BERKELEY LAB

Bringing Science Solutions to the World

SUPERCOMPUTERS AND SUPERINTELLIGENCE

Horst Simon, hdsimon@lbl.gov

17th SIAM Conference on Parallel Processing for Scientific Computing, Paris, April 12 – 15, 2016

“It’s not a human move …”

We just experienced another milestone in machine intelligence:

Alpha Go of Deep Mind (Google) winning Go against Lee Sedol, one of the world’s top go players.

March 11, 2016

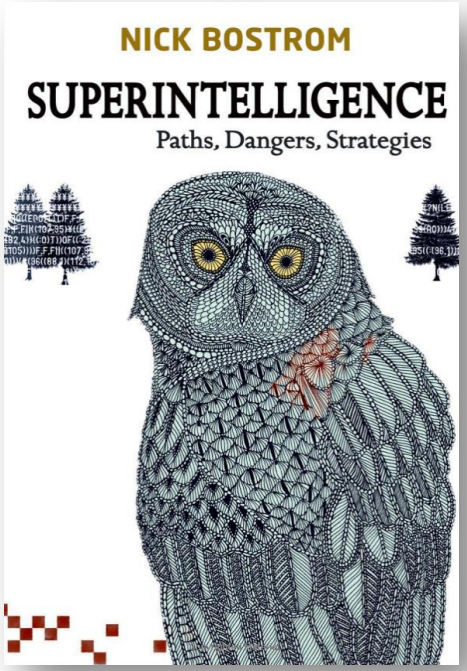
http://www.wired.com/2016/03/sadness-beauty-watching-googles-ai-play-go/?mbid=social_fb

Recent concerns about the “machines” taking over

“Those disposed to dismiss an 'AI takeover' as science fiction may think again after reading this original and well-argued book.”
– Martin Rees, Past President, Royal Society

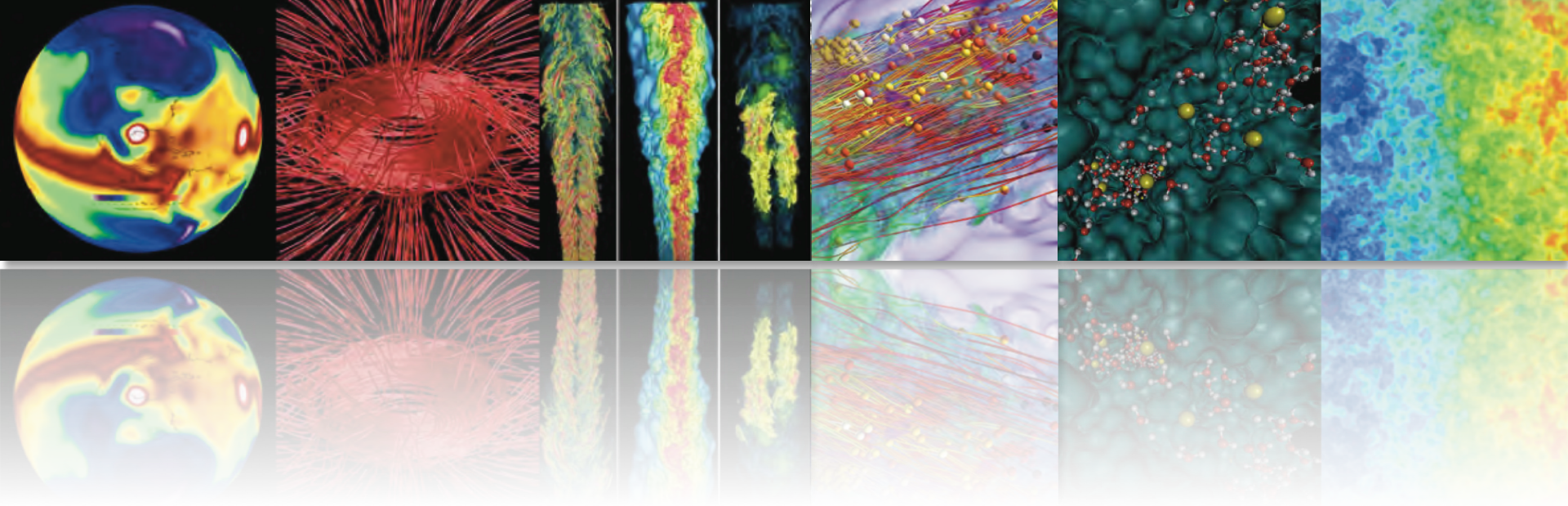
“Worth reading *Superintelligence* by Bostrom. We need to be super careful with AI. Potentially more dangerous than nukes.”
– Elon Musk

“If our own extinction is a likely, or even possible, outcome of our technological development, shouldn't we proceed with great caution?”
– Bill Joy



“Success in creating AI would be the biggest event in human history. Unfortunately, it might also be the last, unless we learn how to avoid the risks.”
– Steven Hawking

Slide adapted from Larry Smarr, UCSD



1 Current Trends in Supercomputing
(The Path to Exascale)

2 Computing and the Brain
(Hype and reality)

3 What Computers Still Can't Do
(and what I think the real dangers are)

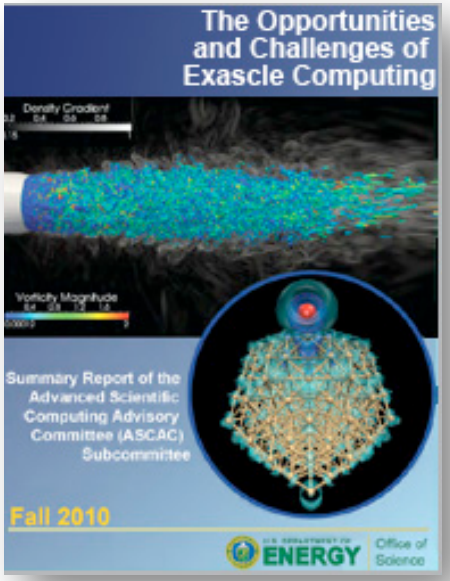
BERKELEY LAB

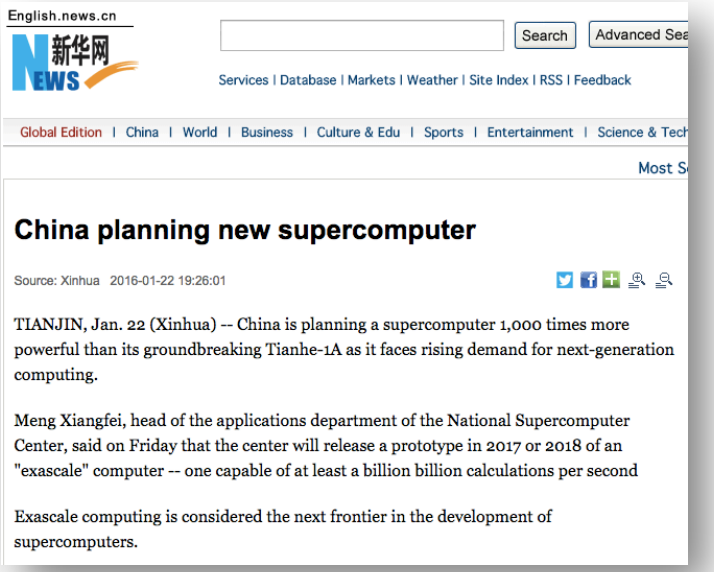
And what about supercomputers?

- At any given time one of the most powerful computers to solve scientific and engineering problems
- Supercomputers and HPC are largely absent from the public discussion about progress in AI

Exascale initiatives are advancing the computational power of supercomputers

- NSCI (National Strategic Computing Initiative) announced by President Obama in June 2015
- Exascale Computing Project ECP started by DOE in the US
- Similar initiatives in Europe, Japan, and China





The TOP500 Project
(by Meuer, Strohmaier, Dongarra, Simon)

TOP500 Supercomputers

Listing of the 500 most powerful computers in the world

Yardstick: Rmax of Linpack

- Solve Ax=b, dense problem, matrix is random
- Dominated by dense matrix-matrix multiply

Updated twice a year:

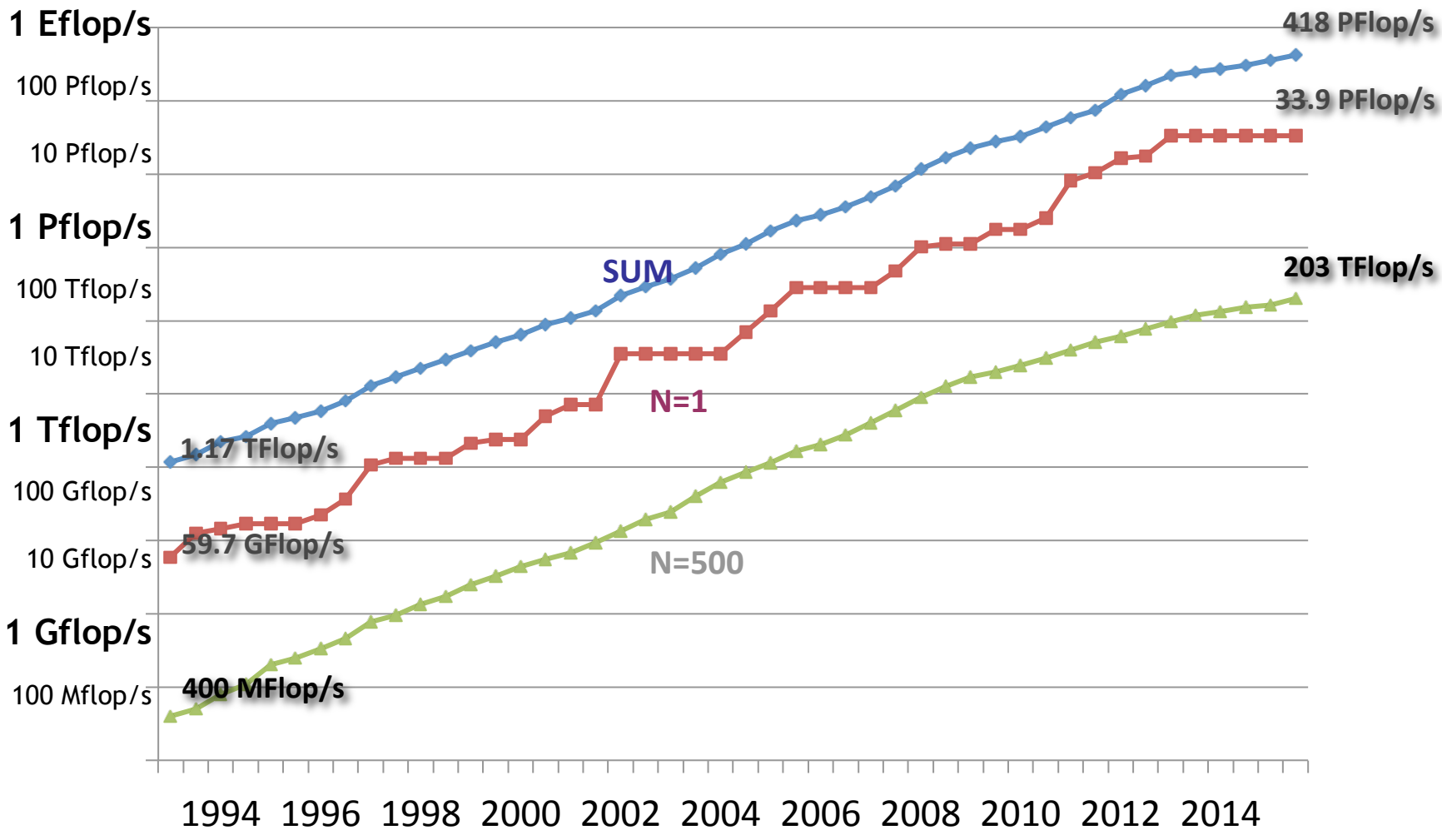
- ISC'xy in June in Germany
- SCxy in November in the U.S.

Hans Meuer (1936 – 2014)

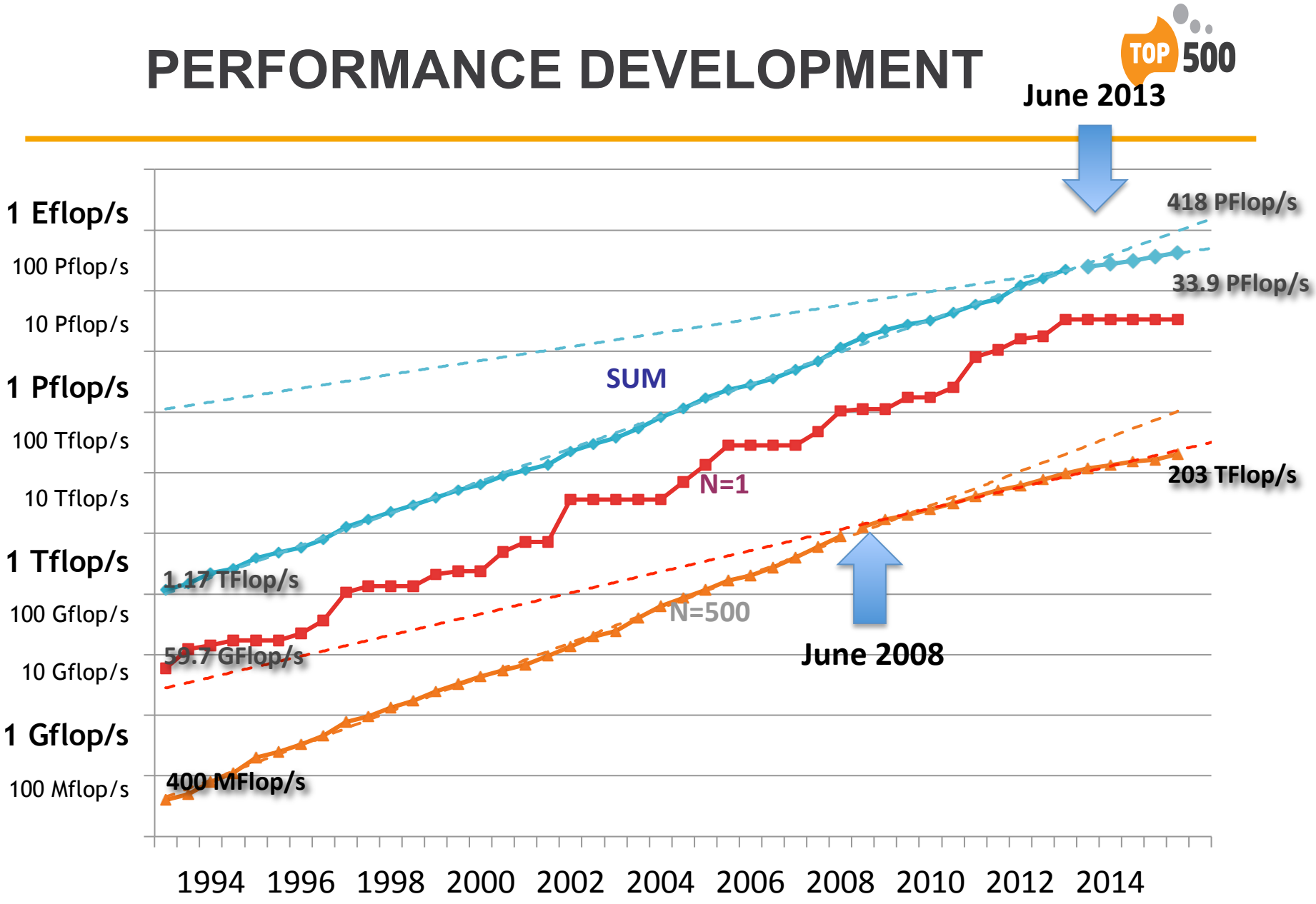
All information available from the TOP500 web site at: www.top500.org

See also Strohmaier, Meuer, Dongarra, and Simon, IEEE Computer, Nov. 2015, pp. 32-39.

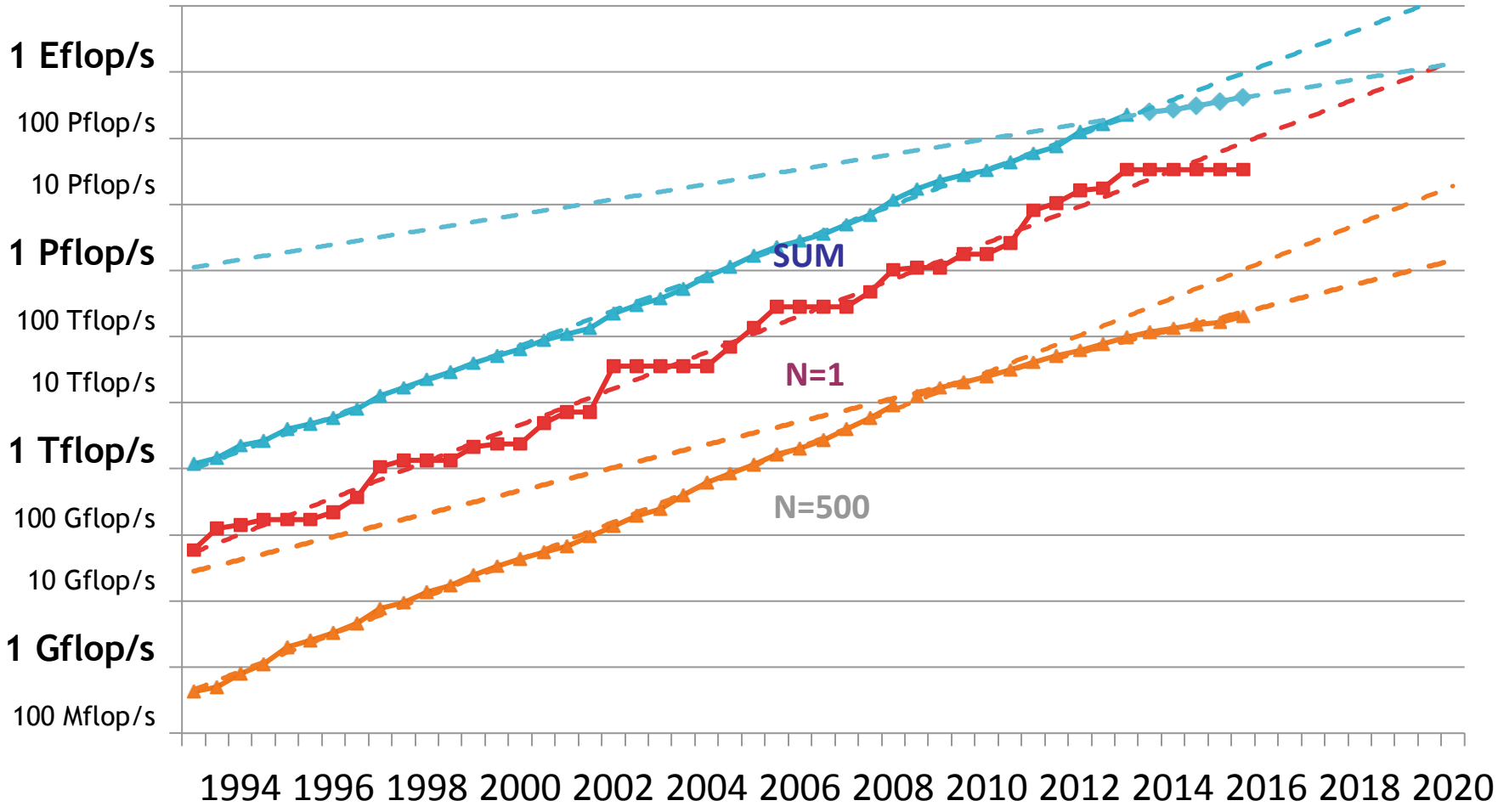
PERFORMANCE DEVELOPMENT

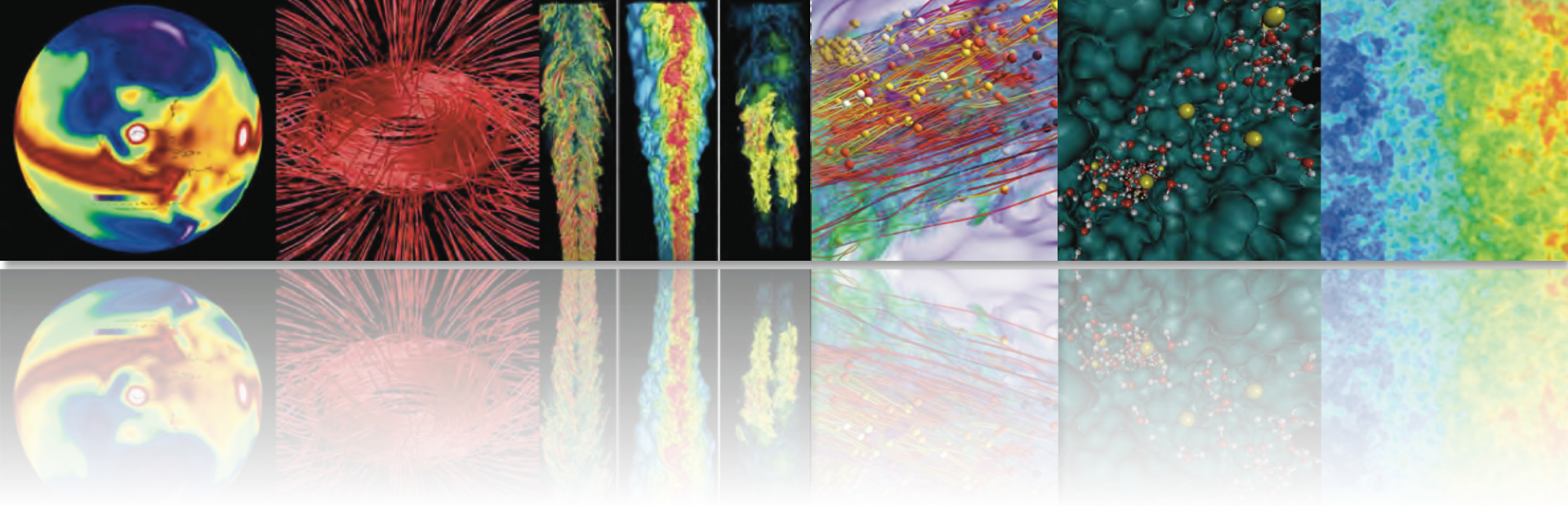


PERFORMANCE DEVELOPMENT



PROJECTED PERFORMANCE DEVELOPMENT





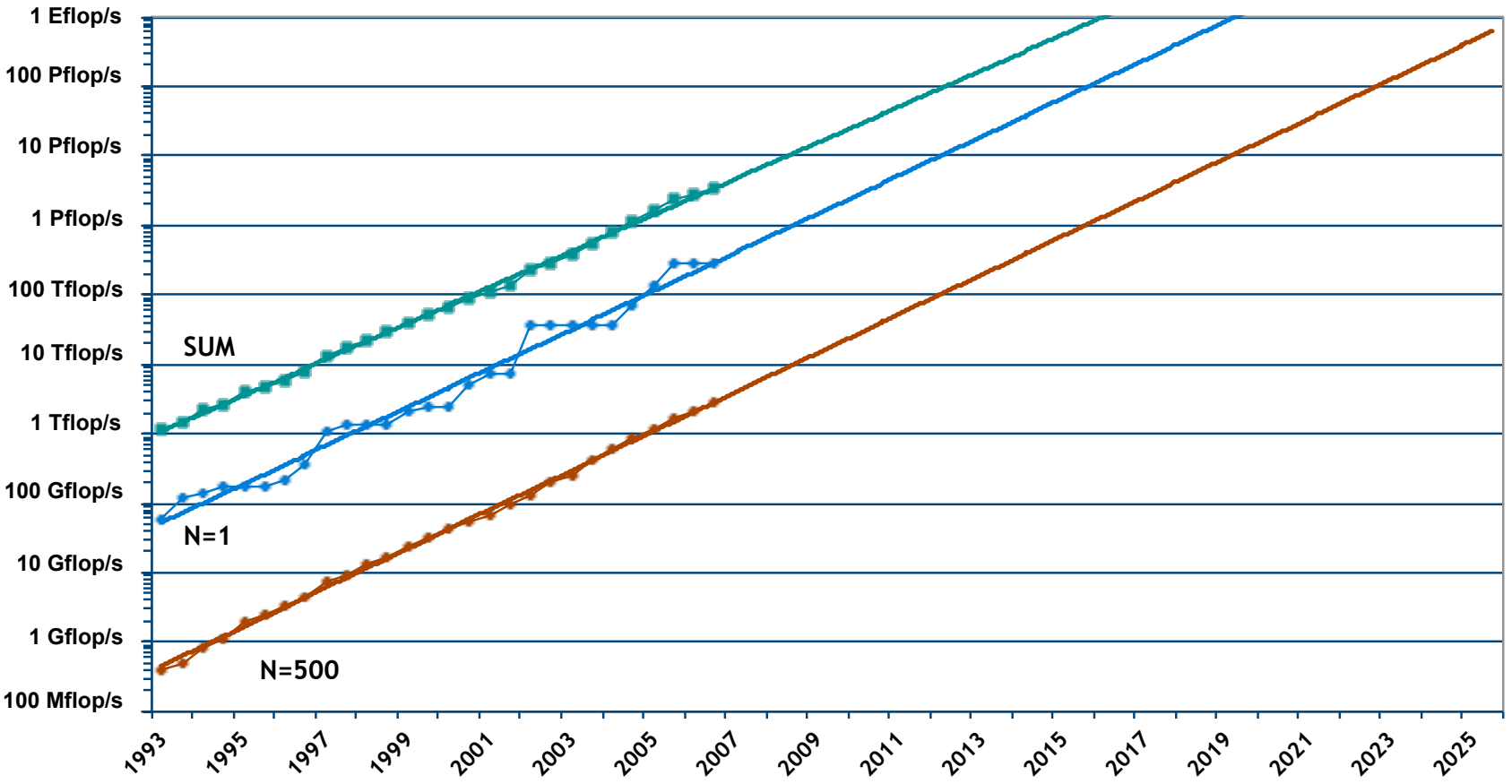
1 Current Trends in Supercomputing
(The Path to Exascale)

2 Computing and the Brain
(Hype and reality)

3 What Computers Still Can't Do
(and what I think the real dangers are)

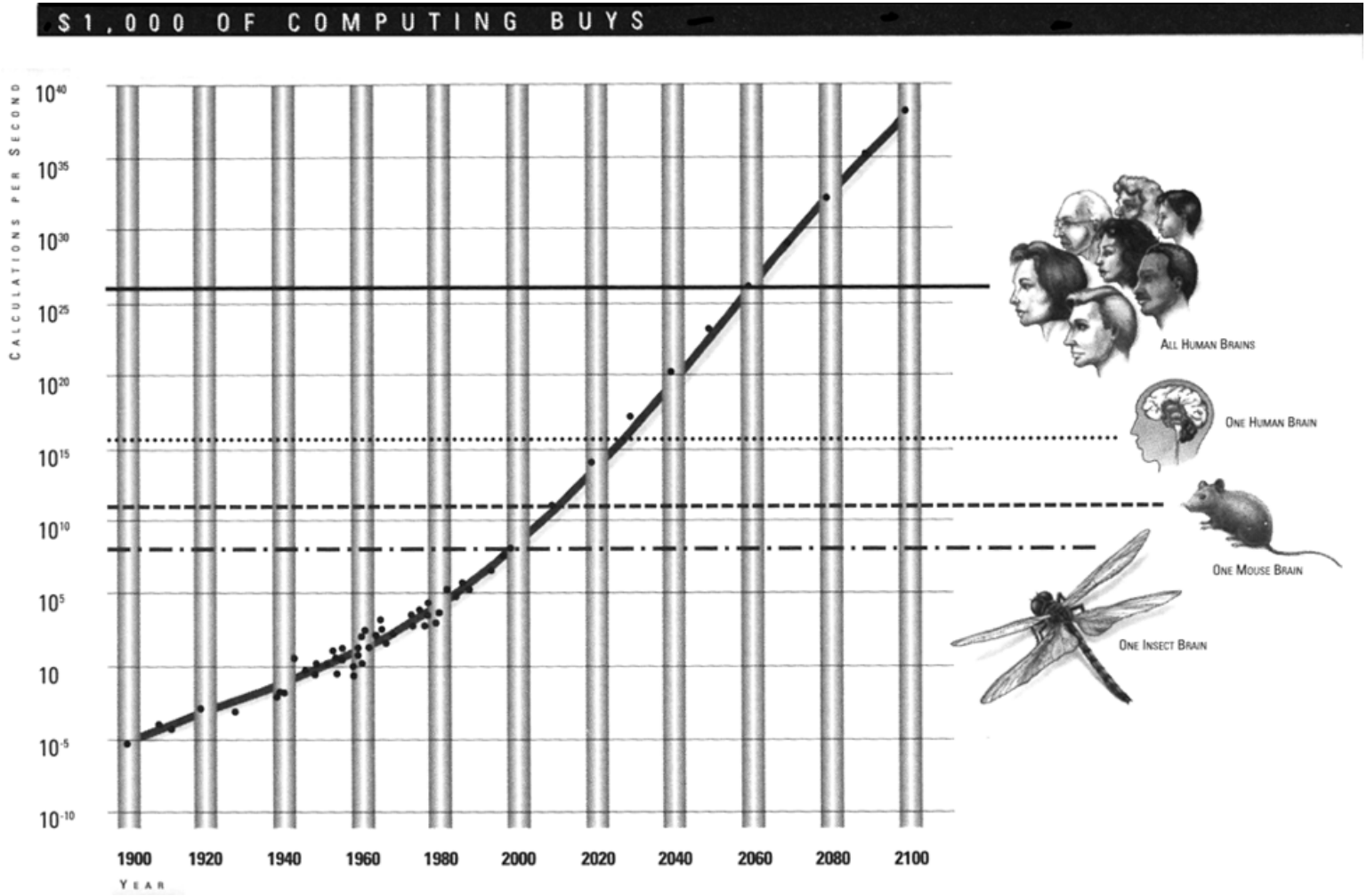
BERKELEY LAB

TOP 500 Performance Projection - The Old Picture From 2007



BERKELEY LAB

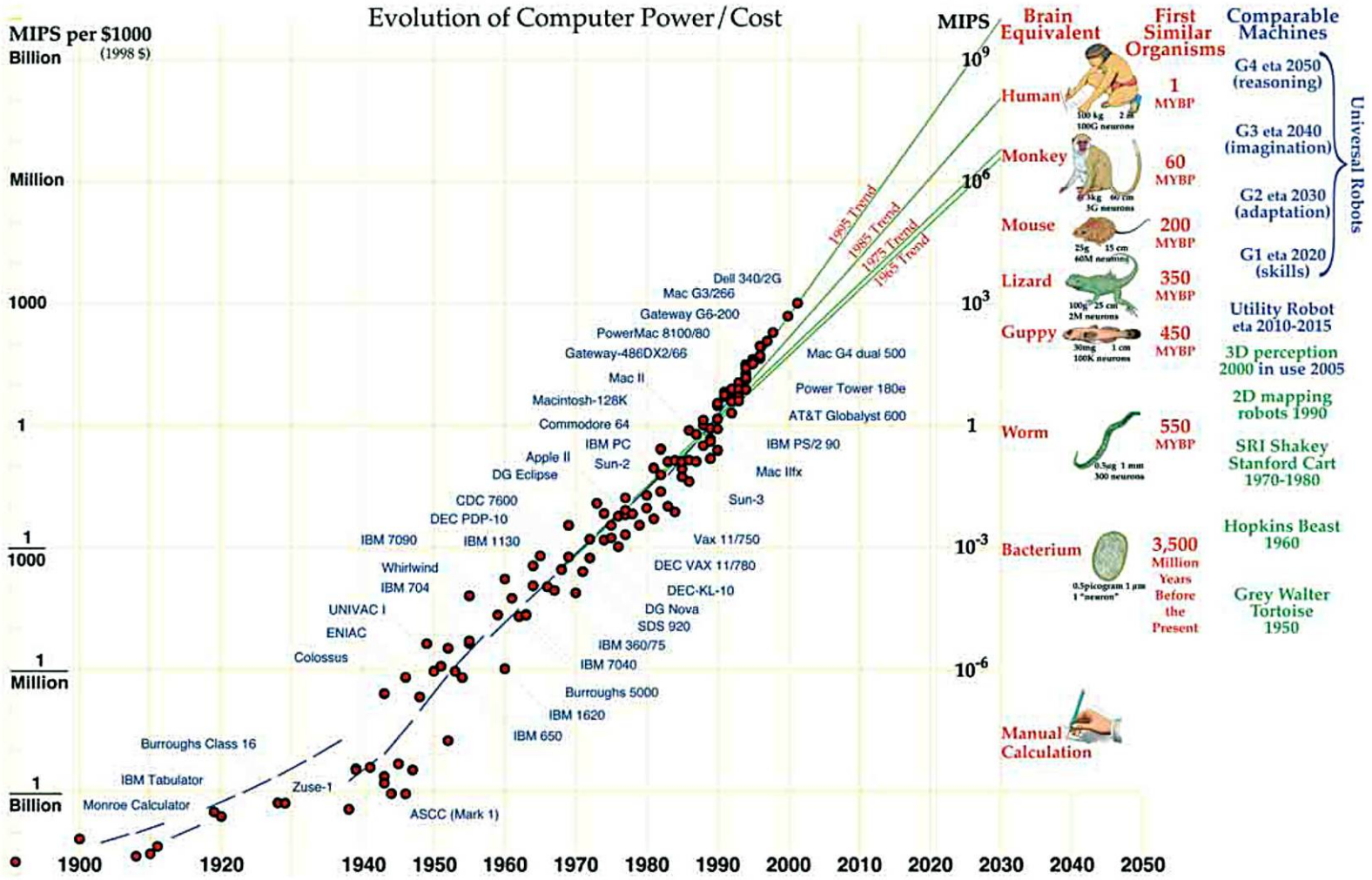
The Exponential Growth of Computing



Adapted from Kurzweil, *The Age of Spiritual Machines.*

BERKELEY LAB

Growth of Computing Power & "Mental Power"



Hans Moravec, *CACM* 10, 2003, pp 90-97.

BERKELEY LAB

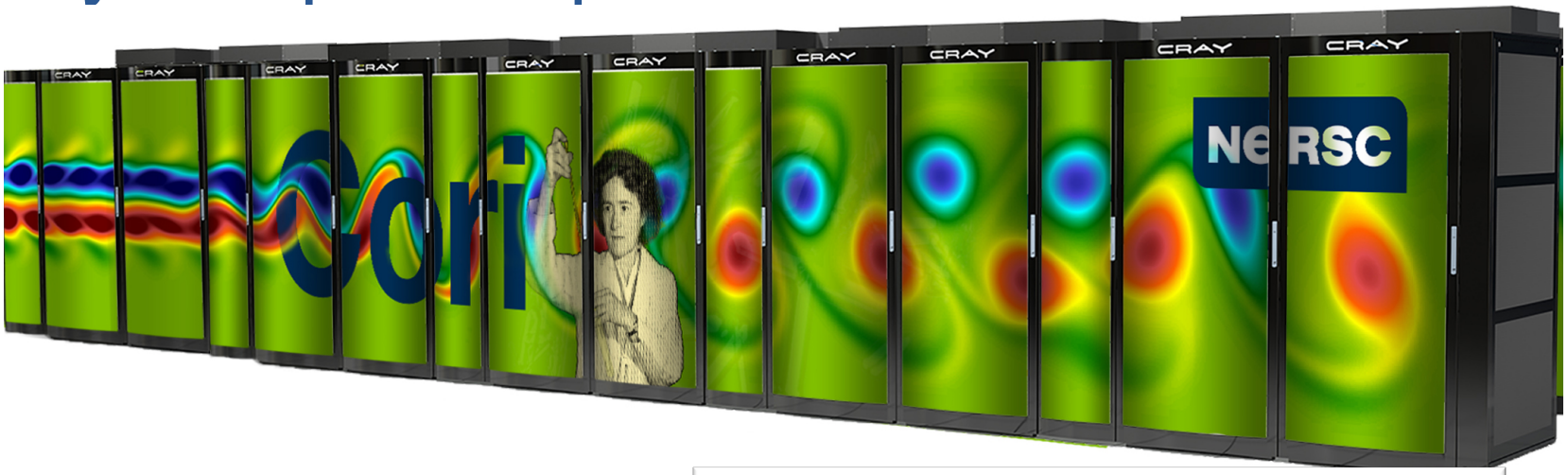
Arguments against this simplistic view

- Naïve extrapolation of current performance trends
  - transition to Post Moore's Law computing not clear
  - HPC systems are focused on excelling in scientific computing
- Scaling of AI and machine learning to millions of cores
  - most powerful computers today are not being used for cognitive tasks
- Successes of machine learning are accomplished through progress in algorithms

History Lesson: 1997

- IBM Deep Blue beats Gary Kasparov (May 1997)
- One of the biggest success stories of machine intelligence,
- However, the chess computer “Deep Blue”, did not teach us anything about how a chess grandmaster thinks
- No further analysis or further developments
- 19 years later the story repeats itself with Go

Today's Supercomputers and the Brain



compute only

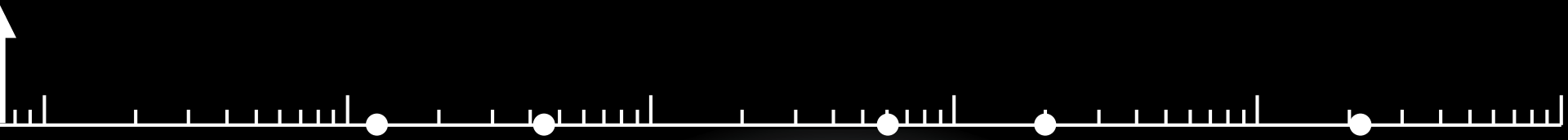
Transistors	Memory	Clock (GHz)	Power (W)	Weight (kg)	Size (ℓ)
$9x10^{12}$	$2x10^{15}$	2	$1.2x10^{6}$	18,800	36,000

$9x10^{10}$	$1x10^{15}$	$10^{-9} – 10^{-5}$	20	1.5
Neurons	**Syn. Conn.**			

Slide from Peter Denes

Modha Group at IBM Almaden

	Mouse	Rat	Cat	Monkey	Human
N:	16×10^6	56×10^6	763×10^6	2×10^9	22×10^9
S:	128×10^9	448×10^9	6.1×10^{12}	20×10^{12}	220×10^{12}



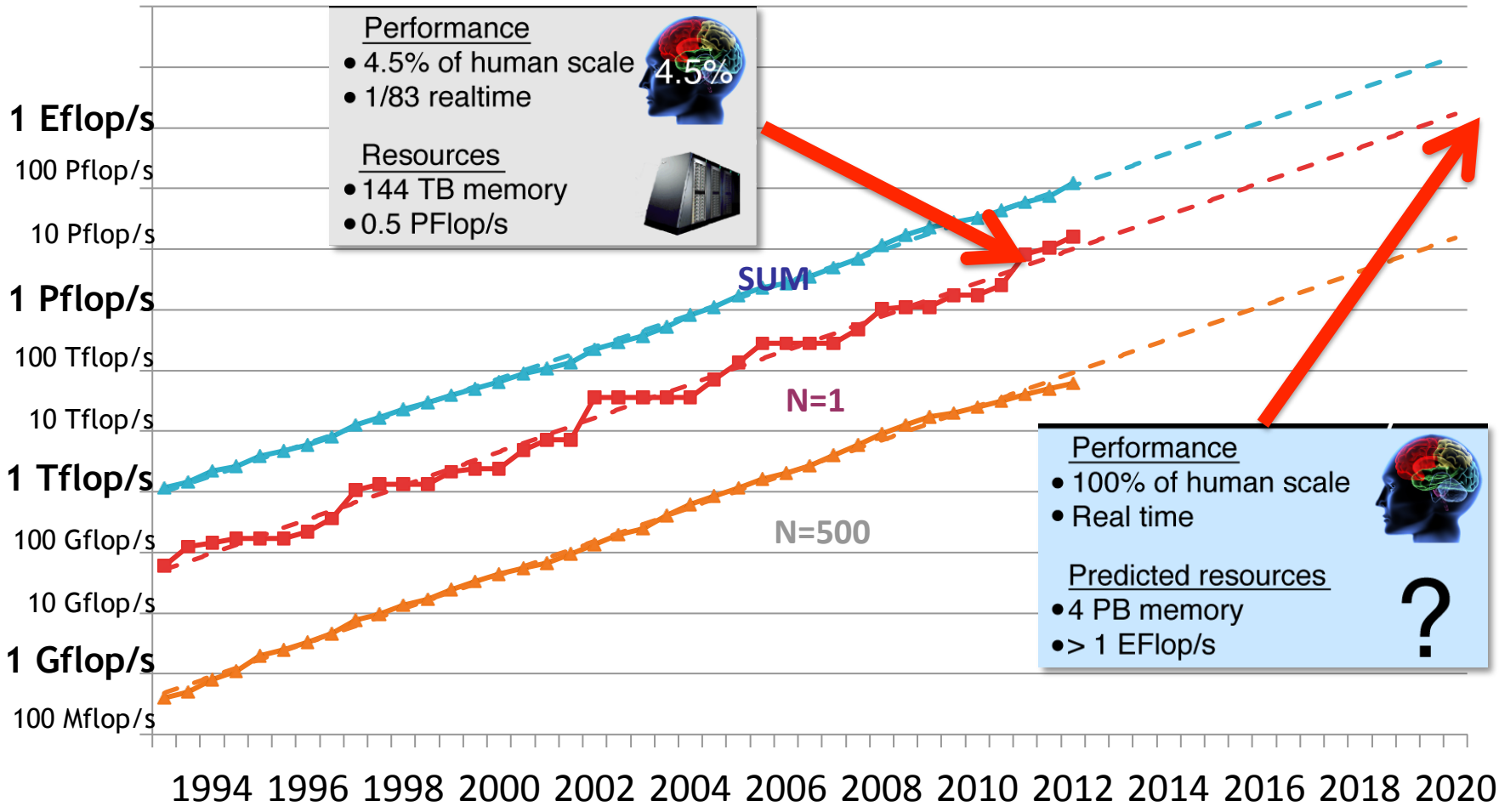
New results for SC12

Almaden	Watson	WatsonShaheen	LLNL Dawn	LLNL Sequoia
BG/L	BG/L	BG/P	BG/P	BG/Q
December, 2006	April, 2007	March, 2009	May, 2009	June, 2012

Latest simulations in 2012 achieve unprecedented scale of $65*10^9$ neurons and $16*10^{12}$ synapses

Real time simulation at the exascale

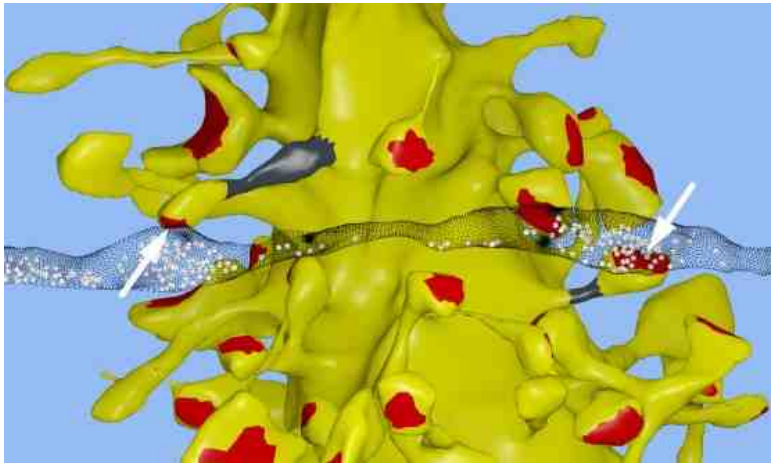
BERKELEY LAB



Ananathanarayanan et al., "The Cat is out of the Bag: Cortical Simulations with 10^9 Neurons and 10^{15} Synapses", Proceedings of SC09.

Recent Evidence for Petabyte Size Memory

- Bartol et al., “Nanoconnectomic Upper Bound on the Variability of Synaptic Plasticity”, Jan. 2016, eLife.
- 26 sizes of synapses corresponds to about 4.7 bits per synapse and thus about 4 – 5 Petabytes



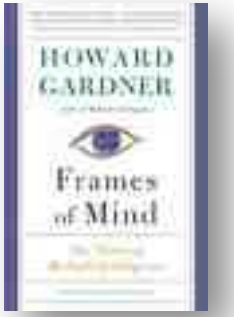
from http://www.salk.edu/news-release/memory-capacity-of-brain-is-10-times-more-than-previously-thought/

Compute Power of the Human Brain

- Estimate of compute power for the human brain is about 1-10 Exaflops and 4-5 Petabytes
- Three different paths lead to about the same estimate
- A digital computer with this performance might be available in about 2024 with a power consumption of at best 20–30 MW (goal of the Exascale project)
- The human brain takes 20 W
- A digital exaflops computer using CMOS technology will still be a factor of a million away from brain power

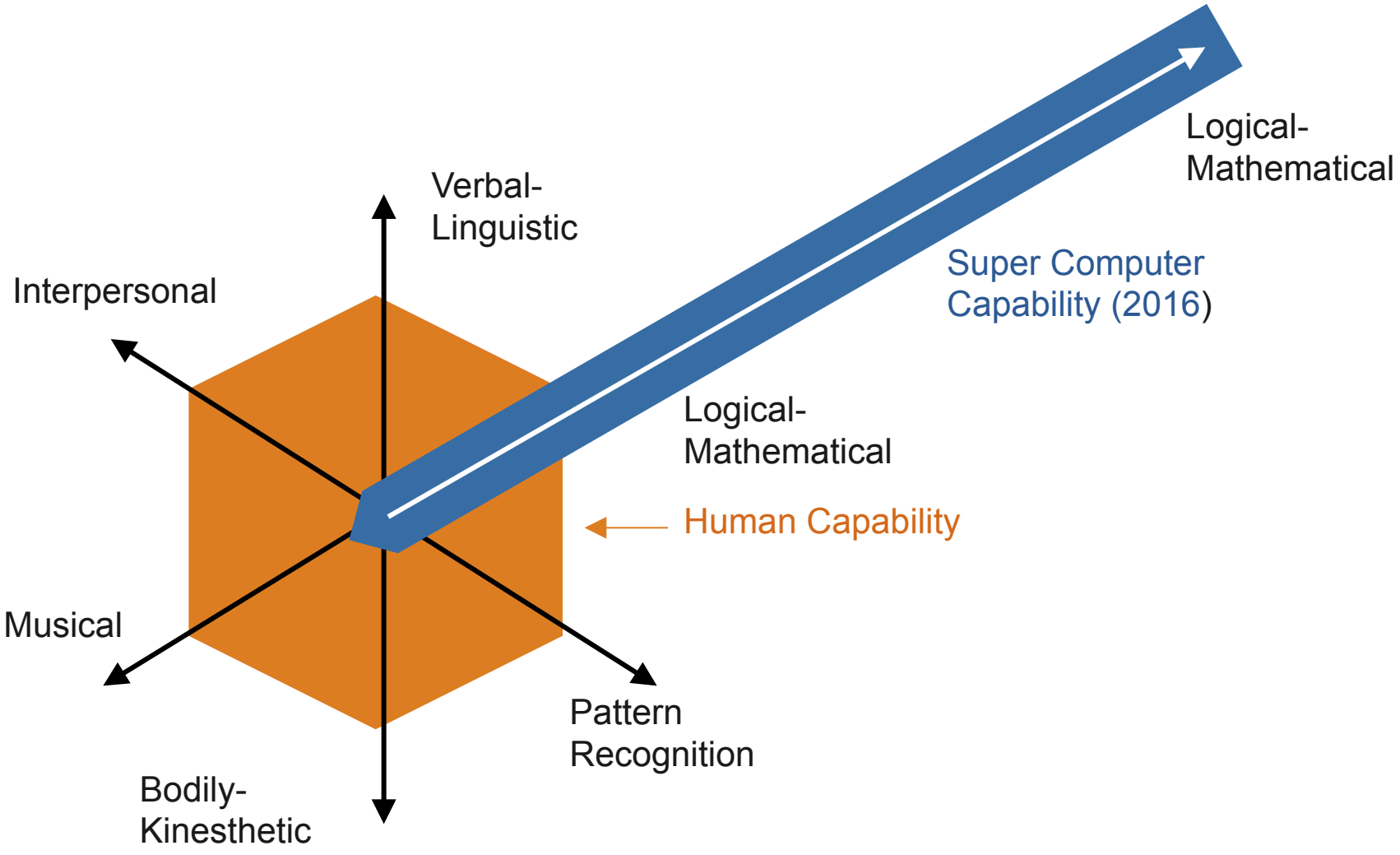
Dimensions of Intelligence

1. Verbal-Linguistic
   - ability to think in words and to use language to express and appreciate complex concepts
2. Logical-Mathematical
   - makes it possible to calculate, quantify, consider propositions and hypotheses, and carry out complex mathematical operations
3. Pattern Recognition
   - capacity to recognize and think about common pattern in our four-dimensional environment
4. Bodily-Kinesthetic
   - ability to manipulate objects and fine-tune physical skills
5. Musical
   - sensitivity to pitch, melody, rhythm, and tone
6. Interpersonal
   - capacity to understand and interact effectively with others



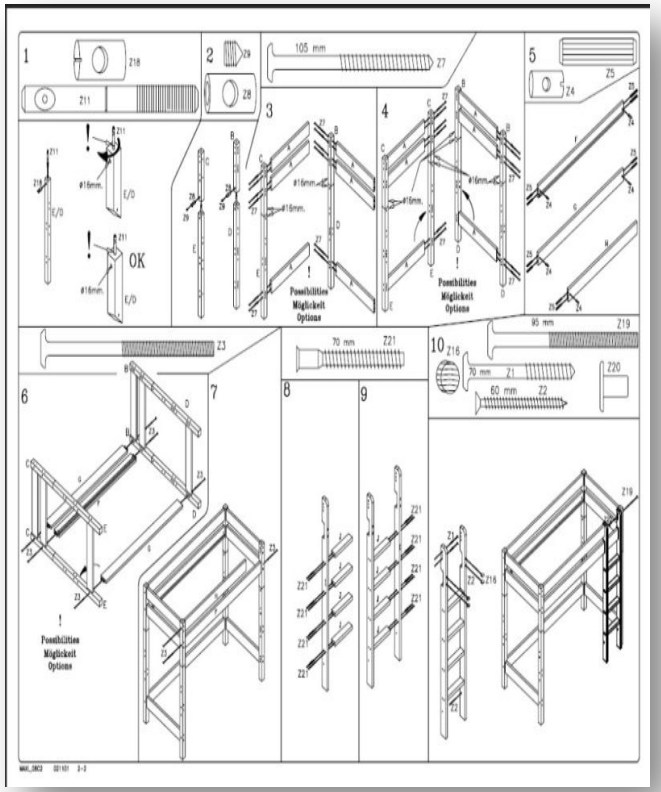
After Howard Gardner. *Frames of Mind: The Theory of Multiple Intelligences*. New York: Basic Books, 1983, 1993.

Current State of Supercomputers



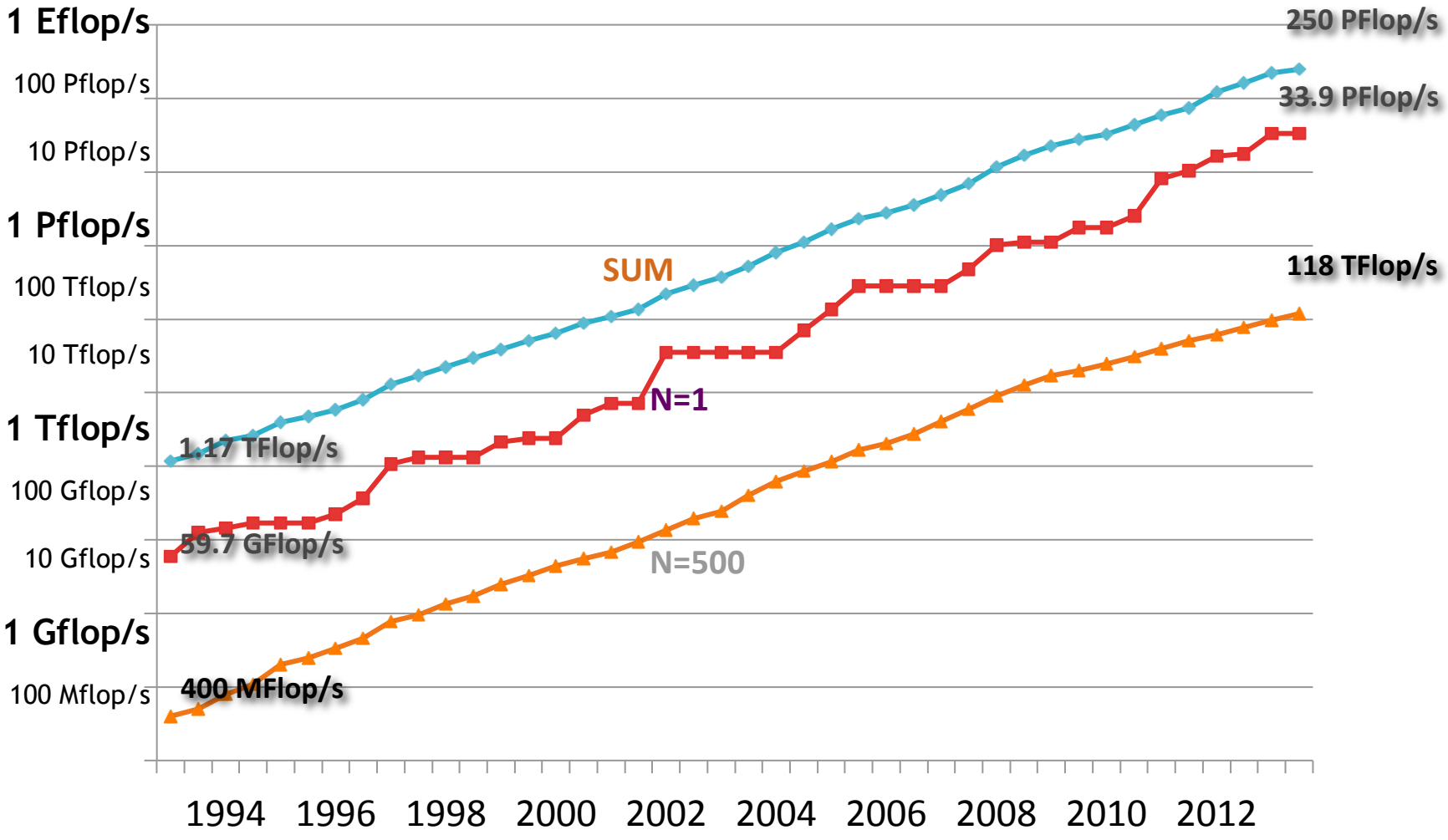
A Comment about the Turing Test

- Test only one dimension – verbal linguistic
  - Need addtional tests that explore other dimensions of human intelligence
  - My favorite is the “Ikea test” for pattern recognition and 3d spatial thinking: have a robot build furniture from the schematic drawing



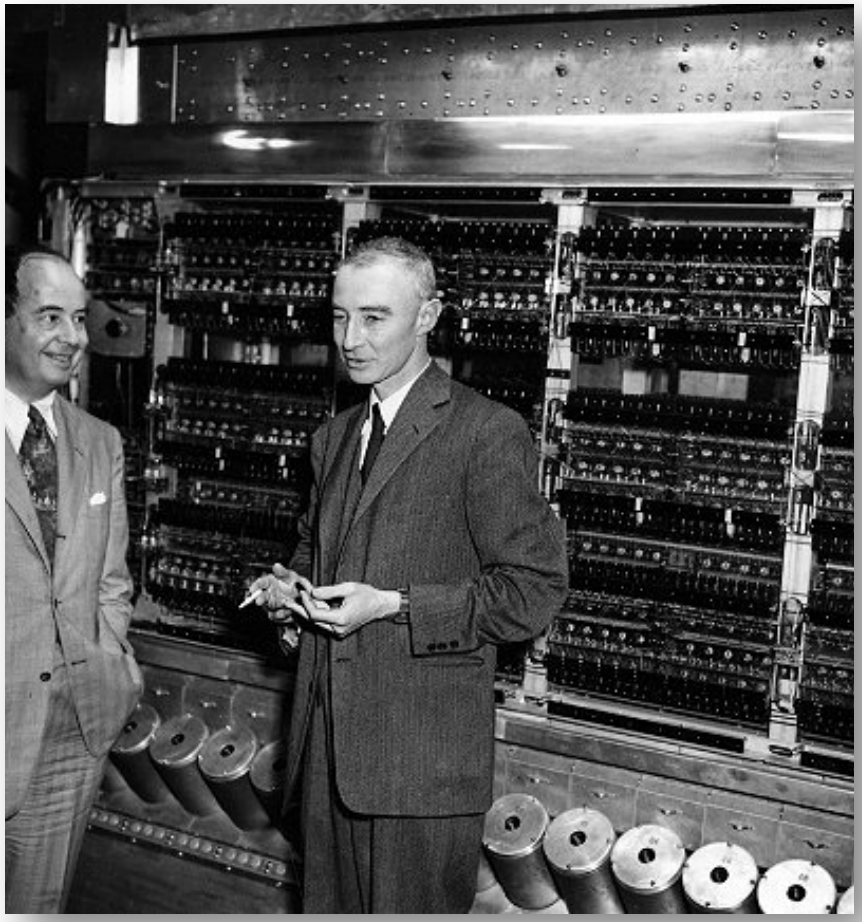
http://io9.gizmodo.com/8-possible-alternatives-to-the-turing-test-1697983985

Performance Development



Source: TOP500 November 2015.

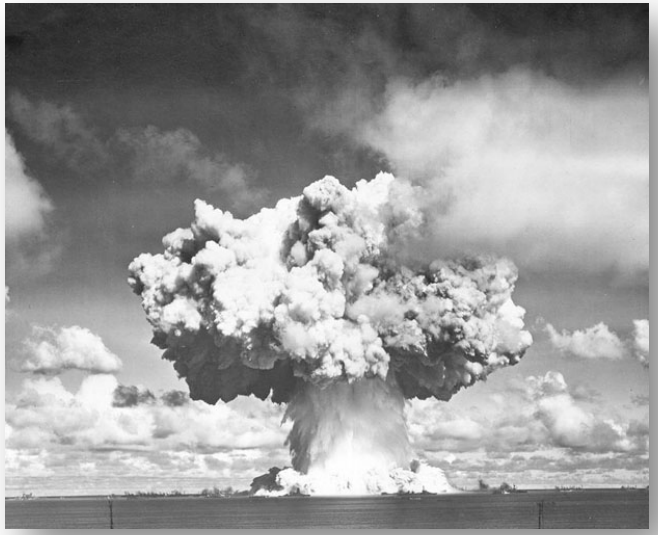
Historical Perspective, or "Why We Are Here Today …."



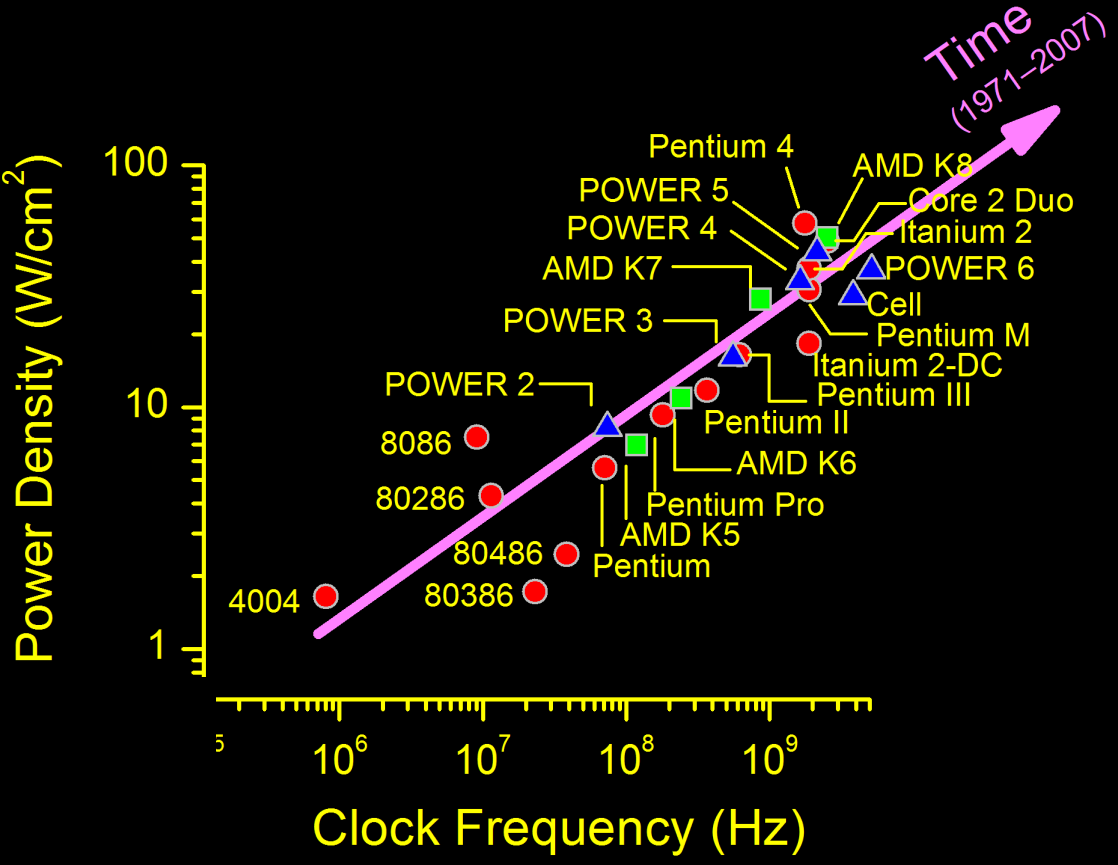
John von Neumann and Robert Oppenheimer, Princeton, IAS, 1952

- 60 years of large scale computational physics applications driving computer development
- Remarkable longevity
- Architecture matched to application: Von Neumann architecture and focus on floating point performance

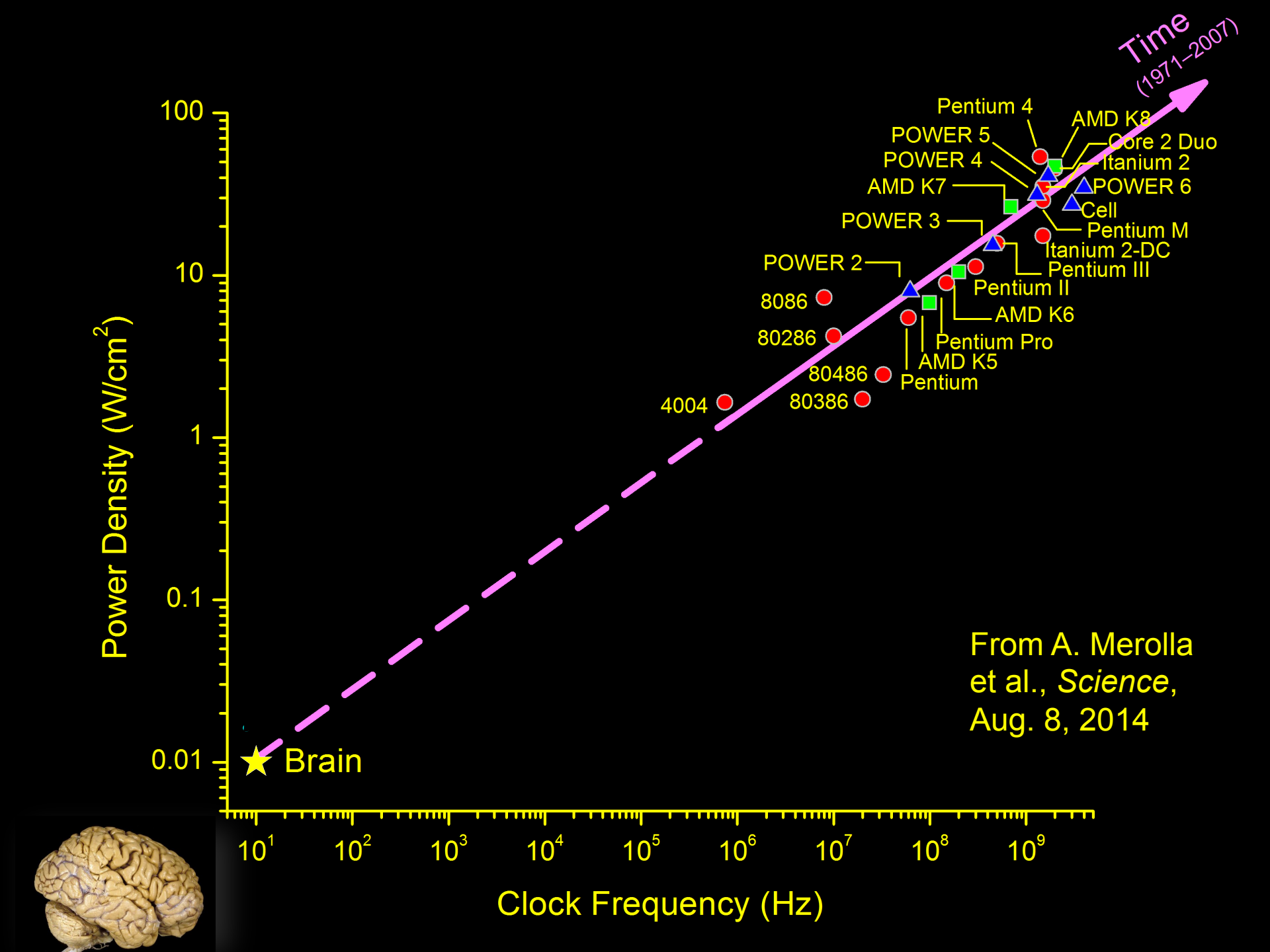
From the Bomb to the Cloud in Sixty Years



- Von Neumann architecture is ideally suited for large-scale, floating point intense, 2D or 3D grid-based, computational physics applications
- Today we are using the same basic architecture for social networking, Web searches, music, photography, etc.



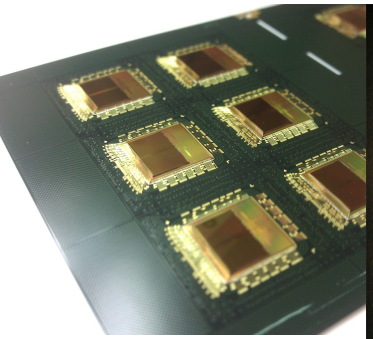
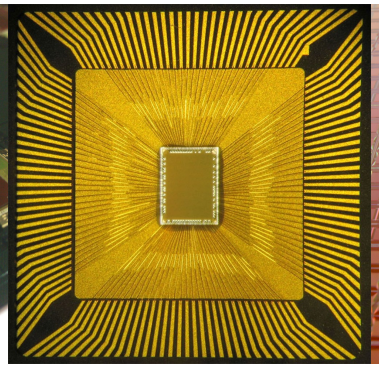
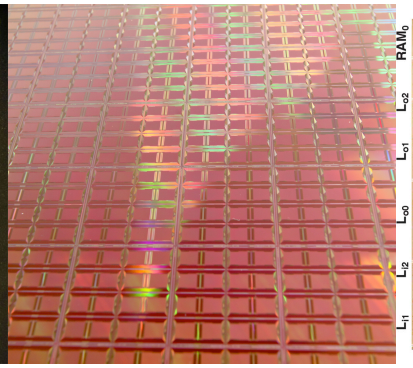
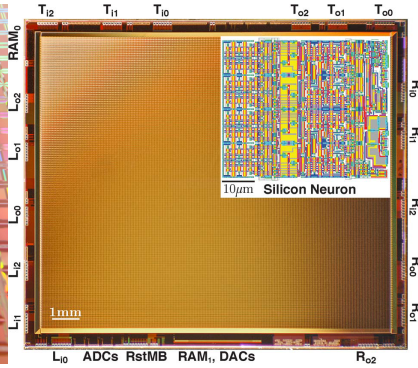
From A. Merolla et al., *Science*, Aug. 8, 2014

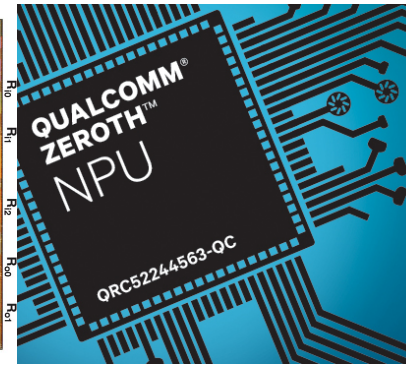

Time
(1971–2007)
Pentium 4
AMD K8
POWER 5
Core 2 Duo
POWER 4
Itanium 2
AMD K7
POWER 6
Cell
POWER 3
Pentium M
Itanium 2-DC
Pentium III
POWER 2
Pentium II
8086
AMD K6
80286
Pentium Pro
AMD K5
80486
Pentium
80386
4004
Brain
100
10
1
0.1
0.01
Power Density (W/cm^2)
10^1
10^2
10^3
10^4
10^5
10^6
10^7
10^8
10^9
Clock Frequency (Hz)
From A. Merolla et al., *Science*, Aug. 8, 2014

Recent Neuromorphic Projects

Five complementary approaches to Neuromorphic Computing (massively parallel, asynchronous communication, configurable):

- Commodity microprocessors (SpiNNaker, HBP)
- Custom fully digital (IBM Almaden)
- Custom mixed-signal (BrainScaleS, HBP)
- Custom subthreshold analog cells (Stanford, ETHZ)
- Custom yybrid (Qualcomm)

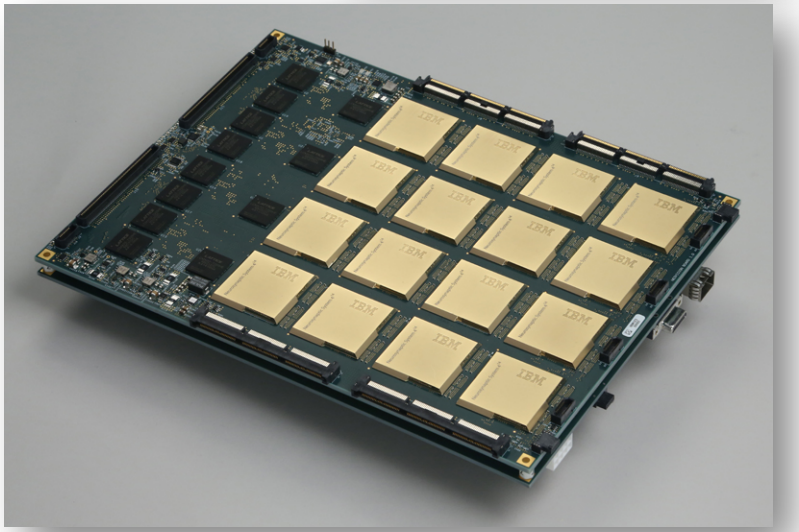





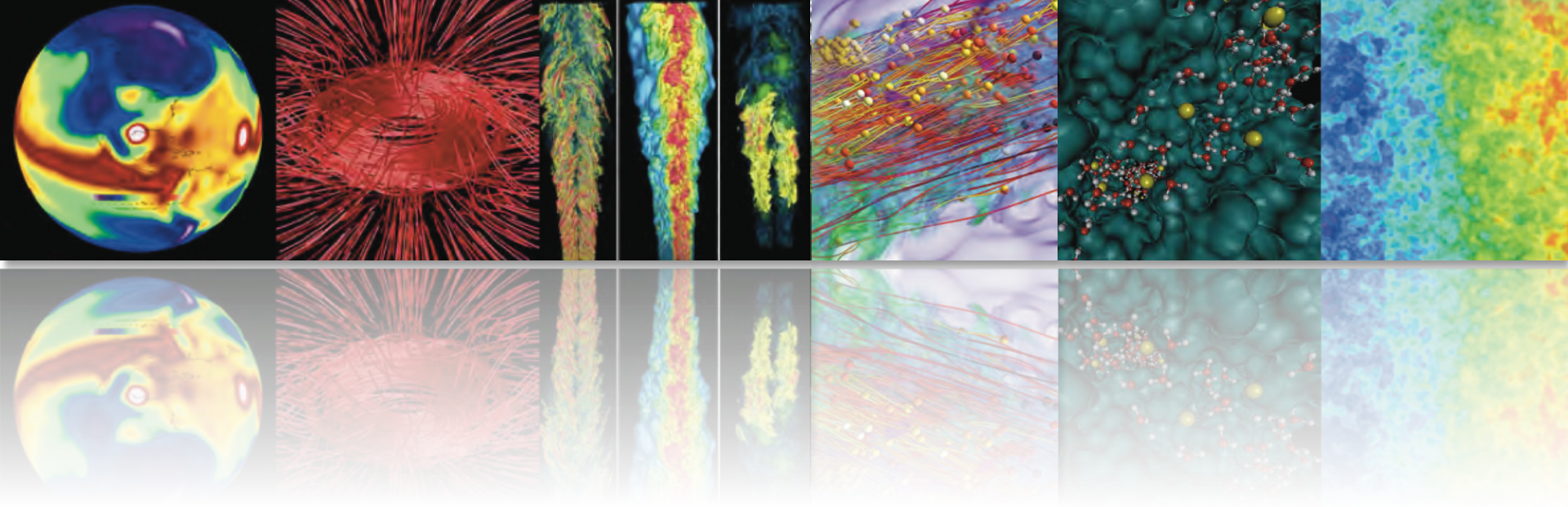


after K. Meier, Nov. 2014

IBM SyNAPSE Project (D. Modha)

- 5.4 B transistor chip TrueNorth,
  4096 neurosynaptic cores, 1 M spiking neurons,
  256 M configurable synapses
- 63 mW power per chip, significantly less energy per event (176,000) when compared to a simulator
- Scalability to large system
- Corelet programming model





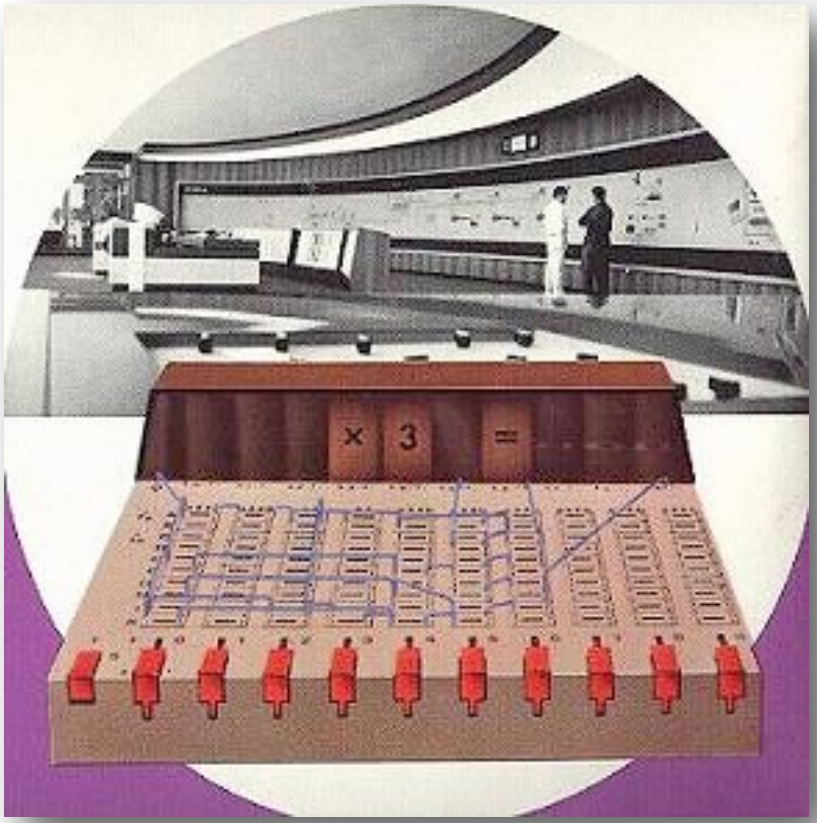
1 Current Trends in Supercomputing
(The Path to Exascale)

2 Computing and the Brain
(Hype and reality)

3 What Computers Still Can't Do
(and what I think the real dangers are)

BERKELEY LAB

About 1967



Kosmos LOGIKUS
"Spielcomputer"

My first computer:

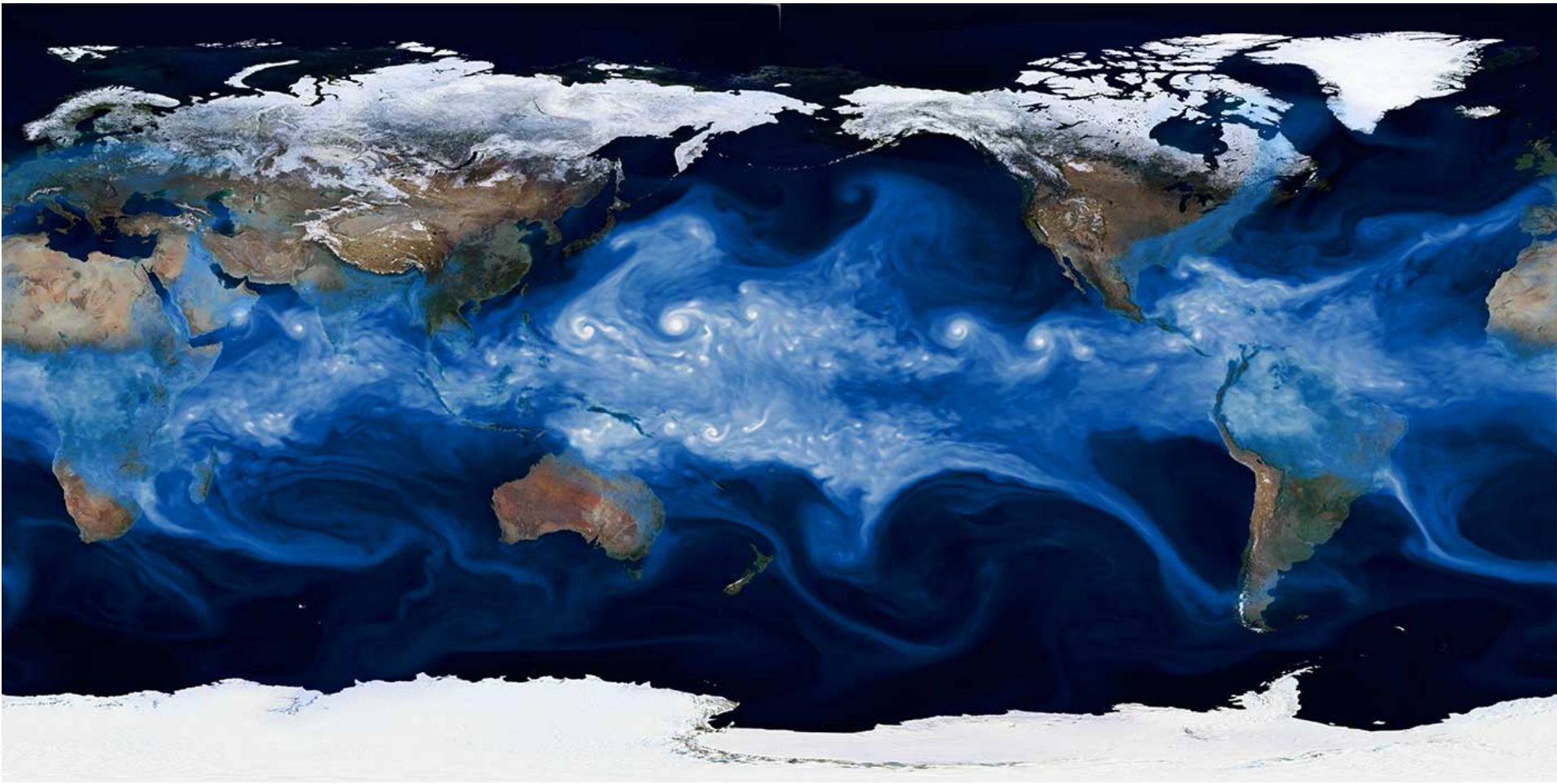
- 10 electric lamps
- 10 switches
- Programming by wiring

http://www.logikus.info/

What I learned:
computers are just machines, wires, batteries, and light bulbs. They cannot have a mental state or experience.
Many years later I realized that this position is equivalent to that I don't believe in strong AI.

A typical supercomputer simulation

… but does Edison@NERSC know what a hurricane is?



BERKELEY LAB

Strong and Weak AI

Strong AI: A physical symbol system can have a mind and mental states.

Does my play computer get upset when I win?

Does the Edison computer at NERSC know what a hurricane IS?

HDS: This is an interesting philosophical question, but we can leave it aside for this discussion.

Weak AI: A physical symbol system can act intelligently.

Is it possible to develop a computer system that performs indistinguishable from a human?

HDS: Yes, in principle, but it will be very, very hard.

Some challenges ahead for modeling the brain

- Unsuitability of current architectures
  - HPC systems are focused on excelling in computing; only one of the six (or eight) dimensions of human intelligence
- Fundamental lack of mathematical models for cognitive processes
  - That's why we are not using the most powerful computers today for cognitive tasks
- Lack of data, standards. That's what the BRAIN initiative in the US should address
- We should be able to model the brain as a complex physical system, but we have barely started

What we should be really concerned about: (1) large scale complex systems controlled by algorithms

Example:

High Frequency Trading (HFT) is now accounting for over 60% of the volume in US equity markets.

The interaction of multiple algorithms create a complex system that we don't understand any longer, yet our prosperity depends on it.

Dow Jones Index on the day of the FLASH crash

See also Center for Innovative Financial Technologies (CIFT) at LBNL:
http://crd.lbl.gov/departments/data-science-and-technology/sdm/current-projects/cift

BERKELEY LAB

What we should be really concerned about: (2) confluence of pattern recognition machines + image analysis + big data + behavioral prediction

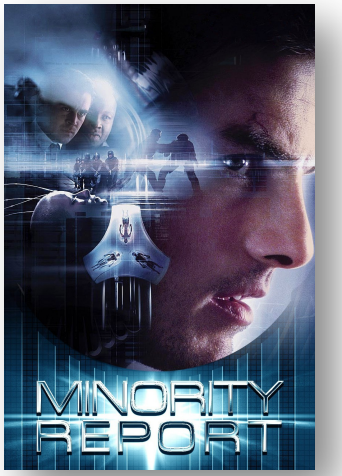
Deployment of neuromorphic processors that excel at pattern recognition inexpensively in large scale (IoT)

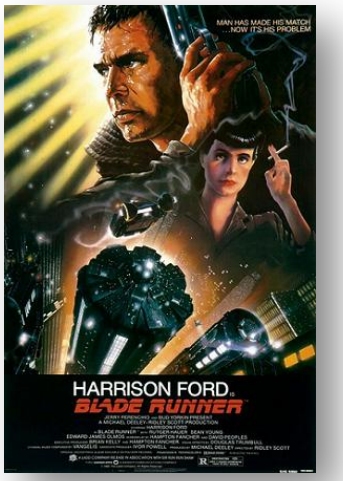
Collection of huge amounts of data and image in the cloud

Capability of real time streaming data analysis

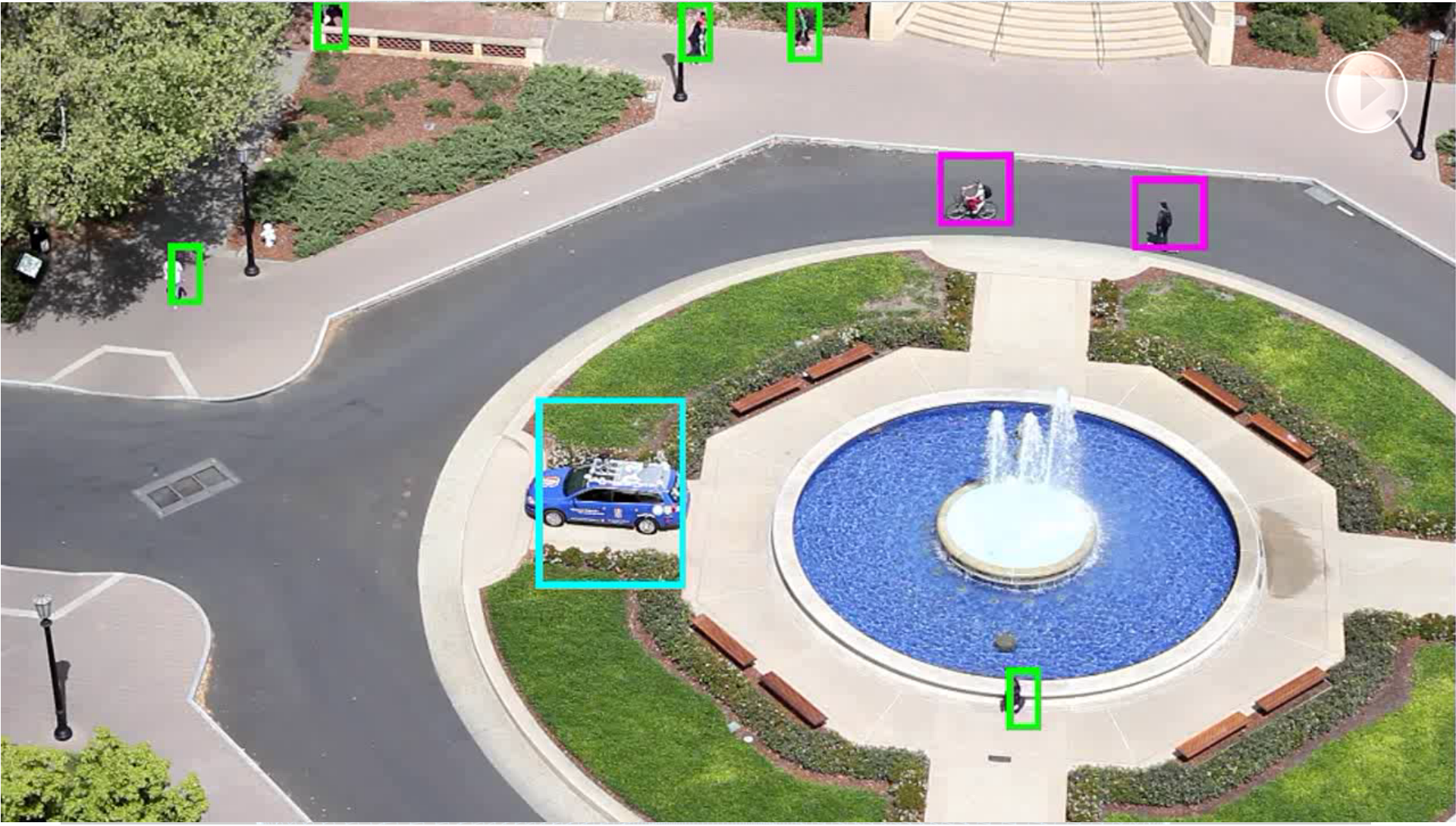
Behavioral prediction

For friends of SF (Philip K. Dick): I am more concerned about "Minority Report Future" than a "Blade Runner Future"





Towards the pattern recognition machine



From D. Modha, IBM Research

Summary – Key Messages

- Current HPC technology is about a factor of 10^9 away from the real time performance of the human brain (10^6 in power, 10^3 in computation, <10 in memory)
- Both technology and architectural innovation are needed to close the gap.
- Given what we know today, an artificial, sentient, superintelligence is unlikey (strong AI).
- Given what we know today, a realistic, highly accurate simulation of brain functions will require major advances in systems, algorithms, and mathematical modeling, in parallel with progress in neuroscience (weak AI).
- The real danger is turning over decisions to systems that we don't understand and control.

Thank You for Contributions

Erich Strohmaier (LBNL)
Jack Dongarra (UTK)
John Shalf (LBNL)
Peter Denes (LBNL)

Michael Wehner and team (LBNL)

Dharmendra Modha and team (IBM)
Karlheinz Meier (Univ. Heidelberg)
and Wikipedia

EXTRA SLIDES

41st List: The TOP10

#	Site	Manufacturer	Computer	Country	Cores	Rmax [Pflops]	Power [MW]
1	National University of Defense Technology	NUDT	Tianhe-2 NUDT TH-IVB-FEP, Xeon 12C 2.2GHz, IntelXeon Phi	China	3,120,000	33.9	17.8
2	Oak Ridge National Laboratory	Cray	Titan Cray XK7, Opteron 16C 2.2GHz, Gemini, NVIDIA K20x	USA	560,640	17.6	8.21
3	Lawrence Livermore National Laboratory	IBM	Sequoia BlueGene/Q, Power BQC 16C 1.6GHz, Custom	USA	1,572,864	17.2	7.89
4	RIKEN Advanced Institute for Computational Science	Fujitsu	K Computer SPARC64 VIIIfx 2.0GHz, Tofu Interconnect	Japan	795,024	10.5	12.7
5	Argonne National Laboratory	IBM	Mira BlueGene/Q, Power BQC 16C 1.6GHz, Custom	USA	786,432	8.59	3.95
6	Los Alamos NL / Sandia NL	Cray	Trinity Cray XC40, Xeon E5 16C 2.3GHz, Aries	USA	301,0564	8.10	
7	Swiss National Supercomputing Centre (CSCS)	Cray	Piz Daint Cray XC30, Xeon E5 8C 2.6GHz, Aries, NVIDIA K20x	Switzerland	115,984	6.27	2.33
8	HLRS – Stuttgart	Cray	Hazel Hen Cray XC40, Xeon E5 12C 2.5GHz, Aries	Germany	185,088	5.64	
9	King Abdullah University of Science and Technology	Cray	Shaheen II Cray XC40, Xeon E5 16C 2.3GHz, Aries	Saudi Arabia	196,608	5.54	2.83
10	Texas Advanced Computing Center/UT	Dell	Stampede PowerEdge C8220, Xeon E5 8C 2.7GHz, Intel Xeon Phi	USA	462,462	5.17	4.51

Tianhe-2 (TH-2) at NUDT, China – #1 on the TOP500 list

Summary of the Tianhe-2 (TH-2) Milkyway 2

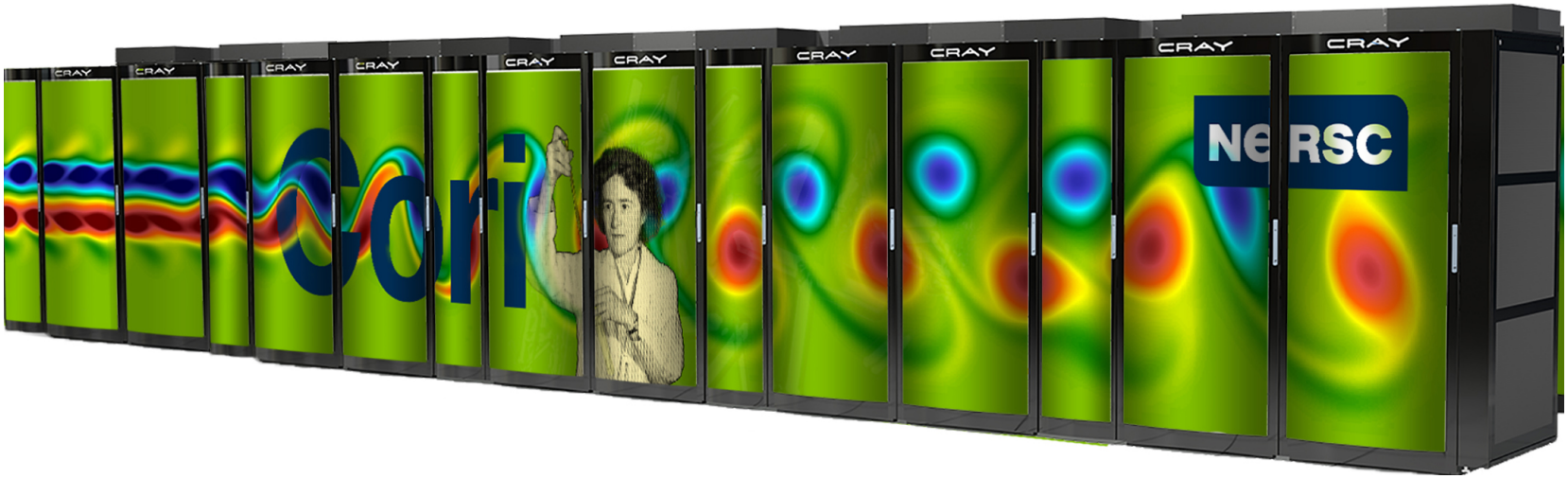
Model	TH-IVB-FEP
Nodes	16,000
Processor	Intel Xeon IvyBridge E5-2692
Speed	2.200 GHz
Sockets per Node	2
Cores per Socket	12
Coprocessers	Intel Xeon Phi 31S1P
Coprocessors per Node	3
Cores per Coprocessor	57
Coprocessors total	48,000
Operating System	Kylin Linux
Primary Interconnect network	Proprietary high-speed interconnecting (TH Express-2)
Peak Power (MW)	17,8
Size of Power Measurements (Cores)	3,120,000
Memory per Node (GB)	64

Summary of all components

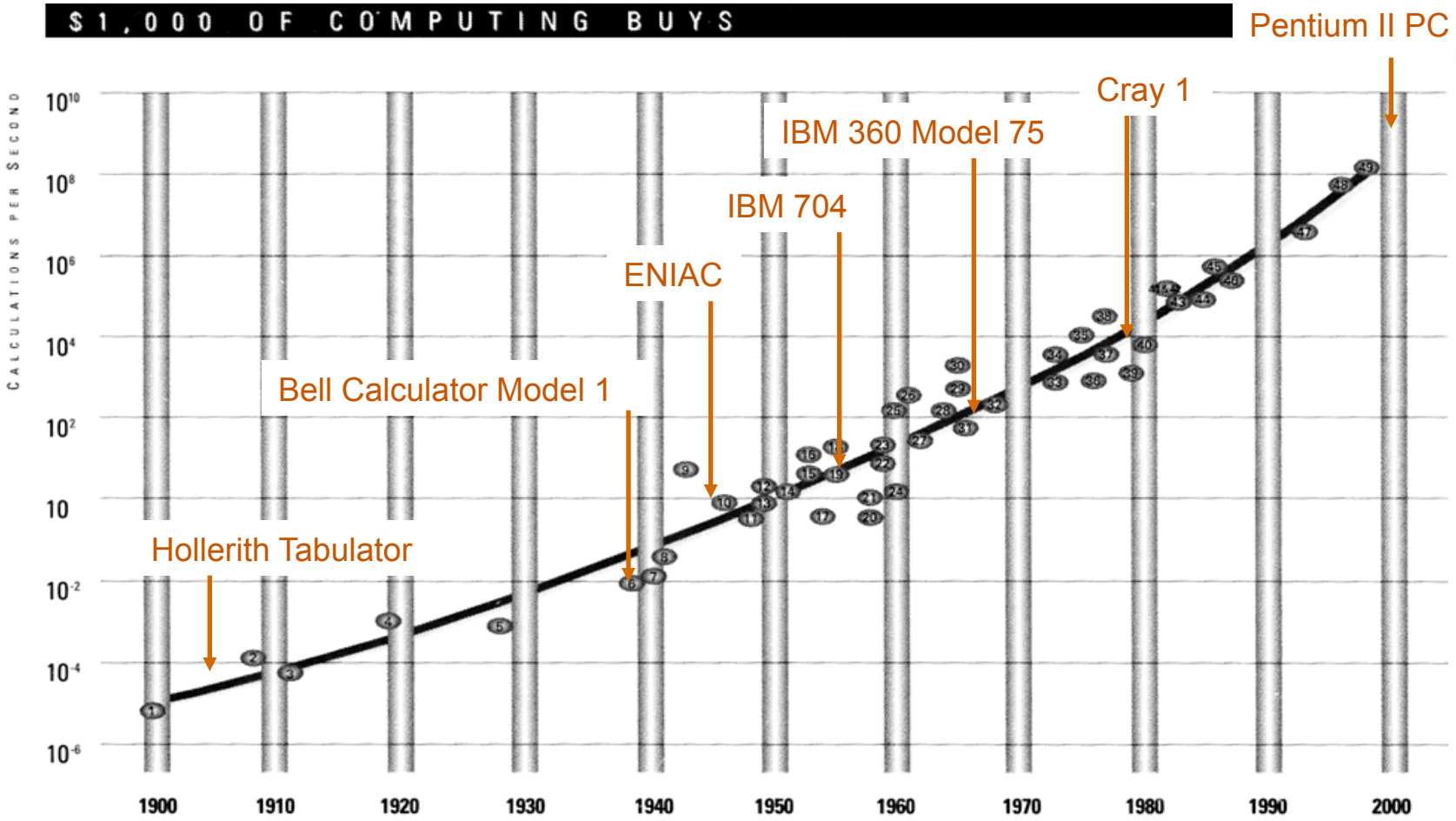
CPU Cores	384,000
Accelerators/CP	48,00
Accelerator/CP Cores	
Memory	1,024,000 GB

Summary of the Tianhe-2 (TH-2) or Milkyway-2	
Items	Configuration
Processors	32,000 Intel Xeon CPU's + 48,000 Xeon Phi's (+4096 FT-1500 CPU's frontend) Peak Performance 54.9 PFlop/s (just Intel parts)
Interconnect	Proprietary high-speed interconnection network, TH Express-2
Memory	1 PB
Storage	Global Shared parallel storage system, 12.4 PB
Cabinets	125 + 13 + 24 = 162
Power	17.8 MW
Cooling	Closed air cooling system

Two Supercomputers in Berkeley



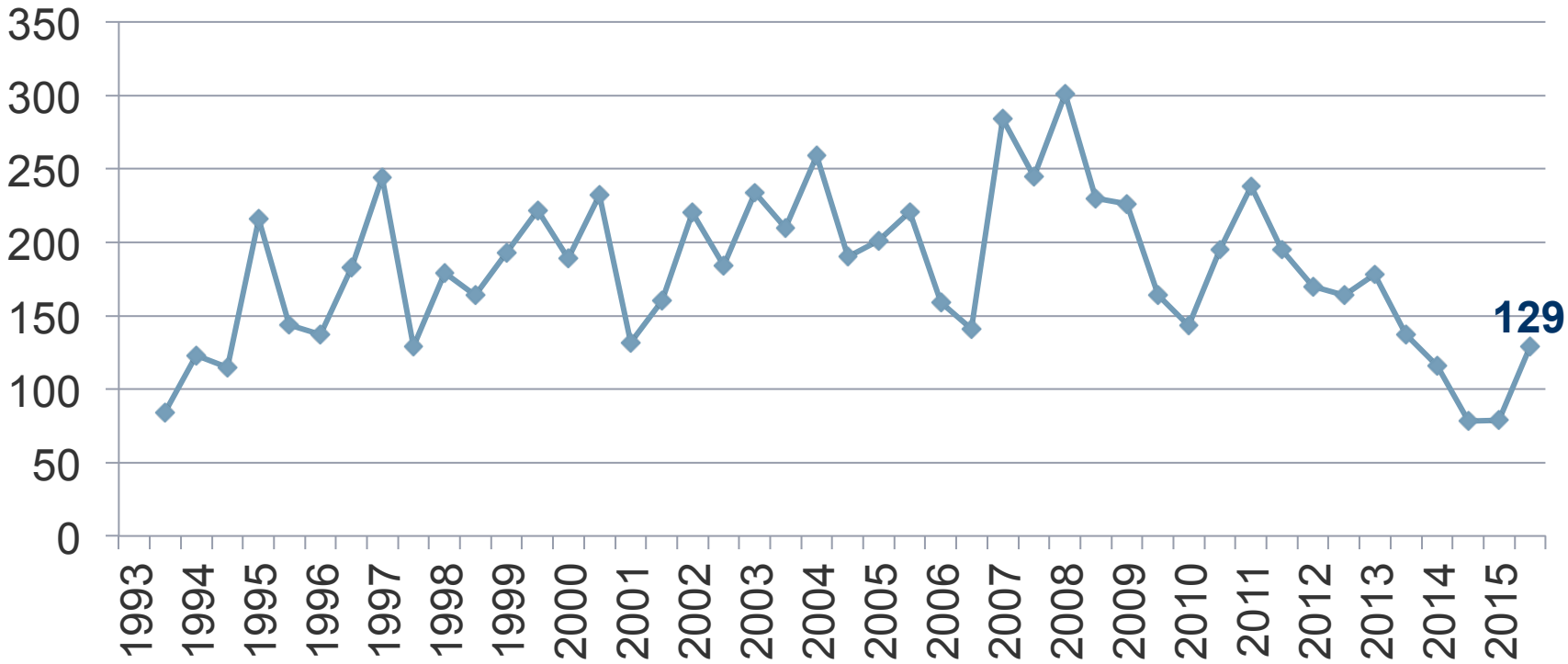
The Exponential Growth of Computing



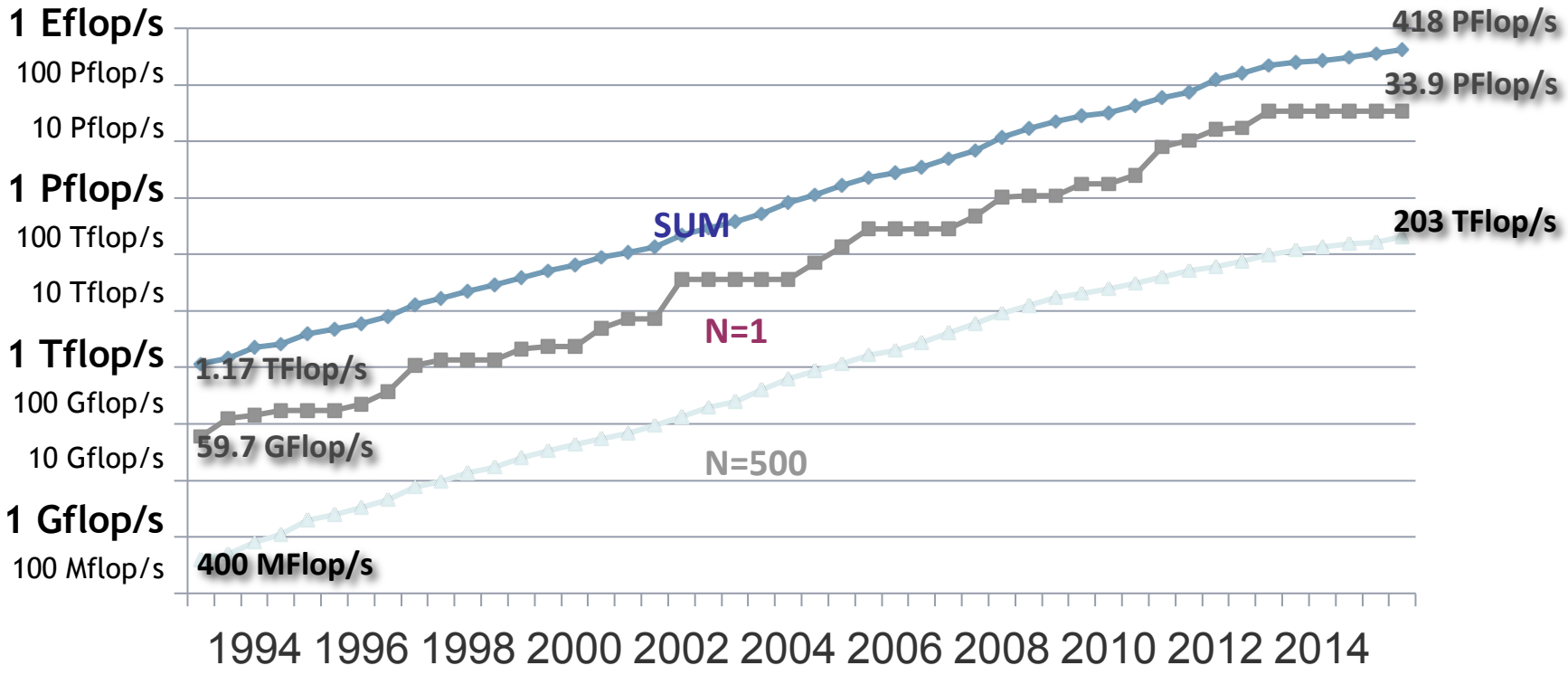
Adapted from Kurzweil, *The Age of Spiritual Machines*

BERKELEY LAB

Replacement Rate

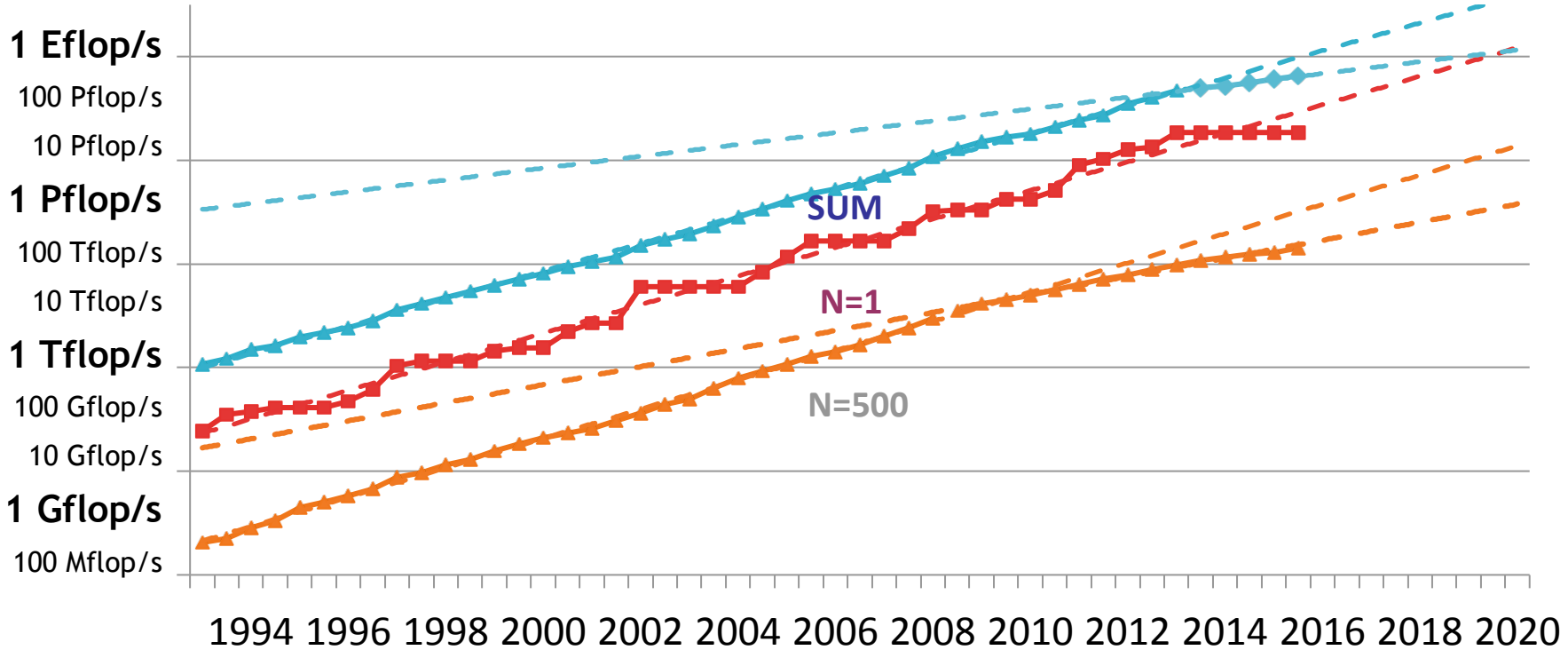


Performance Development



BERKELEY LAB

Projected Performance Development



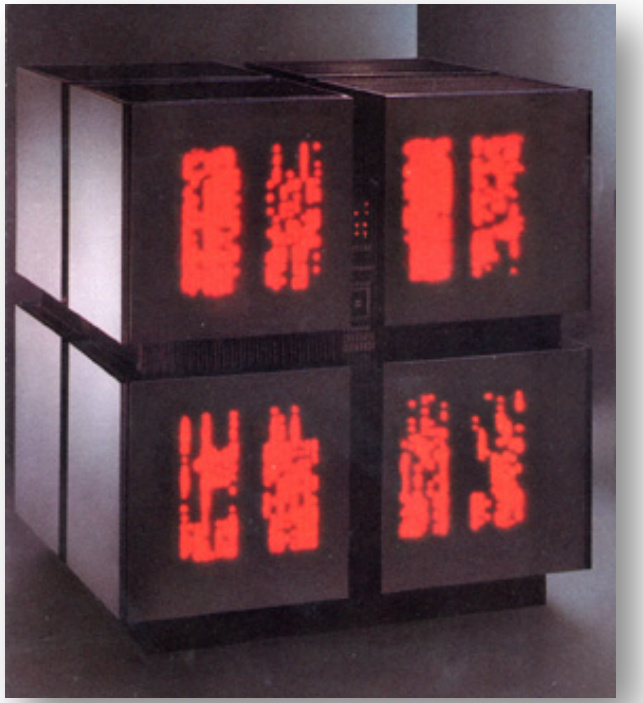
BERKELEY LAB

History Lesson: 1987

"Legendary" CM-2 by Thinking Machines

Architecture evolved into CM-5 (1992) built as MPP for scientific applications

Early history of AI applications on parallel platforms has been lost



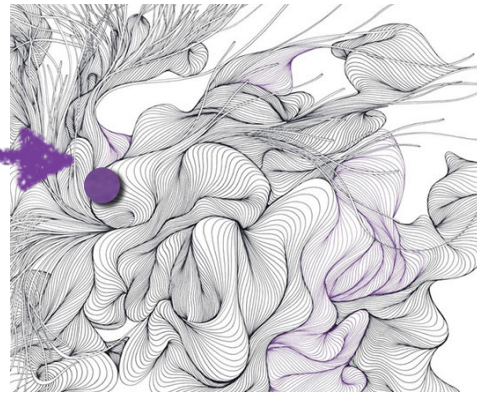
History Lesson: 2011

- IBM Watson beats two best human players
- Still only a narrowly defined game, but Watson demonstrates significant progress in language processing
- Uses supercomputer architecture
- IBM is developing commercial applications based on Watson
- No impact on HPC

DOE / ASCR ➜ *BRAIN* Initiative?

Neuroscience is a vast field

- BRAIN Initiative is a small part
- Driven by measurement, technology and computing



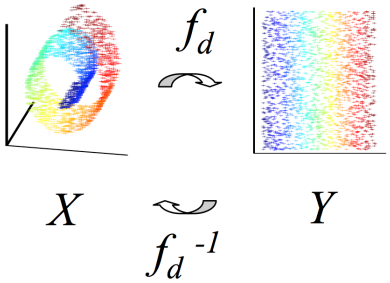
DOE National Labs would bring

- Team science + interdisciplinary integration
- Systems engineering
- Facilities and problems of scale

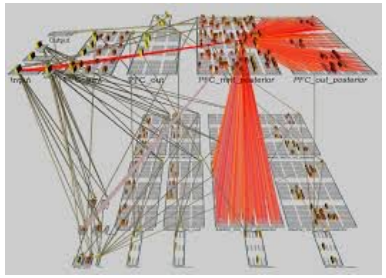
ASCR would bring

Data Management

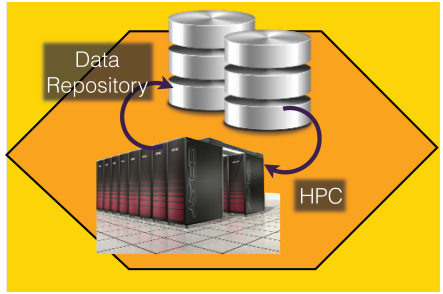
Analysis Methods



Theory and Models



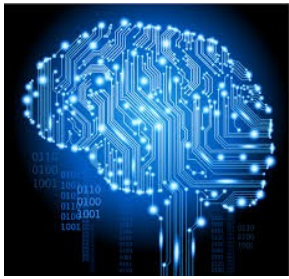
HPC Facilities



BERKELEY LAB

Computing for BRAIN is synergistic with ASCR strategic directions

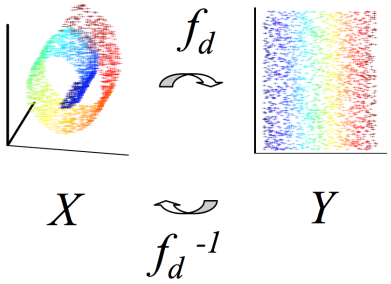
Computing requirements for BRAIN are well aligned with DoE/ASCR data strategy. Investment in applied mathematics and computer science will be synergistic.



Data Management

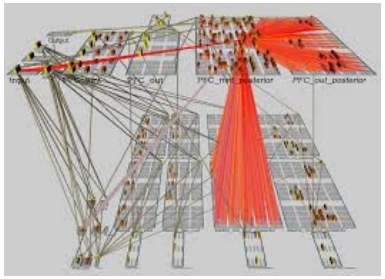
Devise standardizations and models for curation, provenance-tracking, and fusion of multi-modal brain data.

Analysis Methods



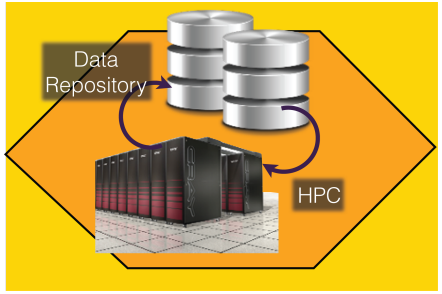
Develop modality agnostic data analytics methods and visualizations to reveal structure of brain data.

Theory and Models



Construct rigorous mathematical theories and simulations to bridge multiple spatio-temporal scales.

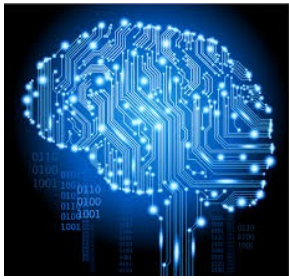
HPC Facilities



Provide community access to centralized data repositories and high-performance computing facilities.

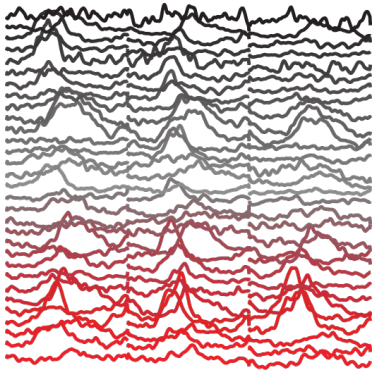
ASCR can uniquely contribute to BRAIN

ASCR can play a unique role in BRAIN computing through advances in applied mathematics and computer science together with HPC facilities.



Function

dynamic data

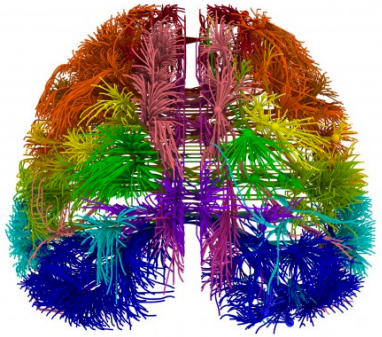


Theory & Models

abstractions

Structure

static data

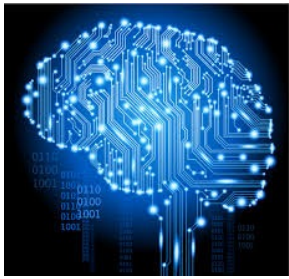


Generation and analysis of raw data

Linking structure to function is a ‘grand challenge’ in general biology and materials

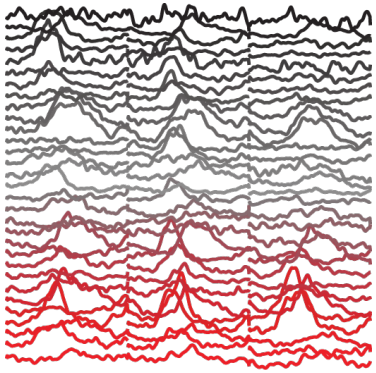
Integration and Synthesis are Crucial

Persistent, open (Web-based) access to multi-modal data to allow the neuroscience community to perform exploratory analysis for data driven discovery (**Data Superfacility**).



Function

dynamic data

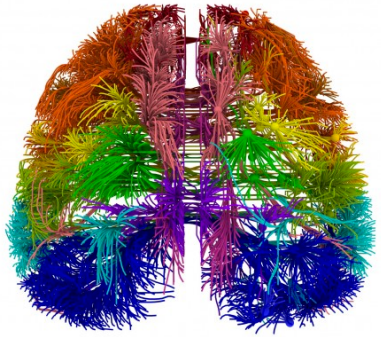


Theory & Models

abstractions

Structure

static data



Derived products

Integration & Synthesis

BRAIN Initiative

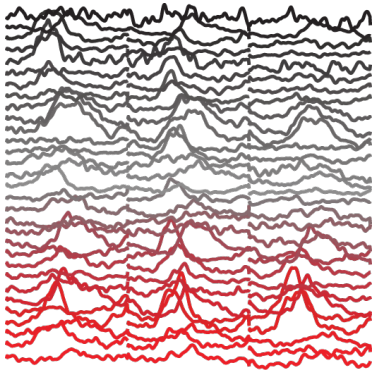
#3. **The brain in action**: Produce a *dynamic picture* of the functioning brain by developing and applying improved methods for large-scale monitoring of neural activity.

#5. **Identifying fundamental principles**: Produce conceptual foundations for understanding the biological basis of mental processes through *development of new theoretical and data analysis tools.*

#2. **Maps at multiple scales**: Generate circuit diagrams that *vary in resolution* from synapses to the whole brain.

Function

dynamic data

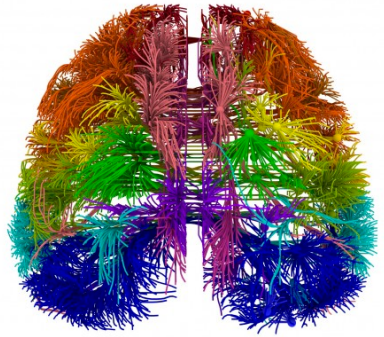


Theory & Models

abstractions

Structure

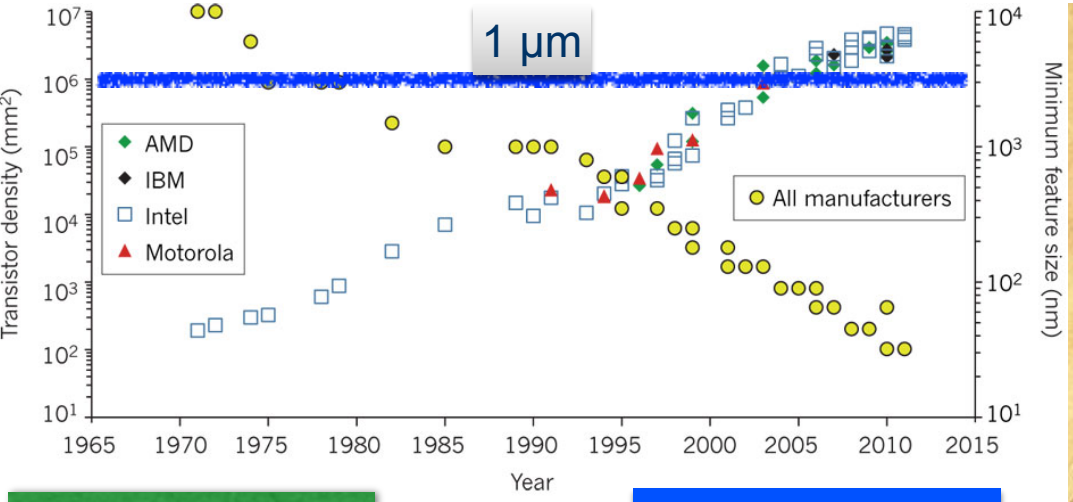
static data



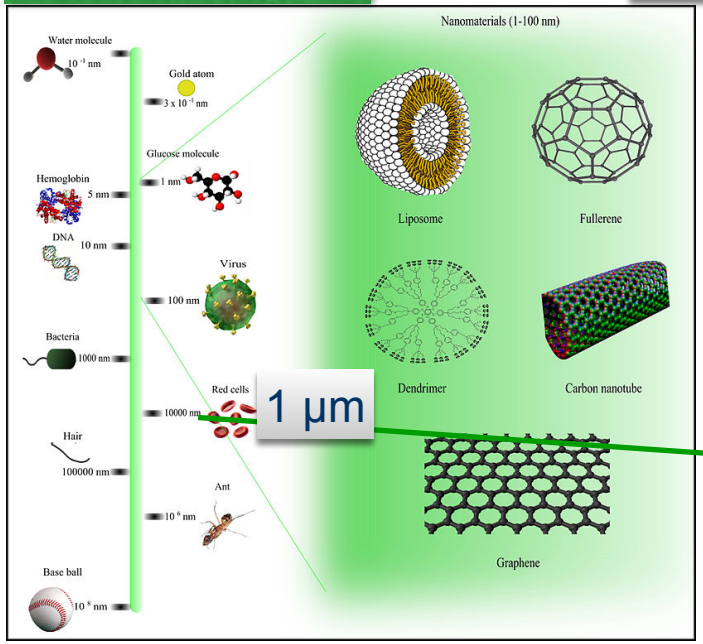
Derived products

Integration & Synthesis

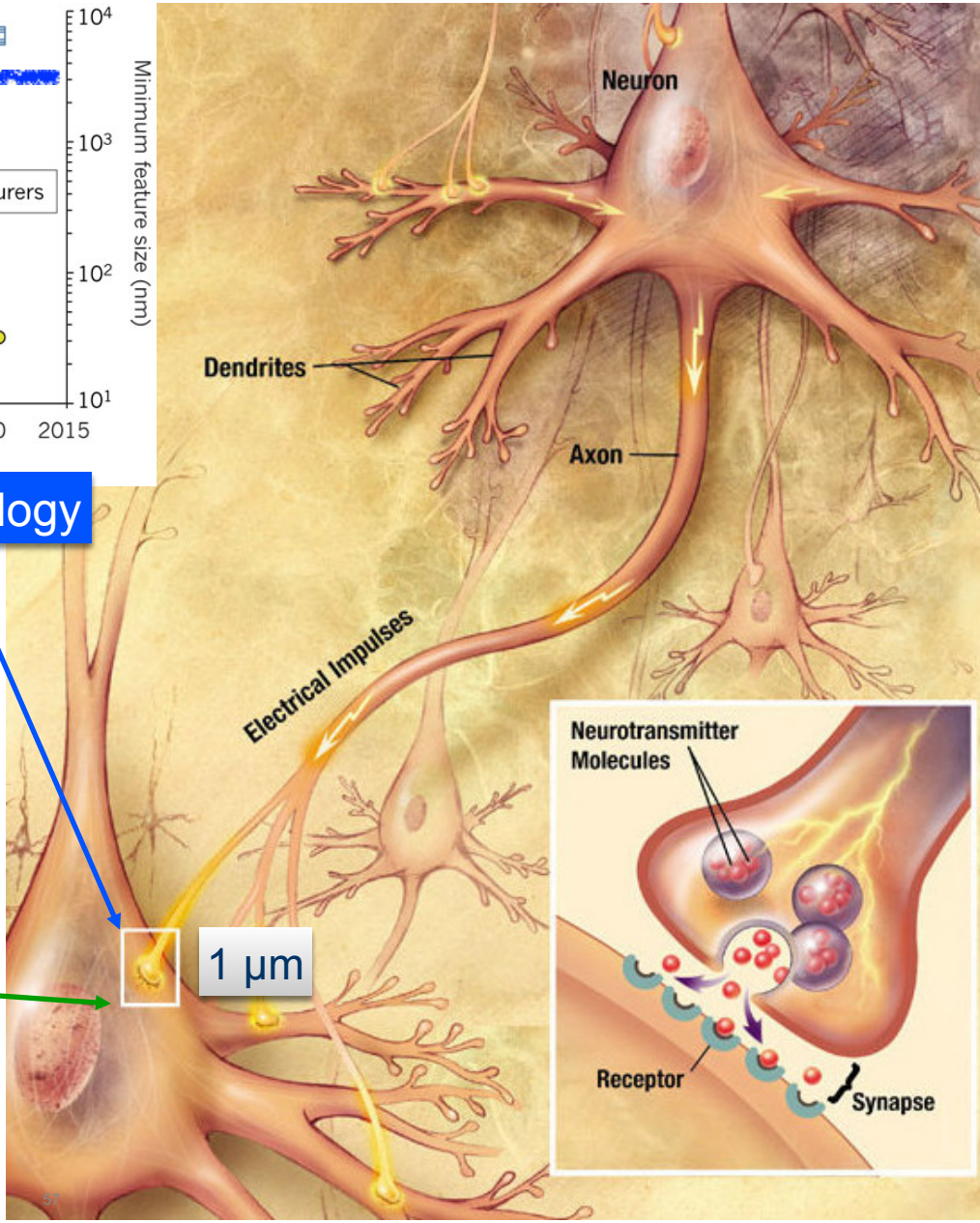
Technology to the rescue



Nanotechnology



Microtechnology



Tri-Institutional Partnership - *a* model

UCSF
University of California
San Francisco

- Clinical
- Neuroscience

Berkeley
UNIVERSITY OF CALIFORNIA

- Neuroscience
- Technology

BERKELEY LAB
Lawrence Berkeley National Laboratory

- Technology, HPC
- Scale

Tri-Institutional Partnership
BRAIN R&D Initiative to support innovative neurotechnology

BERKELEY LAB Lawrence Berkeley National Laboratory | Berkeley UNIVERSITY OF CALIFORNIA | UCSF University of California San Francisco

HOME | ABOUT | SCHEDULE | HOW TO APPLY | NEWS & ANNOUNCEMENTS

Home

NEWS & ANNOUNCEMENTS

New Partnership Launches with Meeting at Berkeley Lab
April 1, 2014

National Lab
Facilities
Capabilities

Universities
Neuroscience
Clinical data

To unite the research strengths of Berkeley Lab, UC Berkeley and UC San Francisco, the three institutions [illegible] Tri-Institutional Partnership as a means to promote co[illegible]utions.

The partnership's first venture will seed collaborative research projects in neurotechnology. The p[illegible] earlier this year with the announcement of a peer-reviewed co[illegible] the participating institutions to catalyze bold, potentially transformative research in neurotechnology at scale.

O[illegible] from all three partners convened at Berkeley Lab for "Proposers' Day," to learn more about the seed program, the opportunities provided by President Obama's BRAIN initiative, and especially to share their research ideas and forge new collaborations.

"Bringing people together across disciplines is difficult, and across institutions even more so," said Graham Fleming, Vice Chancellor for Research at Berkeley. "But the impr[illegible]all three institutions and the energy in the room demonstrates the level of excitement they share in addressing the neurotechnology challenges posed in the BRAIN Initiative."

"I[illegible]ppy to see such a tremendous turnout for our first Proposer's Day. Our scientists were joined by 23 scientists from UC SF and 28 from UC Berkeley," said B[illegible] is demonstrates to me that the Bay Area scientific community is ready for the tri-institutional partnership in order to address new and interdisciplinary scientific challenges."

S[illegible]m will be calibrated on technical excellence, innovation, and the substantive involvement of the collaborative partners across multiple disciplines. To ensure impact, each project must have a clear path from concept to the development of a competitive proposal for outside funding.

Go here to learn more about the Tri-Institutional Partnership and the BRAIN R&D seed-funding project to support innovative neurotechnology.

Challenges of **brain simulation**: link structure to function across scales

Disparate spatiotemporal scales

- nanometers to meters
- picoseconds to years

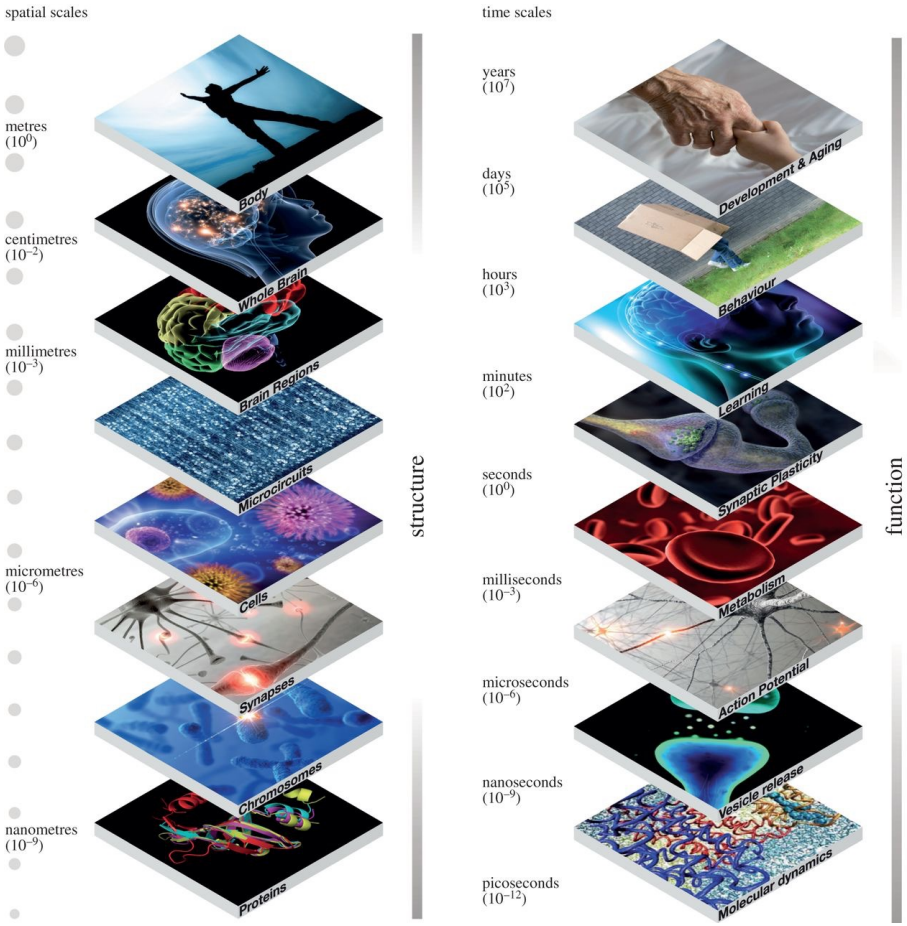
Diverse data types

- genomics to (functional/structural) connectomics
- electrical, optical and other measurements
- behavior, sensory stimuli, perturbations

Complex analysis issues

- fusion of multi-modal data
- inference robust to noise
- provide insight into computations
- statistical prediction of future events
- extraction of important features

Lack of data and models



Human Brain Project, 2012

LBL *BRAIN* Related/Motivated Research

LBL LDRD

Computation

+Machine Learning (K. Bouchard)

+Data model (BRAINformat) (O. Ruebel)

+Graph Analytics (A. Buluc)

+Real-Time Processing & Vis (D. Donofrio & G. Weber)

+Neural Networks (K. Bouchard)

+Neuron Reconstruction at NERSC (Prabhat)

+fMRI analysis (D. Ushizima)

+Neuromorphic Computing (2 LDRDs, FY' 16)

Technology

+High-density electrophysiology (P. Denes)

+Up-Converting Nanoparticles (B. Cohen)

+Opto-acoustic waveguides (P. Schuck)

+Chemical Sensors (C. Chang)

ALLEN INSTITUTE for BRAIN SCIENCE *Fueling Discovery*

IBM

Cal-BRAIN A NEUROTECHNOLOGY PROGRAM FOR CALIFORNIA

Experimentation

+Neurodegeneration (C. McMurray)

+Neurological Aging (W. Jagust)

+Sensorimotor circuits (K. Bouchard)

+Neurocognitive Resilience (A. Wyrobek)

+Toxicant impact on host and microbiome (S. Celniker, A. Snijders, J-H Mao)

+Molecular Basis of Epilepsy (B. Brown)

NIH National Institutes of Health *Turning Discovery Into Health*

Facilities

NERSC

MOLECULAR FOUNDRY

ESnet Energy Sciences Network

ALS ADVANCED LIGHT SOURCE

Tri-Institutional Partnership

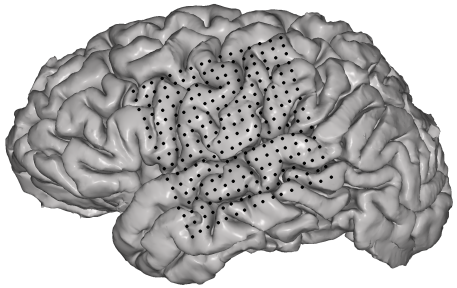
UCSF BERKELEY LAB Berkeley UNIVERSITY OF CALIFORNIA Lawrence Berkeley National Laboratory

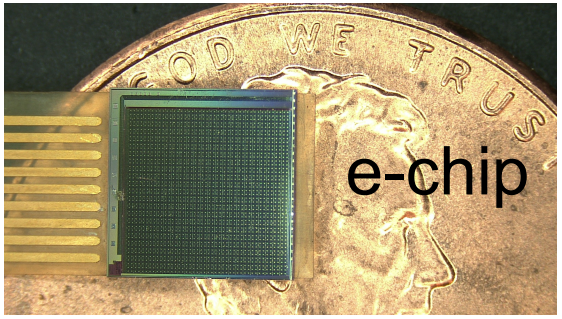
BERKELEY LAB

Electrical Sensors and Optical Emitters for *BRAIN*

Ultra high-density electrophysiology

Go from 10^2 ➜ 10^5 electrodes



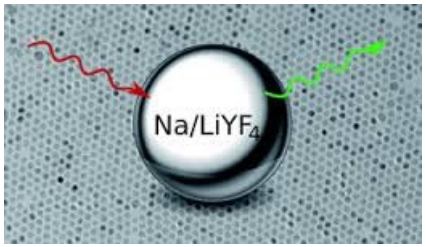


Data volume

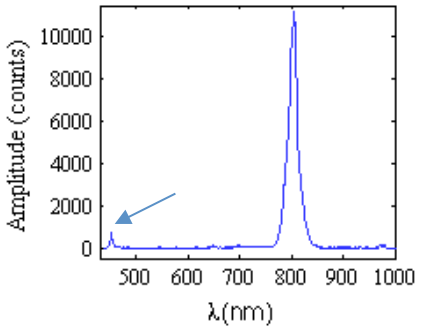
- Today: at limit of "workstation"
- ↑ 3-4 orders of magnitude

Nanocrystral "light-bulbs"

See deeper into brains

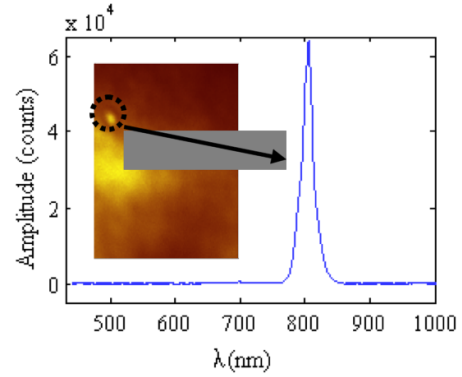


Up **C**onverting **NanoP**articles



Typical Tm/Yb UCNP spectrum

Use blue emission for **optogenetic stimulation**

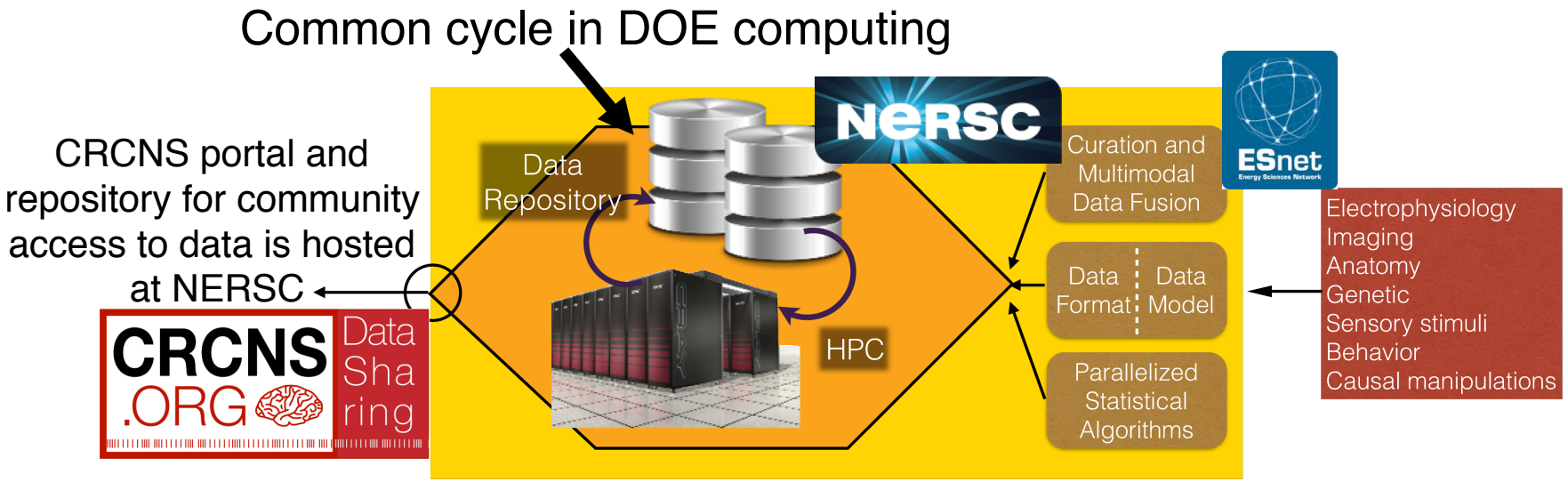


Through 500 µm of mouse brain

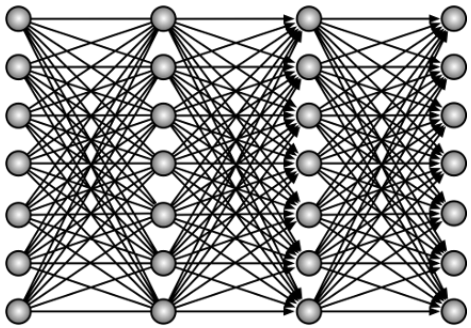
Use red emission for **optical recording**

MOLECULAR FOUNDRY

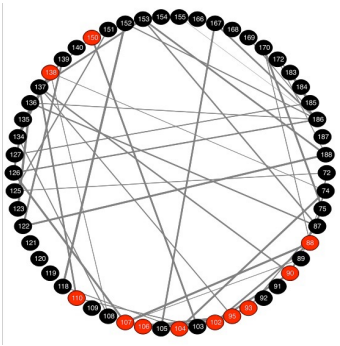
Advanced Computing for **BRAIN** at LBL

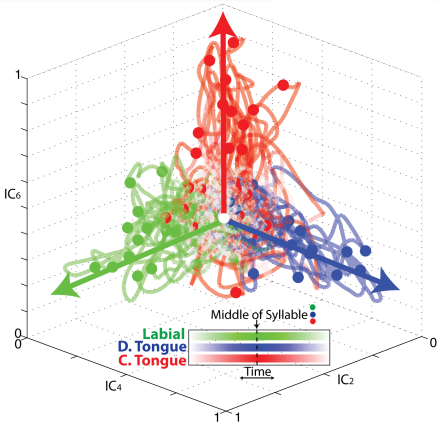


Deep neural networks for decoding brain activity

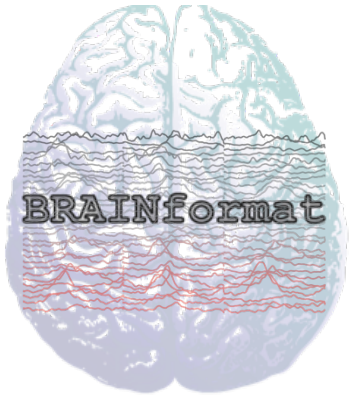


Sparse neural activity for human speech production





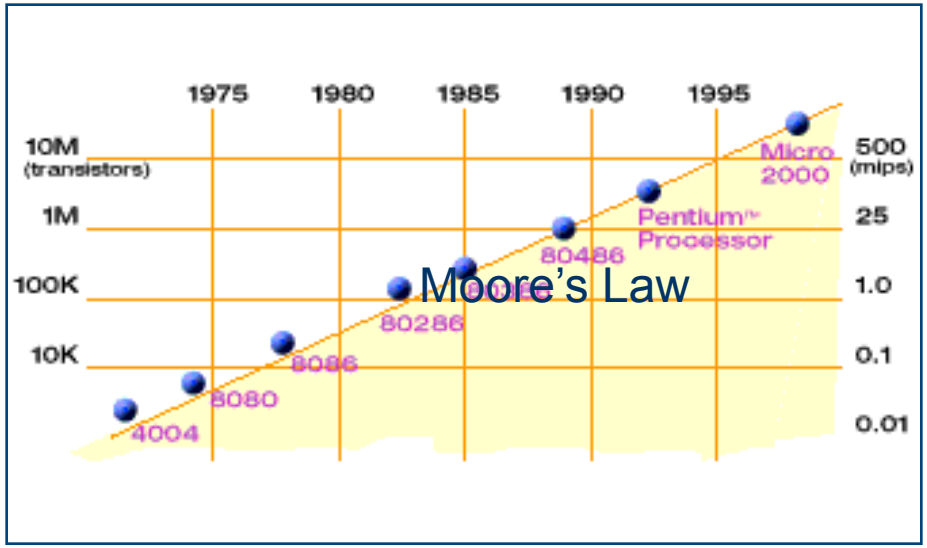
HDF5 format and data model for HPC

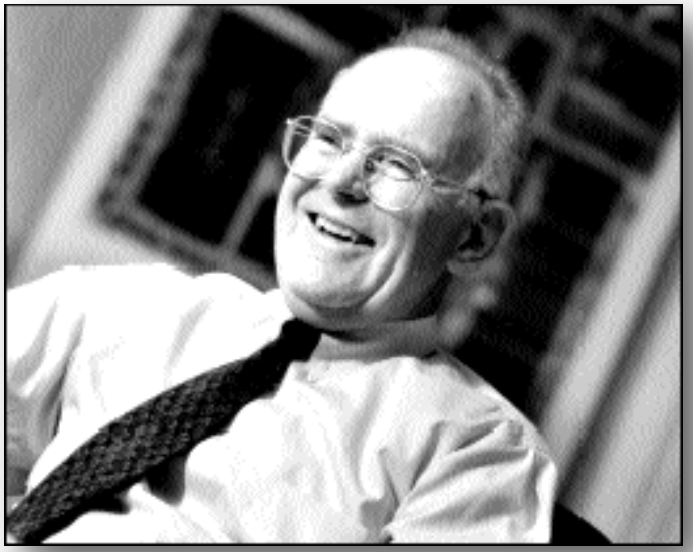


Summary

- Proponents of the rapid development of superintelligence use straightforward extrapolation of current computer performance. This ignores the end of Moore's Law, and the multidimensional nature of human intelligence.
- Simulating the human cortex in real time will require a system with 10 Exaflops, 5 Petabytes,and 20 MW.
- With the human brain taking only 20 W, current technology is at least a factor of a billion away from human brain performance.
- We must investigate new architectures if we want to close this gap.
- We must advance various "brain initiatives" to get the necessary data.

Technology Trends: Microprocessor Capability



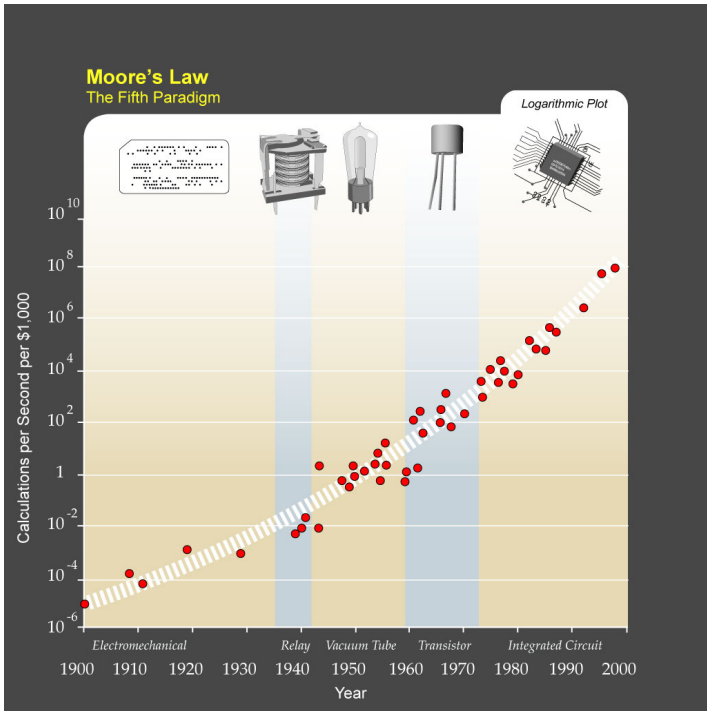


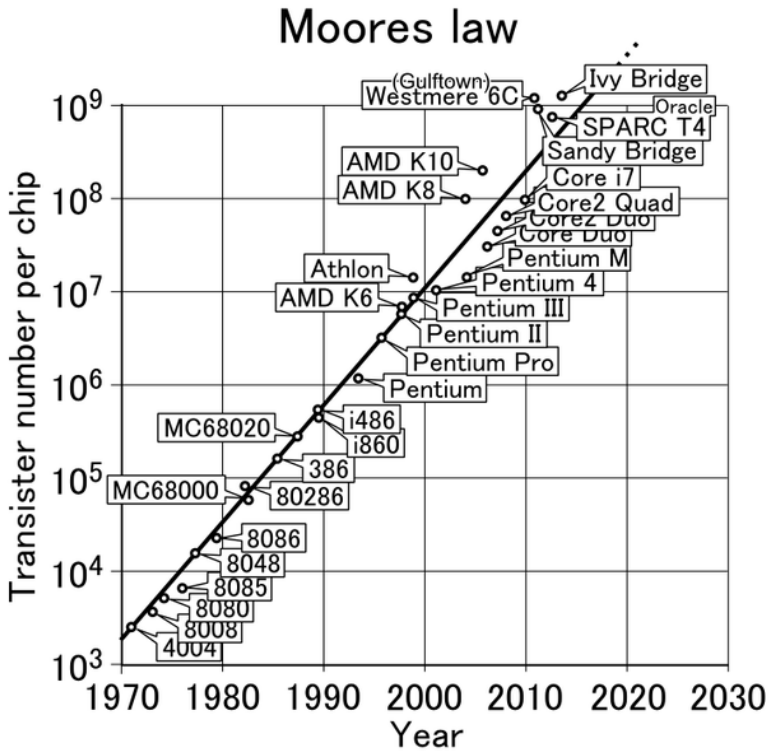
2X transistors/chip every 1.5 years – called "Moore's Law"

Microprocessors have become smaller, denser, and more powerful

Gordon Moore (co-founder of Intel) predicted in 1965 that the transistor density of semiconductor chips would double roughly every 18 months

Sustained Growth of Technology Allowed Us to Ignore Architecture





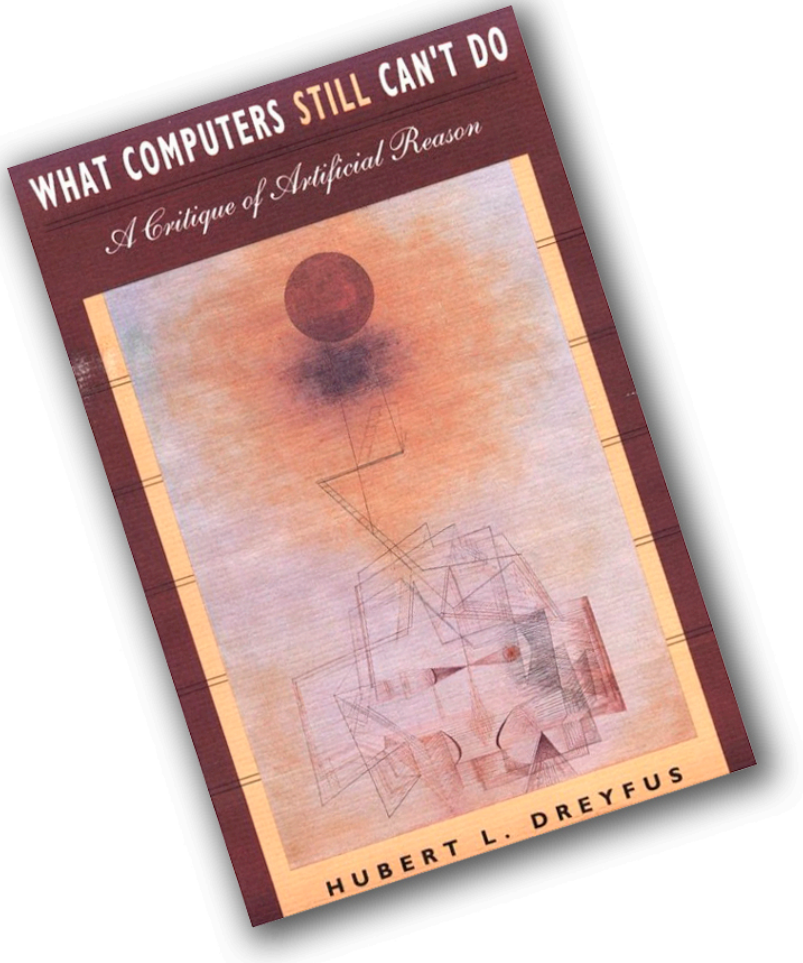
- Two technology transitions since 1940s
- Moore's Law stated for integrated circuits only
- Kurzweil et al. claim accelerated exponential growth across technologies

Motivation for the Title of My Talk

"The computer model turns out not to be helpful in explaining what people actually do when they think and perceive."

Hubert Dreyfus, pg.189

Example: one of the biggest success stories of machine intelligence, the chess computer "Deep Blue", did not teach us anything about how a chess grandmaster thinks



Why This Is Important

- This could be the beginning of the development of a “right-brain” architecture for computers
- The future is a fast pattern recognition machine. It is not relevant if the chips are actually resembling the brain
- Energy efficiency has been demonstrated with conventional 26nm process
- Small systems could be made available easily to wide developer community (think NVIDIA and CUDA, or Raspberry Pi for pattern recognition)
- Potential widespread use in mobile
- Scale up and integrated into HPC.