Information Technology Systems in a 'Post-Silicon' World

Dr. Bernard S. Meyerson, IBM Fellow & Chief Innovation Officer

Progress has been Astonishing

Every generation of technology enabled remarkable outcomes

45 Years,
62M x RAM

How Much More of Moore?

Apollo 11:
2048 words RAM (16-bit word)-> ~4KB
36,864 words ROM

Average Smart Phone:
256 MB – 512 MB Cache
2 GB – 64 GB RAM

For almost 5 decades we have relied upon Moore's Law for exponential improvements in Silicon Technology, however …

If automobiles were like chips …

1970: 15 miles per gallon

1980: 1,500 miles per gallon

1990: 150,000 miles per gallon

2000: 15 Million miles per gallon

2010: 150 Million miles per gallon

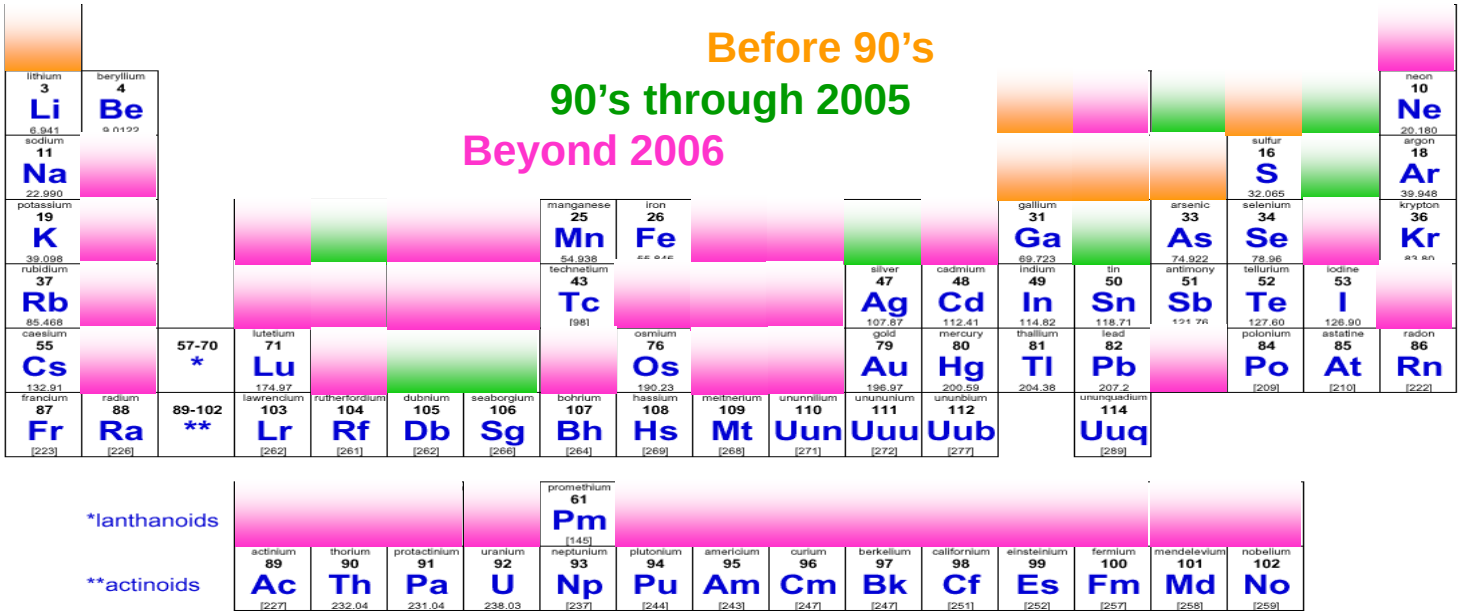
Perhaps this is a bit much??

2020: 1 Billion mpg miles per gallon

After 50 Years, How long before Technology as we know it has run its course?

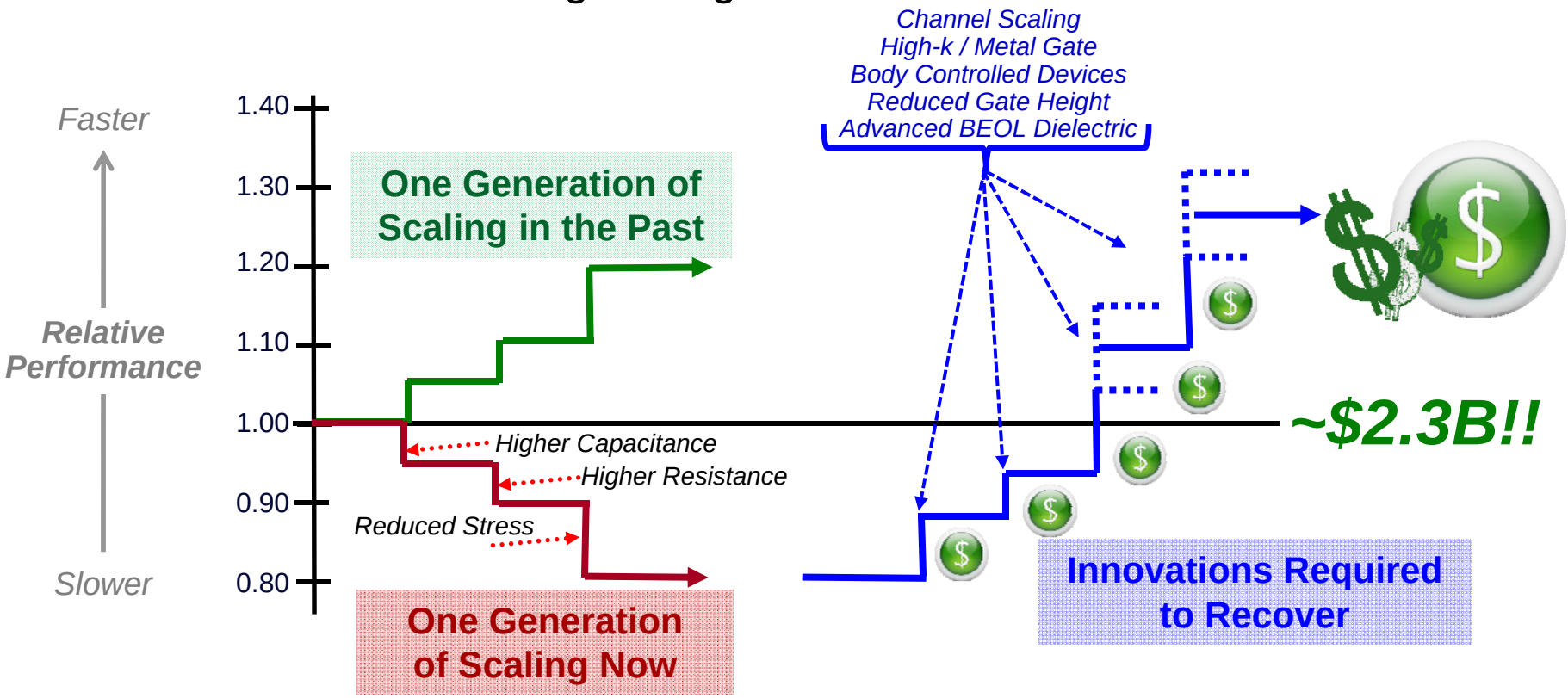
How can you tell the world is now different?

Elements Employed in Silicon Technology

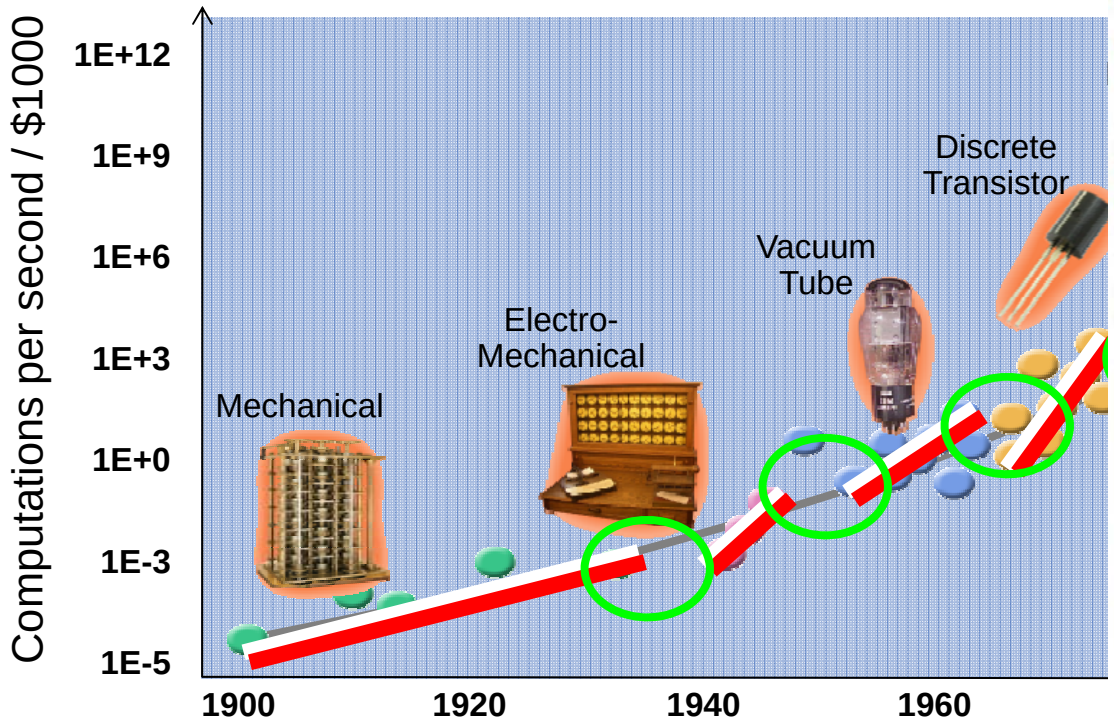


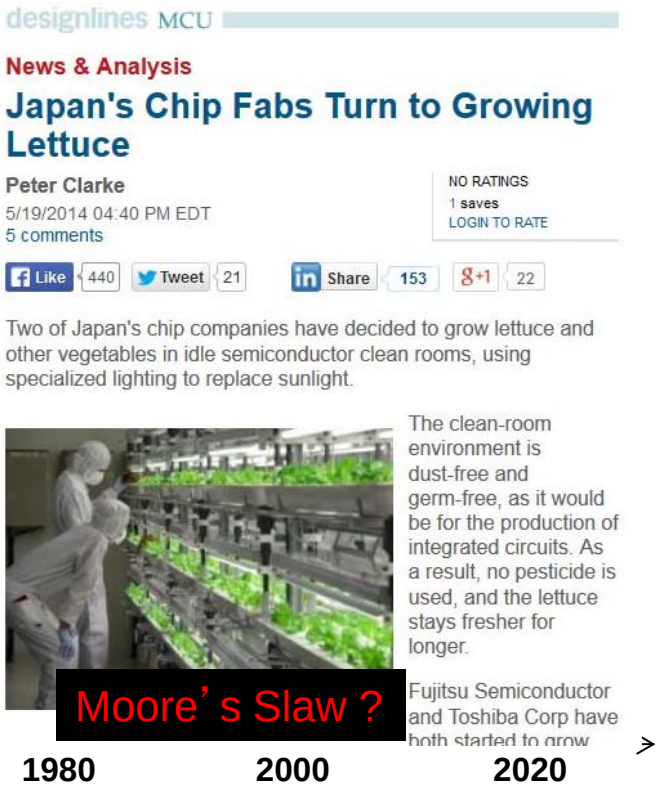
A Discontinuity

Simply scaling technology no longer makes it faster or less costly, although it might set fire to the user



Continuous and Discontinuous Innovation





designlines MCU

News & Analysis

Japan's Chip Fabs Turn to Growing Lettuce

Peter Clarke
5/19/2014 04:40 PM EDT
5 comments

NO RATINGS
1 saves
LOGIN TO RATE

Like 440 Tweet 21 Share 153 +1 22

Two of Japan's chip companies have decided to grow lettuce and other vegetables in idle semiconductor clean rooms, using specialized lighting to replace sunlight.

The clean-room environment is dust-free and germ-free, as it would be for the production of integrated circuits. As a result, no pesticide is used, and the lettuce stays fresher for longer.

Fujitsu Semiconductor and Toshiba Corp have both started to grow

Moore's Slaw ?

Coming “Soon” …

Information Technology in the “Post-Silicon Era”

- Silicon transistors will dominate Information Technology for decades to come, but contribute little to its progress.
- At 186,000 miles per second, light is **far** too slow, so we will need to “fix” that problem.
- Fundamentally new system architectures consisting of specialized hardware, software, and network functionality, will emerge to compensate.
  - **And must be hardened against what is coming.**
    - **Every innovation creates both value and risk.**

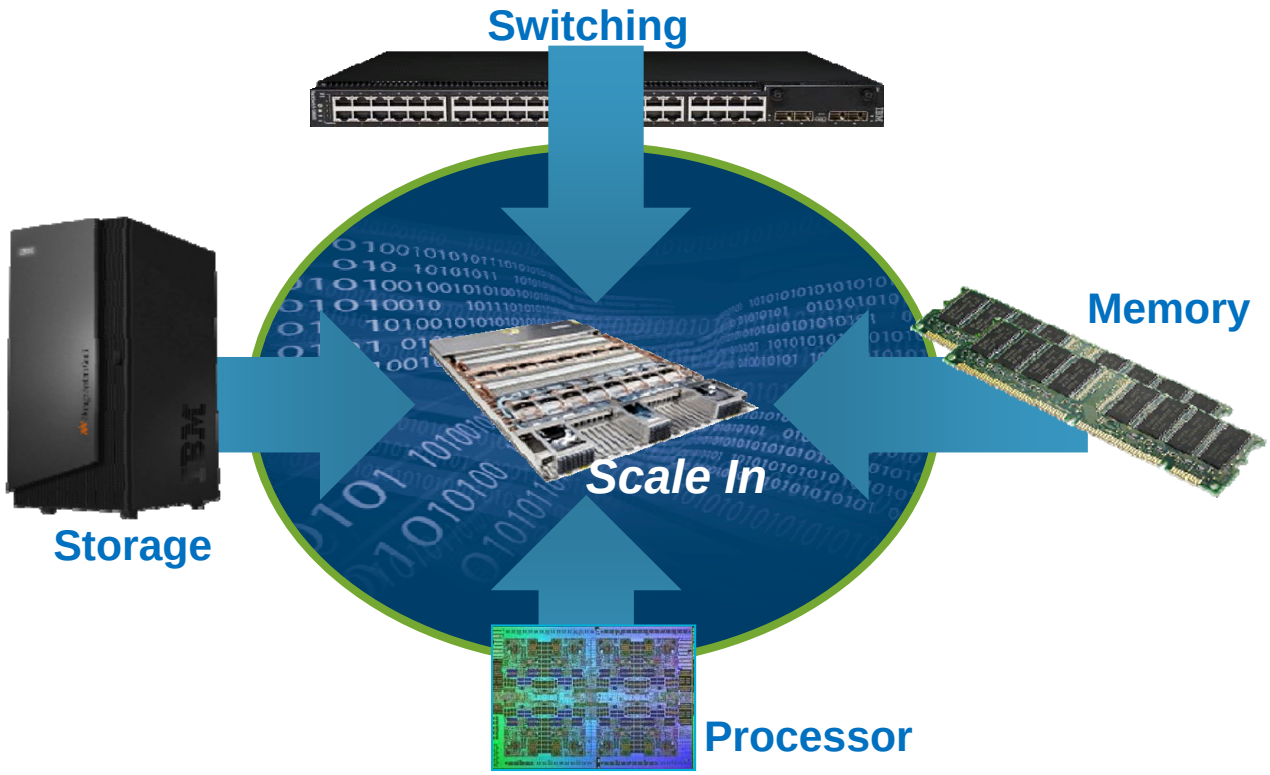
One Example of Innovation;

If you can't make light faster,

Integrate Everything

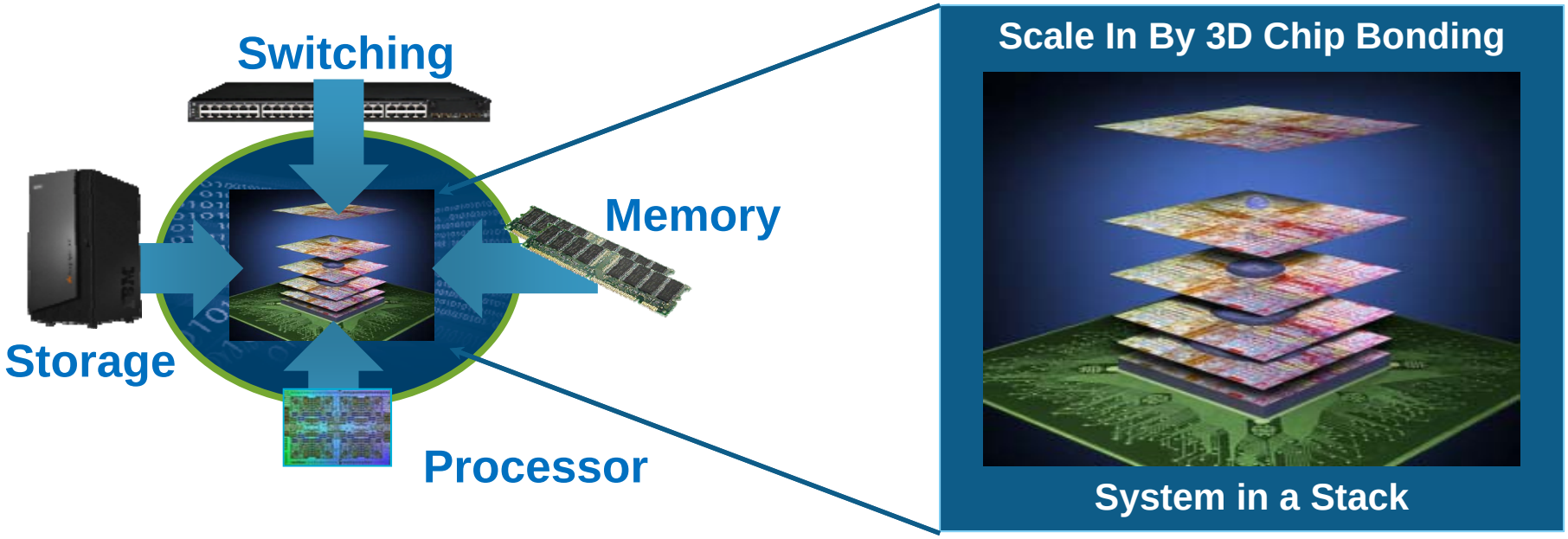
A New, Near Term Technology Strategy

Scale In by System Integration



The Next Several Decades of Innovation;

Scale-in by 3-D Integration

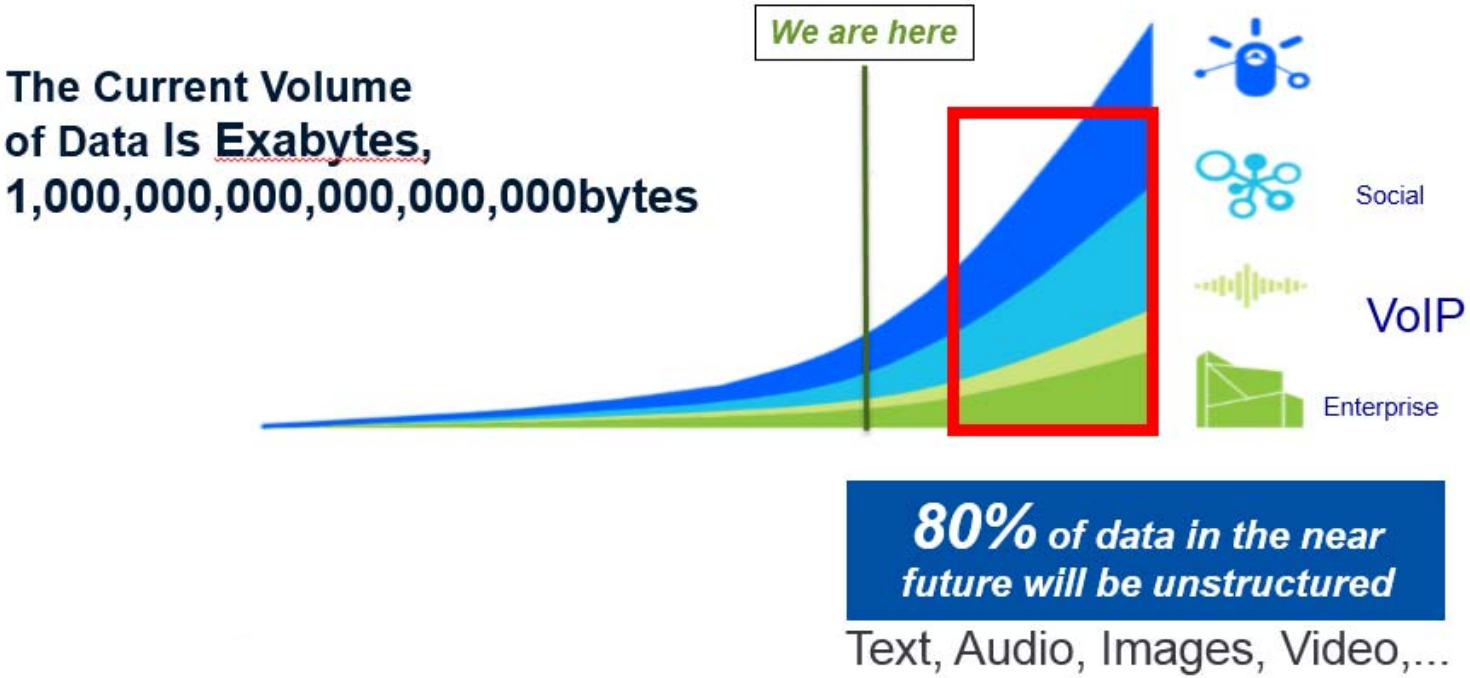


A Sobering Realization

*At the very time we are entering the era of **Big Data**, vast amounts of data being created in an instrumented and interconnected world, silicon technology as we have known it has run its course.*

Data Scale, Velocity, and Veracity, Are ALL Challenges We Must Overcome

Asking The Obvious; Where does the technological horsepower come from to power solutions to these challenges, and WHY is it urgent?



Why Is It Vital We Drive The Ability To Extract Knowledge From “Big Data”

To Save Lives We Must Anticipate The Future And Change It

- In April 2010, the City of Rio faced one of the worst heavy rainfalls in history; flooding and landslides left at least 110 people dead,
- Every summer, the city has to handle the consequences of an intense rainy season, and the mayor and government of Rio De Janeiro determined that the disaster of 2010 must never be repeated.

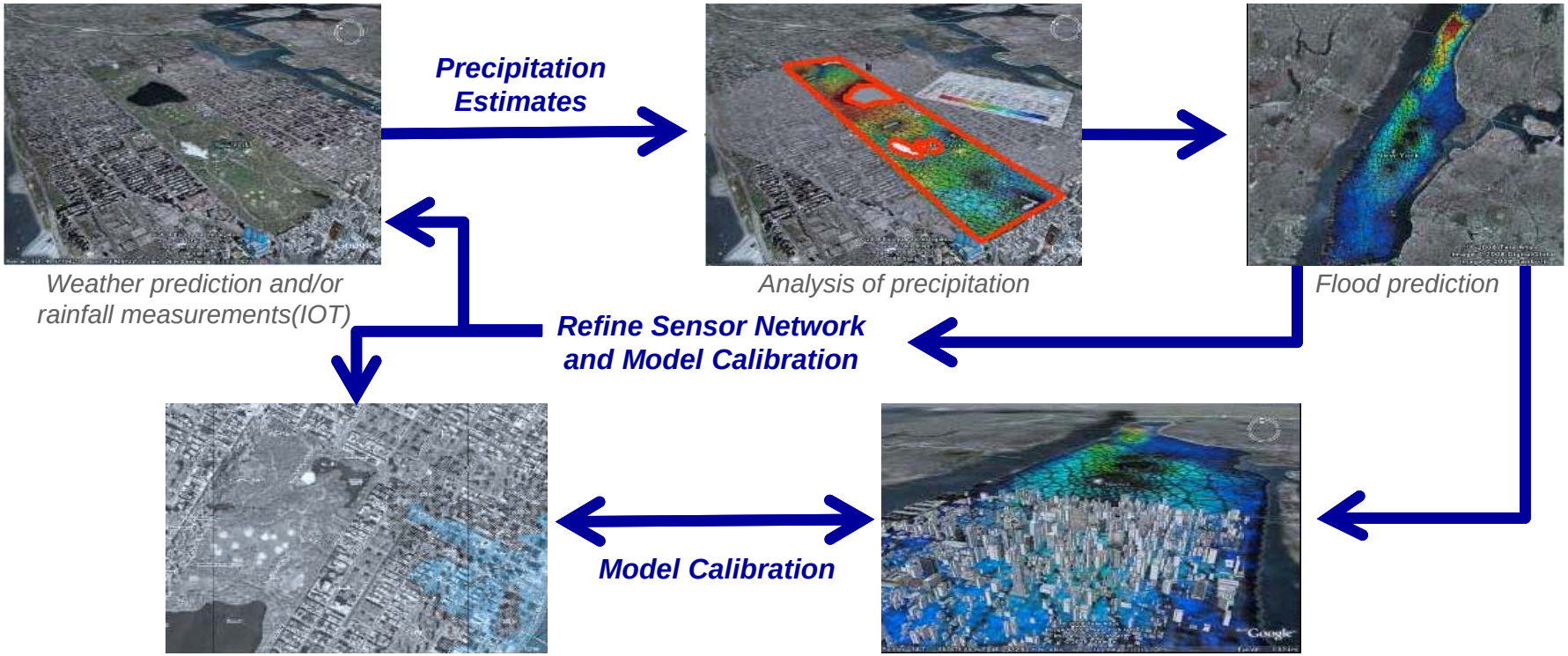
High Performance Computing-Stage One of a Solution

IBM built an advanced high-resolution Weather Prediction System, PMAR - Previsao Meteorologica de Alta Resolucao

Analytics, Predictions, and Proactive Intervention

The Internet of Things is NOT about your toaster reporting in sick.

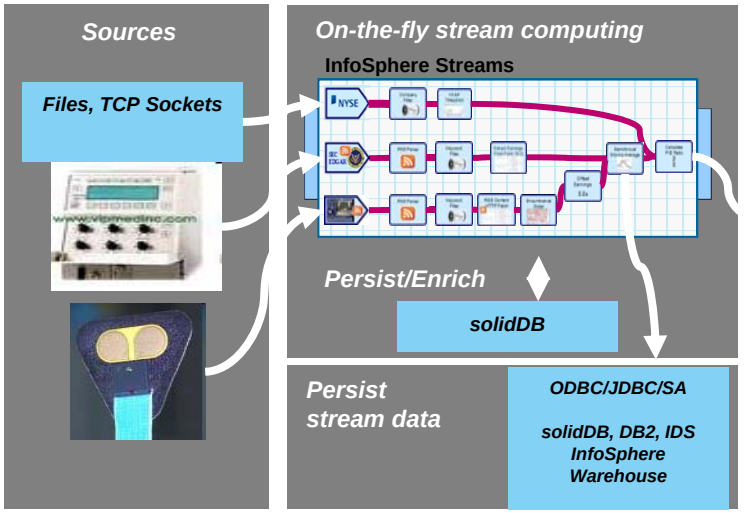
Including an Urban Flood Forecasting System



“Data Baby” Project; Life-Changing Outcomes from Big Data

University of Ontario Institute of Technology *Research Project to monitor and premature infants in the ICU at Sick Children’s Hospital*

- Correlating blood oxygenation with blood pressure to predict “Baby crashing”
- Nosocomial Infection Prediction
  - Monitoring heart rate variability with other information to predict sepsis
  - **System was shown able to warn of life threatening sepsis up to 24 hours earlier than experienced ICU Nurses**



http://www.youtube.com/ibmhealthcare

“Data is the New Oil”

In its raw form, oil has little value. Once processed & refined, it helps power the world.

CNBC

Ann Winbald,
Co-founder Hummer
Winbald Venture Capital

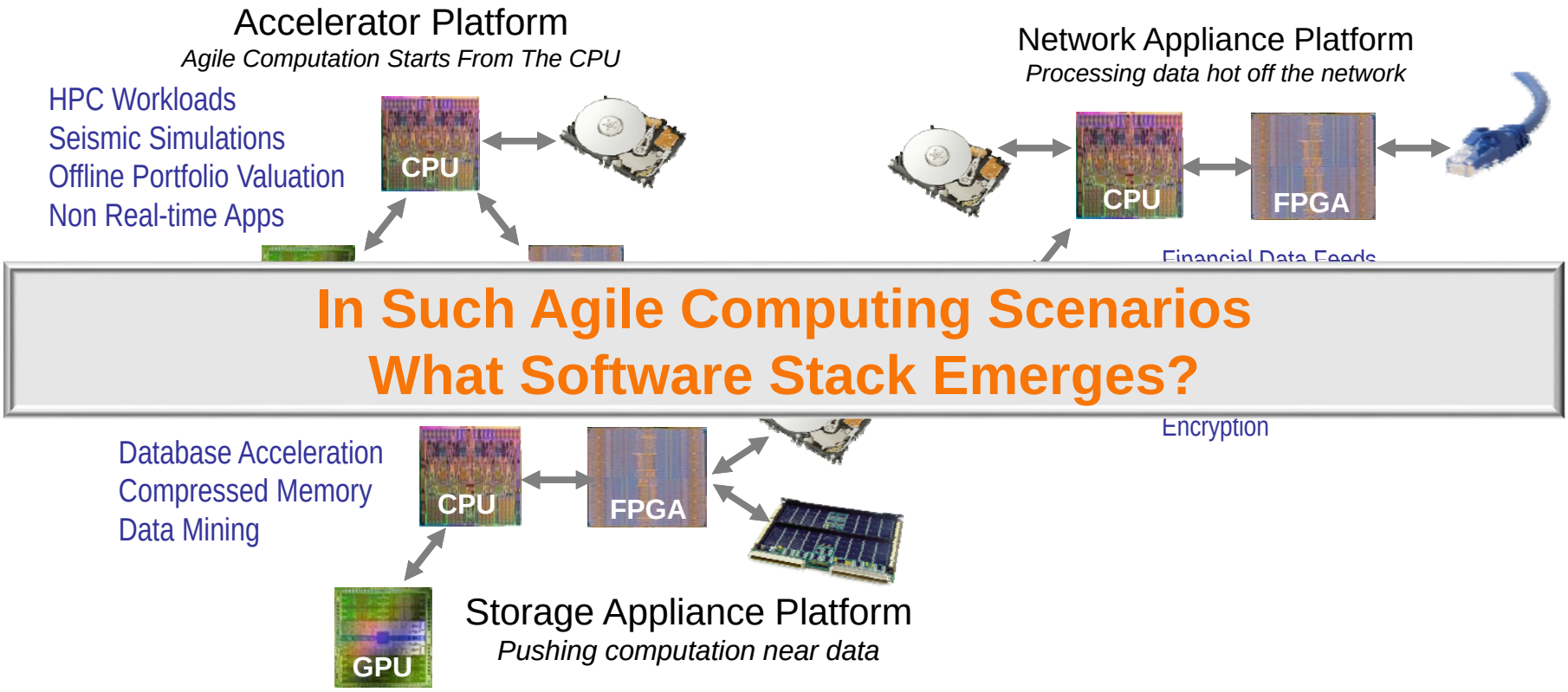
And So Begins an Era of Exponential Change

Prepare for exponential change

Why EXPONENTIAL Change?

- You will collect so much data you that you can't move it.
  - With >20 disparate regulatory environments in Europe alone, good luck moving it legally.
- Not everybody who wants to "move" your data is your friend.
  - Good luck finding where it went and who pilfered it.
- The IOT will connect everything to everything, but is that a good idea?
    - Will we create a data tsunami beyond our ability to deal with it?
        - If so, can intelligence at the edge of the IOT mitigate this issue?
- Are we outrunning our own ability to deal with the very complexity we are creating?
  - Humans have finite "bandwidth", the internet does not.

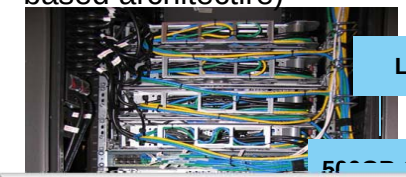
Exponential Change; Providing Vast Computational Horsepower



Exponential Change: A Shift to In-Memory EVERYTHING

Example;12TB, 6K users, 2GB/User Example

Traditional DRAM based architectire)



WWW

10Gb Uplink

Load Balancer

Backup Node

Backup storage req'd for data persistance

26 servers, switches, etc = >1 rack
In-memory database w/ 12TB DRAM
~18kW, ~$145k** cost

10Gb Uplink

POWER8 + CAPI Flash

POWER8 2u Server

4U

A Mainstream Shift from Volatile to Non-Volatile Designs

Coherent CPU/Flash Advantage

- 24:1 physical server consolidation
- >1 rack to 4U Density (>12x)
- 18kW → 1.5kW Energy (12x)
- $24/user → $7.5/user Cost (3.2x)
- 4.5k → 60k Users per rack (13x)*

redis

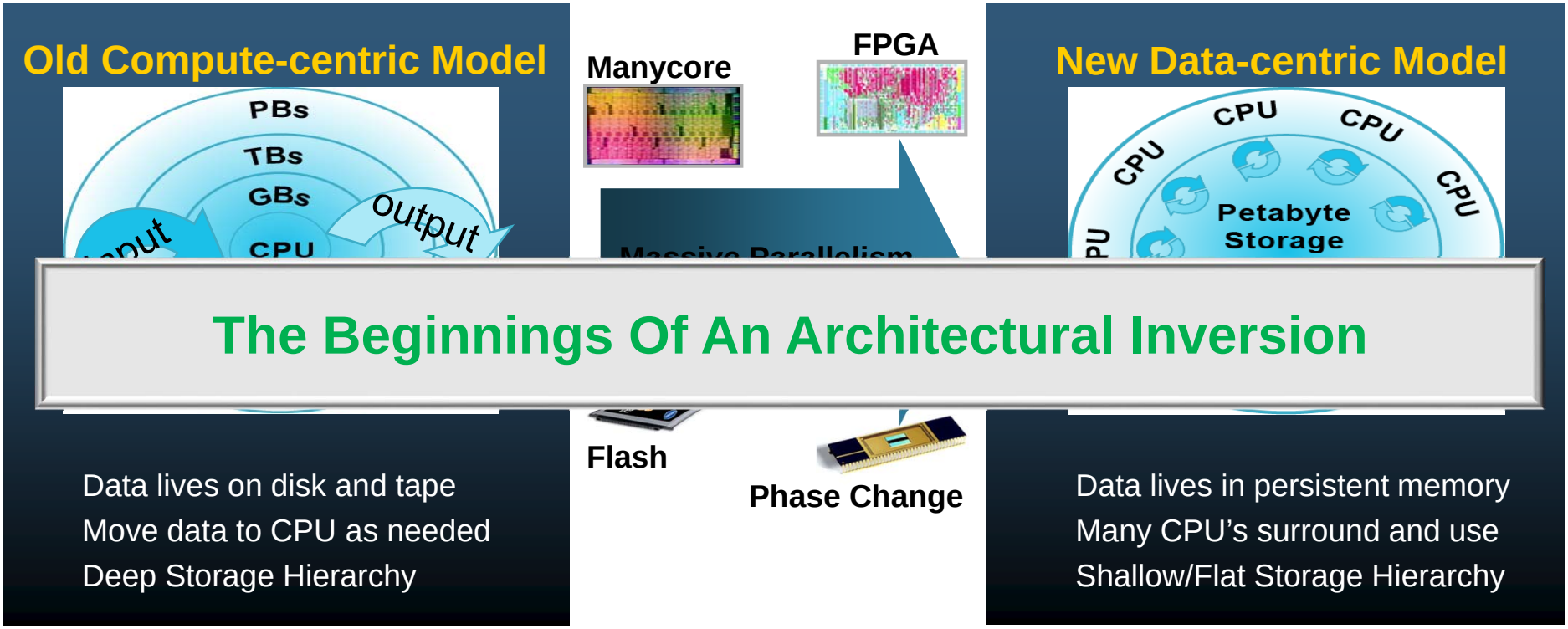
PRELIMINARY SIZINGS

Inbound network limits performance to 1M IOPs in both scenarios, equal capacity (#user, data) in both cases

* Normalized to a single rack
** cost includes 10k$ for 2x 1U switches

Exponential Change; New Data Architectures

We Now Require A Data-Centric Architecture To Avoid Drowning in Data



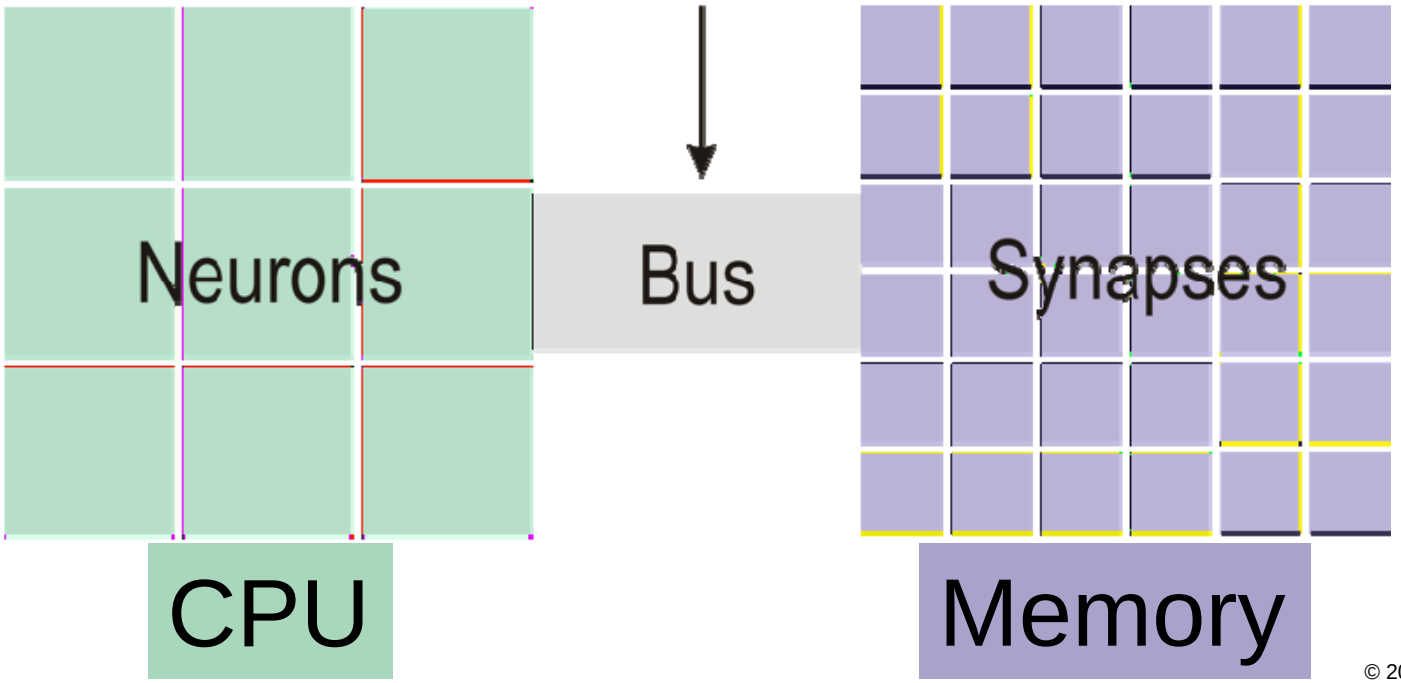
Huge impact on hardware, systems software, and application design

What Happens If We Throw Out All Convention And Go Down Radically Different Technological Paths?

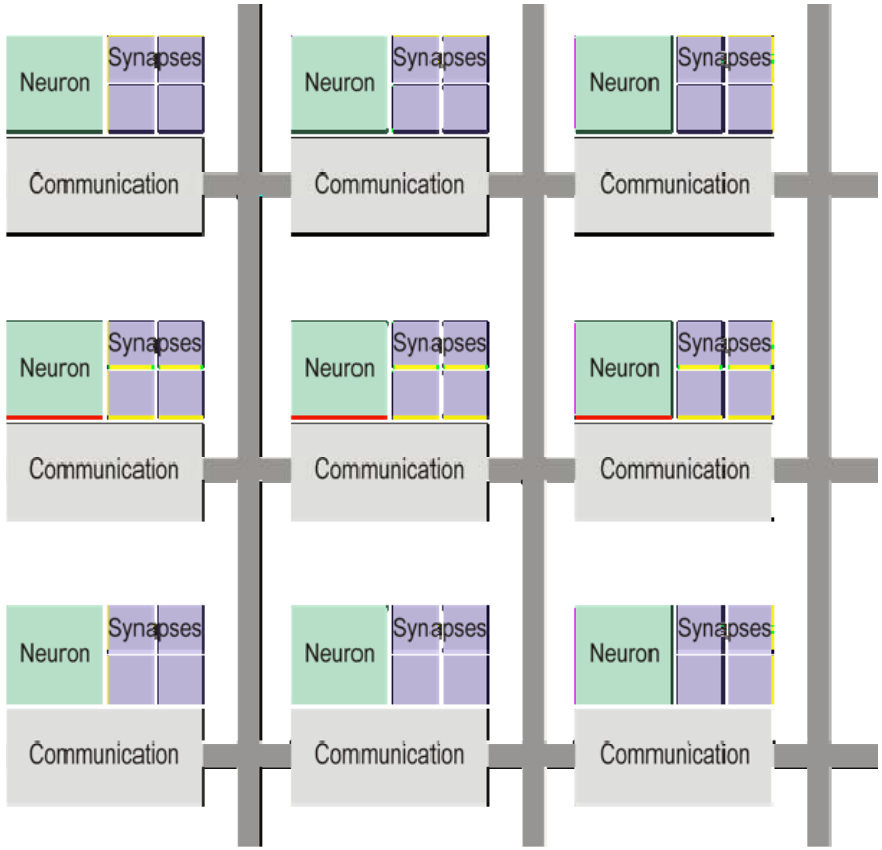
MAGIC

If Humans Were Architected Like Today's Computers, We'd Be Slow and Dumb

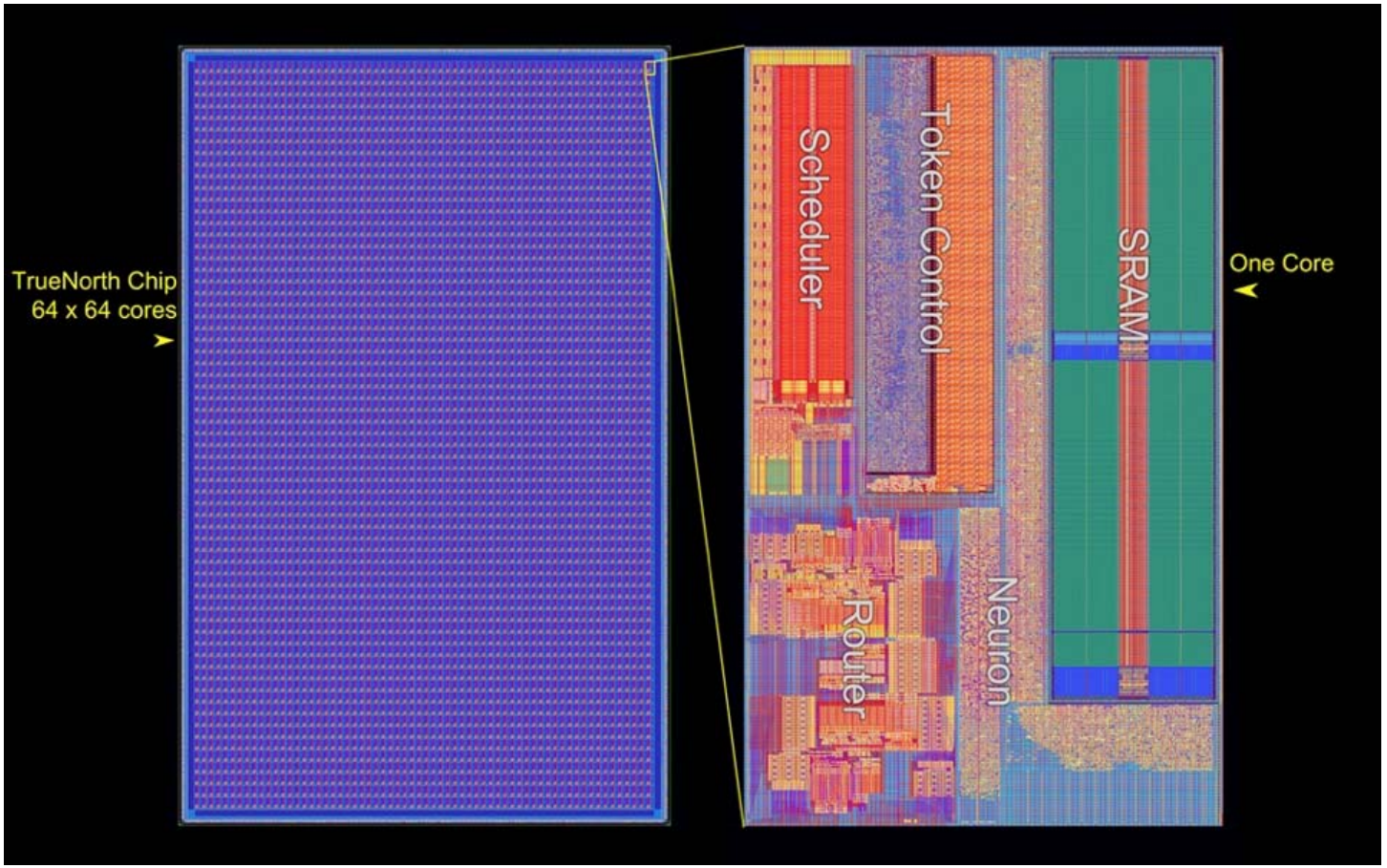
The Memory Bottleneck Found in All "Conventional" System Architectures Would Choke Our Performance



What If You Re-Architected a Microprocessor To More Closely Resemble Human Synaptic Designs?



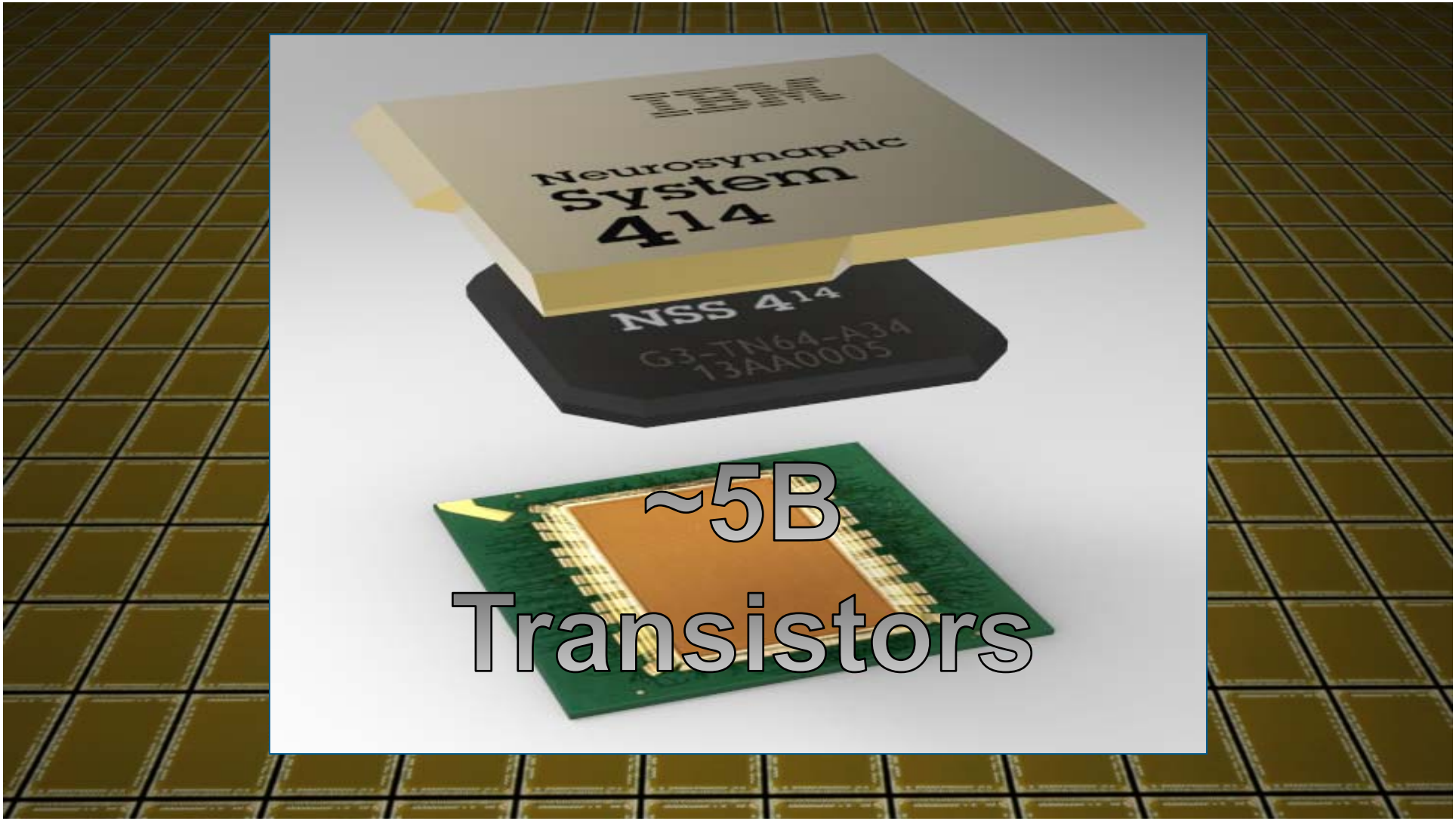
You Would Then Find Yourself at TrueNorth



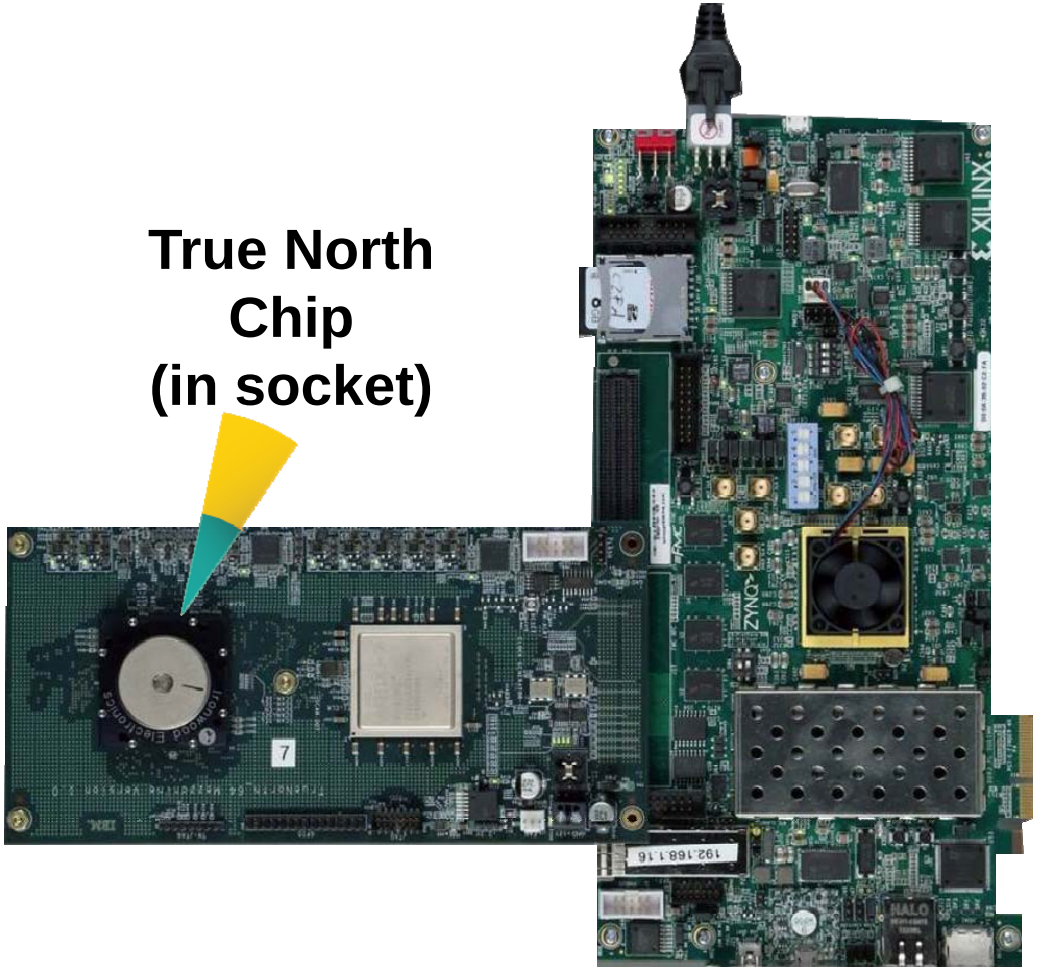
TrueNorth Is The First “At Scale” Neuromorphic Processor

This second generation chip is the culmination of almost a decade of research and development, and is a huge leap forward from the initial single-core hardware prototype developed in 2011.

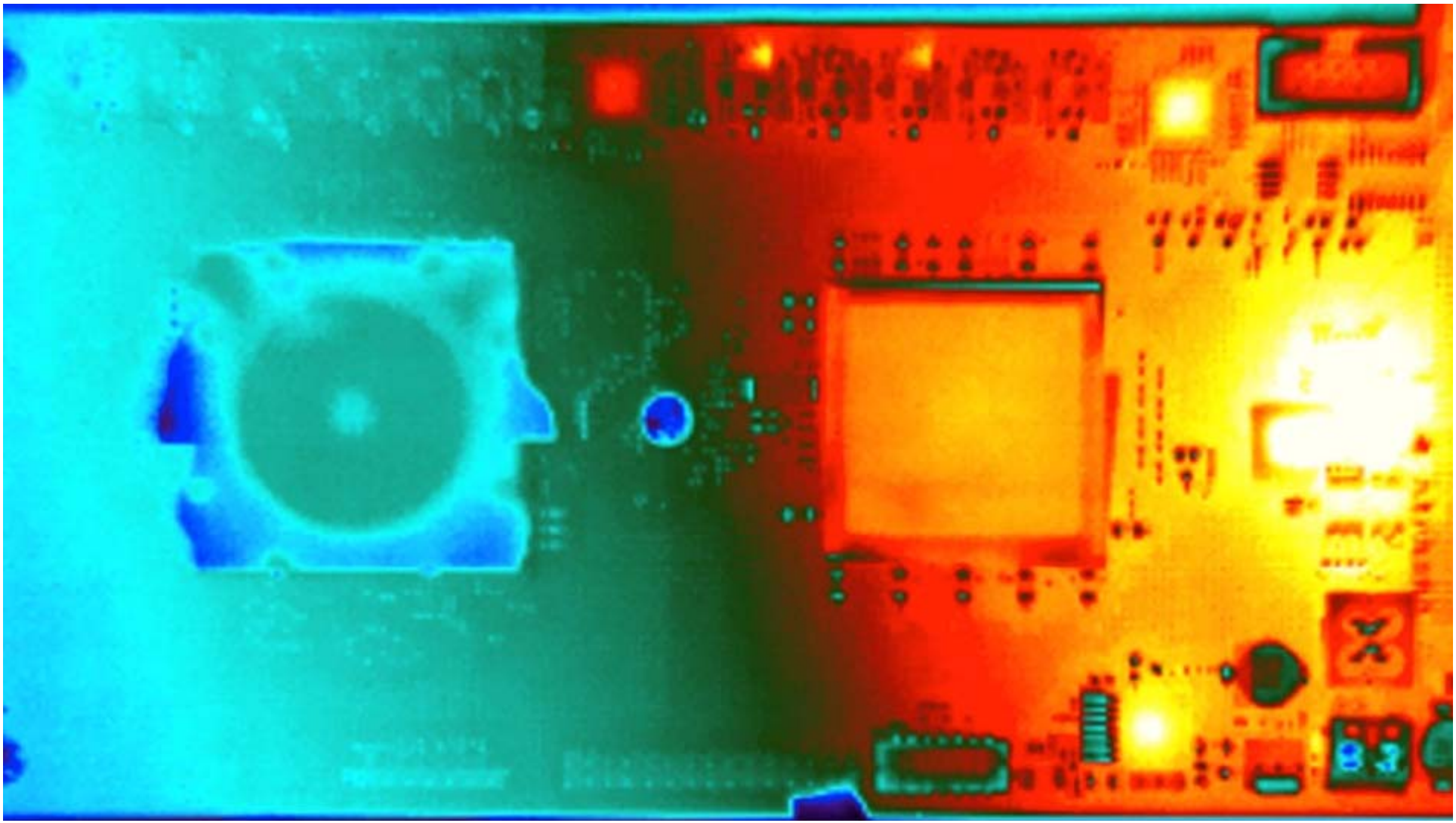
1/10th of a Watt powers the neurosynaptic chip’s 256 million synapses
…with the goal to simulate 1 trillion synapses using only **4 kW of energy**


IBM
Neurosynaptic
System
414
NSS 414
G3-TN64-A34
13AA0005
~5B
Transistors

TrueNorth Ready for “Training” in Detecting and Identifying Objects



Processed by a 4x4 Synaptic Chip 7.2W(2.5W Synaptic Power) board with 16 chips containing 16 million neurons, 4 billion synapses, and “support” chips.

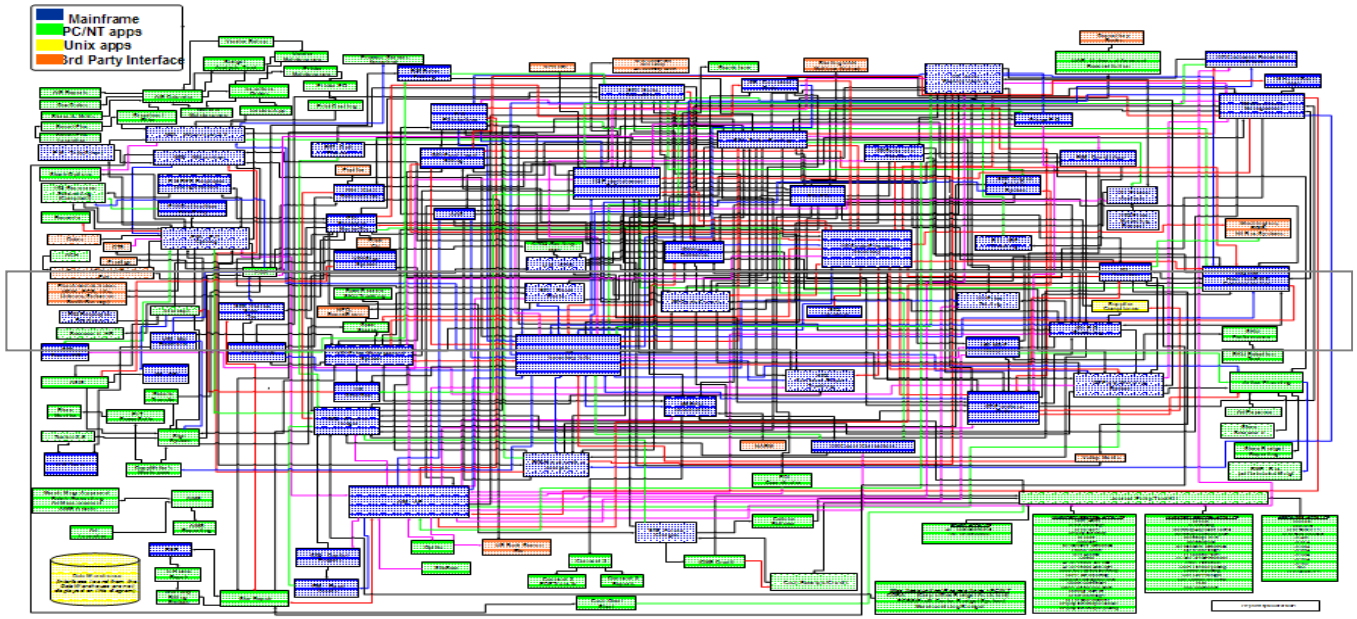


What Happens When Data Complexity and Velocity Overwhelm Us?

Architectural Complexity Makes Problem

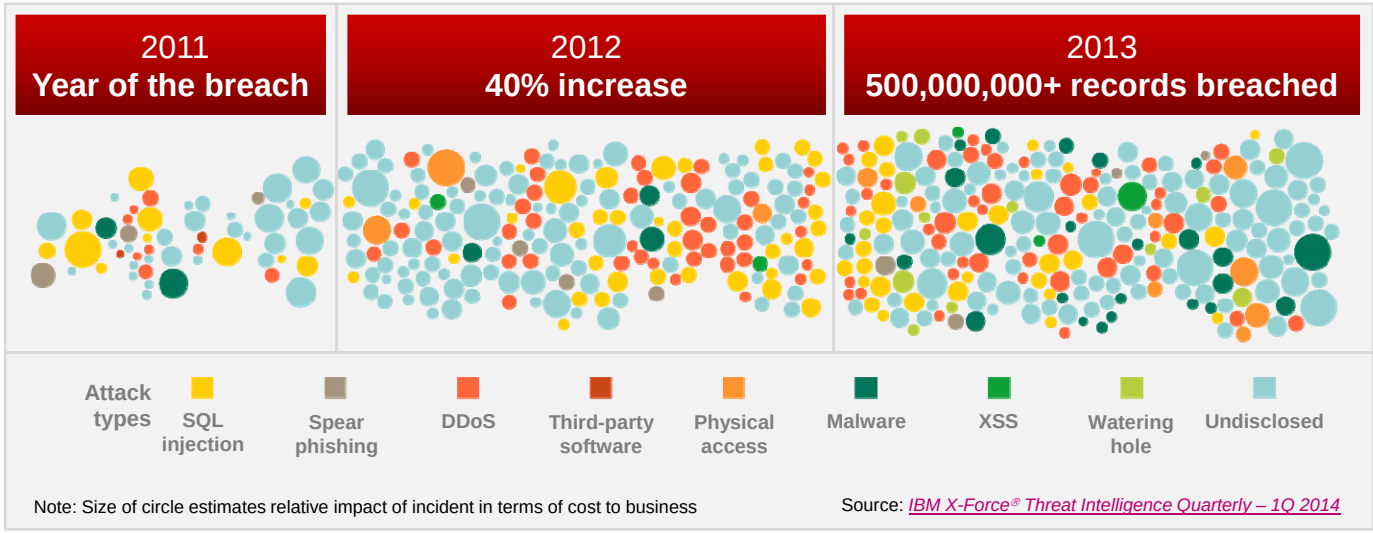
Discovery and Remediation Incredibly Difficult

IT teams cannot identify application and infrastructure dependencies as integration becomes global.



Actual application architecture encountered in engagement with an IBM client

Cyber Attack Complexity and Frequency Are Increasing Exponentially



Note: Size of circle estimates relative impact of incident in terms of cost to business

Source: *IBM X-Force® Threat Intelligence Quarterly – 1Q 2014*

61% of organizations say **data theft and cybercrime** are their greatest threats

2012 IBM Global Reputational Risk & IT Study

$3.5M+ average cost of a **data breach**

2014 Cost of Data Breach, Ponemon Institute

With New Technologies Come New Risks…

Society's Appetite for New Technologies Inevitably Creates New Avenues For Attack That We Must Counter

70%

of security executives have **cloud and mobile concerns**

2013 IBM CISO Survey

614%

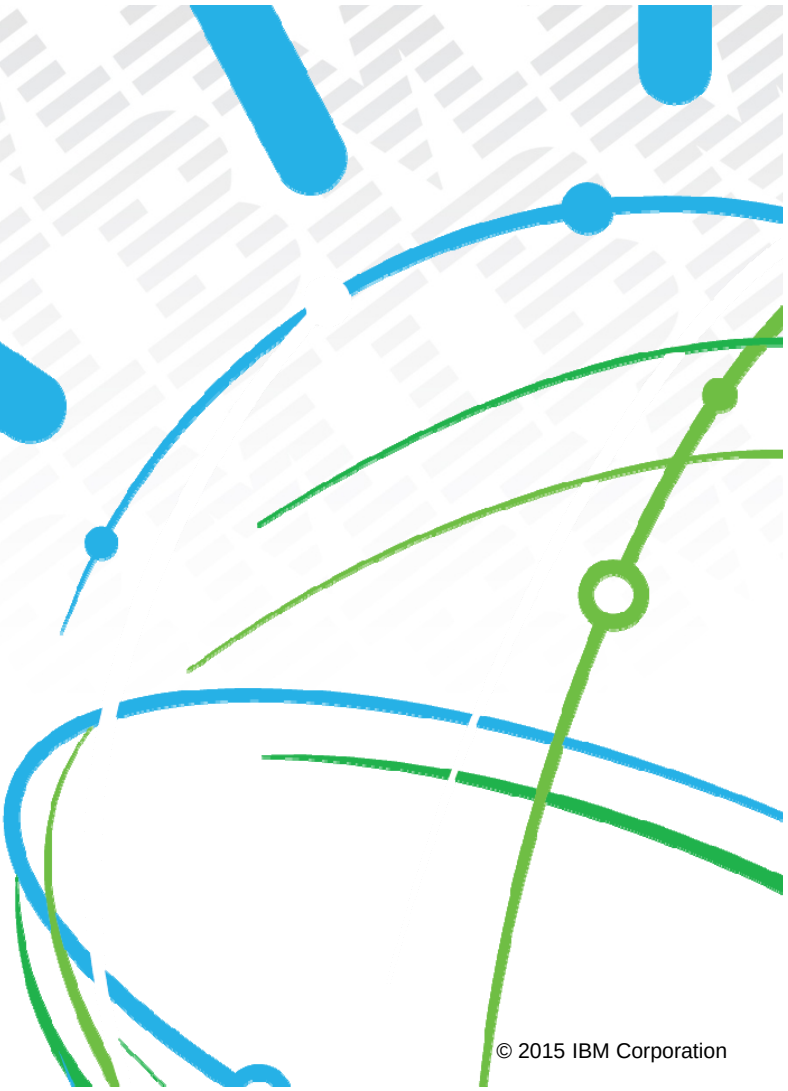
Mobile malware growth in just one year

2012-2013 Juniper Mobile Threat Report

We Will REQUIRE AI To Overcome These Challenges

Where Complex Reasoning and Interaction Extend Human Cognition

Enabling the Era of Accessible Intelligence; Cognitive Computing



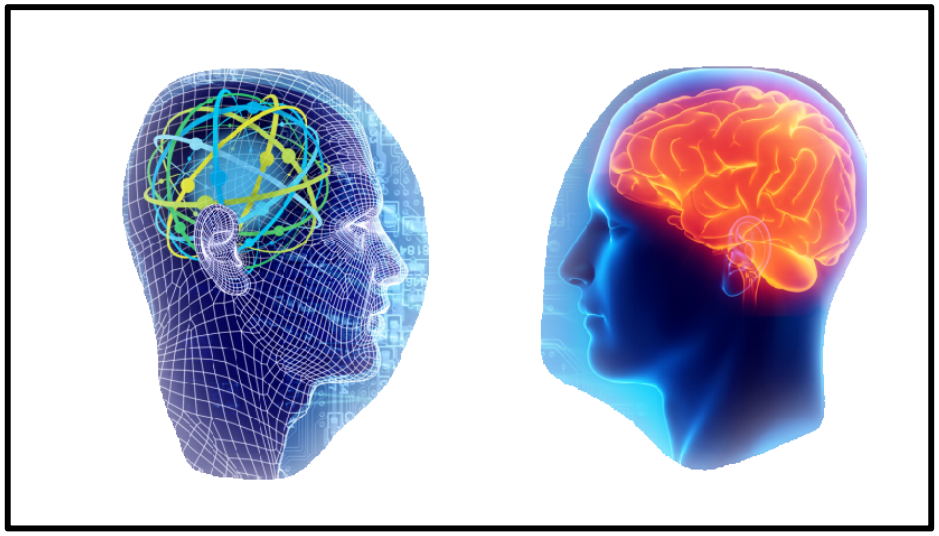
The New Era Of Cognitive Computing Will Transform Our Future

Tabulating Systems Era

Programmable Systems Era

Cognitive Systems Era

Cognitive Systems Expand Our Abilities To Solve Problems



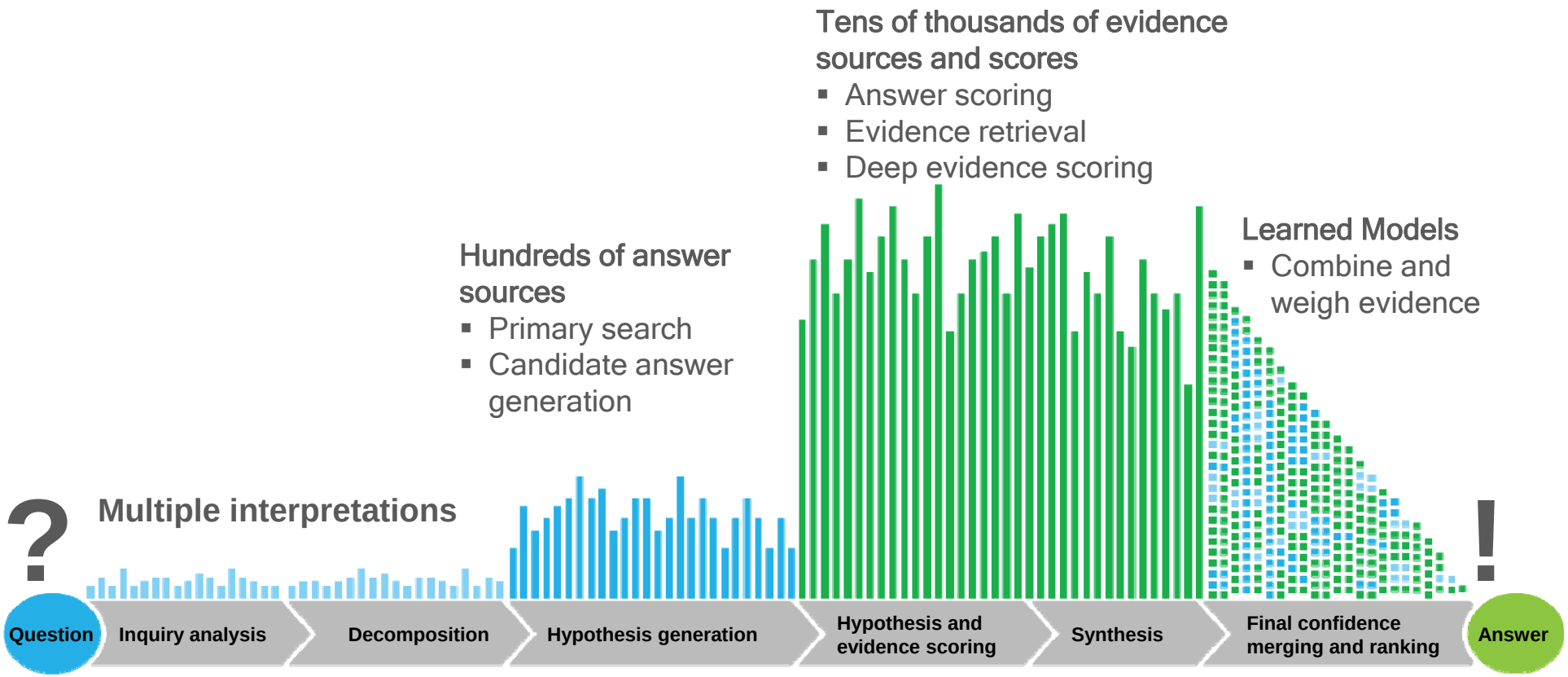
Programmatic Systems

- Leverage traditional data sources
- Follow pre-defined rules (programs)
- Provide the same output to all users

Cognitive Systems

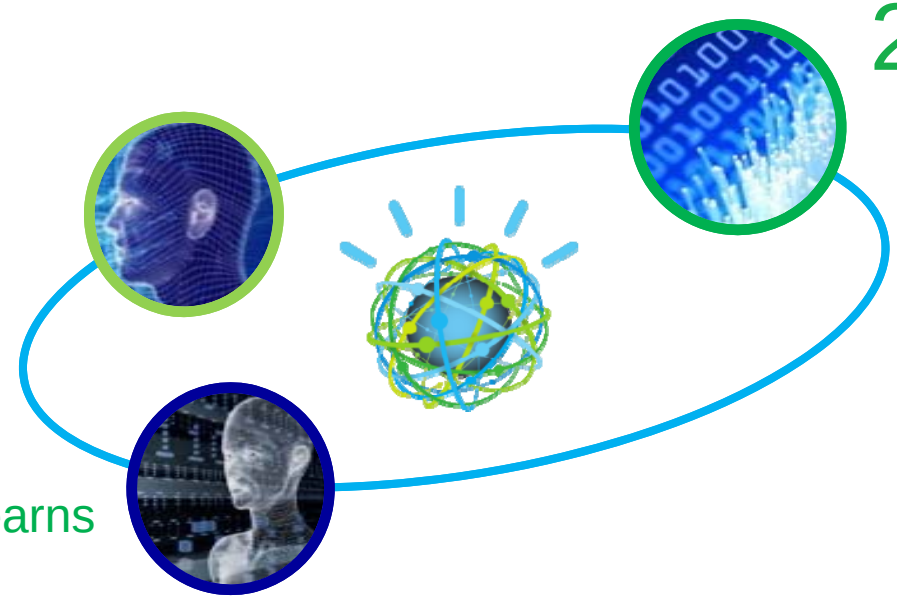
- Are taught, not programmed.
- Learn and improve based on experience
- Interpret sensory and non-traditional data
- Relate to each of us as individuals
- Allow us to expand and scale our own thinking

Cognitive Computing Response Creation



Watson Combines Three Transformational Technologies

1 Understands natural language and human communication



2 Generates and evaluates evidence-based hypothesis

3 Adapts and learns from user selections and responses

AI in Action

- In Criminal Investigations

- In Government Services

- In Healthcare

Live Example; Watson “Solves” Breaking Bad

Breaking Bad
From Wikipedia, the free encyclopedia

“Breaking Bad is an American crime drama television series created and produced by Vince Gilligan. The show originally aired on the AMC network for five seasons, from January 20, 2008 to September 29, 2013.”

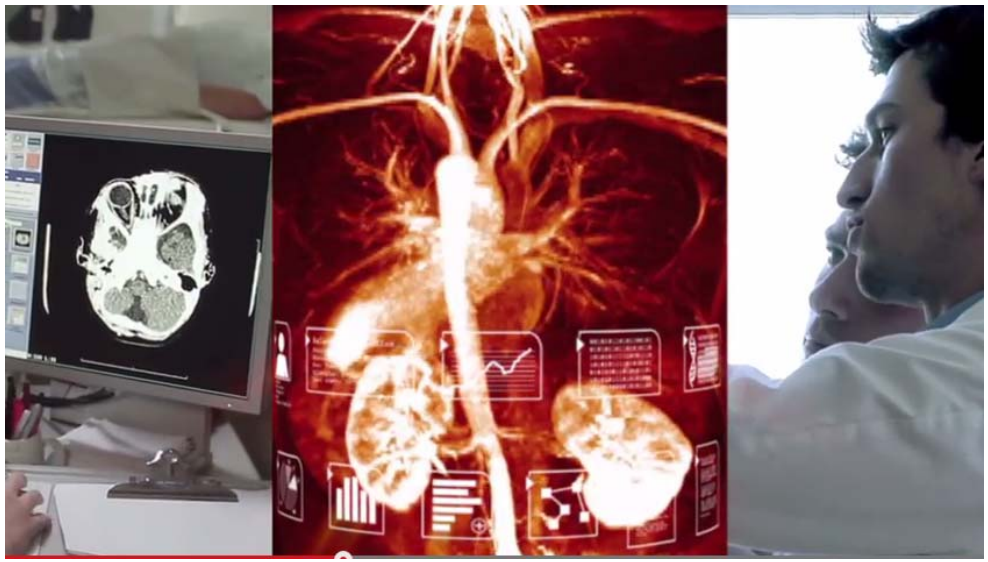
Live Example; A Taxing Experience

“Ask Jasmine”

Ministry of Finance, Singapore

Watson Transforming Healthcare

Teaming With Memorial Sloan Kettering Cancer Center



Memorial Sloan Kettering Cancer Center

IBM

The Bottom Line and the Front Line

- Brute force technology (more of the same) has run its course.
- System level innovation will dominate progress in IT performance.
  - Agile Computing(FGPA's, GPU's, …)
  - Synaptic Architecture
    - Neuromorphic Systems
  - Consumable/Smart IT, e.g. "AI" based systems
- With every new innovation comes new opportunities for progress, and new risks to defend against.
  - You stand on the front lines, to discover and defend vulnerabilities in previously unknown technologies and architectures.