Microsoft

EXTREME COMPUTING GROUP

Defining the future.

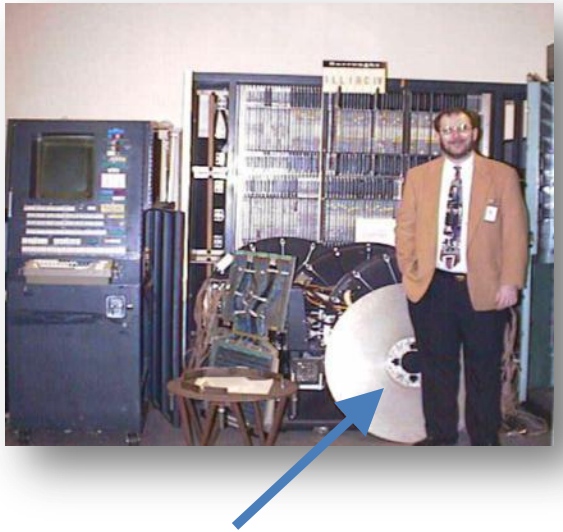
Parallel Computing: Insights for the Future

Dan Reed
Corporate Vice President
Extreme Computing Group &
Technology Strategy and Policy

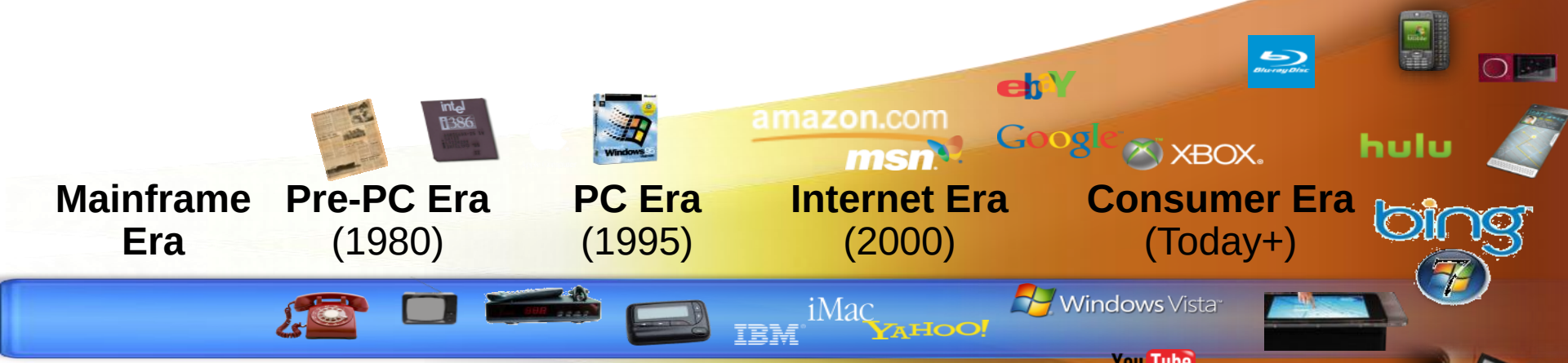
reed@microsoft.com
www.hpcdan.org

You're A Parallel Computing Geezer When …

- The Computer History Museum …
  - … brings back personal memories
- You toggled absolute binary into …
  - … the front panel and read the I/O lights
- You whistled into an acoustic coupler …
  - … and got a carrier signal
- You used an iron to smooth …
  - … a card deck after a thunderstorm



The Good News (So Far) …

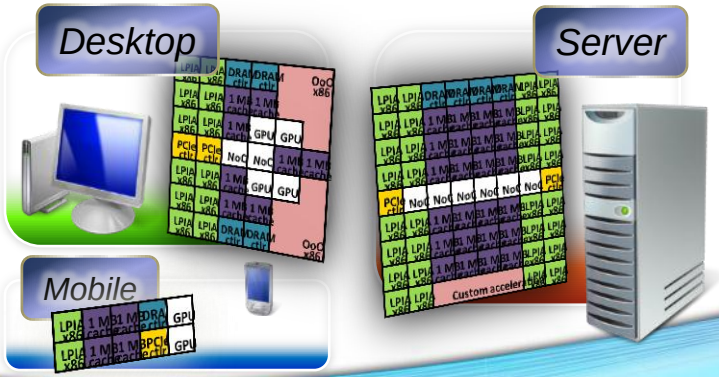


- 21st century implicit computing
  - Increasingly natural interfaces
  - Embedded intelligence
  - Number of cores/person → infinity
    - Consumer parallelism

Microsoft

What Has Changed?

- System on a chip designs
  - Powerful mobile devices
- Graphics processing units
  - High quality graphics
- Explosive data growth
  - Ubiquitous sensors and media
- Inexpensive embedded computing
  - Everyday smart objects, CIP, …
- Wireless spectrum pressure
  - Mobile device growth
- New software models
  - Social networks, clients+clouds …

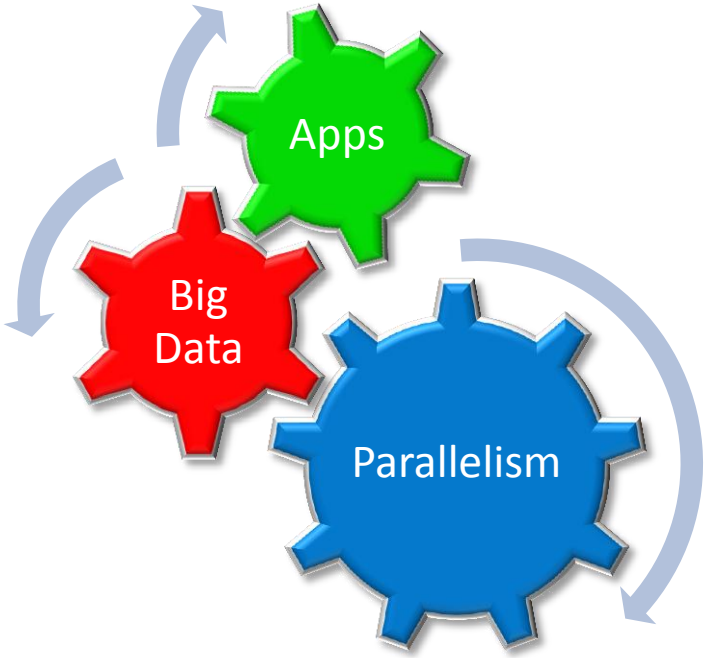


Paucity to Plethora

- Paucity drives certain behaviors
  - Hoarding, conservatism, limitation
- Plethora also drives certain behaviors
  - Speculation, risk taking, profligacy

- Psychologically we still believe in paucity
  - Optimizing for resource minimization
  - Rather than for productivity or simplicity

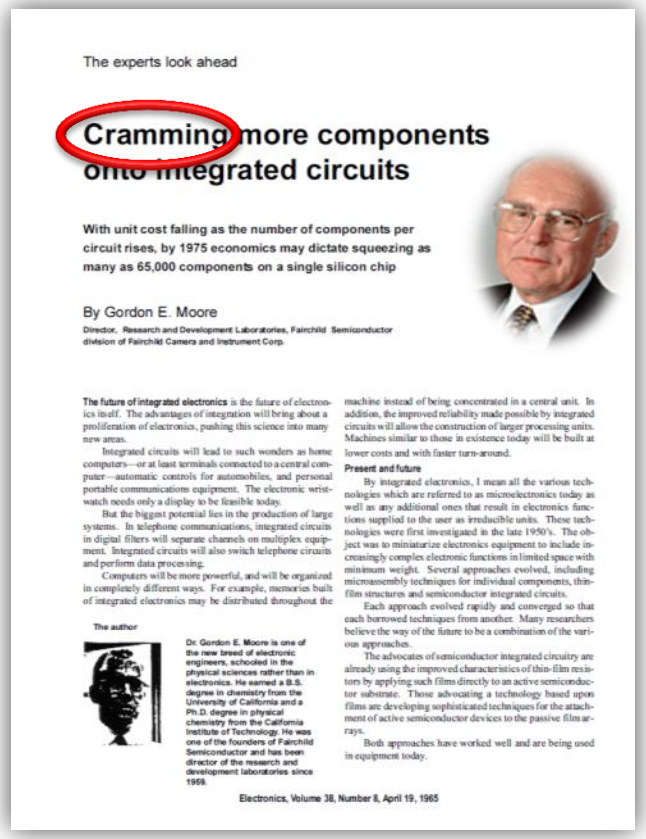
Takeaway Message: Multidisciplinary Convergence



"The future is here, it is just not evenly distributed."

William Gibson

Moore's "Law" and Limiting Exponentials …



The experts look ahead

Cramming more components onto integrated circuits

With unit cost falling as the number of components per circuit rises, by 1975 economics may dictate squeezing as many as 65,000 components on a single silicon chip

By Gordon E. Moore

Director, Research and Development Laboratories, Fairchild Semiconductor division of Fairchild Camera and Instrument Corp.

The future of integrated electronics is the future of electronics itself. The advantages of integration will bring about a proliferation of electronics, pushing this science into many new areas.

Integrated circuits will lead to such wonders as home computers—or at least terminals connected to a central computer—automatic controls for automobiles, and personal portable communications equipment. The electronic wristwatch needs only a display to be feasible today.

But the biggest potential lies in the production of large systems. In telephone communications, integrated circuits in digital filters will separate channels on multiplex equipment. Integrated circuits will also switch telephone circuits and perform data processing.

Computers will be more powerful, and will be organized in completely different ways. For example, memories built of integrated electronics may be distributed throughout the machine instead of being concentrated in a central unit. In addition, the improved reliability made possible by integrated circuits will allow the construction of larger processing units. Machines similar to those in existence today will be built at lower costs and with faster turn-around.

Present and future

By integrated electronics, I mean all the various technologies which are referred to as microelectronics today as well as any additional ones that result in electronics functions supplied to the user as irreducible units. These technologies were first investigated in the late 1950's. The object was to miniaturize electronics equipment to include increasingly complex electronic functions in limited space with minimum weight. Several approaches evolved, including microassembly techniques for individual components, thin-film structures and semiconductor integrated circuits.

Each approach evolved rapidly and converged so that each borrowed techniques from another. Many researchers believe the way of the future to be a combination of the various approaches.

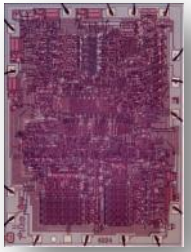
The advocates of semiconductor integrated circuitry are already using the improved characteristics of thin-film resistors by applying such films directly to an active semiconductor substrate. Those advocating a technology based upon films are developing sophisticated techniques for the attachment of active semiconductor devices to the passive film arrays.

Both approaches have worked well and are being used in equipment today.

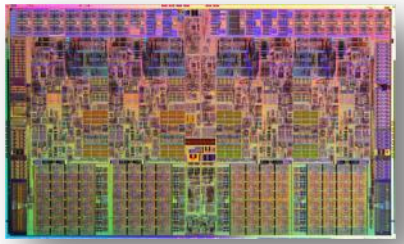
The author

Dr. Gordon E. Moore is one of the new breed of electronic engineers, schooled in the physical sciences rather than in electronics. He earned a B.S. degree in chemistry from the University of California and a Ph.D. degree in physical chemistry from the California Institute of Technology. He was one of the founders of Fairchild Semiconductor and has been director of the research and development laboratories since 1959.

Electronics, Volume 38, Number 8, April 19, 1965

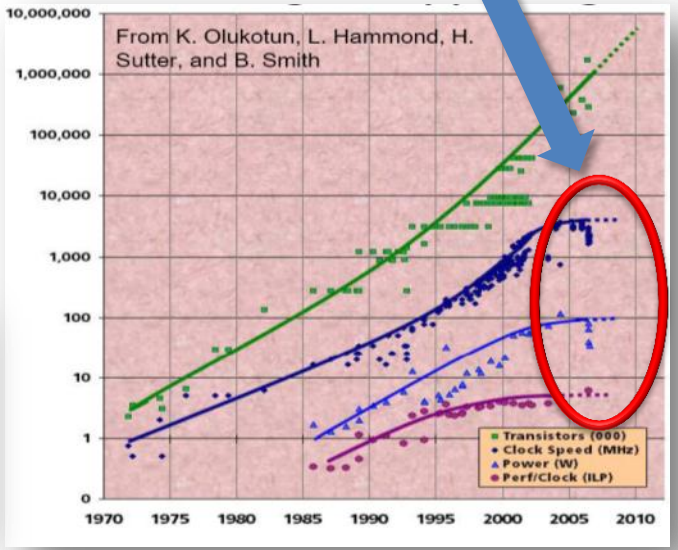


Intel 4004



Intel Core i7

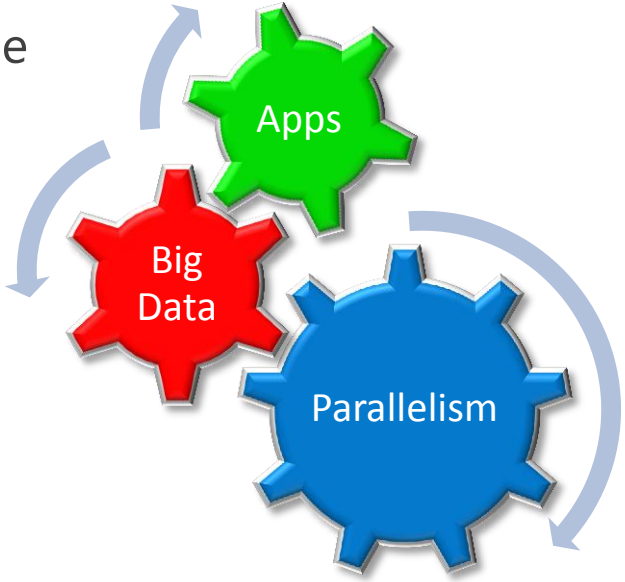
Trouble in River City



Microsoft

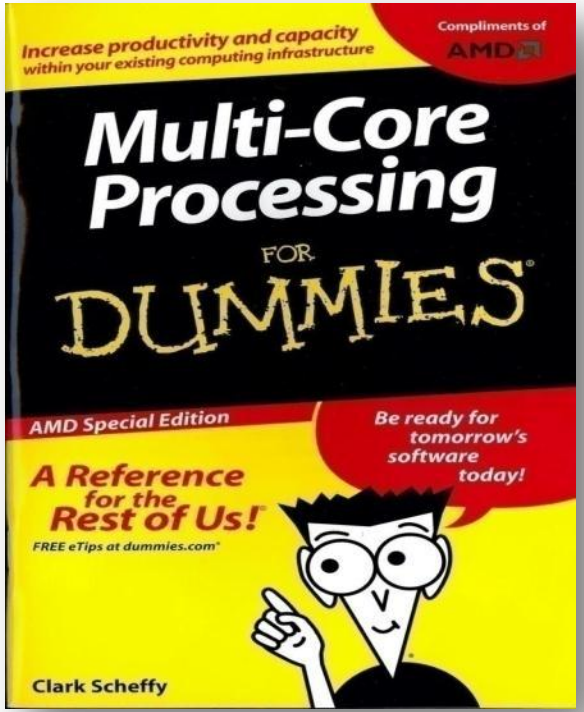
Manycore Challenges The Extant Ecosystem

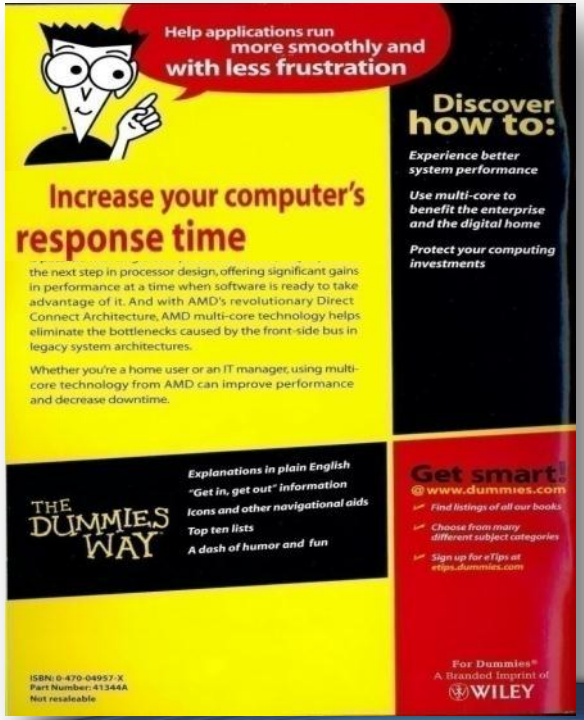
- The free lunch is over
  - Software acceleration via technology alone
  - Clients, servers and infrastructure
- "Surrounded by opportunities"
  - Devices and architectures
  - Programming models and abstractions
  - Algorithms and applications



- From challenge comes opportunity …
  - Old approaches will change or die
  - New applications and systems will arise
    - … including parallel computing

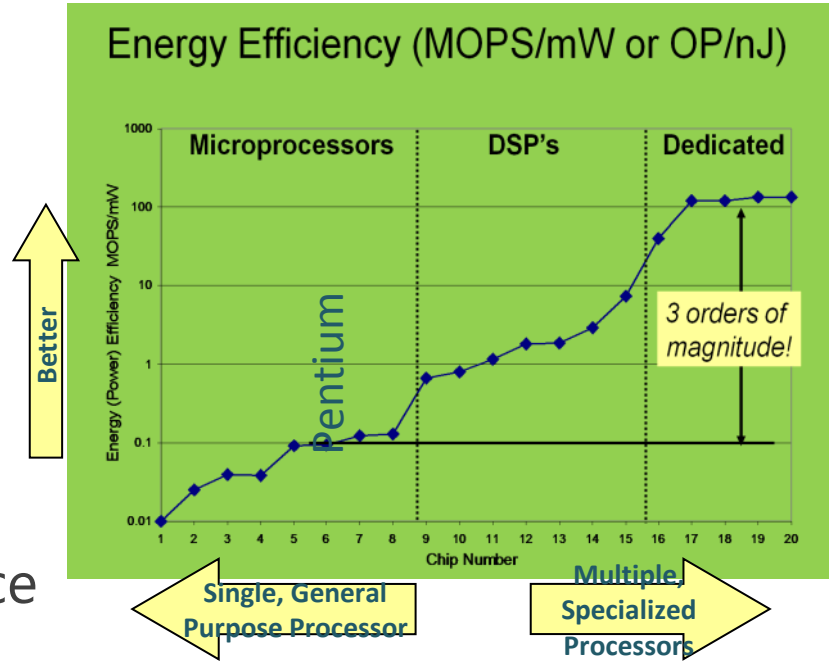
Some Confusion Out There





Rethinking Node Architecture

- Operations/joule
  - Low power, in order wins
- Memory-processor balance
  - Do not fixate on core counts
  - Optimize for workloads
    - TCO, not just ops or FLOPS
- System on a chip (SoC)
  - Learn from the embedded space
  - Embrace heterogeneity
    - Functional and performance



Mihai Budiu, "On The Energy Efficiency of Computation," February 2004

Manycore Resource Management

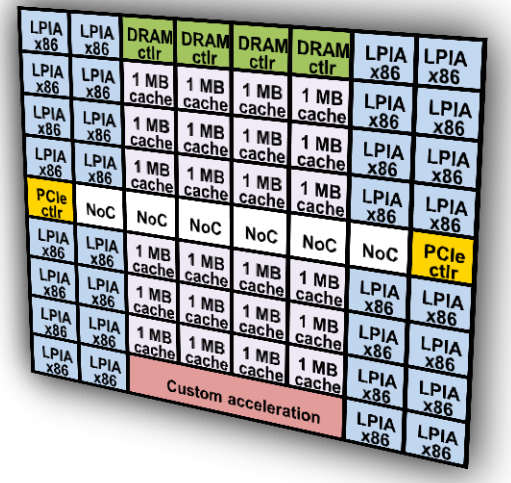
- Sapir–Whorf Hypothesis (SWH)
  - Language influences the habitual thought of its speakers
- The “CPU” model still shapes our world view
  - What is “central” in a mesh of cores?
  - Which is “central” in a heterogeneous manycore chip?
  - Why is a GPU a peripheral?
- Profound implications for software design
  - Resource management, security, …

Think Chocolates *And* Cookies

- Remember Amdahl's Law Speedup = $(S + (1-S)/N)^{-1}$
  - "Amdahl's Law in the Multicore Era," M. D. Hill and M. R. Marty, *IEEE Computer*, July 2008

- Sugar cookies alone
  - Similar, modulo process variation
  - You must eat lots to be satisfied
- Designer chocolates
  - Diversity is a feature
  - Forrest Gump was right

- Multicore implications
  - Legacy and new code
  - Programming heterogeneity
  - System software and services



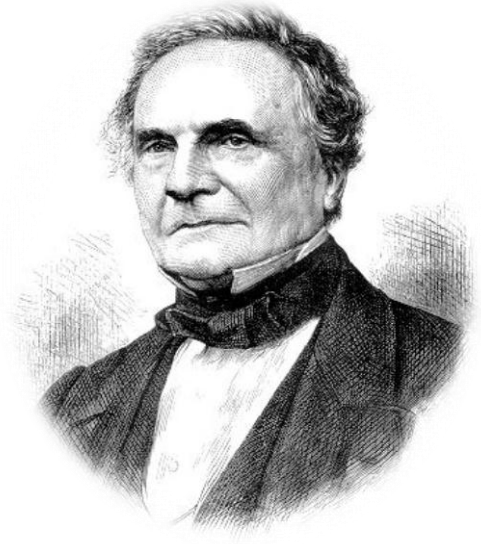
System on a Chip (SoC): The New Motherboard

- Standard building blocks
  - Core(s), memory controller, I/O
- Function-specific accelerators
  - Graphics, communications, sensors, security
- Internet of Things (IoT)
  - Embedded intelligence in everyday objects
  - Experiences and natural user interfaces (NUIs)
  - Resource discovery, security, services, programming

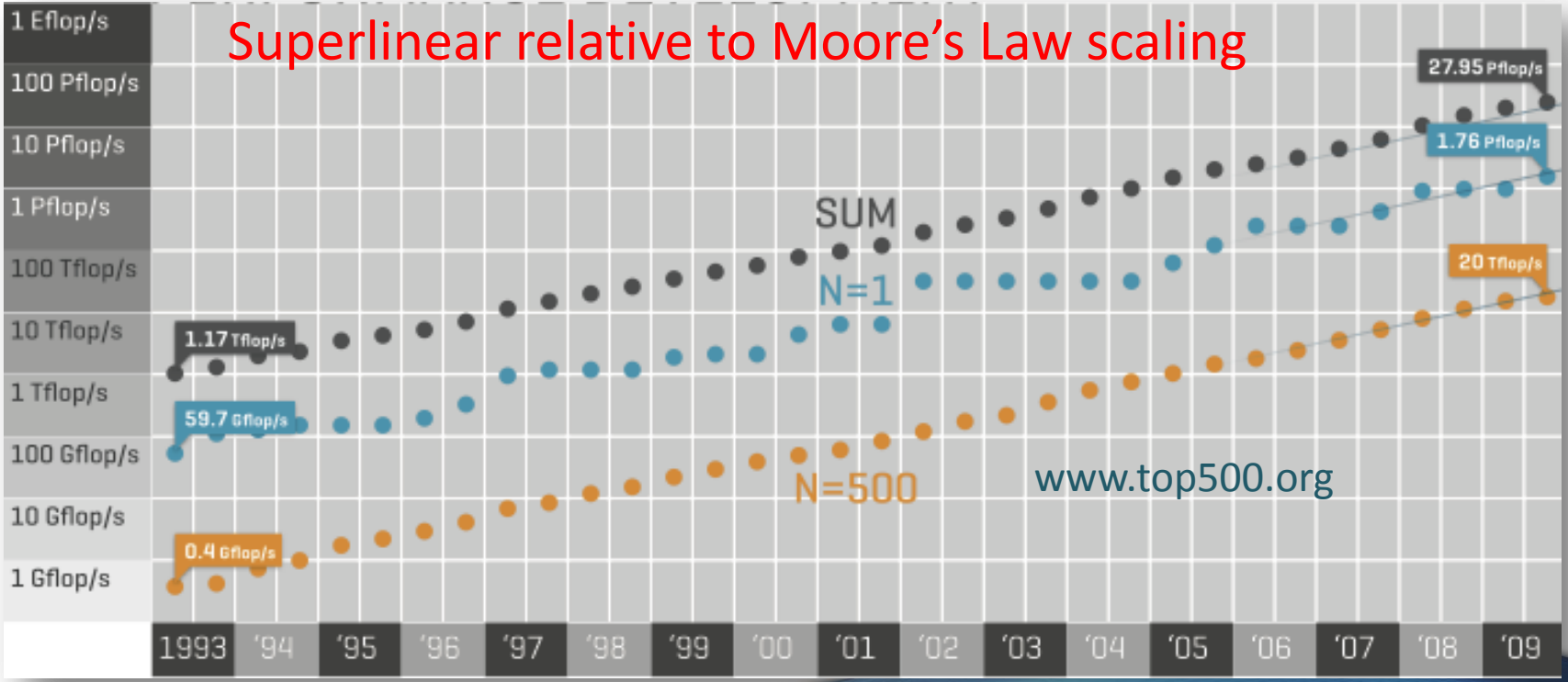
The Performance Psychology

"The most constant difficulty in contriving the engine has arisen from the desire to educe the time in which the calculations were executed to the shortest which is possible."

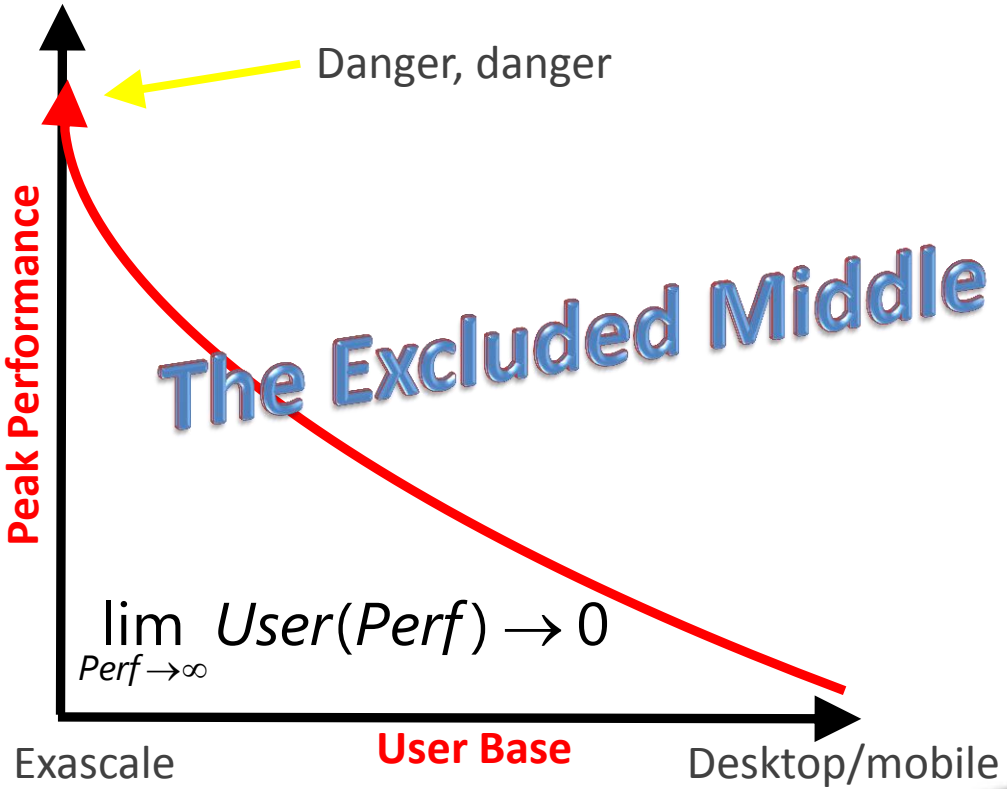
Charles Babbage

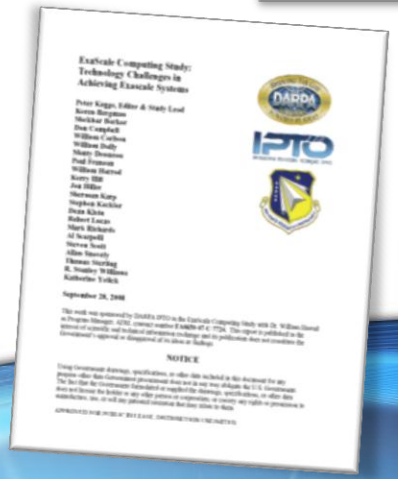


HPC: Teraflop to Petaflop

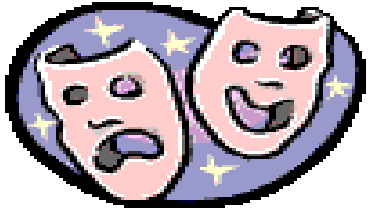


Exascale Exponentials (2018 Extrapolation)





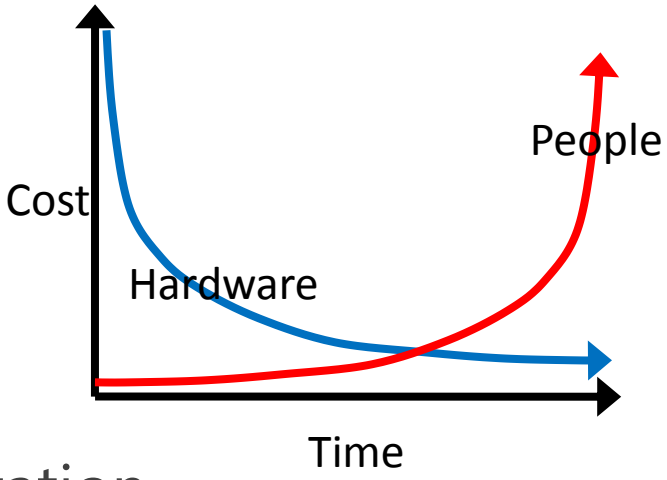
The Siren Call …



- We've seen parts of this movie before
  - Vector processors, systolic arrays, attached processors
- Success requires optimizing for efficiency
  - Data movement, computation and software costs
- Efficient exploitation, in two senses
  - Achieved application performance
    - Holistic assessment, not just application kernels
  - High human and scientific productivity
    - Extant software base, available tools
- We're geeks, we forget the human aspect
- We must raise the abstraction level …

Economic Divergence

- $/compute-year and $/storage-year
  - Declining rapidly even now
- $/developer-year
  - Rising, even in this economy
- High value need not imply high utilization
  - Rapid response changes behavior
- Applications
  - Outlive systems by many years
  - Are rising in complexity
  - Are increasingly multidisciplinary



Increasing Abstraction: Simple Is Good

- Successful technologies are invisible
  - They just work
- Easy beats complex
  - Good enough is, well, good enough
- Abstraction brings simplicity
  - Hiding details
- Simplicity usually means compromise
  - Full complexity is hidden

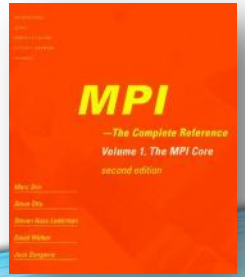
The Big Challenge

I want to build tools so powerful that full professors will use them, and so simple that they can.

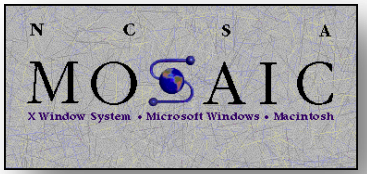
Fred Brooks (rough paraphrase)

Sociology and Community Divergence

- High-end technical HPC
  - Late 1980s
    - Message passing dominates
    - Custom hardware dominates
  - 1990s
    - Commodity clusters emerge
  - Currently
    - Commodity clusters + GPUs dominate
    - MPI dominates



- Distributed systems/services
  - September 1981
    - RFC 791
    - Internet Protocol (IP)
  - December 1993
    - Mosaic atop HTTP
  - Currently
    - Robust services
    - Vibrant industry



What Is an Application?

- An FFT?
  - No, it's an algorithm
- A rendering pipeline?
  - No, it's a software library
- A feature recognition system?
  - No, it's a building block

Microsoft Kinect

- Our notion of "application" is increasingly complex
  - Many integrated and interoperating components
- Our tools must enable creativity accordingly

Creative Empowerment

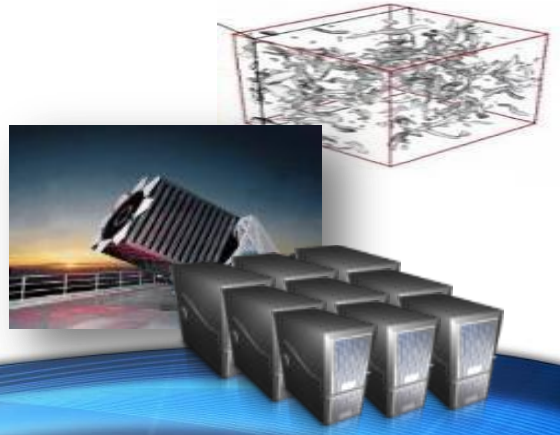
- Very few users love technology itself
  - Clusters and parallel programming
  - Distributed services, grids or clouds
  - Data models and databases

- Optimize for human creativity
  - Invisible and empowering
  - Plethora rather than paucity

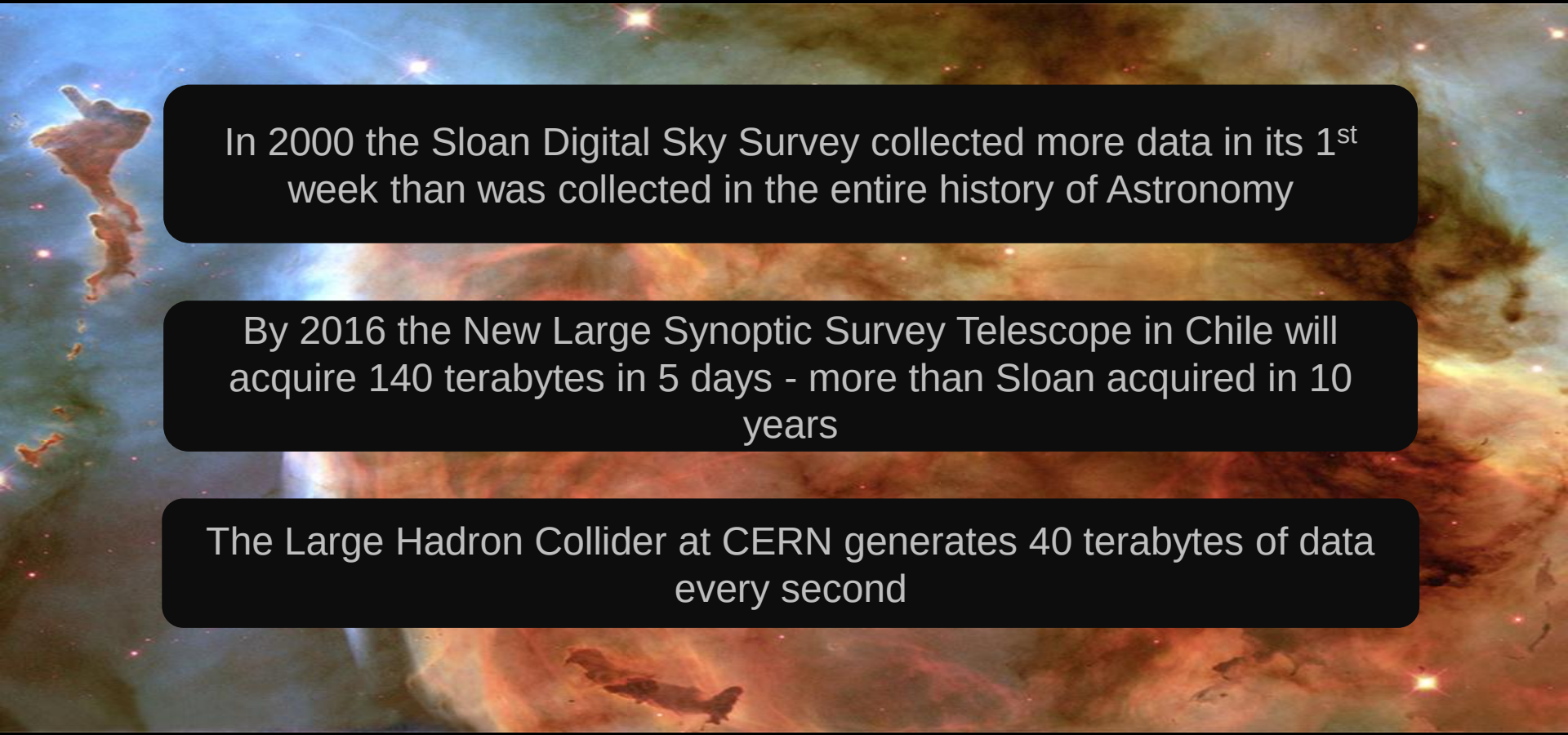
The Changing Nature of Research

- Thousand years ago – Experimental Science
  - Description of natural phenomena
- Last few hundred years – Theoretical Science
  - Newton's laws, Maxwell's equations…
- Last few decades – Computational Science
  - Simulation of complex phenomena
- Today – Data-centric Science
  - Unify theory, experiment and simulation
  - Using data exploration and data mining
    - Data captured by instruments
    - Data generated by simulations
    - Data generated by sensor networks
    - Data generated by humans

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{4\pi G\rho}{3} - \mathrm{K}\frac{c^2}{a^2}$$




1.2×10^{21}
New Bytes of Information in 2010
Source: IDC, as reported in The Economist, Feb 25, 2010

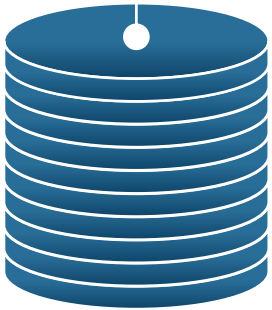


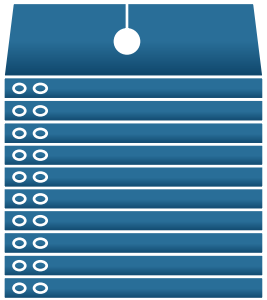
Sources: The Economist, Feb '10; IDC

Economics of Storage

Disk Storage
(per gigabyte)

$$404.0576





$$10.21550

Web Storage
(per gigabyte)

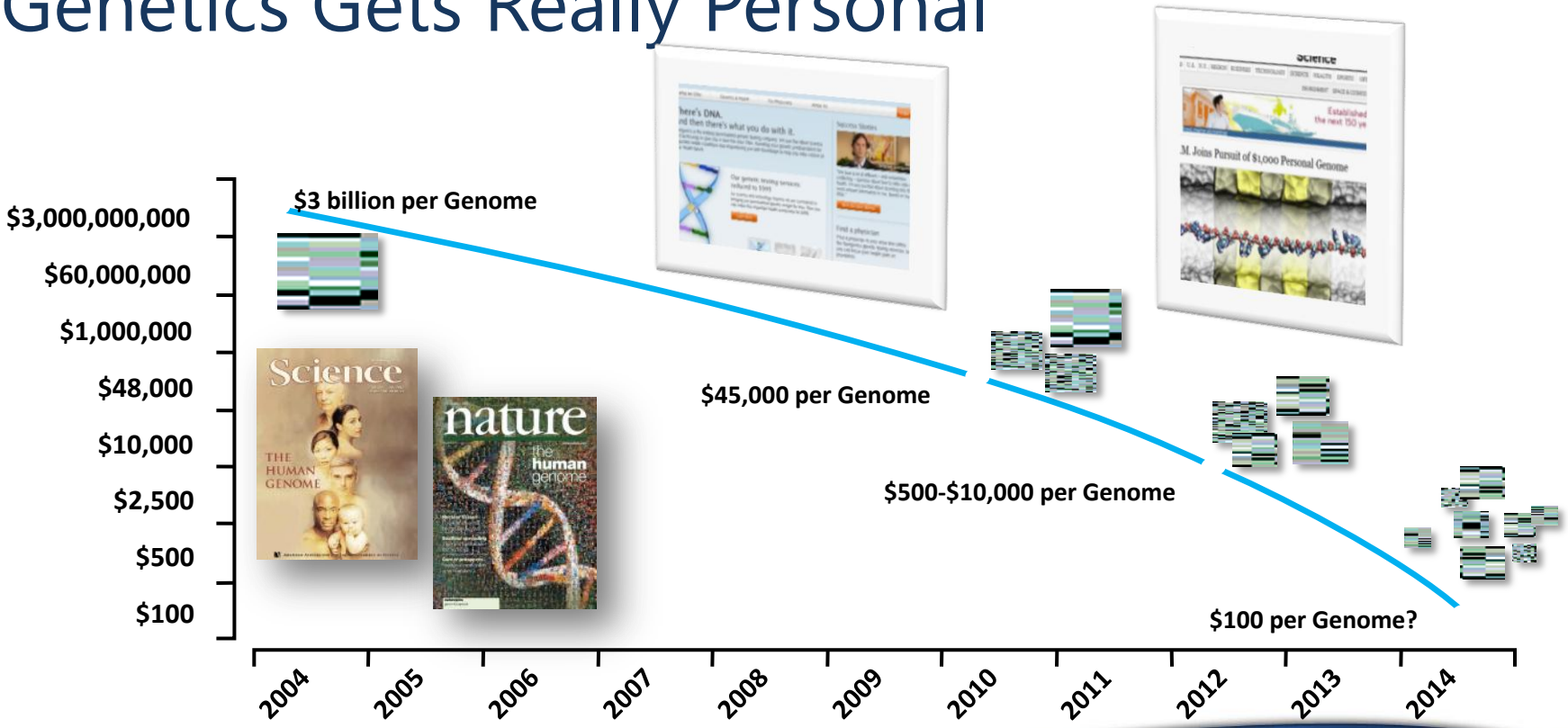
2008

But remember, … free storage is like free puppies

Source: Wired Magazine April 2010, Figures represented in USD

27

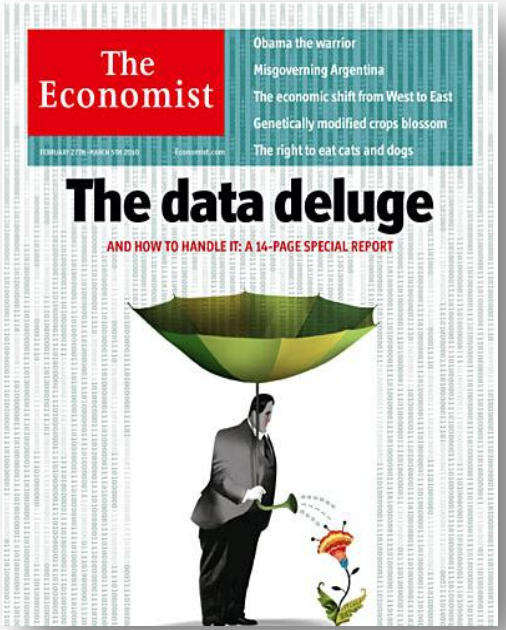
Genetics Gets Really Personal



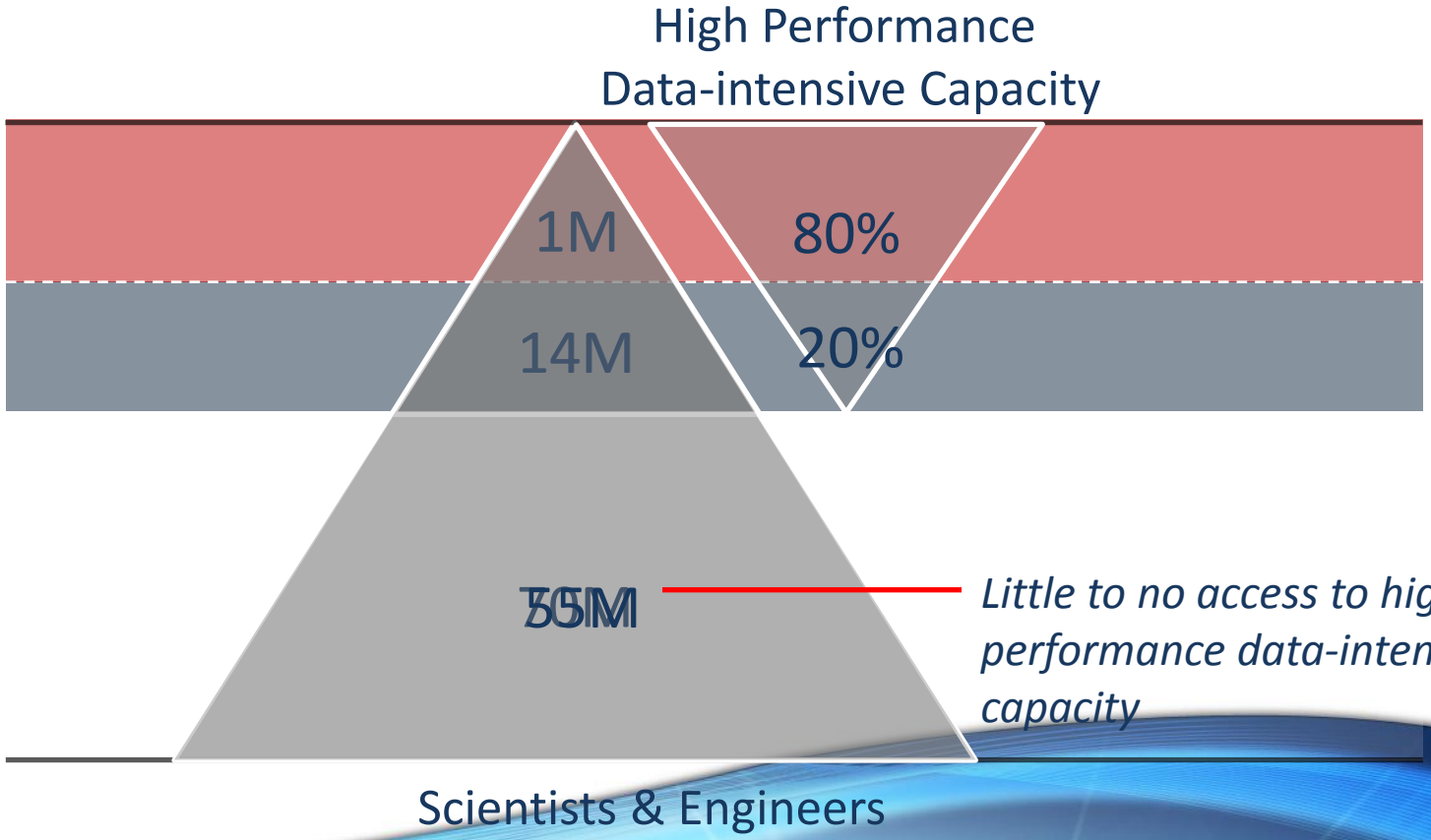
Source: George Church, Harvard Medical School, as reported in *IEEE Spectrum*, Feb '10. Figures represented in USD

Social Implications of the Data Deluge

- Hypothesis-driven
  - "I have an idea, let me verify it."
- Exploratory
  - "What correlations can I glean?"
- Different tools and techniques
  - Rapid exploration of alternatives
  - Data volume and complexity are assets
    - … and challenges
- Simplicity really matters



Lack of Broad Access

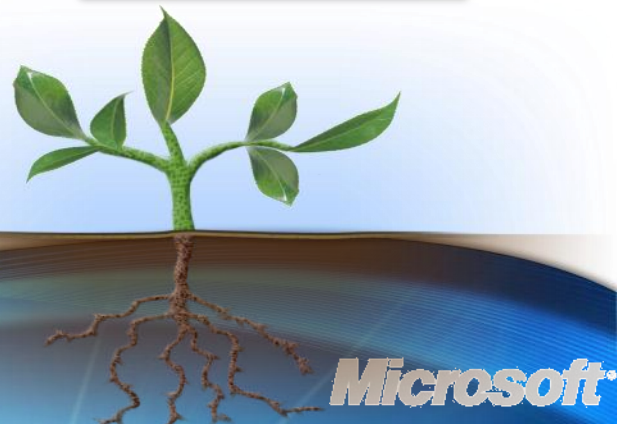


HPC and Clouds: Twins Separated At Birth

- Similar technology issues
  - Node and system architectures
  - Communication fabrics
  - Storage systems and analytics
  - Physical plant and operations
  - Programming models
  - Reliability and resilience
- Differing culture and sociology
  - Design and operations
  - Management and philosophy

Cloud Scaling: Lessons for HPC Exascale

- Environmental responsibility
  - Managing under a 100 MW envelope
  - Adaptive systems management
- Provisioning 100,000 servers
  - Hardware: at most one week after delivery
  - Software: at most a few hours
- Resilience during a blackout/disaster
  - Data center failure
  - Service rollover for 20M customers
- Programming the entire facility
  - Power, environmentals, provisioning
  - Component tracking, resilience, …



Microsoft

A Computer Room:

Neither A Cloud Data Center Nor An Exascale System

Microsoft's Data Center Evolution

2005 2006 2007 2008 2009 2010

Data Center Co-Location Generation 1	Quincy and San Antonio Generation 2	Chicago and Dublin Generation 3	Modular Data Center Generation 4




Deployment Scale Unit

Capacity

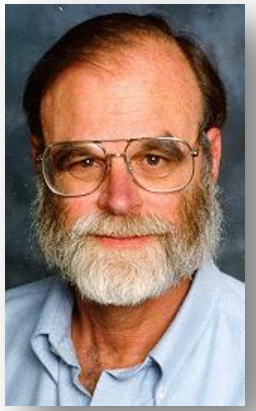
Density and Deployment

Scalability and Sustainability

Time to Market
Lower TCO

Microsoft

Orders of Magnitude Always Matter



- Tools must empower, not frustrate
  - These are systemic problems
  - An insight from Jim Gray …

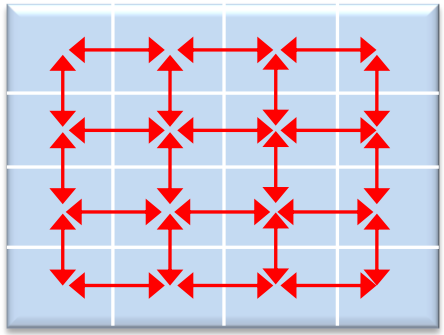
- A computation task has four characteristic demands:
  - Networking – delivering questions and answers
  - Computation – transforming information to produce new information
  - Data access – access to information needed by the computation
  - Data storage – long term storage of information
- The ratios among these *and their costs* are critical

Reliability and Resilience

- Orders of magnitude matter
  - Different bottlenecks appear
- System sizes are rising rapidly
  - The law of large numbers applies
- Failures are frequent
  - Component MTBF is not that high
    - Disks, power supplies, fans, DRAM
- System resilience dominates
  - Components are less important

Domain Decomposition and HPC

- Domain decomposition
  - Spreads vital data across all nodes
  - Each sp[illegible] one memory
    - Exce[illegible]ble ghos[illegible]alo cells
- Single [illegible]de f[illegible]e
  - Caus[illegible] lockag[illegible]en[illegible]simulation
  - Data l[illegible]t and m[illegible]ecovered
- Checkpoi[illegible] *facto* HPC solution
  - Periodically write all data to secondary storage
  - Given failures, one can compute an optimal interval



Listen To Your Grandmother …

- Her good advice
  - Eat your vegetables
  - Do unto others …
  - Always tell the truth
  - Work hard, recognition will come
- "Bad advice" from experts
  - Hardware is expensive and reliable
  - Optimize only for hardware performance
  - MPI is the *lingua franca* of HPC
  - MachoFLOPS trump data analysis
  - Cool your computing center for polar bears
  - Checkpoint frequently to preserve data

Dan's Grandmother Frazier

Microsoft®

Your potential. Our passion.™