter heat

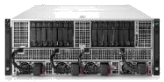
HPC Software

- **HPE Performance Cluster Manager**
- **HPE Message Passing Interface (MPI)**
- **3rd party** solutions offered by HPE



Emerging HPC

HPE Apollo 6500 Gen10



Enterprise platform for accelerated computing

HPE Apollo 70

HPC cluster ready Arm Based server

HPE Apollo 35

Best AMD Performance in Dense HPC Platform

In-memory HPC

HPE Superdome Flex Server

Scale-up, shared memory HPC, combines best of HPE and SGI technologies

HPC Storage Solutions

- HPE Data Management Framework
- Scalable Storage for Lustre
- WekaIO Matrix

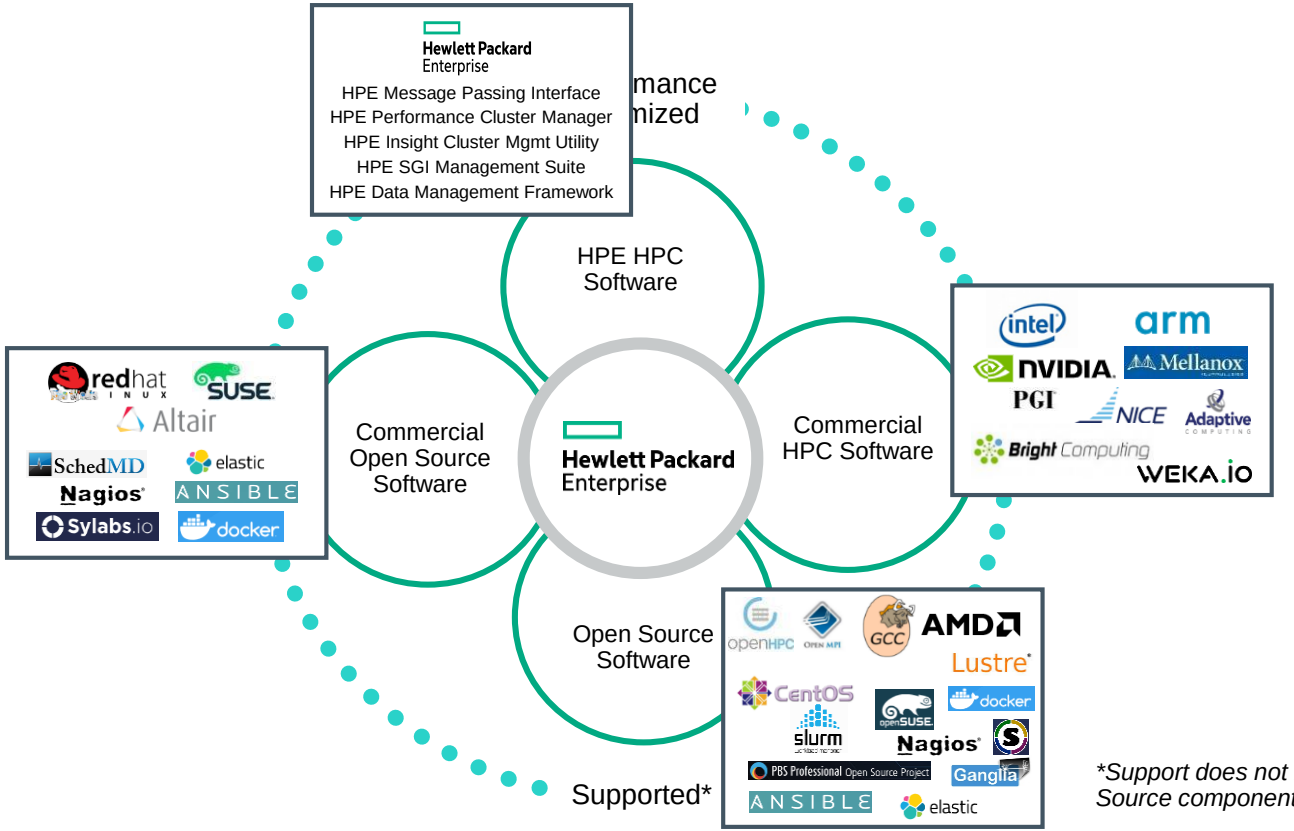
HPE Apollo, ProLiant and JBODs

Networking for HPC

- Intel® Omni-Path Architecture
- Mellanox InfiniBand
- HPE FlexFabric Network

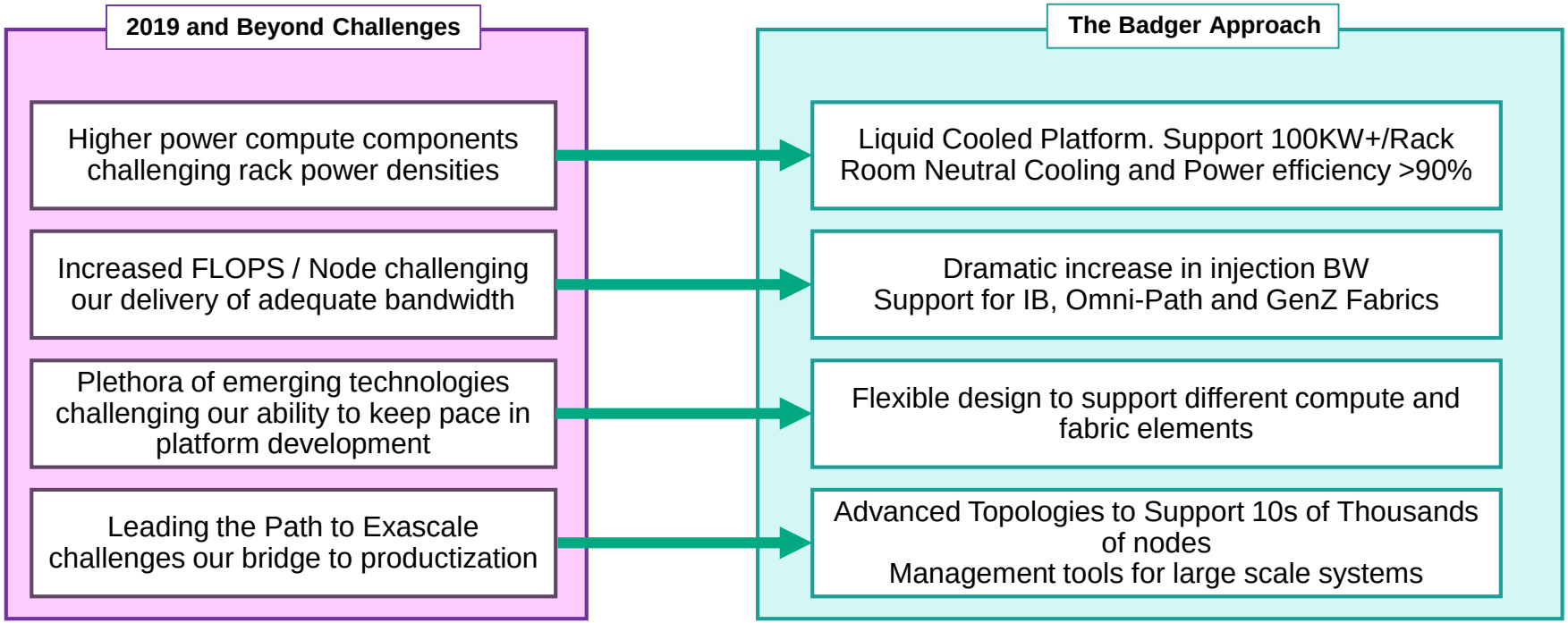
High Performance Computing Software

Building the No. 1 HPC software environment



**Support does not include Open Source components*

Project Badger Approach – A Flexible Platform

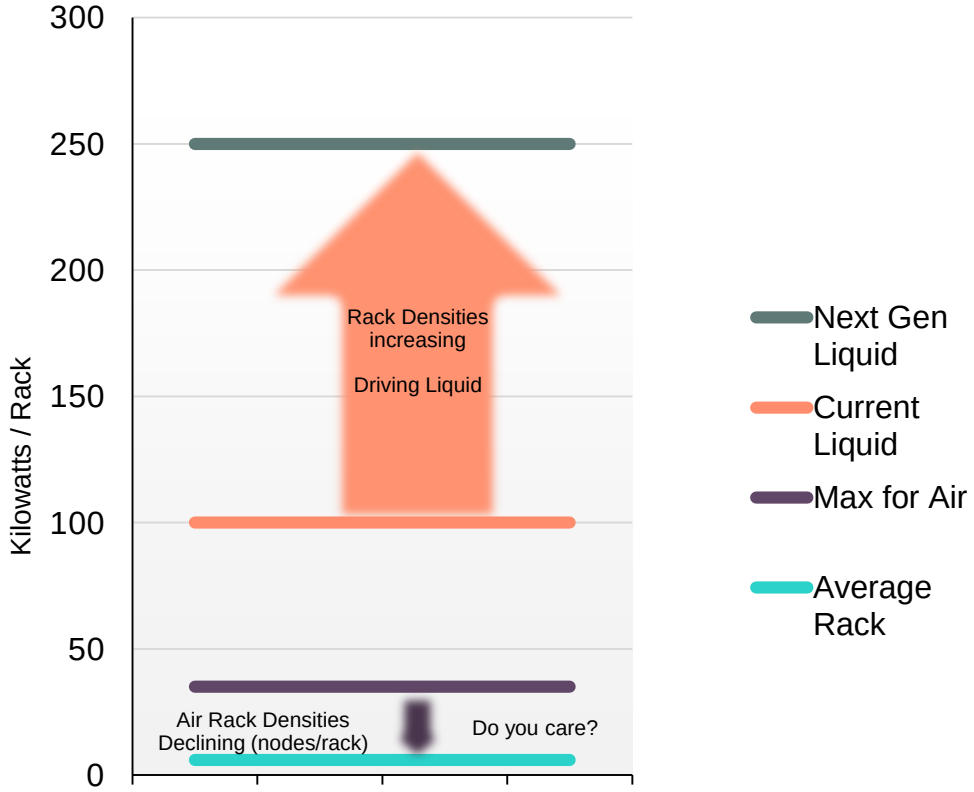


Badger Platform – Why?

Power Density is on the Rise

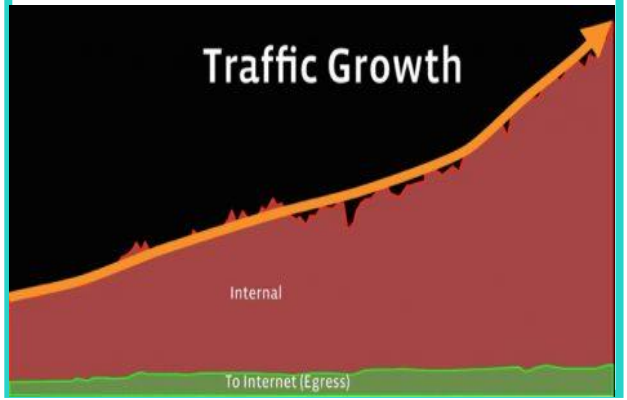
Watts / CPU
250W+

Watts / GPU
400W+

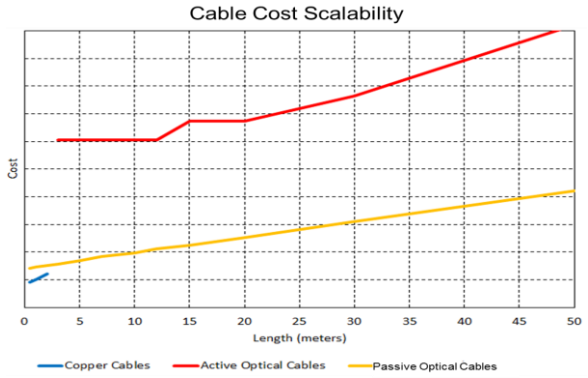


IO Traffic is on the Rise

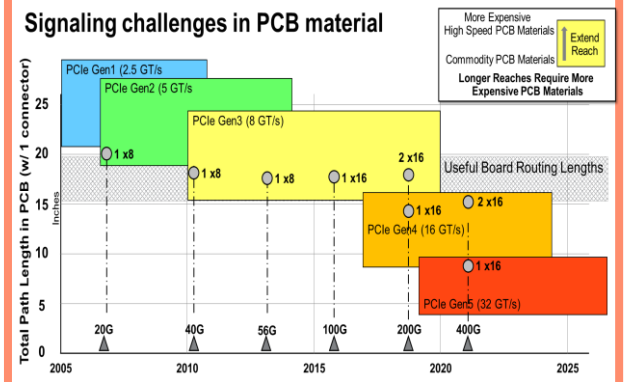
East/West Traffic Is increasing



Connectivity Costs and Complexity increasing



Generational Costs are increasing



Project Badger – HPE Platform for Leadership in HPC

Badger is a production supercomputer, carefully designed with the purpose of providing the users top application performance per dollar invested.

Protect our customer investment and prepare for upcoming technologies

A common infrastructure with choice of processing elements

Outstanding Power Management

Outstanding operational cost / allow effective system management

Flexible, scalable, efficient fabric and topologies

Keep HPE philosophy for HPC

- Badger is a GenZ 'ready' infrastructure design
- Up to 8 planes of extremely high BW fabric w/o burdening the baseline offering
- Design prepared for Optical Interconnect
- Fanless at commodity costs will be a game changer
- Exascale Cluster Management SW
- Reliability improvement with DLC
- Factory Integration and Testing
- Leverage our exascale efforts in other leadership and commercial areas (eg. Optical, GenZ)
- Committed to open and/or commercially available technologies (IB, OPA, GenZ)

Badger is the most flexible, efficient infrastructure. Is our customer choice the right processor, fabric, topology, injection bandwidth to optimize for their budget/workflow and we will support them navigating that path

How we are building it?

Leverage and advance!!!

SGI 8600

- SGI Acquisition - Flagship HPC Product
- Gaining share in TOP 500 and productivity HPC

New Challenges

- Midplane & fabric speeds
- Memory Bandwidth in new processors
- Next Gen PCI Speeds
- Processor Power Footprint

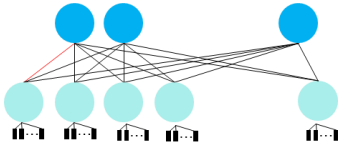
Badger Leverages from SGI 8600 – Proven, solid infrastructure

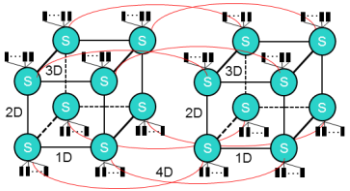
- Architectural Approach
- Liquid Cooling Technology
- Management and Control Infrastructure (HPCM)

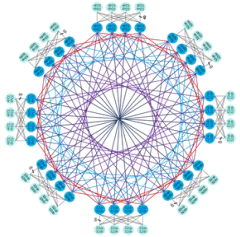
And makes significant advances for the next 4 years:

- De-couple fabric-infrastructure (significant challenge for years!)
- Prepared for next generation optical interconnects
- Support for HDR and NDR fabric speeds, ready for GenZ and Optical
- Increase from 6 to 8+ memory channels/socket and prepare for SOCs&HBM
- Design for PCI Gen4 and PCI Gen5 distance constraints
- Prepare 300W CPU and 500W GPU power footprints and 54V power distributions.
- Driving toward 100% liquid cooled during the product lifecycle

HPE support topology options to match your workloads







Fat-tree

- General purpose fabric
- Relative consistent hop-count and bandwidth resulting in more predictable job performance
- Cost does not scale linearly
  - switch/cable becomes increasingly expensive with cluster sizes that require more than 2-level Fat-tree

Enhanced Hypercube

- Best for appellations with spatial locality
- Low cost, and scales linearly to extremely large systems
- Live integration/extension
- Best captured by aggregate bisection bandwidth of all dimensions

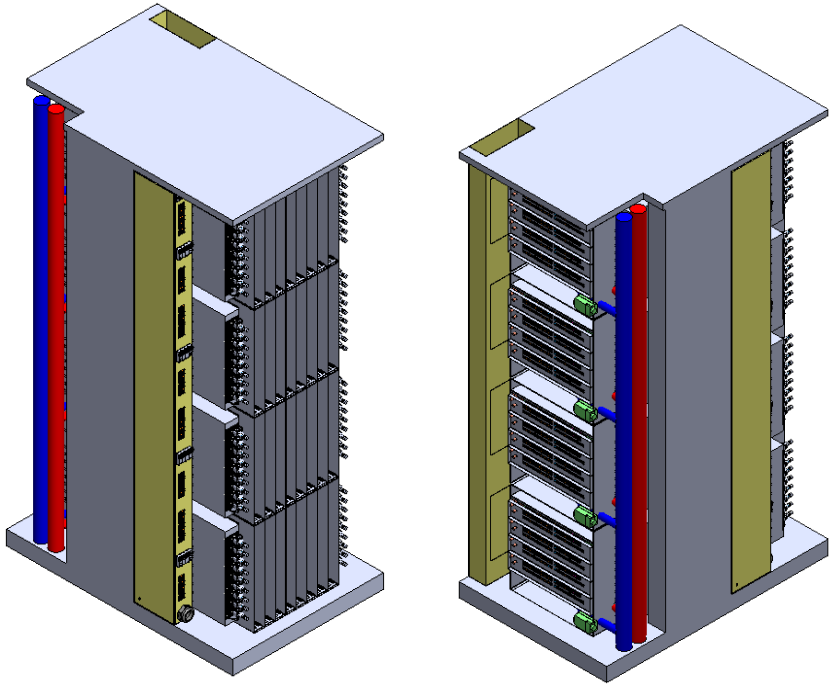
Enhanced Dragonfly (under investigation)

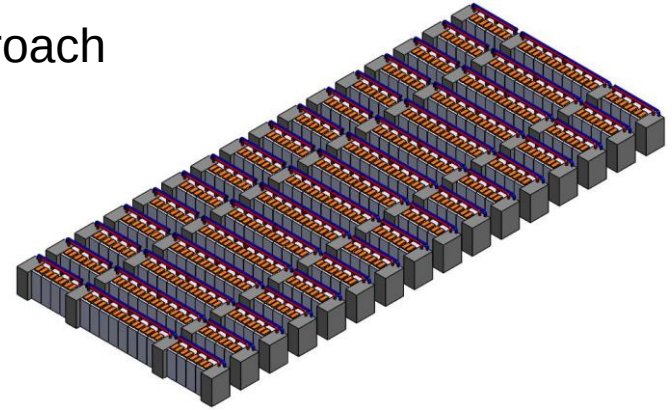
- Group: Fat-tree topology for servers in the same group
- System: Multi-dimensional all-to-all among groups
- Consistent application performance with minimum inter-job interference
- Cost scales linearly for large systems

Badger Infrastructure Overview

The Future of Leadership HPC – Project Badger

Targeting Exascale – With a Balanced System Approach

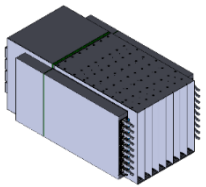




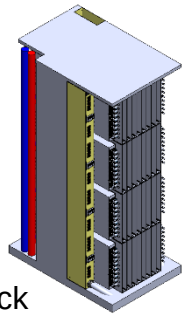
Badger is intended to provide HPE users with a highly efficient, integrated platform capable of delivering leadership performance, energy efficiency, and price/performance at the multi-rack scale systems level.

Badger Infrastructure – Processor and Fabric Agnostic

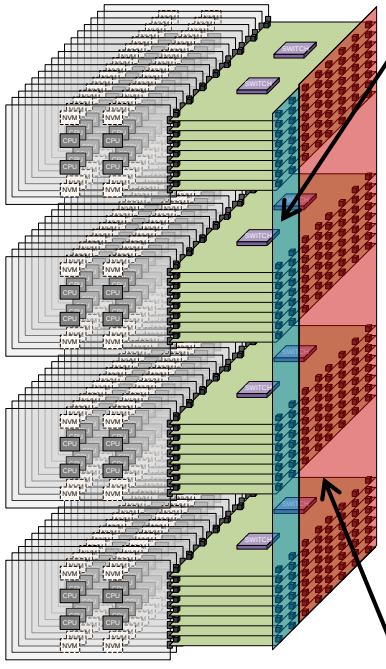
- High-Density and High-Power
  - 100kW to 250kW rack powers over time
  - 256 CPUs and 64 fabric switches per rack typical
    - (4-ports injection BW per CPU)
  - 256 CPUs and 128 fabric switches per rack enabled
    - (8-ports injection BW per CPU)
  - 256 GPUs and 64 CPUs enabled per rack
- Driving toward 100% liquid-cooled
  - Room neutral and fanless cooling solution

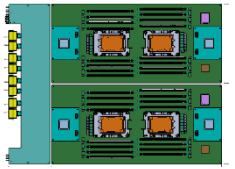


Chassis

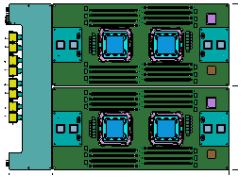


Rack

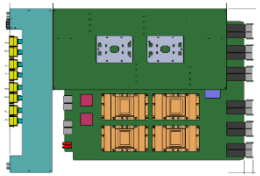




x86 CPUs



ARM CPUs



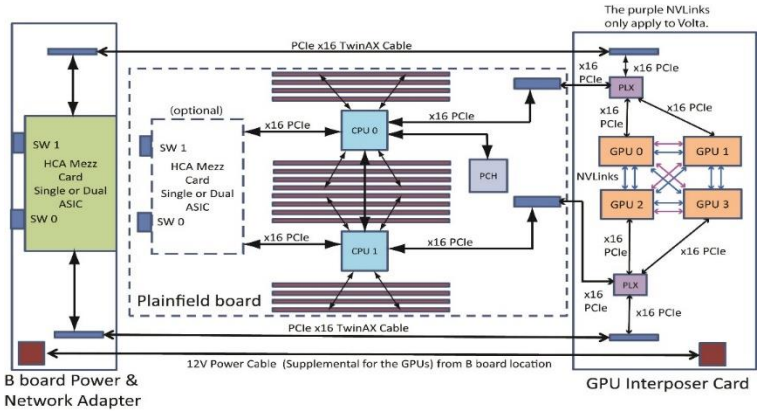
x86 CPUs and GPUs

GPUs in Badger – Build on proven path

Road to Exascale systems is moving in the >>10TF/node direction with one track leading to 2CPUs : 8GPUs
These nodes have power requirements in the order of 3KW/node.

Leadership AI systems (e.g. Tsubame 3.0) will require very high BW rates – up to 1 Omni-Path port per GPU

Badger provides the infrastructure to Power/Cool/Interconnect those nodes at scale.

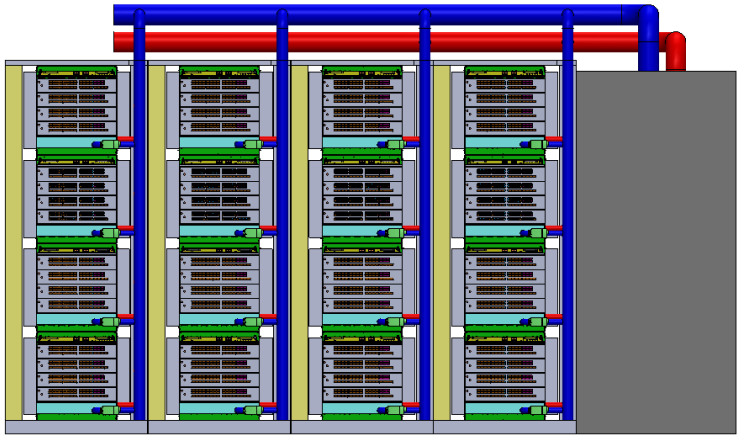


HPE SGI 8600 in production today:
Tsubame 3 - 2160 GPUs

Badger Cooling Equipment Concepts

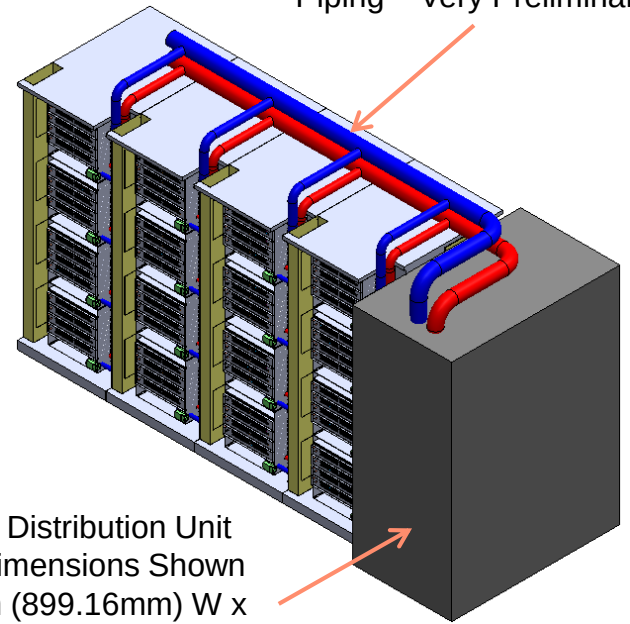
94.70 inch (2405.38mm) H
finished at data center

82.46 inch
(2094.52mm) H
shippable



13.42 Ft (4089.4mm)
Overall Width

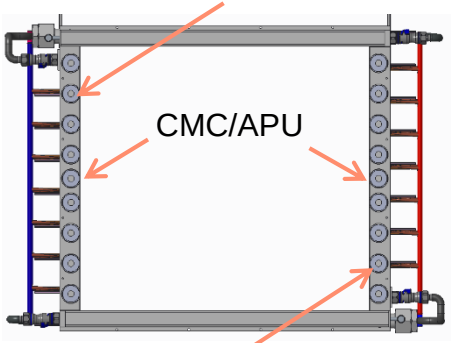
CDU Secondary-Side
Piping – Very Preliminary



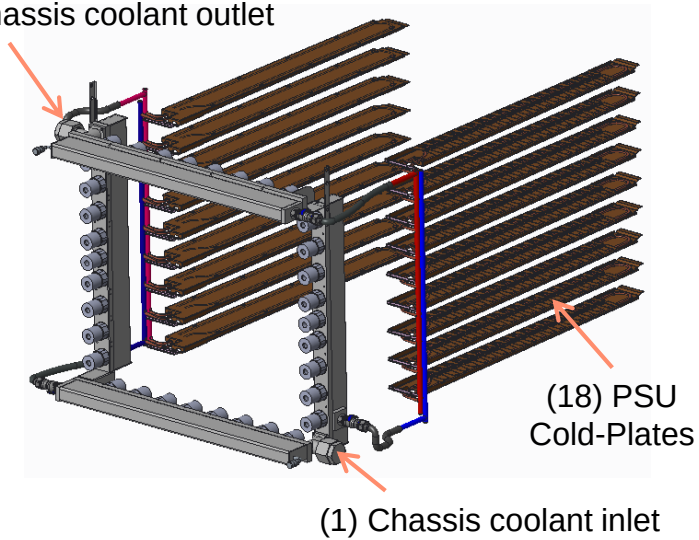
Coolant Distribution Unit
(CDU) Dimensions Shown
35.4 inch (899.16mm) W x
60.0 inch (1524.00mm) D x
80.0 inch (2032.00mm) H

Support top or bottom feed of either water or power

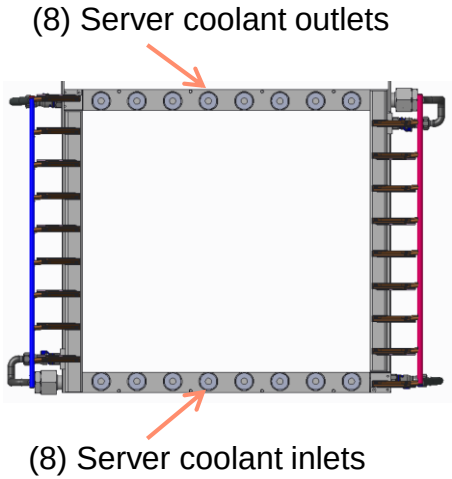
Badger Chassis Coolant Distribution



Rear View



Rear Isometric



Front View

… and what about in Europe? - EuroHPC

https://eurohpc-ju.europa.eu/index.html

- The goal
  - deploying in Europe a world-class supercomputing infrastructure and a competitive innovation ecosystem in supercomputing technologies, applications and skills
  - to develop and maintain in the EU a world-class High-Performance Computing ecosystem, including
    - low-power processor and middleware technologies,
    - algorithms and code design,
    - applications and systems,
    - services and engineering,
    - interconnections,
    - know-how and skills for the next generation supercomputing era
  - two pre-exascale machines and several petascale systems by 2020 for Europe's scientific, industrial and public users
- The budget
  - the initial co-investment with Member States of about EUR 1 billion
  - and EUR 2.7 billion later

Thank you

tibor@silicon.hu