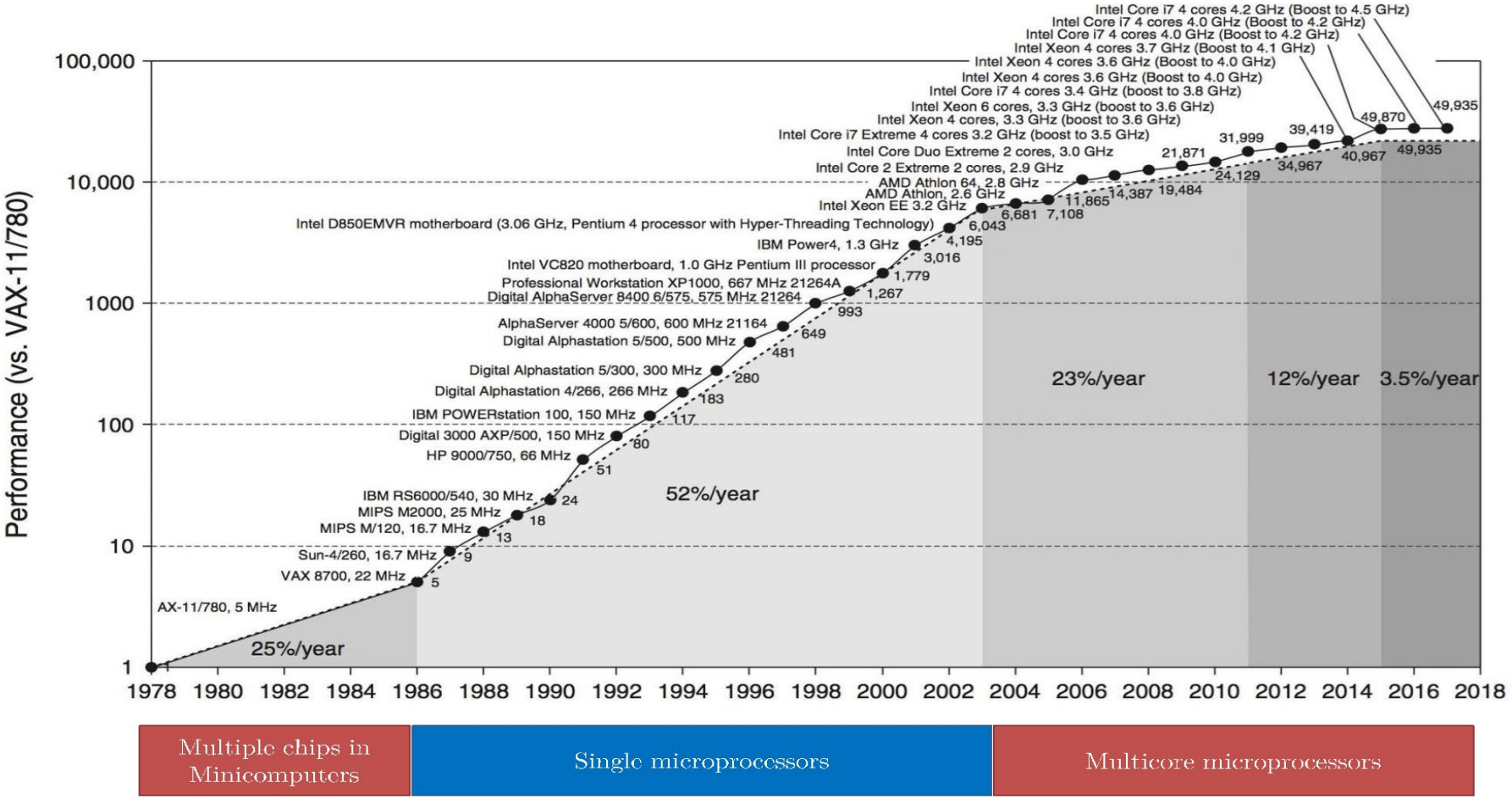
January 24 - 26, 2023
DoubleTree by Hilton San Jose
ChipletSummit.com

Chiplets for HPC

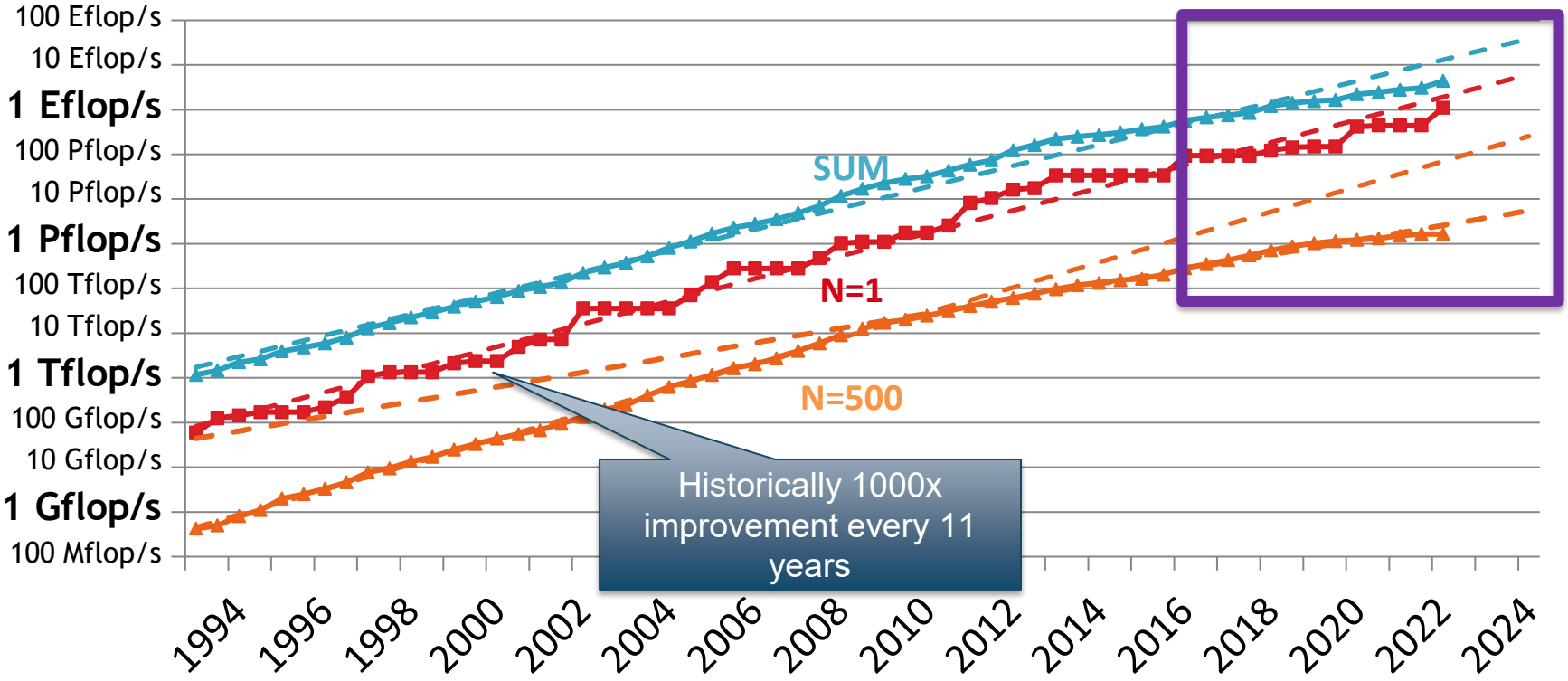
John Shalf
Department Head for Computer Science
Lawrence Berkeley National Laboratory

Moore's Law is Ending *(really it is!)*

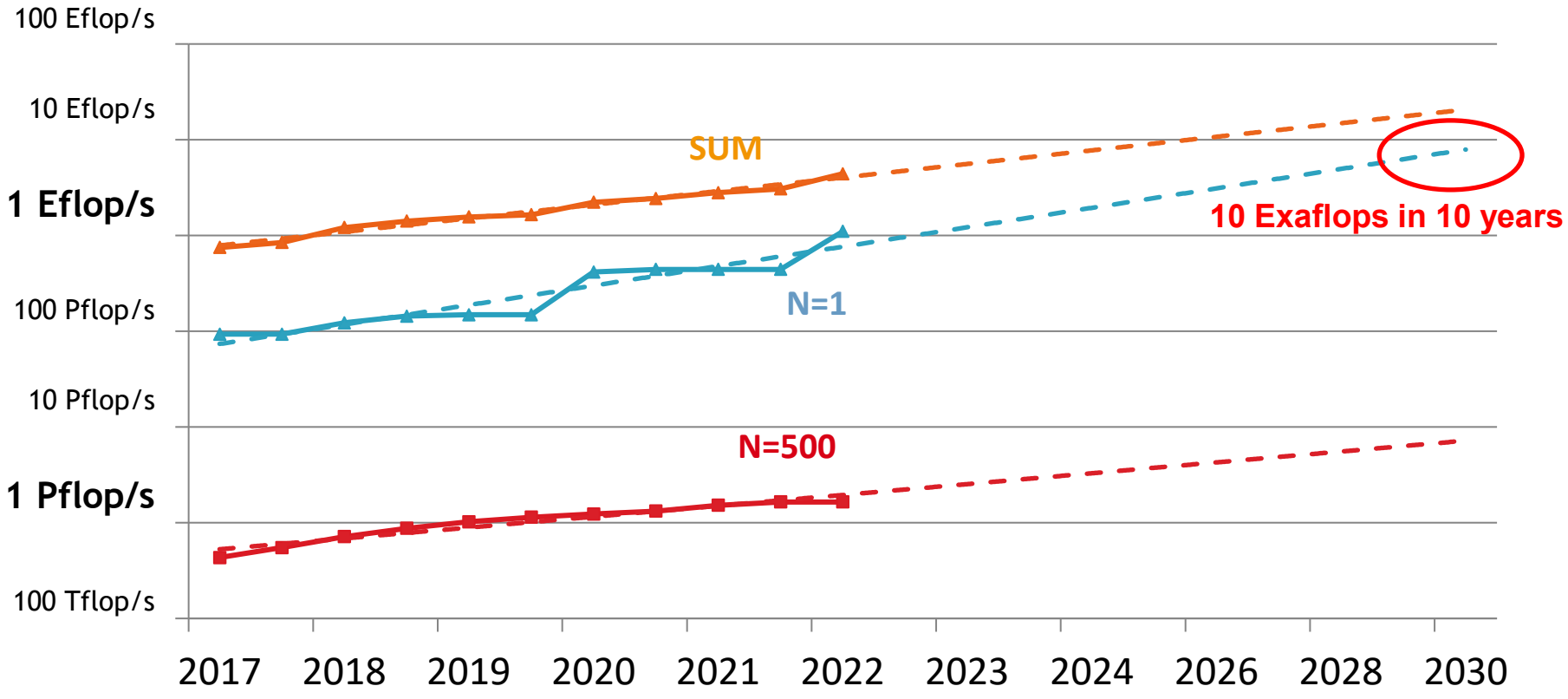
Hennessy / Patterson



Projected Performance Development



Projected Performance Development



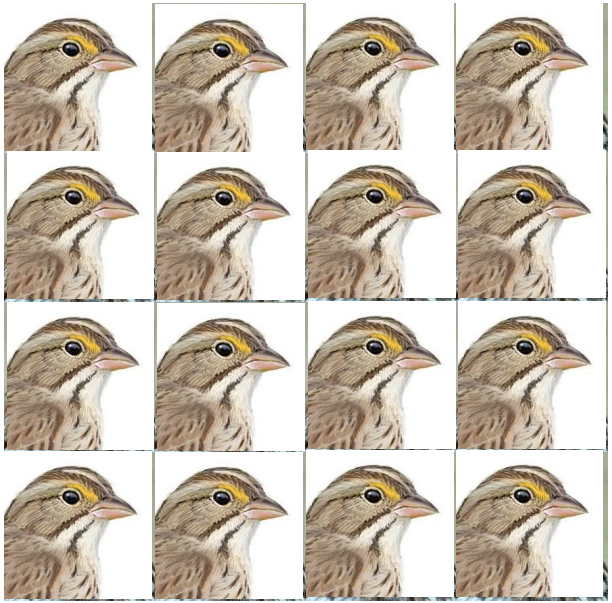
Specialization:

Natures way of Extracting More Performance in Resource Limited Environment

Powerful General Purpose

Xeon, Power

Many Lighter Weight
(post-Dennard scarcity)

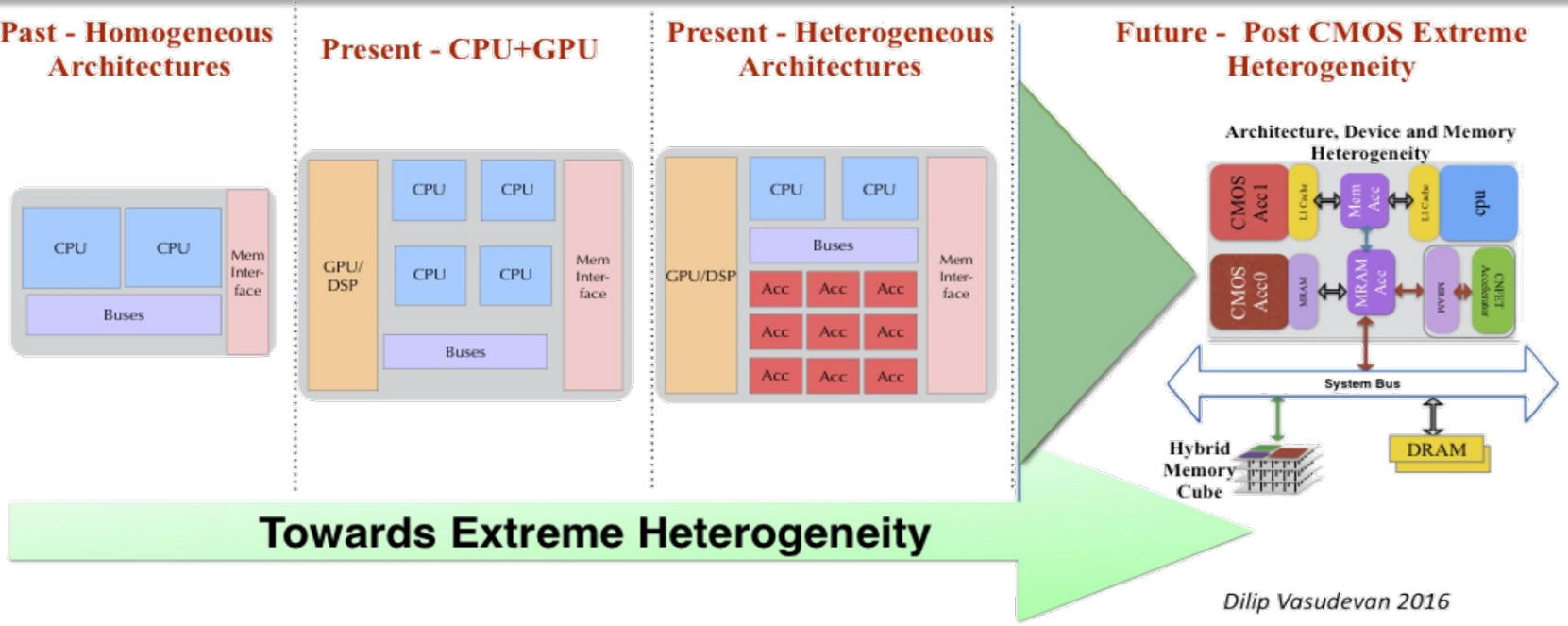


KNL, AMD, Cavium/Marvell, GPU

Many Different Specialized
(Post-Moore Scarcity)

Apple, Google, Amazon

The Future Direction for Post-Exascale Computing



But what are the right specializations to include?

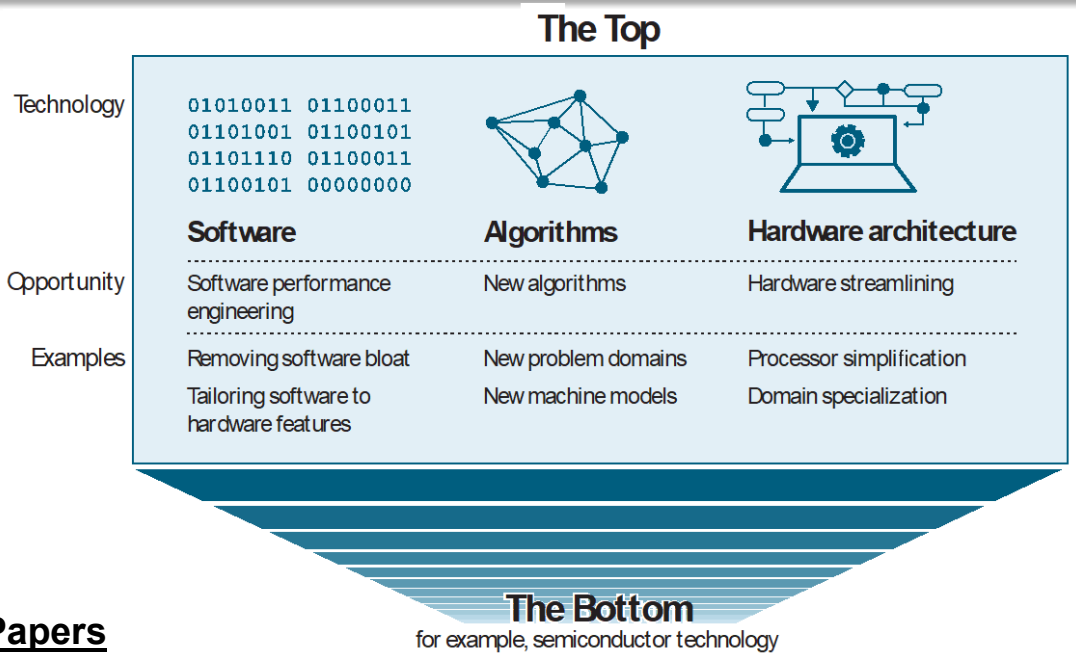
What is the cost model (we know we cannot afford to spin our own chips from scratch)

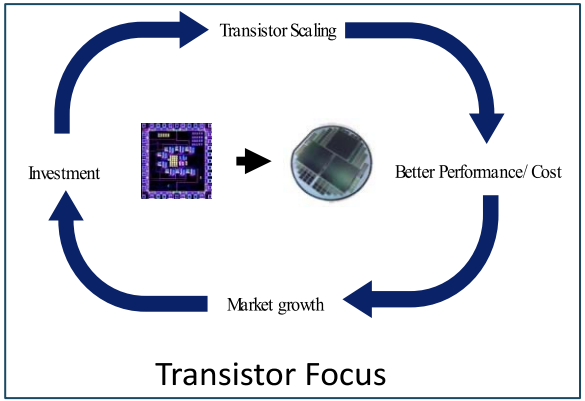
The ARM licensable IP ecosystem : *IP is the commodity (not the chip)*

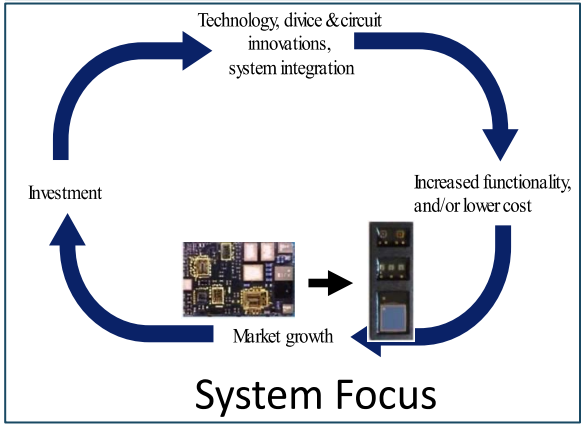
What is the right partnership/economic model for the future of HPC?

Neil Thompson: Economics of Post-Moore Electronics

http://neil-t.com, MIT CSAIL, MIT Sloan School





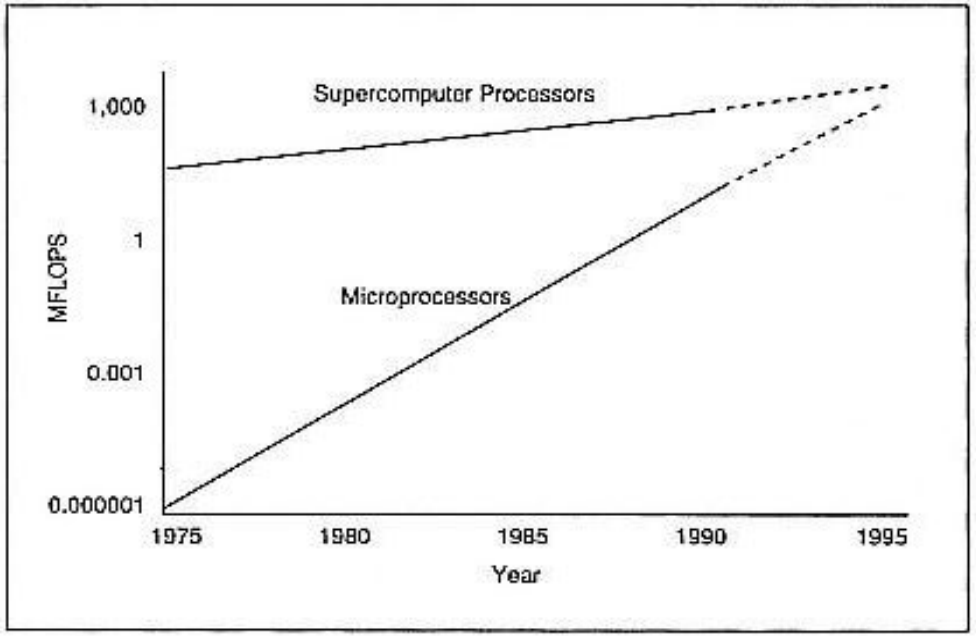


Papers

1. The Economic Impact of Moore's Law
2. There's Plenty of Room at the Top: What will drive computer performance after Moore's Law?
3. The Decline of Computers as a General Purpose Technology

Attack of the Killer Micros 90's

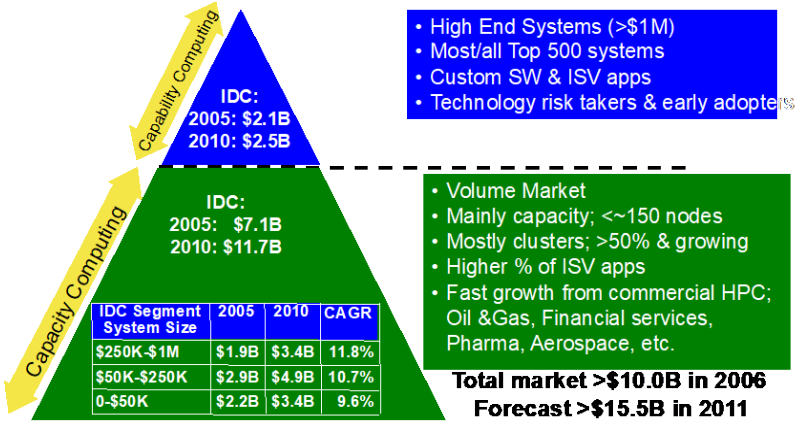
Current Economic Model for COTS HPC



Attack of the killer micros

John Markoff, May 6, 1991

- The move to COST was more about the economic model than technology alone

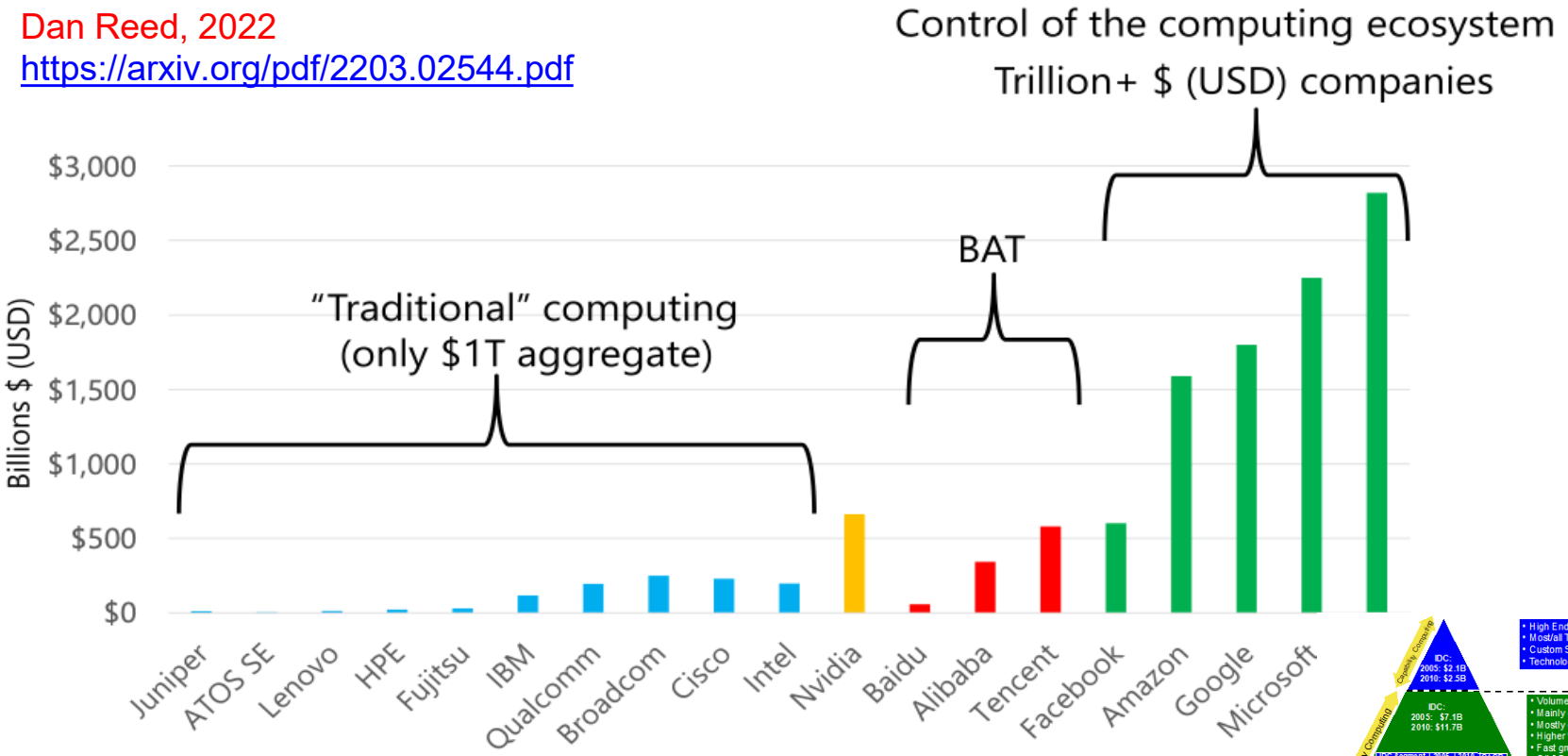


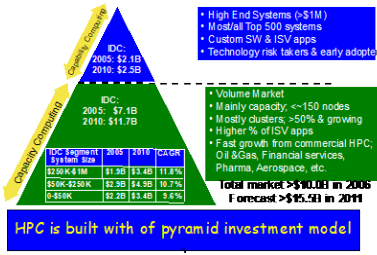
IDC Segment System Size	2005	2010	CAGR
$250K-$1M	$1.9B	$3.4B	11.8%
$50K-$250K	$2.9B	$4.9B	10.7%
0-$50K	$2.2B	$3.4B	9.6%

HPC is built with of pyramid investment model

It is not good enough anymore to understand the Technology Now we must also understand the market context

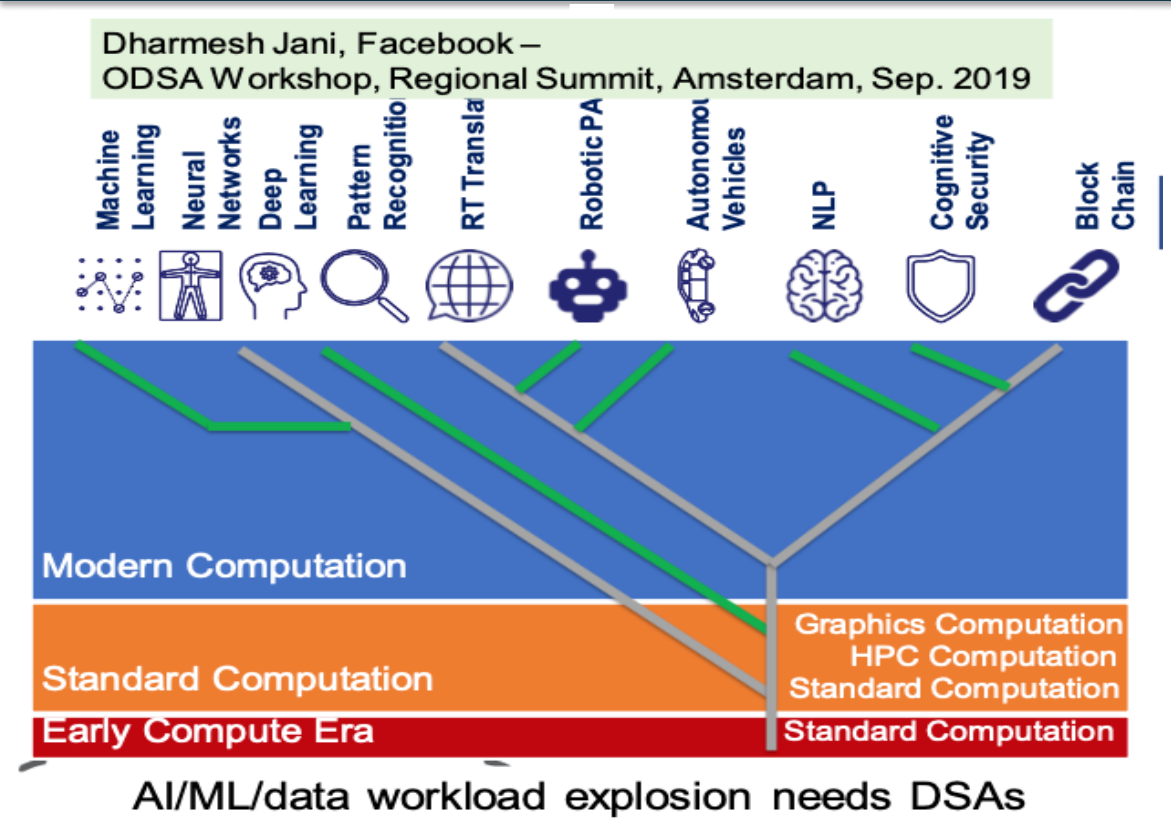
Dan Reed, 2022
https://arxiv.org/pdf/2203.02544.pdf

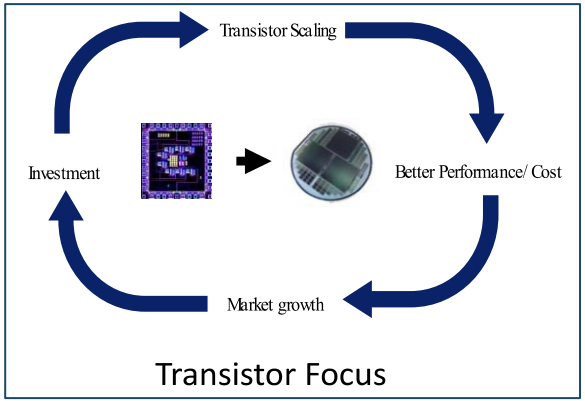


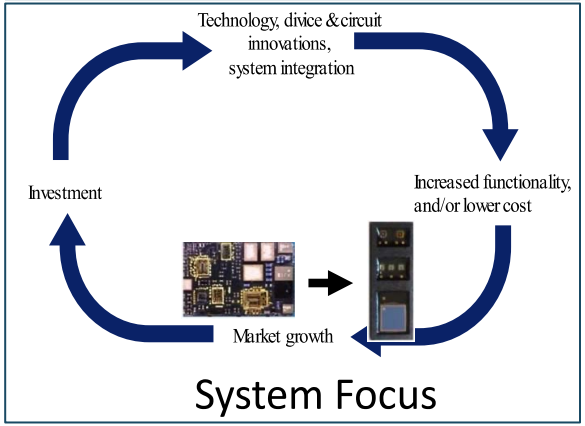


Why? Domain specific Architectures driven by hyperscalers

in response to slowing of Moore's Law (switch to systems focus for future scaling)







Opportunity for HPC: New Economic Model

Open Chiplets Marketplace is forming (ODSA and UCIexpress)

- Licensable IP and assembly by 3rd party lowers that barrier
- Leverage the economic model being created by HyperScale

Leverage this baseline and extend to support HPC

- Smaller incremental cost for HPC to "play"
- *HPC has become "too small to attack the city"*

80:20 Rule: Focus open efforts on what uniquely benefits HPC

- Build up a library of reusable accelerators for HPC.
- Interoperability for sustainability: *Interoperate with Arm IP for commercially supported IP where it exists and focus Open on the 20% that doesn't make commercial sense to license*

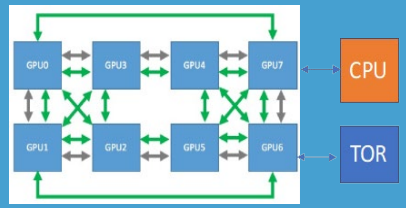
Opportunities for CoPackaged Optics (photonics)

A primer on Resource Disaggregation

Diverse Node Configurations for Diverse Workload Resource Requirements

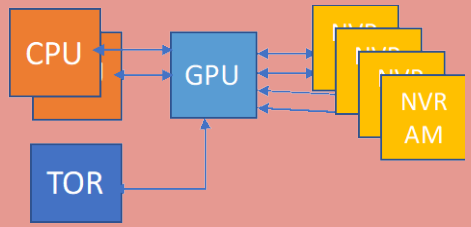
Training

- 8 connections: GPU
- 8 links to HBM (weights)
- 8 links: to NVRAM
- 1 links: to CPU (control)



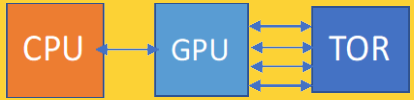
Data Mining

- 6-links: HBM
- 15 links: NVRAM (capacity)
- 4 links: CPU (branchy code)



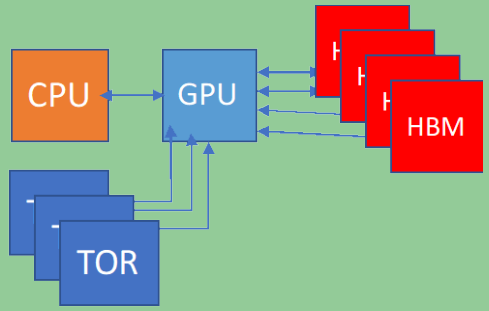
Inference

- 16 links to TOR (streaming data)
- 8 links HBM (weights)
- 1 link: CPU

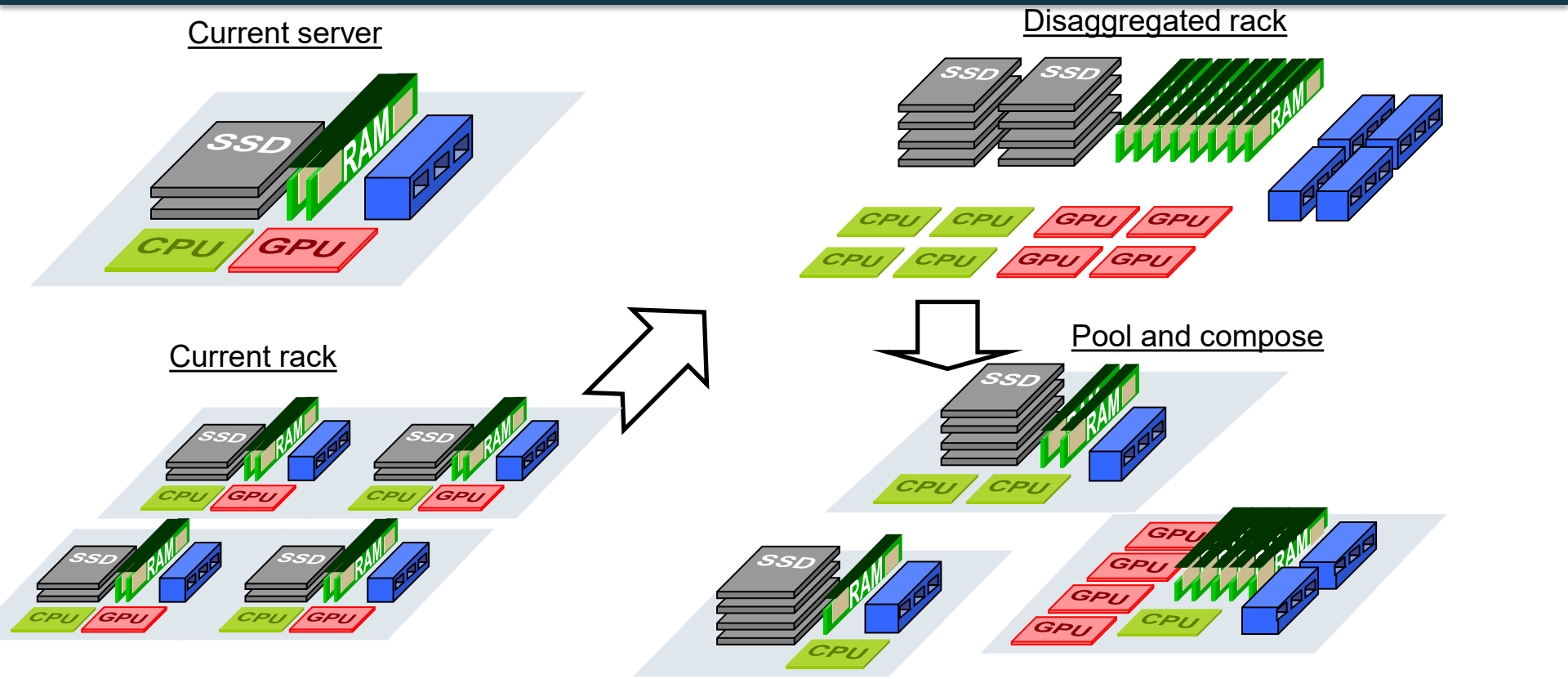


Graph Analytics

- 16 links HBM
- 8 links TOR
- 1 Link CPU

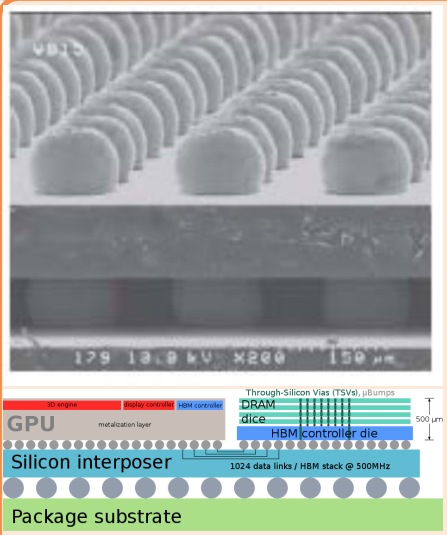


Disaggregated Node/Rack Architecture



Most solutions current disaggregation solutions use Interconnect bandwidth (1 – 10 GB/s)
But this is significantly inferior to RAM bandwidth (100 GB/s – 1 TB/s)

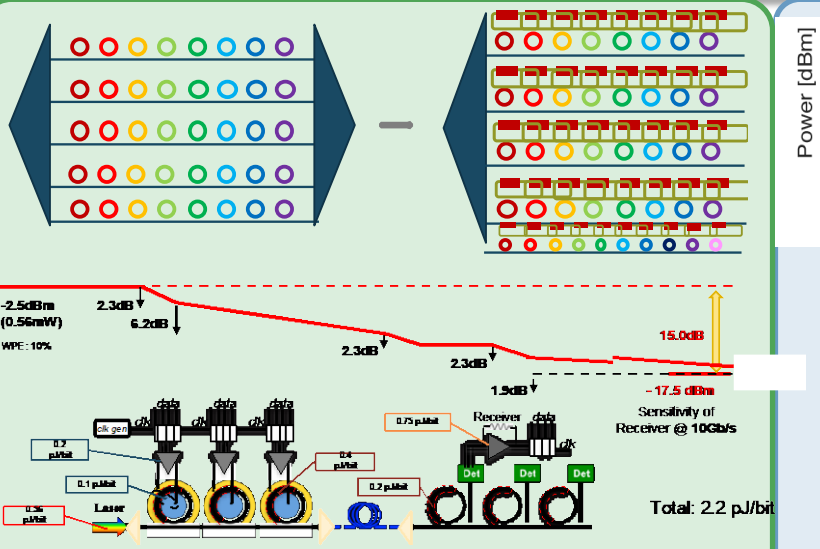
Impedance Matching to Packaging Technology



In-package integration

Solder Microbumps
& Copper Pillars@~10Gbps

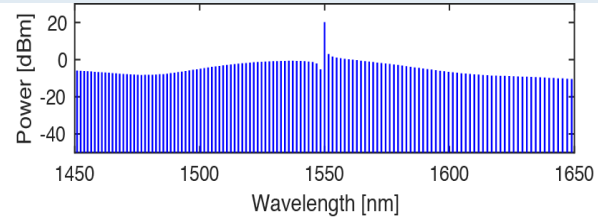
Wide and Slow!

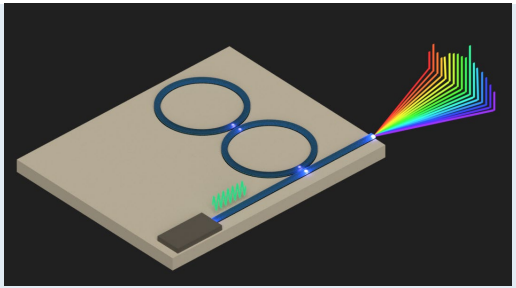


DWDM Using Silicon Photonics

Ring Resonators @ ~10-25 Gb/sec per chan
Many channels to get bandwidth density

Wide and Slow!



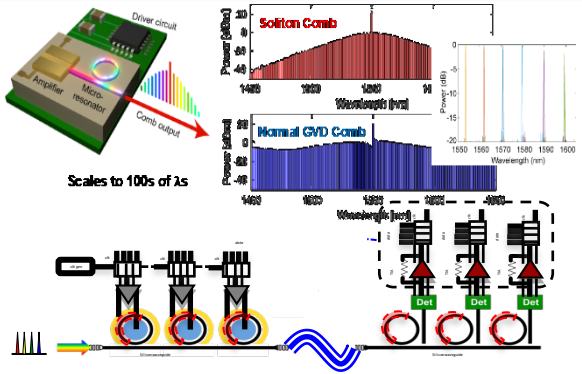


Comb Laser Sources

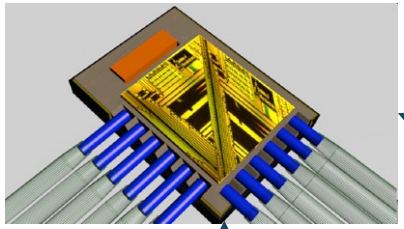
Single laser to efficiently
generate 100s of frequencies

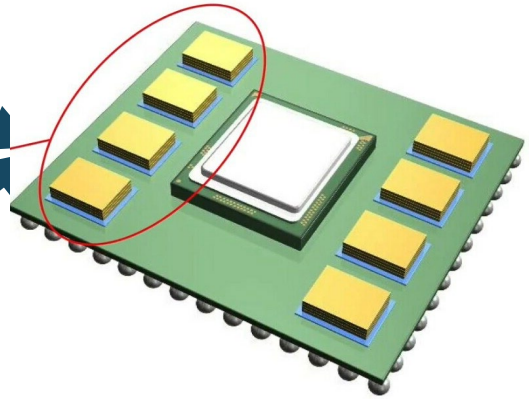
Wide and Slow!

Photonic MCM (Co-Packaged Optics)

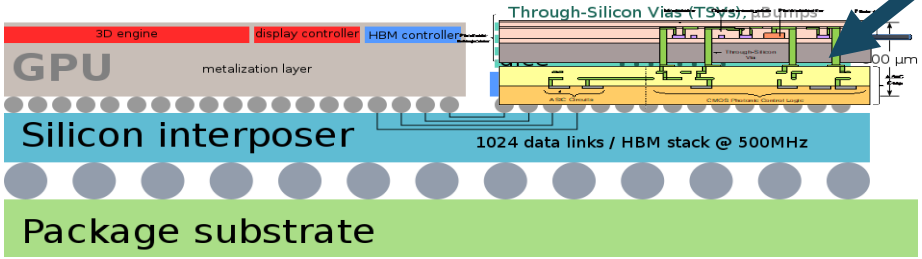


Comb Laser Source with
DWDM Silicon Photonics
Wide-and Slow for high speed links



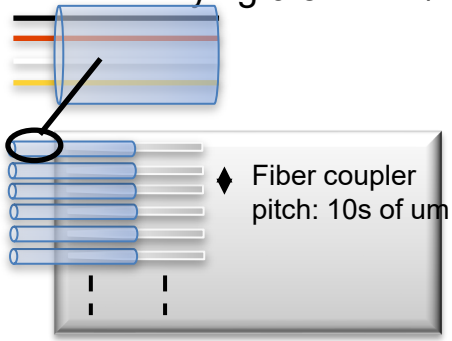


Photonic SiP

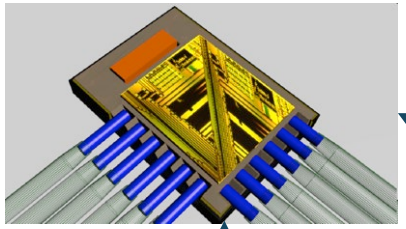


Photonic MCM (Co-Packaged Optics)

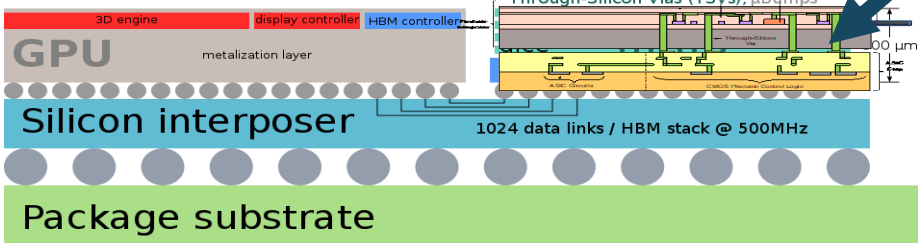
Fiber carrying 0.5 - 1 Tb/s

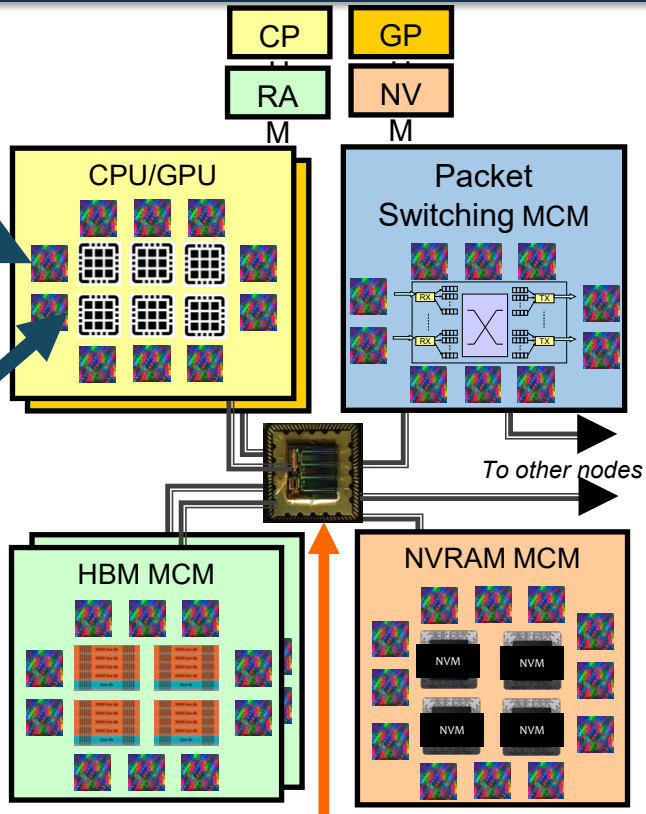


High-Density **fiber coupling array** with 24 fibers = 6-12 Tb/s bi-directional = **0.75 – 1.5 TB/s**



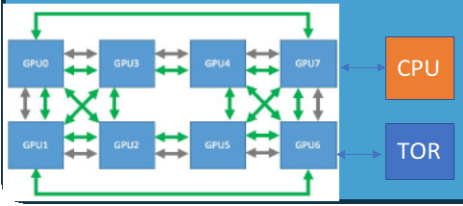
Photonic SiP





Training

- 8 connections: Peer GPU
- 8 links to HBM (weights)
- 8 links: to NVRAM
- 1 links: to CPU (control)

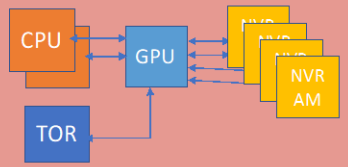


Inference

- 16 links to TOR (streaming data)
- 8 links HBM (weights)
- 1 link: CPU

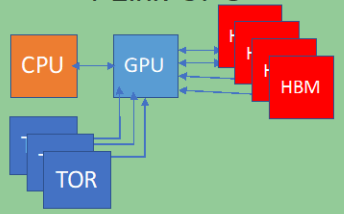
Data Mining

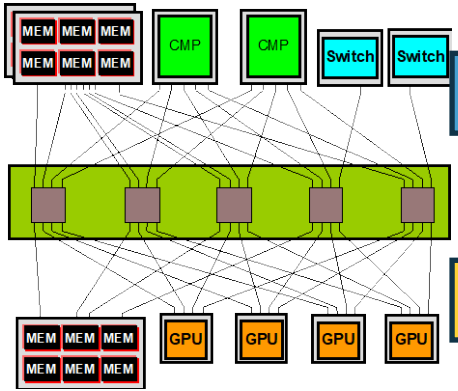
- 6-links: HBM
- 15 links: NVRAM (capacity)
- 4 links: CPU (branchy code)



Graph Analytics

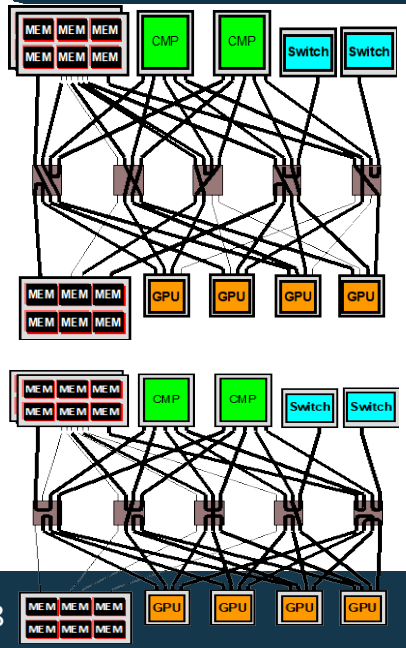
- 16 links HBM
- 8 links TOR
- 1 Link CPU

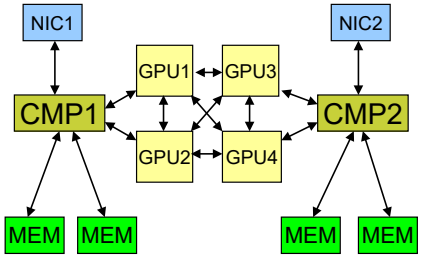


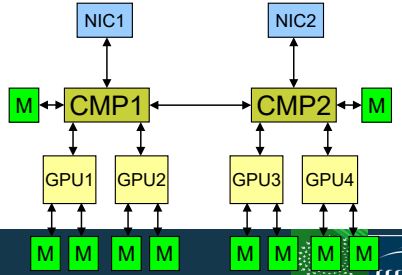


Configure for Training

Configure for Inference







Conclusions

- **Scaling alone is no longer a rational metric for HPC success**
  - After the “Exaflop” there will be no ”Zettaflop” supercomputer
  - We need a different metric for success (more tied to scientific benefit!)
- **Think more seriously about how to use specialization productively for science**
  - Requires deep understanding of applied mathematics and the underlying algorithms to be successful *(chiplets is a way to get there)*
- **Reevaluate the economic model for the design/acquisition of HPC systems**
  - Chiplets enable us to be aligned again with broader industry trends!

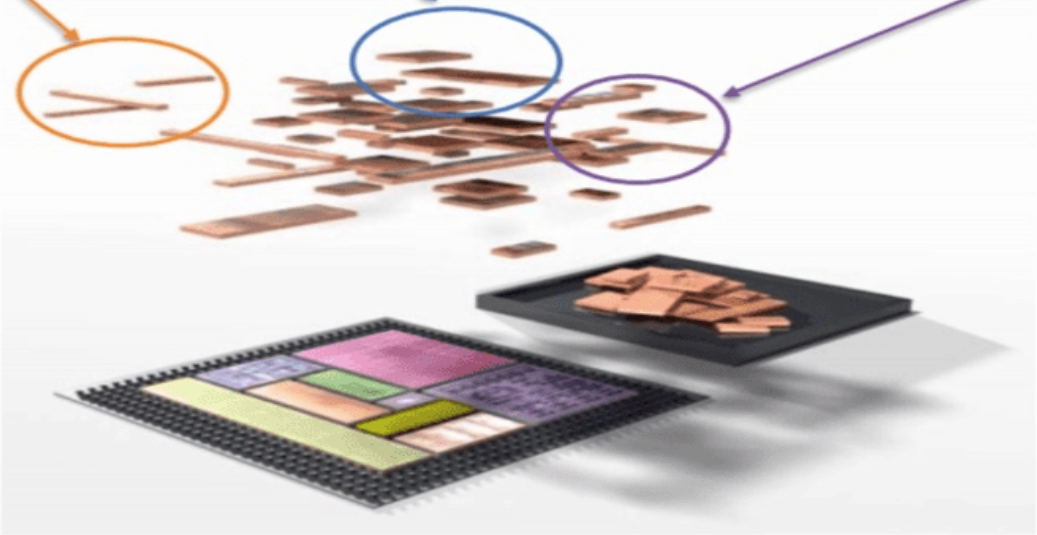
- End

How do chiplets enable domain specialization?

Reusable function blocks
- QR decomposition
- Waveforms
- FFT

Access to Commercial IP
- Memory
- SerDes
- Processors

Big Data Movement
- Image processing
- Machine Learning
- High-speed chiplet networks



CHIPS modularity targets the enabling of a wide range of custom solutions

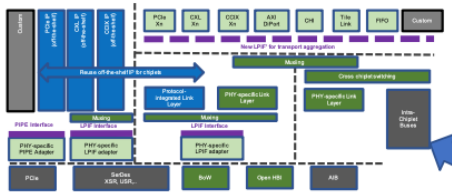
What is Hyperscale Datacenter Strategy

Interconnect on-Chip

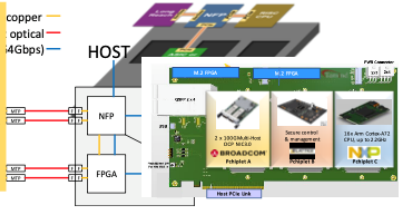
ODSA: Open Domain Specific Architecture

Creating an Open Chiplet Marketplace for Hyperscale Datacenters

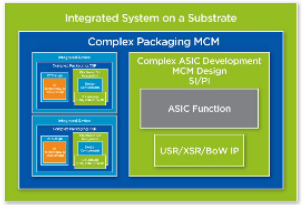
Open D2D Interface
Reduce barrier to interoperation



Reference Designs
Starting point for new designs



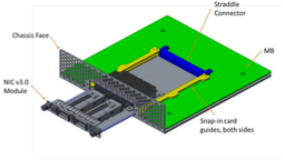
Reference Workflows
Reusable, open practices

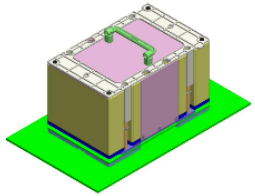


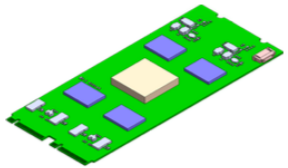
ODSA Activities

Chiplet Marketplace

Integrate best-in-class chiplets from multiple vendors through open interfaces







OCP modular form factors

Co-Sponsors of the ODSA Open Chiplets Marketplace

Semi Vendors
IP providers, EDA
Service providers

Tools, Manufacture, Design,
Test,
Integration

Systems vendors,
End users, ISVs, Service
Providers

Chiplet Bandwidth Roadmap (5 generations of BW doubling)

Table 5: Physical IO Scaling Roadmap for 2D and Enhanced-2D Architectures that use both solder and hybrid interconnects.

Generation Number →		1	2	3	4	5
Raw Linear Bandwidth Density (GBps/mm)		125	250	500	1000	2000
Package Technology	Minimum Bump Pitch (µm)[17]	55	40	30	20	10
	Linear Escape Density (IO/mm)	500	667	1000	2000	4000
	Areal Escape Density (IO/mm^2)	331	625	1111	2500	10000
Signaling Speed (Gbps)		2	3	4	4	4

5.1.2 Area Interconnects for 3D Architectures (see Figure 1)

Table 6: Physical IO Scaling Roadmap for 3D architectures that use both solder and hybrid interconnects.

Generation Number →		1	2	3	4	5
Raw Areal Bandwidth Density (GBps/mm^2)[18]		125	250	500	1000	2000
Package Technology	Minimum Bump Pitch (µm)[19]	40	30	20	15	10
	Areal Escape Density (IO/mm^2)	625	1111	2500	4444	10000
Signaling Speed (Gbps)[20]		1.6	1.8	1.6	1.8	1.6

Industry: Heterogeneous Integration Roadmap

HETEROGENEOUS INTEGRATION ROADMAP

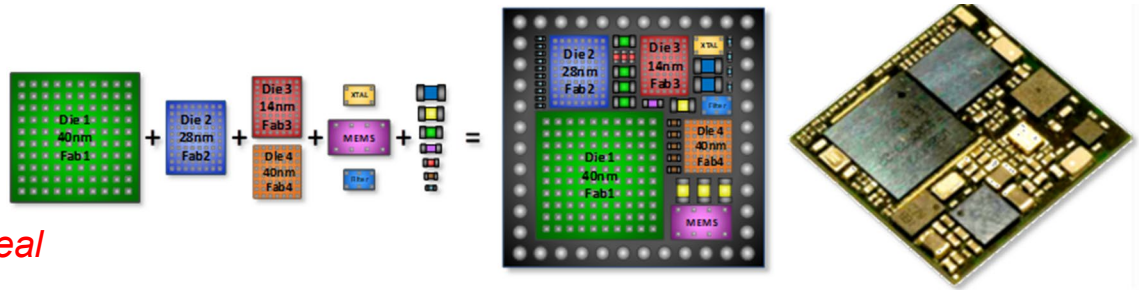
2019 Edition

http://eps.ieee.org/hir

HPC and Megadatacenters is 2nd chapter

Note: leading edge design nodes are not ideal
For every component (e.g. SERDES)

All future applications will be further transformed through the power of AI, VR, and AR.



Die + Heterogeneous

System in Package (SiP)

Conclusions

- **Think more seriously about how to use specialization productively for science**
  - Requires deep understanding of applied mathematics and the underlying algorithms to be successful
- **Reevaluate the economic model for the design/acquisition of HPC systems**
- **Scaling alone (e.g. Zettaflops) is no longer a rational metric for HPC success**
  - What metrics demonstrate effectiveness for science (which should == success?)
  - How to measure *success* in this new environment??? *You can't improve what you can't measure.*
- **Let Us Model Solutions for the Global Climate Crisis without Contributing to Global Warming!**

(Carbon Neutral HPC by 2030! Would not be a bad alternative metric)