Computer Organization

Mihai Udrescu

@MihaiUdrescu

mudrescu@gmail.com

About me

- Computer Architecture and Design
  - Arithmetic devices
  - Cache memory design
  - Hardware description languages (HDL) and FPGA design
  - Evolvable hardware
  - Crypto-chip design for testability
  - Reliability assessment
- Physics of computation
  - Quantum computers and quantum circuit design
  - Quantum algorithms
  - Quantum computer simulation
  - Genetic algorithms in QC

About me

- Network science and science of complexity
  - Social networks
  - Network medicine, network pharmacology, systems biology
  - Cyber-physical systems
- Publications (journals)
  - Nature Scientific Reports
  - PLoS One
  - PeerJ Computer Science, PeerJ
  - Computer Communications
  - Simulation Modeling Practice and Theory
  - International Journal of Computer Mathematics
  - Microelectronics Reliability
  - ACM Transactions on Emerging Technologies in Computing Systems
  - Pharmaceutics, Irish Journal of Medicine, Chest, ERS Journal

About me

- Review activity
  - IEEE Transactions on Computers
  - IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems
  - IEEE Transactions on Evolutionary Computation
  - IEEE Transactions on Computational Social Science
  - ACM Transactions on Embedded Computing Systems
  - Microelectronics Journal
  - Nature Scientific Reports
- Conference committees
  - ACM/IEEE International Conference on Hardware/Software Codesign and System Synthesis (CODES+ISSS)
  - IEEE Congress on Evolutionary Computation
  - IEEE International Conference on Computer and Information Technology
  - IEEE International Symposium on Design & Diagnostics of Electronic Circuits & Systems

About me

- Invited researcher
  - Electrical and Computer Engineering Department – Carnegie Mellon University (2011)
- Fulbright Scholar
  - Electrical and Computer Engineering Department – Carnegie Mellon University (2019-2020)
- Invited Speaker
  - Viterbi School of Engineering, Center for Cyber-Physical Systems and the Internet of Things, University of Southern California (2020)

- Research grants
  - ACM – Association for Computing Machinery
  - ERS – European Respiratory Society
  - Linde Healthcare
  - CNCSIS, UEFICSDI

The computing realm

- ACM/IEEE Curricula
  - Computer Science
    - Using science (math, physics, statistics) to describe the behavior of computer systems
  - Computer Engineering
    - Basic hardware-software, embedded systems, operating systems: lower-level computation issues
    - Computer science + electrical engineering
  - Information Systems
    - Technology-enabled business development
  - Information Technology
    - Managers of computer systems acting as advocates of computing that collaborate with professionals from other fields
  - Software Engineering
    - Techniques and technologies for developing efficient software

Science vs. Engineering

- Richard Feynman

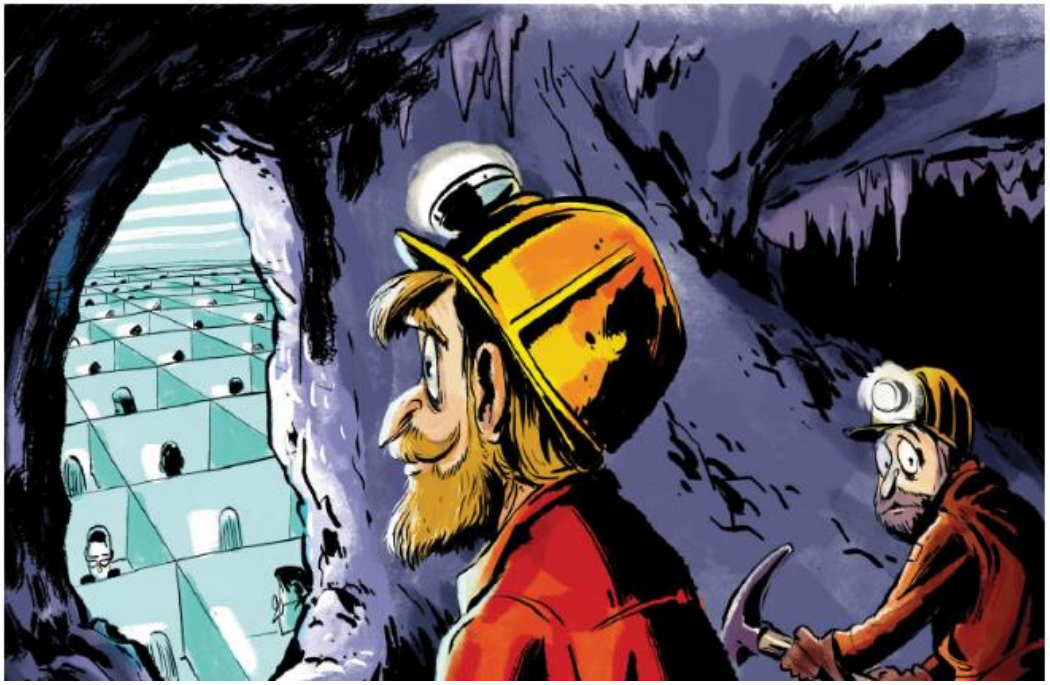
“I don't believe in Computer Science. To me science is the study of the behavior of nature and engineering or applied things is the behavior of things we make.” – Talk at Bell Labs, 1985

- Edsger Dijkstra

“Computer science is no more about computers than astronomy is about telescopes. “

CLIVE THOMPSON BUSINESS 02.08.17 12:30 PM

THE NEXT BIG BLUE-COLLAR JOB IS CODING



Education Hierarchy in Modern Society

Thinkers: Philosophy, Economics, Politics

Thought leaders, VC groups, Heads of State, Thinktanks

↓

Drivers of Society: Scientists and Artists
Cognitive neuroscience, Biomolecular Computing, Quantum Computing
Literature, Art

↓

Worker Bees: Technical skills
Electrical engineering, financial analysts, doctors, lawyers
currently being replaced by AI

Producers: Materials and Products
currently being replaced by AI and 3D printing

Educators/Teachers:
soon to be partially replaced by AI

↓

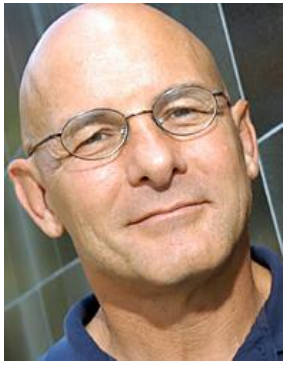
Slaves: Software engineering, Comp Sci, coders
Short career life span, soon to be replaced by AI

Slaves: Soliders, Police Officers, Govt Workers
Manual labour
currently being replaced by AI

Slaves: Poverty sweatshop workers, helpers, assistants
Manufacturing, Manual Labour, Repetitive labour
currently being replaced by AI

Why computer hardware and computer architecture & organization?

- Hardware is the basis of all computation
  - After all, computation is physical
  - The best SW professionals have good HW skills



David Patterson

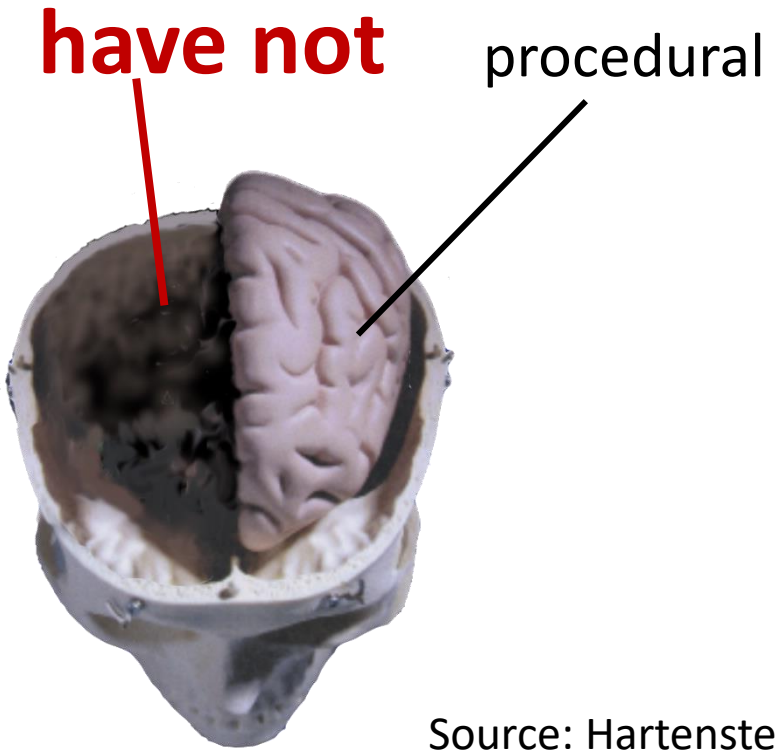
- Understanding HW prepares you for the new era in computation
  - We are at crossroads
  - The computing industry is at deadlock

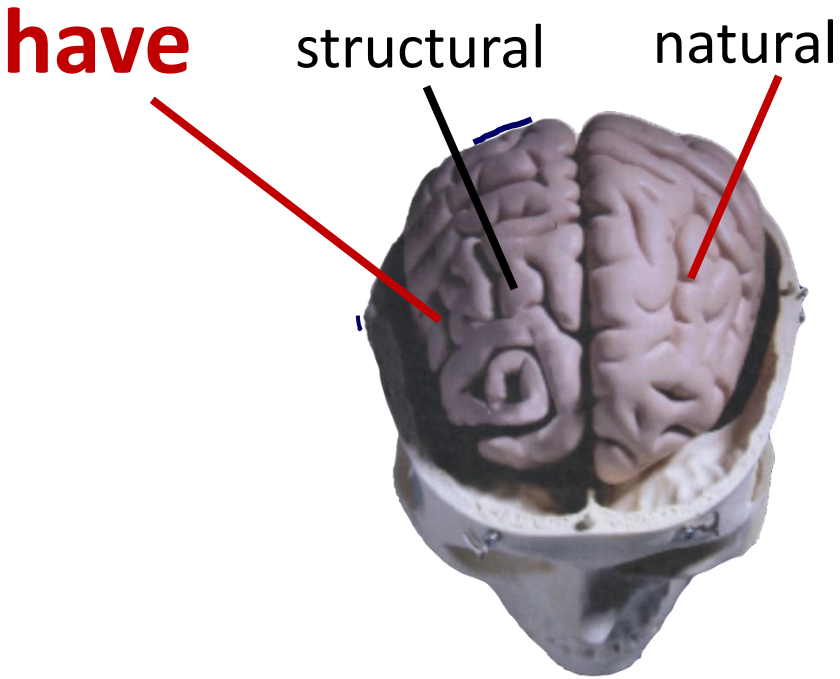
Why not computer architecture & organization?

- Hardware is **hard**

You cannot *teach Hardware to a Programmer

*) efficiently

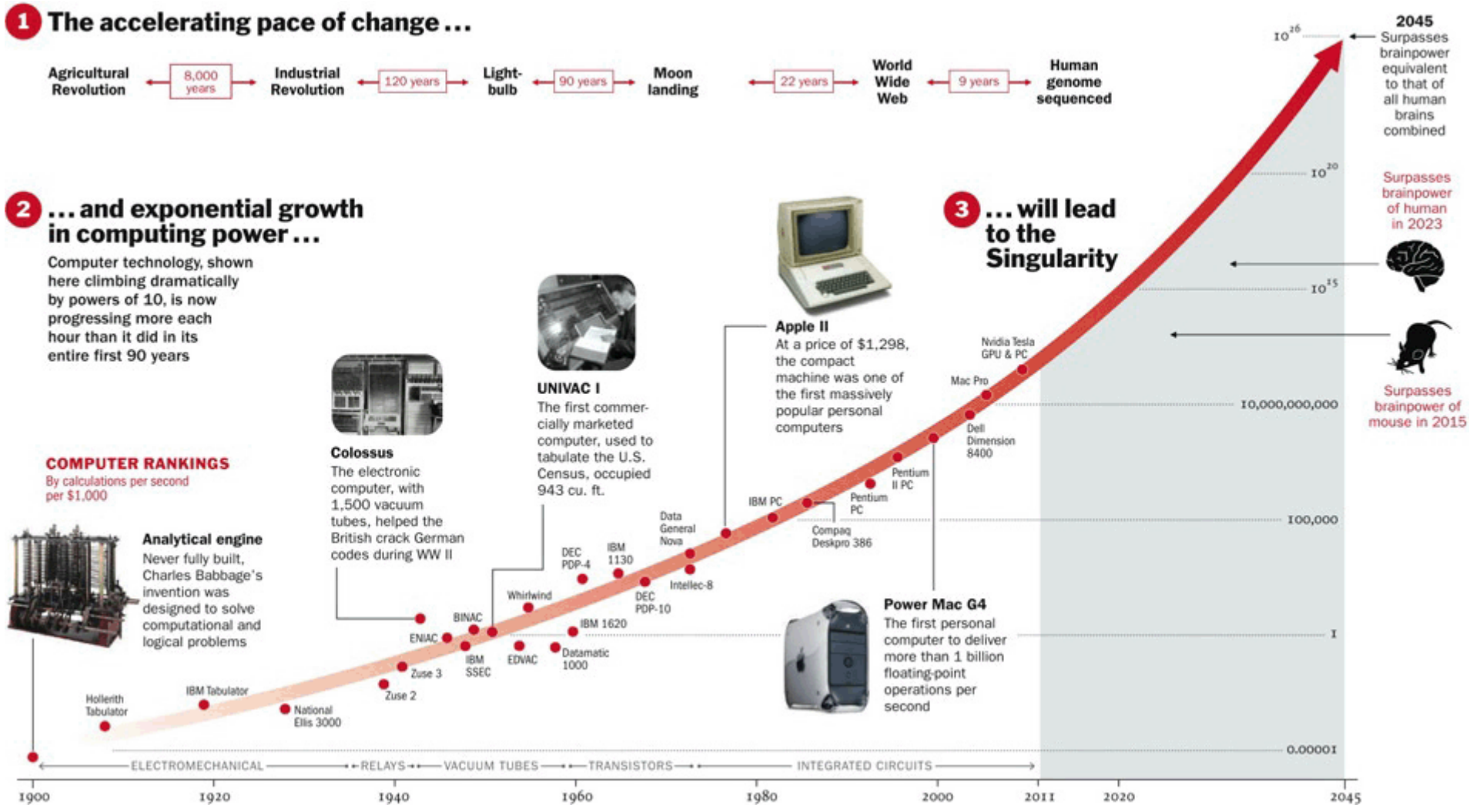


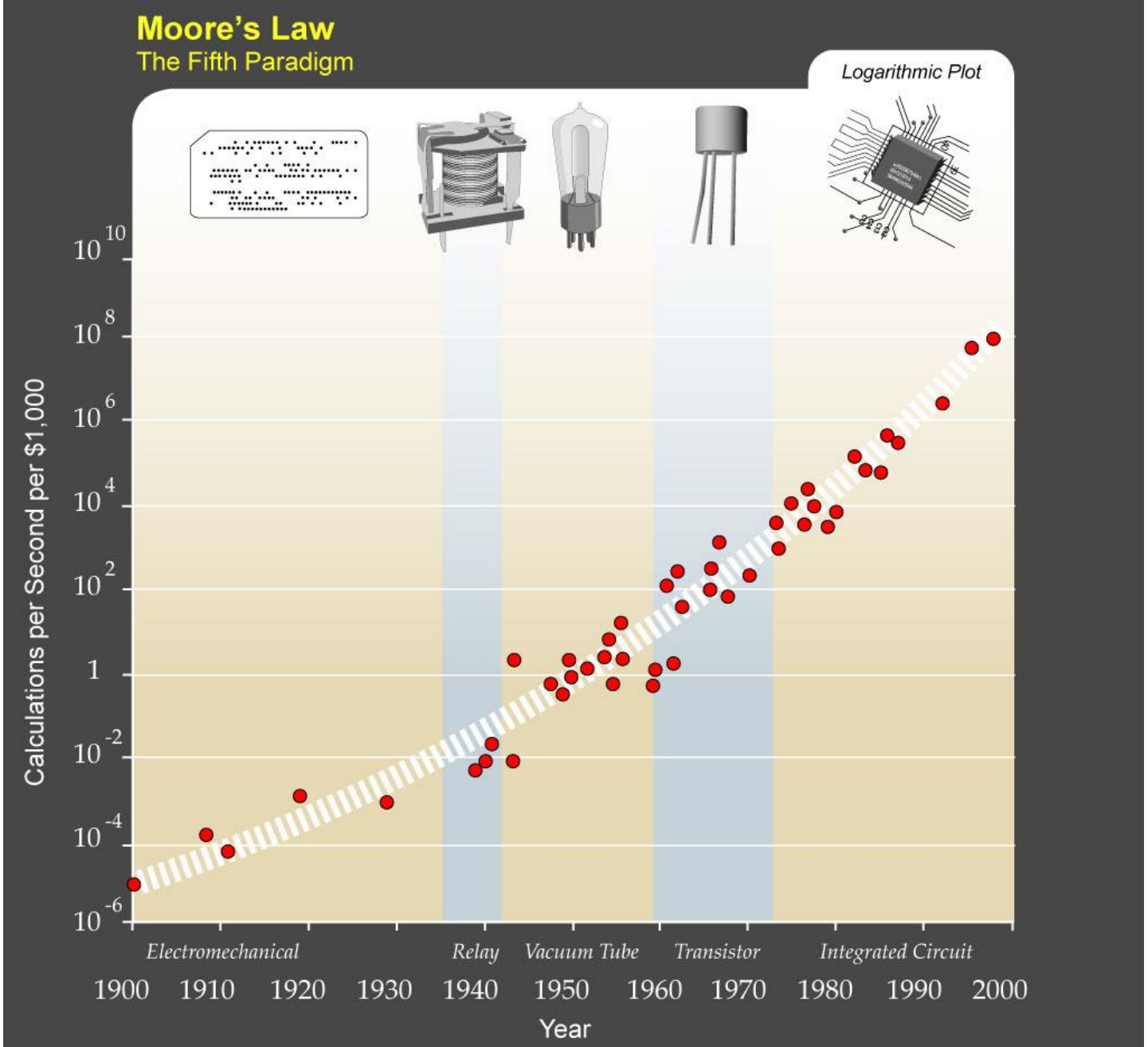


But to a Hardware guy you always can teach Programming

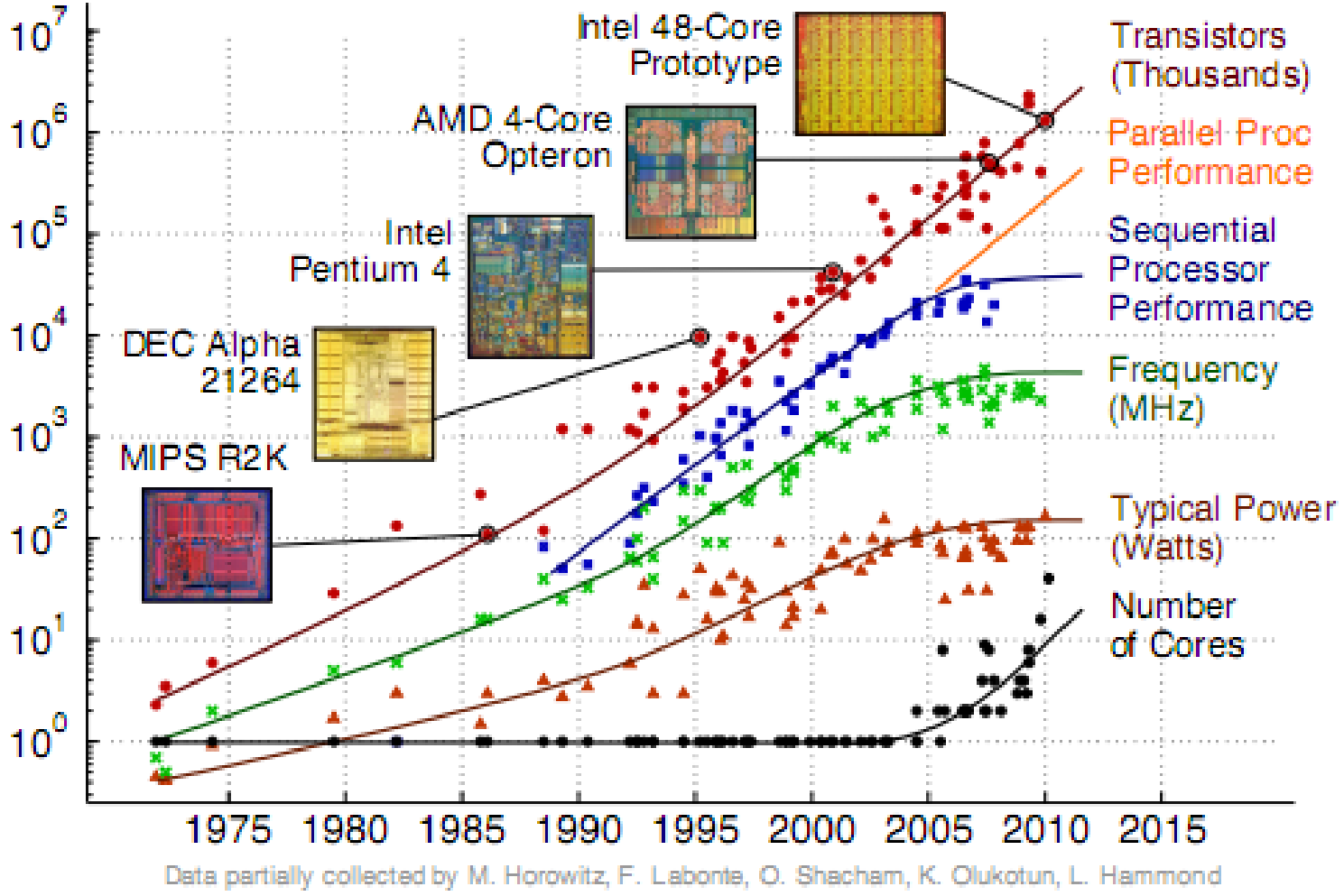
Source: Hartenstein.de

Singularity [John von Neumann]
“Singularity is near” – R. Kurzweil



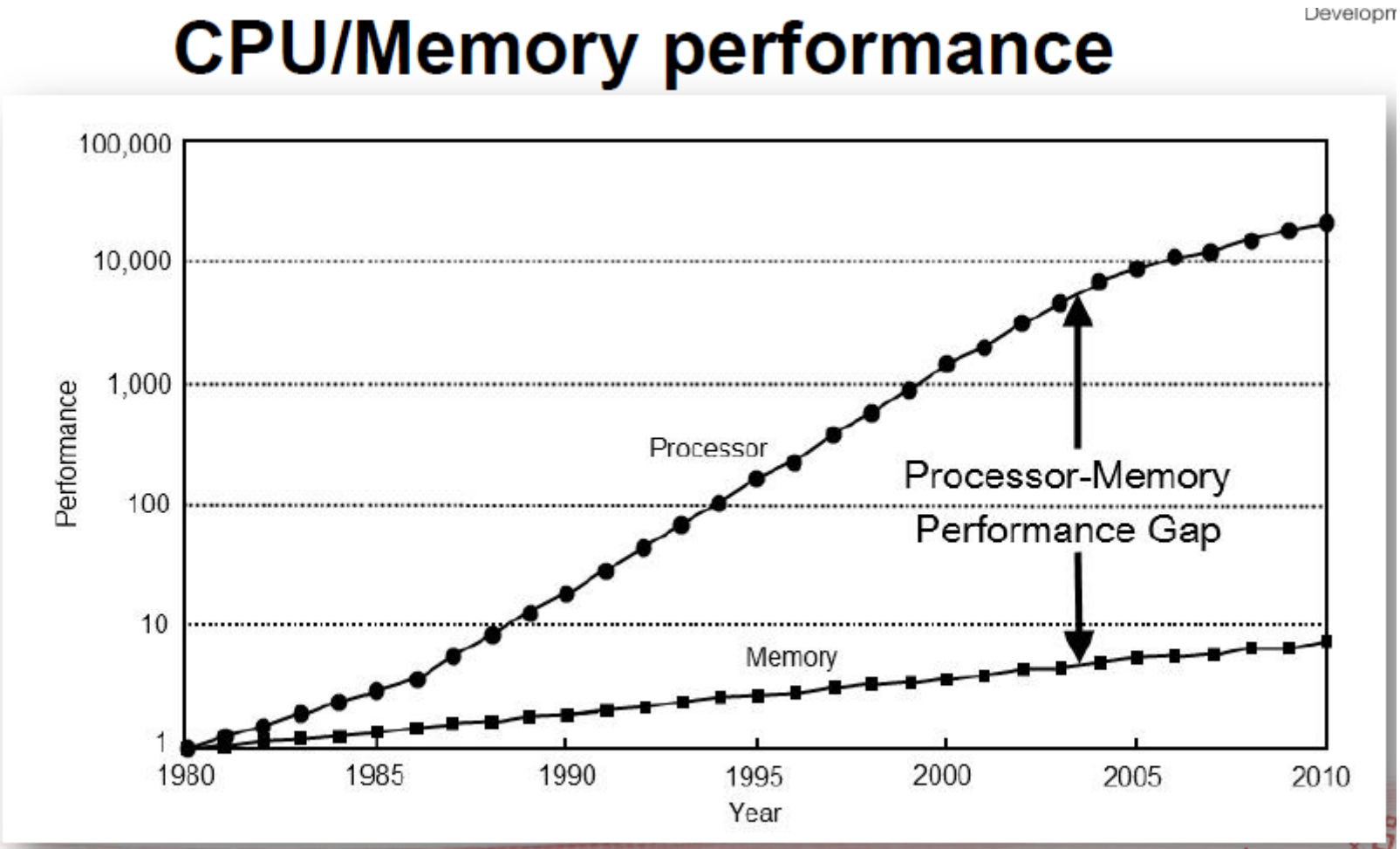

Moore's Law
The Fifth Paradigm
Logarithmic Plot
Calculations per Second per $1,000
10^{10}
10^{8}
10^{6}
10^{4}
10^{2}
1
10^{-2}
10^{-4}
10^{-6}
Electromechanical
Relay
Vacuum Tube
Transistor
Integrated Circuit
1900
1910
1920
1930
1940
1950
1960
1970
1980
1990
2000
Year

Consequences of Power Wall

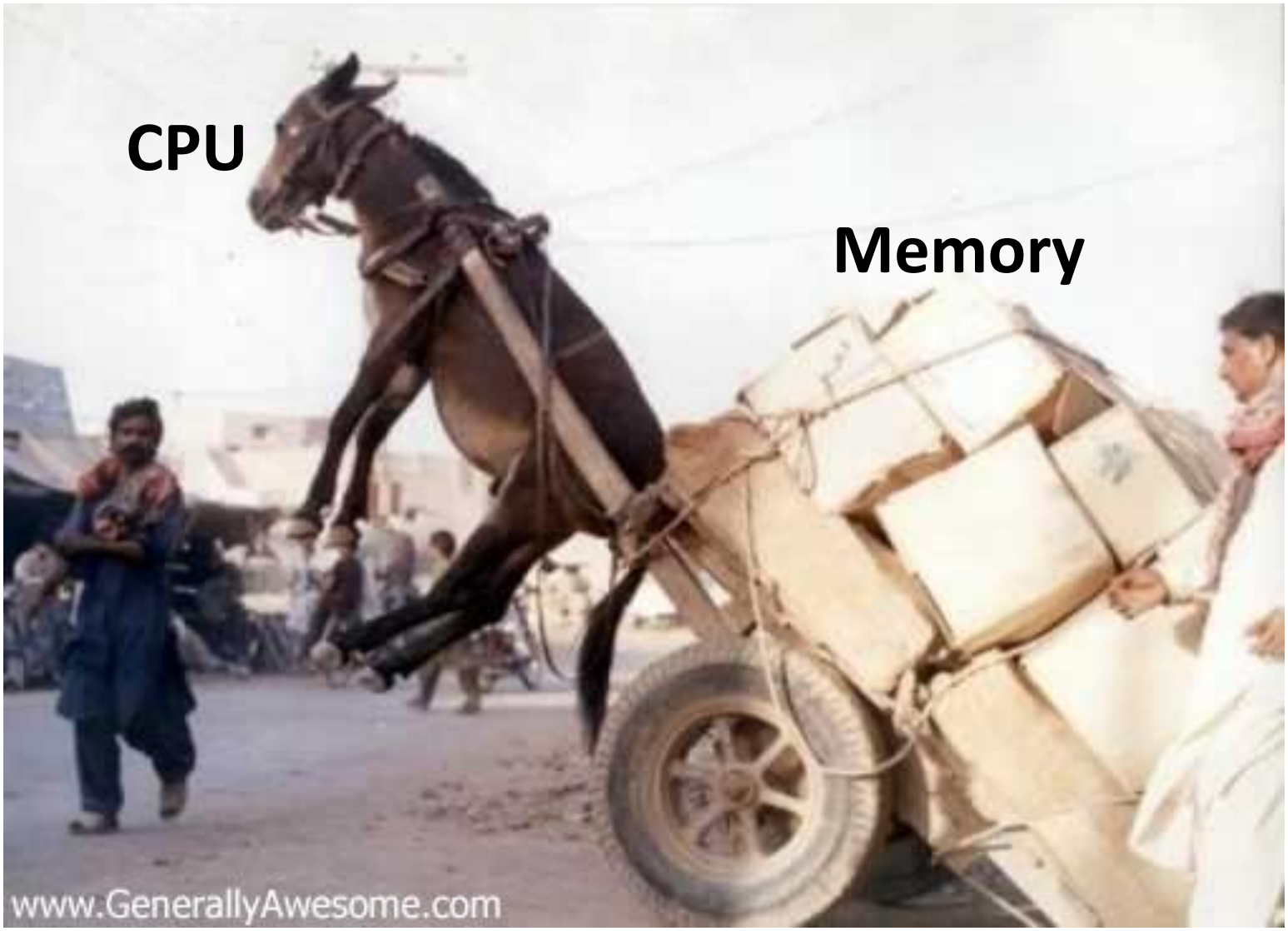


Prepared by C. Batten - School of Electrical and Computer Engineering - Cornell University - 2005 - retrieved Dec 12 2012 - http://www.csl.cornell.edu/courses/ece5950/handouts/ece5950-overview.pdf

Consequences of Memory Wall



CPU vs. Memory



Progress, uninterrupted evolution?

- Scientism vs. science
  - Dogmatic support for “scientific” method
  - Scientific positivism
  - "Twenty years from now, the author envisages the brain builder industry as being one of the world's top industries, comparable with oil, automobile, and construction.“ - Hugo de Garis, 1996

Source: machineslikeus.com

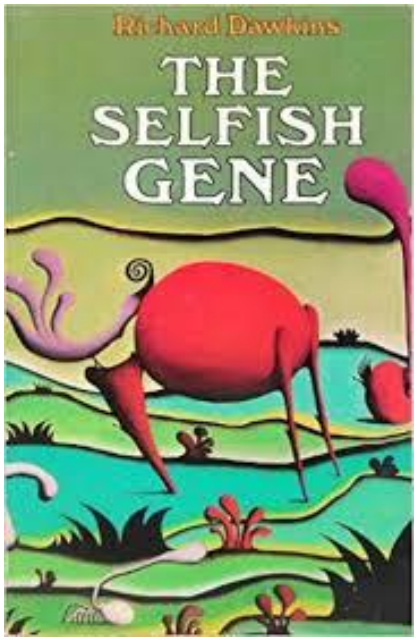
  - Computer industry: always faster, smaller, cheaper

Voices of reason

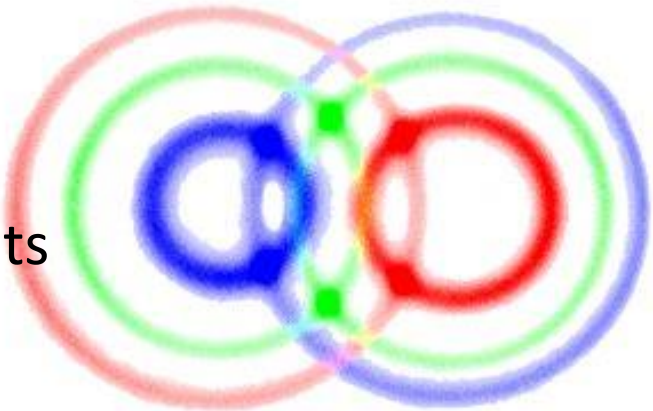
- Nassim Nicholas Taleb
  - Coined the term IYI
  - Exposed bad (bogus) science
  - Exposed badly used statistics

  - http://www.fooledbyrandomness.com/

Scientific method



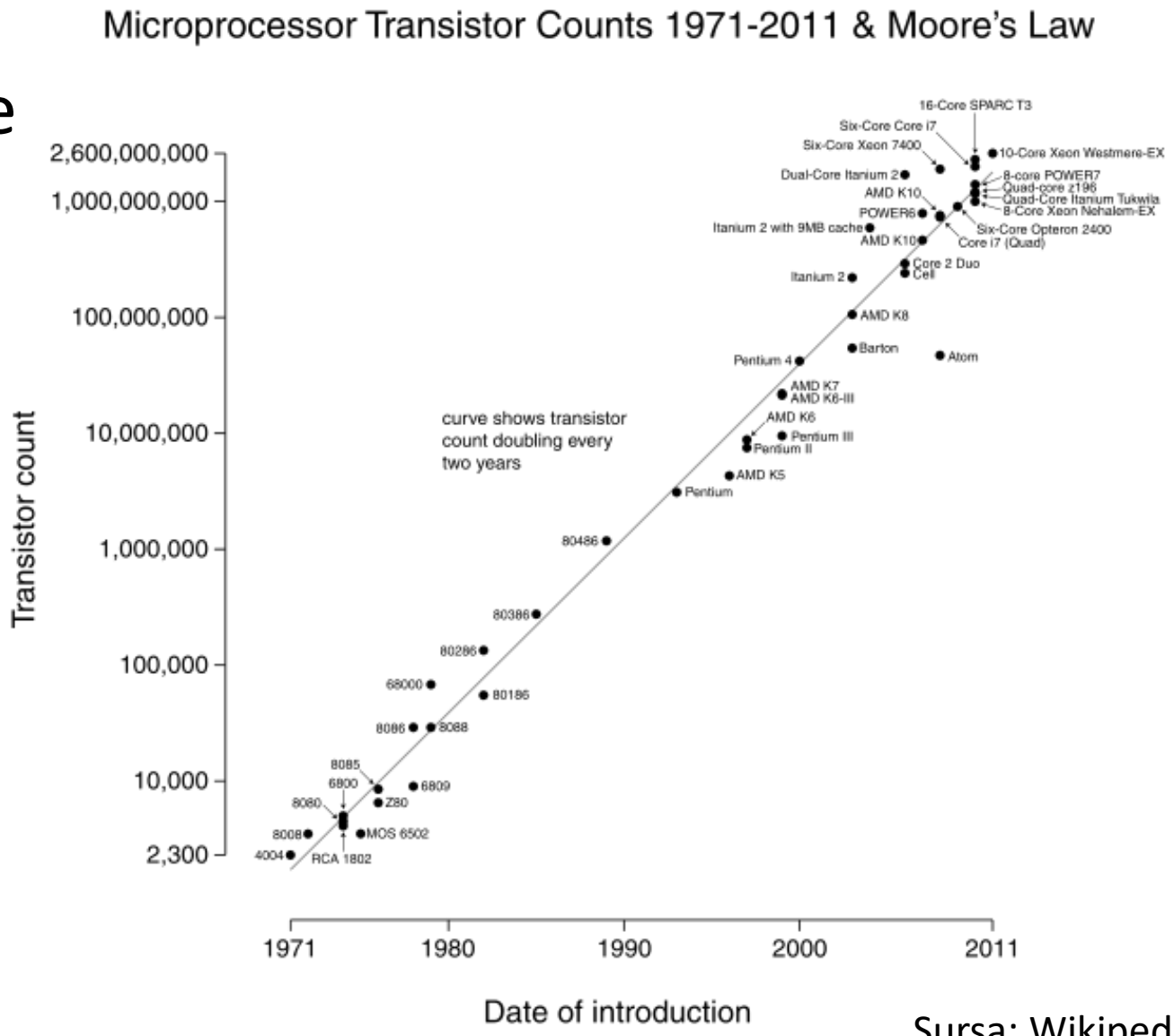
- Reductionism
  - Any complex system is merely the sum of its components
  - Richard Dawkins – hierachical reductionism
    - Living organisms compared with computers
  - Limits of reductionism
    - "The whole is more than the sum of its parts" – Aristotle, Metaphysics
    - Quantum physics – entanglement



Quantum Entanglement of Photon that has decayed-repelled into an entangled pair.

Source of progress in computers

- Consequence of Moore's law



Sursa: Wikipedia

Moore's law has consequences

- 2019-2020, **1 atom/1 bit**
- ITRS – International Technology Roadmap for Semiconductors
  - Quantum computing
  - Bio-inspired computing

Quantum computing?

The Nobel Prize in Physics 2012

Nobel Prize Award Ceremony

Serge Haroche

David J. Wineland

Photo: U. Montan

Serge Haroche

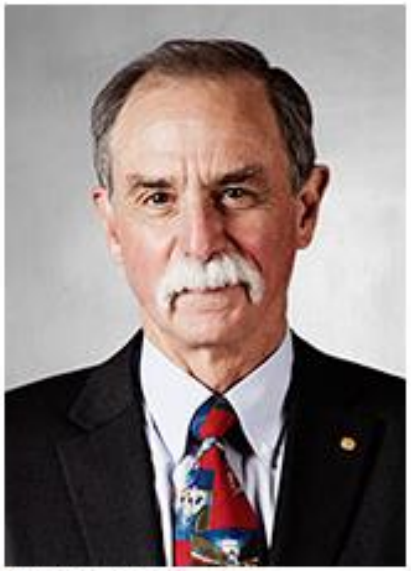
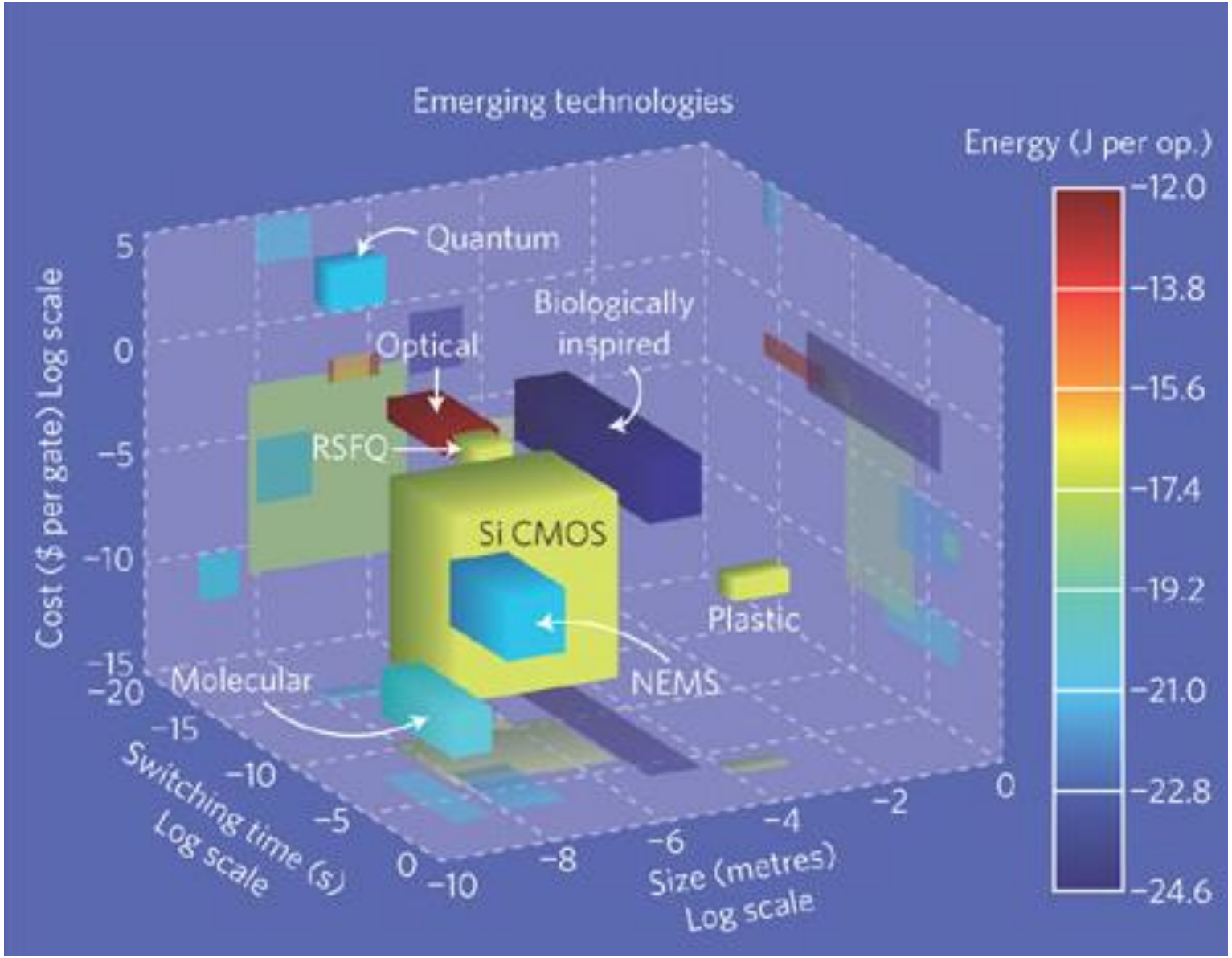


Photo: U. Montan

David J. Wineland

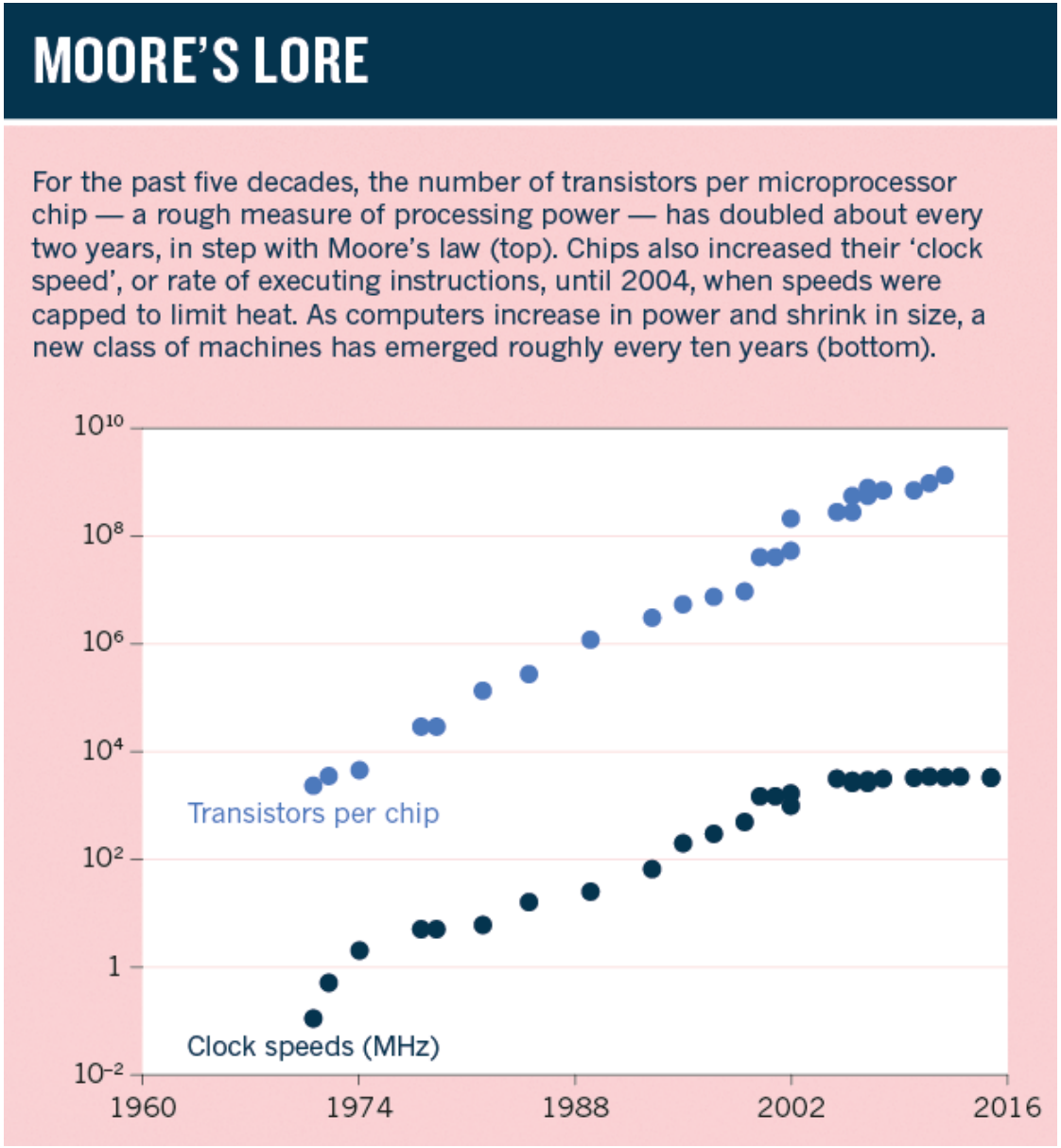
The Nobel Prize in Physics 2012 was awarded jointly to Serge Haroche and David J. Wineland *"for ground-breaking experimental methods that enable measuring and manipulation of individual quantum systems"*

ITRS – Emerging Research Devices



In the meantime

- Moore's law is DEAD
- Today: 14nm
- 2020: 2-3nm
  - 10 atoms across



What’s next?

A multi-core perspective

Press Release

David Patterson *(Peg Skorpinski photo)*

Intel and Microsoft launch parallel computing research center at UC Berkeley

By Sarah Yang, Media Relations | 18 March 2007

BERKELEY – The University of California, Berkeley, is partnering with Intel Corp. and Microsoft Corp. to accelerate developments in parallel computing and advance the powerful benefits of multi-core processing to mainstream consumer and business computers.

ieee spectrum INSIDE TECHNOLOGY

MAGAZINE MU

AEROSPACE BIOMEDICAL COMPUTING CONSUMER ELECTRONICS ENERG

COMPUTING / SOFTWARE

FEATURE

The Trouble With Multicore

Chipmakers are busy designing microprocessors that most programmers can't handle

By DAVID PATTERSON / JULY 2010

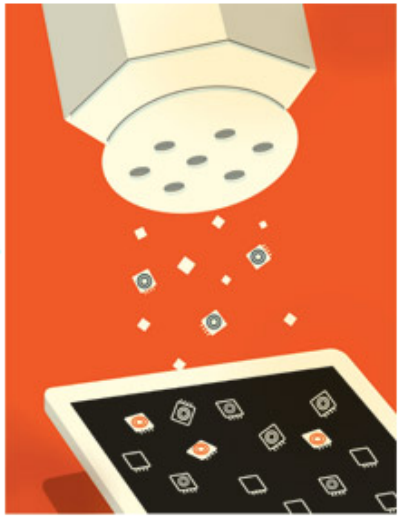


Illustration: Harry Campbell

In 1975, future Hall of Famer Roger Staubach had the football but little else in a playoff game against the Minnesota Vikings. Behind by four points at midfield with 24 seconds to go, the Dallas Cowboys quarterback closed his eyes, threw the ball as hard as he could, and said a Hail Mary. (For you soccer fans, this would be like David Beckham taking a shot on goal from midfield late in injury time.)

His prayer was answered. Staubach's receiver collided with a Viking defender just as the ball arrived but nevertheless managed to pin the football against his leg, scoring the touchdown that took the Cowboys to the Super Bowl. (Imagine Beckham's long ball beating the goalie.) Ever since that game, a desperate pass with little chance of success has been labeled a Hail Mary.

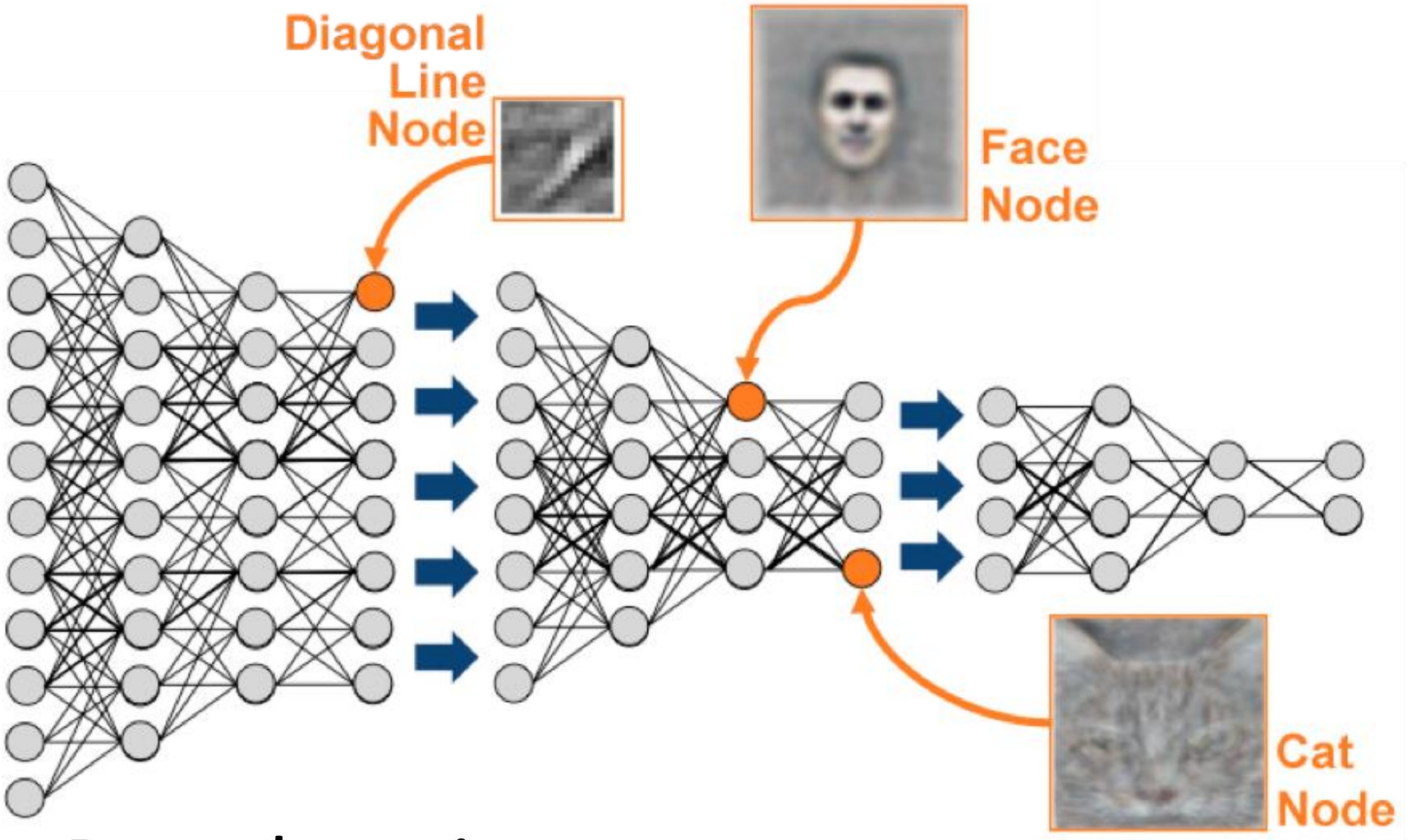
Thirty years later, the semiconductor industry threw the equivalent of a Hail Mary pass when it switched from making microprocessors run faster to putting more of them on a chip—doing so without any clear notion of how such devices would in general be programmed. The hope is that someone will be able to figure out how to do that, but at the moment, the ball is still in the air.

Motivation

- Multi-Core is THE hot top topic as we speak

- David Patterson
  - "The Trouble With Multi-Core" – IEEE Spectrum, vol. 47, no. 7, pp. 24-29 (August 2010)
  - "Chipmakers are busy designing microprocessors that most programmers can't handle"

Feasible solutions



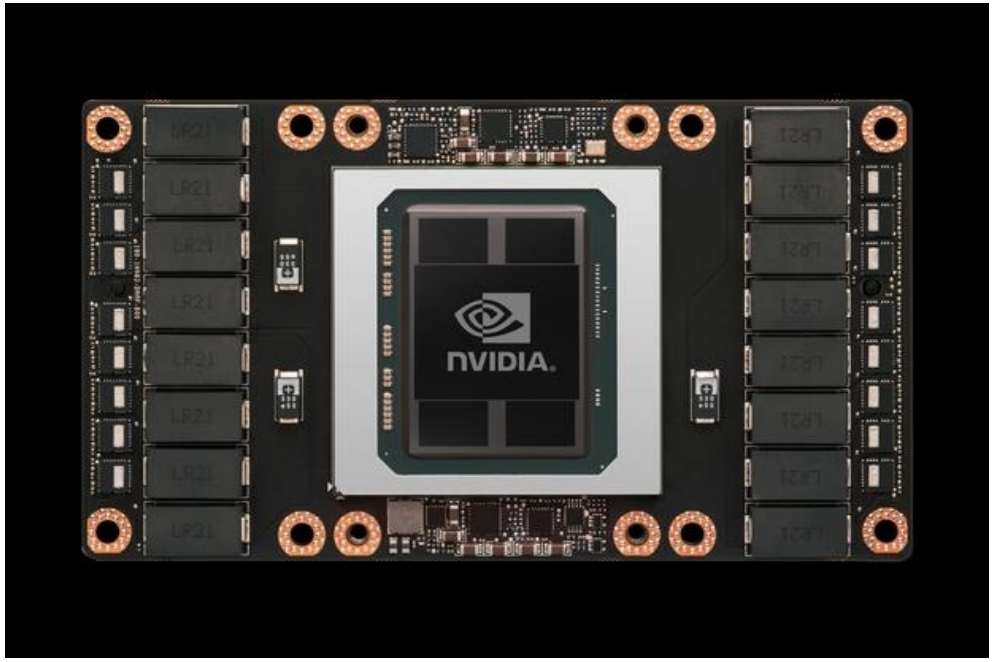
Deep learning, source forbes.com

Deep learning is powerful

- Google's AlphaGo program mauled Lee Sedol, one of the highest-ranking Go players in the word
- A particular case of Machine Learning
  - The computer learns to do something without being specifically programmed for that purpose
    - Recognizing an SUV
    - Some SUV pictures are "shown" to the computer and labeled as SUVs => training
      - The computer builds a model
      - Subsequent model adjustments until the model becomes accurate enough

Deep learning uses neural networks

- Multi-level (layer) artificial neural networks
- Implemented and accelerated with multi-core GPUs
- NVIDIA Tesla



Deep learning applications

LIVESCIENCE NEWS TECH HEALTH PLANET EARTH SPACE STRANGE NEWS ANIMALS

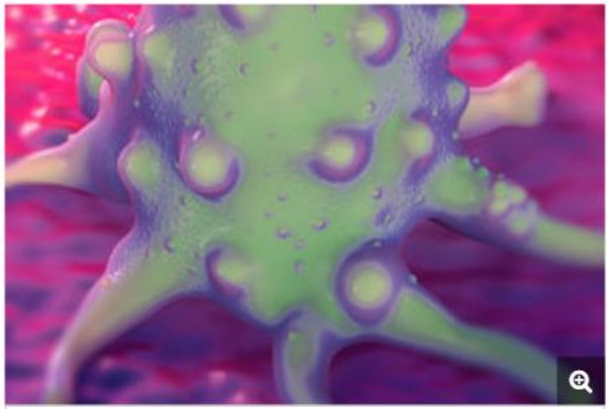
Sponsored by Rekoverr

Live Science > Health

AI Boosts Cancer Screens to Nearly 100 Percent Accuracy

By Christopher Wanjek | June 21, 2016 01:54pm ET

92 63 2 0 58 MORE



Combining artificial intelligence with a human pathologist could boost the accuracy of cancer diagnosis.

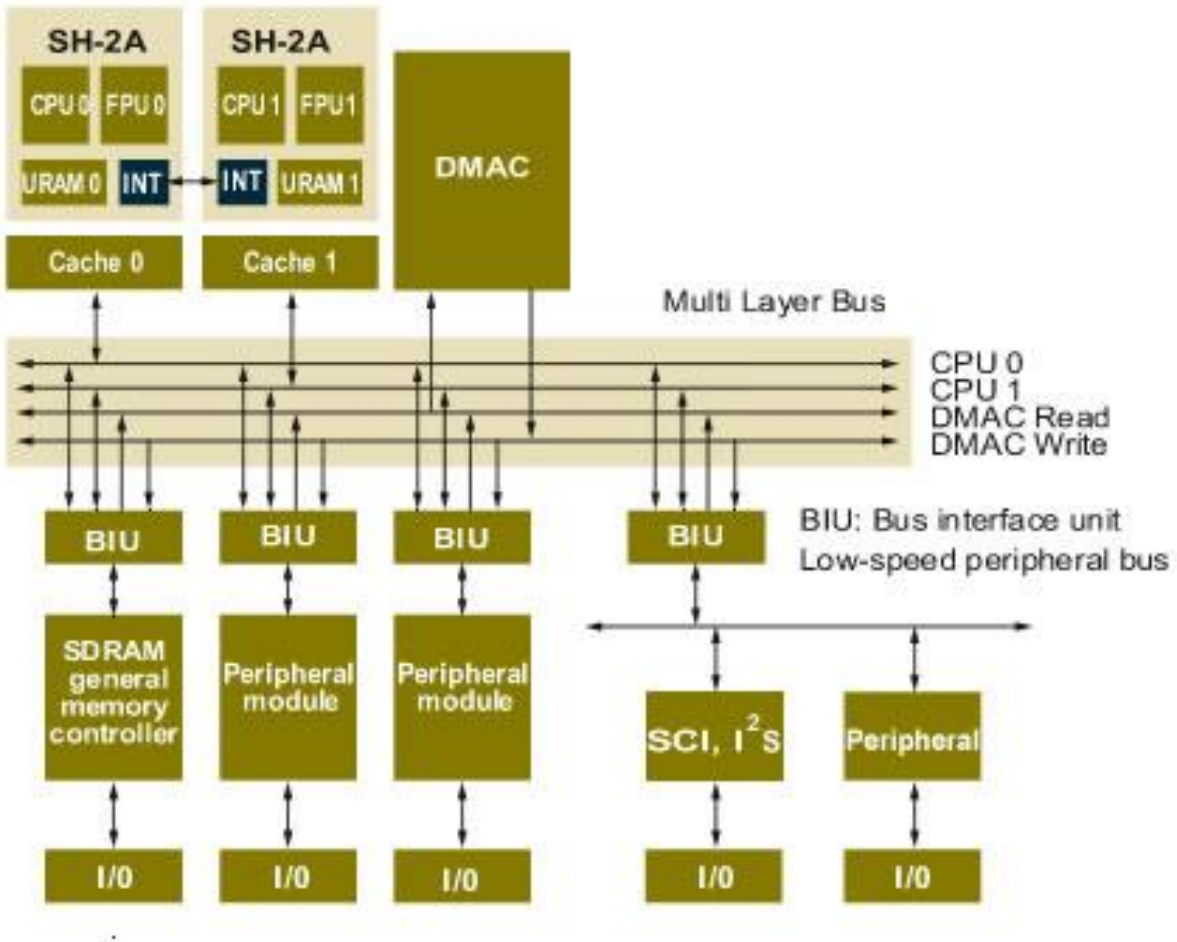
Credit: royaltystockphoto.com / Shutterstock.com

Diagnosing cancer is about to get more accurate, with the help of artificial intelligence.

Pathologists have diagnosed diseases in more or less the same way for the past 100 years, by laboring over a microscope reviewing biopsy samples on little glass slides. Working almost robotically, they sift through millions of normal cells to identify just a few diseased ones. The task is

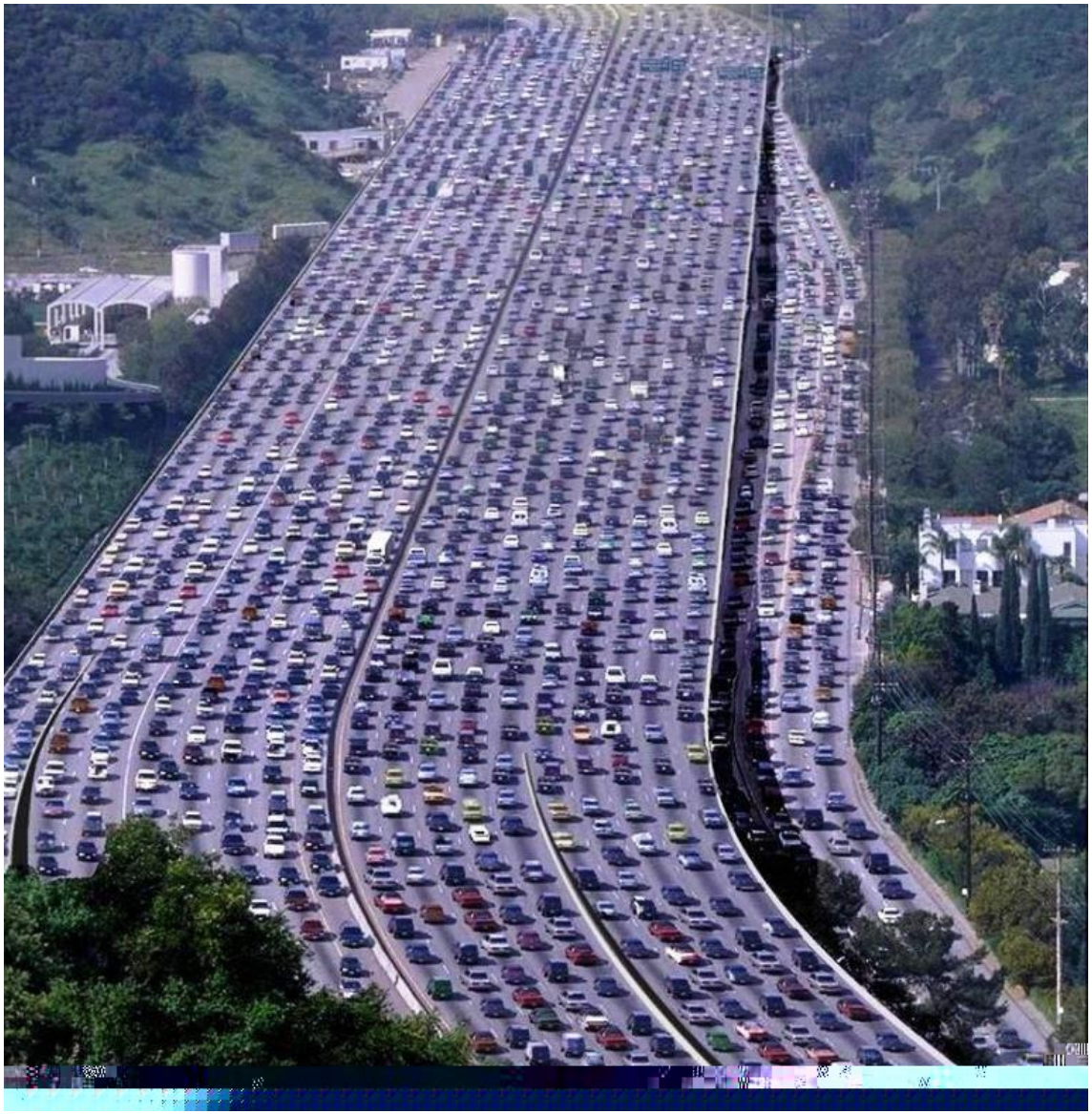
Inter-core communication

- Bus – unique communication resource



Inter-core communication

- Many-core
  (> 100 processors)

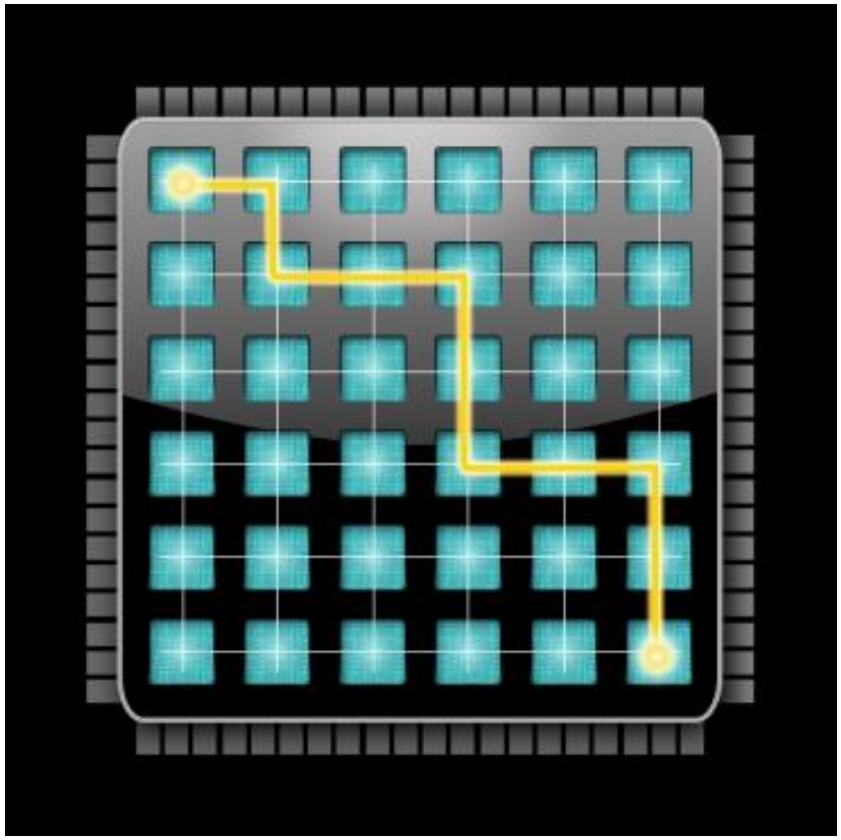


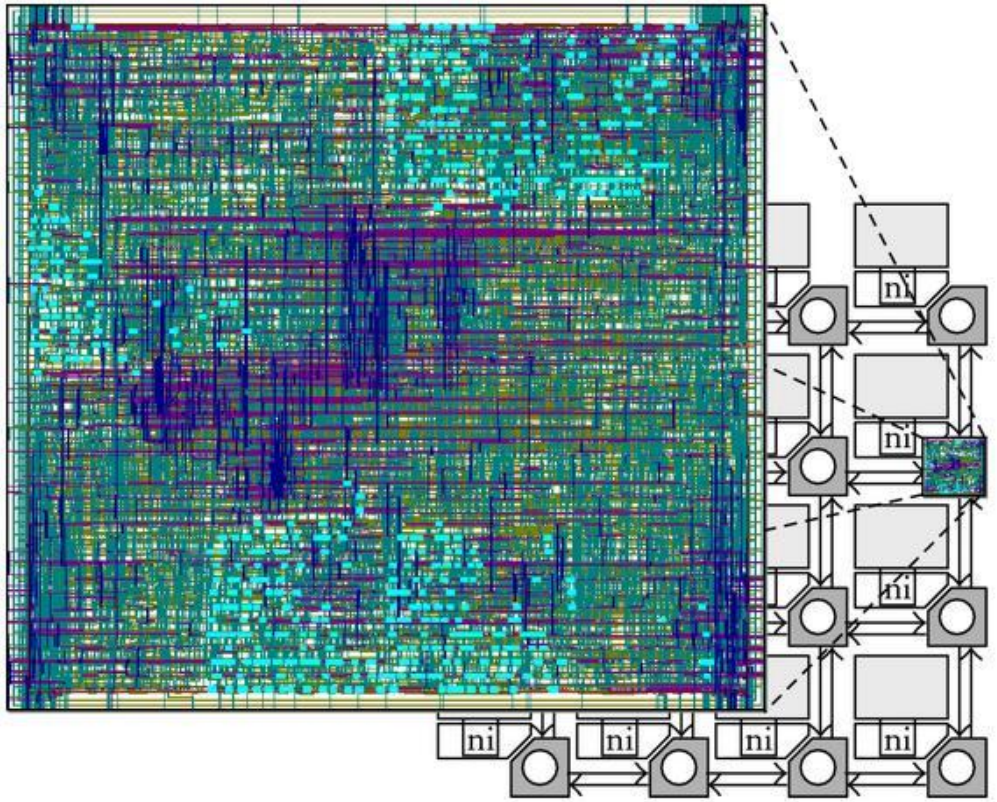
Many-core communication

- Network

Networks-On-Chip

- Topology: regulated, limited by VLSI technology

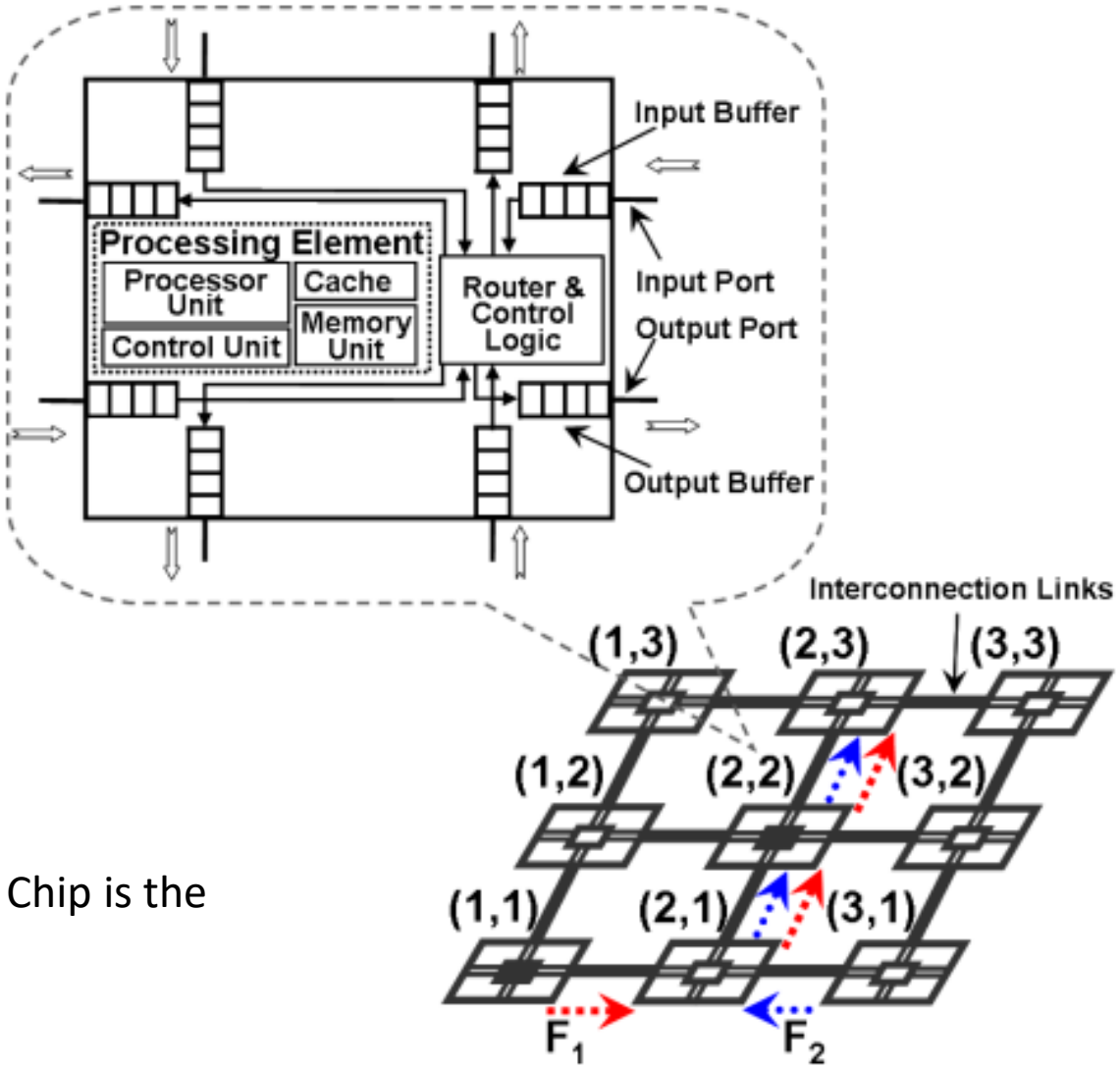




F. Samman, T. Hollstein, M. Glesner. Networks-On-Chip Based on Dynamic Wormhole Packet Identity Mapping Management. VLSI Design (2009)

Communication: dynamic problems

- Congestion



Marculescu & Bogdan, The Chip is the Network, 2009

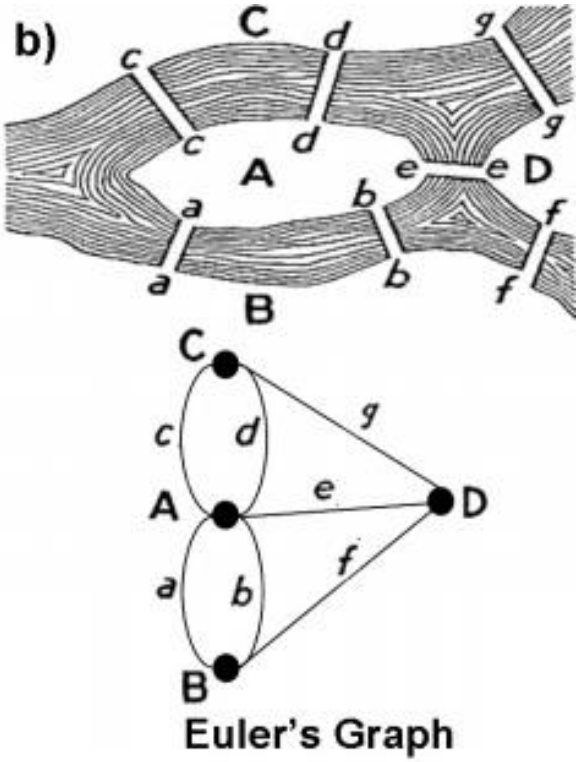
Congestion prediction

- Difficult problem
- Hard to suggest a NoC design which avoids congestion

NoC- design approaches

- Deterministic
- Stochastic
- Statistical physics & complex network science

Leonhard Euler [1707-1783]



Euler's Graph

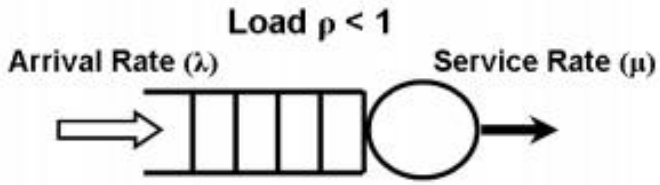
NoC- design approaches

- Deterministic
- Stochastic
- Statistical physics and complex network science

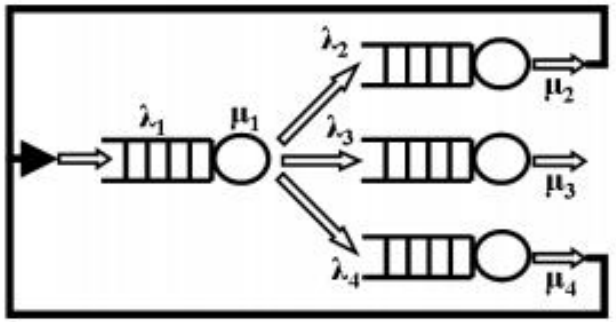
a)

Andrey Markov [1856-1922]

b) Single Queue Problem



Queueing Network Problem



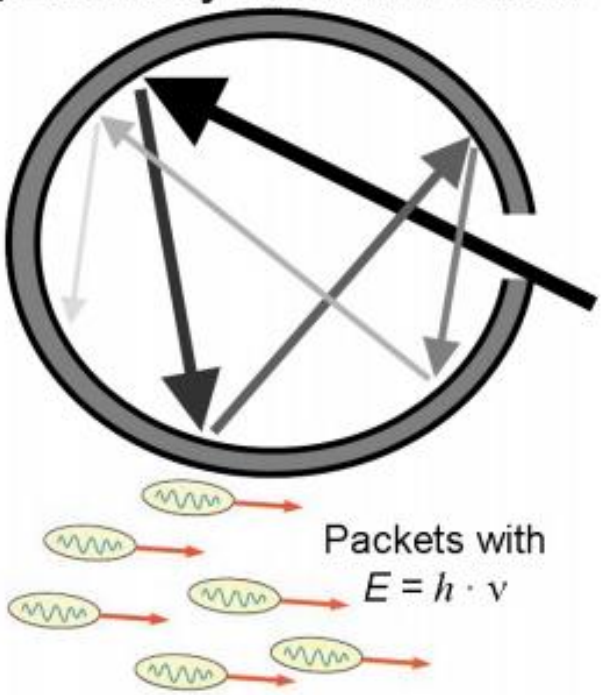
NoC- design approaches

- Deterministic
- Stochastic

a)

Max Planck [1858-1947]

b) Blackbody Radiation Problem



- Statistical physics and complex network science

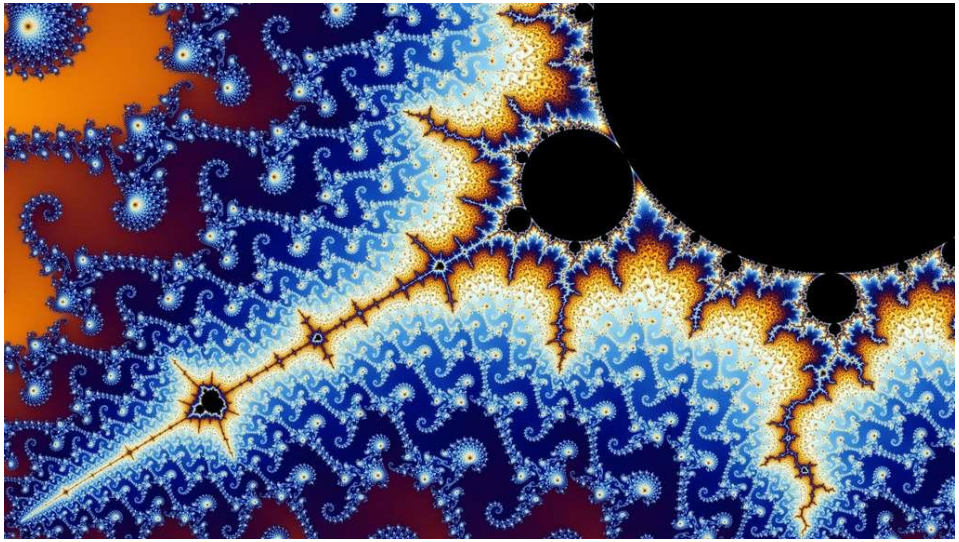
New roles for computers

- Modeling and controlling uncertain systems
  - Nature and society

Social network dynamics

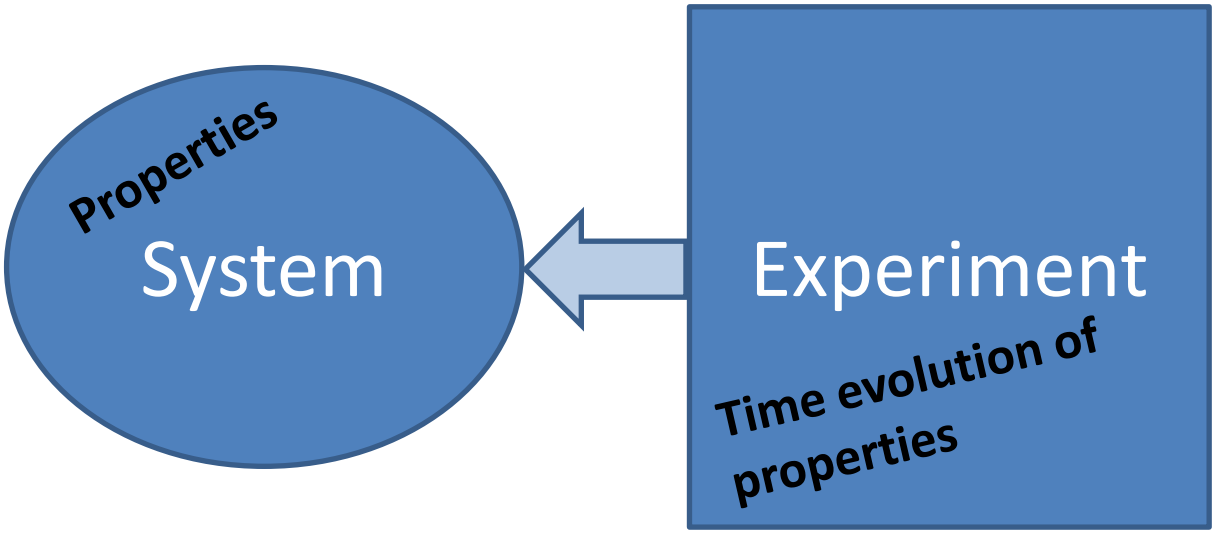
Urban traffic

These are complex systems



What is a complex system?

- "One of the hardest things to explain is why complex systems are actually different from simple systems." – Yaneer Bar-Yam, Why complexity is different.
- Scientific methodology:

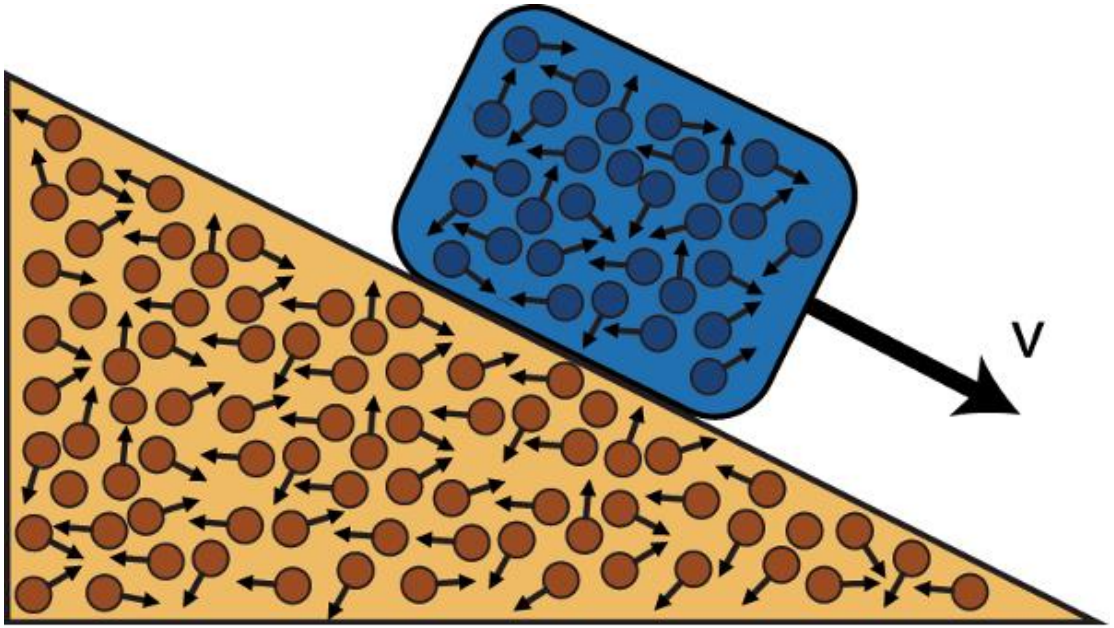


MODEL

- Did we include all properties?
- Properties are changing over time
- Consider as many as possible properties
- **What properties are the most important?**

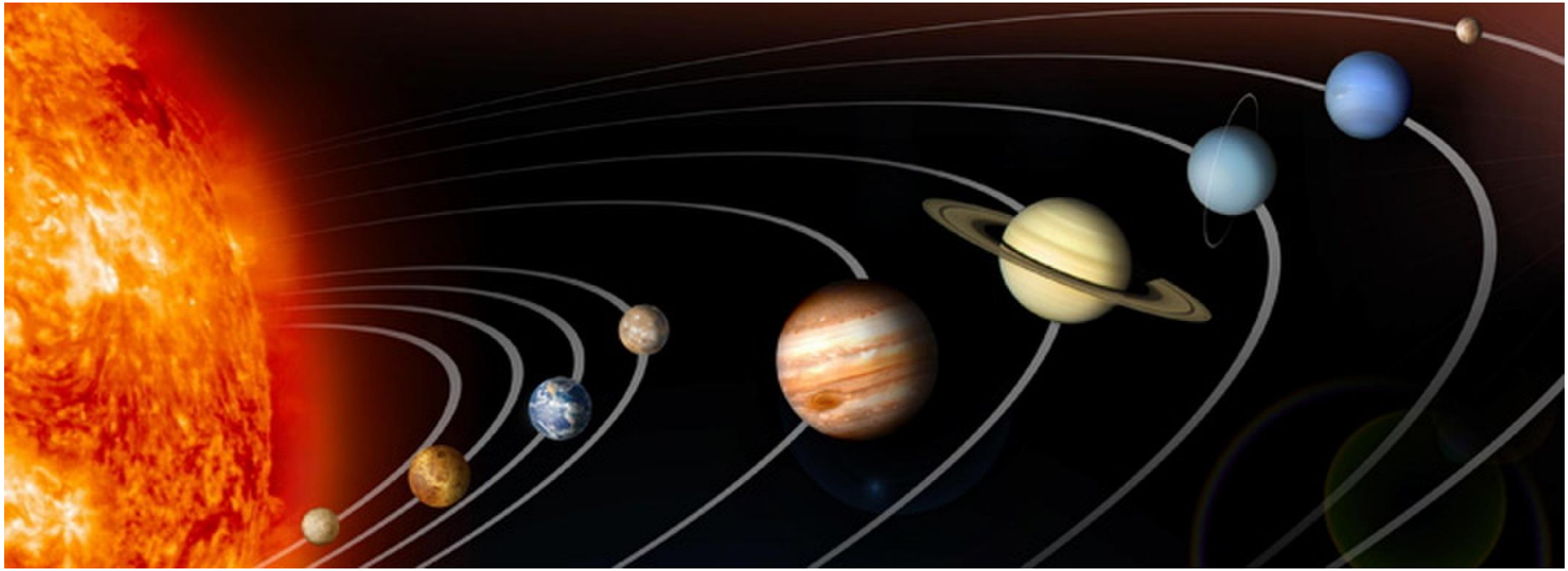
In physics we have

- **Separation of scales**
- Traditionally, any system is charcterized by
  - Scale
  - Average



Sursa: Yaneer Bar-Yam, Why complexity is different

Separation of scales work

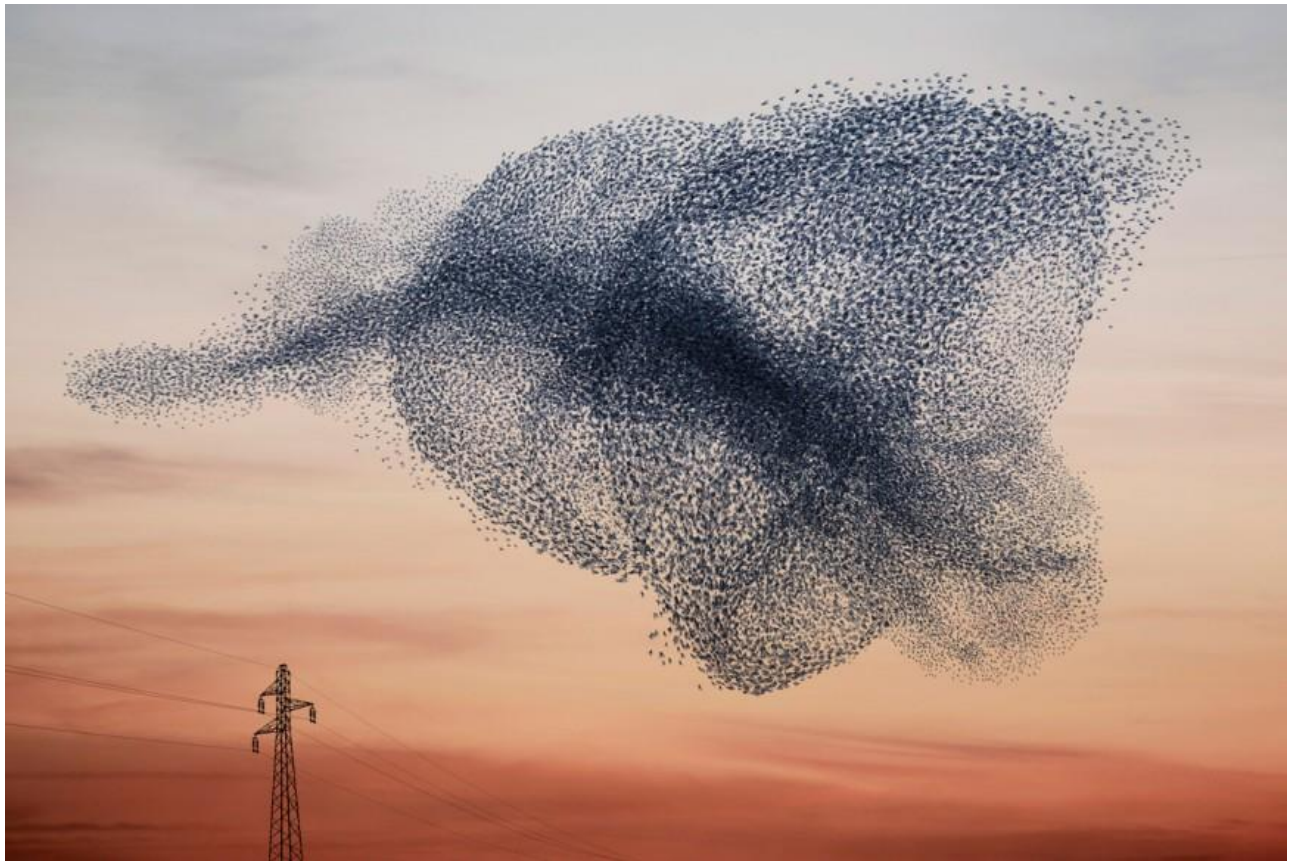


- Insulated solar system
- Interaction with external bodies/objects (e.g., asteroids, comets)

Sursa: space.com

Complex systems

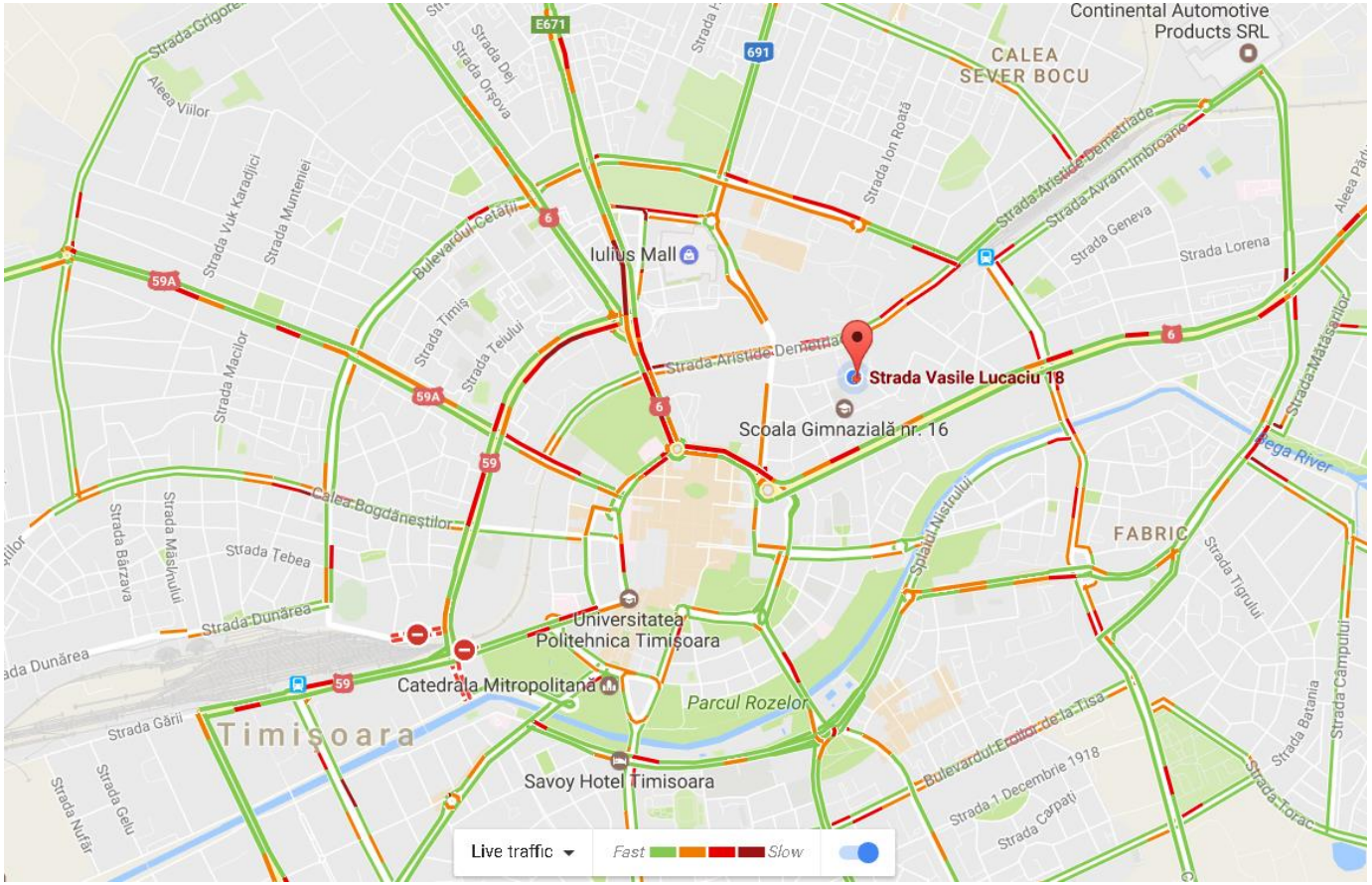
- Cannot be described by separating micro from macro levels



Sursa: http://www.laboiteverte.fr

Other examples

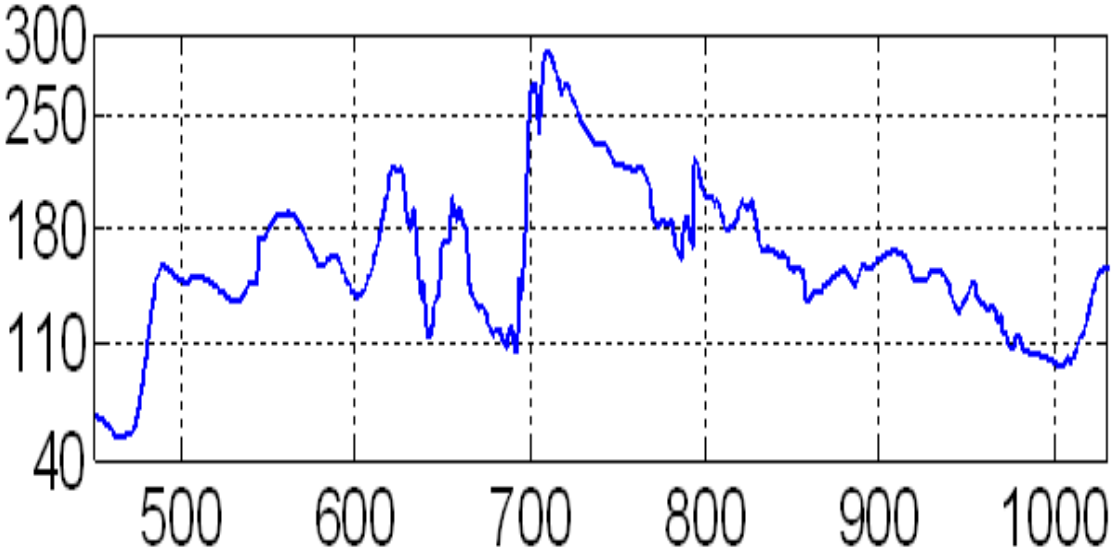
- Urban traffic



Sursa: maps.google.com

Other examples

- Human metabolism



Sursa: M. Ghorbani, P. Bogdan. A cyber-physical approach to artificial pancreas design. CODES+ISSS 2013

Other examples

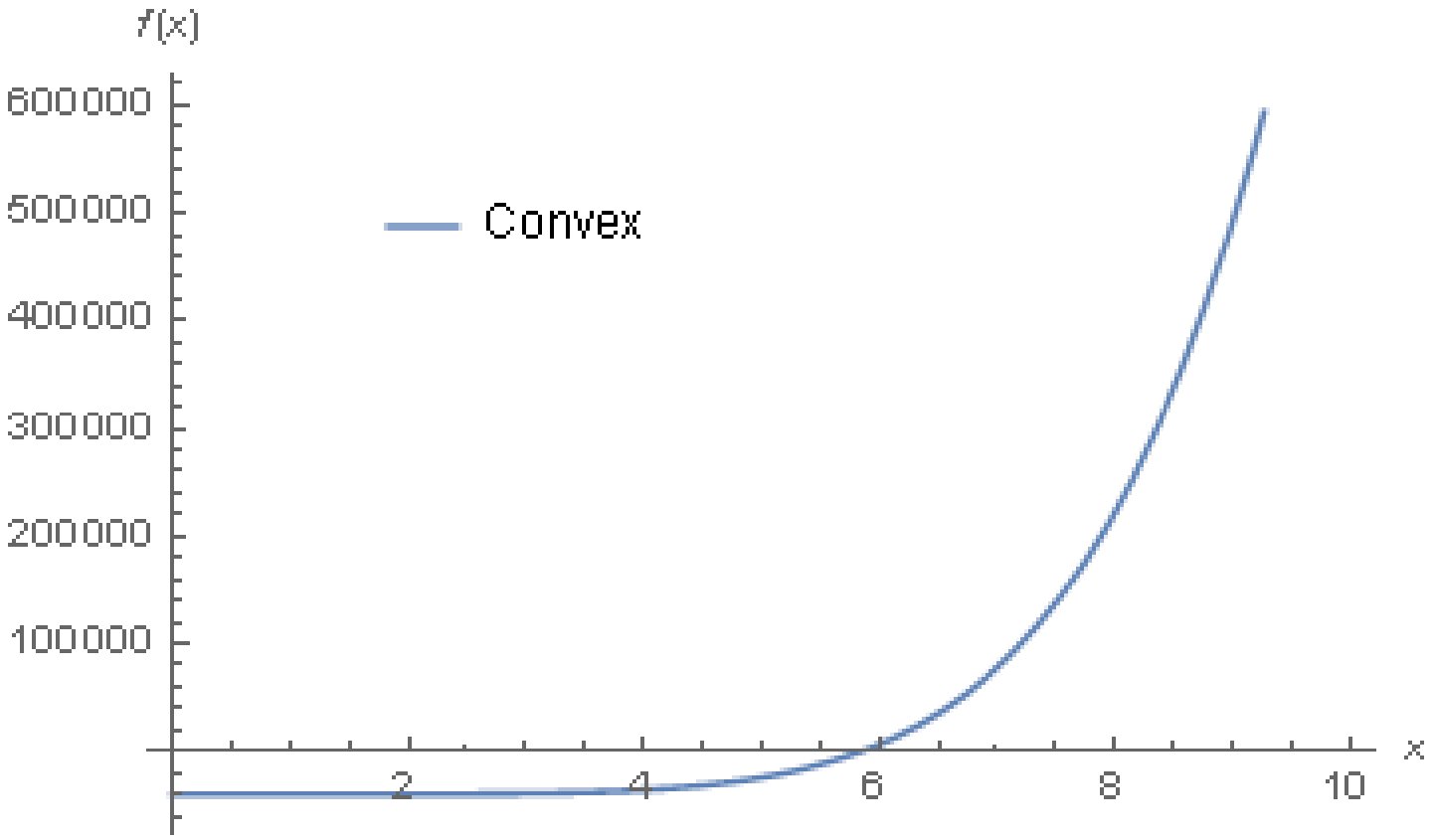
- Markets

Complex systems

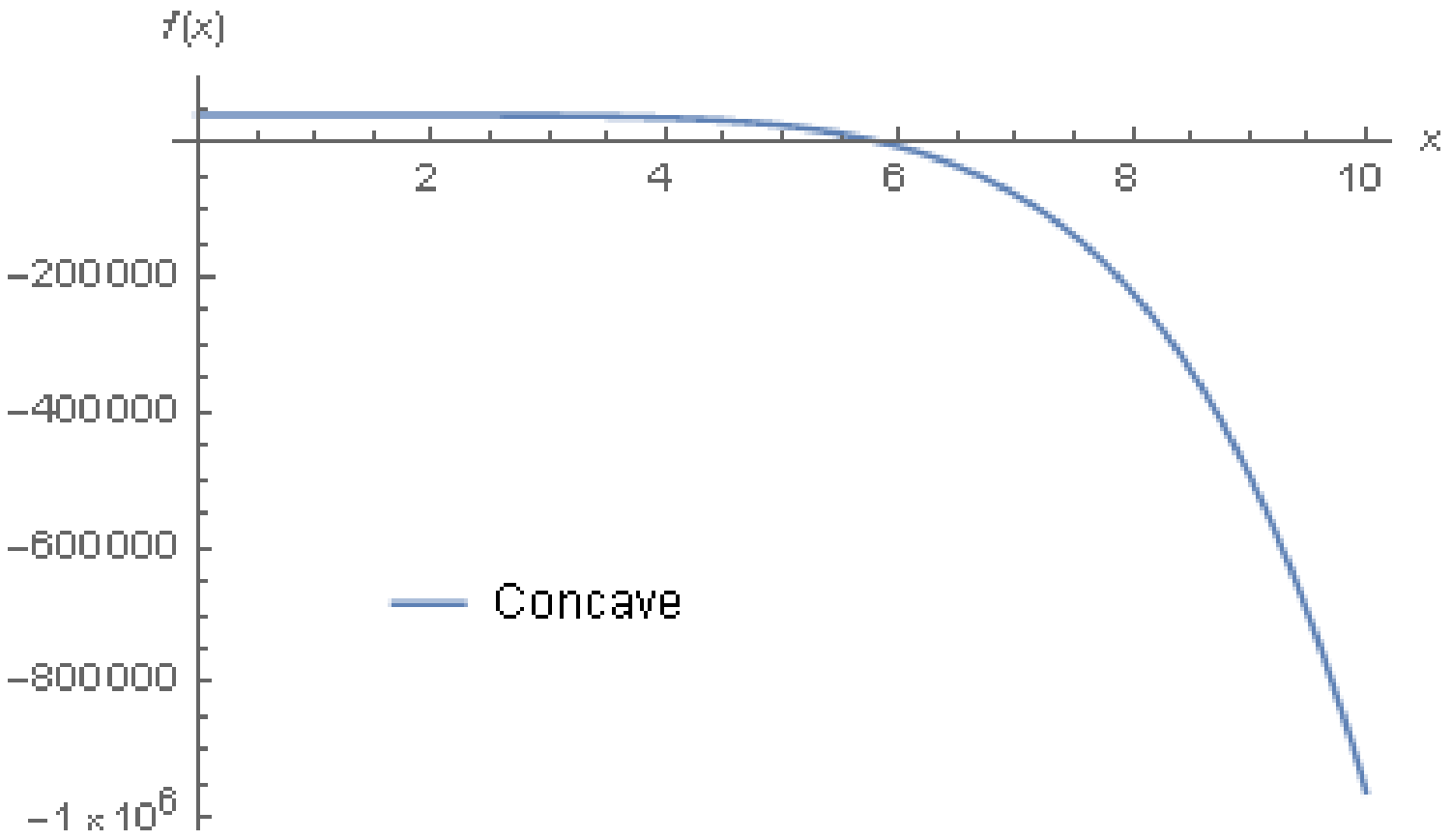
- Components/agents
  - Many
  - Not independent
  - Do not have a coherent behavior
- Aggregate systems
  - Non-linear behavior
  - Asymetric response to interactions (2nd, 3rd order effects)
  - Tend towards antifragility

Gall's Law: *A **complex system** that **works** is invariably found to have **evolved** from a **simple system** that worked. A complex system designed from scratch never works and cannot be patched up to make it work. You have to start over with a working simple system.* – John Gall (1975)

Antifragile-fragile

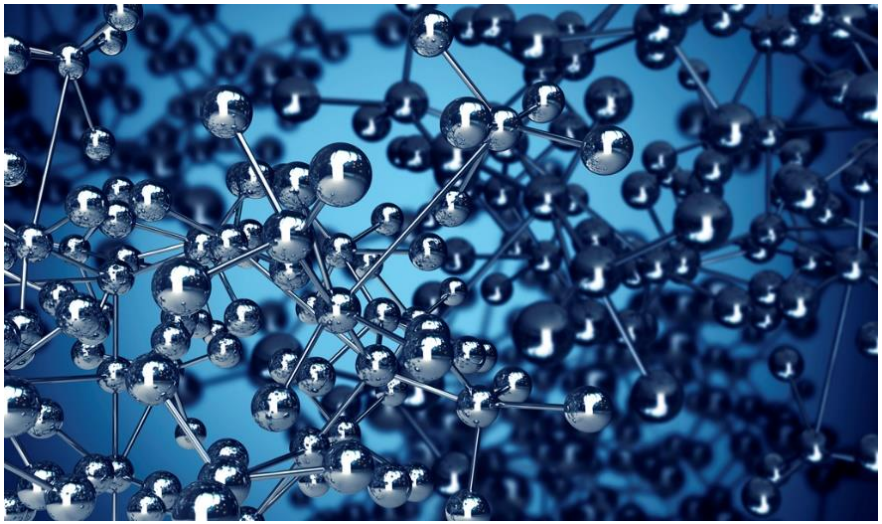


Antifragile-fragile



Statistical physics

- H_2O – properties described by minimizing the free energy of the sysyem, relatively to the observable, macroscopic variables



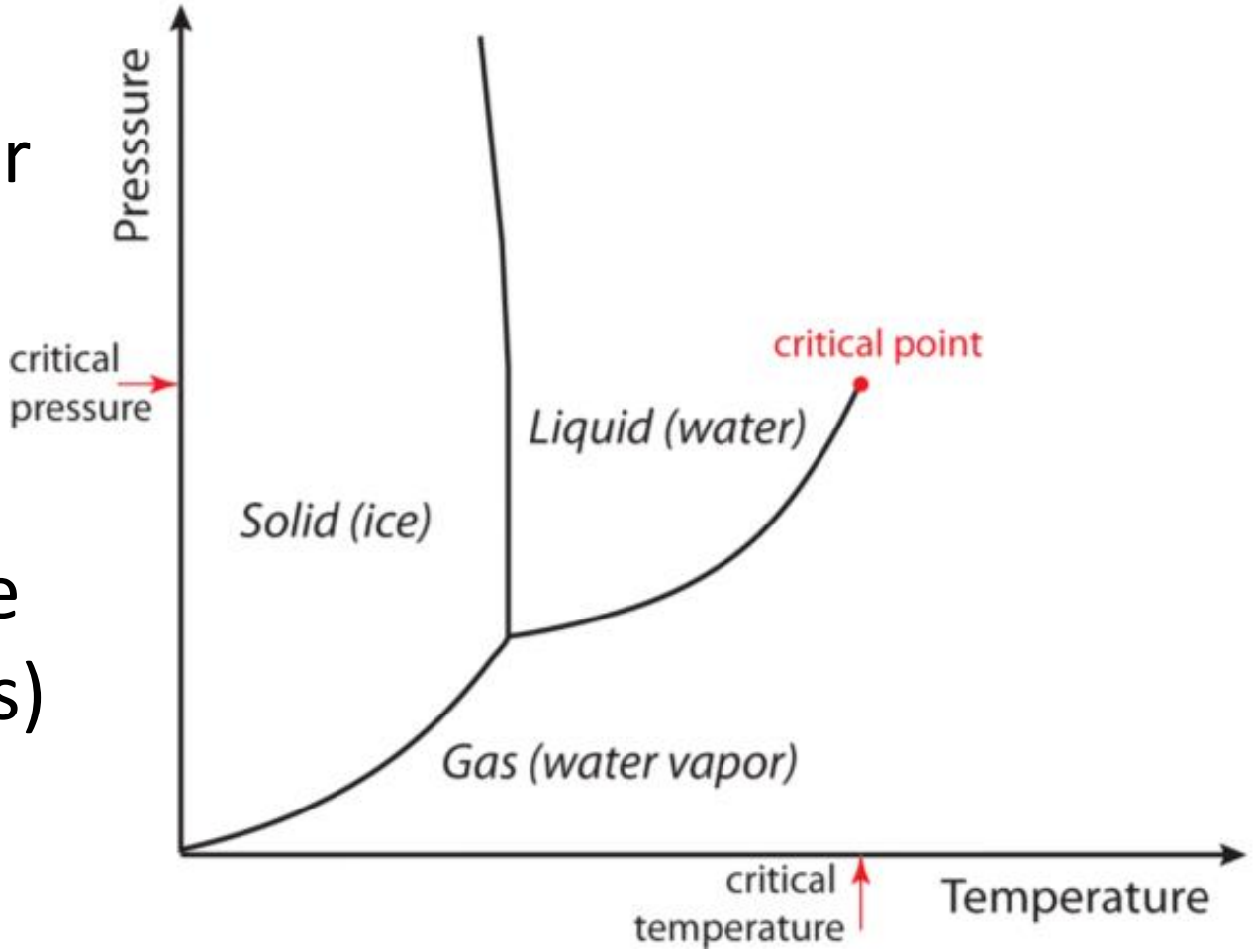
Sursa: http://msutoday.msu.edu

- Pressure
- Temperature
- Volume

These are averages

The average method (***almost***) always works

- Doesn't work for transitions
  - Water – vapor
- 2nd order phase transition points)



Sursa: http://msutoday.msu.edu

Density discontinuity disapears

- Power-law distribution
- Patches (liquid, gas)

$$\rho \propto d^{\gamma}$$

$$\gamma = 0.326$$

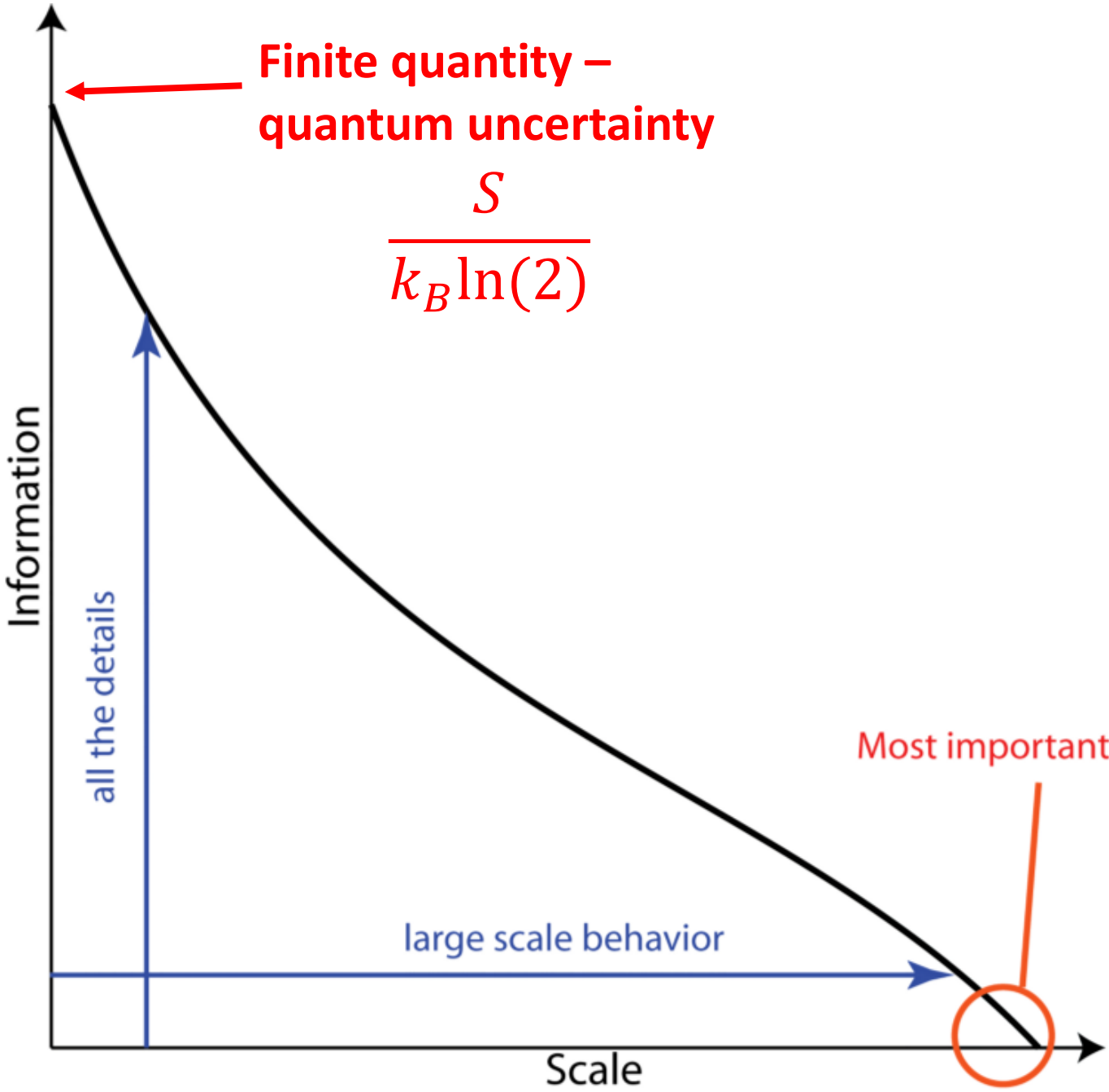
- Landau theory (minimizing free energy): $\gamma = 0.5$
- False premises
  - Matter is homogenous – affects calculus
  - Local averages can be extrapolated – affects statistics

Non-homogenously distributed patches

- **Group renormalization**: a method for identifying relevant parameters for the current scale

Information as function of scale

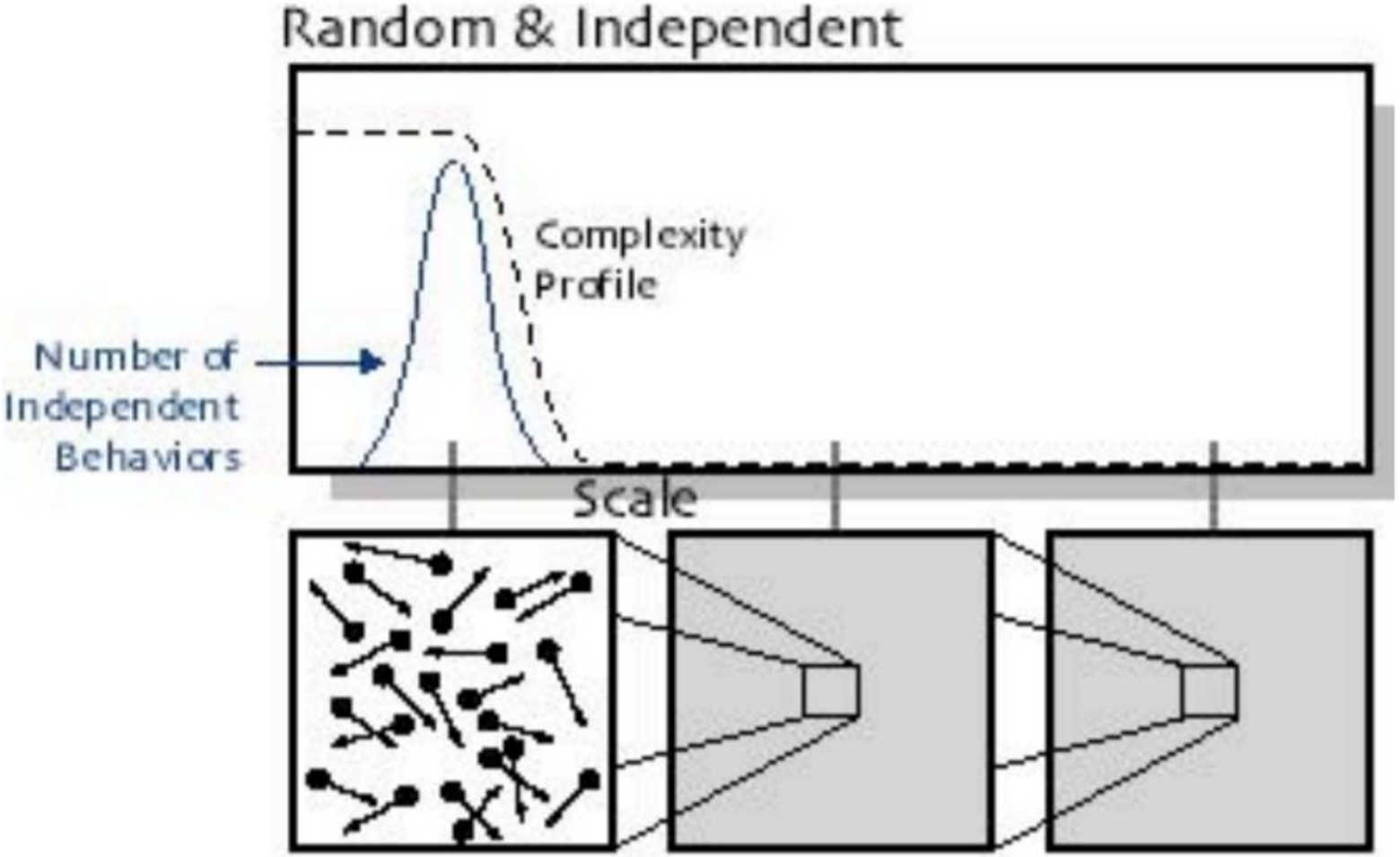
Complexity profile



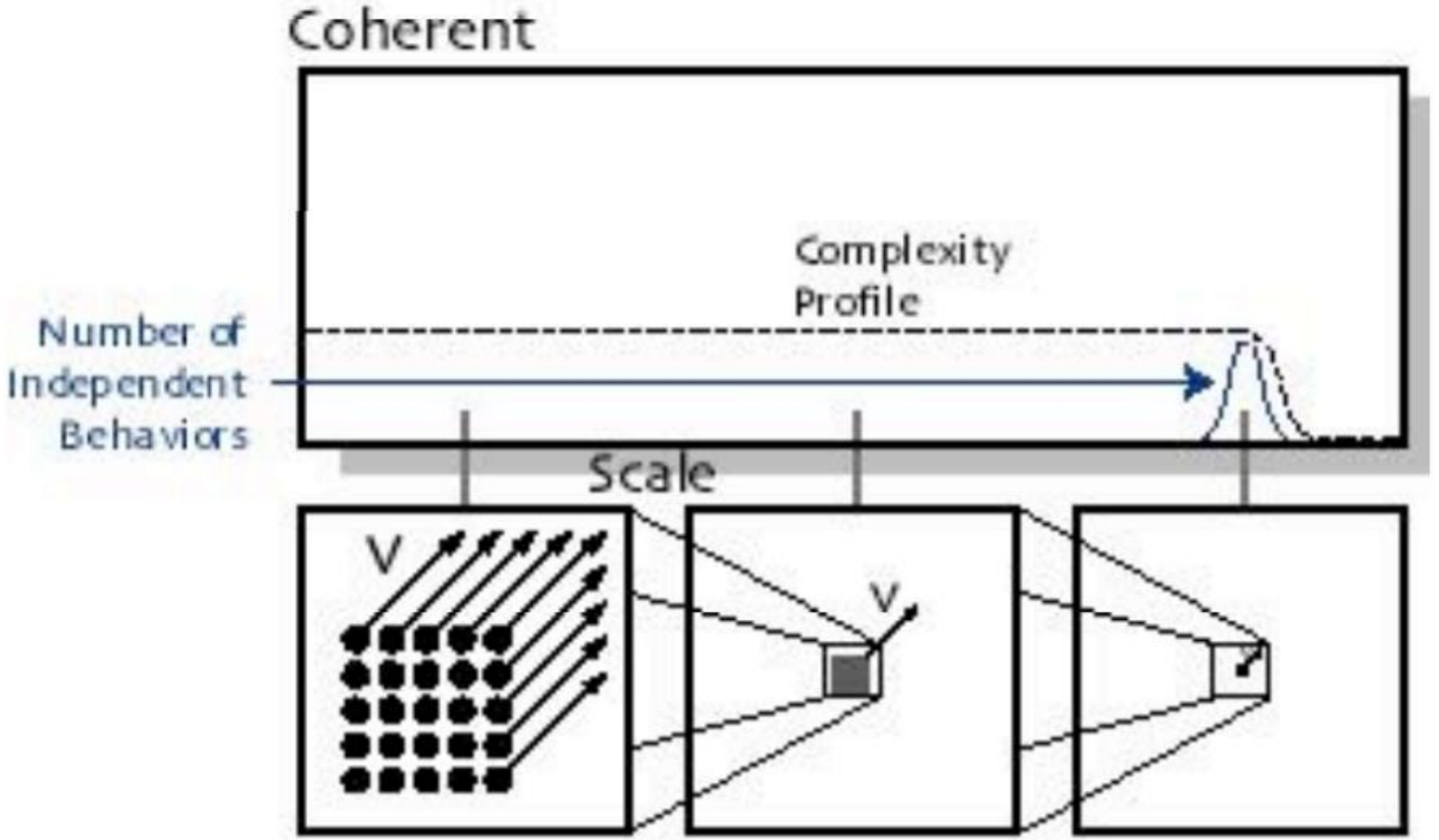
Complexity profile for humans

Y. Bar-Yam, Complexity rising: From human beings to human civilization, a **complexity profile**, Encyclopedia of Life Support Systems, EOLSS UNESCO Publishers, Oxford, 2002

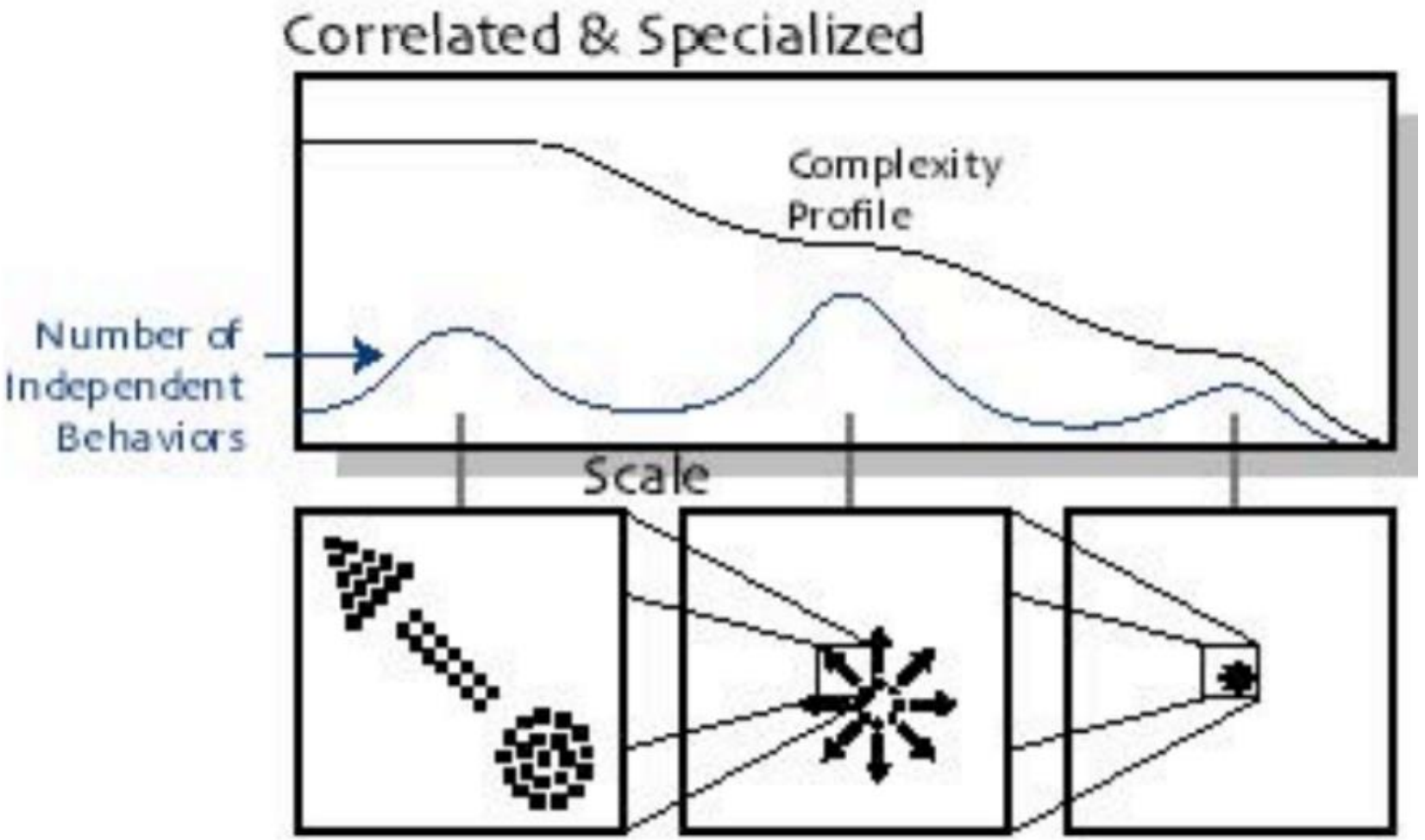
Behavior types



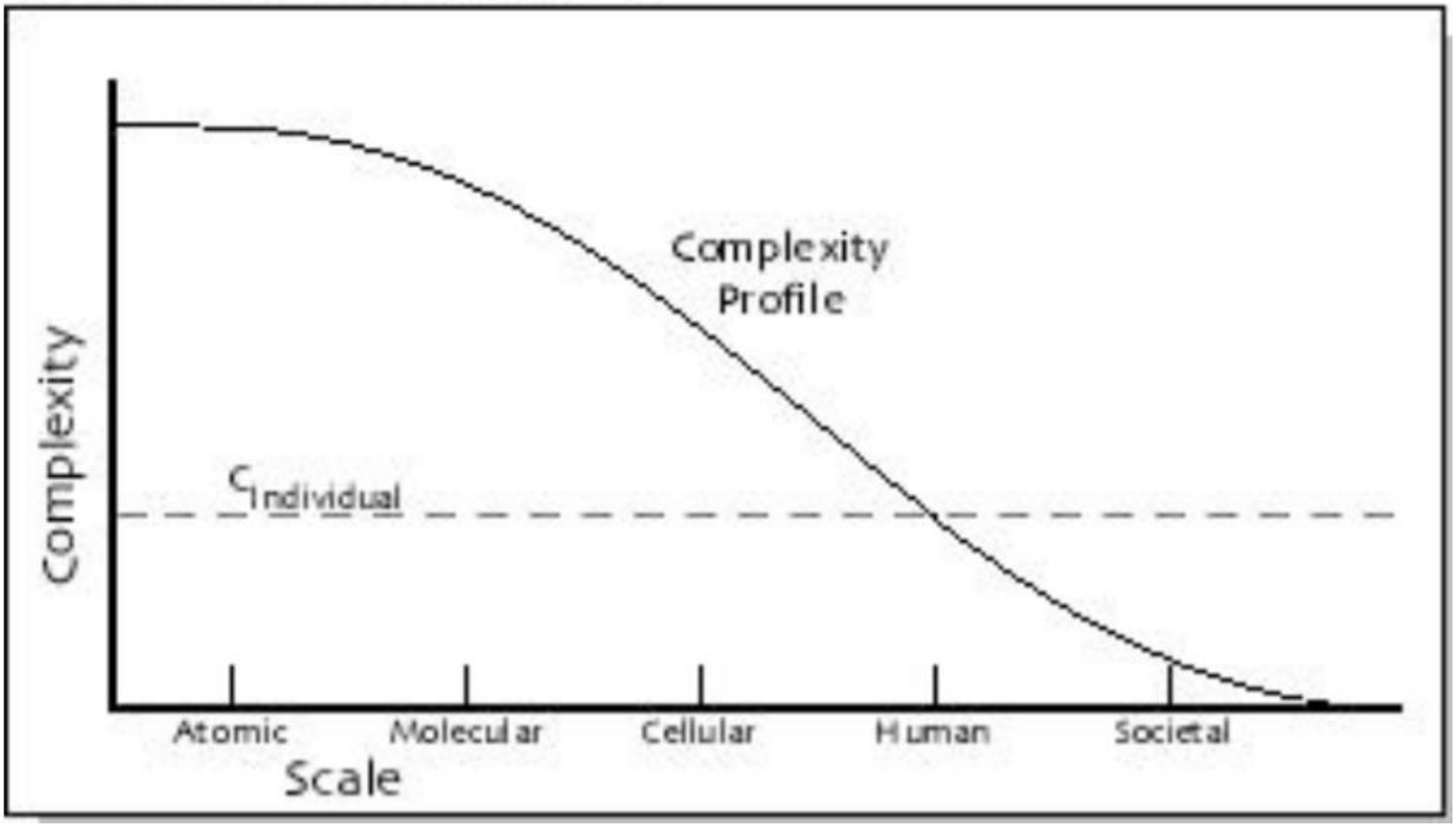
Behavior types



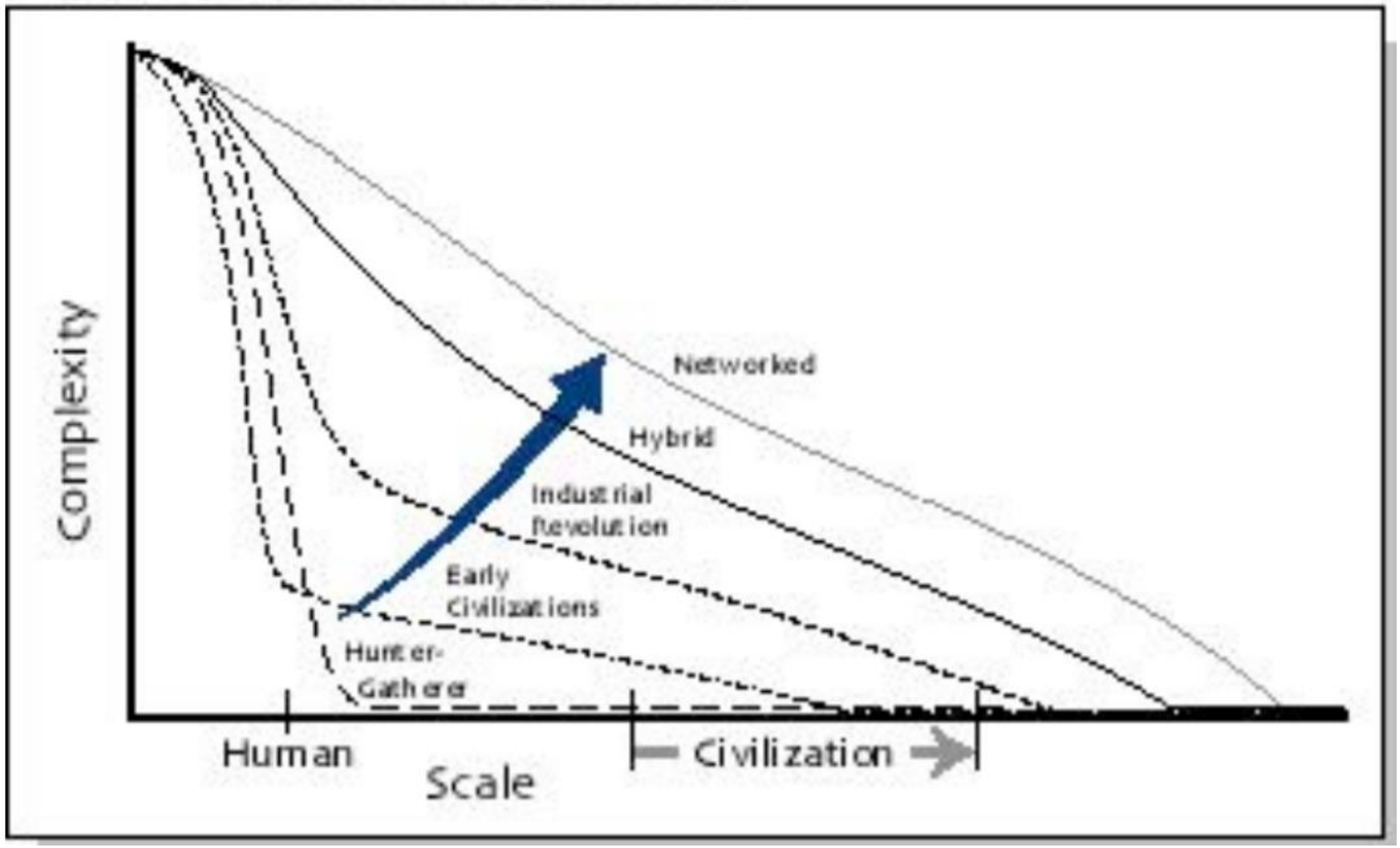
Behavior types



Human complexity



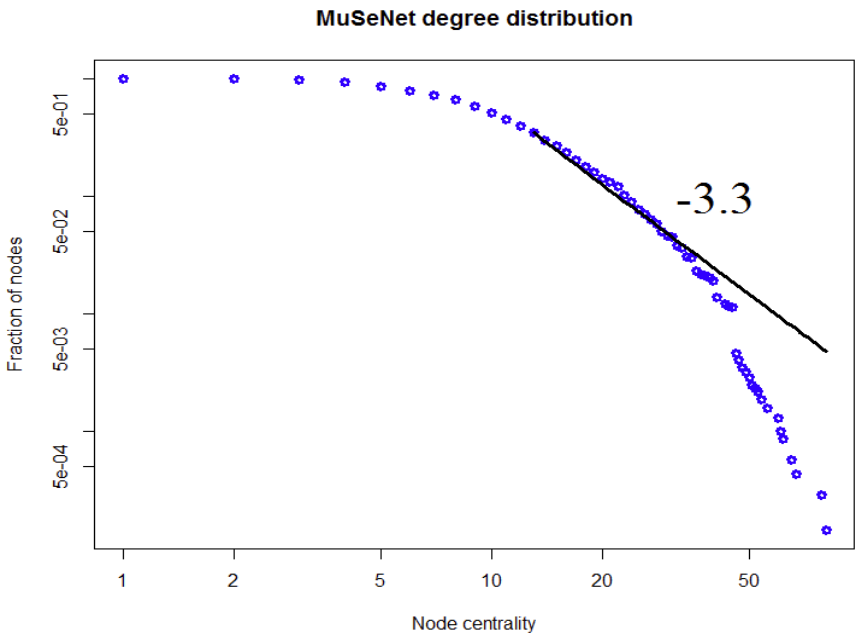
Human complexity evolution

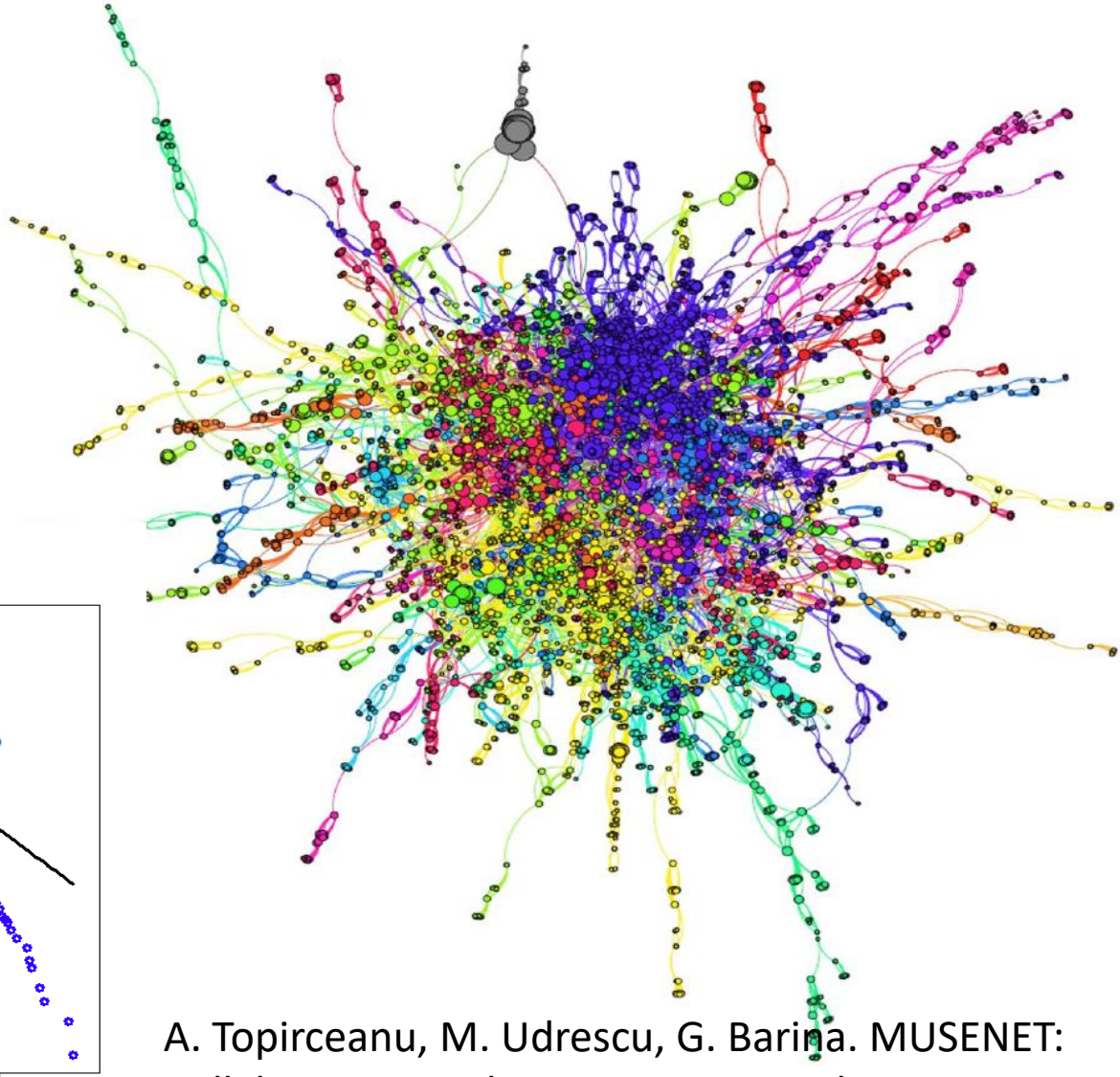


Human collaboration

- MuseNET

$$p(x) = x^{-3.3}$$



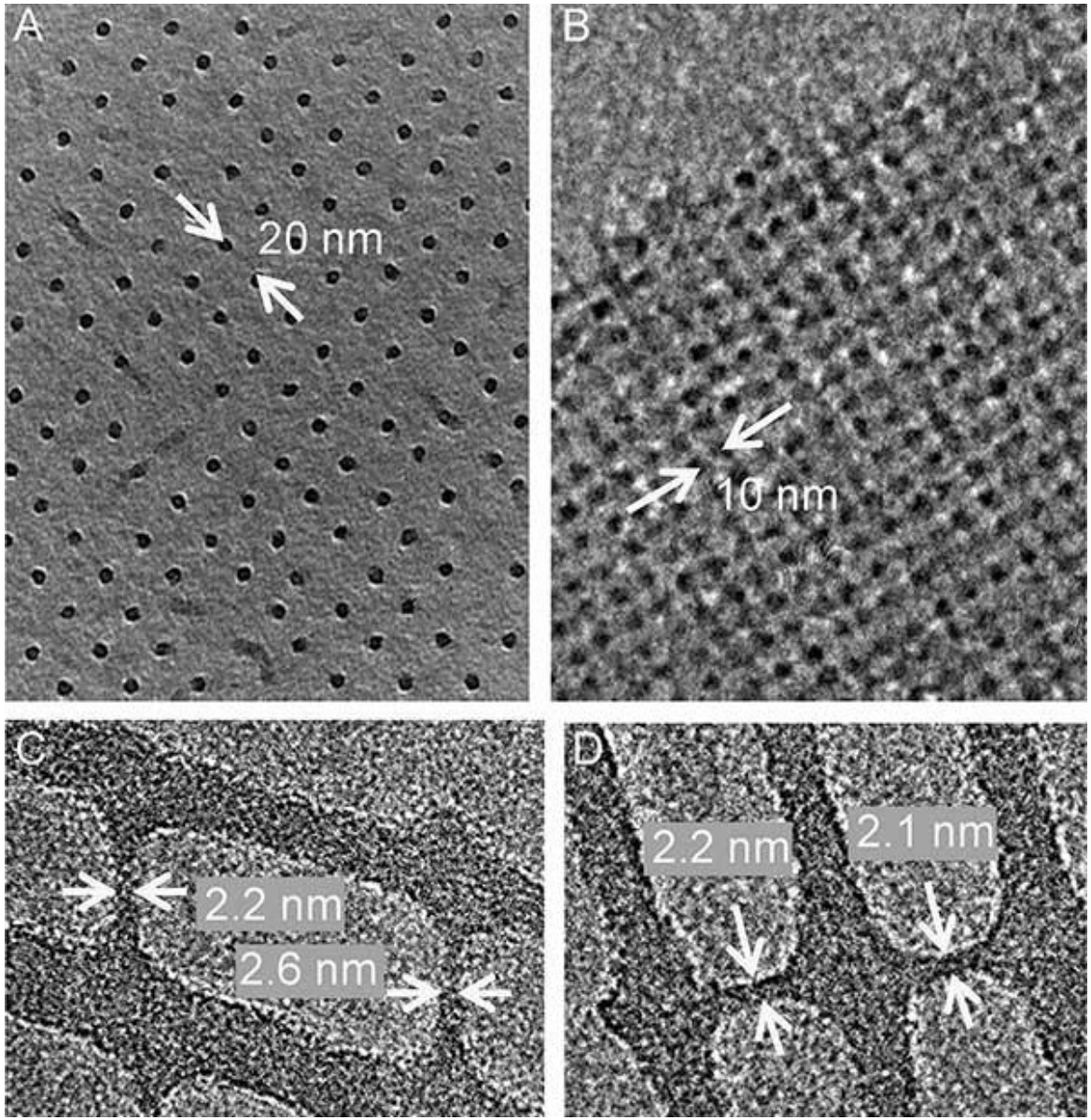


A. Topirceanu, M. Udrescu, G. Barina. MUSENET: Collaboration in the music artists industry. European Conf. Network Intelligence, 2015

Are computers complex systems?

Computer systems: atomic level

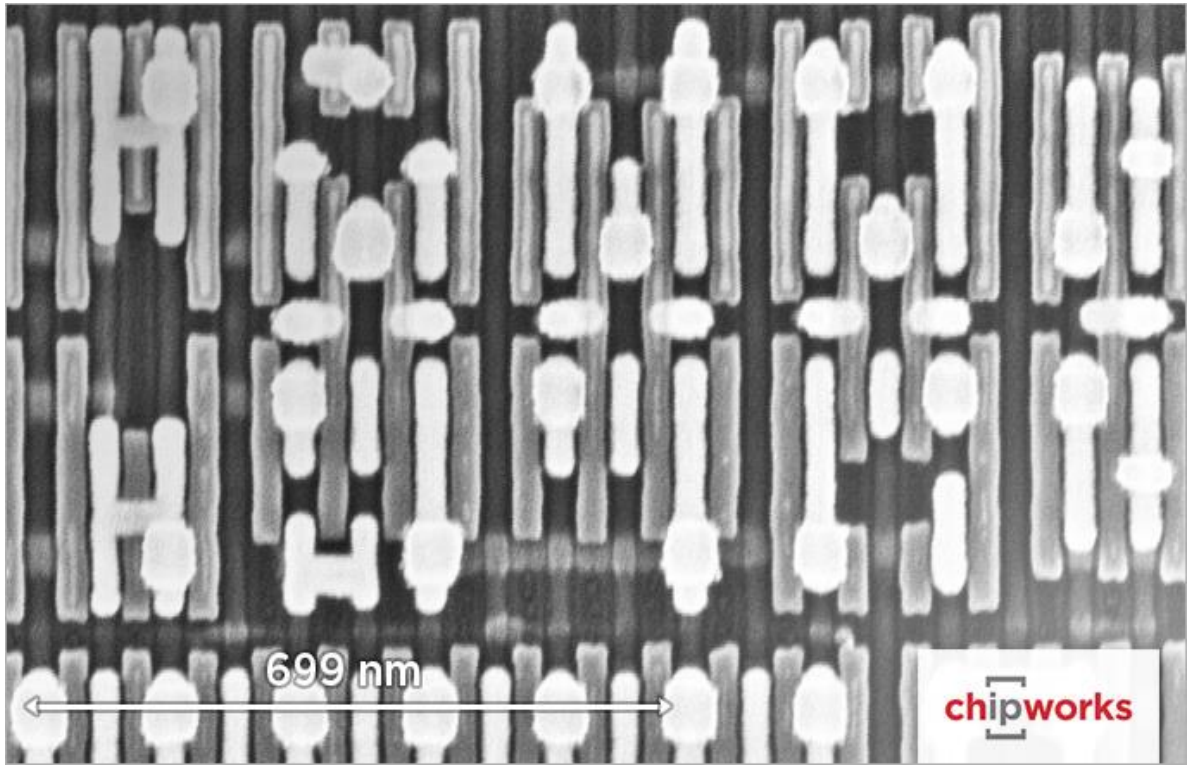
- 1 Si atom:
  $\phi = 0.22$ nm
- Crystallography:
  36 atoms in 14 nm
- AMD Polaris chip:
  $232mm^2$
- **$\approx 10^{12}$ atoms/chip**



Sursa: V. R. Manfrinato et. al. Resolution Limits of Electron-beam Lithography toward the Atomic Scale. MIT Annual Research Report, 2013

Computer systems: transistor-level

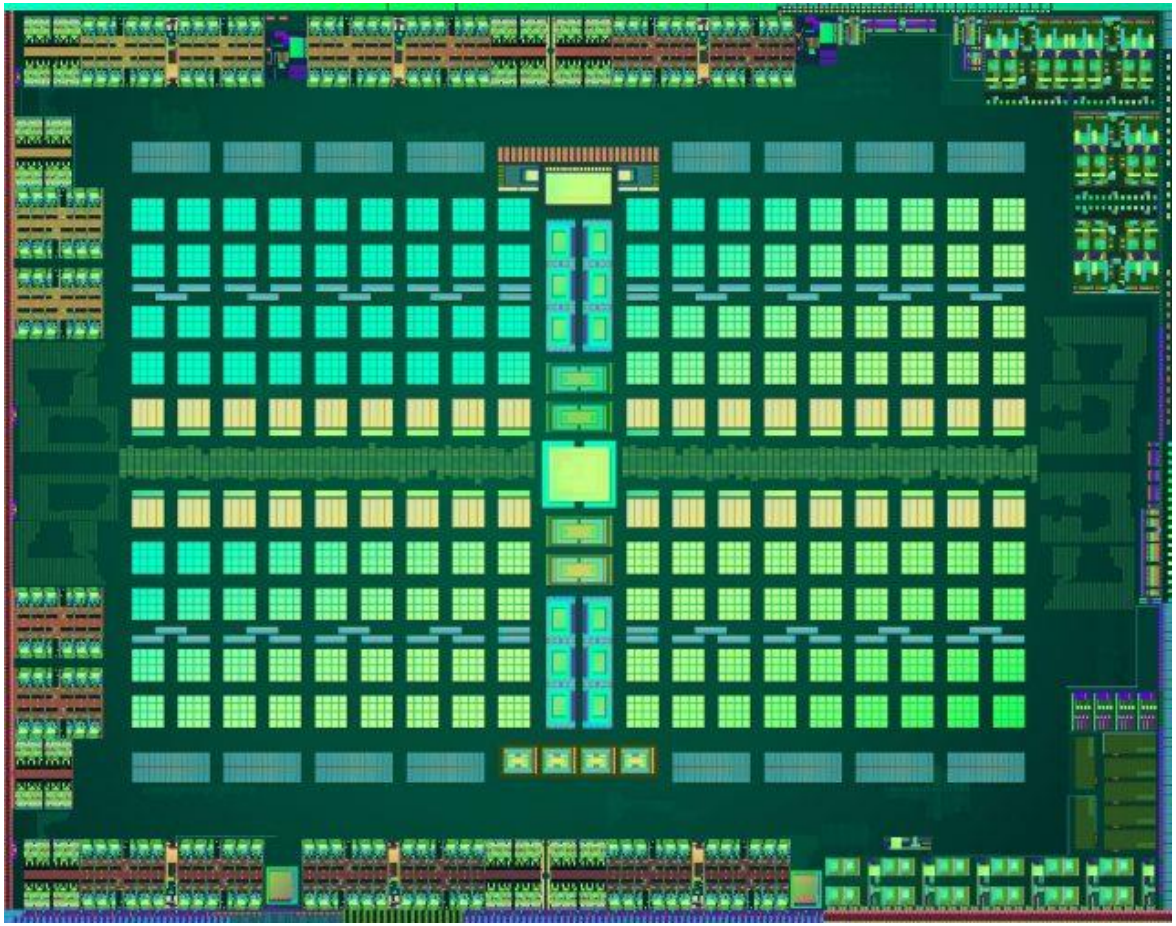
- Tranzistors FinFET 14nm, Broadwell chip
  - **5.7×10^9** in AMD Polaris Radeon RX 480(Ellesmere X)



Sursa: chipworks

Computer system: chip level

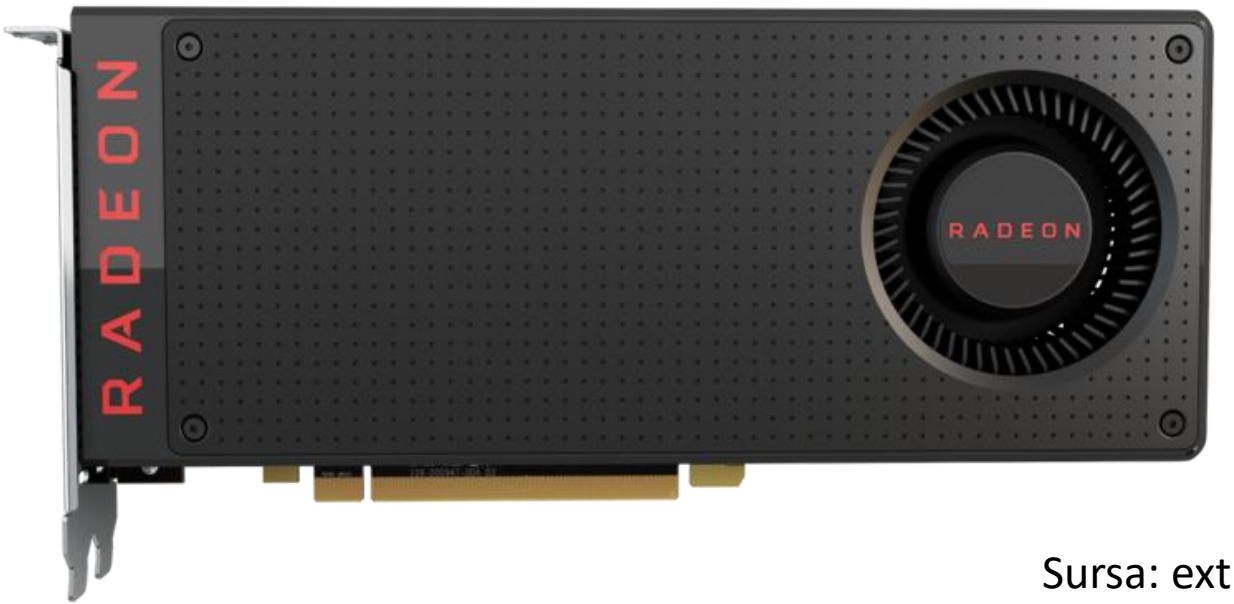
- AMD Polaris: 2304 GPU cores, 232 $mm^2 \approx (16\ mm)^2$



Sursa: extremetech.com

Computer system: system level

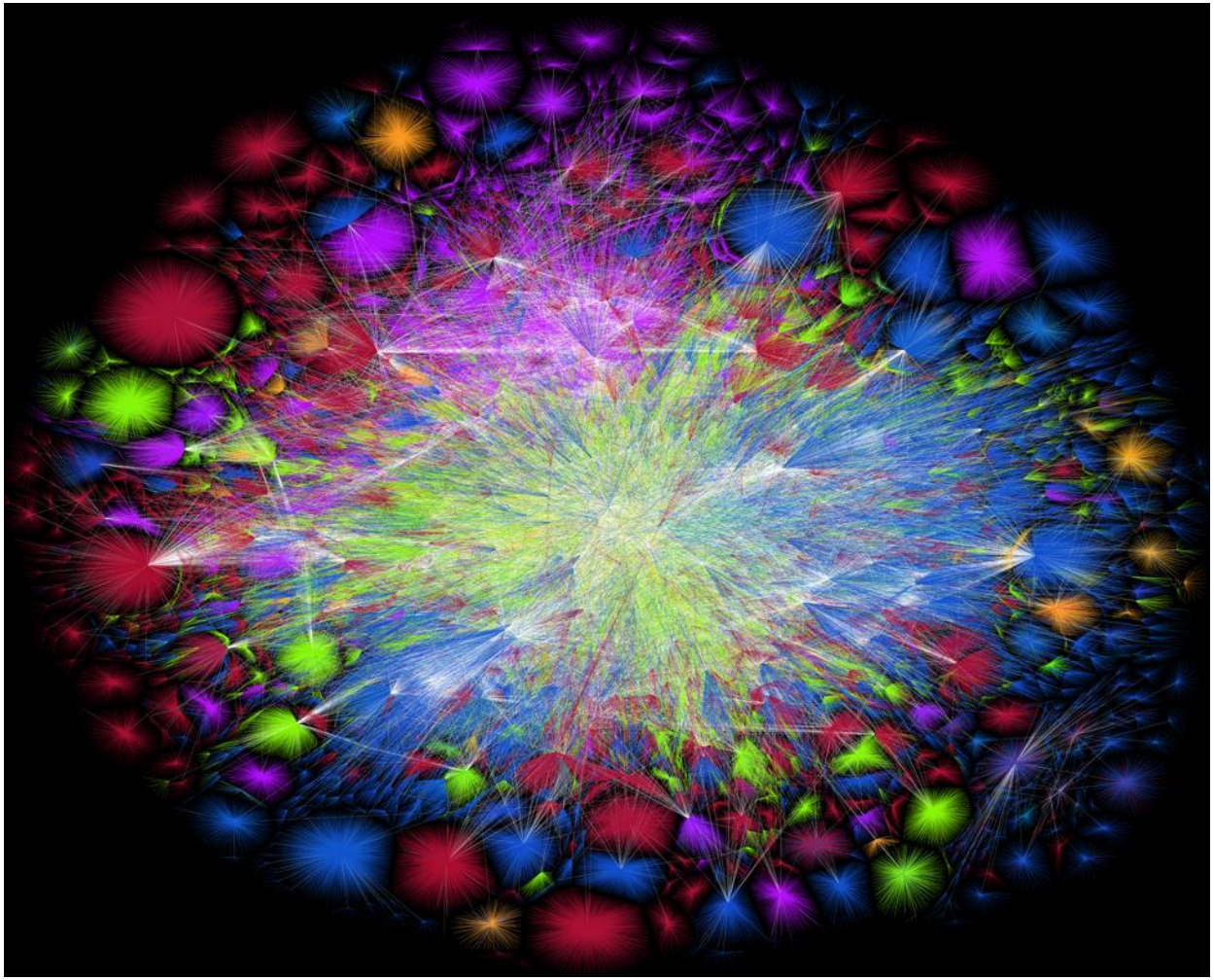
- Performance: 5834 GFLOPS > 5TFLOPS = 5×10^{12}FLOPS
  - Information: 2304 cores×16KB core state ≈ **36×10^{6}**
  - Scală: 0.241m



Sursa: extremetech.com

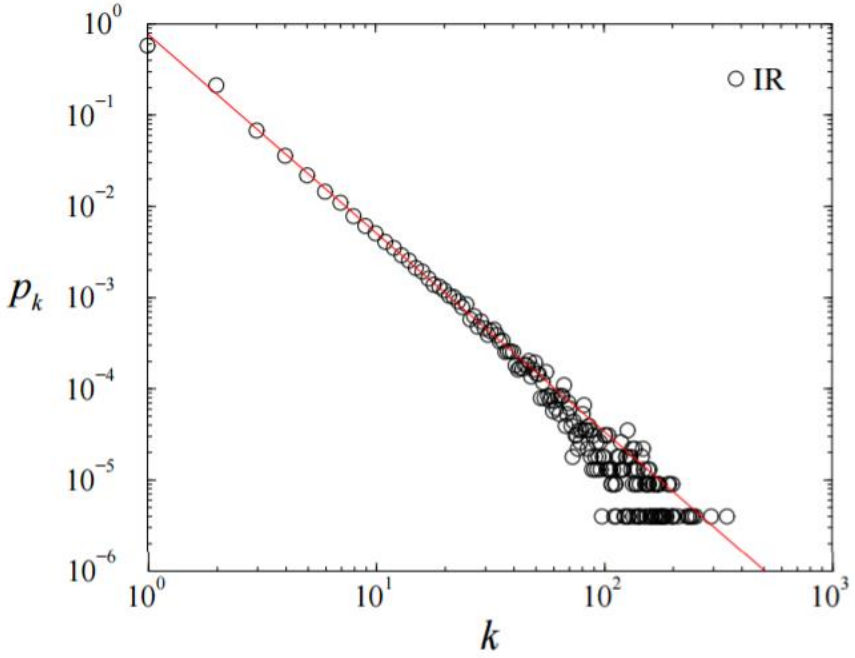
Computer system: network level

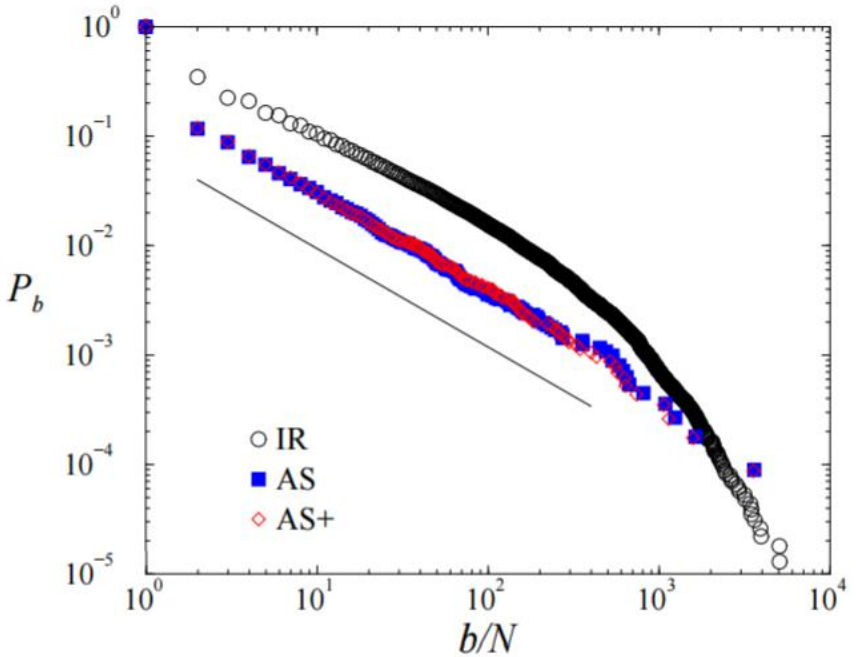
- Internet 2015
- 2015: IPv4 used all adresses
  - $2^{32} = 4 \times 2^{30} \approx 4 \times 10^{9}$
- IPv6
  - $2^{128} \approx 4.3 \times 10^{38}$
- Scale: 12,742 km (Earth diameter)



Sursa: opte.org

Computer system: network level





- Power law de ***degree*** and ***betweenness*** distributions
- $p(x) = x^{-\gamma}$
  - $\gamma = 2.1$ (degree)
  - $\gamma = 1.9$ (betweenness)

Sursa: A. Vazquez et al. Internet topology at the router and autonomous system level . ArXiv 2002

Computer system: global network

- "Temperature" of the Internet
  - Internet bandwidth

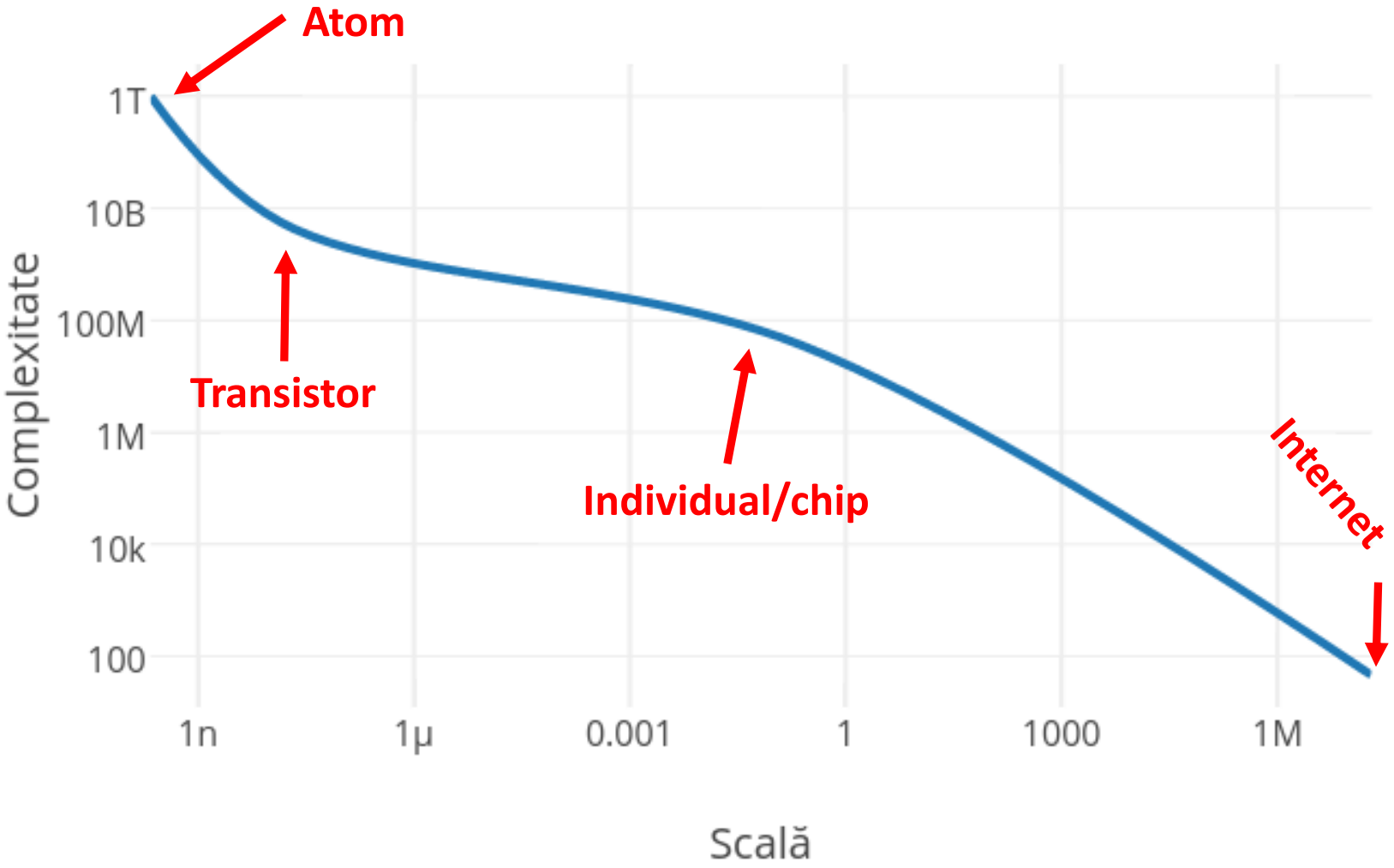
Cisco Visual Networking Index

$$\frac{\text{Data sent in 1 year}}{\text{Seconds in 1 year}} = \frac{1.1\times 2^{70}\ [\text{bytes/year}]}{0.03154\times 2^{30}\ [\text{seconds}]}$$

$=34.8\times 2^{40}$ bytes/second = **34.8 Tbytes**/second

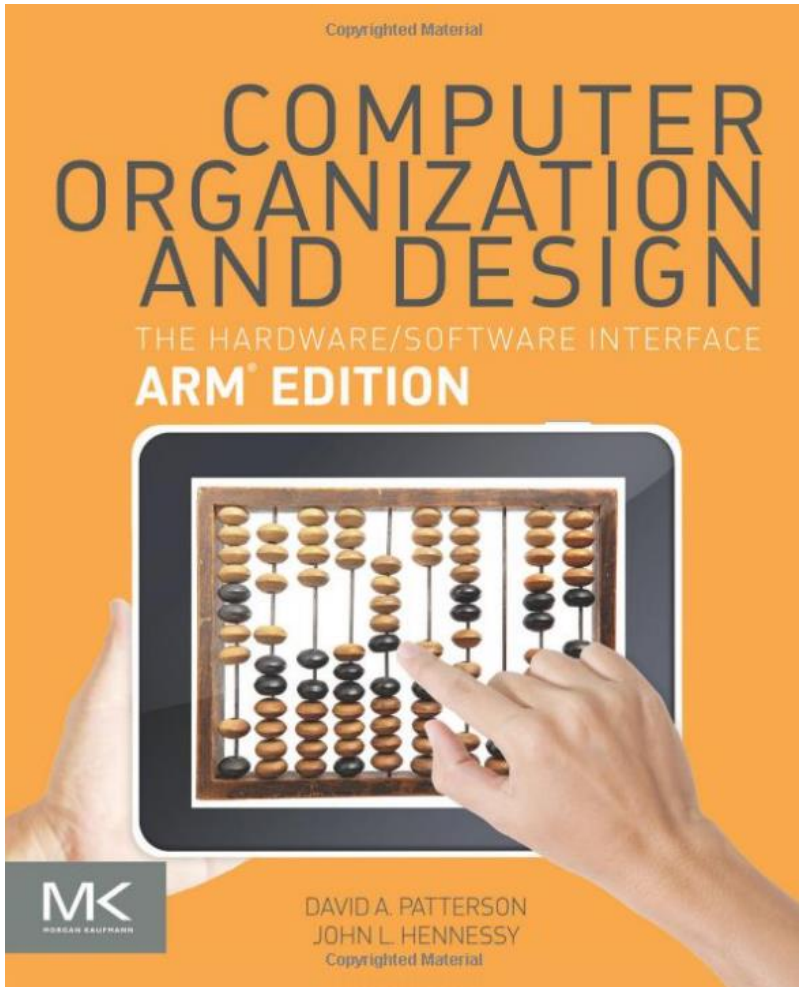
  - M. Hilbert, P. Lopez. The World's Technological Capacity to Store, Communicate, and Compute Information, Science, 2011

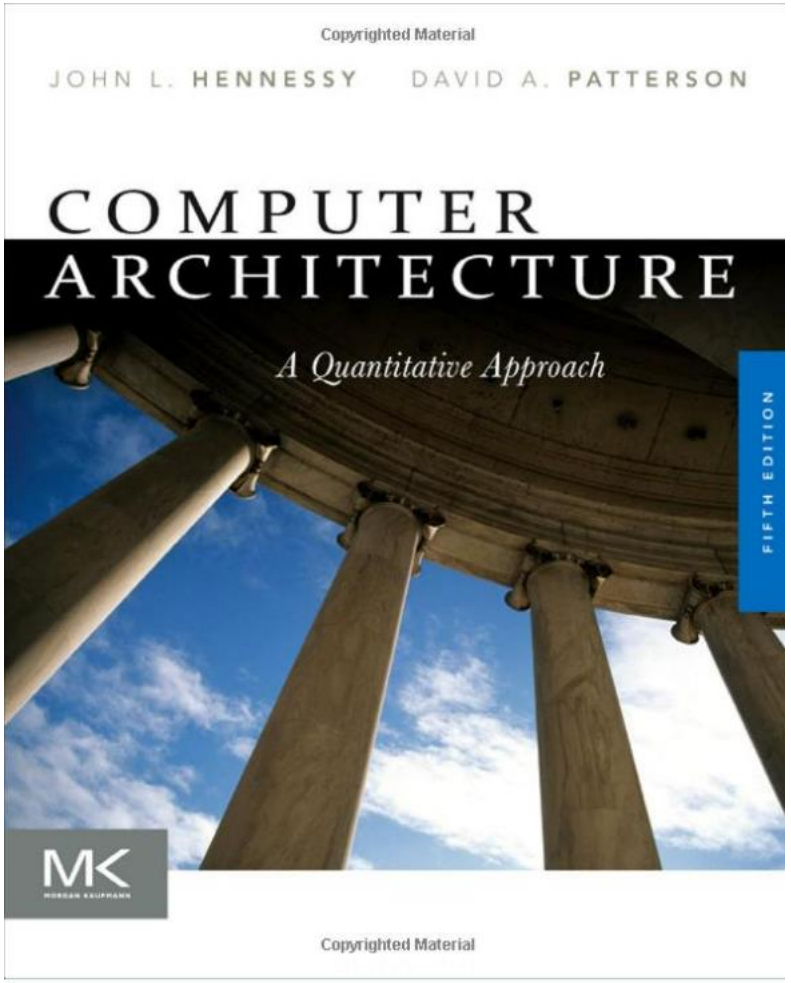
Complexity profile for computer systems



Computers are complex systems

Textbooks


Copyrighted Material
COMPUTER ORGANIZATION AND DESIGN
THE HARDWARE/SOFTWARE INTERFACE
ARM® EDITION
DAVID A. PATTERSON
JOHN L. HENNESSY
Copyrighted Material


Copyrighted Material
JOHN L. HENNESSY DAVID A. PATTERSON
COMPUTER ARCHITECTURE
A Quantitative Approach
FIFTH EDITION
Copyrighted Material