Solving the Information Technology Challenges Beyond Moore's Law

SAND2017-11048C

F. B. "Rick" McCormick
Sandia National Laboratories
fbmccor@sandia.gov

BERKELEY LAB

OAK RIDGE National Laboratory

LAWRENCE LIVERMORE NATIONAL LABORATORY

Sandia National Laboratories

Argonne NATIONAL LABORATORY

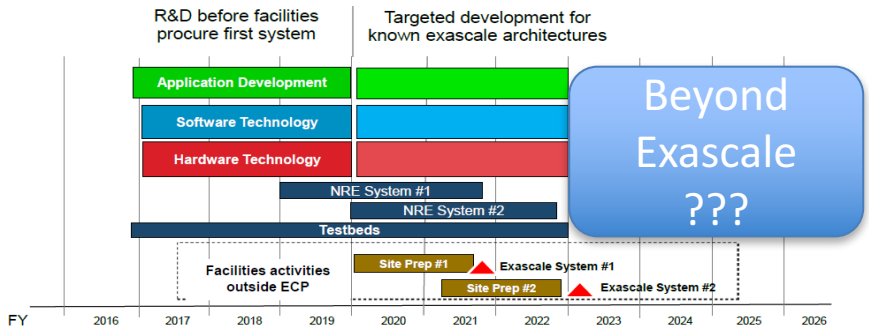
Los Alamos NATIONAL LABORATORY EST.1943

Pacific Northwest NATIONAL LABORATORY

THE Ames Laboratory

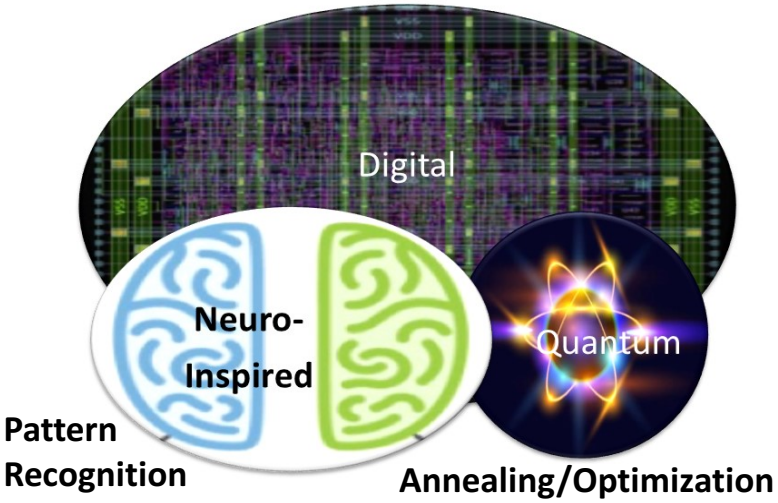
Beyond Moore's Law Computing (BMC)

High-level ECP technical project schedule



https://exascaleproject.org/wp-content/uploads/2016/12/ECP_Update_ASCAC__20161220.pdf

Precision Signal Processing



What is "Beyond Moore's Law"?

- **Litho scaling costs accelerating**
- **Techno-economic agreement dissolving**
- **"More-Moore's Law" progress for the next 5+ years? (HI, 3D, PIM, CMOS specialization)**
- ***Looming disruptive event on the 5-10yr horizon***

Semiconductor industry is huge economic driver

US Semi tool vendor profits now declining, Consolidated: 24 companies ('95)→ 2 today

$56 Billion

Semiconductor R&D and Cap Ex (2013)

Only 1 US leading edge Fab Co (of 4 total) US market share: ~75% ('95)→ ~50% today

$336 Billion

Semiconductor device market (2014)

China/Asia = ~70% of consumer electronics production

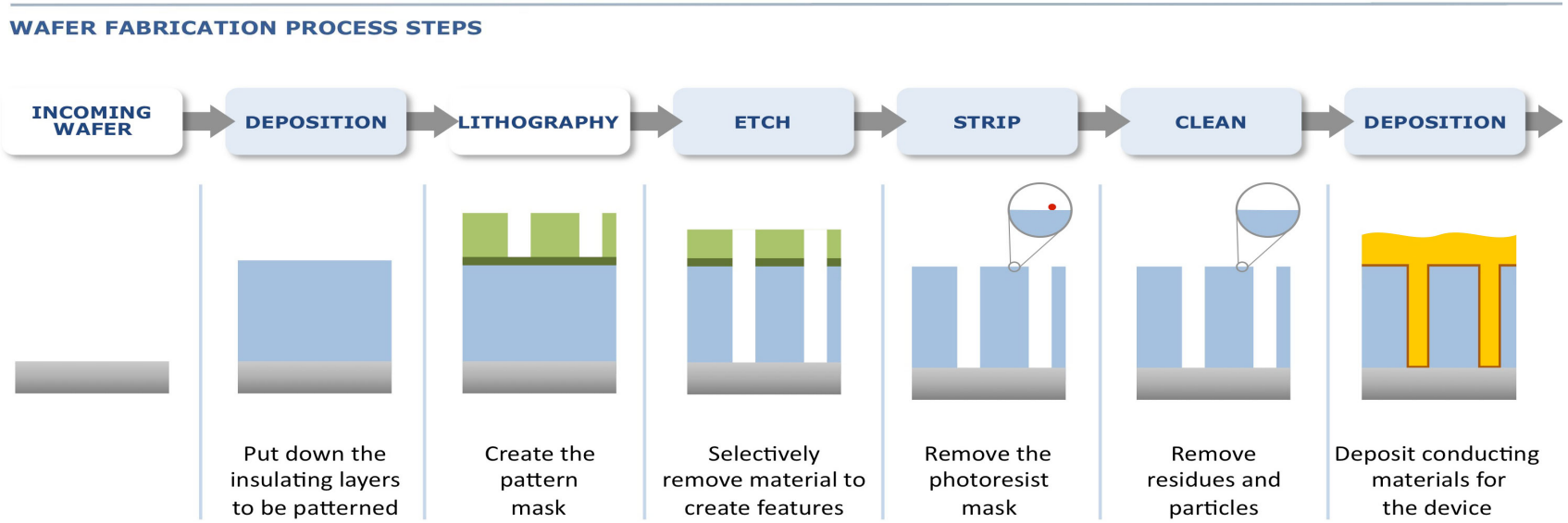
$2.3 Trillion

Global electronics market (2014)

SIA, www.globalsmt.net

The semiconductor industry is crucial to the U.S. economy. Semiconductor sales constitute a $336B/yr. economic engine driving the $2.3T/yr. global electronics industry. Semiconductor products are currently the second largest class of U.S. exports

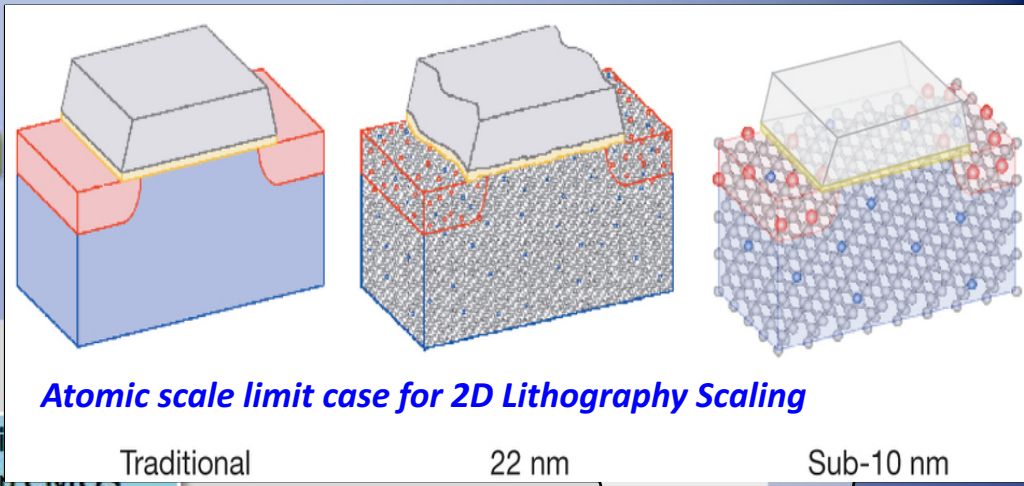
Silicon Integrated Circuit Fab



http://www.lamresearch.com/portals/0/Images/Wafer_Fabrication_Process_Steps.jpg

http://semimd.com/materials-matters/files/2015/08/Costs[illegible]f-wafer-to-IC-chip.jpg

50 years of Sem



Atomic scale limit case for 2D Lithography Scaling

10 um
1 um
100 nm
10 nm
1 nm

Modern CMOS

Beginni
submicron CMOS

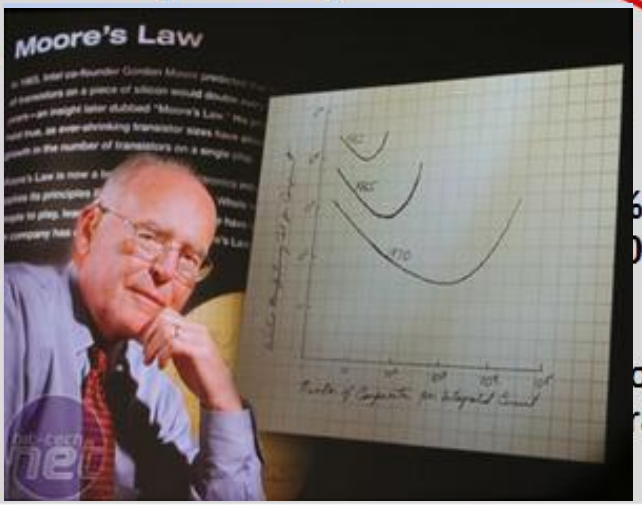
Deep UV litho

65 nm in 2006

28 nm in 2012

Presumed limit to scaling

40 years of scaling history



0%

over the past 40 years
rate is slowing

1970 1980 1990 2000 2010 2020 2027?

5nm

End of Moore's law: technology

Technology Scaling Trends

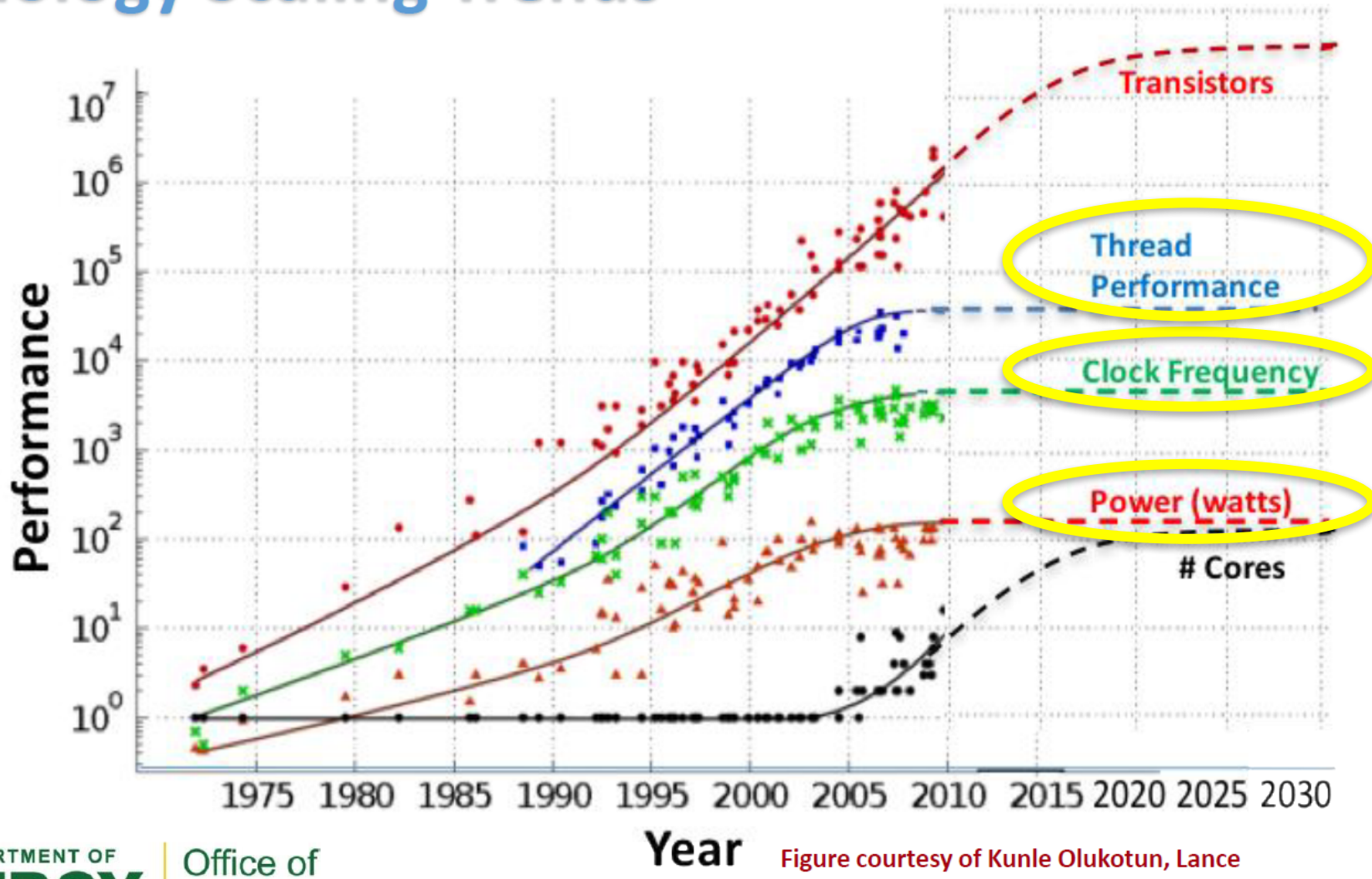
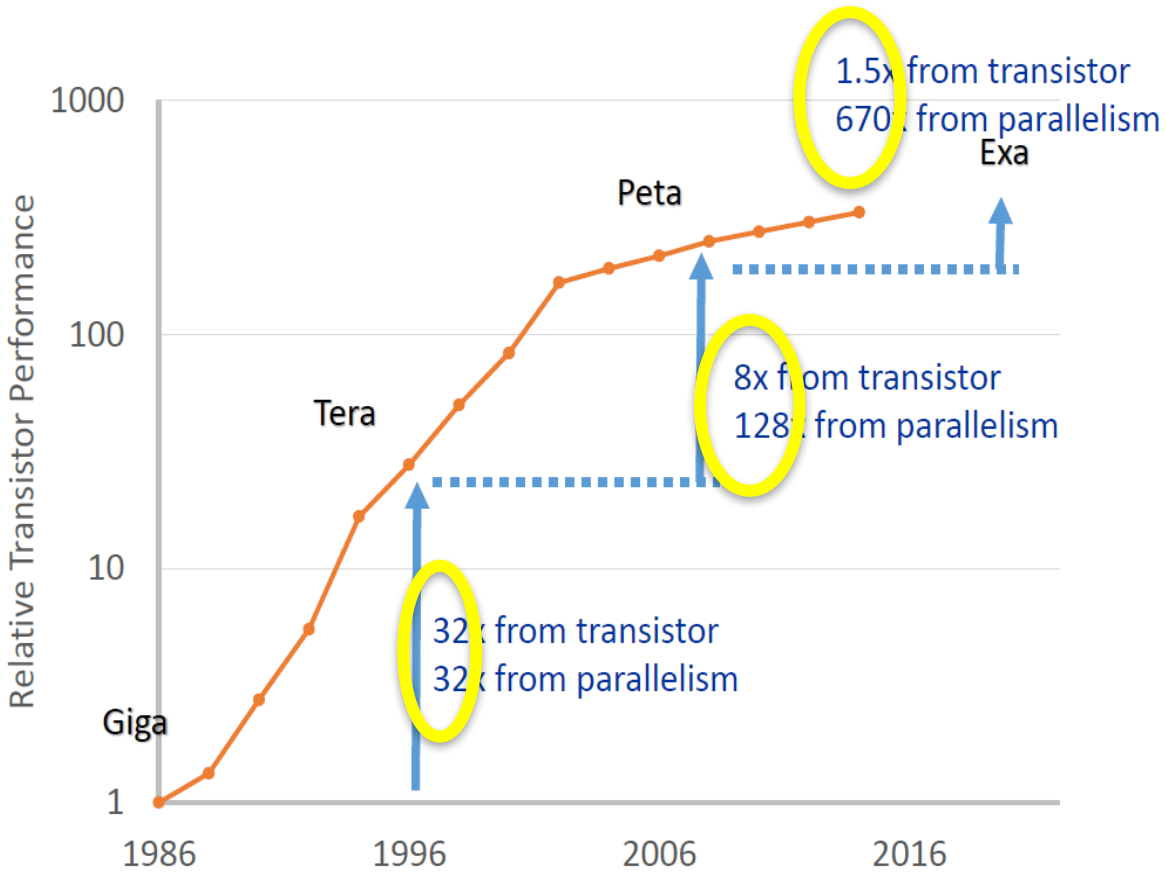


Figure courtesy of Kunle Olukotun, Lance Hammond, Herb Sutter, and Burton Smith

U.S. DEPARTMENT OF ENERGY | Office of Science

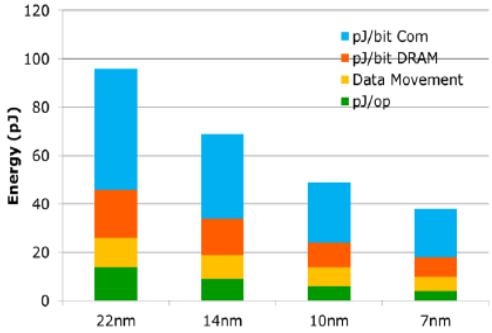
End of Moore's law: technology

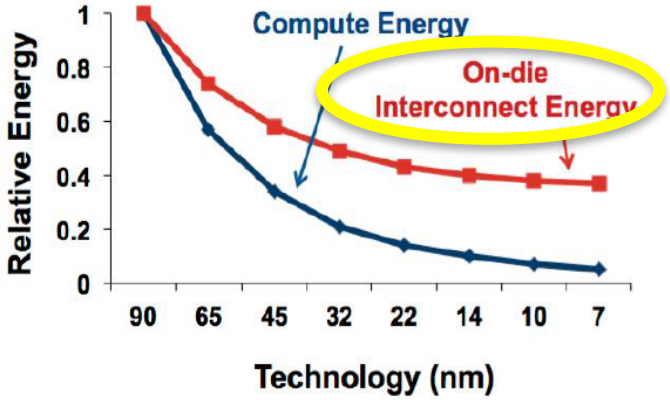
From Giga to Exa, via Tera & Peta



Shekhar Borkar, Intel

Basic compute loop





Shekhar Borkar, Journal of Lightwave Technology, 2013

U.S. DEPARTMENT OF ENERGY | Office of Science

End of Moore's law: economics

designlines MCU

News & Analysis

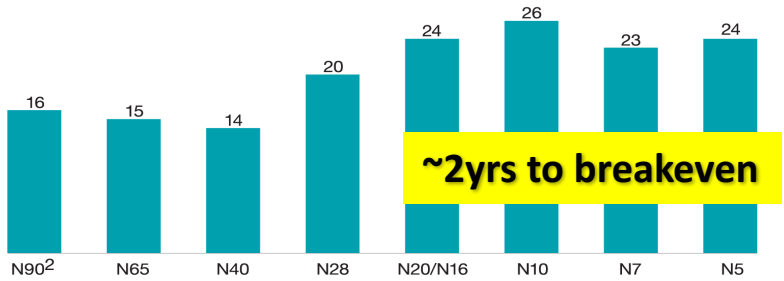
Samsung Breaks Ground on $14 Billion Fab

World's most expensive semi fab

R. Colin Johnson
5/8/2015 03:18 PM EDT
14 comments

Payback time for foundries will be close to 24 months for nodes smaller than 20 nanometers/16 nanometers.

Investment payback time, months[1]

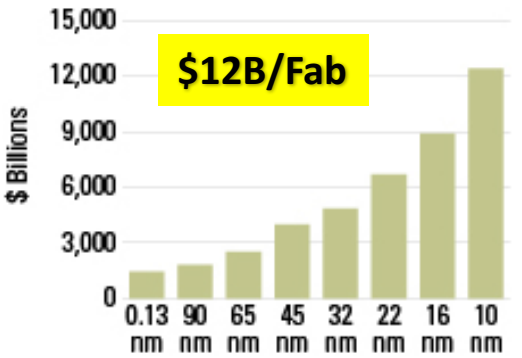


[1]Includes fab capital expenditures and R&D.
[2]N = node.

McKinsey&Company

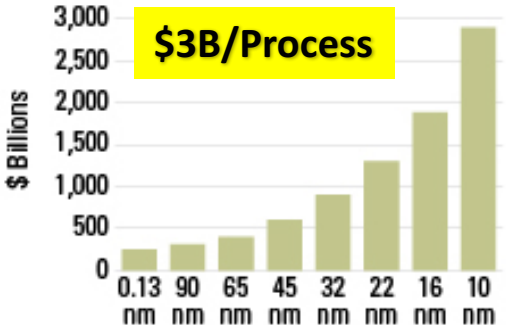
http://www.mckinsey.com/~/media/McKinsey/Industries/Semiconductors/Our%20Insights/Capital%20equipment%20Will%20further%20shrink%20bring%20growth/PNG%20Capital%20equipment%20ex%203.ashx

Figure 4: Fab Costs by Node (in US$ billions)



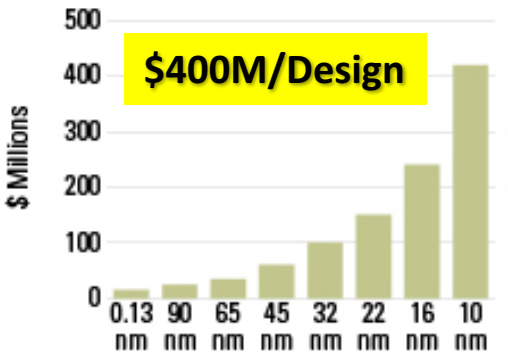
Source: Common Platform Technology Forum 2012 and AlixPartners analysis

Figure 5: Process Technology Development Costs by Node (US$ billions)



Source: Common Platform Technology Forum 2012 and AlixPartners analysis

Figure 6: Chip Design Costs by Node (US$ millions)

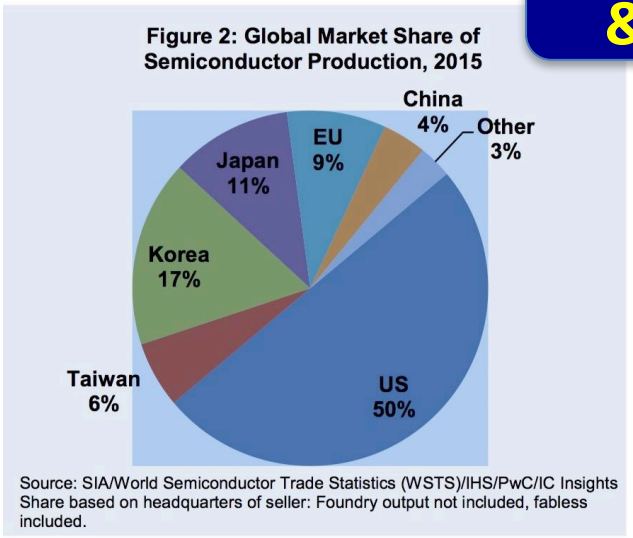


Source: Common Platform Technology Forum 2012 and AlixPartners analysis

http://legacy.alixpartners.com/en/Portals/alix/IC%20Copy%20&%20Images/High-Tech/2014%20Cashing%20in%20with%20Chips/04-05-06-SEMICONDUCTOR-0114.jpg

Beyond Moore's Law Computing (BMC)
Global Drivers…the Problem & the Opportunity

Global Competition & Market Share

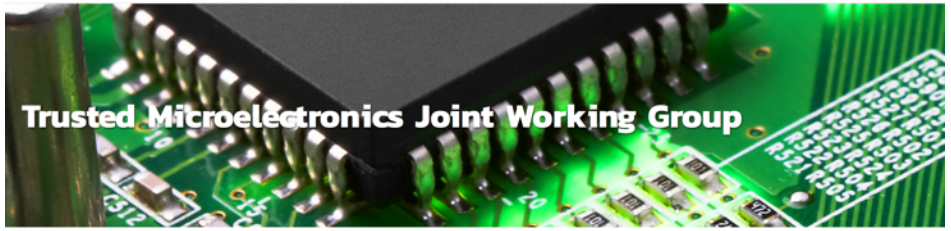


China: $120B-$150B Investment

Sunway TaihuLight : June 2016
Worlds Fastest Computer
An Indigenous Chinese Computer
Dropped Power Consumption by ~25%

National Security

Energy Consumption



Semiconductor Industry Association, 2015-2017

SIA SEMICONDUCTOR INDUSTRY ASSOCIATION

December 18, 2015

President Barack Obama
The White House
1600 Pennsylvania Avenue NW
Washington, DC 20500

Dear Mr. President:

I am writing on behalf of the U.S. semiconductor industry to urge you to take several important actions to help ensure continued U.S. leadership in semiconductor technology. These recommendations are in response to requests from officials in your Administration for suggested actions to address the challenges that recent Chinese policies and actions pose to our industry.

Research: The semiconductor industry has followed Moore's Law to unimaginable levels, enabling smaller, more powerful and energy-efficient devices while reducing costs to consumers. Our industry annually invests about one-fifth of revenue in R&D, a greater share than any other industry. We will continue to **urge government to serve as a partner** in this effort by substantially increasing federal investments in university-based, **pre-competitive research at DARPA, NIST, NSF, and DOE.** We'll also push policymakers to implement and support research programs to advance semiconductor technology and innovation.

U.S. Leadership in High Performance Computing (HPC)

A Report from the NSA-DOE Technical Meeting on High Performance Computing

December 1, 2016

To maintain and extend U.S. leadership in HPC, it is **critical to lead the exploration and development of innovative computing architectures** that will unleash the creativity of the HPC community……participants recommend the U.S. engage in innovative architectural exploration and development…

This will require a strong and sustained commitment to **a variety of architectural approaches**, where each is clearly focused on a reasonable subset of clearly important applications (e.g., **modeling and simulation, large-scale analytics, cybersecurity,** etc.). Multiple teams, each capable of developing novel computing architectures **using a complete co-design approach**, will be required….Fully functional **system prototypes** from these explorations will need to be developed and evaluated for their specific application subset. **Technology advancements** common to many (or all) architectures will also **need to be supported (i.e., new memories, silicon photonics, programming tools and languages, etc.).**

PCAST Study Group, 2017

REPORT TO THE PRESIDENT

Ensuring Long-Term U.S. Leadership in Semiconductors

Executive Office of the President

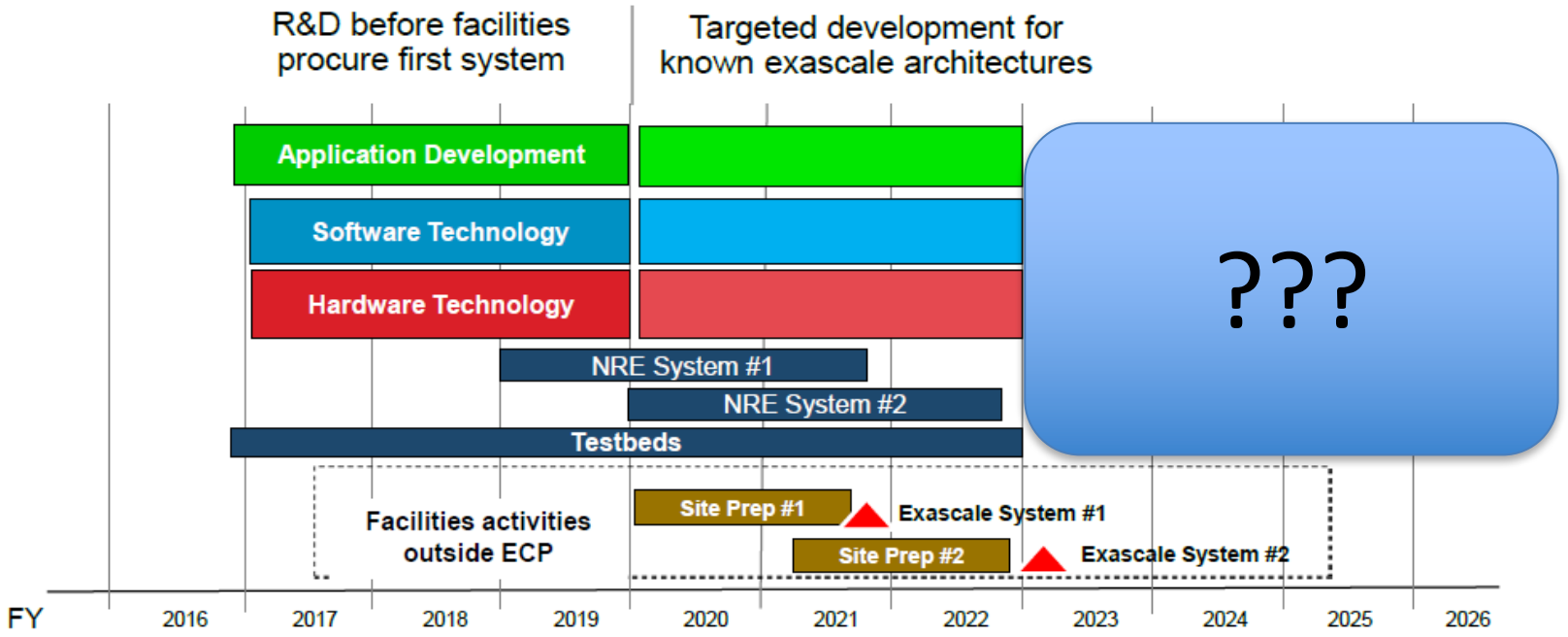
President's Council of Advisors on Science and Technology

January 2017

"The core finding of the report is this:…to mitigate the **threat posed by Chinese industrial policy** and strengthen the U.S. economy. Thus, the report recommends and elaborates on a three pillar strategy to (i) push back against innovation-inhibiting Chinese industrial policy, (ii) improve the business environment for U.S.-based semiconductor producers, and (iii) **help catalyze transformative semiconductor innovation over the next decade. Delivering on this strategy will require cooperation among government, industry, and academia to be maximally effective."`**

How will DOE build HPC AFTER Exascale?

High-level ECP technical project schedule



https://exascaleproject.org/wp-content/uploads/2016/12/ECP_Update_ASCAC__20161220.pdf

Challenges are increasing and Moore's law is ending:

How can we prepare for technology disruption in the next phase?

20fJ/Instruction Impact

PetaScale
2016: 40nJ/FLOP

ExaScale
2022: 20pJ/FLOP

ZettaScale
2030: 20fJ/Instr

Energy-efficient electronics enable a spectrum of key Mission Applications

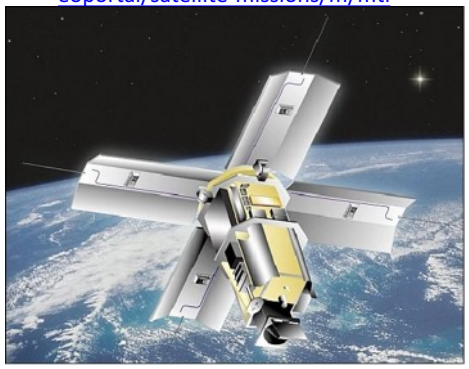
ZettaFLOP HPC in 20MW
Full scale high-fidelity modeling

http://www.deccanchronicle.com/technology/in-other-news/200616/new-chinese-system-named-worlds-top-supercomputer.html

100 PFLOP in 2KW, Design Cluster
Model-based component design

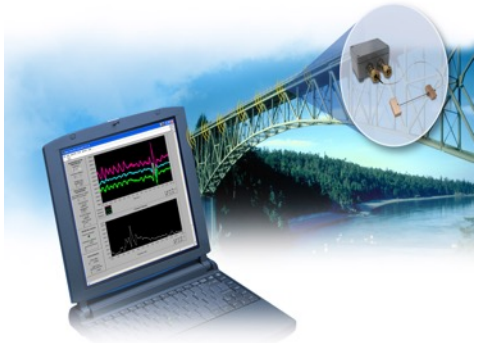
https://blog.cr.yp.to/20140602-saber.html

https://directory.eoportal.org/web/eoportal/satellite-missions/m/mti



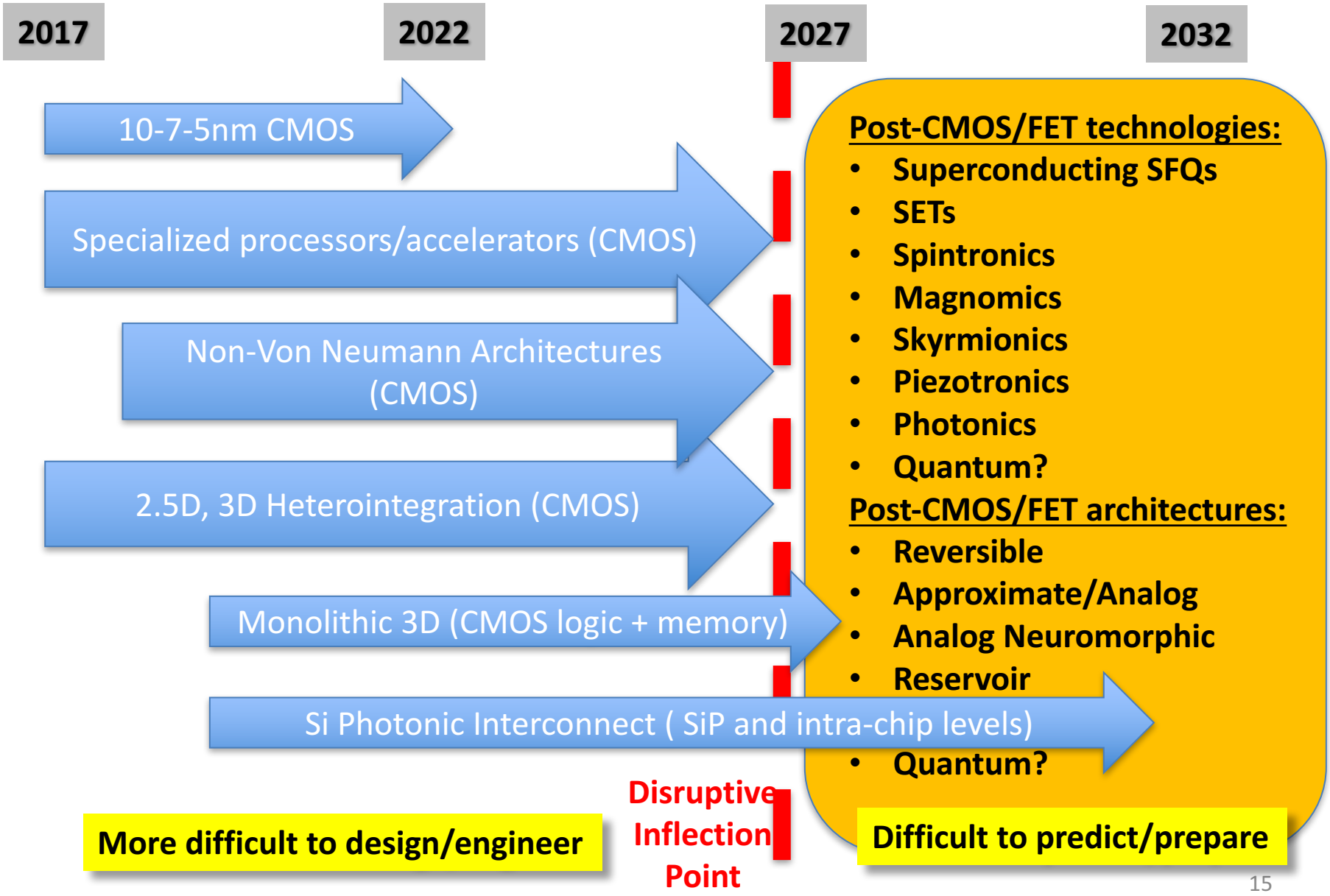
1 PFLOP in 20W
Embedded Cognitive Processor
Autonomous satellite/UAV

1 TFLOP in 20mW
Autonomous remote/embedded sensor

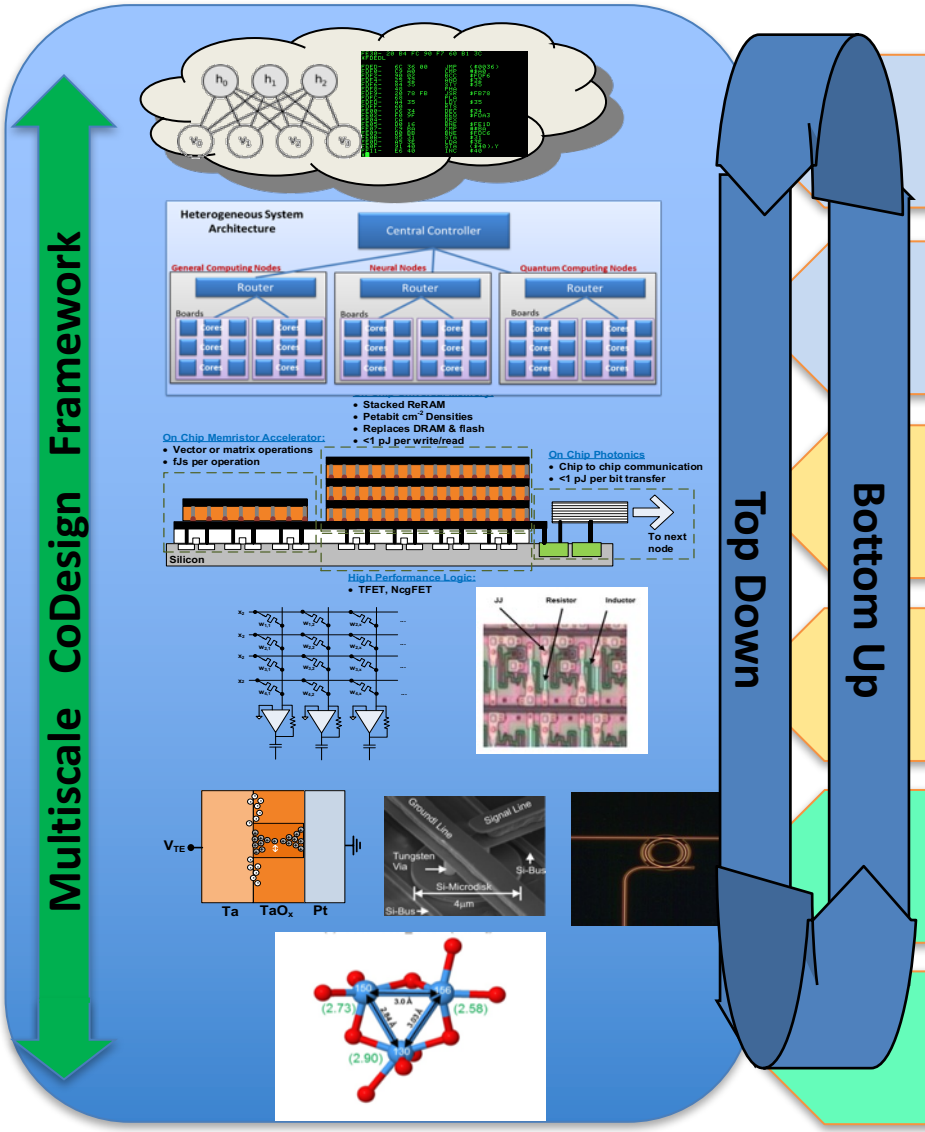


https://www.nsf.gov/discoveries/disc_images.jsp?cntn_id=108363&org=NSF Credit: *MicroStrain, Inc.*

Moore's, "more-Moore's", & BMC



DOE Approach: Multiscale CoDesign



Top Down

Bottom Up

Programming paradigms

System architecture

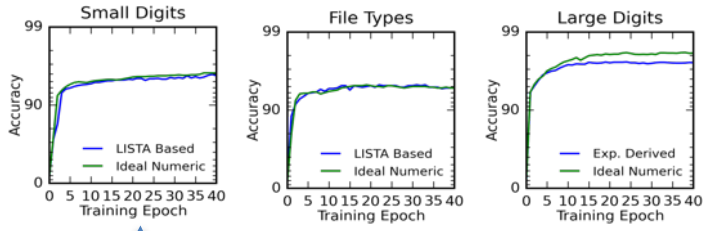
Component integration

Device demonstration

Device physics

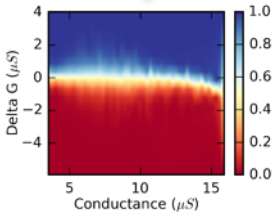
Fundamental science

Novel Device Multiscale CoDesign Example: HAANA Crossbar Accelerator



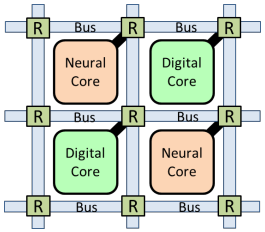
Target Algorithms
- Deep Learning
- Sparse Coding
- Liquid State Machines

Algorithms

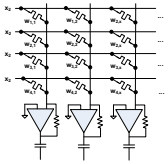


Sandia Cross-Sim: Translates device measurements and crossbar circuits to algorithm-level performance

Architecture

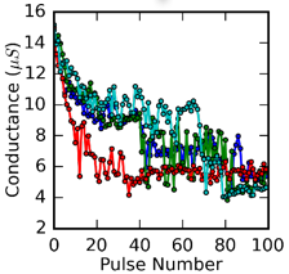


Modified McPAT/CACTI: Model performance and energy requirements



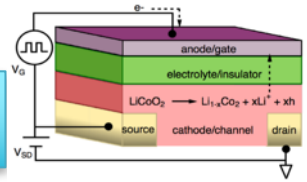
Circuits

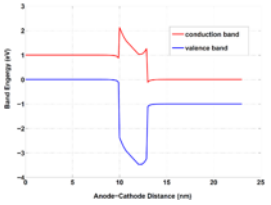
Sandia's Xyce Circuit Sim: Simulate crossbar circuits based on our devices

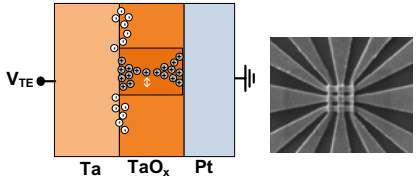


Memristor fabrication and measurements in MESAFab

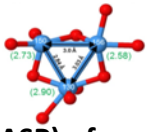
Drift-diffusion model of ReRAM band diagram & transport (REOS, Charon)







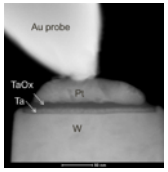
Devices

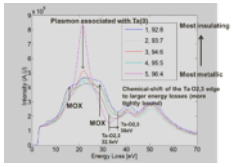


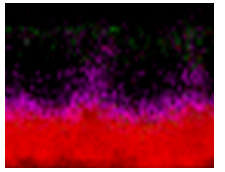
DFT (VASP) of model of oxide physics, bands

Materials

In situ TEM of filament switching: Use DFT model to interpret EELS signature







Beyond Moore Co-design Framework: 5 Working Groups

Modeling

10,000x improvement: 20 fJ per instruction equivalent

Experimental

Advanced Manufacturing

Algorithms

Architectures

Devices

Materials

Algorithms & SW Environments

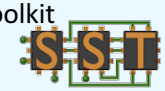
Hardware & Circuit Architectures

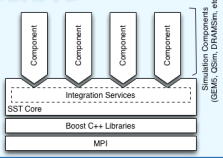
Comm., Memory & Computation Devices

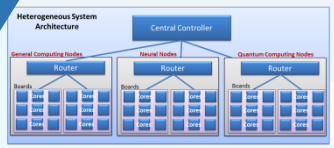
Materials

Algorithms and Software Environments
- Application Performance Modeling

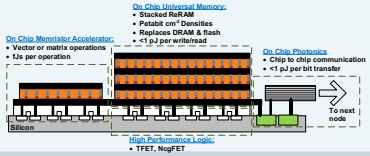
Computer System Architecture Modeling
- Next generation of Structural Simulation Toolkit
- Heterogeneous systems HPC models

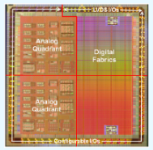






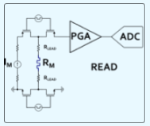
Microarchitecture Models
- McPAT, CACTI, NVSIM, gem5

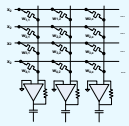


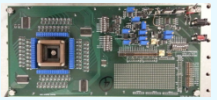


Component Fabrication
- Processors, ASICs
- Photonics
- Memory

Circuit/IP Block Design and Modeling
- SPICE/Xyce model

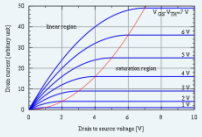


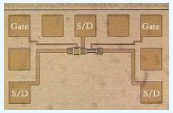




Test Circuit Fab and Measurement
- Subcircuit measurement

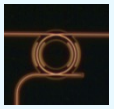
Compact Device Models
- Single device electrical models
- Variability and corner models

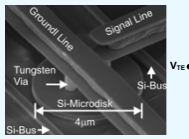


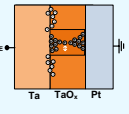


Device Measurements
- Single device electrical behavior
- Parametric variability

Device Physics Modeling
- Device physics modeling (TCAD)
- Electron transport, ion transport
- Magnetic properties

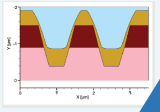


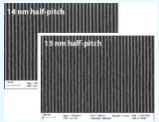


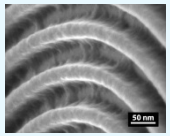


Device Structure Integration and Demonstration
- Novel device structure demonstration

Process Module Modeling
- Diffusion, etch, implant
- Simulation
- EUV and novel lithography models

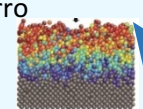




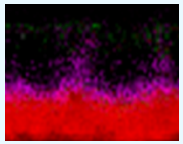


Process Module Demonstrations
- EUV and novel lithography
- Diffusion, etch, implant simulation

Atomistic and Ab-Initio Modeling
- DFT – VASP, Socorro
- MD – LAMMPS



Example activities within a MCF



Fundamental Materials Science
- Understanding Properties/Defects via Electron, Photon, & Scanning Probes
- Novel Materials Synthesis

DOE BMC Key Capabilities & Facilities: Post-Exascale Scientific Computing

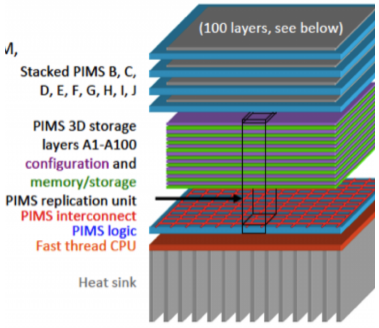
Target Application: Physics Code

A. Rahimi, et al, In *IEEE Design & Test of Computers*, 2015

- ***Novel Interconnect Networks***
- ***Disaggregated memory***

Algorithms

Architecture

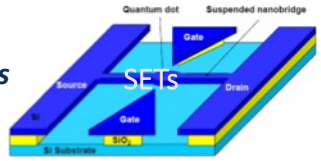


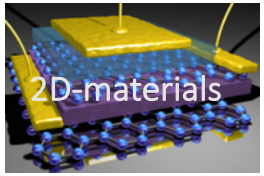
- ***Heterogeneous Integration***
- ***STT circuit Accelerators***
- ***Single-electron Charge Logic***
- ***Photonic Interconnect***

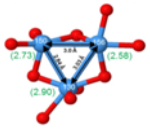
Circuits

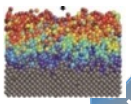
Devices

- ***Photonics & Plasmonic devices***
- ***Single Electron devices***
- ***Magnetic/spin-torque RAM***



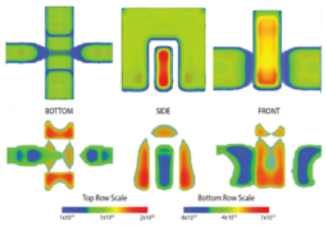






Ab Initio Models

Materials



Universal Charge Device Simulator

Gov’t & Industry BMC Activity Increasing

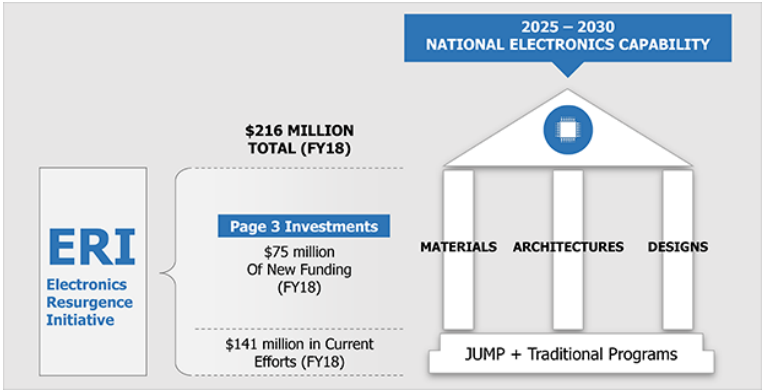
DARPA

arpa·e

SRC Semiconductor Research Corporation

NSF

IARPA
BE THE FUTURE



JUMP

C3

E2CDA

nCORE

Are Photonic Accelerators Ready for Prime Time?

Thank you!

F. B. "Rick" McCormick
Sandia National Laboratories
fbmccor@sandia.gov

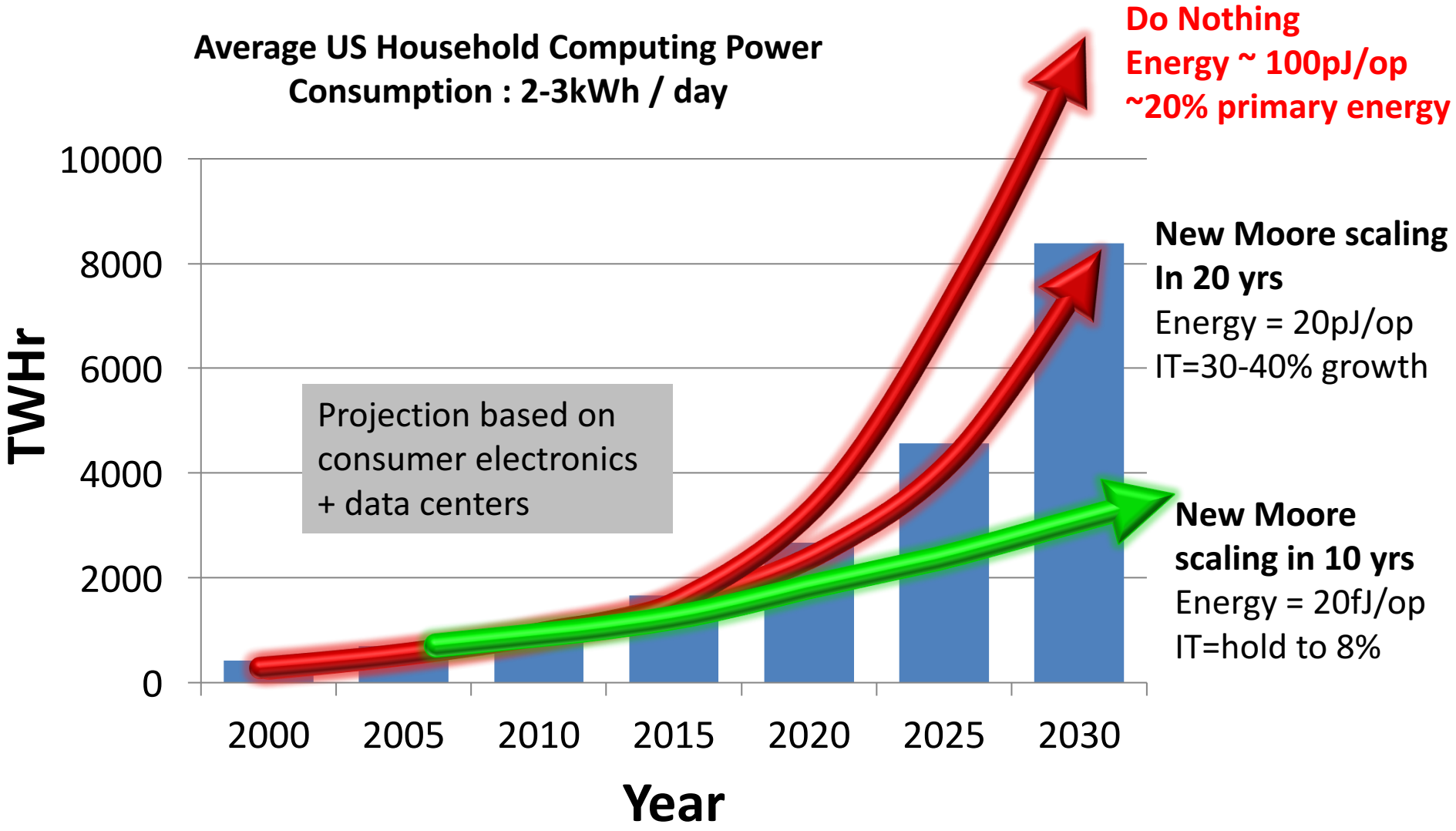
U.S. DEPARTMENT OF ENERGY
NNSA National Nuclear Security Administration

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

Backups

IT challenge for future electricity supply

Global Semiconductor market size ~ $5Trillion by 2030



www.alliancetrustinvestments.com/sri-hub/posts/Energy-efficient-data-centres

www.iea.org/publications/freepublications/publication/gigawatts2009.pdf

From W. Harrod at BIS BMC 7-Lab planning workshop

China: "National Guideline for Development of the IC Industry"

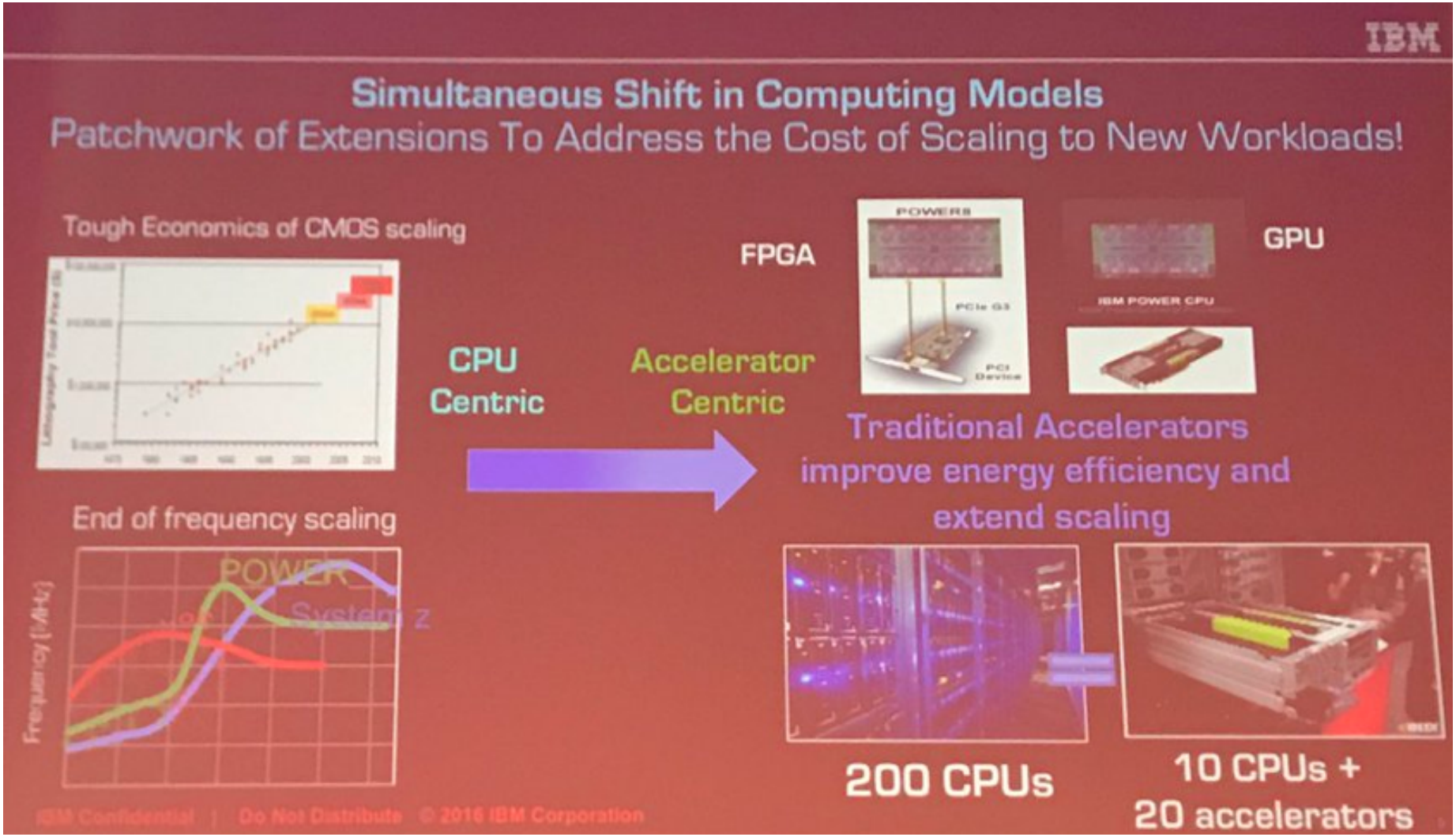
China created a $19.3 billion fund, which will be used to invest in its domestic IC firms. Over the next decade, local municipalities and private equity firms could spend $100 billion across China's IC sector.

In the 1970s, China had several state-run chipmakers, but those concerns were far behind the West in terms of IC technology. So starting in the 1980s, China began to embark on several initiatives to modernize its IC industry.
The latest effort was launched in 2014, when China unveiled what it called the "National Guideline for Development of the IC Industry." The plan is to accelerate China's efforts in a number of areas, such as **14nm finFETs, advanced packaging, MEMS, memory, and various IoT-related processors.**

SMIC, Huawei, imec, and Qualcomm in Joint Investment on SMIC's New Research and Development Company
A Joint Collaboration is Set Up to Build China's Most Advanced Integrated Circuit Research and Development Platform. The current focus will be on **developing 14nm logic technology.**

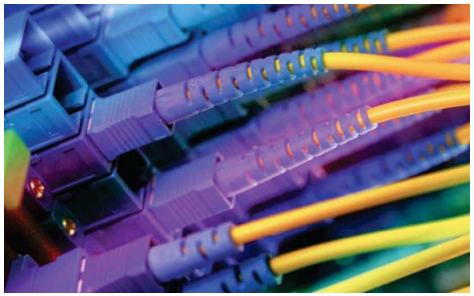
Source: Semiconductor Engineering .:. What China Is Planning

Specialized processors/accelerators



https://pbs.twimg.com/media/C8rNktkXYAAjH54.jpg

Beyond Moore's Law Photonics

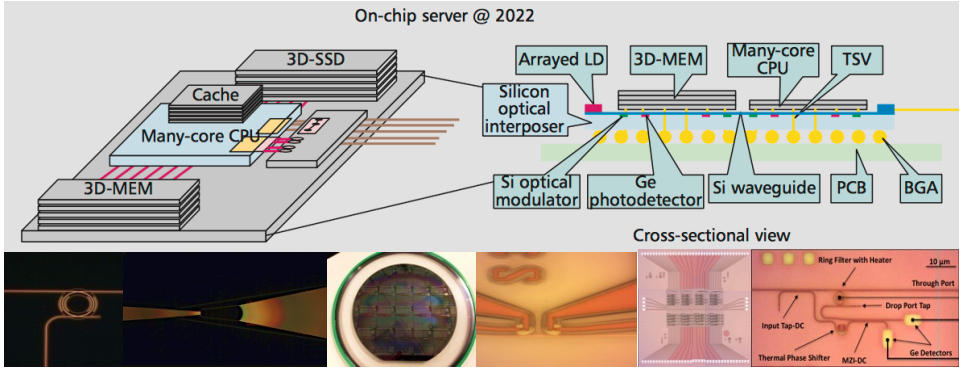


- -> $1/Gb/s
- 25 pJ/bit

Both Too High

Today's Petascale computing utilize optical transceivers for the 'network' interconnect

5 -10 years: Commodity Integrated Photonics



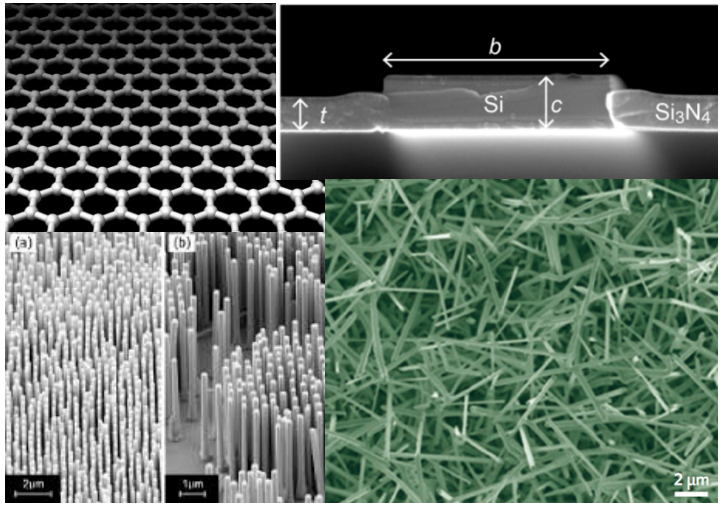
AIM photonics

Primary Institute Focus MRL 4-7

- Hetero-Integrated CMOS photonics manufacturing
- Drastically lower the cost of packaging
- Higher data rate per optical IO, further reduction
- 25 pJ/bit → < 1 pJ/bit optics total (> 25X)
- $1/Gb/s → 10¢ Gb/s ($100/Tb/s) (10X)

10-20 years: Need "Beyond Moore" Photonics

- **Scale Energy and Cost by 10^4**
  - → 1 fJ/bit optics complete (>25,000X equivalent)
  - → 0.01¢ Gb/s ($10¢/Tb/s) (>10,000X equivalent)
- **New Materials, Devices, Components**



- **Interface to LV-CMOS and beyond-Moore's transistor devices**
  - TFETs, Quantum, Neuromorphic, etc.
- **Optical switching, routing, computing**
- **Optical modeling infrastructure**
  - Optical materials, devices → sub-systems → application impact